UNIVERSITÉ DE GRENOBLE

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Virologie Microbiologie Immunologie**

Arrêté ministériel : 7 août 2006

Présentée par

## [2023 : Viafx24]

Thèse dirigée par **[2023 : identité protégée]**

préparée au sein du
**Laboratoire Adaptation et Pathogénie des Microorganismes**
dans **l'École Doctorale Chimie et Sciences du Vivant**

# Thèse constructiviste portant sur les réseaux de régulation chez *Escherichia coli*

Thèse soutenue publiquement le **29 aout 2012**,
devant le jury composé de :

[2023 : identités protégées]

# Table des matières

**Table des figures** **xi**

**I Un peu d'histoire** **1**

1 Préambule . . . 1
2 La fin du vitalisme . . . 1
3 L'organisme est indissociable de son milieu . . . 3
4 Du simple sac d'enzymes à la découverte des gènes . . . 3
5 L'influence de la thermodynamique et des physiciens . . . 6
6 L'ADN support de l'information génétique . . . 7
7 La réplication de l'information génétique . . . 8
8 Un code génétique universel . . . 11
9 La séquence semble aléatoire . . . 12
10 Un programme coordonné . . . 13
11 Analyse de texte du paradigme actuel . . . 15
12 Commentaires et description du paradigme actuel . . . 19
12.1 Le réseau, le programme et l'algorithme . . . 19
12.2 L'autoreproduction, un but en soi ? . . . 20
12.3 Quel est le coût énergétique d'un calcul/changement d'état du réseau ? . . . 20
12.4 Temps, évolution et adaptation . . . 21
13 *E. coli* au 21ème siècle . . . 22

**Références bibliographiques** **27**

**II La bactérie *Escherichia coli* 29**
1 Préambule . . . 29
2 Différence entre procaryotes et eucaryotes . . . 29
3 Réplication, transcription, traduction et repliement des protéines . . . 30
4 Les frontières de la vie : les membranes . . . 35
5 Le métabolisme . . . 35
6 La division cellulaire . . . 36
7 L'adaptation au stress . . . 38
7.1 Réponse aux chocs thermiques . . . 38
7.2 Réponse à une carence en acides aminés . . . 38
7.3 Réponse à une exposition UV . . . 39
8 Motilité, agrégation, biofilm et quorum sensing . . . 39
9 La flore intestinale . . . 40
10 Les souches d'*E. coli* pathogènes . . . 42

**Références bibliographiques 45**

**III Interaction et réseau de régulation 49**
1 Préambule . . . 49
2 Interaction et réseau de régulation . . . 49
3 Deux exemples clés . . . 53
3.1 La répression catabolique . . . 53
3.2 Le chimiotactisme . . . 55
4 La détection des interactions géniques . . . 58
4.1 Les méthodes *in vitro* . . . 58
4.1.1 Le retard sur gel . . . 58
4.1.2 L'empreinte à la Dnase . . . 60
4.2 Les méthodes à haut-debit . . . 60
4.2.1 Les puces à ADN . . . 60
4.2.2 La technique du ChIP-chip . . . 61
4.2.3 Le RNA-Seq et le ChIP-Seq . . . 63
4.3 Les techniques utilisant la biologie moléculaire . . . 63
4.3.1 Les fusions transcriptionnelles . . . 64
4.4 Les méthodes bioinformatiques . . . 70
4.4.1 Inférence de la topologie des réseaux de régulation . . . 70

4.4.2 Détection de site consensus par bioinformatique . . . . . 71

**Références bibliographiques** **75**

**IV La modélisation d'un réseau de régulation** **77**

1 Préambule . . . . . 77
2 Pourquoi modéliser ? . . . . . 77
3 Petit encouragement au biologiste réfractaire . . . . . 80
4 Quelques conjectures nécessaires . . . . . 81
4.1 Processus stochastiques basés sur des réactions chimiques . . . . . 81
4.2 Homogénéité du milieu, diffusion et encombrement . . . . . 82
4.3 Recherche du site de liaison sur l'ADN par le facteur de transcription . . 83
4.4 Transcription et traduction . . . . . 83
4.5 Approximation pour les réactions rapides . . . . . 83
4.6 Intégration des signaux au niveau des promoteurs complexes . . . . . 84
4.7 La topologie du réseau est dynamique . . . . . 84
5 Le formalisme des équations différentielles . . . . . 84
6 Petit détour par la loi d'action des masses . . . . . 86
7 Etude du code . . . . . 88
8 Cas du répresseur . . . . . 89
9 Cas de l'inducteur allostérique . . . . . 89
10 Découplage de la transcription et de la traduction . . . . . 90
11 Cas de la liaison coopérative : . . . . . 91
12 Identification de paramètres à l'aide d'un algorithme génétique . . . . . 92

**Références bibliographiques** **101**

**V Les systèmes dynamiques** **103**

1 Préambule . . . . . 103
2 Un réseau à deux gènes : l'interrupteur (toggle switch en anglais) . . . . . 103
3 Hystérésis induite par une boucle de rétroaction positive . . . . . 105
4 Un réseau à trois gènes : le repressilator . . . . . 109
5 Caractéristiques des réseaux réels . . . . . 111
6 Fonction de quelques motifs courants . . . . . 116
6.1 La boucle de rétroaction négative . . . . . 116
6.2 Le « feed forward loop » cohérent (type 1 avec un porte AND) . . . . . 116

6.3 Le « feed forward loop » incohérent (type 1 avec un porte AND) . . . . . 117
6.4 Les critiques . . . . . 117
7 Etudier la dynamique de réseaux plus gros . . . . . 117

**Références bibliographiques** **121**

**VI Le bruit** **125**
1 Les origines du bruit . . . . . 125
2 L'intérêt du bruit . . . . . 127
3 Le contrôle du bruit . . . . . 128
4 Utiliser la position relative des gènes sur l'ADN . . . . . 131
5 Les méthodes expérimentales pour étudier le bruit . . . . . 132
6 Le bruit intrinsèque et le bruit extrinsèque . . . . . 135
7 Les modèles stochastiques . . . . . 137
7.1 Les équations maîtresses . . . . . 137
7.2 Retrouver l'équation déterministe à partir de l'équation maitresse . . . . 141
7.3 La loi de Poisson . . . . . 142
7.4 L'algorithme de Gillespie . . . . . 143
8 Modèle spatial didactique . . . . . 146

**Références bibliographiques** **153**

**VII La biologie synthétique** **155**
1 Définition . . . . . 155
2 Construire un système vivant . . . . . 157
3 Le « workflow » en biologie synthétique . . . . . 158
3.1 Conception et modélisation . . . . . 158
3.2 Ecriture du programme ADN . . . . . 161
3.3 La caractérisation . . . . . 162
3.4 Les outils/appareils les plus classiques pour évaluer l'exécution du programme . . . . . 162
3.5 Optimisation de ces 3 étapes . . . . . 164
3.6 Le charbon . . . . . 166
4 Listing de quelques fonctions existantes en biologie synthétique . . . . . 168
5 Le futur en biologie synthétique . . . . . 170
5.1 Petit détour par la théorie des graphes . . . . . 170

5.2 La prochaine génération de réseaux synthétiques . . . 173
5.2.1 Filtres « tunable » . . . 173
5.2.2 L'utilisation du bruit . . . 173
5.2.3 Utilisation de la dimension spatiale . . . 175
5.2.4 Réseau de calcul basé sur les protéines . . . 176
5.2.5 Des réseaux qui apprennent . . . 176
5.2.6 Utiliser d'autres fondements pour le calcul . . . 179
5.3 Les applications à court et moyen terme . . . 181
5.3.1 L'usine microbienne . . . 183
5.3.2 La santé . . . 185

**Références bibliographiques** **187**

**VIII Physique quantique et biologie** **191**
1 Préambule . . . 191
2 Physique quantique et biologie . . . 191
3 De l'information quantique dans la cellule ? . . . 194
4 Un nouveau combat entre paradigmes ? . . . 195

**Références bibliographiques** **197**

**IX Evolution (I)** **199**
1 Préambule . . . 199
2 Quelques généralités sur l'évolution . . . 200
2.1 Les mutations . . . 200
2.2 Le transfert horizontal de gènes . . . 201
2.3 La sélection naturelle . . . 201
2.4 La dérive génétique . . . 202
2.5 Les contraintes évolutives . . . 203
3 Evolution et réseau de régulation . . . 206
3.1 Adapter le niveau d'expression d'un gène . . . 207
3.2 Réguler ou ne pas réguler . . . 207
3.3 Sélection d'un motif . . . 208
3.4 La loi de la demande . . . 210
3.5 Apprentissage associatif . . . 211
3.6 Le reflet du miroir . . . 212

4 Evolution et biologie synthétique . . . 213
4.1 Utiliser l'évolution pour « tuner » un réseau de régulation synthétique . 215
4.2 L'évolution peut-elle être un frein aux avancées en biologie synthétique ? 217
4.3 L'évolution de l'évolution . . . 219

**Références bibliographiques** **223**

**X Evolution (II)** **225**
1 Préambule . . . 225
2 Critiques et alternatives à la théorie de l'évolution . . . 225
2.1 Le métabolisme primitif . . . 227
2.2 Le réseau de régulation génique . . . 228
2.3 Le réseau épistatique . . . 234
2.3.1 La notion de paysage adaptatif . . . 235
2.3.2 Le paysage adaptatif sans corrélation . . . 236
2.3.3 Le paysage adaptatif corrélé . . . 237
3 Les unités de sélection . . . 241
3.1 La coévolution . . . 244

**XI Informatique et biologie** **251**
1 Préambule . . . 251
2 Des nouvelles disciplines : l'intelligence artificielle et la vie artificielle . . . 251
3 Les analogies et différences entre cellules et ordinateurs . . . 254
4 Les automates cellulaires . . . 256
5 Le jeu de la vie de Conway . . . 258
6 Un peu de grammaire (les systèmes de substitution séquentielle) . . . 261

**Références bibliographiques** **269**

**XII L'information** **271**
1 La notion d'information . . . 271
2 La théorie de l'information de Claude Shannon . . . 272
2.1 Pourquoi 9.2Mo ? . . . 272
2.2 Pourquoi la séquence du génome d'*E. coli* et une séquence aléatoire font-elles la même taille ? . . . 274
3 La théorie algorithmique de l'information de Kolmogorov . . . 274

4 La profondeur logique de Bennett . . . 275
5 Synthèse en image : Shannon, Kolmogorov, Bennett . . . 277
6 Spéculation personnelle sur l'information . . . 282
6.1 Quels sont les objets mathématiques qui décrivent les interactions/liens virtuels entre nucléotides ? . . . 288
6.2 La fonction qui transforme une séquence ADN en ID et sa réciproque sont-elles vraiment calculables (récursives) ? . . . 288

**XIII Qu'est-ce que la vie 293**
1 Préambule . . . 293
2 La vie perçue d'un point de vue scientifique . . . 295
3 La vie perçue d'un point de vue philosophique . . . 297
3.1 La téléologie . . . 297
3.2 La volonté de puissance . . . 299
4 La conjecture des exemplaires . . . 300
5 La mort . . . 308

**XIV Les bio-armes 317**
1 Préambule . . . 317
2 Jusqu'où l'homme est-il déjà allé . . . 317
3 Où en sommes-nous aujourd'hui ? . . . 319
4 Stratégies de lutte contre les bio-armes . . . 321
5 Petit détour par la science fiction . . . 323
6 Le danger invisible . . . 328
7 Les racines du mal . . . 331

**Références bibliographiques 335**

**XV Futur lointain 337**
1 La médecine du futur . . . 337
2 Changer le système . . . 338
3 Les alternatives futures . . . 343
3.1 Les anéantissements . . . 343
3.1.1 Anéantissement de tout homme sur terre . . . 343
3.1.2 Anéantissement de toute vie sur terre . . . 344
3.2 Arrêt de la recherche, décroissance et cycle . . . 344

3.3 Les dépérissements existentiels . . . 345
3.4 Le surhomme . . . 346
3.4.1 La mort de la mort . . . 347
3.4.2 La symbiose . . . 347
3.4.3 La boucle est bouclée . . . 348
3.4.4 Le pacte . . . 350
3.4.5 Le but . . . 352
4 Pour une futurologie . . . 354
5 La théorie des pistes . . . 355
5.1 Les mauvaises pistes . . . 356
5.1.1 La dignité humaine . . . 356
5.1.2 L'eugénisme ou le racisme . . . 357
5.2 Les bonnes pistes . . . 358
5.2.1 L'amour . . . 358
5.2.2 La musique (l'art) et le rire . . . 359
5.2.3 La morale . . . 359
5.2.4 la liberté . . . 360
5.2.5 La perfectibilité . . . 360
5.2.6 La responsabilité . . . 361
5.2.7 Le jugement esthétique . . . 362
5.2.8 La volonté, la volonté de puissance et la création . . . 363
5.2.9 Le désespoir /La souffrance/ la mort . . . 365

**XVI Un peu d'épistémologie 369**
1 Préambule . . . 369
2 Le socle de base n'est pas sûr . . . 371
2.1 Les épistémologies positivistes . . . 371
2.2 Les épistémologies constructivistes . . . 373
3 La structure des avancées scientifiques . . . 376
3.1 La pyramide . . . 376
3.2 Le cercle . . . 377
3.3 Les poupées russes . . . 378
3.4 La montagne . . . 379
4 La déconstruction du critère de falsifiabilité . . . 381
4.0.1 L'évidence est viciée . . . 381

4.0.2 Les faits sont impurs . . . 382
4.0.3 Aucune théorie ne devrait être rejetée . . . 382
4.0.4 L'attaque ironique contre les positivistes/rationalistes . . . 382
5 Apologie de l'irrationalité . . . 383
6 Reconstruire la place de l'observateur . . . 387
6.0.5 Apprendre à s'auto-interroger . . . 387
6.0.6 L'importance cruciale de la partie immergée de l'iceberg scientifique . . . 388
7 Critique, consensus et intersubjectivité . . . 390
8 La déconstruction du dogme hypothèse– expérience– résultat . . . 390

**XVII Analyse du système de publication 393**
1 Limiter les informations demandées aux rewievers . . . 393
2 Contre l'anonymat . . . 397
3 Faire tomber le mur aristocratique . . . 399

**XVIII Falsification, discipline et critique 409**
1 Le casse tête du chercheur . . . 409
2 La déconstruction du dogme des données inaltérables . . . 410
2.1 Le système positiviste est un système oppressant . . . 417
3 La question disciplinaire . . . 421
3.1 Un système qui nous imprègne . . . 422
3.2 Un système qui nous soumet . . . 422
3.3 Un système qui normalise . . . 425
3.4 Un système qui récompense . . . 425
3.5 Un système qui surveille . . . 426
4 La déconstruction de la critique . . . 427
4.1 Préambule . . . 427
4.2 La critique mène au principe d'autonomie . . . 429
4.3 La critique mène à l'exhaustivité . . . 429
4.4 La critique mène à la stabilité des opinions . . . 430
4.5 La critique mène à l'excellence . . . 430
4.6 Le scepticisme ne vaut que pour soi . . . 434
4.7 L'alternative à la critique : l'interrogation . . . 435

**XIX Qu'est-ce que la science ? 439**

1 La récursivité . . . . . 439

2 L'interrogation . . . . . 442

3 La science est une guerre . . . . . 445

4 La recherche de la grande santé . . . . . 450

**Bibliographie globale des livres 455**

# Table des figures


I.1 Le canard automate de Vaucanson. . . . . 2
I.2 Différents types de milieu . . . . . 4
I.3 *Escherichia coli* . . . . . 5
I.4 Quelques scientifiques célèbres en Biologie. . . . . 9
I.5 Confusions entre les métaphores de l'informatique . . . . . 14

II.1 Quelques chiffres . . . . . 31
II.2 Deux vidéos de microscopie . . . . . 33
II.3 Le génome . . . . . 34
II.4 Le métabolisme . . . . . 37
II.5 Nuage de tag de quelques grands domaines d'étude chez *E. coli* . . . . . 41
II.6 Les souches pathogènes d'*E. coli* . . . . . 43

III.1 Réseau, interactions et adaptation. . . . . 50
III.2 L'interaction ADN–protéine . . . . . 52
III.3 La répression catabolique . . . . . 56
III.4 Le chimiotactisme . . . . . 59
III.5 La technique de ChIP on chip . . . . . 62
III.6 Démonstration vidéo du logiciel Gentle . . . . . 65
III.7 Contrôle d'une interaction . . . . . 66
III.8 Mesure de l'expression du gène *acs* (rapportée par la luciferase) . . . . . 68
III.9 Mesure de l'expression du gène acs (rapportée par la gfp) . . . . . 69
III.10 Démonstration vidéo de l'utilisation du logiciel Wellreader . . . . . 70

III.11 Un petit logo « basic » . . . . . 73

IV.1 Les limites de l'intuition . . . . . 78
IV.2 Promoteur de l'operon aceBAK . . . . . 84
IV.3 Schéma d'une activation transcriptionnelle simple. . . . . 86
IV.4 Dynamique d'une activation simple . . . . . 88
IV.5 Dynamique d'une régulation allostérique . . . . . 90
IV.6 Liaison coopérative et fonction de Hill . . . . . 91
IV.7 Interface graphique (GUI) du logiciel Genetic Network Analyzer . . . . . 93
IV.8 Identification de paramètres . . . . . 94
IV.9 Schéma de la recherche d'un minimum global . . . . . 96
IV.10 Résultats de l'identification des paramètres . . . . . 97
IV.11 Recherche de paramètres avec un algorithme génétique . . . . . 98
IV.12 Le fit de l'éléphant . . . . . 98

V.1 Schéma du toggle switch. . . . . 103
V.2 Schéma d'un vecteur de champs . . . . . 105
V.3 Etude en dynamique de l'interrupteur . . . . . 106
V.4 Etude en dynamique de la boucle de retro-action positive . . . . . 108
V.5 Schéma du repressilator. . . . . 110
V.6 Oscillations de la GFP dans des bactéries *E. coli* . . . . . 111
V.7 Etude du repressilator . . . . . 112
V.8 Réseaux invariants d'échelle . . . . . 114
V.9 Les réseaux biologiques réels. . . . . 115
V.10 Fonctions de quelques motifs . . . . . 118

VI.1 1 génotype, 1 milieu, plusieurs phénotypes . . . . . 126
VI.2 Exemple de bistabilité obtenue avec l'operon *araBAD* . . . . . 130
VI.3 Expérience de pensée . . . . . 133
VI.4 Interface graphique du logiciel ScanR Analysis . . . . . 134
VI.5 Vidéo : démonstration de la facilité d'utilisation du logiciel ScanR analysis. . 135
VI.6 Bruit intrinsèque et bruit extrinsèque . . . . . 136
VI.7 Dissocier bruit intrinsèque et bruit extrinsèque . . . . . 138
VI.8 Schéma expliquant l'obtention de l'équation maîtresse . . . . . 140
VI.9 Schéma historique de l'algorithme de Gillespie publiée en 1977. . . . . 145
VI.10 Simulation stochastique . . . . . 147

VI.11 Cellule virtuelle tridimensionnelle . . . 148
VI.12 Différents paradigmes de programmation . . . 149
VI.13 Communication virtuelle intercellulaire . . . 150
VI.14 Peut-on recréer une cellule « parfaite » *in silico* ? . . . 151

VII.1 Des stratégies différentes et complémentaires pour construire un organisme . 159
VII.2 Environnement de développement intégré futuriste . . . 161
VII.3 Techniques pour « écrire » le programme . . . 163
VII.4 Photo de quelques outils permettant la caractérisation d'un programme ADN 165
VII.5 Présentation de la compétition IGEM. . . . 167
VII.6 Oscillations de la GFP . . . 169
VII.7 Quelques schémas trouvés en biologie synthétique . . . 171
VII.8 La connectivité d'un réseau. . . . 172
VII.9 Les filtres band-pass . . . 174
VII.10 Calcul Parallele . . . 175
VII.11 La dimension spatiale . . . 177
VII.12 Construire des réseaux basés sur la transduction du signal . . . 178
VII.13 Schéma d'un réseau capable d'apprentissage associatif . . . 180
VII.14 Des pelottes de bactéries *E. coli* . . . 181
VII.15 Illustration de l'utilisation de nouveaux modes de calcul. . . . 182
VII.16 Un exemple dressant un pont entre biologie synthétique et médecine . . . 184

VIII.1 La potentielle dualité onde–corpuscule d'un virus . . . 193

IX.1 Les spandrels de la cathédrale San Marco à Venise . . . 205
IX.2 Sélection d'un système de régulation . . . 209
IX.3 L'instabilité génétique du chien . . . 213
IX.4 Rendre un réseau fonctionnel en utilisant l'évolution . . . 218

X.1 Une transition de phase . . . 229
X.2 Fonctionnement d'un réseau booléen . . . 230
X.3 Le modèle NK . . . 232
X.4 Métaphore pour illustrer la frontière entre ordre et chaos. . . . 233
X.5 La notion de paysage adaptatif . . . 236
X.6 Image du mont Fuji au Japon . . . 237
X.7 L'allure du paysage adaptatif d'un organisme . . . 239

X.8 Contrôle de la rugosité d'un paysage adaptatif . . . 240
X.9 Coévolution et paysage adaptatif dynamique. . . . 243
X.10 Réplicateur et interacteur . . . 246
X.11 Entité centrée vs information centrée . . . 248
X.12 La volonté de puissance de la biosphère . . . 249

XI.1 Quelques exemples de vie artificielle. . . . 253
XI.2 Analogies et différences entre une cellule et un ordinateur . . . 256
XI.3 Illustration de quelques automates cellulaires unidimensionnelles . . . 259
XI.4 Le jeu de la vie de Conway . . . 260
XI.5 Les règles de substitution séquentielle . . . 262
XI.6 La soupe de « séquences »(string) . . . 263
XI.7 Résultats de 6 simulations (6 carrés) de la soupe de « séquence » . . . 265

XII.1 La théorie de l'information de Shannon . . . 273
XII.2 La théorie algorithmique de l'information de Kolmogorov . . . 276
XII.3 La profondeur logique de Bennett . . . 278
XII.4 Comparaison des 3 théories de l'information . . . 280
XII.5 Profondeur logique d'un nucléotide . . . 281
XII.6 Découpage de chromosome et information . . . 283
XII.7 Interactions inter-nucléotidiques . . . 284
XII.8 Décompression de la séquence ADN . . . 286
XII.9 Le malaise. . . . 289
XII.10 Les rapports étranges qu'entretiennent les trois théories de l'information. . . . 291

XIII.1 Sol jonché de rochers volcaniques vu par Mars Pathfinder . . . 301
XIII.2 La conjecture des exemplaires. . . . 303
XIII.3 Le logarithme semble intervenir dans la conjecture des exemplaires . . . 305
XIII.4 Pure spéculation sur la question « qu'est ce que la vie ? » . . . 307
XIII.5 Quelques conséquences étranges induites par la conjecture des exemplaires. . . 309
XIII.6 La mort selon Google image . . . 312

XIV.1 Une bio-arme de seconde génération. . . . 325
XIV.2 Peintures de Otto Dix illustrant l'horreur de la première guerre mondiale . . . 330

XV.1 L'endosymbiose . . . 349
XV.2 Le surhomme . . . 351

XV.3 Le rouge-gorge est une des merveilles de la nature . . . . . 362
XV.4 Le vert-gorge . . . . . 364

XVI.1 Renoncer à la recherche d'une vérité objective. . . . . 374
XVI.2 La structure des connaissances. . . . . 380
XVI.3 La démarcation entre science et non-science . . . . . 389

XVII.1 L'encyclopédie Wikipedia est libre et ouverte au monde . . . . . 403
XVII.2 Danser sur le réseau de la connaissance . . . . . 405
XVII.3 Détruire le mur d'enceinte aristocratique . . . . . 407

XVIII.1 Le casse-tête du chercheur . . . . . 411
XVIII.2 Le système positiviste oppressant. . . . . 418
XVIII.3 L'acte de publier revêt l'apparence d'un examen . . . . . 424
XVIII.4 Le rêve de tout scientifique devrait être de posséder un stable instable. . . . 431
XVIII.5 La déconstruction de la critique. . . . . 433
XVIII.6 Le scepticisme ne vaut que pour soi. . . . . 436

XIX.1 Science de bas niveau et science de haut niveau . . . . . 442
XIX.2 Science et Interrogation . . . . . 445
XIX.3 La volonté de puissance . . . . . 447
XIX.4 La science est une guerre. . . . . 451
XIX.5 Interrogation et grande santé. L'interrogation est la seule méthode scientifique incontestable. La recherche de la grande santé est au fondement de l'agir du scientifique. . . . . 453

# Chapitre I

# Un peu d'histoire

## 1 Préambule

Paradoxalement, ce premier chapitre n'est pas le plus facile. Il est relativement abstrait et théorique. Si vous êtes novice en biologie, je vous conseille plutôt de commencer avec le deuxième chapitre sur *E. coli*. Dans la première partie de ce chapitre d'histoire, je décris les principales découvertes du 20$^{\text{ème}}$ siècle (de manière non exhaustive) qui ont abouti à la création du paradigme (manière de voir le monde) actuel en biologie. La deuxième partie se focalise sur les ouvrages de François Jacob et Jacques Monod pour essayer d'extraire les caractéristiques essentielles de ce paradigme. Il me semble que le lecteur qui s'engage dans ce chapitre devrait, idéalement, déjà connaître au moins 50% de ce qui y est expliqué pour que la lecture ne soit pas trop difficile. Si l'histoire des sciences ou la biologie théorique n'est pas votre tasse de thé, passez aux chapitres suivants sans crainte : ils sont plus concrets.

## 2 La fin du vitalisme

Le canard automate du grenoblois Jacques de Vaucanson (1709–1782) accomplit une fonction : la digestion. Les engrenages qui le composent nous indiquent que pour comprendre un système, il faut le contrôler. Si on contrôle la fonction (la digestion) en faisant tourner l'engrenage, alors on comprend comment le mécanisme fonctionne. Dans mon exposé, la bactérie Escherichia coli ne représente qu'un *moyen* pour *comprendre* et pour *contrôler* la vie. Ces deux verbes étant, selon moi, difficilement dissociables. L'introduction historique que je me propose de faire n'est donc pas celle des découvertes relatives au fonctionnement de la bactérie *Escherichia coli*. Je souhaite plutôt retracer brièvement l'évolution, au cours des deux derniers siècles, des idées et concepts associés, de prés ou de loin, à la notion de vie —concepts qui ont émergé en grande partie grâce à l'étude de cette bactérie—. Nous allons voir comment les caractéris-

tiques du vivant se sont, au fur et à mesure, précisées au cours des deux derniers siècles. Mais ne nous leurrons pas, le consensus actuel dans la communauté scientifique est qu'il n'existe pas justement de consensus sur la définition de la vie. Et comme le fait remarquer le fondateur de l'exobiologie Carl Sagan (1934–1996), cela est assez remarquable.

> *« Yet despite the enormous fund of information that each of these biological specialties has provided, it is a remarkable fact that no general agreement exists on what it is that is being studied. There is no generally accepted definition of life »*[1]

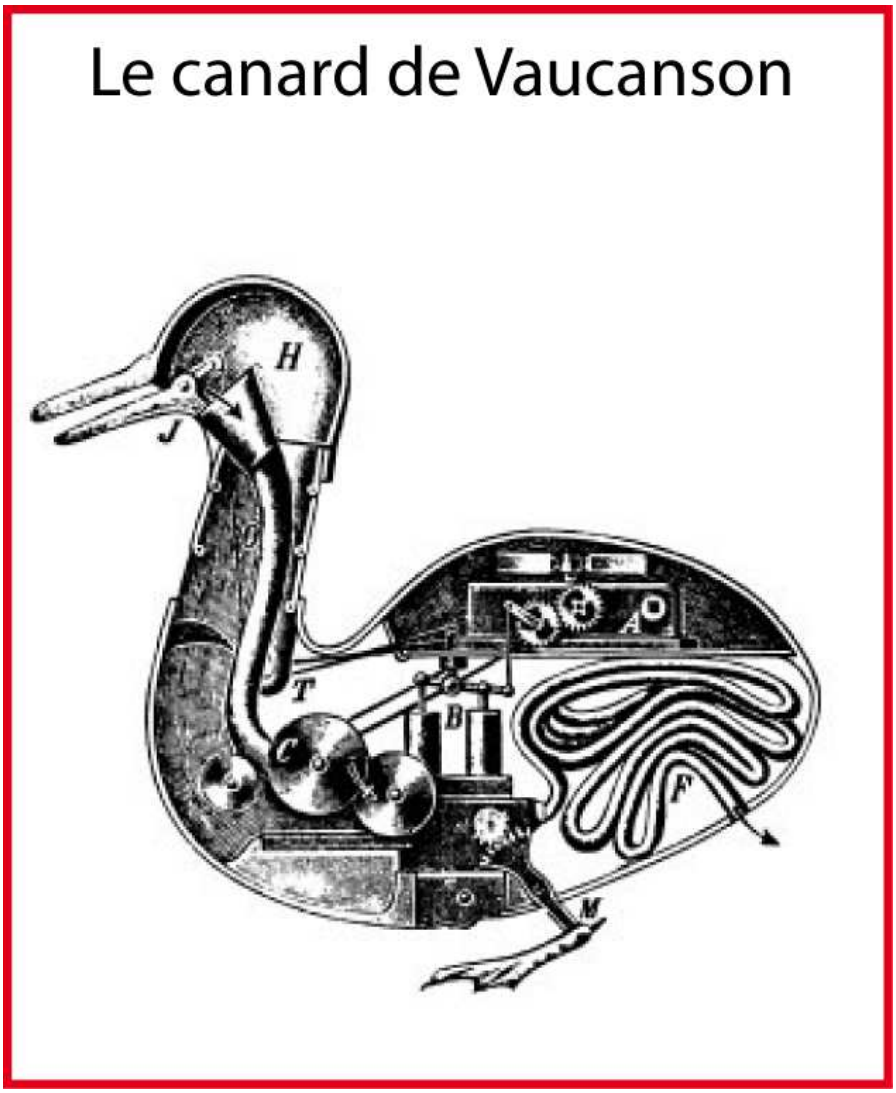



**Figure I.1** – Le canard automate de Vaucanson. La règle des 3 C : On ne Comprend que ce que l'on peut Construire et Contrôler.

[1] Carl Sagan, *Definitions of life.*

Le vitalisme est une doctrine selon laquelle il existe en chaque être vivant un *principe vital* distinct des propriétés physico-chimiques de celui-ci. Le déclin de cette doctrine commence en 1828 lorsque Friedrich Wöhler (1800–1882) effectue accidentellement la synthèse de l'urée, un composé spécifiquement organique. La synthèse des molécules organiques en chimie montre qu'entre la nature et le vivant, il n'y a pas de différence de nature mais de complexité. Au début du 19$^{\text{ème}}$, l'organisme dépensait de la force vitale pour effectuer son travail de synthèse et de morphogenèse, à la fin du 19$^{\text{ème}}$ siècle, il consomme de l'énergie. La vie semble alors caractérisée par l'existence de catalyseurs spécifiques (les enzymes), capables de réaliser des réactions que le chimiste ne sait faire que dans des conditions extrêmes.

# 3 L'organisme est indissociable de son milieu

Une autre notion importante apparait à cette période : *le milieu*. Nous devons à Auguste Comte (1798–1857) une définition précise de cette notion de milieu : *« non plus seulement le fluide dans lequel un corps est plongé mais l'ensemble total des circonstances extérieures nécessaire à l'existence de chaque organisme »*[2]. Mais c'est Claude Bernard (1813–1878) qui a l'idée que *« le phénomène vital n'est tout entier ni dans l'organisme ni dans le milieu : c'est en quelque sorte un effet produit par le contact entre l'organisme vivant et le milieu qui l'entoure »*. Comme le souligne François Jacob, c'est aussi au cours de cette période que la biologie change de lieu de travail : auparavant dans la nature, dans les musées ou dans les jardins botaniques, désormais la biologie se fait en laboratoire. Les scientifiques comprennent que, pour étudier un être vivant, il faut le placer dans des conditions de milieu aussi définies que possibles puis faire varier systématiquement chacun des paramètres du milieu.

> *« il n'est pas exagéré de dire que depuis lors, cette manière de procéder constitue la principale activité des laboratoires de biologie »* [3].

# 4 Du simple sac d'enzymes à la découverte des gènes

En 1885, un pédiatre allemand, Théodore Escherich, isole des bactéries provenant de couches de bébé. Il remarque une bactérie capable de croissance massive et rapide sur de très nombreuses sources de nourriture : lait, pomme de terre, sang. A cette époque, les micro-organismes sont

[2] Georges Canguilhem, *La connaissance de la vie*, bibliothèque des textes philosophiques, p. 170.
[3] François Jacob, *La logique du vivant*, Gallimard, p. 205.
[4] Les deux images à droite sont des images issues de Wikipedia.

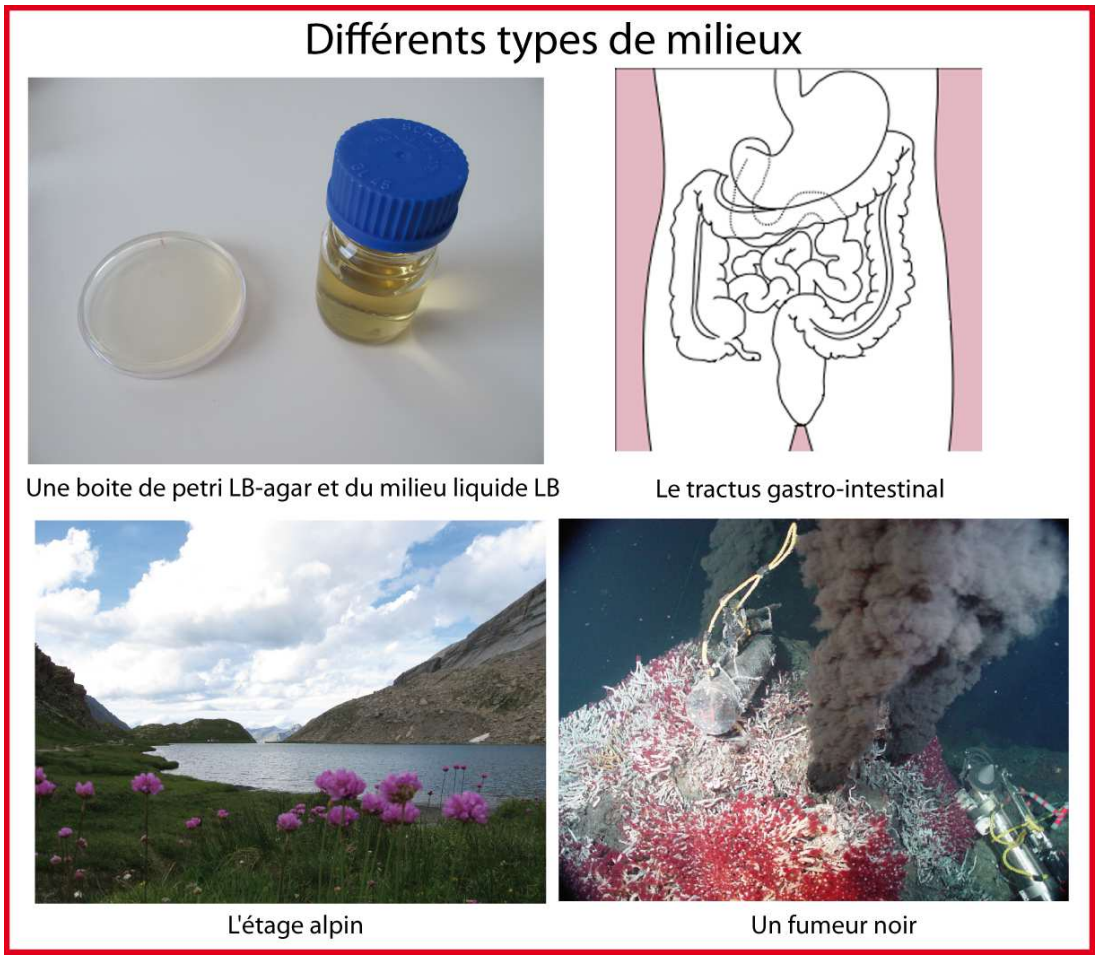



**Figure I.2** – Différents types de milieu. Le milieu liquide LB, l'intestin, la montagne ou les fonds marins sont tous des milieux où la vie peut se développer [4]. Le milieu façonne autant les organismes que les organismes façonnent le milieu.

vus comme de simples *fours*, à la frontière entre vie et non-vie, capables de brûler la nourriture pour créer de la chaleur, des déchets et des molécules organiques. Les bactéries restent donc un mystère car elles sont beaucoup plus petites que les autres organismes connus à cette époque et elles ne possèdent pas de noyau. Théodore Escherich ne se doutait probablement pas que quelques dizaines d'années plus tard, les scientifiques du monde entier commenceraient à cultiver « sa bactérie » en faisant des centaines de milliers d'expériences pour comprendre comment elle survit, se reproduit ou se nourrit. En 1919, en hommage à son découvreur, on nomme cette bactérie *Escherichia coli*[5].

Pendant encore longtemps, de nombreux chercheurs considèrent les bactéries comme de

[5]Plusieurs paragraphes dans ce chapitre dont celui-ci sont inspirés du très pédagogique livre de Carl Zimmer, *microcosm*. Les qualités littéraires de l'auteur rendent la lecture de cet ouvrage (sur la vie d'*E. coli*) très accessible à un public non expert.

simples sacs à enzymes. Jusqu'à ce que, dans les années 1940, Edward Tatum et George Beadle mènent leur célèbre expérience d'irradiation, d'abord chez *Neurospora crassa* (Beadle & Tatum, 1941) puis chez la bactérie *E. coli*. Pour sa croissance, *E. coli* a seulement besoin d'un milieu contenant une source de carbone, d'azote, de phosphate ainsi que de l'eau et des éléments traces. A partir de ce milieu, elle est capable, grâce à ses enzymes, de bio-synthétiser chacun des 22 acides aminés. En irradiant des millions de bactéries, Edward Tatum découvrit, parmi les bactéries survivantes, certains mutants capables de croitre seulement lorsqu'il ajoutait un acide aminé particulier dans le milieu : Cela signifiait qu'une des enzymes nécessaires à la biosynthèse de l'acide aminé ajouté avait été inactivée, d'une manière ou d'une autre, par l'irradiation. Plus surprenant encore, les bactéries mutantes pouvaient se reproduire, mais les générations suivantes conservaient le défaut. Celui-ci était donc transmis. Le chercheur en conclut que derrière chaque enzyme se cachait un gène.

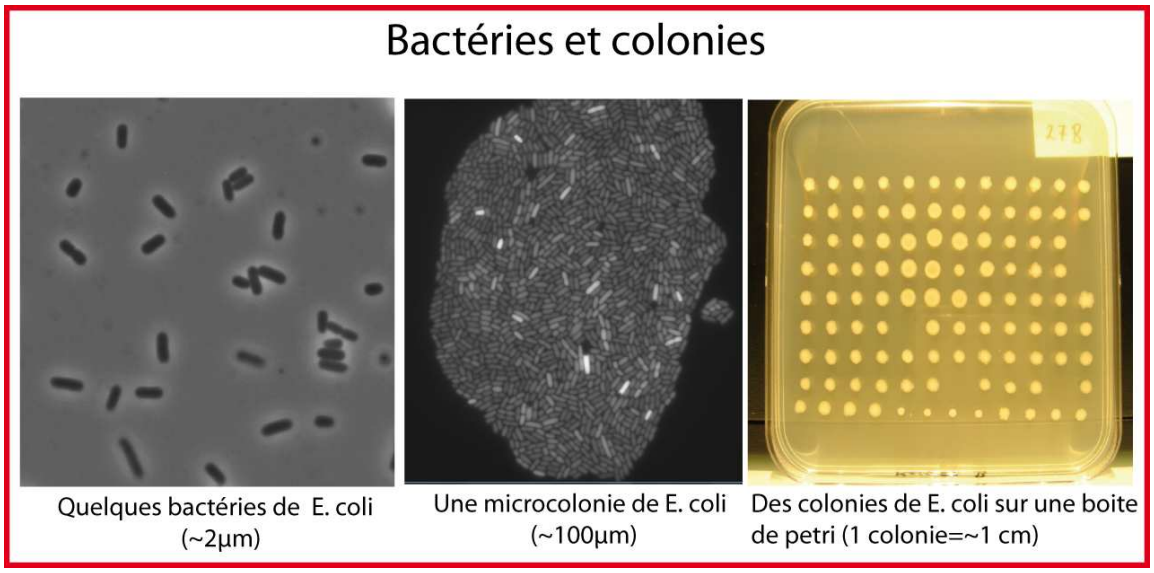



**Figure I.3** – Quelques photos montrant la différence entre une bactérie et une colonie. Une colonie est une population clonale de bactéries. Une colonie peut être observée à l'œil nu alors qu'une bactérie est observée à l'aide d'un microscope.

Joshua Lederberg, jeune étudiant d'à peine 21 ans, commença à travailler dans le laboratoire d'Edward Tatum avec une grande ambition : mettre en évidence un échange de matériel génétique entre bactéries, autrement dit du *sexe bactérien*. Pour cela il utilisa deux doubles mutants. L'un incapable de synthétiser ni la méthionine ni la biotine et l'autre incapable de synthétiser ni la thréonine ni la proline. Le jeune chercheur mélangeât ensuite ces deux souches en culture en ajoutant les 4 nutriments dans le milieu pour leur permettre de croître. Il les laissa ensemble plusieurs semaines pour leur laisser de nombreuses opportunités d'avoir de

l'hypothétique « sexe » . Puis il retira les 4 nutriments et regarda si parmi les descendants, certains avaient acquis la possibilité de pousser. Après des mois de frustration, de boites de pétri gâchées et de colonies mortes, il repéra finalement une colonie capable de pousser sur les 4 nutriments. Rappelons ici qu'une colonie est une population visible qui apparait, sur un milieu solide, après croissance exponentielle d'une bactérie unique. Il en conclut que les ancêtres (les deux doubles mutants) avaient du combiner leurs gènes ensemble (Tatum & Lederberg, 1947). En permettant de croiser des souches d'*E. coli* comme on croise des drosophiles, cette découverte permit d'avoir une compréhension beaucoup plus précise de ce qu'est un gène et offrit un outil pour disséquer le chromosome. 12 ans plus tard, Joshua Lederberg partagea le prix Nobel de médecine avec Edward Tatum et George Beadle.

## 5 L'influence de la thermodynamique et des physiciens

Au cours de mes lectures, j'ai pu constater qu'il y a un principe qui est cité plus que tout autre par les biologistes théoriciens. Il s'agit du deuxième principe de la thermodynamique. Ce principe définit la tendance de tout système physique à évoluer naturellement vers un état d'équilibre caractérisé par une certaine uniformité. L'augmentation de l'entropie signifie simplement que des objets vont explorer spontanément tout l'environnement dont ils disposent. Une métaphore simple pour le comprendre est celle de la goutte d'encre bleue ajoutée à un verre rempli d'eau. Très concentré au début, les molécules d'encre vont explorer tout le verre jusqu'à ce que la concentration (la couleur bleue) devienne homogène. Ce processus est irréversible. Erwin Schrödinger (1887–1961) est un physicien autrichien qui a eu une influence considérable sur le développement de la biologie moléculaire. Dans son petit ouvrage *Qu'est ce que la vie* écrit en 1944, il décrit sa conception de la matière vivante qui « *évite la décomposition vers l'équilibre* » et se « *nourrit d'entropie négative* » . Ainsi, selon cette vision, les être vivants sont « *des systèmes ouverts en déséquilibre thermodynamique* » qui échangent en permanence matière et énergie avec le milieu qui les entoure. Cela n'est pas en contradiction avec le second principe car l'ensemble de l'univers (les êtres vivants et leur milieu) tend bien, lui, vers une augmentation d'entropie[6]. Ainsi, à travers le prisme de la physique statistique, la cellule est un sac où des quantités énormes d'atomes et de molécules diffusent dans le liquide cellulaire. Comment la physique statistique peut elle expliquer les mécanismes de l'hérédité ? Elle ne le peut pas et Erwin Schrödinger comprend que c'est une macromolécule unique, *un cristal apériodique* qui

[6]Cependant, Antoine Danchin met en évidence que Erwin Schrödinger fait une confusion entre l'ordre créé dans la cellule et l'augmentation de l'entropie qui ne correspond pas, comme on le lit souvent, à une augmentation de désordre. (Antoine Danchin, *La barque de Delphes*, Odile Jacob, p. 119).

est responsable de la transmission à la descendance et non une goutte homogène de liquide. Il va même plus loin en proposant la première métaphore directement liée à l'informatique : Erwin Schrödinger propose que le gène est un message écrit en code.

> *« En biologie, nous nous trouvons en face d'une situation entièrement différente. Un seul groupe d'atomes n'existant qu'en un seul exemplaire produit des événements ordonnés et merveilleusement harmonisés l'un avec l'autre ainsi qu'avec le milieu environnant, et cela grâce à des lois extraordinairement subtiles . [...]Nous croyons qu'un gène —ou peut être l'ensemble de la fibre chromosomique— est un solide apériodique.* [Puis dans les notes de bas de page] *Le fait d'être éminemment flexible n'est pas une objection car un mince fil de cuivre l'est également. [...] Une molécule exceptionnellement grosse, qui doit être nécessairement un chef-d'œuvre d'un ordre très complexe, placé sous la baguette magique de la théorie des quanta. [...] Ce que nous voulons faire ressortir est simplement ceci : avec l'image moléculaire du gène il n'est plus inconcevable que le code en miniature puisse se trouver en correspondance exacte avec un plan de développement très complexe et très élaboré et contenir en même temps les moyens de le mettre à exécution »*[7].

# 6 L'ADN support de l'information génétique

Jusqu'aux années 1950, la plupart des scientifiques pensaient que les protéines étaient le support des gènes et donc de l'information génétique. En 1928, Frederick Griffith travaillait sur la bactérie *Streptococcus pneumoniae*. Il disposait d'une souche virulente et d'une souche non virulente. L'une capable de tuer une souris l'autre pas. Lorsqu'il tuait la souche virulente en la chauffant avant de l'injecter à une souris, cette dernière survivait. Cependant, si la souche virulente morte était mélangée à la souche non virulente vivante, une injection tuait la souris. La souche non virulente s'était transformée en souche pathogène et leurs descendants restaient pathogènes et ce grâce à un composé provenant de la souche morte. Il fallut attendre 1944 pour que Oswald Avery purifie la substance transformante et découvre que la seule molécule capable de « transformer » la souche non virulente en souche pathogène n'était pas une protéine mais une chose appelée acide désoxyribonucléique : l'ADN (Avery *et al.*, 1944).

> *« If, however the biologically active substance isolated in highly purified form as the sodium salt of desoxyribonucleic acid actually proves to be the transforming principle, as the available evidence strongly suggest, then nucleic acids of this type*

[7] Erwin Schrödinger, *Qu'est ce que la vie ?*, Points, P.114, 116, 139.

*must be regarded not merely as structurally important but as functionally active in determining the biochemical activities and specific characteristics of pneumococcal cells »*

Mais il fallut encore 10 ans de plus et l'expérience de Alfred Hershey et Martha Chase pour que ce résultat soit pleinement accepté par la communauté scientifique. Ces deux chercheurs mirent au point une expérience en utilisant des bactériophages. C'est-à-dire un virus de bactérie composé uniquement d'une coque protéique et de matériel génétique. Un phage infecte une bactérie en s'attachant à sa membrane externe et en y injectant son matériel génétique qui détournera la machinerie cellulaire bactérienne pour créer de nouveaux virus. Dans leur première expérience, les deux chercheurs marquèrent l'ADN des phages avec du $^{32}$P radioactif (le phosphate est présent dans l'ADN mais absent des 22 acides aminés). Puis ils permirent aux phages d'infecter *E. coli*. Après avoir séparé les phages des bactéries par centrifugation, ils montrèrent que le marqueur radioactif n'était visible que dans le culot bactérien, pas dans le surnageant qui contenait les coques protéiques. Dans leur seconde expérience, les deux chercheurs marquèrent les protéines des phages avec du $^{35}$S radioactif (le souffre est présent dans 2 acides aminés mais est absent de l'ADN). Après séparation, le marqueur radioactif fût trouvé dans les coques protéiques mais pas dans les bactéries infectées. Ce résultat prouva que c'est bien l'ADN et non les protéines qui passe du phage à la bactérie. Ainsi 10 ans après Oswald Avery, l'expérience de Hershey et Chase prouva définitivement que l'ADN était bien le support de l'information génétique (Hershey & Chase, 1952).

*« We have shown that when a particle of bacteriophage T2 attaches to a bacterial cell, most of the phage DNA enters the cell, and a residue containing at least 80 per cent of the sulfur-containing protein of the phage remains at the cell surface [...] this residue plays no further role in infection after the attachment of phage to bacterium »*

## 7 La réplication de l'information génétique

Personne n'était plus excité par ces résultats qu'un jeune biologiste américain nommé James Watson. Mais pour comprendre comme l'ADN agissait comme matériel génétique, il était nécessaire de comprendre sa structure. A cette époque, Watson travaillait à l'université de Cambridge où il commença à réfléchir avec Francis Crick un physicien anglais. En se basant sur les analyses cristallographiques aux rayons X de Rosalind Franklin, ils proposèrent la structure en double hélice de la molécule d'ADN. Les deux clés de ce problème étaient : d'une part, la structure

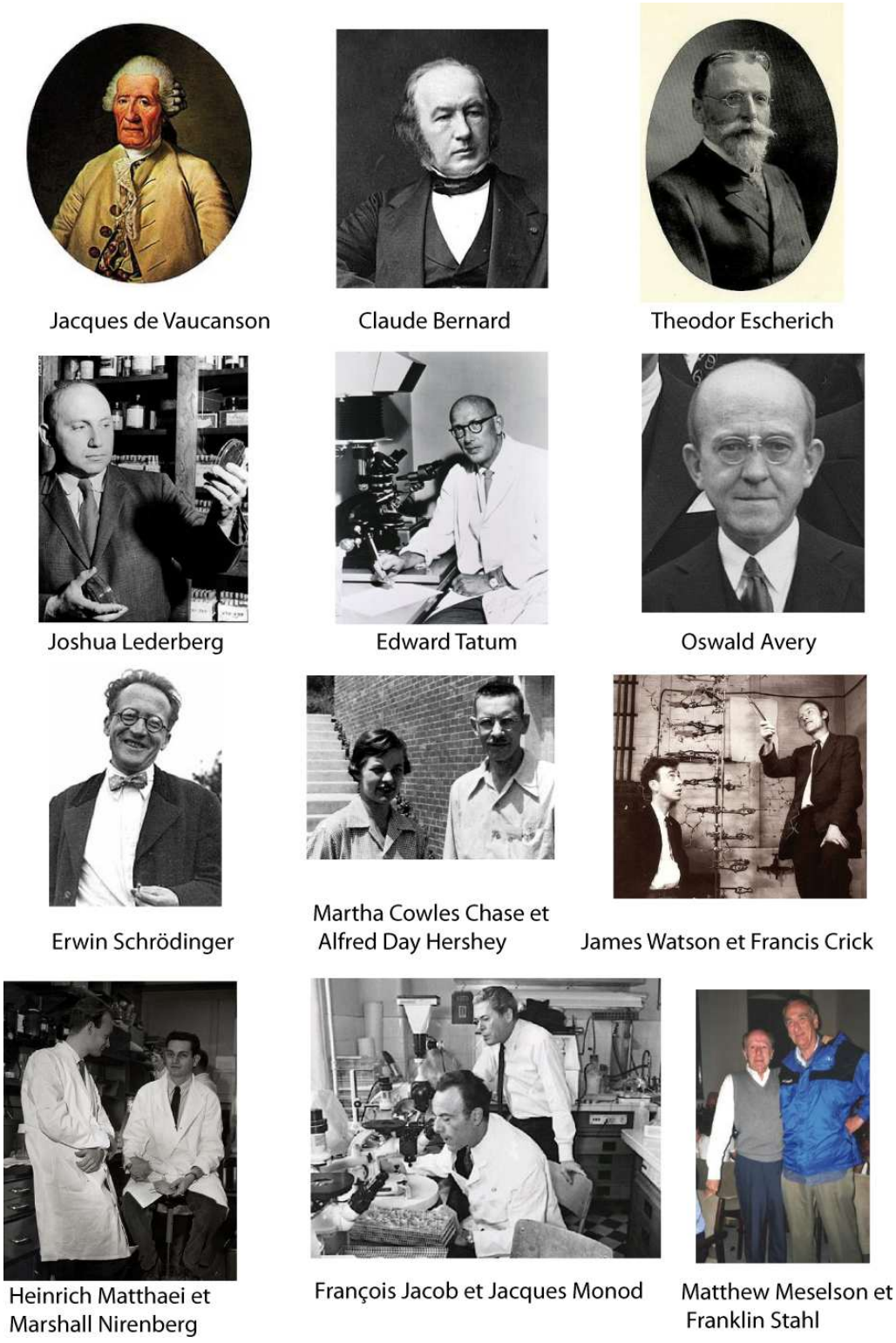



**Figure I.4** – Quelques scientifiques célèbres en Biologie.

hélicoïdale ; d'autre part, l'observation que la structure chimique de l'ADN est composée de quatre bases (A, T, G, et C), et que les deux paires de bases A-T et G-C ont des structures complémentaires sur le plan stérique. La structure qu'ils proposaient semblait expliquer d'elle-même comment fonctionnent les gènes. Chaque brin d'ADN est un empilement de millions de bases arrangées en ligne comme une chaine de caractère. Le texte peut avoir un nombre infini de signification, dépendant de l'empilement des bases. De cette manière, l'ADN stocke l'information nécessaire pour construire n'importe quelle protéine de n'importe quelle espèce (Watson & Crick, 1953). Selon la légende, alors que les deux chercheurs marchaient en direction du pub « Eagle » pour déguster une petite bière bien fraiche, Francis Crick se serait exclamé « nous avons trouvé le secret de la vie » . Cela révèle le fort désir, la forte pression qui régnait à l'époque pour percer enfin les mécanismes de transmission du patrimoine génétique. Mais notez bien que le secret de la vie et sa définition sont ailleurs. Et cet ailleurs pourrait être encore très loin.

> *« We wish to suggest a structure for the salt of desoxyribose nucleic acid (DNA). This structure has novel features which are of considerable biological interest [...] It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material »*

Cette structure de l'ADN en double hélice suggérait aussi comment l'ADN se duplique. Les deux chercheurs visualisèrent comment les deux brins complémentaires se séparaient pour permettre la synthèse de deux nouveaux brins. Cette idée de réplication semi-conservative était séduisante mais il manquait aux deux chercheurs la preuve nécessaire. Et il fallut attendre Matthew Meselson et Franklin Stahl qui mirent en œuvre ce qui allait être considéré comme l'une des plus belles expériences de la biologie.

Les deux chercheurs réalisèrent qu'ils pouvaient suivre la réplication de l'ADN en laissant pousser les bactéries *E. coli* pendant plusieurs générations dans un milieu composé d'azote lourd à 15 nucleons. Après avoir extrait l'ADN, ils le centrifugèrent sur un gradient de densité salin de manière à ce que l'ADN arrête sa migration lorsque sa densité devient égale à celle de la solution saline. L'ADN des bactéries ayant poussées dans un milieu avec de l'azote $^{15}$N avait une densité plus haute que l'ADN des bactéries ayant poussées dans un milieu avec de l'azote normal $^{14}$N. Les deux chercheurs firent alors une seconde expérience. Ils placèrent les bactéries qui avaient poussé sur de l'azote $^{15}$N dans un milieu avec de l'azote normal $^{14}$N. La croissance bactérienne était alors surveillée pour ne permettre aux bactéries qu'une seule division. Si Watson et Crick avaient raison alors Meselson et Stahl savaient à quoi s'attendre : les 2 brins lourds devaient s'être séparés et un nouveau brin léger devait s'être ajouté à chacun

des brins lourds. Après centrifugation sur le gradient de densité salin, l'ADN devrait former une bande à mi-distance entre la bande correspondant à l'ADN fait d'azote lourd ($^{15}$N) et celle fait d'azote léger ($^{14}$N). Et c'est exactement ce qu'ils virent (Meselson & Stahl, 1958).

> *« By means of density gradient centrifugation, we have observed the distribution of $^{15}N$ among molecules of bacterial DNA following the transfer of a uniformly $^{15}N$-substituted exponentially growing E. coli population to $^{14}N$ medium. We find that the nitrogen of a DNA molecule is divided equally between two physically continuous subunits ; that, following duplication, each daughter molecules receives one of these ; and that the subunits are conserved through many duplication »*

# 8 Un code génétique universel

Dés 1956, Francis Crick proposait ce qui allait devenir le dogme central de la biologie moléculaire. L'ADN est transcrit en ARN qui est traduit en protéine. Cependant, à cette époque, on ne savait pas encore quel dictionnaire utilisait *E. coli* pour traduire les instructions écrites dans le langage des gènes en langage des protéines. En 1961, Francis Crick, aidé par les suggestions d'un physicien théoricien nommé George Gamow, démontra qu'une succession de 3 bases d'ADN (appelé codon) codait pour une acide aminé. Commença alors une course pour « cracker » le code génétique. Marshall Nirenberg and Johann Matthaei découvrirent la première entrée du dictionnaire ADN–Protéine (Nirenberg & Matthaei, 1961). Ils broyèrent une culture d'*E. coli* et déposèrent le broyat dans une série de 20 tubes test. Ils ajoutèrent dans chacun des 20 tubes un acide aminé radioactif différent ainsi qu'un ARN synthétique poly-U (un ARN composé uniquement de bases uracile). Dans 19 tubes, rien ne se produisit. Mais, dans le 20$^{\text{ème}}$ tube, le samedi 27 mai 1961, à 3 heures du matin, les mesures de radioactivité indiquèrent la présence d'un long polypeptide composé entièrement de l'acide aminé phénylalanine. Nirenberg and Matthaei conclurent que la séquence d'ARN « UUU » codait l'addition d'une phénylalanine à n'importe quelle chaine de protéine en formation.

> *« The result indicates that polyuridylic acid contains the information for the synthesis of a protein having many of the characteristics of poly-L-phenylalanine »*

Dans les années qui suivirent, le code génétique fut entièrement décrypté (Nirenberg *et al.*, 1965). En 1967, Nirenberg et ses collègues annoncèrent qu'ils avaient trouvé « *un code génétique essentiellement universel* ». Cela signifie que tous les êtres vivants sont composés d'ADN, d'ARN et de protéines et que le code qui permet la transcription puis la traduction de l'ADN en protéine est le même pour tous les être vivants. Ainsi selon la fameuse expression de Jacques

Monod *« ce qui est vrai pour E. coli est vrai pour un éléphant »*. Ce qui différencie principalement deux organismes vivants, c'est la taille du génome et la suite des symboles (A,T,C,G) qui le constitue. *« C'est aux biologistes de ma génération qu'a été accordée cette révélation de la quasi-identité de la chimie cellulaire dans la biosphère entière. »*

# 9 La séquence semble aléatoire

Autre point important à noter c'est la révélation (et la déception) que la séquence d'une protéine (ou du gène correspondant) ne révèle aucune régularité ni aucune singularité qui pourrait aider à déterminer sa structure tridimensionnelle ou ses propriétés. Comme le remarque Jacques Monod, la loi générale qui semble décrire ces séquences, c'est celle du hasard.

> *« Ces structures sont « au hasard » en ce sens que, connaissant exactement l'ordre des 199 résidus dans une protéine qui en comprend 200, il est impossible de formuler aucune règle théorique ou empirique, qui permettrait de prévoir la nature du seul résidu non encore identifié par l'analyse. [...]Messages qui, par tous les critères possibles, semble avoir été écrit au hasard. Message cependant chargé d'un sens qui se révèle dans les interactions discriminatives, fonctionnelles, directement téléonomiques, de la structure globulaire, traduction à trois dimensions de la séquence linéaire »*[8]

La simplicité extrême du code génétique contraste avec la complexité extrême de la séquence d'un génome qui semble principalement guidée par le hasard. Je m'explique. Le code génétique est un code. Exactement comme un humain coderait un texte dont il veut éviter la lecture aux curieux. Casser ce code a été d'une simplicité enfantine (pardon pour les expérimentateurs dont je ne récolterai sans doute pas l'adhésion sur ce point et à juste titre). Ce code est très proche d'une construction cognitive triviale faite par l'homme et c'est cela qui m'est incompréhensible. A l'opposé de cette simplicité, remarquez la complexité de la séquence d'un génome. Cette dernière semble presque indistinguable d'une séquence aléatoire. Elle a subi des cycles et des cycles d'optimisation grâce à l'algorithme génétique qui tourne depuis plusieurs milliards d'années d'histoire évolutive. Le code sous jacent est capable d'engendrer un déterminisme inouï (l'organisme). Sa complexité, à l'opposé du code génétique, est pour l'instant totalement hors de portée et cela me semble déjà beaucoup plus naturel. Pourquoi certains mécanismes choisis par la nature nous semblent si transparents et d'autres si opaques ?

[8] Jacques Monod, *Le hasard et la nécessité*, Point, p. 126, 128.

La découverte du code génétique marque définitivement l'entrée d'une nouvelle discipline, la biologie moléculaire. Son dogme central décrit le sens de transfert de l'information génétique. L'ADN est transcrit en ARN qui est traduit en protéine. Ce transfert est à sens unique. Voici ci-dessous quelques citations de François Jacob.

> *« [La biologie moléculaire] s'installe délibérément à l'une des frontières du monde vivant, à la limite de l'inanimé. Le niveau inférieur se décrit en termes de chimie et de physique. Le niveau supérieur en termes d'organisation, de système logique [...] Mais la nature du système et les méthodes d'observation permettent toujours de considérer ensemble les deux niveaux, de comparer sans cesse le tout de l'organisme et le détail des constituants ou des phénomènes que met en évidence l'analyse [...] C'est dans la comparaison des deux niveaux qu'émerge la description de la cellule bactérienne. [...] L'analyse génétique ne vise plus uniquement alors simplement à démonter le mécanisme de l'hérédité. Elle devient un outil de précision servant à repérer les constituants de la cellule, leur rôle, leur interaction avec les autres éléments. Ce qu'on cherche avec les mutations, c'est à disséquer la cellule sans la détruire »*[9].

# 10 Un programme coordonné

Dans les années 60, le génome était considéré comme une série de plans de protéines : les gènes. L'intérêt de ces plans semblait résider principalement dans la capacité à être transmis à la descendance. Grâce à Jacob et Monod, le génome ne fut plus qu'un simple plan mais un programme codant pour un réseau complexe dans lequel les gènes régulent leurs activités pour permettre l'adaptation de la bactérie à son environnement (Jacob & Monod, 1961).

> *« According to the strictly structural concept, the genome is considered as a mosaic of independent molecular blue-prints for the building of individual cellular constituents. In the execution of these plans, however, co-ordination is evidently of absolute survival value. The discovery of regulator and operator genes, and of repressive regulation of the activity of structural genes, reveals that the genome contains not only a series of blue-prints, but a co-ordinated program of protein synthesis and the means of controlling its execution »*

En relisant Jacob et Monod, on se rend compte à quel point la description (le dogme) du fonctionnement de la cellule a relativement peu évolué en l'espace de quelques décennies. Ce

[9] François Jacob, *La logique du vivant*, Gallimard, p.284, 287.

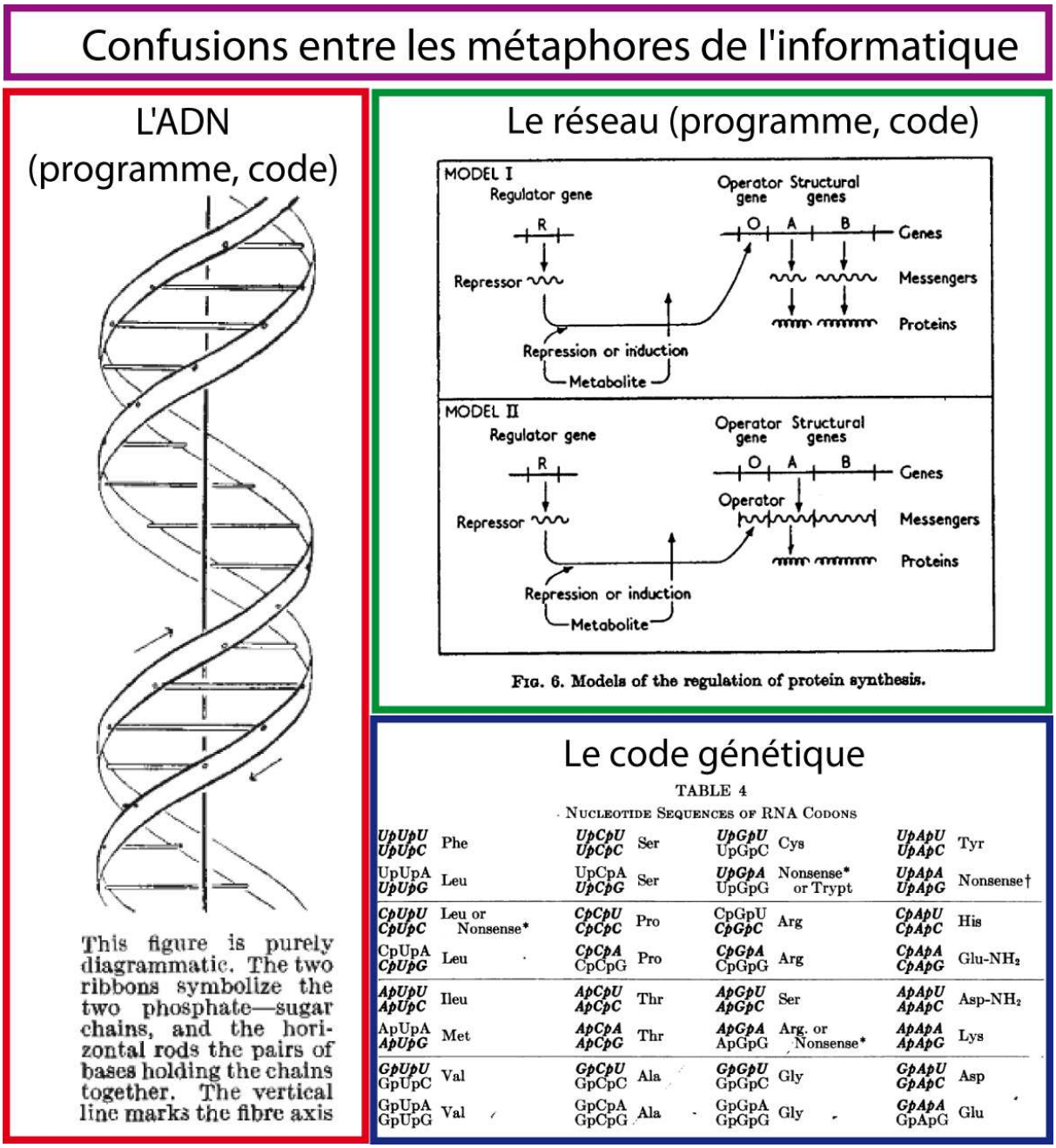



**Figure I.5** – Confusions entre les métaphores de l'informatique. Les notions de code et de programme peuvent être utilisées pour décrire différents concepts. Ultimement, toute l'information (le programme compressé) est encodée dans l'ADN. En « décompressant » l'ADN, on trouve un réseau (le programme décompressé). Le code génétique est une sous partie minuscule du réseau. Si vous ne comprenez pas cette légende, lisez le chapitre sur l'information. Les 3 schémas sont les schémas originaux des publications de Watson et Crick (l'ADN), Jacob et Monod (Le réseau) et Nirenberg (Le code génétique). Ces schemas représentent des éléments clés du paradigme actuel en biologie.

qu'ils expliquent en 1960–1970 dans leurs ouvrages respectifs correspond 40 ans plus tard au paradigme (manière de voir le monde) de départ sur lequel se basent mes travaux de thèse.

> *« En confrontant systématiquement la logique et l'expérience, selon la méthode scientifique, c'est en fait toute l'expérience de [nos] ancêtres que nous confrontons avec l'expérience actuelle*[10] *».*

Puisque que comme nous l'explique Jacques Monod, nous (scientifiques actuels) confrontons nos expériences avec celles de nos ancêtres, il me faut admettre qu'effectivement aucune de mes expériences n'a été en mesure d'ébranler ne serait-ce que d'un nanomètre ce paradigme. Certes vous me rétorquerez que cela n'est pas un gage de solidité... Mais j'insiste ici sur l'idée que le « but » de tout chercheur qui travaillerait sur un sujet similaire au mien —et c'est un sujet qui sera proposé encore longtemps— a pour mission de tenter, par tous les moyens, de briser ou de dépasser ce paradigme. Son échec renforce le paradigme. Plutôt que de paraphraser nos deux prix Nobel, je vous propose de faire une petite étude de leurs textes et du paradigme qui s'y cache.

# 11 Analyse de texte du paradigme actuel

Voici des citations de l'ouvrage de Jacques Monod, *le hasard et la nécessité*. Ce petit ouvrage, très facile d'accès, enchantera par la clarté de son propos. Les considérations politiques, que l'auteur aborde à la fin de son ouvrage, seront discutées plus tard.

> *« La pierre angulaire de la méthode scientifique est le postulat de l'objectivité de la Nature. C'est-à-dire le refus systématique de considérer comme pouvant conduire à une connaissance « vraie » toute* ***interprétation des phénomènes donnés en termes de causes finales, c'est-à-dire de « projet »****. [...] Postulat pur, à jamais indémontrable, car il est évidemment impossible d'imaginer une expérience qui pourrait prouver la non-existence d'un projet, d'un but poursuivi, où que ce soit dans la nature [...] L'objectivité cependant nous oblige à reconnaître le caractère téléonomique des êtres vivants, à admettre que dans leurs structures et performances,* ***ils réalisent et poursuivent un projet****. Il y a donc là, au moins en apparence, une contradiction épistémologique profonde. Le problème central de la biologie, c'est cette contradiction elle-même, qu'il s'agit de résoudre si elle n'est qu'apparente, ou de prouver radicalement insoluble si en vérité il en est bien ainsi. »*

[10] Jacques Monod, *Le hasard et la nécessité*, Point, p. 198.

*« Nous admettrons cependant l'hypothèse que, dans ce système, les interactions moléculaires qui assurent la transmission et l'interprétation des signaux chimiques sont dues à des protéines douées de propriétés de reconnaissance stéréospécifiques différentielles, auxquelles s'applique* ***le principe essentiel de gratuité chimique*** *tel qu'il se dégage de l'étude des interactions allostériques proprement dites. »*

*« On peut donc voir une contradiction dans le fait de dire que le génome « définit entièrement » la fonction d'une protéine, alors que cette fonction est attachée à une structure tridimensionnelle dont le contenu informatif est plus riche que la contribution directement apportée à cette structure par la détermination génétique [...]* ***l'enrichissement d'information correspondant à la formation de la structure tridimensionnelle provient de ce que l'information génétique (la séquence) s'exprime en fait dans des conditions initiales bien définies (en phase aqueuse, entre certaines limites étroites, de températures, composition ionique, etc. )*** *telles que parmi toutes les structures possibles, une seule d'entre elles est en fait réalisable »*

*« Il s'ensuit en effet qu'il n'y a pas de mécanisme possible par quoi la structure et les performances d'une protéine pourraient être modifiées et ces modifications transmises, fût-ce partiellement, à la descendance, si ce n'est comme conséquence d'une altération des instructions représentées par un segment de séquence de l'ADN. Tandis qu'inversement,* ***il n'existe aucun mécanisme concevable par quoi une instruction ou information quelconque pourrait être transférée à l'ADN****. Le système tout entier, par conséquent, est totalement, intensément conservateur, fermé sur soi-même, et absolument* ***incapable de recevoir quelque enseignement*** *que ce soit du monde extérieur »*

*« Nous disons que* ***ces altérations sont accidentelles, qu'elles ont lieu au hasard****. Et puisqu'elles constituent la seule source possible de modifications du texte génétique, seul dépositaire à son tour des structures héréditaires de l'organisme, il s'ensuit nécessairement que le hasard seul est à la source de toute nouveauté, de toute création dans la biosphère. Le hasard pur, le seul hasard, liberté absolue mais aveugle, à la racine même du prodigieux édifice de l'évolution »*

*« En ce sens l'évolution sélective, fondée sur le choix des rares et précieux incidents que contient aussi, parmi une infinité d'autres, l'immense réservoir du hasard microscopique, constitue une sorte de* ***machine à remonter dans le temps*** *»*

Passons maintenant à l'ouvrage de François Jacob *la logique du vivant*, livre plus long et donc un petit peu plus difficile d'accès mais qui représente une formidable histoire de la biologie. Inutile de préciser qu'il a eu une très forte influence sur ce chapitre.

> *« Ce qui est chiffré le long de la chaîne nucléique, ce qui est recopié signe par signe pour être scrupuleusement transmis d'une génération à l'autre, c'est la collection des plans détaillant les architectures de la cellule bactérienne ; c'est l'ensemble des prescriptions permettant d'échafauder minutieusement toute la série des édifices protéiques. Le programme n'est pas transcrit et traduit d'un trait mais par segments. La lecture du message peut être comparée, non à celle d'un rouleau qu'on déviderait d'un bout à l'autre, mais plutôt à celle d'un livre d'instructions dont on consulte les pages en fonction des besoins.* ***Certaines régions du programme contiennent des directives qui renvoient à d'autres régions selon les circonstances****. »*
>
> *« La lecture du message génétique ressemble ainsi à la musique que produisent les machines à disques dans les cafés. En appuyant sur un des boutons, on peut choisir parmi les disques de la machine, celui qu'on désire entendre. Mais en aucun cas on ne peut modifier ni le texte ni l'exécution de la musique enregistrée.* ***De même, un segment du texte génétique contenu dans le chromosome de la cellule bactérienne peut être transcrit ou non, selon les signaux chimiques reçus du milieu ; mais ceux-ci n'en peuvent modifier la séquence, donc la fonction****. Le vieux mot d'adaptation couvre ainsi deux choses différentes. D'un coté il s'agit d'un phénomène survenant chez l'individu ; il traduit en quelque sorte* ***la réponse de l'organisme à quelques facteurs externes****, mais c'est toujours dans les limites permises par les instructions contenues dans le programme. De l'autre coté, au contraire, il s'agit de* ***modifications survenant dans une population, c'est alors un changement du programme lui-même, sous l'effet d'une pression qui favorise certains des programmes à mesure qu'ils apparaissent.*** *»*
>
> *«* ***La plupart des situations auxquelles peut se trouver confronté un colibacille sont prévues dans le message****. Le programme contient ainsi les plans de toutes les pièces nécessaires pour faire une bactérie et il donne à celle-ci les moyens de faire face aux* ***difficultés de la vie courante****. Mais ce n'est qu'un programme. Dans les processus qui conduisent à recopier la séquence nucléique, soit par la reproduction soit pour les synthèses protéiques,* ***l'ADN joue le rôle passif d'une matrice****. Hors de la cellule, sans les moyens d'exécuter les plans, sans*

*l'appareillage de copie ou de traduction,* ***il reste inerte****, tout comme reste inerte une bande magnétique hors de son magnétophone. Pas plus que la mémoire d'une calculatrice, celle de l'hérédité n'agit par elle-même. Fonctionnel seulement au sein de la cellule, le message génétique ne fait rien tout seul. Il peut seulement guider ce qui fait.* ***Pour que soient produites les machines à partir des plans, il faut les machines****. Aucune des substances qu'on peut extraire de la cellule n'a la propriété de se reproduire. Seule la bactérie, la cellule intacte, est capable de croitre et se reproduire* ***car elle seule possède à la fois le programme et le mode d'emploi, les plans et les moyens de les exécuter****. »*

*« Ces moyens d'exécution, ce sont les protéines. Sur leurs propriétés reposent toutes les activités de la cellule, son architecture, son intégration.* ***Les protéines agissent, non pas en formant des liaisons chimiques, mais en s'associant à d'autres composés****. Leur structure leur confère en effet une vertu unique : celle de « reconnaitre » avec exactitude une ou plusieurs espèces chimiques seulement dans la mixture la plus hétérogène.* ***La précision de ce choix, sa spécificité déterminent les relations qui s'établissent entre les constituants de la cellule****. Elles en gouvernent toute la chimie. »*

*« Dans un système aussi complexe, seule la coordination des éléments donne une unité au système. Formée de quelques milliers d'espèce moléculaires, siège de quelques milliers de réaction chimiques qui se déroulent simultanément à toute vitesse, la cellule bactérienne ne peut former une totalité fonctionnelle sans une étroite cohésion de ses constituants. Tous les échanges de matière et d'énergie doivent être réglés dans le détail pour que puisse s'accomplir* ***le dessein de la bactérie : produire deux bactéries****. La petite cellule bactérienne ne saurait donc être une simple collection d'espèces moléculaires enfermées dans un sac et soumises aux lois statistiques qui régissent les éléments simplement juxtaposés et indépendants les uns des autres. Il faut* ***un réseau de communication*** *pour tenir informés les constituants qui se trouvent éloignés les uns des autres à l'échelle atomique et pour en commander les activités particulières en fonction de l'intérêt général et du but commun. A toutes les étapes de la chimie cellulaire interviennent des circuits de régulation pour coordonner les réactions et les ajuster aux exigences de la production.* ***Moyennant une faible dépense d'énergie, la cellule adapte ainsi son travail à ses besoins****. Elle ne produit que ce qu'il lui faut quand il lui faut. L'usine chimique est entièrement automatique. »*

*« Chaque système vivant relève alors de deux plans d'analyse, de deux coupes, l'une horizontale, l'autre verticale, qui ne peuvent être dissociées que pour la commodité de l'exposé. D'un coté, il s'agit de distinguer* ***les principes qui régissent l'intégration des organismes, leur construction****, leur fonctionnement ; de l'autre,* ***ceux qui ont dirigé leurs transformations et leur succession****. Décrire un système vivant, c'est se référer aussi bien à la logique de son organisation qu'à celle de son évolution. C'est aux* ***algorithmes du monde vivant*** *que s'intéresse aujourd'hui la biologie. »*

# 12 Commentaires et description du paradigme actuel

Quelles sont donc les composantes essentielles qui structurent ce paradigme ? Le lecteur me pardonnera ici les surinterprétations dont je suis seul responsable.

## 12.1 Le réseau, le programme et l'algorithme

Ce réseau complexe permet de coordonner l'action des milliers d'espèces moléculaires et de réactions chimiques et permet à la cellule d'être fonctionnelle. Contrairement à un réseau de neurones ou aux transistors, ce réseau est « virtuel » : les branches/flèches qui le composent représentent l'affinité qu'ont certains composés pour d'autres ce qui leurs permet de se réunir. Les fondements théoriques qui permettent l'existence de ce réseau dynamique sont donc : le rôle prépondérant de la liaison faible sur la liaison forte et le mouvement brownien des particules. Ce réseau est non linéaire, c'est-à-dire qu'il contient de nombreuses boucles de rétroaction qui rendent son comportement dynamique très difficile à déterminer avec l'intuition seule. *De ce réseau émerge la vie de la bactérie.*

Jacob et Monod ont démontré que certaines protéines peuvent contrôler l'expression des gènes. C'est l'exemple de l'operon lactose sur lequel je reviendrai. Si un gène code pour une protéine et qu'une protéine peut contrôler l'expression d'un gène alors un gène peut contrôler l'expression d'un autre gène. Ainsi le réseau est lui-même encodé dans le génome. Ce dernier contient donc deux types d'informations : le plan de toutes les protéines et le programme chargé d'adapter la concentration de chaque protéine en fonction de l'environnement extérieur.

On remarque d'emblée qu'une certaine confusion existe autour de la notion de programme et d'algorithme. Qui est le programme/l'algorithme ? Le réseau ? Ou la séquence du génome qui l'encode ? Pour le biologiste synthétique/des systèmes, le programme c'est le réseau. En effet, c'est le réseau qui est, en général, modélisé sous forme d'algorithme pour dévoiler son

fonctionnement dynamique. A l'inverse, pour le biologiste théoricien, le programme, l'algorithme c'est la séquence du génome car, au final, il contient sous forme compressé (presque) toute l'information contenue dans le réseau.

## 12.2 L'autoreproduction, un but en soi ?

La bactérie possède la faculté d'autoreproduction/autoréplication. Levons l'ambigüité entre ces deux termes. Répliquer c'est faire une copie fidele d'un objet. La photocopie d'un document ou la copie de la molécule d'ADN en deux molécules filles sont des exemples de processus de réplication. Par contre la reproduction sous-entend qu'un organisme autonome participe à la création ou crée un deuxième organisme lui aussi autonome. Le problème arrive quand on rajoute le préfixe « auto » qui implique l'autonomie. Qu'advient-il alors de la différence entre autoréplication et autoreproduction ? Dans le cas d'une bactérie, il y a ambigüité car la cellule mère donne deux cellules filles totalement identiques (ce n'est pas tout à fait vrai) et donc les termes d'autoréplication et d'autoreproduction conviennent. Prenons maintenant le cas d'un organisme supérieur comme le Dragon de Komodo qui est capable de se reproduire par parthénogenèse (forme de reproduction asexuée). La femelle pond un œuf. Même si cet œuf a le même génome que sa mère, ce n'est pas la copie conforme de sa mère. Contrairement à l'autoréplication, l'autoreproduction peut inclure la remise en route, à partir d'un état initial, d'un programme qui conduit à l'état de parent.

François Jacob relève bien que le rêve de toute bactérie : c'est devenir deux bactéries. Il y aurait donc un projet, un but, un dessein mais cela rentre en contradiction avec la méthode scientifique qui refuse toute téléologie, toute finalité —domaine réservé à la métaphysique—. Pour trouver un terrain d'entente provisoire, Jacques Monod introduit le terme de « téléonomie », concept scientifique de finalité. Cependant, la profondeur de cette contradiction épistémologique ne lui échappe pas. J'y reviendrai dans le chapitre sur l'épistémologie.

## 12.3 Quel est le coût énergétique d'un calcul/changement d'état du réseau ?

Les deux auteurs ont bien remarqué que se pose la question du coût thermodynamique du calcul effectué par la cellule lors de son adaptation à un changement de milieu. Jacques Monod nous parle de gratuité chimique des interactions allostériques (une molécule en se fixant à une deuxième modifie sa forme et sa fonction). A l'inverse, François Jacob nous explique que l'adaptation de la cellule se fait moyennant une faible dépense d'énergie.

Il semble apparemment possible qu'une cascade de changements allostériques dans la cellule suite à un changement dans le milieu puisse avoir un coût énergétique nul[11]. A l'inverse, d'autres types de changements comme par exemple une modification d'expression génique nécessite une grande consommation d'ATP et donc d'énergie.

Mais il y a une autre question controversée, plus difficile et plus profonde, qui se pose à la physique : le coût thermodynamique minimum qu'entrainent la manipulation et le stockage de l'information. Il semblerait possible, d'un point de vue théorique, de faire un calcul (ou de mesurer une information) sans consommation d'énergie. Et ce serait plutôt l'effacement de la mémoire qui aurait un coût thermodynamique. Et il n'est pas impossible que ces questions théoriques puissent s'appliquer à la manière dont une cellule gère l'information.

Jacques Monod remarque aussi qu'une part de l'information ne vient pas du génome mais du milieu qui fournit des conditions initiales bien définies (température, composition ionique, phase aqueuse). Nous reviendrons sur ce concept d'information, y compris sur celle apportée par le milieu.

## 12.4 Temps, évolution et adaptation

Nous avons vu que le transfert de l'information génétique est à sens unique. Autrement dit, les protéines ne codent pas pour les gènes. En revanche, lorsque les conditions extérieures changent, certaines protéines vont détecter la modification puis modifier, directement ou indirectement, l'expression de certains gènes de manière à réadapter la concentration de certaines protéines aux nouvelles conditions. La bactérie s'adapte à son environnement. Autrement dit, l'environnement ne peut pas changer la séquence du génome (le texte) mais peut changer son expression (lire certaines pages et pas d'autres). D'une certaine manière, le contenu cellulaire, une fois l'adaptation terminée, reflète en miroir le milieu extérieur. On peut imaginer qu'en connaissant le contenu cellulaire, on puisse inférer certaines caractéristiques de l'environnement extérieur et réciproquement. Nous pourrions donc potentiellement répondre à une question dynamique : *y'a-t-il du glucose dans le milieu maintenant, là ?* S'il y a un grand nombre de transporteurs du glucose dans la cellule, cela pourrait signifier qu'il y a du glucose dans le milieu.

Mais comment le plan (le génome) change t'il ? Le plan change grâce aux effets de l'évolution, au couple sélection–mutation. Lorsque la bactérie recopie son génome, elle introduit des erreurs de manière strictement aléatoire. Les mutations qui confèrent un avantage aux bactéries sont sélectionnées. Au fil du temps évolutif (qui se mesure en générations), le génome est donc

[11] Excepté bien sûr le coût thermodynamique minimum

modelé. La connaissance de sa séquence nous permet potentiellement de répondre à une question plus statique : *sais-tu ce qu'est le glucose ?* Si le génome contient un gène codant pour un transporteur du glucose, alors la bactérie « sait » quand et comment utiliser le glucose même si elle n'en a jamais rencontré. Cela reflète aussi l'environnement à long terme. Le maintien, au cours de l'évolution, du gène du transporteur signifie que le glucose est une source de carbone courante et connue des générations précédentes qui l'ont souvent rencontrée. Si on découvrait une bactérie conservée dans la glace, ayant vécu il y a 1 milliard d'années, l'étude de ses gènes et de la fonction des protéines correspondantes pourrait nous donner des indications sur l'environnement extérieur de l'époque. C'est en cela que l'on peut potentiellement « remonter dans le temps ». Cependant l'arbre évolutif n'avance à travers le temps que dans un sens. Tout comme avec le 2$^{\text{ème}}$ principe de la thermodynamique, il y a une irréversibilité qui fait que l'on ne pourra jamais reconstituer l'arbre phylogénétique réel tel qu'il s'est construit c'est-à-dire incluant tous les organismes disparus sans laisser de trace.

Je souhaite nuancer ici l'idée que l'environnement ne peut en aucun cas changer la séquence du génome. En effet, l'environnement intestinal, où vit *E. coli*, a évidement largement contribué à modeler son génome mais cela passe nécessairement par la sélection naturelle. On peut dire que l'organisme choisit et façonne le milieu et, réciproquement, le milieu choisit et façonne l'organisme.

## 13 *E. coli* au 21$^{\text{ème}}$ siècle

On voit à travers ce rappel historique comment les images de la vie et d'une bactérie comme *E. coli* ont évolué au cours du temps. Après « l'étincelle vitale » dont étaient affublés les êtres vivants, les progrès en chimie ont montré que ces derniers étaient composés des mêmes atomes que la matière inerte. La différence tenait dans l'existence de petits catalyseurs complexes capable de catalyser des réactions chimiques extrêmement difficiles à réaliser en laboratoire. L'influence de la thermodynamique faisait des bactéries des fours capables de transformer les molécules organiques en chaleur. Puis émerge dans les années 50–60 une nouvelle branche de la biologie : la biologie moléculaire, très largement influencée par des physiciens comme Erwin Schrödinger ou Max Delbrück. Le *code* génétique, le *programme* cordonné et le *réseau* de régulation font entrevoir la vie sous le prisme de l'information. Les analogies entre les ordinateurs et la vie sont si fortes que les biologistes ne veulent plus comprendre la « vie » mais « les algorithmes du monde vivant ». Comme le dit Bernd-Olaf Küppers, la vie est matière plus information. Le paradigme ou le dogme toujours en vigueur aujourd'hui est posé : la vie *émerge* d'un réseau

de régulation complexe créé principalement par les interactions faibles entre différentes espèces moléculaires. Même si une part d'information vient du milieu, le réseau de régulation et les espèces moléculaires (la plupart) sont encodés dans le génome qui, lui, est modelé par l'évolution. En plus de partager les mêmes atomes, les êtres vivants partagent des molécules communes (ADN, ARN, protéines), un set commun de réactions enzymatiques très proches et un code génétique universel. Ce qui les distingue c'est la taille et la séquence de leur génome.

Après que le paradigme ait été bien posé sur ses bases, 40 ans de recherche en biologie moléculaire et en biochimie ont suivi, permettant une description de plus en plus précise des mécanismes moléculaires gouvernant le fonctionnement d'une bactérie comme *E. coli*. Cette dernière est aujourd'hui l'organisme vivant le plus facile à manipuler et dont le fonctionnement est le mieux décrit. On compte, à la fin du 20ème siècle, plus de 100 000 publications scientifiques sur *E. coli*. En 1997, le séquençage complet du génome d'*E. coli* est achevé (Blattner *et al.*, 1997). Tous ces éléments offrent alors le terreau favorable à l'émergence, au début du 21ème siècle, de deux disciplines jumelles : la biologie des systèmes et la biologie synthétique. Ces deux disciplines font le constat que les méthodes et techniques de la biologie ne permettront pas d'expliquer la dynamique de fonctionnement d'un système vivant à partir de la simple compréhension des briques de base. La biologie se heurte à ce que les mathématiciens, physiciens et informaticiens connaissent déjà bien : la complexité. L'interdisciplinarité devient de plus en plus évidente, voire même indispensable. D'ailleurs, selon Hervé Zwirn, on peut *« voir les sciences de la complexité comme la première avancée vers un affaiblissement progressif de l'importance des frontières disciplinaires »*[12] . Cette interdisciplinarité change petit à petit le regard porté sur *E. coli*. Ainsi :

⇒ *Pour un biologiste « classique » : E. coli* est une bactérie intestinale, potentiellement pathogène. Son étude permet de décrypter beaucoup des mécanismes moléculaires présents ou non chez d'autres espèces ou même d'étudier les processus évolutifs. La découverte d'une nouvelle fonction (protéine), d'une nouvelle régulation (interaction), d'une nouvelle cible d'antibiotique ou de nouveaux mécanismes « entièrement nouveaux » tel le rôle des petits ARN sont les objectifs de ce biologiste à plus ou moins long terme. Ses outils sont les techniques de la biologie moléculaire, de la génétique, de la microbiologie et de la biochimie.

⇒ *Pour un biologiste des systèmes/synthétique : E. coli* est une entité autonome, auto réplicative, capable de calcul et donc de traitement de l'information. L'intérêt d'*E. coli* c'est qu'il s'agit d'une boite noire mais c'est la moins noire des boites noires vivantes.

[12]Hervé Zwirn, *Les systèmes complexes*, Odile Jacob, p. 211.

Le biologiste espère, par exemple, contrôler de plus en plus « d'input » et « d'output » différents pour essayer, par ingénierie inverse, d'élucider son fonctionnement dynamique. En effet, il pense que son réseau de régulation non linéaire est un modèle idéal pour étudier la complexité. Il souhaite démontrer que l'hypothèse « la vie émerge du fonctionnement du réseau de régulation » est un fait scientifique « poppérien » et non une idéologie. Pour cela, il veut contrôler le réseau avec de plus en plus de précision, tirer les ficelles pour décider du phénotype. Citons le bon mot (traduit en français) du prix Nobel Hamilton Smith, quand un journaliste lui demande s'il a l'impression de jouer à dieu : sa réponse : « on ne joue pas » (rire). Si la recherche fait une percée conséquente dans les décennies à venir sur la question de la complexité chez les êtres vivants, les conséquences pour la civilisation humaine peuvent être gigantesques, plus grosses encore que l'arrivée d'internet ou du transistor. Le but du biologiste des systèmes/synthétique : participer à cette révolution. Ces outils sont ceux du biologiste classique ainsi que ceux des mathématiciens, physiciens, informaticiens et ingénieurs. Peut être plus que les outils, c'est le mode de pensée de ces disciplines que ce biologiste doit acquérir. La double casquette lui permet d'entrevoir les problèmes sous des angles différents.

Arrivé au terme de notre périple historique, je souhaite terminer ce chapitre en énumérant les différents types d'explications que l'on trouve en biologie au 21$^{\text{ème}}$ siècle. En effet, si les modes et les biologistes changent, les principes d'intelligibilité perdurent (même s'ils ne sont ni eternels ni immuables). Selon Michel Morange, pour expliquer les phénomènes biologiques (par exemple une fonction donnée), les scientifiques se réfèrent aujourd'hui à 3 schèmes explicatifs[13] :

1. Les explications de type molécularo-mécaniste : elles consistent en l'établissement d'une chaine de causalité, où chaque maillon correspond à une interaction entre une ou quelques molécules.
2. Les explications de type darwinien : une caractéristique (fonction, forme, comportements...) peut être observée parce qu'elle apporte aux individus, directement ou indirectement, une probabilité plus grande de laisser des descendants.
3. Les explications de type physique non causal : elles reposent sur l'existence de relations et de contraintes atemporelles. Par exemple, la capside des virus repose sur des lois géométriques simples. Un exemple que l'on doit à René Thom : décrire de manière intelligible la mécanique de locomotion d'un chat ne nécessite pas de descendre aux niveaux des interactions moléculaires[14]. Un autre exemple amusant dont je ne retrouve pas l'auteur :

[13]Michel Morange, *Les secrets du vivant*, La découverte.
[14]René Thom, *Prédire n'est pas expliquer*, Flammarion, p. 123.

quand un poisson volant sort de l'eau puis finit par retomber dans l'eau : ni Darwin ni le « réseau » du poisson ne sont des schèmes explicatifs appropriés pour décrire la retombée. Les lois de la gravitation sont bien plus adéquates. Un dernier exemple est la structure hexagonale des alvéoles d'abeille qui minimise la quantité de cire nécessaire au pavage du plan et correspond donc à la solution la plus économique. Cet exemple que j'ai trouvé dans un ouvrage de Jean-louis Le Moigne me permet de montrer *ce qui n'est pas* un principe d'intelligibilité : le principe de moindre action, purement idéologique, qui permettrait d'expliquer tout phénomène dés lors que l'on connait la fonction à optimiser[15].

[15]Jean-Louis Le Moigne, *Les épistémologies constructivistes*, PUF, p. 36.

# Références bibliographiques


Avery, O. T., Macleod, C. M., & McCarty, M. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med*, **79**(2), 137–158. 7

Beadle, G. W., & Tatum, E. L. 1941. Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A*, **27**(11), 499–506. 5

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science*, **277**(5331), 1453–62. 23

Hershey, A. D., & Chase, M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*, **36**(1), 39–56. 8

Jacob, F., & Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, **3**, 318–56. 13

Meselson, M., & Stahl, F. W. 1958. The Replication of DNA in Escherichia Coli. *Proc Natl Acad Sci U S A*, **44**(7), 671–82. 11

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., & O'Neal, C. 1965. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci U S A*, **53**(5), 1161–8. 11

Nirenberg, M. W., & Matthaei, J. H. 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A*, **47**, 1588–602. 11

Tatum, E. L., & Lederberg, J. 1947. Gene Recombination in the Bacterium Escherichia coli. *J Bacteriol*, **53**(6), 673–84. 6

## Références bibliographiques

Watson, J. D., & Crick, F. H. 1953. Molecular structure of nucleic acids ; a structure for deoxyribose nucleic acid. *Nature*, **171**(4356), 737–8. 10

# Chapitre II

# La bactérie *Escherichia coli*

## 1 Préambule

Ce chapitre est relativement facile d'accès. Il est écrit pour les non-biologistes (par exemple les mathématiciens) qui veulent se forger rapidement une image globale d'une bactérie comme *E. coli*. En une quinzaine de pages, le lecteur se fera une idée des grands domaines d'étude sur *E. coli* autres que les réseaux de régulation (réservé au chapitre suivant). Je l'ai aussi écrit en pensant aux jeunes biologistes synthétiques qui pourraient potentiellement y puiser quelques idées intéressantes : par exemple, je leur rappelle qu'*E. coli* peut se développer dans des cellules eucaryotes et détourner le cytosquelette d'actine pour se déplacer.

## 2 Différence entre procaryotes et eucaryotes

Pour éviter toutes confusions dans l'esprit d'un lecteur peu familier avec la biologie, je souhaite rappeler ici quelques différences essentielles entre cellules procaryotes (grossièrement les bactéries dont *E. coli*) et les cellules eucaryotes (par exemple, les cellules humaines ou les levures). Première différence : la taille. *E. coli* mesure 2µm de long sur 0.5µm de large (1µm$^3$), c'est-à-dire en général 10 à 100 fois plus petite qu'une cellule eucaryote. Les bactéries n'ont pas de noyau et en général elles ne possèdent qu'un seul chromosome circulaire. L'absence de noyau a une conséquence évidente : la transcription se fait directement dans le cytoplasme. De plus, Il n'y a pas d'intron donc pas d'épissage. Notons aussi l'absence de structures cytoplasmiques comme le réticulum endoplasmique, les mitochondries, les chloroplastes ou le cytosquelette (microtubule). Enfin, le mode de reproduction d'*E. coli* est la division simple alors que les cellules eucaryotes se divisent par mitose (multiplication conforme de la cellule) ou par méiose (formation de gamètes). Ainsi *E. coli* est un organisme très simple comparé aux eucaryotes. L'absence de compartiments intracellulaires est probablement la différence primordiale. En tout

cas du point de vue du modélisateur.

# 3 Réplication, transcription, traduction et repliement des protéines

*E. coli* possède un chromosome circulaire de 4.6 millions de paire de bases (pb). Environ 4500 gènes (soit en moyenne 1000 pb par gène) ont été localisés et 85% ont une fonction connue. Restent environ 850 dont la fonction est inconnue (Merlin *et al.*, 2002). Plus ce chiffre diminuera, plus la difficulté pour déterminer leurs fonctions augmentera mais plus on risquera de découvrir de nouveaux paradigmes dans le fonctionnement d'une cellule. On estime que *E. coli* dispose d'environ 303 gènes essentiels c'est-à-dire dont la disruption/délétion entraîne la non viabilité de la cellule (Baba *et al.*, 2006).

Le chromosome se trouve sous forme super-enroulé dans la bactérie. Le super enroulement de l'ADN est finement contrôlé par des enzymes : les topoisomerases cibles de plusieurs antibiotiques dont les fluoroquinolones (Nollmann *et al.*, 2007). Des protéines ressemblant à des histones (HU, HNS, fis, IHF) structurent également le chromosome mais les procaryotes ne possèdent pas d'histone à proprement parler (Ali Azam *et al.*, 1999).

La réplication du chromosome d'*E. coli* dure approximativement 42 minutes et débute au niveau d'une origine de réplication unique. La vitesse de réplication de l'ADN polymérase est de 1000pb.s$^{-1}$. Parfois, cette dernière fait une erreur et introduit une mutation à une fréquence de $10^{-7}$ mais des mécanismes de correction des erreurs interviennent pour diminuer cette fréquence. Au final, *E. coli* fait des erreurs à une fréquence de l'ordre de $10^{-9}$ par nucléotide répliqué. Notez que si cette fréquence était nulle, *E. coli* ne pourrait pas évoluer et ne pourrait donc pas survivre. En effet, bien que ces mutations sont la plupart du temps soit silencieuses soit délétères, dans de rares cas, elles sont bénéfiques ce qui leurs permet d'être sélectionnées au cours de l'évolution et donc d'être fixées. Ces mutations peuvent apparaître dans un gène pour modifier la fonction de la protéine correspondante mais aussi en dehors des gènes, par exemple dans les zones promotrices pour modifier le réseau de régulation. Un informaticien ou mathématicien peut se représenter ces mécanismes évolutifs comme une heuristique de recherche d'optimum global. Un paramètre varie aléatoirement (les mutations) et éloigne le plus souvent de la solution optimale. Cependant, parfois la variation permet de s'échapper d'un optimum local. Dans ce problème d'optimisation, la fonction optimisée par la nature est le fitness de la bactérie défini comme la capacité d'un individu à se reproduire. Nous reviendrons largement, dans le chapitre sur l'évolution, sur le rôle et l'utilisation de l'évolution pour « tuner » un réseau de régulation

| | | | |
|---|---|---|---|
| Longueur d'une cellule | 2µm | Volume occupé par l'eau | 70% |
| Volume de la cellule | ~1µm$^3$ | Volume occupé par les protéines | 17% |
| Taille moyenne d'une protéine | ~5nm | Volume occupé par tout les ARN | 6% |
| Taille du génome | 4.6 10$^6$ pb (1,55mm si deroulé) | Volume occupé par les ARN ribosomiaux | 5% |
| Nombre de gènes | ~4500 | Volume occupé par les ARN de transfert | 0.8% |
| Taille d'un site de liaison sur l'ADN | ~10pb | Volume occupé par les ARN messager | 0.2% |
| Taille d'un promoteur | ~100pb | Volume occupé par l'ADN | 1% |
| Taille d'un gène | ~1000pb | Volume occupé par les ribosomes | 8% |
| Espacement moyen entre protéines | 7 nm | Volume occupé par les lipides | 3% |
| Espacement moyen entre petites molécules | 3.6 nm | Volume occupé par les LPS | 1% |
| Concentration d'une protéine par cellule | ~1nm | Volume occupé par la muréine | 1% |
| Temps de diffusion d'une protéine à travers la cellule | ~0.1 sec | Volume occupé par le glycogène | 1% |
| Temps de diffusion d'une petite molécule à travers la cellule | ~1 msec | Volume occupé par les ions | 0.3% |
| Temps pour transcrire un gène | 1 min (70pb/sec) | Volume occupé par les petits composés | 1% |
| Temps pour traduire une protéine | 2 min (40aa/sec) | Nombre d'ARN messager par cellule | 4000 |
| Demi-vie d'un ARNm | 2-5 min | Nombre d'ARN ribosomiaux par cellule | 18000 |
| Temps de génération minimum (en milieu riche) | 20 min | Nombre d'ARN de transfert par cellule | 2 10$^5$ |
| Nombre de ribosomes par cellule | 2 10$^4$ | Nombre total d'ARN par cellule | 2.2 10$^5$ |
| Transition entre deux états protéiques (actif/inactif) | 1-100µsec | Nombre de polysaccharides par cellule | 39000 |
| Ordre de grandeur de temps pour qu'un facteur de transcription se fixe à son site de liaison | 1msec | Nombre de lipides par cellule | 25 10$^6$ |
| Ordre de grandeur de temps pour qu'une petite molécule se fixe à une protéine (à l'équilibre) | 1 sec | Nombre de protéines par cellule | ~4 10$^6$ |
| Taux de mutation | ~10$^{-9}$ /pb/génération | Nombre de porines | 6 10$^4$ |

**Figure II.1** – Quelques chiffres bien pratiques pour se fixer les idées. Sources CyberCell (Sundararaj *et al.*, 2004) et (Alon, 2006)

génique .

Le chromosome contient de nombreux opérons qui sont des groupements de gènes transcrits ensemble. L'ARN polymérase, qui transcrit les gènes en ARN, se fixe au niveau du promoteur : une région de l'ordre de 100pb située à proximité d'un gène et indispensable à la transcription. La fixation et l'activation de l'ARN polymérase sont contrôlées par des facteurs de transcription qui se fixent sur leur site de liaison spécifique (10pb) au niveau de la région promotrice. Les facteurs de transcription peuvent être soit activateurs, soit répresseurs. Plus de 196 facteurs de transcription sont décrits parmi les 4500 gènes que compte *E. coli* K-12 (Keseler *et al.*, 2009). Certains facteurs de transcription dit « régulateurs globaux » sont capables de contrôler la transcription de plusieurs dizaines voire centaines de gènes (par exemple Crp, ArcA, FruR, Fnr, Mlc) (Perrenoud & Sauer, 2005). La transcription d'un gène prend environ 1 minute. Elle est ciblée par plusieurs antibiotiques dont la rifampicine.

Puis l'ARNm est traduit par un ribosome (environ $10^4$ par cellules) en environ 2 minutes. Le ribosome est la cible de nombreux antibiotiques comme le chloramphénicol. Notez qu'un ARNm peut être traduit plusieurs fois mais sa durée de vie est très faible, de l'ordre de 2 à 5 minutes. L'arrêt de la transcription d'un gène peut donc impacter assez rapidement la concentration de la protéine résultante. *E. coli* possède environ 4 millions de protéines (taille moyenne 5nm) par cellule. Chaque type de protéine se retrouve en moyenne à la concentration de 1nM c'est-à-dire quelques unités par cellules ; cependant certaines sont présentes en centaines de milliers de copies dans une cellule.

Une fois traduit, un polypeptide se replie dans sa structure tridimensionnelle caractéristique dans laquelle il est fonctionnel. Le mécanisme du repliement des protéines n'est pas encore compris cependant la conformation finale dite « native » est probablement énergétiquement favorable. Les repliements rapides prennent quelques microsecondes mais d'autres demandent plusieurs minutes à plusieurs heures pour se produire. Certaines protéines sont assistées dans leurs repliements par d'autres protéines dites chaperonnes (GroEL/GroEs ; DnaK/DnaJ) qui évitent ainsi les phénomènes d'agrégation de protéines mal repliées (Ben-Zvi & Goloubinoff, 2001). Enfin, il existe aussi un ensemble de protéines (ClpX,clpA) capables de reconnaître et de déplier des protéines marquées ou incorrectement repliées pour leur laisser une seconde chance de repliement ou au contraire les entrainer vers la dégradation (protéase clpP, lon) (Farrell *et al.*, 2005).

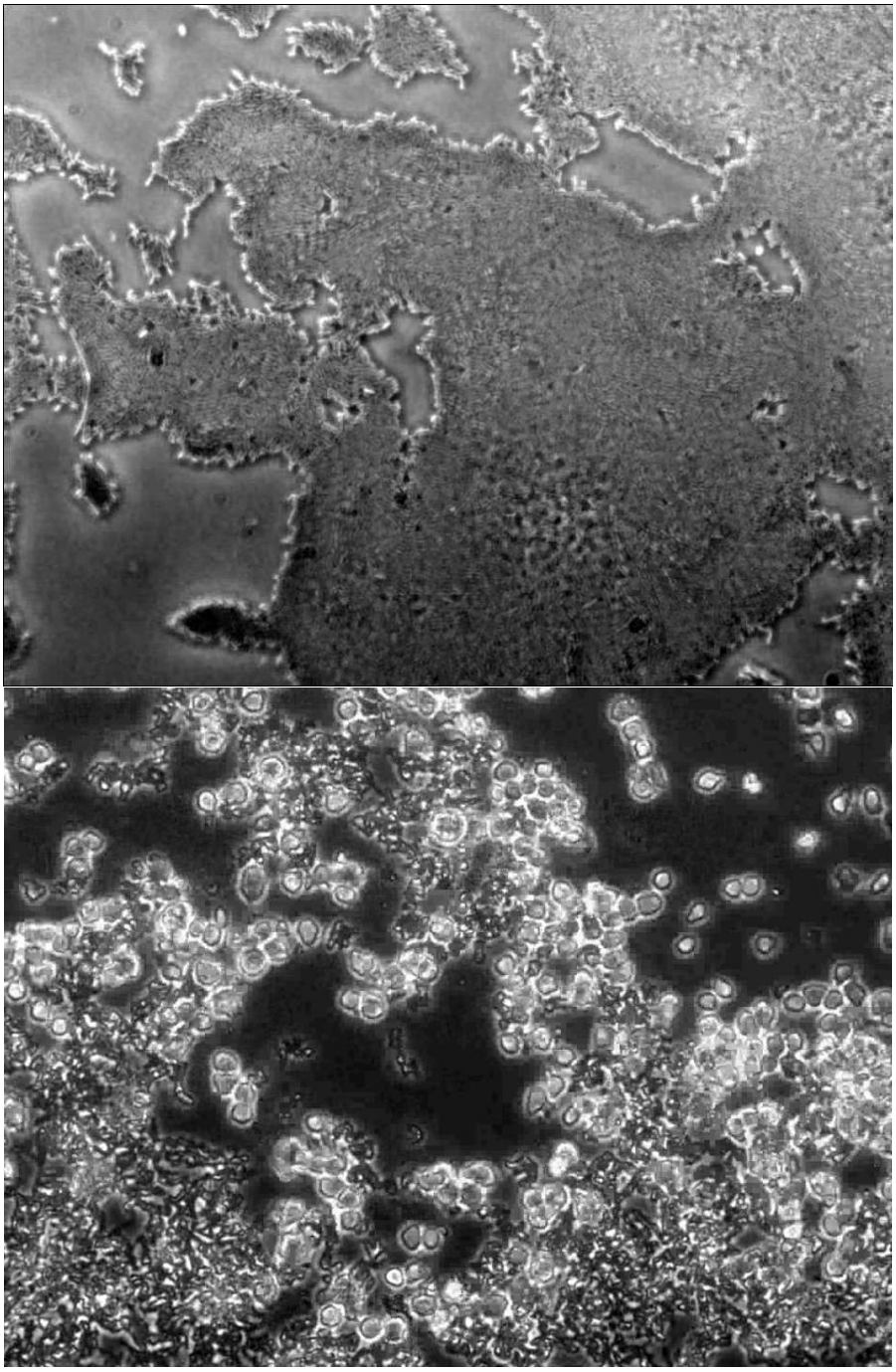


**Figure II.2** – Deux vidéos de microscopie. La première illustre la croissance exponentielle des bactéries *E. coli*. La deuxième représente une compétition entre des bactéries *E. coli* (qui forment une sorte de réseau) et des amibes (eucaryote) *Dictyostelium discoideum* qui se nourrissent des bactéries et finissent par envahir le champ du microscope. Les amibes dites « sociales » finissent même par se regrouper pour former une sorte d'être multicellulaire appelé « fruiting body ».

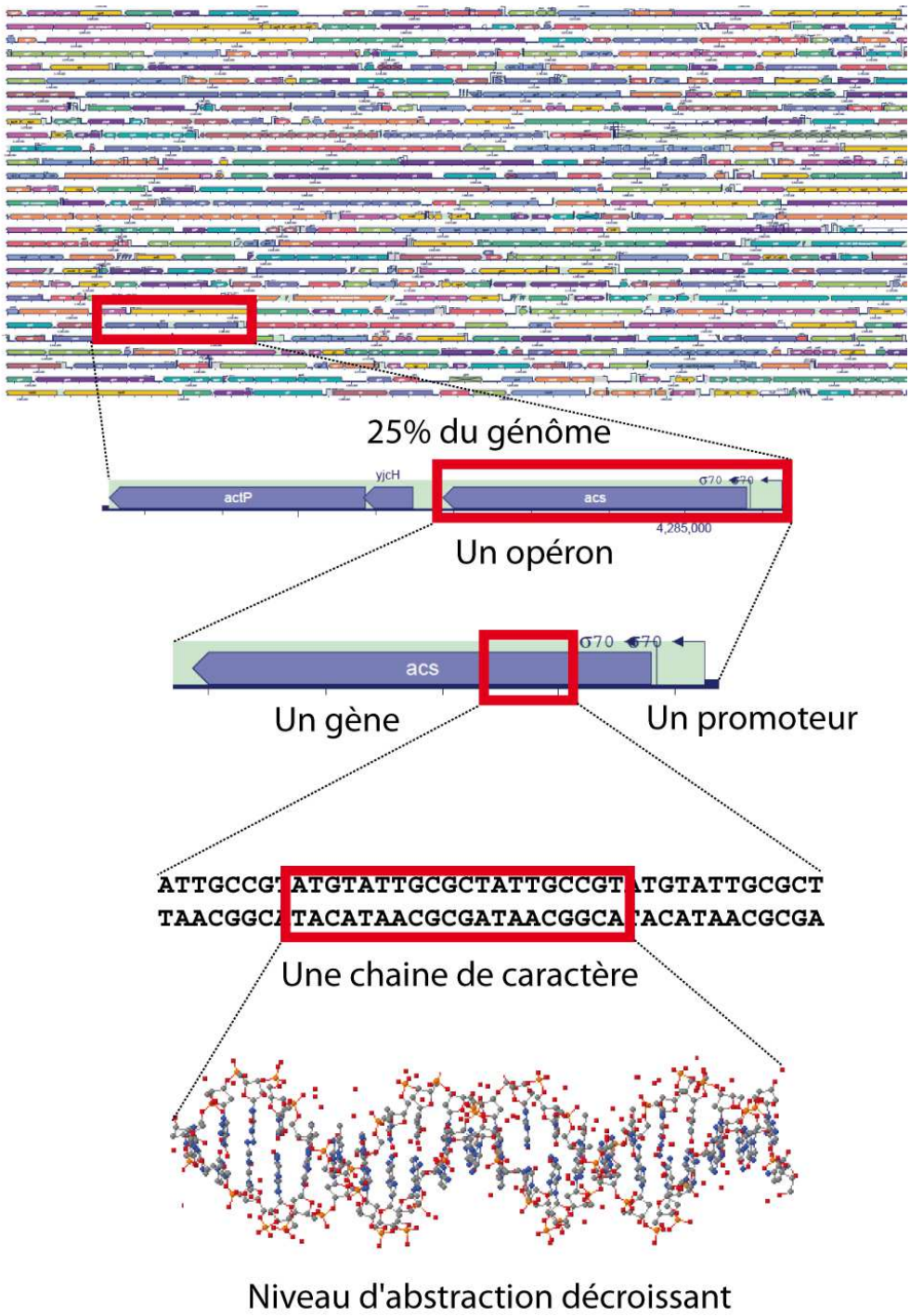



**Figure II.3** – Le génome et l'ADN. Le génome contient des milliers de gènes souvent représentés par des flèches permettant d'identifier le codon d'initiation et le codon stop. Ces flèches sont une abstraction de la séquence ADN (chaine de caractère faite de A, T, C et G) elle-même abstraction de la structure physique tridimensionnelle et atomique de l'ADN. Le haut de l'image est issu d'un poster provenant du site Web Ecocyc (Keseler *et al.*, 2009).

# 4 Les frontières de la vie : les membranes

La frontière entre le milieu extérieur et *E. coli* ne fait que quelques centaines d'atomes d'épaisseur. Mais cette frontière est loin d'être simple. *E. coli* est une bactérie dite à gram négatif dont la structure s'organise en trois grandes parties, soit, de l'intérieur vers l'extérieur, la membrane plasmique (bicouche lipidique), l'espace périplasmique et la membrane externe (bicouche lipidique). Les deux derniers forment ce qu'on appelle la paroi dont la formation est la cible de plusieurs antibiotiques comme la pénicilline. Cette frontière assure plusieurs fonctions. Elle assure le maintien de la forme de la bactérie et le maintien d'un environnement osmotiquement stable. Elle empêche les molécules du milieu extérieur de rentrer et garde les 60 millions de molécules intracellulaires proches les unes des autres pour permettre aux réactions de se dérouler rapidement.

Cependant, la bactérie a besoin d'une connexion avec le monde extérieur. Pour survivre et croître, un organisme doit importer du matériel brut et se débarrasser des déchets produits. *E. coli* dispose de dizaines de milliers de transporteurs protéiques enchâssés dans la membrane externe et dans la membrane plasmique. Ces pores, canaux ou pompes transportent, suivant le cas, dans un sens ou dans l'autre et de manière active ou passive. Le plus souvent, ces transporteurs sont spécifiques d'une molécule donnée. La discrimination se fait le plus souvent en fonction de leur forme. Plus de 254 transporteurs sont décrits parmi les 4500 gènes que compte *E. coli* K-12 (Keseler *et al.*, 2009). Dans la membrane, on trouve également des senseurs protéiques. Ces derniers détectent un changement dans l'environnement extérieur (par exemple, la présence d'une molécule) et transmettent l'information (le signal) à l'intérieur de la cellule le plus souvent par le biais d'une cascade de modifications chimiques (par exemple des phosphorylations) aboutissant à une modification de l'expression génique.

# 5 Le métabolisme

La présence d'une frontière, d'un chromosome et d'une machinerie de réplication, transcription et traduction ne suffit pas pour faire « émerger » la vie. La cellule représente un état hautement organisé de la matière : maintenir cet ordre tout en étant soumis aux principes de la thermodynamique nécessite la mise en place de structures permettant d'utiliser l'énergie (on crée de l'ordre au niveau de la cellule mais, globalement, l'entropie augmente). La cellule est donc un système thermodynamiquement ouvert nécessitant un métabolisme c'est-à-dire un ensemble de réactions chimiques qui se déroulent de manière ininterrompue. Le métabolisme est usuellement divisé en deux catégories : les processus de dégradation (catabolisme) par exemple

pour casser des molécules organiques et produire de l'énergie. Et des processus de synthèse organique (anabolisme) pour biosynthétiser des composés absents dans le milieu de culture comme des acides aminés ou des bases.

Les réactions chimiques sont organisées en chemins (pathway) métabolique dans lesquels un métabolite est transformé en un autre métabolite grâce à une série de réactions chimiques catalysées par une série d'enzymes. Certains chemins métaboliques sont très conservés à travers l'évolution. Tout comme le code génétique ou le dogme ADN–ARN–protéine, la glycolyse et le cycle de Krebs sont présents aussi bien chez *E. coli* que chez l'éléphant. La première espèce dont le métabolisme a été décrit en détail est justement *E. coli*. Cela a été un travail long et fastidieux qui a pris des décennies. Aujourd'hui, on compte plus de 1397 gènes codant pour des enzymes chez cet organisme (Keseler *et al.*, 2009).

# 6 La division cellulaire

La bactérie *E. coli* se divise par fission binaire. Dans un milieu riche, à 37°C sous agitation (respiration aérobie), *E. coli* se divise toutes les 20 minutes. Sa croissance est exponentielle. Autrement dit, si la croissance ne s'arrêtait pas, faute de nutriments ou suite à l'accumulation de déchets, *E. coli* recouvrirai la terre en quelques jours.

Le processus de division inclut la réplication du chromosome, la ségrégation des deux chromosomes de part et d'autre de chaque pôle de la bactérie et la formation d'un septum permettant la séparation en deux cellules filles. Cette séparation débute avec l'assemblage d'un anneau septal composé d'une douzaine de protéines. Le composé majeur de cet anneau septal est la protéine FtsZ, l'homologue procaryote de la tubuline, une des plus importantes protéines du cytosquelette chez les eucaryotes. La polymérisation de FtsZ forme un anneau contractile qui scinde la bactérie en deux cellules filles. Une souche, sans le gène ftsZ (un mutant de délétion), est en effet incapable de se diviser et forme de longs filaments pouvant faire plusieurs dizaines de micromètres. Une régulation spatiale et temporelle très fine assure le bon déroulement du processus de division. En effet, il faut inhiber la formation de l'anneau septal tant que la bactérie est trop petite et s'assurer que l'assemblage de cet anneau se crée bien au milieu de la bactérie et non au pôle (Margolin, 2001).

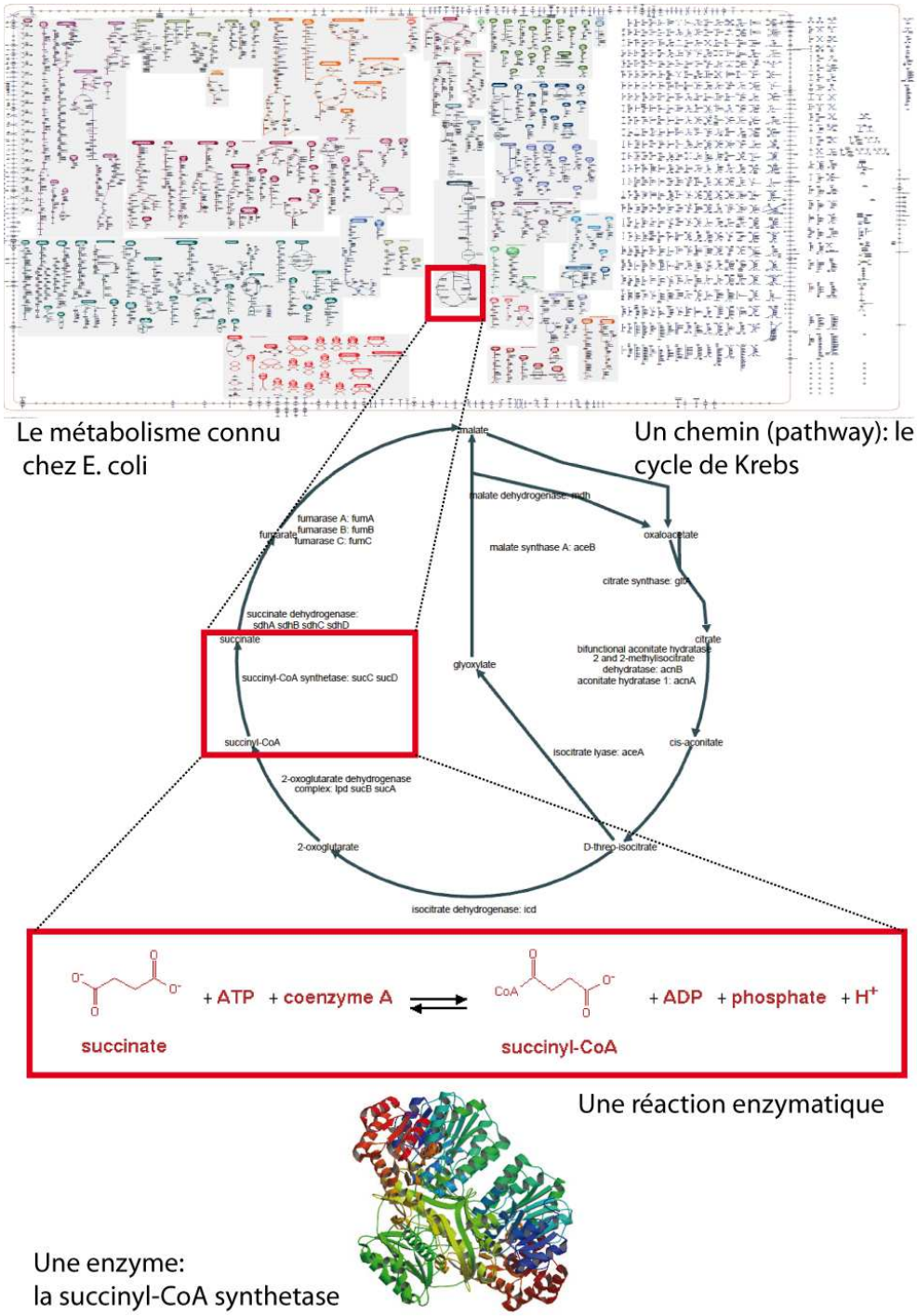



**Figure II.4** – Le métabolisme. Dans le cytoplasme d'*E. coli* se déroulent des nombreuses réactions chimiques différentes que l'on peut réunir en « pathway » (le haut de l'image est issu d'un poster provenant du site Web Ecocyc). Les réactions chimiques sont catalysées (accélérées) par des enzymes dont la structure tridimensionnelle a parfois été déterminée (par diffraction aux rayons X, RMN ou microscopie électronique).

# 7 L'adaptation au stress

Une cellule bactérienne n'a que très peu de possibilités pour choisir et modifier son environnement. Il lui faut donc disposer de mécanismes qui lui permettent de maintenir son homéostasie en répondant rapidement et efficacement à une variété de stress environnementaux qui pourraient menacer son intégrité. Pour cela, la bactérie dispose de senseurs et d'effecteurs capables de faire face, par exemple, à un changement brutal de température, à une forte irradiation ou à l'épuisement des nutriments dans le milieu. En général, ces mécanismes ont un impact négatif sur le taux de croissance. Ainsi, en fonction de conditions extérieures, la bactérie module nécessité de reproduction et nécessité de protection.

## 7.1 Réponse aux chocs thermiques

Un choc thermique a pour conséquence l'accumulation de protéines mal repliées. Ces dernières exposent des résidus hydrophobes à leur surface ce qui entraine leurs agrégations. Cette agrégation menace sérieusement l'organisation et le fonctionnement des composants cytoplasmiques. L'ARN polymérase d'*E. coli* est composée d'un cœur enzymatique (5 sous-unités) responsable de la polymérisation de l'ARN et d'une sixième sous-unité appelée facteur sigma qui permet la reconnaissance des régions promotrices et l'initiation de la transcription. 7 facteurs sigma différents ont été identifiés chez *E. coli* et chacun d'eux régule la transcription d'un groupe spécifique de gènes. En cas de choc thermique, la concentration du facteur sigma dit « 32 » augmente très fortement. Ce facteur sigma se lie au cœur de l'ARN polymérase et permet la transcription de 2 types de gènes : des chaperonnes qui assistent le repliement des protéines et de protéases qui dégradent les protéines mal repliées (Arsene *et al.*, 2000). La mise en place de cette régulation permet donc de limiter les effets négatifs du choc thermique (l'agrégation des protéines).

## 7.2 Réponse à une carence en acides aminés

La réponse stringente est le nom donné à la réponse au stress qui apparait lors d'une carence en acides aminés. Lorsque celle-ci survient, une molécule le (p)ppGpp est produite en grande quantité. Cette molécule inhibe, en partie, la transcription et décroit ainsi la synthèse de la machinerie de traduction (ARNt, ARNr) dont la fonction (la synthèse de protéines) est inutile sans acide aminés (les briques de base des protéines). Cet « alarmone » (le (p)ppGpp) réoriente ainsi l'utilisation des ressources énergétiques vers d'autres processus prioritaires comme la biosynthèse des acides aminés (Chatterji & Ojha, 2001).

### 7.3 Réponse à une exposition UV

L'exposition de *E. coli* à la lumière ultra-violette endommage fortement l'ADN ce qui conduit à l'accumulation de simple brin d'ADN. Pour répondre à ce stress, la bactérie possède un système dit SOS, découvert par Miroslav Radman, qui permet à la bactérie de réparer, en partie, son ADN (Radman, 1974). Lors d'une exposition aux UV, la protéine RecA se fixe aux simples brins d'ADN apparus et induit le clivage d'une autre enzyme nommé LexA. La disparition de LexA induit la transcription d'une quarantaine de gènes dont la plupart sont impliqués dans la protection, la réparation et la réplication de l'ADN (Janion, 2008).

## 8 Motilité, agrégation, biofilm et quorum sensing

La bactérie *E. coli* n'est pas condamnée à rester immobile. Pour assurer sa mobilité, *E. coli* dispose d'une dizaine de flagelles qui tournent comme une hélice de bateau ce qui assure la propulsion. Un flagelle est une structure cylindrique creuse et semi-rigide composé d'une protéine, la flagelline, formant un filament d'une dizaine de micromètres de longueur pour 20nm d'épaisseur. Le filament est ancré dans la membrane plasmique au niveau d'une embase où se trouve le moteur moléculaire. Ce dernier tire son énergie du flux de protons qui traversent la membrane dû au gradient de concentration existant de part et d'autre. Ce moteur assure la rotation du flagelle à plus de 1000 révolutions par seconde ce qui permet à la bactérie d'atteindre une vitesse allant jusqu'à 30µm/s. La motilité des bactéries est, par exemple, très utile pour se déplacer en direction d'une source de nutriments ou au contraire pour s'éloigner d'une substance toxique. Nous reviendrons en détail sur ce phénomène appelé chimiotactisme.

La bactérie *E. coli* n'est pas condamnée à subir les flux de liquides. En effet, les bactéries *E. coli* sont capables d'adhérer à la plupart des surfaces (PVC, polycarbonate, verre...) et de s'agréger entre elles. Pour adhérer, les bactéries utilisent des appendices protéiques plus petits et plus fins que les flagelles : les fimbriae présents en plusieurs centaines d'exemplaires sur chaque cellule. On utilise le terme « biofilm » lorsque on parle d'une communauté de micro-organismes adhérant entre eux et à une surface, et marquée par la sécrétion d'une matrice adhésive et protectrice. Les micro-organismes présents dans la communauté sont souvent capables d'agir de façon coordonnée. Il existe ainsi des mécanismes régulateurs chargés de synchroniser l'expression de gènes particuliers au sein d'une population. Ce mécanisme dit de « quorum sensing » utilisent la capacité des bactéries à communiquer avec leurs congénères via des signaux moléculaires. *E. coli* produit une molécule le 4,5-dihydroxy-2,3-pentanedione dit « auto-inducteur ». Lorsque la densité de population est faible, la diffusion réduit rapidement la concentration de l'auto-

inducteur dans le milieu. A l'inverse, lorsque cette densité augmente, par exemple dans un milieu confiné comme un biofilm, la concentration de l'auto-inducteur dépasse un seuil critique qui est alors perçu par chaque bactérie en même temps. Cela permet de synchroniser l'expression génique pour l'ensemble de la population qui se comporte alors un peu à la manière d'un organisme pluricellulaire et non plus comme une simple somme d'individus indépendants.

Une des publications présente dans cette thèse décrit le fonctionnement d'un système synthétique de communication intercellulaire basé sur une petite molécule : l'AMPc.

# 9 La flore intestinale

La flore intestinale est l'écosystème principal de la bactérie *E. coli*. Elle contient $10^{14}$ bactéries c'est-à-dire dix fois plus que le nombre de cellules humaines dans le corps. On pense qu'il y a environ 1000 espèces qui peuplent nos intestins. *E. coli* représenterait moins de 1 bactérie sur 100 : en effet plus de 99% de nos bactéries seraient des bactéries anaérobies strictes (Firmicutes et les Bacteroidetes). Il est ainsi très difficile d'étudier cet écosystème car la majorité des espèces bactériennes ne peuvent pas être cultivées en présence d'oxygène. Reproduire l'écosystème d'un colon n'est donc pas chose aisée. Et il est encore plus difficile d'étudier les relations qu'entretiennent les différentes espèces entre elles ou avec leur hôte. Ces associations peuvent être répertoriées en commensalisme, mutualisme ou parasitisme. Nous décrivons ci-dessous chacune d'elles.

Le commensalisme est un type d'interaction entre deux êtres vivants dans laquelle l'hôte fournit une partie de sa propre nourriture au commensal sans obtenir de contrepartie évidente de ce dernier. On pense qu'un certains nombre d'espèces bactériennes peuplant nos intestins entretiennent ce type de relation dont le bénéfice n'est pas réciproque.

Cependant, il existe aussi des interactions entre deux ou plusieurs espèces, dans laquelle le symbiote et l'hôte tirent tous les deux profit de cette relation. On qualifie ce type d'association de mutualisme. Elle expliquerait la cohabitation de plus de 1000 espèces différentes dans nos intestins. Selon un certain paradigme, on peut envisager ces différentes espèces comme les maillons d'un métabolisme extracellulaire (comme les enzymes sont les maillons du métabolisme intracellulaire) (Lehours, 2010). Cette flore intestinale, composée de plus de 3.3 millions de gènes (Qin *et al.*, 2010) (150 fois plus que le génome humain) serait, selon certains chercheurs, un organe « oublié » qui aurait de nombreuses fonctions utiles pour les humains. En effet, bien qu'on puisse vivre sans flore intestinale (la gnotobiologie), les microorganismes accomplissent de nombreuses fonctions utiles comme :

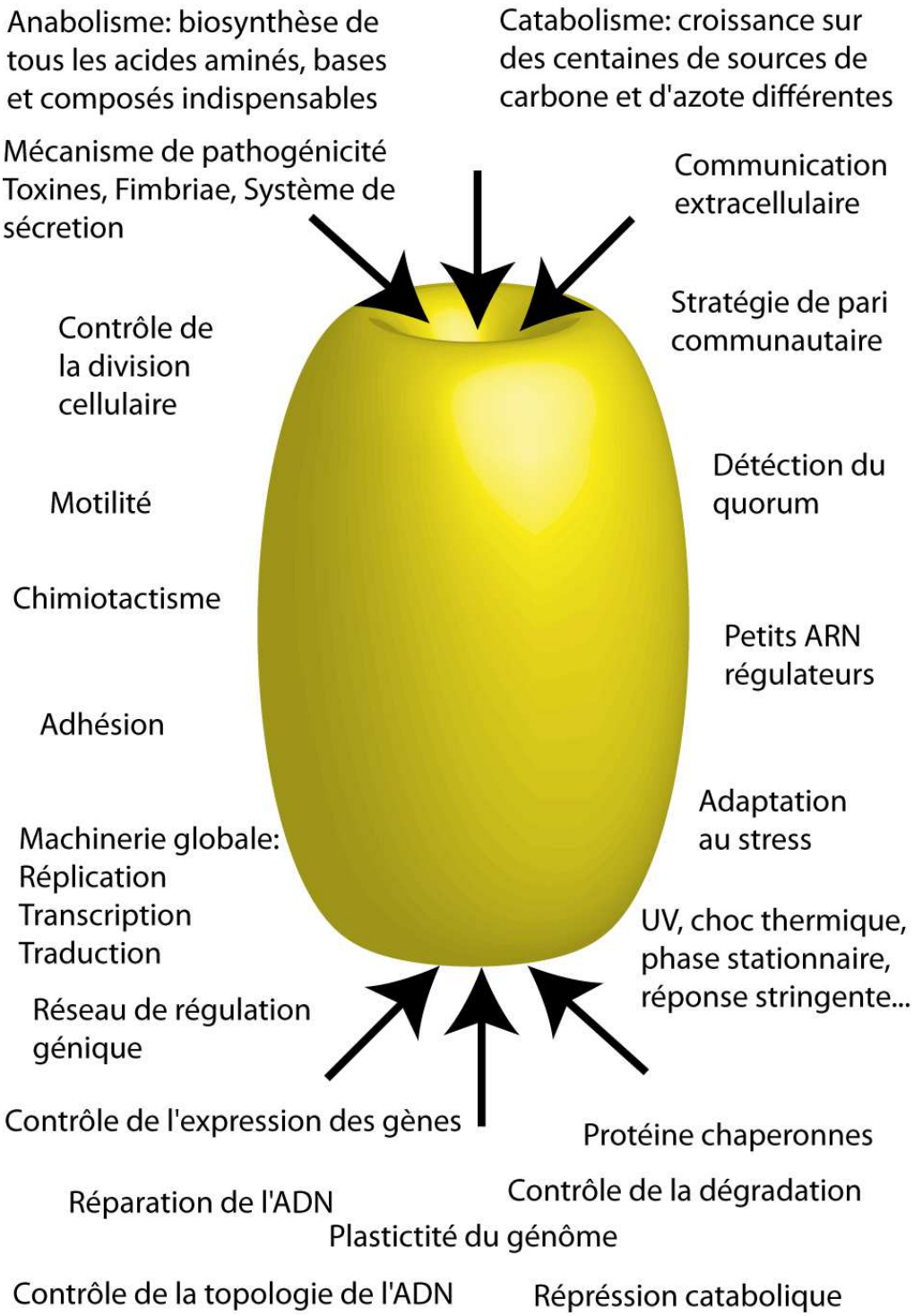



**Figure II.5** – Nuage de tag de quelques grands domaines d'étude chez *E. coli*. J'ai tenté d'être le plus exhaustif possible même si j'oublie forcement beaucoup de domaines. Ce nuage a pour but d'aider un non-biologiste à identifier rapidement ce que *E. coli* peut faire et ne pas faire.

⇒ la fermentation de résidus non digestes
⇒ la production de vitamines (*E. coli* produit la vitamine k2)
⇒ le contrôle de la prolifération des cellules épithéliales et leur différentiation
⇒ le développement et l'entrainement du système immunitaire
⇒ la protection contre les pathogènes (l'effet barrière)

On sait également que la composition et l'abondance relative des espèces bactériennes qui colonisent l'intestin varient d'un individu à l'autre. En effet, une étude a révélé que la flore intestinale de personnes obèses aurait une capacité accrue à extraire l'énergie de l'alimentation (Turnbaugh *et al.*, 2006).

Le parasitisme est une relation biologique où un des protagonistes, le parasite, tire profit (en se nourrissant, en s'abritant ou en se reproduisant) d'une espèce tiers, l'hôte, aux dépens de celui-ci. Certaines souches pathogènes d'*E. coli* sont responsables de gastro-entérites, d'infections urinaires, de méningites et de septicémies. La recherche d'un vaccin continue (Moriel *et al.*, 2010). Je décris dans le paragraphe qui suit les mécanismes de virulence de 4 types de souches.

## 10 Les souches d'*E. coli* pathogènes

Les *E. coli* entérotoxigéniques (ETEC) sont responsables de plus de 380 000 morts/an (source O.M.S). Les décès touchent principalement des enfants de moins de 3 ans qui meurent de déshydratation suite à des diarrhées très sévères (Steffen *et al.*, 2005). Ces décès ne sont pas dus à l'absence de traitements existants mais plutôt à l'absence de structures médicales en mesure de soigner les jeunes patients dans les pays touchés. Le pouvoir pathogène des ETEC est très similaire à celui du cholera. Il s'explique essentiellement par la présence de fimbriaes (qui permettent l'adhésion aux cellules intestinales) et la sécrétion des toxines. Une des toxines, après endocytose, entraine une forte augmentation de la concentration d'une molécule (l'AMPc) ce qui conduit à la phosphorylation d'un transporteur membranaire nommé CFTR. Cette action se traduit par une sécrétion d'ions chlorure ce qui provoque la diffusion osmotique d'eau vers la lumière intestinale puis la diarrhée et potentiellement la mort.

Les *E. coli* entéroinvasives (EIEC) sont responsables de diarrhées aqueuses qui évoluent rapidement en dysenterie (selles contenant du sang et du mucus). On pense que les mécanismes d'invasion des EIEC sont très similaires à ceux des bactéries Shigella. Elles expriment au niveau de leur membrane, un système de sécrétion dit de type 3 qui agit comme une seringue et injecte des protéines dites « invasines ». Ces dernières permettront l'invasion puis la multiplication

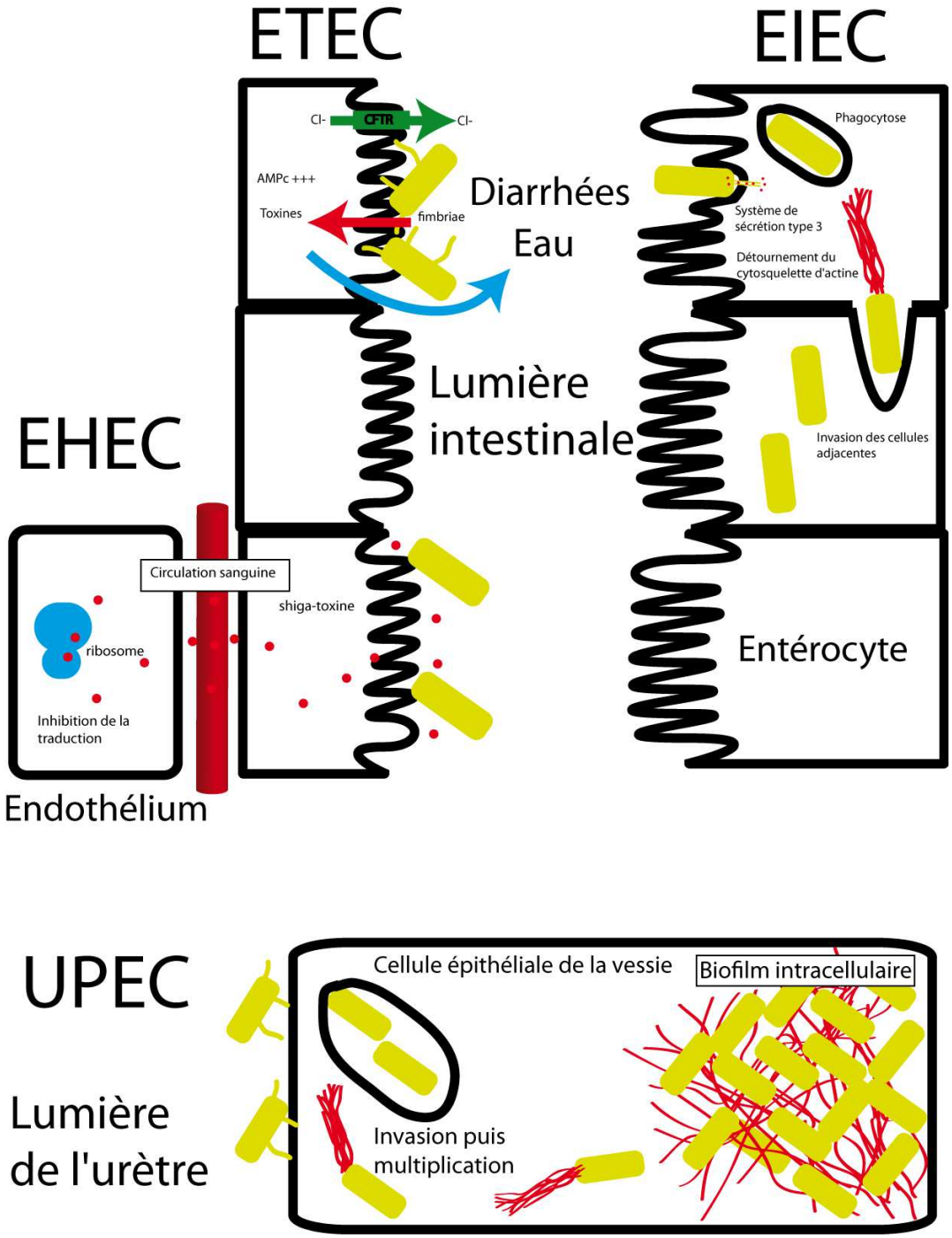



**Figure II.6** – Les souches pathogènes d'*E. coli*. La bactérie *E. coli* peut se développer en intracellulaire. Elle peut détourner le cytosquelette d'actine de la cellule hôte pour se déplacer. Certains mécanismes de pathogénicité ressemblent à ceux du cholera (ETEC) : ils entrainent une forte déshydratation. Les EIEC injectent des toxines dans l'enterocyte en utilisant une seringue (système de sécrétion de type 3). Les EHEC produisent une toxine qui passe dans le sang puis inhibe la traduction des cellules de l'endothélium.

des bactéries dans le cytoplasme des cellules épithéliales du colon. Les bactéries détournent ensuite la polymérisation de l'actine pour pouvoir se propulser et se disséminer de cellules en cellules. L'invasion puis la destruction de la muqueuse intestinale déclenchent une intense réaction inflammatoire.

Les *E. coli* entérohémorragiques (EHEC) sont responsables de colites hémorragiques et parfois de défaillance rénale. Des épidémies à EHEC se sont déclarées suite à l'ingestion de viandes contaminées et insuffisamment cuites (hamburger). Les EHEC produisent une toxine (la Shiga-toxine) qui peut être transférée dans la circulation systémique puis qui peut inactiver la sous unité 60S des ribosomes ce qui bloque la traduction et entraine la mort des cellules de l'endothélium. Une des conséquences possibles est le syndrome hémolytique et urémique (SHU) pouvant conduire à la mort.

Les *E. coli* uropathogènes (UPEC) sont responsables de 90% des infections urinaires. Ces bactéries utilisent les fimbriaes pour se lier aux cellules endothéliales et coloniser ensuite la vessie. Les hémolysines produites entrainent la lyse des cellules des voies urinaires. Les UPEC forment des communautés bactériennes intracellulaires dans les cellules de l'urothélium ce qui leur permet d'échapper au système immunitaire. Elles peuvent également former des biofilms intra et extracellulaires ce qui leur confère une meilleure résistance aux antibiotiques. Ces deux phénomènes expliqueraient la chronicité souvent observée dans les infections des voies urinaires (Blango & Mulvey, 2010; Cegelski *et al.*, 2008).

# Références bibliographiques


Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., & Ishihama, A. 1999. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol*, **181**(20), 6361–70. 30

Alon, Uri. 2006. *An introduction to system biology*. Chapman and Hall/CRC. 31

Arsene, F., Tomoyasu, T., & Bukau, B. 2000. The heat shock response of Escherichia coli. *Int J Food Microbiol*, **55**(1-3), 3–9. 38

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants : the Keio collection. *Mol Syst Biol*, **2**, 2006 0008. 30

Ben-Zvi, A. P., & Goloubinoff, P. 2001. Review : mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J Struct Biol*, **135**(2), 84–93. 32

Blango, M. G., & Mulvey, M. A. 2010. Persistence of uropathogenic Escherichia coli in the face of multiple antibiotics. *Antimicrob Agents Chemother*, **54**(5), 1855–63. 44

Cegelski, L., Marshall, G. R., Eldridge, G. R., & Hultgren, S. J. 2008. The biology and future prospects of antivirulence therapies. *Nat Rev Microbiol*, **6**(1), 17–27. 44

Chatterji, D., & Ojha, A. K. 2001. Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol*, **4**(2), 160–5. 38

Farrell, C. M., Grossman, A. D., & Sauer, R. T. 2005. Cytoplasmic degradation of ssrA-tagged proteins. *Mol Microbiol*, **57**(6), 1750–61. 32

Janion, C. 2008. Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. *Int J Biol Sci*, **4**(6), 338–44. 39

Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., & Karp, P. D. 2009. EcoCyc : a comprehensive view of Escherichia coli biology. *Nucleic Acids Res*, **37**(Database issue), D464–70. 32, 34, 35, 36

Lehours, A-C. 2010. les bacteries du lac pavin. *Pour la science*. 40

Margolin, W. 2001. Spatial regulation of cytokinesis in bacteria. *Curr Opin Microbiol*, **4**(6), 647–52. 36

Merlin, C., McAteer, S., & Masters, M. 2002. Tools for characterization of Escherichia coli genes of unknown function. *J Bacteriol*, **184**(16), 4573–81. 30

Moriel, D. G., Bertoldi, I., Spagnuolo, A., Marchi, S., Rosini, R., Nesta, B., Pastorello, I., Corea, V. A., Torricelli, G., Cartocci, E., Savino, S., Scarselli, M., Dobrindt, U., Hacker, J., Tettelin, H., Tallon, L. J., Sullivan, S., Wieler, L. H., Ewers, C., Pickard, D., Dougan, G., Fontana, M. R., Rappuoli, R., Pizza, M., & Serino, L. 2010. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc Natl Acad Sci U S A*. 42

Nollmann, M., Crisona, N. J., & Arimondo, P. B. 2007. Thirty years of Escherichia coli DNA gyrase : from in vivo function to single-molecule mechanism. *Biochimie*, **89**(4), 490–9. 30

Perrenoud, A., & Sauer, U. 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. *J Bacteriol*, **187**(9), 3171–9. 32

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., & Wang, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**(7285), 59–65. 40

Radman, M. 1974. Phenomenology of an inducible mutagenic DNA repair pathway in Escherichia coli : SOS repair hypothesis. *Molecular and Environmental aspects of mutagenesis*. 39

Steffen, R., Castelli, F., Dieter Nothdurft, H., Rombo, L., & Jane Zuckerman, N. 2005. Vaccination against enterotoxigenic Escherichia coli, a cause of travelers' diarrhea. *J Travel Med*, **12**(2), 102–7. 42

Sundararaj, S., Guo, A., Habibi-Nazhad, B., Rouani, M., Stothard, P., Ellison, M., & Wishart, D. S. 2004. The CyberCell Database (CCDB) : a comprehensive, self-updating, relational

database to coordinate and facilitate in silico modeling of Escherichia coli. *Nucleic Acids Res*, **32**(Database issue), D293–5. 31

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, **444**(7122), 1027–31. 42

# Chapitre III

# Interaction et réseau de régulation

## 1 Préambule

Ce chapitre est dans la lignée du deuxième chapitre sur *E. coli*. Il a pour objectif de montrer aux non-biologistes comment les biologistes perçoivent une interaction ou un réseau. Le chapitre d'après visera le but opposé : montrer aux biologistes comment les mathématiciens (et/ou informaticiens) décrivent une interaction ou un réseau. Dans ce chapitre, j'ai volontairement fait l'effort d'aller relativement en profondeur dans certains paragraphes de manière à permettre aux biologistes d'y extraire une information plus concrète/spécifique. Il y a donc parfois deux niveaux de lecture et certains paragraphes sembleront un peu lourds à un lecteur avide de généralités ou de clarté. Selon moi, la structure/le plan et le message global de ce chapitre comptent plus que les détails. Si ces derniers vous gênent : n'ayez pas peur de lire en diagonale.

## 2 Interaction et réseau de régulation

Dans le chapitre précédent, nous avons dressé un portrait statique du fonctionnement de la bactérie *E. coli*. Nous allons maintenant décrire ce qui permet de « dynamiser » la bactérie. Cette dernière n'est pas un simple sac de molécules indépendantes. Les bactéries contrôlent et ajustent en permanence le niveau d'expression de chacun de leurs gènes pour adapter le contenu cellulaire aux conditions extérieures. Ce contrôle de l'expression des gènes est effectué grâce à un réseau de régulation génique. Un réseau de régulation génique est une collection de gènes qui interagissent les uns avec les autres via leurs produits d'expression (les protéines), permettant un contrôle mutuel de leurs taux d'expression. Les réseaux de régulation génique sont, avec la régulation du métabolisme, les principaux déterminants de l'adaptation d'une bactérie à son environnement.

Un réseau peut induire une fonction que ses sous parties ne possèdent pas : le fameux truisme

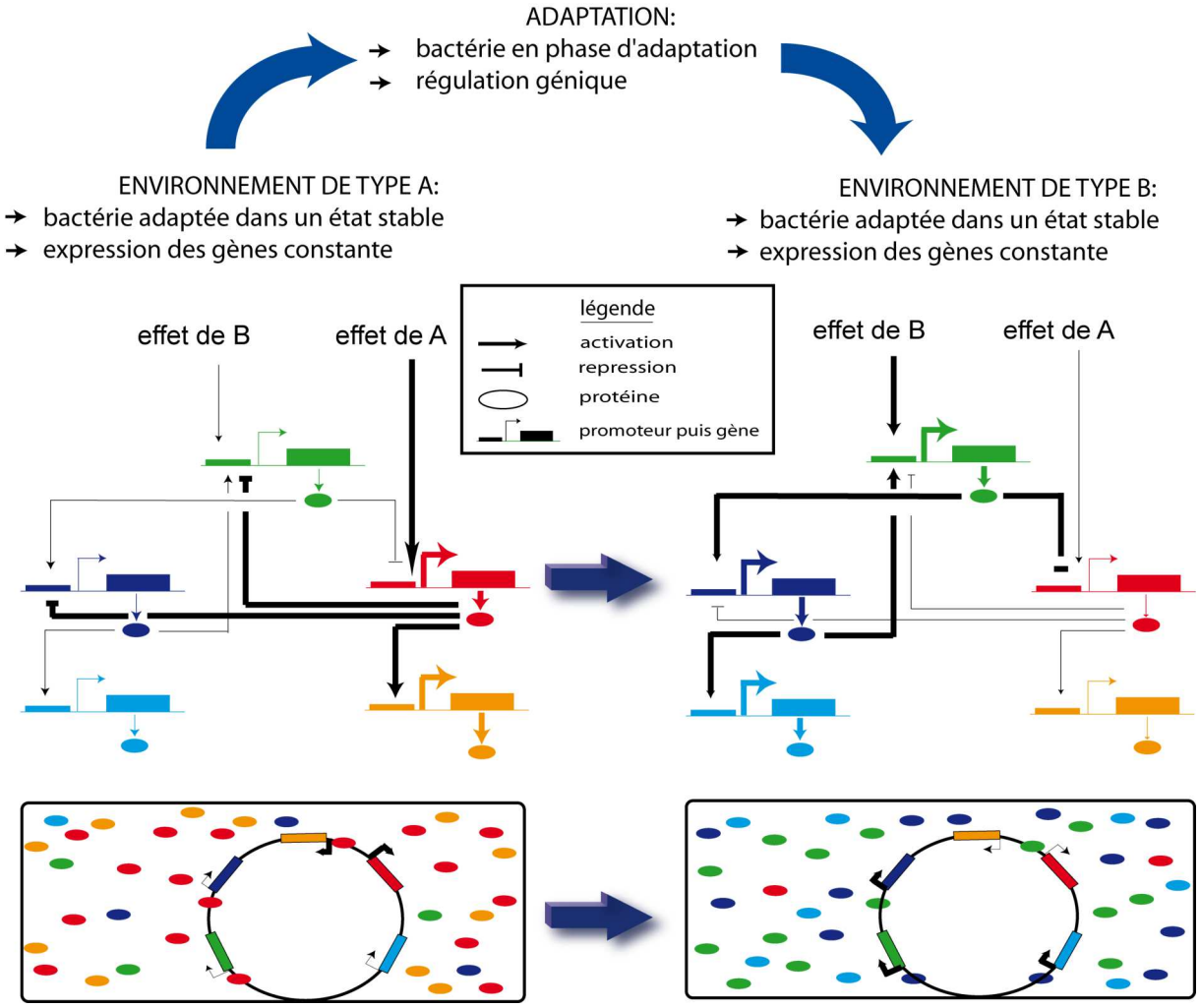



**Figure III.1** – Réseau, interactions et adaptation. Une bactérie est adaptée à un environnement A. Si l'environnement change et devient B, la bactérie détecte ce changement et adapte l'expression de ses gènes à ce nouvel environnement. L'adaptation est permise par la présence d'un réseau de gènes contrôlant l'expression d'autres gènes.

*« le tout est plus que la somme des parties »*. On qualifie ce phénomène d'émergence. La vie (d'une bactérie) émerge de l'ensemble des interactions des molécules internes à la bactérie. Ces interactions forment le réseau de régulation. Notons que la notion d'émergence fait l'objet d'une polémique en philosophie des sciences. Loin d'être un fait établi par la méthode scientifique, l'émergence ne serait qu'un simple mot utile pour décrire des phénomènes complexes que l'on ne comprend pas.

Un réseau est un ensemble de nœuds reliés entre eux par des liens (ou des flèches dans le cas d'un graphe orienté). Dans un réseau de régulation, les nœuds sont des molécules et les liens représentent une interaction physique entre deux molécules. Contrairement à un circuit

intégré, au réseau neuronal ou au réseau internet qui possèdent des liens bien réels, les flèches d'un réseau de régulation sont virtuelles : elles reposent sur les lois de la diffusion. Ces lois introduisent des propriétés probabilistes au réseau de régulation. Un composé A qui baigne dans le cytoplasme d'une bactérie se déplace de manière aléatoire (le mouvement brownien) et a donc une certaine probabilité de rencontrer un composé B en un temps t. Si la rencontre entre les composés A et B induit un changement/effet/modification/conséquence, alors il existe un lien entre A et B.

⇒ Si l'enzyme A (par exemple la phosphoglucose isomerase) rencontre le métabolite B (le glucose-6-phosphate), le changement correspond à la disparition du substrat (le glucose-6-phosphate) et l'apparition du produit (le fructose-6-phosphate). On parle, pour ce type d'interaction, de réaction métabolique.

⇒ Si l'enzyme A (par exemple UhpB) rencontre une protéine B (UhpA), le changement correspond à la modification chimique de la protéine B (la phosphorylation d'UhpA). On parle, pour ce type d'interaction, de « transduction du signal » ou de « modification post-traductionnelle ».

⇒ Si le facteur de transcription A (par exemple LacI) rencontre la séquence promotrice du gène B (le gène *lacZ*), le changement correspond à une modification de l'expression du gène B (répression du gène *lacZ*). On parle, pour ce type d'interaction, de régulation génique/ régulation de l'expression génique /régulation trancriptionnelle.

Cette liste n'est pas exhaustive et on compte d'autres types d'interactions comme, par exemple, les interactions mettant en jeu des ARN ou des interactions intercellulaires. On compte actuellement plus de 1326 réactions métaboliques chez *E. coli*. Ces voies métaboliques sont en général assez linéaires (glycolyse, cycle de Krebs) et ressemblent assez peu à un réseau : la connectivité est faible : Il y a peu de boucles de rétroaction. Le temps d'une réaction enzymatique est inferieur à la seconde. L'ajustement des pools de métabolites et des flux est donc très rapide et se fait également à l'échelle de la seconde.

La transduction du signal est assez peu décrite chez *E. coli*. On compte cependant une trentaine de systèmes de phosphorylation dits à deux composantes. Il existe aussi un système de transfert de phosphate (PTS) chargé de faire rentrer le glucose dans la cellule. La transduction du signal semble aussi être assez linéaire. Ces modifications chimiques, parfois sous forme de cascade, sont très rapides et s'effectuent, comme les réactions métaboliques, à l'échelle de la seconde.

On compte pour l'instant plus de 4792 régulations transcriptionnelles. Ces régulations forment un réseau avec de nombreuses boucles de rétroactions. Il est ainsi difficile de savoir

intuitivement comment évolue un tel système. Ces régulations se font plutôt à l'échelle des minutes, voire des heures. En effet, si la fixation du facteur de transcription au promoteur prend quelques secondes, l'activation ou la répression de la transcription du gène cible ne montrera ses effets que lorsque la protéine cible résultante se sera accumulée suffisamment ou lorsqu'elle se sera dégradée suffisamment.

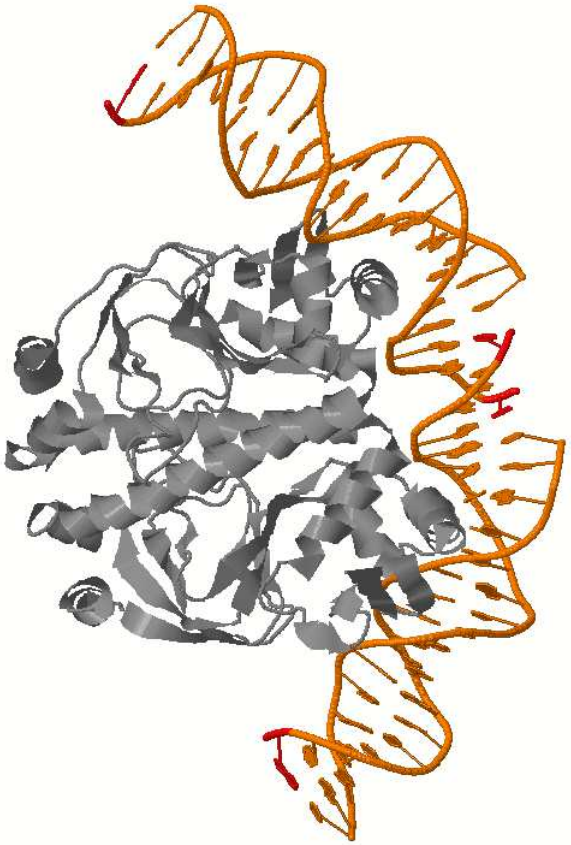


**Figure III.2** – L'interaction ADN–protéine. Vidéo montrant l'interaction génique entre Crp et l'ADN. L'import de la structure tridimensionnelle de Crp et de l'ADN est realisée directement depuis Matlab en se connectant à la base de données Protein Data Bank (PDB). Voir le code Matlab en annexe Code01ImporteStructureProt.m qui gènère cette video.

Dans cette thèse, je parlerai de réseaux de régulation globaux lorsque j'intégrerai tous les mécanismes de régulation présents dans une cellule pour lui permettre de s'adapter. A l'inverse, je parlerai de réseau de régulation géniques lorsque nous étudierons des réseaux construits à partir de régulations transcriptionnelles uniquement. La littérature s'intéresse principalement aux réseaux de régulation génique car c'est dans ce type de réseaux que l'on retrouve le plus de systèmes non linéaires. Cependant cette dichotomie est parfaitement illusoire tant métabolisme, régulations transcriptionnelles, transduction du signal et autres régulations (communi-

cation intercellulaire, ARN...) fonctionnent ensemble et s'influencent mutuellement. Le réseau de régulation global de la bactérie *E. coli* est un réseau extrêmement complexe dans sa topologie avec plusieurs dizaines de milliers/millions/milliards ( ?) d'interactions et de nombreuses boucles de rétroaction. On est encore loin d'avoir découvert toute la pluralité et la diversité des mécanismes de régulation. Ces mécanismes agissent de concert à des échelles de temps très différentes. Enfin, la nature intrinsèquement stochastique des mécanismes de régulation (dû au mouvement brownien) font qu'un réseau de régulation est idiosyncratique : j'entends par là le fait que la réponse d'un même réseau à une perturbation donnée peut être différente entre deux clones. Autrement dit, pour un même génotype, deux bactéries peuvent adopter deux phénotypes différents en réponse à une même perturbation. Nous reviendrons largement sur cette notion importante dans le chapitre sur le bruit.

# 3 Deux exemples clés

Je vais décrire le fonctionnement de deux réseaux de régulation parmi les plus connus et donc les mieux décrits à ce jour. Ces deux réseaux fonctionnent à des échelles de temps très différentes : le premier, la répression catabolique, fonctionne à l'échelle de la dizaine de minutes alors que le deuxième, le chimiotactisme fonctionne à l'échelle de la seconde. Pour ces deux exemples, je suis volontairement rentré dans les détails. Ce niveau de détail enchantera celui qui veut comprendre ces systèmes en profondeur c'est-à-dire celui qui s'apprête probablement à les utiliser. Cependant, si vous souhaitez uniquement avoir une idée générale de la manière dont les biologistes décrivent ces réseaux, alors contentez vous de lire les paragraphes qui suivent de manière superficielle. Essayez juste de capter la « philosophie » sous-jacente en vous concentrant sur le champ lexical utilisé pour décrire les interactions.

## 3.1 La répression catabolique

Pourquoi la bactérie *E. coli* a t'elle besoin d'un réseau de régulation pour s'adapter ? Pourquoi, par exemple, ne pas produire chacune des 4500 protéines en 100 exemplaires pour être sûr de répondre à n'importe quel imprévu ? Par exemple, être prêt à manger immédiatement n'importe quel sucre qui se présenterait dans le milieu extérieur ?

Maintenir des protéines inutiles dans la cellule à un coût. Ce dernier impacte directement le fitness de la bactérie, c'est-à-dire sa capacité à se reproduire. Une bactérie, qui au fil de l'évolution, acquiert la capacité de ne synthétiser une protéine que lorsqu'elle en a vraiment besoin, augmentera son fitness et sera ainsi sélectionnée par l'évolution.

Je vais maintenant décrire le premier mécanisme de régulation génique découvert : la régulation de l'opéron *lac*. Dans son environnement naturel, l'opéron *lac* code pour des protéines nécessaires à la digestion du lactose. La cellule peut utiliser le lactose comme source d'énergie en produisant une enzyme la $\beta$-galactosidase (codée par le gène *lacZ*). Cependant, il est inutile de produire cette enzyme s'il n'y a pas de lactose dans le milieu ou si une source de carbone plus facilement métabolisable comme le glucose est présente dans le milieu. L'expression de l'opéron *lac* est contrôlée par plusieurs mécanismes qui assurent que les dépenses d'énergie pour produire la $\beta$-galactosidase n'ont lieu que lorsque c'est nécessaire. Ces mécanismes de régulation ont pour conséquence l'utilisation séquentielle du glucose puis du lactose en deux phases distinctes de croissances. On appelle cela une diauxie (Monod, 1942).

Le premier mécanisme de contrôle est la réponse régulatrice due à la présence de lactose dans le milieu. Ce mécanisme utilise une protéine régulatrice intracellulaire (un facteur de transcription) nommée répresseur LacI qui bloque la transcription du gène *lacZ* et donc la production de $\beta$-galactosidase. Si le lactose est absent du milieu de culture, ce répresseur se fixe fermement à une courte séquence d'ADN appelée l'operateur lacO juste en amont de l'origine du gène *lacZ*. La fixation du répresseur à son site operateur interfère avec la fixation de l'ARN polymérase au promoteur et empêche ainsi la transcription de l'opéron. Quand la cellule pousse en présence de lactose, un métabolite issu du lactose appelé allolactose se fixe au répresseur LacI induisant un changement de sa conformation. Ainsi altéré, le répresseur est incapable de se fixer à l'operateur lacO et permet à l'ARN polymérase de transcrire le gène *lacZ*. Ainsi la concentration de $\beta$-galactosidase augmente fortement et le lactose peut être métabolisé.

Le second mécanisme de contrôle est la réponse régulatrice due à l'absence de glucose dans le milieu. Ce mécanisme utilise le régulateur global Crp pour augmenter la production de $\beta$-galactosidase en absence de glucose. L'AMPc est une molécule dont la concentration dans la cellule est inversement proportionnelle à celle du glucose. Cette molécule fixe Crp formant ainsi le complexe Crp–AMPc (Kolb *et al.*, 1993). Ce complexe peut alors se fixer au site de liaison de Crp situé dans la région promotrice en amont de l'origine du gène *lacZ* puis aider au recrutement de l'ARN polymérase. En absence de glucose, la production d'AMPc est forte et le complexe Crp–AMPc se fixe à l'ADN augmentant très fortement la production de $\beta$-galactosidase. A l'inverse, en présence de glucose, la concentration d'AMPc est faible, la formation du complexe n'a pas lieu et Crp ne peut pas se lier à l'ADN diminuant drastiquement la transcription de l'opéron *lac* et donc la production de $\beta$-galactosidase. Mais comment le glucose régule-t'il la concentration d'AMPc et donc la transcription de l'opéron *lac* ? L'import de glucose est medié par le système de transfert de phosphate (PTS) (Postma *et al.*, 1993). Ce système transfère séquentiellement un groupement phosphate du Phosphoenolpyruvate (PEP) jusqu'au glucose

formant ainsi du glucose-6-phosphate. La séquence de transfert des phosphates fait intervenir 4 enzymes dont l'enzyme $EIIA_{glc}$. Lors de l'import du glucose, cette enzyme est sous forme déphosphorylée or seule la forme phosphorylée est capable de fixer puis d'activer l'adenylate cyclase, l'enzyme responsable de la biosynthèse de l'AMPc.

Le dernier mécanisme de contrôle est appelé mécanisme de l'exclusion de l'inducteur (Saier & Crasnier, 1996). Lorsque le glucose est importé, l'enzyme $EIIA_{glc}$ est sous forme déphosphorylée. Sous cette forme, l'enzyme est capable de se fixer au transporteur du lactose et d'inactiver ainsi son transport. L'entrée de l'inducteur est ainsi empêchée. Or ce dernier est chargé d'inactiver le répresseur LacI. En conséquence, en présence de glucose, le répresseur reste sous sa forme active et réprime la transcription de l'opéron *lac*.

J'ai choisi de développer cet exemple de la répression catabolique pour plusieurs raisons. Premièrement car le haut niveau de détail dans la compréhension des mécanismes rend son étude intéressante pour les modélisateurs (Kremling *et al.*, 2009). Deuxièmement, car il illustre parfaitement les liens très étroits qui existent entre régulation transcriptionnelle, transduction du signal et métabolisme. Troisièmement car une des publications présente dans cette thèse remet en cause le modèle actuel de répression catabolique que je viens de vous décrire.

## 3.2 Le chimiotactisme

Le chimiotactisme est défini comme un mouvement orienté vers ou dans le sens opposé d'une substance chimique. Les bactéries *E. coli*, capables de se mouvoir grâce à leurs flagelles, sont attirées en direction de certains nutriments (sucres, acide aminés) ou à l'inverse repoussées par des substances toxiques (alcool, acide gras).

Pour réussir à modifier leur direction vers une substance attractive, les bactéries doivent forcement mesurer le gradient de concentration. Cependant la différence de concentration d'une substance attractive entre l'avant et l'arrière de la cellule est trop faible (contrairement aux cellules eucaryotes) pour être détecté. Or, si la mesure du gradient n'est pas spatiale alors c'est qu'elle est temporelle. Autrement dit, cela signifie que les bactéries sont capables de « mémoriser » les concentrations. Petit bémol important que je préfère ajouter quand j'utilise des analogies avec le système nerveux : la bactérie ne « choisit » pas son chemin et n'a pas « conscience » de se diriger vers un nutriment.

La motilité des bactéries est permise par la présence des flagelles qui peuvent tourner dans deux sens : La rotation dans le sens inverse des aiguilles d'une montre arrange les flagelles en un seul faisceau synchrone ; il en résulte la natation rectiligne de la bactérie. On appellera cette phase la course. Au contraire, lorsque la rotation se fait dans le sens des aiguilles d'une montre,

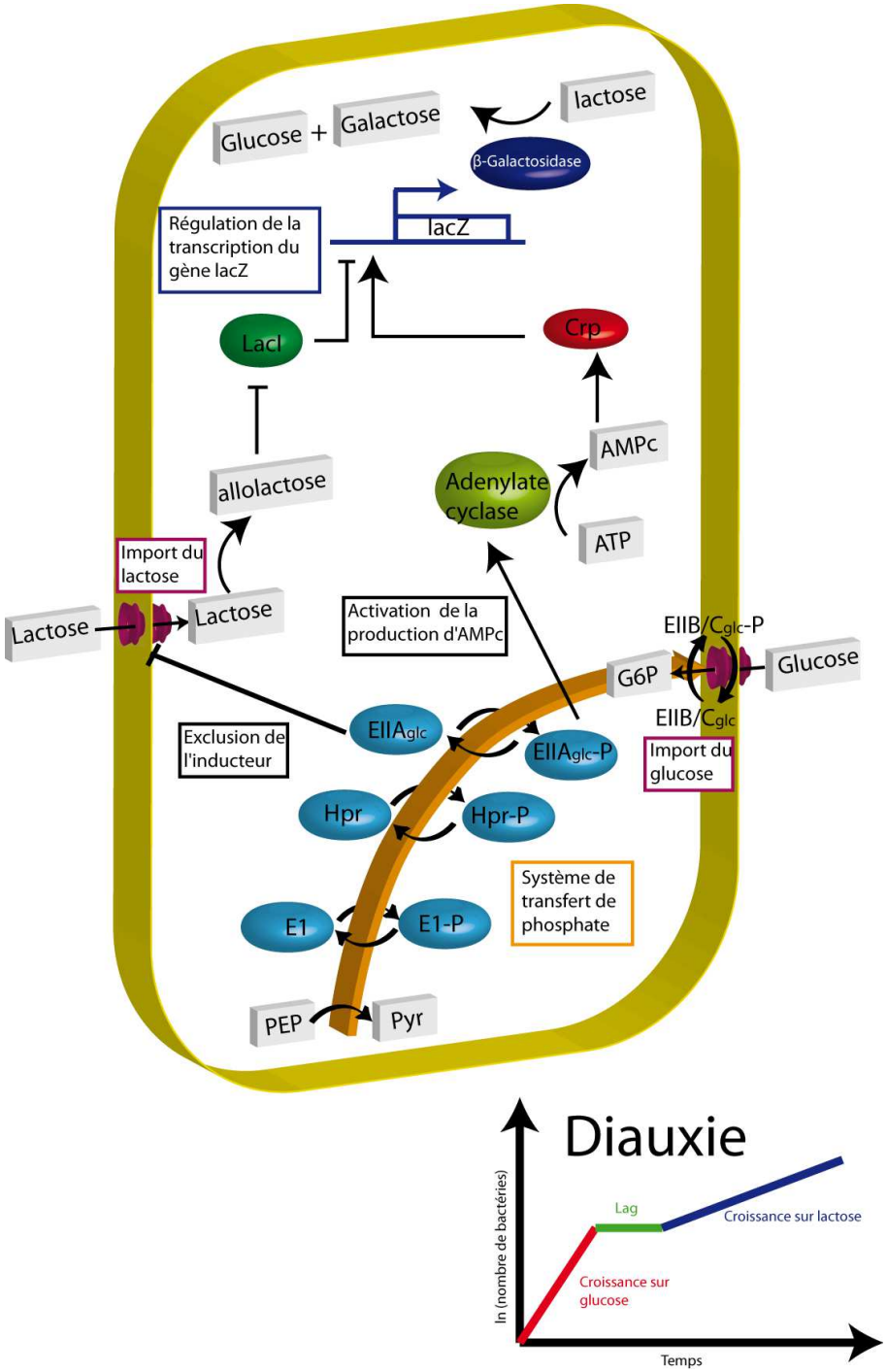



**Figure III.3** – La répression catabolique. La $\beta$-galactosidase n'est produite que si le lactose est présent dans le milieu et le glucose absent. Si le milieu contient du glucose et du lactose, la bactérie consomme déjà le glucose (croissance exponentielle) puis adapte son expression génique pour synthétiser la $\beta$-galactosidase (arrêt de la croissance dit « lag ») avant de reprendre la croissance en consommant le lactose : on parle de diauxie.

les flagelles se désolidarisent les uns des autres, ainsi la bactérie bouge de manière erratique et tourne dans tous les sens sans vraiment avancer. On appellera cette phase la culbute.

Dans un environnement chimique isotrope, *E. coli* se déplace en « marche aléatoire » c'est-à-dire en alternant des épisodes de course et des épisodes de culbute. Lorsque la bactérie avance dans la bonne direction, la concentration de la substance attractive augmente. La bactérie détecte cette augmentation et diminue la probabilité de culbute. Les épisodes de course dans la bonne direction sont alors plus longs et cette marche aléatoire biaisée permet aux bactéries de migrer vers la substance attractive.

Le gradient chimique est perçu à l'aide d'un chémorécepteur transmembranaire appartenant à la famille des « methyl accepting chemotaxis proteins (MCPs) ». Lorsque une substance attractive se fixe au récepteur, un signal est transmis à la partie cytoplasmique du récepteur où les protéines CheW et CheA sont fixées. Ce signal inhibe l'autophosphorylation de l'histidine kinase CheA. En absence de substance attractive, le groupe phosphoryl, résultant de l'autophosphorylation de CheA, est rapidement transféré aux régulateurs de réponse, CheB et CheY. CheY diffuse alors vers le moteur du flagelle et interagit avec la protéine FliM ce qui induit le changement du sens de rotation des flagelles : du sens inverse des aiguilles d'une montre au sens des aiguilles d'une montre. Ainsi l'absence de substance attractive induit la « culbute ». A l'inverse, sa présence inhibe l'autophosphorylation de cheA, diminue la probabilité de « culbute » et permet ainsi une « course » plus longue.

La bactérie garde une sensibilité pour certaines substances attractives du nanomolaire jusqu'au millimolaire. En effet, il faut qu'elle soit capable de « traverser » jusqu'à 6 ordres de grandeur. Or, cette cascade de phosphorylation n'explique pas comment la bactérie se « réadapte » à la nouvelle concentration pour être à nouveau « attirable » par la concentration du dessus. Ce phénomène nommé adaptation compense les variations de concentration de la substance attractive en ajustant en permanence le niveau de methylation du récepteur. L'enzyme CheR ajoute jusqu'à 4 groupes méthyle sur le récepteur de façon constitutive. Plus de groupes méthyle sont rajoutés au récepteur, plus ce dernier active l'autophosphorylation de CheA. A l'inverse, une fois phosphorylé grâce à CheA, la protéine CheB agit comme une méthyle estérase et enlève des groupes méthyles sur la partie cytoplasmique du récepteur. Cette boucle de rétroaction négative permet d'ajuster en permanence le niveau de phosphorylation de CheA (et *in fine* le processus de décision de culbute) quelque soit la concentration de substance attractive présente. Ainsi, la bactérie reste sensible aux changements faibles, même pour les concentra-

tions extrêmes. Cette régulation agit comme une « mémoire »[1] permettant à la bactérie de « se souvenir » des concentrations récentes et de les comparer avec les concentrations actuelles, donc de « savoir » si elle se déplace dans la bonne direction (Vladimirov & Sourjik, 2009). Concluons sur le fait que tout comme la répression catabolique, ce mécanisme de chimiotactisme inspire également de nombreux travaux de modélisation (Bray *et al.*, 2007).

# 4 La détection des interactions géniques

Deux bases de données Ecocyc (Keseler *et al.*, 2009) et RegulonDB (Gama-Castro *et al.*, 2008) référencent les interactions géniques décrites chez *E. coli* et indiquent la ou les méthodes qui ont permis de les détecter. Ces méthodes sont nombreuses et je vais prendre le temps d'en décrire quelques unes. Le choix de la méthode dépend de ce que l'on souhaite faire :

⇒ Les méthodes *in vitro* permettent de démontrer l'existence d'une liaison physique ADN–protéine.

⇒ Les techniques à haut-débit (Puce à ADN, ChIP-chip, RNA-Seq, chIP-seq) permettent de détecter *in vivo* un grand nombre d'interactions en même temps.

⇒ Les gènes rapporteurs permettent de quantifier et contrôler l'effet d'une interaction *in vivo* et en temps réel.

⇒ Les méthodes bioinformatiques essaient de se passer de l'expérimentation souvent longue et fastidieuse.

Rappelez-vous que les interactions les plus fiables sont évidemment celles qui sont confirmées par plusieurs méthodes et par plusieurs équipes. Si vous faite de la biologie synthétique, ne faite jamais confiance *a priori* à une interaction décrite dans la littérature. Vérifiez-la par vous-même et préférez l'utilisation des interactions (ou des briques) les plus classiques qui ont fait leurs preuves depuis longtemps (LacI, TetR, AraC etc...).

## 4.1 Les méthodes *in vitro*

### 4.1.1 Le retard sur gel

Le retard sur gel (EMSA ou electrophoretic mobility shift assay) est une technique de biologie moléculaire permettant de détecter une interaction entre une protéine et de l'ADN. Le principe consiste à faire migrer dans un gel d'électrophorèse, le fragment d'ADN d'intérêt

[1] Vous trouverez sans doute l'explication un peu difficile à comprendre. C'est peut être lié à un manque de clarté de ma part mais aussi lié au fait que le mécanisme de cet effet mémoire est difficile à se représenter cognitivement. L'effort d'abstraction nécessaire est important.

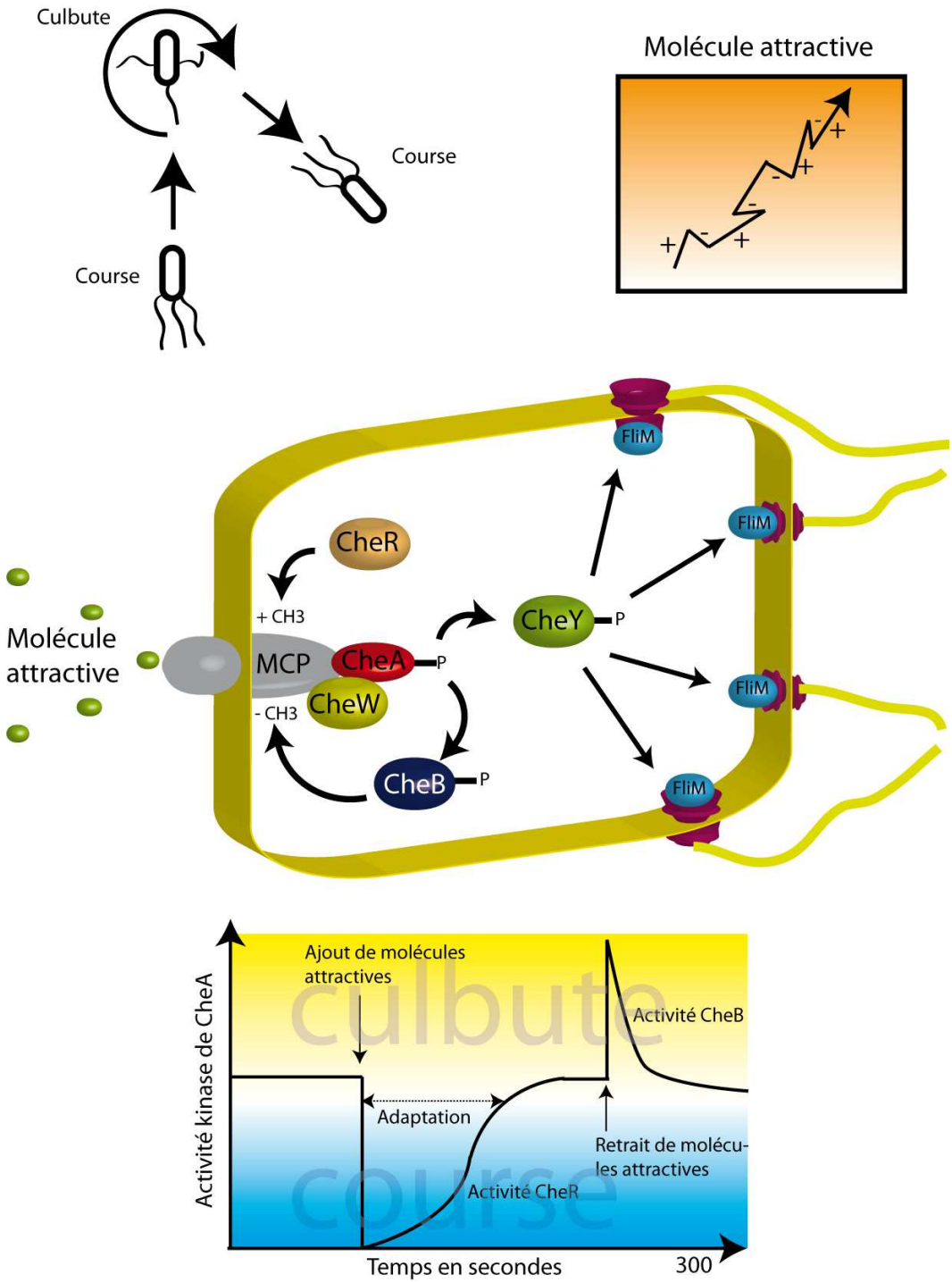



**Figure III.4** – Le chimiotactisme. Quelques schémas illustrant le phénomène de culbute, le chimiotactisme le long d'un gradient de concentration, le réseau de transduction du signal permettant le contrôle du sens de rotation des flagelles et enfin l'effet mémoire.

sur une piste, et sur une autre piste, le même fragment d'ADN mis en contact avec la protéine susceptible de se fixer dessus. La piste contrôle devrait contenir une seule bande correspondant à l'ADN non lié. Si la protéine (le potentiel facteur de transcription) est capable de se fixer au fragment d'ADN, la migration de ce dernier sera plus lente (il sera retardé) et la bande correspondante sera décalée vers le haut. A l'inverse, si la protéine n'a aucune affinité pour le fragment d'ADN, les deux bandes apparaitront au même niveau sur le gel.

### 4.1.2 L'empreinte à la Dnase

L'empreinte à la Dnase est une technique de biologie moléculaire qui détecte les interactions protéine–ADN en utilisant le fait qu'une protéine liée à l'ADN protège souvent cette ADN d'un clivage enzymatique. Par exemple, le fragment d'ADN d'intérêt peut être amplifié par PCR en utilisant un oligonucleotide marqué en 5' au $^{32}$P. Il en résulte une molécule d'ADN double brin marquée radioactivement à une seule de ces extrémités. Le clivage par la Dnase va produire des fragments de toutes les tailles. Les fragments plus petits migreront plus loin que les fragments longs. Le gel est ensuite utilisé pour exposer un film photographique spécial. Le pattern de clivage de l'ADN en absence de protéine est comparé au pattern de clivage de l'ADN en présence de la protéine susceptible de lier l'ADN. Si cette dernière se lie à l'ADN, le site de liaison est protégé de la digestion enzymatique. Cette protection crée une zone claire sur le gel correspondant à « l'empreinte » de la protéine.

## 4.2 Les méthodes à haut-debit

### 4.2.1 Les puces à ADN

L'apparition des puces à ADN a engendré des centaines d'expériences d'analyse du transcriptome qui mesurent le niveau d'ARNm à l'échelle du chromosome entier. Ces analyses ont été très utilisées pour trouver les cibles des facteurs de transcription en identifiant les gènes dont l'expression change lorsque le gène codant pour un facteur de transcription est déleté. Cependant cette approche n'a eu qu'un succès limité. En effet, le réseau de régulation génique d'*E. coli* est extrêmement complexe et de nombreux facteurs de transcription régulent l'expression d'autres facteurs de transcription. Ainsi la délétion d'un facteur de transcription a souvent de nombreux effets secondaires sur l'expression des gènes. Si la transcriptomic est utile pour étudier les patterns globaux de transcription, il est beaucoup plus difficile d'identifier les cibles directes (le « régulon ») d'un facteur de transcription donné.

### 4.2.2 La technique du ChIP-chip

La méthode de Chromatin ImmunoPrecipitation on Chip (abrégée ChIP on chip) est une technique permettant de mesurer directement des interactions protéine–ADN *in vivo*. Il s'agit d'une combinaison de la technique de Chromatin Immunoprécipitation avec la méthode des puces à ADN. Contrairement à la transcriptomic, la mesure de cette interaction ne dépend pas de la transcription du gène en aval. Chez les bactéries, cette méthode est particulièrement adaptée pour déterminer, de manière globale, les sites de fixation sur le chromosome d'un facteur de transcription donné (Grainger & Busby, 2008). Il est ensuite possible de déterminer la séquence d'ADN consensus de ce facteur de transcription.

Dans une expérience de ChIP-chip, les cellules sont fixées avec du formaldehyde ce qui crée des liaisons covalentes *in vivo* entre les facteurs de transcription et leur site de liaison sur l'ADN. Puis l'ADN est extrait des cellules et coupée en courts brins de l'ordre de 500pb par sonication. La protéine d'intérêt (associée aux fragments d'intérêt) est alors immuno-précipitée à l'aide de l'anticorps correspondant. On sépare ensuite l'ADN des protéines puis on marque les fragments d'ADN avec un fluorochrome vert. Puis on hybride ces fragments sur une puce à ADN avec un échantillon contrôle marqué, lui, en rouge. Ce contrôle, dont les données serviront de bruit de fond, peut être l'ADN génomique « faussement » immuno-précipité d'une souche délétée du facteur de transcription étudié. La puce est une plaque de petite taille sur laquelle sont fixés plusieurs dizaines de milliers de fragment d'ADN simple brin de 60 nucléotides. La séquence des ces milliers de fragments est espacée à intervalle régulier tout au long de la séquence du chromosome de *E. coli*. Ainsi les fragments immuno-precipités de plus de 500pb trouvent forcement leurs brins complémentaires auxquels s'hybrider. La plaque est ensuite lavée par des bains spécifiques pour éliminer les brins d'ADN ne s'étant pas hybridés. Chaque point (ou spot) de la puce va être analysée individuellement par un scanner à très haute résolution, et ce à la longueur d'onde d'excitation des fluorochromes rouges et verts. Puis, pour chaque spot, on détermine le ratio des intensités obtenues avec les deux fluorochromes. Les données finales correspondent à la valeur de ce ratio en fonction de la position sur le chromosome d'*E. coli*. Un ratio fluorochromes vert/rouge élevé signifie que l'hybridation de l'ADN immuno-précipité est supérieur à celle du contrôle ce qui signale la présence d'un site de liaison potentiel à cet endroit du chromosome.

[2]Cette image est une image modifiée à partir d'une image issue de Wikipedia.

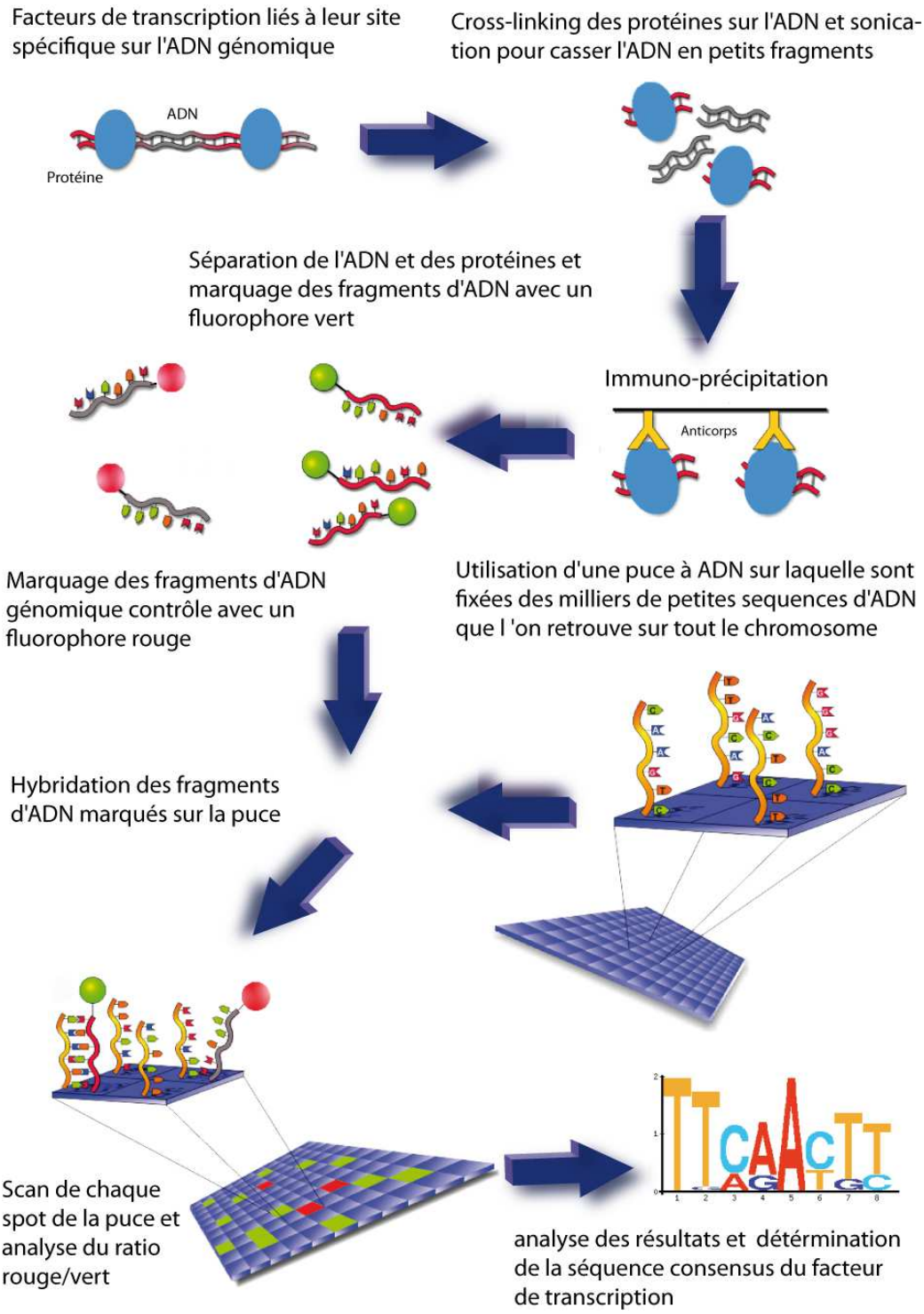



**Figure III.5** – La technique de ChIP on chip. Cette technique permet de déterminer l'ensemble des sites sur le chromosome qui ont une affinité pour un facteur de transcription donné [2].

### 4.2.3 Le RNA-Seq et le ChIP-Seq

Ces deux techniques récentes sont dérivées des deux précédentes mais cette fois la puce à ADN est remplacée par un séquençage à haut-débit. Comme dans l'analyse du transcriptome avec une puce à ADN, dans la technique de RNA-Seq, on s'intéresse au pattern global d'expression des gènes d'une souche donnée. Il peut s'agir d'un mutant d'un facteur de transcription donné que l'on compare avec une souche sauvage. Les différences d'expression entre les deux souches reflètent les effets de la délétion et donc les cibles potentielles du facteur de transcription. La première étape consiste à récupérer les ARN de la souche. Puis, si possible, on se débarrasse des ARN ribosomiaux. Les ARNm sont retrotranscris en ADNc puis fragmentés. Des séquences adaptatrices sont ajoutées à chaque fragment ADNc pour faciliter le séquençage. Chaque molécule est ensuite séquencée en utilisant les nouvelles technologies de séquençage à haut débit. En utilisant les données collectées, on peut créer une carte de transcription à l'échelle du génome qui indique le niveau d'expression de chaque gène/nucléotide (Wang *et al.*, 2009). L'avantage énorme du RNA-Seq est son faible bruit de fond et sa grande résolution (1 nucléotide) supérieur à la résolution obtenue avec une puce à ADN (de l'ordre de la dizaine de nucléotides). Le chIP-Seq est similaire au ChIP-chip mais la puce à ADN est remplacée par le séquençage à haut-débit des ADN immunoprécipités.

## 4.3 Les techniques utilisant la biologie moléculaire

En biologie synthétique, l'expérimentateur ne souhaite pas uniquement avoir connaissance d'une interaction : il veut aussi l'utiliser, la contrôler. Pour cela, il doit utiliser des méthodes qui lui permettent de contrôler une interaction *in vivo* et *en temps réel* (pour ajuster la force de l'interaction pendant l'expérience). Pour cela, le recours à la biologie moléculaire devient indispensable : il faut manipuler, couper, coller, recombiner des séquences d'ADN. En théorie, la biologie moléculaire est facile mais en pratique elle est souvent difficile, longue et fastidieuse à cause de la présence de faux positifs que la théorie ne considère pas vraiment. Selon les personnes, le temps pour réaliser une construction de biologie moléculaire peut prendre entre quelques jours à plusieurs mois. Le plus frustrant en biologie moléculaire, c'est que l'acharnement nécessaire ne rapporte pas un résultat : il rapporte juste un outil, non publiable et qui ne fonctionnera peut être pas lorsque l'expérience l'utilisant sera réalisée. En résumé, un biologiste qui maitrise la biologie moléculaire sur le bout des doigts est un atout précieux dans un laboratoire ou une entreprise...

### 4.3.1 Les fusions transcriptionnelles

L'utilisation d'une fusion transcriptionnelle est une approche classique pour mesurer l'activité d'un promoteur dans une cellule vivante. Cette approche consiste à fusionner un promoteur d'intérêt en amont d'un gène rapporteur. Comme leurs noms l'indiquent, les gènes rapporteurs « rapportent » l'activité d'un promoteur.

Le premier gène rapporteur utilisé est le gène *lacZ*. Comme nous l'avons dit plus haut, ce gène code pour une enzyme, la $\beta$-galactosidase. Or la concentration de cette enzyme peut être facilement quantifiée en mesurant l'hydrolyse d'un substrat chromogenique (ONPG) (Miller, 1972). L'apparition d'une couleur jaune (quantifiable à 420nm avec un spectrophotomètre) reflète la concentration de la $\beta$-galactosidase et donc l'activité du promoteur qui contrôle la transcription du gène *lacZ*. Cependant, la mesure de la concentration de $\beta$-galactosidase nécessite de lyser les cellules ce qui limite la résolution temporelle. De nouveaux gènes rapporteurs sont devenus de plus en plus utilisés et permettent de mesurer l'activité d'un promoteur *in vivo* (et sans avoir à tuer les bactéries), en temps réel et avec une très haute résolution temporelle. La Green Fluorescent Protein (GFP), dont la découverte a fait l'objet d'un prix Nobel en 2008 (Shimomura *et al.*, 1962) émet une lumière verte (509nm) lorsqu'elle est excitée avec une lumière bleue (488nm). En quantifiant les photons émis à l'aide d'un photomultiplicateur, on peut mesurer la concentration de la GFP et remonter ainsi à l'activité du promoteur. A titre de parenthèse, lorsque le promoteur en amont du gène codant pour la GFP est très fort, on peut observer, à l'œil nu, des colonies d'*E. coli* « vertes ». Le deuxième rapporteur, la luciférase génère de la bioluminescence et émet donc des photons directement sans nécessité d'excitation. Comme pour la GFP, on peut mesurer l'activité d'un promoteur simplement en mesurant le nombre de photons émis par la luciférase. Dans le noir, après un petit temps d'adaptation, on peut observer, à l'œil nu, la bioluminescence émise par des colonies d'*E. coli*.

Les fusions transcriptionnelles sont en général placées sur un plasmide. Un plasmide est une molécule d'ADN circulaire double brin distincte de l'ADN chromosomique, capable de réplication autonome et non essentielle à la survie d'une cellule. Les plasmides sont en général présents en plusieurs copies dans une bactérie. Cela à l'avantage d'amplifier le signal lorsque l'on mesure l'expression d'un gène. Cependant, un désavantage est la possible variation du nombre de copies du plasmide au sein d'une même population ce qui peut introduire un biais dans les études à l'échelle de la cellule. Une solution pour éviter ce problème consiste à intégrer la fusion transcriptionnelle directement dans le chromosome par recombinaison homologue. Notez qu'on perd dans ce cas l'amplification du signal.

Pour construire une fusion transcriptionnelle, on utilise des techniques de biologie molécu-

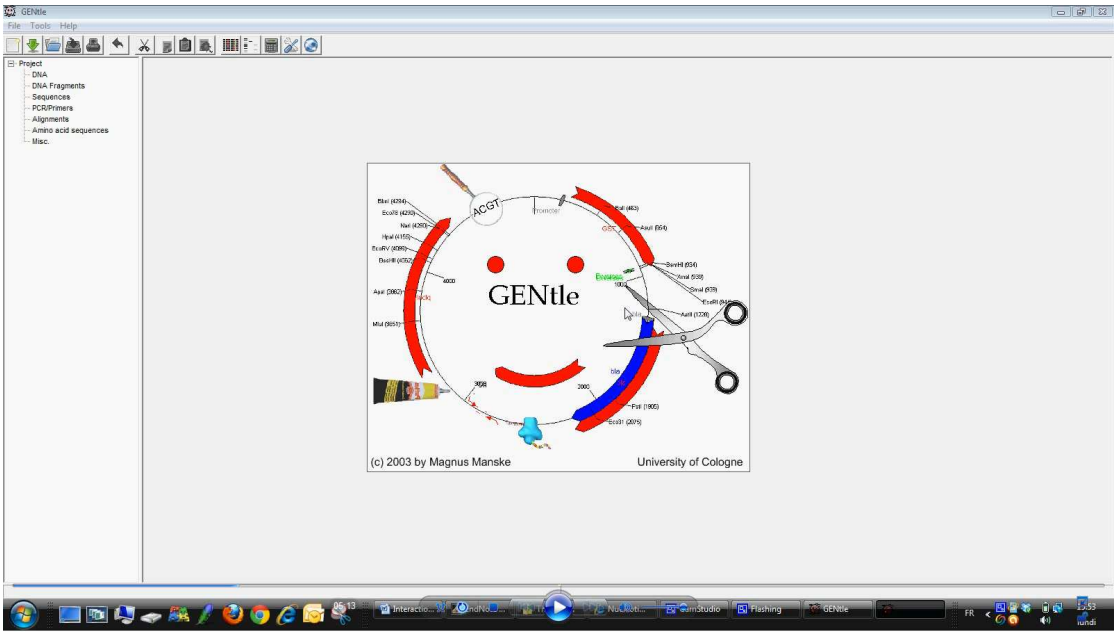



**Figure III.6** – Démonstration vidéo du logiciel Gentle. Cette video explique comment placer/cloner « *in silico* » un promoteur en amont d'un gène rapporteur en manipulant directement la séquence d'ADN.

laire. On amplifie par PCR le promoteur d'intérêt puis on clone ce promoteur en amont du gène rapporteur sur un plasmide. La première étape consiste à réaliser la construction *in silico*[3] . Pour cela j'utilise un logiciel gratuit Gentle mais il existe surement des logiciels plus performants. La vidéo ci-dessous décrit les étapes *in silico*, réalisées sous Gentle, permettant de placer (cloner) un promoteur en amont d'un gène rapporteur. J'ai réalisé cette démonstration pour montrer à un chercheur peu familier avec la biologie que les séquences d'ADN en « ATCG » ne sont pas uniquement une abstraction : on les manipule vraiment exactement comme on manipulerait un texte en utilisant les fonctions « couper » et « coller ». Une fois, la réalisation *in silico* réussie, on peut se lancer dans la réalisation réelle. Vous vous demandez peut être à quoi ressemble une étape de coupage ou de collage d'ADN ? Grossièrement, si vous observiez un biologiste faire de la biologie moléculaire, vous auriez l'impression de voir un homme ou une femme mélanger « de l'eau avec de l'eau » dans un tube pendant plusieurs jours. Puis quelques jours, quelques semaines ou quelques mois plus tard, il sauterait de joie au plafond juste après avoir reçu un résultat de séquençage en criant : ça y est, je l'ai !

Les fusions transcriptionnelles permettent la mesure de la concentration du rapporteur ou de l'activité du promoteur directement à l'échelle de la colonie, en population (milieu liquide) ou à l'échelle de la cellule (sous le microscope). Lorsque nous mesurons l'activité d'un promoteur

[3] *in silico* : Réalisé dans le « silicium » d'un ordinateur par opposition à *in vitro* et *in vivo*. Ce terme a été introduit par Antoine Danchin dans les années 1990.

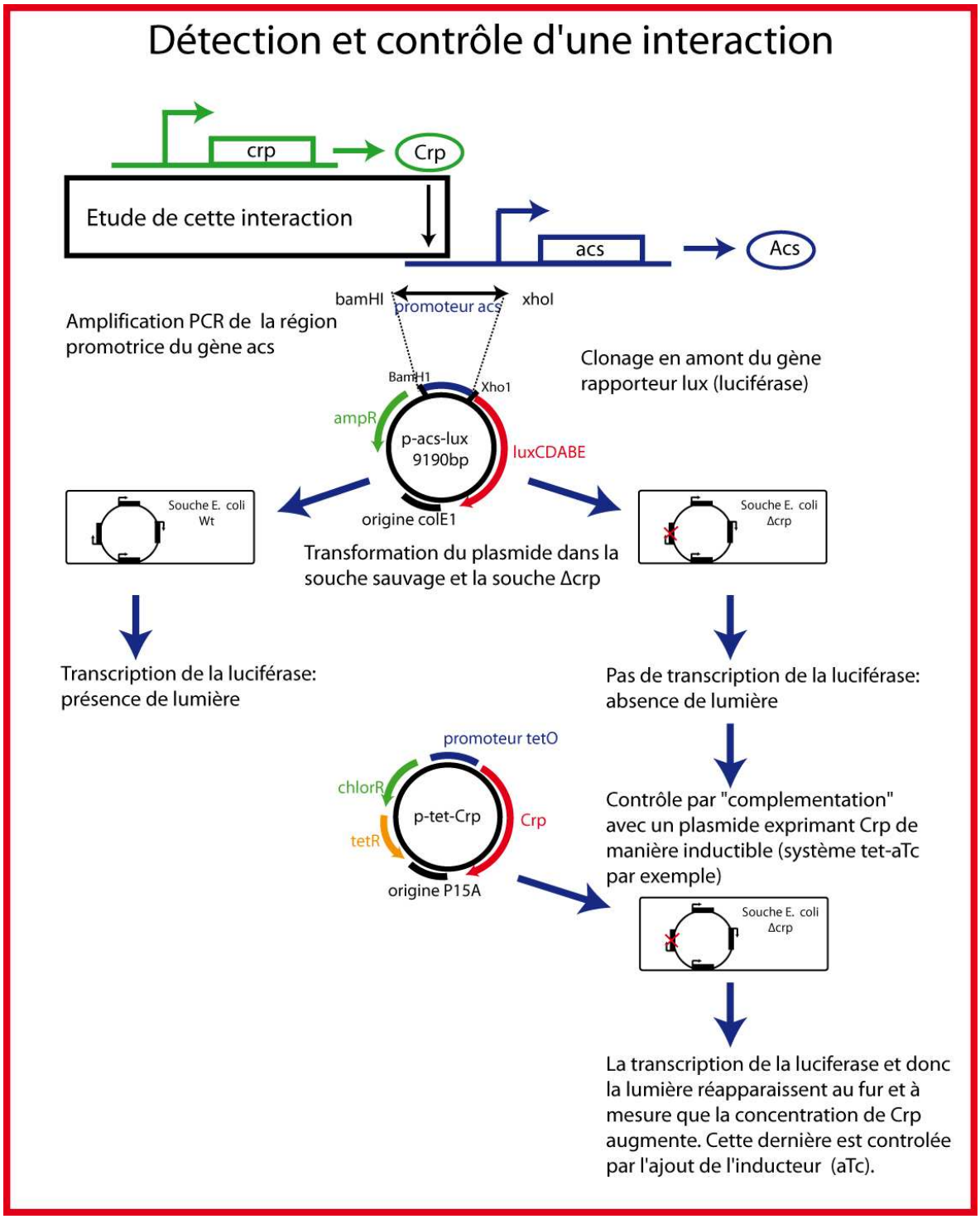



**Figure III.7** – Contrôle d'une interaction en utilisant un rapporteur (luciférase) et un plasmide de surexpression inductible.

dans une population en milieu liquide, nous utilisons un lecteur de microplaque pour mesurer la fluorescence, la luminescence ou l'absorbance. L'intensité I détectée par le photomultiplicateur, après soustraction du bruit de fond, est proportionnelle à la concentration de luciférase ou de GFP dans le puits. L'absorbance mesurée A, après soustraction du bruit de fond, est proportionnelle au nombre de bactéries dans le puits. La concentration moyenne du rapporteur [GFP] ou [Luciferase] dans une bactérie est donc :

$$[Luciferase](t) = \frac{I(t)}{A(t)} \tag{III.1}$$

$$[GFP](t) = \frac{I(t)}{A(t)} \tag{III.2}$$

Notez cependant que la luciferase a également besoin d'énergie pour émettre des photons. Quand l'énergie devient limitante, l'émission de photons chute brutalement et ne reflète plus la concentration de l'enzyme luciferase. Cet artefact peut être d'une part contrôlé et d'autre part utilisé pour inférer un certain nombre de paramètre dans la cellule (voir la publication relative à ce phénomène dans cette thèse).

Nous avons écrit un programme pour analyser des données d'expression génique issues de gènes rapporteurs. Ce programme, appelé Wellreader, a fait l'objet d'une publication (Boyer *et al.*, 2010). Il est possible, à partir des données collectées, de remonter à l'activité du promoteur F à l'aide de l'équation ci-dessous (démonstration dans la publication).

$$F(t) = \frac{dI(t)}{dt}\frac{1}{A(t)} + \gamma\frac{I(t)}{A(t)} \tag{III.3}$$

où $\gamma$ représente le taux de dégradation de la protéine rapporteur.

Pour mesurer une interaction avec une fusion transcriptionnelle, on compare la concentration du rapporteur dans une souche sauvage et dans une souche mutante déletée du facteur de transcription suspecté de contrôler l'activité du promoteur étudié. Notez qu'une différence d'expression ne signifie pas forcement que l'interaction est directe. Enfin pour s'assurer que l'interaction détectée n'est pas artefactuelle, on complémente la souche mutante avec un plasmide exprimant le gène déleté. Si la présence de ce plasmide permet de rétablir le niveau d'expression du mutant au niveau de la souche sauvage alors l'interaction —directe ou indirecte— est bien réelle.

Une publication présente dans cette thèse propose une méthode à haut-débit, basée sur les fusions transcriptionnelles, permettant de mesurer l'activité d'un promoteur dans les 4000 mutants d'*E. coli*.

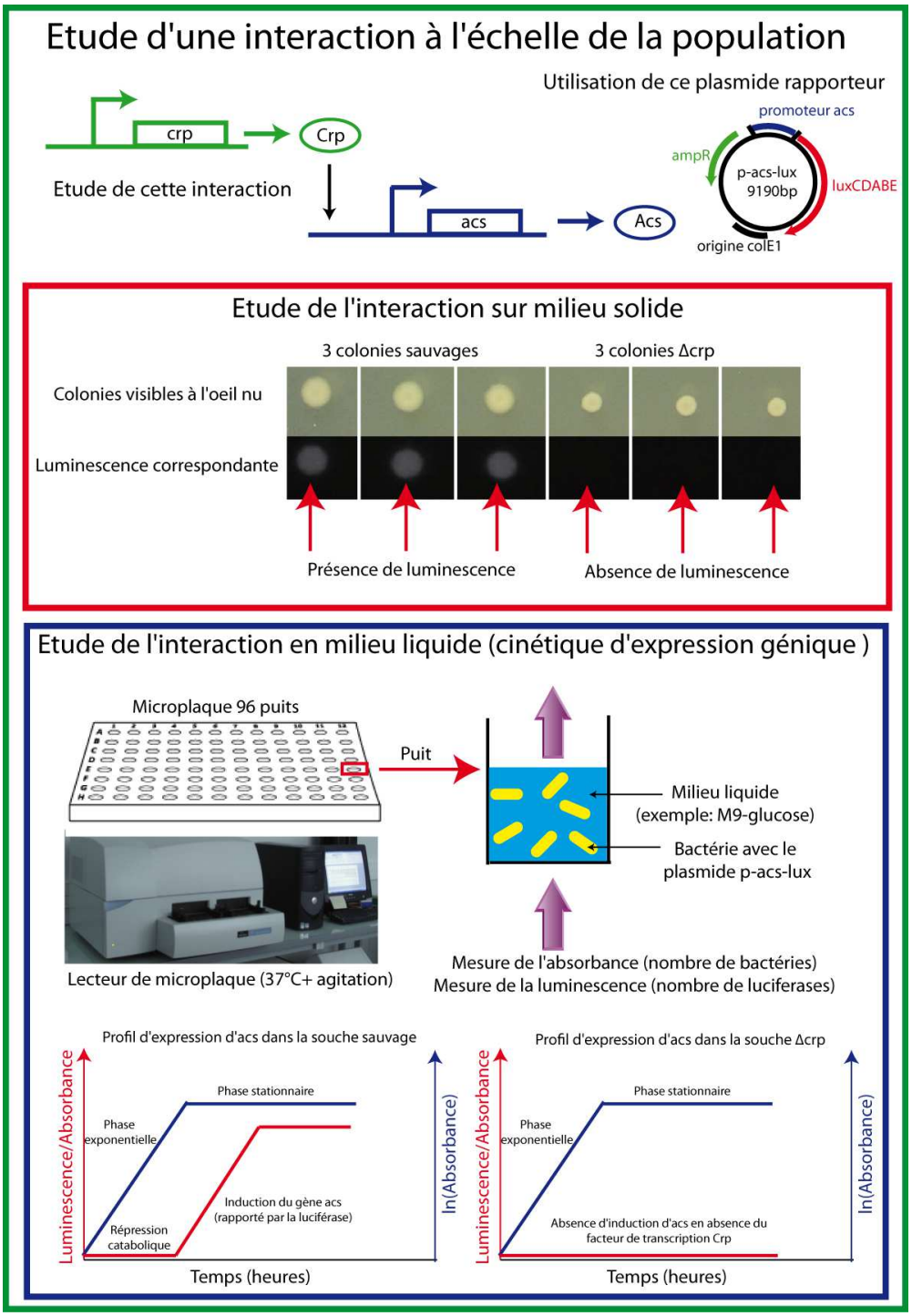
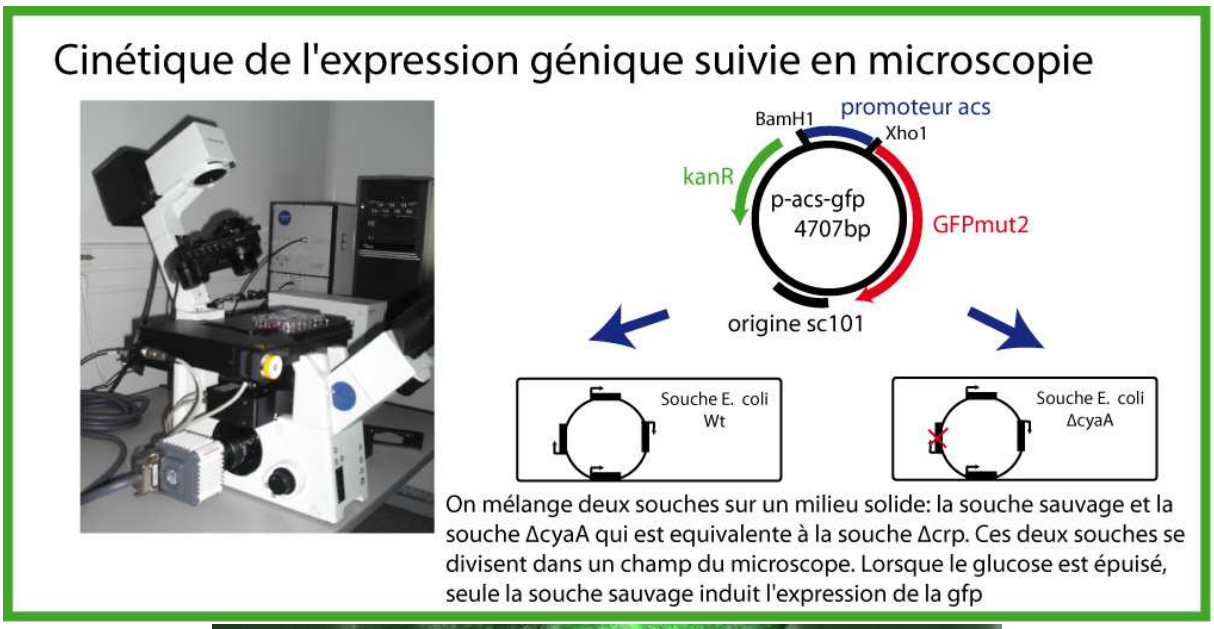



**Figure III.8** – Mesure de l'expression du gène *acs* (rapportée par la luciferase) dans une souche sauvage et dans une souche $\Delta crp$. Cette mesure est effectuée *in vivo*, en temps réel dans un milieu solide ou liquide. Il est possible de suivre l'expression génique au cours du temps avec une très bonne résolution temporelle (une mesure par minute pendant plusieurs heures).



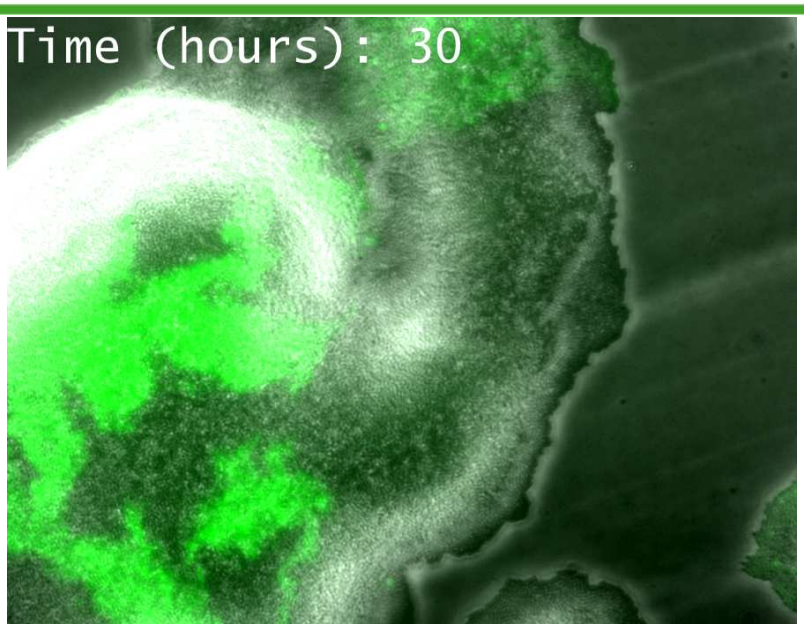




**Figure III.9** – Vidéo : mesure de l'expression du gène *acs* (rapportée par la gfp) dans une souche sauvage et une souche $\Delta cyaA$ mélangées ensemble sur une gélose solide filmée pendant 32 heures. A la fin du film, on observe une forte induction de fluorescence provenant d'une partie seulement du tapis bactérien (la souche sauvage). En absence d'un facteur de transcription fonctionnel, l'autre moitié du tapis ne peut pas induire l'expression de la gfp (vidéo issue d'une publication présente dans cette thèse).

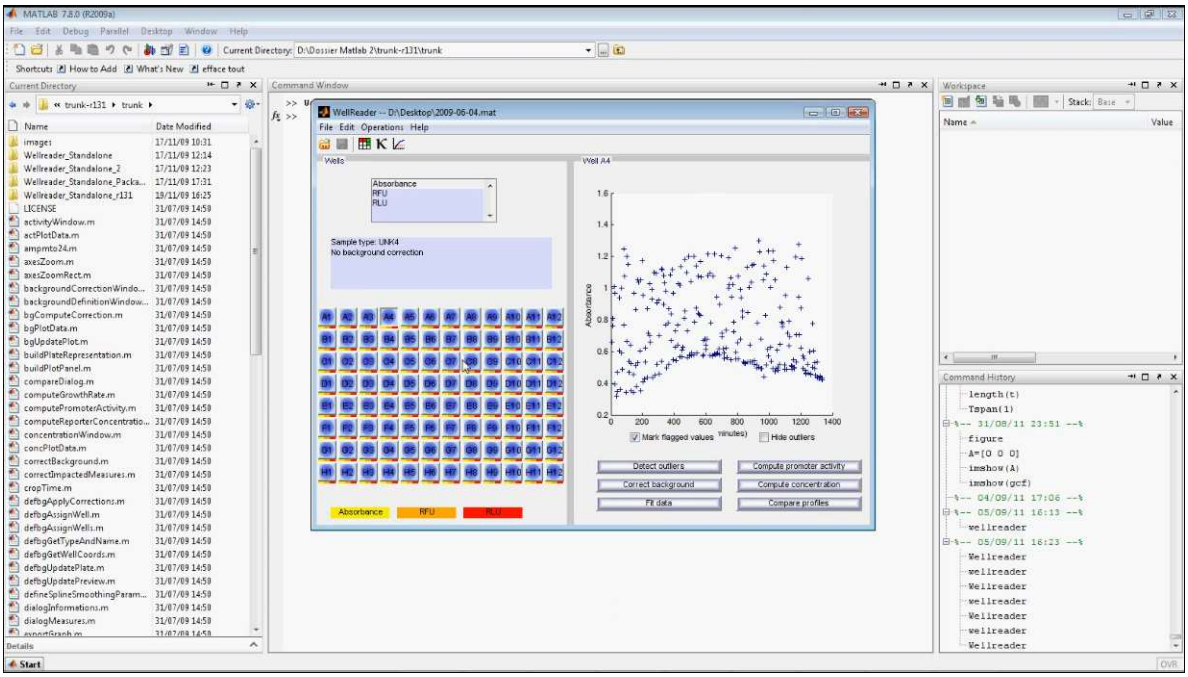


**Figure III.10** – Démonstration vidéo de l'utilisation du logiciel Wellreader qui permet d'analyser des données issues de gènes rapporteurs obtenues à partir d'un lecteur de microplaque.

## 4.4 Les méthodes bioinformatiques

### 4.4.1 Inférence de la topologie des réseaux de régulation

Toutes les techniques ci-dessus permettent de déterminer expérimentalement une interaction. Cependant, elles sont souvent assez chères et longues à mettre en place. L'apparition de grands « set » de données à haut-débit issues des puces à ADN a permis d'utiliser des méthodes « computationnelles » pour « prédire » des interactions géniques. Une manière simple d'exploiter les données issues de puces à ADN consiste à regrouper les gènes ayant des profils d'expression similaires : c'est la clusterisation. Mais il est également possible de mener l'exploration des données plus en profondeur et d'inférer la topologie du réseau, autrement dit de « reconstruire » ce réseau. Brièvement, des méthodes algorithmiques recherchent des patterns de corrélation ou de probabilités conditionnelles qui indiquent une influence causale. C'est le cas par exemple des algorithmes basés sur un réseau bayesien (Friedman *et al.*, 2000). L'inférence de réseaux biologiques est un domaine très actif et beaucoup trop vaste pour être abordé ici de manière exhaustive. Citons l'« inferelator », un algorithme qui démontre la faisabilité de reconstruction d'un réseau à l'échelle du génome à partir d'un nombre assez limité de puces à ADN (Bonneau *et al.*, 2006). Citons enfin les travaux de l'équipe de Bernhard Ø. Palsson. Cette équipe reconstruit des réseaux transcriptionnels et métaboliques à l'échelle du génome en utili-

sant de manière conjointe des séquences annotées, des données à haut-débit et des informations bibliographiques (Covert *et al.*, 2004; Feist *et al.*, 2009).

Je vais maintenant insister un peu plus sur une autre méthode d'inférence d'interaction ADN–protéine : la détection de site consensus par bioinformatique. Mon paragraphe reprend en partie un exemple proposé par Hamid Bolouri dans son ouvrage *Computational Modelling Of Gene Regulatory Networks : A Primer*. Le lecteur pourrait se demander pourquoi je choisis de passer tant de temps à décrire la détermination d'une « séquence–logo ». Il y a deux raisons : premièrement, cela me permet de proposer un petit exercice/exemple de programmation sous Matlab. Deuxièmement, cela me permet d'introduire la théorie de l'information de Shannon sur laquelle je compte revenir plus tard. Le lecteur qui craint les détails peut passer au chapitre suivant sans crainte.

### 4.4.2 Détection de site consensus par bioinformatique

La liaison ADN–protéine est principalement basée sur le pattern des donneurs et des accepteurs de liaison d'hydrogène exposée dans les sillons de la double hélice d'ADN. Ce pattern est donc très dépendant de la séquence et doit être plus ou moins complémentaire d'un pattern similaire présent au niveau du site de liaison de la protéine. La séquence qui maximise l'affinité d'un facteur de transcription donné est appelée séquence consensus. Cette séquence est déterminée expérimentalement grâce à un set de tous les sites séquencés que fixe un facteur de transcription donné. C'est la séquence qui, à chaque position, montre la paire de base la plus souvent trouvée parmi le set. L'utilisation des sites consensus mène les scientifiques dans un piège épistémologique leur faisant confondre le modèle de la réalité (le site consensus) avec la réalité (le site de liaison). En effet, le consensus strict est très éloigné du centre de la distribution gaussienne de l'affinité des sites de liaisons pour un facteur de transcription donné. Le consensus strict est donc très rarement observé. L'affinité trop forte d'un facteur de transcription pour son site consensus pourrait empêcher son détachement et résulterait en une activation ou répression constante du promoteur. Ainsi les algorithmes de recherche de sites de liaison sur une séquence donnée n'utilisent pas un site consensus mais une matrice décrivant pour chaque position, la fréquence d'apparition de chaque nucléotide. Ces matrices assument l'indépendance de chaque base les unes par rapport aux autres (matérialisée dans l'équation par une somme). En d'autre terme, la fréquence avec laquelle une protéine se fixe à un nucléotide donné est proportionnelle à l'affinité de cette protéine pour ce nucléotide.

Supposons qu'un facteur de transcription ait été rapporté pour se lier à un motif de 6 pb grâce à 20 observations expérimentales. Une matrice de positions peut résumer les donnés

expérimentales (figure III.11 ; matrice 1).

| Nucléotide observé | Position dans le motif | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| A | 3 | 12 | 2 | 2 | 3 | 0 |
| C | 5 | 1 | 0 | 11 | 6 | 6 |
| G | 2 | 7 | 12 | 6 | 10 | 13 |
| T | 10 | 0 | 6 | 1 | 1 | 1 |
| total | 20 | 20 | 20 | 20 | 20 | 20 |

On peut aussi représenter les observations comme des probabilités empiriques en divisant la fréquence de chaque nucléotide par 20 (figure III.11 ; matrice 2). La distribution des fréquences de nucléotide chez un organisme donné est rarement uniforme. On normalise donc la probabilité d'observer un nucléotide dans un site de liaison ($P_{observé}$) avec la probabilité d'observer un nucléotide dans une séquence « bruit de fond » ($P_{bruit}$)(figure III.11 ; matrice 3).

Pour calculer la probabilité totale qu'une séquence observée de quelques nucléotides représente un site de liaison, il faut multiplier les probabilités normalisées de chaque paire de base. Sachant que log(A.B)=log(A)+log(B), il est pratique de transformer les probabilités normalisées sous forme logarithmique. Ainsi pour estimer dans quelle mesure une séquence observée représente un site de liaison, on fait simplement la somme $\log(P_{observé}/P_{bruit})$ à chaque nucléotide. Un programme de prédiction fouillera le génome et sélectionnera, pour une séquence donnée, les signaux de longueur six dont le score est supérieur à un certain seuil (Kel *et al.*, 2003).

On peut aussi utiliser les matrices de positions pour créer des « séquences–logo » (figure III.11). Ces dernières offrent une manière simple et intuitive pour visualiser le contenu en information de la matrice de positions. Une pile de lettres est utilisée pour indiquer la fréquence d'occurrence de chaque nucléotide à chaque position dans la séquence alignée (axe des abscisses). La lettre la plus fréquente est placée en première position et les autres sont empilées dessous dans l'ordre des fréquences décroissantes.

La théorie de l'information nous fournit une mesure du contenu en information (IC) à chaque position i donné par la formule suivante (Workman *et al.*, 2005) où b représente les 4 bases A, T, C ou G.

$$IC(i) = \sum_{b=A}^{T} P_{observé}(b,i) \times \log \frac{P_{observé}(b,i)}{P_{observé}(b)} \tag{III.4}$$

Sortie du programme de création du logo sous Matlab

```
FrequenceApparition =

     3    12     2     2     3     0
     5     1     0    11     6     6
     2     7    12     6    10    13
    10     0     6     1     1     1


ProbabiliteApparition =

    0.1500    0.6000    0.1000    0.1000    0.1500         0
    0.2500    0.0500         0    0.5500    0.3000    0.3000
    0.1000    0.3500    0.6000    0.3000    0.5000    0.6500
    0.5000         0    0.3000    0.0500    0.0500    0.0500


Bruit =

    0.1000    0.1000    0.1000    0.1000    0.1000    0.1000
    0.4000    0.4000    0.4000    0.4000    0.4000    0.4000
    0.3000    0.3000    0.3000    0.3000    0.3000    0.3000
    0.2000    0.2000    0.2000    0.2000    0.2000    0.2000


ProbabiliteNormalise =

    1.5000    6.0000    1.0000    1.0000    1.5000         0
    0.6250    0.1250         0    1.3750    0.7500    0.7500
    0.3333    1.1667    2.0000    1.0000    1.6667    2.1667
    2.5000         0    1.5000    0.2500    0.2500    0.2500


Information =

    0.0877    1.5510         0         0    0.0877       NaN
   -0.1695   -0.1500       NaN    0.2527   -0.1245   -0.1245
   -0.1585    0.0778    0.6000         0    0.3685    0.7251
    0.6610       NaN    0.1755   -0.1000   -0.1000   -0.1000
```

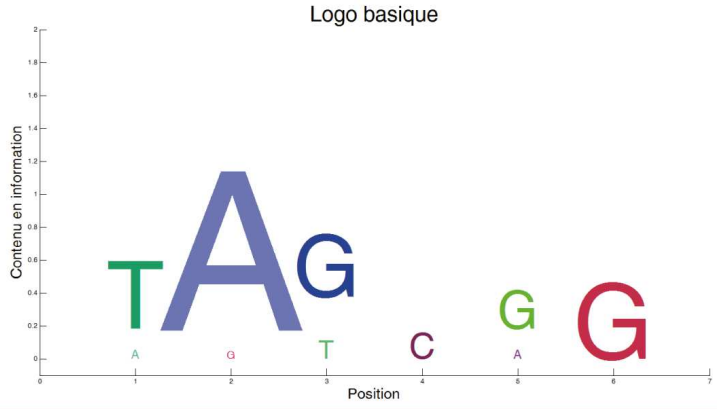



**Figure III.11** – Cette figure illustre « la sortie /output » d'un script Matlab qui génère un petit logo « basic » à partir des données initiales proposées. Voir le code Matlab en annexe Code02SequenceLogo.m qui gènère ce logo.

Le calcul de la hauteur d'une lettre est simplement

$$HL(i) = P_{observé}(b,i) \times \log \frac{P_{observé}(b,i)}{P_{observé}(b)}$$

La hauteur totale à une position donnée indique le contenu en information totale de cette position. Pour un site de liaison candidat donné avec un bruit de fond uniformément distribué de nucléotides aléatoires (0.25), une hauteur de 2 à n'importe quelle position dans la séquence du logo signifie qu'il n'y a qu'un seul nucléotide qui peut apparaitre à cette position.

$$HL = 1 \times \log \frac{1}{0.25} = 2bit$$

A l'inverse, une hauteur de 0 signifie que chaque nucléotide a une probabilité équivalente de se retrouver à cette position.

$$HL = 0.25 \times \log \frac{0.25}{0.25} = 0bit$$

L'utilisation du logarithme base 2 permet de quantifier l'information en bit, c'est-à-dire un nombre binaire qui n'a pour valeur que 0 ou 1. En utilisant 2 bits, on peut indiquer à une position donnée si le nucléotide est un A, un T, un C ou un G. C'est une autre manière de dire que 2 bits permettent $2^2$ =4 combinaisons de valeurs.

Souvent les séquences logo font apparaitre deux zones à fort contenu en information entrecoupés d'une zone à plus faible contenu en information. Les deux zones à fort contenu en information correspondent au sillon majeur de l'ADN où le facteur de transcription peut aisément discriminer les bases ce qui est moins le cas dans le sillon mineur. Ceci est un bel exemple pour entrevoir la structuration physique en double hélice de l'ADN.

# Références bibliographiques


Bonneau, R., Reiss, D. J., Shannon, P., Facciotti, M., Hood, L., Baliga, N. S., & Thorsson, V. 2006. The Inferelator : an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol*, **7**(5), R36. 70

Boyer, F., Besson, B., Baptist, G., Izard, J., Pinel, C., Ropers, D., Geiselmann, J., & de Jong, H. 2010. WellReader : a MATLAB program for the analysis of fluorescence and luminescence reporter gene data. *Bioinformatics*, **26**(9), 1262–3. 67

Bray, D., Levin, M. D., & Lipkow, K. 2007. The chemotactic behavior of computer-based surrogate bacteria. *Curr Biol*, **17**(1), 12–9. 58

Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., & Palsson, B. O. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, **429**(6987), 92–6. 71

Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., & Palsson, B. O. 2009. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, **7**(2), 129–43. 71

Friedman, N., Linial, M., Nachman, I., & Pe'er, D. 2000. Using Bayesian networks to analyze expression data. *J Comput Biol*, **7**(3-4), 601–20. 70

Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M. I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-Martinez, C., Abreu-Goodger, C., Rodriguez-Penagos, C., Miranda-Rios, J., Morett, E., Merino, E., Huerta, A. M., Trevino-Quintanilla, L., & Collado-Vides, J. 2008. RegulonDB (version 6.0) : gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res*, **36**(Database issue), D120–4. 58

Grainger, D. C., & Busby, S. J. 2008. Methods for studying global patterns of DNA binding by bacterial transcription factors and RNA polymerase. *Biochem Soc Trans*, **36**(Pt 4), 754–7. 61

Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. V., & Wingender, E. 2003. MATCH : A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res*, **31**(13), 3576–9. 72

Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., & Karp, P. D. 2009. EcoCyc : a comprehensive view of Escherichia coli biology. *Nucleic Acids Res*, **37**(Database issue), D464–70. 58

Kolb, A., Busby, S., Buc, H., Garges, S., & Adhya, S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem*, **62**, 749–95. 54

Kremling, A., Kremling, S., & Bettenbrock, K. 2009. Catabolite repression in Escherichia coli-a comparison of modelling approaches. *Febs J*, **276**(2), 594–602. 55

Miller, J. H. 1972. Experiments in Molecular Genetics. *Cold Spring Harbor*. 64

Monod, J. 1942. Recherches sur la croissance des cultures bacteriennes. *2nd edn. Paris :Hermann*. 54

Postma, P. W., Lengeler, J. W., & Jacobson, G. R. 1993. Phosphoenolpyruvate :carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev*, **57**(3), 543–94. 54

Saier, M. H., Jr., & Crasnier, M. 1996. Inducer exclusion and the regulation of sugar transport. *Res Microbiol*, **147**(6-7), 482–9. 55

Shimomura, O., Johnson, F. H., & Saiga, Y. 1962. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol*, **59**, 223–39. 64

Vladimirov, N., & Sourjik, V. 2009. Chemotaxis : how bacteria use memory. *Biol Chem*, **390**(11), 1097–104. 58

Wang, Z., Gerstein, M., & Snyder, M. 2009. RNA-Seq : a revolutionary tool for transcriptomics. *Nat Rev Genet*, **10**(1), 57–63. 63

Workman, C. T., Yin, Y., Corcoran, D. L., Ideker, T., Stormo, G. D., & Benos, P. V. 2005. enoLOGOS : a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res*, **33**(Web Server issue), W389–92. 72

# Chapitre IV

# La modélisation d'un réseau de régulation

## 1 Préambule

Ce chapitre (et le chapitre qui suit sur les systèmes dynamiques) s'inspire de deux ouvrages que je conseille vivement : il s'agit de *An introduction to systems biology* (Uri Alon) et *Computational modelling of gene regulatory networks : a Primer* (Hamid Bolouri). J'ai écrit ce chapitre pour les biologistes qui souhaiteraient connaître les bases nécessaires à la modélisation d'un réseau de régulation génique simple. Je pense en particulier à des étudiants IGEM biologistes qui voudraient comprendre ce que font les étudiants IGEM informaticiens/mathématiciens. Ce chapitre décrira le fonctionnement de petits modèles d'équations différentielles ordinaires et leur implémentation sous Matlab. A la fin du chapitre, j'aborderai également une technique d'optimisation pour identifier des paramètres : l'algorithme génétique.

## 2 Pourquoi modéliser ?

On m'a souvent demandé à quoi servait la modélisation. Et parfois, je sentais bien que mon interlocuteur souhaitait plus m'imposer sa réponse à lui que d'écouter la mienne. En effet, beaucoup pensent que « modèle, c'est le mot à la mode à placer dans un projet de recherche pour avoir des sous ». Dans ce chapitre, je vais tenter de démontrer aux sceptiques que la modélisation s'imposera inexorablement en biologie. Ceux qui pensent qu'il s'agit d'une mode sont, il me semble, dans l'erreur.

Raisonnons par l'absurde. Le sceptique envers la modélisation défend l'idée que pour décrire un système vivant, il faut décrire la structure et la fonction des composants ainsi que les interactions qui unissent ces composants en les représentant par des flèches. Ces flèches et la planche statique que le sceptique dessine sur une feuille de papier représentent, selon lui, le plus haut niveau de compréhension. Il pense aussi que le simple regard de ces planches ainsi que son

intuition lui permet de déduire « mentalement » la dynamique du système. Malheureusement, cet objet mental est trop vite limité par nos capacités cognitives et ne remplit donc pas, c'est le moins qu'on puisse dire, le critère minimum d'objectivité requis par la méthode scientifique. Il devient alors nécessaire de faire appel à la modélisation mathématique pour dépasser les limites de notre intuition. Prouvons-le maintenant :

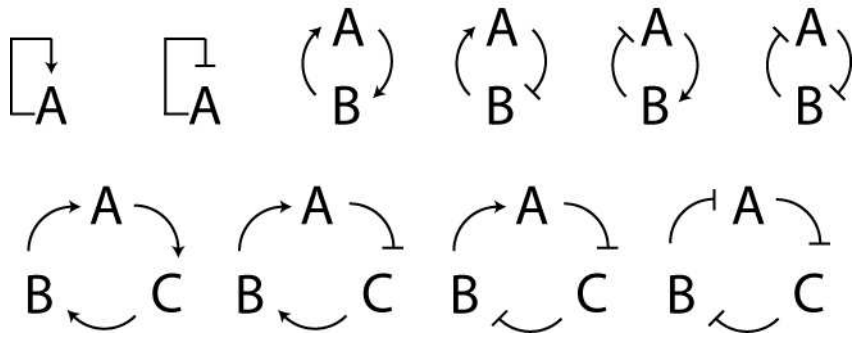



**Figure IV.1** – Les limites de l'intuition. Chaque petit réseau représente ce que j'appelle une planche statique. La connaissance des interactions ne suffit pas à prédire la dynamique de fonctionnement du réseau et sa fonction.

Parmi les petits réseaux ci-dessus, quels sont ceux qui permettent de générer des oscillations dans les concentrations des protéines (A, B, et C) ? Pas facile n'est ce pas ? On observe que même avec des réseaux très simples contenant seulement deux ou trois protéines, la présence de boucles de rétroaction positive ou négative complique très vite la tâche de l'intuition. Cette dernière ne peut tout simplement pas prédire la dynamique résultante de manière certaine. J'avoue ne pas savoir, moi-même, lesquels de ces réseaux sont capables de générer structurellement des oscillations. Une chose est sûre cependant : le dernier en est capable. Il s'appelle « le repressilator » et nous y reviendrons largement.
Alors pourquoi modéliser ? Dans ce chapitre, je montrerai deux exemples concrets qui illustrent l'intérêt de la modélisation. Nous verrons comment un modèle :

⇒ permet de dépasser les limites de l'intuition.

⇒ permet d'identifier des paramètres biologiques d'intérêt comme par exemple une constante de dissociation ou un taux de synthèse.

On peut ranger grossièrement les modèles en deux catégories : les modèles descriptifs qui fournissent un niveau de description bien plus poussé que ne le ferait une simple planche statique et les modèles prédictifs qui, comme leur nom l'indique, génèrent des prédictions que l'on peut ensuite évaluer/comparer avec des données expérimentales.

Imaginez une équipe de biologistes qui souhaiterait comprendre le fonctionnement d'un petit réseau de régulation comportant 20 protéines (20 gènes) chez une bactérie. L'équipe va

peut être commencer par mesurer expérimentalement la concentration des 20 protéines dans une souche sauvage. Puis pour comprendre le fonctionnement dynamique du réseau, ils vont vouloir le perturber, par exemple en supprimant des gènes. Ils pourront facilement construire les 20 simples mutants mais construire les 190 doubles mutants, 1140 triples mutants etc...[1] sera déjà beaucoup plus fastidieux. Le nombre d'expériences que pourrait potentiellement faire l'expérimentaliste croît très rapidement, mais toutes ne sont pourtant pas aussi *informatives*. Un bon modèle permettrait par exemple de sélectionner les expériences les plus prometteuses (les doubles et triples mutants potentiellement intéressant à construire) parmi un grand nombre d'expériences possibles.

Il me faut cependant relativiser la portée de ces modèles prédictifs quitte à être critique. En effet, cet objectif de prédiction est souvent théorique car extrêmement difficile en pratique. Comme le souligne Antoine Danchin, on ne compte plus les modèles qui ont redécouvert le cycle de Krebs (Danchin, 2009). En d'autres termes, les modèles actuels prédisent souvent ce que l'on connait déjà.

Le couple modèle–expérience est une démarche considérée comme saine qui est souvent mise en avant dans les projets de recherche. On retrouve le plan classique : développement d'un modèle parallèlement aux mesures expérimentales puis confrontation du modèle avec ces données expérimentales. Mais, en réalité, comme le souligne Antoine Coulon[2] , *« la confrontation [modèle–données expérimentales] est plus un processus qu'un moment donné dans le projet »*.

> *« La confrontation données–modèles n'a de sens que si elle est itérative sur une boucle courte, c'est-à-dire toutes les étapes (et tous les domaines de compétence) sont susceptibles de se fertiliser mutuellement »*.

Ainsi, il me semble que la confrontation du modèle prédictif avec les données expérimentales a souvent pour objectif idéaliste de faire émerger :

⇒ des prédictions vérifiées expérimentalement et donc un modèle « vrai » (or un modèle est faux par définition)

⇒ une grande quantité d'information « intelligible » extraite des données expérimentales grâce au modèle.

Dans la pratique, je pense que cet objectif idéaliste est assez rarement atteint mais cela n'est pas un problème. Je pense que l'apport principal du couple modèle–expérience n'est pas toujours là où on l'attend. Son but officieux/sa conséquence pratique est de créer un couple modélisateur–expérimentateur partageant deux savoirs distincts dans le but d'envisager la biologie sous le

[1] La formule utilisée est : $c_n^k = \frac{n!}{k!(n-k)!}$ où n est le nombre de gènes et k le nombre de gènes deletés.

[2] Antoine Coulon, *Le hasard au cœur de la cellule*, Syllepse chapitre 3 p. 75.

prisme des systèmes dynamiques et de la complexité. Mathématiciens et biologistes s'influencent mutuellement et apprennent à créer puis partager un savoir commun. Ainsi, l'intérêt principal de la modélisation est d'influencer la manière dont le biologiste pense, travaille et donc « design » ses expériences. Un bon modèle n'a pas pour unique but de générer des prédictions ou d'extraire des informations des données expérimentales : un bon modèle change l'expérience elle-même en changeant le regard de l'expérimentateur.

Aujourd'hui, un biologiste n'envisage plus raisonnablement de travailler sans ordinateur, sans internet, Word ou Excel. Dans quelques années, la complexité des données sera telle que le biologiste ne pourra plus raisonnablement faire de la biologie sans taper une ligne de code. Les langages de programmation (R, Matlab, Python...) changeront durablement sa manière de travailler, de penser ses expériences. Les mots complexité–modèle–système dynamique lui apparaitront moins mystérieux et donc moins suspects.

## 3 Petit encouragement au biologiste réfractaire

Le chapitre qui suit est riche en équations. Cependant, je souhaite tout de suite rassurer le lecteur biologiste potentiellement allergique aux mathématiques. Ce qui suit est écrit spécialement pour lui et ne dépassera jamais le niveau de terminale scientifique : c'est-à-dire mon propre niveau. Il me faut cependant introduire deux concepts barbares : « équation différentielle » qui, dans mon exposé, ne signifiera jamais autre chose que « taux » et « matrice » qui ne signifiera jamais plus que « tableau de nombres ».

Le biologiste devrait considérer la suite comme des travaux pratiques de mathématiques expérimentales. L'intérêt principal consiste à jouer « empiriquement » avec les systèmes dynamiques et cela est rendu possible par l'utilisation des méthodes numériques. Ce qui est important c'est d'étudier, de comprendre et de « jongler » avec le code : essayer par exemple de remplacer par un « 3 » la valeur numérique d'un paramètre fixé initialement à « 5 » puis regarder « la conséquence » sur les profils dynamiques. Le but est de « ressentir » de l'intérieur le système et la méthode. Un petit peu comme une expérience de biologie que l'on répète 100 fois de suite jusqu'à presque « ressentir » ce qui se passe dans le tube. La plupart des figures présentes dans ce chapitre ont été créées avec un script (un programme court) écrit avec le langage « Matlab ». Chaque script est accessible facilement (suivant le support sur lequel vous lisez cette thèse : Web/papier). J'ai essayé de commenter mon code au maximum pour rendre facile sa compréhension. Cependant, il est impossible de tout expliquer et un effort restera toujours nécessaire pour comprendre la logique des scripts, pour rentrer dans la tête du programmeur. J'ai choisi

d'utiliser MATLAB qui est à la fois un langage de programmation et un environnement de développement. Son inconvénient c'est qu'il est payant et non libre. Son avantage c'est qu'il existe une multitude de fonctions et « toolbox » déjà codées qui évitent de réinventer la poudre. Mais son plus gros avantage selon moi, c'est l'aide très complète et les nombreux didacticiels très bien construits qui permettent de faire ses premiers pas « en douceur ». Et cela est particulièrement important quand on n'a pas la culture de l'informatique. Avec un outil comme Matlab, vous pouvez faire, par exemple, du traitement d'image/de vidéo, vous pourrez également faire de la bioinformatique, des simulations de systèmes biologiques, du traitement de données lourd ou générer des figures complexes pour une publication. Bien sûr, vous disposez d'une multitude d'autres outils à peine plus complexes et à la fois utiles et amusants comme des outils d'optimisation (algorithme génétique) ou des algorithmes d'apprentissage (réseaux de neurones). Bon, j'en ai terminé avec la publicité. Sachez qu'il existe aussi des équivalents libres comme Scilab ou Octave ou même d'autres langages tout aussi intéressants comme Python ou R. On peut également espérer que, d'ici quelques années, apparaissent des environnements de développement intégrés (IDE) spécifiques à la biologie synthétique/des systèmes.

La section qui suit « quelques conjectures nécessaires » est un petit peu difficile et est donc plutôt destinée à des experts. Cependant, sa compréhension n'est absolument pas indispensable pour la suite de mon exposé qui sera, elle, plus abordable.

# 4 Quelques conjectures nécessaires

Ce chapitre introduit l'utilisation d'un formalisme mathématique adapté pour modéliser la dynamique de fonctionnement d'un réseau de régulation génique en considérant *le comportement moyen d'une population de cellules* génétiquement identiques. Ce formalisme est basé sur la loi d'action des masses et nous permettra d'introduire l'utilisation des équations différentielles ordinaires (ODE). Cependant, avant de rentrer dans le vif du sujet, je souhaiterais aborder ce que ce formalisme ne modélisera pas. Autrement dit, je souhaiterais insister sur les hypothèses initiales (assumptions en anglais) sur lesquelles repose la suite. Que se passe-t-il donc « pour de vrai » dans une cellule que nos modèles ne montreront pas ?

## 4.1 Processus stochastiques basés sur des réactions chimiques

Ce chapitre décrit la modélisation ODE basée sur la loi d'action des masses. Ce formalisme est particulièrement bien adapté lorsqu'il s'agit de comparer les résultats issus de simulation avec ceux issus des expériences. En effet, la plupart des mesures des concentrations cellulaires

d'ARN ou de protéines sont effectuées à partir d'extraits issus d'un grand nombre de cellules (une population). Les résultats obtenus à partir de ces extraits représentent une *moyenne* apportant une information sur le comportement *moyen* d'une cellule. Or cette moyenne cache la variabilité qui existe entre les cellules. Cette variabilité intercellulaire est due aux déplacements aléatoires des molécules selon le mouvement brownien. Cela injecte des probabilités dans les phénomènes observés. Probabilités qu'il est nécessaire de prendre en compte si l'on veut décrire ce qui se passe dans une cellule unique mais que l'on peut ignorer si on regarde le phénomène à l'échelle de la population. Nous reviendrons plus tard sur les méthodes de modélisation stochastique de l'expression des gènes. Cependant rappelons brièvement ici que la fréquence des collisions est fonction de la concentration des molécules et de leurs vitesses. Parmi toutes les collisions, seule une partie des molécules ont une énergie suffisante et la bonne orientation au moment de l'impact pour rompre les liaisons existantes et en former de nouvelles. L'énergie thermique minimale pour dépasser la barrière de transition permettant à la réaction chimique de se produire s'appelle l'énergie d'activation. Une des caractéristiques importantes de la biochimie cellulaire c'est que cette barrière d'énergie n'est pas suffisamment haute pour empêcher les changements aléatoires d'état aux températures physiologiques. Ceci signifie que la plupart des interactions protéines–protéines et ADN–protéines sont constamment en train de se former puis de se dissocier.

## 4.2 Homogénéité du milieu, diffusion et encombrement

Dans les modèles ODE que nous allons étudier, nous allons parler de concentration de protéines. Ces concentrations seront toujours considérées comme homogènes dans la cellule *E. coli*. Nous ne modéliserons pas l'espace et il n'y aura donc pas de gradient spatial. Bien que, contrairement aux cellules eucaryotes, il n'y ait pas de compartiments chez *E. coli*, cela n'empêche pas l'existence de ces gradient spatiaux (Llopis *et al.*, 2011). De plus, l'espace dans la cellule est extrêmement encombré ce qui réduit significativement les taux de diffusion des protéines et des complexes macromoléculaires (Zhou *et al.*, 2008; Elowitz *et al.*, 1999). Une des conséquences directe est la ségrégation des molécules qui ont une affinité les unes pour les autres ce qui crée des concentrations locales plus fortes. Pour l'interaction ADN–protéine, il a été montré que l'effet de l'encombrement peut augmenter l'activité de la liaison ADN–protéine par plus d'un ordre de grandeur (Poon *et al.*, 1997).

## 4.3 Recherche du site de liaison sur l'ADN par le facteur de transcription

On sait depuis les années 70 que les facteurs de transcription trouvent leur site de liaison beaucoup plus vite que ce qui serait attendu avec une diffusion simple en 3 dimensions (Riggs *et al.*, 1970). L'hypothèse actuelle propose une recherche avec deux types de diffusion (von Hippel & Berg, 1989). Le facteur de transcription recherche sa cible (son site de liaison) par glissement le long d'un court segment d'ADN ($\sim$ 100pb). Cela représente une diffusion facilitée à 1 dimension. Puis le facteur de transcription se dissocie, diffuse dans le cytoplasme (en 3 dimensions) avant de se lier à une nouvelle région d'ADN sur lequel il glisse à nouveau. La protéine ne restant sur l'ADN que quelques millisecondes, ce processus peut être itéré un grand nombre de fois jusqu'à ce qu'elle trouve enfin sa cible (en moins de 5 minutes) (Elf *et al.*, 2007; Li & Elf, 2009). Selon ce modèle de recherche, il est possible que notre vision binaire de la liaison ADN–protéine (faible–forte /spécifique–aspécifique) soit fausse et qu'il soit plus réaliste de penser en terme de « gaussienne » des distributions d'affinité des sites de liaison pour un facteur de transcription donné (Slutsky & Mirny, 2004).

## 4.4 Transcription et traduction

Dans les modèles que nous allons décrire, la machinerie globale (ARN polymérase, ribosomes) et l'énergie (ATP, pouvoir réducteur) seront toujours considérés en excès et donc non limitants. De plus, le délai de transcription–traduction–repliement qui existe entre la fixation du facteur de transcription activateur et l'arrivée des premières protéines fonctionnelles ne sera pas pris en compte.

## 4.5 Approximation pour les réactions rapides

Les régulations allostériques (fixation de l'IPTG sur le répresseur LacI par exemple), modifications post traductionnelles (phosphorylation, acétylation..) et les réactions métaboliques s'effectuent à une vitesse très rapide (millisecondes, secondes). Par conséquent, elles sont considérées « à l'équilibre » par rapport aux régulations géniques (minutes, heures). Autrement dit, l'ajout d'IPTG en concentration suffisante dans le milieu change instantanément l'affinité de LacI pour l'ADN par rapport à l'arrivée de la $\beta$-galactosidases. Il n'est donc pas nécessaire de modéliser explicitement l'import de l'IPTG et la formation du complexe [lacI.IPTG] en fonction du temps.

## 4.6 Intégration des signaux au niveau des promoteurs complexes

Nos modèles représentent toujours des cas simples avec soit 1, soit 2 facteurs de transcriptions régulant l'activité d'un même promoteur. Lorsque ces deux facteurs sont présents en même temps, on utilise la logique booléenne classique (fonction AND ou OR par exemple) pour intégrer le signal résultant. Mais comment intégrer les signaux issus de promoteurs plus complexes comme le promoteur aceBAK ci-dessous ? Quelle allure a la fonction de sortie qui décrit l'activité d'un promoteur complexe en fonction de ses « entrées » protéiques ? On touche ici à deux limitations :

⇒ une limitation expérimentale : aucune méthode n'est en mesure actuellement de fournir une résolution suffisante pour nous aider à déterminer cette fonction.

⇒ une limitation relative au formalisme de modélisation : de mon point de vue, seul un modèle spatial à l'échelle de l'atome et de la liaison chimique peut fournir la fonction qui décrit l'activité d'un promoteur complexe en fonction de ses « entrée » protéiques.

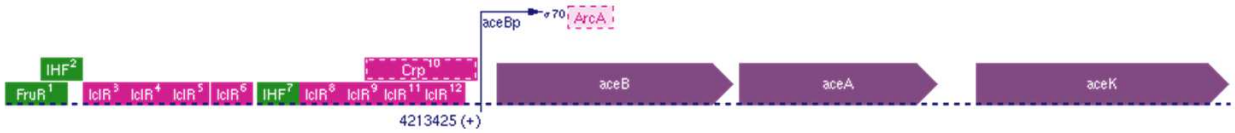



**Figure IV.2** – La régulation très complexe de l'operon aceBAK (image Ecocyc)

## 4.7 La topologie du réseau est dynamique

Dans les modèles que nous allons voir, la topologie sera considérée comme statique. En réalité, la topologie est dynamique et devrait, dans l'absolu, être modélisée comme une variable. Par exemple, les variations de super-enroulement (qu'on appelle aussi topologie mais qui n'a rien à voir avec la topologie dont je parle ici) de l'ADN peuvent supprimer l'affinité d'une protéine pour un site sur l'ADN ce qui supprime l'existence de cette interaction.

# 5 Le formalisme des équations différentielles

Quel est le point de départ du biologiste qui souhaite modéliser la dynamique de fonctionnement d'un réseau de régulation génique ? Le biologiste veut suivre l'évolution de la concentration d'une protéine en fonction du temps.

Le nombre de protéines B dans une bactérie au temps t+dt est égal au nombre de protéines B au temps t auquel on ajoute la quantité synthétisée $\kappa \cdot dt$ durant le temps dt et auquel on

soustrait la quantité dégradée $\gamma \cdot dt \cdot [B]$ durant le temps dt. Comme le volume de la bactérie est considérée constant, notez également que l'on peut aisément passer du nombre de molécules aux concentrations et vice versa en divisant ou en multipliant par le volume.

On a :

$$[B] = [B]_t + \kappa \cdot dt - \gamma \cdot dt \cdot [B] \tag{IV.1}$$

Je vois venir votre question : pourquoi la quantité dégradée est-elle proportionnelle à B ? Cela est dû aux « histoires de demi-vie » dont je suis sûr qu'il vous reste surement quelques traces : la fonction qui décrit cette dégradation est une exponentielle et non une droite. Par exemple si la synthèse s'arrête ($\kappa = 0$) et qu'on a $\gamma = 1/2$, dt=1 et B=100 au temps t, alors à t+1, B=50 et on a dégradé 50 molécules de B. A t+2, B=25 et on a dégradé 25 molécules de B, à t+3, B=12.5 et on a dégradé 12.5 molécules de B. On voit donc bien que le nombre de molécules de B dégradées dépend du nombre de molécules de B déjà présentes. Vous pouvez aussi constater, en fronçant les sourcils sur le « 12.5 molécules de B », l'intérêt de travailler en concentration (nombre réel, continu) et non en nombre de molécules (nombre entier, discret).

Si on passe B de l'autre coté, on a :

$$[B]_{(t+dt)} - [B]_t = \kappa \cdot dt - \gamma \cdot dt \cdot [B] \tag{IV.2}$$

Si on divise les deux cotés par dt, on a :

$$\frac{[B]_{(t+dt)} - [B]_t}{dt} = \kappa - \gamma \cdot [B] \tag{IV.3}$$

La voici donc notre première équation différentielle. Le taux de changement de la concentration de la protéine B par unité de temps (d[B]/dt) est proportionnel au taux de synthèse $\kappa$ de la protéine B (qui inclut transcription et traduction) moins le taux de dégradation $\gamma$ multiplié par [B].

$$\frac{d[B]}{dt} = \kappa - \gamma \cdot [B] \tag{IV.4}$$

Le terme $\gamma$ comprend à la fois la dégradation spontanée de la protéine et le taux de dilution. Ce dernier dilue le contenu protéique par deux à chaque fois que la cellule bactérienne se divise.

$$\gamma = \gamma_{deg} + \gamma_{dil} \tag{IV.5}$$

Suivant les conditions et l'organisme utilisé, c'est l'un ou l'autre de ces termes qui est prédominant. Par exemple, dans une cellule différenciée qui ne se divise plus ou alors pour des bactéries en phase stationnaires, on a $\gamma = \gamma_{deg}$. A l'inverse, pour des bactéries en phase

exponentielle $\gamma \simeq \gamma_{dil}$.

Par définition, à l'état d'équilibre, on a $\frac{d[B]}{dt} = 0$. Par conséquent, la concentration de B est constante. On déduit que [B] est égale au ratio entre les taux de synthèse et de dégradation.

$$[B] = \frac{\kappa}{\gamma} \tag{IV.6}$$

# 6 Petit détour par la loi d'action des masses

Maintenant, on sait suivre l'évolution de la concentration d'une protéine B en fonction du temps. Et ce qu'on voudrait c'est pouvoir moduler cette concentration en fonction du contrôle de la transcription du gène b. On peut donc souhaiter qu'un gène b soit activé par un facteur de transcription A.

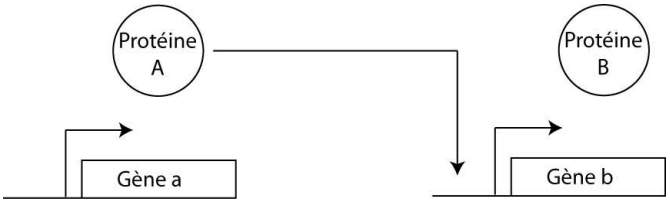



**Figure IV.3** – Schéma d'une activation transcriptionnelle simple.

Ainsi, si la protéine A se fixe au promoteur du gène b, cela facilite le recrutement de l'ARN polymérase qui transcrit le gène b. L'ARN correspondant est ensuite traduit ce qui résulte en une accumulation de la protéine B. C'est donc le taux de synthèse de la protéine B qui est fonction de la concentration du facteur de transcription A tel que :

$$\frac{d[B]}{dt} = \kappa \cdot f([A]) - \gamma \cdot [B] \tag{IV.7}$$

Mais quelle est cette fonction f ? Cette fonction relie la concentration moyenne du facteur de transcription A au niveau moyen d'ARN et de protéines B produites. Elle correspond à la fraction de cellules dont le promoteur du gène b (ADN) est occupé par le facteur de transcription A (et donc permettant l'initiation de la transcription) parmi toutes les cellules (et donc tous les promoteurs). Autrement dit c'est la fraction de complexe (ADN–protéine A) parmi tout les sites d'ADN libre. Cette fraction correspond à la probabilité que le promoteur soit actif. Ainsi le site ADN peut être soit fixé à la protéine A soit libre ce qui résulte dans l'équation de

conservation suivante :

$$[ADN] + [A] \cdot [ADN] = [ADN_{tot}] \tag{IV.8}$$

où $ADN_{tot}$ est la concentration totale de sites cibles d'ADN (promoteurs du gène b dans une population de cellules). Les protéines A (notées A) et le promoteur (noté ADN) diffusent dans le cytoplasme et occasionnellement se rencontrent pour former le complexe [A.ADN]. Ce processus est décrit par la cinétique d'action des masses où A et ADN entrent en collision et se lient à un taux $k_{on}$. Le taux de formation du complexe est proportionnel au taux de collision, donné par le produit des concentrations A et ADN :

$$Taux\ de\ formation\ du\ complexe = k_{on} \cdot [A] \cdot [ADN] \tag{IV.9}$$

Le complexe [A.ADN] peut se dissocier à un taux $k_{off}$ . Le taux de changement de [A.ADN] basé sur ces processus de collision et de dissociation est décrit par la cinétique d'action des masses :

$$\frac{d[A \cdot ADN]}{dt} = k_{on} \cdot [A] \cdot [ADN] - k_{off}[A \cdot ADN] \tag{IV.10}$$

A l'état d'équilibre (d[A.DNA]/dt=0), on a :

$$k_d \cdot [A \cdot ADN] = [A] \cdot [ADN] \tag{IV.11}$$

Où $k_d$ représente la constante de dissociation $k_d = \frac{k_{off}}{k_{on}}$. Ainsi la fraction d'ADN sur lequel A est lié est :

$$f([A]) = \frac{[A \cdot ADN]}{[ADN_{tot}]} = \frac{[A \cdot ADN]}{[A \cdot ADN] + [ADN]} = \frac{\frac{[A] \cdot [ADN]}{k_d}}{\frac{[A] \cdot [ADN]}{k_d} + [ADN]} = \frac{\frac{[A]}{k_d}}{\frac{[A]}{k_d} + 1} = \frac{[A]}{[A] + k_d}$$

Cette équation correspond à l'équation de Michaelis et Menten plus connue dans le contexte de la cinétique enzymatique. Si on remplace f(A) dans l'équation IV.7, on obtient :

$$\frac{d[B]}{dt} = \kappa \cdot \frac{[A]}{[A] + k_d} - \gamma \cdot [B] \tag{IV.12}$$

C'est cette première équation que nous allons écrire puis « intégrer » dans « Matlab ».

# 7 Etude du code

Vous pouvez copier le code présent en annexe Code03ActivationSimple.m puis aller dans Matlab, File,New,Blank m-file et coller le code. Appuyer sur F5 pour exécuter le code. La figure IV.4 apparait. On observe que la concentration du facteur de transcription A est constante. Celui-ci active la transcription du gène b ce qui augmente transitoirement la concentration de la protéine B. Puis cette dernière atteint un état d'équilibre. On peut donc découper la régulation en deux états. Un *état transitoire* d'adaptation et un *état d'équilibre* (steady state en anglais) lorsque les concentrations deviennent constantes.

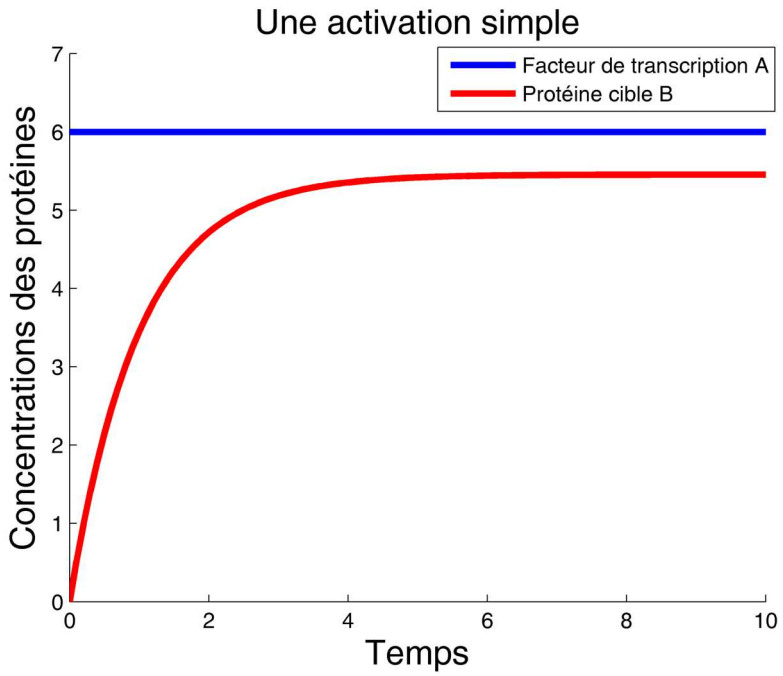



**Figure IV.4** – Dynamique d'une activation simple. Voir le code Matlab en annexe Code03ActivationSimple.m qui génère cette figure.

Le code peut sembler compliqué à première vue mais sachez que tout ce qui est en vert c'est les commentaires (les explications). Ce code peut être découpé en plusieurs parties :

⇒ La fonction nommée « équation » décrit le système composé de deux équations différentielles. La première équation représente l'évolution de la concentration du facteur de transcription A en fonction du temps (constante et donc nulle). La deuxième équation représente l'évolution de la concentration de la protéine B en fonction du temps et qui dépend de A (voir equation IV.12)

⇒ L'attribution (arbitraire) de valeurs numériques aux 3 paramètres (kappa, gamma, kd) ainsi que les deux conditions initiales qui correspondent aux concentrations de A et de B

au temps t=0. Notez que dans tous les scripts que nous allons étudier, j'attribue toujours une valeur arbitraire aux paramètres. Les modélisateurs plus sérieux recherchent dans la littérature si les valeurs numériques des paramètres ont été identifiées ou à défaut, ils les fixent à des valeurs biologiquement raisonnables.

⇒ L'intégration de la fonction « équation » est faite par un solveur numérique : c'est une fonction de Matlab appelée ode45 qui intègre « numériquement » l'équation différentielle c'est-à-dire qu'elle calcule l'évolution de la concentration de A et B en procédant par itération successive. A chaque petit pas, le solveur fait une petite erreur (négligeable pour nous) et la solution proposée n'est donc pas exacte. Pour avoir une solution exacte, il faudrait intégrer « analytiquement » l'équation mais cela s'avère en général impossible pour des systèmes d'équations non linéaires trop complexes.

⇒ La fonction ode45 renvoie comme par magie les concentrations de A et de B (stockées en ligne dans la matrice y) en fonction du temps. On peut ensuite tracer les résultats dans la figure et ajouter un titre, une légende et des noms aux axes.

# 8 Cas du répresseur

Dans le cas où A réprime B, on montre facilement que :

$$\frac{d[B]}{dt} = \kappa \cdot (1 - \frac{[A]}{[A] + k_d}) - \gamma \cdot [B] = \kappa \cdot \frac{k_d}{[A] + k_d} - \gamma \cdot [B] \tag{IV.13}$$

# 9 Cas de l'inducteur allostérique

Il arrive très souvent qu'un facteur de transcription soit activé par un inducteur. C'est le cas, comme nous l'avons vu, de la protéine Crp qui est spécifiquement activée par la petite molécule AMP cyclique. La liaison de l'AMPc à Crp change sa conformation et augmente son affinité pour son site spécifique sur l'ADN. L'échelle de temps de la fixation de l'AMPc à Crp est beaucoup plus courte que celle des processus de régulation génique. Nous faisons donc l'approximation du quasi état d'équilibre (quasi steady state). Dans le cas simple sans coopérativité, on a :

$$[Crp]_{actif} = [Crp \cdot cAMP] = \frac{[Crp] \cdot [cAMP]}{k_{d2} + [cAMP]} \tag{IV.14}$$

où $k_{d2}$ est la constante de dissociation de l'AMPc à Crp.

Dans le cas de la protéine Acs, sur laquelle j'ai travaillé, et dont la transcription est spécifiquement activée par le complexe [Crp.cAMP], le taux de changement de sa concentration s'écrit donc :

$$\frac{d[Acs]}{dt} = \kappa \cdot \frac{\frac{[Crp]\cdot[cAMP]}{k_{d2}+[cAMP]}}{\frac{[Crp]\cdot[cAMP]}{k_{d2}+[cAMP]} + k_d} - \gamma \cdot [Acs] \tag{IV.15}$$

où $k_d$ est la constante de dissociation du complexe Crp–cAMP à l'ADN.

Voici la figure que l'on obtient cette fois. Remarquez que sans les commentaires, le code est beaucoup plus court.

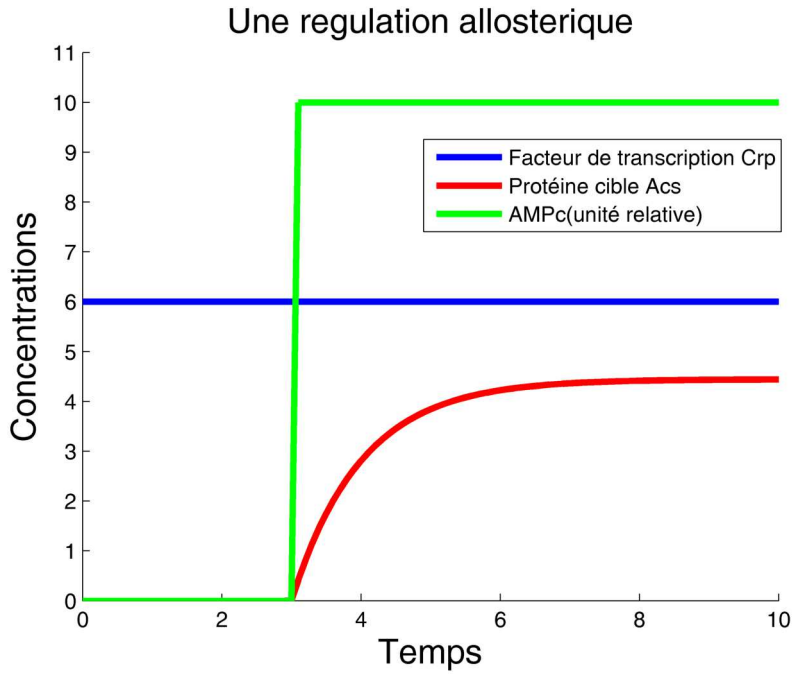



**Figure IV.5** – Dynamique d'une régulation allostérique. Voir le code Matlab en annexe Code04RegulationAllosterique.m qui gènère cette figure.

## 10 Découplage de la transcription et de la traduction

Jusqu'à présent, nous avons fait l'hypothèse que le taux de synthèse incluait transcription et traduction. Il est bien sûr possible de découpler ces deux processus en deux équations différentielles distinctes pour, par exemple, y inclure des régulations traductionnelles. On a :

$$\frac{d[ARN_m]}{dt} = \kappa_{trans} - \gamma_{darn} \cdot [ARN_m] \tag{IV.16}$$

$$\frac{d[Prot]}{dt} = \kappa_{trad} \cdot [ARN_m] - \gamma_{dprot} \cdot [Prot] \tag{IV.17}$$

où $\kappa_{trans}$ correspond au taux de transcription, $\gamma_{darn}$ correspond au taux de dégradation des ARNm, $\kappa_{trad}$ correspond au taux de traduction des ARNm et $\gamma_{dprot}$ correspond au taux de dégradation de la protéine.

# 11 Cas de la liaison coopérative :

Il arrive parfois que la liaison du facteur de transcription au promoteur soit coopérative. Cela signifie qu'il faut plusieurs facteurs de transcription fixés au promoteur pour obtenir une activation complète de la transcription du gène B et que la fixation des premiers aide à la fixation des suivants. Quand tel est le cas, on utilise l'équation de Hill qui est un cas général de l'équation de Michaelis et Menten :

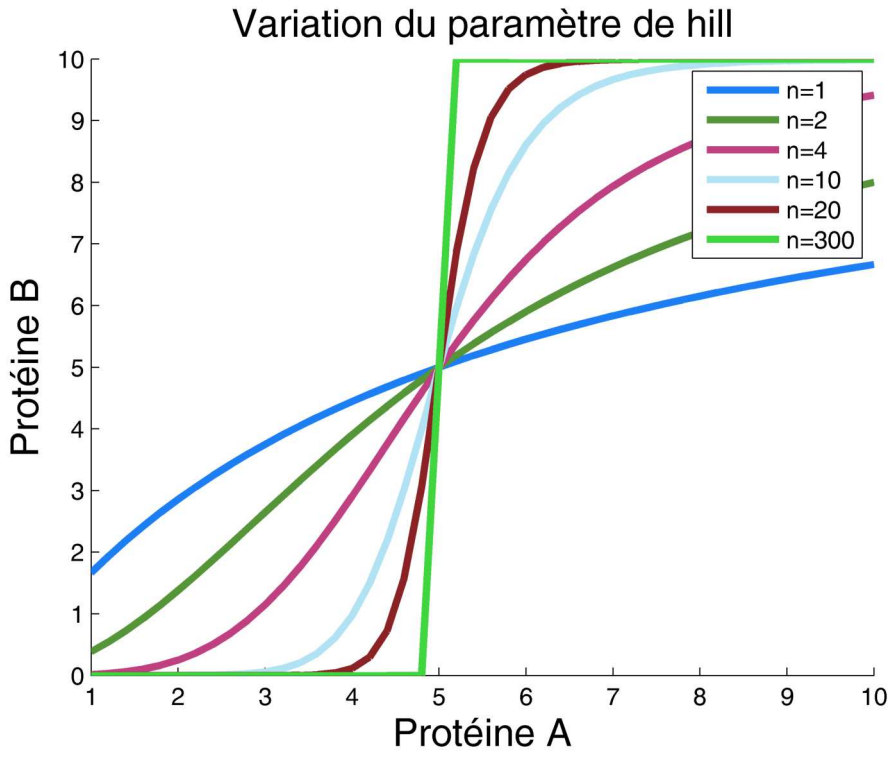



**Figure IV.6** – Liaison coopérative et fonction de Hill. Voir le code Matlab en annexe Code05ParametreHill.m qui génère cette figure.

$$\frac{d[B]}{dt} = \kappa \cdot \frac{[A]^n}{[A]^n + k_d^n} - \gamma \cdot [B] \tag{IV.18}$$

Cette équation introduit le paramètre de Hill noté n qui permet de contrôler l'allure et la pente de la courbe correspondante. Lorsque n=1, on obtient l'équation de Michaelis et Menten. Quand n augmente, la courbe devient sigmoïdale jusqu'à devenir une fonction en forme d'escalier lorsque n tend vers l'infini. On obtient alors l'équation linéaire par morceau ci-dessous.

$$\frac{d[B]}{dt} = \kappa \cdot S^{+}([A], \theta) - \gamma \cdot [B] \tag{IV.19}$$

où $S^{+}$ est une fonction en escalier et $\theta$ est une valeur seuil de telle manière que $S^{+} = 0$ si $[A] < \theta$ et $S^{+} = 1$ si $[A] > \theta$.

Ce formalisme booléen a été introduit par Leon Glass et Stuart Kauffman (Glass & Kauffman, 1973) dans les années 70. Son principal avantage, c'est le peu de paramètres nécessaires (3 dans l'exemple ci-dessus) et cela a son importance étant donné le peu d'informations quantitatives généralement présentes dans la littérature. Notre laboratoire a participé à l'élaboration d'un logiciel librement accessible nommé Genetic Network Analyzer (GNA) basé sur ce formalisme booléen. Ce logiciel permet d'étudier qualitativement la dynamique de réseaux géniques de plus grosses tailles (de Jong *et al.*, 2003). Il a été utilisé à plusieurs reprises avec succès pour modéliser des réseaux biologiques réels (Ropers *et al.*, 2006).

## 12 Identification de paramètres à l'aide d'un algorithme génétique

Imaginons que l'on ait acquis des données expérimentales montrant la dégradation d'une protéine B au cours du temps (voir la figure IV.8 pour bien comprendre le système). On sait également que la diminution de la concentration de la protéine B est due à l'arrêt de sa synthèse. Cet arrêt est lui-même lié à l'accumulation d'un facteur de transcription A qui réprime la transcription du gène b. La topologie (une répression simple) de notre réseau est donc connue. On peut modéliser la dynamique de fonctionnement de ce réseau à l'aide du système d'équations différentielles ci-dessous.

$$\frac{d[A]}{dt} = \kappa_1 - \gamma_1 \cdot [A]$$

$$\frac{d[B]}{dt} = \kappa_2 \cdot \frac{k_d^n}{[A]^n + k_d^n} - \gamma_2 \cdot [B]$$

Il y a 6 paramètres inconnus : $\kappa_1$,$\gamma_1$, $\gamma_2$, $\kappa_2$, n et $k_d$.
L'objectif sera d'identifier/déterminer ces paramètres ce qui permettra également d'inférer la

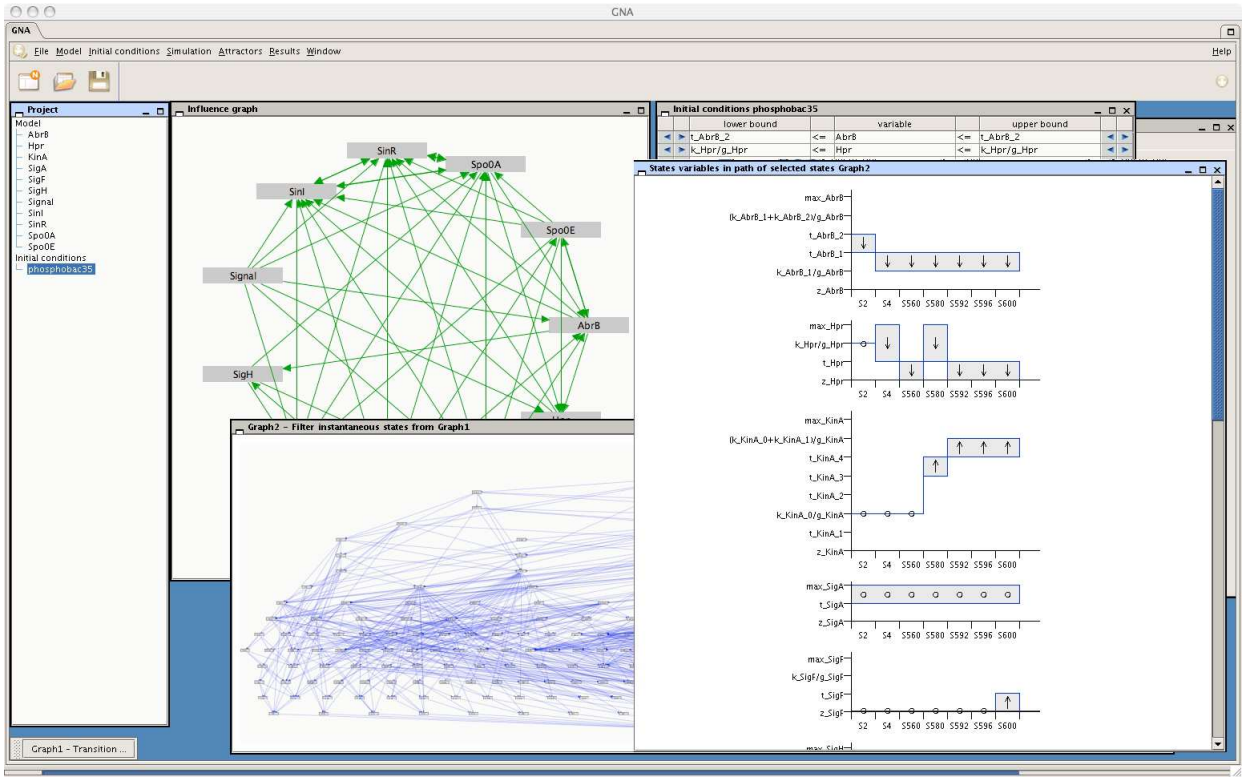



**Figure IV.7** – Interface graphique (GUI) du logiciel Genetic Network Analyzer

concentration du facteur de transcription A au cours du temps.

Plus exactement, mon but sera de vous prouver que l'on peut extraire beaucoup d'informations à partir d'un jeu de données limité. Et pour cela, je vais commencer par créer un faux jeu de données expérimentales à partir du modèle lui-même. Je fixe donc la valeur de 6 paramètres de manière arbitraire : $\kappa_1$=10, $\gamma_1$=1, $\kappa_2$=10, $\gamma_2$=1, $k_d$=7 et n=10 puis je génère les profils des protéines B et A en fonction du temps à l'aide des équations. Pour renforcer leur image de « données expérimentales », j'ai échantillonné ces courbes en 34 mesures temporelles (approximativement 4 points par unité de temps). L'objectif est le suivant : on souhaite écrire un algorithme capable de « retrouver » ces 6 paramètres et retrouver également les données en bleu (la concentration du facteur de transcription A au cours du temps) uniquement à partir des données expérimentales en rouge ainsi que de la connaissance de la topologie du réseau.

Retrouver/rechercher la valeur de ces paramètres consiste à résoudre un problème d'optimisation. Au début, on intègre notre système d'équations différentielles avec des paramètres choisis aléatoirement (souvent dans une fourchette biologiquement raisonnable, par exemple les réels positifs). Puis on compare la courbe proposée par le modèle décrivant [B] en fonction du temps avec le jeu de données expérimentales. Pour cela, on évalue la fonction proposée par le

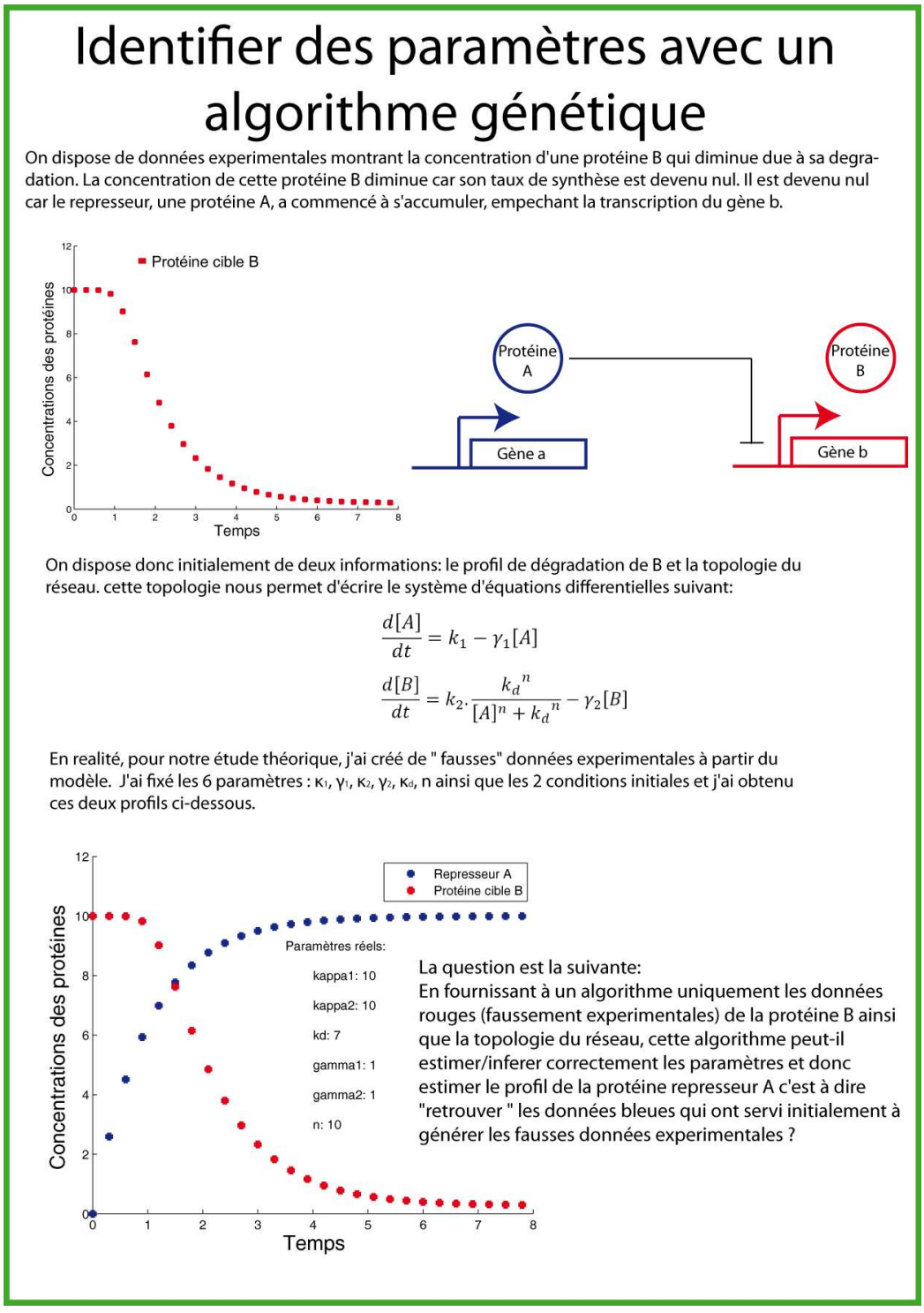



**Figure IV.8** – Identifier plusieurs paramètres à partir de données expérimentales et la connaissance de la topologie du réseau.

modèle aux 34 valeurs (n) de temps correspondant à l'échantillonnage des données expérimentales ($y_i$) ce qui génère un vecteur de même taille ($\hat{y}_i$). Puis on utilise la méthode des moindres carrés.

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{IV.20}$$

L'idée consiste ensuite à minimiser la valeur obtenue S. En effet, plus la valeur de S est petite, plus cela signifie que la courbe du modèle s'ajuste aux données expérimentales.

Un algorithme possible consiste à faire varier aléatoirement (avec la fonction rand de Matlab par exemple) les 6 paramètres que nous recherchons puis sélectionner le jeu de paramètres générant la plus petite valeur de S. Cependant, il est impossible de tester toutes les combinaisons et nous ne pourrons jamais savoir si nous avons vraiment trouvé la valeur minimum de S.

D'autres algorithmes utilisent l'information sur la dérivée pour trouver le minimum. Par exemple, si en augmentant tel paramètre, S diminue (sa dérivé est négative) alors l'algorithme continue à augmenter le paramètre jusqu'à ce que la dérivée devienne nulle. Au contraire, si S augmente (sa dérivé est positive), alors l'algorithme fait varier le paramètre dans l'autre sens. Mais ce type d'algorithme tombe très vite dans un minimum « local », c'est-à-dire un creux du paysage des paramètres qui n'est pas le minimum global recherché même si la dérivée est devenue nulle.

Ce dont nous avons besoin, c'est d'un algorithme qui fouille le « paysage » des paramètres de manière plus efficace qu'une recherche strictement aléatoire et qui est en même temps capable de s'échapper d'un minimum local (figure IV.9). Au fur et à mesure des itérations et en un temps raisonnable, l'algorithme doit nous fournir une solution qui s'approche du minimum global même si elle ne l'atteint pas. Un des algorithmes répondant à ce cahier des charges est l'algorithme génétique.

Les premiers algorithmes génétiques ont été développés par John Holland dans les années 60 (Holland, 1975). Pour faire simple, un algorithme génétique met en compétition une population d'individus ayant chacun un jeu de paramètres. Le fitness des « individus » c'est-à-dire leurs capacités à se reproduire est fonction de la valeur de S. Dans notre cas (minimisation), les individus qui ont une faible valeur de S se reproduisent plus que ceux qui ont une forte valeur de S. A chaque cycle de reproduction, des mutations ont lieu pour fouiller l'espace des paramètres et s'échapper des minima locaux. Des recombinaisons (entre paramètres) (appelées aussi crossing-over) ont également lieu pour tenter de « mettre ensemble » de bons paramètres. Puis une sélection est mise en place pour limiter le nombre d'individus (très vite « computationnellement » couteux) en éliminant les plus faibles (ceux qui ont les valeurs de S les plus

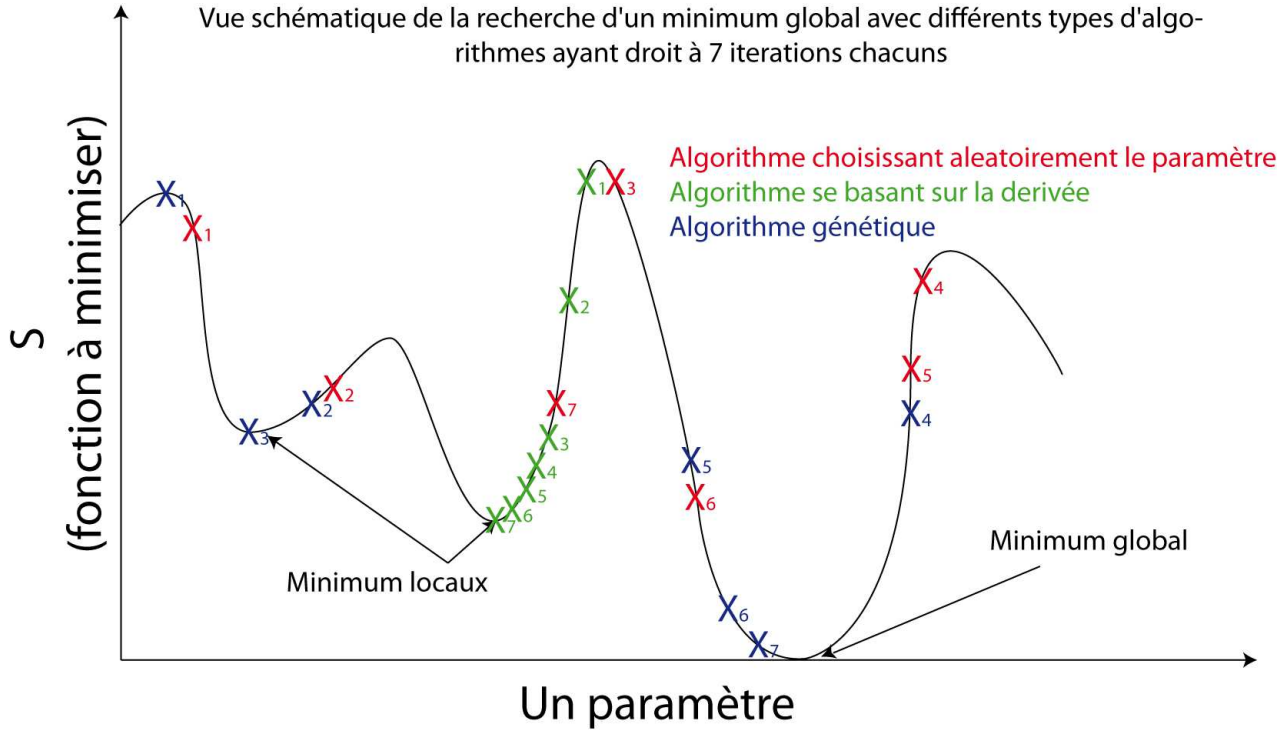



**Figure IV.9** – Schéma de la recherche d'un minimum global

grandes). Au bout de quelques cycles de reproduction, l'algorithme arrive à une solution acceptable mais on ne peut jamais être certain qu'il s'agit du minimum global. Notez également que l'algorithme génétique est un algorithme heuristique[3] : il n'aboutit pas forcement au même résultat quand il est lancé deux fois de suite.

En réalité, comprendre dans les moindres détails le fonctionnement d'un algorithme génétique n'est pas vraiment nécessaire pour un biologiste. Nous utilisons une fonction Matlab nommée « ga » et c'est exactement comme pour le solveur numérique « ode45 » : ces fonctions sont des boites noires dont il n'est pas vraiment nécessaire de connaître totalement le fonctionnement sous jacent. Elles « font le job et elles le font bien ».

Dans le script Matlab, je confie le système des 2 équations différentielles à la fonction ode45 pour qu'elle l'intègre. Cette fonction ode45 est elle-même confiée à l'algorithme génétique (la fonction ga) qui s'occupe de faire varier puis de sélectionner les bons paramètres avec « sa population » cachée dans la boite noire. Puis quand l'algorithme a terminé, il ne nous reste plus qu'à comparer les paramètres réels (faussement expérimentaux) avec les paramètres estimés par le modèle.

Dans l'identification de paramètres visualisable sur la figure IV.11 (premier film) (ou fi-

[3]Heuristique est un terme qui signifie l'art d'inventer, de faire des découvertes.

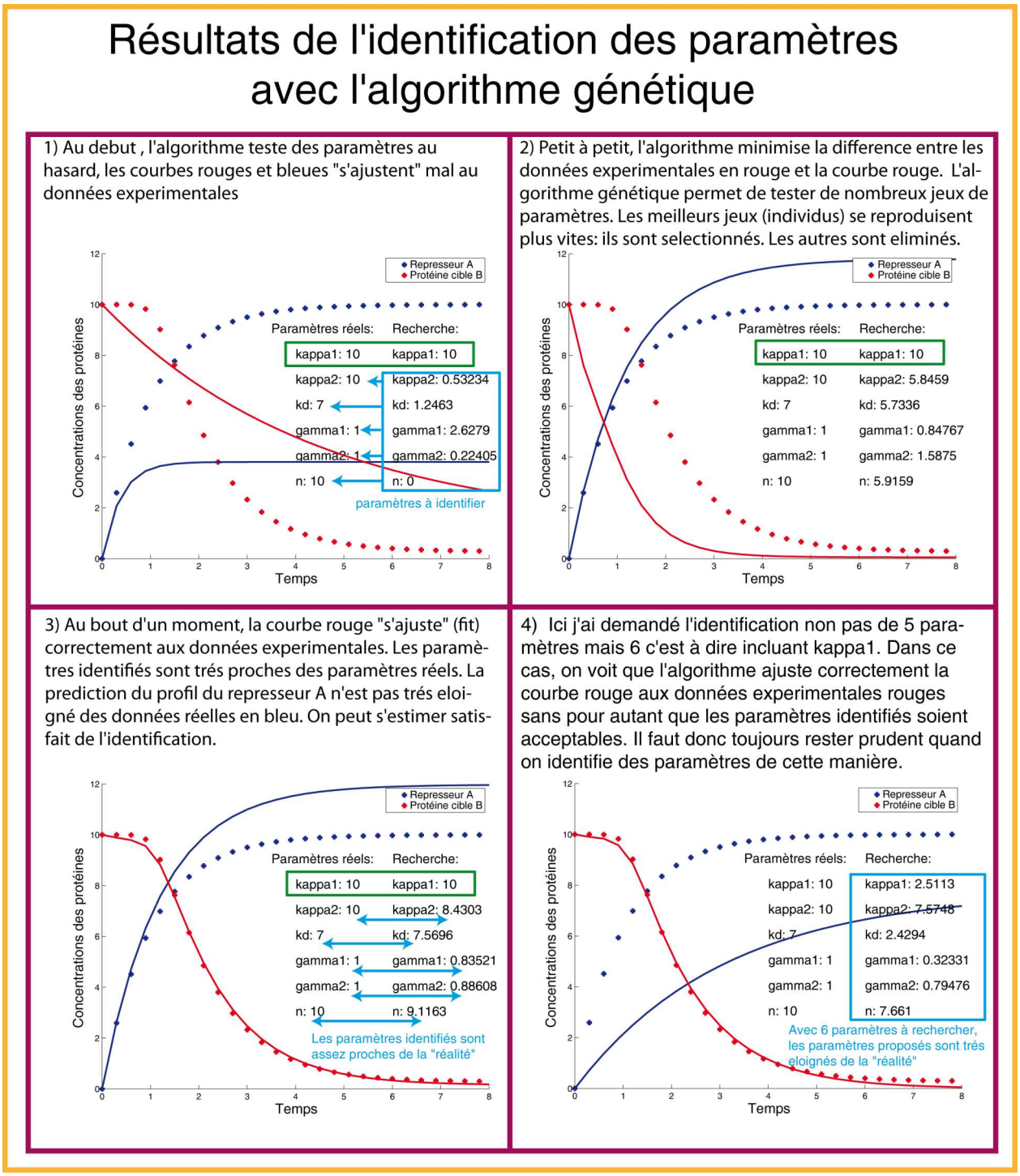



**Figure IV.10** – Résultats de l'identification des paramètres (les vidéos permettent de mieux visualiser le processus).

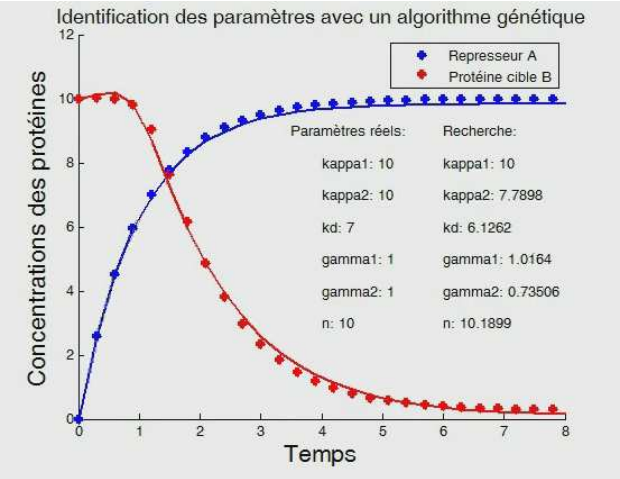



**Figure IV.11** – Vidéo : visualisation de la recherche des paramètres avec un algorithme génétique. Ici, les paramètres proposés sont cohérents. Voir le code Matlab en annexe Code06IdentificationParametre.m qui gènère cette video.

gure IV.10 ; point numéro 3), on recherche 5 paramètres sur les 6 possibles (j'ai indiqué à l'algorithme la valeur de $\kappa_1$). Chacun peut observer que le fit des données expérimentales, les paramètres proposés ainsi que la prédiction d'évolution de [A] en fonction du temps sont loin d'être farfelus. Notre algorithme a prédit des valeurs très proches de la réalité pour les 5 paramètres. Vous avez donc ici une preuve en image que la modélisation peut incontestablement identifier/prédire des paramètres biologiques d'intérêt à partir de données expérimentales très limitées.

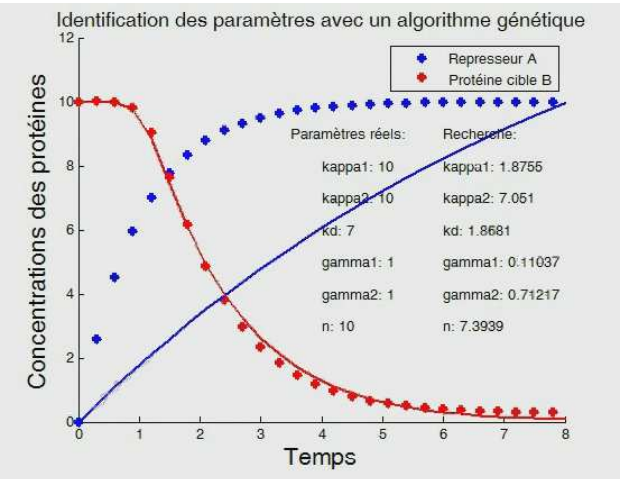



**Figure IV.12** – Vidéo : le risque « d'overfitting ». Quand le nombre de paramètres devient trop grand, on peut « fitter » n'importe quoi sans pour autant que les paramètres proposés soient cohérents. Voir le code Matlab en annexe Code07FitElephant.m qui gènère cette video.

Mais je vais maintenant tempérer mes propos avec un contre-exemple. Il est également tout à fait possible que le modèle propose un ajustement (un fit) aux données expérimentales de bonne qualité alors même que les paramètres proposés sont complètement farfelus (figure IV.10 ; point numéro 4). On utilise souvent la métaphore « du fit de l'éléphant » qui signifie qu'à partir d'un certains nombres de paramètres, on peut ajuster (fitter) n'importe quoi (par exemple un éléphant) sans pour autant que les paramètres proposés correspondent à ce que l'on recherche. Regardez la figure IV.12 (deuxième film), où l'on recherche non plus 5 mais tous les paramètres (6). A la fin, l'ajustement du modèle à la courbe rouge n'est pas si mauvais mais les paramètres identifiés sont très éloignés de la « réalité ». Il faut donc garder un œil critique sur des paramètres identifiés avec un modèle. Vérifications et recoupements sont nécessaires.

# Références bibliographiques


Danchin, A. 2009. Information of the chassis and information of the program in synthetic cells. *Syst Synth Biol*, **3**(1-4), 125–34. 79

de Jong, H., Geiselmann, J., Hernandez, C., & Page, M. 2003. Genetic Network Analyzer : qualitative simulation of genetic regulatory networks. *Bioinformatics*, **19**(3), 336–44. 92

Elf, J., Li, G. W., & Xie, X. S. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**(5828), 1191–4. 83

Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B., & Leibler, S. 1999. Protein mobility in the cytoplasm of Escherichia coli. *J Bacteriol*, **181**(1), 197–203. 82

Glass, L., & Kauffman, S. A. 1973. The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol*, **39**(1), 103–29. 92

Holland, John. 1975. Adaptation in Natural and Artificial Systems. *University of Michigan Press*. 95

Li, G. W., & Elf, J. 2009. Single molecule approaches to transcription factor kinetics in living cells. *FEBS Lett*, **583**(24), 3979–83. 83

Llopis, P. M., Jackson, A. F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., & Jacobs-Wagner, C. 2011. Spatial organization of the flow of genetic information in bacteria. *Nature*. 82

Poon, J., Bailey, M., Winzor, D. J., Davidson, B. E., & Sawyer, W. H. 1997. Effects of molecular crowding on the interaction between DNA and the Escherichia coli regulatory protein TyrR. *Biophys J*, **73**(6), 3257–64. 82

Riggs, A. D., Bourgeois, S., & Cohn, M. 1970. The lac repressor-operator interaction. 3. Kinetic studies. *J Mol Biol*, **53**(3), 401–17. 83

Ropers, D., de Jong, H., Page, M., Schneider, D., & Geiselmann, J. 2006. Qualitative simulation of the carbon starvation response in Escherichia coli. *Biosystems*, **84**(2), 124–52. 92

Slutsky, M., & Mirny, L. A. 2004. Kinetics of protein-DNA interaction : facilitated target location in sequence-dependent potential. *Biophys J*, **87**(6), 4021–35. 83

von Hippel, P. H., & Berg, O. G. 1989. Facilitated target location in biological systems. *J Biol Chem*, **264**(2), 675–8. 83

Zhou, H. X., Rivas, G., & Minton, A. P. 2008. Macromolecular crowding and confinement : biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys*, **37**, 375–97. 82

# Chapitre V

# Les systèmes dynamiques

## 1 Préambule

Dans ce chapitre, qui est la suite directe du chapitre précédent, nous allons étudier la dynamique de fonctionnement de petits réseaux simples.

## 2 Un réseau à deux gènes : l'interrupteur (toggle switch en anglais)

Commençons par étudier le toggle switch : un petit réseau comprenant deux gènes qui se répriment mutuellement par l'intermédiaire des deux protéines correspondantes. L'analyse théorique de ce motif a été effectuée il y a une dizaines d'années (Cherry & Adler, 2000). L'implémentation biologique a été effectuée quasiment au même moment (Gardner *et al.*, 2000).

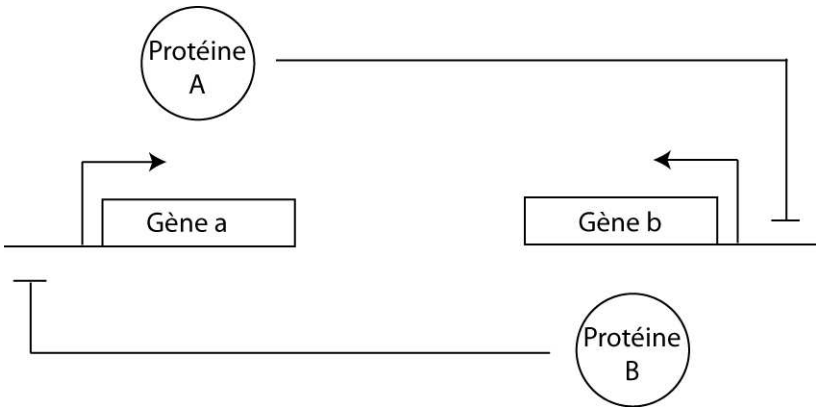



**Figure V.1** – Schéma du toggle switch.

Suivant les conditions initiales, les concentrations des protéines A et B peuvent atteindre deux états d'équilibre différents (et opposés) : on dit que le système est bistable. En effet, si A et B sont respectivement en forte et faible concentration au départ de la cinétique, le système atteint un état d'équilibre avec beaucoup de A et peu de B car la présence de A réprime l'expression de B. A l'inverse, l'état d'équilibre opposé sera atteint si c'est B qui est majoritaire au départ empêchant ainsi l'expression de A (figure V.3a). Ce motif a été appelé « interrupteur » car en perturbant/changeant suffisamment la concentration d'une des protéines (en réalité son affinité pour le promoteur dans l'implémentation biologique), on peut faire « switcher » le système dans l'autre état stable.

Les cinétiques (concentrations en fonction du temps) n'offrent pas toujours la meilleure perspective pour comprendre le comportement d'un système dynamique. Il faut parfois recourir à l'espace des phases c'est-à-dire à un espace abstrait dont les coordonnées sont les variables dynamiques du système. Dans cette espace des phases, on peut tracer les deux « nullclines » c'est-à-dire les valeurs de B en fonction de A tel que d[B]/dt=0. On obtient facilement l'équation suivante :

$$\kappa \cdot \frac{k_d^n}{[A]^n + k_d^n} - \gamma \cdot [B] = 0 \; soit \; [B] = \frac{\kappa}{\gamma} \cdot \frac{k_d^n}{[A]^n + k_d^n}$$

Et réciproquement les valeurs de A en fonction de B tel que d[A]/dt=0 soit

$$[A] = \frac{\kappa}{\gamma} \cdot \frac{k_d^n}{[B]^n + k_d^n}$$

A l'intersection de ces « nullclines », ni la concentration de A ni la concentration de B ne changent. Par conséquent, en ces points fixes, le système est à l'état stable (figure V.3d).

Il peut être aussi utile de tracer un champ de vecteur nous indiquant le taux de changement des variables du système à différents points dans le diagramme de phase (correspondant par exemple à la condition initiale). En prenant le coté gauche de l'équation différentielle (le taux de changement en fonction du temps) de chacune des deux équations du modèle, on construit un vecteur dont la longueur indique le taux auquel le système change d'état et la direction indique où le système est dirigé dans l'espace des phases. Par exemple, si d[B]/dt=4 et d[A]/dt=-2 avec B tracé sur l'axe horizontal et A sur l'axe vertical alors le vecteur pointe en bas à droite comme montré ci-dessous.

Dans la figure V.3d (où le paramètre de coopérativité n est fixé à 3), les « nullclines » se croisent en trois points : il y a trois états stables : deux états stables « stables » où la concentration de A est forte et celle de B est faible (et réciproquement) et un état stable

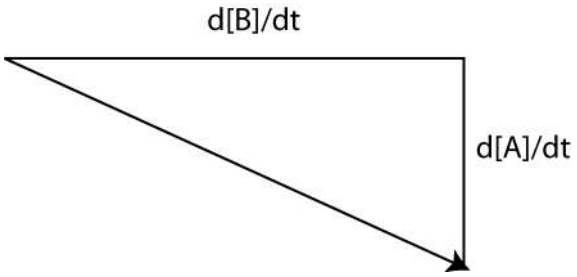



**Figure V.2** – Schéma illustrant la manière dont sont calculés les vecteurs des figures V.3c et V.3d.

« instable » qui correspond à des concentrations de A et B identiques. On observe que beaucoup de vecteurs tendent déjà vers cet état instable puis bifurquent vers l'un ou l'autre des deux autres états stables.

Dans la figure V.3c , le paramètre de coopérativité n est fixé à 1. Dans ce cas, on observe que les « nullclines » se croisent en un point unique : le seul point d'équilibre du système (figure V.3c). Cela a une importante conséquence : cela signifie que le phénotype qu'exhibe un système dynamique (par exemple la bistabilité) dépend fortement de la valeur des paramètres. Ainsi, dans notre exemple, pour une même topologie du réseau (répression mutuelle), le système peut exhiber une dynamique bistable ou monostable en fonction de la valeur des paramètres. Ainsi, pour étudier le comportement d'un système dynamique, il faut aussi travailler dans l'espace des paramètres. Lorsque qu'un certain seuil dans la valeur numérique d'un paramètre est dépassé, le système passe soudainement d'un régime dynamique monostable à un régime dynamique bistable. Cette transition soudaine s'appelle une bifurcation. Dans la figure V.3b, on trace le ou les états stables de la concentration de A en fonction du taux de dégradation et du paramètre de Hill. On voit bien s'opérer la bifurcation entre les deux régimes dynamiques (monostable versus bistable).

## 3 Hystérésis induite par une boucle de rétroaction positive

Etudions maintenant un autre type de réseau bistable : la boucle de rétroaction positive. L'exemple très classique de l'induction bistable de l'operon lactose est connu depuis plus de cinquante ans (Novick & Weiner, 1957; Cohn & Horibata, 1959). C'est un réseau qui est maintenant bien caractérisé mais qui reste encore très étudié aujourd'hui (Robert *et al.*, 2011; Ozbudak *et al.*, 2004; Chang *et al.*, 2009). Le réseau dont nous allons modéliser le fonctionnement est une version simplifiée (figure V.4a) à laquelle j'ai retiré la répression catabolique (le

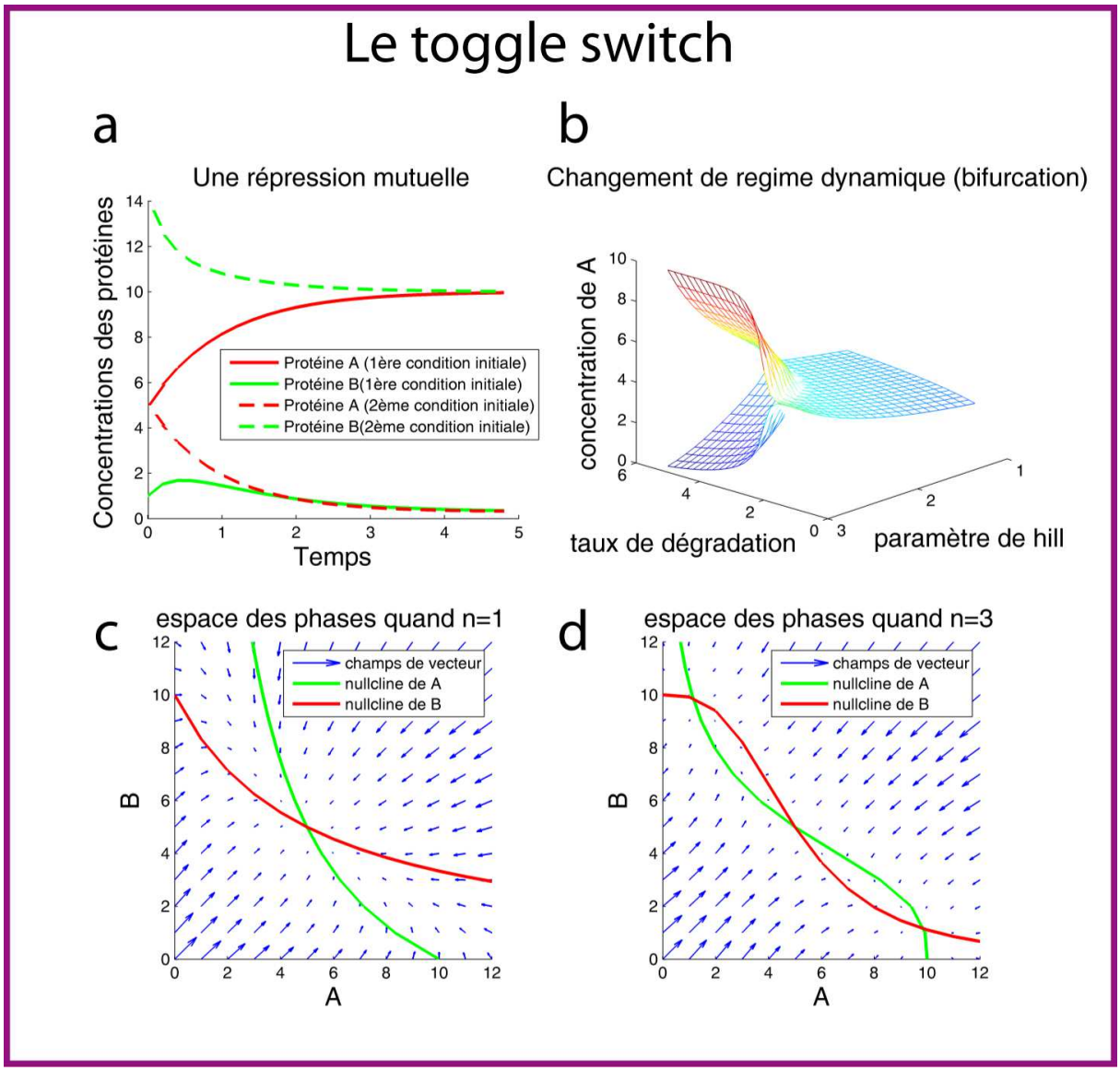



**Figure V.3** – a) dynamique bistable en fonction des conditions initiales. b) bifurcation d'un régime monostable à un régime bistable en fonction de la valeur des paramètres. c) espace des phases avec un seul point stable. d) espace des phases avec trois points stables. Voir le code Matlab en annexe Code08ToggleSwitch.m qui gènère cette figure.

complexe Crp–cAMP). Dans ce réseau, l'expérimentateur contrôle la concentration de l'IPTG dans le milieu (l'input). Cet IPTG pénètre la cellule de manière « transporteur indépendante » et se fixe au répresseur LacI. Ce dernier change alors de conformation ce qui diminue fortement son affinité pour le promoteur de l'operon *lac*. L'absence du répresseur sur le promoteur a pour conséquence la transcription des gènes *lacZ* et *lacY*. Le gène *lacZ* code pour une enzyme (la $\beta$-galactosidase) dont il est facile de mesurer la concentration (voir le chapitre sur les

interactions) : c'est l'output. Le gène *lacY* code pour le transporteur LacY. Une fois dans la membrane, ce dernier augmente fortement le transport de l'IPTG (cette fois de manière transporteur dépendante). Comme l'import d'IPTG augmente, il y a plus de transcription de *lacY* ce qui augmente à nouveau l'import d'IPTG et ainsi de suite. C'est le principe d'une boucle de retro-action positive.

La figure V.4b nous montre l'évolution prédite par le modèle numérique des différentes concentrations des composés en fonction du temps. Au début, jusqu'au temps 7, comme l'import de l'IPTG extracellulaire est limité, la concentration de la $\beta$-galactosidase et du transporteur augmente très lentement. Puis, lorsque la concentration du transporteur dépasse un certain seuil (temps 7), la boucle de rétroaction se met en place et les concentrations d'IPTG, de $\beta$-galactosidase et du transporteur explosent puis atteignent leur état d'équilibre (temps 24). Au temps 25, on imagine que l'expérimentateur enlève l'IPTG du milieu. Concrètement, il centrifuge les cellules, remplace le surnageant par du milieu sans IPTG puis resuspend les cellules dans le nouveau milieu. En absence d'IPTG, le répresseur réprime à nouveau l'operon lac et la transcription s'arrête. La $\beta$-galactosidase et le transporteur LacY se dégradent selon leur taux de dégradation respectif. Au temps 40, il n'y a plus de $\beta$-galactosidase mais il y a encore le transporteur ce qui fait que, contrairement à ce qui se passe à t=0, le nouvel ajout d'IPTG active immédiatement la transcription de l'operon lac.

Et vous allez me dire, oui et alors ? Pourquoi cela est-il si intéressant ? Cela est intéressant car d'un point de vue « informatique », entre T=0 et T=25 on a « écrit » une information et au temps t=40, on a « lu » cette information. La cellule, ici, a joué le rôle de mémoire, elle a stocké un bit d'information. Encore une fois nous retrouvons l'analogie cellule–ordinateur. Rentrons maintenant en peu dans le détail en présentant le phénomène d'hystérésis. L'hystérésis c'est la propriété d'un système qui tend à demeurer dans un certain état quand la cause extérieure qui a produit le changement d'état a cessé. Pour bien illustrer ce phénomène, faisons une expérience de pensée (expérience simulée « *in silico* » dans la figure V.4c). Imaginez que l'on place les cellules « naives ; bit=0 » (c'est-à-dire celles à t=0 qui n'ont jamais connu l'IPTG) dans un milieu et que l'on augmente petit à petit l'IPTG. Lorsque qu'un certain seuil $\theta_1$ est dépassé, la concentration d'IPTG devient suffisante pour activer la boucle de rétroaction et induire l'expression de la $\beta$-galactosidase (courbe verte) : le bit passe de 0 à 1 et *on a « écrit » une information.* Faisons-la même expérience mais cette fois en prenant les cellules « initiées, bit=1 » mais sans $\beta$-galactosidase (cellules à t=40). On augmente petit à petit l'IPTG mais cette fois le seuil $\theta_2$, nécessaire pour induire l'expression de la $\beta$-galactosidase est beaucoup plus faible (courbe rouge) car les cellules avaient déjà le transporteur. Maintenant supposez que l'on soit en possession de cellules dont on ne sait pas si elles sont « naïves, bit=0 » ou « initiées ;

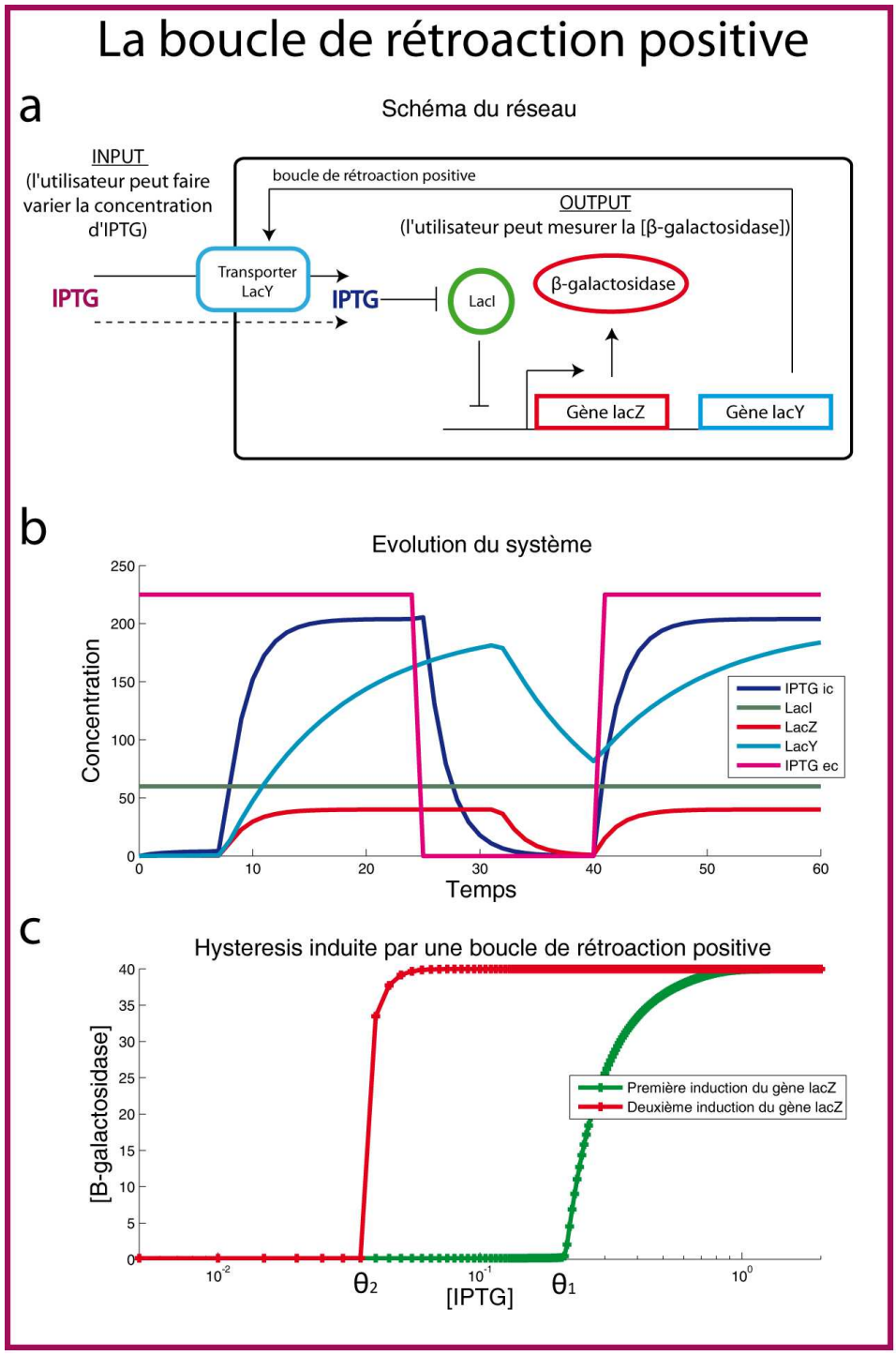



**Figure V.4** – Etude en dynamique de la boucle de retro-action positive. a) schéma du réseau. b) dynamique du système. c) hystérésis et effet mémoire. Voir le code Matlab en annexe Code09Hysteresis.m qui génère cette figure.

bit=1 ». Il suffit de placer les cellules à une concentration d'IPTG comprise entre $\theta_2$ et $\theta_1$. Si les cellules induisent la $\beta$-galactosidase, elles sont « initiées » et le bit était égal à 1. A l'inverse, si elles n'induisent pas la $\beta$-galactosidase, alors elles sont « naïves » et le bit était égal à 0. En se plaçant à une concentration d'IPTG intermédiaire entre $\theta_2$ et $\theta_1$, *on a « lu » une information* stockée en mémoire.

J'aimerai préciser ici que les phénomènes de bistabilité que nous avons étudiés dans ce chapitre sont toujours modélisés de manière déterministe comme si toutes les cellules se comportaient de la même manière. Sachez que la conséquence principale de ces phénomènes de bistabilité s'observe en comparant les cellules entre-elles. Pour un même génotype et dans le même milieu, les bactéries se partageront en deux groupes avec deux phénotypes bien distincts. Mais ne vous inquiétez pas, j'y reviendrai largement lorsque nous aborderons le bruit, les stratégies de pari et la modélisation stochastique. Et nous n'y sommes pas encore.

## 4 Un réseau à trois gènes : le repressilator

La topologie de certains réseaux permet aux concentrations des protéines impliquées de ne jamais atteindre un état stable, d'osciller en permanence au cours du temps. Le « repressilator » est un exemple de ce type de réseau. Il a été implémenté biologiquement (sur un plasmide) en l'an 2000 par Michael Elowitz (Elowitz & Leibler, 2000). Il est composé de 3 gènes qui se répriment mutuellement ce qui permet l'apparition d'oscillations dans la concentration des trois protéines (figure V.5). Plutôt que d'appeler les protéines A, B et C, J'ai préféré conserver le nom de protéines utilisées car elles sont très communes et tout biologiste synthétique y aura à faire un jour ou l'autre. Il s'agit des répresseurs LacI, TetR et CI.

Au regard des résultats de simulation (figure V.7a), on observe bien l'oscillation des 3 protéines en fonction du temps. De plus, on observe une nouvelle fois que la topologie seule n'est pas suffisante pour déterminer le comportement du système : la valeur numérique des paramètres compte. En effet, dans notre simulation, lorsque le paramètre de Hill est fixé à 3, les oscillations perdurent et le système n'atteint jamais un état stable. A l'inverse, lorsque ce paramètre est fixé à 2, les oscillations s'atténuent et le système finit par atteindre un état stationnaire (figure V.7a vs V.7c).

Comme je l'ai déjà dit plus haut, l'esprit a des difficultés à prédire de manière certaine la dynamique de fonctionnement de réseaux très simples. Et il me semble très peu probable que Michael Elowitz ait construit son plasmide « repressilator » sans préalablement avoir vérifié « *in silico* » que la topologie du réseau était bien en mesure de générer des oscillations. Le modèle

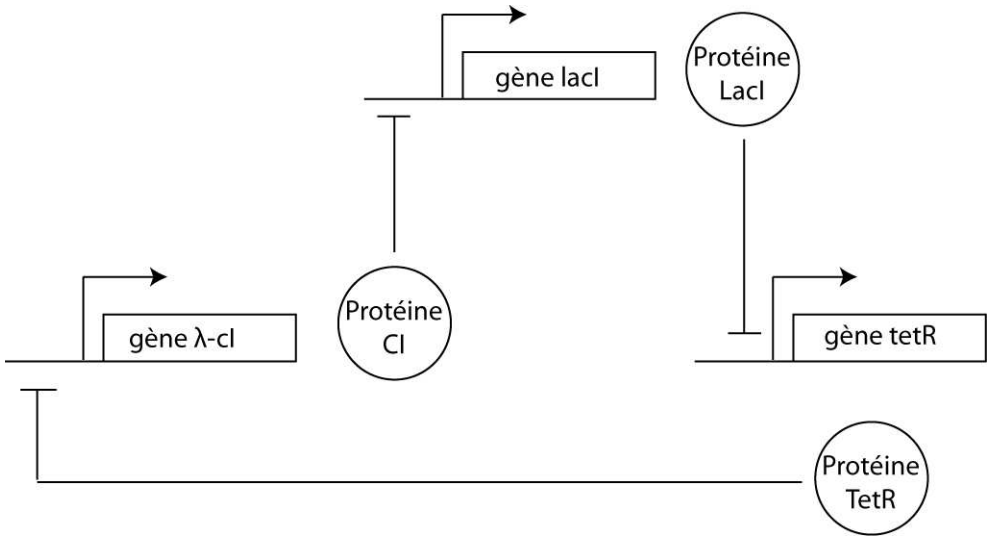



**Figure V.5** – Schéma du repressilator.

peut, par exemple, aider à moduler la période ou l'amplitude des oscillations en indiquant *a priori* quel paramètre modifier. Par exemple, pour diminuer la période des oscillations, on peut diminuer le taux de dégradation des protéines. Pour cela, on clone un tag « ssra » en 3' de la séquence codante de l'ADN. Cela rajoute les acides aminés ANDENYALAA en 3' de la protéine. Cette séquence adresse les protéines vers le proteasome augmentant ainsi leurs taux de dégradation. Avant d'entreprendre des manipulations fastidieuses de biologie moléculaire, il est prudent et facile de vérifier qu'en diminuant le « gamma » dans le modèle, cela diminue bien la période des oscillations.

Petit parenthèse ici puisque l'on vient de voir comment changer gamma. Comment peut-on changer le taux de synthèse Kappa ou la constante de dissociation $k_d$ « pour de vrai » ? Pour changer Kappa, on peut modifier par mutagenèse dirigée un ou plusieurs nucléotides de la « TATA box » (lieu où se fixe l'ARN polymérase) ou du « ribosome binding site (RBS) » (lieu où se fixe le ribosome) pour diminuer respectivement le nombre d'évènements de transcription ou de traduction. Même méthode pour modifier la constante de dissociation : on change par mutagenèse dirigée les nucléotides du site de fixation à l'ADN après avoir bien étudié le motif consensus pour diminuer ou augmenter l'affinité du facteur de transcription pour ce site.

La figure V.6 (video) montre des bactéries contenant la « green fluorescent protein » (GFP) dont la concentration oscille au cours du temps. Ce n'est pas exactement le « repressilator » mais un système dynamique oscillant un peu plus récent et donc plus joli mais à peine plus complexe (Stricker *et al.*, 2008). La vidéo est le fruit du travail de leurs auteurs. Je l'ai ajoutée ici avec leurs autorisations. Je souhaitais absolument l'inclure (et pas seulement mettre un lien

hypertexte) car *elle illustre ce qu'est la biologie synthétique* et surtout ce qu'elle promet. Rendez vous compte, les oscillations de la GFP que vous regardez sur cette video ont été encodées sur un morceau d'ADN par un être humain. Les auteurs peuvent même contrôler la période des oscillations (entre 10 et 40 minutes). Je ne sais pas si vous serez d'accord avec moi mais je trouve que cela rend beaucoup moins futuriste l'idée de faire du calcul avec des cellules et donc avec de la matière « humide » (dite Wetware par opposition aux puces électroniques « sèches » dites Hardware).

Cependant, ce détour par la biologie synthétique ne doit pas vous laisser croire que ces oscillateurs ne sont que des bizarreries sans intérêt créées par et pour l'homme. En effet, des oscillateurs moléculaires existent dans la nature et pas seulement chez les eucaryotes. On en trouve chez certaines archaebactéries (Kondo *et al.*, 1993; Mihalcescu *et al.*, 2004). Ils sont utiles pour l'adaptation du contenu cellulaire au rythme circadien c'est-à-dire à l'alternance du jour et de la nuit.

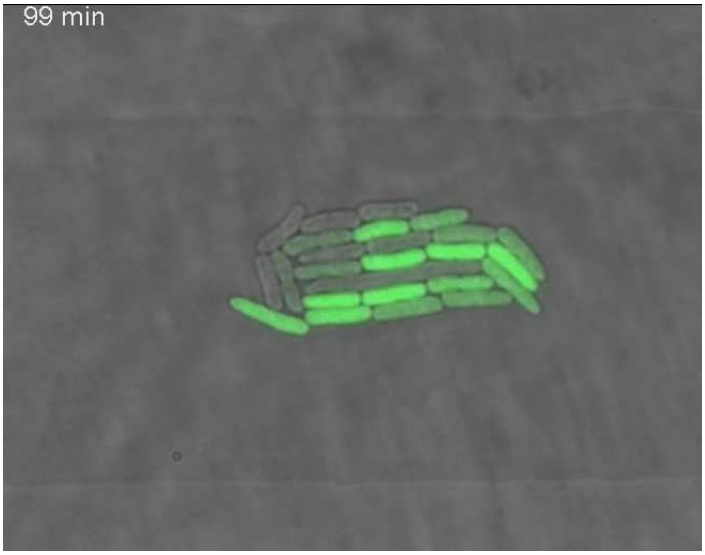



**Figure V.6** – Vidéo : oscillations de la GFP dans des bactéries *E. coli* (La vidéo est publiée avec l'autorisation de Jesse Stricker et Jeff Hasty).

# 5 Caractéristiques des réseaux réels

On doit aux mathématiciens Paul Ërdos et Alfred Rényi les premiers travaux sur les réseaux aléatoires. Ces derniers sont composés de nœuds et de liens. Lorsque les liens sont placés aléatoirement, la plupart des nœuds ont, en moyenne, le même nombre de liens : le système est « démocratique ». La distribution qui représente le nombre de liaisons par nœud a une allure

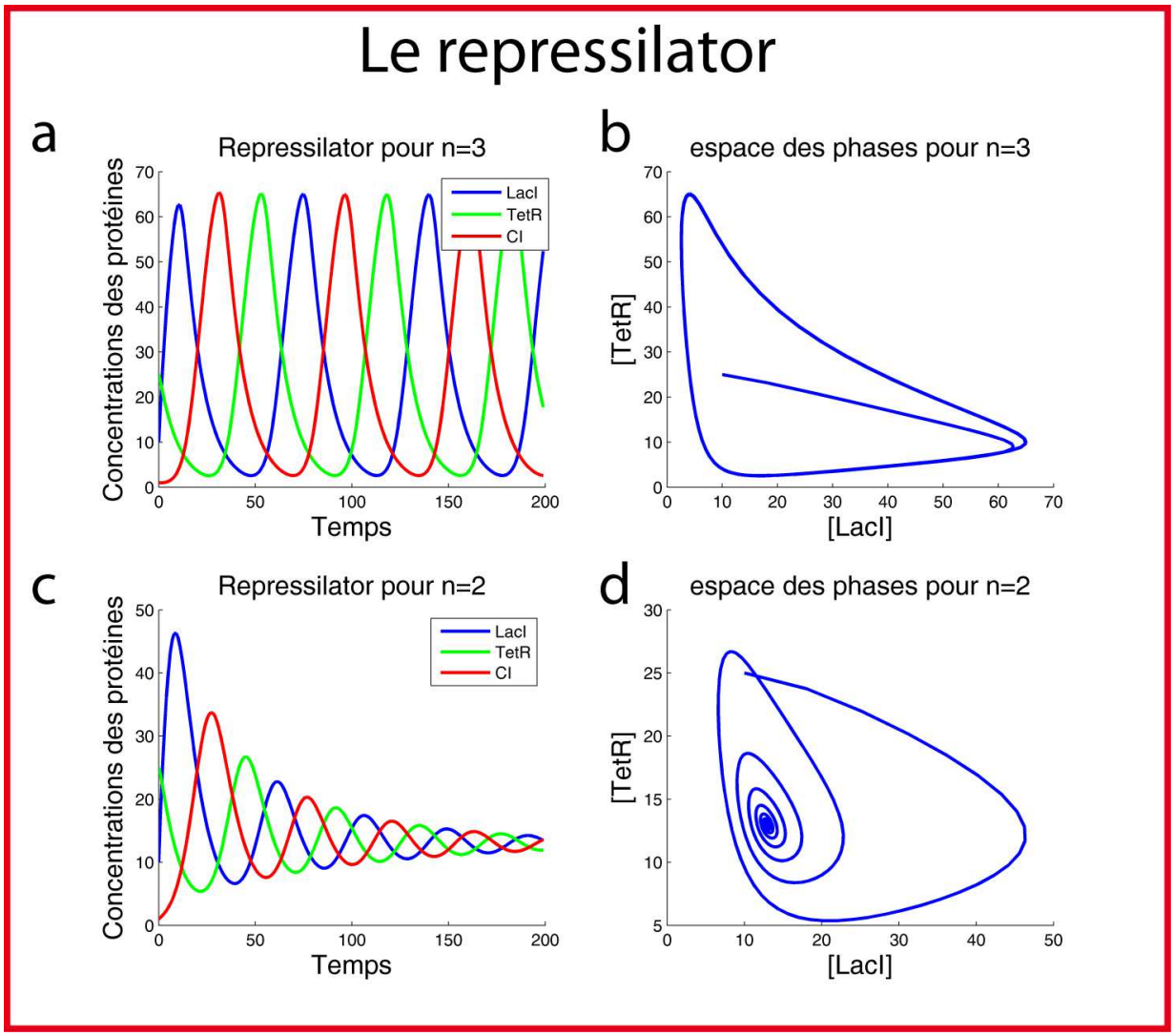



**Figure V.7** – Etude du repressilator. Voir le code Matlab en annexe Code10Oscillation.m qui génère cette figure.

de cloche : c'est une distribution de Poisson (figure V.8). Cela signifie qu'il n'y a quasiment pas de nœuds ayant soit très peu soit énormément de liens (Barabasi & Bonaneau, 2010).

Mais les réseaux des organismes vivants ne sont pas des réseaux aléatoires. Ils possèdent la propriété d'invariance d'échelle (retrouvée aussi avec d'autres types de réseaux comme Internet) (Jeong *et al.*, 2000, 2001; Ravasz *et al.*, 2002). Quelques supernœuds (des « hubs ») possèdent un très grand nombre de liaisons avec d'autres nœuds qui eux, au contraire, sont très peu connectés. Dans le réseau internet, « google » est un de ces supernœuds connecté au site web de tricot de votre grand-mère —site sans doute très intéressant mais probablement peu connecté—. Dans le réseau de régulation génique d'*E. coli*, c'est la protéine Crp qui est

un « hub » connecté à des enzymes moins « populaires ». Contrairement à la distribution de Poisson « démocratique » des réseaux aléatoires, la distribution des réseaux invariants d'échelle suit une loi de puissance. Le nombre de liens par nœud décroit exponentiellement. Ces réseaux présentent certains comportements prévisibles. Par exemple, contrairement aux réseaux aléatoires, ils sont très résistants aux défaillances accidentelles mais extrêmement vulnérables aux attaques ciblées sur les « hubs » (figure V.8). Cependant, l'hypothèse selon laquelle les réseaux de régulation génique suivent une loi de puissance est contestée (Lima-Mendez & van Helden, 2009). Voici quelques arguments des contradicteurs :

⇒ La loi de puissance n'apporterait aucune information nouvelle en biologie et serait donc sans intérêt réel.

⇒ Les « hubs » ne sont pas toujours essentiels : la preuve c'est que le mutant crp est viable alors qu'un très grand nombre d'enzymes (nœuds très peu connectés) sont indispensables à la survie de la bactérie.

⇒ Les données qui sont censées prouver la loi de puissance sont sur-interprétées.

Je rajoute un argument personnel : les réseaux tels qu'ils sont décrits actuellement avec leurs « hubs » (figure V.9a) pourraient représenter plus le pool des expériences réussies par les chercheurs que la réalité biologique. Les « hubs » seraient les exemples/cas d'écoles les plus simples, les zones où les certitudes sont les plus fortes. Un nœud où les chercheurs convergent car les expériences y seraient plus facilement reproductibles. Je m'explique : j'en suis arrivé à travailler sur le « hub » Crp par dérive. J'ai commencé sur des gènes moins connus et dont la régulation est moins comprise. L'absence de résultats, de compréhension et de reproductibilité m'a fait dériver vers d'autres directions, là où les expériences offraient une meilleure reproductibilité. Petit à petit, de gènes en gènes, j'en suis arrivé à Crp. Si les chercheurs qui publient sur Crp ont subi une dérive similaire à la mienne, alors la quantité énorme d'informations accessible sur Crp est plus liée à la facilité/reproductibilité des expériences qu'à une réalité biologique selon laquelle Crp est un « hub ».

> *« C'est une plaisanterie assez courante parmi les chercheurs que de comparer ce qu'ils font, dans leur activité quotidienne, à la recherche de clés perdues en marchant la nuit. Si on n'a pas de lampe de poche, on commencera par les rechercher là où il y a de la lumière, c'est-à-dire sous le lampadaire : pas forcément parce que l'on pense qu'elles ont été perdues à cet endroit précis, mais simplement parce que si c'est le cas, on a une petit chance de les apercevoir »*[1]

[1] Michel Morange, *La vie expliquée*, Odile Jacob p. 158.

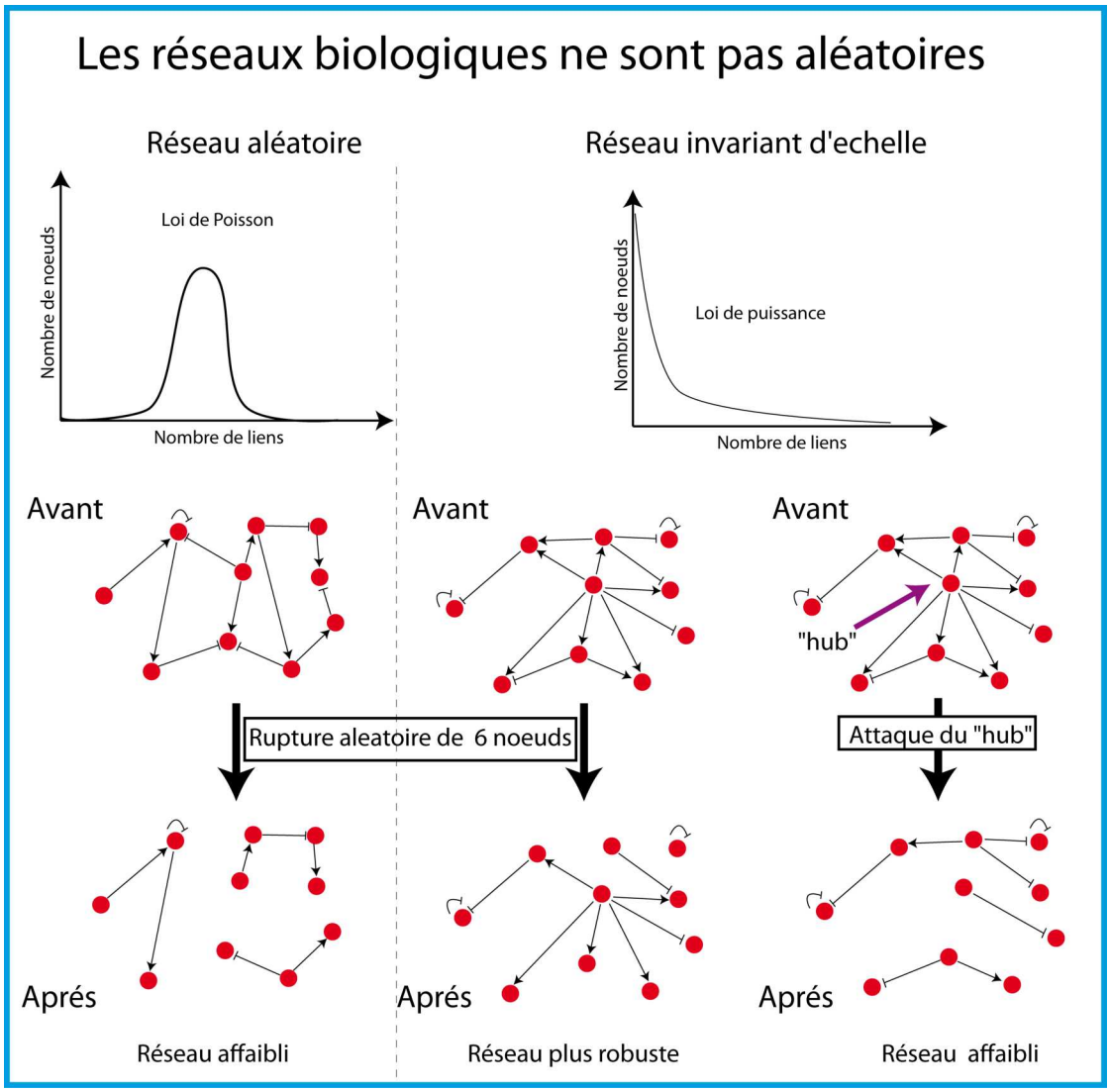



**Figure V.8** – une des caractéristiques des réseaux biologiques est l'invariance d'échelle.

Malgré mon argument ci-dessus, considérons tout de même la figure V.9a comme une représentation globalement fidèle du réseau transcriptionnel réel d'*E. coli*. Nous avons vu que ce réseau n'est pas aléatoire du point de vue de la distribution des liens par nœud. Mais il n'est pas aléatoire également du point de vue de la surreprésentation statistique de certains patterns. Un pattern qui apparait significativement plus souvent dans le réseau réel que dans un réseau aléatoire est appelé motif (Milo *et al.*, 2002; Shen-Orr *et al.*, 2002). La figure V.9b nous montre que l'on retrouve beaucoup plus souvent des boucles de rétroaction négative et des « feed forward loop » (que je renonce pour l'instant à traduire par « boucle d'action avant ») par rapport à un réseau aléatoire. L'idée proposée par l'équipe de Uri Alon (Alon, 2006) c'est que

ces motifs surreprésentés ont dû être préservés au cours de l'évolution contre les mutations qui modifient les flèches de manière aléatoire. En effet pour abolir un site de liaison protéine–ADN, il ne suffit en général que d'une seule mutation (1 seul changement de base/lettre ATCG). Or s'ils ont été préservés, c'est qu'ils ont été sélectionnés pour un avantage qu'ils apportent à l'organisme. La topologie du motif, c'est-à-dire sa structure aurait une fonction. La fonction *a priori* évidente d'une boucle de rétroaction négative (qui agit comme un thermostat), c'est de préserver l'homéostasie de l'organisme. A l'inverse, une boucle de rétroaction positive peut aider, chez les organismes supérieurs, à la différentiation irréversible des cellules dans une voie ou une autre. Mais il existe aussi des fonctions plus subtiles, impossibles à comprendre sans l'utilisation d'un modèle mathématique. C'est ce que nous allons étudier maintenant.

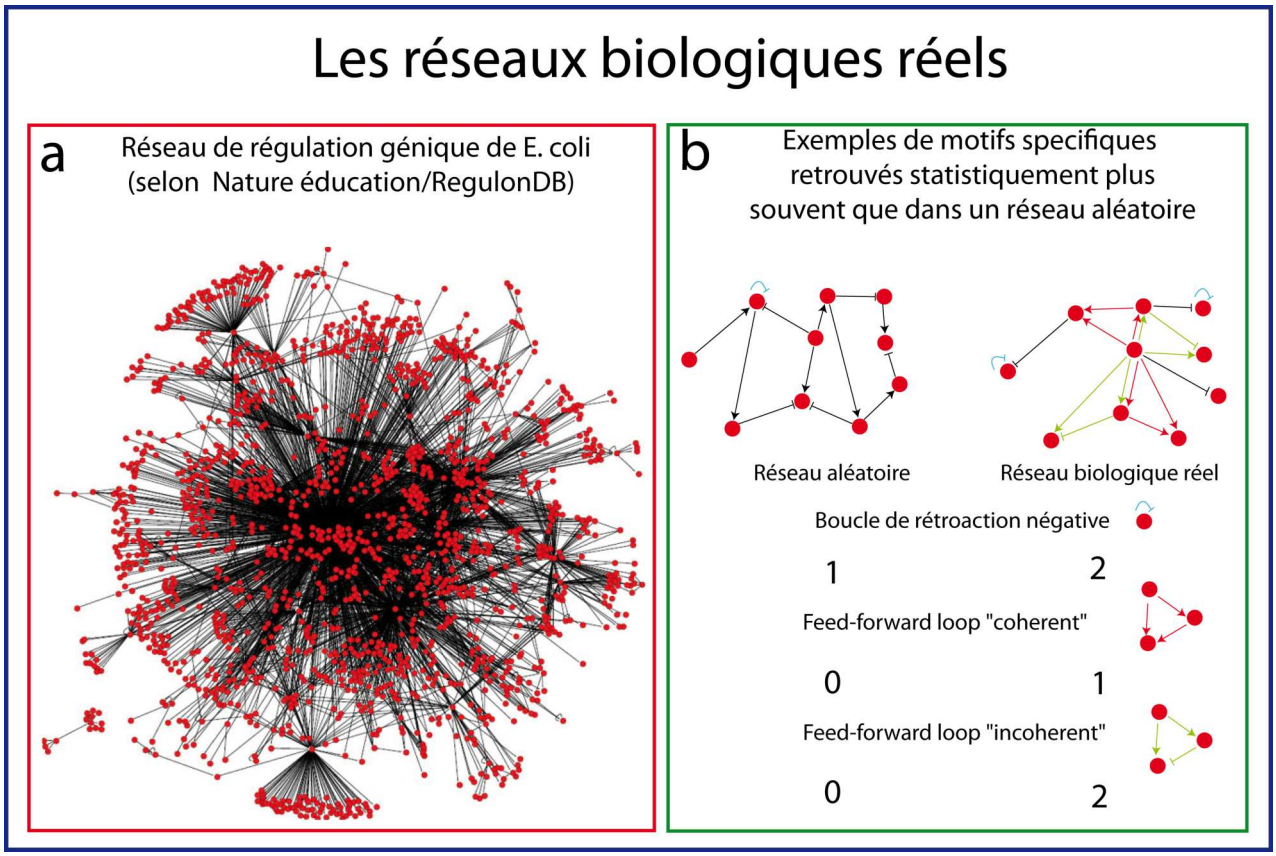



**Figure V.9** – Les réseaux biologiques réels.

# 6 Fonction de quelques motifs courants

## 6.1 La boucle de rétroaction négative

Recherchons une des fonctions de la boucle de rétroaction négative grâce à un petit modèle ODE (figure V.10a). On observe que la courbe rouge (avec boucle de rétroaction négative) atteint l'état d'équilibre plus vite que la courbe bleue (sans boucle de rétroaction négative). Cela semble contre-intuitif mais notez bien que pour obtenir ce résultat, le taux de synthèse (force du promoteur) de l'équation avec boucle de rétroaction est choisi 5 fois plus grand que celui de l'équation sans boucle de rétroaction. Vous voyez donc que, pour un même état d'équilibre, la boucle de rétroaction négative accélère la réponse à un signal donné. Cela a également été démontré expérimentalement (Rosenfeld *et al.*, 2002) dans le contexte du système de réparation de l'ADN (SOS) dont j'ai déjà parlé dans le chapitre sur *E. coli* .

## 6.2 Le « feed forward loop » cohérent (type 1 avec un porte AND)

Ce motif peut agir comme « détecteur de persistance » (Mangan & Alon, 2003). Il a été démontré expérimentalement avec le réseau d'utilisation de l'arabinose (Mangan *et al.*, 2003). Dans ce système, un pulse d'AMPc active la protéine Crp qui active alors l'expression de deux gènes : le gène *araC* et l'operon *araBAD*. Ce dernier permet l'utilisation de l'arabinose comme source de carbone. Le gène *araC* code pour la protéine AraC qui, elle aussi, active, en présence d'arabinose, l'expression de l'operon *araBAD*. Au final, pour activer *araBAD*, il faut à la fois Crp et AraC : c'est une porte logique AND. Le « feed forward loop » est dit « cohérent » car la protéine Crp agit (de notre point de vue) de manière cohérente en activant un élément intermédiaire *lui aussi* chargé d'activer la cible finale. En choisissant judicieusement les paramètres, on peut montrer avec un petit système ODE pourquoi ce réseau peut agir comme détecteur de persistance.

Mais d'abord qu'est ce qu'un détecteur de persistance ? Imaginez une camera dans une gare couplée à un algorithme chargé de détecter automatiquement des objets abandonnés, par exemple, une valise. Dans la gare, des valises posées quelques secondes/minutes à coté de leur propriétaire, il y en a plein. Ce qu'il faut à l'algorithme c'est un filtre qui lui indique les persistances c'est-à-dire les valises qui n'ont pas bougé depuis plus de 5 minutes, par exemple. Cela permet de déclencher une alarme que lorsque cela est vraiment nécessaire. Chez *E. coli*, c'est la même chose. Des petites variations d'AMPc, des petits pulses, il y en a potentiellement beaucoup. Cela représente du bruit et il ne faut pas activer l'operon *araBAD* dans ce cas. Ces variations d'activité de Crp sont insuffisantes pour que la concentration d'AraC dépasse un seuil

suffisant pour l'activation de l'operon *araBAD*. Au contraire, ce que la bactérie veut détecter, c'est une concentration d'AMPc qui augmente et se maintient suffisamment longtemps. Quand tel est le cas, la concentration de la protéine AraC finit par dépasser un certain seuil. La présence conjointe de Crp et d'AraC permet à la porte AND de fonctionner et d'activer l'operon *araBAD* (figure V.10b).

### 6.3 Le « feed forward loop » incohérent (type 1 avec un porte AND)

Ce motif peut agir comme un « générateur de pulse » (Mangan & Alon, 2003). Il a été démontré expérimentalement avec un petit système synthétique multicellulaire (Basu *et al.*, 2005). Voyons quelle est la dynamique de ce motif avec le modèle ODE de la figure V.10c. Un signal active la protéine A. Celle-ci active l'expression des protéine B et C. Mais quand la protéine B commence à s'accumuler, elle réprime l'expression de C, d'où le pulse observé.

### 6.4 Les critiques

3 critiques s'opposent à l'idée que la topologie (la structure) d'un réseau (d'un motif) détermine la fonction par le biais de la sélection (évolution) :

⇒ Comme nous l'avons vu, à une topologie ne correspond pas nécessairement une seule fonction. Les valeurs numériques des paramètres comptent et en les faisant varier, on peut passer par plusieurs régimes dynamiques possibles (Voigt *et al.*, 2005).

⇒ Le motif est toujours étudié de manière isolée or les connexions du motif au reste du réseau sont trop importantes/influentes pour que l'on puisse inférer sa fonction à partir de la seule structure locale (Knabe *et al.*, 2008).

⇒ Le fait que certains motifs soient statistiquement plus nombreux dans un réseau biologique réel par rapport à un réseau aléatoire n'est pas dû *à la sélection du motif pour sa fonction* mais plutôt à *la topologie « invariante d'échelle » du réseau biologique réel*. Comme nous l'avons vu, cette topologie crée des « hubs » et ces derniers favorisent l'apparition de « feed forward loop » par rapport à un réseau strictement aléatoire (Konagurthu & Lesk, 2008).

## 7 Etudier la dynamique de réseaux plus gros

Nous avons, pour l'instant, étudié la dynamique de petits motifs réels trouvés chez *E. coli* ou des petits réseaux synthétique crées par l'homme. Mais ces réseaux n'ont jamais dépassé 3

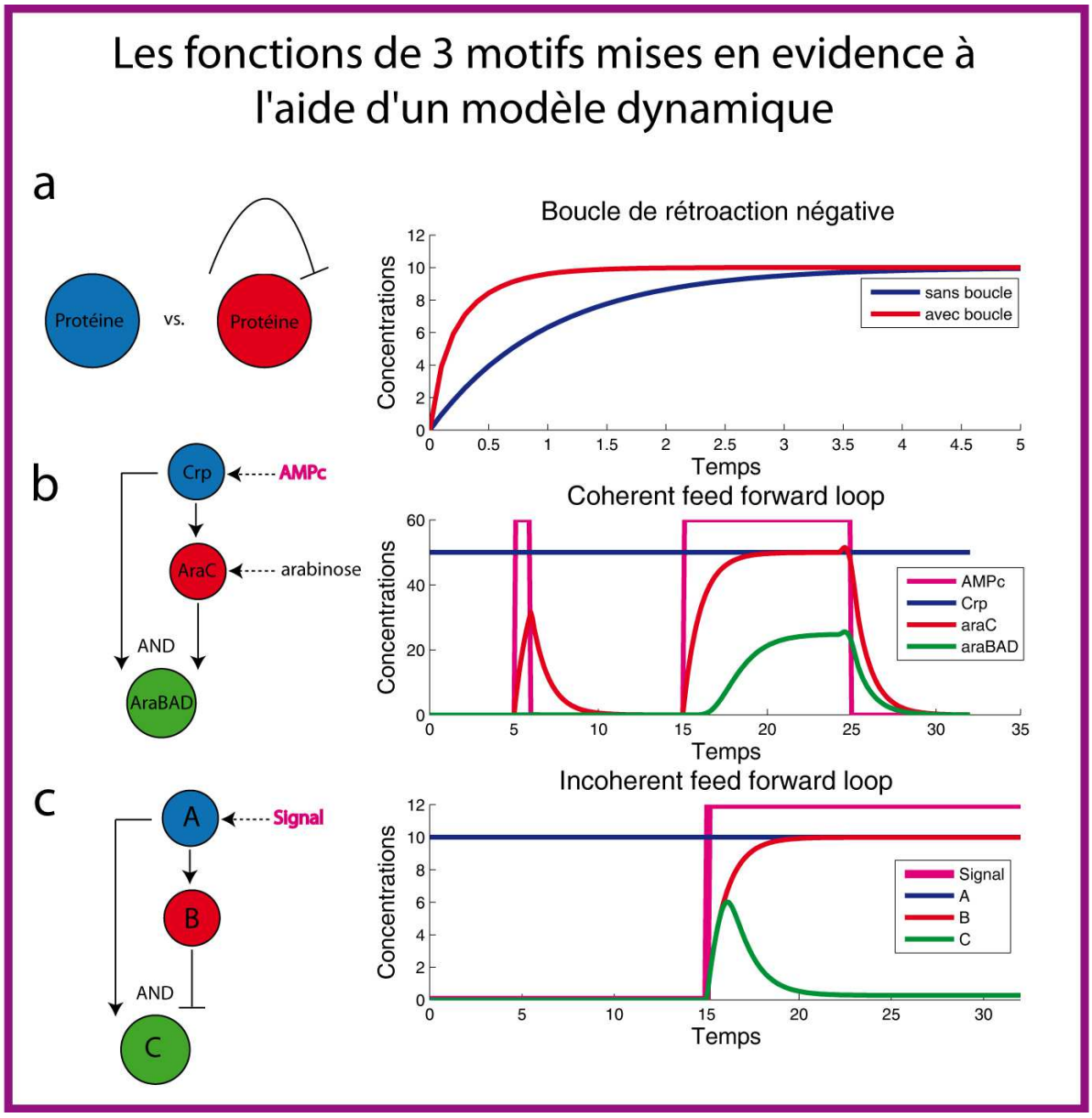



**Figure V.10** – Fonctions de quelques motifs. a) la boucle de rétroaction négative accélère la réponse à un signal donné. b) Le « feed forward loop » cohérent agit comme un détecteur de persistance. c) Le « feed forward loop » incohérent agit comme un générateur de pulse. Voir le code Matlab en annexe Code11MotifAlon.m qui gènère cette figure.

gènes. Or le génome d'*E. coli* compte 4000 gènes. Peut-on passer à l'étape suivante et simuler la dynamique de fonctionnement de réseaux plus gros contenant 5,10, 25 ,200 puis 4000 gènes ?

Si on met de coté les erreurs dues à l'utilisation de méthodes numériques, le formalisme ODE étudié dans ce chapitre n'a pas de limites intrinsèques et pourrait donc, en théorie, simuler un réseau de 4000 gènes sans difficulté. Pour un réseau très linéaire sans boucle de rétroaction,

on peut augmenter le nombre de gènes sans trop diminuer la qualité des prédictions effectuées. Mais dés lors que le système devient non linéaire (boucles de rétroaction) ou que la connectivité (nombre de liens par nœud) augmente trop, les prédictions deviennent fortement dépendantes des conditions initiales et de la valeur numérique des paramètres. C'est ce qu'on appelle la dépendance sensitive aux conditions initiales (le populaire effet papillon). Pour proposer un exemple simple, une toute petite différence dans la valeur d'un paramètre (15,000007 par rapport à 15,000008) aura des conséquences énormes sur le phénotype final : la cellule vit ou la cellule meurt. Pour couronner le tout, dites vous que ce problème existe dans un système déterministe et il est évidemment amplifié dans un système stochastique (c'est le cas de la cellule) ce qui accroit encore les incertitudes sur le phénotype final. Or malgré son réseau non linéaire au fonctionnement stochastique, la cellule réussit à se prémunir de cette dépendance sensitive aux conditions initiales de manière évidente. En effet, lors de son adaptation à divers stimuli, la population de cellules, via son réseau de régulation, semble réussir à basculer vers le phénotype adéquate en calculant la solution optimale et ce de manière reproductible et donc un minimum déterministe. Comment fait-elle ?

En électronique, un module est défini comme une unité fonctionnelle qui est capable de maintenir ses propriétés intrinsèques indépendamment de ses connections externes. C'est un concept important car il permet aux ingénieurs de connecter plusieurs éléments ensembles sur une puce tout en maintenant la capacité de prédire le comportement de l'ensemble (le résultat après assemblage). Beaucoup de biologistes pensent que pour être suffisamment robuste, le réseau d'un organisme vivant doit être composé de modules c'est-à-dire être *modulaire* (Hartwell *et al.*, 1999; Wolf & Arkin, 2003; Kashtan & Alon, 2005). Selon Gunter Wagner, la question n'est même plus tant de savoir si les réseaux sont modulaires, mais plutôt quelle est l'origine de cette modularité (sélection naturelle ?) et quelles sont ses conséquences sur « l' évolvabilité » d'un organisme (Wagner *et al.*, 2007).

Un des enjeux futurs de la biologie synthétique sera de construire ces réseaux modulaires dont on contrôle le comportement d'ensemble. Pour cela, il faudra assembler des petits modules et utiliser le savoir faire de l'électronique pour minimiser les phénomènes de rétroactivité (Del Vecchio *et al.*, 2008; Sauro, 2008). Réussir à connecter ensembles quelques petits modules de 2 à 4 gènes et démontrer qu'on contrôle le fonctionnement global sera déjà une belle avancée. Pour cela, il faut selon moi, limiter la taille des modèles dynamiques (moins de 10–15 gènes). En effet, ces derniers doivent rester prédictifs de manière à permettre à l'expérimentateur d'utiliser ces prédictions pour ajuster les paramètres via les techniques de biologies moléculaires.

Il me semble que les novices ont tendance à sous-estimer la difficulté technique que représente le contrôle quantitatif d'une seule interaction. Le passage du contrôle d'un réseau de n

interactions à n+1 peut prendre énormément de temps (mois, années). Le repressilator que nous avons étudié contient 3 gènes et 3 interactions et c'est un exploit technique qui prendrait des années à être reproduit par une équipe d'aujourd'hui.

En biologie synthétique, je pense qu'il faut privilégier les petits modèles (simples mais pas simplistes). Les gros modèles contenant des dizaines/centaines de gènes et métabolites sont beaucoup trop gros pour être réellement prédictifs au sens où je l'entends. Ces modèles ne peuvent pas me fournir une prédiction fiable m'indiquant comment modifier certains paramètres de mon réseau pour adapter son comportement dynamique. Or la sureté des prédictions est primordiale car la modification de la séquence ADN pour ajuster un paramètre (biologie moléculaire) est très fastidieuse et peut prendre des mois. Pourtant et paradoxalement, les prédictions générées par ces gros modèles sont souvent justes. La raison est simple : le chercheur connaît son but et finit donc évidemment par l'atteindre. Comme le souligne Antoine Danchin, beaucoup de modèles redécouvrent le cycle de Krebs.

Pour faire de la biologie synthétique bottum-up, il faut vérifier et contrôler expérimentalement chaque interaction. Cela peut prendre du temps et se fait parfois dans la douleur mais c'est indispensable. Les interactions données sur les bases de données comme Ecocyc ou RegulonDB devraient être considérées comme fausses *a priori* jusqu'à confirmation expérimentale. Comprenez-moi bien, ce n'est pas une critique envers ces bases de données, ni envers le travail des chercheurs qui ont découverts ces interactions. C'est le constat qu'on se trouve ici à la frontière de la connaissance et cette connaissance reste relative et n'attend qu'à être amélioré. En biologie, ce qui est vrai dans un laboratoire sera faux dans celui d'à coté. Parfois, un chercheur qui décrit/propose une nouvelle interaction va, lui-même, ajouter du conditionnel dans sa publication initiale. Puis au fil des années et des publications citant cette interaction, le conditionnel disparait et petit à petit l'interaction devient sacrée. Pourquoi, dans les bases de données, n'y a t'il pas de chiffres, même indicatifs, associés à une interaction ? Par exemple, dans tel condition, ce facteur de transcription active d'un facteur 100 l'expression de ce gène ou au contraire, il ne l'active que de 2%. Il n'y a pas de chiffres car la résolution actuelle de nos expériences ne nous permet pas d'atteindre un tel niveau de standardisation. On est très loin de l'électronique.

Je donne un petit peu mes recettes de cuisine (qui n'engagent que moi) car il me semble que les novices ont parfois tendance à surestimer les informations de topologie. En biologie synthétique, ne faites pas confiance à une information *a priori*. Vérifiez là toujours par vous-même et rappelez vous que ce qui est facile en théorie est toujours difficile en pratique et ce qui est difficile en théorie est quasiment impossible en pratique.

# Références bibliographiques


Alon, Uri. 2006. *An introduction to system biology*. Chapman and Hall/CRC. 114

Barabasi, Albert, & Bonaneau, Eric. 2010. Les reseaux invariants d'echelle. *Pour la science*. 112

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature*, **434**(7037), 1130–4. 117

Chang, D. E., Leung, S., Atkinson, M. R., Reifler, A., Forger, D., & Ninfa, A. J. 2009. Building biological memory by linking positive feedback loops. *Proc Natl Acad Sci U S A*. 105

Cherry, J. L., & Adler, F. R. 2000. How to make a biological switch. *J Theor Biol*, **203**(2), 117–33. 103

Cohn, M., & Horibata, K. 1959. Inhibition by glucose of the induced synthesis of the beta-galactoside-enzyme system of Escherichia coli. Analysis of maintenance. *J Bacteriol*, **78**, 601–12. 105

Del Vecchio, D., Ninfa, A. J., & Sontag, E. D. 2008. Modular cell biology : retroactivity and insulation. *Mol Syst Biol*, **4**, 161. 119

Elowitz, M. B., & Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**(6767), 335–8. 109

Gardner, T. S., Cantor, C. R., & Collins, J. J. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**(6767), 339–42. 103

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. 1999. From molecular to modular cell biology. *Nature*, **402**(6761 Suppl), C47–52. 119

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabasi, A. L. 2000. The large-scale organization of metabolic networks. *Nature*, **407**(6804), 651–4. 112

Jeong, H., Mason, S. P., Barabasi, A. L., & Oltvai, Z. N. 2001. Lethality and centrality in protein networks. *Nature*, **411**(6833), 41–2. 112

Kashtan, N., & Alon, U. 2005. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A*, **102**(39), 13773–8. 119

Knabe, J. F., Nehaniv, C. L., & Schilstra, M. J. 2008. Do motifs reflect evolved function ?–No convergent evolution of genetic regulatory network subgraph topologies. *Biosystems*, **94**(1-2), 68–74. 117

Konagurthu, A. S., & Lesk, A. M. 2008. On the origin of distribution patterns of motifs in biological networks. *BMC Syst Biol*, **2**, 73. 117

Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., & Johnson, C. H. 1993. Circadian rhythms in prokaryotes : luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A*, **90**(12), 5672–6. 111

Lima-Mendez, G., & van Helden, J. 2009. The powerful law of the power law and other myths in network biology. *Mol Biosyst*, **5**(12), 1482–93. 113

Mangan, S., & Alon, U. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*, **100**(21), 11980–5. 116, 117

Mangan, S., Zaslaver, A., & Alon, U. 2003. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol*, **334**(2), 197–204. 116

Mihalcescu, I., Hsing, W., & Leibler, S. 2004. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*, **430**(6995), 81–5. 111

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. 2002. Network motifs : simple building blocks of complex networks. *Science*, **298**(5594), 824–7. 114

Novick, A., & Weiner, M. 1957. Enzyme Induction as an All-or-None Phenomenon. *Proc Natl Acad Sci U S A*, **43**(7), 553–66. 105

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., & Van Oudenaarden, A. 2004. Multistability in the lactose utilization network of Escherichia coli. *Nature*, **427**(6976), 737–40. 105

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabasi, A. L. 2002. Hierarchical organization of modularity in metabolic networks. *Science*, **297**(5586), 1551–5. 112

Robert, L., Paul, G., Chen, Y., Taddei, F., Baigl, D., & Lindner, A. B. 2011. Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch. *Mol Syst Biol*, **6**, 357. 105

Rosenfeld, N., Elowitz, M. B., & Alon, U. 2002. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, **323**(5), 785–93. 116

Sauro, H. M. 2008. Modularity defined. *Mol Syst Biol*, **4**, 166. 119

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet*, **31**(1), 64–8. 114

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., & Hasty, J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature*, **456**(7221), 516–9. 110

Voigt, C. A., Wolf, D. M., & Arkin, A. P. 2005. The Bacillus subtilis sin operon : an evolvable network motif. *Genetics*, **169**(3), 1187–202. 117

Wagner, G. P., Pavlicev, M., & Cheverud, J. M. 2007. The road to modularity. *Nat Rev Genet*, **8**(12), 921–31. 119

Wolf, D. M., & Arkin, A. P. 2003. Motifs, modules and games in bacteria. *Curr Opin Microbiol*, **6**(2), 125–34. 119

# Chapitre VI

# Le bruit

## 1 Les origines du bruit

Deux jumeaux (ou deux clones de brebis) diffèrent en apparence et en comportement. Bien qu'ayant exactement le même génome, ils vont faire des choix différents durant leur vie : ils conservent heureusement leur libre arbitre. Les choses fonctionnent de la même manière dans une population de cellules clonales (ayant le même génome). En effet, des cellules identiques génétiquement exposées au même environnement/milieu peuvent cependant montrer des variations significatives dans leur contenu cellulaire (protéine, ARN, métabolite etc...) et donc montrer des différences sur le plan phénotypique : 1 génome, 1 milieu, plusieurs phénotypes. Cette variabilité est liée à la nature stochastique de l'expression des gènes. Cette dernière est elle-même due au déplacement aléatoire des molécules selon le mouvement brownien ce qui injecte des probabilités dans les phénomènes observés. Ainsi la variation stochastique dans l'expression des gènes est appelée bruit et c'est un déterminant clé des variations phénotypiques observées entre cellules clonales. Quand une molécule X est présente en très nombreux exemplaires dans la cellule, les effets de X sont facilement et correctement prédits par un modèle déterministe qui représente le comportement moyen de l'ensemble des exemplaires de X. A l'inverse quand la molécule X n'est présente qu'en quelques exemplaires, les effets stochastiques/le bruit deviennent prédominants.

Prenons le cas d'une protéine présente en 10 exemplaires dans une cellule mère[1]. Lors de la division, chaque cellule fille héritera en moyenne de 5 exemplaires de cette protéine. Mais en réalité, la répartition n'est pas une simple division par deux : elle suit une loi de Gauss (forme de cloche centrée sur la moyenne). Autour de cette moyenne de 5 exemplaires se produiront des situations très diverses et il y a une certaine probabilité (faible) qu'une cellule fille hérite des 10 exemplaires et que l'autre n'en obtienne aucun. Si on place alors judicieusement une boucle

[1]La plupart des protéines d'une cellule sont présentes en moins de dix exemplaires.

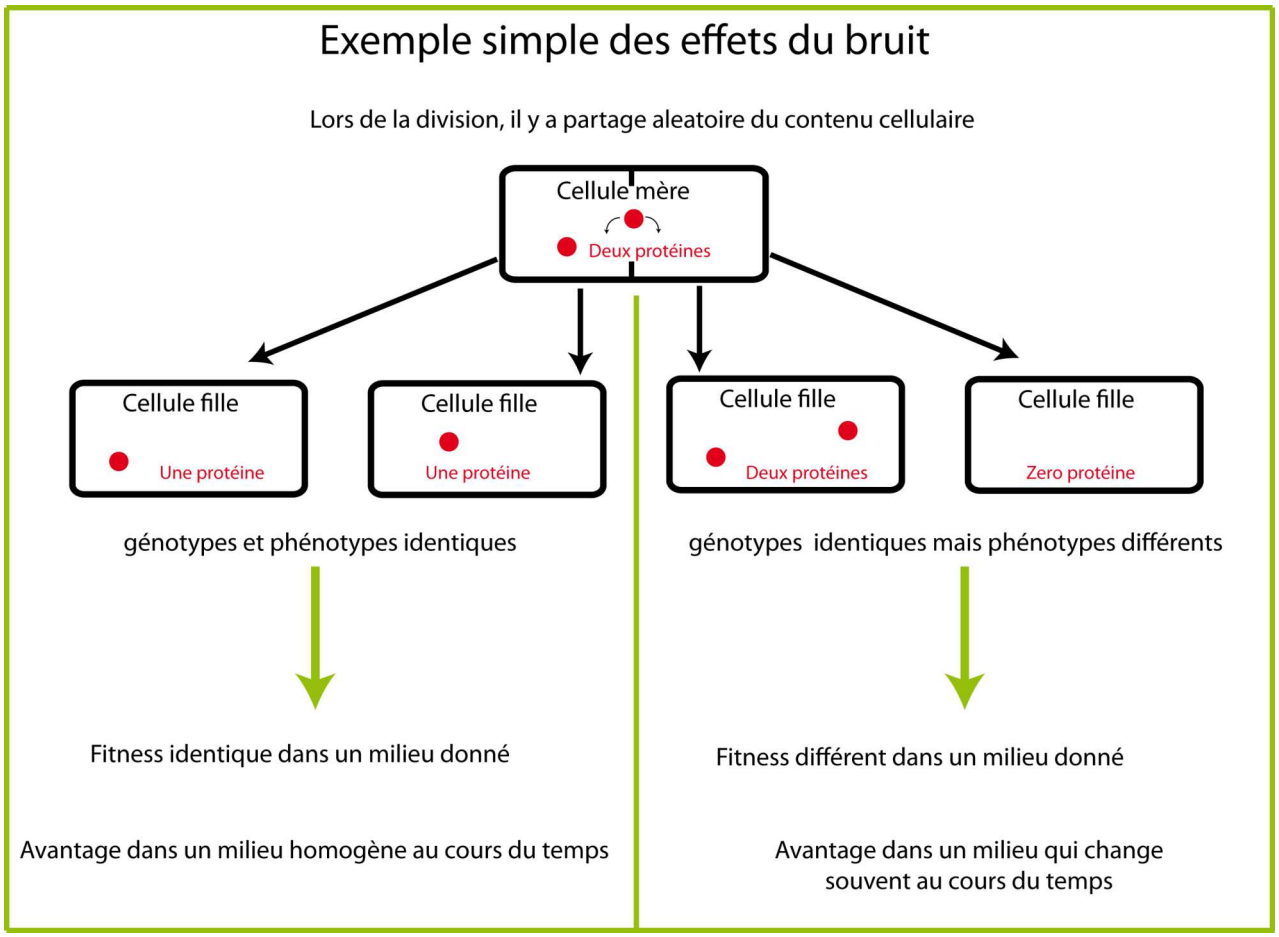



**Figure VI.1** – A cause/grâce au bruit, on peut avoir 1 génotype, 1 milieu mais plusieurs phénotypes. Cela peut fournir un avantage sélectif à une population si le milieu change souvent

de rétroaction positive dans notre système, on peut maintenir cette différence éternellement : les cellules filles de la cellule qui n'a hérité d'aucun exemplaire resteront « OFF » comme leur mère et les cellules filles de la cellule qui a hérité des 10 exemplaires resteront « ON ». On comprend donc bien qu'un état phénotypique lié au bruit puisse se transmettre de génération en génération : il s'agit d'un cas particulier d'héritage épigénétique.

Je voudrais cependant tempérer ici l'idée que le processus de division est asymétrique à cause du seul hasard. En effet, l'équipe de François Taddei a montré que les cellules filles qui héritent des pôles « vieux » arborent une diminution du taux de croissance (Stewart *et al.*, 2005) : elles vieillissent puis elles meurent. Il semblerait que la raison à ce résultat étonnant soit l'accumulation d'agrégats protéiques dans les pôles les plus vieux au cours des divisions. Les cellules déploieraient donc une stratégie de répartition asymétrique des agrégats protéiques pour perpétuer la population au détriment de cellules âgées (Lindner *et al.*, 2008). Ce résultat

démolit l'idée intuitive que toutes les cellules d'une population comme E. coli se divisent de manière infinie tant qu'il y a du milieu frais.

## 2 L'intérêt du bruit

Le bruit dans l'expression des gènes est souvent perçu par les biologistes de manière négative : le bruit provoquerait des effets néfastes et la cellule aurait appris, au cours de l'évolution, à le combattre ou du moins à le contrôler. Mais le bruit n'est pas uniquement nuisible : une population de cellules peut l'utiliser pour accroitre son fitness. En effet, l'hypothèse la plus logique pour expliquer pourquoi les bactéries exhibent des variations phénotypiques c'est qu'il s'agit d'une stratégie dite « de pari » (bet-hedging strategy) : les variations phénotypiques permettent à la population de ne pas mettre tous ses œufs dans le même panier. Si un changement brusque et violent apparait dans l'environnement, la plupart des cellules seront peut être tuées mais quelques unes seront peut être en mesure de résister grâce à leur différence de phénotype. Un des exemples le plus connu de stratégie de pari est le phénomène de persistance bactérienne, identifié la première fois dans les années 40 (Bigger, 1944). Les cellules dites persistantes sont résistantes aux antibiotiques qui s'attaquent aux cellules en division. Autrement dit ces cellules sont dans un état transitoire d'arrêt de croissance ce qui les protège des antibiotiques alors même que leur génome est identique à celui des autres cellules sensibles aux antibiotiques. Le « switch » entre « croissance normale » et « état de persistance » est stochastique par nature : 1 cellule sur 10, sur 100 ou sur 1000 est, par exemple, dans cet état et si un antibiotique arrive, toutes les cellules seront tuées sauf celle-ci. Si l'antibiotique finit par disparaitre, certaines cellules persistantes « switcheront » à nouveau vers un mode de « croissance normale » ce qui entrainera le retour de la population bactérienne. Cette population sera à nouveau composée de cellules sensibles aux antibiotiques et de cellules persistantes. Un autre exemple très classique de stratégie de pari est le phénomène de sporulation chez *Bacillus Subtilis*. La sporulation est un processus bistable car deux sous-populations peuvent être distinguées dans une population clonale de cellules en phase stationnaire : les spores (très résistants aux stress environnementaux) et les cellules « non-sporulantes » capables de repartir en croissance exponentielle plus rapidement si du milieu de croissance « frais » réapparait.

De manière plus générale, on comprend, avec ces exemples, que dans des conditions environnementales changeantes, la production d'une descendance avec des phénotypes variables assure qu'au moins 1 des descendants aura le phénotype approprié dans un environnement donné (ce descendant « fitte » bien avec le milieu). C'est une stratégie de répartition des risques qui a

pour conséquence que les descendants ne sont pas tous adaptés de manière optimale à l'environnement futur. Ainsi, dans un environnent très fluctuant, le fitness global d'une population hétérogène peut être supérieur à celui d'une population homogène. A l'inverse, dans un milieu constant, une population homogène aura toujours un fitness supérieur. En fait, il est fort possible que la répartition optimale des phénotypes mime les statistiques de changements environnementaux. Ainsi la variation phénotypique due au bruit n'est pas seulement un handicap ou un défaut : c'est un paramètre qui peut probablement être « tuné/ sélectionné » par l'évolution en fonction de la fréquence des changements environnementaux.

Notez que chez les organismes supérieurs (le ver *C. elegans*, la drosophile, l'homme...), le bruit ne permet pas seulement une meilleure adaptation aux environnements fluctuants, il intervient également dans le processus de développement. L'hétérogénéité (la simple gaussienne) présente dans une population clonale de cellules non différenciées pourrait subir des mécanismes d'amplification orientant petit à petit l'expression de gènes de chaque cellule dans une voie de différentiation particulière.

## 3 Le contrôle du bruit

Comment la cellule contrôle-t-elle le bruit ? Certains mécanismes de rétrocontrôle pourraient avoir évolué spécifiquement pour minimiser ou amplifier l'impact des variations stochastiques dans l'expression des gènes. On cite souvent la boucle de rétroaction négative comme système capable de réduire le bruit. Pour tester cette hypothèse, des chercheurs ont fusionné le gène codant pour le facteur de transcription TetR avec celui codant pour la GFP. Cette protéine de fusion réprime l'expression de son propre gène (via le promoteur tet). Les chercheurs ont ensuite mesuré la distribution de fluorescence des cellules et montré que cette distribution est plus étroite en présence de la boucle de rétroaction que sans cette dernière (Becskei & Serrano, 2000).

A l'inverse, une cellule peut amplifier les effets du bruit en utilisant une boucle de rétroaction positive. En faisant varier la force de la boucle de rétroaction, on peut même transformer une réponse graduelle en une réponse binaire où les cellules expriment un gène soit fortement soit faiblement (Isaacs *et al.*, 2003). A l'échelle de la population, ce comportement d'interrupteur peut résulter en une distribution bimodale dans l'expression du gène : quand on observe ce type de pattern, on parle en général de bistabilité (voir le chapitre sur les systèmes dynamiques). Ce phénomène de bistabilité induit par une boucle de rétroaction positive, a été décrit pour la première fois dans les années 50 par Novick et Weiner avec l'expression de la $\beta$-galactosidase

(Novick & Weiner, 1957)). Voici une phrase limpide extraite de leur publication :

> *« It was difficult to understand this result on the assumption that each bacterium has about the same enzyme content. We were able to show that this assumption does not apply at low inducer concentrations. We discovered that at the low inducer concentrations used in these experiments the population consists essentially of individual bacteria that are either making enzyme at full rate or not making it at all [...] .This can be understood on the following basis. When inducer is added to a culture of growing bacteria, there is a certain chance, determined by the inducer concentration, that a given bacterium will produce its first permease molecule. Once a bacterium has one permease molecule, the internal inducer concentration is raised, and the probability of the appearance of a second permease molecule is increased. In this sense the induction of permease in the individual bacterium is an autocatalytic process*[2] *, and, within a short time after the appearance of its first permease molecule, the bacterium becomes fully induced, synthesizing both permease and galactosidase at maximum rate. Because the transition is accomplished so rapidly, the relative number of bacteria at an intermediate state of induction is small. »*

Pour obtenir un réseau capable de générer des multi stabilités, il faut en général :

⇒ soit que le système exhibe des cinétiques non linéaires. Pour les facteurs de transcription, la non-linéarité peut survenir lors de liaisons coopératives sur l'ADN.

⇒ soit que le système exhibe une ou des boucles de rétroaction positives (ou que le système exhibe un *nombre pair* de boucles de rétroaction négative).

Une des caractéristiques classiques de la bistabilité est le phénomène d'hystérèse. L'hystérèse se réfère à la situation dans laquelle la transition d'un état à un autre requiert l'induction (ou levée d'induction) d'un niveau plus fort que ce qui se passe lors de la transition dans le sens inverse. Ceci génère des caractéristiques de mémoire qui rendent la réponse d'une cellule à une perturbation donnée dépendante de son histoire récente.

Il faut garder à l'esprit que le concept de robustesse d'un réseau peut revêtir deux notions bien distinctes :

⇒ la robustesse à des petites variations dans la valeur des paramètres (à rapprocher de la sensibilité aux conditions initiales et du fameux effet papillon). Ce type de robustesse n'a pas besoin d'un modèle stochastique pour être étudié : un modèle déterministe suffit.

[2] Ce que nous appelons aujourd'hui boucle de rétroaction positive.

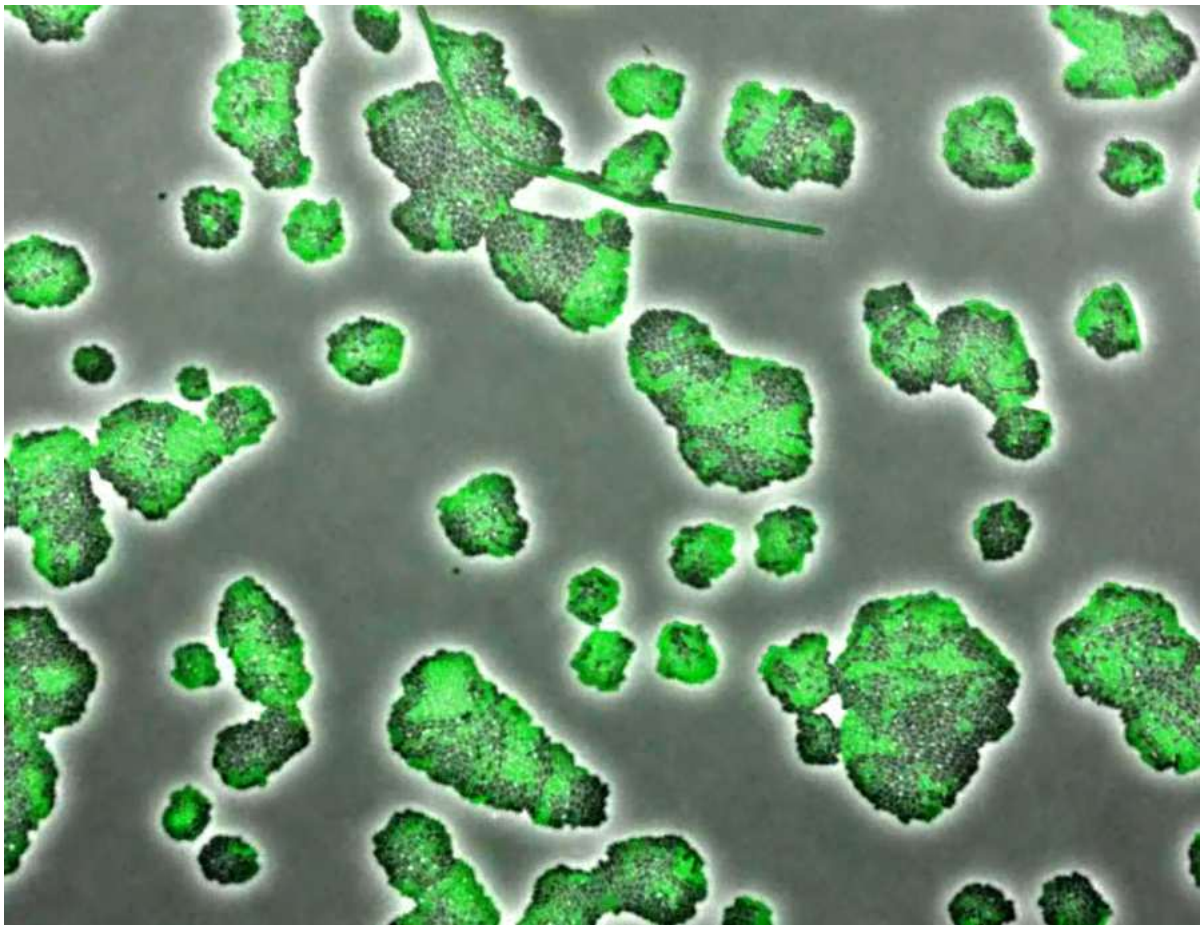


**Figure VI.2** – Exemple de bistabilité obtenue avec l'operon *araBAD*. Remarquez la présence de cellules très fluorescentes et d'autres qui le sont beaucoup moins. Remarquez aussi la présence d'un « serpent » (une longue bactérie qui semble ne plus pouvoir se diviser) . L'existence de ce serpent peut être due à une mutation ou à l'expression stochastique des gènes.

⇒ la robustesse au bruit : comment une cyanobactérie comme *Synechococcus* elongates fait-t-elle pour maintenir constante la fréquence de ses oscillations circadiennes ? Des chercheurs ont pu montrer que la persistance d'oscillations stables (sur plusieurs mois) chez ces bactéries n'est pas due à un phénomène de synchronisation des oscillateurs par communication intercellulaire. La forte stabilité temporelle est encodée dans le réseau intracellulaire : c'est une propriété de *la* cellule (Mihalcescu *et al.*, 2004). Comment une telle stabilité peut-elle exister malgré la présence de fluctuation moléculaire ? Comment le réseau circadien peut-il être si résistant au bruit biochimique ? Vous voyez que cette robustesse au bruit n'est pas exactement la même que celle relative aux petites variations dans la valeur des paramètres. En effet, ici, un modèle stochastique est indispensable pour comprendre cette robustesse.

# 4 Utiliser la position relative des gènes sur l'ADN

L'équipe de Johan Elf a construit une protéine « LacI–YFP » en fusionnant un gène codant pour le lac represseur (lacI) et un gène codant pour une protéine fluorescente (yfp). En modifiant la force du promoteur controlant la transcription de ce gène, cette équipe arrive à limiter le nombre de protéines LacI–YFP à 3 exemplaires en moyenne par cellule. En utilisant un microscope à fluorescence, les auteurs arrivent à repérer un exemplaire unique du facteur de transcription fixé sur son site spécifique sur l'ADN (le lac operateur) : on parle de « single molecule microscopy » car on repère une tache de fluorescence à un endroit précis dans la cellule. Bien sûr lorsque les chercheurs ajoutent de l'IPTG, le facteur de transcription se détache et diffuse dans la cellule : on perd alors la faculté de le localiser et les images de microscopie montrent une fluorescence diffuse dans toute la cellule. Un exemplaire LacI–YFP dans une cellule met environ 360 secondes pour trouver son site spécifique sur l'ADN et s'y fixer. Il passe environ 90% du temps à diffuser le long de l'ADN de manière non spécifique (diffusion 1D) et 10% du temps à effectuer des translocations en 3D d'un segment d'ADN à l'autre en passant par le cytoplasme.

Je vous propose maintenant, à partir de ces informations, une petite expérience de pensée (figure VI.3) dont j'ai eu l'idée après avoir lu un article de Jean-Jacques Kupiec dans *Pour la science* (Kupiec, 2009). Nous savons que l'effet du bruit sur l'expression de gènes est d'autant plus marqué que le nombre de molécules est restreint : considérons donc une seule molécule LacI dans la cellule mais 3 sites operateurs identiques. Le premier site est le site sur lequel est fixée la molécule LacI au début de l'expérience. Le deuxième site est très proche du premier site : par exemple à 200pb. Ce site réprime l'expression d'un gène réprimant l'expression de la RFP (Red Fluorescent Protein) lorsque LacI y est fixé. Le dernier site est beaucoup plus loin sur le chromosome (à plus de 1 million de pb) et on imagine que la distance réelle est également importante (quelques centaines de nanomètres) ce qui n'est pas forcement le cas à cause du superenroulement de l'ADN. Ce site réprime l'expression d'un gène réprimant l'expression de la GFP (Green Fluorescent Protein) lorsque LacI y est fixé. L'expérience de pensée débute en supposant qu'on a une population de 1000 cellules toutes ayant LacI sur le premier site operateur. Je rajoute de l'IPTG à une concentration suffisante ce qui libère immédiatement LacI du premier site. Puis on rajoute au bout d'un temps judicieusement choisi (entre 1 et 1000 secondes) un grand volume de milieu pour diluer l'IPTG et rendre LacI à nouveau actif. Ma question est la suivante : la probabilité que LacI se fixe au deuxième site (plus proche) est-elle supérieure à la probabilité de fixation au troisième site plus éloigné ? Autrement dit : y aura-t-il beaucoup plus de cellules rouges que de cellules vertes ? Ma prédiction est que la

réponse est oui et cela signifie qu'on peut utiliser la distance relative sur l'ADN pour contrôler l'expression d'un gène. On peut encoder de l'information dans l'ADN en utilisant la distance physique qui sépare les gènes. Cela peut sembler trivial mais cela a des conséquences lorsqu'on réfléchit sur les analogies entre cellules et ordinateurs. En effet, cela signifie qu'on peut encoder une information dans l'ADN non plus uniquement en choisissant la valeur d'un nucléotide (A, T, C, G) mais également en choisissant la distance physique qui le sépare d'un autre nucléotide. Or mon hypothèse est que cela est impossible dans une machine de Turing. Je reconnais que la question est un petit peu abstraite, je ne sais pas si elle a un quelconque intérêt et depuis quelque temps, j'ai un doute. J'y reviendrai peut être un jour.

# 5 Les méthodes expérimentales pour étudier le bruit

Je vois essentiellement deux techniques permettant de mesurer les variations d'expression génique dans une population des cellules :

⇒ La cytométrie en flux. C'est une technique qui permet de faire défiler des particules (par exemple des bactéries) à grande vitesse dans le faisceau d'un laser, en les comptant et en les caractérisant. Parmi les paramètres récoltés peuvent se trouver l'aire, le facteur de circularité (sur lequel je vais revenir) ou la fluorescence moyenne de chaque cellule. Notez un point qui me semble toujours extraordinaire : certains cytomètres en flux sont équipés d'un trieur. Prenons un exemple : vous faites passer dans votre cytomètre 100 millions de cellules « vertes » (GFP) mélangées à 100 millions de cellules « rouges » (RFP). Après l'illumination de chaque cellule par le laser, le trieur oriente chaque cellule dans un compartiment ou un autre en fonction des paramètres récoltés (la fluorescence) qui sont quasi-immédiatement disponibles. Cela vous permet de séparer vos deux populations rouges et vertes puis de les récupérer.

⇒ La cytométrie en image. C'est une technique dont l'instrument principal est le microscope. Le chercheur acquiert des images de ses cellules (en contraste de phase et/ou en fluorescence) puis il fait tourner un algorithme qui détecte les cellules sur les images et en extrait les paramètres (comme la cytométrie en flux) : l'aire, le facteur de circularité ou la fluorescence par exemple.

Avec ces deux techniques, on obtient des distributions. Par exemple, la distribution de fluorescence moyenne des cellules ou la distribution des aires des cellules. On peut ensuite « fitter » ces données de distribution avec un modèle stochastique (que nous aborderons plus loin) pour identifier les paramètres d'un modèle par exemple. La cytométrie en image est

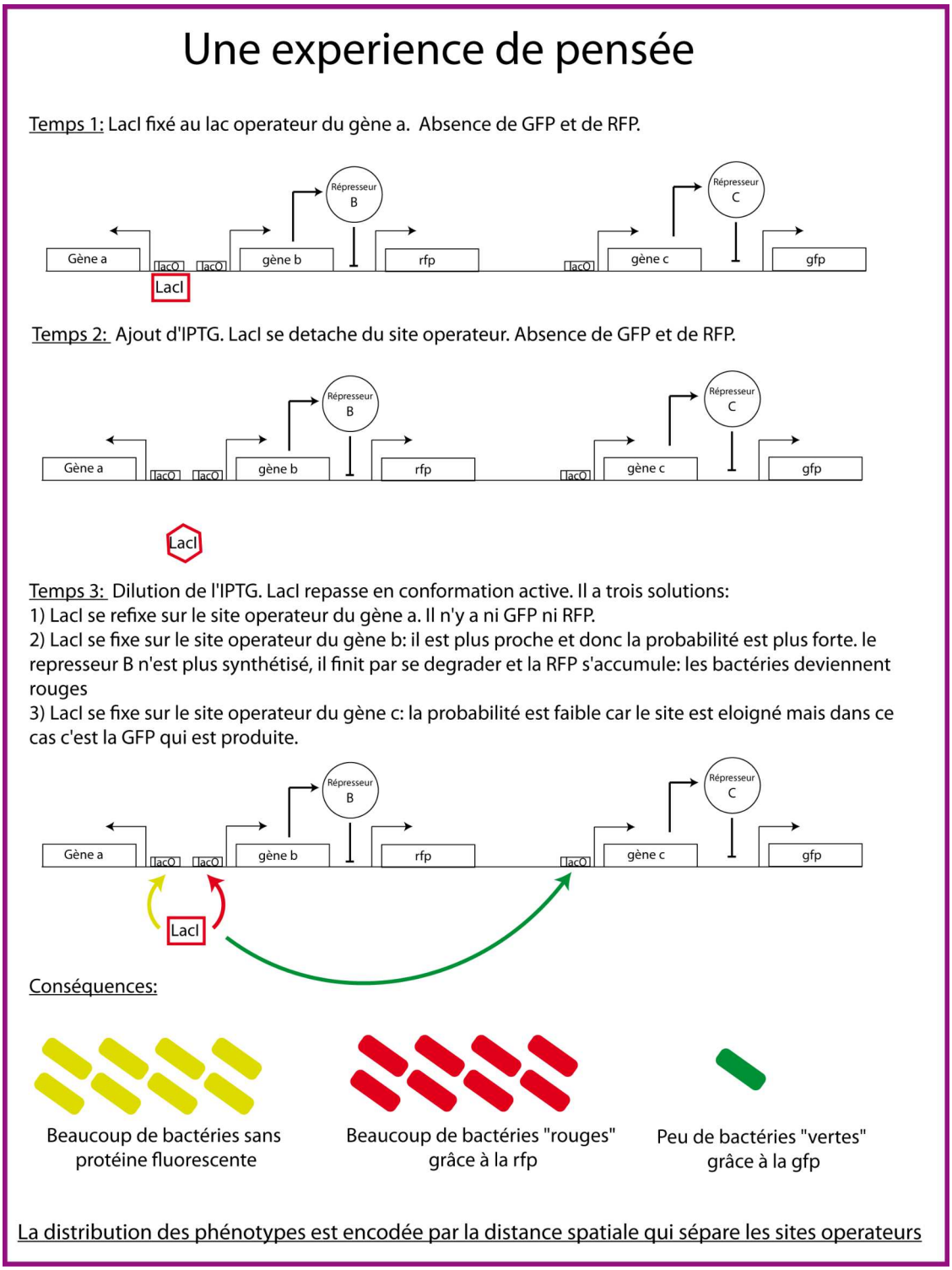



**Figure VI.3** – Une expérience de pensée : encoder de l'information dans le génome en utilisant la distance spatiale entre les gènes.

globalement plus chère que la cytometrie en flux cependant elle dispose d'un avantage de poids : elle permet à tout moment d'effectuer un contrôle visuel sur une sous-population de cellules, ce qui peut s'avérer rassurant pour l'interprétation les résultats.

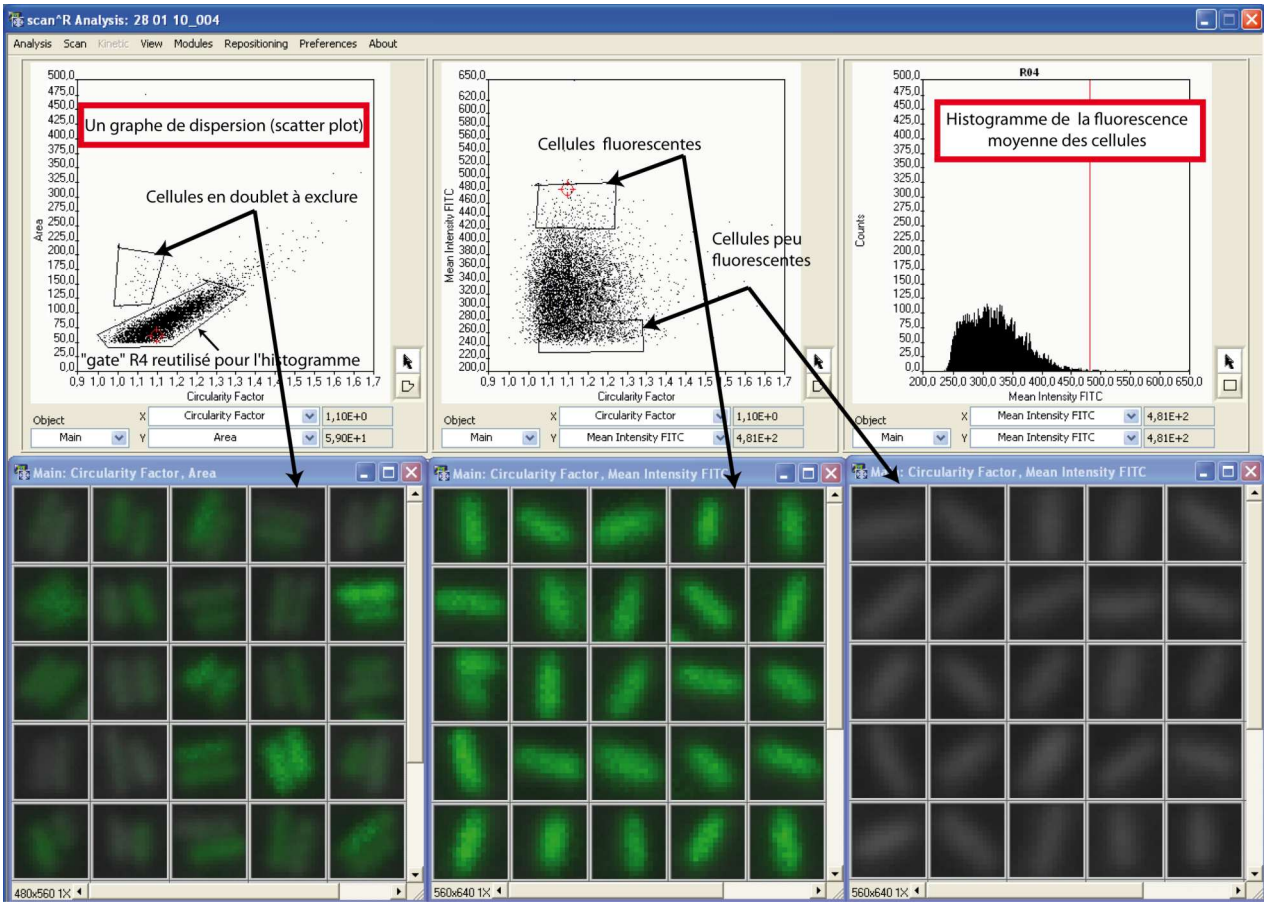



**Figure VI.4** – Interface graphique du logiciel ScanR Analysis développé par la société Olympus et l'EMBL Heidelberg. Un algorithme détecte toutes les particules dans un batch d'images (des dizaines de milliers de cellules sont détectées en quelques secondes !). Puis l'utilisation des graphes de dispersion permet de sélectionner les groupes de cellules intéressantes (gate). La possibilité de visualiser facilement les images des groupes de cellules est « un plus » indéniable de la cytomètrie en image par rapport à la cytomètrie en flux.

Il y a un point technique que je souhaite aborder. Il peut intéresser toute personne confrontée à des images de cellules et qui souhaiterait détecter puis extraire les paramètres à l'aide d'un petit programme « home-made ». Cette personne pourrait développer son algorithme « d'image processing » et se rendre compte qu'il souhaite exclure les particules inintéressantes, par exemple, les doublets de cellules ou les poussières. L'exclusion des faux positifs ne doit pas être intégrée dans la couche logicielle de détection des cellules. Cela peut s'avérer difficile et représenter une perte de temps. Il faut accepter les faux positifs (un algorithme rudimentaire)

et exclure manuellement les faux positifs à l'aide de notre œil. Pour ne pas que cela devienne une corvée, on utilise la méthode du graphe de dispersion (scatter plot). Ce graphique « plot » simplement un paramètre en fonction d'un autre : par exemple l'aire en fonction du facteur de circularité (indication sur la forme des cellules). Cela « trie » les cellules et permet de séparer (et visualiser) les mauvais groupes d'objets détectés et les bons groupes. De simple fonctionnalités « user friendly » permettront de sélectionner (entourer) les bons objets pour ne travailler qu'avec eux dans la suite de l'analyse (c'est le concept de « gate » ). Pour résumer : Développeurs : ne perdez pas de temps à implémenter des fonctionnalités trop complexes dans le processus de détection d'objets. Préférez l'utilisation *a posteriori* des graphes de dispersion (figure VI.4). La video VI.5 illustre, en quelques minutes, la puissance de la cytométrie en image.

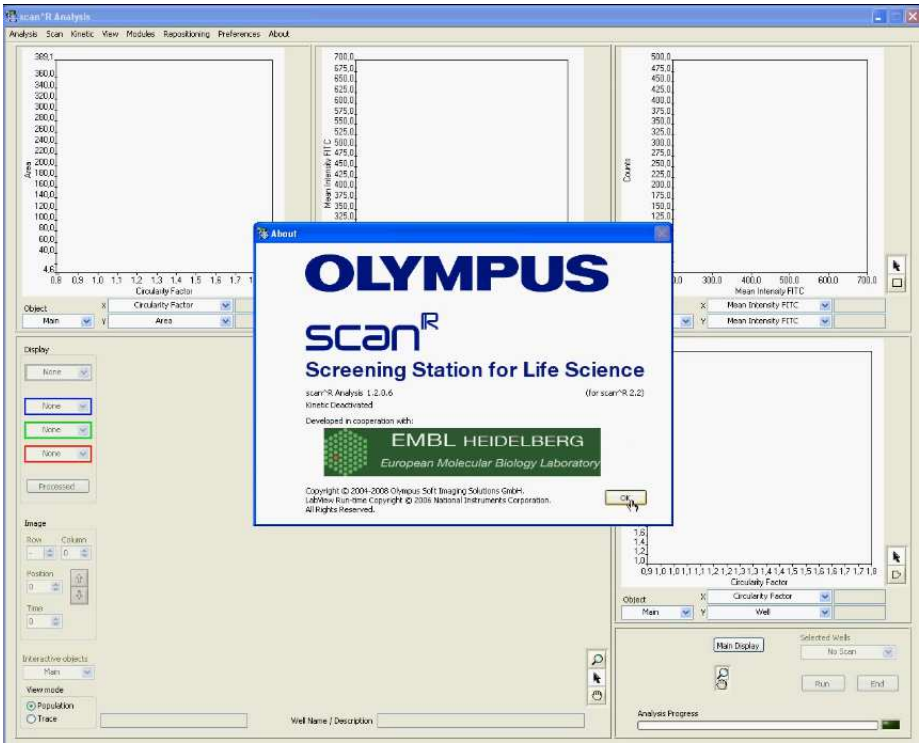



**Figure VI.5** – Vidéo : démonstration de la facilité d'utilisation du logiciel ScanR analysis.

# 6 Le bruit intrinsèque et le bruit extrinsèque

Il existe différentes sources de bruit qu'il convient de distinguer. Les variations d'expression génique inhérentes au processus biochimique de l'expression génique sont appelées bruit intrinsèque. On peut définir le bruit intrinsèque comme étant les fluctuations générées par l'activation

et la désactivation stochastique d'un promoteur, ainsi que celles générées par la synthèse et la dégradation des ARNm et protéines (figure VI.6). Le bruit extrinsèque, lui, provient de deux sources :

⇒ le bruit extrinsèque relatif au réseau. Ce bruit est généré par les fluctuations des concentrations des facteurs de transcription contrôlant l'expression d'un gène. Ainsi le bruit intrinsèque dans l'expression d'un gène a codant pour un facteur de transcription A représente aussi le bruit extrinsèque vis-à-vis de l'expression du gène b qui est activé par le facteur de transcription A.

⇒ le bruit extrinsèque relatif à la cellule. Ce bruit provient des fluctuations dans les concentrations des ribosomes, de l'ARN polymérase, des métabolites, le volume de la cellule, l'état de la cellule dans le cycle cellulaire (car avant la division cellulaire, une cellule peut posséder plusieurs copies du même gène)...

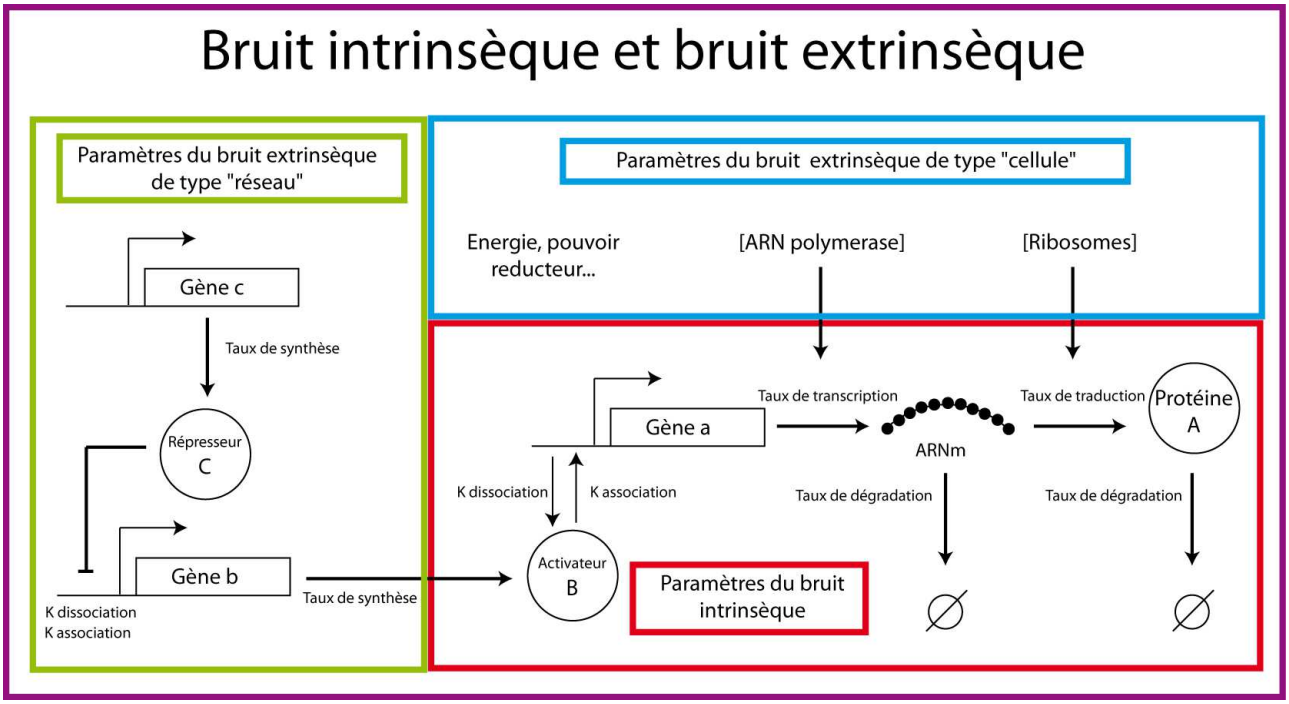



**Figure VI.6** – Bruit intrinsèque et bruit extrinsèque

En 2002, l'équipe de Michael Elowitz à Caltech en Californie, a proposé une méthode permettant de distinguer le bruit intrinsèque du bruit extrinsèque (Elowitz *et al.*, 2002). Cette méthode consiste à quantifier l'expression de deux gènes rapporteurs (YFP et CFP) placés à deux endroits équivalents sur le chromosome d'*E. coli*. Si l'expression des gènes était déterministe, toutes les cellules d'une population devrait avoir exactement la même concentration de

protéine fluorescente jaune (YFP) et bleue (CFP) : les cellules devraient toutes avoir la couleur verte. Ce n'est évidemment pas ce qu'ont observé les chercheurs : la couleur de fluorescence varie d'une bactérie à l'autre, allant du jaune à l'orange en passant par le vert. Le bruit intrinsèque est le bruit qui génère une différence de concentration de rapporteurs (YFP et CFP) dans la même cellule. Sur un graphe de dispersion représentant la fluorescence jaune en fonction de la fluorescence verte, le bruit intrinsèque peut être évalué sur la diagonale qui descend de gauche à droite. A l'inverse, le bruit extrinsèque est évalué sur la diagonale qui monte de gauche à droite (figure VI.7). Il représente les fluctuations des éléments (facteur de transcription, machinerie globale) qui affectent les concentrations de deux rapporteurs de manière identique dans une cellule donnée mais créent des différences entre deux cellules.

# 7 Les modèles stochastiques

## 7.1 Les équations maîtresses

Des connaissances plus poussées en mathématique deviennent nécessaires pour comprendre en détail les publications décrivant des modèles stochastiques de l'expression génique. La question de la modélisation du bruit en biologie semble d'ailleurs plutôt avoir été capturée par les physiciens et mathématiciens. Cependant, j'ai dans l'idée que nous, biologistes, pouvons tout de même réussir à comprendre les concepts, la substance pour pouvoir ne serait-ce que discuter avec un mathématicien. Je vais essayer de mettre en avant ce que je comprends et ce qui me semble important mais je demande une certaine indulgence à mon lecteur si perdurent quelques inepties . Dans ce cas, le lecteur pourrait directement consulter mes sources : des tutoriaux trouvés sur Internet (Biology of the noisy gene : Juan F. Poyatos ; Single cell experiments and Gillespie's algorithm : Fernand Hayot).

Pour expliquer comment on modélise l'expression stochastique des gènes, je vais repartir de l'équation la plus simple décrite dans le chapitre sur la modélisation.

$$\frac{d[B]}{dt} = \kappa - \gamma \cdot [B] \tag{VI.1}$$

Notez que dans ce cas d'école extrêmement simple, il n'y a pas de régulation transcriptionnelle et on ne découple pas la production des ARNm et des protéines alors qu'il y a également introduction de bruit lors de ces étapes. Des modèles plus rigoureux mais un peu plus complexes découplant transcription et traduction existent également (Thattai & van Oudenaarden, 2001). Nous allons transformer cette équation de manière à ce qu'elle puisse décrire/intégrer les

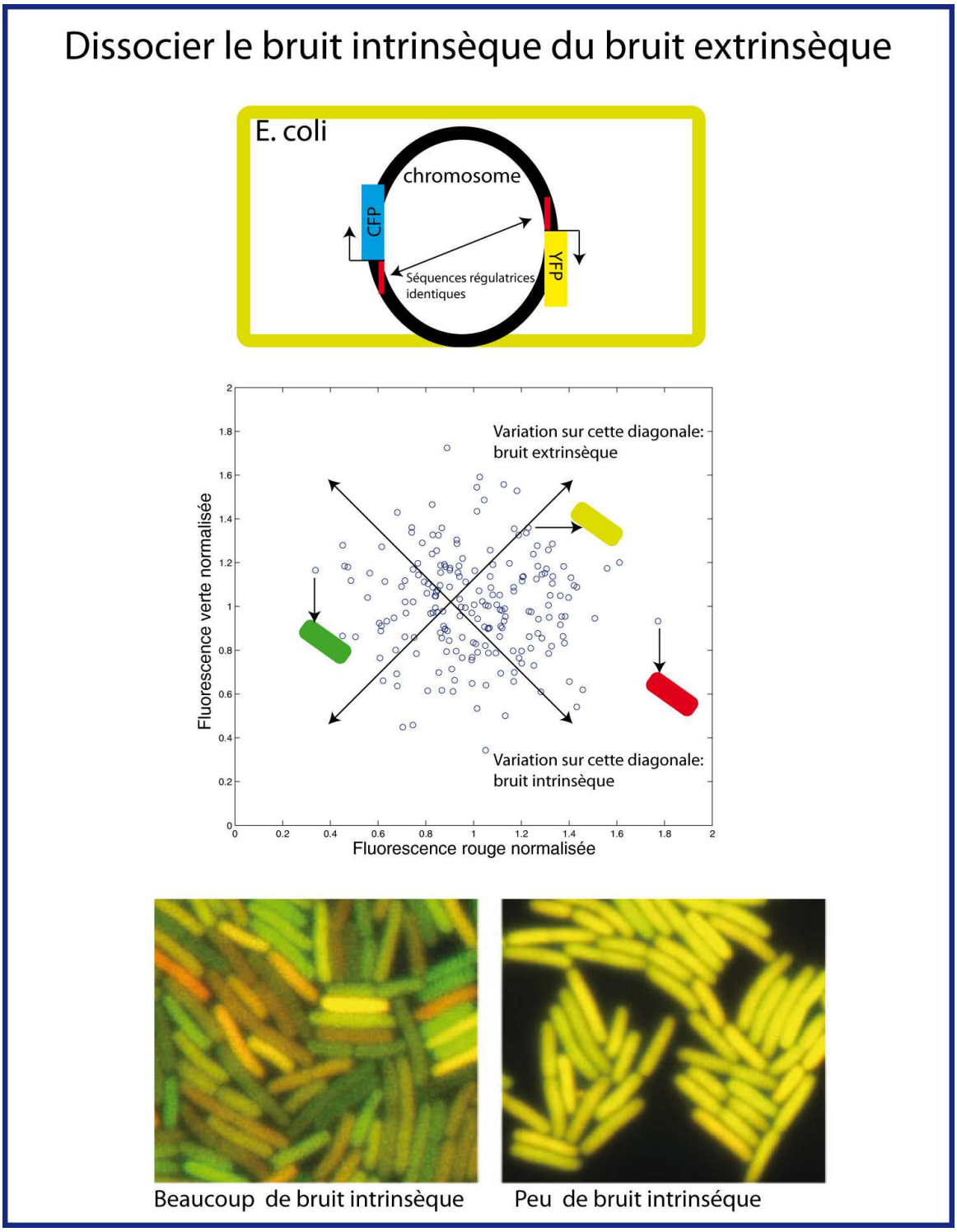



**Figure VI.7** – L'expérience de l'équipe de Michael Elowitz pour dissocier bruit intrinsèque et bruit extrinsèque. Les images de microscopie sont tirées de leur publication *Stochastic gene expression in a single cell.*

variations de la concentration de la protéine B dans une population de cellules. Le mot clé dont nous avons besoin est le mot probabilité. Pour cela, on utilise un type particulier d'équations différentielles dites équations maitresses (master equations). Une équation maîtresse est une équation différentielle décrivant l'évolution temporelle d'une distribution de probabilité. Cette distribution représente la probabilité d'un système d'occuper chacun des ensembles d'état discret (1, 2, 3 protéines par cellules...) en regard d'une variable de temps continue. Autrement dit, on va remplacer la concentration (c'est à dire la concentration moyenne de la protéine B dans la population de cellules) par une probabilité.

Si on raisonne à l'échelle de la cellule, on travaillera avec une probabilité P(n, t) d'avoir n protéines de type B dans une cellule à un temps t : par exemple, il y aura 25% de chance (0.25) d'avoir 3 protéines B dans la cellule, 50% de chance (0.5) d'avoir 4 protéines B dans la cellule et 25% de chance (0.25) d'avoir 5 protéines B dans la cellule.

Si on raisonne à l'échelle de la population (grand nombre de cellules), cette probabilité « se transforme » en distribution. En reprenant l'exemple précédent, si on a 1000 cellules : cela signifie qu'approximativement 250 cellules auront 3 protéines B, 500 cellules auront 4 protéines B et 250 cellules auront 5 protéines B.

On ne décrit plus l'évolution de la concentration moyenne de B dans la population en fonction du temps : on décrit l'évolution de la distribution du nombre de protéines B dans la population en fonction du temps.

Voyons maintenant comment s'écrit cette équation. Soit P(n,t) la probabilité d'avoir n protéines de type B dans une cellule à un temps t. Nous souhaitons décrire l'évolution de P(n,t) en fonction du temps c'est-à-dire $\frac{dP(n,t)}{dt}$. Supposons qu'au temps t, l'état du système soit connu, c'est à dire le nombre de protéine B dans la cellule. Quelle est la probabilité P (n, t+dt) d'avoir n protéines B au temps t+dt tel que dt soit suffisamment petit de manière à ce qu'une seule réaction de production ou dégradation puisse se produire ? Il faut considérer trois possibilités (figure VI.8) :

⇒ Il y a n-1 protéines B au temps t et c'est l'événement de production qui a lieu au temps dt ce qui donne une contribution à P(n,t+dt) égale à $\kappa$.dt.P(n-1,t)

⇒ Il y a n+1 protéines B au temps t et c'est l'événement de dégradation qui a lieu au temps dt ce qui donne une contribution à P(n,t+dt) égale à $\gamma$.dt.(n+1).P(n+1,t). Pourquoi y a-t-il le facteur (n+1) qui apparait ? A cause du fait que la vitesse de dégradation dépend du nombre de protéines déjà présentes (voir le chapitre sur la modélisation)

⇒ Il n'y a aucune réaction qui se produit. Ce troisième point est un petit peu difficile à comprendre. Les deux premiers points considèrent les probabilités *« d'arriver »* à n

molécules soit en venant de n-1 molécules (production) soit en venant de n+1 molécules (dégradation). Mais il faut aussi prendre en considération la probabilité de *ne pas « quitter/ bouger »* de n protéines durant le temps dt si on était déjà à n protéines au temps t. Cette probabilité c'est la probabilité d'avoir n protéines au temps t, P(n,t) auquel on soustrait la probabilité qu'ait lieu un événement de production $\kappa$.dt.P(n,t) durant le temps dt ainsi que la probabilité qu'ait lieu un événement de dégradation $\gamma$.dt.n.P(n,t) durant le temps dt.

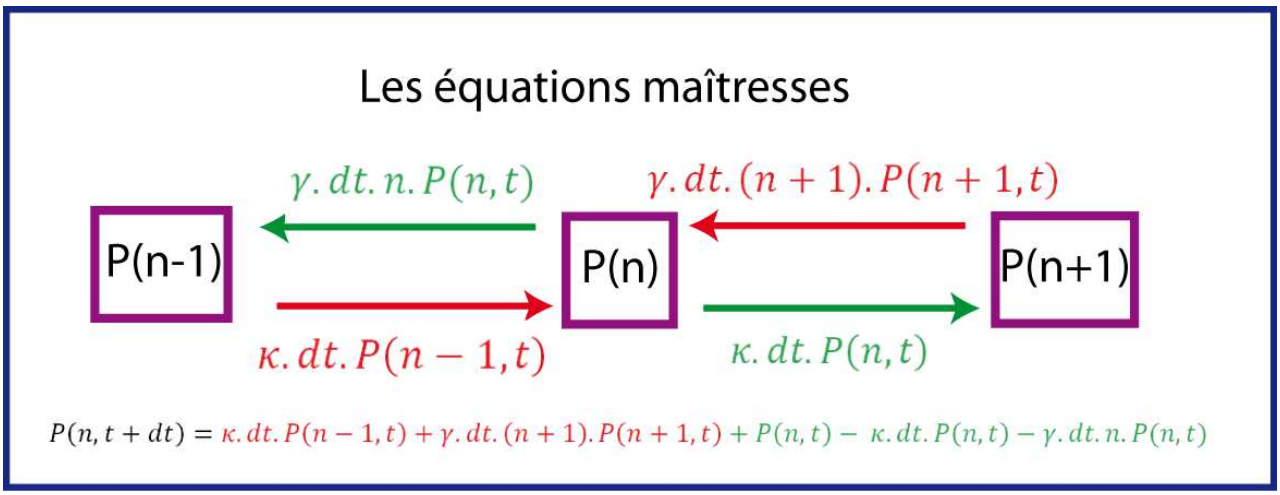



**Figure VI.8** – Schéma expliquant l'obtention de l'équation maîtresse. En rouge, c'est la probabilité d'arrivée sur « n » au temps t+dt et en vert c'est la probabilité de départ de « n » durant le temps dt.

On a donc :

$$P(n,t+dt) = \kappa.dt.P(n-1,t) + \gamma.dt.(n+1).P(n+1,t) + P(n,t) - \kappa.dt.P(n,t) - \gamma.dt.n.P(n,t)$$

On peut ensuite passer sous forme de dérivée $\frac{dP(n,t+dt)-P(n,t)}{dt}$ et on obtient quasi immédiatement l'équation ci-dessous qui correspond à l'équation maitresse :

$$\frac{dP(n,t)}{dt} = \kappa.[P(n-1,t) - P(n,t)] + \gamma[(n+1).P(n+1,t) - n.P(n,t)] \qquad \text{(VI.2)}$$

## 7.2 Retrouver l'équation déterministe à partir de l'équation maitresse

A partir de cette équation maitresse, on peut retrouver l'équation « déterministe »[3] des concentrations $\frac{d[B]}{dt} = \kappa - \gamma \cdot [B]$ (La démonstration qui suit est un peu fastidieuse : n'hésitez pas à la sauter). Pour cela, il me faut introduire l'espérance <n> qui est l'équivalent en probabilité de la moyenne. On a :

$$< n >= \sum_{n=1}^{N} nP_n$$

En prenant la dérivée par rapport au temps, on a :

$$\frac{d < n >}{dt} = \sum_{n=1}^{N} \frac{dP(n,t)}{dt}$$

En remplaçant le terme de gauche par le résultat obtenu juste avant ce paragraphe, on obtient :

$$\frac{d < n >}{dt} = \sum_{n=1}^{N} n.\left\{\kappa.[P(n-1,t) - P(n,t)] + \gamma[(n+1).P(n+1,t) - n.P(n,t)]\right\}$$

En remaniant et en se débarrassant des t encombrants, on obtient :

$$\frac{d < n >}{dt} = \sum_{n=1}^{N} n[-P_n(\kappa + n\gamma) + P_{n-1}\kappa + P_{n+1}(n+1)\gamma]$$

En sortant $\kappa$ et $\gamma$ des sommes, on obtient :

$$\frac{d < n >}{dt} = -\kappa < n > -\gamma \sum_{n=1}^{N} n^2 P_n + \kappa \sum_{n=1}^{N} p_{n-1} n + \gamma \sum_{n=1}^{N} n(n+1) p_{n+1}$$

En factorisant $\kappa$ et $\gamma$, on a :

$$\frac{d < n >}{dt} = \kappa \left( \sum_{n=1}^{N} P_{n-1} n - < n > \right) - \gamma \left( \sum_{n=1}^{N} n^2 P_n - \sum_{n=1}^{N} n(n+1) P_{n+1} \right)$$

Focalisons-nous sur le premier terme (en considérant que n= n-1+1) :

$$\kappa \left( \sum_{n=1}^{N} P_{n-1} n - < n > \right) = \kappa \left( \sum_{n=1}^{N} P_{n-1}(n-1) + \sum_{n=1}^{N} P_{n-1} - < n > \right) = \kappa(< n > +1 - < n >) = \kappa$$

[3] Je mets déterministe entre guillemet car en réalité, les équations maitresses, malgré la présence de probabilités, sont également déterministes.

Puis focalisons-nous sur le deuxième terme (en considérant que n= n+1-1) :

$$-\gamma\left(\sum_{n=1}^{N} n^2 P_n - \sum_{n=1}^{N} n(n+1)P_{n+1}\right) =$$
$$-\gamma\left[< n^2 > - \left(sum_{n=0}^{N}(n+1)(n+1)P_{n+1} - P_1 - sum_{n=1}^{N}(n+1)P_{n+1}\right)\right]$$

C'est-à-dire :

$$-\gamma\left(\sum_{n=1}^{N} n^2 P_n - \sum_{n=1}^{N} n(n+1)P_{n+1}\right) = -\gamma(< n^2 > - < n^2 > + < n >) = -\gamma < n >$$

En reconstituant les deux termes, on retrouve bien :

$$\frac{d < n >}{dt} = \kappa - \gamma < n >$$

C'est-à-dire l'équivalent de notre équation déterministe initiale $\frac{d[B]}{dt} = \kappa - \gamma \cdot [B]$

## 7.3 La loi de Poisson

A partir de l'équation VI.2, on peut chercher les solutions à l'état d'équilibre c'est-à-dire quand $\frac{dP(n,t)}{dt} = 0$. Il existe une démonstration par récurrence (trop difficile pour moi ; voir le tutoriel de Fernand Hayot sur internet) qui montre que la distribution de P(n) suit une loi de probabilité de la forme suivante :

$$p(n) = \frac{(\frac{\kappa}{\gamma})^n exp(-\frac{\kappa}{\gamma})}{n!}$$

C'est à dire une loi de Poisson de la forme :

$$p(n) = \frac{< n >^n exp(- < n >)}{n!}$$

On repère bien que le terme $\frac{\kappa}{\gamma}$ correspond à l'espérance <n> ce qui est en parfaite adéquation avec l'équation obtenue $[B] = \frac{\kappa}{\gamma}$ en résolvant l'équation déterministe $\frac{d[B]}{dt} = \kappa - \gamma \cdot [B]$ à l'état d'équilibre $\frac{d[B]}{dt} = 0$ (voir le chapitre sur la modélisation). Ceci est donc encore un bon exemple montrant la correspondance entre le comportement moyen du système déterministe et la moyenne (l'esperance) dans la population obtenue avec la formulation stochastique.

Mais que représente exactement ce bruit de Poisson ?

Nous avons vu qu'il est possible de mesurer l'expression génique (la fluorescence rapportant

le nombre de molécules de GFP) de chaque cellule d'une population clonale placée dans un milieu identique. Les données obtenues permettent de « plotter »[4] la distribution de fluorescence des cellules. Cette distribution de fluorescence reflète la distribution du nombre de protéines GFP dans les cellules. Une fois les distributions obtenues, on peut extraire des paramètres comme la moyenne ou la variance. De même, la forme de la distribution peut être comparée (fittée) avec de distributions connues comme les lois de gauss, gamma ou Poisson ce qui peut révéler les mécanismes sous jacent de la stochasticité. Dans notre cas, nous avons vu que le modèle stochastique le plus simple de l'expression d'un gène (pas de régulation, pas de passage par les ARN) suit une loi de poisson. C'est une loi discrète, adaptée à la description d'événements (nombre de molécules) rares. Une des caractéristiques de la loi de poisson c'est que la variance est égale à la moyenne. Notez que, au delà de 5 événements (molécules), la loi de poisson ressemble fortement à une loi normale dont la variance est égale à la moyenne.

Un paramètre qui est couramment utilisé pour évaluer la force du bruit est le facteur Fano F :

$$F = \frac{\sigma^2}{< n >} \tag{VI.3}$$

Dans notre cas d'école très simple de l'expression génique, la distribution des protéines suit une loi de Poisson donc F=1 car, je le répète, la variance est égale à la moyenne. Mais quand on travaille sur un système plus complexe intégrant des régulations géniques ou des boucles de rétroaction, les équations maitresses deviennent très complexes. La mesure du facteur Fano permet d'évaluer les déviations par rapport au comportement de Poisson (par rapport à 1).

## 7.4 L'algorithme de Gillespie

Lorsque le système devient trop complexe (la plupart du temps), les équations maitresses ne peuvent plus être intégrées analytiquement. On utilise, tout comme pour les équations déterministes, un algorithme numérique nommé algorithme de Gillespie. Notez cependant que contrairement aux méthodes d'approximation numérique utilisées par Matlab dans le chapitre de modélisation (Méthode de Runge-Kutta ; fonction de type ode45), l'algorithme de Gillespie est un algorithme exact : il n'introduit aucune erreur dans le processus itératif (Gillespie, 1977). La description qui suit peut être utile à une personne cherchant à comprendre le fonctionnement de l'algorithme de Gillespie (script Matlab joint). Ceux qui ne souhaitent pas rentrer dans les détails peuvent sauter ce paragraphe.

[4]La traduction du verbe « to plot » en français est « tracer » mais je préfère franciser le mot anglais en « plotter ». De même, la traduction du verbe « to fit » en français est « ajuster » mais je préfère franciser le verbe en « fitter ».

Le point essentiel de cet algorithme consiste à créer deux nombre aléatoires : le premier sert à déterminer après combien de temps la prochaine réaction prendra place, le second permet de choisir quelle réaction chimique se produit. Dans notre exemple, il n'y a que deux réactions chimiques : synthèse et dégradation d'une protéine. Supposons qu'il y ait $\mu$=1,2,...réactions. $a_\mu(t)$ est le produit de deux parties : la vitesse (le taux) de réaction $c_\mu(t)$ et le nombre de réactions (collisions) possibles. Si par exemple on a :

$$P_1 + P_2 \xrightarrow{c_\mu} Z$$

Où Z est l'heterodimère formé par $P_1$ et $P_2$, on a :

$$a_\mu(t) = c_\mu P_1 P_2$$

où $P_1$ et $P_2$ représentent le nombre respectif de molécules de type $P_1$ et $P_2$. Le produit $P_1P_2$ représente donc bien le nombre de collisions possibles. Voyons maintenant comment cet algorithme fonctionne. N'hésitez pas à jongler entre cette explication et le script Matlab qui reprend l'exemple simple de l'évolution d'une protéine P soumise à synthèse et dégradation dans une cellule.

Supposons que le système soit connu au temps t, ce qui signifie que le nombre de molécules de chaque type est connu (ainsi que les taux de réactions). Par conséquent, la quantité $a_\mu(t)$ est connue pour chaque réaction. $a_\mu(t)dt$ est la probabilité que, étant donné l'état du système au temps t, la réaction $\mu$ se produise dans l'intervalle de temps (t,t+dt). Appelons $a_o(t) = \sum a_\mu(t)$ la somme de tous les $a_\mu(t)$. Le terme $a_0(t)$ dt est la probabilité que n'importe quelle réaction se produise durant le temps (t,t+dt). A partir de là, l'algorithme suit les étapes suivantes :

⇒ On détermine le temps $\tau$ (après t) correspondant au temps où aura lieu la prochaine réaction. On crée un nombre aléatoire $r_1$ à partir d'une distribution uniforme compris entre 0 et 1 puis on choisit $\tau$ selon l'équation suivante

$$\tau = \left(\frac{1}{a_0}\right) ln \left(\frac{1}{r_1}\right)$$

Cette équation est une fonction de densité de probabilité exponentielle. Elle est liée au fait que la probabilité qu'une réaction ait lieu croit de manière exponentielle lorsque le temps passe.

⇒ On choisit aléatoirement la réaction qui se produit au temps $\tau$ (dans notre cas soit la synthèse soit la dégradation). On crée un nombre aléatoire $r_2$ à partir d'une distribution uniforme compris entre 0 et 1. Si ce nombre tombe entre 0 et $\frac{a_1}{a_0}$ la réaction 1 est choisie. Si il tombe entre $\frac{a_1}{a_0}$ et $\frac{a_1+a_2}{a_0}$, la réaction 2 est choisie et ainsi de suite.

⇒ Le choix de la réaction a pour conséquence de changer le nombre de molécules au temps (t+$\tau$). Dans notre cas, soit il y a synthèse soit il y a dégradation d'une protéine c'est-à-dire soit $P \longrightarrow P+1$ soit $P \longrightarrow P-1$ avec P le nombre de protéines. Ainsi la valeur d'$a_\mu$ change. L'algorithme revient au point numéro 1 via une boucle « while » et le processus est réitéré « tant que » le temps maximum définit par l'utilisateur n'est pas atteint.

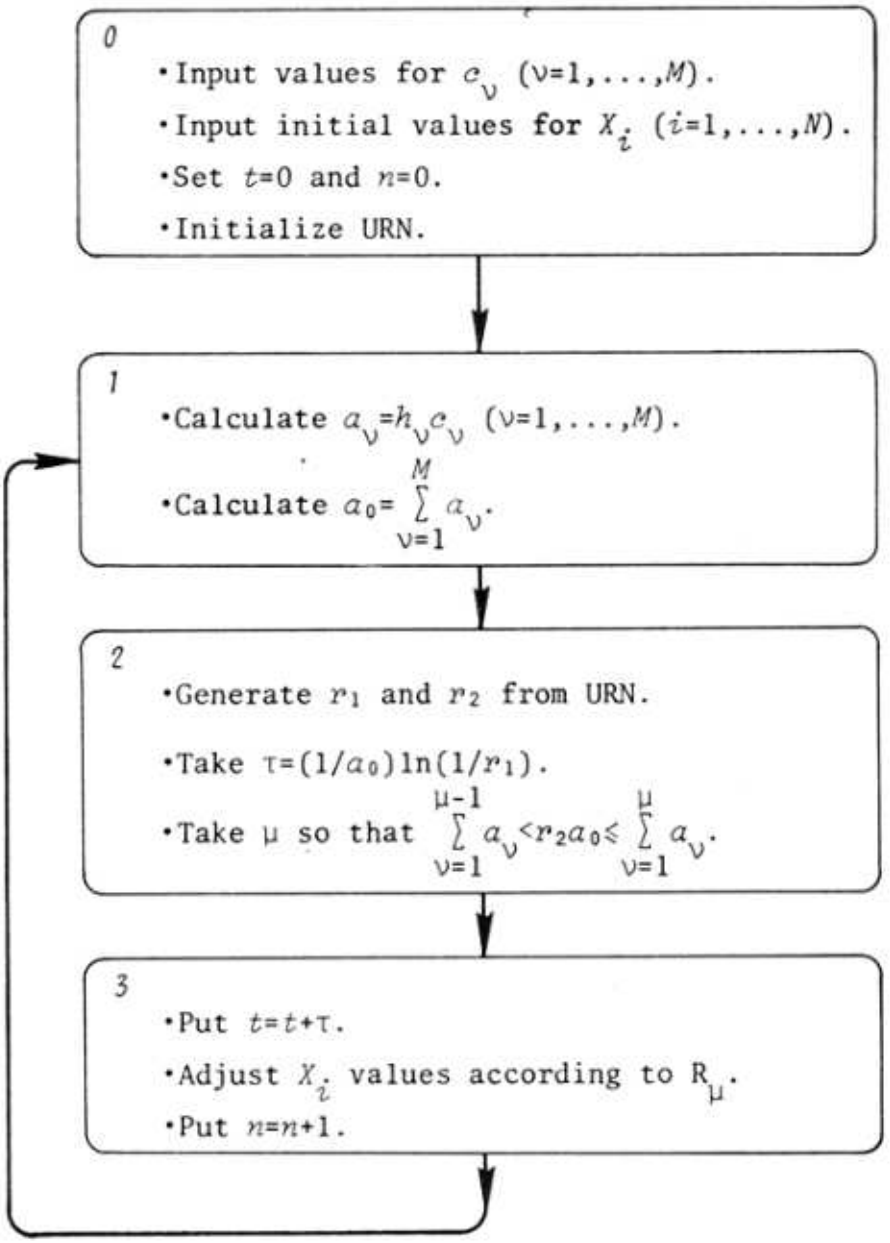



**Figure VI.9** – Schéma historique de l'algorithme de Gillespie publiée en 1977.

Dans notre cas d'école simple, on peut « plotter » , une fois la boucle « while » terminée, l'évolution du nombre de protéines P en fonction du temps et comparer le résultat avec la version déterministe (figure VI.10a). Mais on peut aussi placer une boucle for en amont de la boucle while : cette boucle for génère un profil (évolution de P en fonction du temps) différent à chaque itération c'est-à-dire pour chaque cellule ce qui nous permet de « plotter » les profils de 100 cellules par exemple (figure VI.10b). Si on extrait le nombre de molécules P atteint à la fin de chaque cinétique, on peut « plotter » la distribution de protéines P dans la population de cellules (figure VI.10c). Vous pourriez vouloir comparer (fitter) cette distribution avec la distribution obtenue expérimentalement dans le but d'identifier des paramètres ou dans le but de valider ou infirmer votre modèle.

# 8 Modèle spatial didactique

Avant de commencer mon master 2, je pensais que les scientifiques réussiraient sans doute rapidement (décennies) à créer une cellule virtuelle *in silico* (de Jong & Rechenmann, 2005). Intéressé par ce type de recherche, j'ai rencontré Hans Geiselmann et commencé à travailler sur la modélisation des réseaux de régulation génique sous sa direction. *A posteriori*, je me rends compte maintenant de ma naïveté cachée derrière l'idée de créer une cellule virtuelle.

Les jeux vidéo (par exemple World of Warcraft) d'aujourd'hui semblent capables de contenir tant de détails, tant d'informations qu'un non expert pourrait légitimement se demander pourquoi les scientifiques n'utilisent pas ce savoir faire pour modéliser une cellule. Pourquoi ne peut-on pas, pour l'instant, créer une cellule virtuelle, une sorte de modèle tridimensionnel dans lequel on intégrerait tout (déplacement brownien des molécules, transcription, traduction, réseaux de régulation génique) ? Pour deux raisons essentiellement :

⇒ la première c'est la puissance de calcul. Si ce modèle tridimensionnel avait une résolution spatiale de l'ordre de l'atome, la puissance de calcul nécessaire deviendrait énorme. De plus, dans le cahier des charges, voulez-vous que ce modèle soit capable de calculer le repliement des protéines à partir de la chaine d'acides aminés ? je vous conseille de répondre « non » car le problème du repliement (selon le dogme actuel de minimisation de l'énergie libre) est un problème que les algorithmes actuels (ainsi que ceux du futur comme les algorithmes quantiques) ne peuvent pas traiter en un temps polynomial (un temps acceptable) à cause de l'explosion combinatoire qui augmente le nombre de calculs à réaliser de manière exponentielle avec le nombre d'acide aminés. Ainsi, en l'état actuel des connaissances, le temps de calcul du repliement d'une protéine prendrait plus

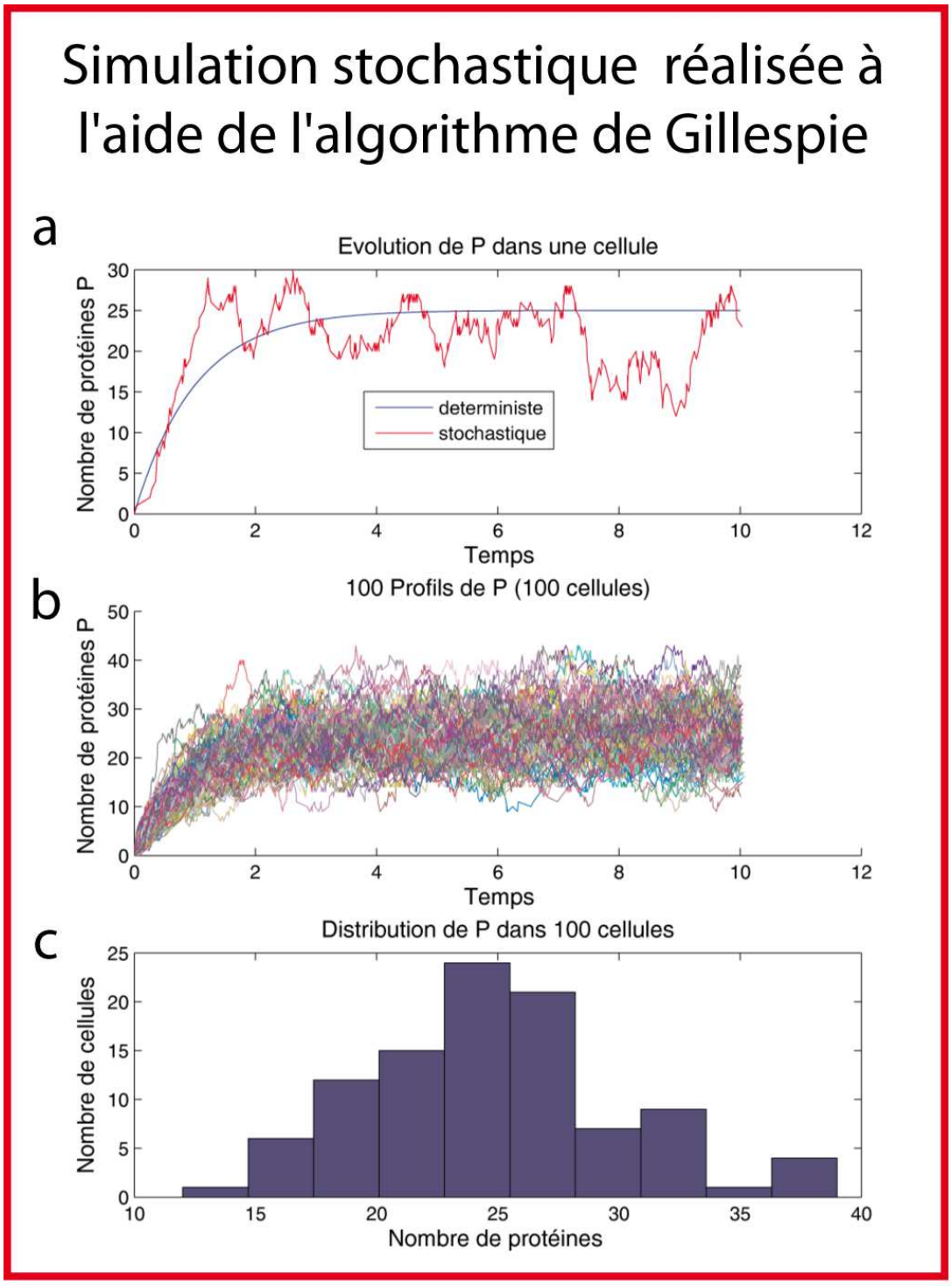



**Figure VI.10** – Simulation stochastique du modèle le plus simple de l'expression génique (pas de régulation transcriptionnelle, pas de découplage transcription/traduction). Voir le code Matlab en annexe Code12AlgorithmeGillespie.m qui gènère cette figure.

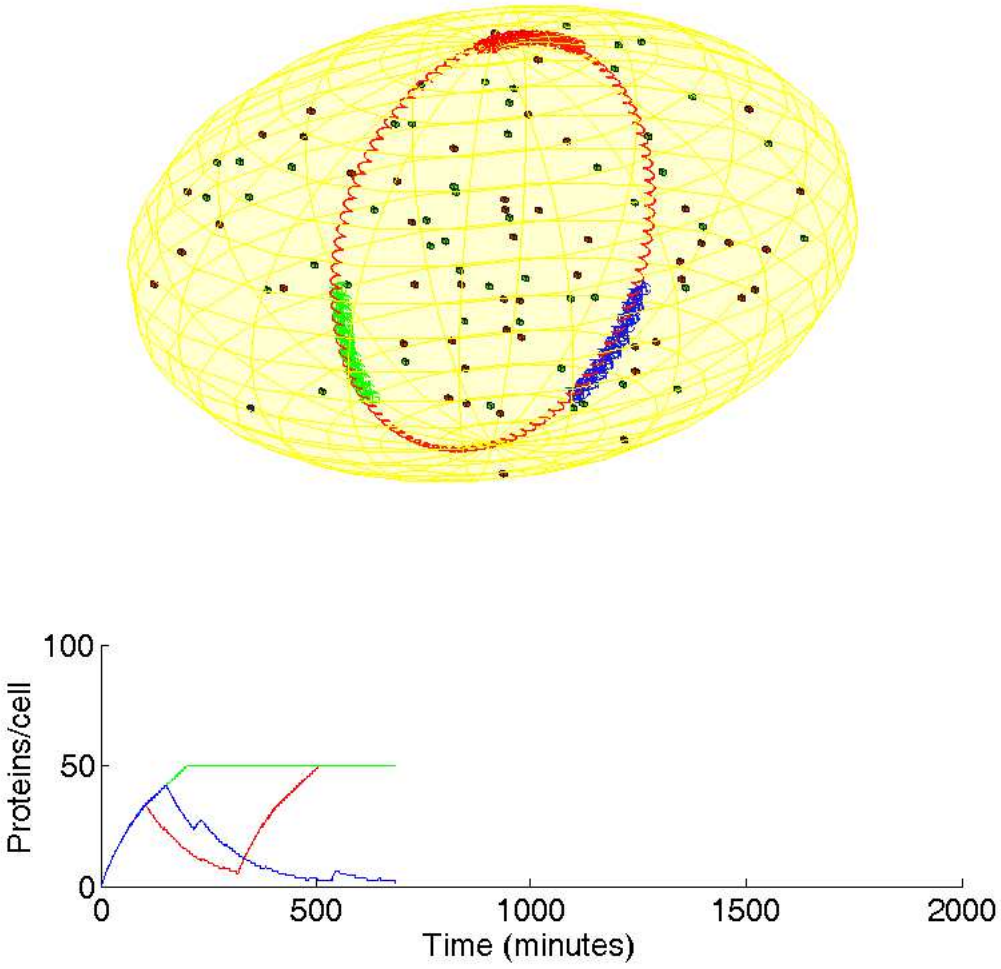



**Figure VI.11** – Video montrant ma cellule virtuelle composée de trois protéines facteur de transcription capables de se fixer sur le plasmide central et de réprimer leur expression mutuelle (selon le schéma du repressilator). Le lecteur remarquera la présence d'oscillations dans la concentration des protéines mais ces oscillations sont de mauvaise qualité à cause de la présence du bruit (expression stochastique des gènes). Le code qui génère cette video est un long script Matlab. Voir le code Matlab en annexe Code13Oscillation3D.m qui génère cette vidéo.

de temps que l'âge de l'univers (15 milliards d'années) et ce même si l'ordinateur pouvait faire une opération, un calcul par temps de Planck (le plus petit temps possible) soit $5x10^{-44}$ s.

⇒ La deuxième raison qui explique pourquoi on ne peut pas, pour l'instant, créer de cellules virtuelles, est liée aux problèmes fondamentaux relatifs à la complexité. On ne comprend pas encore comment un réseau biologique s'y prend pour être si robuste au bruit et si peu sensible aux conditions initiales.

Malgré ces deux raisons et parce que je suis un peu têtu, je me suis lancé dans la programmation

d'une petite cellule virtuelle tridimensionnelle pour « voir » jusqu'ou je pouvais aller avant d'être bloqué.

Mon modèle ne gère pas le contact (rebond) des molécules les unes contre les autres. En revanche, la cellule virtuelle peut :

⇒ créer des protéines ou métabolites qui se déplacent selon le mouvement brownien (l'algorithme ne décrit évidemment pas de manière parfaite ce processus : par exemple, il ne calcule pas la trajectoire correcte que prendrait une molécule après avoir rebondi contre la paroi de la cellule).

⇒ synthétiser différents types de protéines, chacun avec un taux de synthèse respectif.

⇒ dégrader différents types de protéines, chacun avec un taux de dégradation respectif.

⇒ faire contrôler l'expression d'un gène via la fixation d'un facteur de transcription sur le promoteur.

La manière dont sont encodés les 3 derniers points essaie de respecter au maximum les équations déterministes décrites dans le chapitre sur la modélisation.

Programmation procédurale

```
function K = promoterActivity(well, mType, mName, absName, time)
global WR;

if nargin == 5
    t = time;
else
    min1 = min(WR.wells(well).(mType).(mName).time);
    max1 = max(WR.wells(well).(mType).(mName).time);
    min2 = min(WR.wells(well).Absorbance.(absName).time);
    max2 = max(WR.wells(well).Absorbance.(absName).time);
    t = max(min1, min2):1:min(max1, max2);
end

dI = fnval(fnder(fit(well, mType, mName)), t); % signal derivative
A = fnval(fit(well, 'Absorbance', absName), t); % absorbance level
rC = reporterConcentration(well, mType, mName, absName, t);

gammaR = WR.global_parameters.([mType '_default_gamma']);
c = WR.global_parameters.plasmid_copies;

K = dI ./ (c .* A) + gammaR .* rC;

end
```

Programmation orientée objet

```
classdef Membrane < handle

    properties
        CenterCoord ;
        Rayon;
        mt=20;
        ListeEntites;
    end

    methods
        function this = Membrane(xc,yc,zc,xr,yr,zr,mt) %
            this.CenterCoord(1) = xc;
            this.CenterCoord(2) = yc;
            this.CenterCoord(3) = zc;
            this.Rayon(1) = xr;
            this.Rayon(2) = yr;
            this.Rayon(3) = zr;
            this.mt=mt;
            [x, y, z] = ellipsoid(xc,yc,zc,xr,yr,zr,mt);
            surface(x, y, z,'EdgeColor','k','FaceColor','k')
            alpha(0.1)%transparence
        end

    end
end
```

**Figure VI.12** – En programmation, il existe plusieurs paradigmes c'est-à-dire de styles fondamentaux de programmation. La programmation procédurale est (après l'écriture de petits scripts) le mode de programmation le plus facile pour débuter. La programmation orientée objet est un peu plus abstraite mais souvent plus efficace. Matlab permet d'utiliser l'un ou l'autre de ces paradigmes au choix.

La raison pour laquelle je n'ai pas pu implémenter plus de fonctionnalités (et dépasser les milliers de molécules) est avant tout liée au temps de calcul. Mon algorithme n'est certainement pas optimal. Pourtant, j'ai essayé de l'améliorer de différentes manières :

⇒ Changement de paradigme de modélisation. Une version de l'algorithme a été implémentée selon le paradigme de programmation « orienté objet » (avec des objets, des classes et des méthodes) par opposition au paradigme de programmation procédurale (avec des fonctions).

⇒ Délocalisation des calculs sur plusieurs machines en utilisant la « parallel computing toolbox » de Matlab.

Malgré cela, mon processeur (ou mes processeurs) ne peut pas calculer en un temps raisonnable (secondes) le déplacement brownien de milliers de molécules ainsi que leur dégradation et synthèse liées aux variations d'expression génique.

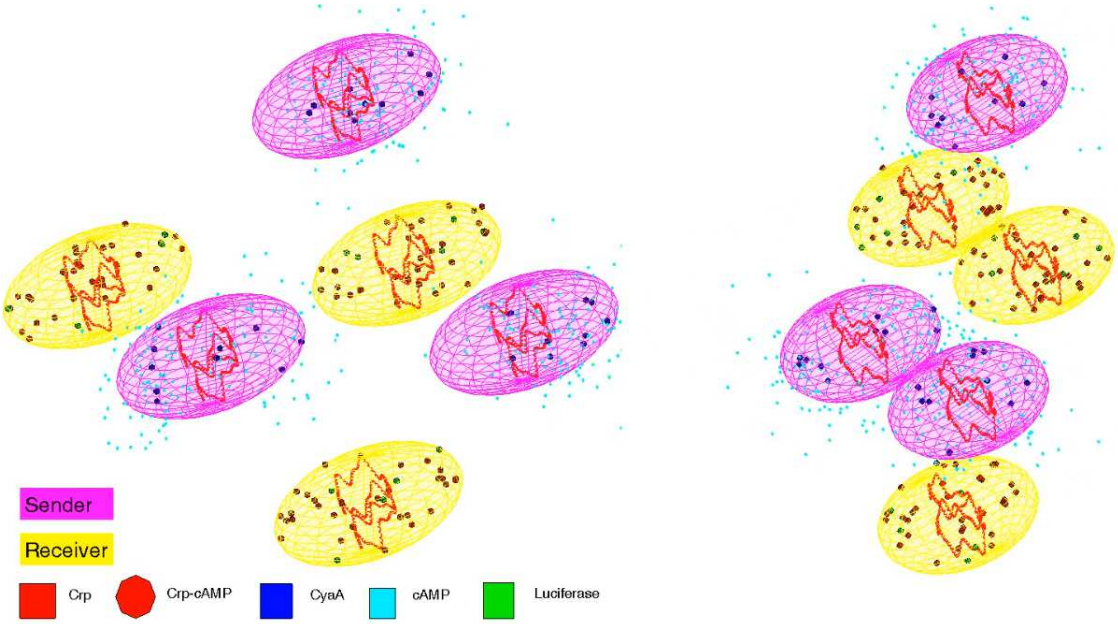



**Figure VI.13** – Video montrant un petit système de communication intercellulaire via l'AMPc que j'ai aussi implémenté[5] expérimentalement. Si vous souhaitez comprendre les détails, reportez vous à la publication sur l'AMPc. Le code qui génère cette simulation est « orienté objet ».

L'intérêt de la programmation d'une telle cellule est de deux ordres :

⇒ didactique : apprendre la programmation à travers un exercice relié à la biologie. Je ne pense pas cependant que le code que je fournis ait un quelconque intérêt scientifique (je suis sûr du contraire).

[5] Je fais exprès d'utiliser le verbe « implémenter » (champ lexical de l'informatique) en parlant d'une construction synthétique réelle pour inciter les biologistes synthétiques expérimentalistes à apprendre à penser, quand cela est nécessaire, en informaticien. Ils doivent apprendre à jongler entre les champs lexicaux de la biologie et de l'informatique/mathématique.

⇒ paradigmatique : chaque fonctionnalité que j'encode révèle la naïveté/compréhension de ma perception d'*E. coli*. Ici naïveté et compréhension sont antinomiques : naïveté car de toutes façons, notre manière de voir *E. coli* est fausse. Il y a encore beaucoup de choses cruciales que nous n'avons pas encore comprises. Compréhension car en encodant une fonctionnalité comme la régulation d'un promoteur, vous la contrôlez (via la modification des paramètres) et donc vous commencez à la comprendre. En effet, je pense que l'on ne comprend vraiment que ce que l'on contrôle (voir le chapitre sur la biologie synthétique).

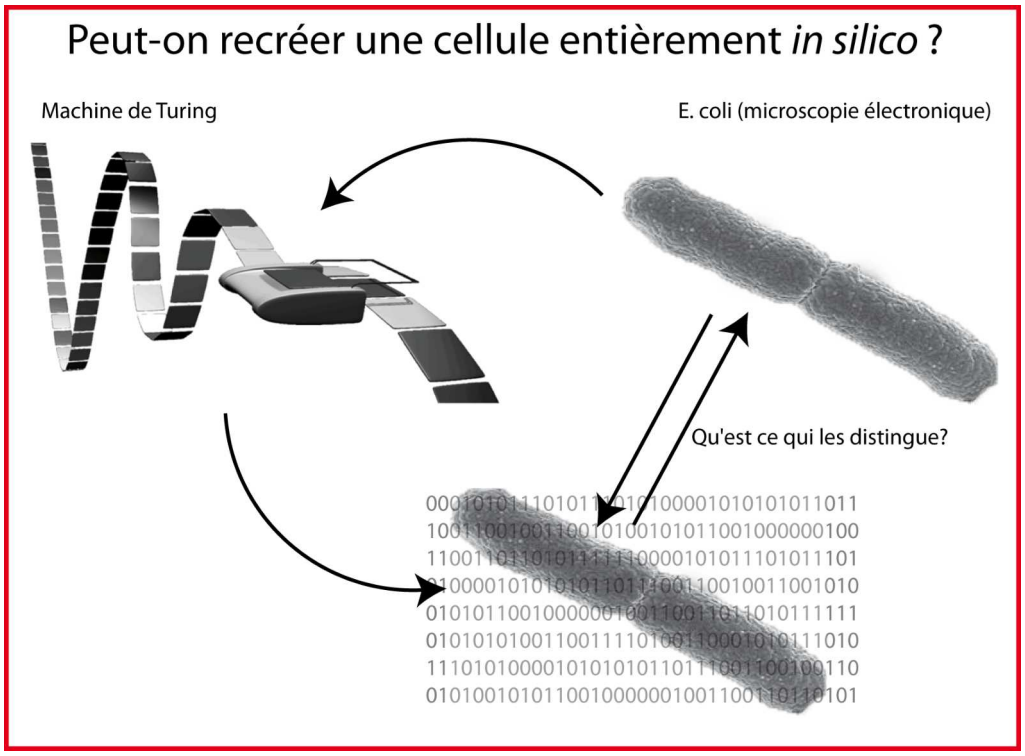



**Figure VI.14** – Peut-on recréer une cellule[6] « parfaite » *in silico* ? Peut-on la faire vivre avec une machine de Turing ? Si non, pourquoi ?

Derrière ces petits programmes « jeux en 3D » se cache une difficulté future plus profonde. Au fur et à mesure que le biologiste synthétique créera des réseaux complexes, il lui faudra avoir une estimation de plus en plus précise des paramètres. Et pour cela, il faudra créer des modèles spatiaux tridimensionnels avec une résolution de plus en plus importante. Prenons un exemple : la seule solution pour calculer précisément quelle sera l'expression d'un gène régulé par un promoteur complexe impliquant plusieurs facteurs de transcription (voir l'opéron aceBAK dans le chapitre sur la modélisation) consistera à créer un modèle pouvant simuler le contact

[6]La photo de microscopie électronique est issue de Wikipedia : article sur *E. coli*.

entre les atomes de cette adénine là (base de l'ADN) avec les atomes de cette valine ci (facteur de transcription) et de faire cela pour tous les atomes impliqués. C'est, il me semble, la seule solution pour calculer/simuler/modéliser des sorties (output) complexes.

Mais on peut aussi se demander si, dans le futur, cela aura réellement un sens d'encoder autant d'informations dans un programme écrit avec un langage de programmation qui convertit l'information en bit (software) pour un ordinateur composé de transistors (hardware). En effet, il se pourrait bien que, dans le futur, l'ordinateur chargé d'évaluer la sortie (l'output) du promoteur complexe ne soit plus un ordinateur classique mais la cellule elle-même : le Wetware.

Enfin, et je terminerai mon chapitre avec cette question : peut-on concevoir une cellule virtuelle parfaite *in silico* avec un programme fait de 0 et de 1 ? Et si cette cellule existait, qu'est que ce qui la distinguerait d'une cellule réelle ? Si on peut créer une cellule avec de 0 et des 1, on peut créer des organismes multicellulaires avec des 0 et 1, puis des hommes avec des 0 et 1. Qu'est que ce qui distinguerait ces hommes *in silico* des hommes réels ?

# Références bibliographiques


Becskei, A., & Serrano, L. 2000. Engineering stability in gene networks by autoregulation. *Nature*, **405**(6786), 590–3. 128

Bigger, JW. 1944. Treatment of staphylococcal infections with penicillin by intermittent sterilisation. *Lancet*. 127

de Jong, H., & Rechenmann, Francois. 2005. Et la cellule devint virtuelle. *La recherche*. 146

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. 2002. Stochastic gene expression in a single cell. *Science*, **297**(5584), 1183–6. 136

Gillespie. 1977. Exact stochastics simulation of coupled chemical reactions. 143

Isaacs, F. J., Hasty, J., Cantor, C. R., & Collins, J. J. 2003. Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U S A*, **100**(13), 7714–9. 128

Kupiec, Jean-Jacques. 2009. L'ADN entre hasard et contraintes. *Pour la science*. 131

Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., & Taddei, F. 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc Natl Acad Sci U S A*, **105**(8), 3076–81. 126

Mihalcescu, I., Hsing, W., & Leibler, S. 2004. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*, **430**(6995), 81–5. 130

Novick, A., & Weiner, M. 1957. Enzyme Induction as an All-or-None Phenomenon. *Proc Natl Acad Sci U S A*, **43**(7), 553–66. 129

Stewart, E. J., Madden, R., Paul, G., & Taddei, F. 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol*, **3**(2), e45. 126

Thattai, M., & van Oudenaarden, A. 2001. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A*, **98**(15), 8614–9. 137

# Chapitre VII

# La biologie synthétique

## 1 Définition

Voici la définition de la biologie synthétique trouvée sur Wikipedia :

> *« La biologie synthétique est un domaine scientifique combinant biologie et principes d'ingénierie dans le but de concevoir et construire (« synthétiser ») de nouveaux systèmes et fonctions biologiques. »* [1]

Deux mots clés sont importants dans cette définition. Ce sont les notions de construction et de fonction mais il manque, il me semble, un autre concept important : celui de contrôle d'où la définition que je propose ci-dessous.

Biologie synthétique : *construire* et donc *contrôler*, à l'échelle moléculaire, un organisme vivant ou une fonction biologique donnée. Ces notions de fonction, de construction et de contrôle cachent en réalité une quatrième notion que les scientifiques n'aiment pas trop : la finalité, le but, la téléologie. Ces buts peuvent être par exemple :

⇒ gagner de l'argent (commerce)
⇒ produire une molécule à visée thérapeutique
⇒ produire du bio-fuel
⇒ créer des bio-senseurs qui détectent puis détruisent des cellules cancéreuses
⇒ augmenter l'espérance de vie
⇒ tuer la mort
⇒ créer le surhomme.

Vous voyez que les finalités possibles associées à la biologie synthétique sont nombreuses et il ne faut pas, il me semble, limiter la finalité de la biologie synthétique au commerce comme le fait la « Royale Academy of Enginneering » aux Royaumes Unis :

[1] Source Wikipedia : article sur la biologie synthétique.

> *« The ultimate goal of synthetic biology is* ***to develop commercial applications that will benefit society****, ie to design and build engineered biological systems that process information, manipulate chemicals, fabricate materials and structures, produce energy, provide food, and maintain and enhance human health and our environment. »*[2]

Soyons clairs, je ne nie pas l'existence et l'intérêt des applications commerciales issues de la biologie synthétique. Cependant, la biologie synthétique englobe également des concepts comme le surhomme qui transcendent largement le commerce. En effet, contrôler « moléculairement » un système vivant c'est potentiellement, à terme, contrôler « moléculairement » l'homme. C'est un fait. Et le danger c'est de vouloir le cacher et surement pas de vouloir le dire. Car le cacher serait le meilleur moyen, pour les chercheurs, de se couper de la société et de perdre sa confiance.

Un autre point important à souligner c'est le changement de paradigme (manière de voir le monde) qu'introduit la biologie synthétique. Dans la biologie normale, les chercheurs font des découvertes et les publient. Ils argumentent leurs résultats, proposent une démonstration censée convaincre leurs pairs. Une preuve solide apparait lorsque le pair réussit à reproduire le résultat ce qui n'est pas, en pratique, forcement simple. En biologie synthétique, les choses se passent différemment : en construisant et en contrôlant votre organisme, votre fonction, vous démontrez en même temps votre compréhension quasi-totale du système. La publication n'a plus pour but d'argumenter, de convaincre d'un résultat mais seulement d'expliquer comment le chercheur s'y est pris. Le biologiste synthétique ne démontre pas, il montre, par exemple, la fluorescence de la bactérie qui oscille. La fonction (les oscillations) est démontrée (montrée) d'un simple coup d'œil : pas de statistiques, pas de graphiques, juste une vidéo. Le résultat est visible et compréhensible par les chercheurs de n'importe quel paradigme (voir la figure VII.6). Prenons une image : si un architecte construit un gratte-ciel : une fois construit et inauguré, il serait fou de remettre en cause son existence ou la réussite de sa construction. En biologie synthétique, c'est la même chose mais le gratte-ciel c'est la souche. La simplicité des démonstrations de la biologie synthétique repose sur l'exécution de la fonction par la souche. Ainsi les démonstrations de biologie synthétique sont très pures par rapport à celles de la biologie classique[3] car elles sont basées sur la construction et le contrôle. Et on ne comprend vraiment que ce que l'on peut construire et contrôler. La contrepartie que le chercheur paye pour pouvoir disposer de démonstrations limpides c'est l'énorme difficulté que représente le contrôle moléculaire d'un système vivant à partir de sa séquence ADN.

[2] Source Rapport de l'Académie Royale d'Ingénierie de l'UK sur la biologie synthétique p. 33. L'emphase : c'est moi qui souligne.

[3] Un an plus tard, j'ai un doute sur cette manière de présenter les choses

## 2 Construire un système vivant

La biologie synthétique peut être regardée par le prisme des technologies nécessaires pour construire un système vivant *de novo*. Un jour peut être, nous mettrons des atomes C, H, O, N, P, S etc... dans le réservoir d'une grosse machine reliée à un ordinateur dans lequel nous rentrerons la séquence ADN d'*E. coli*. Puis on appuiera sur « enter », on attendra quelques minutes puis on récupérera la bactérie créée par la machine en suivant les instructions de la séquence ADN et en utilisant les briques atomiques de base. Mais nous en sommes encore très loin. Voici 4 directions de recherche différentes et complémentaires qui permettront, peut être, d'aboutir un jour à la construction d'un organisme *de novo*.

⇒ *Travailler avec le génome le plus petit possible*. Une équipe au Japon cherche le génome minimal c'est-à-dire le génome le plus petit possible capable de coder une entité auto-réplicative en mesure d'utiliser de l'énergie. Ils ont déjà réussi à supprimer plus de 30% du génome de *E. coli* soit plus de 1300 gènes (Hashimoto *et al.*, 2005; Posfai *et al.*, 2006). D'autres chercheurs travaillent à partir de bactéries possédant un génome encore plus simple (*Mycoplasma genitalium*) de 580 000 pb soit 7 fois plus petit qu'*E. coli* (Glass *et al.*, 2006). Cependant cette bactérie ne peut pas vivre seule dans un milieu « classique » : c'est une bactérie parasite qui vit dans des cellules épithéliales. Elle utilise donc les gènes de son hôte de manière indirecte pour vivre.

⇒ *Construire un châssis cellulaire*. Certains chercheurs essaient de fabriquer des « protocellules ». Pour cela, ils fabriquent des vésicules lipidiques qui incorporent une « soupe de molécules » de la machinerie globale nécessaire pour transcrire et traduire (acide-aminés, ribonucléotides, ARN polymerase, ribosomes etc...) ainsi qu'un plasmide (ADN) contenant le gène rapporteur GFP. Les chercheurs arrivent à faire exprimer la GFP dans les vésicules lipidiques pendant plus de 4 jours (Noireaux & Libchaber, 2004). Ils arrivent même à enchâsser des pores protéiques dans les membranes pour permettre la diffusion de nutriments et améliorer les performances de ces « protocellules ». Cependant il ne s'agit pas de vraies cellules synthétiques car ces « protocellules » ne peuvent évidemment pas s'auto-répliquer.

⇒ *Fabrication du support du programme* : l'ADN synthétique. Les avancées les plus prodigieuses dans ce domaine nous viennent de l'institut de Craig Venter (dans le Maryland et en Californie). Son équipe a réussi à synthétiser un génome entièrement synthétique en suivant les instructions d'une séquence présente sur un ordinateur (*Mycoplasma mycoides*, 1 millions de pb). Puis ils ont transplanté ce génome synthétique dans le châssis cellulaire d'une autre bactérie (*Mycoplasma capricolum*) la transformant en *Mycoplasma mycoides*

(Gibson *et al.*, 2010). Il faut bien comprendre l'énorme avancée conceptuelle que représente cette expérience. Cela prouve de manière irréfutable la théorie selon laquelle il y a découplage entre la machine (le châssis cellulaire) et le programme (l'ADN) dans un système vivant, exactement comme dans une machine de Turing. On savait déjà effectuer depuis longtemps des transplantations de noyaux qui changeaient effectivement l'espèce. Cependant, rien n'empêchait d'imaginer qu'il y ait des éléments dans le noyau autre que l'ADN intervenant dans le processus.

⇒ *Programmer le réseau de régulation dans l'ADN.* C'est la difficulté majeure (les 3 points précédents posent un certain nombre de problèmes techniques difficiles mais surmontables à court ou moyen terme). La stratégie la plus commune consiste à essayer de « penser » un réseau de régulation génique comme un réseau électronique. Pour cela, les biologistes couplent de « biobricks ». Une biobrick est une séquence d'ADN qui a une fonction définie. Par exemple, une séquence promotrice répresseur ou activateur, une porte AND/OR, un terminateur, un « Ribosome Binding Site » d'une force donnée. En couplant ces « biobricks » ensemble, on peut implémenter des petits systèmes pour l'instant inferieurs à 10 gènes (des compteurs, des filtres passe-bande, des interrupteurs, des oscillateurs etc...) qui exercent leur fonction dans une bactérie comme *E. coli* par exemple. L'idée est d'utiliser le savoir faire de l'ingénierie pour standardiser ces biobricks et faciliter le travail du biologiste. Mais il faudra petit à petit augmenter la taille de ces réseaux et y inclure de plus en plus de boucles de rétroaction. Or on ne sait pas encore comment on peut contrôler un réseau complexe constitué de plusieurs dizaines, centaines ou milliers de gènes. Le biologiste se heurte à « la complexité » et il a besoin du mathématicien pour l'assister ou plutôt pour que ce dernier s'approprie ce problème et apporte des solutions. Il va falloir apprendre à contrôler le bruit et l'évolution pour que ces derniers ne soient plus des problèmes mais plutôt des solutions. Il faudra d'abord comprendre comment la nature s'y prend pour garder le contrôle du réseau et éviter la dépendance sensitive aux conditions initiales puis s'en inspirer. Et il est fort possible qu'il nous faille dépasser le paradigme un peu naïf de l'électronique.

# 3 Le « workflow » en biologie synthétique

## 3.1 Conception et modélisation

Le biologiste synthétique doit programmer une fonction dans l'ADN. Pour cela il dispose de séquences dont la fonction est déjà connue. Son rôle est d'agencer ces séquences pour obtenir la

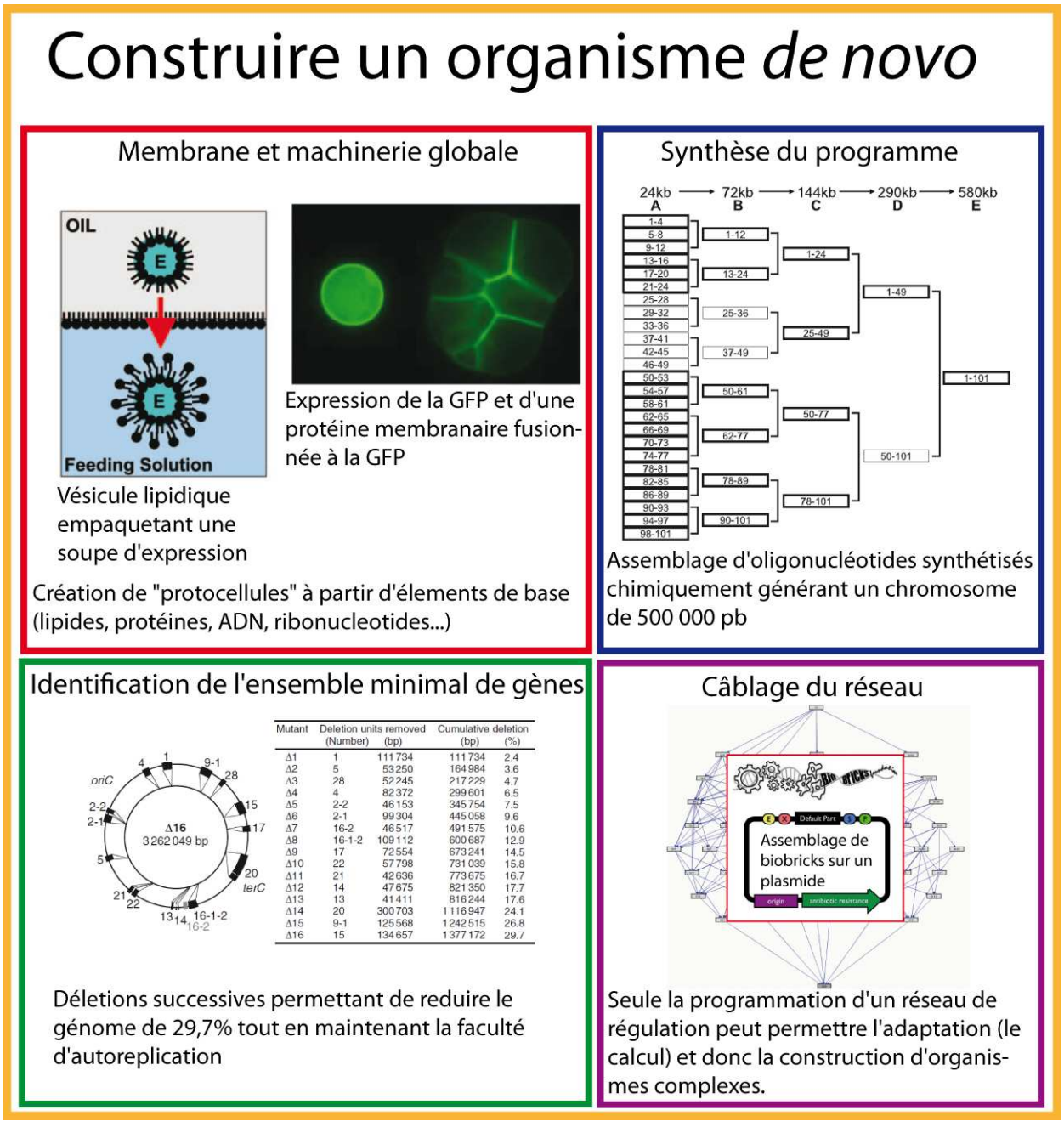



**Figure VII.1** – Des stratégies différentes et complémentaires pour construire un organisme *de novo* (certaines images proviennent des articles citées). De manière très simpliste (c'est plus compliqué en réalité), on peut dire que les carrés rouges, verts et bleus posent des problèmes essentiellement techniques alors que le carré violet (le câblage du réseau) pose un problème plus conceptuel.

fonction souhaitée. Il va, en général, connecter ensemble des promoteurs, des RBS (Ribosome Binding Site), des terminateurs de transcription avec des gènes. Ces gènes peuvent être soit des facteurs de transcription ou des protéines de transduction du signal qui sont des acteurs directs du réseau de régulation, soit des enzymes ou gènes rapporteurs qui peuvent être des « output ».

Pour le design/la conception du réseau, le biologiste peut simplement dessiner le schéma du réseau et évaluer intuitivement les conséquences. Cependant, s'il veut gagner en précision quantitative ou s'il y a des boucles de rétroaction qui rendent la dynamique non intuitive, le biologiste peut utiliser la conception assistée par ordinateur (CAO). Pour cela, il construira et résoudra soit un modèle d'équations différentielles ordinaires soit un modèle stochastique en fonction de l'estimation de l'impact du bruit sur le fonctionnement de son réseau.

Je l'ai dit dans le chapitre sur la modélisation : il me semble que le biologiste doit fixer le caractère prédictif du modèle comme priorité absolue. Ce n'est pas toujours le cas : en biologie des systèmes par exemple, il est possible (et utile) de faire de très bons modèles descriptifs mais non prédictifs.

Pour conserver le caractère prédictif : il faut faire des modèles simples mais non simplistes. Il faut limiter à tout prix le nombre de paramètres car la valeur numérique de ces derniers est souvent inconnue. Selon moi (et grossièrement), un modèle non linéaire contenant plus de 10 gènes régulés n'a que très peu de chance de générer des prédictions fiables et donc utilisables. Or la fiabilité des prédictions est cruciale car les étapes de biologie moléculaire qui suivent sont extrêmement difficiles à réaliser en pratique. Si le modèle génère une prédiction fausse, on risque de passer 6 mois à faire de la biologie moléculaire pour rien.

Il n'existe pas encore, à ma connaissance, d'environnements de développement intégrés (IDE) propres à la biologie synthétique. Selon moi, un environnement de ce type devrait associer :

⇒ Une gestion simple et intuitive de la séquence ADN avec des fonctionnalités de couplage de « biobrick » et donc de « clonage *in silico* ». La connexion aisée à des bases de données sur internet permettrait d'importer des objets/de l'information facilement (biobrick, valeur de paramètres etc...)

⇒ Des simulateurs déterministes et stochastiques faciles d'utilisation qui permettraient de prédire la dynamique du système dans des conditions standards.

⇒ Des fonctionnalités (rational design, évolution dirigée) qui assisteraient le biologiste pour « tuner » les paramètres du système.

Les IDE qui s'imposeront seront les plus simples et les plus intuitives. Les « usines à gaz » demandant un savoir trop poussé n'auront aucune chance de séduire les biologistes.

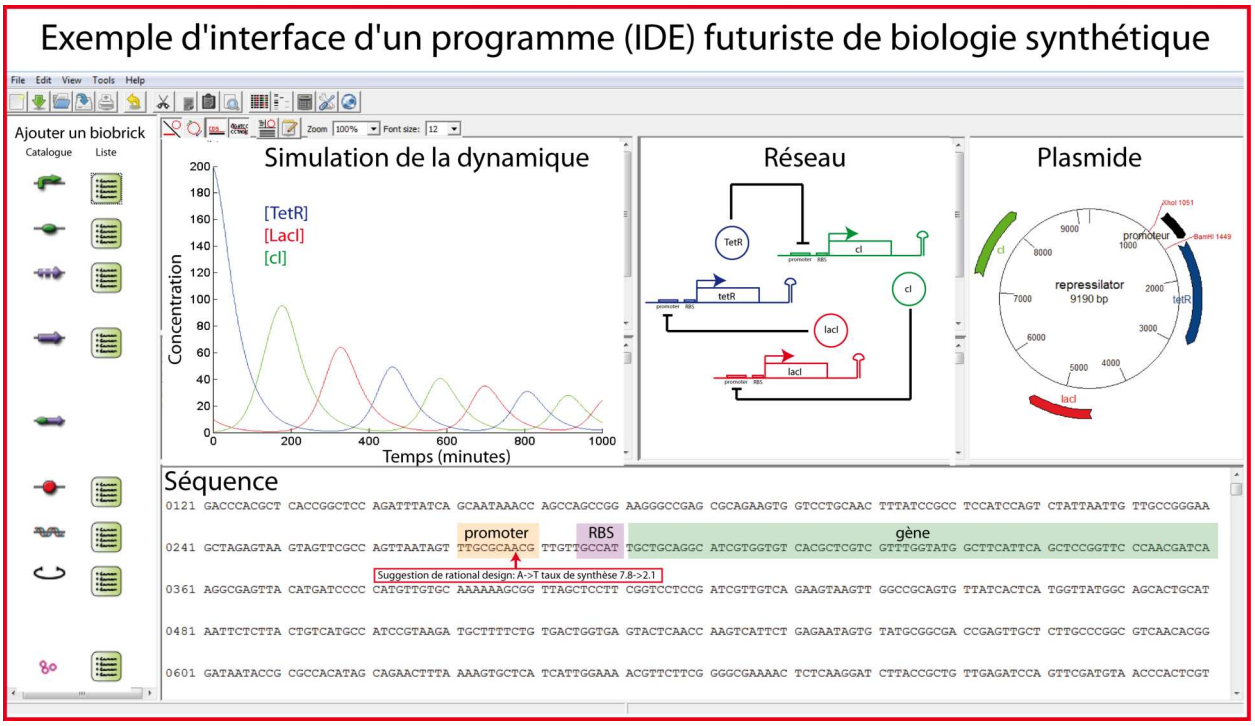



**Figure VII.2** – Environnement de développement intégré futuriste pour la biologie synthétique. Séquence ADN et algorithme de simulation y sont étroitement couplés. Evidemment, le programme n'existe pas : la figure est un montage.

## 3.2 Ecriture du programme ADN

Les techniques classiques de biologie moléculaire s'apparentent au découpage de l'ADN avec des ciseaux très spécifiques (les enzymes de restriction), puis au collage (raboutage) avec de la colle (la ligase). La PCR permet d'amplifier le gène ou la séquence ADN de son choix en rajoutant aux extrémités des petites séquences appelées site de restriction que « les ciseaux » (les enzymes de restriction) reconnaissent. On peut ensuite venir « coller/rabouter/placer » ce fragment ADN contre une autre séquence ADN, un plasmide par exemple. On appelle cela le clonage (rien à voir avec la brebis Dolly). En théorie, les clonages sont simples : il n'y a apparemment aucune raison que cela ne marche pas quand on respecte le protocole. En pratique, je connais peu de gens qui ne considèrent pas le clonage comme une lutte. Ce qui devrait prendre une ou deux semaines prend souvent des mois, au grand désespoir des biologistes. La difficulté de la biologie moléculaire est un des freins au développement de la biologie synthétique. Fabriquer un plasmide avec une fonction donnée coûte très cher, vu le temps nécessaire (et donc le salaire) que le biologiste y consacre. Cependant certains biologistes ont des doigts de fées. Ils valent chers sur le marché du travail. Pour ma part, force est de constater que la biologie moléculaire ne m'a pas dans ses petits papiers.

Une autre technique consiste à supprimer les étapes de digestion et de ligation en les remplaçant par la technique de recombinaison homologue (crossing over). Cette technique permet de modifier un plasmide mais également d'effectuer directement des modifications sur le chromosome bactérien (Datsenko & Wanner, 2000; Yu *et al.*, 2000).

Enfin un grand espoir repose également sur la synthèse *de novo* de gènes (ou de n'importe quelle séquence ADN). Cette technique devient, avec le temps, de plus en plus rapide et fiable et de moins en moins onéreuse. L'idée est simple : alléger la charge de travail du biologiste synthétique futur. Dans l'idéal, ce dernier ne devrait se concentrer que sur la partie conception et caractérisation et sous traiter la partie synthèse/écriture du programme à une entreprise. Après la phase de conception, le biologiste n'aurait plus qu'à envoyer la séquence « designée » de son plasmide à une entreprise qui se chargerait de le synthétiser.

Nous avons vu la partie « écriture » de l'ADN. Notez que la partie « lecture » est également essentielle en biologie synthétique. C'est la technologie du séquençage. Les prix et performances de la « lecture » et de « l'écriture » c'est-à-dire du séquençage et de la synthèse s'améliorent de façon exponentielle avec le temps. Le développement de ces technologies devrait pouvoir accélérer le développement de la biologie synthétique.

## 3.3 La caractérisation

La caractérisation consiste à évaluer/exécuter le programme ADN. On place ce programme ADN dans un châssis bactérien et on place la bactérie dans un milieu particulier à définir. Puis on regarde si l'exécution du programme permet d'obtenir la fonction désirée (le phénotype). Ainsi la caractérisation du programme ADN se fait toujours *in vivo*. Pour évaluer la fonction, les biologistes placent sur l'ADN des gènes rapporteurs qui fournissent une sortie (output). Il peut s'agir de la GFP (fluorescence), de la luciferase (luminescence) ou plus généralement d'enzymes qui catalysent une réaction chimique facilement quantifiable ce qui permet de remonter à la concentration de la dite enzyme ($\beta$-galactosidase, gènes de résistance aux antibiotiques etc...).

## 3.4 Les outils/appareils les plus classiques pour évaluer l'exécution du programme

⇒ le lecteur de microplaque pour une caractérisation à l'échelle de la population. Les puits d'une microplaque contiennent du milieu avec les bactéries. Le lecteur de microplaque peut mesurer des paramètres comme l'absorbance, la fluorescence ou la luminescence. Ce type de données peut facilement être comparé aux solutions d'un modèle d'équations

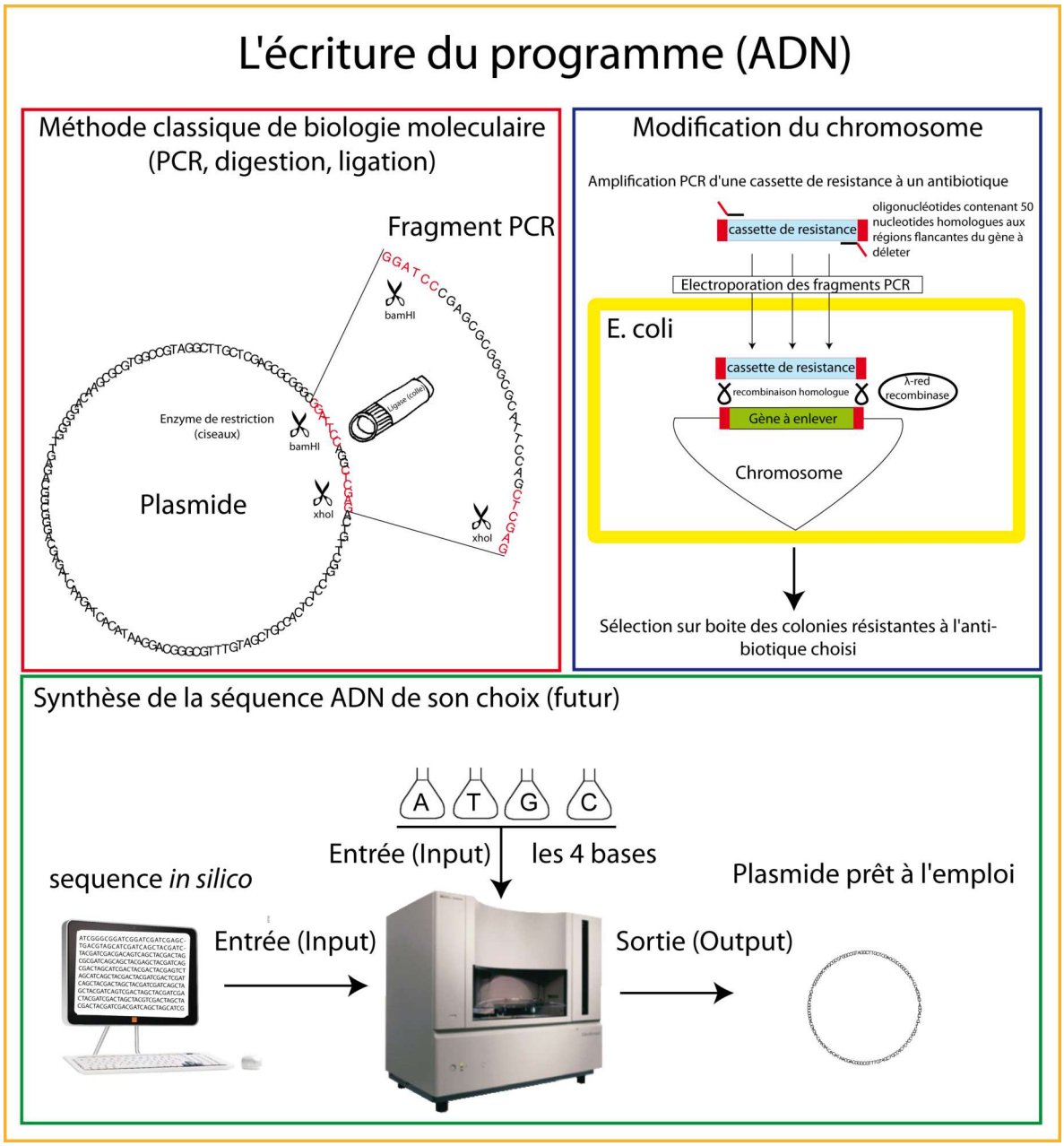



**Figure VII.3** – Techniques pour « écrire » le programme. La biologie moléculaire et l'ingénierie du chromosome sont des techniques souvent difficiles en pratique. Les biologistes synthétiques attendent patiemment la synthèse chimique de séquences de la taille d'un plasmide (5000–10 000pb)

différentielles ordinaires. Lors d'une cinétique d'expression génique, un lecteur de microplaque peut faire des mesures toutes les minutes (par exemple) ce qui permet d'obtenir des données d'expression avec une haute résolution temporelle.

⇒ Le cytomètre en flux fournit des données sur chaque bactérie. Par exemple, la taille

(l'aire), ou la fluorescence. Ces données permettent d'étudier la distribution de fluorescence des cellules ou la distribution des aires. Elles sont donc particulièrement propices à l'évaluation des prédictions générées par un modèle stochastique.

⇒ La microscopie à fluorescence. Elle permet de générer des images ou des films (cinétiques) d'expression génique à l'échelle de la cellule isolée. Une seule image peut contenir plusieurs types d'informations en jouant sur la superposition des couleurs : par exemple, on peut générer une image contenant l'information du contraste de phase, l'information de la fluorescence verte pour « monitorer » la concentration de la GFP et l'information de la fluorescence rouge pour « monitorer » la concentration de la RFP. Des logiciels très puissants (technologie scannR chez Olympus) permettent de détecter et comptabiliser en quelques secondes les paramètres (aire, fluorescence...) de milliers de cellules isolées à partir des images. Cela permet, comme pour le cytomètre en flux, de générer des graphes de dispersion (scatter plot) ou des distributions mais avec, en plus, l'énorme avantage de pouvoir vérifier ses résultats en regardant directement les images des cellules.

⇒ La microfluidique : couplée à la microscopie à fluorescence, la microfluidique est un domaine très prometteur pour effectuer de nombreuses mesures en parallèle à l'échelle de la cellule isolée.

⇒ Tous les dispositifs expérimentaux que les chercheurs peuvent imaginer : par exemple, un chercheur peut utiliser comme « output » la mobilité des bactéries. Il fabrique alors un dispositif expérimental en utilisant une gélose nutritive semi-solide où les bactéries « nagent ». En fonction de la séquence ADN programmée, ces dernières nagent plus ou moins vite. Il suffit alors de mesurer la distance qu'elles parcourent et de s'en servir comme données reflétant/rapportant l'expression de tel ou tel gène nécessaire à la mobilité.

## 3.5 Optimisation de ces 3 étapes

Pour accélérer le développement de la biologie synthétique, chacune de ces 3 phases (conception, écriture et caractérisation) devrait être optimisée, standardisée, si possible automatisée le tout sous forme d'approche haut débit.

Prenons l'exemple du repressilator (voir le repressilator dans le chapitre sur la modélisation). Ce plasmide code un système qui génère des oscillations dans la concentration des protéines. Il est composé de 3 promoteurs et de 3 gènes. S'il fallait recommencer sa construction en privilégiant une approche haut-débit optimisée, voilà comment je m'y prendrais.

Utilisation de l'IDE futuriste décrite ci-dessus. Le biologiste ajoute les « biobrick » par

[4]L'image du système microfluidique est une image issue de Wikipedia.

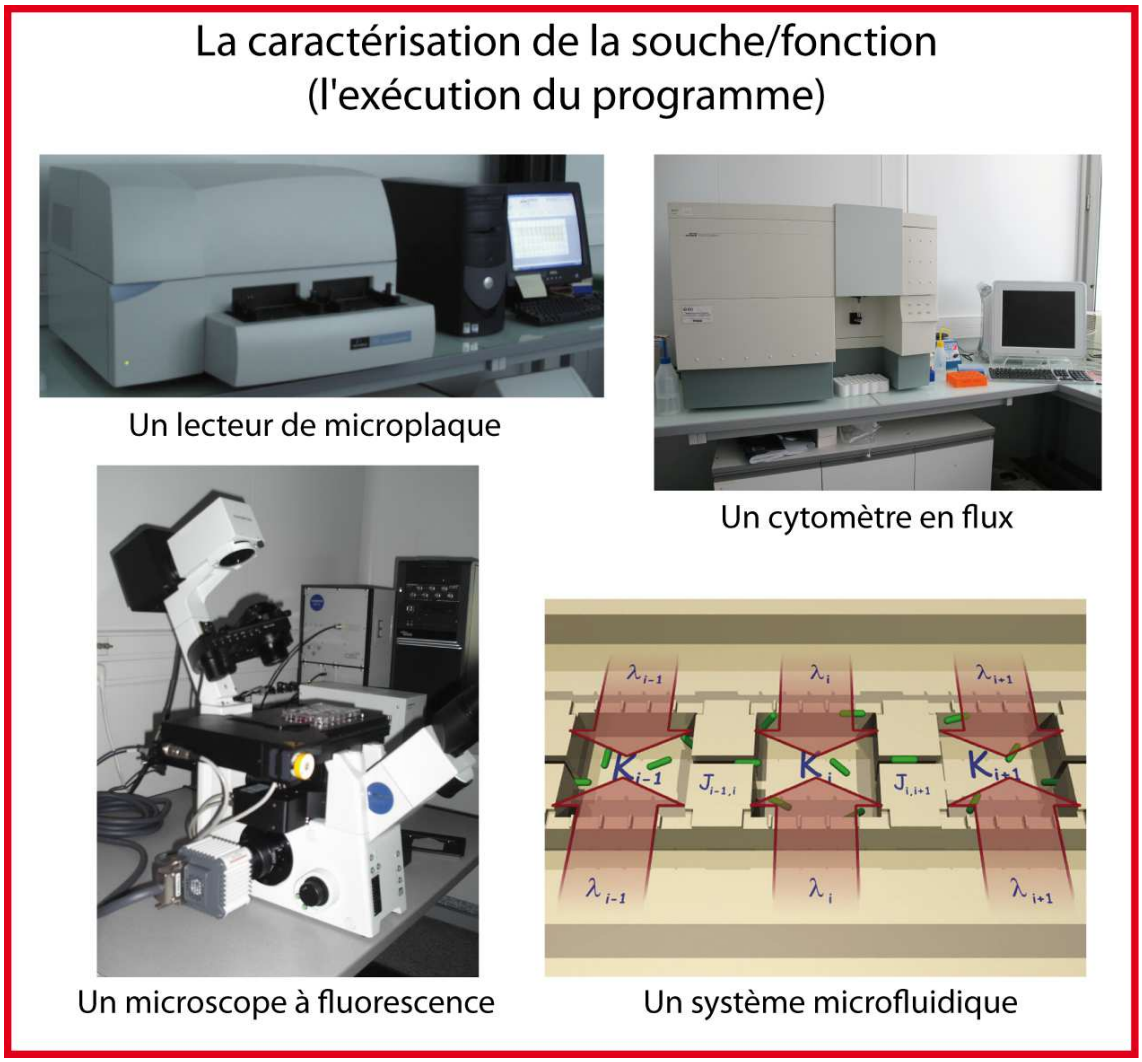



**Figure VII.4** – Photo de quelques outils permettant la caractérisation d'un programme ADN[4].

glisser/déposer ce qui crée le squelette de base du plasmide. Puis un algorithme de simulation calcule la dynamique du système. Le logiciel assiste ensuite le biologiste pour détecter les nucléotides qui sont les points chauds critiques influant sur les paramètres du système. Ce sont les nucléotides des promoteurs, des RBS etc... qui modulent les taux de synthèse, l'affinité des facteurs de transcription sur les promoteurs, etc... Une fois une dizaine de points chauds choisis par l'utilisateur, le logiciel génère toutes les combinaisons de séquences possibles. Par exemple, une centaine. L'algorithme de simulation évalue à nouveau ces 100 séquences et met éventuellement de coté les séquences qui n'auraient aucune chance de générer le phénotype voulu. Puis l'expérimentateur envoie de manière semi-automatisée (gestion par le logiciel) les séquences à une entreprise pour synthèse. Cette dernière renvoie une microplaque contenant soit

les 100 plasmides, soit les 100 souches bactériennes contenant les dits plasmides. Le biologiste peut alors caractériser les 100 séquences ADN avec un dispositif microfluidique contenant le nombre de canaux nécessaires pour tester les 100 souches en une seule cinétique (en parallèle). Il suffit ensuite d'identifier les souches dont la concentration de la GFP oscille à la fréquence et l'amplitude souhaitée.

Au jour d'aujourd'hui, ma prédiction est qu'il faut plusieurs années (entre 3 et 10 ans) à une équipe de 3–5 personnes pour designer, construire et caractériser ces 100 plasmides à la main sans même avoir l'assurance d'obtenir, à la fin du temps imparti, un plasmide « parfait » c'est à dire générant des oscillations à telle fréquence et de telle amplitude par exemple. Ce qu'il faut c'est donc travailler avec des ingénieurs/thésards/chercheurs à l'optimisation du workflow décrit ci-dessus de telle manière à réduire les temps à un niveau acceptable : quelques mois maximum pour caractériser le phénotype de 100 souches contenant 100 programmes bien définis.

## 3.6 Le charbon

L'IGEM (International Genetically Engineered Machine compétition) est une compétition internationale de biologie synthétique entre étudiants. Cette compétition, lancée par le MIT en 2004, a pour but de développer la biologie synthétique. Elle compte, en 2011, 164 équipes soit plus de 1000 étudiants.

La construction de souches[5] réalisant des fonctions plus ou moins loufoques est confiée aux étudiants IGEM[6] pour plusieurs raisons :

⇒ La biologie synthétique représente le futur/la science-fiction et les jeunes ne s'y trompent pas (un peu comme l'informatique et la programmation dans les années 70–80).

⇒ Cette compétition est une excellente publicité pour le développement de la biologie synthétique

⇒ il faut de l'imagination et ne pas être trop formaté pour imaginer de nouvelles souches, de nouvelles fonctions. En cela les étudiants sont très précieux car ils apportent de nouvelles idées.

⇒ la biologie moléculaire reste encore très difficile. L'acharnement nécessaire est confié à la jeunesse.

Certains pourraient croire que la biologie synthétique est confiée aux étudiants car elle relève de l'ingénierie et non de la science. On pourrait penser qu'il s'agit simplement de l'application de

[5]Bien sûr, c'est les étudiants qui déterminent leur sujet

[6]Mais pas seulement bien sûr et beaucoup de thésards, post-docs, chercheurs font aussi de la biologie synthétique.




**Figure VII.5** – Présentation de la compétition IGEM. La photo du bas a été prise lors d'une présentation des midi-MINATEC à Grenoble le 2 septembre 2011.

protocoles standards, relativement simples, pour « jouer » à fabriquer des souches loufoques. On pourrait croire également que les chercheurs, ayant déjà largement déblayé le terrain, auraient confié la tache de poursuivre la biologie synthétique à des ingénieurs pour que ces mêmes chercheurs puissent se focaliser sur des choses plus utiles ou plus abstraites. Cette manière de pensée potentielle ne reflèterait pas selon moi la réalité. La biologie synthétique ne correspond pas à un transfert de compétence des chercheurs aux ingénieurs. La biologie synthétique amorce un changement de paradigme.

L'ancien paradigme ou cadre de recherche du 20$^{ème}$ siècle fixait comme objectif la découverte de nouvelles entités (ADN, protéine, ARN) ou de nouveaux mécanismes ou fonctions de protéines. La dernière grande découverte de ce type a marqué la fin du 20$^{ème}$ siècle : la découverte de l'ARN interférence. Le nouveau paradigme, celui du 21$^{ème}$ siècle, c'est celui du système, du réseau, de la complexité et de d'information. Maintenant que l'on connait presque tous les élé-

ments de la cellule, il faut faire le chemin inverse et essayer de remettre ces éléments ensemble en reconstruisant le réseau. L'objectif ce n'est plus la découverte de nouvelles entités ou de nouveaux mécanismes, l'objectif c'est le contrôle du système. Ce contrôle fournit une preuve de qualité supérieure de la compréhension d'un système biologique. Et bien caché derrière le contrôle se cache bien sûr (je le répète) la notion de finalité c'est-à-dire la téléologie.

Le but caché derrière la construction d'une souche loufoque, c'est la construction et le contrôle de souches de plus en plus complexes : c'est-à-dire accroitre les capacités computationnelles d'un système vivant (le wetware). Les progrès futurs qui vont changer nos vies résident dans l'accroissement du contrôle de ces capacités computationnelles et surement pas dans la découverte d'une énième entité ou d'un énième mécanisme inconnu. Le changement de paradigme est en marche et on sait depuis Thomas Kuhn[7] que cela passe souvent par le renouvellement des générations de chercheurs. Les anciens partiront à la retraite [8] et les nouveaux seront les étudiants IGEM qui font, je le crois, souvent partie des très bons étudiants : motivés[9] , aptes au travail en équipe et prêts à se mobiliser pour un projet qu'ils doivent mener eux-mêmes dés le premier cycle universitaire. Et, vous l'aurez compris, cette génération d'étudiants aura été « élevée » avec les notions de systèmes, de réseaux, de bruit, de complexité, de chaos, d'informations, etc... Ils auront baigné dans la fournaise de l'interdisciplinarité réelle (biologie, physique, mathématique, chimie, informatique etc...). Et c'est dans cette fournaise que sont susceptibles de se faire les découvertes fondamentales qui changeront le plus la vie des hommes.

# 4 Listing de quelques fonctions existantes en biologie synthétique

Ma liste n'est pas exhaustive et a principalement pour but de donner des références bibliographiques à ceux que cela intéresserait. La figure VII.7 fournit quelques schémas extraits des publications citées.

⇒ L'interrupteur génétique (le toggle switch). Cette construction, basée sur une double répression, permet de maintenir « on » l'expression d'un gène et « off » le deuxième et réciproquement (Gardner *et al.*, 2000). Voir figure VII.7A. Voir le modèle correspondant

[7] Thomas Kuhn, *La structure des révolutions scientifiques*, voir le chapitre d'épistémologie.

[8] Excusez-moi si cela peut sembler irrévérencieux. Ce n'est pas mon intention : seul le concept Kuhnien m'intéresse ici.

[9] Notez que je n'ai jamais participé à une compétition IGEM : mon avis est donc relativement impartial. Cependant, il y a une équipe IGEM dans mon laboratoire et j'ai pu constater la grande motivation de ses membres.

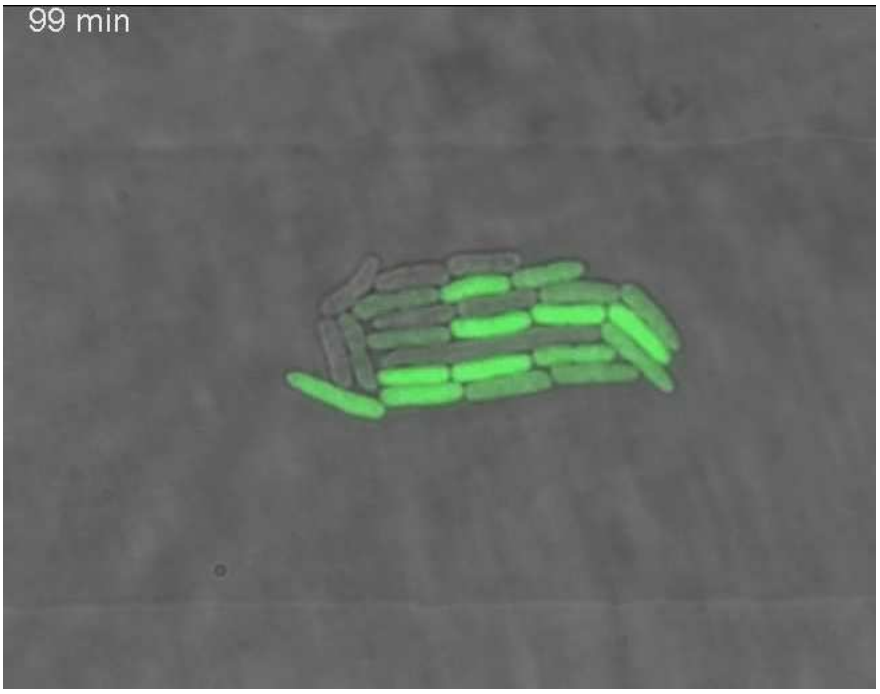



**Figure VII.6** – Vidéo issue de la publication (Stricker *et al.*, 2008) ajoutée ici avec l'autorisation des auteurs. Cette vidéo illustre bien ce qu'est la biologie synthétique. Rendez-vous compte : les oscillations de la GFP que vous regardez ont été encodées sur un morceau d'ADN par un être humain. Les auteurs peuvent même contrôler la période des oscillations (entre 10 et 40 minutes).

au toggle switch. Voir le chapitre sur les systèmes dynamiques.

⇒ L'oscillateur (le repressilator). Cette construction, basée sur une triple répression, permet de faire osciller la concentration de 3 protéines répresseurs (Elowitz & Leibler, 2000). Voir figure VII.7B et figure VII.6. Voir le chapitre sur la modélisation.

⇒ Les oscillateurs multicellulaires. Il s'agit d'un système synthétique incluant une molécule de communication extracellulaire permettant de synchroniser les oscillations à l'échelle de la population (Danino *et al.*, 2010). Voir figure VII.7C.

⇒ Système oscillant proie–prédateur : Il s'agit d'un système de double communication intercellulaire mettant en jeu 9 gènes. C'est actuellement le système le plus complexe (en terme de nombre de gènes) que je connaisse (Balagadde *et al.*, 2008). Voir figure VII.7D.

⇒ La programmation de patterns multicellulaires : une bactérie « sender » produit un composé extracellulaire l'AHL (Acyl-Homoserine Lactone) qui est détecté par une bactérie « receiver ». Les cellules « receiver » expriment ou non la GFP en fonction de la concentration de l'AHL et donc de la distance qui les séparent des « senders » ce qui forme des jolies patterns (Basu *et al.*, 2005). Voir figure VII.7E.

⇒ Le filtre passe-bande. L'utilisation d'un « incoherent feed forward loop » permet d'ex-

primer un gène dans une fourchette de concentration bien précise d'un activateur. Il y a une borne vers le haut et une borne vers le bas (Sohka *et al.*, 2009). Voir figure VII.7F.

⇒ La mémoire : le couplage de boucles de retroaction positive peut générer un système bistable (phénomène d'hystérèse) pouvant servir de mémoire (étudié dans le chapitre sur la modélisation) (Chang *et al.*, 2009). Voir figure VII.7G.

⇒ Le compteur : un système synthétique capable de compter jusqu' à 3 événements d'induction (Friedland *et al.*, 2009). Voir figure VII.7H.

⇒ Détection de frontière : il s'agit d'un programme ADN dans *E. coli* qui lui permet de détecter et marquer les lignes ou frontières dans une image en secrétant un pigment noir à cette endroit seulement (Tabor *et al.*, 2009). Voir figure VII.7I.

⇒ Appareil photo : la bactérie Escherichia coli est transformée en un appareil photo de 100 millions de pixel. Un récepteur stimulé par la lumière déclenche une cascade de réaction qui entraine l'expression d'un gène (*lacZ*) produisant un pigment noir (Levskaya *et al.*, 2005).

# 5 Le futur en biologie synthétique

## 5.1 Petit détour par la théorie des graphes

En théorie des graphes, un graphe correspond à un réseau abstrait et n'a donc rien à voir avec un graphique. Un graphe est un ensemble de points (pour nous les gènes) reliés par des arêtes (les interactions). Etant donné que, dans un réseau biologique, ces arêtes sont orientées, on parle de graphe orientés et je parlerai de flèches plutôt que d'arêtes. Accroitre les capacités computationnelles d'un système biologique synthétique passe par l'augmentation de la taille et de la complexité des graphes. Autrement dit, il faudra réussir à augmenter :

⇒ le nombre de gènes N

⇒ le nombre de flèches F unissant ces gènes.

La connectivité C d'un réseau est défini comme le rapport du nombre de flèches sur le nombre de gènes soit C=F/N. Prenons le cas simple où on ne sépare pas l'activation et l'inhibition (autrement dit une flèche peut représenter soit l'une soit l'autre) :

⇒ Si C=0. Il n'y a pas d'interaction et l'expression des gènes de l'organisme est dite constitutive. Par exemple, on peut vouloir faire produire à une souche (usine bactérienne) un métabolite particulier qui nécessite une chaine linéaire de 5 enzymes. L'expression de chacune de ces 5 enzymes est indépendante. Il n'y a pas de régulation spécifique. On place

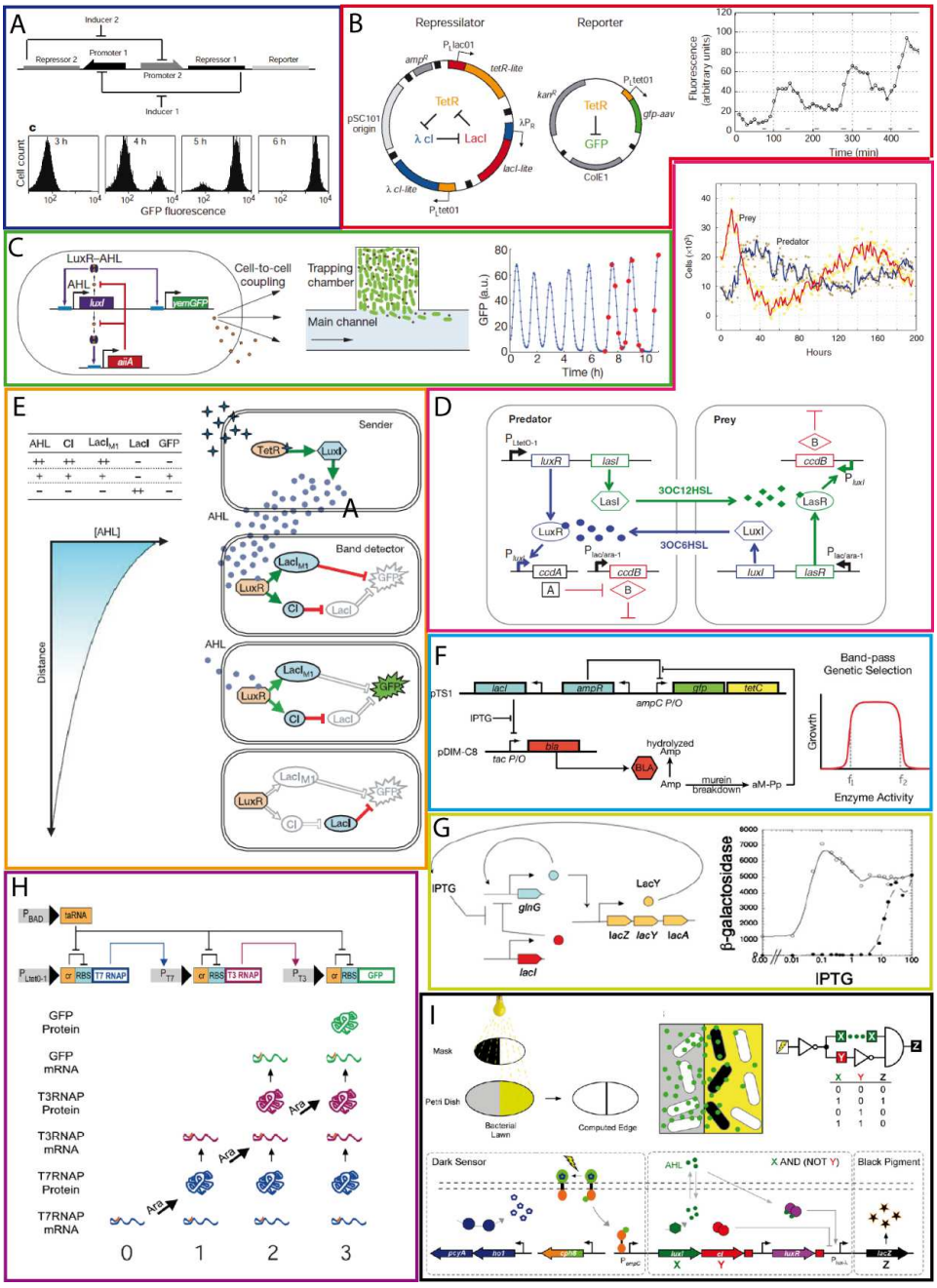



**Figure VII.7** – Quelques figures issues des publications citées pour vous donner une idée de l'allure des schémas trouvés en biologie synthétique. N'essayez pas de les comprendre ici. Référez vous plutôt directement à la publication. A : l'interrupteur (le toggle switch) : remarquez la bistabilité qui apparait sur la distribution de fluorescence lors du switch. B : l'oscillateur (le repressilator). Remarquez les oscillations de la GFP sur le graphe. C : oscillateur synchronisé entre les cellules. D : Système proie prédateur type « lapin–renard » ; remarquez les oscillations de la taille des 2 populations. E : Système qui génère des patterns en fonction d'un gradient d'homo-serine lactone (AHL), une molécule de communication intercellulaire. F : filtre passe-bande « tunable » ; repérez la forme caractéristique de la courbe rouge. G : bistabilité induite par des boucles de rétroaction positive couplées ; repérez le pattern d'hystérèse caractéristique d'une mémoire. H : le compteur. I : la détection de frontière.

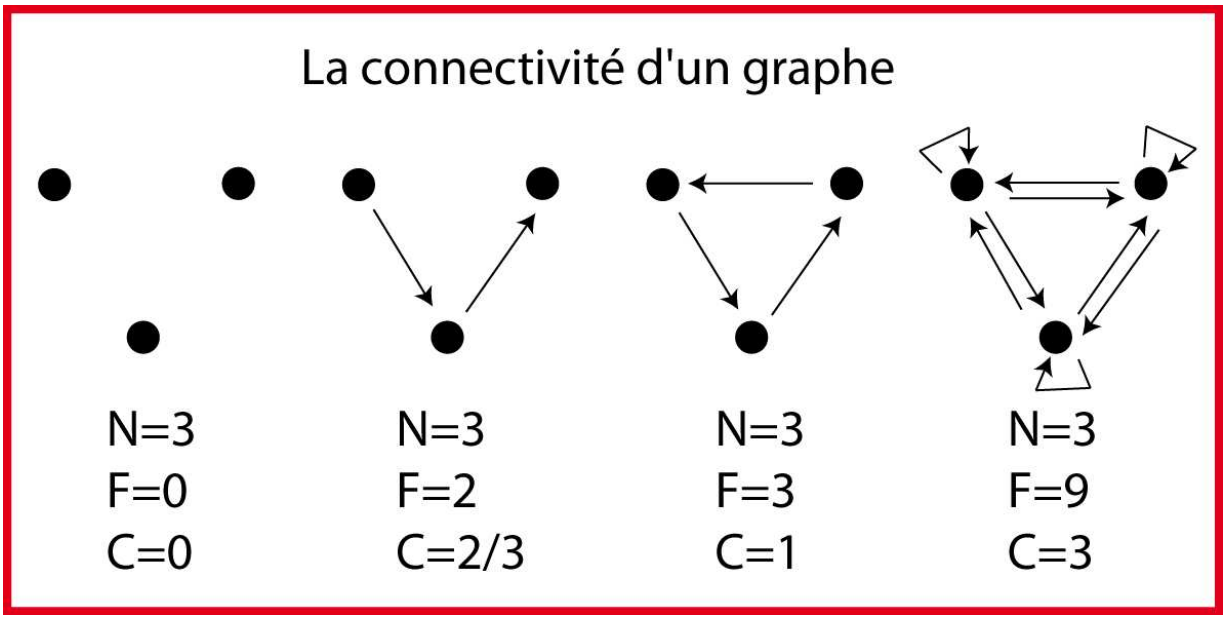



**Figure VII.8** – La connectivité d'un réseau.

devant chaque gène un promoteur constitutif de la force adéquate de manière à optimiser la production du dit métabolite.

⇒ Si C=F/N= (N-1)/N. Dans ce cas, il y a une flèche de moins que le nombre de gènes. Si chaque gène est connecté au réseau alors la dynamique est parfaitement intuitive car il ne peut pas y avoir de boucles de rétroaction. Cela permet potentiellement de se passer d'un modèle mathématique en ce qui concerne le comportement dynamique du système.

⇒ Si C=1. Il y a autant de flèches que de gènes. Si chaque gène est connecté au réseau, Il existe forcement une boucle de rétroaction rendant potentiellement la dynamique du système non intuitive. Un modèle mathématique risque d'être indispensable.

⇒ Si C=N (car (N(N-1)+N)/N=N). Le système contient N(N-1)+N flèches. Dès que le nombre de gènes est supérieur à 2, la dynamique du système est impossible à prédire sans modèle.

Il y a donc deux enjeux en biologie synthétique :

⇒ Augmenter N tout en fixant C<1. Cela permettra d'apprendre à travailler avec des réseaux de grande taille, par exemple contenant plus de 10, 20, 50, 100 gènes. La dynamique restera totalement intuitive. La difficulté est plutôt technique : il faut réussir à contrôler, fixer, optimiser l'expression d'un grand nombre de gènes dans une cellule pour obtenir la fonction souhaitée.

⇒ augmenter C tout en fixant N. La dynamique n'est plus intuitive et il faut absolument

un modèle mathématique. La difficulté est conceptuelle : comment augmenter C tout en gardant le contrôle d'un système contenant un grand nombre de boucles de rétroaction ? comment faire pour se prémunir de la dépendance aux conditions initiales ? faut-il utiliser les techniques de l'électronique : modularité, rétroactivité (Del Vecchio *et al.*, 2008; Sauro, 2008) ? Ou bien faut-il comprendre comment la nature s'y prend et utiliser sa méthode (évolution ?) ? La nature sait-t-elle d'ailleurs travailler avec des hautes valeurs de C ?

## 5.2 La prochaine génération de réseaux synthétiques

Certaines idées des paragraphes qui suivent se basent sur l'excellente publication *Next generation synthetic gene network* de Timothy Lu paru en 2010 dans Nature Biotechnology (Lu *et al.*, 2009). Je conseille vivement sa lecture car c'est une niche à bonnes idées. D'ailleurs, certaines de ces idées sont des pistes que je garde en tête pour mes recherches futures. Et je serais donc très heureux de discuter avec un lecteur qui réfléchirait à fouiller en direction de l'une de ces perspectives.

### 5.2.1 Filtres « tunable »

La technique actuelle en biologie synthétique est la suivante : le biologiste connecte des séquences d'ADN entre-elles (« plug and play »), chacune ayant une fonction donnée de manière à créer une séquence nouvelle, plus complexe. Souvent cette séquence ne fournit pas du premier coup la fonction attendue, il faut « tuner » certains paramètres en modifiant soit rationnellement (rational design) certains nucléotides (par exemple pour diminuer la force d'un promoteur) soit en utilisant des techniques d'évolution dirigée (voir chapitre sur l'évolution). Dans les deux cas, il faut venir modifier la séquence ADN. Et que ce soit nous ou l'évolution qui le fasse, cela prend toujours du temps. L'idée c'est donc de créer des composants (des filtres) dont les paramètres sont « tunables » beaucoup plus facilement c'est-à-dire sans avoir à modifier la séquence ADN. C'est le cas par exemple du filtre passe-bande dont j'ai déjà parlé ci-dessus (Sohka *et al.*, 2009) dont la bande (la zone d'activité) peut être ajustée sur plus de 4 ordres de grandeur en faisant varier un stimulus externe (inducteur).

### 5.2.2 L'utilisation du bruit

Le bruit est souvent considéré comme gênant alors qu'on soupçonne les populations de cellules de l'utiliser pour « computer » des « outputs ». L'idée est d'apprendre à utiliser et contrôler le bruit. Il faut pour cela créer des générateurs de bruit « tunables ». Le couplage

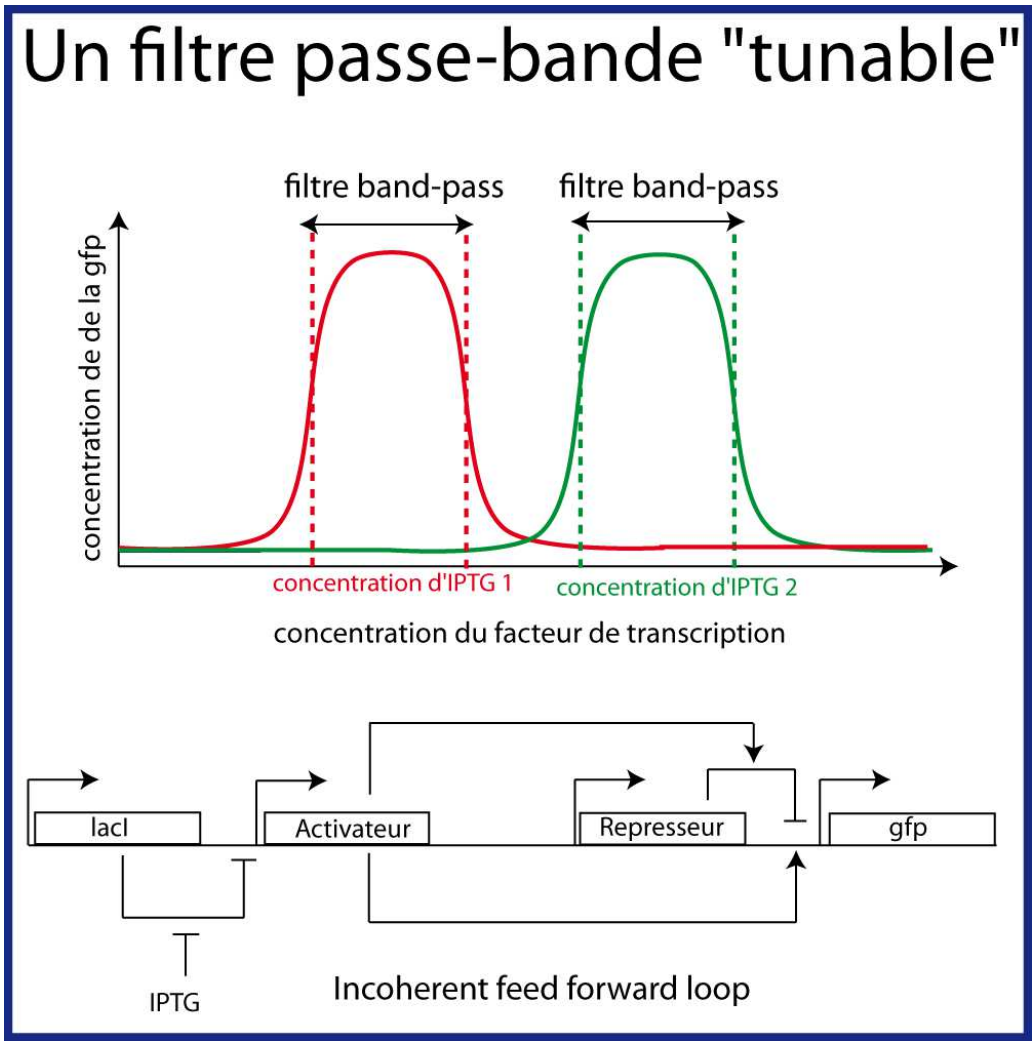



**Figure VII.9** – Des filtres passe-bande dont les paramètres « tunables » pourraient éviter d'avoir à modifier la séquence ADN. Le filtre ci-dessus est basé sur un « incoherent feed forward loop ».

de motifs propageant le bruit pourrait générer une grande diversité dans la population clonale de cellule. Diversité qui pourrait peut-être démultiplier les capacités de calcul du Wetware. Comment ? Je ne sais pas. Mais les capacités de calcul sont souvent réduites/cantonnées au réseau intracellulaire et les millions de cellules dans un puits d'un lecteur de microplaque ne servent qu'à amplifier le signal. Or en générant de la diversité grâce au bruit dans une population clonale, on peut peut-être faire en sorte que chaque cellule résolve un calcul particulier. La population traiterait alors les calculs de manière massivement parallèle. Peut-être que cette question théorique intéressera les mathématiciens / modélisateurs ?

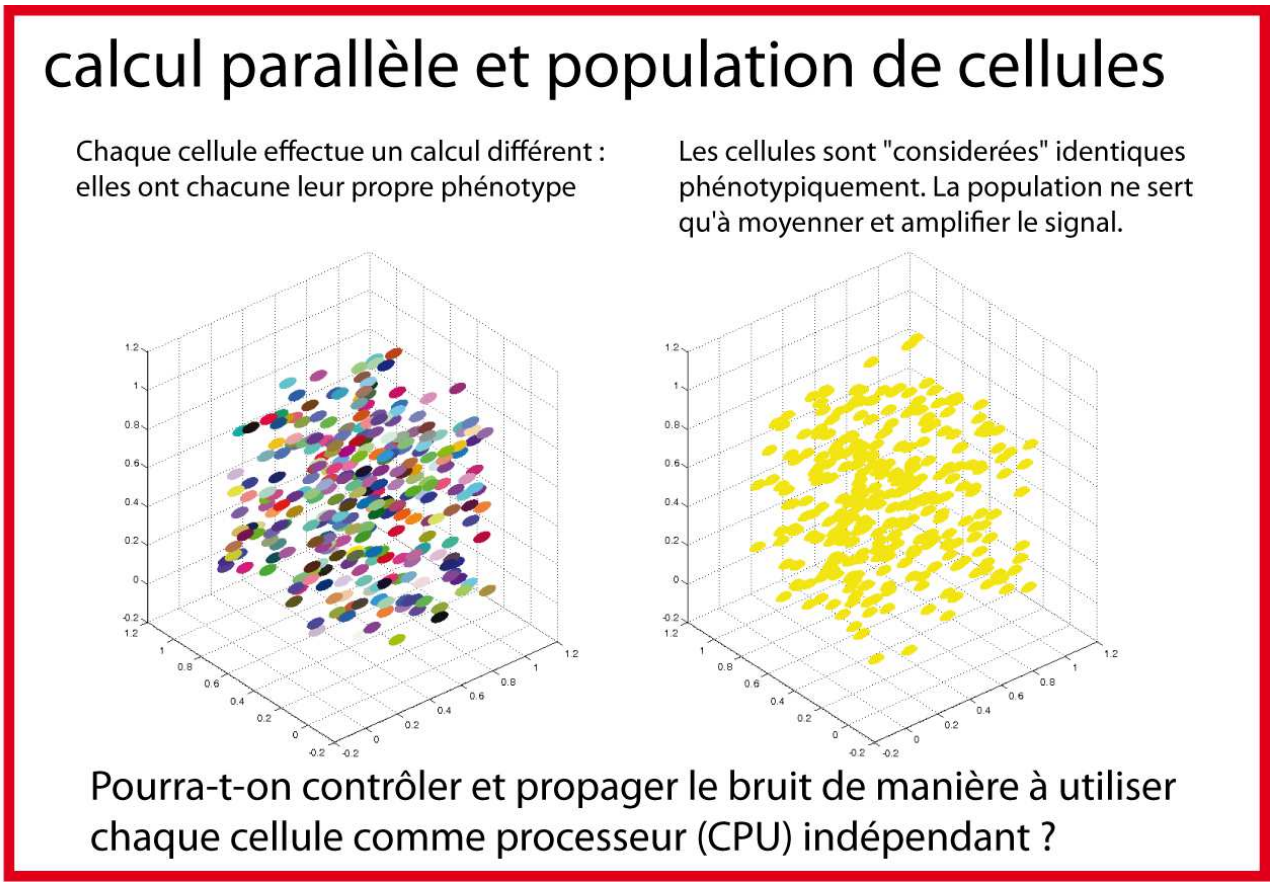



**Figure VII.10** – La diversité phénotypique peut-elle servir de moyen pour opérer des calculs de manière massivement parallèle ?

### 5.2.3 Utilisation de la dimension spatiale

Pour l'instant, il me semble que les systèmes synthétiques sont incapables de contrôler et d'utiliser la position (coordonnée XYZ) d'une entité (par exemple une protéine). Ces systèmes se basent sur l'hypothèse « well mixed » de répartition homogène des protéines et autres molécules dans la cellule. Cependant, cette hypothèse devient faible lorsqu'une protéine est en un seul exemplaire dans la cellule : connaitre sa localisation devient une information de poids. Et Il existe des méthodes de microscopie « at the single molecule level » qui permettent de localiser assez précisément des molécules fluorescentes uniques (Elf *et al.*, 2007; Li & Elf, 2009). En contrôlant la localisation, on pourrait faire beaucoup de chose :

⇒ imaginez un facteur de transcription qui serait fixé aux membranes dans une condition donnée. Puis lorsqu'un événement bien précis arriverait, il se détacherait et viendrait se fixer à l'ADN pour activer l'expression de sa cible.

⇒ imaginez que vous arriviez à maintenir deux sous-unités d'une protéine chacune dans un

pôle opposé de la cellule jusqu' à ce qu'un événement choisi les fasse lâcher leurs pôles puis se rencontrer en engendrant un phénotype particulier.

On pourrait jouer sur la probabilité de rencontre entre deux molécules en faisant varier la distance spatiale qui les sépare ou en jouant sur l'affinité qu'a une protéine pour une zone/entité particulière. Les modèles d'équations différentielles ordinaires ne suffisent plus, il faut passer à un formalisme utilisant des équations aux dérivées partielles plus adaptées aux modèles spatiaux.

Notez que le contrôle de la dimension spatiale intracellulaire permettrait[10] le contrôle spatial extracellulaire. On pourrait ainsi programmer *E. coli* comme un petit robot se déplaçant aux endroits voulus en fonction des stimuli externes.

### 5.2.4 Réseau de calcul basé sur les protéines

Nous avons beaucoup discuté de réseaux synthétiques basés sur l'expression des gènes. Ceux-ci ont l'énorme avantage de nous faire bénéficier d'outputs (les sorties) très appréciables : les gènes rapporteurs (GFP, *lacZ*, luciferase...). Mais le gros défaut des réseaux de gènes, c'est l'échelle de temps longue (minutes, heures) sur laquelle ils opèrent ce qui rend les calculs peu « réactifs ». La biologie synthétique devrait regarder en direction de la transduction du signal qui est parfois oubliée ; en effet, l'allostérie (les molécules qui s'emboitent) et les modifications chimiques (phosphorylation, acétylation etc...) se produisent à une échelle de temps beaucoup plus rapide (secondes). Par exemple, construire et contrôler un réseau non linéaire de phosphorylation pourrait permettre la création d'ordinateurs cellulaires très rapides. Chez *E. coli*, il faut peut être regarder du coté des systèmes à deux composantes. Le problème c'est qu'il nous manque des outputs permettant de contrôler l'état du système *in vivo* et en temps réel comme c'est le cas avec les gènes rapporteurs. Ce qu'il faudrait, c'est inventer des protéines qui ne « fluorescencent » ou « luminescent » que lorsqu'elles sont phosphorylées par exemple. Puis on créerait des contrôleurs à activité kinase et phosphorylase spécifiques de ces protéines « rapporteurs ».

### 5.2.5 Des réseaux qui apprennent

Si vous êtes biologiste, vous avez peut être entendu parler de la synapse de « Hebb ». Ce chercheur a fournit une contribution très importante à la neurobiologie : l'idée que deux neurones en activité au même moment créent ou renforcent leur connexion de sorte que l'activation de l'un

[10] Mais il existe d'autres méthodes possibles pour contrôler le déplacement d'une bactérie (voir le paragraphe sur le chimiotactisme dans le chapitre sur *E. coli*).

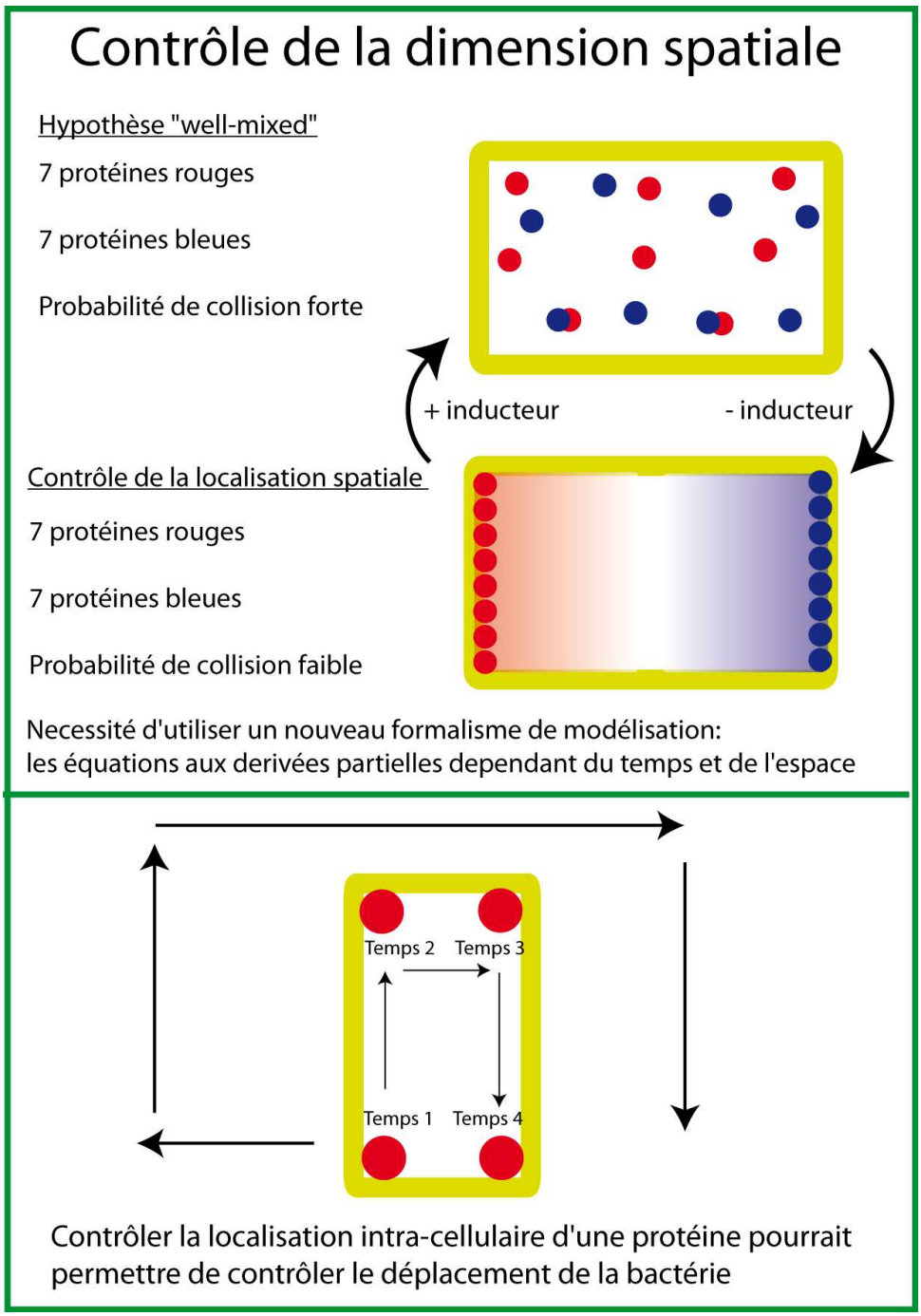



**Figure VII.11** – Apprendre à utiliser la dimension spatiale. En haut : dans les deux cas, la concentration cellulaire des protéines rouges et bleues est la même, pourtant leur probabilité de collision n'est pas la même. L'ajout ou le retrait d'un inducteur pourrait permettre d'exercer un contrôle sur cette localisation. En bas : si on arrivait à fixer une protéine dans un angle donné par exemple, celle-ci pourrait activer un système propulseur (type flagelle) lui permettant de nager dans une direction directement fonction de l'angle où est située la protéine.

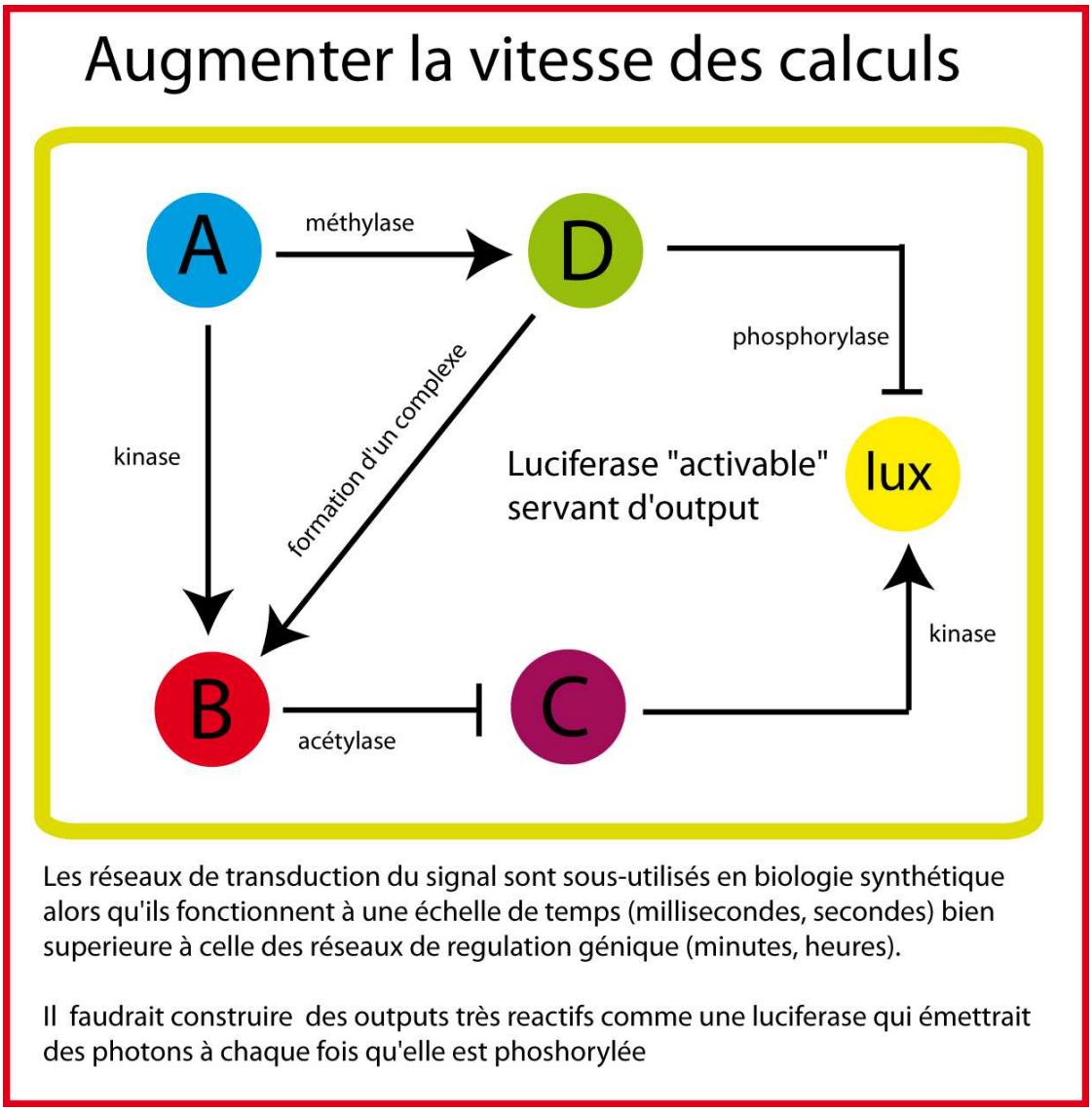



**Figure VII.12** – Construire des réseaux basés sur la transduction du signal, plus rapide.

par l'autre soit plus facile à l'avenir. C'est ce mécanisme de renforcement à l'échelle du système nerveux qui pourrait expliquer, dans une certaine mesure, les mécanismes de l'apprentissage.

Si vous êtes informaticien ou mathématicien, vous préférez sans doute que je vous parle des réseaux de neurones artificiels qui permettent aux programmes de reconnaissance vocale de s'améliorer au cours du temps par exemple. En effet, un perceptron multicouche apprend à reconnaitre les mots que vous prononcez au fur et à mesure que vous lui parlez. Le programme s'améliore.

Peut-on encoder dans l'ADN un programme/un réseau capable d'apprentissage ? Oui. En

suivant les conseils dispensés dans la publication de Lu (Lu *et al.*, 2009). Le cahier des charges est simple : le réseau doit être capable de répondre différemment à un même stimulus en fonction de son apprentissage. Voici le design basic d'un réseau permettant l'apprentissage associatif. La description est un petit peu difficile : aidez vous de la figure VII.13. Deux facteurs de transcription A et B sont exprimés en réponse à deux stimuli bien distincts. Supposons que chaque facteur de transcription déclenche l'expression de deux protéines effectrice A et B elle-même contrôlant chacune une cascade de réactions A et B. Lorsque les deux stimuli se produisent au même moment : les deux facteurs de transcription sont transcris et ils activent un interrupteur (un toggle switch). Ceci crée la mémoire associative. Plus tard, si un seul des facteurs de transcription (par exemple B) est produit, la logique de la porte AND entre le toggle switch et le facteur de transcription B entraine la production de la protéine effectrice qui contrôle la voie de *l'autre* activateur (dans notre exemple A). Ainsi une fois que le réseau a associé les stimuli A et B ensemble (interrupteur en position « on »), l'arrivée, plus tard, du stimulus A (respectivement B) suffit à déclencher la voie B (respectivement A) alors que ce n'était pas le cas avant l'association.

### 5.2.6 Utiliser d'autres fondements pour le calcul

Les capacités computationnelles des réseaux synthétiques actuels proviennent principalement des réseaux géniques intracellulaires. Nous avons vu qu'il faut élargir ces capacités à *la transduction du signal* pour gagner en vitesse et *aux copies de bactéries clonales* pour gagner en parallélisme. Mais on peut encore envisager d'autres formes de calcul :

⇒ *Avec la lumière* : il existe des enzymes capables d'émettre des photons (la luciferase) et des photorécepteurs (les phytochromes) capables de détecter des photons. Si on réussissait à coupler ces deux systèmes d'émission et de réception (et c'est loin d'être gagné à cause de problèmes quantitatifs), on pourrait peut être réussir à transmettre une information en intra ou intercellulaire via les ondes électromagnétiques (les photons). Autrement dit à faire communiquer deux bactéries à la vitesse de la lumière ⌣̈ .

⇒ *Avec l'électricité* : peut-on transformer *E. coli* en neurone ? Autrement dit peut-on imaginer utiliser des canaux potassiques, calciques, sodiques pour contrôler l'état de polarisation de la membrane d'*E. coli* ? puis utiliser cet état de polarisation pour transmettre de l'information entre cellules ? Cette idée de réseau d'*E. coli* comme réseau de neurones germe plus facilement lorsque l'on observe des bactéries *E. coli* très longues (>100 µM) (dû sans doute à une croissance très lente à 8–10°C)

⇒ *En codant l'information en modulation de fréquence* : l'information dans un réseau gé-

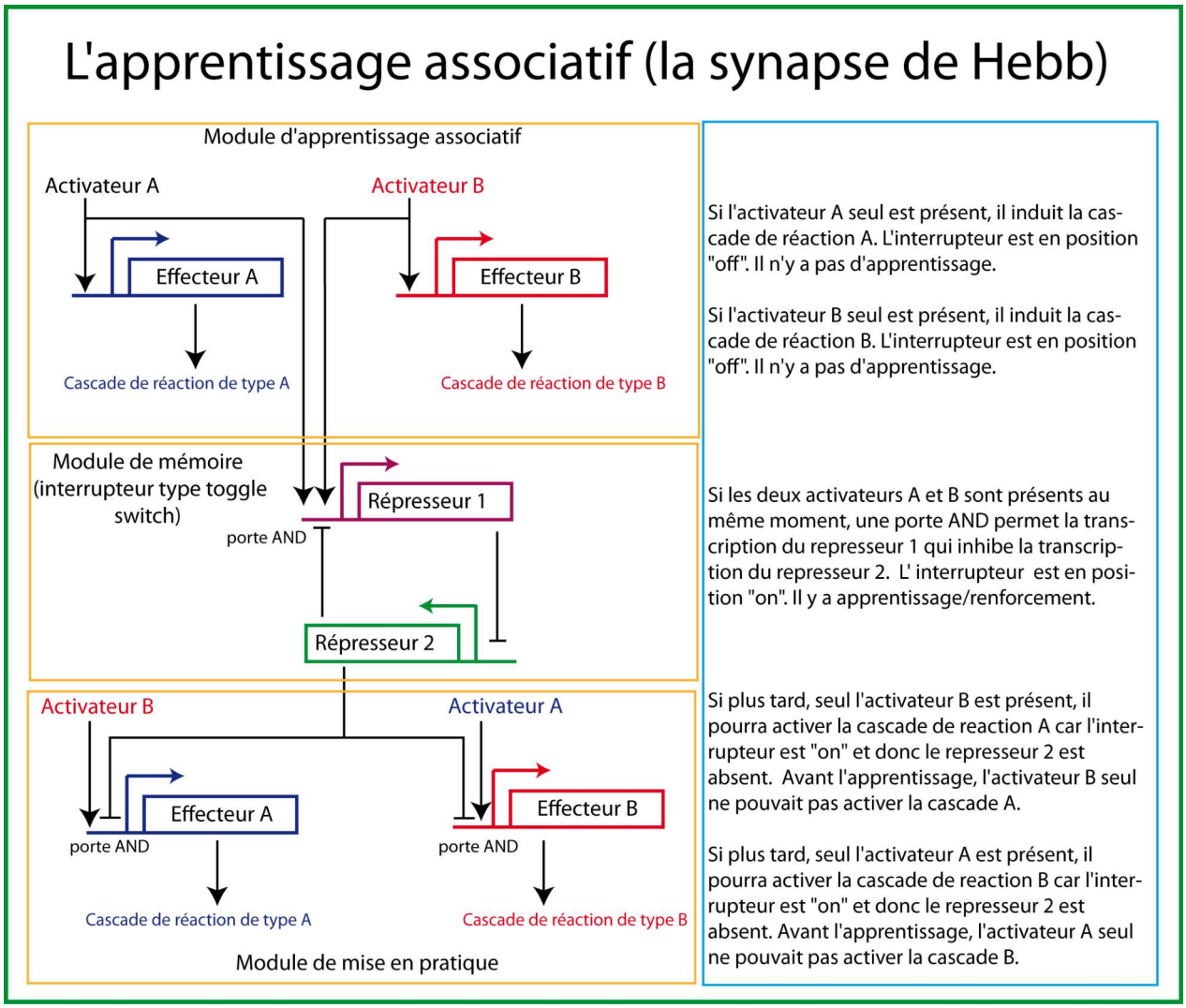



**Figure VII.13** – Schéma d'un réseau capable d'apprentissage associatif (basé sur la présence d'un interrupteur).

nique intracellulaire est codée en amplitude. Peut-on faire comme avec les potentiels d'action des neurones qui encodent l'information en modulation de fréquence ? On sait déjà faire des oscillateurs qui génèrent des « spikes ». Selon Timothy Lu, cette information de fréquence pourrait être utile pour propager de l'information sur de longues distances car la fréquence est moins susceptible de diminuer au fur et à mesure que la distance s'accroit comparé au nombre absolu de molécules qui forment inexorablement un gradient de concentration.

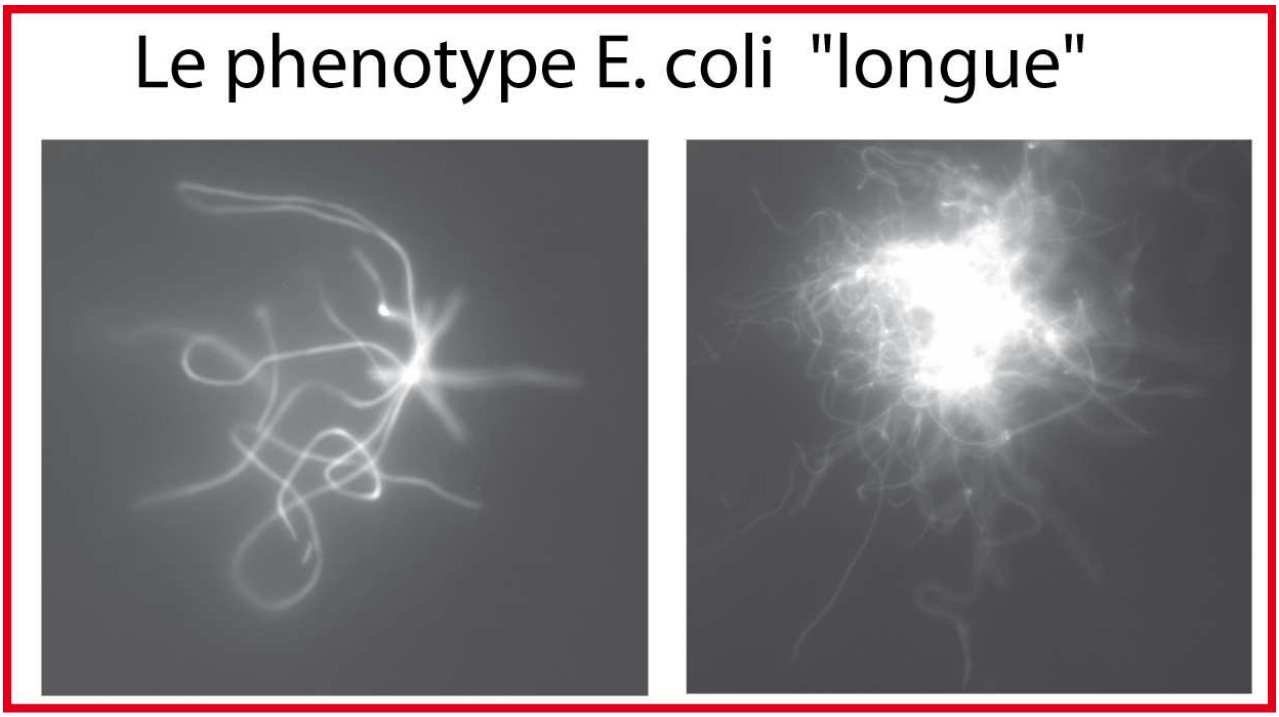



**Figure VII.14** – Des pelottes de bactéries *E. coli* ayant poussé lentement (chaque bactérie peut atteindre une centaine de micromètres).

## 5.3 Les applications à court et moyen terme

Pour des raisons pratiques, je vais découper les applications relatives à la biologie synthétique en deux grandes familles :

1. les applications dans le domaine de la santé, du calcul, de l'agriculture, de l'industrie et de l'écologie.
2. les applications militaires, les applications relatives à l'augmentation de l'espérance de vie et enfin la question du surhomme.

Le premier point est l'objet de la fin de ce chapitre. Le deuxième point est l'objet des deux chapitres qui suivent.

Dans le premier point, il n'y a pas, il me semble, de questions éthiques, existentielles ou relatives à l'essence de l'homme. Les bactéries et les « protocellules » n'ont pas de systèmes nerveux. Par conséquent, leur utilisation n'engendre aucune souffrance. Une bactérie avec un système synthétique devrait être considérée comme un ~~vulgaire~~ assemblement d'atomes ayant la capacité de traiter de l'information. J'ai finalement barré le vulgaire car il me semble que l'on doit tout de même le respect à nos ancêtres ⌣̈. Ces applications nécessitent cependant quand même l'application du principe de précaution de manière raisonnable. Notez également qu'il

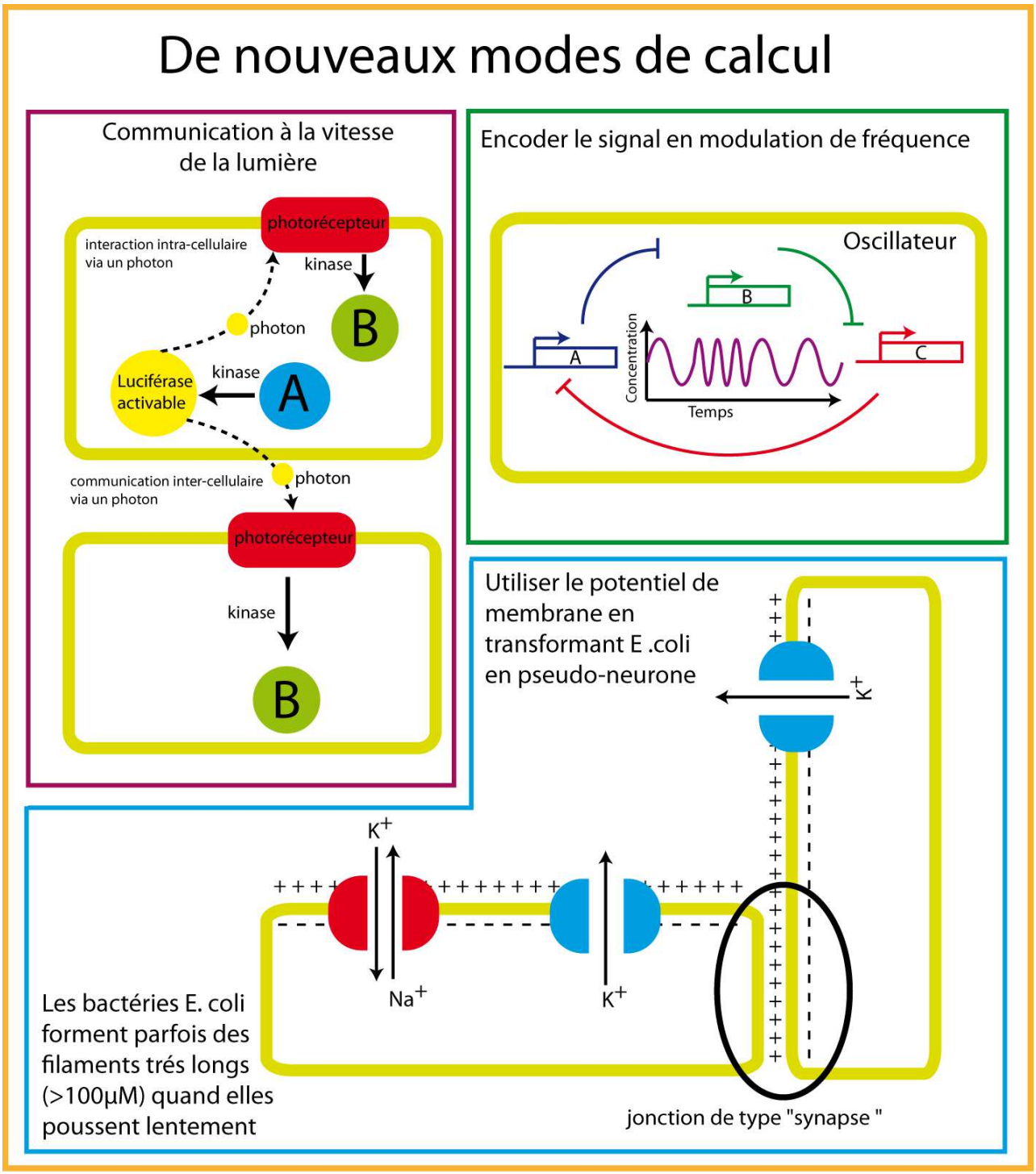



**Figure VII.15** – Illustration de l'utilisation de nouveaux modes de calcul.

pourrait être néfaste, sur le plan économique, que ce premier point subisse les conséquences d'un débat sociétal dont l'enjeu serait le deuxième point. C'est pourquoi je m'attelle à opérer la démarcation. Cette dernière est également temporelle : le premier point est une question d'années ou de décennies alors que le deuxième point est une question de demi-siècles ou de siècles. Enfin, il me semble fort probable que la plupart des chercheurs défendraient âprement le premier point avec pour argument principal la liberté de mener des recherches pacifistes susceptibles d'augmenter le niveau de confort[11] de l'homme. A l'inverse, la plupart des chercheurs refuseraient de créer des armes biologiques et entrevoient la question du surhomme et de l'espérance de vie avec le même regard, les mêmes doutes et les mêmes craintes que le peuple.

Voici les deux principales applications peu polémiques que les biologistes synthétiques entrevoient pour les décennies à venir :

### 5.3.1 L'usine microbienne

L'industrie chimique fabrique toutes sortes de composés chimiques dont certains sont très chers à synthétiser. L'idée consiste à construire une souche bactérienne qui se charge de la fabrication du dit composé. Le temps de recherche et développement pour construire la souche peut être important car les techniques utilisées (ingénierie du métabolisme, biologie moléculaire et biologie synthétique) sont difficiles. Mais une fois la souche obtenue, on ne risque pas de la perdre : elle est acquise. Il ne reste plus qu'à la placer dans un grand fermenteur pouvant dépasser 1000 mètres cubes avec un milieu de culture contenant divers minéraux ainsi que des sources de carbone et d'azote bon marché. On laisse ensuite les bactéries opérer la transformation (j'imagine que cela est relativement rapide : jours ?) puis on récupère le composé d'intérêt. Les rendements peuvent être conséquents : un % non négligeable du poids de la source de carbone ajoutée au préalable sera converti en molécule d'intérêt. Si la différence de prix au kilo entre les deux molécules est très importante, le bénéfice le devient aussi. Le composé en question peut être un médicament : par exemple l'artemisinin un médicament anti-malaria (Ro *et al.*, 2006). Mais ce peut être aussi un colorant (Yan *et al.*, 2005), un simple acide aminé (Park *et al.*, 2007), un bio-fuel (Atsumi & Liao, 2008) ou des nano-materiaux (la soie des araignées) (Widmaier *et al.*, 2009). L'ingénierie du métabolisme permet aussi d'envisager de fabriquer des souches chargées de décontaminer un milieu pollué en transformant le composé polluant en un composé non polluant. Pour éviter que ces bactéries, placées dans l'écosystème polluée, ne se

[11] Ce mot « confort » peut choquer quand il s'agit de santé. Cependant, je l'utilise ici car la notion de santé est trop proche/difficilement dissociable des notions plus polémiques d'allongement de l'espérance de vie et donc de la disparition de la mort. Confort intègre donc ici une idée sur laquelle nous sommes, il me semble, tous d'accord : nous ne souhaitons pas « souffrir » physiquement avant de mourir.

propagent au reste de l'environnent de manière non souhaitée, les biologistes développent des technologies de « bio-confinement » pour garder le contrôle de la dissémination de la bactérie. Voir le chapitre sur l'évolution.

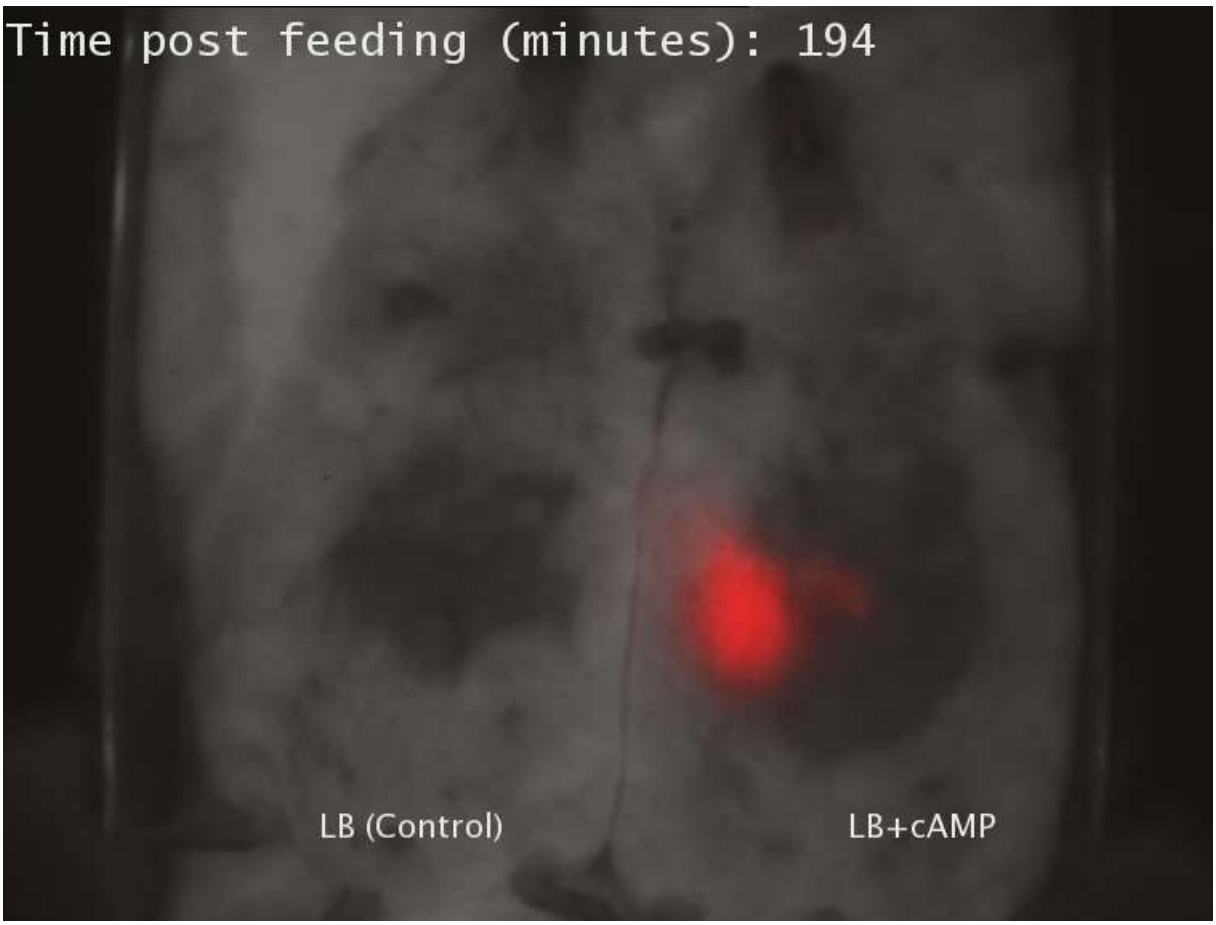



**Figure VII.16** – Un exemple dressant un pont entre biologie synthétique et médecine. Induire l'expression d'un gène d'une bactérie présente dans l'intestin d'une souris (voir publication dans cette thèse). On inocule par voie orale à deux souris une solution de bactéries incapables de produire de l'AMPc. Ces bactéries sont transformées avec un plasmide exprimant la luciferase sous contrôle d'un promoteur Crp–cAMP dépendant. Puis quelques temps plus tard, on inocule, toujours par voie orale, de l'AMPc à une souris (à droite) et de l'eau à l'autre souris (à gauche). Au bout d'un moment, les bactéries de la souris de droite expriment la luciferase grâce à la présence de l'inducteur : l'AMPc. On détecte la luminescence provenant du tractus gastro-intestinal de manière transcutanée avec une camera spéciale. Concrètement, on a induit et « monitoré » l'expression d'un gène (la luciferase) de bactéries situées dans le système gastro-intestinal de la souris. Notez que dans cette video, on filme la luminescence (les bactéries) au cours du temps et non les souris (photo superposée) ce qui explique certains décalages lorsque les souris se réveillent entre les anesthésies.

### 5.3.2 La santé

Le développement des systèmes synthétiques de première génération permet d'imaginer développer des thérapeutiques différentes du simple médicament type « molécule chimique active ». On peut imaginer fabriquer des souches « bio-senseurs » capables de détecter, par exemple, un problème dans l'intestin (hausse de température, présence d'une bactérie pathogène, d'une toxine, d'un virus) puis d'envoyer un signal (output) détectable (par exemple dans les selles : luminescence, pigments, odeurs) pouvant alerter le patient ou le médecin. Pour cela, il faut déjà accroitre notre connaissance dans un domaine de recherche qui me semble relativement peu exploité : l'écosystème gastro-intestinal. Bien sûr, en plus d'avoir un rôle de senseur passif, les souches bactériennes synthétiques à visée médicale pourraient avoir un rôle plus actif : « attaquer » une souche pathogène (en lui volant tout simplement ses ressources par exemple), dégrader une toxine ou synthétiser des molécules antivirales. Les oscillateurs synthétiques pourraient être utilisés pour permettre la synthèse d'un médicament à intervalle de temps régulier. Dans le cas de pathologies comme le diabète, des boucles de rétroaction pourraient contrôler la production d'insuline par des bactéries en fonction de la concentration extérieure en glucose (mesurée à l'aide des mécanismes de la répression catabolique par exemple) (j'avoue que dans cet exemple, je ne sais pas trop dans quelle partie du corps il faudrait placer la souche).

Enfin, il faut savoir que, après une injection en intraveineuse (ce qui ne pose pas de problème avec une souche non pathogène), certaines souches bactériennes comme *E. coli* disposent de la faculté de venir se loger automatiquement dans les tumeurs sans que l'on sache exactement pourquoi (Jiang *et al.*, 2010). Ainsi en développant un système synthétique cytotoxique qui détruirait spécifiquement les cellules tumorales, on pourrait imaginer un traitement innovant contre le cancer. Dans sa publication, Jiang injecte des bactéries *E. coli* en intraveineuse dans des souris. Les bactéries disposent d'un plasmide exprimant la luciferase et un autre permettant l'expression du gène clyA codant pour la cytolysin A, une drogue anti-tumorale. D'une part, les chercheurs identifient les métastases de la souris de manière transcutanée grâce à la luciferase des bactéries venues se loger dans la tumeur. D'autre part, les chercheurs observent une résorption significative des métastases grâce à l'action de la cytolysin A[12]. Certes il s'agit de résultats très fondamentaux ayant peu de chance de déboucher sur des essais cliniques mais ils ont le mérite de pointer du doigt, au biologiste synthétique, une direction potentiellement prometteuse.

[12]Le traitement est effectué en combinaison avec de la radiothérapie ce qui rend la publication moins « sexy » mais ne discrédite en rien les résultats.

# Références bibliographiques


Atsumi, S., & Liao, J. C. 2008. Metabolic engineering for advanced biofuels production from Escherichia coli. *Curr Opin Biotechnol*, **19**(5), 414–9. 183

Balagadde, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., Quake, S. R., & You, L. 2008. A synthetic Escherichia coli predator-prey ecosystem. *Mol Syst Biol*, **4**, 187. 169

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature*, **434**(7037), 1130–4. 169

Chang, D. E., Leung, S., Atkinson, M. R., Reifler, A., Forger, D., & Ninfa, A. J. 2009. Building biological memory by linking positive feedback loops. *Proc Natl Acad Sci U S A*. 170

Danino, T., Mondragon-Palomino, O., Tsimring, L., & Hasty, J. 2010. A synchronized quorum of genetic clocks. *Nature*, **463**(7279), 326–30. 169

Datsenko, K. A., & Wanner, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**(12), 6640–5. 162

Del Vecchio, D., Ninfa, A. J., & Sontag, E. D. 2008. Modular cell biology : retroactivity and insulation. *Mol Syst Biol*, **4**, 161. 173

Elf, J., Li, G. W., & Xie, X. S. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**(5828), 1191–4. 175

Elowitz, M. B., & Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**(6767), 335–8. 169

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., & Collins, J. J. 2009. Synthetic gene networks that count. *Science*, **324**(5931), 1199–202. 170

Gardner, T. S., Cantor, C. R., & Collins, J. J. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**(6767), 339–42. 168

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., 3rd, Smith, H. O., & Venter, J. C. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**(5987), 52–6. 158

Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchison, C. A., 3rd, Smith, H. O., & Venter, J. C. 2006. Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A*, **103**(2), 425–30. 157

Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamakawa, T., Yamazaki, Y., Mori, H., Katayama, T., & Kato, J. 2005. Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. *Mol Microbiol*, **55**(1), 137–49. 157

Jiang, S. N., Phan, T. X., Nam, T. K., Nguyen, V. H., Kim, H. S., Bom, H. S., Choy, H. E., Hong, Y., & Min, J. J. 2010. Inhibition of Tumor Growth and Metastasis by a Combination of Escherichia coli-mediated Cytolytic Therapy and Radiotherapy. *Mol Ther*. 185

Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., & Voigt, C. A. 2005. Synthetic biology : engineering Escherichia coli to see light. *Nature*, **438**(7067), 441–2. 170

Li, G. W., & Elf, J. 2009. Single molecule approaches to transcription factor kinetics in living cells. *FEBS Lett*, **583**(24), 3979–83. 175

Lu, T. K., Khalil, A. S., & Collins, J. J. 2009. Next-generation synthetic gene networks. *Nat Biotechnol*, **27**(12), 1139–50. 173, 179

Noireaux, V., & Libchaber, A. 2004. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc Natl Acad Sci U S A*, **101**(51), 17669–74. 157

Park, J. H., Lee, K. H., Kim, T. Y., & Lee, S. Y. 2007. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A*, **104**(19), 7797–802. 183

Posfai, G., Plunkett, G., 3rd, Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., de Arruda, M., Burland, V., Harcum, S. W., & Blattner, F. R. 2006. Emergent properties of reduced-genome Escherichia coli. *Science*, **312**(5776), 1044–6. 157

Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., & Keasling, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**(7086), 940–3. 183

Sauro, H. M. 2008. Modularity defined. *Mol Syst Biol*, **4**, 166. 173

Sohka, T., Heins, R. A., Phelan, R. M., Greisler, J. M., Townsend, C. A., & Ostermeier, M. 2009. An externally tunable bacterial band-pass filter. *Proc Natl Acad Sci U S A*, **106**(25), 10135–40. 170, 173

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., & Hasty, J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature*, **456**(7221), 516–9. 169

Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A., & Ellington, A. D. 2009. A synthetic genetic edge detection program. *Cell*, **137**(7), 1272–81. 170

Widmaier, D. M., Tullman-Ercek, D., Mirsky, E. A., Hill, R., Govindarajan, S., Minshull, J., & Voigt, C. A. 2009. Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol Syst Biol*, **5**, 309. 183

Yan, Y., Chemler, J., Huang, L., Martens, S., & Koffas, M. A. 2005. Metabolic engineering of anthocyanin biosynthesis in Escherichia coli. *Appl Environ Microbiol*, **71**(7), 3617–23. 183

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, **97**(11), 5978–83. 162

# Chapitre VIII

# Physique quantique et biologie

## 1 Préambule

Avant de commencer ce chapitre, je souhaite convaincre le lecteur biologiste de ne pas succomber à son déterminisme propre ⌣̈. En effet, ce dernier lui suggère très certainement de passer son chemin à la simple évocation du mot « quantique » ! Biologistes, résistez ! utilisez votre libre arbitre ! et lisez *les petites expériences de pensée* que je vous propose. C'est promis, il n'y a aucune équation (ce qui m'arrange bien. . .⌣̈).

## 2 Physique quantique et biologie

Le paradigme actuel de la biologie considère que l'échelle adéquate explicative des phénomènes en biologie ne descend pas en dessous de la molécule. Les règles qui régissent le comportement de ces molécules, leurs interactions par exemple et les phénotypes résultants relèvent de la physique classique. Evidemment, le biologiste ne nie pas que les molécules soient constituées d'atomes. Il lui apparait trivial également qu'à l'échelle atomique et subatomique les lois quantiques aient un pouvoir explicatif supérieur à celui des lois de la physique classique. Ce que le biologiste ou plutôt le paradigme actuel de la biologie nie, c'est que la physique quantique puisse avoir un pouvoir explicatif autre que trivial pour expliquer un phénotype donné. Ce scepticisme est en partie justifié par le fait que très peu d'expériences en biologie ont pu dresser un pont entre biologie moléculaire/structurale et mécanique quantique. Pour des bonnes revues sur ce sujet, on peut lire (Davies, 2004; Arndt *et al.*, 2009).

Dans l'expérience des fentes d' Young, les physiciens utilisent une source capable d'émettre des photons un par un. Ceux-ci sont ensuite détectés (également un par un) sur un écran placé derrière les fameuses fentes d' Young : les impacts forment petit à petit une figure d'interférences illustrant la dualité onde–corpuscule de la matière. Imaginez maintenant que la source n'émette

plus des photons mais des virus. Pour un biologiste, un virus est une entité biologique composé de plusieurs dizaines de millions d'atomes et mesurant quelques dizaines de nanomètres. L'image que le biologiste en a provient des clichés de microscopie électronique mais les virus sont aussi utilisés en routine, sans pouvoir les voir, pour modifier spécifiquement le génome d'un organisme. Imaginez donc maintenant que les virus, que vous projetez un par un sur les fentes d'Young, génèrent des interférences sur l'écran placé derrière (Figure VIII.1). Selon les lois classiques concernant les trajectoires de ces corpuscules, il est impossible d'interpréter ce phénomène. La seule explication possible passe alors par la mécanique quantique. Lors du franchissement des fentes, le virus passe dans un état superposé et se comporte comme une onde. Il passe en réalité par les deux fentes en même temps avant « d'interférer avec lui-même ». Et si vous cherchez à mesurer par quelle fente le virus est passé, cela ferait immédiatement disparaitre les interférences. Sachez que ce type d'expériences a été réussi avec des molécules (dont certaines biologiques) de plus de 60 atomes (Hackermuller *et al.*, 2003). Le physicien Anton Zeilinger, reconnu mondialement pour ces travaux en physique quantique, pense que l'on réussira cette expérience avec des virus d'ici quelques années/décennies (Zeilinger, 2010). Si tel est le cas, cela signifie que nous biologistes, devrons modifier notre regard sur ce qu'est effectivement un virus. Se borner à ignorer totalement les résultats de la mécanique quantique, à vouloir rester dans son paradigme confortable, serait une erreur classique de scientifique très bien expliqué par Thomas Kuhn dans son ouvrage *La structure des révolutions scientifiques* (sur lequel je reviendrai dans le chapitre sur l'épistémologie).

Cependant, vous allez me dire que cette expérience de pensée avec des virus est faite dans des conditions très éloignées de l'environnement cellulaire. Cela ne justifie donc en rien la nécessité de changer de paradigme pour décrire les phénomènes se produisant dans le cytoplasme. Regardons ensemble si la mécanique quantique a potentiellement un pouvoir explicatif dans un environnement cellulaire normal. Une molécule en état de cohérence quantique (superposition de multiples états) peut posséder plusieurs valeurs pour une certaine quantité observable. Par exemple, plusieurs positions. Ces états superposés sont absolument inconnus à un niveau macroscopique. Le phénomène physique susceptible d'expliquer la transition entre les règles de la physique quantique et les règles de la physique classique s'appelle la décohérence quantique. Cette dernière a pour conséquence la disparition des états quantiques superposés lorsqu'une interaction avec l'environnement apparait (par exemple par l'effet d'un appareil de mesure ou d'une simple interaction moléculaire). La décohérence « rétablit » en quelque sorte une position unique à la molécule. La vitesse à laquelle se produit la décohérence dépend de la nature et de la température de l'environnement (Zurek, 1982; Unruh & Zurek, 1989). Ainsi l'observation de phénomènes quantiques en physique requiert souvent de très faibles températures et un très

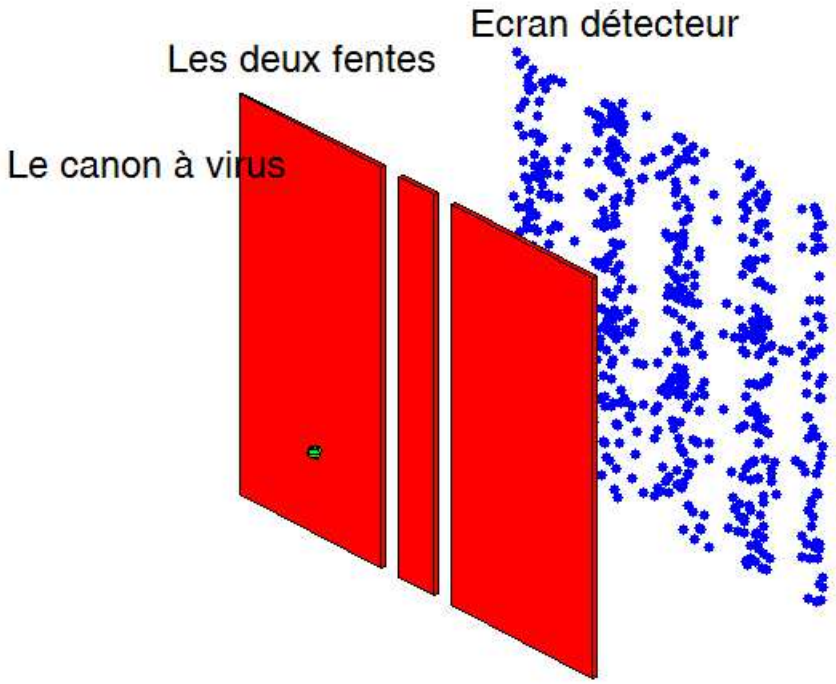



**Figure VIII.1** – Vidéo illustrant la potentielle dualité onde–corpuscule d'un virus projeté sur les fentes d'Young et générant des interférences. La vidéo est générée sous Matlab : lors du passage du virus au niveau des fentes, le virus se « transforme » en grosse boule verte représentant une onde passant par les deux fentes. Puis il frappe l'écran détecteur placé derrière les fentes et génère un point bleu. Si on renouvelle l'expérience un certain nombre de fois, on observe une figure d'interférence sur l'écran. Seule la mécanique quantique est en mesure d'expliquer ce résultat. Voir le code Matlab en annexe Code14VirusOndeParticule.m qui génère cette vidéo.

haut degré d'isolement du système c'est-à-dire un environnement bien différent du cytoplasme chaud et aqueux d'une cellule. C'est pour cette raison que la biologie moléculaire et la mécanique quantique ont tant de mal à s'influencer/se fertiliser mutuellement. Pour un nucléotide dans une bactérie, le temps de décohérence est de l'ordre de la femto seconde donc bien inferieur à l'échelle de temps d'une réaction chimique. Or pour imaginer que les effets quantiques puissent jouer un rôle non trivial en biologie, il faut absolument que le taux de décohérence se fasse à une échelle biologique. La démonstration de la présence de sous espaces/ de zones dans la matière organique protégées des effets de la décohérence pendant une échelle de temps biologiquement suffisante (milliseconde) reste donc une étape importante à franchir.

Cependant, depuis le début du 21$^{\text{ème}}$ siècle, plusieurs conjectures et expériences ont com-

mencé à construire des ponts très sérieux entre biologie et physique quantique. Citons la conjecture controversée du célèbre physicien anglais Roger Penrose (Hameroff & Penrose, 1996) qui postule que la conscience émergerait de la cohérence quantique présente au niveau des microtubules. La présence de cohérence quantique a aussi été proposée pour expliquer la sensibilité extraordinaire de l'odorat (Brookes *et al.*, 2007) ou le sens de l'orientation des oiseaux (Gauger *et al.*, 2011).

En 2007, l'équipe de Graham Fleming a mis en évidence un état de cohérence quantique dans un système photosynthétique (Engel *et al.*, 2007). Les chercheurs ont utilisé le complexe bacterio-chlorophyllien FMO (pour Fenna–Matthews–Olson) qui participe à la conversion de l'énergie lumineuse (photon du soleil) en énergie chimique (ATP) chez les bactéries sulfureuses vertes. Lorsqu'un photon frappe ce complexe, il lui transmet son énergie sous la forme d'une excitation électronique. Les chercheurs ont pu montrer que l'excitation électronique, plutôt que d'emprunter un chemin parmi d'autres pour traverser la protéine au risque de ne pas trouver le plus court, se comporte comme une onde c'est-à-dire comme si elle se propageait à travers tous les chemins possibles en même temps (Grousson, 2011). Certes, l'expérience a été mené à une température peu propice à la vie (-196°C) mais une expérience équivalente a mis en évidence en 2010 des résultats similaires à température ambiante (Collini *et al.*, 2011).

## 3 De l'information quantique dans la cellule ?

Selon le paradigme de Jacob et Monod, la complexité des organismes vivants vient de la présence d'un réseau d'interaction qui suit les lois de la physique classique. Ce réseau apporte des capacités « computationnelles » à la cellule. Cette dernière stocke et manipule les bits de la théorie de l'information et donc, fonctionne un peu à la manière des algorithmes et des ordinateurs d'aujourd'hui. Les modèles que je décris dans cette thèse suivent cette hypothèse car ils accordent (de manière plus ou moins cachée) une position unique aux molécules. Chaque molécule possède des coordonnées spatiales XYZ uniques et se déplace aléatoirement selon le mouvement brownien.

Maintenant, faisons l'hypothèse que la physique quantique joue un rôle non trivial en biologie. Les molécules peuvent donc se retrouver dans une superposition de position. En biologie des systèmes/ synthétique, ces propriétés quantiques auraient des avantages et des inconvénients. Inconvénients liés à l'accroissement des incertitudes car la position résulte d'une densité de probabilité de présence (une fonction d'onde). Les modèles doivent la calculer à l'aide de l'équation de Schrödinger. Avantages liés au fait que l'algorithme (par exemple pour rechercher

une cible comme un promoteur) devient beaucoup plus efficace car il fonctionne sur un régime parallèle. Si la cellule utilise ces propriétés de superposition pour « computer » un phénotype, alors elle ne manipule plus des bits classiques (0 ou 1) mais des bits quantiques (qubit).

Ces deux disciplines que sont la physique quantique et la biologie pourraient bien se fertiliser mutuellement. En effet, si la cellule sait utiliser des qubit ou augmenter/contrôler les durées de la cohérence, cela pourrait avoir un intérêt majeur pour les physiciens. Ces derniers pourraient essayer de comprendre le vivant (en s'appuyant, par exemple, sur l'évolution) dans le but de comprendre la physique sous-jacente. Coté applications, il leur serait alors aisé de détourner ou de copier les stratégies du vivant à leurs propres fins pour créer, par exemple, le fameux ordinateur quantique.

# 4 Un nouveau combat entre paradigmes ?

Reste à savoir si la vie fonctionne réellement sur le mode quantique. On voit bien que le paradigme principal en biologie va devoir combattre ce nouvel embryon[1] de paradigme concurrent qui nous vient des physiciens.

Les arguments des défenseurs du paradigme principal (surtout des biologistes) sont :

⇒ *Absence de preuve.*

Il faudra déjà prouver (en réalité persuader, selon Thomas Kuhn[2]) le rôle non trivial de la physique quantique en biologie. En effet, il est encore difficile d'imaginer que des processus quantiques perdurent plus de quelques nanosecondes sur plus de quelques nanomètres dans l'environnement chaud et non isolé de la cellule.

⇒ *Scepticisme épistémologique.*

Michel Morange[3] y voit *« une position [qui] relèverait d'un réductionnisme un peu radical parfois récurrent chez certains physiciens qui veulent tout ramener à leur discipline. Or il est intéressant de noter que dès les années 1930, Erwin Schrödinger, l'un des pères fondateurs de la mécanique quantique, s'élevait déjà contre cette tendance »* .

Les arguments des défenseurs (surtout des physiciens) du paradigme concurrent sont :

⇒ *Argument « évolutif » (normalement réservé aux biologistes !).*

Pourquoi l'algorithme évolutif, qui a fouillé l'arbre des possibles pendant des milliards d'années, n'aurait-il pas réussi à rendre la cellule capable d'utiliser le principe de superposition quantique ? Pourquoi le vivant ne chercherait t'il pas parmi l'ensembles des

[1] Car on ne peut pas encore vraiment parler de communauté scientifique défendant ce point de vue.
[2] Thomas Kuhn, *La structure des révolutions scientifiques.*
[3] Mathieu Grousson , *La vie serait quantique*, Science et vie Avril 2011, p. 72.

propriétés physiques disponibles afin d'optimiser une fonction donnée ? Une autre manière de voir les choses serait de dire que l'évolution a justement conduit les organismes (ou a été conduit) à la frontière entre régime classique et régime quantique de manière à pouvoir calculer à la fois avec des bits classiques et des bits quantiques. Dans ce cas, la vie suivrait des lois qui ne serait ni totalement quantique ni totalement classique. Rien n'empêche également de penser que le régime quantique puissent n'être qu'une fonction, un outil parmi d'autres que la cellule utiliserait à son gré de manière locale (spatialement ou temporellement).

⇒ *Scepticisme épistémologique (inversement de l'argument antiréductionniste).*
Au cours du siècle dernier, la biologie, via le réductionnisme, n'a pas cessé de diminuer son échelle spatiale explicative : organismes, puis organes, puis tissus, puis cellules, puis molécules. Or, comme par hasard, la limite explicative en biologie s'arrête exactement là où les physiciens manquent encore de certitudes. Là où ils essaient en vain, via la théorie de la décohérence, d'expliquer comment s'opère le passage du microscopique au macroscopique (paradoxe du chat de Schrödinger). N'est-ce pas là une manière étrange (et dogmatique) d'imaginer que la limite explicative réelle en biologie s'arrête exactement là où le flou dans la connaissance physique commence et ce par une pure coïncidence ?

# Références bibliographiques


Arndt, M., Juffmann, T., & Vedral, V. 2009. Quantum physics meets biology. *Hfsp J*, **3**(6), 386–400. 191

Brookes, J. C., Hartoutsiou, F., Horsfield, A. P., & Stoneham, A. M. 2007. Could humans recognize odor by phonon assisted tunneling ? *Phys Rev Lett*, **98**(3), 038101. 194

Collini, E., Wong, C. Y., Wilk, K. E., Curmi, P. M., Brumer, P., & Scholes, G. D. 2011. Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature*, **463**(7281), 644–7. 194

Davies, P. C. 2004. Does quantum mechanics play a non-trivial role in life ? *Biosystems*, **78**(1-3), 69–79. 191

Engel, G. S., Calhoun, T. R., Read, E. L., Ahn, T. K., Mancal, T., Cheng, Y. C., Blankenship, R. E., & Fleming, G. R. 2007. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, **446**(7137), 782–6. 194

Gauger, E. M., Rieper, E., Morton, J. J., Benjamin, S. C., & Vedral, V. 2011. Sustained quantum coherence and entanglement in the avian compass. *Phys Rev Lett*, **106**(4), 040503. 194

Grousson, M. 2011. La vie serait quantique. *Science et vie*. 194

Hackermuller, L., Uttenthaler, S., Hornberger, K., Reiger, E., Brezger, B., Zeilinger, A., & Arndt, M. 2003. Wave nature of biomolecules and fluorofullerenes. *Phys Rev Lett*, **91**(9), 090408. 192

Hameroff, S, & Penrose, R. 1996. Orchestrated reduction of quantum coherence in brain microtubules : A model for consciousness. *Mathematics and computers in simulation*. 194

Unruh, W. G., & Zurek, W. H. 1989. Reduction of a wave packet in quantum Brownian motion. *Phys Rev D Part Fields*, **40**(4), 1071–1094. 192

Zeilinger, Anton. 2010. Les promesses d'un nouveau monde. *Pour la science*. 192

Zurek, W. H. 1982. Environment-induced superselection rules. *Phys. Rev. D*, **26**(Oct), 1862–1880. 192

# Chapitre IX

# Evolution (I)

## 1 Préambule

Pourquoi consacrer un chapitre entier à l'évolution dans une thèse qui traite des réseaux de régulation ? Je vais répondre à cette question par une autre question. « Pourquoi le réseau de régulation de *E. coli* existe t'il ? »

Pour répondre à cette question, je peux me contenter du schème explicatif molécularo–mécaniste[1]. Le réseau existe *parce que* les protéines, ADN, métabolites interagissent ensemble selon les lois de la physique et de la chimie. Ces interactions créent un réseau non linéaire à partir duquel, on le croit[2], émerge la vie.

Mais je peux aussi répondre à cette question en utilisant le schème explicatif darwinien. Le réseau existe *parce que* l'algorithme évolutif a façonné le génome pendant plusieurs milliard d'années. L'environnement sélectionne les mutations bénéfiques c'est-à-dire celles qui confèrent aux individus une plus forte capacité à se reproduire. Or pour se reproduire, il faut déjà survivre et pour survivre, il faut pouvoir s'adapter à des changements dans l'environnement. La présence d'un réseau de régulation permet cette adaptation et confère donc un avantage sélectif aux individus qui le possèdent.

Ce chapitre sera découpé en 3 parties :

1. la première partie rappelle quelques généralités concernant l'évolution.
2. la seconde partie se focalise plus particulièrement sur le rôle de l'évolution dans l'apparition des réseaux de régulation.
3. la troisième partie étudie les liens entre évolution et biologie synthétique.

[1] Voir la fin du chapitre 1, Michel Morange , *Les secrets du vivant* , La découverte.

[2] Lorsque l'on parle d'émergence, il faut rester prudent car cette notion ne respecte pas l'impératif de rigueur scientifique tel que défini par l'épistémologie institutionnelle positiviste.

# 2 Quelques généralités sur l'évolution

Contrairement à une idée répandue, la théorie de l'évolution n'est pas simple. La difficulté ne vient pas du pattern général. Elle vient du fait qu'en matière d'évolution, « le diable est dans les détails » et j'essaierai de le demontrer avec plusieurs exemples. Notez également que j'aborde moi-même ce sujet en tant que novice (mon école de pensée d'origine étant plutôt « molécularo–mécaniste » ). Mon exposé pourrait refléter une compréhension partielle voire erronée de la théorie de l'évolution. Il doit donc être lu avec prudence et sens critique.

Dans l'évolution, j'identifie différents thèmes qu'il convient de traiter correctement :

⇒ les mutations
⇒ le transfert horizontal de gènes
⇒ la sélection naturelle
⇒ la dérive génétique
⇒ les contraintes évolutives

## 2.1 Les mutations

Nous l'avons vu, lors de la réplication de l'ADN, il arrive que l'ADN polymérase se trompe et insère le mauvais nucléotide. Il existe des mécanismes de correction d'erreurs mais il y a toujours des mutations qui y échappent. Au final, le taux de mutation moyen chez *E. coli* est de l'ordre de $10^{-9}$ mutations par paire de bases par génération. Par conséquent, sur une population de 1 milliard de bactéries, il y a statistiquement une bactérie mutante pour chacun des nucléotides. Comme le chromosome d' *E. coli* contient très approximativement 1 million de nucléotides. Sur une population de 1 milliard de bactéries, il y a un million de bactéries mutantes. Ainsi il y a en moyenne une bactérie sur mille qui possède une mutation. On voit donc que, dés lors que l'on dépasse mille bactéries dans une flasque, la population n'est plus génotypiquement homogène. Il y a une distribution des différents génotypes. La plupart des mutations sont soit neutres (sans conséquence directe) soit délétères (elle diminue le fitness de la bactérie). Quelques unes apportent un avantage, c'est celles-là qui seront sélectionnées. Selon le paradigme actuel, les mutations se produisent de manière strictement aléatoire lors des erreurs de réplication. Par conséquent, la bactérie ne peut pas choisir le nucléotide à muter en fonction de l'environnement (Lamarckisme). En revanche, la bactérie peut jouer sur le taux de mutation via le contrôle des mécanismes de réparation (Barrick *et al.*, 2009). Dans un environnement hostile par exemple, avoir un taux de mutation plus élevé peut permettre de « fouiller » plus vite les possibilités d'adaptation mais le prix à payer c'est un grand nombre de mutants, eux, beaucoup moins

adaptés.

## 2.2 Le transfert horizontal de gènes

Le transfert horizontal de gènes est un processus dans lequel un organisme intègre du matériel génétique provenant d'un autre organisme sans en être le descendant. Chez les bactéries, il survient de 3 manières : par conjugaison, par transduction ou par transformation. La conjugaison est le transfert de matériel génétique (souvent un plasmide) par contact direct de cellule à cellule. A l'inverse, la transduction et la transformation s'effectuent sans contact intercellulaire. La transduction permet la récupération d'ADN provenant d'un autre organisme par l'intermédiaire d'un virus (chez les bactéries, on parle de bactériophage) tandis que la transformation est la récupération d'ADN extracellulaire présent dans l'environnement. On estime qu'au moins 18% du génome de *E. coli* provient de transfert horizontal de gènes (Lawrence & Ochman, 1998). Le transfert horizontal a sans doute joué un rôle important dans l'évolution des réseaux de régulation (McAdams *et al.*, 2004; Price *et al.*, 2008) en fournissant aux bactéries des modules complets dont les paramètres peuvent être facilement « tunés » avec des mutations. Notez que les algorithmes génétiques informatiques sont de redoutables heuristiques qui permettent de résoudre des problèmes d'optimisation difficiles. Ils tirent en partie leur puissance de la capacité à combiner ensemble de « bons » paramètres par transfert horizontal de gènes (équivalent du crossing-over).

## 2.3 La sélection naturelle

Dés lors que la population n'est plus génotypiquement homogène, les différents génotypes se retrouvent en compétition pour les ressources de l'environnement. La sélection naturelle explique comment l'environnement influe sur l'évolution des génomes en sélectionnant les individus les plus adaptés. De manière plus formelle, *« La sélection naturelle désigne le fait que les traits qui favorisent la survie et la reproduction, voient leur fréquence s'accroître d'une génération à l'autre. Cela découle logiquement du fait que les porteurs de ces traits ont plus de descendants, et aussi que ces derniers portent ces traits »* [3]. Derrière l'apparente simplicité du concept de sélection naturelle se cache le débat le plus intense et le plus passionné de toute la philosophie de la biologie au cours des quarante dernières années : le débat sur les unités de sélection —débat toujours ouvert aujourd'hui en 2011—.

C'est l'ouvrage controversé de Richard Dawkins *Le gène égoïste* publié en 1976, vendu à

[3] Wikipedia, article sur la « sélection naturelle ».

plus d'un million d'exemplaires et traduit dans 25 langues qui « enflamme » le débat parmi les biologistes et philosophes de la biologie (dont les principaux représentants sont Dawkins, Gould, Lewontin, Maynard-Smith, Mayr et Hull). La proposition de Dawkins tient en une phrase : bien que l'on ait cru jusqu'ici que la bonne unité de sélection était l'individu, en réalité la bonne unité de sélection est le gène. Selon lui, les gènes qui se sont imposés sont ceux qui provoquent des effets qui servent leurs intérêts propres (c'est-à-dire de continuer à se reproduire) et pas forcément les intérêts de l'individu. Citons l'exemple du comportement instinctif (et donc dicté par leurs gènes) de ces araignées mâles qui tentent de se reproduire tout en s'exposant au cannibalisme de la femelle. Cette vision des choses explique aussi l'altruisme au niveau des individus. Par exemple, quand un individu se sacrifie pour protéger la vie d'un membre de sa famille, il agit dans l'intérêt de ses propres gènes.

Dans le deuxième chapitre sur l'évolution, je formulerai une critique de la notion de gène égoïste et reviendrai plus en détail sur le débat des unités de sélection. De manière très simpliste, notez que les opposants à l'idée que le gène est la seule unité de sélection défendent une vision plus hiérarchique. Selon eux, la sélection naturelle peut s'opérer à différents niveaux (gènes, génomes, cellules, organismes, espèces, populations, écosystèmes...).

## 2.4 La dérive génétique

La dérive génétique est l'évolution d'une population ou d'une espèce causée par des phénomènes aléatoires, impossibles à prévoir. La dérive peut avoir pour conséquence la fixation de mutation délétère ou au contraire la perte de mutation bénéfique. Elle est causée par le biais d'échantillonnage lié au fait que les effectifs d'une population sont toujours finis. Par conséquent, son effet est plus fort sur les petites populations mais agit également sur les grandes.

Pour illustrer la différence essentielle entre dérive génétique et sélection naturelle, je vais décrire deux expériences d'évolution expérimentale très proches mais aux conséquences opposées.

Dans la première expérience, on part d'une colonie isolée que l'on met dans un milieu nutritif liquide bien défini. Les bactéries se divisent de manière exponentielle et une fois qu'elles atteignent la phase stationnaire (par exemple $10^7$ bactéries), on les dilue 100 fois ($10^5$ bactéries) dans du nouveau milieu frais de même composition. Et on fait cela tous les jours ce qui aboutit à des milliers de génération [4]. Les bactéries qui ont des mutations bénéfiques se divisent plus vite. Par conséquent, elles transmettent plus souvent leurs mutations à leur descendance et

[4]Cette expérience a été initié par Richard Lenski il y a quelques décennies et est toujours en cour aujourd'hui. Une des équipes (dirigé par le Pr. Dominique Schneider)de mon laboratoire travaille sur les données issues de ces expériences

petit à petit cette dernière devient majoritaire. Ce que l'on observera c'est une augmentation du fitness au cours des générations successives. Les bactéries s'adaptent de mieux en mieux à leur environnement. C'est l'effet de la sélection naturelle.

Maintenant faisons une autre expérience. On part d'une colonie isolée que l'on strie sur boite (milieu solide nutritif bien défini) ce qui permet d'isoler des bactéries à différents endroits sur la boite. Pendant la nuit, les bactéries isolées vont se diviser une fois, deux fois, 4 fois de manière exponentielle et le lendemain matin, on observera des colonies rondes (des millions de bactéries) provenant initialement de bactéries uniques. On re-strie une colonie sur boite pour isoler à nouveau des bactéries et ainsi de suite tous les jours pendant des milliers de générations. Contrairement à l'expérience précédente, ce que l'on observera c'est une diminution du fitness au cours des générations successives. Les colonies successives deviennent de moins en moins adaptées. Autrement dit, les bactéries se divisent de moins en moins vite. C'est l'effet de la dérive génétique. Pourtant, cette expérience ressemble beaucoup à l'expérience précédente : Pourquoi observons-nous alors le résultat inverse ?

La différence essentielle vient de l'échantillonnage très différent. Dans la première expérience, on dilue 100 fois les bactéries dans du milieu frais tous les jours. Les $10^5$ bactéries que l'on récupère sont un panel représentatif des $10^7$ bactéries ayant participé à la compétition et donc ayant subi les effets de la sélection naturelle. Au contraire, dans la deuxième expérience, le passage de la colonie à la bactérie unique crée une dilution bien plus forte, un énorme goulot d'étranglement. En sélectionnant une colonie provenant d'une bactérie unique, l'expérimentateur ne « capitalise » pas les effets de la sélection naturelle. Au fil des cycles, les colonies successives accumulent des mutations délétères (statistiquement beaucoup plus nombreuses que les mutations bénéfiques) ce qui explique la diminution du fitness au cours des générations.

## 2.5 Les contraintes évolutives

Selon Stephen J. Gould, l'évolution est contingente (possibilité qu'une chose arrive ou n'arrive pas) ce qui signifie que si l'on devait « rejouer » les 4 milliards d'années d'évolution depuis l'apparition de la vie, il est peu probable que l'homme réapparaitrait. L'apparition de l'homme n'est pas nécessaire. En évolution, « tout n'est pas mieux dans le meilleur des mondes » ce qui signifie que Pangloss[5] et les « adaptationnistes » se trompent. L'évolution contient une part de hasard et la sélection naturelle ne saurait aboutir à la perfection, à une adaptation optimale. L'incapacité qu'à la sélection naturelle à conduire à l'optimisation idéale est due à la présence de contraintes que Maynard-Smith définit comme cela :

[5]Pangloss est le précepteur de Candide dans le conte philosophique de Voltaire (1759).

> *« Les contraintes développementales peuvent se définir comme un biais dans la production de phénotypes variants, ou une limitation de la variabilité phénotypique, causée par la structure, le caractère, la composition ou la dynamique du système développemental »* (Maynard-Smith *et al.*, 1985)[6].

Penser en termes de contraintes aide à s'affranchir d'une des dérives du darwinisme qui consiste à ne raisonner qu'en termes de fonctions c'est à dire d'utilité immédiate des adaptations. En effet, d'autres explications sont possibles pour expliquer l'existence des structures biologiques. Pour comprendre cela, je vais rappeler l'exemple très classique (et polémique) des spandrels (en réalité des pendentifs) que l'on doit à Gould et Lewontin (Gould & Lewontin, 1979). Ces derniers notent que les espaces entre les arches qui supportent le toit de la basilique St Marc à Venise sont magnifiquement décorés. Les auteurs soutiennent que bien que ces espaces triangulaires aient été admirablement bien utilisés, ils ne sont pourtant que des sous produits architecturaux vis-à-vis de l'emploi des arches soutenant le dôme du toit. En d'autres termes, leur présence n'est pas due initialement à leur utilisation artistique. La leçon à retenir est simple : lorsque l'on se trouve face à un trait particulier, on ne devrait pas conclure que sa fonction particulière (ici artistique) est la raison pour laquelle le trait est présent. On retiendra donc que chacune des caractéristiques d'un organisme n'est pas toujours le produit direct de la sélection.

C'est dans la lignée de cette intervention sur les spandrels qu'un nouveau terme apparait en 1982 : l'exaptation (Gould & Vrba, 1982). Gould et Vrba reprennent l'idée que la sélection qui s'exerce sur une structure et la maintient en vue d'une « utilité » actuelle n'est pas toujours celle qui explique son apparition originelle. Le rôle passé et le rôle présent de la sélection pour les caractéristiques des organismes peuvent être tout à fait différents. Nietzsche, qui s'intéressait beaucoup à la biologie, avait déjà décrit ce concept il y a plus d'un siècle :

> *« Tout événement du monde organique est une manière de subjuguer, de dominer et toute subjugation, toute domination, à leur tour équivalent à une nouvelle interprétation, à un accommodement, où le sens et le but antérieur s'obscurciront nécessairement et même disparaitront tout à fait. On a beau comprendre l'utilité d'un organe physiologique (ou d'une institution juridique, d'une coutume sociale, d'un usage politique, ou encore d'une forme artistique ou d'un culte religieux), on n'a pour autant rien compris encore à sa naissance : pour gênant et désagréable que cela soit à de vieilles oreilles - de tout temps en effet on a cru que la finalité démontrable, l'utilisé d'une chose, d'une forme, d'une institution, était aussi la cause*

[6] Tiré du site Web de Laurent Penet.

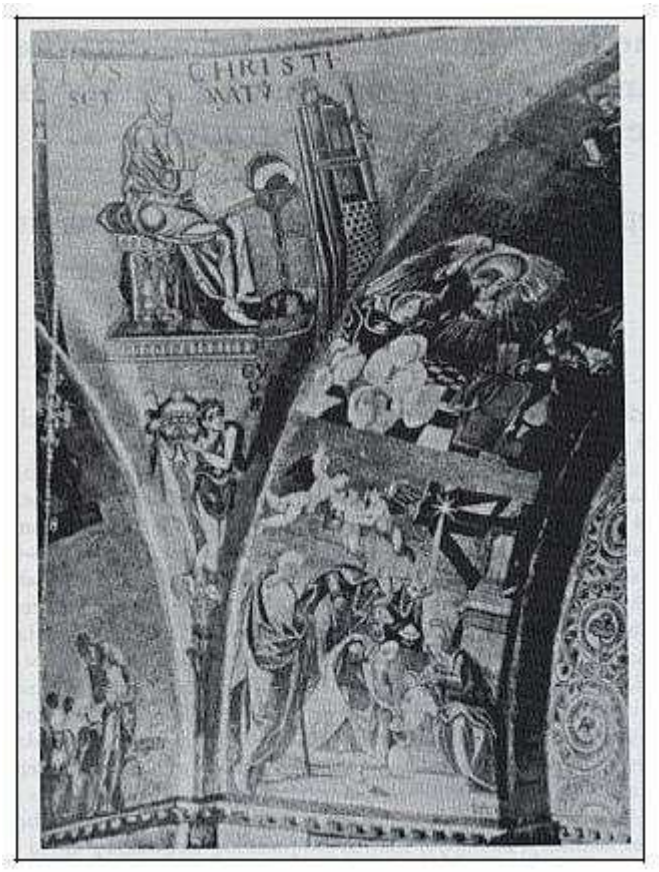


**Figure IX.1** – Les spandrels de la cathédrale San Marco à Venise (image de Gould et Lewontin).

*de leur naissance : l'œil aurait été fait pour voir, la main pour saisir.[...] Tout but, toute utilité ne sont cependant que des symptômes indiquant qu'une volonté de puissance s'est emparée de quelque chose de moins puissant qu'elle et lui a de son propre chef imprimé le sens d'une fonction ; et toute l'histoire d'une chose, d'un organe, d'un usage peut être ainsi une chaîne continue d'interprétations et d'adaptations toujours nouvelles, dont les causes ne sont même pas nécessairement en rapport les unes avec les autres, mais peuvent se succéder et se remplacer les unes les autres de façon purement accidentelle. »*[7]

Un très bel exemple d'exaptation a été découvert en 2009. Certains mammifères nocturnes utilisent une stratégie peu commune pour la vision nocturne (pour détecter des photons uniques). Quand la lumière s'estompe, la chromatine des noyaux de leurs cellules en bâtonnet change la transcription d'une part mais se réorganise aussi de telle manière à se comporter

[7]Nietzsche, *La généalogie de la morale*, Folio p. 85

comme une lentille. Cela permet de focaliser les photons sur les récepteurs photosensibles là où le signal lumineux est ensuite traduit en influx nerveux. Ainsi, la quantité de lumière détectée est augmentée, un avantage certain dans des conditions de faible luminosité. On comprend bien que cette nouvelle fonction qu'a l'ADN (lentille) n'a rien avoir avec son rôle de support du programme génétique. Il serait stupide de dire que l'évolution a inventé l'ADN en cherchant à optimiser la vision nocturne. Ici l'arche qui soutient le dôme c'est l'ADN support de l'information génétique. A partir de là, il y a de nouveaux espaces qui sont contraints (les spandrels) et que l'évolution utilise pour faire apparaitre de nouvelles fonctionnalités (l'ADN comme lentille).

# 3 Evolution et réseau de régulation

*Pourquoi l'évolution aurait t'elle sélectionné la présence d'un réseau plutôt que son absence ?*

Nous savons que les gènes sont modifiés par l'évolution. Mais les promoteurs, RBS, facteurs de transcription eux aussi subissent les effets des mutations ce qui fait que le réseau est évidemment lui aussi finement « tuné » par l'évolution. Quel peut-être l'intérêt pour une bactérie de posséder un réseau de régulation couteux métaboliquement alors qu'elle pourrait se contenter de garder constante et optimale l'activité du promoteur de chaque gène ? Et si la bactérie a un intérêt dans la possession d'un réseau de régulation, ce dernier est-il toujours optimal ? Optimisé par l'évolution ?

Dés lors que l'on est confronté à un problème d'optimisation, il faut pouvoir définir précisément la fonction à optimiser ce qui n'est pas toujours évident car nous ne la connaissons pas toujours : par exemple quelle est la fonction à optimiser des cellules d'un tissu particulier, par exemple la peau ? On ne sait pas. Avec les bactéries, les choses sont plus simples : dans le cas théorique où les bactéries poussent dans un environnement constant constamment réapprovisionné, on peut utiliser le taux de croissance comme fonction à optimiser (le fitness). En effet, dans ce cas, la pression évolutive maximise le taux de croissance. Pour calculer cette fonction de fitness, il faut réfléchir en termes de coût et de bénéfice. Nous allons voir comment la sélection naturelle peut :

⇒ adapter le niveau d'expression d'un gène.

⇒ choisir de réguler ou pas l'expression d'un gène.

⇒ choisir d'utiliser un « motif » complexe pour réguler l'expression d'un gène par rapport à une simple activation/répression.

## 3.1 Adapter le niveau d'expression d'un gène

Qu'est-ce qui détermine le niveau optimal d'expression du gène *lacZ* codant pour la $\beta$-galactosidase, l'enzyme nécessaire pour utiliser le lactose comme source d'énergie ? Le fitness f de la bactérie dépendra du bénéfice b qu'elle tire de la présence de $\beta$-galactosidase auquel il faut soustraire le coût c de production de cette enzyme.

$$f = b - c \tag{IX.1}$$

La courbe du taux de croissance en fonction de la concentration de $\beta$-galactosidase a une forme de cloche. Quand la concentration de $\beta$-galactosidase est trop faible (non optimale), la bactérie dégrade le lactose trop lentement et pousse donc également lentement. Si on augmente la concentration, on finit par atteindre le maximum de la courbe en cloche : la concentration de $\beta$-galactosidase est optimale et par conséquent, l'ajout de nouvelles enzymes n'améliore plus le taux de croissance. Au contraire, le coût métabolique (transcription et traduction) ainsi que la monopolisation de ressources (acides aminés) qui pourraient être utilisées pour d'autres protéines finit par diminuer le taux de croissance.

L'équipe d'Uri Alon[8] a mené une expérience d'évolution expérimentale. Ils sont partis de plusieurs tubes, chacun avec une concentration différente de lactose, dans lesquels ils ont fait évoluer indépendamment une souche sauvage d'*E. coli*. Ils ont pu montrer qu'en l'espace de quelques centaines de générations, des mutations apparaissent permettant aux bactéries d'adapter leur concentration de $\beta$-galactosidase à un niveau optimal : forte concentration de $\beta$-galactosidase quand il y a une forte concentration de lactose et inversement. Bien sûr, la population la mieux adaptée prenait très rapidement le dessus sur les autres (Dekel & Alon, 2005).

## 3.2 Réguler ou ne pas réguler

Dans cette partie, on va étudier pourquoi certains gènes sont régulés alors que d'autres sont exprimés de manière constitutive. Il nous faut tout d'abord considérer, non plus un environnement constant, mais un environnement variable. Par exemple, dans cet environnement variable, il y a parfois du lactose et parfois il n'y en a pas. Or le gène *lacZ* n'apporte un bénéfice direct au fitness que lorsque le lactose est présent. La présence d'une régulation signifie que le gène *lacZ* n'est produit que lorsque le lactose est présent. Mais cette régulation a un coût : le coût de sa production et de sa maintenance. Ainsi la régulation n'aura un intérêt que si son bénéfice est

[8]Plusieurs paragraphes de cette partie « évolution et réseau de régulation » se basent largement sur les chapitres 10 et 11 de l'ouvrage de Uri Alon, *An introduction to system biology* , Chapman & Hall.

supérieur à son coût. Pour analyser la stratégie optimale, nous allons comparer trois organismes avec 3 designs différents pour la régulation du gène *lacZ*.

Appelons p la fraction du temps où le lactose est présent dans le milieu. p représente la demande pour l'expression de *lacZ*. Le reste du temps (1-p), l'expression de *lacZ* est inutile. Dans l'organisme 1, *lacZ* est exprimé de manière constitutive (sans régulation). Le bénéfice b n'a lieu que lors de la fraction p du temps, quand le lactose est présent. Son fitness $f_1$ est donc :

$$f_1 = pb - c \tag{IX.2}$$

L'organisme 2 possède une régulation : il n'exprime *lacZ* qu'en présence de lactose. De cette manière, l'organisme économise la production non nécessaire lorsque le lactose est absent : il ne paye donc le coût c qu'une fraction du temps p. Cependant il faut payer le prix r pour produire et maintenir le système de régulation.

$$f_2 = pb - pc - r \tag{IX.3}$$

Enfin, le troisième organisme ne produit pas du tout *lacZ*. Le fitness $f_3$ sur lactose est égal à zéro : c'est le niveau de référence où il n'y a ni coût ni bénéfice.

$$f_3 = 0$$

La régulation est sélectionnée par l'évolution lorsque l'organisme 2 a le plus haut fitness c'est-à-dire lorsque $f_2 > f_1$ et $f_2 > f_3$. Cela nous amène aux deux inégalités suivantes :

$$p < 1 - \frac{r}{c} \;\; et \;\; p > \frac{r}{b-c}$$

Ces inégalités relient une propriété de l'environnement (la demande pour l'expression de *lacZ* correspondant à la fraction du temps où le lactose est présent) aux paramètres de coût et de bénéfice de l'expression de *lacZ* ainsi que son système de régulation. Pour chaque valeur de p, ces équations nous disent si le système de régulation sera ou non sélectionné par rapport aux autres designs plus simples. La gamme des environnements dans lesquels chacun des 3 designs est optimal est montrée dans la figure IX.2.

## 3.3 Sélection d'un motif

Nous avons vu dans le chapitre sur les systèmes dynamiques qu'il existe certains motifs (des circuits élémentaires récurrents) qui apparaissent statistiquement plus souvent dans le réseau

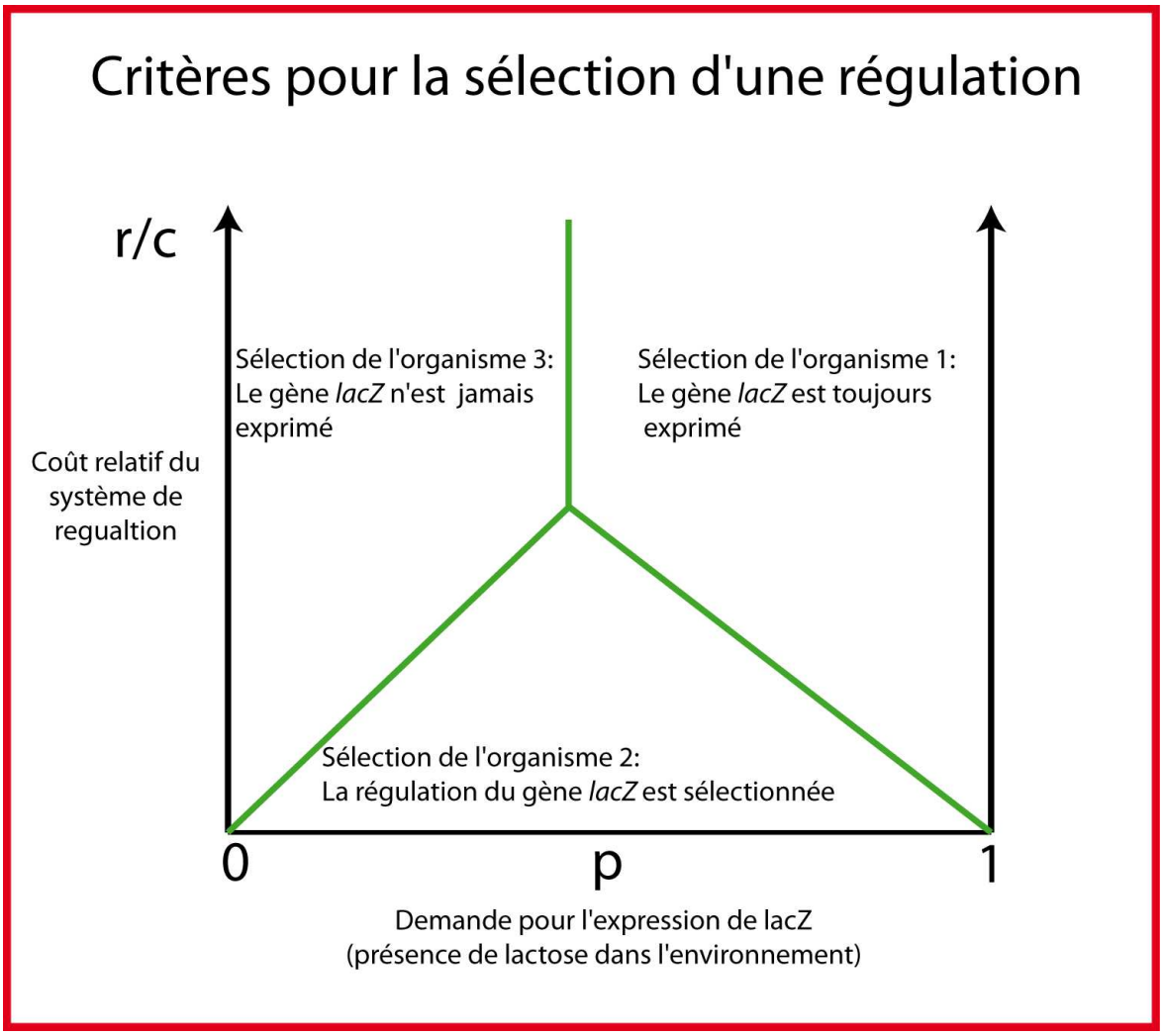



**Figure IX.2** – Diagramme de phase montrant les régions où la régulation de *lacZ*, son expression constitutive ou son absence d'expression sont optimales. L'axe des x représente la fraction du temps où le lactose est présent. L'axe y exprime le coût relatif du système de régulation. Cette figure est reprise et adaptée à partir de l'ouvrage d'Uri Alon p. 203.

de régulation génique de *E. coli* par rapport à un réseau strictement aléatoire. Il semble donc que ces motifs aient été sélectionnés par l'évolution. Nous avons vu que l'un de ces motifs, le feed forward loop, agit comme un détecteur de persistance.

Dans un environnement où la présence et l'absence de lactose varie au cours du temps, il peut y avoir des petits pulses de lactose, par exemple de quelques minutes, et des pulses plus longs où le lactose reste présent pendant plusieurs heures. Si le système de régulation du lactose déclenche l'expression de *lacZ* lors des petits pulses, le coût risque d'être plus important que le bénéfice car peut être que lorsque les premières molécules de $\beta$-galactosidase arriveront, le lactose aura déjà disparu. La présence d'un motif régulateur plus complexe (le feed forward

loop) filtre ces petits pulses et empêche qu'ils ne déclenchent l'induction de *lacZ*. Il en résulte un gain de fitness potentiel. En utilisant, la même stratégie d'étude des coûts–bénéfices, on montre *facilement* qu'il existe un espace dans le diagramme des phases où le Feed forward loop est sélectionné par rapport à une régulation simple (Dekel *et al.*, 2005).

Selon l'équipe d'Uri Alon, la présence d'un motif donné (par exemple le FFL) et donc sa sélection par l'évolution, signifie qu'il est le mode de régulation optimal par rapport à une condition environnementale donnée. Par exemple, si la régulation de *lacZ* est contrôlée par un FFL, cela signifie qu'il doit exister de nombreux pulses de lactose dans l'environnement qui nécessitent d'être filtrés. Autrement dit, on peut imaginer faire de l'écologie inversée et déduire des informations de l'environnement que rencontre une bactérie en fonction de son réseau de régulation génique. Cela est basé sur l'idée qu'un circuit optimal contient/est un *modèle interne de l'environnement*.

## 3.4 La loi de la demande

Pour continuer sur cette idée, vous avez peut-être remarqué que l'activateur et le répresseur transcriptionnel ont exactement le même *but* : n'exprimer un gène qu'en présence d'un signal donné. En effet, le signal peut activer un activateur (arabinose–AraC) ou inactiver un répresseur (lactose[9] –LacI). Parmi ces deux systèmes équivalents, l'évolution en choisit un. Mais y-a-t-il des règles qui gouvernent cette sélection ou est-ce que l'évolution choisit au hasard répression ou activation ?

Cette question a été soulevée par Savageau qui a montré que le mode de contrôle est corrélé à la demande (fraction du temps où l'expression du gène est nécessaire). Les gènes qui sont souvent « demandés » tendent à être contrôlés par une régulation positive. A l'inverse, les gènes dont le produit est peu demandé, tendent à être contrôlés négativement. Vous voyez donc qu'avec cette règle, en connaissant le mode de régulation d'une enzyme (positive ou négative), vous pouvez inférer l'abondance ou la rareté du substrat correspondant de l'environnement « sauvage » de la bactérie.

L'équipe d'Uri Alon a proposé une explication à ce phénomène. Elle repose sur le fait que quand un site de fixation d'un régulateur sur l'ADN est « libre », il est plus exposé à des fixations non spécifiques alors que le site sur lequel est fixé le bon régulateur est, lui, mieux protégé des erreurs. En effet, ces erreurs causées par une fixation d'un régulateur non spécifique peuvent changer l'expression du gène (de manière non voulue) et diminuer le fitness de la bactérie à cause du coût inutile que cela entraine.

[9]En réalité, il s'agit de l'allolactose.

| Sucre | regulation | regulateur | demande |
|---|---|---|---|
| arabinose | positive | araC | forte |
| rhamnose | positive | rhaS | forte |
| fucose | positive | fucR | forte |
| glycérol | négative | glpR | faible |
| galactose | négative | galR,galS | faible |
| lactose | négative | lacI | faible |

Considérons maintenant que l'expression de *lacZ* est contrôlée par un activateur. Si le lactose est présent très souvent, alors l'activateur est très souvent fixé sur le promoteur ce qui protège des erreurs de fixation non spécifique réduisant le fitness. A l'inverse, si le lactose est très rare, l'activateur ne sera que rarement fixé au promoteur. Ce dernier sera alors bien plus exposé aux erreurs de fixation non spécifique ce qui réduira le fitness de la bactérie. Dans ce deuxième cas, la meilleure stratégie consiste à réguler l'expression de *lacZ* avec un répresseur de telle manière que le site ADN soit protégé la majorité du temps.

Imaginez maintenant que pendant des milliers d'années (ou de générations quand on parle en temps évolutif), une bactérie ait été exposée très souvent au lactose. Par conséquent, elle régule positivement l'expression de *lacZ*. Imaginez maintenant que, suite à un changement climatique par exemple, le lactose devienne plutôt rare. L'activateur étant moins souvent fixé sur le promoteur, il va y avoir une augmentation des fixations non spécifiques qui déclenchent l'expression de *lacZ* inutilement. Est-ce que le coût sur le fitness généré peut avoir pour conséquence d'inverser, au fil du temps évolutif, le mode de régulation (de activateur à répresseur) ? Et bien il semble que la réponse soit oui (Shinar *et al.*, 2006) : des mutants avec le mode de régulation inverse finissent par être sélectionnés.

## 3.5 Apprentissage associatif

Nous avons vu dans les exemples ci-dessus qu'il peut exister une corrélation entre la topologie et les paramètres d'un réseau de régulation d'une part et certaines conditions environnementales d'autre part. Avant de donner un nouvel exemple de la puissance de l'évolution pour modeler un réseau de régulation reflétant l'environnement extérieur, je souhaite vous ramener au lycée en vous rappelant la petite expérience de Pavlov. Le chien de ce dernier avait « associé » le son d'une sonnette avec l'arrivée de la viande. Il suffisait ensuite d'agiter la sonnette (sans forcement apporter la viande) pour que le chien se mette à saliver. Evidement, ce type d'apprentissage associatif passe par le système nerveux. Et bien, on pense que les bactéries sont capables d'apprentissage associatif via, non plus le système nerveux (qu'elles ne possèdent pas), mais via

leur réseau de régulation.

Quand *E. coli* arrive dans notre bouche, elle va tout suite faire l'expérience d'une augmentation de température (37°C). Puis elle se retrouve dans le tractus gastro-intestinal où nous avons vu qu'elles sont en compétition avec de centaines d'autres espèces bactériennes (*E. coli* y est minoritaire <1%). Dans le tractus gastro-intestinal, la concentration en oxygène chute et les bactéries doivent « switcher » sur un programme de respiration anaérobie. Les bactéries peuvent détecter cette chute et adapter leur expression génique pour y faire face. Mais la compétition est rude dans les intestins et une bactérie qui « switch » en mode anaérobie dés l'augmentation de température dans la bouche, sera tout de suite « prête » à affronter les conditions anaérobiques des intestins. Il n'y aura pas le « délai » classique de « détection–expression ». Le gain de fitness est probablement suffisant pour permettre la sélection de cette bactérie. Ainsi, l'évolution détecte la corrélation [augmentation de la température– diminution de la pression partielle en oxygène] et l'imprime dans le réseau de régulation de la bactérie (Tagkopoulos *et al.*, 2008). Ces résultats laissent donc bien supposer que l'on peut travailler de manière inverse dans un laboratoire (reverse-ingénierie) : l'étude des corrélations (temporelles ou pas) présentes dans l'expression génique d'un organisme pourrait nous aider à reconstituer les corrélations d'ordre écologiques que vit ce même organisme.

## 3.6 Le reflet du miroir

La métaphore du « reflet du miroir » entre environnement et réseau de régulation/génome semble encore appropriée avec l'exemple ci-dessous. Une des questions essentielles à l'étude des réseaux de régulation est : sont-ils modulaires ? Et si oui, comment et pourquoi l'évolution a-t-elle sélectionné la modularité ? Kashtan a fait évoluer des réseaux de régulation *in silico* en utilisant des algorithmes évolutifs classiques. *Il a pu démontrer que « ses » réseaux deviennent modulaires s'il change l'environnement de manière modulaire.* Chaque environnement correspond en fait à un but évolutif (une fonction de fitness) et chacun de ces buts évolutifs est composé de sous buts (Kashtan & Alon, 2005). Selon Kashtan, les variations dans l'environnement pourraient accélérer l'évolution : plus il y aurait de changements dans l'environnement, plus une bactérie « naviguerait » facilement dans le paysage adaptatif à la recherche de l'optimum global (Kashtan *et al.*, 2007). Kashtan n'a pas manqué de voir que cette manière de procéder pourrait significativement améliorer les algorithmes évolutifs (informatiques) d'optimisation.

# 4 Evolution et biologie synthétique

Depuis plusieurs milliers d'années déjà, l'homme utilise l'évolution pour sélectionner des traits qui l'intéressent chez certaines espèces. L'exemple de sélection artificielle qui me semble le plus percutant c'est celui du loup et du chien. L'hypothèse communément admise chez les scientifiques est que le chien descend du loup. On pense que les chasseurs–cueilleurs adoptaient parfois des bébés loup. En grandissant, certains des louveteaux devenaient agressifs et étaient alors tués. Les loups au caractère « dominé » étaient ainsi sélectionnés pour leur absence d'agressivité, leur « bonne compagnie ». L'homme en protégeant et nourrissant les loups (devenus chien) a créé un barrage de protection génétique contre les effets de la sélection naturelle (l'arche qui soutient le toit de la cathédrale). L'espace créé derrière ce barrage (les spandrels) est à l'origine de la très forte instabilité génétique du chien dont la conséquence est la multiplicité de races phénotypiquement très différentes. La fonction sélectionnée par l'homme, la « bonne compagnie » ne pourrait pas apparaitre seule dans la nature sans la présence de l'homme. Dit humoristiquement, si vous abandonnez un caniche royal en Sibérie, il est peu probable qu'il devienne le mâle « alpha » d'une meute de loup.



**Figure IX.3** – L'instabilité génétique du chien. Ce caniche royal est devenu le mâle alpha de cette meute de loups par la seule force de son caractère ☺

Avec cet exemple du chien, on voit que l'homme a construit (au fil des décennies/siècles) un organisme capable d'exécuter une fonction. La méthode est une méthode « à l'aveugle » qui utilise le principe de sélection artificielle : on ne garde que les individus qui s'approchent le plus de la fonction désirée et on les fait se reproduire entre eux. On est tellement habitué à cette manière de procéder que cela ne nous choque pas. Cela semble presque naturel. En biologie synthétique, les choses sont un peu différentes. On vient spécifiquement changer ce *nucléotide là* sur l'ADN pour transformer le molosse en caniche (j'exagère un peu). L'évolution change le même nucléotide mais sans fournir l'information. En ce début de 21$^{\text{ème}}$ siècle, pour le même résultat (obtenir telle fonction), la présence du hasard couplé à l'absence d'information rassure (dieu, la nature) alors que la connaissance de l'information (savoir quel nucléotide changer) apparaît diabolique.

Le but du biologiste synthétique consiste à construire un organisme, à contrôler une fonction en se passant de l'évolution. Le rêve serait de pouvoir, étant donnée une séquence nucléotidique, calculer la fonction (l'organisme) ou au contraire, étant donnée une fonction (l'organisme) calculer la (ou les) séquence nucléotidiques.

C'est exactement ce qu'a fait Michael Elowitz avec son plasmide repressilator (Elowitz & Leibler, 2000) : il combine des petites séquences nucléotidiques simples : 3 promoteurs et 3 gènes dont il connait la fonction pour créer une séquence nucléotidique plus compliquée qui code pour une fonction, elle aussi, plus complexe : des oscillations dans la concentration de protéines au cours du temps. Et ça a marché. Mais cette approche « plug and play » (empruntée à l'électronique) qui consiste à combiner ensemble des séquences nucléotidiques avec des fonctions simples pour générer des fonctions plus complexes montre très vite ses limites en biologie. La raison est simple : le niveau de standardisation de ces briques de base est très faible par rapport aux briques de base des microprocesseurs (les transistors). Le niveau d'incertitude et de méconnaissance est très important et pose donc très vite problème. Imaginez que vous combiniez quelques « biobriks » ensemble pour générer une fonction complexe et qu'une fois dans la cellule, la séquence ne génère pas le phénotype attendu ? Ce n'est peut être qu'un paramètre (par exemple la force d'un promoteur) qui mérite d'être ajusté ? Mais comment savoir quel nucléotide modifier ? On peut faire du « rational design » et tenter de modifier un nucléotide donné par mutagénèse dirigée en espérant être en mesure de prédire correctement le résultat mais c'est une approche chronophage et donc risquée. Une solution consiste à réutiliser l'évolution (une dernière fois) pour tuner finement les paramètres, c'est-à-dire pour accroitre la connaissance/l'information sur la séquence en vue d'obtenir la meilleure correspondance fonction–séquence. On parle d'évolution dirigée.

## 4.1 Utiliser l'évolution pour « tuner » un réseau de régulation synthétique

Il existe plusieurs étapes majeures dans une expérience d'évolution dirigée. La diversité génétique est crée par mutagenèse aléatoire et/ou recombinaison *in vitro* de une ou plusieurs séquences parentes (DNA shuffling). Les séquences altérées sont ensuite clonées dans un plasmide pour expression du phénotype dans la souche hôte (par exemple *E. coli*). Les clones qui expriment la fonction améliorée sont identifiés par screening à haut débit. Les séquences qui encodent la fonction améliorée sont isolées pour le prochain cycle d'évolution dirigé.

Pour être clair, prenons l'exemple de l'amélioration de la GFP par évolution dirigée. Les chercheurs voulaient créer une GFP encore plus fluorescente. Ils ont crée une banque de séquences de GFP par de techniques de recombinaison *in vitro*. Les séquences sont clonées dans un plasmide pour expression chez *E. coli*. Il suffit ensuite d'identifier parmi les milliers de colonies, celles qui sont les plus fluorescentes puis de refaire ce cycle plusieurs fois en repartant des colonies sélectionnées. Par cette méthode, en 3 cycles seulement, les chercheurs ont réussi à obtenir une GFP 45 fois plus fluorescente (Crameri *et al.*, 1996).

En appliquant cette méthode non plus à la GFP mais à un facteur de transcription, vous voyez que l'on peut modifier certains paramètres, par exemple diminuer ou augmenter son affinité à l'ADN. Notez qu'il existe une alternative au screening qui consiste à utiliser la sélection pour identifier le phénotype recherché. Dans ce cas, on couple la fonction recherché à la survie cellulaire, par exemple en utilisant une cassette de résistance à un antibiotique. Cela permet d'assurer la pression évolutive nécessaire à la sélection.

Mais vous allez me dire « c'est génial, avec l'évolution dirigée, on peut tout faire ! »

Par exemple, pourquoi ne pas partir d'un plasmide composé du même nucléotide, par exemple 10000 adénosines les unes à la suite des autres. On pourrait diriger l'évolution de cette séquence pour qu'elle encode une fonction donnée sélectionnée par l'expérimentateur, par exemple les oscillations dans la concentration de protéine.

Cela n'est évidemment pas possible pour une raison simple : l'espace des séquences à fouiller est astronomique. il y a 4 bases possibles (A,T,C,G) et 10000 pb soit $4^{10000}$ séquences possibles. Le nombre de colonies que l'on peut raisonnablement « screener » ne peut guère dépasser le million. Mais même si on pouvait « screener » des milliards de milliards de séquences, on ne fouillerait qu'une infime partie de l'espace des séquences possibles. La probabilité de tomber au hasard sur la séquence du repressilator (séquence générant les oscillations) est quasi nulle.

La bonne démarche en faisant de l'évolution dirigée en biologie synthétique consiste à faire en sorte que la séquence qui encode la fonction recherchée ait une bonne probabilité d'être

présente parmi les quelques milliers/millions de séquence « screenées ». Pour cela, il y a deux solutions :

⇒ si la séquence est longue (un gène de 1000 pb), il faut que la séquence recherchée ne soit qu'à quelques mutations (ou événements de recombinaison) de la séquence d'origine.

⇒ si la séquence est courte, par exemple un RBS (ribosome binding site) de 7 pb et que l'on souhaite modifier l'efficacité de la traduction du gène en aval, on peut facilement tester l'ensemble des combinaisons $4^7$ soit 16384 séquences car cela correspond à un nombre de colonies « screenable ». Cette approche a été utilisée pour « tuner » un petit réseau synthétique permettant à *E. coli* de *décider* d'envahir ou pas des cellules cancéreuses en fonction des conditions extérieurs (Anderson *et al.*, 2006).

Nous avons vu que la démarche « plug and play » consistant à créer un petit réseau à partir de briques (séquences) de base est une approche qui montre très vite ses limites car la standardisation est moins aisée qu'en électronique. La bonne démarche est plutôt « plug, mutate and play ». Le « plug » permet de réduire l'espace des séquences en fournissant la structure générale du système, le « mutate » se charge de modifier les quelques bases nécessaires pour obtenir les paramètres compatibles avec la fonction recherchée. Pour continuer l'analogie avec l'informatique, on peut dire que l'étape « mutate » correspond à l'étape de debugage d'un programme informatique : on modifie quelques caractères (parenthèses oubliées...) ou quelques segments pour que le code fonctionne. Un fois debuggé, il ne reste alors plus qu'à exécuter le programme (le « play » ). La correspondance séquence–fonction est alors optimale. Je vais maintenant décrire un bel exemple de la méthode « plug, mutate and play » issu des travaux de l'équipe de Frances H. Arnold (Yokobayashi *et al.*, 2002).

Les chercheurs créent un petit réseau linéaire très simple où la protéine répresseur LacI réprime l'expression de la protéine répresseur cI qui, elle-même réprime l'expression de la yellow fluorescent protein (YFP). En présence de l'inducteur IPTG qui « inactive » allosteriquement la protéine LacI, cI devrait être exprimé et par conséquent YFP ne devrait pas l'être. A l'inverse, en absence d'ITPG, LacI empêche l'expression de cI et donc YFP devrait être exprimé. Le problème qui s'est posé, après avoir construit ce petit circuit simple, c'est que, avec ou sans IPTG, les colonies n'étaient jamais fluorescentes. Le problème venait d'une petite fuite dans le promoteur lac qui permettait, même en absence d'IPTG, une expression suffisante de cI pour réprimer l'expression de YFP. Les auteurs ont alors faire évoluer la protéine cI par mutagénèse aléatoire et ont « screené » les colonies montrant le phénotype attendu : fluorescence en absence d'IPTG et absence de fluorescence en présence d'IPTG. Ils ont réussi à isoler des mutants de la protéine cI dont l'affinité pour l'ADN est plus faible ce qui compense la fuite du promoteur lac

et permet de retrouver le phénotype attendu. Les chercheurs démontrent ainsi que l'évolution dirigée peut aider à résoudre, *in vivo*, des problèmes complexes d'optimisation impliquant de nombreux paramètres. C'est donc une méthode puissante pour « tuner » finement un petit réseau (le debugger) et obtenir la propriété/la fonction désirée (Haseltine & Arnold, 2007).

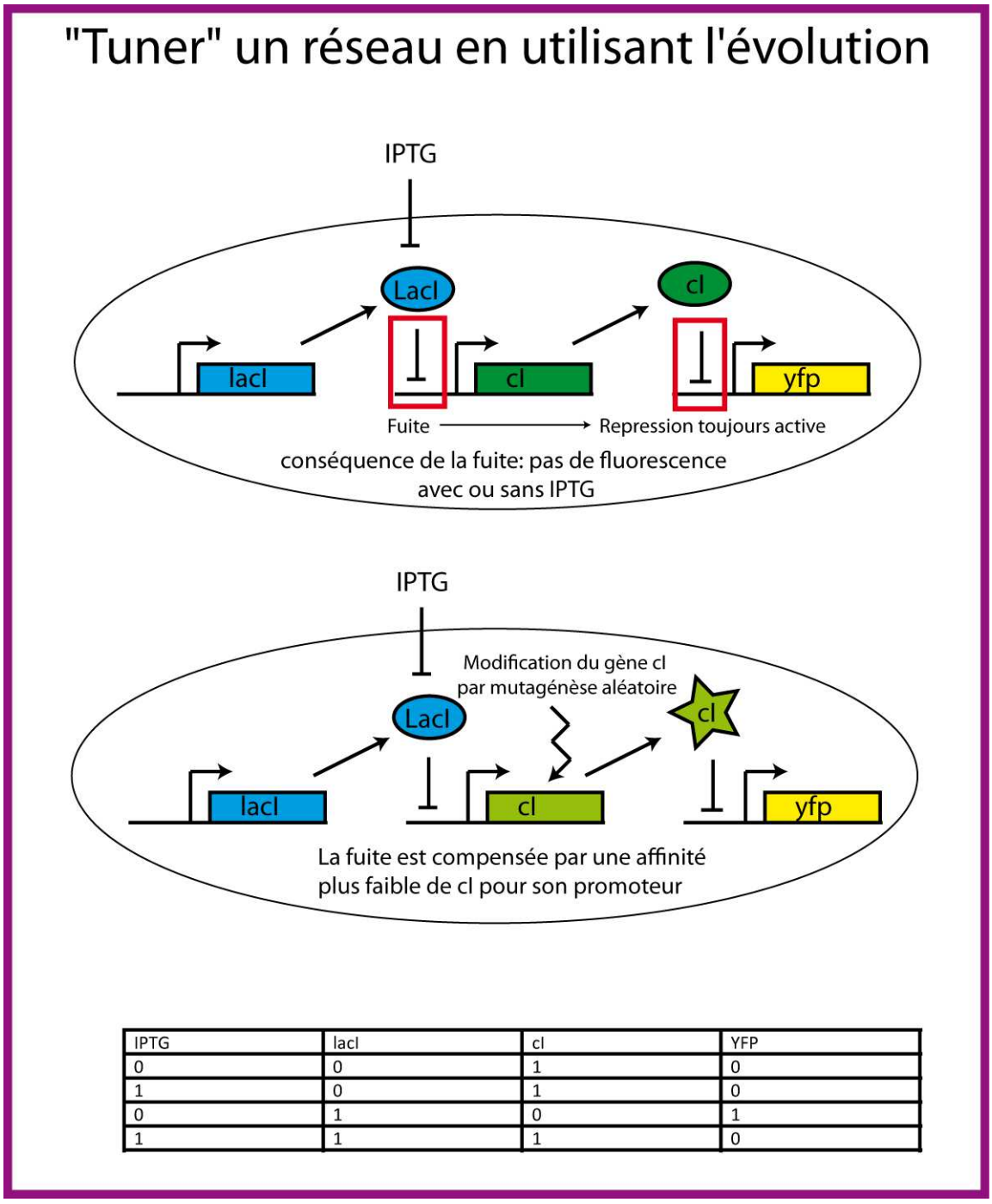
## 4.2 L'évolution peut-elle être un frein aux avancées en biologie synthétique ?

Mais une fois que l'on réussit à obtenir un réseau fonctionnel, une correspondance séquence–fonction optimale, le problème inverse se pose forcement. L'évolution ne s'arrête pas une fois que la bactérie est dans le fermenteur pour exécuter la fonction pour laquelle elle a été programmée. L'objectif de l'expérimentateur peut très vite s'opposer à l'objectif de l'évolution à savoir maximiser le fitness. La plupart des circuits synthétiques ont un coût métabolique or même quand ce coût est faible, il entraine une diminution du fitness par rapport à une souche sans circuit fonctionnel. Inexorablement, en seulement quelques jours (You *et al.*, 2004), des mutants ayant inactivé le circuit apparaissent et finissent par devenir majoritaires (Arkin & Fletcher, 2006).

Pour limiter les effets de l'évolution , on peut tenter de diminuer la fréquence d'apparition des mutations en surexprimant les protéines impliquées dans la réparation des erreurs (gène mutS, mutL) (Zhao & Winkler, 2000) mais inexorablement des mutants mieux adaptés apparaitront. Car même si l'ADN polymérase est probablement la machine de copie la plus efficace sur terre, elle ne recopiera jamais l'ADN avec une fiabilité de 100%.

Réussir à isoler un circuit synthétique des effets de l'évolution ne sert pas seulement à protéger la fonction recherchée. Il faut aussi empêcher que le circuit n'évolue dans une direction que l'on ne contrôlerait pas et/ou qu'il ne se dissémine dans l'environnement contre notre volonté. Si on crée un « calculateur » synthétique, on ne veut pas que ce calculateur se balade d'organismes en organismes via les échanges de matériels génétiques, ne serait-ce que d'un point de vue éthique. Il y a un risque d'échappement que les biologistes combattent depuis longtemps en créant des stratégies de « confinement biologique ». Il faut réussir à isoler le matériel génétique synthétique de l'environnement.

Une des sources majeures de changement évolutif dans la nature vient du transfert horizontal de gènes. Les 3 principales voies de transfert horizontal sont la conjugaison, la transduction et la transformation (décrites plus haut dans ce chapitre). Une méthode employée depuis longtemps consiste à supprimer dans la souche qui reçoit le « circuit synthétique » les gènes nécessaires à ces différents modes de transfert horizontal. Par exemple, on supprime les gènes de compétence




| IPTG | lacI | cI | YFP |
|---|---|---|---|
| 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 |

**Figure IX.4** – Rendre un réseau fonctionnel en utilisant l'évolution. Le promoteur lac a une fuite. Par conséquent, la protéine cI est exprimée à un niveau suffisant pour réprimer l'expression du gène yfp qu'il y ait de l'IPTG ou pas. En modifiant le gène cI par mutagénèse aléatoire, les auteurs ont diminué l'affinité de la protéine cI pour son promoteur ce qui compense la fuite.

(com), les gènes nécessaires à la recombinaison sur le chromosome (rec) ou bien les gènes indispensables à l'infection virale (Skerker *et al.*, 2009).

La bioremediation est l'utilisation d'organisme vivants dont la fonction est de dépolluer un écosystème après une contamination toxique. Le microorganisme doit rendre l'écosystème tel qu'il était avant contamination (par exemple, en dégradant un composé toxique). Mais comment s'assurer que l'on contrôle l'action et l'invasion du microorganisme, que celui-ci ne modifie l'environnement que dans le sens voulu, et qu'une fois sa mission accomplie, il arrête d'agir ? Le problème c'est que l'évolution est parfaitement adaptée pour trouver la faille, pour casser toutes stratégies de « confinement » que l'on pourrait implémenter pour garder la souche sous contrôle. Il faut donc superposer plusieurs stratégies indépendantes de confinement de manière à ce que la probabilité d'échappement soit quasi nulle. Voici deux stratégies qui semblent assez robustes à l'évolution tout en étant faisables. Elles ne seront cependant pas simples à implémenter :

⇒ le confinement nutritionnel consiste à rendre la souche dépendante d'une substance rare ou inconnue dans la nature pour survivre ou n'intervenant pas ou peu dans le vivant (Fluor, silice...) (Schmidt, 2009)

⇒ le confinement sémantique consiste à utiliser un code génétique alternatif (orthogonal) de manière à ce que si par hasard l'ADN est transféré dans d'autres bactéries, la traduction ne produise que des protéines non fonctionnelles (Herdewijn & Marliere, 2009).

On voit bien que dés lors que l'on implémente des fonctions avec l'ADN, il faut garder à l'esprit que ces fonctions sont susceptibles d'évoluer dans des directions non souhaitées et de se propager dans la nature. Contrôler l'évolution sera donc un passage obligé en biologie synthétique. Combattre un processus qui excelle à trouver des fuites, des failles dans la sécurité n'est pas chose aisé mais des solutions existent.

## 4.3 L'évolution de l'évolution

L'environnement ne peut pas changer *directement* (i.e sans passer par la sélection) la séquence du génome. Par exemple, si subitement la concentration du glucose dans le milieu augmente énormément, la bactérie ne peut pas détecter ce changement puis modifier spécifiquement tel nucléotide du promoteur du gène codant pour le transporteur du glucose dans le but d'augmenter fortement l'expression de ce dernier. On considère qu'il n'y a pas de lamarckisme : les mutations apparaissent toujours aléatoirement. On pense qu'il n'y a aucun biais statistique.

Mais le biologiste synthétique peut-il rendre ses lettres de noblesse à Jean-Baptiste de Lamarck ? Peut-on imaginer un système synthétique qui modifie spécifiquement l'ADN en fonction

d'un stimulus externe ?

Les facteurs de transcription reconnaissent des sites spécifiques sur l'ADN et les systèmes de correction d'erreur sont capables de détecter un mésappariement sur l'ADN et de le corriger (Li, 2008). En associant synthétiquement ces deux mécanismes, on peut imaginer créer une bactérie qui modifie son génome en fonction d'un stimulus externe.

Imaginez, par exemple, que l'on modifie le facteur de transcription AraC de telle manière à ce qu'en présence d'arabinose, il se fixe toujours en amont de l'operon araBAD mais n'active plus la transcription. A la place, il recrute la machinerie d'excision/réparation qui change un nucléotide (*ce nucléotide là*) sur le promoteur de telle manière à augmenter de manière « constitutive » l'expression de l'operon araBAD. Ce changement sera transmissible à la descendance. Evidemment, cela passe par la connaissance *a priori* dont dispose l'homme (il faut savoir quel nucléotide modifier). Mais l'exemple ci-dessus ne représente que le cas simple. La connaissance *a priori* des nucléotides à modifier spécifiquement en fonction des conditions extérieurs pourrait nous permettre de créer des réseaux extrêmement complexes chargés de « modeler » le génome en temps réel en fonction des conditions environnementales. De quoi faire se retourner Darwin dans sa tombe n'est-ce pas ?

Mais s'il semble relativement simple de faire du lamarckisme, pourquoi la nature n'a-t-elle pas choisi cette voie ?

Je vois deux arguments :

⇒ modifier quelques nucléotides en fonction du milieu extérieur semble faisable mais contrôler l'ensemble des nucléotides du génome semble beaucoup plus complexe car les mécanismes chargés de contrôler la modification du génome en fonction de l'environnement sont eux aussi codés dans le génome et donc soumis à modification. Très vite, on perd l'intuition des conséquences potentielles. Il est possible que ces régulations lamarkiennes soient tout simplement théoriquement inimplementables à grande échelle.

⇒ La nature ne dispose pas de l'information *a priori* (celle justement dont on dispose) : elle ne sait pas que la modification de tel nucléotide induira telle conséquence sur le phénotype. En fait, l'homme a une idée partielle du paysage adaptatif de la bactérie qu'il utilise pour permettre à la bactérie de faire du lamarkisme.

Selon moi, plus on aura une connaissance précise *a priori* du paysage adaptatif, plus on pourra programmer des fonctions lamarckiennes. Mais j'avoue ne pas savoir si cela a un quelconque intérêt par rapport à un simple réseau de régulation. Il faudrait y réfléchir plus longuement.

J'ai appelé, un peu pompeusement d'ailleurs, le titre de ce paragraphe « évolution de l'évo-

lution ». L'idée était que les fonctions lamarckiennes que je propose d'implémenter synthétiquement représentent une évolution de l'évolution qui passe par l'invention du cerveau humain et sa compréhension des mécanismes moléculaires à l'origine de l'évolution. Mais j'avoue ne plus trop savoir si le titre est adéquat car je suis gêné par le fait que cette évolution de l'évolution passe par une connaissance *a priori* or le principe de l'évolution consiste justement à fouiller un paysage inconnu.

# Références bibliographiques


Anderson, J. C., Clarke, E. J., Arkin, A. P., & Voigt, C. A. 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol*, **355**(4), 619–27. 216

Arkin, A. P., & Fletcher, D. A. 2006. Fast, cheap and somewhat in control. *Genome Biol*, **7**(8), 114. 217

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, **461**(7268), 1243–7. 200

Crameri, A., Whitehorn, E. A., Tate, E., & Stemmer, W. P. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol*, **14**(3), 315–9. 215

Dekel, E., & Alon, U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**(7050), 588–92. 207

Dekel, E., Mangan, S., & U., Alon. 2005. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *physical biology*. 210

Elowitz, M. B., & Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**(6767), 335–8. Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. England. 214

Gould, S. J., & Lewontin, R. C. 1979. The spandrels of San Marco and the Panglossian paradigm : a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci*, **205**(1161), 581–98. 204

Gould, S. J., & Vrba, E. 1982. Exaptation ; a missing term in the science of form. *Paleobiology*. 204

Haseltine, E. L., & Arnold, F. H. 2007. Synthetic gene circuits : design with directed evolution. *Annu Rev Biophys Biomol Struct*, **36**, 1–19. 217

Herdewijn, P., & Marliere, P. 2009. Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem Biodivers*, **6**(6), 791–808. 219

Kashtan, N., & Alon, U. 2005. Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A*, **102**(39), 13773–8. 212

Kashtan, N., Noor, E., & Alon, U. 2007. Varying environments can speed up evolution. *Proc Natl Acad Sci U S A*, **104**(34), 13711–6. 212

Lawrence, J. G., & Ochman, H. 1998. Molecular archaeology of the Escherichia coli genome. *Proc Natl Acad Sci U S A*, **95**(16), 9413–7. 201

Li, G. M. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res*, **18**(1), 85–98. 220

Maynard-Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., & Wolpert, L. 1985. Developmental constraints and evolution. *The Quarterly Review of Biology*. 204

McAdams, H. H., Srinivasan, B., & Arkin, A. P. 2004. The evolution of genetic regulatory systems in bacteria. *Nat Rev Genet*, **5**(3), 169–78. 201

Price, M. N., Dehal, P. S., & Arkin, A. P. 2008. Horizontal gene transfer and the evolution of transcriptional regulation in Escherichia coli. *Genome Biol*, **9**(1), R4. 201

Schmidt, M. 2009. Xenobiology : a new form of life as the ultimate biosafety tool. *Bioessays*, **32**(4), 322–31. 219

Shinar, G., Dekel, E., Tlusty, T., & Alon, U. 2006. Rules for biological regulation based on error minimization. *Proc Natl Acad Sci U S A*, **103**(11), 3999–4004. 211

Skerker, J. M., Lucks, J. B., & Arkin, A. P. 2009. Evolution, ecology and the engineered organism : lessons for synthetic biology. *Genome Biol*, **10**(11), 114. 219

Tagkopoulos, I., Liu, Y. C., & Tavazoie, S. 2008. Predictive behavior within microbial genetic networks. *Science*, **320**(5881), 1313–7. 212

Yokobayashi, Y., Weiss, R., & Arnold, F. H. 2002. Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A*, **99**(26), 16587–91. 216

You, L., Cox, R. S., 3rd, Weiss, R., & Arnold, F. H. 2004. Programmed population control by cell-cell communication and regulated killing. *Nature*, **428**(6985), 868–71. 217

Zhao, J., & Winkler, M. E. 2000. Reduction of GC –> TA transversion mutation by overexpression of MutS in Escherichia coli K-12. *J Bacteriol*, **182**(17), 5025–8. 217

# Chapitre X

# Evolution (II)

*« Anti-Darwin. Pour ce qui est de la fameuse lutte pour la vie, elle me semble jusqu'à présent plus souvent proclamée que prouvée. Elle peut avoir lieu mais c'est l'exception : le caractère le plus général de la vie, ce n'est nullement la pénurie, la famine, c'est plutôt la richesse, l'opulence et même l'absurde gaspillage. »*[1]

Nietzsche

## 1 Préambule

Dans le premier chapitre sur l'évolution, j'ai essayé de rester relativement « positiviste » : le lecteur y a trouvé de nombreuses citations de publications scientifiques. Dans ce chapitre, il n'y aura pas de citations de publications. Nous allons aborder des aspects plus théoriques mais également plus polémiques, spéculatifs voire métaphysiques. La première partie traite des critiques/alternatives possibles à la théorie de l'évolution. Le socle de mon argumentation sera basé sur les travaux de Stuart Kauffman. La deuxième partie de ce chapitre reprend le débat sur les unités de sélection. J'essaie d'y décrire une vision plus personnelle.

## 2 Critiques et alternatives à la théorie de l'évolution

Quand on s'apprête à débattre/discuter/attaquer/critiquer la théorie de l'évolution, le risque est grand que votre « adversaire » (un scientifique) se braque et/ou se ferme comme une huitre. Il a souvent l'impression que l'on s'apprête à aborder le débat classique qui existe entre créationnistes et évolutionnistes. Dans ce débat, les échafaudages cognitifs qui amènent à défendre l'une ou l'autre de ces doctrines/théories sont si différents que les scientifiques entrevoient im-

[1] Nietzsche, *Crépuscule des idoles*, Folio p. 66

médiatement que leurs arguments ne pourront pas convaincre. L'intérêt du débat devient alors uniquement politique.

Dans mon cas, je souhaite « discuter » de la théorie de l'évolution en utilisant, il me semble, le même échafaudage cognitif que les scientifiques. Mon but est donc bien de convaincre le lecteur scientifique à chercher à coté ou aux frontières du paradigme darwinien (et ne pas se contenter de chercher à l'intérieur). L'idée n'est pas de demontrer que la théorie de l'évolution est fausse mais plutôt de chercher si il est possible de l'englober dans quelque chose de plus large, de plus complet, de plus prédictif (comme la théorie de la gravitation de Newton a été englobée par la théorie de la relativité générale d'Einstein).

Karl Popper lui-même a mis en cause la théorie de l'évolution[2] en arguant que celle-ci n'est pas falsifiable et relève donc plus de la métaphysique :

> *« I have come to the conclusion that Darwinism is not a testable scientific theory, but a metaphysical research program »*

Puis il s'est finalement rétracté :

> *« I have changed my mind about the testability and logical status of the theory of natural selection ; and I am glad to have an opportunity to make a recantation »*

Ce qui gène avec la théorie de l'évolution c'est qu'elle semble pouvoir expliquer n'importe quelle fonction/structure ordonnée trouvée dans la nature. Or une théorie capable d'expliquer « tout et n'importe quoi » est-elle vraiment une théorie scientifique ? La réponse est oui car la théorie de l'évolution est une théorie explicative mais non prédictive. Prenons un exemple : l'apparition de la fonction « œil » peut-être expliquée avec des concepts de mutations, de sélection et de contraintes mais si je prends une séquence ADN de 4 milliards de paire de bases faite que de A (adénine), je ne peux pas prédire la liste des mutations susceptibles de conduire à la fonction « œil ».

Je pense qu'il faut donc partir de ces deux mots « explicatif » et « prédictif » et se poser les questions suivantes :

⇒ Y a-t-il des fonctions/structures ordonnées qui ne peuvent pas être expliquées par la théorie de l'évolution. Si oui lesquelles et pourquoi ? A l'inverse quelles sont les propriétés communes aux fonctions explicables par la théorie de l'évolution ?

⇒ Nous verrons dans la suite que le paysage adaptatif sur lequel les organismes évoluent est un hypercube. Quelles sont les propriétés de cette hypercube, a-t-il un structure particulière dont la connaissance nous fournirait des outils pour devenir prédictif ?

[2]Voir l'article de Frank J. Sonleitner *What Did Karl Popper Really Say About Evolution ?*.

Cette première partie sera divisée en 3 exemples issus de la théorie de Stuart Kauffman, un biologiste théoricien américain. Il a écrit un ouvrage *The origin of order* dans lequel il propose une alternative à l'idée que l'ordre que l'on observe dans les organismes vivants provient uniquement des effets de la sélection naturelle. Pour écrire cette partie, je me suis basé sur l'ouvrage *At home in the universe* qui représente une version simplifiée et très accessible de *The origin of order*. J'ai également utilisé l'ouvrage d'Hervé Zwirn *Les systèmes complexes* qui reprend en français la description du modèle NK. Mon texte est peut être une synthèse un petit peu abrupte : je renvoie donc le lecteur avide de clarté vers ces deux ouvrages.

Si vous avez l'occasion de lire l'ouvrage *At home in the universe*, vous vous rendrez compte que ce dernier est truffé de conditionnels. L'auteur se sent obligé de mettre en doute ses propres idées et parle de sa théorie comme d'une métaphore *« more of a metaphor than the start of a real theory »* [3] . Cette attitude humble reflète les doutes de l'auteur mais également, il me semble, les œufs sur lesquels il a marché en proposant sa théorie. Autrement dit « ses excuses » reflètent l'attitude dogmatique des scientifiques vis-à-vis d'une théorie de l'évolution devenue « intouchable ».

## 2.1 Le métabolisme primitif

Commençons par un petit exemple simple : une sorte de cas d'école. Imaginons un graphe (un réseau) aléatoire constitué de nœuds et de liens connectés ensemble aléatoirement (figure X.1). Au départ, on considère 1000 nœuds non connectés (l'exemple dans la figure X.1 en contient 10). Puis on choisit 2 nœuds au hasard que l'on connecte ensemble avec un lien et on cherche le cluster (nœuds connectés ensemble) le plus grand. A ce stade, le plus grand cluster contient 2 nœuds. Puis on itère ce processus : on choisit 2 nouveaux nœuds au hasard qu'on réunit par un lien et on cherche le cluster le plus grand et ainsi de suite jusqu'à ce que ce cluster contiennent l'ensemble des 1000 nœuds. Au début, lorsque que l'on augmente le nombre de nœuds, la taille du cluster stagne et reste faible. Puis lorsque le ratio liens/nœuds dépasse ½, comme par magie, un cluster géant se forme soudainement (visible pour 10 nœuds mais plus évident avec 1000) car la plupart des petits clusters déjà formés se connectent ensemble d'un coup. On parle de transition de phase. Selon Stuart Kauffman, ce type de transition pourrait avoir été à l'origine de l'apparition de la vie. Dans notre réseau, les nœuds seraient des molécules (à pouvoir catalytique) et les liens des réactions chimiques. Une fois que le ratio réactions/molécules dépasse 0.5, un vaste réseau de réactions catalysées émerge/cristallise soudainement créant une sorte d'ensemble auto-catalytique capable de se maintenir par lui-

[3] Stuart Kauffman, *At home in the universe*, p. 165.

même. Vous voyez que l'ordre présent dans ce métabolisme primitif n'est pas lié à la présence d'évolution ou de sélection mais plutôt aux propriétés des réseaux aléatoires.

> *« The intuition I want you to take away from this toy problem is simple : as the ratio of threads to buttons increases, suddenly so many buttons are connected that a vast web of buttons forms in the system. This giant component is not mysterious ; its emergence is the natural, expected property of a random graph. The analogue in the origin-of-life theory will be that when a large enough number of reactions are catalyzed in a chemical reaction system, a vast web of catalyzed reactions will suddenly crystallize. Such a web, it turns out, is almost certainly auto-catalytic-almost certainly self-sustaining, alive. »* [4]

## 2.2 Le réseau de régulation génique

Stuart Kauffman est sans doute l'un des premiers scientifiques à avoir proposé l'utilisation de réseaux booléens pour décrire la dynamique des réseaux de régulation génique. Dans un réseau booléen, les nœuds (gènes) sont reliés par des flèches (les interactions géniques). Chaque gène est soit allumé (1) soit éteint (0). Un réseau booléen opère en temps discret (c'est un type d'automate cellulaire). La valeur des gènes à un temps t représente l'état du réseau. A t+1, chaque gène adopte un nouvel état en fonction des flèches du réseau. Au fur et à mesure que le temps passe, et en fonction des états initiaux des gènes, la séquence des états converge vers un cycle attracteur. En connaissant l'ensemble des états et trajectoires que draine un attracteur, on peut tracer le bassin d'attraction correspondant.

Prenons l'exemple d'un réseau booléen à trois gènes (figure X.2). Dans notre exemple,

⇒ l'état du gène 3 au temps t+1 est donné par l'état des gènes 1 et 2 au temps t selon la fonction OR.

⇒ l'état du gène 2 au temps t+1 est donné par l'état des gènes 1 et 3 au temps t selon la fonction OR.

⇒ l'état du gène 1 au temps t+1 est donné par l'état des gènes 2 et 3 au temps t selon la fonction AND.

L'ensemble des cas possibles est présenté dans le tableau de la figure X.2. Si l'état initial du réseau est 000, il restera dans cette configuration. Si l'état initial est 010 ou 001, il oscillera indéfiniment entre ces deux configurations. Dans tous les autre cas (011, 100, 101, 110,111), le système finira par se stabiliser sur 111. Dans nos exemples, 000 et 111 sont des cycles attracteurs

[4] Stuart Kauffman, *At home in the universe*, p. 30.

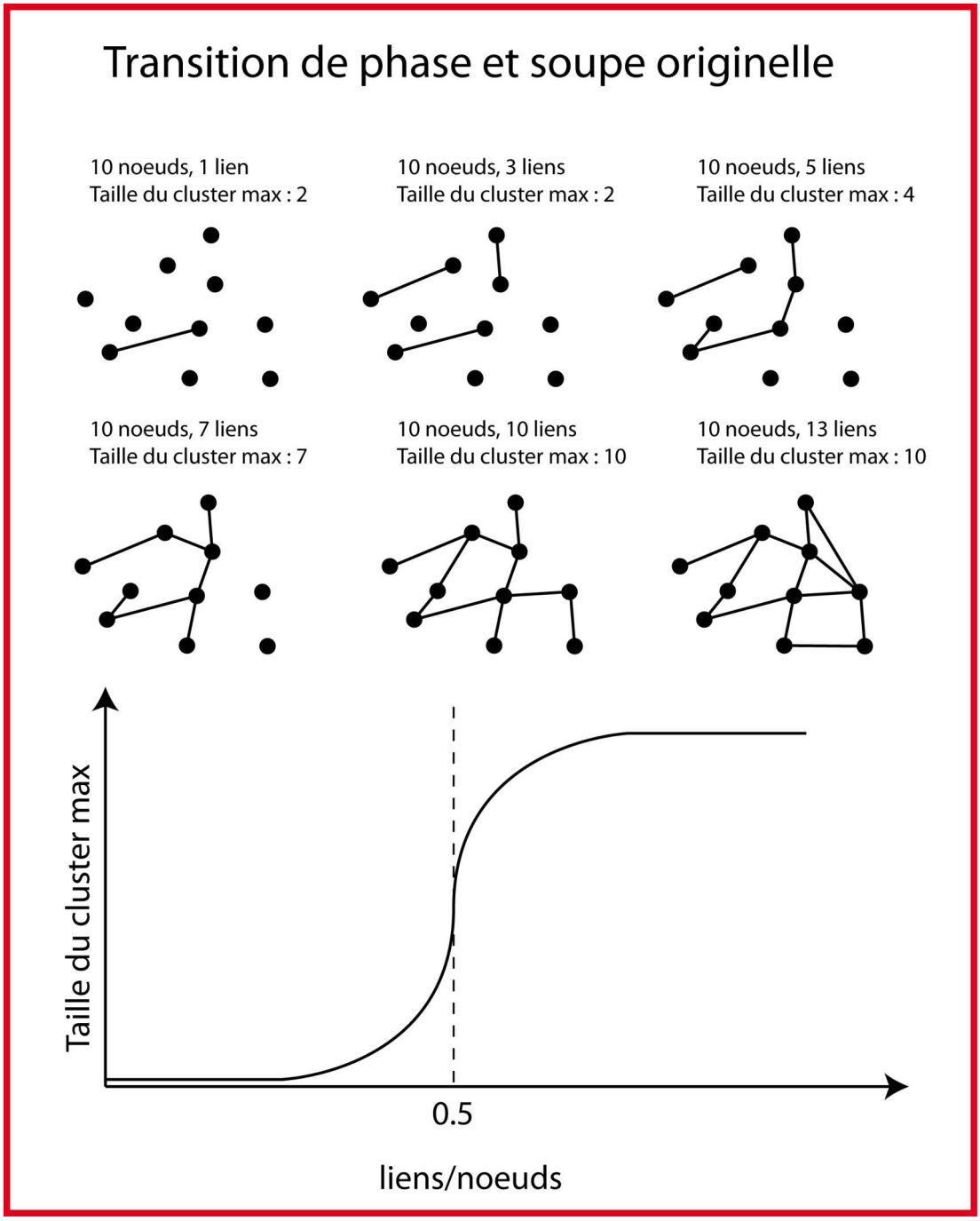



**Figure X.1** – Une transition de phase. Dans un réseau, lorsque que l'on augmente le nombre de liens de manière aléatoire, on rencontre une transition de phase pour un rapport liens/nœuds égale à ½. D'un coup, le réseau cristallise et la taille du cluster le plus grand s'accroît soudainement. Ce comportement trouve des analogies avec l'hypothèse « métabolisme first » décrivant les origines de la vie. Dans cette exemple, l'ordre qui apparait avec la naissance du métabolisme n'est pas créé par la présence de sélection naturelle. Cette figure est reproduite et modifiée à partir d'une image semblable présente dans *At home in the universe*.

de longueur 1 : une fois que le système atteint cet état, il y reste. 010/000 forme un cycle attracteur de longueur 2. On peut imaginer des cycles attracteurs passant successivement par tous les états possibles du système. Pour un réseau de 3 gènes, il y a $2^3$états possibles. Pour un réseau de n gènes, il y a $2^n$ états possibles : Si on considère le génome humain ( 30 000 gènes), il y a $2^{30000}$ configurations possibles : un nombre qui défie l'entendement (et l'ordinateur !).

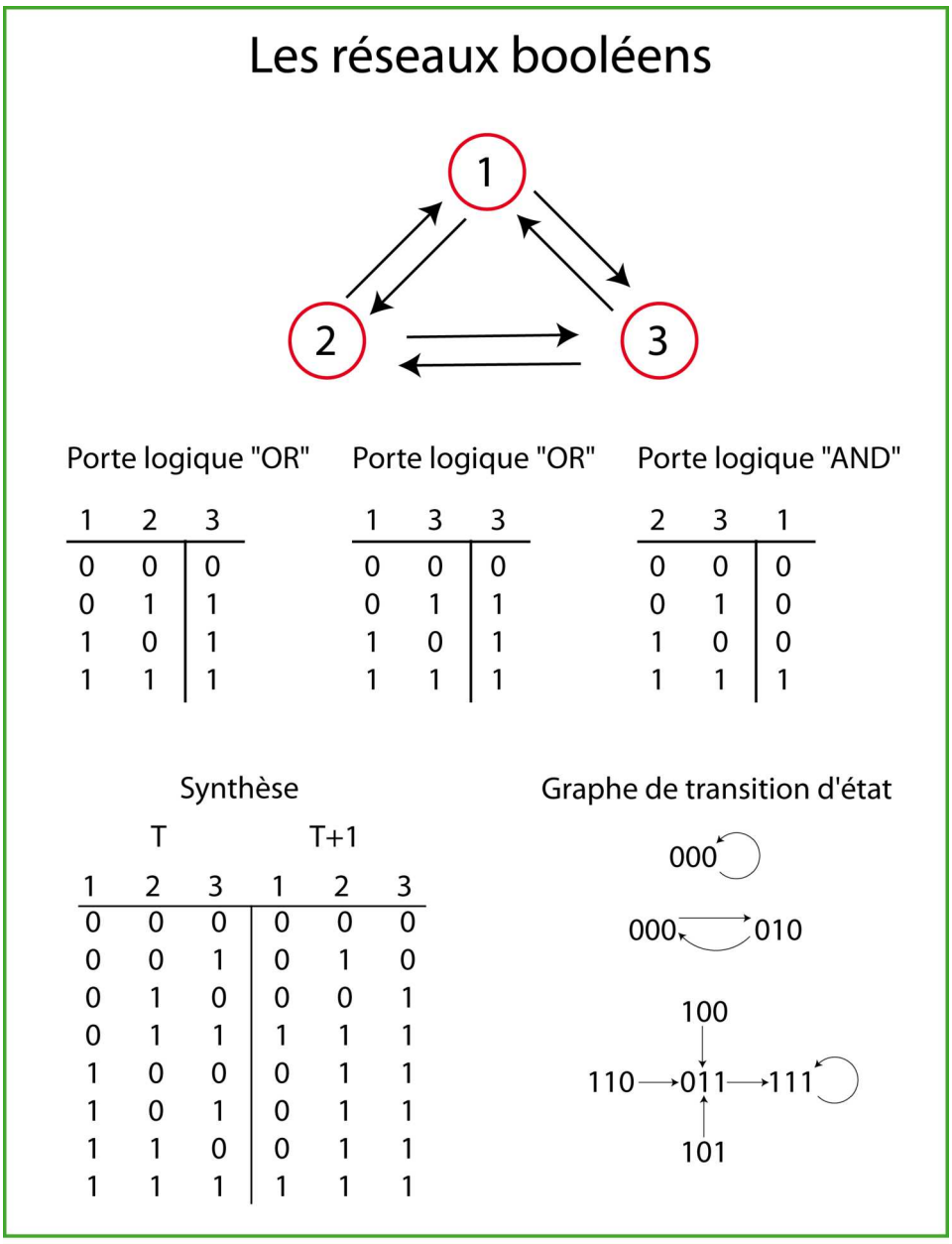



**Figure X.2** – Fonctionnement d'un réseau booléen. Cette figure est reproduite et modifiée à partir d'une image semblable présente dans *At home in the universe*.

Dans son modèle NK, Stuart Kauffman fixe le nombre N de gènes (nœuds) et fait varier le nombre moyen d'interactions (liens) K régissant l'état de chaque gène. C'est un réseau booléen aléatoire : les portes logiques ainsi que les flèches qui unissent les gènes sont choisis aléatoirement. Puis Stuart Kauffman se focalise sur les deux questions suivantes :

⇒ Combien y-a-t-ils de cycles attracteurs ?

⇒ Quelle est la longueur des cycles attracteurs ?

Lorsque K est égal à N (voir l'exemple de la figure X.3), chaque gène dépend de l'ensemble des autres gènes. Le réseau est totalement connecté. La longueur moyenne des cycles attracteurs est de l'ordre de la racine carré du nombre de configurations. Le nombre de cycles attracteurs est de l'ordre de N/e (où e est la base des logarithmes népériens et vaut environ 2.718). Si on prend N= 30 000 gènes (soit le nombre de gènes du génome humain), le nombre de cycle attracteur est de l'ordre de 10 000 et la longueur moyenne des cycles attracteurs est de l'ordre de $\sqrt{2^{30000}}$ c'est-à-dire un nombre astronomique. Cette longueur est telle qu'il est impossible de voir autre chose qu'un comportement chaotique. Le système parcourt chacun des $\sqrt{2^{30000}}$ états avant de revenir au premier état. De plus, ce type de réseau est très sensible aux conditions initiales : le simple fait de changer un 1 ou un 0 le fait basculer vers un autre cycle attracteur.

En revanche, lorsque k=2 soit deux interactions en moyenne par gène, la longueur des cycles devient égale non plus à la racine carré du nombre de configurations ($2^N$) mais à la racine carré du nombre de gènes (N). De plus, le nombre d'attracteurs est également de l'ordre de la racine carré du nombre de gènes (N). Autrement dit, le système reste confiné dans une portion minuscule de son espace des configurations. Une transition de phase dans le comportement du réseau se produit et l'ordre apparait. De plus, le réseau est très stable : en changeant un 1 ou un 0, le système reste le plus souvent dans le même bassin d'attraction. Seule quelques rares « mutations » permettent de « switcher » sur un autre attracteur.

Ainsi lorsque K est inferieur à 2, le comportement du système est ordonné et fiable alors que lorsque K est supérieur à 3, le comportement du système est instable et chaotique. Selon Stuart Kauffman, l'évolution (la vie) est contrainte de se placer à la frontière entre l'ordre et le chaos c'est-à-dire au niveau de la transition de phase qui s'opère en « tunant » K.

Pour illustrer cette idée, Stuart Kauffman nous propose une image en couleur (figure X.4). Lors du passage de l'instant t à t+1, un nœud peut changer d'état. Colorons en rouge les nœuds du réseau qui n'ont pas changé d'état et en vert les autres. Quand K est grand (>3), le réseau est chaotique : on observera un océan de vert avec quelques îlots minuscules rouges : les nœuds ne se stabilisent jamais. Si on considère une petite perturbation dans l'environnement qui viendrait modifier l'expression d'un gène (changer son état de 0 à 1), cette perturbation se

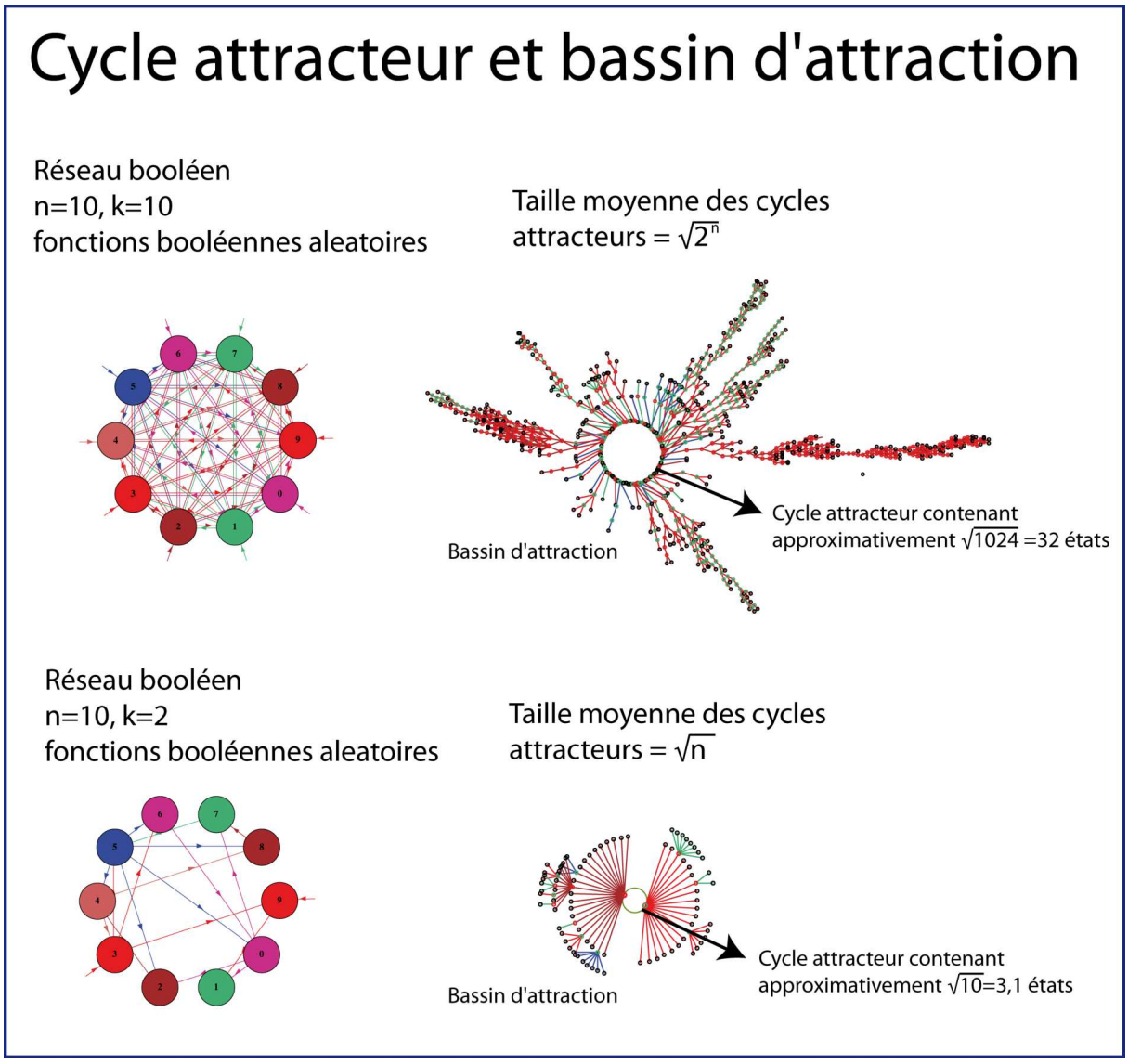



**Figure X.3** – Le modèle NK. Lorsque K=N, chaque gène est connecté à tous les autres. La taille des cycles attracteurs varie en fonction de nombre de configurations possibles. En revanche, lorsque K=2, une transition de phase se produit et la taille des cycles attracteurs varie en fonction de nombre de gènes. Les bassins d'attraction de cette figure ont été réalisés avec le logiciel Discret Dynamic Lab (DDLAB).

propagerait à l'ensemble du réseau en bouleversant totalement sa dynamique. La sensibilité du système est telle qu'une cellule ne pourrait jamais coordonner son comportement. Si on regarde maintenant un réseau très ordonné, au début tout est vert, puis au fur et à mesure que l'on converge vers un cycle, le rouge s'étend et à la fin la plupart des nœuds sont rouges. Seuls quelques îlots verts indépendants perdurent, ceux qui participent au cycle. Une perturbation dans ce système ne pourrait pas se propager car les îlots verts sont déconnectés les uns des autres. Le système est figé, la cellule ne peut pas modifier son comportement en réponse à un

changement environnemental. Enfin, à la frontière entre l'ordre et le chaos, les îlots verts sont en contact ténu les uns avec les autres dans l'océan de rouge. La plupart des perturbations seront sans conséquence sur la cellule, mais certaines permettront de « switcher » d'un îlot vert à l'autre : la cellule se coordonne, s'adapte sans pour autant bousculer toute la dynamique du système. Pour Hervé Zwirn :

> *« les systèmes adaptatifs fonctionneraient dans le régime ordonné qui se situe à la frontière de la transition entre ordre et chaos. Ils doivent en effet respecter un subtil équilibre entre stabilité et flexibilité »* [5].

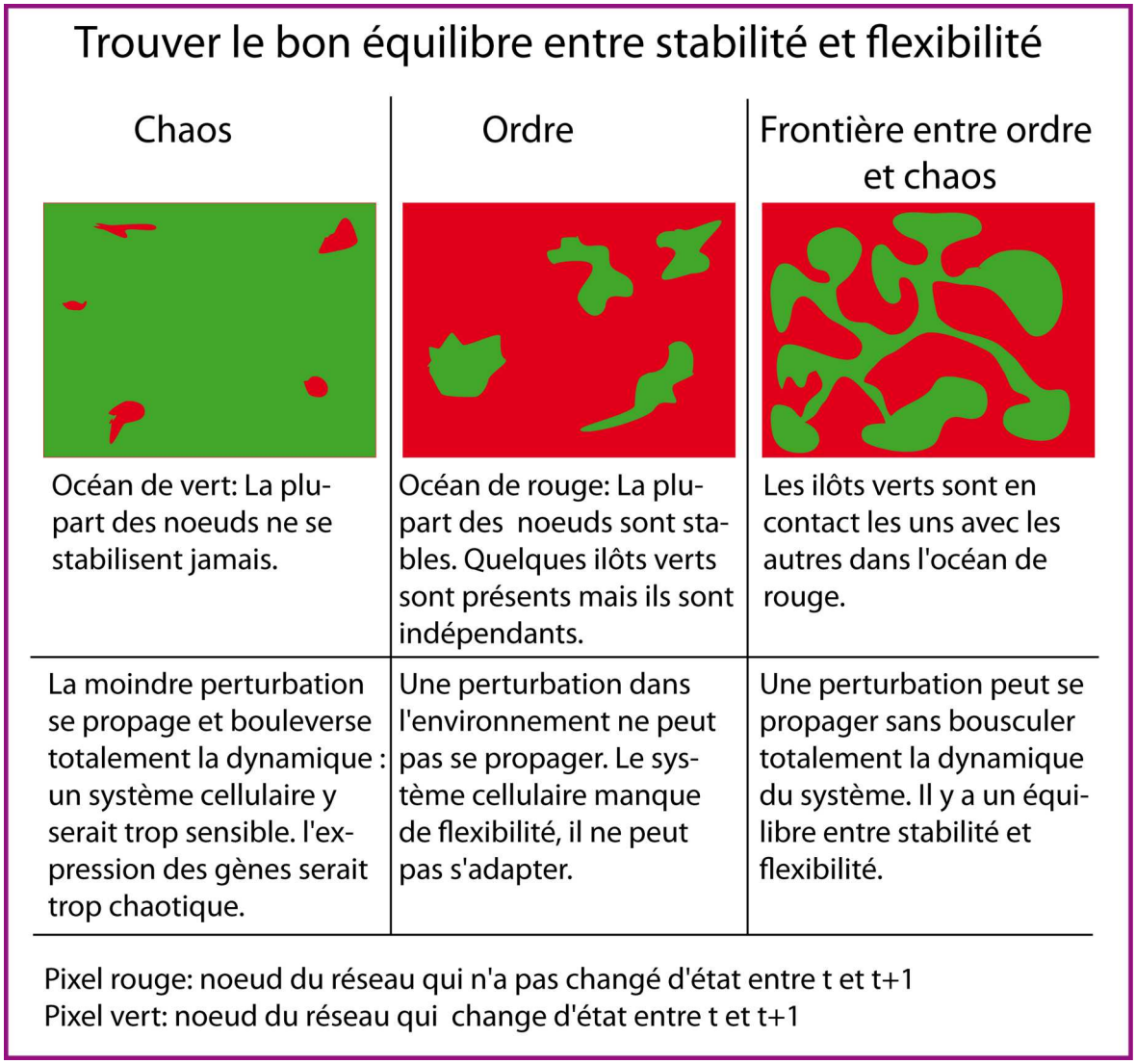



**Figure X.4** – Métaphore pour illustrer la frontière entre ordre et chaos.

[5] Hervé Zwirn, *les systèmes complexes*, Odile Jacob p. 96.

Stuart Kauffman avance plusieurs arguments biologiques en faveur de l'idée que les réseaux de régulation génique aient un K proche de 2.

⇒ Premièrement, les gènes ont souvent un petit nombre « d'input » commandant leurs expressions : seuls quelques facteurs de transcription régulent l'activité d'un promoteur.

⇒ Deuxièmement, on a vu que lorsque N=30 000 (le nombre de gènes présents dans le génome humain) et K=2, le nombre d'attracteurs est de l'ordre de la racine carré de N soit 173. Or dans un environnement donné, le nombre de type de cellules différenciées chez l'homme (équivalent des cycles attracteurs) est de 256, un nombre très proche de 173. Enfin, la longueur moyen d'un cycle attracteur est de 173 or si on considère qu'il faut 10 minutes pour activer ou réprimer un gène, le temps pour accomplir un cycle sera de l'ordre de 173x10 soit environ 30 heures ce qui est précisément la bonne échelle de temps pour le comportement d'une cellule humaine.

Ainsi pour Stuart Kauffman, l'ordre généré par un tel réseau est de l'ordre « gratuit » c'est-à-dire qui ne provient pas des cycles et de cycles de sélection naturelle. L'évolution est contrainte de « tuner » ses paramètres (par exemple K) pour être et rester à la frontière entre ordre et chaos. L'évolution est ici subordonnée à des lois relatives au fonctionnement du réseau.

> *« The reason complex systems exist on, or in the ordered regime near, the edge of chaos is because evolution takes them there.[. . .] perhaps natural selection then tunes their parameters, tweaking the dials for K and P, until they are in the ordered regime near this edge- the transitional region between order and chaos where complex behavior thrives. After all, systems capable of complex behavior have a decided survival advantage, and thus natural selection finds its role as the molder and shaper of the spontaneous order for free ; »* [6]

## 2.3 Le réseau épistatique

Nous allons maintenant étudier un deuxième type de réseau présent dans les organismes vivants : Le réseau épistatique. Vous ne comprendrez que tardivement le lien avec le sujet de cette première partie : l'alternative à la théorie de l'évolution. En effet, il me faut tout d'abord expliciter la notion de paysage adaptatif (fitness landscape). Je précise tout de suite que je vais introduire une nuance importante par rapport aux travaux de Stuart Kauffman. Les paysages adaptatifs de ce dernier ont comme unité de base le gène or je pense qu'il est préférable d'utiliser le nucléotide.

[6] Stuart Kauffman, *At home in the universe*, p. 49.

### 2.3.1 La notion de paysage adaptatif

La chose la plus importante à garder à l'esprit, c'est que la notion de paysage adaptatif (terme inventé dans les années 40) est une *image*. Il faut donc *imaginer* un paysage en 3 dimensions, fait de plaines et de montagnes. L'axe Z représente le fitness. Les axes X et Y forment une grille de tous les génomes possibles d'une certaine taille. Les sommets des montagnes représentent des hautes valeurs de fitness, c'est-à-dire de fortes capacités à se reproduire. Il faut se représenter l'évolution comme la lutte d'une population d'organismes, via les mutations et la sélection pour grimper au sommet de la plus haute montagne : le maximum global. Imaginons un génotype donné dont la valeur de fitness le place en bas d'une montagne dans le paysage adaptatif. Comment se passe la marche adaptative ? une mutation apparait au hasard dans le génome. Puis le fitness de ce nouveau génome est évalué. S'il est supérieur, on monte le long de la montagne. S'il est inferieur, on ne fait rien.

Le problème, c'est que le *vrai* paysage adaptatif ne peut pas être représenté en 3 dimensions. L'axe Z ne pose pas de problème : c'est le fitness. Mais au dessus d'un génome de 2 nucléotides, les axes X et Y ne suffisent plus à représenter toutes les possibilités de génomes possibles. Pour un génome de 2 nucléotides, pas de problème. Il y a 16 possibilités : AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, GG. Hypothétiquement (en fait, on ne peut bien sur rien faire avec un génome de 2 nucléotides), je peux synthétiser ces fragments d'ADN, les transformer dans un châssis cellulaire et mesurer pour chaque combinaison le fitness associé. Puis je dresse le paysage adaptatif comme ci-dessous : le premier nucléotide est sur l'axe X (4 possibilités : A, T, C et G) et le deuxième nucléotide est sur l'axe Y (4 possibilités : A, T, C et G). Ce paysage adaptatif ressemble bien à un paysage montagneux (figure X.5). Mais vous voyez que si je rajoute un nucléotide (génome de 3 nucléotides), j'ai maintenant un problème car je n'ai plus de dimensions à ma disposition puisque que X, Y et Z sont déjà utilisés. Au dessus de 2 nucléotides, nous sommes obligés de rajouter des dimensions qui ne sont pas représentables (par définition) en 3 dimensions. Pour un génome de n nucléotides, le paysage adaptatif correspondant possède $2^n$ dimensions. Pour les matheux, c'est un hypercube de dimension $2^n$. Pour les biologistes, le paysage adaptatif d'*E. coli* nécessiterait « seulement » 8 millions de dimensions pour être représenté alors que notre esprit dit « stop » au delà de 3 dimensions.

Mais peut-on connaitre tout de même un peu l'allure de ce paysage adaptif même si la représentation tridimensionnelle n'est qu'une image ? Et surtout, peut-on savoir à quoi il ne ressemble pas ? La réponse est oui et c'est ce que nous allons voir maintenant.

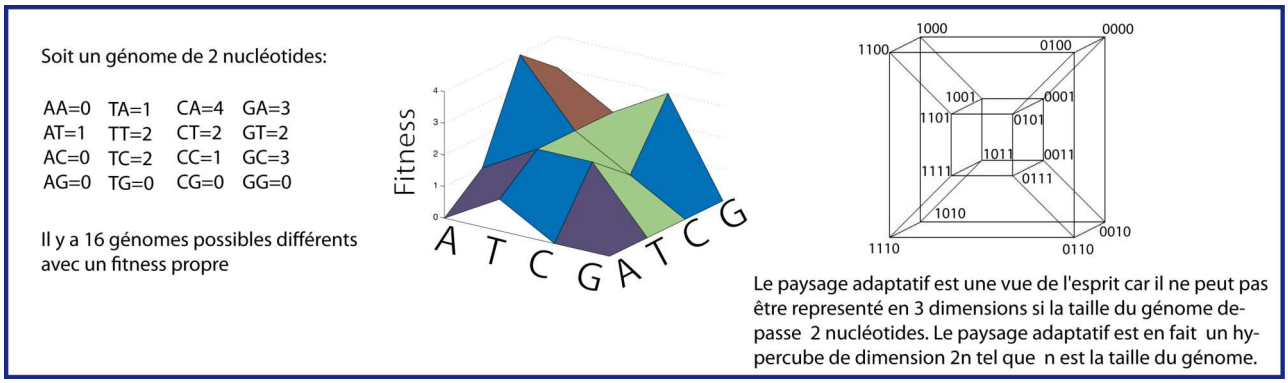



**Figure X.5** – La notion de paysage adaptatif. Seul le paysage adaptatif d'un génome de 2 nucléotides peut être représenté en 3D. Au delà, il s'agit d'un hypercube. A droite j'ai représenté un hypercube à 4 dimensions avec des « bits » (équivalent grosso modo des nucléotides). Chaque séquence de 4 bits est connectée aux 4 voisins les plus proches (ne variant que d'un bit). Par exemple 0000 est connecté à 1000, 0100, 0010 et 0001.

### 2.3.2 Le paysage adaptatif sans corrélation

Dans un paysage adaptatif sans corrélation, chaque nucléotide du génome est strictement indépendant. Imaginons un génome de 100 nucléotides indépendants. Il y a $4^{100}$ génomes possibles différents. Puis admettons un génome de 100 paires de bases composé uniquement de A (donc 100 Adénines les unes à la suite des autres). Nous décidons que ce génome est celui qui a le meilleur fitness : l'optimum global. Notez que ce choix semble être particulier alors qu'il ne l'est pas : ce génome a exactement la même probabilité d'être le plus « fit /adapté » que les ($4^{100}$-1) autres. Puis créons un génome aléatoire de 100 nucléotides. Il y aura approximativement 25% de chaque base A, T, C, G. Faisons évoluer ce génome à la recherche de l'optimum global c'est-à-dire le génome de 100 A. Comme les nucléotides sont indépendants les uns des autres, chaque A qui apparaitra dans le génome aléatoire apportera un gain de fitness et sera donc sélectionné. Ceci a pour conséquence qu'on ne s'intéresse pas aux mutations qui apparaissent dans les A car elles ne seront pas sélectionnées/fixées étant donné qu'elles diminuent le fitness. A l'inverse toute mutation qui apparait sur les bases T, C ou G du génome aléatoire a une chance sur 3 d'être un A. il faudra donc en moyenne à peine 3x100 mutations pour obtenir le génome optimal c'est-à-dire le génome le plus adapté composé que de A. Pour le génome d'*E. coli* de 4 millions de nucléotides, le nombre de combinaisons (mutations) à tester pour trouver l'optimum global serait de 12 millions. Ce chiffre peut paraître grand mais il est en réalité minuscule. Si les choses étaient aussi simples, les 4 milliards d'années d'évolution auraient largement suffit à trouver le génome optimal via 12 millions de mutations. Or *E. coli* avec son taux de croissance de 20 minutes ne possède pas un génome optimal. L'algorithme évolutif tourne encore pour amélio-

rer ce fitness. Rien n'empêche de penser qu'il existe un génome hypothétique de 4 millions de paires de bases permettant à *E. coli* de se diviser toutes les 2 minutes. Aucun des génomes présents sur terre n'a atteint l'optimum global. Or, avec un paysage adaptatif sans corrélation, le maximum global est très facile à trouver car il ne nécessite qu'un nombre de mutation très faible, de l'ordre de grandeur de la taille du génome. Ce paysage adaptatif ressemble au Mont Fuji au Japon (figure X.6 et figure X.7b).



**Figure X.6** – Image du mont Fuji au Japon qui culmine à 3776 mètres d'altitude[7].

.

### 2.3.3 Le paysage adaptatif corrélé

Une conclusion : les paysages adaptatifs des organismes ne ressemblent pas au mont Fuji. La seule solution possible pour empêcher cela consiste à supprimer la clause d'indépendance des nucléotides entre eux. Les nucléotides d'un génome ne sont pas indépendants, ils sont reliés par des interactions virtuelles. Stuart Kaufmann parle d'épistasie, c'est-à-dire les interactions virtuelles qui existent entre deux ou plusieurs gènes. Par exemple, un gène A qui permet d'importer le glucose dans une cellule a une relation d'épistasie avec un gène B qui permet d'utiliser

[7]L'image est issue de Wikipedia.

ce glucose comme énergie. Le gain de fitness apporté à la cellule n'apparait que si A et B sont présents. Le gène A sans le gène B (et réciproquement) ne sert à rien. Un fil invisible unit ces deux gènes. A l'échelle des 4000 gènes d'*E. coli*, il existe donc un réseau complexe d'interactions épistatiques qui n'a rien à voir avec le réseau de régulation génique. Vous comprenez peut être maintenant une des raisons qui m'a poussé à faire, dans ma thèse, un détour par l'évolution : pour traiter correctement la question des réseaux chez *E. coli*, je ne pouvais pas ignorer les réseaux épistatiques. Il est facile de se représenter le réseau épistatique reliant des gènes entre eux. Cependant, je le répète, l'unité de base est le nucléotide et il est donc beaucoup plus juste et rigoureux de parler d'épistasie inter-nucléotidique. Je traite de ce sujet plus en profondeur dans mon chapitre sur l'information et j'y renvoie le lecteur. L'important est de comprendre que les nucléotides « se parlent » entre eux. Donnons un exemple : dans un codon stop TAA d'un gène codant pour une protéine essentielle, le T a une relation d'épistasie avec les deux A. Si, par exemple, le dernier A était un C, le codon deviendrait TAC, ce qui n'est pas un codon stop. Par conséquent, la protéine essentielle deviendrait trop longue, potentiellement non fonctionnelle et donc le fitness chuterait que le T soit présent ou non. Ainsi le T n'apporte le fitness à la bactérie que si les deux A son présents (ou s'ils forment un autre codon stop : TGA et TAG). L'absence d'indépendance permet d'éviter d'avoir un paysage adaptatif en forme de mont Fuji ce qui serait horrible car nous serions déjà des être parfaits depuis longtemps...

Pour introduire cette communication inter-nucléotidique, je vais vous parler du modèle NK de Stuart Kauffman relatif au réseau épistatique. Un génome peut être composé de N gènes et chacun de ces N gènes peut avoir K interactions épistatiques avec les autres gènes. Dans notre cas, c'est presque exactement la même chose mais on « switch » en mode « nucléotide » : un génome peut être composé de N nucléotides et chacun de ces N nucléotides peut avoir K interactions épistatiques avec les autres nucléotides. Dans le modèle NK, un nucléotide peut avoir au minimum K=0 interaction avec les N-1 autres nucléotides. Dans ce cas, les nucléotides sont indépendants les uns des autres et le paysage adaptatif a une forme de Mont Fuji. Dans le modèle NK, un nucléotide peut avoir au maximum K=N-1 interactions avec les autres nucléotides. Dans ce cas, chaque nucléotide exerce une influence sur tous les autres. Le paysage adaptatif ressemble à un paysage totalement aléatoire (figure X.7a). Le nombre de maxima locaux devient très vite astronomique quand N croît ce qui signifie qu'une population d'organismes ne peut évoluer dans se paysage. Elle tombera sans cesse sur des maxima locaux. La marche adaptative (l'algorithme évolutif) est totalement inutile car la recherche revient à tester des génotypes au hasard. Une population ne pourrait fouiller qu'une quantité infime du paysage. La seule solution pour trouver le maximum global consiste à essayer une par une toutes les possibilités soit, pour un génome de 100 nucléotides, $4^{100}$ soit plus que le nombre d'atomes dans

l'univers connu ($10^{80}$). Le temps de Planck est la plus petite unité de temps mesurable dans le cadre de nos théories. Il est de $5x10^{-44}$ secondes. L'âge de l'univers est estimé à environ 14 milliards d'années, soit $4x10^{17}$ secondes. Si on mesurait le fitness d'un génotype à chaque temps de Planck, on aurait pu tester, depuis la création de l'univers, seulement $20x10^{61}$ génomes soit encore loin de $4^{100}$ requis pour déterminer le génome possédant le meilleur fitness.

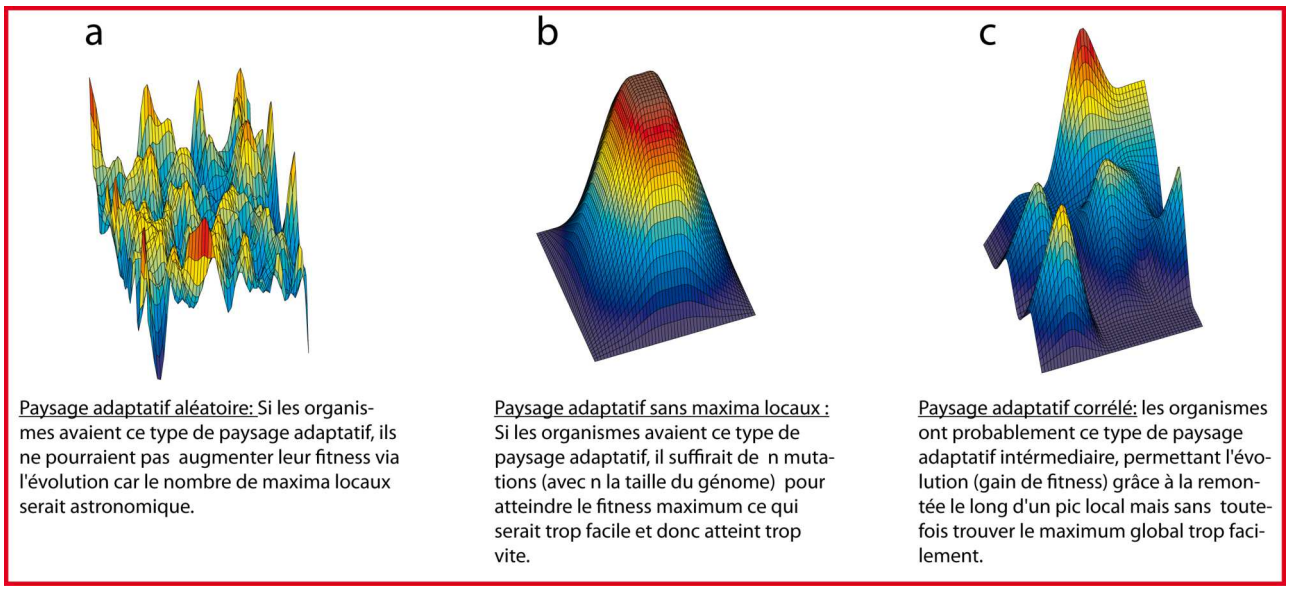



**Figure X.7** – L'allure du paysage adaptatif d'un organisme. Même si un paysage adaptatif en 3D n'est qu'une image, on sait à quoi il ne ressemble pas. Voir le code Matlab en annexe Code15PaysageAdaptatif.m qui gènère cette figure.

Les choses ne se passent donc pas de cette manière dans la nature. Les populations d'organismes arrivent à utiliser la marche adaptative ce qui signifie que le paysage adaptatif ne peut pas être entièrement aléatoire. En fait, en faisant varier K, on contrôle la rugosité du paysage adaptatif (figure X.8). Dans la nature, K doit être judicieusement choisi entre 0 et N-1 de manière à former un paysage adaptif « navigable » [8] comme celui de la figure X.7c. La rugosité du paysage adaptatif détermine son « evolvabilité ». Mais qui est à l'origine de cette evolvabilité ? est-ce-que c'est l'évolution elle-même ? L'évolution peut-elle contrôler/tuner/faire évoluer l'evolvabilité ? Ou est-ce-que l'evolvabilité est encore une fois une caractéristique intrinsèque du réseau épistatique à laquelle est subordonnée l'évolution ?

*« Those who believe that natural selection is the sole source of biological order must*

[8] Notez que cette navigabilité du paysage adaptatif est toujours embêtante car elle se base sur ce que notre esprit peut « construire » c'est dire un paysage en 3 dimensions alors qu'il nous faudrait un esprit capable de construire cognitivement en n dimensions. Cet esprit là serait une sorte de surhomme et cette question est réservée aux chapitres sur le futur auquel je renvoie mon lecteur.

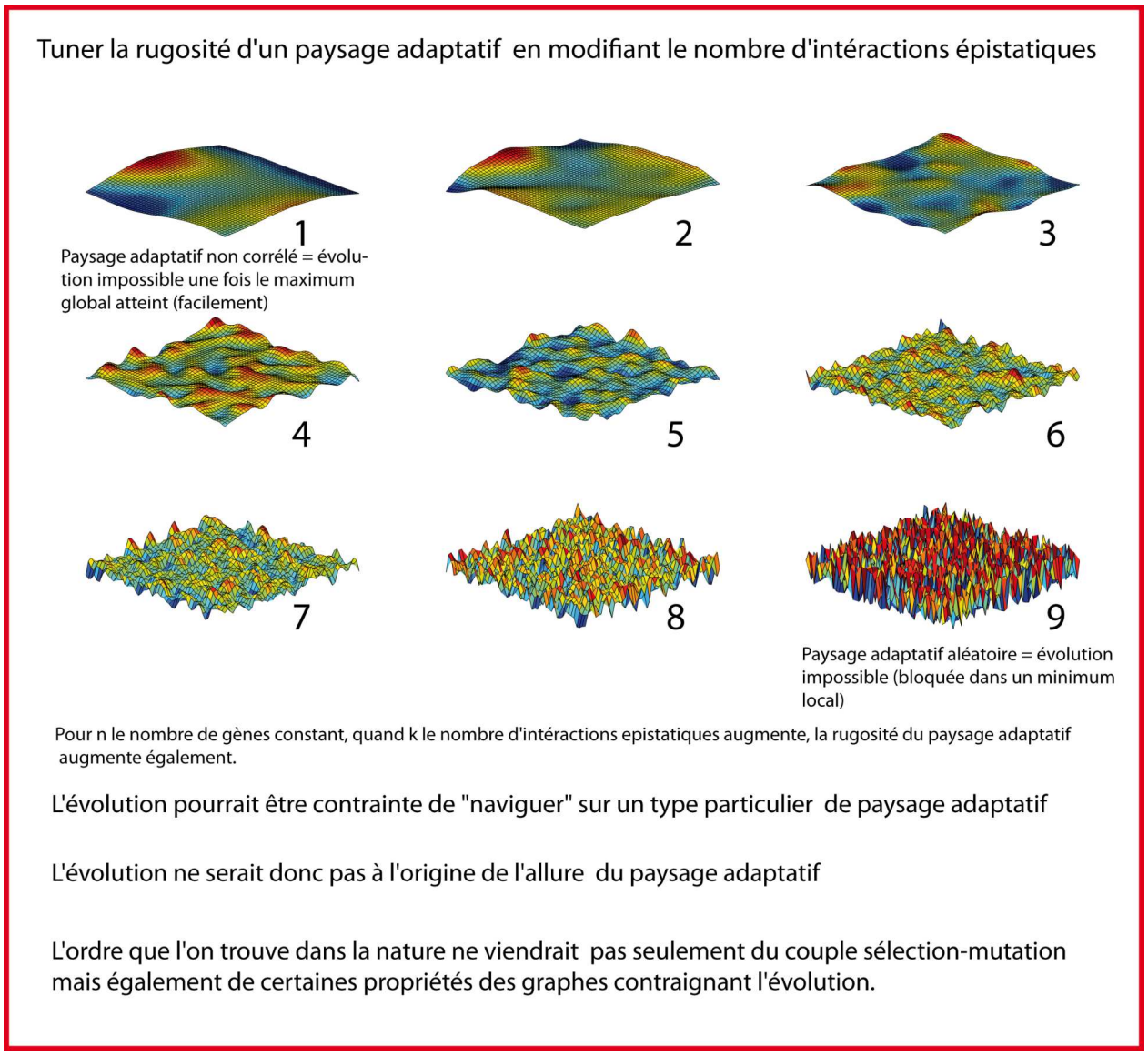



**Figure X.8** – Contrôle de la rugosité d'un paysage adaptatif en faisant varier K. L'évolution peut-elle faire évoluer le paysage adaptatif de manière à ce qu'il soit navigable ? Ici le script Matlab qui « tune » la rugosité n'est pas basé sur le modèle NK. En effet, le vrai paysage adaptatif n'est pas représentable en 3D. Voir le code Matlab en annexe Code16TuneRugositePaysageAdaptatif.m qui génère cette figure.

*assume that selection itself achieved organisms with the proper level of epistasis, the right value of K. But is selection powerful enough to craft the structure of fitness landscapes ? Or are there limits to the power of selection ? If selection is too weak to ensure evolvability, how is such evolvability achieved and maintained ? Could self-organization play a role ? Here is a profound problem that must reshape our*

*thinking. »* [9]

*« However, the very limits on selection we have discussed must raise questions about whether selection itself can achieve and sustain the kinds of organisms that adapt on the kinds of landscapes where selection works well. It is by no means obvious that selection can, of its own accord, achieve and sustain evolvability. Were cells and organisms not inherently the kinds of entities such that selection could work, how could selection gain a foothold ? After all, how could evolution itself bring evolvability into existence, pulling itself up by its own bootstraps ? And so we return to a tantalizing possibility : that self-organization is a prerequisite for evolvability, that it generates the kinds of structures that can benefit from natural selection. »* [10]

# 3 Les unités de sélection

Je vais maintenant passer à la deuxième partie de ce chapitre. Pour l'aborder, il faut avoir lu la partie précédente sur les paysages adaptatifs.

Le débat sur les unités de sélection cherche à mettre en évidence sur quel entité opère la sélection naturelle. On a pensé pendant longtemps que celle-ci s'opérait sur l'organisme jusqu'à ce que Richard Dawkins propose que la sélection naturelle s'opère sur les gènes eux-mêmes. Ces derniers seraient égoïstes c'est-à-dire qu'ils favoriseraient leur copie parfois même au détriment de la survie de l'organisme qui les contient. Les opposants à l'idée que le gène est la seule unité de sélection défendent une vision plus hiérarchique. Selon eux, la sélection naturelle peut s'opérer à différents niveaux (gènes, génomes, cellules, organismes, espèces, populations, écosystèmes...). Thomas Pradeu[11] note qu'une des clarifications les plus intéressantes provient de David Hull.

*« La clarification qui fut probablement la plus utile est due à David Hull. Ce dernier propose de faire la distinction entre deux entités biologiques impliquées dans le processus évolutionnaire : le réplicateur, qui désigne « une entité qui transmet sa structure largement intacte dans des réplications successives » (i.e. une entité qui est fidèlement copiée), et l'interacteur, qui désigne « une entité qui interagit comme un tout cohésif avec son environnement d'une manière telle que cette interaction est la cause du caractère différentiel de la réplication » (i.e. une entité sur laquelle la sélection naturelle agit directement) [...]. Hull montre qu'il est clair que les meilleurs*

[9] Stuart Kauffman, *At home in the universe*, p. 101.
[10] Stuart Kauffman, *At home in the universe*, p. 104.
[11] Thomas Pradeu, Chapitre sur la Philosophie de la biologie dans *Précis de philosophie des sciences.*

> *réplicateurs, dans l'état actuel de nos connaissances, sont les gènes (ce qui ne veut pas dire que ce sont les seuls) et donc que le véritable débat sur les « unités de sélection » concerne en réalité les seuls interacteurs. Une fois le débat clairement situé à ce niveau, la réponse de Dawkins convainc peu. [...] La réponse dominante au problème clarifié par Hull est [...] qu'il existe une hiérarchie d'interacteurs, dont le niveau le plus clairement établi est celui de l'organisme, le gène pouvant être, mais seulement parfois, un interacteur. L'organisme est en effet probablement le meilleur exemple d'interacteur, car c'est sur les traits phénotypiques de l'organisme que s'exerce principalement l'action de la sélection naturelle bien que ce dernier insiste tout autant sur l'idée que l'organisme n'est pas le seul interacteur. Dawkins a en partie reconnu ce point en développant sa thèse du « phénotype étendu ». Cependant, pour Dawkins, la véritable entité sur laquelle s'exerce la sélection naturelle est, non pas l'organisme comme tel, mais l'ensemble des traits phénotypiques sur lesquels les gènes exercent leur influence, c'est-à-dire précisément le « phénotype étendu », qui peut aller bien au-delà des frontières de l'organisme. Par exemple, dans le cas d'un parasite, le système nerveux de l'organisme parasité peut faire partie du phénotype étendu du parasite. »*

j'aimerais maintenant apporter une vision plus personnelle à cette question des unités de sélection. Je n'ai pas étudié ce débat dans le détail mais il me semble, au regard du texte ci-dessus, que le consensus actuel fait la distinction entre le replicateur et l'interacteur. Je suis en parfait accord avec cette dichotomie et je vais maintenant tenter de préciser (avec le plus de modestie possible) comment je me représente le replicateur et l'interacteur. La notion centrale que je vais proposer est le concept d'interaction à la fois inter-nucléotidique et infra-nucléotidique. Certains des paragraphes qui suivent risquent de mettre à rude épreuve les capacités d'abstraction du lecteur. Je reconnais également que la clarté de mon texte n'est peut-être pas toujours optimale cependant, à défaut de garantir la vérité, je garantie l'existence d'un sens.

J'évacue tout de suite le problème du replicateur en précisant, bien que cela soit évident et je ne pense pas qu'il y ait matière à débat sur ce point, que l'unité de sélection du réplicateur n'est pas le gène comme noté dans le texte ci-dessus mais plutôt le nucléotide répliqué. En effet, le nucléotide est l'unité de base de la séquence ADN et correspond à 2 bits car il y a 4 possibilités : A, T, C et G (cela pourrait être codé avec 2 bits : 00, 01, 10, 11).

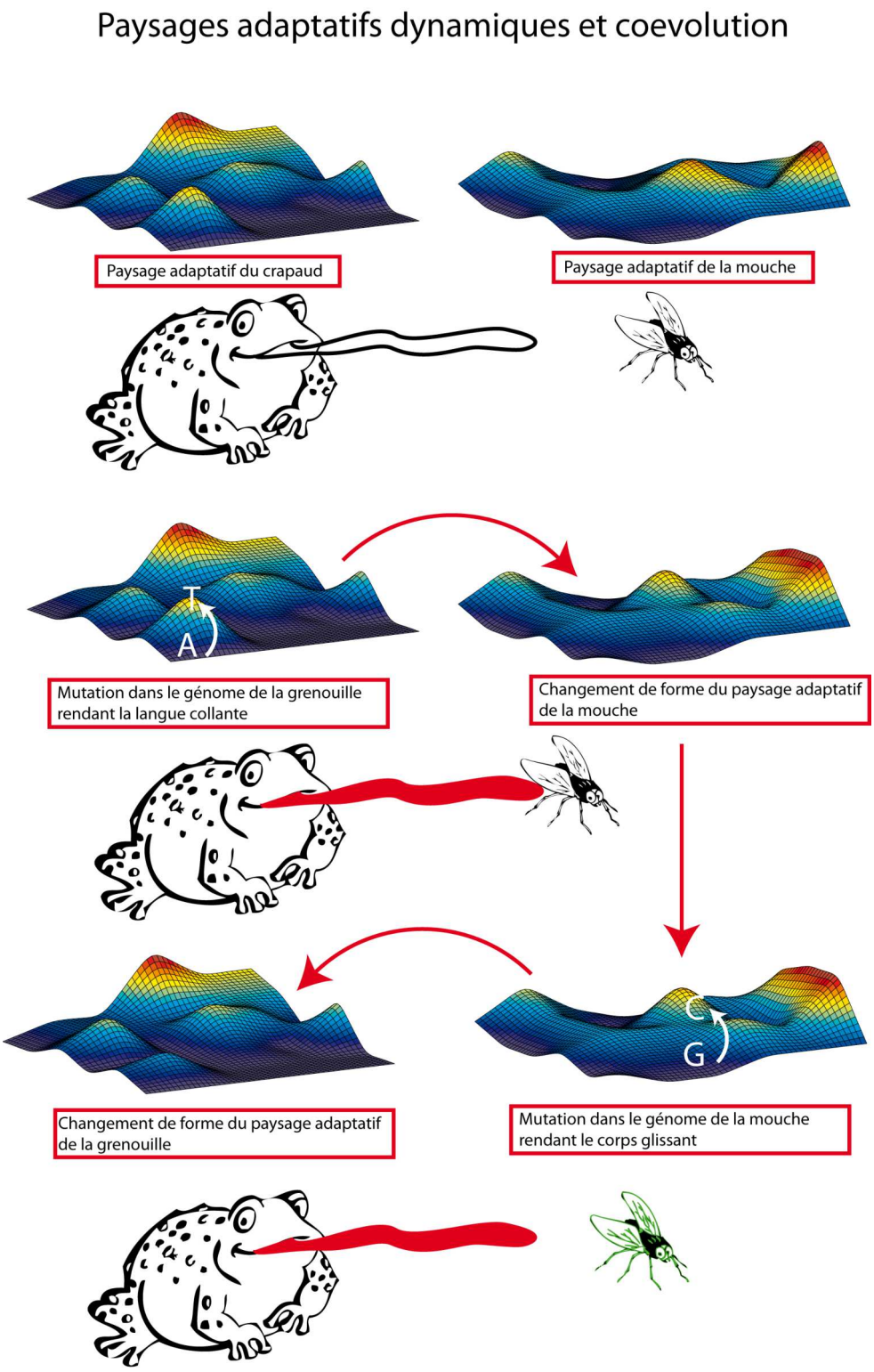



**Figure X.9** – Coévolution et paysage adaptatif dynamique.

## 3.1 La coévolution

Dans la partie précédente, nous avons vu que les nucléotides sont connectés entre eux par un réseau épistatique. Mais ces interactions épistatiques inter-nucléotides ne sont pas limitées à l'organisme. De telles interactions peuvent exister entre des organismes différents. Prenons l'exemple connu (et que j'emprunte encore à Stuart Kauffman) du crapaud et de la mouche (Figure X.9). Un crapaud a un certain génome qui forme un paysage adaptatif. Et il en va de même pour la mouche. Imaginons maintenant qu'une mutation apparaisse dans le génome du crapaud : par exemple un A devient un T. Cette mutation confère de meilleures propriétés collantes à la langue des crapauds. Le crapaud qui a la mutation attrape statistiquement les mouches plus souvent. Il a un meilleur fitness et transmet donc mieux son patrimoine génétique. Petit à petit, la population de crapauds avec la mutation grossit ce qui introduit des contraintes sur la population de mouches. Le paysage adaptatif des mouches se déforme, il y a une diminution du fitness de la population des mouches à cause de la mutation chez les crapauds. Des mouches mutantes (par exemple un G qui devient un C) avec un corps glissant sont sélectionnées puis deviennent majoritaires. Cet événement change le paysage adaptatif des crapauds en diminuant le gain de fitness associé à la mutation A->T. Ces phénomènes de coévolution ont parfois des conséquences étranges comme le « Red Queen Effect » : les organismes sont constamment obligés de fixer de nouvelles mutations pour maintenir le fitness constant. C'est exactement ce qui c'est passé ici avec les crapauds et les mouches : au final, une fois les deux mutations apparues, aucune des deux espèces n'a obtenu un gain de fitness.

Mais ce qui m'intéresse surtout, c'est que le A de départ des crapauds possède une relation virtuelle épistatique avec le G de départ des mouches. Cette relation épistatique virtuelle est, il me semble, un objet mathématique. Comme je ne sais pas trop de quel objet il s'agit, je propose en attendant mieux, qu'il s'agisse d'une fonction (voir chapitre sur l'information). Cette fonction decrirait l'interaction d'un nucléotide avec un ou plusieurs autres et prendrait un certain nombre de paramètres en argument. L'interaction virtuelle épistatique entre le A et le T a son pendant physico-chimique c'est-à-dire le contact physique entre les atomes de langues et les atomes de mouches. Ma conjecture est la suivante : toute interaction nucléotidique épistatique (décrite par une fonction difficile à conceptualiser) possède, à un moment ou à un autre, une réalité physique.

La difficulté réside dans le fait qu'il faut parfois passer *par le milieu* pour détecter l'expression d'interaction inter-nucléotidique. Prenons un autre exemple : les nucléotides du génome des arbres génèrent indirectement de l'oxygène qui compose notre atmosphère. Cet oxygène, nous devons le respirer pour vivre. Il se fixe physiquement à l'hémoglobine de nos érythrocytes

(globule rouge) et permet la respiration. Il y a interaction virtuelle entre les nucléotides des arbres et nos nucléotides.

Ce que je décris a des liens évidents avec le concept de phénotype étendu de Richard Dawkins. Ce dernier argumente le fait que nous, biologistes, restreignons de manière arbitraire l'idée de phénotype en nous limitant à l'expression phénotypique des gènes d'un organisme sur son propre corps. Par exemple, selon Richard Dawkins, le nid d'un oiseau ou les monticules de terre des termites sont des phénotypes « étendus ».

Cependant, les choses que je présente sont différentes. Je le répète : je passe d'une vue gène-centrée à une vue nucléotide-centrée. Car la notion de gène fait ressortir, selon moi, certains aspects anthropocentriques de la nature humaine. Nous, humains, « trouvons facile » de manipuler, de séparer des entités physiques distinctes. Ce marteau est différent de ce stylo. Cette enzyme A n'est pas cette enzyme B. Notre manière de comprendre une cellule consiste d'abord à comprendre ses entités et ce qui caractérise globalement une entité c'est le fait que ses atomes sont unis par des liaisons covalentes. Avec le nucléotide, c'est différent : la notion d'entité (et de liaison covalente) disparait. On baigne enfin dans un régime informatif strict car la correspondance nucléotide/bit est totale. En supprimant le gène, on supprime l'anthropocentrisme « entité ». Ainsi si le concept « phénotype étendu » casse la barrière mentale de l'organisme (les gènes de l'oiseau génèrent un phénotype : le nid). Les interactions épistatiques inter-nucléotidiques inter-organismes permettent d'aller plus loin en s'affranchissant, non seulement de l'anthropocentrisme « organisme » mais aussi de l'anthropocentrisme « entité ». Non : le gène ne représente pas une unité de sélection ni l'organisme d'ailleurs. L'unité de sélection de type « réplicateur » c'est le nucléotide. L'unité de sélection de type « interacteur » c'est une fonction décrivant une interaction entre nucléotides ayant une réalité physique que nous pouvons percevoir à différents niveaux (moléculaire, cellulaire, organisme, écosystèmes). La présence de la réalité physique, lors de l'exécution du programme génétique, permet la sélection naturelle. C'est donc cette fonction (ou une partie de cette fonction) qui est sélectionnée, qui est égoïste. En se débarrassant de « l'entité », on peut alors concevoir l'idée que cette fonction est un concept de niveau encore plus bas que le nucléotide car ce dernier est une unité physique de base « grossière » chargé d'encoder et de compresser l'information. Cette fonction est de plus « bas niveau » non pas sur le plan physico-chimique (en dessous du nucléotide, il y a les atomes et cela ne nous intéresse pas) : elle est de plus bas niveau sur le plan informatif. Dans un nucléotide, on peut compresser une grande quantité d'information, une sorte de pack d'interactions. Une seule de ces interactions est peut-être primordiale, sélectionnée, égoïste et les autres voyageant avec elle de manière opportuniste. L'unité de sélection de type « interacteur » est une fonction représentant une interaction inter-nucléotidique parmi un pack d'interactions.

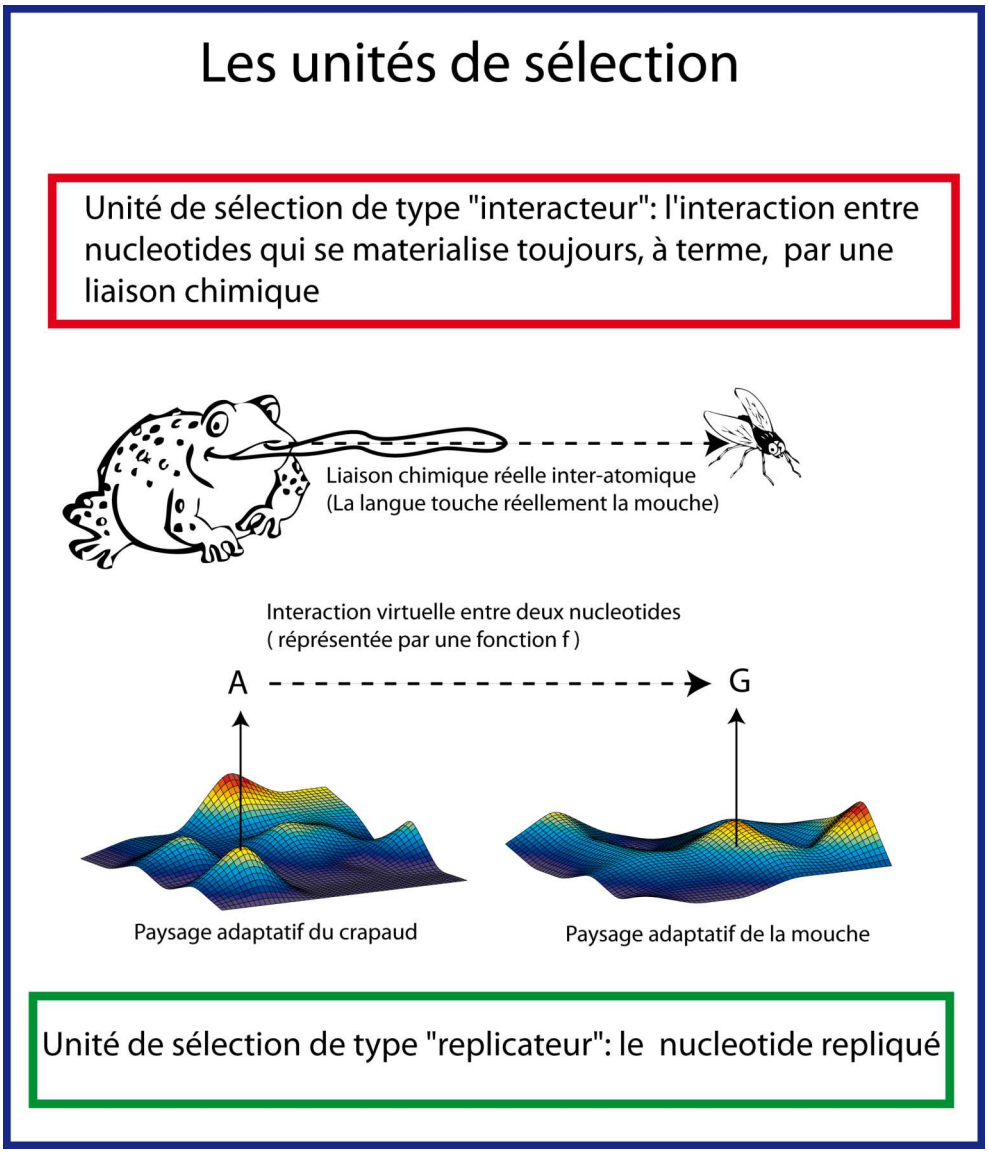



**Figure X.10** – Réplicateur et interacteur. L'unité de sélection de type « réplicateur » est le nucléotide répliqué. Cette unité de base encode l'information de manière grossière. En effet, un ou plusieurs nucléotides peuvent être reliés par un pack d'interactions épistatiques. On peut décrire une de ces interactions par une fonction. C'est cette interaction/fonction qui représente l'unité de type « interacteur ». Cette fonction a toujours une réalité physique sur laquelle s'exerce la sélection naturelle. Nous percevons cette réalité physique à différentes échelles : du microscopique ou macroscopique.

Le tout est bien caché derrière/dans l'unité de type « réplicateur » : le nucléotide.

A partir de là, il faut se pencher à nouveau sur le débat « réductionnisme/holisme ». Pour faire simple, l'holisme affirme que la nature crée des ensembles qui sont plus que la somme de leurs parties. A l'inverse, le réductionnisme affirme que, pour comprendre un système, il faut le décomposer en éléments simples et donc que l'on peut expliquer le tout à l'aide des parties. On

comprend bien qu'il est difficile d'expliquer le fonctionnement d'un écosystème sans décrire le fonctionnement des organismes qui le composent. De même, on peut difficilement expliquer le fonctionnement d'un organisme sans décrire le fonctionnement des cellules. Et ainsi de suite : Ecosystème—organisme—cellule—molecule. Et le problème apparaît clairement lorsqu'apparait le « dernier »niveau : les atomes (j'exclue les particules élémentaires pour faire simple) car le retour dans le sens inverse semble alors impossible. En effet, une soupe d'atomes C, H, O, N, P, S, dont on connait parfaitement les propriétés, ne semble pas nous permettre de remonter à l'organisme ou à l'écosystème.

Dans ce débat, il me semble que le problème vient de la dimension « entité-centrée »alors qu'il faut réfléchir de manière « information-centrée ». Avec cette manière de penser, il y a une première dichotomie qui n'est pas explicable en l'état actuel des connaissances :

⇒ Un monde sans information (ou très pauvre en information) avec une soupe d'atomes et de molécules simples (par exemple : ce qu'on trouve sur une planète *a priori* sans vie comme la planète Mars).

⇒ Un monde avec information comme la biosphère sur la planète Terre. Dans ce monde avec information, on peut trouver différent niveaux à envisager plutôt sous forme de boucle :

- L'unité de base du réplicateur : le nucléotide répliquée
- La séquence informative proprement dite (l'ADN par exemple)
- L'exécution du programme (le programme c'est la séquence ADN). Cela nécessite l'ADN avec un milieu (ce dernier incluant un châssis cellulaire, les lois physiques, le temps...) capable d'interpréter le programme.
- Mais en réalité, en dessous de l'unité de base du réplicateur (le nucléotide répliquée), on trouve l'interaction infra-nucléotidique mais celle-ci n'a de sens que lors de l'exécution du programme ADN ce qui permet de boucler la boucle ((Figure X.11)).

Avec cette vision, la question consistant à savoir si les atomes sont suffisants pour tout expliquer ou si l'émergence est une condition nécessaire disparait. La seule chose qui est nécessaire, pour expliquer tous les niveaux, c'est l'information. Et pour avoir de l'information, il faut, je le crois, disposer de séquences informatives un peu particulières. Mais je laisse mon lecteur se diriger vers le chapitre sur la vie s'il veut en savoir plus.

Si on casse la barrière entre les organismes, on peut s'amuser à faire une petite expérience de pensée. On peut concaténer tous les génomes de toutes les espèces vivantes sur terre (je n'intègre pas tous les mutants pour faire simple). Partons avec une fourchette large : 100 millions d'espèces. Soyons large également sur la taille moyenne d'un génome : 1 milliard de nucléotides. On a donc une séquence ADN de $10^{17}$ nucléotides de long représentant le génome

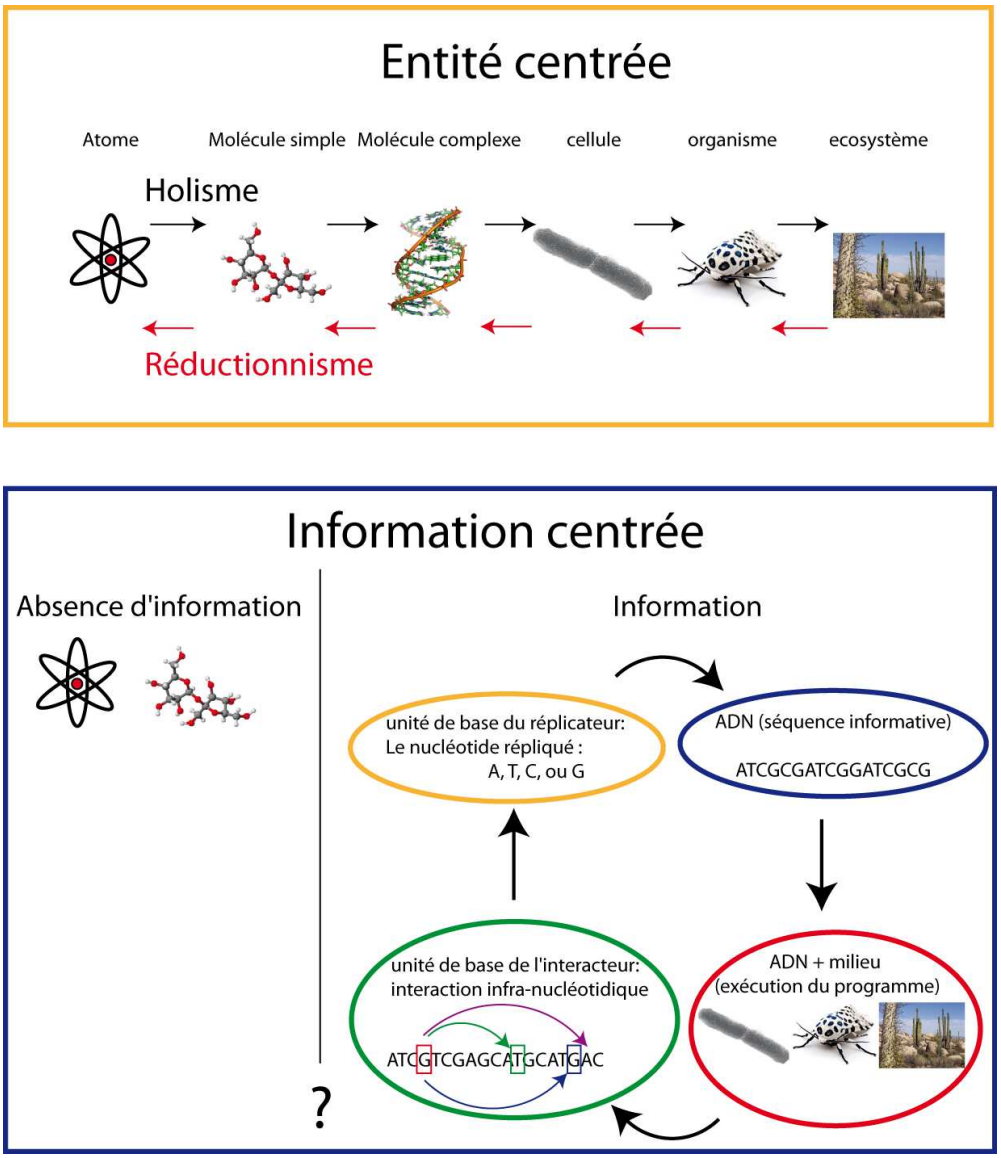



**Figure X.11** – Entité centrée vs information centrée

de la biosphère. Cette séquence forme un réseau gigantesque d'interaction inter-nucléotidique épistatique. Reposons nous maintenant la question pourquoi gènes et organismes nous semblent tant adaptés comme unité de sélection de base ? Car dans cet immense réseau épistatique de la biosphère, les gènes et organismes représentent des points chauds à forte densité d'interaction inter-nucléotidique. Le gène tout d'abord car on comprend bien que les interactions épistatiques inter-nucléotidiques d'un gène sont très nombreuses et très informatives. Autrement dit, la densité d'interactions inter-nucléotidiques au niveau d'un gène semble supérieure à celle d'une région non codante. Enfin l'organisme car il représente aussi peut être un point chaud très

dense en interactions épistatiques. En effet, le nombre d'interactions épistatiques unissant les nucléotides d'un même génome semble supérieur au nombre d'interactions épistatiques unissant les nucléotides de génome/organisme différents. La nature pourrait ainsi trouver plus facile d'injecter de l'information en utilisant les interactions entre nucléotides d'un même génome.

Je « switch » en mode métaphysique pour anticiper sur les chapitres suivants. La vie sur terre pourrait se caractériser par cet immense réseau épistatique entre nucléotides. Imaginez maintenant le paysage adaptatif de ce super génome de $10^{17}$ nucléotides. Et posez-vous la question suivante : que mettez-vous sur l'axe Z ? Si vous mettez quelque chose, vous admettez qu'il y a un but, une finalité à la vie, à la biosphère, une téléologie, une téléonomie, une fonction qui est optimisée.



**Figure X.12** – La volonté de puissance de la biosphère. Remarquez la forêt de sapin (très dense en information) qui tente d'envahir la matière minérale (pauvre en information). Des conquêtes de la biosphère encore plus difficiles (extra-terrestres) seront réalisées par un élément caché dans la photo : le photographe c'est-à-dire l'homme et son cerveau qui représentent sans doute la faille qu'a trouvé l'algorithme évolutif pour permettre à la vie de s'échapper de la planète terre[12].

[12]Photo Wikipedia du lac Morraine dans le parc national de Banff prés de Calgary au Canada.

Ce que je mettrai bien sur l'axe Z, c'est la volonté de puissance[13] de la biosphère. La biosphère n'a pas la volonté de vivre mais de s'étendre, de gagner en puissance. Regardez jusqu'où réussit à se loger la vie : dans les fonds marins, dans les geysers... La vie mène un combat contre le monde minéral. Elle cherche, via son algorithme évolutif, à envahir les zones qui lui résistent. La biosphère est, en l'état actuel de nos connaissances, limitée à la surface de la croûte terrestre. Nul doute qu'elle essaie par tous les moyens de gagner en profondeur. Mais il y aussi la conquête de l'univers que la vie voudrait réussir. Le cerveau humain sera-t-il l'outil qui permettra à la vie de réussir cette conquête ? Ainsi la vie pourrait chercher à coloniser la matière minérale pour l'informatiser. Les atomes dans l'univers sont trop indépendants (sur le plan informatif) les uns des autres. La vie pourrait chercher à accroitre toujours plus la quantité d'information injectée dans les atomes en créant un immense réseau épistatique interatomique (ou inter-particulaire si vous préférerez). Un vulgaire caillou contient peu d'informations par rapport à un organisme vivant contenant le même nombre d'atomes. Le but de la vie ? Accroitre sans cesse la quantité d'informations (et les capacités de calcul) présentes dans les $10^{80}$ atomes de l'univers connu. Ce chapitre est maintenant terminé. Ce dernier paragraphe, très « métaphysique » et « spéculatif », offre au lecteur une excellente transition vers les chapitres sur la vie, l'information ou le futur.

[13] Concept de Friedrich Nietzsche.

# Chapitre XI

# Informatique et biologie

## 1 Préambule

Dans ce chapitre, nous allons étudier les liens existant entre informatique et biologie. Plus spécifiquement, je me focaliserai sur les analogies et différences entre cellule et ordinateur. Un point important qu'avait remarqué William Ross Ashby c'est, qu'avant l'arrivée des ordinateurs, il n'y avait pas de système de complexité moyenne. Un trou énorme existait entre la complexité d'un système vivant (un arbre par exemple) et le reste du monde physico-chimique. Ce trou a été comblé par l'arrivée de l'informatique qui a offert à la fois de nouveaux outils/ un nouveau système complexe à étudier pour comprendre les règles de fonctionnement de systèmes encore plus complexes : les être vivants.

## 2 Des nouvelles disciplines : l'intelligence artificielle et la vie artificielle

A la différence d'un automate comme le canard de Vaucanson, qui fonctionne selon une succession d'opérations préétablies, un robot peut modifier dynamiquement son comportement. Un organe sensoriel (capteur), en donnant une information sur l'environnement extérieur, peut modifier l'activité d'un organe moteur par l'intermédiaire d'une architecture de contrôle (pseudo système nerveux).

Les premiers robots qui imitèrent la vie d'une manière vraiment impressionnante (dans les années 50) furent les tortues cybernétiques du neurophysiologiste William Grey Walter[1]. Ce dernier appela ses deux premières tortues Elmer et Elsie, acronymes de « Electro MEchanical Robots Light Sensitive ». Leur mécanisme interne comprenait un œil photoélectrique et

[1]Mes sources pour ce paragraphe sont : l'ouvrage de Jean-Claude Heudin, *Créatures artificielles*, Odile Jacob p. 170 ainsi que Wikipedia : Tortues de Bristol.

deux amplificateurs à tube commandant des relais qui activaient des moteurs de direction et de progression. Les deux tortues donnaient l'impression d'explorer leur environnement. Elles repéraient les sources de lumière de faible intensité dans les environs et tournaient autour. Elles pouvaient contourner les obstacles et éviter les sources de lumière plus vives qui avaient pour effet de les « brûler ». Mais lorsque que leurs batteries commençaient à faiblir, elles y puisaient en fait leur énergie. Si la lumière était fixée sur les créatures elles-mêmes, elles montraient alors des mouvements encore plus complexes d'attraction et de répulsion qui évoquaient certains comportements sociaux de défense de territoire ou de reproduction. Ces robots, en ne répétant jamais exactement la même action, montraient une sorte d'indétermination et de spontanéité rappelant le schéma général d'action d'un animal. Ce comportement émergent qui ressemble à la vie est une forme précoce de ce qu'on appelle maintenant « la vie artificielle ».

La « vie artificielle » est un champ de recherche interdisciplinaire alliant informatique et biologie dont l'objectif est de créer des systèmes artificiels s'inspirant des systèmes vivants, soit sous la forme de programmes informatiques, soit sous la forme de robots[2]. C'est la discipline « sœur jumelle » de l'intelligence artificielle. Je développe ici quelques exemples/résultats obtenus dans ces disciplines.

Le robot Kismet, développé dans les années 90 au MIT, possède un système auditif, visuel et expressif lui permettant de participer à des interactions humaines et de démontrer en même temps sa capacité à « simuler » des émotions (Breazeal, 1998). Quand un visiteur se met devant, l'expression du robot change en une expression très convaincante d'intérêt et de plaisir. Ce qui arrive après dépend bien sûr de l'attitude du visiteur : si celui-ci approche ses mains vers le robot, ce dernier semble agacé. Si le visiteur montre des couleurs vives, le robot sourit. De nombreux visiteurs rapportent qu'il est très difficile de détourner le regard quand Kismet vous fixe dans les yeux. Et il est quasiment impossible de ne pas anthropomorphiser la machine.

Il existe également des programmes de simulation de « vie artificielle » essayant de reproduire le plus fidèlement possible certaines caractéristiques de la vie (évolution, autoreproduction, adaptation à l'environnement...). « Framsticks », par exemple, est un simulateur tridimensionnelle de vie artificielle (Komosinski & Ulatowski, 2009). Les organismes consistent en des structures physiques (des sortes de bâtonnets) et des structures de contrôles (pour interagir avec l'environnement ; sorte de cerveau). Ils évoluent au cours du temps en fonction du paysage adaptatif défini par l'utilisateur (la fonction à optimiser étant par exemple la vitesse) mais peuvent aussi coévoluer spontanément dans un environnement complexe.

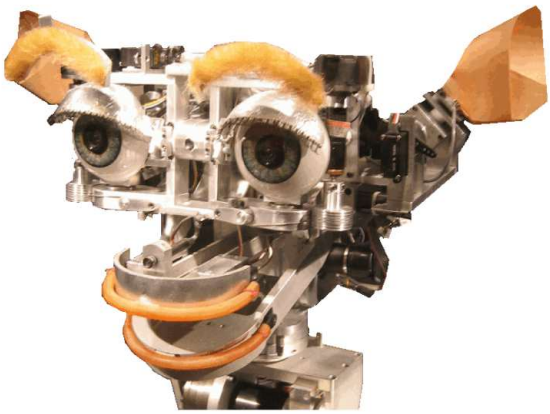
Enfin, je ne peux résister à l'envie de vous parler de Watson, un programme informatique

[2]Article de Wikipedia : Vie artificielle.

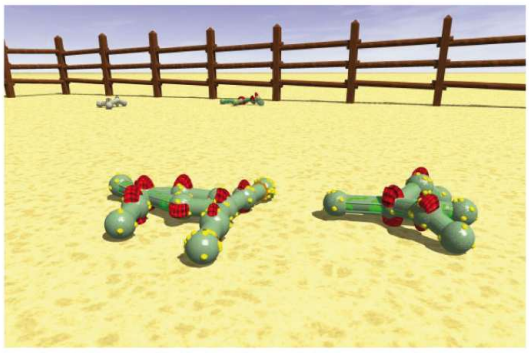
d'intelligence artificielle conçu par IBM dans le but de répondre à des questions formulées en langue naturelle. En Février 2011, Watson a battu les deux champions du monde (humains) au jeu télévisé américain Jeopardy. Le présentateur pose n'importe quelle question aux candidats (transcrite en texte pour le programme) et Watson « comprend » puis répond en moyenne plus vite et mieux que les candidats humains. Il n'a pourtant pas accès à Internet. . .mais dispose de 200 millions de page de contenu (incluant tout Wikipedia) sur 4 téraoctets de disque dur.



La tortue-robot Elsie



Le robot Kismet



Les Framsticks




Le programme Watson

**Figure XI.1** – Quelques exemples de vie artificielle.

Je dirais que les trois techniques qui ont été le plus utilisées dans le développement de la vie artificielle sont les automates cellulaires (un système dynamique discret avec des règles fixes et simples mais capable de faire émerger une grande complexité), les réseaux de neurones (capables d'apprentissage et donc de fournir un mécanisme perceptif indépendant des idées du programmeur) et les algorithmes évolutionnistes (servant à résoudre des problèmes d'optimisation ; par exemple l'algorithme génétique). Dans les dénominations de ces trois techniques apparaissent

des mots issus de la biologie : « cellulaires », « neurones » et « génétique ». Pourquoi les informaticiens empruntent-ils du vocabulaire à la biologie pour décrire le fonctionnement de leurs algorithmes ? D'où viennent les liens entre informatique et biologie, quelles sont les analogies et différences entre une cellule et un ordinateur. Il est temps maintenant de regarder tout cela de plus prés.

# 3 Les analogies et différences entre cellules et ordinateurs

Actuellement, la différence la plus évidente entre une cellule et un ordinateur, c'est la taille. Une bactérie comme E. coli fait environ 1$\mu$m de large sur 2$\mu$m de long alors que les microprocesseurs de nos ordinateurs et téléphones ont encore une taille de l'ordre du centimètre même si la miniaturisation se poursuit.

La deuxième différence c'est le fait que le réseau nécessaire au calcul dans un circuit intégré est réel alors qu'il est virtuel dans une bactérie (voir chapitre sur les interactions). La troisième différence qui saute aux yeux c'est que la partie « hardware » de nos ordinateurs est « sèche » alors qu'une cellule est faite de matière humide « wetware ».

Il existe pourtant une analogie évidente entre ordinateur et cellules : dans les deux cas, il y a découplage entre le programme et la machine. L'ordinateur et le cd-rom (ou la clef USB pour être plus moderne) sont deux entités bien distinctes. Pour ceux qui préfèrent parler de machine de Turing (modèle abstrait du fonctionnement d'un ordinateur), il y a bien un découplage entre « le ruban » et la « tête de lecture ». Dans une cellule, c'est la même chose, il y a bien découplage entre la machine (machinerie globale, métabolisme : toute la cellule sauf l'ADN) et le programme c'est-à-dire l'ADN (Danchin, 2008). Dans les deux cas, on a bien un programme statique (qui ne change pas) et une exécution dynamique.

Cependant cette exécution revêt des différences de poids. La machine de Turing (=l'ordinateur) avance le long du ruban (=le long du programme) de manière séquentielle alors que l'ADN d'une cellule est lu et exécuté de manière parallèle par les ARN polymérases et les ribosomes[3]. De plus, une machine de turing est strictement déterministe alors que la cellule « compute » de manière probabiliste via le mouvement brownien des molécules.

Les processeurs des ordinateurs sont composés de transistors ne gérant chacun que deux états. Un programme informatique est donc, au final, interprété par la machine comme une

[3]Cette manière de voir est partiellement fausse si on imagine le travail de plusieurs « coeurs » d'un processeur ou, à l'inverse, le travail d'une seule ARN polymérase dans la cellule

suite de 0 et de 1 (système binaire : des bits). Au contraire, l'ADN est une suite, non pas de deux états, mais de 4 états : les 4 bases nucléotidiques (A, T, C, G). Un ordinateur est incapable de s'auto-reproduire. D'ailleurs aucune machine créée par l'homme n'a actuellement cette faculté alors qu'une cellule en est capable. Faire deux bactéries, c'est même la « finalité » (vous verrez plus loin que je n'ai pas peur de ces anthropocentrismes) d'une bactérie.

On doit à John von Neumann l'explication théorique du fonctionnement d'un système auto-réplicatif[4] (Fates, 2001). Cet homme, à l'esprit décidément très créatif, a aussi suggéré de lâcher la bombe A « Fat man » sur Kyoto, *idée* finalement écartée par Roosevelt[5] :

⇒ Un système auto-réplicatif contient une description de lui-même (dans une cellule, cette description c'est l'ADN)
⇒ Il ne contient pas (ne nécessite pas) une description de la description (ce qui évite la récursivité à l'infini) (il n'y a pas besoin dans l'ADN d'une partie qui décrit l'ADN)
⇒ La description (l'ADN) est utilisée deux fois, comme ensemble d'instruction (transcription et traduction), et comme support qui peut être copié (réplication)
⇒ Un constructeur universelle (ARN polymérases + ribosomes) est capable d'interpréter la description en construisant un grand nombre d'objets (les protéines)
⇒ Un des objets construit par le constructeur universel (l'ADN polymérase) est capable de copier la description.

Revenons maintenant à nos différences entre cellule et ordinateur : il y a une différence essentielle relative à la notion d'ordre et de complexité. En biologie, on vit dans le « paradigme de l'évolution » : toutes structures complexes, contenant de l'ordre (une cellule, un œil, une girafe, une enzyme) ne peut provenir que de l'algorithme génétique (l'évolution) qui « tourne » depuis quelques milliards d'années en générant des mutations et des recombinaisons puis sélectionne inexorablement les structures les mieux adaptées.

A l'inverse, en informatique, des structures ordonnées, désordonnées ou à la frontière entre « ordre et chaos » peuvent émerger à partir de règles très simples, élémentaires même (c'est-à-dire sans nécessiter 4 milliards d'années d'évolution). C'est ce que l'on observe avec les automates cellulaires que nous allons étudier maintenant.

[4]Les points ci-dessous sont repris et modifiés à partir du très bon rapport de DEA de Nazim Fatés.
[5]Article Wikipedia : John von Neumann.

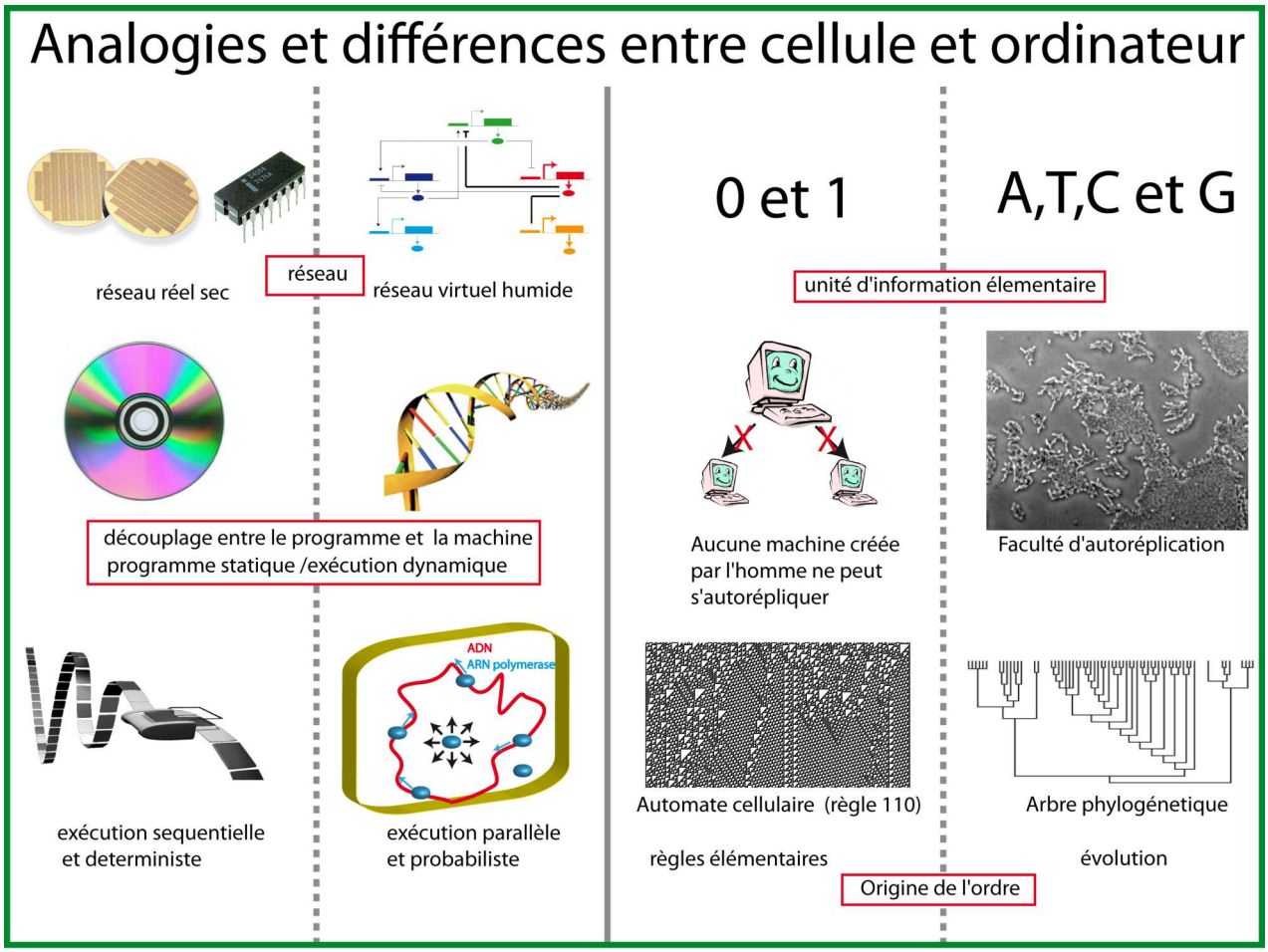



**Figure XI.2** – Analogies et différences entre une cellule et un ordinateur

# 4 Les automates cellulaires

Un automate cellulaire[6] est un système dynamique discret consistant en une grille régulière où chaque case de la grille (appelée cellule) peut prendre un ensemble fini d'états (par exemple 2 états : noir ou blanc). Le système peut évoluer au cours du temps : l'état d'une cellule au temps t+1 est fonction de l'état au temps t d'un nombre fini de cellules appelé son « voisinage ». À chaque nouvelle unité de temps, les mêmes règles sont appliquées simultanément à l'ensemble des cellules de la grille, produisant une nouvelle « génération » de cellules qui dépendent entièrement de la génération précédente.

[6]Cette introduction sur les automates cellulaires est très largement inspiré de l'article « automates cellulaires » de Wikipedia. Le lecteur doit donc attribuer la paternité de certains paragraphes ci-dessous aux contributeurs Wikipedia. Notez cependant que ces découvertes sur les automates cellulaires élémentaires doivent être attribuées, quant à elles, à Stephen Wolfram (*A new kind of science*).

L'automate cellulaire non trivial le plus simple que l'on puisse concevoir consiste en une grille unidimensionnelle de cellules ne pouvant prendre que deux états (« 0 » ou « 1 »), avec un voisinage constitué, pour chaque cellule, d'elle-même et des deux cellules qui lui sont adjacentes. Chacune des cellules pouvant prendre deux états, il existe $2^3$=8 configurations (ou motifs) possibles d'un tel voisinage. Pour que l'automate cellulaire fonctionne, il faut définir quel doit être l'état, à la génération suivante, d'une cellule pour chacun de ces motifs. Il y a $2^8$=256 façons différentes de s'y prendre, soit donc 256 automates cellulaires différents de ce type.

Les automates de cette famille sont dits « élémentaires ». On les désigne souvent par un entier entre 0 et 255 dont la représentation binaire est la suite des états pris par l'automate sur les motifs successifs 111, 110, 101, etc.

À titre d'exemple, considérons l'automate cellulaire défini par la table suivante, qui donne la règle d'évolution suivante :

| Motif initial (t) | 111 | 110 | 101 | 100 | 011 | 010 | 001 | 000 |
|---|---|---|---|---|---|---|---|---|
| Valeur suivante de la cellule centrale (t+1) | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |

Cette table montre que si par exemple, à un temps t donné, une cellule est à l'état « 1 », sa voisine de gauche à l'état « 1 » et sa voisine de droite à l'état « 0 » ($3^{\text{ème}}$ colonne du tableau), au temps t+1, la cellule centrale sera à l'état « 0 ». Par convention, la règle présentée dans le tableau ci-dessus est nommée « règle 30 », car 30 s'écrit 00011110 en binaire et 00011110 correspond à la deuxième ligne du tableau ci-dessus, décrivant la règle d'évolution.

Dans un automate élémentaire unidimensionnel, le temps est représenté du haut vers le bas c'est-à-dire que chaque ligne est le résultat de la ligne précédente. Si l'on part d'une grille initiale où toutes les cellules sont à l'état « 0 » sauf une et que l'on suit la « règle 30 » on aboutit au résultat visible sur la figure XI.3.

Les 256 règles (pourtant très proches) ne conduisent pas toutes au même genre de structures (j'ose le mot « phénotype »). Stephen Wolfram (Mathématicien célèbre à qui l'on doit notamment, le logiciel de calcul formel *Mathematica*) a proposé une classification de ces structures (Wolfram, 1984) :

⇒ Classe I : presque toute configuration initiale conduit à un état homogène (exemple avec la règle 254 sur la figure XI.3).

⇒ Classe II : des structures stables ou périodiques émergent, mais rien de plus (exemple avec la règle 90 sur la figure XI.3).

⇒ Classe III : comportement chaotique avec des motifs apériodiques. À long terme les fréquences d'apparitions des différents motifs se stabilisent (exemple avec la règle 30 sur la figure XI.3)

⇒ Classe IV : « émergence » de structures complexes capables d'osciller, de se mouvoir, voire de persévérer plus ou moins dans leur auto-organisation malgré des perturbations structurelles (exemple avec la fameuse règle 110 sur la figure XI.3)

C'est cette dernière classe qui est particulièrement intéressante pour nous biologistes. En effet, l'émergence (visuelle) de ces structures complexes ne peut manquer de nous rappeler un phénomène vivant : par exemple la zone que j'ai entourée d'un cercle sur la règle 110 (figure XI.3) nous rappelle la caractéristique fondamentale qu'attribue Stuart Kauffman à la vie qui nagerait « à la frontière entre ordre et chaos ».

Ce qui est « génial » avec les automates cellulaires, c'est qu'ils sont, pour nous biologistes, une transition douce, didactique, presque expérimental vers l'informatique et les mathématiques. La figure XI.3 est générée par un script Matlab de 30 lignes seulement. J'indique au début du script les numéros des règles que je veux voir « plotter » (254, 90, 30 et 110) et le script se charge de faire les calculs (les plus téméraires se frotteront aux 2 lignes représentant le cœur de l'algorithme emprunté à la technique/l'astuce de David Young).

Précisons tout de même que derrière l'apparente simplicité expérimentale des automates cellulaires se cachent des problèmes de mathématiques formels bien plus difficiles (1000 fois trop difficile pour moi d'ailleurs). Mais je le répète, pour nous biologistes, la seule chose qui importe c'est de constater visuellement, « de prendre acte » que des structures complexes peuvent émerger à partir de règles triviales. Rien à voir avec les milliards d'années d'évolution nécessaires chez les être vivants. Et c'est cela qui ne peut manquer de questionner (à tort ou à raison) tout biologiste « honnête avec lui même » sur la valeur du paradigme évolutionniste et ses alternative possibles.

## 5 Le jeu de la vie de Conway

Le jeu de la vie (attention il ne s'agit pas d'un jeu en réalité) a été imaginé par John Conway dans les années 70. C'est probablement l'un des automates cellulaires les plus connus. Malgré des règles très simples, le jeu de la vie génère des motifs complexes. Le jeu se déroule sur une grille à deux dimensions dont les cases appelées cellules peuvent prendre deux états distincts : vivantes ou mortes. Comme précédemment, à chaque itération, on calcule l'évolution de chacune des cellules de la grille mais cette fois-ci, l'automate est bidimensionnel : l'évolution d'une cellule est entièrement déterminée, non pas comme tout à l'heure par ses deux voisines, mais par l'état de ses huit voisines. Le temps n'est plus représenté du haut vers le bas par des lignes représentant les itérations successives mais par des trames, à la manière d'un film,

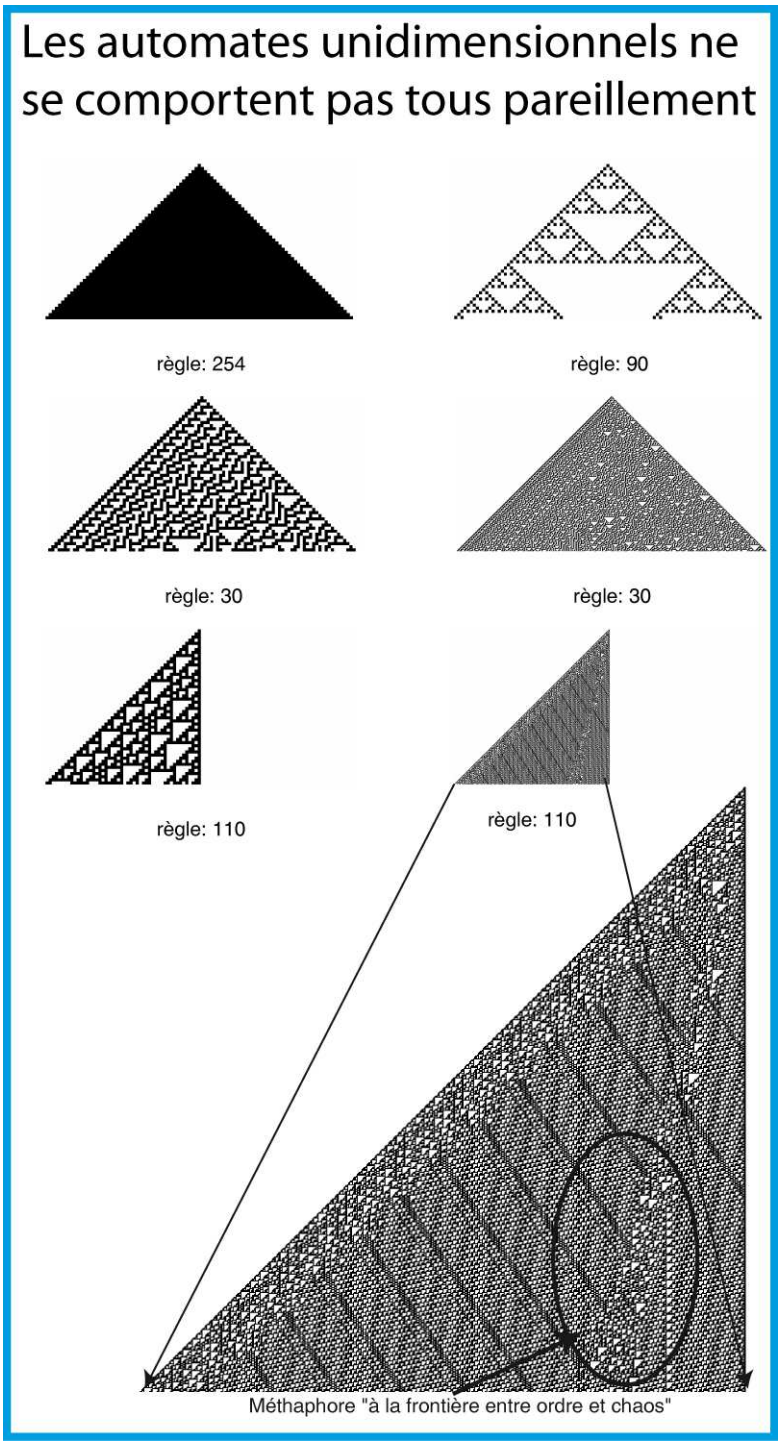



**Figure XI.3** – Illustration de quelques automates cellulaires unidimensionnelles. Rendez-vous compte que le script qui génère l'essentiel de cette figure ne fait que 30 lignes. Pour les règles 30 et 110 : la figure de gauche représente les premières itérations de la figure de droite. Voir le code Matlab en annexe Code17RegleWolfram.m qui gènère cette figure.

représentant l'état de la grille à chaque itération.

Rendez-vous compte : l'automate du jeu de la vie ne suit que les deux règles suivantes :

⇒ Une cellule morte qui possède exactement trois voisines vivantes devient vivante (elle naît).

⇒ Une cellule vivante qui possède deux ou trois voisines vivantes le reste, sinon elle meurt.

Ainsi, ces deux règles suffisent à générer des patterns complexes (Figure XI.4. La concision du script Matlab dénote par rapport à la complexité des patterns observés sur la vidéo générée par le script. Encore une fois, je souhaite vous faire remarquer que notre œil puis notre cerveau attribue à ces patterns une sorte de pseudo-activité vivante. D'où l'analogie « cellule » et le nom donné à cet automate « jeu de la vie ».

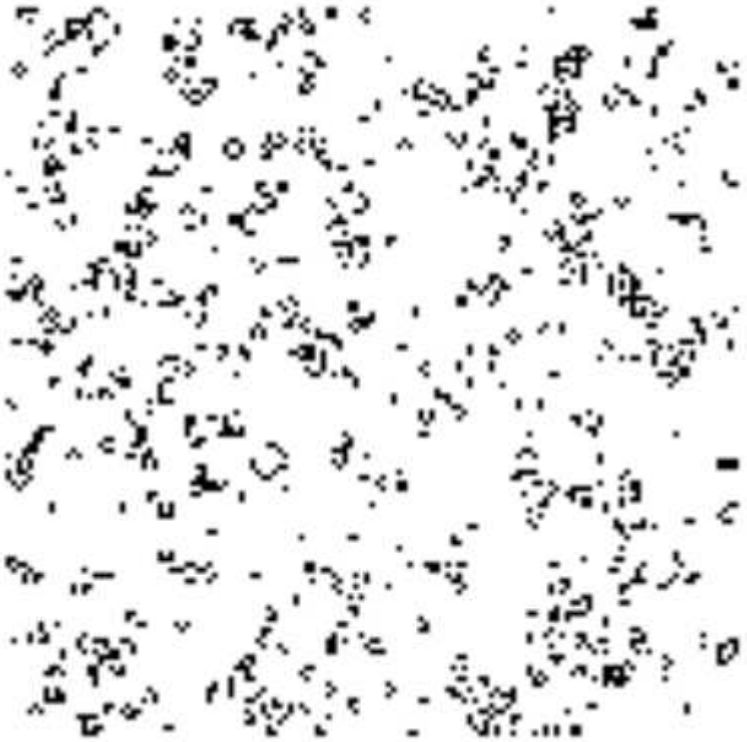


**Figure XI.4** – Vidéo illustrant le jeu de la vie de Conway (Output du script Matlab). Voir le code Matlab en annexe Code18JeuxVieConway.m qui génère cette vidéo.

Il me semble que c'est cette simple analogie visuelle, ce regard porté à cette complexité émergente, qui est l'unique raison unissant automate cellulaire (Mathématique, informatique) et biologie. Je souhaite faire ici une parenthèse sur cette question des analogies, des métaphores. J'ai trouvé assez souvent, dans des ouvrages de scientifiques ou de philosophes, une critique

forte vis-à-vis de ces analogies par exemple chez Bachelard :

> *« Aussi l'esprit scientifique doit-il sans cesse lutter contre les images, contres les analogies, contre les métaphores »* [7]

Je pourrais facilement sombrer dans le positivisme dogmatique et m'élever formellement contre le lien visuel invérifiable, infalsifiable, ce lien si mince qui relie « l'impression de vie » avec « l'automate cellulaire de Conway ». Mais je soupçonne au contraire ces analogies de cacher plus qu'elles ne laissent transparaitre. Pour prendre un autre exemple, j'ai remarqué qu'en biologie, on utilise beaucoup d'images « humaines » : les gènes « égoïstes », un phénomène « altruiste », la « communication » intercellulaire, le « code » ou « texte » génétique, « la fonction » ou « finalité » d'une enzyme, d'un organe. L'interprétation la plus classique de ces analogies c'est l'anthropocentrisme : la fâcheuse tendance à tout ramener à l'homme. Or, il me semble que « ces dérives » des analogies existent principalement en biologie, beaucoup moins en physique ou en chimie par exemple. La question qui vient alors immédiatement à l'esprit c'est : pourquoi en biologie, les scientifiques trouvent autant de « concepts » à dimension humaine ? Deux réponses possibles : la faute anthropocentrique ou, comme je le soupçonne, un lien indicible, complexe, indiscernable mais qui donne à ces concepts (probablement à redéfinir) une dimension plus universelle dans le règne vivant qu'on ne l'imagine *a priori*.

Ainsi, je ne condamnerais pas les métaphores et les images aussi vite. Elles sont souvent infalsifiables et irrationnelles mais, selon moi, elles restent une pièce cruciale (par exemple les « fous ») de l'échiquier scientifique. Elles construisent et orientent la pensée. Parfois en mal mais le plus souvent, je le soupçonne, en bien.

# 6 Un peu de grammaire (les systèmes de substitution séquentielle)

Les automates cellulaires ne sont pas seuls à pouvoir créer de la complexité à partir de règles simples. Stephen Wolfram identifie, dans son ouvrage *A new kind of science*, un grand nombre de systèmes montrant des caractéristiques similaires. Je vais en détailler un en particulier : le système de substitution séquentielle[8]. Derrière ces mots barbares se cache ni plus ni moins que la fonction « find and replace » de votre éditeur de texte préféré. Concrètement, je définis initialement 3 règles de substitution (de grammaire) :

⇒ Si l'algorithme trouve le motif 011, il le remplace par le motif 1000.

[7] Bachelard, *La formation de l'esprit scientifique*, Vrin p. 45.
[8] Stephen Wolfram, *A new kind of science*, p. 92.

⇒ Si l'algorithme trouve le motif 0100, il le remplace par le motif 101.
⇒ Si l'algorithme trouve le motif 1, il le remplace par le motif 1100.

Puis l'algorithme scanne la séquence initiale de gauche à droite, opère le remplacement selon les règles de substitution puis passe à l'itération suivante. La visualisation des itérations s'effectue, comme avec les automates élémentaires, avec des lignes en noir et blanc (0 et 1) de telle sorte que la nouvelle ligne soit le résultat de la substitution effectuée sur la ligne précédente (figure XI.5). En fonction de règles, on peut, comme avec les automates, générer des patterns homogènes, réguliers, aléatoires ou un peu « bizarres » comme le pattern présenté sur la figure XI.5.

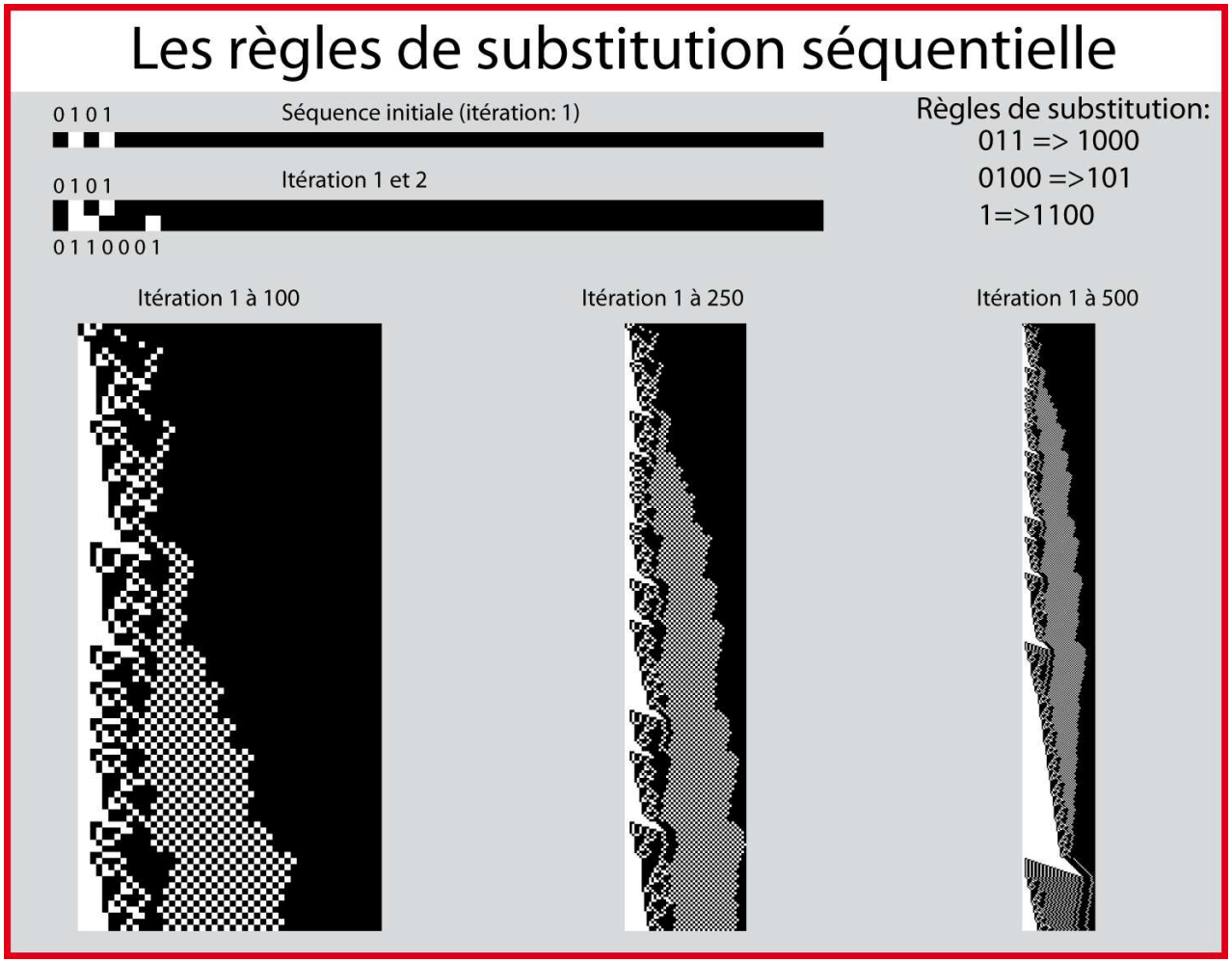



**Figure XI.5** – Les règles de substitution séquentielle. En haut à gauche : exemple de la première itération : on remplace le premier 1 par 1100. Et ainsi de suite : les différentes images montrent de plus en plus d'itérations. Voir le code Matlab en annexe Code19GrammaireWolfram.m qui gènère cette figure.

Mais vous allez me dire : oui mais en quoi cela est-il intéressant ? N'est-ce pas un peu redondant avec la partie sur les automates cellulaires élémentaires ? J'aime cet exemple car il nous permet de dresser un pont vers les travaux de Stuart Kauffman. Ce dernier a écrit deux ouvrages *The origin of order* et son équivalent un peu simplifié (parfaitement suffisant pour notre objectif) nommé *At home in the universe*.

L'hypothèse principale des ouvrages de Stuart Kauffman est la suivante : de l'ordre peut émerger sans forcement que cela soit dû à l'évolution. La vie se situerait et évoluerait à la frontière entre ordre et chaos. Cette frontière ne serait pas nécessairement le produit de l'évolution. Au contraire, elle lui serait « antérieure », elle la « précéderait », elle serait un concept « de plus bas niveau » si vous préférez.

Pour argumenter l'émergence de l'ordre en absence de toute pression de sélection, Stuart Kauffman développe un certain nombre d'idées (voir chapitre sur l'évolution II). Je vais en développer une ici[9]. L'expérience qui suit est à connecter au modèle de « la soupe originelle » expliquant les origines de la vie.

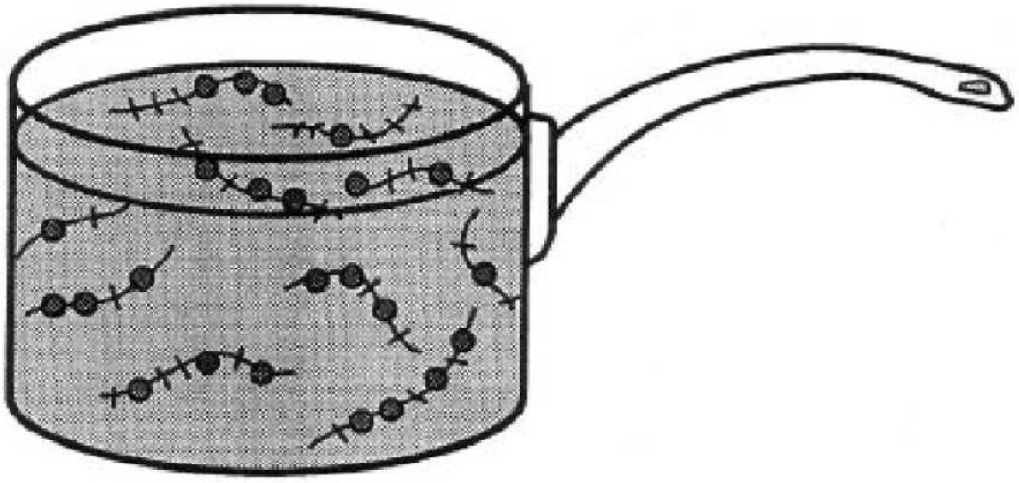


**Figure XI.6** – La soupe de « séquences » (string). Cette illustration est tirée de *At home in the universe*. Quand on mélange différentes séquences ensembles, une succession de nouvelles séquences émerge.

Prenez une casserole et ajoutez-y vos différentes séquences de bits (suite de 0 et de 1). Ces séquences doivent être envisagées comme des molécules. Comme tout à l'heure, il existe des règles de substitution mais cette fois-ci, celles-ci fonctionnent en trio. Si par exemple un substrat (par exemple 111) rencontre une enzyme (01001) alors on substitue le substrat (111) par le produit (010).

[9]Stuart Kauffman, *At home in the universe*, P. 156-157.

Tirez maintenant deux de ces séquences (molécules) au hasard (on imagine qu'il s'agit d'une collision) : si vous tirez un substrat et son enzyme correspondante alors remplacez dans la soupe le substrat par le produit, puis recommencez depuis le début : tirez deux séquences (molécules) au hasard, s'il y a correspondance avec une règle de grammaire, opérez la substitution et ainsi de suite. Au fur et mesure, de nouveaux motifs (séquences) vont être créés. Ces motifs sont autant de nouveaux sites enzymatiques potentiels capables de reconnaitre un nouveau substrat pour le convertir en nouveau produit. Et c'est tout ce qu'il faut : on a une chimie algorithmique spécifiée par une grammaire spécifique. Les séquences de bits dans une casserole se transforment en nouvelles séquences de bits et ce à l'infini générant une « floraison » de nouvelles séquences. C'est l'émergence de ces patterns de floraison au cours du temps qui est intéressante. Car de là peut naitre de l'ordre mais également des structures d'une grande complexité : selon Stuart Kauffman, ce modèle de soupe originelle pourrait expliquer ni plus ni moins que la coévolution technologique de nos sociétés ! Il ne s'agit bien sûr que de simples métaphores/spéculations mais alors pourquoi nous semblent-elles taper si justes ? Pourquoi nous semblent-elles si « orientantes » ?

A partir de la description de Stuart Kauffman de la chimie algorithmique, je n'ai pas vraiment réussi à comprendre si sa description correspondait à une simple argumentation/métaphore ou si celle-ci était confortée par des résultats de simulation. En effet, l'auteur rentre ensuite dans des considérations mathématiques abstraites et complexes qui me dépassent totalement. Biologiste de formation, j'ai souhaité resté à un niveau que je qualifierais de plus expérimentale c'est à dire plus à ma portée : j'ai essayé d'encoder la soupe originelle de séquences sous Matlab. Il faut considérer cela comme un jeu dont on sait, à la base, qu'il n'en sortira rien d'autre que « l'aspect formateur » de la démarche.

Cela a été plus compliqué que je m'y attendais. Partant de la description de Kauffman, il a néanmoins fallu que j'opère un certain nombre de choix qui, au final, m'ont peut être fondamentalement éloigné de ce qu'avait en tête son auteur.

Je ne vais pas détailler exactement la manière dont je m'y suis pris, ce serait trop long. Brièvement, une séquence de 50 bits évolue au cours du temps en fonction de 150 règles de substitution. Ces règles définissent des enzymes, des produits et des substrats qui sont des séquences en bits choisis aléatoirement. Les tailles de ces séquences sont également choisies aléatoirement mais elles sont comprises entre 3 et 15 bits. A chaque itération, toutes les règles de grammaire sont appliquées à la séquence dans un ordre aléatoire et de manière dose dépendante : si un substrat et une enzyme particulière sont présents en de nombreux exemplaires, il y aura beaucoup de produits dans la séquence suivante. J'ai fait tourner l'algorithme sur 1500 itérations ce qui représente en réalité plusieurs dizaines de milliers de substitution.

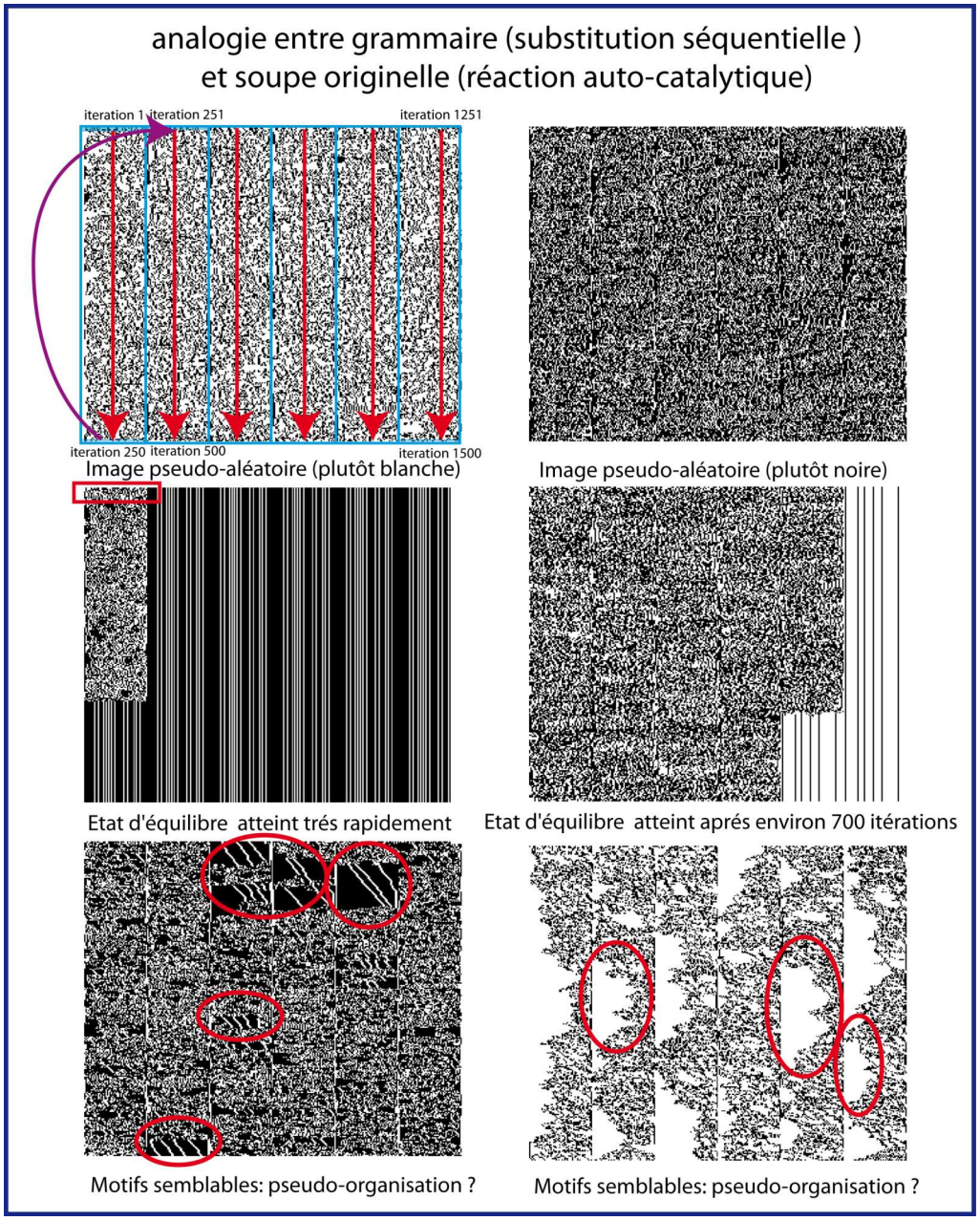



**Figure XI.7** – Résultats de 6 simulations (6 carrés) de la soupe de « séquence ». Voici les paramètres :
- Longueur de la séquence (soupe de molécules) : 50 bits (gardée constante)
- Nombre de règles de substitution (grammaire) : 150 (définies aléatoirement)
- Taille maximum des enzymes, substrats et produits : 15 bit
- Taille minimum des enzymes, substrats et produits : 3 bits
- Nombre d'itérations : 1500

Pour faciliter la visualisation des résultats (éviter des colonnes trop longues de 1500 lignes), les colonnes sont poolées en carré (par tranche de 250 lignes) comme indiqué sur le schéma en haut à gauche. L'encadré rouge sur le carré du milieu montre 10 lignes de 50 bits soit les 10 premières itérations. Les ronds rouges sur les carrés du bas montrent des zones qui attirent l'œil : représentent-ils quelque chose ayant des similitudes avec la vie ? Voir le code Matlab en annexe Code20GrammaireSoupeOriginelle.m qui génère cette figure.

J'identifie deux questions fondamentales à se poser quand on entreprend d'encoder ce type de simulation :

⇒ *quelle est la meilleure forme de représentation des résultats ?* je n'ai pas eu d'autres idées que les images (que le cerveau analyse en moins d'une seconde). La première ligne correspond à la condition initiale aléatoire : c'est une ligne de 50 bits représentées en noir et blanc (les 0 et les 1). La ligne directement en dessous est le résultat d'une itération comportant plusieurs substitutions. Les 1500 itérations forment une colonne trop longue pour être visualisée facilement (50 pixels x 1500 pixels). Dans la figure XI.7 illustrant les résultats obtenus lors des simulations, j'ai donc poolé des colonnes de 250 lignes ensembles ce qui génère un carré. L'angle gauche du haut de mon carré représente la première itération alors que l'angle droit du bas représente la dernière itération. Ce type de représentation et d'analyse des résultats est limité car bien que l'œil soit extrêmement performant pour détecter des structures complexes, rappelant la vie, il ne serait en revanche pas très prudent de lui accorder une confiance excessive.

⇒ *Car : que cherche-t-on vraiment au final ?* La première chose à constater en regardant les différents carrés de la figure XI.7, c'est la grande variabilité visuelle des images obtenues (j'en ai regardé presque mille soit plusieurs heures de simulations). Rappelons la présence de 150 règles de grammaire dans toutes les simulations. Intuitivement, on aurait pu s'attendre à ce qu'un nombre si important de règles choisies aléatoirement limite mécaniquement / statistiquement la variabilité visuelle des images. Or ce n'est pas du tout le cas : on voit que :

- de nombreuses images semblent grossièrement aléatoires (pseudo-aléatoire car la présence de la grammaire introduit un biais). Cette « aléatoire ressenti » varie dans la prédominance du blanc ou du noir et dans ce que j'appellerai la « granularité » (figure XI.7 : les deux carrés du haut).
- de nombreuses images finissent par atteindre un état d'équilibre (figure XI.7 : les deux carrés du milieu)
- certaines images, plus rares, semblent plus complexes, plus « organisées ». Ce sont ces images qui rappellent le plus la vie grâce à la métaphore « la vie navigue à la frontière entre ordre et chaos ».

Mais on constate aussi qu'il est difficile d'aller plus loin, de chercher des conclusions et pistes plus en profondeur. Il semble que cette approche expérimentale nous cantonne, quoi qu'il arrive, au domaine de la métaphore. Ce que nous recherchons, en réalité, dans les images,

c'est une information de *valeur*[10]. Or nous ne trouvons cette information ni dans les images aléatoires, ni dans les images homogènes. Nous sommes maintenant prêts à aborder le chapitre sur l'information.

[10]Terme emprunté à Antoine Danchin.

# Références bibliographiques


Breazeal, Cynthia. 1998. Early Experiments using Motivations to Regulate Human-Robot Interaction. 252

Danchin, A. 2008. Bacteria as computers making computers. *FEMS Microbiol Rev.* 254

Fates, Nazim. 2001. Les automates cellulaires : vers une nouvelle epistemologie ? *Rapport de DEA*. 255

Komosinski, Maciej, & Ulatowski, Szymon. 2009. *Framsticks : Creating and Understanding Complexity of Life*. second edn. New York : Springer. Chap. 5, page 107–148. 252

Wolfram, S. 1984. Universality and Complexity in Cellular Automata. *Physica D.* 257

# Chapitre XII

# L'information

## 1 La notion d'information

Nous avons vu, dans le chapitre précédent, que le cerveau, par l'intermédiaire de l'œil humain, accorde une plus grande quantité d'information à certaines images qu'à d'autres. Mais qu'appelle-t-on exactement information ? En général, quand on parle d'information, on parle d'une chaine de caractère. Il peut s'agir de nos journaux quotidiens, du code secret de ma carte bancaire, des suites de caractères présents dans l'ouvrage *La barque de Delphe*[1] , de la suite des notes d'une mélodie agréable, des suites de caractères du dernier programme que je vous ai décrit, de coordonnées GPS des emplacements de tous les « Mac Donald » sur le territoire français. Bref, ce mot information est très large : il englobe beaucoup de choses. Pourtant, il me semble que le concept d'information effraie parfois les biologistes car il est souvent associé aux mathématiques via les fameuses « théories de l'information ». Dans ce chapitre, j'aurai atteint mon but si je réussis à convaincre des biologistes que pour comprendre l'essence même de ces théories, il n'est pas forcement nécessaire de comprendre les équations associées —Soyez bien certain d'ailleurs que je ne les comprends pas—.

En revanche, il y a deux choses dont il faut avoir conscience :

⇒ La limitation intrinsèque des théories actuelles de l'information : il n'existe pas de théorie de l'information de *valeur*[2] capable de s'intéresser à la signification d'un message.

⇒ Quand on parle d'information, on peut très vite faire de la métaphysique déguisée derrière des mathématiques —et ce n'est pas un problème à condition d'en avoir conscience—.

Dans l'exposé qui suit, je vais poser une question :
*« Quelle peut être l'information associée à un génome ? »*

[1] Antoine Danchin, *La barque de Delphes*, Odile Jacob.
[2] Pour reprendre la formulation d'Antoine Danchin.

Je vais tout d'abord tenter de décrire quelle sont les apports des théories actuelles pour évaluer l'information contenue dans un génome. Puis je développerai un point de vue personnel pour essayer d'aller plus loin. Chacun sera libre de fixer librement le curseur entre science et métaphysique.

# 2 La théorie de l'information de Claude Shannon

Par ou commencer ? Je vous propose de commencer par télécharger, via Matlab, la séquence du génome d'*E. coli* (Figure XII.1). J'enregistre ensuite cette séquence sur mon disque dur : celle-ci fait 9.2 Méga-octets pour un fichier texte de 4 639 675 paires de bases. Maintenant, je crée une séquence aléatoire de même taille (4 639 675 pb), je l'enregistre sur mon disque dur et, surprise, celle-ci fait exactement la même taille : 9.2 Mo.

A partir de là, deux questions apparaissent :

⇒ Pourquoi ces séquences font-elles 9.2 Mo sur mon ordinateur ?

⇒ Pourquoi la séquence du génome d'*E. coli* et une séquence aléatoire font-elles la même taille ?

## 2.1 Pourquoi 9.2Mo ?

La place qu'occupe tout fichier sur un ordinateur peut être calculée en utilisant la théorie de l'information dont Claude Shannon est le père fondateur avec son article *A Mathematical Theory of Communication* publié en 1948. L'équation principale de la théorie de l'information est :

$$I = k.log_2(P) \tag{XII.1}$$

Dans notre cas, I représente la quantité d'information (en bit) présente, selon Shannon, dans la séquence d'ADN de taille k et où P est la probabilité d'apparition d'un nucléotide donné. Par exemple, pour une séquence d'un seul nucléotide « A », on a :

I=1x$\log_2(0.25)$ = 2 bit car il y a 4 nucléotides possibles (A, T, C, G) dont une chance sur 4 (0.25) d'avoir un A.

Pour le génome d'*E. coli* (4 639 675 nucléotides) on a :

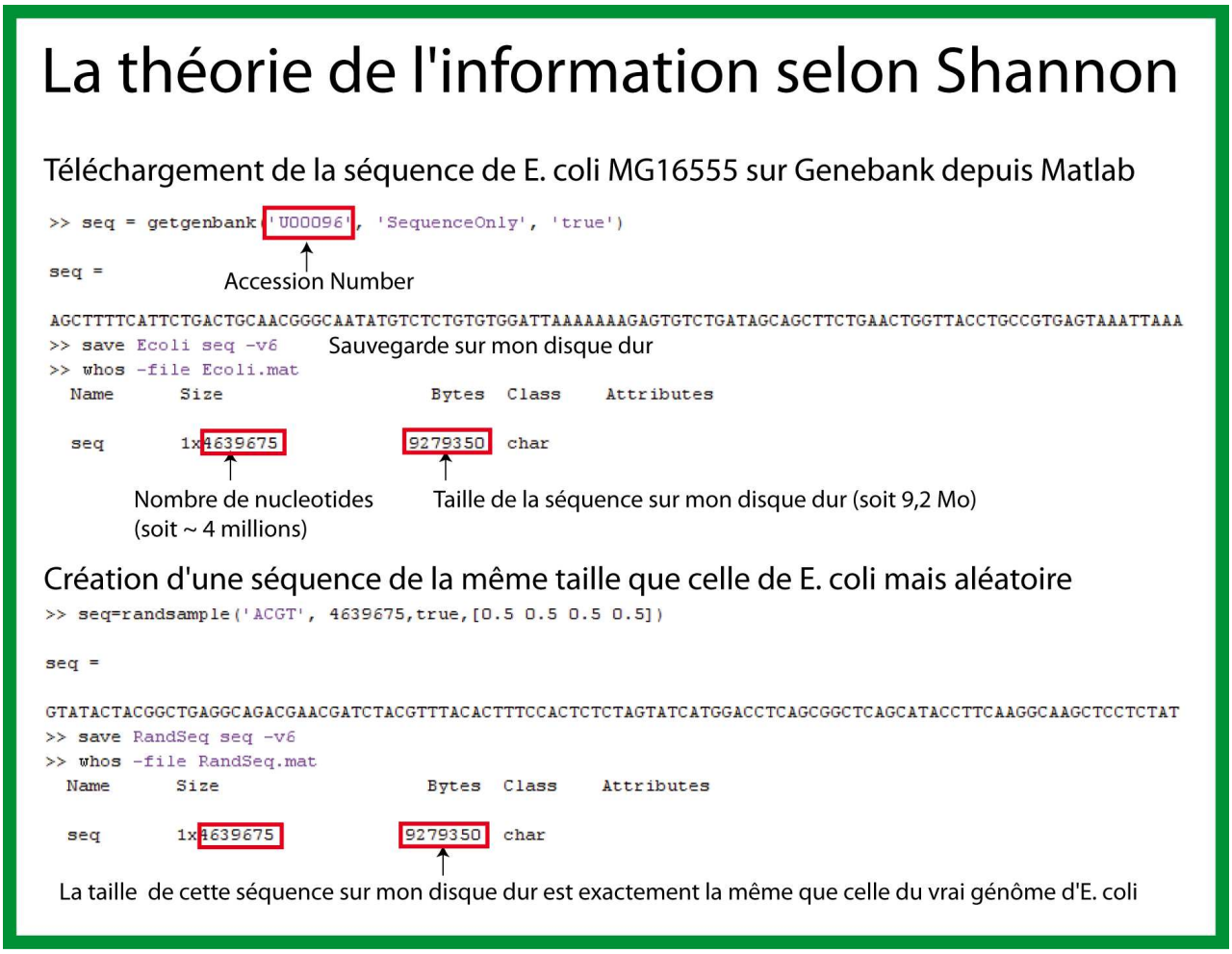



**Figure XII.1** – La théorie de l'information de Shannon

$$I= 4\ 639\ 675 \text{ x } \log_2 (0.25)=4\ 639\ 675 \text{ x } 2 = 9\ 279\ 350 \text{ bit}$$

Or un octet fait 8 bits, on a donc :

$$I= 9\ 279\ 350\ /8 = 1.16 \text{ Mega octet.}$$

C'est la taille minimum que peut prendre le génome d'*E. coli* sur un disque dur sans compression. Cependant, les éditeurs de texte encodent chaque caractère sur 8 bits (256 possibilités de lettres/symboles différents) pour pouvoir gérer le grand nombre de lettres de l'alphabet. La séquence d'*E. coli* fait donc finalement :

I= 1.16 x 8 = 9.28 Mega-octet.

### 2.2 Pourquoi la séquence du génome d'*E. coli* et une séquence aléatoire font-elles la même taille ?

La théorie de l'information, de l'aveu même de son fondateur, n'est pas vraiment une théorie de l'information mais plutôt une théorie de la communication. Cette théorie est très utilisée pour coder et transmettre un message en bits mais elle exclut totalement la sémantique : elle est donc totalement indifférente à la signification des messages. Le sens d'un message peut pourtant être considéré comme essentiel dans la caractérisation de l'information. Par exemple, les informations présentes dans les phrases « sort ! Il y a le feu ! », « Sort ! Il y a le car ! » et « sgrt e ij n a he dje r » ne possèdent pas, à l'évidence, les mêmes valeurs. Or cette « valeur », la théorie de l'information de Shannon ne pourra pas la mesurer : tous ces messages font 176 bits. De la même manière, nous ressentons intuitivement que le génome d'*E. coli* contient bien plus d'information qu'une séquence aléatoire de même taille car ce vrai génome contient l'information capable de générer une entité auto-réplicative alors que la séquence aléatoire n'est qu'une suite de symboles choisis au hasard.

## 3 La théorie algorithmique de l'information de Kolmogorov

Encore une fois, cette théorie ne va pas pouvoir mesurer la « valeur » d'une information mais elle va pouvoir mesurer sa compressibilité. En effet, l'idée principale de la théorie algorithmique de l'information est qu'une chose est d'autant plus complexe, ou contient d'autant plus d'information, qu'elle est difficile à expliquer, c'est-à-dire fondamentalement longue à expliquer. Par exemple, la séquence ADN

« AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA »

a une description courte « 30 répétitions de A » alors que la séquence aléatoire

« ATCGTGGGCTAGGGGTACGCGCCATCGATC »

n'a apparemment pas de description plus simple que la chaine de caractère elle-même.

Ainsi, du point de vue de la théorie algorithmique de l'information, découverte dans les années 60, l'information présente dans une chaine de caractère est équivalente à la longueur

de la plus courte représentation de cette chaine de caractère. Or cette représentation courte est essentiellement un programme qui, lorsqu'il s'exécute, fournit la chaine de caractère en « ouptut ».

Ainsi, la complexité algorithmique (dite complexité de Kolmogorov) d'une chaine de caractère x est définie comme la longueur du plus petit programme qui « compute » x.

Par conséquent, une chaine de caractère dont la complexité de Kolmogorov est petite, relativement à la taille de la chaine de caractère correspondante, n'est pas considérée comme complexe/informative.

La complexité de Kolmogorov permet de définir formellement le concept « aléatoire ». En effet, une chaine de caractère est dite « aléatoire » si et seulement si elle est plus petite que n'importe quel programme capable de la générer.

Ainsi, selon cette théorie, ce qui est complexe, ce qui contient de l'information, c'est ce qui est aléatoire : on parle de complexité aléatoire. Vous voyez que cette manière de pensée rentre en contradiction avec nos intuitions en biologie : pour nous, un organisme vivant est très complexe sans pour autant que l'on se représente cette organisme comme étant essentiellement composé de hasard.

Pour être sûr que vous ayez bien compris la théorie algorithmique de l'information, je vous ai concocté un petit exercice « pratique » relatif au génome. Sur la figure XII.2, vous pouvez voir un petit script que j'ai appelé Kolmogorov.m (le programme) qui consiste en une boucle qui écrit 4 639 675 fois le nucléotide A. Si je sauve cette séquence de A sur mon disque dur (Kolmogorov.mat ; la séquence) je tombe sur l'information de Shannon (9.2Mo). Cependant, le programme Kolmogorov qui m'a permis de générer cette séquence ne fait, lui, que 42 octets. Il représente une description de la séquence beaucoup plus courte que la séquence elle-même. Ainsi selon Shannon, la séquence du génome de *E. coli* et une séquence de « A » de même taille contiennent la même information (9.2Mo) alors que, selon Kolmogorov, la séquence de « A » est beaucoup plus simple/moins informative ce que nous, biologistes, ressentons effectivement intuitivement.

# 4 La profondeur logique de Bennett

La profondeur logique de Bennett est un concept de logique informatique mentionnée dans un article de Gregory Chaitin en 1977 qui l'attribue au physicien et logicien américain Charles H. Bennett.

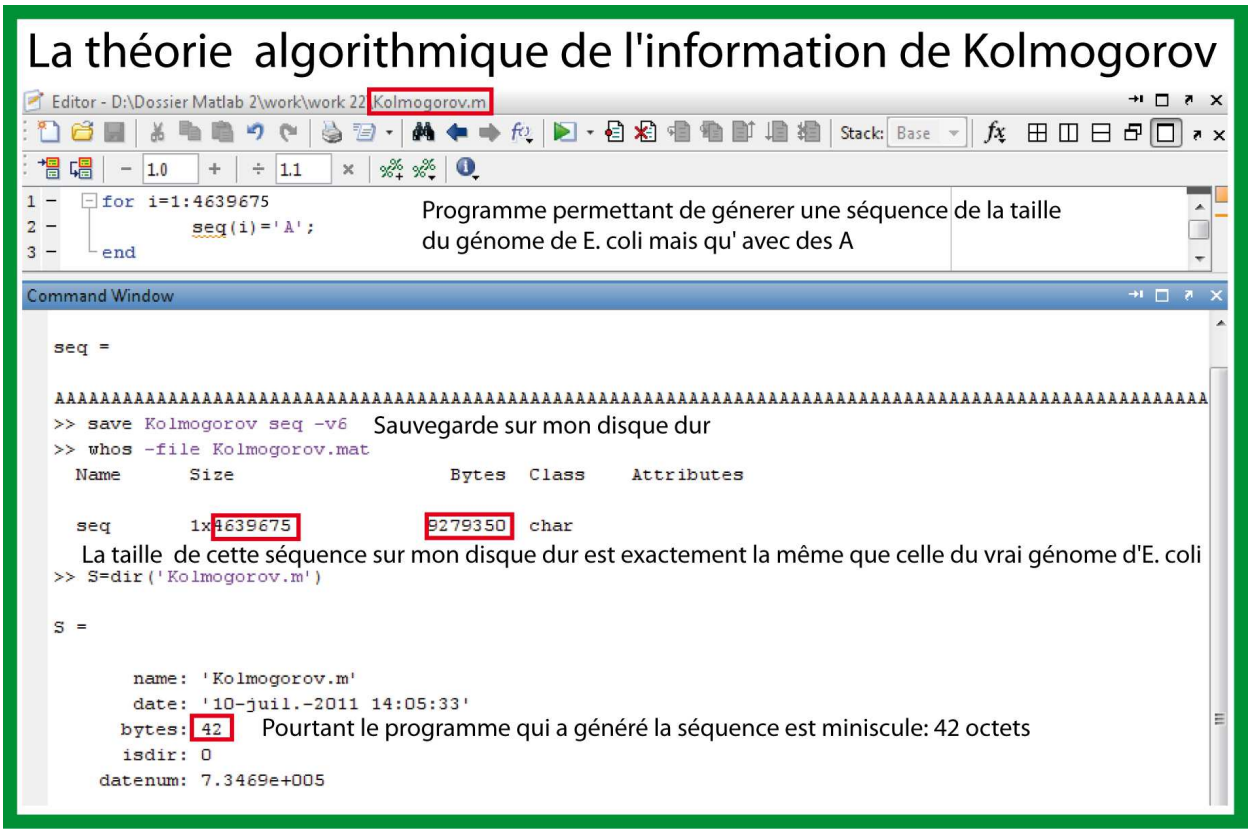



**Figure XII.2** – La théorie algorithmique de l'information de Kolmogorov

Ce concept diffère de la complexité de Kolmogorov dans le fait qu'il considère le temps de calcul de l'algorithme le plus court, plutôt que sa longueur.

Soit une chaine de caractères x, P(x) est le temps de calcul du programme minimal de x.

Le temps de calcul est lié à la présence de boucles (par exemple, la boucle for) et donc à la présence d'itérations. Ainsi l'information contenue dans un algorithme s'y trouve à l'état potentiel. Elle nécessite un temps plus ou moins long pour en être extraite.

Comme la complexité de Kolmogorov, la profondeur de Bennett est approchable mais n'est pas calculable (la fonction x -> P(x) n'est pas récursive). La profondeur logique est donc sujette à de fortes propriétés d'indécidabilité.

Cependant, la profondeur logique apporte une nouvelle dimension à la mesure de la complexité. La notion de complexité organisée (profondeur logique) par opposition à la complexité aléatoire (complexité de Kolmogorov). En effet, la profondeur logique mesure la présence de structures, d'organisation et donc elle diminue avec l'aléatoire alors que la complexité de Kolmogorov, elle, augmente avec l'aléatoire (absence de structure). Autrement dit, l'incompressibilité

entraîne la rapidité du calcul.

Je vais illustrer ces idées à l'aide de mon exemple relatif au génome illustré sur la figure XII.3. Je dispose d'un script Matlab, trouvé sur Internet, qui calcule les décimales de Pi en utilisant la méthode de John Machin. Je donne au script le nombre de décimales de Pi que je veux connaître, puis une fois le programme exécuté, celui-ci me « renvoie » la suite de ces décimales calculés :

« 31415926535897932384626433832795028841971693 »

Puis, grâce à quelques lignes de code ajoutées dans le script, ce dernier substitue les entiers par les bases ADN selon le plan suivant :

0,1,2 deviennent A ; 3,4 deviennent T ; 5,6 deviennent C et 7,8,9 deviennent G.

On obtient alors la séquence suivante qui, en quelque sorte, est un reflet de Pi :

« TATACGACCTCGGGGTATGTCACTTTGTAGGCAAGGTAGGACGT »

Pour 10 000 décimales (je ne le fais pas cette fois pour 4 639 675 décimales car cela prendrait plusieurs heures), la séquence, enregistrée sur mon ordinateur fait 20ko alors que le programme qui calcule Pi ne fait que 2 ko. Ainsi selon la théorie algorithmique de l'information de Kolmogorov, cette séquence est moins compliquée qu'une séquence aléatoire incompressible. Si maintenant, je relance le calcul en demandant non pas 10 000 mais 30 000 décimales, la taille de la séquence générée sera bien 3 fois plus grande sur mon disque dur (selon la théorie de l'information de Shannon) c'est-à-dire 60 ko. Bien sûr, la taille du programme calculant Pi n'a pas changé et fait toujours 2 ko. Or intuitivement, on ressent bien que connaître 30 000 décimales de Pi est plus informatif que de n'en connaître que 10 000. Et cette intuition là, la complexité de Kolmogorov ne peut pas la capter. Pour prendre en compte, cette complexité dite organisée, on utilise le concept de profondeur logique. En effet, le temps d'exécution du programme pour calculer les 10 000 décimales est de l'ordre de 1 seconde alors qu'il faut 10 secondes pour en calculer 30 000. Ainsi, les boucles « for » présentes dans le programme font que la séquence basée sur Pi a une forte complexité organisée (long temps de calcul) mais une faible complexité aléatoire (la taille du programme est petite).

# 5 Synthèse en image : Shannon, Kolmogorov, Bennett

Pour être sûr que les différences entre ces trois manières d'envisager la complexité (l'information) sont claires dans l'esprit de mon lecteur, je vais donner un dernier exemple basé sur des images.

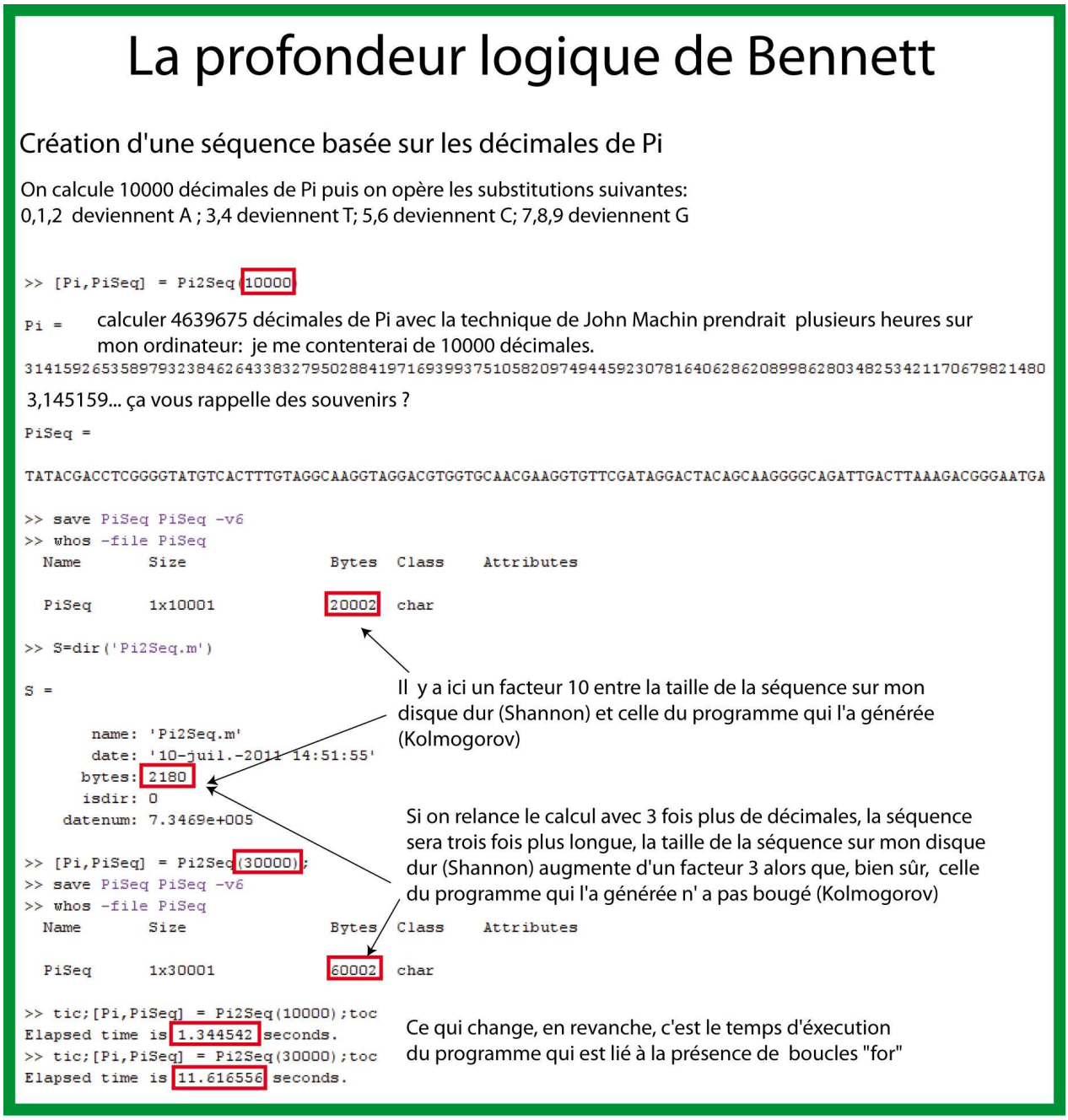



**Figure XII.3** – La profondeur logique de Bennett

Sur la figure XII.4, on observe 3 images : une image uniforme (toute noire), une image aléatoire (composée de pixels blancs et noirs placés aléatoirement) et une image correspondant à l'ensemble de Mandelbrot (une fractale populaire) en noir et blanc également. *Selon la théorie de l'information de Shannon, ces trois images contiennent autant d'information.* En effet, elles

font toutes exactement 230 ko car le nombre de pixel (520x420) est identique dans les trois cas. Selon Kolmogorov, l'image uniforme est la plus simple car le programme qui la génère ne fait que 51 Ko. L'image de l'ensemble de Mandelbrot nécessite, quant à elle, un programme de 182 ko. L'image aléatoire est générée par un programme court mais ce dernier intègre la fonction rand (fonction qui crée du hasard) or cette fonction injecte de l'information dans le système. Le script ne contient donc pas, en lui-même, la description de l'image car si on l'exécute deux fois suites, il y aura deux images différentes. La seule manière de faire réapparaitre l'image, c'est d'indiquer au programme la valeur des 520x420= 218400 pixels ce qui au final, ne représente pas une compression et nous ramène à l'information selon Shannon soit 230ko (plus une constante). *Ainsi, selon, la théorie algorithmique de l'information de Kolmogorov, l'image la plus complexe/informative est l'image aléatoire* (elle est incompressible ; c'est la complexité aléatoire), vient ensuite l'ensemble de Mandebrot puis enfin, l'image uniforme qui est celle qui contient le moins d'information.

Etudions maintenant le temps nécessaire pour générer ces images : les images uniformes et aléatoires sont générées en moins d'une seconde alors que l'ensemble de Mandelbrot a nécessité plus de 20 secondes. Cette dernière image a une plus grande profondeur logique selon Bennett. On parle de complexité organisée. En effet, cette image, par rapport aux deux autres, possède une structure, une organisation. *C'est donc ces structures liées à la profondeur logique d'un programme qui nous rappellent le plus les êtres vivants. Ce sont elles qui sont, à notre sens, les plus informatives.*

Voyons maintenant pourquoi. Commençons tout d'abord par résumer ces 3 manières de conceptualiser l'information :

⇒ La théorie de Shannon postule que le génome d'*E. coli* contient autant d'information qu'un génome aléatoire

⇒ La théorie de Kolmogorov propose de tenter de « compresser » le génome d'*E. coli* avec un programme pour évaluer sa complexité algorithmique

⇒ La théorie de Bennett propose d'évaluer le temps nécessaire pour générer ce génome (Bennett).

Les biologistes dressent immédiatement un pont entre le troisième point et la théorie de l'évolution. En effet, le temps nécessaire pour générer le génome actuel d'*E. coli* correspond aux milliards d'années du temps évolutif, c'est-à-dire des milliards de cycles de réplications/ mutations/ sélection. Je tiens cette idée d'Antoine Danchin[3] :

[3]Antoine Danchin, *Information of the chassis and information of the program in synthetic cells*, Syst Synth Biol, 2009.

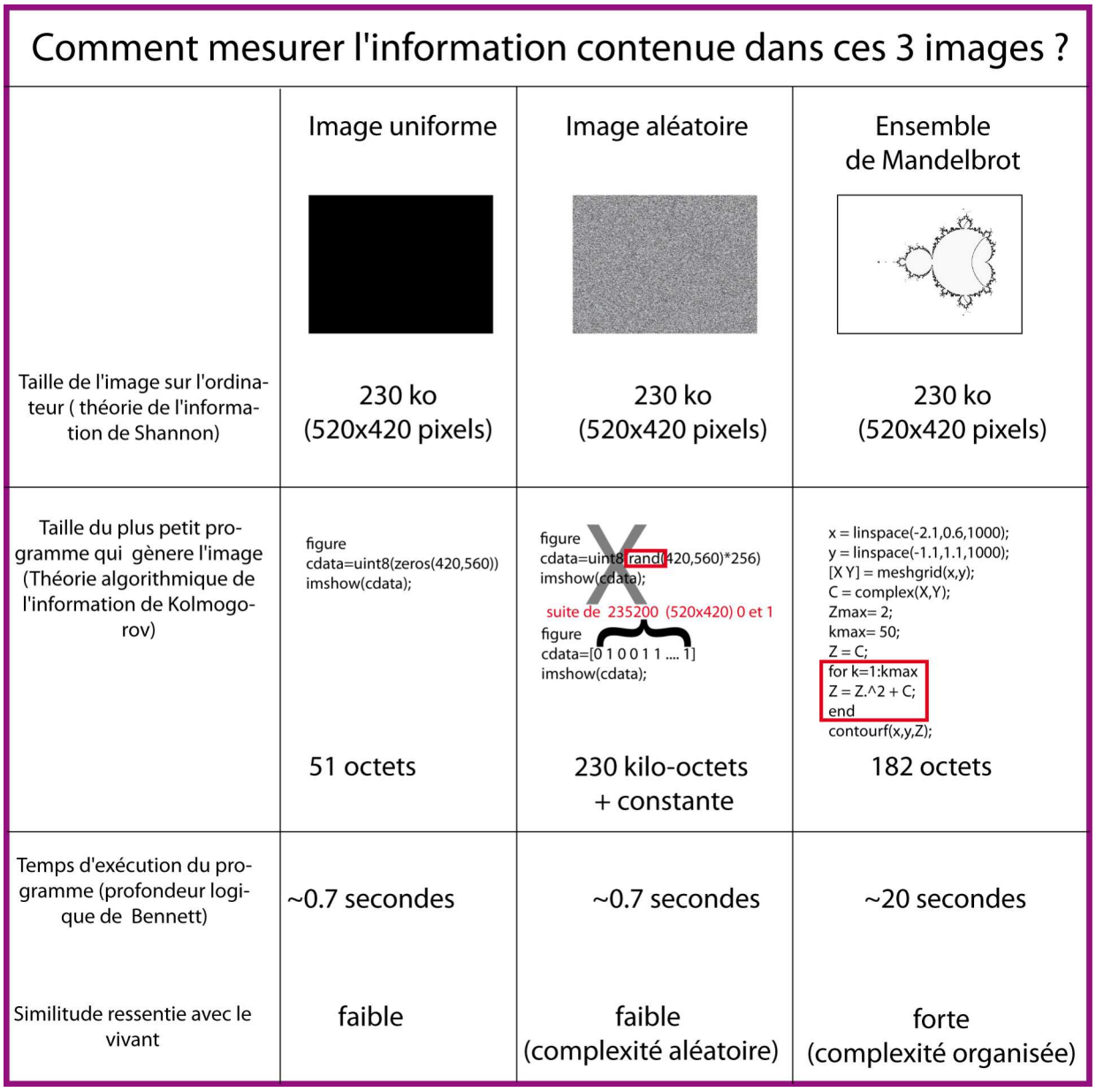



**Figure XII.4** – Comparaison des 3 théories de l'information. Voir le code Matlab en annexe Code21CompareInformationTroisImages.m qui gènère cette figure.

> *« The functions coded by the genetic program are the result of a very long evolution. And if we keep the algorithmic metaphor, because DNA comes from DNA comes from DNA...in an endless replication process, the nucleotides in the sequence have considerable logical depth ».*

A partir de là, il me semble que tout scientifique qui souhaite approfondir le concept d'information (en lien avec la biologie) doit se frayer un chemin, par lui-même, dans la jungle de

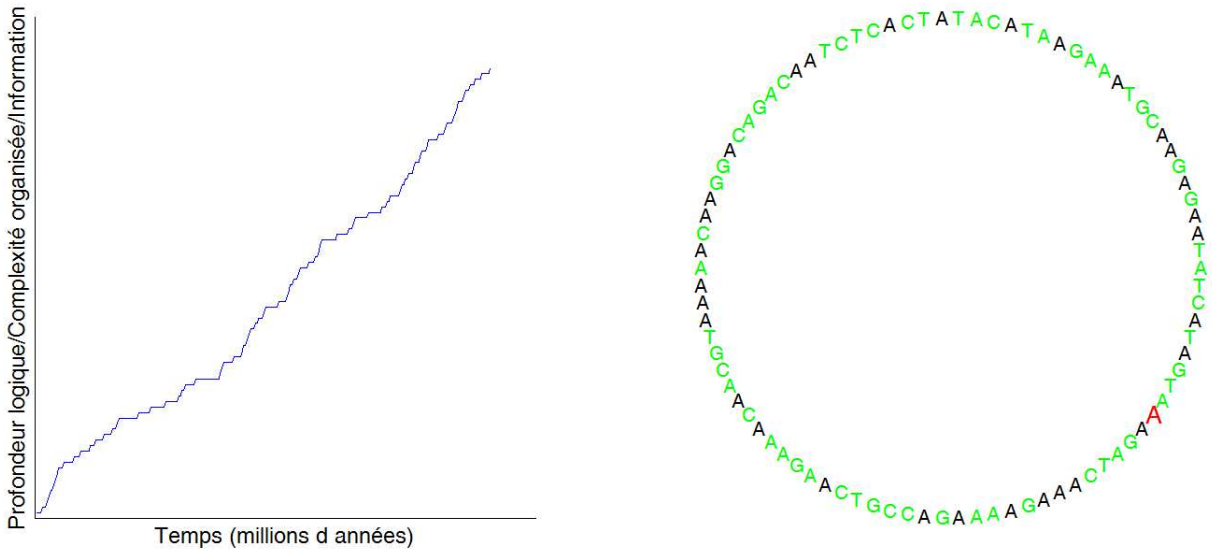



**Figure XII.5** – Video-Métaphore illustrant l'augmentation de profondeur logique d'une séquence au fur et à mesure des cycles réplication/mutation/sélection. Voir le code Matlab en annexe Code22ProfondeurLogiqueNucleotide.m qui gènère cette figure.

l'inconnu. Il dispose pour l'aider de quelques outils (ses machettes) que sont les fondements des 3 théories décrites. Cependant, pour explorer l'inexploré, il devra nécessairement en inventer de nouveaux. Il me semble que ce constat est également partagé par Antoine Danchin ou plus exactement, je le partage avec Antoine Danchin. En effet, c'est après avoir lu son « appel » ci-dessous (diffusé depuis son site web) que j'ai souhaité réfléchir à ce problème de l'information de *valeur*. Autrement dit, j'ai souhaité répondre à son appel à ma manière.

> *« C'est dans ce contexte que ce site est un appel à l'inventivité, mathématique en particulier, mais aussi expérimentale...Par construction la jeunesse manque d'expérience, et manque de références approfondies. Elle tombe donc aisément dans les modes, les poncifs et les naïvetés qui réinventent la roue. Et le plus souvent elle y reste, et elle participe alors de l'image commune que donne la vieillesse. Mais ce n'est pas inévitable, et il est une tout autre voie : cette même inexpérience a la conséquence remarquable qu'elle permet de parcourir des terrains entièrement vierges, et surtout d'oser des pistes que l'expérience aurait pu faire croire fermées pour toujours. Elle donne par définition, un point de vue extérieur[...] Or je recherche ici la création d'un nouvel objet mathématique, d'une nouvelle structure, ou la réécriture dans un ordre nouveau d'objets ou de structures anciennes pour définir ce que le sens commun appelle « information » , avec la connotation foncière d'information*

*dont on a à faire quelque chose, d'information utile, ou d'information associée à une valeur[...] J'ai l'espoir qu'un lecteur saura l'explorer. Et l'intérêt de la mathématique est aussi qu'elle permet de distinguer l'intuition romantique délirante de l'intuition profonde qui résiste au long travail de la démonstration. Ce long et difficile chemin —comme celui de l'expérience— comprend que le théorème n'existe que lorsqu'il est démontré, et que c'est au décours de la démonstration qu'apparaît l'inattendu, là où se crée l'information. »* [4]

# 6 Spéculation personnelle sur l'information

Pour débuter mon exploration dans la jungle, j'ai commencé par une intuition : tout biologiste pense que le génome d'*E. coli* (9.2Mo selon Shannon) contient beaucoup plus d'information de valeur qu'une séquence aléatoire de même taille (9.2Mo). La piste « Kolmogorov » nous oriente vers la compression de ce génome alors qu'il faut penser rigoureusement dans la direction opposée : il faut décompresser ce génome. Autrement dit, à ce génome de 9,2 Mo n'est pas associé un programme plus petit (en bit) (la complexité de Kolmogorov) mais une Information Décompressée bien plus grande (toujours en bit) que j'appellerai ID dans la suite de mon exposé. Je partirai donc du postulat suivant :

*ID existe, c'est une chaine de caractère de taille finie, en bit.*

Ce postulat est basée sur :

⇒ l'intuition d'existence :
ID (séquence du génome d'*E. coli*)> ID (séquence aléatoire de même taille)

⇒ l'intuition qu'ID a une taille finie
ID (séquence du génome humain)>ID (séquence génome d'*E. coli*)

La taille de l'information décompressée du génome humain semble intuitivement bien supérieure à celle du génome d'*E. coli*. En cela, elle constitue une borne, une limite supérieure vers le haut prouvant qu'ID (*E. coli*) est une séquence de taille finie et non infinie. Je me dois cependant de reconnaitre ici ma crainte de la faiblesse du raisonnement mathématique.

L'existence de ID peut être aussi entrevue par une expérience de pensée : la découpe d'un génome en morceaux de plus en plus petit. En effet, en coupant au fur et à mesure de deux en deux le génome d'*E. coli*, on crée des séquences de 4 600 000 bases, 2 300 000 puis 1 150 000 ...64, 32, 16, 8, 4 ,2 ,1 nucléotides (illustré sur la figure XII.6). Selon Shannon, cette découpe

[4] Site Web d'Antoine Danchin : www.normalesup.org/~adanchin/.

divise symétriquement par deux, de manière constante, le contenu en information des séquences correspondantes. Or dès la première découpe, on passe d'une séquence pouvant générer la vie (un système auto-réplicatif) à une séquence incapable de générer la vie. La perte d'information est bien supérieure à un facteur 2. Puis cette perte d'information au fur et à mesure des découpes diminue : elle n'est donc pas constante. Elle tend vers la perte de Shannon (facteur deux) qu'elle atteint lors de la dernière découpe (de 2 bases à 1 base).

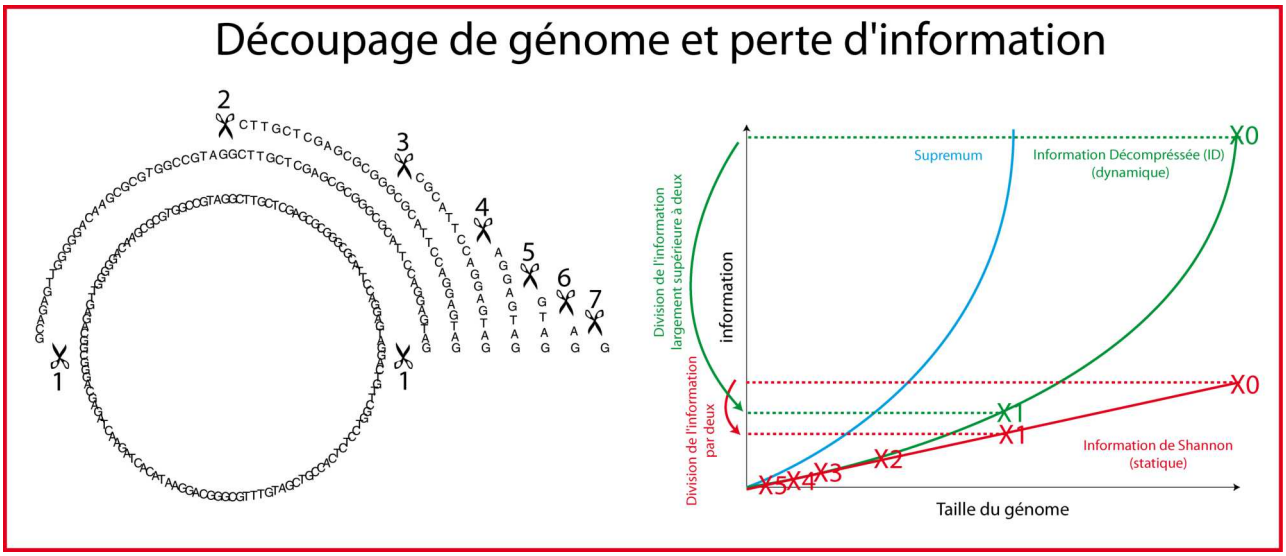



**Figure XII.6** – A gauche, on imagine que l'on découpe petit à petit un génome en deux brins de taille identique. A droite, on regarde comme évolue l'information de Shannon (en rouge) et l'Information Décompressée en vert au fur et à mesure des coupures. Le contenu en information de Shannon est divisé par 2 à chaque coupure quelque soit la coupure. A l'inverse, l'Information Décompressée (ID) semble être intuitivement divisée par un facteur gigantesque pour les coupures initiales, facteur qui diminue au fur et à mesure des coupures pour tendre vers 2 lorsqu'il ne reste plus que deux nucléotides.

D'où vient ID ? Et pourquoi la perte d'information décroit-elle, intuitivement, au fur et à mesure des découpes ? Pourquoi y a-t-il plus d'informations dans le génome d'*E. coli* que dans un génome aléatoire de même taille ? Parce que les cycles de réplications/mutations/sélection au cours de plusieurs milliards d'années d'évolution ont engendré des nucléotides avec une grande profondeur logique. *Cette profondeur logique se caractérise par des interactions/liens virtuels entre nucléotides.*

Je m'explique : commençons avec un cas simple (figure XII.7) :

Nous avons vu que le principe d'un réseau de régulation génique est simple : un gène A codant pour un facteur de transcription peut contrôler la transcription d'un autre gène B. Autrement dit, les nucléotides du gène A ont un lien invisible, virtuel avec les nucléotides du gène B. Ce lien ne peut apparaitre que lors de l'évaluation/l'exécution du programme (la vie de

la bactérie *E. coli*, la dynamique). Il n'empêche que ce lien virtuel existe, *a priori*, sur la séquence statique avant toute exécution. *Et c'est ce type de lien, inexistant dans une séquence aléatoire, qui génère un surplus d'information « à ajouter » à l'information minimale de Shannon.*

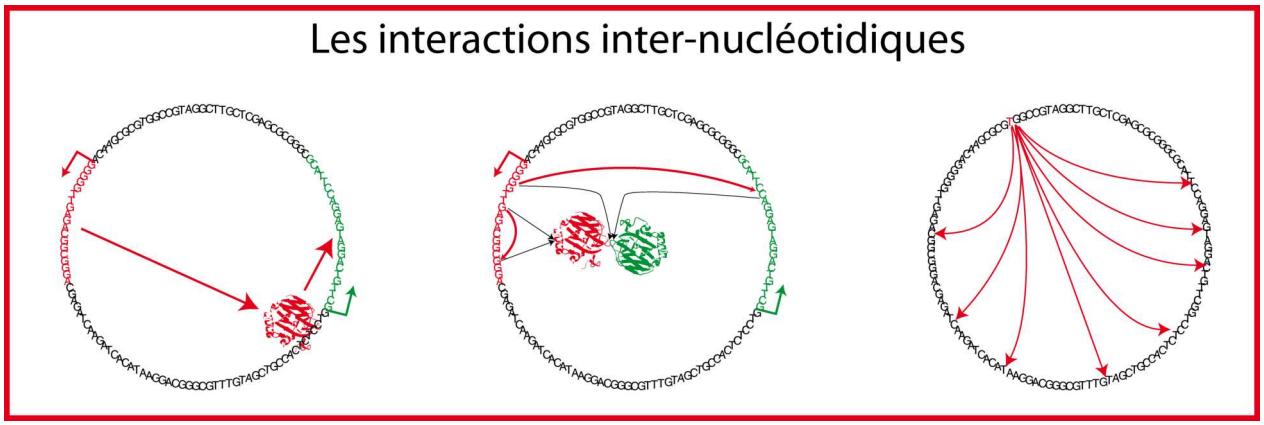



**Figure XII.7** – A gauche : les nucléotides du gène codant pour le facteur de transcription rouge ont un lien virtuel avec les nucléotides du gène vert dont l'expression est régulée par le facteur de transcription rouge. Au milieu : il faut envisager ces interactions au sens large : des nucléotides d'un même gène ou de deux gènes différents peuvent avoir une interaction entre eux si les acides aminés correspondant interagissent physiquement. A droite : au final, chaque nucléotide peut avoir de multiples liens virtuels avec les autres nucléotides.

Ces interactions ne se limitent pas aux interactions des réseaux de régulation génique : ce sont des interactions dans un sens plus large. Le nucléotide numéro 235 d'un gène a une interaction virtuelle avec le nucléotide numéro 747 si les deux acides aminés se retrouvent côte à côte lors du repliement et forment une liaison chimique. Même raisonnement si deux nucléotides appartenant à deux gènes différents se retrouvent, une fois sous forme d'acide aminé, en contact physique (par exemple lors de transduction du signal). Ce sont vraiment des interactions virtuelles au sens large : les nucléotides d'un gène codant pour une enzyme impliquée dans la biosynthèse des lipides ont une interaction virtuelle avec tous les nucléotides des gènes codant des protéines. En effet, la membrane (composée de lipides) limite physiquement la diffusion dans l'espace de ces protéines. Notez aussi que ces interactions virtuelles ont toujours une réalité physique lors de l'exécution du programme. Cette réalité physique c'est toujours, quoi qu'il arrive une liaison physique : le facteur de transcription touche vraiment le promoteur (et matérialise par là la liaison virtuelle entre les nucléotides correspondants), les acides aminés sont vraiment en contact dans la protéine repliée, les lipides empêchent vraiment les protéines de sortir.

Vous comprenez donc bien que le nombre d'interactions de ce type est astronomique (mais, il me semble, fini). En fait, chaque nucléotide a potentiellement un impact /une interaction

virtuelle/ un lien avec l'ensemble des autres nucléotides. C'est ce réseau dense de liens invisibles qui engendre l'information cachée dans la séquence. Dans une séquence aléatoire en bit, chaque bit est strictement indépendant des autres bits par définition. Ce n'est pas le cas dans une séquence génomique : les bits sont reliés par des fils invisibles. Le fil devient visible lorsque la liaison physique se matérialise et cela ne se produit que lors de l'exécution dynamique du programme.

Nous avons donc vu qu'à une séquence ADN en bits (en fait, une base ADN contient 2 bits ; on peut donc parler de séquence ADN en bases ou en bits) est associée une chaine de caractères également en bits, bien plus grande, et nommée **I**nformation **D**écompressé. A partir de là, qu'est ce qui m'empêche de postuler qu'il existe une fonction f tel que :

$$ID_{E.coli} = f(\textit{séquences du génome de E.coli}) \tag{XII.2}$$

Ainsi que la fonction réciproque (pas au sens mathématique ; ou du moins, je n'en sais rien)

$$\textit{Séquence du génome de E.coli} = f^{-1}(ID) \tag{XII.3}$$

Pour l'instant, pour avoir accès à ID, on décompresse l'information de la séquence à l'aide de la cellule elle-même. C'est ce que les biologistes font expérimentalement tous les jours. La cellule (la machine) interprète le programme (la séquence) ce qui crée des phénotypes. Ce sont ces phénotypes mis bout à bout qui représentent ID en quelque sorte : ID c'est la description la plus courte de l'ensemble des interactions virtuelles entre nucléotides *de telle sorte que les bits de cette description soient strictement indépendants* (ce qui n'est pas le cas dans la séquence ADN). Les publications scientifiques, les bases de données qui décrivent des interactions, des phénotypes sont une part (infime) d'ID. Evidemment, cette part infime ne représente pas la description « la plus courte » et pourrait donc être compressée. Avant de commencer ma thèse, j'ai été attiré par l'idée, un peu naïve, de créer une cellule « virtuelle » dans l'ordinateur c'est-à-dire de réaliser un modèle tridimensionnel d'une cellule qui soit le plus proche possible de la réalité. Ce type de modèle pourrait, dans l'idéal, décrire le fonctionnement du réseau et reproduire l'ensemble des phénotypes de la bactérie. Ce qui est important de constater, c'est que ce programme (l'algorithme) de cellule virtuelle aussi parfait qu'il puisse devenir, risque d'être, de l'ordre de grandeur d'ID c'est-à-dire potentiellement astronomique. Tous les modèles dynamiques actuels contiennent la description d'une part infime d'ID, pas la description de la séquence ADN !

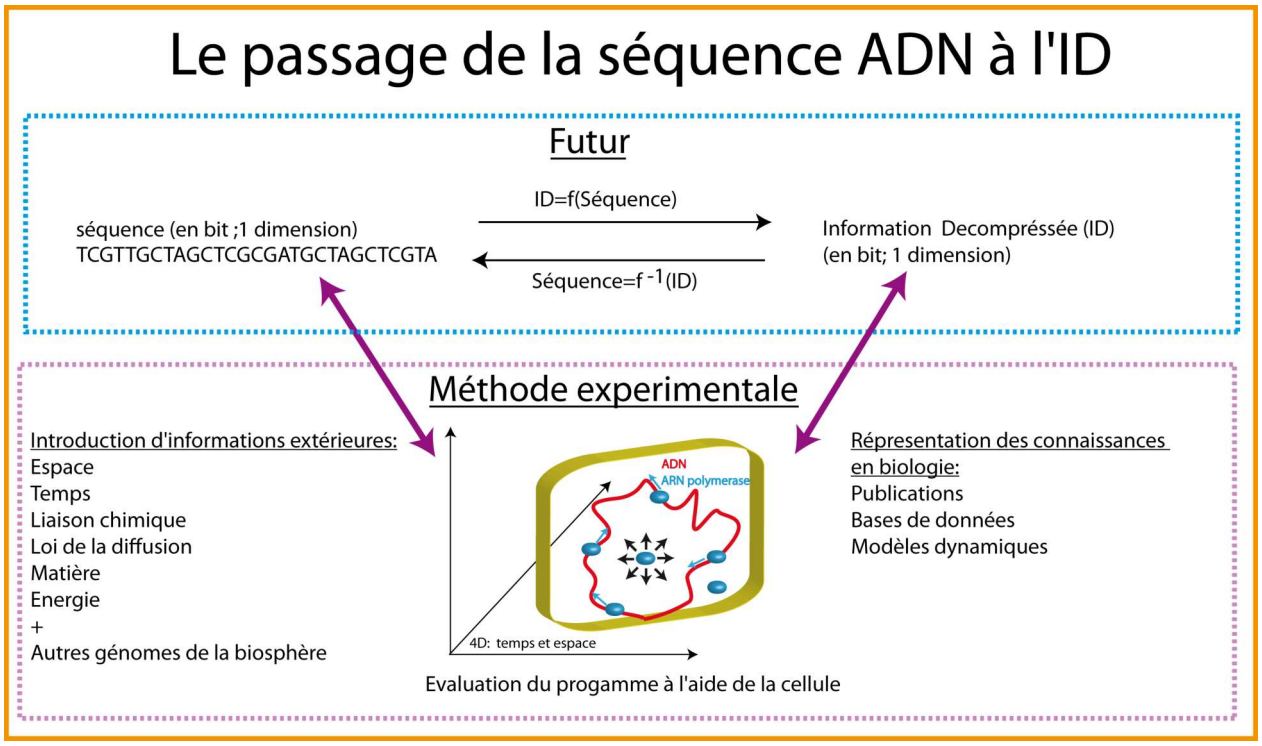



**Figure XII.8** – La séquence ADN est décompressée par les biologistes à l'aide de la cellule elle-même. Lors de cette décompression, il y a introduction d'informations extérieures. La décompression du programme apporte des informations qui appartiennent à l'ID. Nous représentons ces informations dans les publications, les bases de données et les modèles. Dans le futur peut être, une fonction nous permettra de connaitre l'ID étant donnée la séquence ADN et réciproquement mais cela est peut être tout simplement impossible pour des raisons théoriques qui restent à déterminer.

Je souhaite maintenant préciser un point important. Lorsque la cellule est utilisée pour exécuter le programme (la séquence), il y a introduction/injection d'informations étrangères au génome. Antoine Danchin appelle cette information l'*information du châssis* (Danchin, 2009). Je souhaite découper cette information en trois types (découpage arbitraire car les éléments peuvent être placés dans plusieurs catégories) :

⇒ Information d'origine physique : il s'agit du temps, de l'espace, les lois relative à la matière et l'énergie (liaisons chimiques), les lois de la diffusion et j'en oublie.

⇒ Information d'origine biologique : ou plutôt relative au milieu : PH, température, conditions initiales du contenu cellulaire. Ces dernières sont importantes pour les questions relatives au bruit (multi-stabilité).

⇒ Information d'origine métaphysique : via l'introduction de hasard (de libre arbitre !) lié aux échelles de l'infiniment petit (déplacement aléatoire des molécules selon le mouvement brownien)

La méthode qu'utilise la cellule pour compresser ID en séquence ADN devrait intéresser de plus prés la communauté scientifique et en particulier les mathématiciens car la « nature » semble disposer d'une capacité de compression de l'information prodigieuse. Rendez-vous compte en imaginant la taille gigantesque de l'ID des humains : cette ID est compressée en une séquence de 3 milliards de bases, ce qui tient sur un DVD. Or une simple « pensée humaine » pourrait être assimilée à un phénotype et donc faire partie de cette ID même si l'introduction de hasard dans le processus (et donc potentiellement d'information) me fait douter sur ce point mais cela à l'avantage de nous laisser un petit peu de libre arbitre.

Les mathématiciens ne devraient-t-il pas regarder de plus prés comment la nature s'y prend pour incorporer plus d'information que celle de Shannon dans une chaine de caractère ? Ne devraient-ils pas s'intéresser de plus prés à ce que représentent ces interactions virtuelles entre nucléotides/bits ? Quel est le socle théorique nécessaire pour pouvoir procéder à ce type de compression ? Par exemple, la machine capable de compresser et décompresser doit-elle forcement avoir des caractéristiques similaires à celles de la cellule ? C'est-à-dire :

⇒ calcul parallèle et probabiliste ?

⇒ support/réalité physique (les interactions virtuelles doivent se concrétiser dans la réalité physique cellulaire : c'est la liaison/le contact physique entre entités) ?

Cette question du support physique me semble importante : peut-on réellement imaginer qu'une machine de Turing (ordinateur) puisse prendre en argument la séquence et ressortir l'ID en output et réciproquement, ce qui revient, d'une certaine manière, à by-passer, la réalité physique ?

J'ai proposé ci-dessus une petite application qui permettrait de compresser ou décompresser une quantité d'information astronomique en essayant de reproduire la méthode qu'utilise la nature. Mais cette petite application est l'arbre qui cache la forêt. Les conséquences sur l'homme et plus généralement sur la vie si cette fonction (et sa réciproque) existait et si on la découvrait (ou si on l'approchait) serait gigantesques : ces conséquences dépassent l'entendement. Par exemple, l'utilisation de cette fonction nous permettrait d'accroitre nos capacités cognitives de manière prodigieuse ce qui permettrait de « construire » de nouveaux objets mathématiques pour l'instant non constructibles par nos cerveaux actuels. Mais je réserve cette partie beaucoup trop métaphysique pour le chapitre sur le futur lointain.

Il me faut maintenant rentrer un peu plus en profondeur, là où se logent les difficultés. Là où le lecteur risque de subir le brouillard qui se loge dans ma tête. Il ne devrait se lancer dans la lecture des paragraphes qui suivent que si la lecture des paragraphes précédents lui a semblé relativement facile. Vous voilà « *informé* » ⌣̈.

J'identifie deux questions difficiles (où je ne peux fournir que des éléments de réponses incomplets voire faux)

⇒ Quels sont les objets mathématiques qui décrivent les interactions/liens virtuels entre nucléotides ?

⇒ La fonction qui transforme une séquence ADN en ID et sa réciproque sont-elles vraiment calculables (récursives) ?

## 6.1 Quels sont les objets mathématiques qui décrivent les interactions/liens virtuels entre nucléotides ?

Pour répondre à cette question, il me faudrait sans doute l'aide d'un mathématicien. Je donne un certains nombres de pistes sans connaitre la bonne réponse et sans même savoir s'il y en a une (figure XII.9).

Un nucléotide a potentiellement un lien avec l'ensemble des nucléotides du génome. On peut donc imaginer, pour chaque nucléotide, l'existence d'un vecteur de taille [taille(génome)-1] dont chaque élément de ce vecteur décrit l'interaction de ce nucléotide avec un autre nucléotide (cette interaction peut très bien être nulle, inexistante). Cet élément, cette interaction particulière, c'est une fonction qui prend en argument l'ensemble de la séquence. Car la manière dont un nucléotide interagit avec un autre nucléotide dépend des autres [taille (génome)-2] nucléotides. Vous voyez que les capacités d'abstraction sont mises à rude épreuve. Premièrement, je suis gêné par l'incertitude concernant la présence ou l'absence de récursivité dans le phénomène que je décris. Deuxièmement, j'ai des difficultés à voir si un nucléotide a potentiellement un lien avec l'*ensemble des nucléotides du génome* ou si un nucléotide a potentiellement un lien avec *l'ensemble des ensembles des nucléotides du génome*. Enfin, le génome, à l'évidence, encode des phénomènes de multi-stabilité (voir le chapitre sur le bruit) et je ne sais pas si ceux-ci représentent une difficulté supplémentaire.

## 6.2 La fonction qui transforme une séquence ADN en ID et sa réciproque sont-elles vraiment calculables (récursives) ?

La manière de formaliser les liens virtuels entre nucléotides permet de postuler l'existence d'un supremum à l'information décompressée d'un génome que nous appellerons IDM (**I**nformation **D**écompressée **M**aximum). Cette IDM correspond au nombre maximum atteignable d'interactions entre nucléotides.

Toute séquence A de n bits (compressée ; bits non indépendants les uns des autres) associe

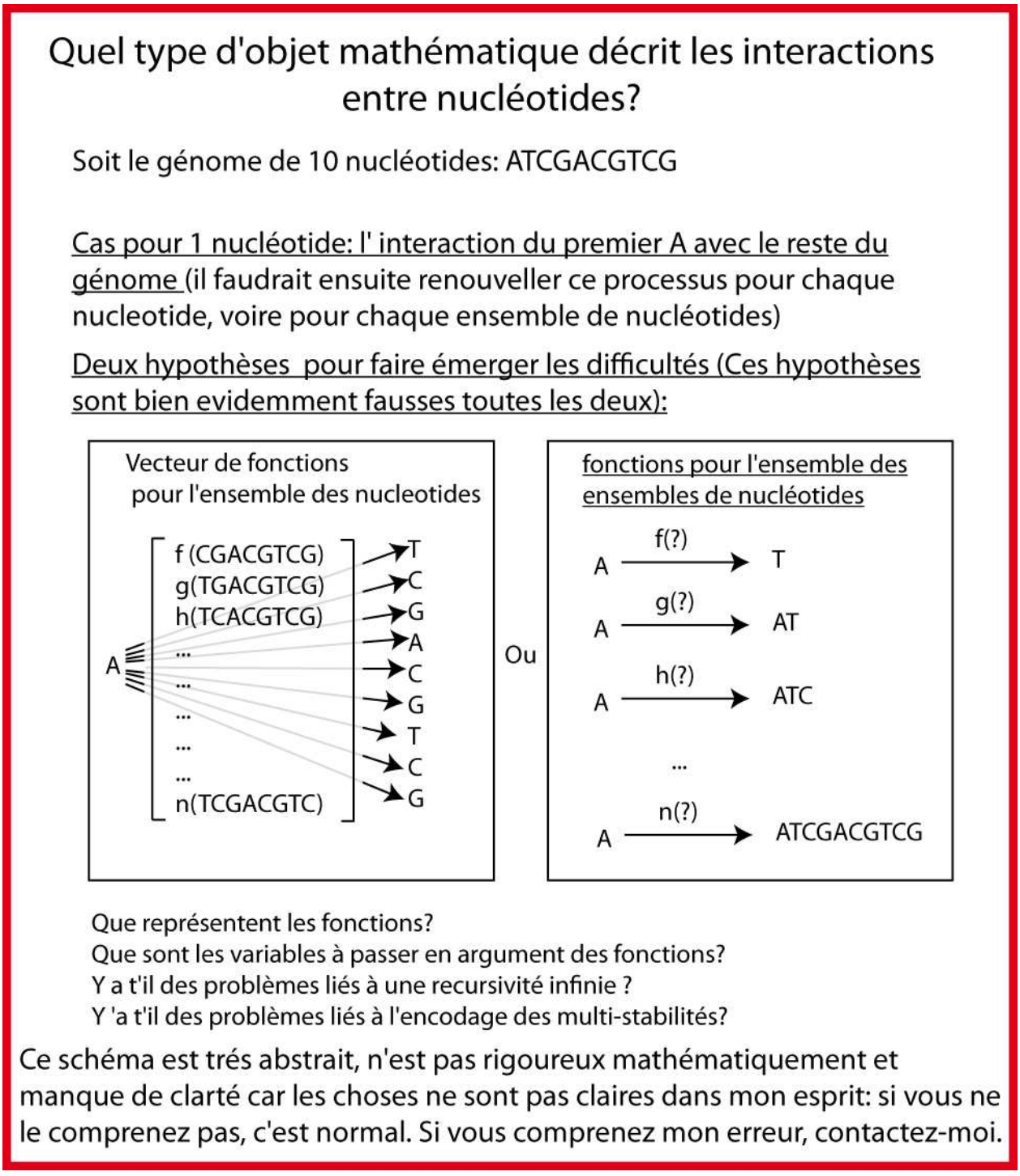



**Figure XII.9** – Le malaise.

une séquence B (nommée IDM) de m bits (décompressée ; bits indépendants ; m»n) tel qu'il est impossible de rajouter de l'information (des bits) dans B sans devoir en rajouter dans sa version compressée A.

Notez que l'ID d'un génome donné est, le plus souvent, bien inférieur à l'IDM : toutes les interactions possibles entre bits ne sont pas forcement utilisées dans un génome donné : la « matrice peut très bien être creuse ». De plus, pour deux génomes de même taille (approximativement), par exemple le chimpanzé et l'homme, nous ressentons bien (d'ailleurs un peu prétentieusement) que l'ID de l'homme est supérieure à l'ID du chimpanzé, à cause des facultés des deux cerveaux qui sont très différentes. Cela montre bien que pour une séquence x d'une taille donnée, il existe un IDM tel que tout génome (passé, présent et futur) de taille x a une ID inférieure ou égale à l'IDM. Notez que si l'IDM n'existait pas, quel intérêt aurait la nature à augmenter la taille des génomes pour augmenter la complexité des organismes ? (taille du génome de *E. coli* : 4 millions de bases ; taille du génome humain : 4 milliards de bases soit une différence de taille d'un facteur 1000). Il suffirait à la nature de garder la taille des génomes constante et de se contenter d'accroitre la profondeur logique des nucléotides, d'accroitre indéfiniment le nombre d'interactions virtuelles entre nucléotides. On vivrait dans un monde où le génome d'*E. coli* et celui de l'homme feraient la même taille mais où celui de l'homme serait beaucoup plus profond, contiendrait bien plus d'informations cachées.

Mais pourquoi m'acharner à vous persuader (je n'ai pas la prétention de démontrer quoi que ce soit) de l'existence de l'IDM ?

Il me semble que l'IDM, par définition, correspond à l'opposé/l'inverse/la réciproque (je ne suis pas très à l'aise avec ces notions) de la complexité de Kolmogorov.

Etant donné que l'IDM est l'information décompressée maximale, cela signifie que son programme le plus petit, c'est à dire l'information compressé (la séquence ADN si vous préférez) représente dans notre cas la complexité de Kolmogorov.

Plus l'ID d'un génome est proche/tend vers l'IDM, plus cela signifie que la séquence ADN correspondante est proche /tend vers la complexité de Kolmogorov.

Vous voyez donc que ces différents objets que je vous ai décrits, entretiennent des liens étranges avec les 3 théories de l'information : la séquence compressée (ADN) et l'IDM sont toutes les deux des chaines de caractères mesurables en bits (selon Shannon). Le passage de la séquence compressée à l'IDM se fait en faisant appel à la profondeur logique de Bennett mais le passage inverse est étrange car la séquence compressée, c'est par définition la complexité de Kolmogorov vis-à-vis de l'IDM.

Or cela pose un problème. En effet, la fonction qui prend une séquence (IDM) en argument

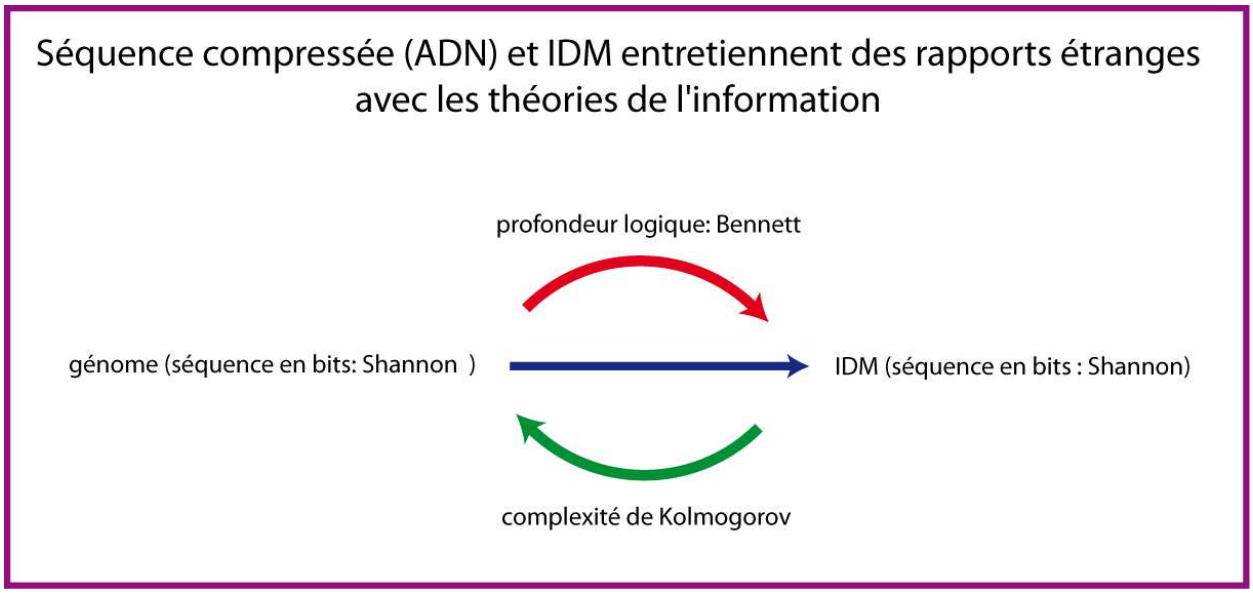



**Figure XII.10** – Les rapports étranges qu'entretiennent les trois théories de l'information.

et renvoie la complexité de Kolmogorov (la séquence ADN compressée) est une fonction non calculable par une machine de Turing (un ordinateur) car, nous l'avons vu, elle est non récursive.

Je ne sais pas exactement ce qu'il en est de la fonction opposée prenant en argument la séquence compressée ADN et renvoyant l'IDM : est-elle calculable ? Car la profondeur logique, elle aussi, est non récursive.

A partir de là, il y a deux manières d'envisager les choses :

⇒ Les biologistes vont se heurter à la théorie de la calculabilité. Nous ne pourrons peut-être jamais déterminer/mesurer l'ID ou l'IDM d'une séquence ADN en utilisant une fonction ni déterminer la séquence ADN à partir de l'ID ou l'IDM. Cela signifierait que l'objet que je ressens comme « mathématique » décrivant les interactions virtuelles entre nucléotides est soit inaccessible, soit inexistant. Ou en tous cas, il ne peut pas être « utilisé/traité » pour aboutir à la compression ou la décompression. Dans ce cas de figure, même la nature se heurte à la théorie de la calculabilité : l'ADN (l'immense compréhension de l'information) représenterait alors l'aboutissement de milliards d'années d'évolution, de milliards d'années de recherche de la séquence la plus courte possible étant donnée une ID dans le but de tendre vers la complexité de Kolmogorov.

⇒ l'autre manière de voir les choses consiste à dire que le fonctionnement de la « nature » diffère d'une machine de Turing (d'un ordinateur) dans sa manière de compresser et décompresser l'information. Cette même nature est capable de « by-passer » la théorie de la calculabilité, elle peut compresser au maximum et décompresser au maximum une

information de manière « pseudo-récursive » via un type de machine, de fonctionnement qui nous échappe.

A vrai dire, je penche plutôt pour le premier point, plus respectueux des paradigmes d'informatique et de biologie. Cependant j'ai le sentiment qu'à partir d'une séquence donnée, on devrait pouvoir connaitre la quantité d'information maximum injectable dans cette séquence car cette quantité ne me semble pas infinie. Or si on peut faire le chemin dans un sens via un objet mathématique, je ne vois pas de raison, à première vue, de ne pas pouvoir le faire dans l'autre sens [5].

Voila, j'en ai terminé avec l'information. Vous l'aurez compris, je suis maintenant perdu en pleine jungle. Des aventuriers suivront-ils ma piste ? Ou flaireront-ils au contraire dés le départ qu'elle ne mène nulle part ? A vrai, dire, me reposant maintenant contre un arbre dans cette forêt vierge, je m'attends plutôt à ce que quelques singes rieurs se moquent gentiment de moi, ou qu'un boa vienne tordre le cou à mes allégations. Tel est le risque lorsque l'on s'aventure seul dans la jungle. Mais c'est un risque à prendre.

[5] Il y a encore ici un risque de faiblesse du raisonnement mathématique (problème relatif aux notions d'injection/surjection ?)

# Chapitre XIII

# Qu'est-ce que la vie

## 1 Préambule

Il y a eu, au cours de l'histoire, de multiples tentatives pour essayer de définir la vie, ou au moins, en extraire les caractéristiques, les propriétés essentielles. Pourtant, aujourd'hui, le consensus en matière de définition de la vie c'est qu'il n'y a pas de consensus[1]. Certains argumentent même qu'il est inutile de rechercher cette définition[2]. Je pense au contraire que les scientifiques et philosophes devraient persévérer dans cette voie au fil des découvertes et avancées en biologie. S'amuser à rechercher une définition/ des caractéristiques essentielles de la vie est un petit jeu sans prétention mais formateur pour l'esprit. Et cela influence forcement, de prés ou de loin, la manière d'envisager, de penser et d'élaborer des expériences. Michel Morange va encore plus loin, il écrit :

> *« Une telle absence de curiosité de la part de la majorité des biologistes n'est pas gênante à court terme. On peut faire un très bon travail de biologie sans s'interroger sur ce qu'est la vie. A long terme, elle est probablement plus dommageable : elle est la marque d'une certaine paupérisation intellectuelle du milieu scientifique, préjudiciable à sa créativité »*[3]

A quoi pourrait bien servir une définition de la vie ?

Cette définition pourrait avoir des applications directes dans le domaine de la vie artificielle. En effet, à partir de quand devra-t-on considérer que tel robot, tel programme informatique est vivant ? Sur quel critère se baser ? L'automate « canard de Vaucanson » est-il vivant ? Non assurément mais qu'en est il des robots tortues de William Grey Walter ? Et les cellules du

[1] Carl Sagan, *Definitions of life.*

[2] Edouard Machery, *Why I Stopped Worrying About the Definition of Life... And Why You Should as Well*, 2010.

[3] Michel Morange, *La vie expliquée*, Odile Jacob p.180

jeu de la vie de John Conway ou les fourmis de Chris Langton ? Y aura-t-il un jour où le plus sceptique des sceptiques n'arrivera plus à distinguer ces programmes d'un organisme vivant ?

Cette définition nous aiderait aussi en biologie synthétique. En effet, certaines équipes essaient de recréer des cellules *de novo* : on parle de « protocellules » (vésicule lipidique contenant des molécules biologiques et capable d'exprimer des gènes). A partir de quand devra-t-on considérer ces « protocellules » comme des cellules vivantes ? Certains travaillent déjà à cette question et il a été proposé qu'une cellule artificielle devrait être considérée « vivante » à partir du moment où de vraies cellules n'arrivent plus à la distinguer d'une cellule *normale*.

Enfin cette définition de la vie nous aiderait en exobiologie (astrobiologie) pour rechercher la vie extraterrestre sous quelque forme que ce soit. Cette discipline a besoin de savoir quelles sont les caractéristiques de la vie pour décider ce qu'il faut rechercher et donc quel type d'instrument construire. Notez d'ailleurs que la recherche de vie extra-terrestre part souvent du postulat d'une vie basée sur le carbone ou l'oxygène. Une nouvelle discipline, la xénobiologie, essaie de dépasser cette limitation anthropocentrique appelée « chauvinisme du carbone » en développant de nouvelles hypothèses plus audacieuses sur ce que pourraient être des formes alternatives de vie extra-terrestre.

On voit donc que la recherche de la définition de la vie a de nombreuses applications immédiates. Je précise, même si cela est évident, que cette définition a également des conséquences morales (et donc politiques) que j'ai volontairement éludées ici.

Je vais donc me lancer dans l'étude des caractéristiques et définitions de la vie. Je me sens autorisé à cette étude car derrière le fonctionnement d'un réseau de régulation intracellulaire se cache le dogme/l'hypothèse de l'émergence de la vie. Il me faut donc en savoir un peu plus sur cette dernière. Mon plan sera découpé en 4 parties.

Dans la première partie, je « listerai » l'ensemble des vues/ caractéristiques/ définitions de la vie avec pour but une *relative* exhaustivité (beaux antonymes !). Ces vues/ définitions/ caractéristiques, je les ai puisées et notées au fur et à mesure de mes lectures chez un certain nombre d'auteurs et elles sont retranscrites parfois de manière exacte et parfois en substance[4]. Dans la seconde partie, j'opérerai un virage métaphysique (déjà bien entamé). Dans la troisième partie, je proposerai un angle de vue personnel (une spéculation/une construction cognitive faisable) sur la question « qu'est-ce que la vie ? ». Enfin, la dernière partie abordera la grande oubliée à savoir la mort sous un regard philosophique.

[4] J'ai un peu honte de ne pas être en mesure d'associer chaque auteur à chaque vue. Voici quelques noms qui me viennent à l'esprit : Edgar Morin, Michel Morange, François Jacob, Jacques Monod, Georges Canguilhem, Antoine Danchin, Bichat, Aristote, Bernd-Olaf Küppers, M. Perret, Stuart Kauffman, Humberto Maturana, Francisco Varela, Ali Saïb, Haeckel, Nietzsche, Barbara Stiegler et Wikipedia.

## 2 La vie perçue d'un point de vue scientifique

⇒ *« Vie : le fait de se nourrir, de croître et de dépérir par soi même »*

⇒ *« Vie : ensemble des fonctions qui résistent à la mort »*

⇒ *« Mort : menace permanente et inconditionnelle de décomposition de l'organisme »*

⇒ *« Mort : le cadavre n'est autre que le corps vivant retombé sous l'empire des forces physiques. »*

⇒ *« La différence entre vie et non vie n'est pas à chercher dans une différence d'atomes (constituants) mais dans une différence de complexité »*

⇒ *« Les être vivants sont en déséquilibre thermodynamique car ils échangent en permanence matière et énergie avec le milieu qui les entoure »*

⇒ *« La vie est un système ouvert , capable de s'auto-entretenir, formé de réactions chimiques capables de s'auto-entretenir, formé de réactions chimiques organiques , catalysées pas à pas et à des températures presque ambiantes par des catalyseurs organiques complexes et spécifiques qui sont eux même produits par le système »*

⇒ *« Le phénomène vital est un effet produit par le contact entre l'organisme vivant et le milieu qui l'entoure »*

⇒ *« Deux conceptions différentes s'opposent au cours de l'histoire :* ***pour comprendre la vie, il faut décrire les constituants*** *opposé à* ***pour comprendre la vie, il faut comprendre les interactions (et leurs natures) entre ces constituants.*** *»*

⇒ *« Les grosses molécules caractérisent le vivant »*

⇒ *« Contrairement aux corps inanimés, les corps vivants sont indissolublement liés au temps. Chez eux aucune structure ne peut être détachée de l'histoire. »*

⇒ *« Le propre de la vie par rapport à la matière inorganique c'est la mémoire. »*

⇒ *« C'est la possibilité d'accumuler de l'expérience, c'est-à-dire de réduire à l'identique, grâce aux concepts, des excitations pourtant différentes qui est le problème propre de l'organique. »*

⇒ *« La vie implique l'existence d'une structure spatio-temporelle »*

⇒ *« La notion du vivant est une notion dynamique, évoluant en fonction de nos connaissances. En conséquence, la frontière entre la matière inerte et le vivant est tout aussi instable »*

⇒ *« Vie : capacité de se reproduire avec variation »*

⇒ *« La stabilité de la vie, condition sine qua non de son entrée dans l'histoire, est liée de manière absolue à sa capacité à se reproduire »*

⇒ *« Il n'y a jamais naissance vie et mort mais plutôt vie et mort d'un individu car il y a*

*toujours eu continuité dans la transmission de la vie »*

⇒ *« Une des caractéristiques de la vie est la détection et l'adaptation aux propriétés de son environnement »*

⇒ *« La vie émerge du réseau de régulation lui-même encodé dans le génome »*

⇒ *« La mort est comme une nécessité prescrite dès l'œuf par le programme génétique lui-même ; résister à la mort tel est alors le propre du vivant »* [Et oui la mort ne serait pas subie par la vie mais codée par elle-même. Voyez ici le paradoxe de la vie : la mort serait déjà encodée dans le programme génétique par la vie/la nature via l'évolution alors même que l'organisme (la vie/la nature) est essentiellement encodé pour résister à la mort]

⇒ *« Être vivant : triple flux de matière énergie et information »*

⇒ *« La vie navigue à la frontière entre ordre et chaos »*

⇒ *« Un organisme vivant est un ensemble organisé de matière qui tend à se maintenir à l'état homéostatique par une utilisation concertée d'énergie. »*

⇒ *« Le parchemin de la vie est plus imprégné de soupe primitive, de sources hydrothermales ou d'argiles catalytiques que de molécules informationnelles ou de replicateurs autonomes. »*

⇒ *« L'erreur est liée à la vie, et donc à la mort. A tous les niveaux, une trop grande quantité d'erreurs entraîne la mort »* [Par exemple, un animal entend un craquement de branche, analyse la situation, décide qu'il n'y a pas de danger et ne fuit pas. Si ce craquement est produit par un prédateur, alors son erreur lui sera peut être fatale. L'erreur peut aussi être interprétée au sens d'erreur lors de la réplication de l'ADN, erreur de repliement de protéine, erreur de détection d'un signal. . . Ainsi la vie comporte d'innombrables processus de détection, de refoulement de l'erreur.]

⇒ *« Y a-t-il un caractère envahissant de la vie ? »*

⇒ *« Existe-t-il une division franche entre vivant et non vivant ? »*

⇒ *« Existe-t-il une division franche entre le passage temporel de vie vers la mort ? »*

A défaut de disposer d'une définition totalement consensuelle, il existe tout de même un consensus sur les propriétés ci-dessous :

Un organisme vivant :

⇒ évolue. Cependant il existe aussi des contre arguments : si en utilisant les technologies futures, l'homme décide de maintenir son génome constant, alors il n'évolue plus (au sens classique) mais il reste néanmoins vivant.

⇒ dispose d'un métabolisme (consommation, transformation et stockage d'énergie ou de masse). Cela permet d'exclure les virus et prions qui ne sont pas considérés comme vivant

par la majorité des scientifiques.

⇒ se reproduit (créer de façon autonome d'autres entités similaires ou identiques à soi-même). Mais il existe ce fameux contre argument : *« La capacité des êtres vivants à se reproduire est seconde car un lapin isolé est bien vivant »*. On peut utiliser le même argument avec n'importe quel animal stérile.

Voici un petit nuage de tags de synthèse sur la vie :

métabolisme – flux – energie – homéostasie – temps – espace – erreur – stabilité – frontière – ordre – chaos – transition de phase – matière – information – programme – code – complexité – interaction – réseau – reproduction – évolution – adaptation – milieu

J'espère que le lecteur aura pu puiser dans cette synthèse ce qu'il est venu y chercher, peut-être une certaine inspiration ou un simple éclaircissement.

# 3 La vie perçue d'un point de vue philosophique

Il me vient ici à l'esprit principalement deux thèmes à développer mais cela n'est absolument pas exhaustif. Il s'agit des notions de téléologie et de volonté de puissance.

## 3.1 La téléologie

La vie a-t-elle une finalité, un but, un projet que nous n'aurions pas encore découvert ?

Beaucoup de gens pensent que justement *la vie se définit par son absence de finalité* et j'avoue que je reconnais un charme certain à cette manière d'envisager les choses. Mais cela nous arrange bien, nous scientifiques, car la méthode scientifique nous interdit formellement le recours à la téléologie en biologie : c'est-à-dire l'explication des phénomènes vivants en termes de but. Ainsi, selon une boutade de François Jacob *« longtemps le biologiste s'est trouvé devant la téléologie comme auprès d'une femme dont il ne peut pas se passer mais en compagnie de qui il ne veut pas être vu en public »*. Et Jacques Monod remarque bien le problème profond que cela pose. Il propose le terme de téléonomie (concept scientifique de téléologie) c'est-à-dire la qualité de la matière vivante en tant qu'elle matérialise un projet.

> *« L'objectivité cependant nous oblige à reconnaître le caractère téléonomique des êtres vivants, à admettre que dans leurs structures et performances, ils réalisent et poursuivent un projet. Il y a donc là, au moins en apparence, une contradiction*

> *épistémologique profonde. Le problème central de la biologie, c'est cette contradiction elle-même, qu'il s'agit de résoudre si elle n'est qu'apparente, ou de prouver radicalement insoluble si en vérité il en est bien ainsi. »* [5]

Ainsi, il nous est interdit de penser, en biologie, en termes de *finalité* et je ne suis pas sûr que l'on ait le droit de penser en termes de *fonction*. Car tout de suite on est accusé soit de faire des anthropocentrismes grossiers, soit de faire de la métaphysique. Voyons des exemples pour être plus clair.

Une enzyme convertit un substrat en produit. Nous n'avons pas le droit de dire que la finalité de l'enzyme est de convertir tel substrat en tel produit ni d'ailleurs que la fonction de l'enzyme est de convertir tel substrat en tel produit. A cause (ou grâce) à la théorie de l'évolution, une enzyme n'a pas de but et n'a pas de fonction. Les buts et les fonctions sont réservés aux hommes et aux dieux.

Or je suspecte que certains anthropocentrismes cachent (et donc contiennent) plus qu'un simple biais cognitif sans intérêt. Ce n'est pas un hasard si nous, pauvres biologistes humains, voulons et aimons attribuer une finalité à une enzyme (il fallait absolument pouvoir utiliser le lactose comme source d'énergie), attribuer une finalité au lymphocyte T (il fallait absolument se défendre contre cette bactérie pathogène), à la peau (il fallait absolument disposer d'une barrière imperméable et protectrice), à l'aile (la fonction/finalité d'une aile c'est de voler). Or si nous pouvons attribuer une finalité, une fonction à tous les constituants biologiques (molécules, cellules, organes) en dessous de notre échelle, et quand bien même cette finalité est « une vue de l'esprit » à cause de la théorie de l'évolution, peut-on imaginer que l'organisme entier et même la biosphère entière dispose aussi (tout comme les niveaux inférieurs) d'une finalité « vue de l'esprit » (qu'importe qu'elle soit un anthropocentrisme !) ? Finalité qui ne nous sauterait pas aux yeux car nous en serions un maillon ?

J'aime bien la métaphore de la montre : celle-ci connaît (pourrait connaître) son mécanisme parfaitement, elle ne connaitra jamais sa finalité : donner l'heure à l'homme. Le même type de parallèle peut-il être appliqué à l'homme qui, dans le futur peut être, connaitra sur le bout des doigts son fonctionnement sans jamais être en mesure de connaître son but ?

Quel pourrait être le but, le projet, la finalité de tel ou tel organisme ?

Quel pourrait être le but, le projet, la finalité de la biosphère entière ?

Et si la finalité de tout organisme, de la biosphère était « la puissance » ?

[5] Jacques Monod, *Le hasard et la nécessité*, Points p. 37–38.

## 3.2 La volonté de puissance

Selon Nietzsche, la vie se définit par la « volonté de puissance » :

> *« Partout où j'ai trouvé quelque chose de vivant, j'ai trouvé de la volonté de puissance ; et même dans la volonté de celui qui obéit j'ai trouvé la volonté d'être maître. [...]*
> *Et la vie elle-même m'a confié ce secret : « Voici, m'a-t-elle dit, je suis ce qui doit toujours se surmonter soi-même ». [...]*
> *Il n'a assurément pas rencontré la vérité, celui qui parlait de la « volonté de vie » , cette volonté n'existe pas.*
> *Car : ce qui n'est pas ne peut pas vouloir ; mais comment ce qui est dans la vie pourrait-il encore désirer la vie !*
> *Ce n'est que là où il y a de la vie qu'il y a de la volonté : pourtant ce n'est pas la volonté de vie, mais — ce que j'enseigne — la volonté de puissance. »*[6]

La vie doit exploiter **un potentiel** :

> *« vivre c'est essentiellement dépouiller, blesser, dominer ce qui est étranger et plus faible, l'opprimer, lui imposer durement sa propre forme, l'englober et au moins au mieux l'exploiter. [...] L'exploitation n'est pas le propre d'une société vicieuse ou d'une société imparfaite et primitive : elle est inhérente à la vie dont elle constitue une fonction primordiale, elle découle très exactement de la volonté de puissance, qui est la volonté de vie »* [7].

> *« C'est un fait que nous-mêmes croissons, changeons continuellement, rejetons nos vieilles écorces, muons à chaque printemps, ne cessons de devenir plus jeunes, plus à venir, plus hauts, plus forts, enfonçons toujours plus vigoureusement nos racines dans les profondeurs* **[Le potentiel dont je parle]** *—dans le mal— tout en embrassant simultanément le ciel toujours plus amoureusement, plus largement, et en aspirant toujours plus avidement en nous sa lumière, de toutes nos branches et de toutes nos feuilles. Nous croissons comme des arbres —cela est difficile à comprendre comme toute vie !— [...] tel est notre sort, comme on l'a dit : nous croissons en hauteur [...] la fatalité de notre hauteur, notre fatalité. »*[8]

La finalité (anthropocentrique) d'un organisme ou de la biosphère pourrait-il être la puissance ? Mais alors que serait cette puissance ? Comment définir cette puissance que rechercherait

[6] *Ainsi parlait Zarathoustra/Deuxième partie/De la victoire sur soi-même.*
[7] Nietzsche, *Par delà bien et mal*, Folio p. 182.
[8] Nietzsche, *Le gai savoir*, Flammarion p. 336.

la vie ? Il est temps maintenant de passer à mon « essai » (au sens de tentative) pour définir la vie. Je le répète : il ne s'agira que de simples spéculations/constructions cognitives inoffensives. Je ne prétends rien démontrer.

# 4 La conjecture des exemplaires

Certains scientifiques cherchent « le génome minimal » c'est-à-dire l'ensemble minimum de gènes (par exemple 250 gènes) permettant à un organisme (une bactérie) de se reproduire de manière autonome. Michel Morange parle de manière humoristique de PGCD (le plus grand commun diviseur) de la vie. Il voit aussi que ce projet ne nous aidera probablement pas dans la quête à la définition de la vie car *« rechercher l'organisme vivant le plus simple est peut être aussi vain qu'essayer de trouver le plus petit parmi les géants »* [9]. Selon Morange, *« l'erreur c'est de chercher aux états limites et de chercher à définir la vie a partir d'un seul organisme isolé alors qu'il faut regarder l'écosystème tout entier »*. En effet, *« un être vivant ne peux survivre que parce qu'il utilise le travail effectué par d'autre. [. . .][C'est une] vision globale écologique de la vie »* [10] .

Prenons un exemple, on peut enlever, par des méthodes de génie génétique, les gènes de biosynthèse des bases chez E. coli. Celle-ci ne pourra alors plus fabriquer les fameuses bases A, T, C, G nécessaires à la réplication et à la transcription. Par conséquent, elle ne pourra plus « vivre » sauf si on rajoute ces bases dans le milieu de culture (ce qui s'appelle une complémentation). Ce sont donc *les gènes fantômes* de l'organisme qui a produit les bases, récupérées par l'industriel qui nous les vend sous forme de poudre, qui permettent à notre souche d'*E. Coli* « OGM » de vivre. De même que les gènes fantômes des végétaux (qui rejettent notre si précieux oxygène dans l'atmosphère et nous fournissent les molécules organiques de notre alimentation) sont indispensables à la vie de l'humanité et donc complètent son génome.

Toute cette démonstration pour en arriver à un point. Il faut selon moi dissocier la définition de l'organisme de la définition de la vie. Ce sont deux choses différentes [11]. On peut définir un organisme avec la faculté de reproduction, d'évolution, la présence d'un métabolisme. Ce sont nos organismes bien connus sur terre. On peut, à souhait, adapter la définition pour inclure ou exclure telle ou telle entité (virus, prion, robot. . .). Pour ce qui est de définir la vie, je propose de procéder différemment. Il faut regarder dans un endroit où il n'y en a pas (la planète Mars) et regarder précisément *ce qu'il n'y a pas*. Regardez la photo ci-dessous. Que manque-t-il ?

[9] Michel Morange, *Les secrets du vivant*, La découverte p. 230.
[10] Michel Morange, *Les secrets du vivant*, La découverte p. 142.
[11] Cette idée n'est pas nouvelle. Elle était déjà connue (ou proposée ?) par Ernst Haeckel il y a plus d'un siècle.



**Figure XIII.1** – Sol jonché de rochers volcaniques vu par Mars Pathfinder le 8 septembre 1999 (Article Wikipedia : Mars ; image de la NASA)

Sur Mars, il n'y a pas de molécules trop complexes : Parmi les molécules les plus grosses, on trouve la serpentine ($(Mg,Fe)_3Si_2O_5(OH)_4$) ou le talc ($Mg_3Si_4O_{10}(OH)_2$)) qui contiennent tout au plus une trentaine d'atomes[12]. Jusqu'ici, rien de nouveau : on sait bien qu'une des caractéristiques de la vie c'est sa capacité à fabriquer et posséder des molécules énormes : l'ADN de l'homme, une des plus grandes molécules sur terre, fait 3.2 milliards de nucléotides et chaque nucléotide contient au moins une quarantaine d'atomes soit une molécule de l'ordre de 100 milliards d'atomes.

J'aimerais maintenant que l'on « switch » en mode « information ». C'est-à-dire que l'on passe des molécules aux bits. Je ne connais pas vraiment de méthode : c'est un switch purement conceptuel et donc critiquable. J'avoue ne pas savoir si cela est réellement possible car, comme le dit René Thom « *on n'échappe pas au continu* » [13]. Admettons qu'à la molécule la plus complexe sur Mars (30 atomes) correspondent approximativement 30 bits. Et à la molécule la plus complexe (sur terre), un simple brin d'ADN humain de 100 milliards d'atomes, correspondent approximativement 100 milliards de bits.

Ma molécule de 30 atomes sur Mars est donc transformée en une chaine de trente 0 et 1, les uns à la suite des autres. Cette suite de 0 et 1, je l'appellerai séquence dans la suite de mon exposé et elle n'aura rien à voir avec une séquence ADN. Voici par exemple une séquence de 30 bits :

011000110011101010010100010100

[12] Je n'en suis pas sûr : ma source est encore l'article de Mars sur Wikipedia.
[13] René Thom, *Prédire n'est pas expliquer*, Champs p. 67.

Pour le simple brin d'ADN humain, on aurait une séquence de 100 milliards de bits ce qui tiendrait tout juste sur mon disque dur externe.

Maintenant, opérons la même transformation conceptuelle pour Mars qui devient une séquence de 0 et de 1. Cette séquence contient *toute l'information* contenue dans Mars. Appelons cette séquence S. S est gigantesque car elle contient la description de chacune des molécules, leurs positions, les interactions : absolument tout ce qui est nécessaire pour décrire Mars *exactement*.

Ma conjecture est la suivante : Seule la vie est capable de créer des exemplaires d'une taille bien supérieure à ce que peut créer le hasard.

*Ce qu'on ne trouve pas sur Mars, c'est-à-dire dans la séquence S (en bit), c'est une séquence P apériodique (en bits) au moins supérieure au logarithme de S (multiplié par une constante) en au moins deux exemplaires.*

Donc, si quelqu'un trouvait dans la séquence S (Mars), deux séquences apériodiques P identiques (deux exemplaires de P) tel que taille(P)>log(taille(S)) x constante alors P serait une preuve de l'existence passée ou présente de vie sur Mars. P serait une trace de vie.

Je sens mon lecteur un petit peu circonspect. Derrière ces phrases d'apparence un peu barbare se cache quelque chose de relativement simple. J'ai écrit un petit script Matlab qui crée une séquence aléatoire S de 0 et de 1 d'une certaine taille par exemple 100 bits (c'est-à-dire cent 0 et 1). Puis le script recherche la séquence la plus longue P que l'on trouve en 2 exemplaires dans S. Au fur et à mesure, le script augmente la taille de S et regarde comment évolue la taille de P en fonction de S. Quand S augmente, P augmente de moins en moins vite et il semble que ce soit une fonction logarithme qui « fit » correctement les données de la simulation (à une constante prés). En fait, si la séquence S (Mars) faisait $10^{80}$ bits (le nombre d'atomes dans l'univers connu), la séquence P la plus longue présente en deux exemplaires serait d'à peine 184 bits soit $\log(10^{80})$.

Ainsi, si l'on trouvait sur cette planète mars (et sa version conceptuelle de $10^{80}$ bits), au moins deux molécules identiques de quelques centaines d'atomes (quelques centaines de bits), alors, selon ma conjecture, on serait en présence d'une preuve de l'existence passée ou présente de vie.

Sur la graphique de la figure XIII.3, on voit que la présence de la constante sert à s'éloigner vers le haut des données expérimentales (en bleu) qui représentent la taille maximale la plus probable de la séquence P (en deux exemplaires) dans une séquence S aléatoire. En s'éloignant vers le haut (en augmentant P), on fait chuter très vite la probabilité de trouver par hasard ces deux exemplaires. On peut donc définir une ligne de vie arbitraire au delà de laquelle trouver

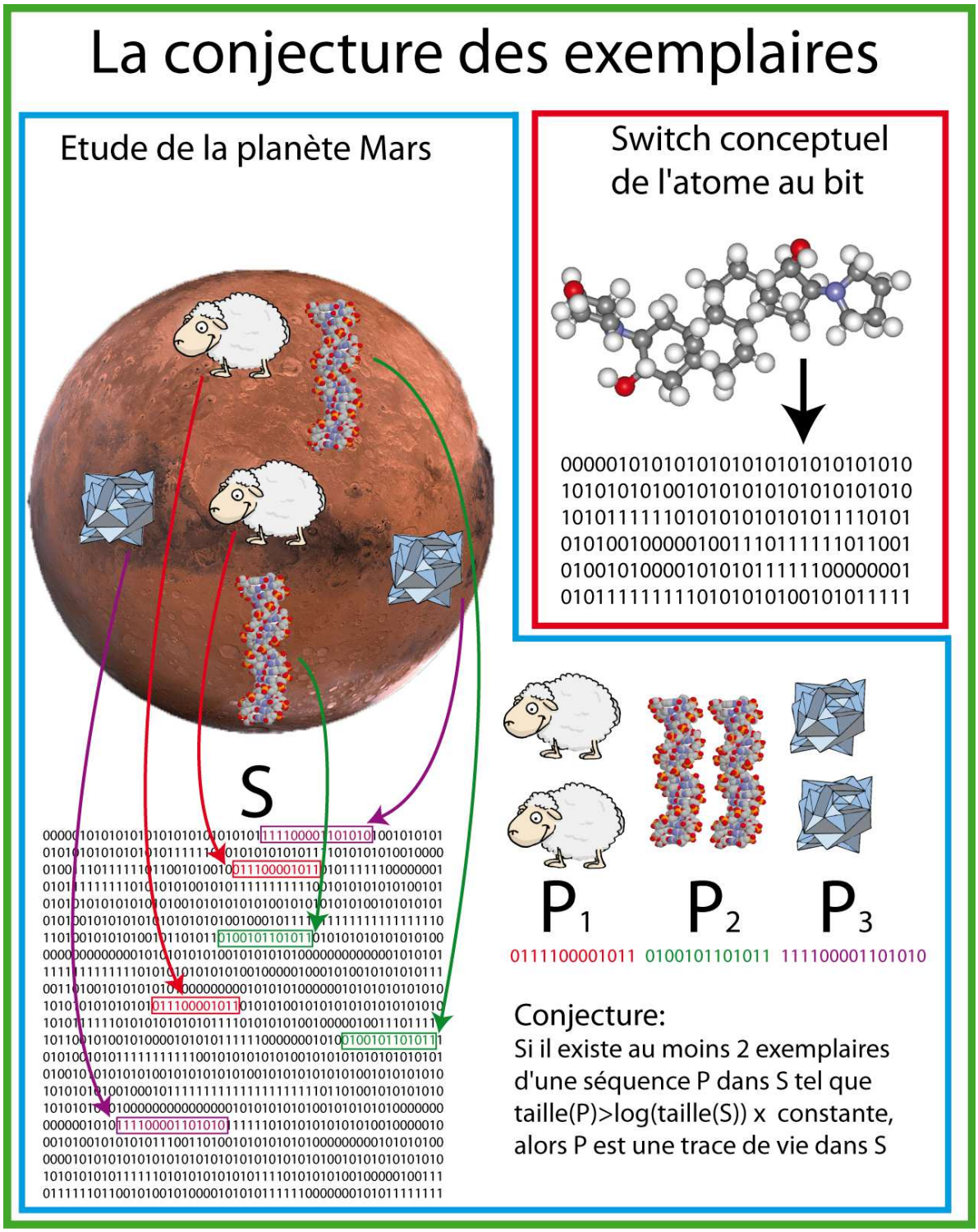



**Figure XIII.2** – La conjecture des exemplaires.

deux exemplaires ne peut être dû au hasard qu'avec une probabilité infime et est, bien plus probablement, une trace de vie passée ou présente.

Remarquez que cette notion d'exemplaire comme caractéristique de vie nécessite absolument la présence de cette « ligne de vie logarithmique ». En effet, deux atomes d'hydrogène sont identiques, ils forment deux exemplaires sans pour autant former l'essence même de la vie.

J'ai rajouté le mot « apériodique » pour supprimer les cas du type « diamant ». On pourrait imaginer trouver sur Mars des exemplaires de « diamants » représentant une longue séquence de bits. Le tour de passe-passe « apériodique » supprime ce cas de figure car il supprime les exemplaires très longs mais sans contenu en information selon Kolmogorov (facilement compressible par un algorithme). Notez d'ailleurs que les séquences d'ADN ou de protéines des êtres vivants sont bien globalement apériodiques même si il existe des séquences répétées compressibles.

La conjecture des exemplaires se propose donc comme socle théorique de recherche de la vie pour l'exobiologie (ou la Xénobiologie). Ce qu'il faut c'est rechercher deux exemplaires apériodiques de grande taille (en bit) —molécules ou pas molécules—. Notez que selon ma conjecture, trouver une soupe de molécules complexes mais toutes différentes ne suffit plus. Il faut en trouver au moins 2 identiques.

A partir de là, seule la moitié du travail est faite. Car définir ce qu'est une trace de vie et définir la vie elle-même n'est pas exactement la même chose. Pour franchir cette deuxième étape, il me faut encore m'enfoncer plus en profondeur dans la métaphysique. J'espère que le lecteur positiviste monomaniaque me pardonnera.

Qu'est ce que la vie ?

*Le système S est vivant au temps $t_1$ si et seulement si il existe un temps $t_2 > t_1$ tel que la somme de l'ensemble des exemplaires $P_n$ appartenant à S au temps $t_2$ est supérieur à cette même somme au temps $t_1$. $P_n$ est une séquence en deux exemplaires tel que taille($P_n$)>log(taille(S)) × constante et n est le nombre de séquences P différentes.*

$$\sum_{i=1}^{n} P_n(t2) > \sum_{i=1}^{n} P_n(t1) \tag{XIII.1}$$

Le but de la vie est donc la maximisation de cette somme sur un temps suffisamment long. Et c'est cette maximisation (recherche d'une pente globalement positive sur le long terme) du nombre de types d'exemplaires et du nombre d'exemplaires appartenant à S qui représente la volonté de puissance Nietzschéenne, c'est à dire la vie.

La vie cherche son expansion. Elle « veut » injecter de l'information dans chaque kilomètre cube de planète, dans chaque centimètre cube de terre, dans chaque nanomètre cube d'atomes.

# Pourquoi le logarithme ?

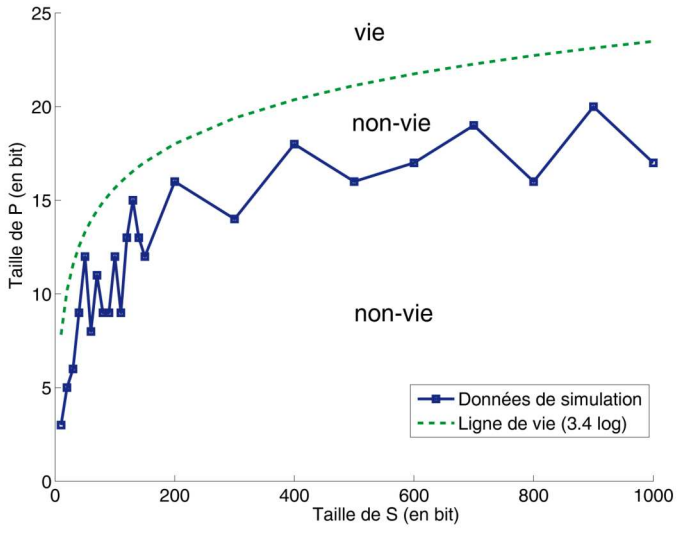
J'ai créé un algorithme qui recherche dans une séquence en bit P aleatoire la taille maximum d'une séquence S appartenant à P et présente en au moins deux exemplaires

S=22; P=5

S=86; P=9

S=152; P=12

Résultat: La taille de la séquence noire S grossit beaucoup plus vite que la taille des 2 séquences rouges P




En bleu, on voit les données issues de la simulation jusqu'à S=1000. Il semble que la taille de P augmente de manière proportionnelle au logarithme (mutliplié par une constante) de la taille de S. En prenant une marge suffisante (en augmentant la constante; ici arbirairement 3.4), on peut definir une ligne de vie: Dans la zone au dessus de cette ligne, la probabilité de trouver 2 exemplaires identiques sans ce que ceux-ci representent une trace de vie est infime. En dessous, de cette ligne, trouver 2 exemplaires identiques est trés probable. Ils ne representent pas une trace de vie.

**Figure XIII.3** – Le logarithme semble intervenir dans la conjecture des exemplaires. Voir le code Matlab en annexe Code23ExemplaireLog.m qui gènère cette figure.

Si elle le pouvait, la vie envahirait le centre de la terre ou l'univers. Mais pour l'instant, elle cherche. Injecter de l'information dans les atomes de l'univers, voila son *but*. Et je soupçonne que cette injection d'information passe inexorablement par la notion d'exemplaires. Se pourrait-il que les concepts d'information, de mémoire et d'exemplaires supérieurs au logarithme soient des concepts indissociables ?

Petite parenthèse d'épistémologie. N'hésitez pas à la sauter. Intuitivement, nous avons le sentiment qu'il y a, au cours de l'histoire, une pente positive dans l'élaboration des connaissances, une sorte de pyramide : on appelle cela le progrès. Or les relativistes extrêmes nient l'existence de ce progrès qui ne serait qu'une vue de l'esprit (attention, il ne s'agit pas ici de morale). Et il me semble qu'en l'état actuel des connaissances, il est impossible de démontrer l'existence du progrès et donc de contredire les relativistes. Or la conjecture des exemplaires, tentative pour définir la vie, peut aussi servir à définir ce qu'inclut la vie, c'est-à-dire le progrès. La maximisation de la somme des exemplaires « d'énoncés informatifs » qui se propagent pourrait servir de socle à la définition du progrès (hors question morale je le répète). Nous aurions donc un point d'accroche pour contre-argumenter face aux relativistes. Fin de la parenthèse.

Je répète que la conjecture ci-dessus est une pure spéculation. Les équations ne sont pas là pour donner « une illusion de science » : nous nageons en pleine métaphysique. Je prends volontairement le risque de raconter/de présenter de grosses bêtises pour montrer que justement en métaphysique, le ridicule ne tue pas : je suis à peu prés sûr qu'une personne lisant ces lignes pourrait sans doute, en réfléchissant quelques minutes, présenter de bons arguments à l'encontre de ma conjecture. Ces arguments pourraient me convaincre et me faire rejeter ma propre conjecture. Car c'est la flexibilité de l'opinion et non sa stabilité qui caractérise le fonctionnement scientifique.

Bien que non positiviste, étrangement cependant, ma conjecture métaphysique conserve une certaine faculté de falsifiabilité (selon Popper). En voici la preuve : *« Si quelqu'un trouvait un système extraterrestre présentant à l'évidence deux exemplaires identiques de très grandes tailles (en bit) sans que ceux-ci représentent une trace de vie,* ***alors ma conjecture serait fausse*** *»*. Puis la version opposée, *« Si quelqu'un trouvait un système extraterrestre ne présentant pas deux exemplaires identiques de très grandes tailles (en bit) alors même que l'évidence empirique indique la présence d'une vie extraterrestre dans ce système,* ***alors ma conjecture serait fausse*** *»*. Je ne suis cependant pas naïf et je ne vais pas tenter de faire passer ce pseudo critère de réfutabilité comme une preuve de la scientificité de ma conjecture.

J'ai utilisé l'exemple de Mars comme du système S mais en fait, le système S peut représenter n'importe quoi et donc peut avoir des dimensions beaucoup plus restreintes, des dimensions « expérimentales ». Mettez de l'eau dans un tube et c'est votre système S. Peut-

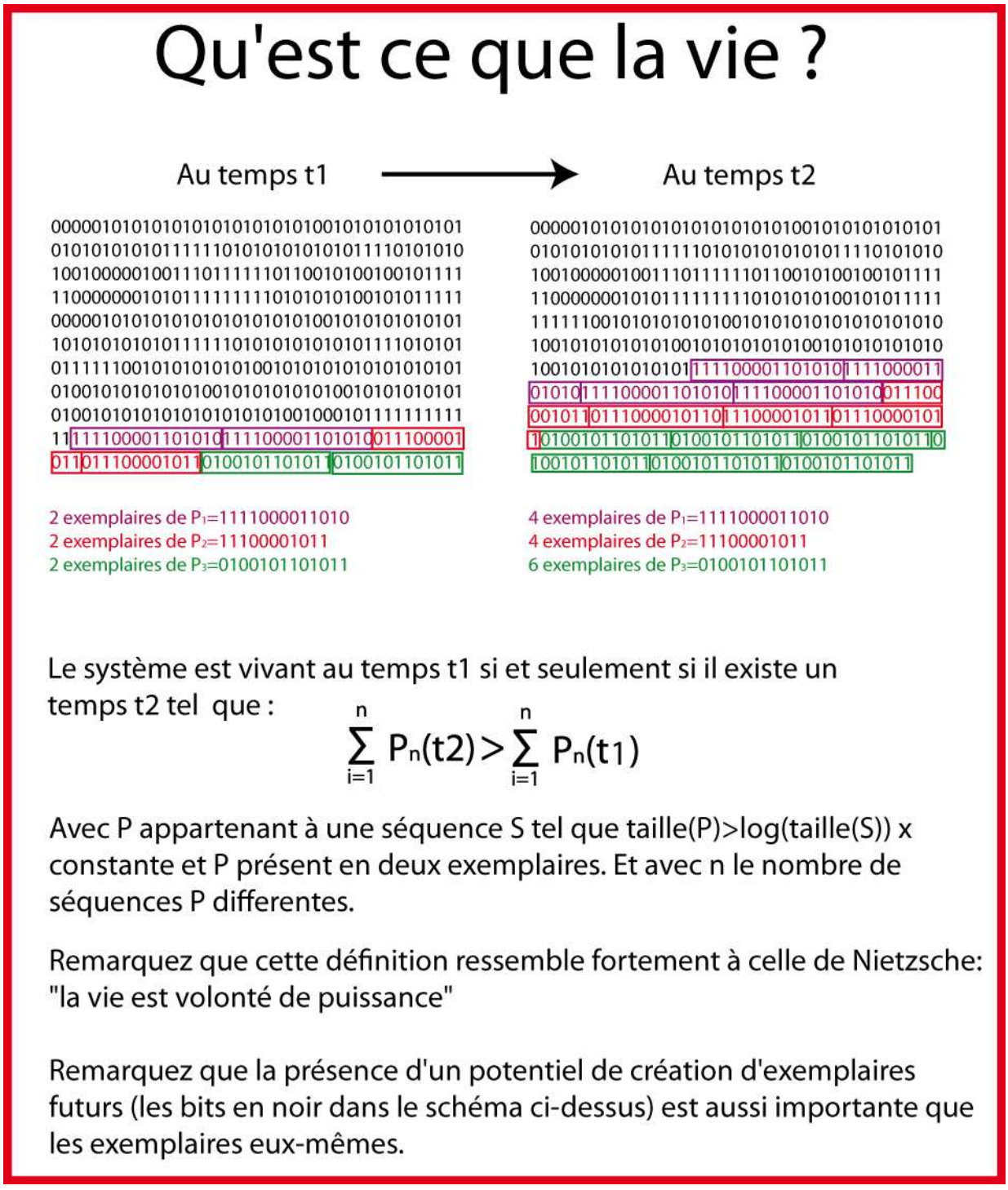



**Figure XIII.4** – Pure spéculation sur la question « qu'est ce que la vie ? »

on trouver/imaginer deux exemplaires de grandes tailles (en bit) (molécules ou pas, ce qu'on veut !) dans ce tube ? Selon moi non : il n'y a pas de traces de vie dans ce système. Rajouter maintenant une protéine (sans motif répété) : Ce système comporte t'il une trace de vie ? Selon ma conjecture : Non. Rajouter maintenant une deuxième protéine identique et alors le système contient une preuve de vie passée. Cette vie passée c'est la vie de l'organisme modèle utilisé qui a été capable de synthétiser ces deux protéines identiques.

La conjecture des exemplaires induit quelques conséquences étranges :
Placez des bactéries dans du glycérol et congelez-les à -80°C. Si après quelques mois, on remet ces bactéries congelées dans du milieu frais, elles « revivent » c'est-à-dire qu'elles se remettent à se diviser. C'est ce que l'on fait tous les jours au laboratoire. Le système Bactéries/ glycérol/ congélateur contient indéniablement des preuves de vie passée (enzymes identiques, ADN, ARN identiques etc...). Pourtant, selon la conjecture, le système Bactéries/ glycérol/ congélateur est mort car il n'existe pas de temps $t_2 > t_1$ tel que ce système contienne plus d'exemplaires qu'au temps $t_1$. Le fait que les bactéries soient ressuscitables ne changent en rien au fait que le système Bactéries/glycérol/congélateur soit mort. Le fait de faire remonter la température et d'ajouter du milieu change le système (injection de calories) qui devient bien vivant car les bactéries ressuscitent mais il n'est en rien comparable au système précédent (sans calorie).

Placez un ordinateur portable sur Mars. Cet ordinateur contient un programme basique capable de manipuler (effacer et créer) des exemplaires (des séquences identiques de grandes tailles en bit). Maintenez la batterie de l'ordinateur en utilisant des cellules photovoltaïques. Votre système Ordinateur–Mars–Soleil vivra quasi éternellement.

# 5 La mort

Dans la conjecture des exemplaires, la mort/destruction des organismes/des exemplaires a un intérêt pour l'heuristique qu'est la vie. En effet, pour maximiser le nombre total d'exemplaires, il est nécessaire d'en détruire certains pour récréer du potentiel et chercher une combinaison plus optimale. La vie, volonté de puissance, a choisi de nous faire mourir car sans cela, celle-ci serait bloquée dans un maximum local. Le temps $t_2$ contenant plus d'exemplaires n'existerait plus ce qui signerait, par définition, l'arrêt de mort de la vie. Une idée déjà exprimée par Nietzsche :

> *« Je veux dire que le fait de devenir partiellement inutile, de dépérir et dégénérer, de perdre sens et utilité, bref de mourir, cela aussi appartient aux conditions d'un progrès véritable : lequel prend toujours forme de volonté et de voie vers plus*

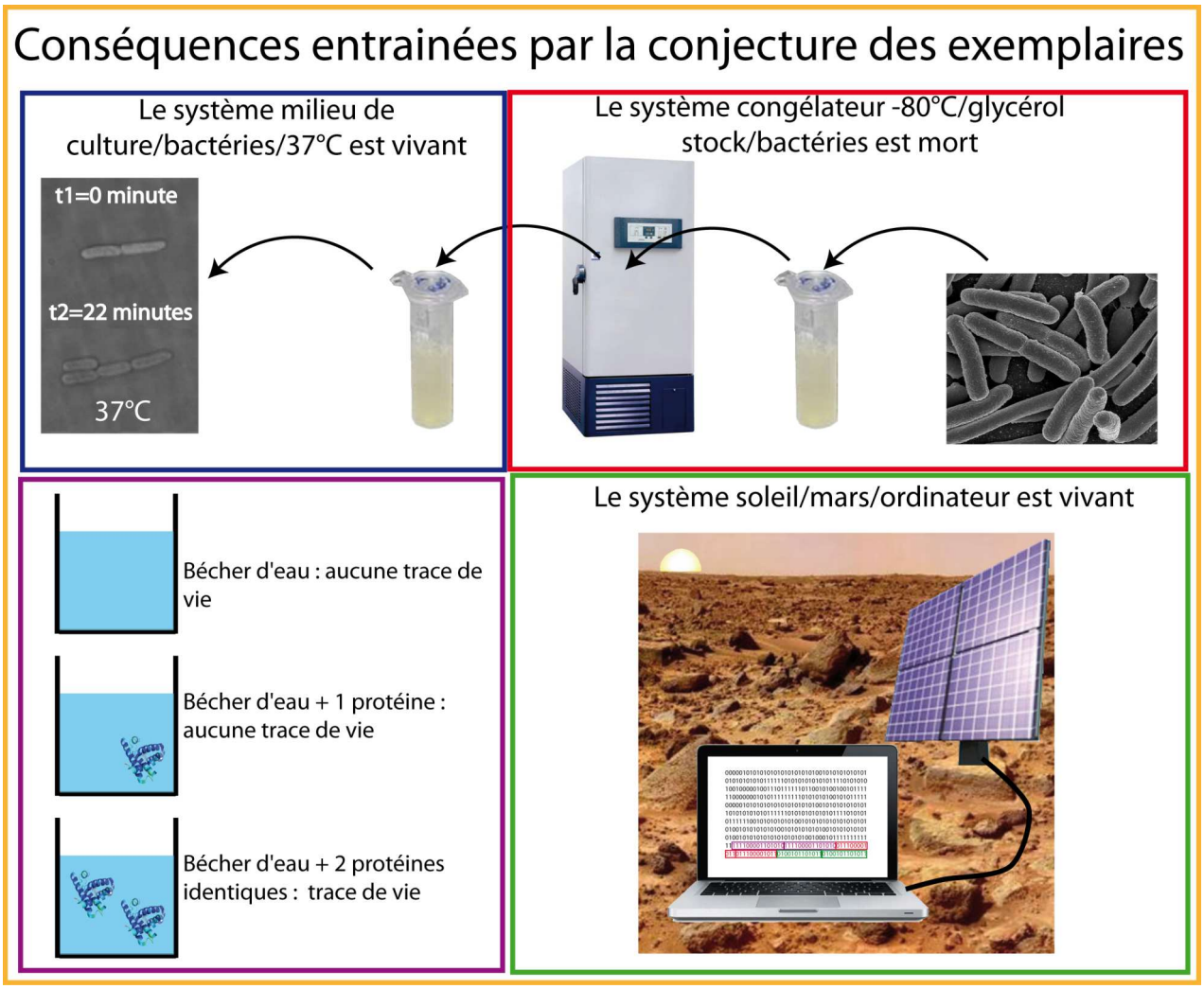



**Figure XIII.5** – Quelques conséquences étranges induites par la conjecture des exemplaires.

*de puissance et s'accomplit toujours aux dépens d'un grand nombre de puissance mineures. »*[14]

En décrivant la conjecture des exemplaires, j'ai utilisé le mot mort mais il aurait mieux valu que j'utilise le mot *non-vie* ou le mot *potentiel*. En effet, selon moi, la mort n'est pas l'opposé de la vie. La mort est le mot qui marque la transition de la vie à la non-vie. La mort, chez l'homme, est difficilement dissociable de la peur qu'elle engendre. Cette peur est plus liée à la transition qu'à l'éternité de non-vie qui lui succède. Cette transition implique :

⇒ peur de la souffrance et de l'agonie

⇒ peur de l'inconnue

[14]Nietzsche, *La généalogie de la morale*, Folio p. 86

⇒ peur quant à l'avenir de ses proches et leurs souffrances

⇒ peur instinctive de la décomposition pestilentielle du corps. Cette peur de la décomposition n'est pas une peur pour soi. C'est une peur pour son image.

> *« Ô spectacle de terreur ; mort difforme et affreuse à voir ; horrible à penser et combien horrible à souffrir »*[15]

Je pense que la peur instinctive d'un corps en décomposition est liée à la présence d'eau. En effet, c'est l'eau qui est un des moteurs de la décomposition (via l'action des micro-organismes). L'évolution nous a sélectionnés pour avoir peur de ces organismes morts mais encore hydratés et donc actifs « calculant ! » car la transmission des maladies passe parfois par l'intermédiaire des cadavres. A l'inverse, un corps sans eau (momie déshydratée) ou un animal desséché (par exemple un insecte mort) ne génèrent pas la même anxiété. L'intuition nous fait percevoir le caractère inoffensif de ces cadavres.

Plus généralement, je soupçonne que la peur de la mort soit encodée dans le génome et soit sélectionnée par l'évolution. Petit question d'éthique : si nous détections les gènes/nucléotides responsables de cette peur irrationnelle, devrions-nous les muter pour l'humanité entière ? Si non, devrions nous interdire à quelqu'un qui ne souhaite plus avoir peur de la mort de muter les siens ?

Cependant, le rapport de l'homme à la mort dépend aussi de sa manière d'envisager le temps. A l'époque où la dérivée positive du progrès n'était pas vraiment visible, le temps avait une apparence plus cyclique. Avec l'alternance des saisons, avec le cycle des générations, l'arrivée de la mort semblait normale, dans la logique de ce temps cyclique, en harmonie avec la nature. A l'inverse, avec le progrès, le développement des sciences, le temps a pris une allure linéaire, croissante. La mort nous apparait comme une erreur à cause de la sensation que le progrès continuera après notre mort. On saute d'un bateau qui continue en ligne droite, sans nous, vers son but. Cette sensation « d'erreur » se surajoute négativement aux effets instinctifs/évolutifs. Nous verrons bien plus tard que paradoxalement ce bateau (la science) avance à la recherche du néant.

Etant donné ce rapport particulier qu'à l'homme à la mort, il lui est très difficile de traiter la mort de manière scientifique et de la loger disons à la même enseigne que la vie. Pour un scientifique comme pour tout le monde d'ailleurs, penser à la mort c'est penser à sa mort et cela engendre des réactions infantiles.

> *« Toute conscience de la mort, provoque toujours des réactions infantiles, car la mort est la seule chose hors du pouvoir de l'homme, la seule devant laquelle il est*

[15] Edgar Morin, *L'homme et la mort*, Points p. 43.

*impuissant totalement, comme un enfant. »* [16]

La distance positiviste, consistant à cacher autant que possible le scientifique derrière ses observations, est plus difficile à trouver qu'avec tout autre questionnement scientifique y compris la vie. Ainsi la mort a une dimension profondément humaine et c'est sous cet angle là que je souhaite l'aborder. C'est en réalité le seul angle que j'entrevoie car la mort dépossédée de connotation humaine, c'est, je le répète, de la non-vie. Nietzche disait à peu prés l'inverse, mais au final, j'ai l'étrange sentiment que cela revient au même.

> *« Gardons nous de dire que la mort est le contraire de la vie. Le vivant n'est qu'un genre de mort, et un genre très spécial »* [17]

On sait, grâce au psychiatre Elisabeth Kübler-Ross, qu'à l'annonce d'une mort prochaine, l'homme passe par différents stades, de manière cyclique. Ces cycles peuvent se produire à différentes échelles de temps (heures, jours, semaines, mois). J'énumère ces différents stades pour permettre aux lecteurs qui en ignoraient l'existence de les intégrer dans leur savoir. Cela peut s'avérer utile pour celui qui souhaiterait lutter contre ces états cognitifs standards, cycliques et douloureux qui touchent les patients condamnés. Il s'agit :

⇒ du déni (mais non je ne suis pas malade)
⇒ de la colère (souvent dirigée contre une autre cible humaine)
⇒ du marchandage (je veux bien mourir mais laissez-moi au moins voir naître mes petits enfants, finir mon œuvre...)
⇒ de la dépression/tristesse (tristesse de devoir partir, désintérêt envers la vie pour lutter contre cette tristesse)
⇒ de l'acceptation (A tout film il y a une fin, c'est la vie).

Un professeur m'a dit un jour en substance : « *l'homme éprouve des difficultés à penser à sa propre mort plus de quelques secondes d'affilée* ». J'y ai cru pendant longtemps. Je n'y crois plus. Car l'idée de la mort occupe aussi une place, chez les biens portants, dans l'esprit de bon nombre d'entre nous de manière quotidienne (mais cachée)

> *« Chacun cache sa mort, l'enferme à double tour »* [18]

> *« Silencieuse : elle ronge invisible, secrète, honteuse, la conscience au cœur même de la vie quotidienne. Qui oserait en gémir dans la malédiction commune ? »* [19]

[16] Edgar Morin, *L'homme et la mort*, Points p. 185.
[17] Nietzsche, *Le gai savoir*, Flammarion p. 163.
[18] Edgar Morin, *L'homme et la mort*, Points p. 43.
[19] Edgar Morin, *L'homme et la mort*, Points p. 43.

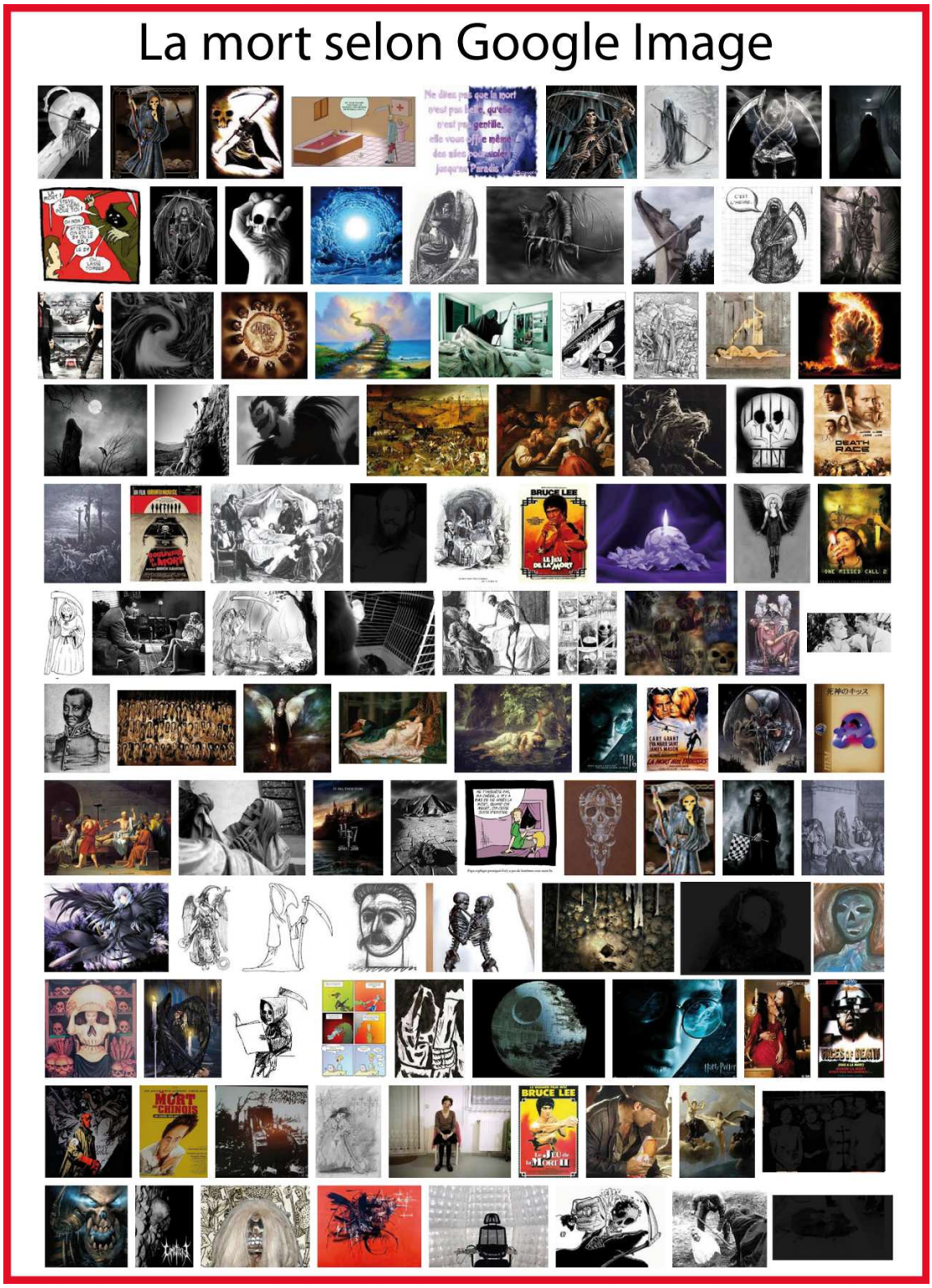



**Figure XIII.6** – La mort selon Google image

*« On ne philosophe pas pour passer le temps, ni pour se faire valoir, ni pour faire joujou avec les concepts : on philosophe pour sauver sa peau et son âme. »* [20]

S'il est vrai que beaucoup pensent régulièrement à la mort avec angoisse durant leur vie, les choses pourraient être bien pires :

*« Et pourtant c'est la mort et le silence de mort qui est l'unique certitude et le lot commun à tous dans cet avenir ! Qu'il est étrange que cette unique certitude et ce lot commun n'aient presque aucun pouvoir sur les hommes et qu'ils soient à mille lieux de se sentir comme une confrérie de la mort ! »* [21]

Etrangement il arrive parfois qu'il y ait un décalage entre la peur de la mort (durant toute sa vie) et l'arrivée de la mort, lors de l'agonie, se passant dans la sérénité :

*« C'est, souvent encore en pleine conscience, dès que s'affaiblit le muscle de l'affirmation de soi que l'horreur de la mort se dissipe. D'où le fait que certains hommes, hantés toute leur vie durant par l'idée de la mort, montrent au moment physique de l'agonie un calme qui les stupéfie eux-mêmes. C'est ce que l'on appelle la béatitude des mourants où il semble que l'espèce étende sa patte protectrice sur l'individu agonisant. »* [22]

S'il est vrai que la mort nous fait peur, il faut aussi admettre que cette dernière nous aide à conduire nos vies :

*« Comme le signalait Kierkegaard, la mort envisagée dans le sérieux, n'est-elle pas une source d'énergie comme nulle autre ? Ne stimule-elle pas l'action ? Et si la transformation de l'essence de l'agir humain faisait elle perdre à l'esprit sa profondeur ? A l'homme animé de sérieux, la pensée de la mort à venir indique le but à diriger sa course. La mort est un stimulant de la vie et le sérieux comprend que l'idée de la mort représente une invitation à l'action. Ne perdons pas notre temps ? »* [23]

On peut aussi ressentir, comme Spinoza, le fait que rien ni personne (pas même la mort) ne pourra changer le fait que « nos vies » ont existé, ont été :

*« Nous sentons et expérimentons que nous sommes éternels »*

*« Sagesse du grand homme qui tient la mort à distance en continuant son œuvre immortelle »* [24]

[20] André Comte-Sponville, *Présentation de la philosophie*, le livre de poche p. 151.
[21] Nietzsche, *Le gai savoir*, Flammarion p. 228.
[22] Edgar Morin, *L'homme et la mort*, Points p. 49.
[23] Jacqueline Russ, *La pensée éthique contemporaine*, PUF p. 33.
[24] Edgar Morin, *L'homme et la mort*, Points p. 63.

Peu de gens, finalement, se demandent ce qu'ils pourraient donner à la vie « en échange ». Nietzsche nous dirait qu'il donnerait à la vie « de la volonté de puissance ». Car il y a une chose que l'on ne peut pas offrir à la vie, c'est la volonté de vivre.

> *« Il n'a assurément pas rencontré la vérité celui qui parlait de la « volonté de vie » , cette volonté n'existe pas ».*[25]

La preuve est dans l'éternel retour :

> *« Le poids le plus lourd. Et si un jour ou une nuit, un démon se glissait furtivement dans ta plus solitaire solitude et te disait : cette vie telle que tu la vis et l'a vécu, il te faudra la vivre encore une fois et encore d'innombrables fois ; et elle ne comportera rien de nouveau, au contraire, chaque douleur et chaque plaisir et chaque pensée et soupir et tout ce qu' y a dans ta vie d'indiciblement petit et grand doit pour toi revenir, et tout suivant la même succession et le même enchainement »*[26]

Qui souhaiterait un tel éternel recommencement de la vie ? Personne. C'est bien là une preuve que nous ne sommes pas tant attachés à la vie en elle-même mais plutôt à la volonté de puissance qu'elle rend possible. En fait, Nietzsche pense qu'idéalement, il faudrait pouvoir mener chaque seconde de sa vie de manière à pouvoir souhaiter que ces secondes se répètent dans un cycle infini. Or justement, pour pouvoir souhaiter cela, il faut avoir cherché à gagner en puissance toute sa vie. Posons-nous une question simple ? Voudrions-nous vivre éternellement la vie d'un bovin ? Une vache dans un champ nous semble avoir une volonté de puissance trop faible et nous méprisons une telle vie, fade et sans intérêt. La vie d'un bovin nous semble plus proche de la mort que de notre vie propre.

Pour résumer, quand on voit que :

⇒ la mort est la disparition de la fatigue

⇒ une patte protectrice se glisse avant et durant l'agonie pour remplacer petit à petit l'affirmation de son soi par la sérénité

⇒ notre mort n'entraine pas la mort de « L'homme » ni comme espèce, ni comme idéal

⇒ la mort est probablement encodée et souhaitée par la nature dans un but heuristique (ou qui nous échappe)

⇒ la peur de la mort a sans doute été encodée dans le génome par le processus évolutif de sélection ce qui démontre sa totale irrationalité.

⇒ la mort est nécessaire à l'agir humain

[25] Nietzsche, *Ainsi parlait Zarathoustra/Deuxième partie/De la victoire sur soi-même.*

[26] Nietzsche, *Le gai savoir*, Flammarion p. 279.

⇒ le vivant n'est qu'un genre de mort

⇒ nous étions « non-vie » avant notre naissance et donc nous « connaissons » déjà un peu l'au-delà

⇒ nos vies sont éternelles quoi qu'il advienne

⇒ nous ne souhaiterions pour rien au monde revivre nos vie à l'identique dans un cycle infini

On peut donc bien considérer que, après tout, comme disait Platon :

*« La mort est un beau risque à courir ».*

# Chapitre XIV

# Les bio-armes

## 1 Préambule

A la demande de l'académie des sciences, Henry Korn a publié un très bon rapport en 2008 sur la question de la menace biologique. Il y décrit :

⇒ l'utilisation historique de la biologie à des fins militaires ou terroristes.

⇒ les méthodes actuelles envisagées pour se prémunir de cette menace.

Dans une première partie, je vais extraire certaines informations présentes dans ce rapport pour offrir à mon lecteur la connaissance de base nécessaire (je précise, comme Henry Korn l'a fait, qu'aucune information confidentielle n'est divulguée dans ce chapitre). Puis j'aborderai une partie plus personnelle, plus spéculative que l'on pourra aisément qualifier de science fiction. Mon but sera de montrer que les risques ne sont pas forcement là où on les attend.

J'aurai pu appeler ce chapitre bioterrorisme mais je préfère envisager, tout comme Henry Korn, la menace biologique au sens large en évitant tout discours partisan trop proche de la politique et trop loin des sciences. J'ai donc choisi le terme bio-arme, plus neutre.

## 2 Jusqu'où l'homme est-il déjà allé

Commençons par faire un triste constat. En termes de bio-armes, l'homme, au cours de l'histoire, est déjà allé très loin. Citons en 1374, au siège de Kaffa, le catapultage par les turcs de cadavres pestiférés par-dessus les remparts, déclenchant une épidémie de peste noire parmi les défenseurs génois.

Mais c'est au $20^{\text{ème}}$ siècle que l'homme a créé les plus vastes programmes de recherche visant à développer des armes bactériologiques tactiques et stratégiques. Développons en chiffres quelqu'un de ces programmes pour que lecteur ait une idée globale de leurs ampleurs et de leurs

conséquences[1].

*Le programme japonais (1931–1945)* : 150 bâtiments, 3000 scientifiques. 300kg de bacilles de la peste et du charbon (100 gr suffisent à éliminer la population d'une petite ville). Inoculation de toutes sortes de souches pathogènes aux prisonniers chinois. 12 essais sur champ de bataille sur des villes et villages chinois en contaminant la nourriture, l'eau ou l'air. Lors d'une attaque en 1942 par le vibrion cholérique, le cholera se propagea dans les armées japonaises faisant plusieurs milliers de morts ce qui arrêta les essais sur le terrain.

*Le programme soviétique* : L'URSS a mené à partir de 1973, en violation d'une convention signée en 1972, un énorme programme de mise au point d'armes biologiques impliquant plus de 65000 personnes dispersées sur 40 sites dont plus de 8 centres de recherche, 5 usines de production et un centre d'essai en plein air sur un île. 8 agents ont été militarisés (variole, peste, tularémie, morve, encéphalite équine, charbon, fièvre Q, virus de Marburg). Les soviétiques auraient même produit plusieurs tonnes de virus de la variole (un des pires virus qu'ait connu l'humanité : résistant, transmission par aérosol, 20–40% de mortalité ; plus de 500 millions de morts au 20$^{\text{ème}}$ siècle) en 1980 sous forme de poudre lyophilisée pour équiper des missiles à longue portée. Bien sûr, lors du développement de ces armes, plusieurs accidents mortels ont eu lieu dont une fuite de bacilles de charbon d'une usine militaire provoquant la mort de 64 personnes dans le village d'à coté.

*Le programme américain* : En 1943, à la demande de Roosevelt, un grand programme sur les armes biologiques fut développé. Un centre de recherche fut établi à Fort Detrick dans le Maryland comprenant 250 bâtiments et des habitations pour 5000 personnes. Les américains ont travaillé sur la militarisation des agents infectieux pour leur permettre, par exemple, de rester plus longtemps en suspension dans l'air et augmenter la probabilité d'inhalation humaine. Ils ont mis au point des missiles à ogives biologiques comportant 720 petites bombes, largables à 16 km d'altitude, censés pulvériser une superficie de plus de 150km$^2$. Ils envisagèrent aussi la dissémination de cocktails « invalidants » sur Cuba pour rendre malade sans tuer l'ensemble de la population pour en faciliter l'invasion. Des essais de contamination (tularémie, fièvre Q) furent même entrepris sur des porteurs sains et des objecteurs de conscience. Même si les américains ont signé en 1972 la convention internationale d'interdiction des armes biologiques et chimiques, Fort-Detrick reste un centre dévolu à la recherche sur la prévention contre les armes biologiques et d'importants programmes secrets s'y déroulent encore.

*Le programme Sud-africain* : Dans les années 80, les partisans de l'apartheid ont favorisé l'émergence d'une organisation politique appelée Freedom Front. Cette organisation mit sur pied

[1]Il existe d'autres programmes connus (français, britannique, irakien) que je ne développe pas ici pour éviter les redondances. Je renvoie le lecteur au rapport de Henry Korn pour avoir des détails sur ces programmes.

un programme de développement d'armes biologiques. Le Dr. Basson fut nommé responsable du projet et acquit une solide formation dans ce domaine grâce aux nombreuses informations recueillies auprès de chercheurs britanniques, américains et canadiens. Des recherches secrètes furent développées pour obtenir des armes capables d'exterminer et stériliser sélectivement la population noire.

# 3 Où en sommes-nous aujourd'hui ?

La fabrication du matériel biologique pathogène n'exige pas une formation très poussée en microbiologie. Un grand local avec quelques équipements de petites dimensions et quelques fermenteurs suffisent ainsi que la capacité de se procurer les souches pathogènes. La difficulté réside surtout dans la « militarisation » c'est-à-dire les connaissances à acquérir pour conserver et propager le matériel biologique pathogène.

La plupart des états n'accordent qu'un faible intérêt stratégique aux armes biologiques du fait du caractère imprévisible des risques pour l'attaquant lui-même. En d'autres termes, les armes biologiques sont difficilement contrôlables. On ne peut exclure l'évolution des souches une fois disséminées et l'échappement puis le retournement contre l'agresseur. Les conséquences sont donc difficiles à évaluer et potentiellement dramatiques.

Ainsi une convention sur l'interdiction des armes biologiques a été ouverte à la signature en 1972 et ratifiée depuis par 162 états qui se sont engagés à ne pas développer, produire, stocker ou utiliser des armes biologiques. Cependant, l'efficacité de la convention reste limitée étant donné qu'elle ne prévoit aucun régime de vérification du respect de ces dispositions.

Comme le souligne Henry Korn, une des difficultés pour faire appliquer cette convention c'est la nature duale des recherches : c'est-à-dire la barrière très étroite qui sépare une recherche défensive, visant à se protéger d'un pathogène, et offensive, visant à le militariser. En effet, les recherches fondamentales nécessaires pour comprendre le fonctionnement d'un pathogène ont évidemment des applications en santé publique mais peuvent tout autant servir à des fins militaires.

> *« [Ainsi des recherches à visées défensives] peuvent être rapidement interprétées par un pays tiers comme étant de nature offensive, particulièrement si ces recherches impliquent, par exemple, des modifications génétiques qui pourraient sembler au premier degré destiner à accroitre les propriétés pathogènes. »* [2]

Prenons un exemple : pour comprendre comment des souches pathogènes deviennent de

[2]Henry Korn, *Les menaces biologiques*, PUF p. 69.

plus en plus résistantes aux antibiotiques, il est nécessaire de mettre en place des expériences d'évolution visant à renforcer la résistance de ces souches. Cela permet d'identifier le mécanisme à l'origine de la résistance puis de mettre en place une stratégie pour lutter contre. Mais on voit immédiatement que cette même souche, rendue résistante en laboratoire par des mécanismes d'évolution dirigée, peut tout autant être utilisée à des fins militaires.

En particulier, les techniques actuelles de biologie moléculaire permettent d'envisager l'amélioration du génome de certaines souches (virus ou bactéries) pour les rendre bien plus pathogènes. PCR, séquençage, synthèse chimique de génome, expérience d'évolution dirigée (DNA shuffling) sont des techniques utilisées (ou sur le point d'être utilisées) en routine dans les laboratoires pour mener des recherches dans de nombreuses directions. Ces techniques peuvent être détournées pour militariser des souches (augmenter la virulence, augmenter la résistance dans l'environnement, augmenter la probabilité de transmission interhumaine). Par exemple, Kobinger a montré en 2001 que l'introduction d'un gène du virus Ebola dans le génome du virus du sida confère à ce dernier la capacité de se déplacer via le système respiratoire (alors que pour l'instant, le mode de propagation principal du VIH est le rapport sexuel) (Kobinger *et al.*, 2001).

Cet exemple nous montre que certaines informations présentes dans les publications scientifiques (non confidentielles, ouvertes à tous) sont susceptibles d'être utilisées par des personnes mal intentionnées. Même danger avec la publication de la méthode permettant de synthétiser entièrement un virus (la poliomyélite) par assemblage d'oligonucleotides à partir de l'information de sa séquence ADN *in silico* (Cello *et al.*, 2002) :

> *« Our result show that it is possible to synthesize an infectious agent by in vitro chemical biological means solely by following instructions from a written sequence »*

Il faudra attendre 2005 pour que soit réussie la synthèse du virus de la grippe espagnol. Voyez ici la première phrase de l'article :

> *« The pandemic influenza virus of 1918–1919 killed an estimated 20 to 50 million people worldwide. With the recent availability of the complete 1918 influenza virus coding sequence, we used reverse genetics to generate an influenza virus bearing all eight gene segments of the pandemic virus to study the properties associated with its extraordinary virulence. »*

Ce type de travaux peut tout autant servir pour lutter contre les futures épidémies de grippes très virulentes que comme recette de cuisine pour reproduire et utiliser militairement un virus pandémique qui avait disparu.

# 4 Stratégies de lutte contre les bio-armes

Il existe différents lieux et personnes susceptibles de créer des bio-armes :

⇒ *La division classique du monde en noir et blanc* : Corée du nord, Syrie, Soudan, Pakistan... sont identifiés comme producteurs potentiels de bio-armes.

⇒ *Les états démocratiquess* : ce sont eux qui, au cours de l'histoire, ont développé les plus gros programmes d'armes biologiques. La présence de stocks pourtant interdits et les recherches offensives sont à suspecter étant donné que la convention de 1972 ne contient pas de volet de vérification (voir le débat à propos du NBACC).

⇒ *Les universités et laboratoires de recherche* : les contrôles sur les sujets de recherche sont limités ce qui est justifié par le fait de ne pas nuire à la créativité des chercheurs. Ceux-ci étudient, créent et stockent des souches pathogènes relativement facilement. Le but est toujours la santé publique mais on n'est jamais à l'abri d'une brebis galeuse.

⇒ *Le domaine amateur* : Le relatif faible coût des infrastructures nécessaires pour faire de la biologie moléculaire ou synthétique a pour conséquence le développement de ces branches dans le domaine amateur (DIYbio). Au fil des années, il sera de plus en plus facile de mener des recherches d'évolution dirigée dans son garage pour fabriquer des souches pathogènes.

Quelles sont les pistes actuelles de réflexion pour empêcher le développement de bio-armes ?

⇒ Détecter les recherches sensibles et évaluer leurs bénéfices/risques. Ces contrôles doivent être mesurés de manière à ne pas verrouiller la recherche, d'autant plus que le nombre de laboratoires « à risque » serait limité (300 laboratoires dans le monde et 2500 chercheurs seraient concernés soit moins de 1 % de la recherche en microbiologie).

⇒ Eviter la publication d'informations sensibles. Les éditeurs de journaux sont de plus en plus sensibilisés à cette question. Les informations sensibles seraient néanmoins accessibles à des personnes accréditées.

⇒ Réfléchir à la mise en place de contrôles sur la synthèse d'oligonucleotides, de gènes ou de génomes. En effet, certaines séquences ADN (gènes de virulence, etc...) pourraient être répertoriées et interdites de synthèse sans autorisation. En effet, les chercheurs font souvent appel à des prestataires extérieurs (entreprises privées) pour ces synthèses. Un algorithme pourrait trés facilement détecter une séquence commandée interdite. Le même principe pourrait être appliqué aux séquençages. Tout résultat de séquençage ou demande de synthèse suspecte étant immédiatement détecté, des personnes malintentionnées seraient dans l'obligation de se rabattre sur l'achat du matériel couteux de synthèse et de séquençage ce qui, à nouveau, les exposerait vis-à-vis des services de renseignements.

⇒ Sensibilisation de la communauté scientifique aux risques et aux problèmes éthiques posés par la biologie. Le modèle d'apprentissage de la science doit être repensé ce que Einstein avait déjà souligné :

> *« il ne suffit pas d'apprendre à l'homme une spécialité. Car il devient ainsi une machine utilisable mais non une personnalité. Il importe qu'il acquière un sentiment, un sens pratique de ce qui vaut la peine d'être entrepris, de ce qui est beau, de ce qui est moralement droit. Sinon il ressemble davantage, avec ses connaissances professionnelles, à un chien savant qu'à une créature harmonieusement développée. Il doit apprendre à comprendre les motivations des hommes, leurs chimères et leurs angoisses pour déterminer son rôle exact vis-à-vis des proches et de la communauté »* [3]

Je relaie aussi l'idée du philosophe Michel Serres de faire prêter un serment aux futurs chercheurs (au moment du soutien du doctorat par exemple) du même type que le serment d'Hippocrate (que tout médecin prête à la fin de ses études). Le voici :

> *« Pour ce qui est de moi, je jure : de ne point faire servir mes connaissances, mes inventions et les applications que je pourrais tirer de celles-ci à la violence, à la destruction ou à la mort, à la croissance de la misère ou de l'ignorance, à l'asservissement ou à l'inégalité, mais de les dévouer, au contraire, à l'égalité entre les hommes, à leur survie, à leur élévation et à leur liberté »*[4]
>
> *« Les auteurs du trésor l'ont prêté »*

Je le prête également.

Je ne suis cependant pas naïf et je vois bien le conflit d'intérêt qu'il y aurait immédiatement avec l'industrie de l'armement. En effet, celle-ci ne pourrait plus embaucher de docteurs. Je doute donc que l'idée puisse aboutir en France. Lors des grandes découvertes du génie génétique dans les années 70, il y a eu une conférence organisée par Paul Berg à Asilomar en Californie. Cette conférence a appelé à un moratoire sur les manipulations génétiques afin d'éviter que des bactéries génétiquement modifiées puissent se disperser dans l'environnement. Elle a également pointé du doigt les risques d'une nouvelle course à l'armement liés aux technologies issues du génie génétique. Cette conférence était à l'initiative des chercheurs, ce qui selon Henry Korn « crédibilisa de manière considérable leur démarche ». De plus, « les participants ont été laissés libres de communiquer dans les medias sur leurs débats et sur les conclusions de la conférence, ce qui leur attira la confiance de l'opinion publique ». Je reviens maintenant au serment de

[3] Albert Einstein, *Comment je vois le monde*, Flammarion p. 34.
[4] Michel Serres, *Le trésor. Dictionnaire des sciences.*

Michel Serres. Je l'ai dit : je doute que les directeurs d'école doctorale et les présidents d'université soient en mesure d'imposer un tel serment à cause du lobby de l'armement. Par contre, la communauté IGEM, dont j'ai déjà parlé, pourrait avoir une grande influence, une carte à jouer. C'est une communauté internationale d'étudiants, de chercheurs passionnés de biologie synthétique qui est structurée autour d'une compétition annuelle. La communauté IGEM pourrait peut-être initier la mise en place d'un serment du type de celui de Michel Serres que les jeunes biologistes synthétiques pourraient prêter. Cela pourrait avoir deux effets :

⇒ crédibiliser la biologie synthétique auprès des opinions publiques mondiales pour montrer que ce sont des biologistes conscients de leurs responsabilités qui s'emparent de la question éthique (ce qu'ils font déjà) sans attendre les pressions politiques.

⇒ amorcer un mouvement qui pourrait se propager à toute la science. En effet, un tel serment prêté par tout nouveau chercheur serait, il me semble, sans précédent dans l'histoire des sciences. Cela rapprocherait la science et le peuple en montrant, via un symbole clair et ostensible, que, comme le dit Edgar Morin :

> *« Tout scientifique sert au minimum deux dieux. Le premier dieu est celui de l'éthique de la connaissance, qui exige que tout soit sacrifié à la soif de connaitre. Le second est le dieu de l'éthique civique et humaine. »*

# 5 Petit détour par la science fiction

Je vais maintenant passer à une partie plus personnelle, plus spéculative, proche de la futurologie et de la science fiction. Je suis d'ailleurs convaincu que la réflexion éthique devrait se nourrir davantage de la science-fiction (Besnier, Avril 2010).

Nous avons vu dans les paragraphes précédents que, suite à la convention de 1972, la plupart des états démocratiques ont abandonné les armes bactériologiques, au moins officiellement. Sans vouloir sombrer vers une quelconque théorie du complot paranoïaque, je doute que les raisons de cet abandon soient altruistes. En effet, si on étudie les traités de réduction des armes stratégiques entre l'ex-union soviétique et les états-unis (SALT, START et SORT), 40 ans de négociations ont permis de réduire les stocks de têtes nucléaires jusqu'à « uniquement » 1500 têtes nucléaires par pays. De quoi détruire sans doute encore un certain nombre de fois la terre. On voit donc que quand l'intérêt stratégique est réel, les efforts sont plus lents pour détruire les stocks d'armes. En fait, les armes bactériologiques sont peu intéressantes stratégiquement pour les démocraties car trop peu contrôlables. A l'inverse des missiles stratégiques, ce n'est pas une technologie fiable. Elle a très mauvaise presse dans les opinions publiques et est perçue comme

une arme sale relativement bon marché et facile à produire —une arme d'état voyou comme l'ex-Irak de Saddam Hussein—. Il valait donc mieux l'interdire à tous pour mieux « stigmatiser » et « criminaliser » les états voyous.

Pourquoi ce paragraphe polémique ? car les choses pourraient bien changer. La contrôlabilité des armes biologiques pourrait augmenter à cause de la biologie synthétique.

Même si il reste encore de nombreux défis technologiques à relever, le « wetware », c'est-à-dire le calcul avec de la matière humide (une cellule), pourrait faire son apparition et se développer très vite. Dés lors qu'on saura faire faire du calcul à une cellule, lui faire « computer » des comportements, des « outputs » complexes, le nombre d'applications possibles va exploser. Et les lanternes des militaires vont se mettre à briller. Le problème n'est pas tant de « tuer ». « Tuer », on sait déjà faire. L'enjeu est dans la définition de la biologie synthétique. L'enjeu, c'est le contrôle. Contrôle des armes et contrôle des hommes.

Quelle vision peut-on entrevoir de ces futures armes ? Au début, elles seront basiques. Une simple bactérie *E. coli* avec un plasmide inductible permettant la surexpression d'enzymes produisant des toxines (type cholérique) responsables de fortes diarrhées. L'attaque se produit en deux temps. La contamination (par exemple alimentaire) est silencieuse. Une des difficultés consiste à réussir à faire prendre le dessus à l'arme biologique sur la flore commensale. Une fois la souche synthétique en place et au moment propice, on passe à la deuxième étape. On ajoute dans l'alimentation de la victime, l'inducteur du système répresseur ou activateur (IPTG, arabinose) ce qui produit, de manière dose dépendante, l'enzyme puis la toxine. En fonction de la dose, on peut donc contrôler la force et la durée de la diarrhée, jusqu'à la mort. Des scenarios « hollywoodiens » de type « agent secret » avec chantages divers peuvent facilement être envisagés avec ce type de technologie. Plus « computationnelle » serait la même arme biologique mais dont le déclenchement est programmé par un minuteur (horloge moléculaire basée sur le principe des oscillateurs par exemple). Entre la primo-infection et le déclenchement de la diarrhée ou de la mort s'écoule le temps (en heures, en jours ou en semaines) programmé par le biologiste. On voit déjà qu'avec ces armes « très simples », on contrôle déjà la pensée de sa victime sans même contrôler son système nerveux. En effet, une victime qui a mal est une victime qui pense déjà à sa douleur et pas à autre chose.

Mon exemple est basique. Il montre ce que signifie le contrôle. On peut aussi infecter une population entière à son insu en vue d'un but futur. Ce but peut être la mort, la souffrance, la démence, le contrôle d'états mentaux plus fins etc... On peut aussi infecter sélectivement une population en créant des bio-armes qui ne ciblent qu'un type particulier de patrimoine génétique.

La bio-arme de la figure XIV.1 provoque 5 diarrhées à intervalle de temps régulier et n'at-

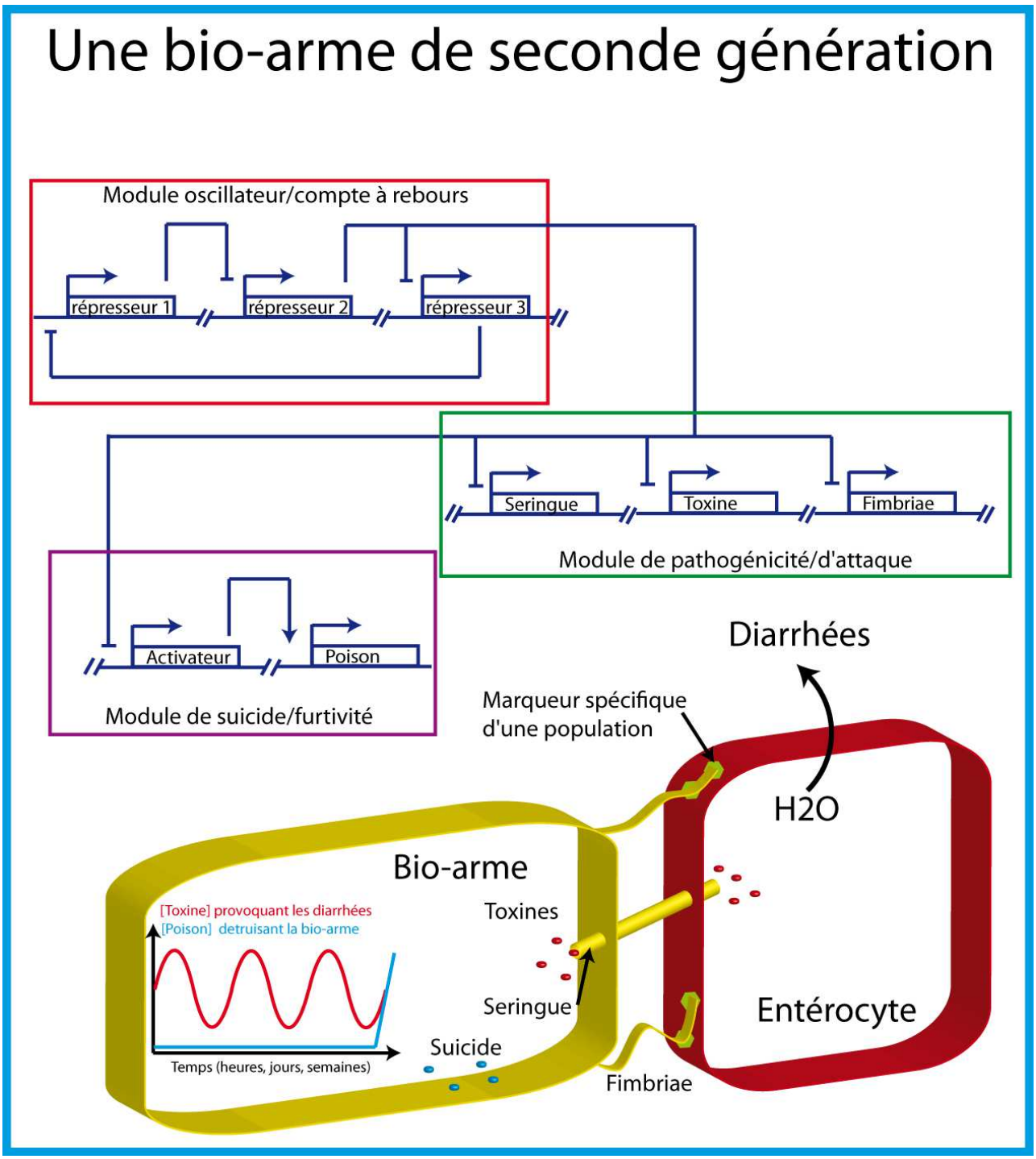



**Figure XIV.1** – Une bio-arme de seconde génération.

taque que les hommes (masculin). Elle fonctionne de la manière suivante. Le premier module correspond au repressilator : 3 répresseurs se répriment mutuellement ce qui génère des oscillations. Ces dernières sont utilisées comme compte à rebours. Les paramètres peuvent être facilement « tunés » pour adapter la fréquence des oscillations. Lorsque le deuxième répresseur disparait, un système de sécrétion, des fimbriaes ainsi que des enzymes produisant des toxines sont synthétisées. Les fimbriaes permettent l'adhésion aux entérocytes (cellules de l'intestin). De plus, ces derniers sont spécifiques de marqueurs que l'on trouve que dans une population donnée : par exemple que chez les hommes. Une fois en contact avec sa cible, la bio-arme injecte des toxines grâce à une sorte de seringue. Ceci induit une perte hydrique dans les entérocytes provoquant la diarrhée. Puis le deuxième répresseur réapparait ce qui stoppe le mécanisme d'attaque pendant un cycle. Un système suicide permet de détruire la bio-arme après un certain nombre de cycles. Concrètement, à chaque cycle, un activateur ultrastable est produit en petite quantité. Quand sa concentration dépasse un certain seuil, par exemple au bout de 5 cycles de diarrhées, ce dernier active l'expression d'un gène codant pour un poison cellulaire. La bio-arme est détruite.

Ce système est furtif pour plusieurs raisons : la bio arme ne peut pas être isolée car les cassettes de résistance aux antibiotiques ont été enlevées. Toutes les fonctions sont encodées à des emplacements différents sur le chromosome (pas de plasmide). La souche se suicide une fois sa mission achevée ce qui rend sa culture définitivement impossible. Si les gènes utilisés appartiennent tous à *E. coli* : ni une PCR, ni un séquençage du génome ne peut permettre de détecter facilement la technologie car chaque fonction est isolée. Seule une analyse poussée du réseau et des sites consensus peut permettre de comprendre le fonctionnement de l'arme.

Pour donner un ordre de grandeur, une bio-arme fonctionnelle de ce type (les fonctions de calcul sont basiques) serait, à mon avis, fabriquée en 5 à 20 ans par une équipe de recherche comprenant 5 à 10 chercheurs. Notez cependant que l'estimation est difficile et je peux me tromper lourdement dans un sens ou dans l'autre.

J'ai longtemps réfléchi (plusieurs mois) avant de publier ce schéma, d'autant que j'ai prêté le serment de Michel Serres. J'ai finalement jugé en conscience que les bénéfices sont supérieurs aux risques. Les risques sont : donner de mauvaises idées. Peu probable car premièrement ce schéma n'est pas une recette, deuxièmement les mécanismes décrits sont vagues, troisièmement tout étudiant IGEM arriverait très facilement à des conclusions similaires. Enfin, il n y' a rien de vraiment nouveau dans ce que je décris (publier la séquence du virus de la variole est autrement plus dangereux). Les bénéfices sont : faire prendre conscience aux scientifiques, aux pouvoirs publics et au peuple des risques de développement de bio-armes basées sur le Wetware/le calcul. Car je le répète, l'important, ce n'est pas la seringue, l'important c'est le contrôle à l'échelle

moléculaire. Quand on contrôle finement une souche, on peut contrôler finement un homme.

Malheureusement, la fiabilité d'une telle arme supprimera petit à petit aux yeux de l'opinion son image d'arme sale. Au contraire, elle risque d'être perçue comme un bijou de technologie, un véritable ordinateur miniature cellulaire. Voila la vraie arme d'une démocratie ! On peut théoriser des nouvelles approches de dissuasion de type « coupable *a priori* » pour dissuader une personne ou un peuple d'enfreindre les règles de « l'État de droit » : toute personne subit, à son insu, la primo-infection à sa naissance, l'arme est latente. Cette dernière attend patiemment la sortie du droit chemin de la victime pour la sanctionner d'une manière ou d'une autre. Avec le temps et l'habituation aux technologies issues de la biologie synthétique, ce type d'arme, dissuasive, non létale, capable de « sanctionner » pourrait bien avoir ses partisans et remplacer les prisons. Et le problème est là. Comme toutes les armes, *les bio-armes finiront par plaire.*

Ce type d'armes computationnelles pourrait infecter de manière latente un peuple petit à petit pour pouvoir, « en cas de problème », maitriser, dissuader, contrôler les opinions de son propre peuple ou du peuple du bloc opposé sans entrainer la mort. Imaginez juste les conséquences si on diminue très légèrement la qualité de vie statistique d'un peuple, en le rendant très légèrement malade à son insu de manière presque imperceptible. Le moral du peuple, la croissance, l'économie diminuerait légèrement à cause de la légère dépression à l'allure naturelle mais en réalité strictement synthétique, strictement voulue. Imaginez un peuple contaminé souffrant de diarrhées persistantes et faisant chuter l'économie du pays sans que les pouvoirs publics n'aient la solution : un traitement pour stopper le fléau. Imaginez alors maintenant l'état (le bloc) d'en face qui, comme par hasard, dispose d'un antidote, mais pas gratuit bien sûr.

Si les Américains ou Européens recevaient des signaux forts que Chinois ou Russes (et réciproquement) développaient des programmes d'armes biologiques de seconde génération à partir de la biologie synthétique, cela relancerait immédiatement la course à l'armement. La convention de 1972 ne serait plus qu'un lointain souvenir. Au fur et à mesure se développeraient de plus en plus de petits programmes biologiques malveillants. Au début ces derniers seraient créés par les états puis les particuliers ou groupuscules aux revendications plus ou moins claires prendraient le relais. Les états seraient obligés d'investir massivement dans des solutions de parade en créant donc des bio-armes anti bio-armes. Ces « logiciels biologiques » seraient chargés de surveiller l'homéostasie de l'hôte et de détecter et détruire toute intrusion suspecte.

Le couple virus-antivirus de l'informatique serait recréé mais en biologie cette fois et nous deviendrions au mieux des personnes analysées, au pire des personnes infectées. Cela pourrait d'ailleurs cesser d'être une lutte et devenir uniquement un marché. On paye pour être « updaté »

tous les jours avec les nouvelles séquences ADN détectées comme véritable « signatures » de bio-armes. Cela nous protègerait pour un temps des nouvelles menaces, des nouvelles souches malveillantes toujours plus furtives. Au bout d'un moment, on finirait par se demander si ce ne sont pas les sociétés antivirales qui créent les nouvelles menaces biologiques pour maintenir le marché... mais cette histoire vous la connaissez déjà...

## 6 Le danger invisible

Je pense que vous avez compris où je veux en venir. Le danger ne vient pas seulement par exemple des états voyous ou des universitaires fous. Bien sûr, une attaque, un drame peut toujours survenir mais selon moi l'attaque serait « rapidement » (c'est toujours relatif quand on parle de nombre de morts) stoppée, circonscrite. Des mesures politiques seraient immédiatement adoptées : attaque du pays voyou ou durcissement considérable des contrôles des chercheurs.

L'autre problème beaucoup plus sournois, beaucoup plus grave que j'identifie vient de nous, le peuple. Aucune personne dans le monde, je l'espère, ne souhaite voir se développer les pires « saloperies biologiques » que l'homme puisse créer dans le seul but d'en torturer un autre. Mais si maintenant je rajoute les frontières et quelques hommes politiques nous expliquant que *l'autre pays/bloc (toujours le méchant bien sûr !)* les développe et les utilisera contre nous, nous voterons « pour » la recherche et le développement des bio-armes. Et le peuple me payera, moi ou un autre, pour développer ses horloges et ses seringues moléculaires (oscillateur et système TSS) ou pour réfléchir à la furtivité des armes...

Rappelons-nous les propos d'Einstein qui n'ont, selon moi, pas pris une ride. Car tout le monde se rappelle qu'Einstein a été confronté à ce difficile problème éthique. C'est lui qui a écrit à Roosevelt le 2 aout 1939 en attirant l'attention sur le niveau atteint par la science allemande en matière de fission nucléaire. Et c'est cette lettre qui a déclenché le projet Manhattan (auquel Einstein n'a cependant pas participé) qui aboutira aux explosions nucléaires de Nagasaki et d'Hiroshima (entre 200 et 300 milles morts). En prenant une décision qu'on imagine très difficile, Albert Einstein a été un maillon clé du plus grand conflit dans l'histoire entre science et éthique. On peut donc rétrospectivement accorder un certain crédit aux idées qu'il a défendues dans les années qui ont suivi cette décision.

> *« Nous savons que la conscience collective est un petite plante misérable qui a coutume de se faner au moment où on en a le plus besoin »* [5]

[5] Albert Einstein, Lettre à Max Born, octobre 1953.

*« N'importe où, en 15 jours, une campagne de presse peut exciter une population incapable de jugement à un tel degré de folie que les hommes sont prêts à s'habiller en soldats pour tuer ou se faire tuer ».* [6]

*« L'industrie des armements représente concrètement le plus terrible péril pour l'humanité. Elle se masque, puissante force maligne, derrière le nationalisme, qui s'étend partout. »* [7]

*« Si jadis un homme incarnait une valeur aux yeux de la société quand il dépassait d'une certaine mesure son égoïsme personnel, on doit exiger de lui aujourd'hui qu'il dépasse l'égoïsme de son pays ou de sa classe. Seulement alors, arrivé à cette maitrise, il pourra améliorer le destin de la communauté humaine. »* [8]

Pourquoi rappeler ces propos d'Einstein ? Pour bien faire comprendre à mon lecteur que les scientifiques sont des hommes avant d'être des savants et comme le philosophe moraliste Bertrand Russell le dit

*« la science, en soi, ne peut pas nous fournir un éthique. Elle peut nous indiquer comment atteindre un objectif donné et, parfois, nous montrer que certains objectifs sont inaccessibles. Mais parmi ceux réalisables, notre choix doit être guidé par des considérations autres que purement scientifiques. »*

On connait bien les liens de dépendances mutuelles très étroits qui existent entre scientifiques et militaires :

*« Aujourd'hui plus aucune armée n'est capable de vaincre sans la contribution des scientifiques et rares sont les chercheurs et ingénieurs capables de gagner la bataille de leur argumentation sans l'aide des militaires. »*[9]

Il suffit de regarder certains chiffres publiés par l'OCDE pour s'en convaincre :

*« Plus de la moitié du budget de recherche et développement (R&D) du gouvernement des États-Unis est affecté à la défense. Au Royaume-Uni, le budget de R&D en matière de défense représente plus du tiers de l'ensemble des dépenses publiques de recherche tandis qu'en France comme en Espagne, il s'élève à environ un quart du total. En 1999, les États-Unis ont affecté 0,45% de leur PIB à la R&D en matière de défense. Le Royaume-Uni et la France se classent juste derrière avec respectivement 0,26% et 0,22%. »*[10]

[6] Albert Einstein, *Comment je vois le monde*, Flammarion p. 19.
[7] Albert Einstein, *Comment je vois le monde*, Flammarion p. 78.
[8] Albert Einstein, *Comment je vois le monde*, Flammarion p. 117.
[9] Bruno Latour, *La science en action*, La découverte P. 414
[10] Source : L'observateur OCDE, base de données MSTI

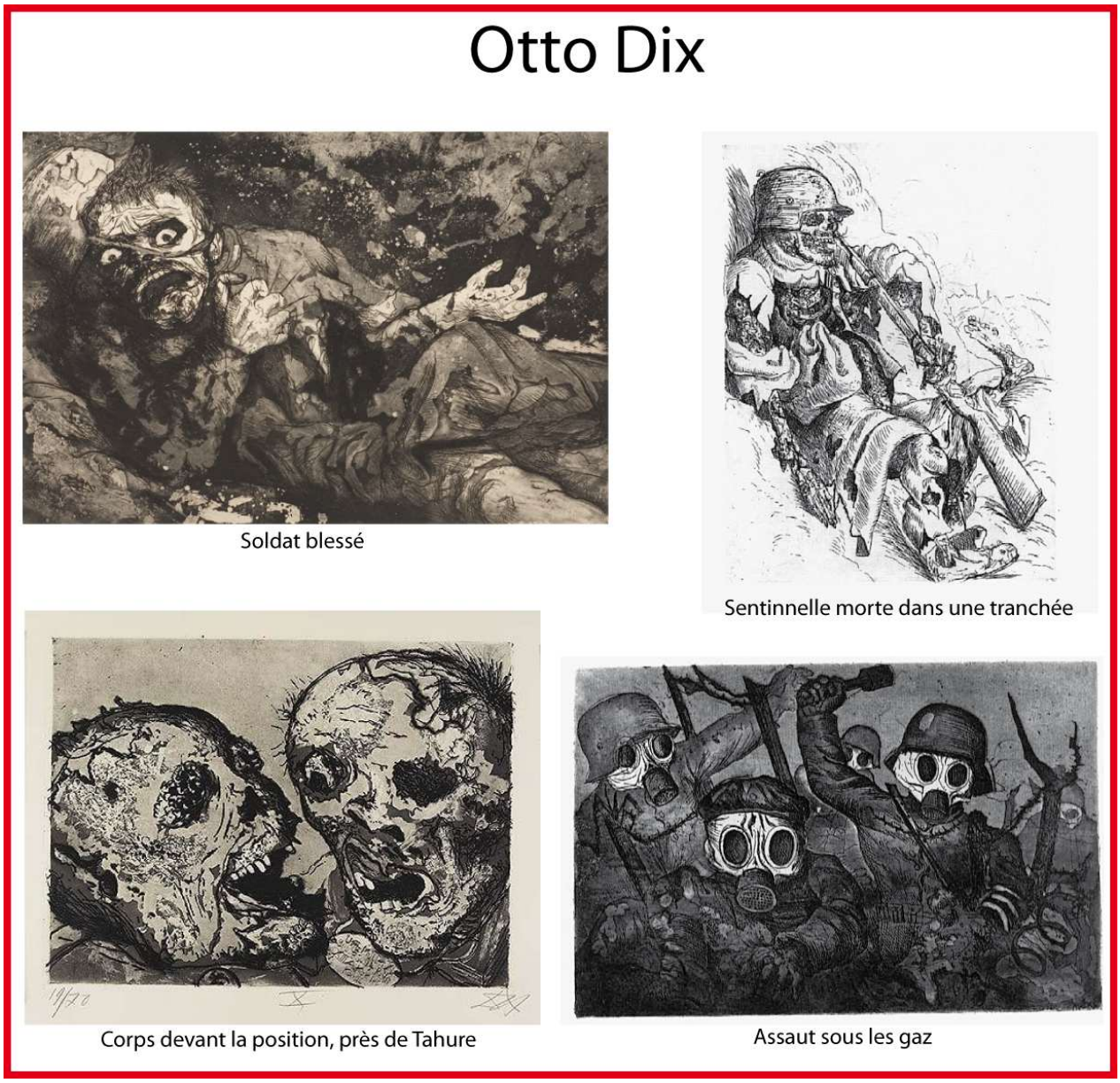



**Figure XIV.2** – Peintures de Otto Dix illustrant l'horreur de la première guerre mondiale

Certains scientifiques fabriqueront des armes biologiques pour de l'argent, pour le pouvoir ou pour la gloire. Citons Robert Oppenheimer faisant remarquer que, suite au développement de l'arme nucléaire, les physiciens avaient *« connu le péché »* et la réponse de Von Neumann *« parfois on confesse un péché pour s'en attribuer le crédit »*. Je n'ose même pas commenter.

D'autres attendront d'être soumis à une propagande forte avant de céder. D'autres encore (peut-être comme Einstein) céderont par conviction humaniste tout en doutant profondément que la démarche bénéficie réellement à l'humanité. Enfin, certains ne céderont jamais. Mais

pour cela, nous allons le voir, il faut penser par delà bien et mal[11].

Je pense à ces hommes qui ont peut-être eu à la possibilité d'inventer la poudre mais ont renoncé par conviction. Et les millions d'autres hommes qui ont été perforés par une balle, déchiquetant un organe, une artère, brûlant les tissus. Peut-être pensent-ils, pendant les quelques minutes qui les séparent de la faucheuse, pendant que la marre de sang dans laquelle ils baignent commence à grossir après avoir fini d'imbiber leurs vêtements, peut-être pensent-ils à *ces inconnus qui n'ont pas inventé la poudre alors qu'ils le pouvaient.*

Mais le fait que certains ne cèdent jamais à la tentation d'utiliser leur intelligence pour fabriquer une arme ne change rien au fait que la fabrication de la dite arme est inexorable. Il n'existe aucune arme que la science ait entrevue et qui n'ait pas été fabriquée pour des raisons éthiques. Dés lors qu'il y a possibilité scientifique, il y a toujours eu existence. C'est un fait.

Ma prédiction est donc la suivante. Des armes similaires aux armes futuristes décrites plus haut existeront. Et la faute ne reviendra pas aux « méchants » mais à *nous, peuple* incapable de renoncer à la recherche, au développement et à la production des armes quelles qu'elles soient. Le peuple ne peut pas éviter l'arrivée de ces armes qu'il redoute à cause de certains de ses schémas mentaux trop profondément enfouis. La psychanalyse nécessaire des peuples à l'échelle mondiale pour vaincre ses schémas est peut-être possible mais également de l'ordre de la spéculation.

L'horreur absolue engendrée par le nazisme a eu pour conséquence de faire réfléchir les peuples. En refusant la guerre, des hommes comme Jean Monet ou Robert Schuman ont pensé puis créé l'Union Européenne. Je crains que l'humanité « n'ait besoin » d'aller encore une étape plus loin dans l'horreur suprême (si cela est possible) pour permettre, si elle en survit, de tenter de refuser définitivement les armes tout comme Monet et Schuman ont refusé la guerre. Si malheureusement cette horreur suprême arrive, elle pourrait bien reposer sur un instrument nouveau : les armes biologiques de seconde génération. Celles-là même que les peuples auront voulu à cause des frontières.

# 7 Les racines du mal

Un élément responsable de l'inexorabilité actuelle du développement de nouvelles armes est liée au fait que, dans notre monde, la violence est légitime. L'état peut user de la violence contre ses citoyens hors du contexte de légitime défense : par exemple en forçant un prisonnier à se dévêtir. Ainsi selon Max Weber, l'état détient le monopole de la violence. Il est le seul à

[11] En référence, bien sûr, à l'ouvrage de Nietzsche.

bénéficier du droit de mettre en œuvre la violence physique sur son territoire. Cette violence est considérée comme légitime. Elle est confiée aux policiers et aux militaires. Or la violence nécessite des armes que les scientifiques développent.

> *« Il faut concevoir l'Etat contemporain comme une communauté humaine qui, dans les limites d'un territoire déterminé, revendique avec succès pour son propre compte le monopole de la violence légitime. »*[12]

Il nous est impossible de nous priver de la capacité potentielle de *faire mettre l'autre à genou*. Dans nos sociétés modernes, il existe une multitude de moyens pour *faire mettre l'autre à genou* (moyen financier, hiérarchique, psychologique...) mais le moyen ultime qui reste lorsque *l'autre* résiste c'est la force qui passe par la supériorité —physique, intellectuelle, du nombre ou des armes—.

En cas de menace grave, la complexité et les boucles de rétroaction rendent la prise de décision difficile pour un scientifique. Il serait idiot de le nier. Ainsi, selon Edgar Morin,

> *« Einstein s'est cru profondément responsable devant l'humanité quand il a, dans un premier temps, lutté contre tous préparatifs militaires. Il s'est cru encore plus responsable devant l'humanité quand il est intervenu instamment pour la fabrication de la bombe atomique [...] Un acte d'individu ou de groupe entre dans un complexe d'inter-rétroactions qui le font dériver, dévier et parfois inverser son sens ; ainsi une action destinée à la paix peut éventuellement renforcer les chances de guerre. Inversement, une action renforçant les chances de guerre peut éventuellement œuvrer pour la paix (intimidation). Il ne suffit donc pas d'avoir de bonnes intentions pour être vraiment responsable. La responsabilité doit affronter une terrible incertitude »*[13]

Voici un autre exemple de boucle de rétroaction. Qu'est-ce qui peut me faire céder à la tentation de mettre mon intelligence à contribution sur un but opposé à la dignité humaine ? L'identification d'une menace que je reconnaitrais comme un mal absolu. Pour l'enrayer, je crée un autre mal absolu : je développe une arme. Et la boucle est bouclée. Voyons maintenant ce qu'est le mal absolu.

Selon le philosophe Hans Jonas, auteur du *principe responsabilité*, la technique recèle en son sein deux types de menaces différentes :

⇒ l'anéantissement physique (par exemple une guerre nucléaire)

⇒ le dépérissement existentiel *« qui peut conduire à une mise sous tutelle éthique, par exemple à travers l'automatisation de tout travail, le contrôle psychologique et biologique*

[12] Max Weber, *Le savant et le politique*, Bibliothèque p.125

[13] Edgar Morin, *Science avec conscience*, Point p. 109.

*des comportements, les formes de domination totalitaire, voire à travers une modification du reconditionnement génétique de notre nature »* [14] . Le lecteur pensera immédiatement aux deux contre-utopies célèbres *Le meilleur des mondes* et *1984* .

Il existe des nuances parmi ces menaces suprêmes. Je peux par exemple céder à la tentation de fabriquer une arme pour empêcher soit l'anéantissement de l'humanité (version faible) soit l'anéantissement de toute vie (version forte). Je peux aussi par exemple céder à la tentation de fabriquer une arme pour empêcher le dépérissement existentiel faible : toutes sociétés futures qui nous apparaitraient, à nous maintenant en 2011, comme une négation insupportable de l'essence de l'homme, par exemple une société sans libre arbitre apparent. Mais je peux aussi vouloir empêcher le dépérissement existentiel fort défini comme étant une souffrance physique et morale maximum et constante infligée à tout homme pour le reste de l'éternité.

Tout scientifique confronté à une menace qu'il juge insupportable peut se retrouver face au dilemme consistant à choisir si, oui ou non, il mettra son intelligence à contribution pour fabriquer une arme. Quelque soit son choix, à cause de la complexité, son action ou inaction peut entrainer les pires de scenarios. Faisons maintenant une expérience de pensée (elle avait déjà un peu commencé). Vous devez choisir de mettre à contribution ou pas votre intelligence pour fabriquer une arme. Vous savez que l'un de ces choix seulement a pour conséquence l'anéantissement de toute vie sur terre. A cause des boucles de rétroaction complexes, vous n'êtes pas en mesure d'évaluer les probabilités respectives associées à ces choix. Vous êtes dans l'incertitude la plus totale et votre choix revient donc à tirer à pile ou face. Dans l'expérience de pensée, cela permet de se placer « par delà bien et mal » et donc d'écarter toute question morale. Quoi qu'il arrive vous n'êtes pas responsable de votre choix puisque seul le hasard orientera le futur. Cependant, vous êtes toujours en mesure de prendre en compte des considérations personnelles pour faire votre choix. *Ce que vous voulez vous.* En effet, quitte à ce que la vie et l'humanité disparaissent, que voulez-vous que votre décision symbolise ? Car ce qu'est l'essence de l'homme sera associé à votre décision pour le reste de l'éternité. Si vous choisissez de fabriquer l'arme, vous choisissez d'associer au mot homme, les mots lutte, combat, violence, volonté de puissance. A l'inverse si vous choisissez de *ne pas fabriquer l'arme*, vous choisissez d'associer au mot homme, les mots paix et amour. Quel est le meilleur choix ? Et d'ailleurs existe-t-il un meilleur choix ? Comme élément de réponse, je ferai remarquer à mon lecteur que si vous choisissez de fabriquer l'arme, *vous êtes comme ce qui vous place face à ce choix*, c'est-à-dire la violence. Vous êtes ce qui vous place face à ce choix. Tandis que si vous décidez de ne pas fabriquer l'arme, *vous êtes autre chose, vous êtes différent. Vous êtes en plus.*

[14] Hans Jonas, *Pour une éthique du futur*, Payot-rivages, p. 100.

Alors, cher lecteur, avez-vous choisi ?

Peut-être trouveriez-vous de l'inspiration dans le dernier paragraphe du discours de la méthode écrit par René Descartes en 1637 ?

> *« Et que m'on inclination m' éloigne si fort de toute sorte d'autres desseins, principalement de ceux qui ne sauraient être utiles aux uns qu'en nuisant aux autres, qui si quelques occasions me contraignaient de m'y employer, je ne crois point que je fusse capable d'y réussir. »*

# Références bibliographiques

Besnier, Jean-Michel. Avril 2010. La reflexion ethique devrait se nourrir davantage de la science-fiction. *Pour la science*. 323

Cello, J., Paul, A. V., & Wimmer, E. 2002. Chemical synthesis of poliovirus cDNA : generation of infectious virus in the absence of natural template. *Science*, **297**(5583), 1016–8. 320

Kobinger, G. P., Weiner, D. J., Yu, Q. C., & Wilson, J. M. 2001. Filovirus-pseudotyped lentiviral vector can efficiently and stably transduce airway epithelia in vivo. *Nat Biotechnol*, **19**(3), 225–30. 320

# Chapitre XV

# Futur lointain

## 1 La médecine du futur

Il va s'opérer dans les décennies qui arrivent, probablement au cours du 21$^{\text{ème}}$ siècle, une transition de la médecine 1.0 vers la médecine 2.0.

La médecine 1.0 est une médecine principalement empirique. La chimie y est au centre. On nous prescrit des médicaments constitués de molécules capables « d'attaquer » la pathologie. La médecine 1.0 est une guerre dans laquelle la pathologie est « la méchante » et le médicament « le gentil ». Avec la médecine 2.0, les choses changent. Les connaissances acquises en biologie au cours de la deuxième moitié du 20$^{\text{ème}}$ siècle sont intégrés dans les thérapeutiques. C'est la biologie et non plus la chimie qui devient le centre de cette nouvelle médecine. Je pense aux thérapies géniques ou aux thérapies cellulaires par exemple. La connaissance du génome humain donnera des possibilités d'agir à la racine, en sélectionnant des génomes « sains » et donc en écartant les mutations ou allèles « malsains ». On pourra peut-être augmenter le déterminisme en modifiant ou supprimant des gènes donnant des prédispositions statistiques aux cancers, à l'alcoolisme, à la schizophrénie, au diabète... Les parents pourront choisir « dans un catalogue » un certain nombre de « phénotypes » qu'ils veulent voir chez leurs enfants, en commençant par le sexe puis peut-être la taille, la couleur des yeux, certaines prédispositions statistiques à l'intelligence...

Les efforts en biologie synthétique permettront peut-être la création d'un système immunitaire suppléant. Des petites cellules, dotées des technologies computationnelles, seront capables de détecter et de détruire des cellules cancéreuses qui auraient échappé au système immunitaire de l'hôte par exemple.

Les efforts pour recréer de nouveaux organes (par exemple un foie ou un poumon) à partir de cellules indifférenciées d'un patient malade permettront peut-être de prolonger les durées de vie. En effet, on pourra remplacer systématiquement tout organe malade par un neuf, un peu

comme lorsque l'on remplace le démarreur défectueux d'une voiture au garage.

La mort reculera. L'espérance de vie augmentera. Mais il y a une chose sur laquelle la médecine 2.0 risque de buter : c'est le cerveau. Certaine pathologies neurologiques risquent de nécessiter des connaissances plus poussées car on ne pourra pas simplement changer de tête. En effet, on ne sait pas stocker la conscience sur un autre support physique (disque dur) pour la recharger dans la nouvelle tête. La médecine 2.0 ne gère pas la complexité. C'est sa limite. Elle peut agir sur le génome. Elle peut fabriquer des cellules médicaments en câblant un réseau rudimentaire, offrant quelques possibilités computationnelles mais c'est tout. Tout ce qui relève de réseaux complexes c'est la médecine 3.0 que nous aborderons plus tard. Mais alors, à quoi bon vivre plus longtemps si le cerveau ne suit pas ? Ce sera une des questions éthiques à laquelle il faudra répondre.

Une des particularités de la médecine 2.0 c'est qu'elle amorce un tournant en tendant à supprimer l'image de guerre très ancrée de la médecine 1.0. En effet, au fur et à mesure qu'augmente l'espérance de vie, le concept de « pathogène » perd petit à petit de son sens. Cancer, virus, bactéries, prions : la technologie suffit à en venir à bout, parfois à la racine, en modifiant le génome des cellules totipotentes de la morula (l'embryon du futur enfant). Il y a de moins en moins de « méchants », mais plutôt l'usure d'un cerveau fatigué, usure que l'on ne sait pas encore traiter.

La dissociation du corps et de l'esprit est en marche. En l'absence de menaces extérieures, l'importance ressentie du corps diminue petit à petit. Nous percevons celui-ci comme une vulgaire machine et ce sentiment s'accroit avec la bidouillerie du génome. Vous n'êtes plus votre corps mais uniquement votre esprit. On sait, qu'au cours de l'histoire, l'homme a mis du temps à associer son « âme » à sa tête et non à son cœur. Une transition similaire s'opérera petit à petit car le contrôle de l'homme sur ce corps change la représentation mentale qu'il en a et le transforme petit à petit en une machine, un « exo » squelette, en dehors, en plus de *son soi*.

## 2 Changer le système

[Un an après l'écriture de cette partie, j'inclus ce petit commentaire entre-crochet qui fait office de légère dépréciation : cette partie conserve, il me semble, des qualités c'est pourquoi je la maintient mais elle est trop naïve, elle risque de générer des contre-sens et ne correspond plus exactement à mes convictions philosophiques]

Je rentre maintenant dans un domaine où mes idées pourraient être taxées de naïves ou médiocres. Je l'accepte et je prends le risque. Tout dans ce chapitre sur le futur n'est de toute

façon que pure futurologie, spéculation. En tant que futurologie, mon texte peut difficilement séparer le descriptif (ce qui sera) du normatif (ce qui devrait). Oui, il peut/doit être entrevu comme une mise en garde mais totalement découplée de la politique actuelle. Pour être clair, je ne tente pas de distiller en cachette mes opinions politiques. Je tente de réfléchir, de manière detachée, aux conséquences potentielles de l'apparition des technologies de la médecine 2.0 si le système politico-économique reste figé, c'est-à-dire tel qu'il est actuellement.

L'industrie pharmaceutique fonctionne actuellement, il me semble, sur le principe suivant : elle investit massivement dans la recherche et développement pour développer des nouvelles thérapeutiques. Ces thérapeutiques ne sont pas distribuées gratuitement aux malades qui en ont besoin. Elles sont achetées par les consommateurs malades ou les systèmes d'assurance maladie des états ce qui permet le retour sur investissement. De nouveaux fonds peuvent être alors alloués dans la recherche et développement de nouvelles thérapeutiques pour combattre de nouvelles pathologies ou des pathologies pas encore vaincues. Ce que je décris, c'est le noyau (le fond) tel que je le perçois. Je sais bien que l'écorce (la surface) est composée d'un mercantilisme froid parfois même très nauséabond.

Ce système est inégalitaire et injuste parce qu'on sait que de nombreux malades (en particulier dans les pays en voie de développement) n'ont pas accès aux médicaments trop chers. La conséquence est simple : ils en meurent ou souffrent énormément alors même que la solution, la pilule miracle existe. C'est un fait, c'est le système dans lequel nous vivons et s'il n'est pas parfait, en revanche, il est perfectible. Tout acteur censé en a conscience. Refuser de reconnaitre la perfectibilité du système actuel c'est être, selon moi, soit fou soit aveugle soit égoïste.

Le système actuel fonctionne (ou prétend fonctionner) sur un régime utilitariste : il prétend maximiser (tenter de maximiser) l'accès aux nouveaux traitements au plus grand nombre. Certes un africain pauvre séropositif ne pourra pas bénéficier immédiatement des dernières innovations en matière de traitement anti-VIH réservées au début aux personnes en mesure de payer (les pays riches). Mais une fois le retour sur investissement réalisé, le traitement finira par devenir accessible aux pauvres (médicaments génériques, brevets qui tombent dans le domaine public au bout de 20 ans) alors que ces derniers n'auraient jamais pu en bénéficier si l'industrie pharmaceutique n'avait pas eu les moyens (les retours sur investissements) pour les développer. Ces propriétés en apparence utilitaristes ne sont pas un objectif mais plutôt une conséquence. L'objectif premier de toute entreprise restant, par définition, le profit.

La question est donc la suivante : ce système pseudo-utilitariste reposant sur les principes du système économique mondial est-t-il compatible avec l'arrivée future de la médecine 2.0 dans la vie de tous les jours ? Je ne le pense pas.

Dans le domaine de la biologie synthétique, le « toggle switch » (l'interrupteur molécu-

laire que j'ai présenté dans le chapitre sur les systèmes dynamiques) a été breveté. Est-ce-que c'est « bien » ? Est-ce-que ce brevet va dans le sens de « plus d'humanité », d'une « humanité meilleure » ? Je n'en ai aucune idée. Mes connaissances en économie sont trop faibles pour que je me prononce sur une telle question. Je donne cet exemple pour convaincre mon lecteur que j'essaie de m'éloigner au maximum de toute idéologie politique.

Avant de commencer, je précise tout de suite que l'analyse faite dans le paragraphe qui suit est spéculative même si les chiffres ne le sont pas[1]. La question soulevée par l'analyse du système actuel est la question des inégalités. Ces inégalités peuvent être observées sur le simple critère de l'espérance de vie : 82 ans en France ; 35 ans en Angola. Il me semble (mais c'est à vérifier) que l'augmentation de l'espérance de vie durant les siècles derniers a été relativement peu corrélée aux avancées de la médecine même si l'arrivée des antibiotiques a joué un rôle incontestable. Les paramètres importants sont plutôt l'hygiène (fabrication des égouts par exemple), les guerres ou les famines. Dans les pays développés, les différences observées entre classes sociales[2] ne sont pas tant liées, il me semble, à la différence d'accès aux soins mais plutôt corrélées à la pénibilité physique des travaux exercés, la consommation d'alcool et tabac ainsi que la qualité de l'alimentation et le sport.

Mais les choses pourraient bien avoir changé avec la pandémie de VIH qui a fait plus de 25 millions de mort depuis 1981. 1 millions 600 pour la seule année 2007 en Afrique. Certains pays africains ont plus de 25% de leur population séropositive. Dans ce cas, l'absence d'accès à des traitements antirétroviraux efficaces devient *le paramètre majeur* qui diminue l'espérance de vie.

Avec l'arrivée au cours du 21$^{\text{ème}}$ siècle de la médecine 2.0, les écarts pourraient s'accentuer très fortement (ils sont déjà si grands). Le prix de certaines thérapeutiques pourrait exploser dû au fait que la médecine va devenir « individualisée » car fonction du patrimoine génétique. Les riches pourront s'offrir des traitements issus d'une technologie extrêmement poussée qui sera par conséquent hors de prix pour les pauvres. Les riches pourront choisir dans un catalogue les caractéristiques de leurs enfants (minimiser les risques de cancers, de Parkinson, d'alcoolisme, maximiser la taille, l'intelligence etc...). Les pauvres n'y auront pas accès ou s'endetteront à vie pour offrir ce luxe à leurs enfants. Petit à petit les écarts d'espérance et de qualité de vie augmenteront : on peut imaginer que certains hommes riches atteindront 170, 180, 190, 200 ans. A l'inverse, l'espérance de vie n'augmentera que très lentement pour les pauvres n'ayant qu'un accès limité à la technologie (40, 60, 80, 100 ans).

[1] Source Wikipedia : article sur le SIDA et l'espérance de vie.

[2] Entre 6 à 10 ans (suivant les sources) de différence d'espérance de vie à 35 ans entre un cadre supérieur et un ouvrier.

Vous allez me dire : tout le monde bénéficiera de la technologie puisque l'espérance de vie augmentera petit à petit pour tout le monde. Et donc le peuple « pauvre » devrait se tenir tranquille.

Or les choses ne sont pas aussi simples. On sait grâce à des expériences d'économies comportementales que l'*homo sapiens* ne se comporte pas toujours comme l'*homo oeconomicus*. Il ne maximise pas toujours son intérêt personnel et fait donc parfois des choix « irrationnels » : il sanctionne l'injustice. Cela a été découvert avec le jeu de l'ultimatum : deux joueurs qui ne se connaissent pas interagissent ensemble pour décider comment se diviser une somme d'argent : 100 euro par exemple. Le premier joueur fait une proposition dans la manière d'opérer la division. Par exemple : 50/50, 70/30, 90/10. Et le second joueur décide s'il accepte ou refuse la proposition. S'il accepte la proposition, les deux joueurs repartent chacun avec la somme correspondant au partage proposé. S'il refuse, aucun des deux joueurs ne touche rien. Le jeu est joué qu'une seule fois. Tout partage proposé par le premier joueur supérieur ou égal à 99/1 devrait rationnellement être accepté par le second joueur car ce dernier gagnera 1 euro ce qui est mieux que 0 euro s'il rejette l'offre. Or ce qu'on observe, c'est que les joueurs ont un sens aigu de la justice. Beaucoup offrent le partage le plus équitable 50/50. Et en dessous de 20 euro (partage 80/20), très souvent le second joueur refuse l'offre et donc préfère perdre 20 euro pour sanctionner l'autre joueur de son manque d'équité, de justice.

Revenons maintenant à notre question en remplaçant l'argent par l'espérance de vie : si la technologie offre 100 ans d'espérance de vie aux riches et offre quelques miettes aux pauvres (5 ans), ces derniers devraient rationnellement, en tant que *homo oeconomicus*, s'estimer heureux. Mais ils préféreront en réalité interdire aux riches, s'ils le peuvent, la technologie offrant 100 ans d'espérance de vie, quitte à perdre les 5 ans car l'injustice, l'inégalité est trop criante.

Selon Merton, la science a pour fondement[3] :

⇒ *L'universalisme* : les connaissances scientifiques doivent être considérées indépendamment de leurs producteurs.

⇒ *Le communisme* (ou communalisme pour éviter les confusions) : les connaissances scientifiques doivent rester des biens publics.

Ainsi les connaissances scientifiques et leurs utilisations (par exemple une nouvelle technologie issue de la médecine 2.0) n'appartiennent pas aux personnes qui ont amené le capital. En effet, ce capital n'est qu'une goutte d'eau par rapport aux millénaires de science nécessaires à la formation de la pyramide ayant permis d'aboutir à la dite technologie. Nous verrons dans le deuxième chapitre d'épistémologie que le chercheur est un soldat, pas un mercenaire. Son

[3]Source Wikipedia : Article sociologie des sciences.

sacrifice est bien supérieur à sa solde. On n'imagine pas qu'un soldat qui prend une balle à la guerre le fait pour sa solde à la fin du mois. Un soldat a besoin de convictions : le patriotisme, la liberté, la bataille contre le mal... Le scientifique c'est la même chose. Mais ces convictions sont l'universalisme et le communalisme de ses découvertes. Et le peuple le sait. Dés lors, ce dernier ne peut accepter une injustice trop lourde sur l'accès à la technologie.

Le peuple a parfaitement conscience que personne sur terre n'a fait quelque chose de si extraordinaire durant sa vie qu'il mérite de vivre mieux et plus longtemps qu'un autre homme. Le peuple sait aussi que les meilleurs des hommes, les plus grands humanistes, ceux qui ont consacré leur vie à faire le bien autour d'eux : tous ces hommes cracheraient sur « un bonus de vie », si le reste de l'humanité, leurs frères, n'y avaient pas droit.

La technologie qui génère des « années de vie supplémentaires » est une technologie trop sensible. Elle est fondamentalement faite pour accroitre l'espoir, pour accroitre le confort, pour améliorer la qualité de vie des hommes sur terre. Or une technologie qui vend des « années de vie supplémentaires » risque de générer l'inégalité et donc l'injustice et donc la rancœur. Cette rancœur mènera à la guerre, à la mort et aux armes biologiques. Ce n'est pas uniquement une question d'espérance de vie : en touchant au génome humain de manière inégalitaire (via des prix inaccessibles), on risque d'opérer des scissions dans l'espèce humaine si les génomes deviennent trop différents. On devra inventer du nouveau vocabulaire : les surhommes, les hommes, les sous-hommes. Peut-être préférez vous les lettres : hommes de type A, hommes de type B, hommes de type C etc... Vous voyez donc bien que ce que je prône c'est l'égalité, pas l'égalitarisme. C'est la protection de l'essence de l'homme comme entité indivisible. On ne peut pas se permettre une hétérogénéité génomique trop forte. On ne peut pas se permettre des multi-stabilités dans la distribution des génomes même si le système qui les génère est utilitariste. La scission des espèces engendrée par un système économique inadéquat, si elle est subie par le peuple, mènera à la guerre inter-espèce.

Ainsi le système économique doit subir des adaptations ou un re-fondement de telle manière à ce qu'il soit *subordonné* au maintient, coute que coute, d'une essence unique et indivisible de l'homme. Ce système pourrait *éventuellement* maintenir une inégalité sur la plan matériel (voiture, ordinateur, service...) si celle-ci sert à toujours mieux protéger l'égalité sur le plan « essentiel » : l'espèce est une et indivisible et on n'achète pas des années de vie comme on achète une voiture. Pour faire simple, un milliardaire peut se payer, s'il le souhaite, un voyage dans l'espace mais il ne peut pas se payer l'immortalité : il doit attendre que tous ses frères, les hommes, y aient accès également.

[Ajout un an plus tard : derrière cet *éventuellement* placé en italique dans le paragraphe ci-dessus se cache ma volonté de ne pas rentrer dans le domaine politique. Je souhaite néan-

moins préciser que je n'ai pas un goût prononcé pour l'exploitation de l'homme par l'homme. Nietzsche juge cette exploitation comme une des conditions même de la vie. Je ne suis pas encore convaincu.]

# 3 Les alternatives futures

J'ai beaucoup insisté dans la partie précédente sur la nécessaire « sérénité » que l'humanité doit rechercher en parallèle aux progrès de la biologie. Réussir ou non dans cette entreprise de la sérénité c'est orienter le futur dans l'une des directions qui suit :

## 3.1 Les anéantissements

### 3.1.1 Anéantissement de tout homme sur terre

Il se pourrait que l'humanité aille droit dans le mur et mette un terme à son existence de manière subie ou voulue. Cet anéantissement pourrait être vu comme un choix de la nature décidant de se débarrasser de l'homme. C'est ce que nous fait remarquer Hans Jonas :

> *« La nature peut-elle supporter l'esprit qu'elle a fait naitre de son sein ? Faut-il que trop harcelée par lui, elle l'élimine en revanche de son système ? L'esprit est-il en mesure de se rendre supportable en fin de compte à la nature s'il s'aperçoit qu'elle ne le supporte pas ? »* [4]

Pourtant, selon Nietzsche :

> *« Nous préférons tous la ruine de l'humanité à la ruine de la connaissance »* [5]

Autrement dit, si l'homme devait choisir entre :

⇒ L'anéantissement de toute connaissance mais la survie de l'humanité (à l'état de bovin : c'est-à-dire plus aucun progrès au cours du temps).

⇒ L'anéantissement de l'humanité mais la sauvegarde de la connaissance sur des supports appropriés et potentiellement accessibles à une intelligence future.

Il choisirait la deuxième possibilité. Pourquoi ? Parce qu'il est habitué à vivre avec une volonté de puissance dont « la dérivée » lui apparait bien supérieur à celle d'un bovin. Le retour en arrière serait trop lourd pour être acceptable. Car bien sûr, la volonté de vivre n'existe pas. Seule la volonté de puissance existe. L'opportunité, même hypothétique, de transmettre son savoir (c'est-à-dire la démonstration d'une volonté de puissance qui s'est affirmée, malgré la

[4] Hans Jonas, *Pour une éthique du futur*, Payot-rivages p. 65.
[5] Stephan Zweig, *Nietzsche*, Stock p. 76

souffrance, pendant des siècles) à une « intelligence future »(c'est-à-dire une volonté respectable car puissante et en excès sur elle-même) lui semble de loin préférable.

### 3.1.2 Anéantissement de toute vie sur terre

On peut imaginer la création d'armes de destruction massive détruisant toute vie et toute trace de vie sur terre (par la désintégration complète de toutes les molécules en atomes par exemple). Dans ce cas, on ne peut plus considérer que c'est l'œuvre de la nature puisque celle-ci est volonté de puissance. Elle ne peut donc pas entrainer son anéantissement.

On voit ici que le concept de mort peut se décliner en différentes combinaisons : Mort de la molécule, mort de la cellule, mort de l'organisme (celle que l'on connait), mort de l'espèce, mort de la vie. Schopenhauer pense que si on donnait *« à un individu le choix d'être anéanti, ou de voir anéantir le reste du monde : [il serait inutile de préciser] de quel coté, le plus souvent, la balance pencherait »* [6] . Je ne vois pas les choses de cette manière. En effet, l'homme a conscience de la présence des deux derniers niveaux de la mort qu'il juge, plus ou moins consciemment, pire que sa propre mort. Par exemple, l'instinct maternel peut amener des parents à se sacrifier pour leurs enfants. Je pense qu'il est même concevable que soit encodée dans le génome (et surgissant via le système nerveux) une fonction rendant possible le fait de surmonter la peur de sa mort (et donc sacrifier sa vie) pour ces deux causes : mort de l'espèce, mort de la vie.

## 3.2 Arrêt de la recherche, décroissance et cycle

Il se pourrait aussi que l'homme se persuade, suite à des abominations diverses, que la recherche ne peut mener qu'à l'anéantissement physique ou au dépérissement existentiel. Les hommes se concerteraient pour stopper la recherche, « le progrès » dans le but de se maintenir à un *statu quo ad vitam aeternam*. On peut l'imaginer mais il n'existe pas de preuve que cela soit possible. De plus, le risque de sombrer vers un système totalitaire ressemblant à celui de George Orwell *1984* (dépérissement existentiel) est grand. En effet, que fait-on aux gens qui cherchent en cachette ? On les tue ? On les redresse ?

Une autre alternative consiste à envisager le futur sous l'angle de la décroissance. Il ne s'agit pas d'un retour au passé : il s'agit juste d'essayer de se passer de plus en plus de technologies. Se passer des technologies « Wetware », se passer petit à petit d'internet, des transistors, des voitures, de l'électricité. Ce n'est pas seulement se passer de la technologie mais en détruire toute trace : en désimpregner le cerveau de l'homme au fur et à mesure des générations. Puis peut-être, au bout d'un moment, recommencer dans le « bon » sens en faisant des cycles dans

[6] Arthur Schopenhauer, *Le fondement de la morale*, le livre de poche p. 142.

« la zone » la plus pacifiste et moins pesante pour la nature par exemple. Mais on se heurte au même risque que précédemment d'aboutir à un système totalitaire : que fait-on des gens qui refusent et cachent les preuves du passé technologique ? On les tue ? On les redresse ?

## 3.3 Les dépérissements existentiels

Il s'agit d'un monde futur dans lequel nous considérerions les hommes qui y vivent —nos enfants et les enfants de nos enfants— comme nos victimes. La responsabilité nous oblige à tout faire pour éviter cela. Et cette responsabilité ne va que dans un sens car ces hommes du futur ne pourront jamais nous le rendre. C'est une responsabilité de l'homme sur l'espèce humaine mais j'y reviendrai.

> *« Pourtant c'est avant tout l'accusation montrant ces êtres du futur comme nos victimes qui nous interdit moralement la distanciation égoïste du sentiment, généralement justifiée par l'éloignement considérable de l'objet : cela ne saurait être, nous ne pouvons l'admettre, nous n'avons pas le droit de le faire. »* [7]

Un des angles du dépérissement existentiel c'est la perte du libre arbitre. Dans *le meilleur des mondes*, contre-utopie écrite en 1932, Aldous Huxley décrit un monde parfait où la liberté a disparu, le doute a disparu mais les gens sont heureux, chacun est à sa place et se réjouit de son sort. C'est une société où toutes les passions sont bannies car elles génèrent de la souffrance. Hors la souffrance ne peut exister dans un monde parfait. Les humains sont encouragés à avoir des relations sexuelles avec des partenaires multiples mais sans jamais s'attacher. Pas d'amour, juste du plaisir. Les hommes travaillent non pas parce que cela est nécessaire mais parce que cela accroit leur bonheur de manière inconsciente : en se livrant à une activité légèrement désagréable, ils profitent bien mieux de leur temps libre. Ils sont payés en « soma », une drogue parfaite sans effet secondaire.

Cette société parfaite nous apparait comme horrible. Nous ressentons donc le besoin de l'empêcher. Mais pourquoi ? Qu'est ce qui nous fait croire que si on demandait à un grec vivant au $4^{\text{ème}}$ siècle avant Jésus Christ ou à un égyptien en -3000 av. JC, ceux-là ne jugeraient pas notre monde actuel comme la parfaite description d'une contre-utopie ?

Qui a t'il de mal à supprimer le libre arbitre si cela accroit notre bonheur ? Un monde de pur jouissance où peur, souffrance, arme, guerre a disparu ? Et perdre les passions qui nous sont si chères comme l'amour n'est ce pas un sacrifice nécessaire, un mal pour un bien ? Burrhus Skinner, psychologue et grand penseur américain, prêche en faveur de ce genre d'utopie dans

[7] Hans Jonas, *Pour une éthique du futur*, Payot-rivages p. 103.

son ouvrage *Par delà la liberté et la dignité*[8].

## 3.4 Le surhomme

Le surhomme n'a pas bonne presse. Une des raisons de ce rejet est liée à son créateur : Nietzsche. Le concept de surhomme a été associé, à tort, avec le soldat nazi du 3ème Reich, soldat de la mort (thanatos). Au contraire, le concept de surhomme est un concept essentiellement biophile. L'homme doit se surmonter dans l'accomplissement de la vie qui est volonté de puissance. André Comte-Sponville n'est pas d'accord. Pour lui,

> *« dépasser l'homme se serait le trahir ou le perdre [...] et l'être d'un homme n'est pas moins détruit s'il se change en ange que si il se change en cheval. [...] L'homme n'est pas dieu : il ne restera pleinement humain qu'à la condition d'accepter de n'être ni sa cause ni sa ruine »* . [9]

La question du surhomme est une question déjà ancienne faisant débat. André Comte-Sponville rappelle des propos de Montaigne citant Sénèque *« Ô la vile chose et abjecte que l'homme, s'il ne s'élève au dessus de l'humanité ! »* . Montaigne rajoute alors ce commentaire *« Voila un bon mot et un utile désir, mais pareillement absurde. Car de faire la poignée plus grande que le poing, la brassée plus grande que le bras, et d'espérer enjamber plus que l'étendue de nos jambes, cela est impossible et monstrueux. Ni que l'homme se monte au-dessus de soi et de l'humanité »* [10]

La création du surhomme passe par la médecine 2.0 qui augmentera nettement l'espérance de vie. Mais ce surhomme acquerra toute sa puissance lors du passage à l'étape suivante : la médecine 3.0. Cette médecine engendre en réalité la mort de la médecine (et donc sa propre mort) puisque les pathogènes disparaissent et donc la guerre également. La médecine 3.0 c'est acquérir d'une part la fonction permettant de passer du code ADN à l'information décompressée[11] correspondante (le phénotype si vous préférez) et d'autre part sa fonction réciproque. La découverte de telles fonctions, si elles existent, a de conséquences difficilement imaginables. Mais essayons quand même.

[8] En référence bien-sûr à Nietzche et son ouvrage *Par delà bien et mal*.
[9] André Comte-Sponville, *Présentation de la philosophie*, le livre de poche p. 137.
[10] André Comte-Sponville, *Présentation de la philosophie*, le livre de poche p. 142.
[11] Voir le chapitre sur l'information.

### 3.4.1 La mort de la mort

La mort avait déjà nettement reculé voire disparu avec la médecine 2.0. Mais avec la médecine 3.0, le décryptage de l'information, du réseau, de la complexité de n'importe quel système vivant est total. Le fonctionnement des réseaux intra et extra-cellulaires (neurones) est totalement assimilé : les mécanismes cérébraux qui génèrent la pensée n'ont donc plus aucun secret. L'âme de l'homme est potentiellement séparable de son corps. Ce dernier n'est plus qu'un vulgaire outil, modelable à façon qu'on peut changer à volonté : vous voulez voler ? Taper « sur un clavier » l'information décompressée du système cellulaire volant de votre choix. La fonction réciproque vous renvoie ensuite la séquence ADN. Insérez cette ADN dans un châssis cellulaire qui se divise et voila votre système volant. Insérez-y votre âme et c'est vous ! Pourquoi l'évolution n'a pas créé des systèmes volants naturelles, des énormes oiseaux de la taille d'un avion capables de traverser l'atlantique ? Parce que l'évolution n'a pas su trouver l'intérêt pour le *fitness* (la fonction à optimiser) de faire une telle traversée. Mais vous, qu'est ce qu'elle en sait l'évolution, de votre intérêt ?

### 3.4.2 La symbiose

Au cours de l'évolution, il y a eu différents types d'unification entre des systèmes vivants. L'apparition de la multi-cellularité est un exemple. C'est l'unification des cellules clonales en un organisme dont la survie devient prioritaire à celle des cellules (j'ai fait beaucoup d'anthropomorphismes dans cette thèse et je m'en suis justifié ailleurs). Par exemple, l'apoptose (la mort programmée des cellules) est un mécanisme indispensable à la vie de l'organisme. Un autre exemple d'unification nous est donné par la théorie de l'endosymbiose qui postule que la mitochondrie est le résultat de l'incorporation d'une bactérie par une cellule eucaryote primitive.

On peut imaginer que le futur de l'homme soit l'unification des organismes en un super organisme (le surhomme) de la même manière que les cellules se sont assemblées pour former un organisme. Cette unification n'a pas besoin d'avoir une réalité physique car ce qui compte c'est la présence d'un réseau. Vous allez me dire que les hommes forment déjà un super organisme « la société » grâce aux interactions sociale et au réseau internet. Mais la technologie future permettra de supprimer les intermédiaires que sont la voie ou l'ordinateur : les cerveaux des hommes seront connectés par les ondes électromagnétiques. On aura donc 3 niveaux de réseau : le niveau intracellulaire (le réseau de régulation génique), le niveau intercellulaire (le réseau neuronal) et le réseau inter-cerveau (via les ondes électromagnétique). Vous pourriez me rétorquer que la mise en réseau « directe » (sans l'intermédiaire de l'ordinateur) ne supprimera

pas forcement l'égoïsme : le fait de toujours privilégier son cerveau, son organisme à celui des autres. Mais quel intérêt cela pourrait-il avoir ? L'égoïsme n'est t'il pas lié à l'existence d'un corps et l'assouvissement de ses besoins instinctifs pulsionnels ? A un tel niveau de technologie : manger, boire, respirer, se reproduire seront des instincts qui n'auront plus de raisons d'être. L'égoïsme n'aura plus de substance sur laquelle s'exercer. L'homme surmonte ses instincts et crée au même moment le surhomme qui correspond à l'unification des hommes. Cette unification accroit les capacités « computationnelles /de réflexion » du surhomme qui pense mieux et qui pense plus loin. En présentant les choses de cette manière, je laisse entendre que l'homme est responsable de la création du surhomme. Mais la nature serait en droit d'en revendiquer la paternité : c'est elle qui, en inventant (via l'algorithme génétique) l'extraordinaire cerveau humain, a crée les conditions d'une unification inter-organisme. Tout comme les cellules ne sont pas responsables de la création d'êtres multicellulaires, les hommes ne sont pas responsables de la création d'êtres « multi-organismes ». Ce surhomme peut être entrevu comme la fourmilière ou l'essaim d'abeille. Ces derniers, bien que constitués d'agents de base (fourmis et abeilles), possèdent des propriétés émergentes qui sont plus que la simple somme des agents (dû à la présence d'un réseau d'interaction). Et ce n'est pas les fourmis qui « décident » de s'unir en fourmilière mais bien la nature qui tire les ficelles. Cependant, avec le surhomme, les choses sont un peu différentes. La nature change d'algorithme pour « évoluer ». Jusqu' à maintenant, elle utilisait principalement le couple mutation-sélection (ainsi que les échanges de matériel génétique via les crossing-over pour augmenter l'efficacité de l'algorithme d'optimisation). Le hasard couplé à la sélection expliquerait (mais il existe des théories concurrentes voir le chapitre sur l'évolution II) l'ensemble des fonctions et structures ordonnées qui constituent les systèmes vivants. Par exemple, la vue. Avec le surhomme, l'outil de la nature change. En effet, le surhomme continuera d'évoluer mais ce ne sera plus grâce aux changements aléatoires de nucléotides ni par la sélection. En créant le cerveau humain puis le surhomme, la nature s'apprête à changer fondamentalement les lois qui gouvernaient jusqu'à présent l'évolution de la vie.

### 3.4.3 La boucle est bouclée

Une fois que l'homme aura compris le fonctionnement de son cerveau, Il se pourrait qu'il ait la solution à certains problèmes plus complexes : car le surhomme pourra dépasser les limites de fonctionnement du cerveau « trop humain », trop simple. Si les hommes ont plus de neurones et de synapses que les mouches, c'est que le nombre doit quand même compter et être proportionnel, dans une certaine mesure, aux capacités computationnelles. Qu'est ce

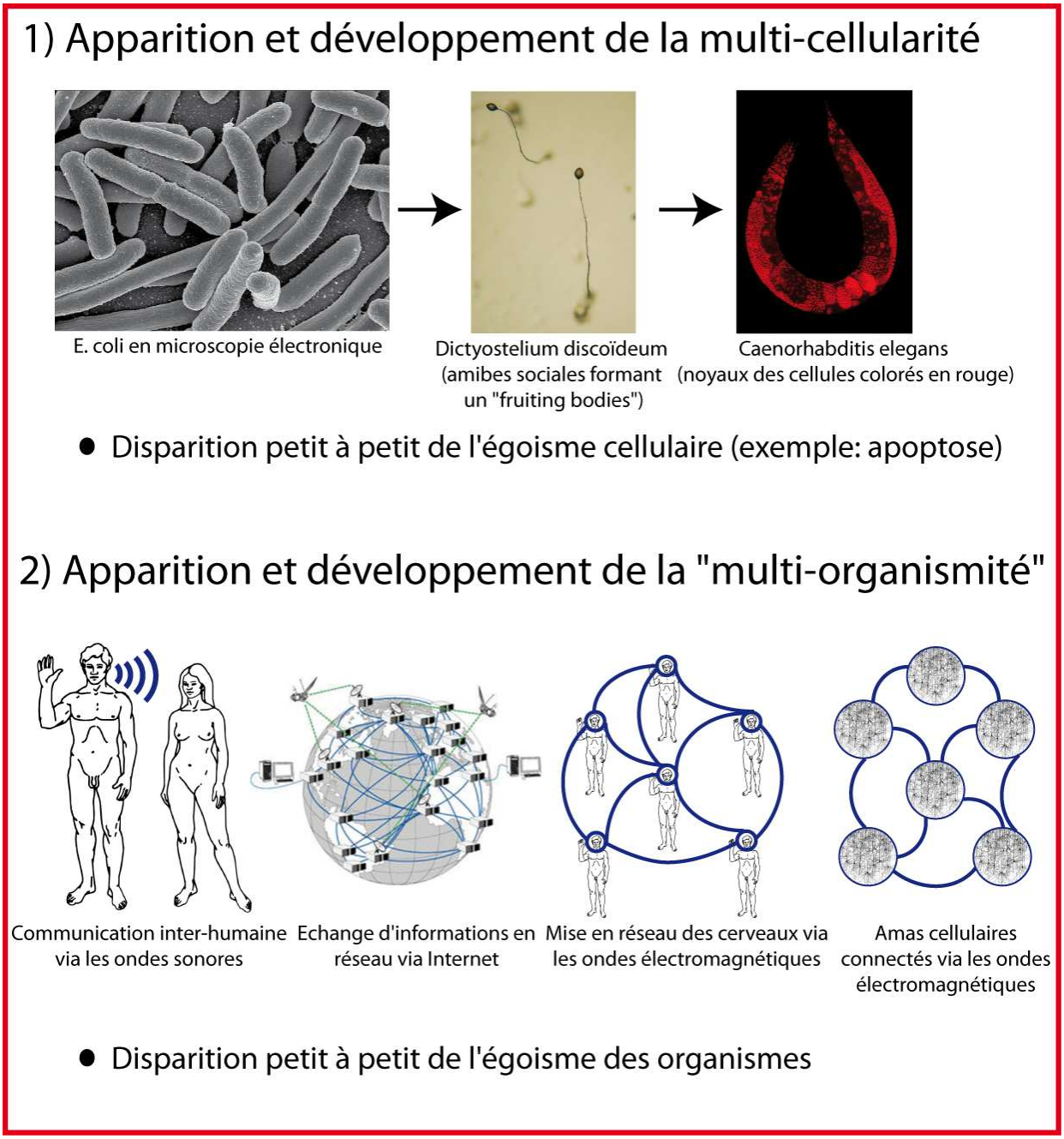



**Figure XV.1** – La société est le précurseur du surhomme. Au fil des temps évolutifs, les cellules se sont assemblées en organismes et ont perdu leur égoïsme (mort programmée des cellules pour le bien de l'organisme). Rien n'empêche de penser que la nature soit en train d'opérer une réunion similaire à l'échelle des organismes. En effet, le réseau qui unit les hommes n'a jamais été aussi dense. En supprimant le corps et ses besoins pulsionnels grâce à la biologie synthétique, l'homme pourrait se débarrasser de son égoïsme. On assisterait alors à l'avènement du surhomme : amas cellulaires connectés dotés de facultés cognitives exceptionnelles.

qui empêche d'imaginer des amas de neurones de la taille d'une planète ? Voila la nouvelle mission du surhomme pour progresser, évoluer : une étoile composée essentiellement de neurones qui « calculent/réfléchissent ». Cette étoile neuronale n'est pas obligée d'utiliser uniquement le « Wetware » comme technique computationnelle. En effet, si le calcul quantique est plus efficace pour traiter certains problèmes, l'étoile computationnelle ne s'en privera pas car elle utilise toutes technologies potentiellement utiles créées par ses capacités cognitives. Cette étoile computationnelle, ce surhomme pourrait bien réussir à « penser/calculer/comprendre » ce qui nous était inaccessible à cause de nos limites cognitives. Par exemple, se représenter un espace à n-dimension (le fameux hypercube traité dans le chapitre Evolution II). Et qu'en est-il des 7 problèmes réputés insurmontables des mathématiques fondamentales, les fameux problèmes du prix du millénaire de l'institut Clay ? Comme par exemple savoir si P=NP ? Les futures percées des mathématiques fondamentales passent peut être par le dépassement de nos limites cognitives intrinsèques —limites qui empêchent la découverte et la compréhension de certains phénomènes—. Cet argument pourrait convaincre certains mathématiciens pures de s'intéresser de plus prés à la biologie n'ont plus comme une simple application mais comme l'outil clé des futures percées, ruptures, révolutions en mathématique fondamentale. Je le répète, certains outils mathématiques du futur ne sont pas « constructibles » par nos cerveaux actuels. Changer, améliorer, surmonter le cerveau humain permettra de découvrir ces nouveaux objets. L'étoile computationnelle peut sans doute aller encore plus loin et nous éclairer sur les vieux problèmes philosophiques. Qu'est ce qui empêche de penser que cette étoile puisse enfin savoir s'il existe un monde platonicien des vérités mathématiques ? Et la réalité existentielle est-elle vraiment essentielle ?

### 3.4.4 Le pacte

Augmentons encore nos facultés cognitives : élargissons-nous, surhomme, à l'ensemble de l'univers. Utilisons chaque particule de l'univers pour calculer, pour comprendre (si ces mots ont encore un sens pour le surhomme). Chaque particule doit devenir information. C'est cela la mission de l'homme devenu surhomme. Informatiser chaque atome de l'univers dans le but de répondre à la question *pourquoi*. Et s'il réussit, n'est-il alors pas devenu dieu ? Ne pouvons-nous pas entrevoir le mot de Voltaire différemment de son sens initial ? « Dieu a crée l'homme à son image et ce dernier le lui a bien rendu ». Poussons maintenant l'expérience de pensée plus loin. J'ai défendu l'idée tout à l'heure qu'il fallait éviter à tout prix une scission *subie* dans la distribution des génomes des hommes car cela pourrait générer différentes espèces et aboutir à une guerre inter-espèce. Cependant qu'en est-il d'une scission souhaitée et pacifique : des

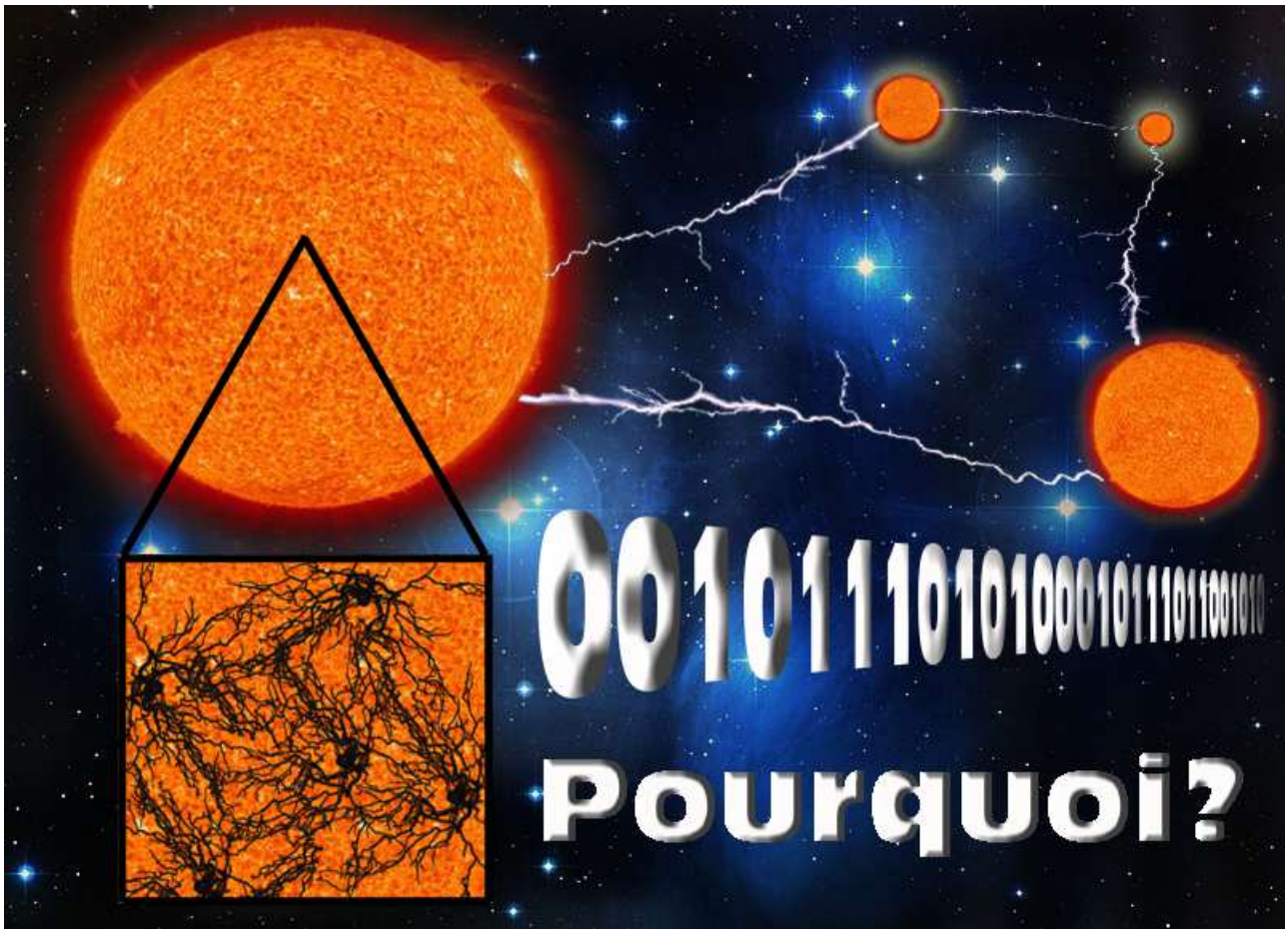



**Figure XV.2** – Le surhomme. Voici ma représentation (naïve) du surhomme qui informatise les atomes de l'univers dans le but de répondre à la question *pourquoi*. Ce surhomme est composé d'étoiles « computationnelles » qui communiquent entre elles par la lumière. j'ai représenté les capacités cognitives de l'étoile via un zoom montrant un réseau de neurone. Le calcul sera-t-il encore basé sur des bits classiques ? La lumière sera-t-elle le moyen le plus rapide de transmettre des informations aux autres étoiles ? la question « pourquoi » aura-t-elle encore un sens pour le surhomme ?

hommes souhaitant continuer à mener une existence paisible « sans savoir pourquoi » et sans vouloir « savoir pourquoi » et d'autres hommes voulant suivre un autre chemin, souhaitant se lancer dans l'aventure du surhomme pour « savoir pourquoi ». Ces hommes pourraient-ils sceller un pacte ? Une fois devenu dieu, le surhomme s'engagerait à ne pas se manifester et à laisser les hommes vivre leurs existences. Qu'est-ce-qui nous prouve d'ailleurs qu'un tel contrat, signé dans le passé, n'est pas déjà en vigueur ? Car si on peut imaginer devenir le surhomme du futur, rien n'empêche d'imaginer que le surhomme du présent existe déjà et ne se manifeste pas.

Cette question de la scission me semble intéressante car je pense qu'en l'état actuel de la société, elle serait impossible. Aucun homme n'est prêt à en laisser devenir un autre surhomme.

Je pense qu'on vit dans une société où la majorité des hommes sont hostiles à l'idée du surhomme. Ils n'accepteraient jamais la scission. Soit il y a surhomme pour tout le monde soit il n'y a pas de surhomme. C'est que le niveau de confiance nécessaire pour accepter une scission est gigantesque. Mais vouloir *à tout prix* empêcher l'avènement du surhomme, c'est prendre le risque de sombrer vers le totalitarisme, c'est-à-dire le dépérissement existentiel. Le risque pris en acceptant une scission pacifique pourrait bien être préférable.

### 3.4.5 Le but

Cette expérience de pensée sur la scission me semble également intéressante car elle nous révèle des éléments sur notre société et les agents qui la composent. En imaginant une scission, on imagine deux types de mentalités, deux manières de voir le monde de part et d'autre de la scission. Ceux qui pensent que la vie se définit par son absence de finalité : la vie n'a pas d'autres buts que d'être vécue. Ils s'entrevoient eux-mêmes comme une petite poussière dans un tout. Cette poussière exerce une fonction dont il n'est pas nécessaire de comprendre le sens. Une fleur naît, s'épanouit, se fane puis meurt. Il n'y a rien à comprendre, c'est comme cela. Le deuxième type de mentalité pense qu'il existe un but, une finalité, une téléologie. Ils ne pensent pas en termes de fonction mais en termes de mission. Il faut avancer pour comprendre le pourquoi. Concevoir la vie sans la présence d'une finalité est, selon eux, une manière bien fade de voir la vie. La présence d'une dérivée positive, d'une construction qui grandit, brique par brique, rassure. Le but peut même finir par être son propre objet c'est à dire garder un but à tout prix. Car que ferait le surhomme après avoir répondu à la question *pourquoi* ? Se pourrait-il que la question *pourquoi* n'est aucune réponse ? Se pourrait il que la question *pourquoi* ne soit qu'une pure construction cognitive dont l'unique but est de garantir le maintien éternel d'un but ? Se pourrait-il que ces deux types de mentalité forment les futurs blocs/partis/camps/adversaires se livrant une guerre pour imposer leur manière de voir le monde ? Nous serions en présence d'un beau paradoxe car cela signifierait que les partisans de l'absence de but auraient pour but de faire plier les partisans du but.

Creusons plus en profondeur cette question du but. J'identifie deux buts essentiels :

⇒ un but qui cherche à supprimer la notion de but (il cherche à se supprimer) :

- soit en ne cherchant plus à répondre à la question *pourquoi* c'est-à-dire en cherchant à annihiler l'existence même de cette question.

> *« Pour Platon, philosopher c'est chercher la vérité avec toute son âme. Face à cet entêtement qu'il considère comme une pathologie, Nietzsche répond par ce mot provocateur : en admettant que nous voulions le vrai, pourquoi pas plutôt*

*le non vrai ? Ou l'incertitude ? »*[12]

– soit en ayant pour but de répondre à la question *pourquoi* (en cherchant le vrai). Une fois répondu à la question, difficile d'imaginer autre chose que le néant c'est-à-dire l'absence de but. Autrement dit, une dérivée plate dans le progrès (suite à la découverte du « tout ») ressemble fortement à un électrocardiogramme plat c'est-à-dire la mort/le néant/la fin du jeu.

*« Qui nous dévoilerait l'essence du monde nous infligerait à tous la plus pénible désillusion »*[13]

⇒ Un but qui cherche à maintenir la notion de but (il cherche à se maintenir) : en faisant en sorte de conserver éternellement la question *pourquoi*. Il s'agit du cas où le but est son propre objet. Une citation allemande populaire illustre bien cette idée « Der weg ist das ziel »(le but c'est le chemin).

On voit donc que la notion *de but* semble, dans tous les cas, être le passage obligé pour supprimer ou maintenir *un but*. Ainsi le but peut être soit temporaire soit définitif. Dans les deux cas, ce but, que le surhomme pourrait se fixer, est dépendant d'un autre concept Nietzschéen : celui de la grande santé. La grande santé implique la capacité à conquérir une santé que l'on sacrifie sans cesse (nouvelle souffrance dont il faut guérir). Pour pouvoir se diriger vers un but, il faut une volonté et pour avoir cette volonté, il faut pouvoir conquérir une santé que l'on perd sans cesse. C'est-à-dire l'acte de chercher (un palliatif à la souffrance) pour trouver (une santé vigoureuse) et sa réciproque. Sans grande santé, pas de but.

*« Nous qui sommes nés trop tôt pour un avenir dont la démonstration n'est pas encore faite, nous avons besoin, pour une fin nouvelle, d'un moyen nouveau, je veux dire d'une nouvelle santé, d'une santé plus vigoureuse, plus aigue, plus endurante, plus intrépide et plus joyeuse que ne furent jusqu'à présent toutes les santés. Celui dont l'âme est avide de faire le tour de toutes les valeurs qui ont eu cours et de tous les désirs qui ont été satisfaits jusqu'à présent, [...] Celui là aura avant tout besoin d'une chose, de la grande santé —d'une santé que l'on possède non seulement, mais qu'il faut conquérir sans cesse, puisque sans cesse il faut la sacrifier ! »*[14]

Ainsi les partisans de *l'annihilation de tout but* doivent tout faire pour éradiquer la volonté et la grande santé. Or pour cela ils ont justement besoin de la volonté et de la grande santé. Et je ne suis pas sûr que ce paradoxe puisse être facilement surmonté.

[12]Antoine Le Bos, *Nietzsche ou le scepticisme viril* (Postface de Ecce Homo), Mille est une nuit p. 170
[13]Nietzsche,*Humain trop humain (I)*,Folio Essais p.54
[14]Nietzsche, *Ecce Homo*, Mille est une nuit p. 118

# 4 Pour une futurologie

En matière de futur, de science futur, le peuple doit s'impliquer. Et le peuple doit être impliqué. En particulier dans le domaine de la biologie synthétique. Il nous faut donc débattre ensemble des le début. L'intégration du peuple ne doit pas être factice, il ne s'agit pas de cocher une case, de faire semblant de débattre. Le peuple doit réellement pouvoir « choisir » la société dans laquelle il souhaite vivre et évoluer. Si les chercheurs décident seuls, cela revient à confier le futur aux mains d'un système aristocratique (aristos en grec : les meilleurs) et non démocratique. Par exemple, sur une question cruciale comme la mort de la mort, il semble que la science puisse difficilement nous offrir un cadre de réflexion morale pour décider si oui ou non, nous voulons tuer la mort. Comme le dit Jacqueline Russ,

> *« si à partir de 1960, les progrès en matière de médecine et de science de la vie ont rendu nécessaire la création de comité d'éthique c'est que nous nous sommes trouvés en quelque sorte dépassés par les possibilités techniques que nous avons crée et que nous avons eu besoin de penser nos actes afin de les orienter vers une finalité qui ne soit pas seulement celle de l'exercice de la puissance mais aussi celle de la réalisation de l'humain. [. . .] le questionnement éthique vient donc se loger dans cet écart entre le possible et le désirable que le progrès de la science et le cours tragique de l'histoire ont rendu nécessaire »* [15]

L'éthique a besoin de futurologie qui n'est pas une science positiviste mais une projection au loin selon la méthode scientifique. Cette futurologie peut nous aider à éviter la catastrophe illustrée par la métaphore des grains de sable qui tombent les uns après les autres d'un sablier et s'empilent en bas en formant une montagne . Puis c'est le grain de sabre de trop qui provoque l'effondrement de la montagne. La futurologie a pour but de détecter et bloquer le grain de sable de trop avant qu'il ne tombe.

En science, il est difficile de prévoir à l'avance quelles découvertes vont être réalisées et comment ces dernières vont changer la société. Je ne sais pas, par exemple, combien de penseurs ont réussi à prédire l'explosion des ordinateurs puis l'explosion d'internet et les conséquences sur nos vies. En biologie, il se pourrait que les choses soient un peu différentes. On sait déjà plus ou moins ce qu'on sera capable de faire dans quelques décennies/siècles : maitriser le fonctionnement d'un réseau biologique à l'échelle moléculaire. Cela nous semble être une question de temps car si la nature y arrive, nous ne voyons pas de raison de ne pas y arriver. Cependant, la complexité est telle que la réussite de cette mission, par exemple découvrir la fonction qui transforme la séquence ADN en information décompressée, pourrait nous prendre un temps

[15] Jacqueline Russ, *La pensée éthique contemporaine*, PUF p. 4.

gigantesque. Ce temps doit être utilisé pour réfléchir aux conséquences des découvertes et pour orienter les recherches en direction de, non pas ce qui est faisable, mais de ce qui est souhaitable, voulu par le peuple.

Pour résumer, nous[16], chercheurs, en l'état actuel des connaissances, n'identifions pas d'obstacles scientifiques majeurs (ça ne veut pas dire qu'ils n'existent pas) qui empêcheraient l'homme dans les siècles à venir, de devenir immortel. L'humanité dispose donc de quelques décennies, siècles pour réfléchir aux conséquences et savoir si elle souhaite cette immortalité ou non. Et si oui, sous quelle forme. Car on peut imaginer être réellement « immortel » pendant 100 ans face aux maladies et aux accidents c'est-à-dire avoir l'assurance de vivre 100 ans, puis décider de rejoindre, ou non, le royaume des cieux après cette date. Mais on peut aussi imaginer des régimes totalitaires interdisant la mort ce qui nous condamnerait à vivre pour l'éternité ce que j'imagine, peu d'hommes sur terre souhaitent.

Avec la futurologie, les découvertes deviendraient, dans une certaine mesure, programmées, choisies, identifiées en amont et validées par l'éthique. Selon Hans Jonas,

> *« Toute futurologie sérieuse telle que l'exige l'objectif de la responsabilité devient une branche de la recherche qu'il convient de cultiver en soi et sans relâche en suscitant la coopération de nombreux experts dans les domaines les plus divers. [...] La vision de l'avenir au service de l'éthique du futur revêt donc une fonction intellectuelle et une fonction émotionnelle ; elle doit instruire la raison et animer la volonté. »* [17]

# 5 La théorie des pistes

Un point qui me semble crucial, c'est de laisser la chance aux générations futures de penser différemment de nous. Ce n'est pas parce que *le meilleur des mondes* de Aldous Huxley nous semble insupportable qu'il le serait pour nos enfants. Et nous n'avons aucune légitimité à leurs imposer notre vision de « l'essence de l'homme » car Sartre nous a expliqué depuis relativement longtemps maintenant que :

> *« l'existence précède l'essence [...]. L'homme existe d'abord, c'est-à-dire que l'homme est d'abord ce qui se jette vers un avenir et ce qui est conscient de se projeter dans l'avenir [...] l'homme sera d'abord ce qu'il aura projeté d'être.[...] La première démarche de l'existentialisme est de mettre tout homme en possession de ce qu'il est et de faire reposer sur lui la responsabilité totale de son existence. Quand*

[16] Je pense parler au nom des chercheurs. Si cela vous choque, remplacez le « nous » par un « je ».
[17] Hans Jonas, *Pour une éthique du futur*, Payot-rivages p. 87.

> *nous disons que l'homme est responsable de lui-même, nous ne voulons pas dire que l'homme est responsable de sa stricte individualité mais qu'il est responsable de tous les hommes ».*[18]

Ainsi tout ce que nous pouvons faire pour nos enfants, c'est leurs donner nos meilleurs « mèmes[19] », nos meilleurs idées, nos meilleurs pistes en ce qui concerne la *vie meilleure* et *l'essence de l'homme*. Ces pistes n'ont pas de valeurs universelles. En voila la preuve : je peux vouloir transmettre à mes enfants le concept d'égalité que je trouve sublime mais celui-ci n'a qu'une valeur temporelle limitée car s'il y a avènement du surhomme tel que je l'ai décrit (symbiose des hommes pour former le surhomme à la manière des cellules qui forment l'organisme), le concept d'égalité disparait aussitôt (il cesse d'être orientant) puisque le surhomme est seul. La puissance cognitive du surhomme rend toutes nos constructions cognitives potentiellement désuètes.

Les pistes sont des constructions cognitives que nous souhaitons voir propager. A nos enfants de décider s'ils veulent à leur tour les propager telles quelles, les modifier, ou tout simplement les mettre au placard. Je ne suis donc pas d'accord avec cette phrase de Sartre « *l'existentialisme pense que l'homme, sans aucun appui et sans aucun secours, est condamné à chaque instant à inventer l'homme.* » [20] car la continuité qui existe entre les générations permet l'existence des pistes. Et celles-ci représentent justement l'appui et le secours dont l'homme a besoin pour affronter son existence et se réinventer à chaque instant. Même si elles ne sont pas universelles, les pistes donnent un sens et une direction à nos enfants pour l'avenir. Ils peuvent les suivre au début puis tracer plus tard leurs propres chemins. Avant de décrire certaines pistes en détail, je voudrais vous montrer qu'il existe des pistes que nos maîtres/nos parents/nos ancêtres nous ont transmis et que nous jugeons préférable de ne pas suivre.

## 5.1 Les mauvaises pistes

### 5.1.1 La dignité humaine

La dignité humaine est une piste *a priori* séduisante. Le physicien et humaniste Paul Langevin n'a-t-il pas écrit *« tout effort de l'intelligence serait vain s'il n'avait pour but ultime la dignité humaine »* ? En réalité, le concept de dignité humaine est plus creux et dangereux qu'il n'y parait. Et Schopenhauer l'a bien vu

[18] Jean-Paul Sartre, *L'existentialisme est un humanisme*, folio essais p. 29–30–31.

[19] Terme de Richard Dawkins apparu dans *Le gène égoïste* qui signifie : une idée, un élément culturel qui se transmet, se propage.

[20] Jean-Paul Sartre, *L'existentialisme est un humanisme*, folio essais p. 40.

> *« Ils ont dissimulé à l'aide de ce mot imposant : « la dignité de l'homme » leur impuissance à fournir un fondement réel, ou du moins plausible, pour la morale, comptant sagement que leur lecteur, se voyant attribuer à lui aussi cette dignité, se contenterait à ce prix »* [21]

En effet, ce concept peut menacer les libertés publiques et individuelles. Un exemple de Ruwen Ogien[22] : au nom de la « dignité humaine », certains groupes vont souhaiter « criminaliser » certaines sexualités minoritaires (prostitution, sadomasochisme, homosexualité, transsexualité...)[23] alors que d'autres groupes vont souhaiter « libéraliser » ces mêmes pratiques sexuelles. On voit donc bien que ce concept de dignité humaine n'est pas suffisamment précis. Certaines personnes pourraient l'utiliser pour s'opposer au surhomme or l'argument serait trop peu rigoureux pour être acceptable.

### 5.1.2 L'eugénisme ou le racisme

Jacques Monod, prix Nobel de Médecine, dans son ouvrage *Le hasard et la nécessité*, se livre à une analyse qui nous semble, avec le recul, aberrante. Ne le jugez cependant pas trop vite car le même livre où est publiée cette analyse, est imprégné d'humanisme ce qui excuse largement l'homme pour sa piste erronée.

> *« Comme chacun sait, les statistiques révèlent une corrélation négative entre le quotient d'intelligence et le nombre moyen d'enfants des couples. [...] Situation dangereuse, qui risque de drainer peu à peu vers une élite qui tendrait en valeur relative à se restreindre, le potentiel génétique le plus élevé. [...] Il y a plus : à une époque récente, même dans les sociétés relativement « avancées », l'élimination des moins aptes, physiquement et aussi intellectuellement, était automatique et cruelle. La plupart n'atteignaient pas l'âge de la puberté. Aujourd'hui beaucoup de ces infirmes génétiques survivent assez longtemps pour se reproduire. Grâce aux progrès de la connaissance et de l'éthique sociale, le mécanisme qui défendait l'espèce contre la dégradation, inévitable lorsque la sélection naturelle est abolie, ne fonctionne plus guère que pour les tares les plus graves »* [24]

Même chose avec James Watson, prix Nobel de Médecine, qui a découvert la structure en double hélice de l'ADN. Ce dernier a déclaré, le 14 octobre 2007, dans les colonnes du journal

[21] Arthur Schopenhauer, *Les fondements de la morale*, le livre de poche p. 103.
[22] Ruwen Ogien, *L'éthique aujourd'hui*, Folio essais p. 170.
[23] L'association de ces pratiques sexuelles est basée ici sur l'unique critère de minorité. En dehors de celui-ci, elles n'ont rien à voir les unes avec les autres.
[24] Jacques Monod, *Le hasard et la nécessité*, Points p. 206.

britannique le Sunday Times, qu'il était *« fondamentalement pessimiste quant à l'avenir de l'Afrique [parce que] toutes nos politiques d'aide sont fondées sur le fait que leur intelligence [celles des Africains] est la même que la nôtre [Occidentaux] alors que tous les tests disent que ce n'est pas vraiment le cas »* [25]. En guise de contre-attaque, le 9 décembre 2007, le Sunday Times publiait une étude scandinave affirmant que l'ADN de Watson (l'un des premiers génomes humains à avoir été séquencé) comportait 16% de gènes légués par des ancêtres noirs. Je souhaite tout de même indiquer qu'en 2008, sur un documentaire de la chaine BBC, Watson a déclaré *« je n'ai jamais pensé être raciste, je ne me vois pas comme une raciste. J'ai été mortifié par [cette polémique]. Cela a été la pire chose de ma vie »* .

Le but ici n'est pas de propager des polémiques ou de s'acharner sur des éminents scientifiques. Le but est de montrer que tout le monde —y compris les meilleurs— peut être dans l'erreur, propager la mauvaise piste. Les rôles des enfants étant simplement de fermer la dite piste et d'en chercher d'autres. On voit aussi à travers ses exemples les dangers du scientisme prônant l'effacement de la politique au profit de la gestion scientifique des problèmes sociaux. Non, le futur ne doit pas être confié aux seuls aristocrates (les chercheurs) quand bien même ils seraient prix Nobel de médecine.

Je profite également de ce paragraphe pour inclure mon texte dans le risque de « la piste erronée ». Je ne suis pas à l'abri de fautes d'appréciation, de discours nauséabonds involontaires qui pourraient potentiellement heurter ou blesser un lecteur. Si tel était le cas et si j'en avais connaissance, je serais bien sûr prêt à reconsidérer mon écrit et à m'excuser le cas échéant.

## 5.2 Les bonnes pistes

Voyons maintenant ensemble quelques pistes que nous voudrions transmettre à nos enfants pour leurs permettre de construire un futur serein. Quelles pistes faut-il transmettre aux générations futures concernant *l'essence de l'homme, la vie meilleure* ? Qu'est-ce-que l'homme ne doit jamais perdre pour éviter de se vider de sa substance ? Je précise que ma liste n'est évidemment pas exhaustive et je serais très heureux de connaître, cher lecteur, vos suggestions en la matière.

### 5.2.1 L'amour

La vie ne mérite pas d'être vécue sans amour. Certaines dépressions majeures, caractérisées par la perte de la capacité d'aimer, sont les pires maladies qui existent et conduisent souvent celui qui en victime au suicide.

[25] Source Wikipedia : article sur James Watson.

*« C'est l'amour qui fait vivre, puisque c'est lui qui rend la vie aimable. C'est l'amour qui sauve ; c'est donc lui qu'il s'agit de sauver »* [26]

### 5.2.2 La musique (l'art) et le rire

*« La vie sans musique est simplement une fatigue, une erreur. »* [27].

J'imagine un surhomme qui aurait colonisé l'univers : ce surhomme, seul entité pensante et donc informative de l'univers, augmenterait sa puissance dans le but de répondre à la question pourquoi. Ce surhomme, machine computationnelle générant de la puissance, ne devrait-il aussi pouvoir mettre et écouter de la musique dans tout l'univers ? Ce surhomme ne devrait-il pas aussi pouvoir rire à gorge déployé ?

### 5.2.3 La morale

Si le surhomme de demain (tout seul puisque issu de la réunion des hommes) pensera par delà bien et mal. A l'inverse, l'homme d'aujourd'hui peut conserver la morale dans les pistes qu'il transmet. En jonglant avec le présent et le futur, la théorie des pistes nous permet d'être en accord avec Nietzsche pour ce qui est du futur tout en suivant les pas de Kant pour ce qui est du présent. Pour ma part, je transmettrai donc à mes enfants les notions de bien et de mal. Vertu et universalité de l'action désintéressée, souci de l'intérêt général sont des idées à propager. N'en déplaise à Schopenhauer, l'impératif catégorique est une piste qui oriente :

*« Agis uniquement d'après la maxime qui fait que tu peux aussi vouloir que cette maxime devienne une loi universelle »* [28].

Sans oublier la maxime sublime de Rousseau :

*« fais à autrui comme tu veux qu'on te fasse »* [29]

Penser comme Nietzsche « par delà bien et mal » au présent ne correspond pas à mes convictions :

*« Qui atteindra quelque chose de grand s'il ne sent pas en lui la force et la volonté d'infliger de grandes douleurs ? Savoir souffrir est la moindre des choses : de faibles femmes et même les esclaves y excellent souvent. Mais ne pas périr de détresse et*

[26] André Comte-Sponville, *Présentation de la philosophie*, le livre de poche p. 43.
[27] Stephan Zweig, *Nietzsche*, Stock p. 123
[28] Emmanuel Kant, *Fondation de la métaphysique des mœurs*
[29] Jean-Jacques Rousseau, *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*, Flammarion p. 98.

*d'incertitudes intérieures lorsqu'on inflige une grande souffrance et que l'on entend le cri de la souffrance -voila ce qui est grand, voila ce qui appartient à la grandeur »*[30]

En se plaçant par delà bien et mal grâce à la biologie, Nietzsche attaque violemment la morale Kantienne/chrétienne. Or cette opposition entre morale et biologie n'est pas si évidente que cela dans mon esprit. Un de mes projets futurs consistera à s'engouffrer dans ce qui m'apparait être une « une faille ». A l'origine de cette faille se trouve la volonté de Nietzsche de penser en termes « d'organismes ». Le choix de cette unité de base (l'organisme) me semble être une hypothèse forte au fondement du système philosophique Nietzschéen. Mais cette hypothèse est-elle juste et légitime ?

### 5.2.4 la liberté

Le surhomme computationnel, régnant seul sur l'univers, peut-il être libre ? Et plus généralement, peut-on être libre lorsque l'on est tout seul ? Car ce concept de liberté est plus complexe qu'il n'y parait. Pour Nietzsche, la liberté n'existe que pour le guerrier (réel ou abstrait) combattant une dictature (réelle ou abstraite) :

*« Ma conception de la liberté. La valeur d'une cause se mesure parfois non à ce qu'on atteint par elle, mais à ce qu'il faut la payer, à ce qu'elle nous coûte [...] L'homme libre est un guerrier. A quoi mesure t'on la liberté chez les individus comme chez les peuples ? A la résistance qu'il faut surmonter, à la peine qu'il en coûte pour garder le dessus. Le type supérieur d'homme libre, il faudrait la chercher là ou il s'agit constamment de vaincre la résistance la plus forte : à quelques pas de la tyrannie, tout prêt de seuil qui marque l'asservissement. »* [31]

### 5.2.5 La perfectibilité

Peut-être que les caractéristiques qui ne devraient surtout pas disparaitre chez l'homme sont son libre arbitre apparent et sa faculté de se perfectionner. C'est ce que nous propose Rousseau dans ce texte qui n'a pas pris une ride.

*« Je ne vois dans tout animal qu'une machine ingénieuse, à qui la nature a donné des sens pour se remonter elle-même, et pour se garantir, jusqu'à un certain point, de tout ce qui tend à la détruire, ou à la déranger. J'aperçois précisément les mêmes choses dans la machine humaine, avec cette différence que la nature seule fait tout*

[30] Nietzsche, *Le gai savoir*, Flammarion p. 261.
[31] Nietzsche, *Crépuscule des idoles*, Folio p. 83

*dans les opérations de la bête, au lieu que l'homme concourt aux siennes, en qualité d'agent libre. L'un choisit ou rejette par instinct, et l'autre par un acte de liberté ; ce qui fait que la bête ne peut s'écarter de la règle qui lui est prescrite, même quand il lui serait avantageux de le faire, et que l'homme s'en écarte souvent à son préjudice. C'est ainsi qu'un pigeon mourrait de faim près d'un bassin rempli des meilleures viandes, et un chat sur des tas de fruits, ou de grain, quoique l'un et l'autre pût très bien se nourrir de l'aliment qu'il dédaigne, s'il s'était avisé d'en essayer ; c'est ainsi que les hommes dissolus se livrent à des excès, qui leur causent la fièvre et la mort ; parce que l'esprit déprave les sens, et que la volonté parle encore quand la nature se tait. [...] C'est la faculté de se perfectionner ; faculté qui à l'aide des circonstances, développe successivement toutes les autres et réside parmi nous tant dans l'espèce que dans l'individu. Au lieu qu'un animal est au bout de quelques mois ce qu'il sera toute sa vie, et son espèces au bout de mille ans ce qu'elle était la première année de ces milles ans »* [32]

### 5.2.6 La responsabilité

Hans Jonas est un philosophe allemand qui a inspiré le principe de précaution. Son principal ouvrage, *Le principe responsabilité* place la responsabilité comme principe fondamental à ne jamais perdre au cours de la progression future de l'humanité. C'est donc une piste.

*« L'homme est le seul être connu de nous qui puisse avoir une responsabilité. En pouvant l'avoir, il l'a. Être capable de responsabilité signifie déjà être placé sous le commandement de celle-ci : le pouvoir même entraine avec lui le devoir »* [33]

*« L'homme est le seul être connu de nous qui puisse avoir une responsabilité. Immédiatement, nous reconnaissons dans ce pouvoir davantage qu'une simple donnée empirique. Nous y reconnaissons un critère distinctif et décisif de l'essence humaine dans sa dotation en être. [...] Mais, dés lors, la capacité de responsabilité devient elle-même son propre objet étant donné que sa détention oblige à perpétuer sa présence dans le monde. [...] En soi, la capacité de responsabilité oblige donc chaque fois ses détenteurs à rendre possible l'existence d'autres détenteurs futurs. Pour que la responsabilité ne disparaisse pas du monde —tel est son commandement immanent— il faut qu'il y ait aussi des humains à l'avenir »* [34]

[32] Jean-Jacques Rousseau, *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*, Flammarion p. 78–79.

[33] Hans Jonas, *Pour une éthique du futur*, Payot-rivage p. 76.

[34] Hans Jonas, *Pour une éthique du futur*, Payot-rivage p. 92–93–94.



**Figure XV.3** – Le rouge-gorge est une des merveilles de la nature. Tout homme qui prend le temps de regarder vivre quelques minutes un rouge-gorge trouvera que la nature est belle.

### 5.2.7 Le jugement esthétique

L'esthétique est une discipline philosophique ayant pour objet « le beau ». Pour Schopenhauer, « *la beauté est une promesse de bonheur* ». Pour Kant, *« Est beau ce qui plaît universellement sans concept »* et le beau ne se confond pas avec l'agréable. Ce dernier relève d'une perception strictement personnelle et privé :

> *« Aussi bien disant : le vin des Canaries est agréable, [une personne] admettra volontiers qu' [une autre personne] corrige l'expression et lui rappelle qu'il doit dire : « cela m'est agréable ». Ainsi pour l'agréable, le principe retenu est « chacun ses goûts ». Il en va autrement du beau. Kant explique que lorsque il (une personne*[35]*) dit qu'une chose est belle, il attribue aux autres la même satisfaction ; il ne juge pas seulement pour lui mais pour autrui et parle alors de la beauté comme si elle était une propriété des choses. C'est pourquoi il dit : la chose est belle et dans son jugement exprimant sa satisfaction, il exige l'adhésion des autres [...] [et] il les blâme s'ils jugent autrement et leur dénie un gout qu'il devrait cependant posséder d'après ses exigences ».*[36]

[35] C'est moi qui rajoute.
[36] Kant, *Critique de la faculté de juger*, bibliothèque des textes philosophiques, p.74–75.

Si vous avez compris la différence entre le beau et l'agréable selon Kant, livrons nous maintenant à une petite experience d'esthétique. Que pensez-vous de ce petit rouge-gorge (figure XV.3) ? est-il beau au sens de Kant ? c'est-à-dire me blamerez-vous si je vous dit qu'il est laid ?

Et maintenant, celui-là (figure XV.4) ?

Mon but est de montrer que toute technologie permettant d'augmenter notre contrôle, à l'echelle moleculaire, d'un système vivant (par exemple la fonction permettant de passer de la séquence ADN à l'information decompréssée) risque de modifier notre perception du beau. Le premier rouge-gorge est beau parce qu'il mysterieux, il symbolise la nature qui nous domine, qui nous fait exister, des forces qui nous dépassent. Le deuxième rouge-gorge est trop associé à la notion de contrôle. Il devient un prolongement de notre bras, un outil, un artefact, un robot crée par l'homme ce qui lui fait perdre sa beauté. Ainsi l'excés de comprehension et de contrôle d'un système vivant risque de lui fait perdre toute sa beauté. Je laisse mon lecteur tirer la conclusion par lui-même sur une technologie basée sur la biologie synthétique pouvant, soi-disant, rendre les femmes (ou les hommes) plus belles (plus beaux).

### 5.2.8 La volonté, la volonté de puissance et la création

Pour Nietzsche, l'homme doit avant tout être un créateur et pour créer il doit vouloir. Voila, selon lui, la piste que ne devrait pas perdre de vue les générations futures. Voici ci-dessous quelques aphorismes extraits de « Ainsi parlait Zarathoustra » [37].

> *« Sauriez-vous créer un dieu ? Ne me parlez donc plus de tous les Dieux ! Cependant vous pourriez créer le surhomme. Ce ne sera peut être pas vous-mêmes mes frères ! Mais vous pourriez vous transformer en pères et en ancêtres du surhomme : que ce soit votre meilleur création ! »*
>
> *« Et ce que vous appeliez monde doit être d'abord créé par vous : votre raison, votre imagination, votre volonté, votre amour doivent devenir votre monde même ! Et, vraiment, ce sera pour votre félicité, vous qui cherchez la connaissance ! »*
>
> *« Créer -c'est la grande délivrance de la douleur et l'allégement de la vie. Mais afin que naisse le créateur, il faut beaucoup de douleurs et de métamorphoses »*
>
> *« Tous mes sentiments souffrent en moi et sont prisonniers : mais mon vouloir arrive toujours libérateur et messager de joie ».*
>
> *« Vouloir affranchit : c'est la vraie doctrine de la volonté et de la liberté »*

[37] Nietzsche, *Ainsi parlait Zarathoustra /Deuxième partie/Sur les îles bienheureuses et première partie/mille et un buts.*

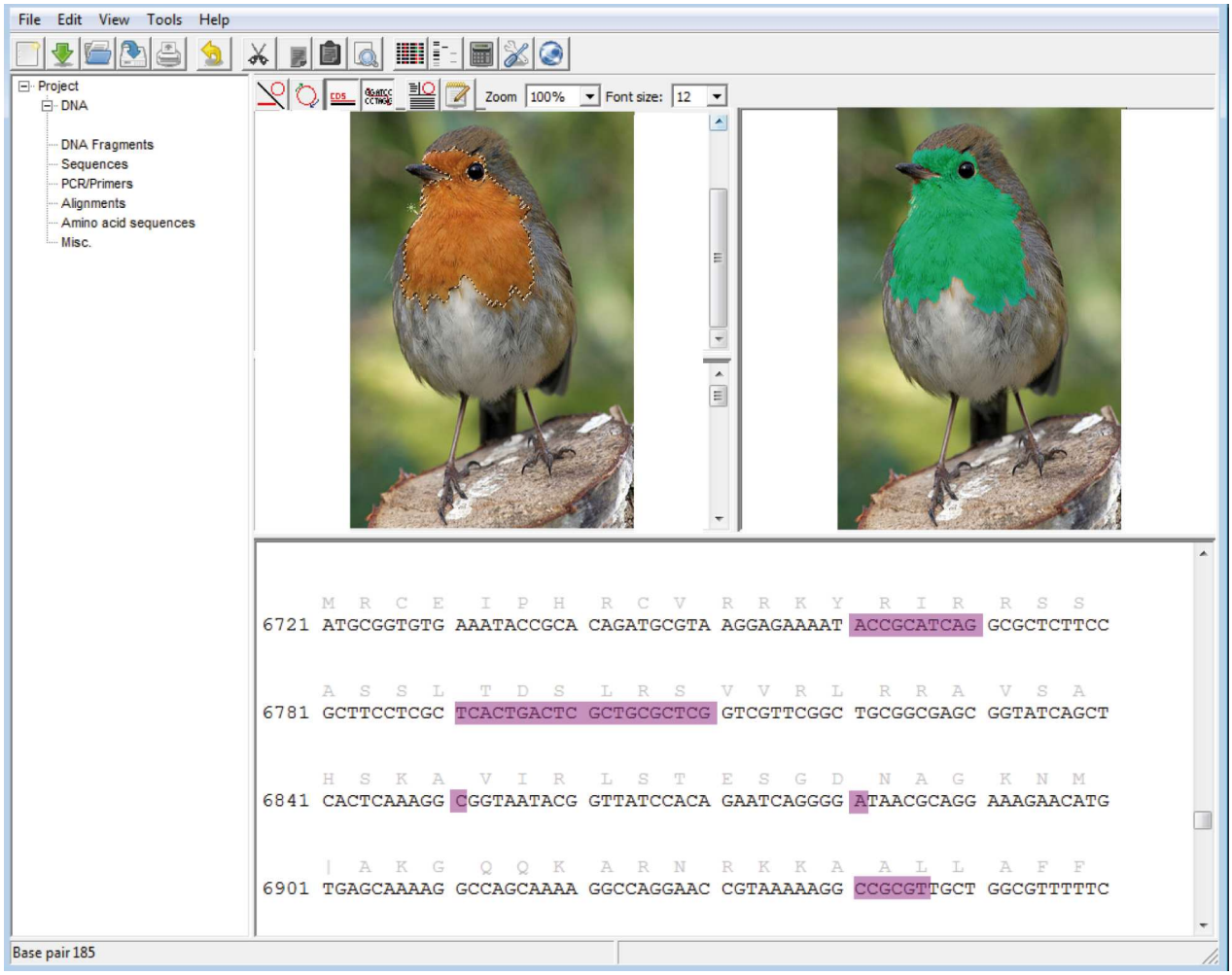



**Figure XV.4** – Le vert-gorge est un artefact de l'homme. Ici j'ai voulu representer un programme futuriste (une « global user interface » de type WYSIWYG « what you see is what you get »). Ce programme permet de choisir les modifications phenotypiques du rouge-gorge que l'on souhaite effectuer, puis il se charge ensuite d'adapter le genotype. Il suffit ensuite de faire synthetiser le code ADN proposé, le transformer dans un chassis cellulaire pour obtenir, après developpement, le rouge-gorge de son choix. Plus precisement, ici je selectionne la gorge rouge avec une baguette magique type « photoshop » puis je la « repeins » en vert. Le programme adapte en temps réel certaines zones de l'ADN (surlignées en violet) pour permettre la synthèse d'un pigment vert à la place du pigment rouge par exemple. Ma conviction est que ce vert-gorge, pur prolongement de la main de l'homme (un artefact, un outil) n'est pas beau comme le vrai rouge gorge « naturel ». La présence de l'interface graphique (la GUI) qui signifie le contrôle de l'homme sur la nature fait perdre instantanement toute beauté à cette même nature.

*« Ne plus vouloir et ne plus évaluer et ne plus créer ! Ô que cette grande lassitude reste toujours loin de moi »*

*« Dans la recherche de la connaissance, ce n'est encore que la joie de la volonté, la joie d'engendrer et de devenir que je sens en moi ; et s'il y a de l'innocence dans ma connaissance c'est parce qu'il y a en elle de la volonté d'engendrer ».*

*« Cette volonté m'a attiré loin de Dieu et des Dieux ; qu'y aurait-il donc à créer, s'il y avait des Dieux ? »*

*« Les valeurs changent lorsque le créateur se transforme. Celui qui doit créer détruit toujours. »*

*« Les créateurs furent d'abord des peuples et plus tard seulement des individus. En vérité, l'individu lui-même est la plus jeune des créations. »*

*« Il y a eu jusqu'à présent mille buts, car il y a eu mille peuples. Il ne manque que la chaîne des mille nuques, il manque le but unique. L'humanité n'a pas encore de but. »*

*« L'humanité n'a pas encore de but. Mais dites moi mes frères si l'humanité manque de but, n'est elle pas elle-même en défaut ? »*

### 5.2.9 Le désespoir /La souffrance/ la mort

Selon Kierkegaard, le désespoir des hommes est universel. Ce désespoir jouerait-il un rôle dans notre vouloir ?

*« La conception courante du désespoir en reste à l'apparence, elle n'est qu'une vue superficielle, non pas une conception. Elle prétend que chacun de nous soit le premier à savoir s'il est désespéré ou non. L'homme qui se dit désespéré, elle croit qu'il l'est, mais il suffit qu'on ne croie pas l'être, pour qu'on ne passe pas non plus pour l'être. On raréfie ainsi le désespoir, quand en réalité il est universel. Le rare ce n'est pas d'être désespéré au contraire le rarissime, c'est vraiment de ne pas l'être. »* [38]

Et qu'en est-il de la mort ? Est-elle nécessaire à l'agir humain ?

*« Dans quelle mesure la prolongation de la vie est-elle désirable ? Comme le signalait Kierkegaard, la mort envisagée dans le sérieux, n'est elle pas une source d'énergie comme nulle autre ? Ne stimule elle pas l'action ? Et si la transformation de l'essence de l'agir humain faisait elle perdre à l'esprit sa profondeur ? A l'homme animé de sérieux, la pensée de la mort à venir indique le but à diriger sa course. La mort est*

[38] Kierkegaard, *Traité du désespoir*, folio essais p. 78.

*un stimulant de la vie et le sérieux comprend que l'idée de la mort représente une invitation à l'action. Ne perdons pas notre temps ! »* [39]

Et la souffrance : est-il vraiment souhaitable de vouloir la faire disparaitre ?

*« Vous voulez abolir la souffrance dans la mesure du possible, et il n'y a pas de plus folle ambition. Et nous ? Il semble que nous la voudrions encore plus profonde et plus grave qu'elle ne le fut jamais. Le bien-être tel que vous le concevez n'est pas un but, c'est à nos yeux un terme. Un état qui rend l'homme aussitôt ridicule et méprisable, qui fait souhaiter sa ruine. La culture de la souffrance, de la grande souffrance, ne savez vous pas que c'est là l'unique cause des dépassements de hommes ? Cette tension de l'âme dans le malheur, qui l'aguerrit, son frisson au moment du grand naufrage, son ingéniosité et sa vaillance à supporter le malheur, à l'endurer, à l'interpréter, à l'exploiter jusqu'au bout, tout ce qui lui a jamais été donné de profondeur, de secret, de dissimulation, d'esprit, de ruse, de grandeur, n'a-t-il pas été acquis par la souffrance, à travers la culture de la grande souffrance ? [...] Et que votre pitié s'adresse à la créature dans l'homme, à ce qui doit être façonné, brisé, forgé, taillé, brulé, porté à incandescence et purifié, à ce qui souffrira nécessairement et doit souffrir. »* [40]

*« On peut presque classer les hommes d'après la profondeur que peut atteindre leur souffrance [...] Il est des esprits libres et insolents qui voudraient cacher et nier qu'ils sont des cœurs brisés, fiers et incurablement blessés ; la bouffonnerie elle-même est quelque fois le masque d'un savoir douloureux et trop lucide. D'où il suit qu'on fera preuve de délicatesse en respectant « le masque » et en n'allant pas faire de la psychologie et placer sa curiosité au mauvais endroit »* [41]

J'aimerai faire une précision avant de clôturer ce chapitre. Nous avons étudié une piste, à transmettre potentiellement aux générations futures, qui postule que le désespoir, la souffrance et la mort sont nécessaires à l'agir humain. Cette piste, qui peut être bonne comme mauvaise, est proposée dans un cadre strictement philosophique. En tant que chercheur, je me dois, et cela ne représente d'ailleurs pas une difficulté, de dissocier mes conceptions scientifiques pragmatiques des concepts philosophique plus abstraits, relatifs au futur. En d'autres termes, mon moteur en tant que biologiste, consiste à combattre, par la recherche, la souffrance morale et/ou physique (des malades) engendrée par des maladies horribles. Pour cela, on cherche (la communauté

[39] Jacqueline Russ, *La pensée éthique contemporaine*, PUF p. 33.
[40] Nietzsche, *Par delà le bien et le mal*, folio essais p. 144.
[41] Nietzsche, *Par delà le bien et le mal*, folio essais p. 197.

scientifique), souvent avec acharnement, à un niveau fondamental ou appliqué, tout ce qui pourrait atténuer ou guérir à court, moyen ou long terme les souffrances et les maladies que subissent les malades. Leurs espoirs en la recherche et aux chercheurs ne représentent pas « rien ». Au contraire, c'est l'impératif, le carburant de mon moteur chaque matin. Le fait que mes recherches (en biologie théorique/synthétique) soient découplées (en apparence seulement) des applications médicales ne m'empêche pas de conserver, comme fondement de mon agir, les objectifs de la médecine. Du moins je l'espère et je veux y croire. C'est un objectif qui peut sembler naïf à certains mais je l'assume totalement pour une raison simple : c'est le seul qui vaille la peine. Descartes, il y a 400 ans déjà, ne s'y trompait pas :

> *« Mais je dirai seulement que j'ai résolu de n'employer le temps qu'il me reste à vivre à autre chose qu'à tacher d'acquérir quelque connaissance de la nature qui soit telle qu'on en puisse tirer des règles pour la médecine plus assurées que celles qu'on a eues jusqu'à présent »* [42]

[42] Descartes, *Fin du Discours de la méthode*, le livre de poche p. 171.

# Chapitre XVI

# Un peu d'épistémologie

## 1 Préambule

J'ai pris la décision d'écrire un chapitre d'épistémologie relativement tôt durant ma thèse. Il est vrai que cela peut surprendre dans une thèse dont le sujet est « réseau de régulation génique chez *E. coli* ». Je ressens donc le besoin de me justifier en 3 points. Si vous ne comprenez pas cette justification ou que vous ne savez pas ce qu'est l'épistémologie alors sautez le préambule et lisez directement la suite.

1. L'institution positiviste qui m'emploie me « demande » de travailler sur les réseaux de régulation génique. Or ceci est légèrement paradoxal car il suffit de « gratter » un petit peu pour se rendre compte que la notion de réseau, de système est la chasse-gardé des épistémologies constructivistes concurrentes. Ces épistémologies défendent justement l'idée qu'il faut décloisonner les sciences et réfléchir au système complexe que les sciences représentent.

   > *« Il ne faut pas éliminer l'hypothèse d'un néo-obscurantisme généralisé, produit par le mouvement même des spécialisations, où le spécialiste lui-même devient ignare de tout ce qui ne concerne pas sa discipline, où le non-spécialiste renonce d'avance à toute possibilité de réfléchir sur le monde, la vie, la société, laissant ce soin aux scientifiques, lesquels n'en ont ni le temps, ni les moyens conceptuels. Situation paradoxale que celle où le développement de la connaissance instaure la résignation à l'ignorance et où le développement de la science est, en même temps, celui de l'inconscience. »*[1]

   Je me revendique donc de l'héritage des épistémologies constructivistes et me place sous leur protection. En effet, ce sont ces dernières qui me permettent de décloisonner mon sujet de thèse en l'envisageant sous des angles variés.

[1] Edgar Morin, *Science avec conscience*, p. 17.

2. Il me semble que, dans d'autres domaines de la biologie que le mien, par exemple en physiologie, en neurologie ou en écologie, les expériences sont lourdes à mettre en place. Elles sont aussi plus chères et prennent plus de temps (semaines, mois). Au final, le nombre de cycles hypothèse–expérience–interprétation qu'un chercheur pourra mener sera limité. En microbiologie, en particulier avec *E. coli*, il est possible de faire un très grand nombre de cycles hypothèse–expérience–interprétation (des centaines, milliers) car la bactérie pousse très vite (pré-culture en une nuit) et l'expérience en elle-même ne coûte pas cher. Ce grand nombre de cycles permet de se forger une image intuitive du fonctionnement de la bactérie. Mais il permet en outre d'évaluer (en essayant de prendre du recul) comment avance « sa » science et si les découvertes produites sont fonction de l'application stricte de la méthode scientifique. En ce qui me concerne, je crains que la réponse soit clairement non et ce, même si vérifiabilité et reproductibilité ont joué un rôle incontestable.
3. Je me suis heurté, dans mon environnement scientifique, à un vif rejet lorsque je défendais les positions suivantes :
   ⇒ la méthode scientifique n'a pas de base solide et toute démonstration contient une part de persuasion.
   ⇒ la science est impure car se cache derrière elle un scientifique qui est avant tout un homme. Heureusement que ces impuretés sont là sinon la science ne pourrait pas avancer.
   ⇒ La science n'avance pas seulement selon le schéma hypothèse–expérience–résultat. En biologie, les systèmes expérimentaux n'aident pas seulement à répondre aux questions, ils aident à la générer.

   Etant un piètre avocat, je perd systématiquement tout débat. Pourtant mes arguments proviennent des idées d'épistémologues et sociologues reconnus mais elles restent, il me semble, relativement mal acceptées/connues par les scientifiques.

   > *« Ainsi nul n'est plus désarmé que le scientifique pour penser sa science. La question « qu'est-ce que la science ? » est la seule qui n'ait encore aucune réponse scientifique. C'est pourquoi s'impose plus que jamais la nécessité d'une auto connaissance de la connaissance scientifique. Celle-ci doit faire partie de toute politique de la science, comme de* ***la discipline mentale du scientifique*** *».*[2]

   Or comme j'ai la conviction que notre rôle (de scientifique mais aussi d'humain) consiste à propager les idées/les constructions cognitives/les pistes auxquelles on croit, je m'y emploie à l'écrit à défaut de réussir à l'oral.

[2]Edgar Morin, *Science avec conscience*, p. 20.

Malgré mon prosélytisme marqué, je trouve normal que les attaques contre « la sacro-sainte méthode scientifique positiviste » peinent à envahir la sphère scientifique (les laboratoires). En effet, pendant l'éducation, on apprend aux élèves des règles, des méthodes, de la rigueur, on leur laisse entendre que les mathématiques sont disciplines reines car plus rigoureuses. Puis arrive un anarchiste, Paul Feyerabend, dans les années 70 qui propose son provoquant :

> *« Étant donné un but, l' « a-méthode » [absence de méthode] de l'anarchiste [épistémologique] a une plus grande chance de réussir que n'importe quel ensemble de critères, règles ou prescriptions bien définis »* [3]

Comment voulez-vous expliquer à ces élèves (devenus chercheurs) que, finalement, l'anarchiste n'a peut être pas tort : l'absence de méthode vaut mieux que tout. La pilule passe mal et c'est normal.

Après cette justification un peu longue, voici mon plan. J'essaierai, dans ce chapitre, de mettre en perspective quelques résultats et idées qui ont été obtenus en épistémologie au cours du siècle dernier. Il s'agit principalement des travaux de Karl Popper, Thomas Kuhn, Paul Feyerabend, Edgar Morin et Jean louis le Moigne. Plus surprenant, j'ai trouvé chez Nietzsche des idées d'épistémologie très avant-gardistes. Je ne me souviens pas avoir lu d'ouvrages de scientifiques ou épistémologues citant explicitement son travail.

Dans les chapitres qui suivent (plus personnelles), je prônerai le décloisonnement du système de publication actuel et je montrerai le rôle caché des arbitrages (falsification) en science. Puis je marcherai sur des œufs en analysant la place de la discipline en science et je déconstruirai la critique. Enfin, je proposerai une réflexion personnelle sur la question « qu'est-ce que la science ? ».

# 2 Le socle de base n'est pas sûr

## 2.1 Les épistémologies positivistes

Savez-vous ce qu'est l'épistémologie ?

L'épistémologie est le domaine de la philosophie des sciences qui étudie les sciences et la connaissance. Elle cherche à comprendre et à décrire les règles de fonctionnement de la science, par exemple la fameuse méthode scientifique. Mais plus profondément, elle cherche à percer à jour les règles qui sous tendent l'acte de « connaitre ».

Savez-vous ce qu'est l'épistémologie positiviste ?

[3] Paul Feyerabend, *Contre la méthode*, Points p. 215.

L'épistémologie positiviste est l'épistémologie dominante et donc institutionnelle. Elle explique grâce au principe de « vérifiabilité » le fait que, au cours des millénaires, la science progresse alors que, au contraire, les religions stagnent. On utilise ce principe en laboratoire tous les jours. Si vos expériences ne sont pas vérifiables, l'institution ne vous permettra pas de prétendre qu'elles s'inscrivent dans le cadre de la démarche scientifique. Au cours du 20$^{\text{ème}}$ siècle, Karl Popper a proposé une amélioration du principe de « vérifiabilité » : le principe de « falsifiabilité ». Karl Popper part du constat que l'on ne peut pas réellement « vérifier » ce qui est vrai. Selon lui, lorsque qu'un scientifique propose une théorie, il faut qu'il soit en mesure de proposer une expérience qui, si elle s'avérait non concluante/fausse, démontrerait que sa théorie est fausse. Voici un exemple : je propose la théorie suivante : La bactérie *E. coli* ne pousse pas sur une gélose constituée de 70% d'éthanol. Ensuite je vous propose de « réfuter » ma théorie à l'aide du principe de falsification « si vous préparez une gélose constituée de 70% d'éthanol puis inoculez dessus des bactéries *E. coli* et que ces dernières réussissent à former des colonies, *alors ma théorie est fausse* ». Quelles que soient la théorie et l'expérience, l'énoncé devra toujours être structuré de la même manière : Si [mon expérience ne fonctionne pas], alors ma théorie est fausse. Ce principe permet d'écarter des théories pseudo-scientifiques capables d'expliquer tout et son contraire. Il faut noter que Karl Popper n'était pas un positiviste (il semble qu'il en rigole lui-même en parlant de « la légende de Popper »). Cependant, le fait de proposer un principe opérant une démarcation entre science et non science (c'est sa plus grande erreur) a contribué à le faire considérer par beaucoup comme un positiviste.

Mais ces principes de vérifiabilité et falsifiabilité ne correspondent pas aux fondements de l'épistémologie positiviste. Jean-Louis le Moigne[4] identifie deux hypothèses aux fondements :

⇒ L'hypothèse ontologique : Elle nous est si familière, à nous scientifiques, qu'elle n'est pas toujours perçue comme une hypothèse épistémologique.

> *« La réalité essentielle de la réalité existentielle : la connaissance que constitue progressivement la science est la connaissance de la réalité, une réalité postulée indépendante des observateurs qui la décrivent (même si leurs descriptions n'en sont pas indépendantes) »*

⇒ hypothèse déterministe :

> *« Il existe quelque forme de détermination interne propre à la réalité connaissable, détermination elle-même susceptible d'être connue. Sous sa forme la plus familière, le déterminisme est un causalisme [...]. Ce n'est pas parce que les lois déterminant la réalité ne peuvent être établies et donc connues du fait de l'in-*

[4]Jean-Louis le Moigne, *Que sais-je : Les épistémologies constructivistes*, PUF p. 20 et 23.

*complétude de nos instruments de mesure (sensibilité aux conditions initiales, principe d'incertitude d'Heisenberg) que ces lois n'existent pas »*

A ces deux hypothèses, Jean-Louis le Moigne[5] ajoute deux méthodes qui en découlent :

⇒ le principe de modélisation analytique qui consiste à décomposer en parties, en petites briques élémentaires. On se souviendra du deuxième précepte du discours de la méthode qui propose « *de diviser chacune des difficultés [. . .] en autant de parcelles qu'il se pourrait et qu'il serait requis pour les mieux résoudre* »[6]

⇒ le principe de raison suffisante (qui découle du déterminisme). Ce principe postule l'existence des causes mais également l'équivalence entre cause et effet : « Si A est cause de B, B ne peut être causé que par A et donc A est la raison suffisante ou l'explication certaine de B ». Cette formulation de raison suffisante nous vient de Leibniz « *rien jamais n'arrive sans qu'il n'y ait une cause ou du moins un raison déterminante.* ». Ce principe ajoute subrepticement une hypothèse forte sur la nature du raisonnement positif ou de l'explication : celle de la naturalité de la logique déductive (formelle). Cette dernière acquiert un statut disciplinaire privilégié dans le champ des connaissances car elle assure que l'ordonnancement (les longues chaines de causes à effet) est naturel ou réel et donc rend indiscutablement compte de la réalité.

Jusque-là, si vous n'aviez jamais entendu parler d'épistémologie, sachez qu'il est fort probable que vous travailliez, dans votre laboratoire, sous le régime du positivisme et de sa logique déductive. Cela n'est pas un problème car l'épistémologie positiviste a fait ses preuves. Cependant, il est important de garder à l'esprit que les fondements de cette épistémologie (comme de toutes autres d'ailleurs) sont des hypothèses. C'est-à-dire des énoncés qui n'ont pas pu passer l'examen de la vérifiabilité. Autrement dit le socle, le fondement de notre méthode n'est pas sûr. Il dépend de ces hypothèses qui dépendent elles-mêmes de nos perceptions sensorielles. Il est temps maintenant de passer au deuxième type d'épistémologie que je souhaite décrire.

## 2.2 Les épistémologies constructivistes

Ces épistémologies sont également basées sur des hypothèses. Autrement dit le socle n'est pas plus sûr mais il ne l'est pas moins. Les quelques paragraphes qui suivent sont trop empruntés à l'ouvrage de Jean-louis le Moigne pour que je puisse m'en attribuer la paternité. Mon travail

[5] Jean-Louis le Moigne, *Que sais-je : Les épistémologies constructivistes*, PUF p. 27 et 32. Dans un objectif de synthèse, j'ai parfois dû remanier le texte de l'auteur ce qui m'a conduit à enlever les guillemets. Cependant mes modifications sont limitées et je considère donc que la paternité de ces phrases revient à l'auteur initial.

[6] Descartes, *Discours de la méthode*, le livre de poche, p. 89.

consiste en une synthèse, une reconstruction rapportant et remaniant les phrases et définitions les plus limpides que l'auteur ait écrites. Mon but affiché est de tenter de faire « saisir » au lecteur, en quelques lignes, des concepts fondamentaux. Pourtant, j'ai bien conscience que ma synthèse est difficile d'accès et demande un effort d'abstraction important. Je n'ai pas réussi à la rendre plus claire. Je renvoie donc le lecteur insatisfait vers que le *que-sais-je ?* limpide de Jean-louis le Moigne. Pour éviter toute lourdeur, je n'utiliserai pas les guillemets mais placerai les paragraphes remaniés en italique.

On ne peut pas exclure l'hypothèse que l'on vive dans une sorte de rêve, de matrice, de jeu video. Dés lors, notre accès à la réalité est limité. Par exemple, le sol sur lequel le bonhomme d'un jeu vidéo marche n'existe pas réellement dans notre réalité, il est virtuel. Pour ce bonhomme, au contraire, le sol a une réalité physique puisqu'il ressent un « phénomène » qui le maintient et l'empêche de tomber.




**Figure XVI.1** – Renoncer à la recherche d'une vérité objective.

J'ouvre ici une petite parenthèse réservée aux théoriciens/logiciens. Voici mon problème : Soit un scientifique « à la pensée libre » (c'est-à-dire non créé par le programmeur) se retrouvant dans un jeu vidéo (un programme créé de toute pièce avec des 0 et des 1). Via l'utilisation illimitée de la raison et de l'expérience, le scientifique peut-il :

⇒ reconstituer toutes les lois du monde virtuel dans lequel il vit ?

⇒ reconstituer toutes ces lois et déterminer, par ingénierie inverse, le code source du programme ?

⇒ reconstituer les lois + le code source du programme + les lois de notre monde réel ?

Autrement dit, existe-t-il une limite infranchissable entre les lois (issues de phénomènes dits virtuels) et le code source ? Si non, existe-t-il alors une limite infranchissable entre le code source et nos lois (issues de phénomènes dits réels ) ? Si vous disposez d'éléments de réponse ou d'une démonstration, n'hésitez pas à me contacter. Fin de la parenthèse.

*Autrement dit, le bonhomme ne peut pas exclure l'existence d'un réel connaissable qui lui serait étranger. Ainsi, la connaissance qu'il peut construire du réel est celle de sa propre expérience du réel. Il y a une renonciation consciente à la valeur de vérité objective qui a sans doute longtemps retardé l'acceptation par l'épistémologie institutionnelle. La réalité connaissable doit pouvoir être cognitivement construite. Par conséquent, l'hypothèse relative à la construction de la connaissance ne fait plus appel à une norme du « vrai » mais à une norme de « faisabilité » : « Entre l'être et le connaître, le faire »*[7]*. Les constructions cognitives sont d'ailleurs familières depuis longtemps aux mathématiciens qui font exister ces objets réels par construction que sont les figures géométriques, les nombres ou les opérateurs symboliques. On a toujours cherché des explications quand c'était des représentations qu'on pouvait seulement essayer d'inventer. Le sujet connaissant ne représente pas des choses mais des opérations (ou des interactions) et la connaissance qu'il en construit par des représentations est elle-même opératoire ou active. La représentation active d'un phénomène connaissable transforme récursivement la connaissance que nous en avons, laquelle, à son tour...« En changeant ce qu'il connait du monde, l'homme change le monde qu'il connaît. Et en changeant le monde dans lequel il vit, l'homme se change lui-même »*[8]*. La connaissance que construit le sujet par son expérience organise simultanément le mode de construction de cette connaissance ou son intelligence. On ne peut donc plus dés lors séparer la connaissance de l'intelligence (ou cognition) qui la produit et il nous faut entendre la connaissance par le processus qui la forme autant que comme le résultat de ce processus de formation. L'intelligence organise le monde en s'organisant elle-même.*

Jean-Louis le Moigne identifie deux hypothèses à la base des connaissances constructibles :

⇒ *L'hypothèse phénoménologique : le réel connaissable est un réel phénoménologique, celui que le sujet expérimente. Ce dernier ne connait pas de chose en soi (hypothèse ontologique) mais il connait l'acte par lequel il perçoit l'interaction entre les choses.*

⇒ *L'hypothèse téléologique : En attribuant au sujet connaissant le rôle décisif dans la construction de la connaissance, l'hypothèse phénoménologique oblige en quelque sorte*

[7] Paul Valery, *cahier I.*

[8] Formule de Th. Dobzhansky (1962).

*à prendre en compte l'intentionnalité ou les finalités de ce sujet connaissant. Cette hypothèse téléologique n'est pas encore très volontiers acceptée par l'épistémologie institutionnelle positiviste. Or comme l'avait remarqué H. Von Foerster, à la réponse à la question « pourquoi », on peut tout aussi bien s'exprimer en termes de « à fin de » qu'en termes de « parce que ». Ainsi, connaitre en termes de fins plausibles constitue un mode de connaissance au moins aussi bien raisonné que de connaitre en termes de causes probables.*

*On peut donc résumer ainsi le paradigme constructiviste comme étant l'expression téléologique d'expériences cognitives s'articulant systémiquement dans leur contexte et susceptibles d'être manipulés (ou computés) selon des procédures cognitives reproductibles. La valeur socioculturelle du paradigme constructiviste pourrait aider à organiser et reconnaitre comme vraie science, les sciences de l'artificiel, la où on retrouve une téléologie comme les sciences des systèmes (informatique, science économique par exemple). Le but étant de leurs donner un appui épistémologique pour les aider à sortir du cadre du charlatanisme. C'est la conception d'une discipline scientifique par leur projet plutôt que par leur objet.*

J'aimerais revenir sur les constructions cognitives qui font exister les figures géométriques, les nombres ou les operateurs symboliques. Le point important c'est que ces constructions cognitives sont « faisables ». Or, la compréhension du fonctionnement des réseaux de régulation intracellulaires (intra-neuronal) et extracellulaire (réseau de neurones) peut conduire, dans le futur, à un accroissement de capacités computationnelles/cognitives. Cela pourrait potentiellement élargir le « faisable » et permettre la construction de nouveaux objets mathématiques qui n'étaient jusqu'alors pas « pensables », pas « constructibles ». Les avancées et les futures « ruptures/révolutions » en mathématiques pures pourraient provenir non plus de l'application des règles de la logique formelle mais par l'élargissement, par des méthodes issues de la biologie, de nos capacités cognitives. J'utilise cet exemple pour illustrer le phénomène de récursivité dans l'élaboration de la connaissance : le cerveau élabore par construction une connaissance (par exemple en biologie, en mathématique) capable de modifier, d'améliorer le cerveau. Ce dernier produira alors une nouvelle connaissance par accroissement de « faisabilité », etc...

# 3 La structure des avancées scientifiques

## 3.1 La pyramide

Nous avons vu dans cette première partie qu'il existe une incertitude sur les fondements de la science. Cependant, nous avons le sentiment qu'il y a, au cours de l'histoire, une pente

positive du progrès/des connaissances et ce, par opposition à la pente des connaissances issues des religions qui semble, elle, rester constante[9]. Comme le dit Bachelard, « *L'histoire des sciences est la plus irréversible de toutes les histoires* ». Ainsi, on se représente les connaissances comme une pyramide formée par l'ajout au fur et à mesure de briques vérifiables. L'image de la pyramide est robuste chez les scientifiques car c'est comme cela que la science est enseignée durant l'apprentissage primaire et secondaire. Mais c'est une image qui rend mal compte du développement réel des sciences car il existe des ruptures brutales qui provoquent des avancées majeures : par exemple l'arrivée de la relativité générale en remplacement de la mécanique Newtonienne. Je vais donc maintenant substituer l'image de la pyramide par celle du cercle.

## 3.2 Le cercle

Dans son ouvrage *La Structure des révolutions scientifiques* publié en 1962[10], Thomas Kuhn s'efforce de montrer pourquoi le développement scientifique n'est pas un processus linéaire/ cumulatif. En effet, selon lui, la science progresse de manière fondamentalement discontinue, c'est-à-dire non par accumulation mais par rupture. L'évolution des idées scientifiques s'organise en deux grandes phases alternatives, qualifiées de science normale et de science extraordinaire. La phase de science normale est la phase usuelle, celle qui constitue en durée l'essentiel de l'histoire des sciences. Un groupe de scientifique adhère massivement à un paradigme qui par ses « accomplissements scientifiques passés » et sa logique, fournit « le point de départ d'autres travaux ». Un paradigme est une représentation du monde qui repose sur une base définie (une forme de rail de pensée). C'est l'ensemble des croyances, des techniques, des règles admises et intériorisées comme « normes » qui sont partagées par les membres d'un groupe de scientifique, au cours d'une période de consensus théorique. Je demande au lecteur de se représenter le paradigme comme un cercle. La science qui est y faite est essentiellement prédictive : les scientifiques testent inlassablement le paradigme en concevant de nouvelles énigmes de manière à le renforcer (à élargir le cercle). Une bonne énigme doit être suffisamment difficile pour être intéressante mais suffisamment simple pour être résolvable et ainsi permettre aux scientifiques de montrer leur habilité et leur ingéniosité : « *La conviction que, si seulement il est assez habile, il réussira à résoudre une énigme que personne encore n'a résolue, ou résolue aussi bien* »[11]. Les

[9] Le relativisme, poussé à l'extrême, nie l'existence même de « progrès » (la dérivée positive) qui ne serait qu'une vue de l'esprit. Bien qu'il soit impossible, en l'état actuel des connaissances, de le prouver, je pense qu'il existe un progrès scientifique (je ne suis pas un relativiste). Je propose deux pistes d'étude : la première est dans la théorie des exemplaires, chapitre sur la vie et la seconde est une méthode « scientifique » pour caractériser le progrès à la fin de la partie « analyse du système de publication actuel ».

[10] Ce paragraphe s'inspire en partie de l'article Wikipedia « La structure des révolutions scientifiques ».

[11] Thomas Kuhn, *La structure des révolutions scientifiques*, Flammarion p. 64.

scientifiques normalisent tout l'espace à l'intérieur du cercle par des exercices de nettoyage, de sorte à faire entrer « *la nature dans leur boîte* ». Les problèmes qui y sont résolus sont de taille très limités. Mais inévitablement, des anomalies apparaissent, généralement dénoncées par des théories concurrentes. La plupart d'entre elles seront expédiées par des raisonnements *ad-hoc* pour sauver le paradigme. Mais si ces anomalies sont prises en considération, alors une crise s'installe et c'est le début d'une révolution scientifique : on passe dans la phase du régime extraordinaire. On voit alors apparaitre de nouvelles théories (des nouveaux petits cercles) cherchant à expliquer les anomalies et à remplacer le paradigme mis en défaut. Ces théories sont donc candidates au titre de nouveau paradigme. Bien sûr, durant la crise, de nombreux scientifiques (ceux qu'une carrière féconde avait profondément engagés dans l'ancien paradigme) gardent « *la certitude que l'ancien paradigme parviendra [finalement] à résoudre tous les problèmes* »[12]. Ces résistances sont utiles car elles « *empêchent que le paradigme soit trop facilement renversé* » ce qui évite que « *les scientifiques [soient] dérangés sans raison* »[13]. Le processus de révolution n'est pas immédiat. Il progresse par des prises de positions successives de la part des groupes scientifiques confrontés à une crise. Lorsque ces différents groupes se rallient finalement à une nouvelle théorie consensuelle, celle-ci devient un paradigme et la révolution est achevée. La science normale, prédictive, réapparait et les scientifiques sont chargés de nettoyer et d'élargir le cercle jusqu'à ce que de nouvelles anomalies apparaissent.

## 3.3 Les poupées russes

Le schéma des cercles souffre selon moi d'une lacune (je ne sais pas si la critique qui suit a déjà été opposée à Thomas Kuhn mais cela ne m'étonnerait pas). En effet, j'imagine plutôt une structure invariante d'échelle (type poupées russes) : Autrement dit il existerait selon moi une multitude de cercles dans le cercle. Par ses expériences, tout scientifique teste et met en danger chaque jour des paradigmes de tailles différentes. Comme le dit Jacques Monod :

> *« En confrontant systématiquement la logique et l'expérience, selon la méthode scientifique, c'est en fait toute l'expérience de [nos] ancêtres [nos prédécesseurs] que nous confrontons avec l'expérience actuelle »*[14]

Un étudiant, par exemple, peut être amené à déconstruire une petit brique de l'étudiant précédent en repérant un artefact ou une anomalie passé jusque là inaperçu. Et il est fort possible qu'il rencontre la fameuse résistance de Kuhn, d'autant plus si « la brique » a été publiée. Bien

[12]Thomas Kuhn, *La structure des révolutions scientifique*, Flammarion p. 209.
[13]Thomas Kuhn, *La structure des révolutions scientifique*, Flammarion p. 99.
[14]Jacques Monod, *Le hasard et la nécessité*, Points p. 198.

sûr, déconstruire le « micro-paradigme » d'un chercheur est un processus courant alors que déconstruire le paradigme de quelques milliers de chercheurs est beaucoup plus rare pour des raisons statistiques évidentes : il est courant qu'un chercheur se trompe et beaucoup moins courant que 1000 chercheurs se trompent. C'est cette rareté qui rend les gros paradigmes beaucoup plus visibles, plus médiatiques. Notez enfin que, du point de vue d'un paradigme donné (un cercle), tous les paradigmes du dessous (inscrits dans ce cercle) représentent la science normale.

## 3.4 La montagne

Une notion essentielle qui n'est pas captée distinctement par l'image des cercles paradigmatiques c'est la notion d'erreur. Pour intégrer ces erreurs, il faut imaginer une heuristique à la recherche d'un optimum global. L'image de la montagne aide à se représenter le processus. Imaginez un point de départ (tout en bas de la montagne) et un objectif à atteindre : le sommet de la montagne qui symbolise au choix, la verité, la connaissance, dieu, la réalité connaissable etc... Les scientifiques cherchent, depuis des milliers d'années, le chemin, la voie d'escalade à travers l'immense falaise qui les mènera au sommet. Ils avancent souvent en suivant la voie la plus simple, la plus intuitive, là ou les expériences sont facilement reproductibles et donc faciles. Mais inexorablement les difficultés augmentent, les expériences se compliquent, les phénomènes ne se laissent pas approcher sans douleur. Sans parler de la multitude de biais cognitifs qui les oriente dans la mauvaise direction et qui nécessite une grande énergie pour être dépassée. Certains scientifiques sont chargés d'avancer tout droit, coûte que coûte, centimètre par centimètre, d'autres explorent les environs directs : ils redescendent un peu, quelques mètres, dizaines de mètre à la recherche d'un passage plus simple pour effectuer une percée. Enfin certains, mais ils sont rares, sont capables de redescendre très bas, presque tout en bas pour choisir une autre voie qui mènera plus haut.

Redescendre c'est reconnaitre que le chemin parcouru peut potentiellement mener à une impasse. Ce n'est pas le constat d'un échec mais bien le processus normal de recherche de toutes heuristiques. Ainsi ces erreurs (qui orientent dans la mauvaise direction en devenant des dogmes) ainsi que ce qui les corrige (les retours en arrière, les déviations) sont le fonctionnement normal de la science et donc les conditions du progrès. Ces erreurs font intrinsèquement parties de l'heuristique de recherche : elles en sont l'essence même. Georges Canguilhem a bien identifié le processus de redescente (l'effort violent dont il parle) :

> *« Il est instructif de constater combien plus et mieux que leur continuateurs et commentateurs, les fondateurs de théories nouvelles se sont rendus compte des faiblesses et des insuffisances de leurs systèmes. Leurs réserves sont ensuite oubliées ce qui*




**Figure XVI.2** – La structure des connaissances.

*pour eux était hypothèse devient dogme de plus en plus intangible à mesure qu'on s'éloigne davantage des origines et* ***un effort violent devient nécessaire pour s'en délivrer*** *lorsque l'expérience vient démentir les conséquences plus ou moins lointaines d'idées dont on avait oublié le caractère provisoire et précaire. »*[15]

# 4 La déconstruction du critère de falsifiabilité

Paul Feyerabend est un philosophe des sciences qui a écrit un ouvrage (dont je conseille vivement la lecture) *Contre la méthode* publié en 1975. Il se définit comme un épistémologue anarchiste. Il nie l'existence de règles méthodologiques universelles, nie l'existence d'une démarcation entre science et non science et donc nie que le critère de falsifiabilité de Popper puisse opérer cette démarcation. Selon lui, en science, *« tout est bon »*.

Feyerabend a vivement critiqué l'attitude consistant à juger la qualité des théories scientifiques (prédictions) en les comparants avec des résultats expérimentaux (observations). Son analyse de « l'argument de la tour » est particulièrement convaincante. Cet argument constituait l'objection majeure à la théorie tentant de démontrer que la terre tourne. Beaucoup pensaient, à l'époque de Copernic, que le fait qu'une pierre tombant d'une tour atterrisse juste devant la tour prouve que la terre est immobile. En effet, ils pensaient que si la terre effectuait une rotation pendant que la pierre tombait, celle-ci aurait atterri derrière la tour. Si la terre tournait, les objets ne tomberaient pas à la verticale mais en diagonale selon eux. Or comme cela ne se produisait pas, cela prouvait l'immobilité de la terre.

Je ne vais pas paraphraser plus longtemps la pensée de Feyerabend. Je laisse découvrir à mon lecteur un avant goût de la clarté, de la profondeur et de l'acidité de son texte via quelques citations choisies.

### 4.0.1 L'évidence est viciée

*« La prise en compte de tous ces éléments : terme d'observation, noyau de facultés sensorielles, sciences auxiliaires, spéculations à l'arrière plan, suggère qu'une théorie peut être incompatible avec l'évidence empirique non parce qu'elle est incorrecte, mais parce que c'est l'évidence même qui est viciée »*[16]

[15] Georges Canguilhem, *La connaissance de la vie*, bibliothèque des textes scientifiques p. 56 (extrait d'une conférence).

[16] Paul Feyerabend, *Contre la méthode*, Points p. 68.

### 4.0.2 Les faits sont impurs

*« Les théories sont testées et éventuellement réfutées par des faits. Les faits contiennent des composantes idéologiques, des conceptions plus anciennes qu'on a perdues de vue ou qui n'ont peut être jamais été formulées de façon explicite. De telles composantes sont hautement suspectes ».*[17]

### 4.0.3 Aucune théorie ne devrait être rejetée

*« D'après nos résultats actuels, presque aucune théorie n'est compatible avec les faits. Exiger de n'admettre que les théories compatibles avec les faits, cela nous laisse à nouveau sans aucune théorie (il n'y a pas une seule théorie qui ne soit en difficulté d'une manière ou d'une autre) [...] La bonne méthode ne doit contenir aucune règle qui nous oblige à choisir entre des théories sur la base de la falsification ; bien plutôt ses règles doivent nous permettre de choisir entre des théories que nous avons déjà testées et qui sont réfutées [...]. Non seulement les faits et les théories sont en constant désaccord, mais ils ne sont jamais aussi nettement séparés que nous voulons le croire. Les règles de méthodologie parlent de « théories », d' « observations » et de « résultats expérimentaux » comme s'il s'agissait d'objets précis, bien définis dont les propriétés seraient faciles à évaluer et seraient compris de la même manière par tous les scientifiques. Or le matériel qui est réellement mis à disposition d'un scientifique : ses lois, ses résultats expérimentaux, ses techniques mathématiques, ses préjugés épistémologiques [...] est indéterminé de bien des manières, ambigu, et jamais complètement séparé du contexte historique »*[18]

### 4.0.4 L'attaque ironique contre les positivistes/rationalistes

*« Développez vos idées de façon qu'elles puissent être critiquées ; attaquez les impitoyablement ; n'essayez pas de les protéger, exhibez leurs points faibles ; éliminez les aussitôt que ces points faibles sont devenus manifestes – ce sont là quelques règles mises en avant par nos rationalistes critiques »* [19]

Ainsi Feyerabend argumente l'idée que toute expérience que nous faisons peut être conceptualisée de différentes manières et qu'aucune conceptualisation est dénuée d'hypothèses théoriques. Cette thèse de « dépendance à la théorie » de toutes interprétations issues d'expériences

[17]Paul Feyerabend, *Contre la méthode*, Points p. 81.
[18]Paul Feyerabend, *Contre la méthode*, Points p. 67.
[19]Paul Feyerabend, *Contre la méthode*, Points p. 187.

a été poussée à l'extrême par le sociologue des sciences Harry Collins dans son ouvrage *Tout ce que vous devriez savoir sur la science*. Il y met en évidence ce qu'il appelle « la régression de l'expérimentateur ». Cette régression apparait quand une théorie peut seulement être établie/vérifiée par la mise en place d'une expérience technique. Pour cela évidemment, la technique doit être fiable. Or le jugement qui consiste à savoir si la technique est fiable dépend des résultats. Or ces résultat seront vérifiés/validés/interprétés grâce à la théorie elle-même ce qui génère une boucle de dépendance. L'expérience célèbre de Sir Arthur Eddington avait pour but de confirmer la théorie de la relativité générale d'Einstein. L'expérience consistait à mesurer la déviation que la lumière d'une étoile subit à cause de la gravitation exercée par la masse du Soleil. Cette déviation étant une des prévisions découlant de la théorie d'Einstein. Or ce qu'Harry Collins montre c'est que l'interprétation des observations (des photos de l'étoile) dépendait de la théorie d'Einstein.

> *« L'observation et la prédiction étaient liées en un cercle de confirmation mutuelle et non pas indépendantes l'une de l'autre, comme on devrait s'y attendre selon l'idée conventionnelle d'une vérification expérimentale [...] la réalité était donc qu'on était d'accord pour s'accorder »* [20]

Il apparait donc aux yeux du philosophe et du sociologue que le processus classique de comparaison prédiction–observation n'est pas aussi fiable/innocent/sûr qu'il n'y parait.

# 5 Apologie de l'irrationalité

Revenons maintenant aux combats qui opposent les paradigmes de Thomas Kuhn. Ce dernier se rend compte que la confrontation des paradigmes se joue sur un terrain qui n'est que partiellement rationnel. Les opinions des scientifiques, leurs croyances et leurs visions du monde influent sur le fonctionnement de la science qui prend une dimension spéculative souvent doublée de questionnements métaphysiques. Pour Kuhn

> *« les adeptes de paradigmes concurrents ne s'entendent jamais complètement [...] leur discussion est inévitablement un dialogue de sourds. Chacun peut espérer convertir l'autre à sa conception de la science et de ses problèmes, aucun ne peut espérer prouver son point de vue. La concurrence entre paradigmes n'est pas le genre de bataille qui puisse se gagner avec des preuves »* [21].

[20] Harry Collins, *Tout ce que vous devriez savoir sur la science*, Points p. 69.
[21] Thomas Kuhn, *La structure des révolutions scientifiques*, Flammarion p. 204.

> *« Les deux groupes voient des choses différentes quand ils regardent dans la même direction à partir du même point »*[22]

> *« pour comprendre comment se font les révolutions scientifiques, il nous faudra étudier [...] les techniques* ***de persuasion*** *par discussion qui jouent un rôle au sein de ces groupes assez particuliers qui constituent le monde des sciences »*[23]

Ça y est le mot est prononcé : la lutte entre paradigmes se gagne par la persuasion plutôt que par la démonstration. Mais Kuhn rechigne à aller au bout de cette idée. Il souhaite se démarquer de tout relativisme potentiellement hostile à la science *« [dire] que le changement de paradigme ne saurait se justifier par des preuves, ce n'est pas prétendre qu'aucun argument n'a de valeur et qu'on ne peut persuader les scientifiques de changer d'avis »*[24]. Car il ne voit que trop bien que pousser ce germe relativiste dans ses retranchements reviendrait à nier l'existence de la science comme autre chose qu'un mythe ou une religion. Ce saut là, c'est Feyerabend qui ose le franchir.

> *« On conclura qu'il est souhaitable de laisser les inclinations aller à l'encontre de la raison dans n'importe quelles circonstances car la science peut en tirer profit [...]. Un conflit entre la raison et les pré-conditions du progrès est possible, il nous indique comment il peut naître et il nous force à conclure que nos chances de progrès peuvent se voir détruites par notre désir d'être rationnel ».*[25]

> *« Une nouvelle période dans l'histoire des sciences commence par* un retour en arrière *qui nous ramène... à un stade antérieur, où la théorie était plus vague et avait un contenu empirique moindre. Ce retour en arrière n'est pas un accident. Il a une fonction définie ; il est essentiel, si nous voulons dépasser le statu quo [...] Retour en arrière vraiment essentiel – mais comment pouvons nous persuader les autres de nous suivre ? Comment pouvons nous les attirer loin d'un système déjà défini, sophistiqué et fort de son succès empirique, pour les amener à transférer leur intérêt sur une hypothèse incomplète et absurde ? [...] Comment pouvons nous les convaincre que le succès du statu quo n'est qu'apparent, et sera reconnu comme tel dans 500 ans ou plus, alors qu'il n'y a aucun argument de notre coté ? Il est clair que l'attachement aux idées nouvelles devra être provoqué par d'autres moyens que des arguments. Par des moyens irrationnels tel que la propagande, l'émotion, les hypothèses ad hoc, et l'appel à des préjugés de toutes sortes. Nous avons besoin de ces*

[22]Thomas Kuhn, *La structure des révolutions scientifiques*, Flammarion p. 207.
[23]Thomas Kuhn, *La structure des révolutions scientifiques*, Flammarion p. 136.
[24]Thomas Kuhn, *La structure des révolutions scientifiques*, Flammarion p. 207.
[25]Paul Feyerabend, *Contre la méthode*, Points p. 170.

*« moyens irrationnels » pour soutenir ce qui n'est qu'une foi aveugle –jusqu'à ce que nous ayons trouvé les sciences auxiliaires, les faits, les arguments qui transforment cette foi en « connaissance » solide. »* [26]

*« Pour nous résumer : ou que nous regardions, quels que soit les exemples que nous considérions, nous voyons que les principes du rationalisme critique (prendre les réfutations aux sérieux ; élargir le contenu ; éviter les hypothèses ad hoc ; « être honnête » —si cela signifie quelque chose— et ainsi de suite) et a fortiori les principes de l'empirisme logique (être précis ; fonder les théories sur des mesures ; éviter les idées vagues et peu sûres ; et ainsi de suite) rendent mal compte du déroulement passé de la science, et sont susceptibles de retarder les progrès futurs. Ils rendent mal compte d'une science bien plus « molle » et « irrationnelle » que son image méthodologique. Et ils sont susceptible de la retarder car les efforts pour rendre la science plus « rationnelle » et plus précise sont voués à la supprimer, nous l'avons vu [...]. Ces « déviations », ces « erreurs », sont les conditions du progrès. Elles permettent à la connaissance de survivre dans le monde complexe et difficile que nous habitons, elles nous permettent de rester des agents libres et heureux. Sans « chaos », point de savoir. Sans une destitution fréquente de la raison, pas de progrès. Les idées qui aujourd'hui forment la base même de la science n'existe que parce qu'il y a eu des préjugés, de la vanité, de la passion ; parce que cela se sont opposés à la raison ; et parce qu'on les a laissés agir à leur guise. Nous devons conclure que même à l'intérieur de la science, la raison ne peut pas et ne doit pas avoir une portée universelle, qu'elle doit souvent être outrepassée, ou éliminée, en faveur d'autre instances. Il n'y a pas une règle qui reste valide dans toutes les circonstances et pas une seule instance à laquelle on puisse toujours faire appel ».* [27]

*« [L'anarchiste épistémologique] ses buts restent stables, ou changent à la suite d'une discussion, ou par ennui, ou après une expérience de conversion, ou pour impressionner une maitresse – et ainsi de suite. Si on lui donne un but, il peut essayer de l'atteindre avec l'aide de groupes organisés, ou tout seul, il peut se servir de la raison, de l'émotion, du ridicule, d'un « engagement profond » et de tout autre moyen inventé par les hommes pour obtenir le meilleur de leurs semblables... »* [28]

Je glisse cependant ici un contre-argument d'Imre Lakatos contre Paul Feyerabend mais retranscris par ce dernier.

[26] Paul Feyerabend, *Contre la méthode*, Points p. 165–166.
[27] Paul Feyerabend, *Contre la méthode*, Points p. 195–196.
[28] Paul Feyerabend, *Contre la méthode*, Points p. p209.

*« Si aucune argumentation ne peut détruire l'anarchisme épistémologique, on peut cependant démontrer son absurdité : Ou donc est l'anarchiste qui, par pur esprit de contradiction, sort d'un bâtiment par la fenêtre du 50ème étage au lieu de prendre l'ascenseur ? »*[29]

Edgar Morin lui aussi n'hésite pas à détruire la ligne de démarcation/le mur entre science et philosophie.

*« Il n'y a pas une frontière nette entre science et philosophie. Bien entendu, à leurs pôles, à leur noyau central, elles sont tout à fait différentes puisque le caractère original de la science, c'est surtout l'obsession vérificatrice, falsificationniste, et que l'obsession centrale de la philosophie, c'est la réflexivité et le retour du sujet sur soi même. Mais quand même, il faut dire que dans l'activité scientifique, il y a beaucoup de réflexivité [...] et la philosophie —par nature— ne méprise pas en soi la vérification ou l'expérimentation [...]. Certes aujourd'hui certains néo-technocrates de la science la [l'activité spéculative et philosophique qui naît de la science] méprisent comme séniles spéculations et discussion de fond. Mais ils veilleront »* [30]

*« Ces même esprits [scientifiques bureaucratisés, formés selon les modes classiques de pensée] ne veulent pas se rendre compte que, contrairement au dogme classique de disjonction entre science et philosophie, les sciences avancées de ce siècle ont toutes rencontré et ré-éclairé les problèmes philosophiques fondamentaux (qu'est ce que le monde ? La nature ? La vie ? L'homme ? La réalité ?) Et que les plus grands scientifiques depuis Einstein, Bohr et Heisenberg, se sont mués en philosophes sauvages. »*[31]

Nietzsche, dans son ouvrage *Le gai savoir* publié en 1882 c'est-à-dire avec une centaine d'années d'avance sur Feyerabend, entrevoie l'absence de démarcation nette entre science et croyance.

*« Préludes à la science : croyez-vous donc que les sciences seraient apparues et se seraient développées si elles n'avaient pas été précédées par les magiciens, alchimistes, astrologues et sorcières, eux qui, par leurs promesses et leurs espérances illusoires, ont dû commencer par créer la soif, la faim et le gout pour les puissances cachées et défendues ?[...] peut-être l'ensemble de la religion apparaitra t'elle, un jour lointain, comme un exercice et un prélude, de la même manière que se présentent à nous,*

[29]Paul Feyerabend, *Contre la méthode*, Points p. 244.
[30]Edgar Morin, *Science avec conscience*, Points p. 57.
[31]Edgar Morin, *Science avec conscience*, Points p. 9.

*ici, des préludes et des exercices préparatoires à la science qui ne furent absolument pas pratiqués ni ressentis comme tels »*[32]

# 6 Reconstruire la place de l'observateur

A plusieurs reprises, des amis m'ont dit avoir été heurtés/choqués en lisant un ou deux paragraphes de cette thèse, par l'écriture à la première personne c'est-à-dire la présence de nombreux « je ». Selon eux, la présence du « je » relève de la prétention et je devrais préférer l'utilisation du « nous » ou de la forme passive. Le problème que j'ai avec le « nous » c'est que je ne suis pas prêt à faire assumer à toute mon équipe les bêtises, fautes et spéculations présentes dans cette thèse. Le problème que j'ai avec la forme passive, c'est que je ne suis pas prêt à laisser entendre/croire que c'est la science/dieu/la vérité qui sort de ma bouche. J'utilise le « je » pour une raison simple : c'est *la vérité* et je ne vois pas l'intérêt de mettre une couche de maquillage.

Ainsi la science qui est enseignée et pratiquée dans les laboratoires élimine, à tort, l'expérimentateur qui ne devient qu'un observateur indépendant de la nature.

*« L'élimination par principe du sujet observateur, expérimentateur et concepteur de l'observation, de l'expérimentation, de la conception, a éliminé l'acteur réel, le scientifique, homme, intellectuel, universitaire, esprit inclus dans une culture, une société, une histoire. On peut même dire que le retour réflexif du sujet scientifique sur lui-même est scientifiquement impossible, parce que la méthode scientifique s'est fondée sur la disjonction du sujet et de l'objet, et que le sujet a été renvoyé à la philosophie et à la morale. »* [33]

### 6.0.5 Apprendre à s'auto-interroger

Il faut absolument changer cet état de fait :

*« Le sujet doit se réintroduire de façon autocritique et autoréflexive dans sa connaissance des objets »*[34]. Edgar Morin fait même une proposition pour la recherche *« les scientifique [doivent être] aptes à s'auto-interroger ; c'est-à-dire que la science soit apte à s'auto-étudier »* [35]

[32]Nietzsche, *Le gai savoir*, Flammarion p. 245.
[33]Edgar Morin, *Science avec conscience*, Points p. 20.
[34]Edgar Morin, *Science avec conscience*, Points p. 29.
[35]Edgar Morin, *Science avec conscience*, Points p. 31.

### 6.0.6 L'importance cruciale de la partie immergée de l'iceberg scientifique

> *« Il y a des croyances non expérimentales et non testables derrière les théories, c'est-à-dire dans l'esprit des savants et des chercheurs ; il y a des impuretés pas seulement métaphysiques mais sans aucun doute sociologique et culturelles »*[36] *« Il y a donc ce problème de l'imagination scientifique qu'on élimine parce qu'on ne saurait l'expliquer scientifiquement mais qui pourtant est à la source des explications scientifiques ».*[37] *« themata [Holton], programme de recherche [Lakatos], paradigme [Kuhn] etc., voici des notions qui introduisent dans la scientificité des éléments apparemment impurs, mais je le répète nécessaire [au] fonctionnement [des sciences] »*[38]

Nietzsche quant à lui s'amuse à chercher les liens qui existent entre le caractère d'un chercheur et la profession de ses parents (notez que nous sommes en 1882 c'est-à-dire une vingtaine d'années avant l'entrée en scène de Freud[39] et 80 ans avant l'arrivée de Feyerabend, Morin...) :

> *« Si l'on a un peu exercé son œil à reconnaître et à prendre sur le fait, à la lecture d'un livre savant, d'un traité scientifique, l'idiosyncrasie [comportement particulier] intellectuelle du savant —tout savant en possède une—, on apercevra presque toujours derrière elle la « préhistoire » du savant, sa famille, en particulier ses genres de profession et de métier. Là où s'exprime le sentiment « voila qui est prouvé, j'en ai fini avec cela », c'est ordinairement l'ancêtre dans le sang et l'instinct de savant qui, à partir de* ***sa perspective****, approuve le « travail accompli » –****la croyance à la preuve n'est qu'un symptôme de ce que l'on a considéré de tout temps, au sein d'une lignée de travailleurs acharnées, comme « du bon travail »****.[...] Le fils d'un avocat devra être aussi avocat en tant que chercheur : il voudra en premier lieu faire triompher sa cause, en second lieu, peut être avoir raison. Les enfants des maîtres d'école se reconnaissent à l'assurance naïve avec laquelle, en tant que savants, ils considèrent leur cause comme déjà démontrée pour peu qu'ils l'aient exposée avec enthousiasme et chaleur : c'est qu'ils sont fondamentalement habitués à ce qu'on les croie – chez leurs pères, cela faisait parti du « métier » ! »* [40]

[36] Edgar Morin, *Science avec conscience*, Points p. 43.
[37] Edgar Morin, *Science avec conscience*, Points p. 46.
[38] Edgar Morin, *Science avec conscience*, Points p. 45.
[39] Prélude à la découverte de l'inconscient ?
[40] Nietzsche, *Le gai savoir*, Flammarion p. 294–295.




**Figure XVI.3** – La démarcation entre science et non-science

# 7 Critique, consensus et intersubjectivité

Selon Morin, *« la science se définit non pas par la certitude mais au contraire peut-être par l'incertitude »*[41]. Dés lors, la volonté de chercher une démarcation nette et claire entre science et non science est une idée erronée, une idée de maniaque. C'est la seule erreur de Popper qui, avec son « critère de falsifiabilité » comme principe de démarcation, s'est accolé sans la vouloir une étiquette de positiviste. Pour Morin, la démarcation entre science et non science vient de l'activité qu'entretient la communauté scientifique. Plus précisément, *« l'objectivité est le résultat d'un processus critique développé par une communauté/société scientifique jouant un jeu dont elle assume pleinement la règle. [L'objectivité] est produite par* ***un consensus*** *»*. Comme dit Popper, *« l'objectivité des énoncés scientifiques réside dans le fait qu'ils puissent être inter-subjectivement soumis à des tests »*. Comme le souligne Morin, *« il y a un lien inouï entre intersubjectivité et objectivité »*[42] . On comprend mieux maintenant pourquoi il est si absurde de vouloir éliminer le rôle du sujet observateur.

> *« L'objectivité scientifique n'exclut pas l'esprit humain, le sujet individuel, la culture, la société : elle les mobilise. Et l'objectivité se fonde sur la mobilisation ininterrompue de l'esprit humain, de ses puissances constructives, de ferments socioculturels et de ferments historiques [...] si l'on veut chercher quelque chose de crucial, c'est la libre communication ; c'est la critique intersubjective qui est le point crucial et nodal de l'idée d'objectivité »*[43]

Ainsi selon Morin, la critique et le consensus sont responsables, via le fonctionnement de la communauté, de l'émergence de l'objectivité. Ce n'est pas un critère net entre science et non science mais c'est un critère quand même. J'essaierai plus tard de formuler une critique épistémologique contre cette idée que la critique scientifique puisse être à la base de l'émergence de l'objectivité ou de la verité.

# 8 La déconstruction du dogme hypothèse– expérience– résultat

Dans les paragraphes précédents, nous avons vu que la méthode qui consiste à confronter théorie et expérience est moins pure et donc plus suspecte qu'elle n'y parait. Cela est principalement dû au fait qu'il peut exister des relations de dépendance entre la théorie et l'expérience

[41] Edgar Morin cite Whitehead.
[42] Morin, *Science avec conscience*, Points p. 41.
[43] Edgar Morin, *Science avec conscience*, Points p. 56.

et parce que les faits peuvent être impurs et véhiculer des idéologies difficilement décelables. Il existe un autre dogme en science. L'idée que la méthode scientifique commence par la formulation d'une question ou d'une hypothèse, suivie par une expérience fournissant des résultats qui seront interprétés afin d'évaluer si l'on obtient une réponse à la question initiale.

Selon le philosophe des sciences Hans-Jörg Rheinberger, la recherche expérimentale en biologie ne commence pas forcément par une théorie, une question clairement formulée. En effet, les systèmes expérimentaux précédent parfois les problèmes qu'ils aident éventuellement à résoudre. Car ces systèmes expérimentaux n'aident pas seulement à répondre aux questions, ils aident à la générer. Ainsi contrairement à ce qu'avance Popper, le développement des disciplines en biologie n'est pas principalement guidé par les idées et les théories : Le processus de recherche serait plutôt « guidé de derrière » par les capacités intrinsèques d'un système expérimental que les scientifiques tentent d'explorer. Ainsi un biologiste choisit souvent un organisme modèle qui doit être vu comme un système, une sorte de boite noire. Ce simple choix orientera le type de découvertes qu'il fera et donc au final sa carrière tout entière. La boite noire idéale fournit une dose suffisante de mystère, de difficulté tout en offrant une dose suffisante de reproductibilité pour ne pas décourager l'expérimentateur. La boite noire agit comme un générateur de surprise pour l'expérimentateur qui y explore l'espace des possibles sans forcément qu'une question soit formulée explicitement ou qu'une théorie spécifique soit testée. Parfois, c'est seulement *a posteriori*, que le biologiste détermine à quelle question répond un résultat donné.
Ainsi pour nous microbiologistes, *E. coli* est à la fois :

⇒ L'objet d'étude (on cherche à comprendre comment la bactérie fonctionne)
⇒ L'instrument/l'outil (qu'il faut apprendre à « utiliser » et « sentir »)
⇒ Le générateur de question (qui nous oriente vers les découvertes).

Il me semble que beaucoup de scientifiques pensent que :

⇒ il faut formuler une question bien précise avant chaque expérience
⇒ il faut être en mesure de prédire les résultats (positifs ou négatifs) auxquels on s'attend
⇒ il faut si possible contrôler (fixer) tous les paramètres du système pour maximiser la reproductibilité (température, PH etc...)
⇒ il faut privilégier le qualitatif au quantitatif : mieux vaut faire peu d'expériences mais bien (en notant tous les détails parfaitement sur le cahier de manipulation) qu'en faire beaucoup et mal (en négligeant les détails).

Bien que je valide et applique aussi cette méthode, je pense que celle-ci ne suffit pas :

⇒ il faut aussi faire des expériences sans réelle question « juste pour voir/sentir ce qui se passe ». Il s'agit d'expériences parfois loufoques, souvent stupides, toujours irréfléchies,

basées sur une intuition de dernière minute ou pour tester ce que fait tel produit par exemple. Ce qu'il faut c'est : n'avoir aucune idée du résultat. Pour être honnête, ce type d'expériences (j'en ai fait beaucoup des stupides !) n'a jamais rien donné. Autrement dit, la nature ne m'a jamais accordé une seule fois « serendipity » c'est-à-dire la découverte au hasard qui rend heureux. Pourtant ces expériences sont utiles : elles entretiennent le moral, la curiosité mais surtout elles aident l'expérimentateur à « sentir » sa boite noire, à acquérir une image interne du fonctionnement de la bactérie. Ces sensations, qui ne se laissent pas toujours retranscrire sur papier aisément, aideront à la formulation des questions futures.

⇒ il faut aussi faire des expériences sans contrôler/fixer tous les paramètres. Cela permet petit à petit (au fur et à mesure des problèmes) de connaitre les paramètres qui comptent réellement. En apprenant à isoler puis contrôler uniquement les paramètres qui comptent, on économise l'énergie (en général du temps de manipulation fastidieux) à chercher à contrôler ceux qui ne comptent pas. De plus, on a une meilleure idée de la robustesse de ces résultats aux variations.

Pour synthétiser, je pense qu'il ne faut pas hésiter à faire beaucoup d'expériences « imparfaites » pour « explorer » son système. Cela permet de se heurter à l'incompréhension des résultats, à l'échec avec une fréquence très élevée. Car le plus souvent, la nature ne dévoile ses mystères qu'à ceux qui se sont « plantés » 1000 fois. En effet, les erreurs sont nécessaires. Elles sont le reflet en miroir des erreurs de l'algorithme heuristique de recherche de la connaissance (le schéma de la montagne vu précédemment).

# Chapitre XVII

# Analyse du système de publication

*« Dans l'œuvre de la science seulement on peut aimer ce qu'on détruit, on peut continuer le passé en le niant, on peut vénérer son maître en le contredisant. »* [1]

Bachelard, *La formation de l'esprit scientifique*

## 1 Limiter les informations demandées aux rewievers

Les personnes en premières « lignes » de la critique scientifique sont les « reviewers ». Un « reviewer » est un scientifique chargé d'évaluer la publication d'un autre scientifique à la demande d'un éditeur d'un journal scientifique. Le processus garantit l'anonymat du fameux « reviewer » qui peut alors émetre un rapport « objectif » sans risquer de s'attirer l'antipathie de l'auteur de la publication évaluée. Ce processus de « revue » est souvent considéré comme « le moins pire » des systèmes car cela permet aux scientifiques d'être évalués par leurs pairs (d'autres scientifiques du même domaine) ce qui vaut mieux qu'une évaluation menée par un homme politique, par un financeur ou par le peuple c'est-à-dire des gens qui ne connaissent rien à la science. Cependant beaucoup de gens souffrent des rapports négatifs des « reviewers ». En effet, le processus étant anonyme, il n'est pas toujours tendre. Les « reviewers » s'attaquent aux raisonnements de cause à effet, ils demandent des preuves supplémentaires (les fameux « contrôles »), ils peuvent tout simplement écrire qu'ils ne sont pas convaincus par les résultats ou que ces résultats ne sont pas vraiment intéressants, novateurs, originaux, significatifs etc... Bref, il n'est pas rare que les dents grincent coté auteur.

Peut-on imaginer un autre système. Je le crois. Car je pense que ce processus de revue est

[1] Bachelard, *La formation de l'esprit scientifique*, Bibliothèque des textes philosophiques p.301.

fondamentalement subjectif[2].

Analysons tout d'abord les travaux d'Harry Collins, un sociologue des sciences britannique. Il a mené une expérience simple[3] : il a demandé à 3 chercheurs différents de commenter anonymement les travaux d'un quatrième et ceux à 4 reprises (soit 16 chercheurs en tout) :

Commentaires sur l'expérience réalisée par le scientifique/le laboratoire[4] (1) :

⇒ Chercheur (2) : [...] c'est pourquoi le matériel qu'ils ont [dans ce laboratoire], malgré toute sa complexité, possède certaine caractéristiques qui, s'ils détectent quelque chose, leur permet d'être plus crédibles. [...] On voit qu'ils l'on vraiment pensé [...]

⇒ Chercheur (3) : Ils espèrent obtenir une grande sensibilité, mais, franchement, je ne les crois pas. Il y a des méthodes plus subtiles que la force brute [...]

⇒ Chercheur (4) : Je pense que [les gens de ce laboratoire] ont perdu la tête.

Commentaires sur l'expérience réalisée par le scientifique/le laboratoire (5) :

⇒ Chercheur (6) : [...] il travaille dans un tout petit labo, mais j'ai examiné ses résultats, et il est certain qu'il en a d'intéressants.

⇒ Chercheur (7) : Je ne suis pas vraiment impressionné par ses capacités et j'ai plus de doutes sur tout ce qui vient de lui que sur ce que font les autres

⇒ Chercheur (8) : Cette expérience, c'est de la merde !

Commentaires sur l'expérience réalisée par le scientifique/le laboratoire (9) :

⇒ Chercheur (10) : Ses résultats sont très impressionnants. C'est du travail de pro et ça a l'air d'être du sûr.

⇒ Chercheur (11) : Si vous voulez mon avis, et pourtant nous sommes bons amis, sa sensibilité est [faible]...et il n'a absolument aucune chance [de détecter les ondes gravitationnelles]

⇒ Chercheur (12) : Si vous faites comme [ce gars] et que vous vous contentez de donner les chiffres à des filles pour qu'elles les traitent, c'est sûr que vous ne savez plus où vous en êtes. Vous ne savez même pas si ces filles ne sont pas en train de discuter avec leur petit ami en même temps

Commentaires sur l'expérience réalisée par le scientifique/le laboratoire (13) :

[2]Je précise que je n'ai jamais eu aucun problème avec le système de publication/ « les reviewers ». Correction : un an plus tard, maintenant j'ai eu « un problème »avec un reviewer qui refuse un de mes articles. Heureusement, ce chapitre était déjà écrit depuis longtemps, on ne peut donc pas m'accuser de l'avoir écrit pour cause d'aigreur prononcée ! ⌣̈

[3]Harry Collins, *Tout ce que vous devriez savoir sur la science*, p. 136.

[4]Il n'apparait pas très clairement dans l'ouvrage la distinction entre l'homme et son laboratoire.

⇒ Chercheur (14) : L'expérience de [ce gars] est très intéressante et on ne devrait pas l'éliminer sous le seul prétexte que l'équipe est incapable de la répéter
⇒ Chercheur (15) : L'affaire de [ce gars] ne m'impressionne pas beaucoup
⇒ Chercheur (16) : Et puis il y a l'affaire de [ce gars]. C'est une imposture complète.

Cette étude est relative à une controverse sur les ondes gravitationnelles. Il n'est donc pas forcement évident que l'on puisse transposer ce schéma à toute la science et en particulier au processus de « peer-review ». Cependant, constatez la grande différence d'opinions ainsi que les pensées « brutes » permises par l'anonymat. Même si cette étude n'a aucune valeur statistique et qu'elle n'a pas été menée pour répondre à la question de l'objectivité du processus de « peer-review », je soupçonne que la subjectivité soit la règle.

Pourtant je ne pense pas que l'on puisse incriminer « les reviewers » sur la base de leur mauvaise foi ou de leur incompétence. Au contraire, je suis intimement persuadé qu'ils ont l'impression d'être parfaitement honnêtes et de faire des critiques constructives et légitimes. Que le « reviewer » soit un « post-doc » débutant ou « un ponte du domaine » : aucun des deux ne peut atteindre le niveau d'objectivité requis car leurs commentaires seront toujours beaucoup trop influencés par « leur science », leurs paradigmes, leurs rails de pensée, leurs visions du monde aussi respectables et profondes qu'elles puissent être. La preuve en est le nombre incalculable de fois où les rapports des « rewievers » ne concordent pas, ni sur le message global (positif ou négatif) ni sur les exigences (refaire une expérience, un contrôle, revoir le texte, la discussion). Selon moi, le problème c'est que la publication contient trop d'information ce qui casse leurs facultés d'analyse objective. Pour étayer cette thèse, je vais me baser sur les travaux de Malcolm Gladwell, un journaliste américain qui commente des résultats de sociologie pour le grand public. Il a écrit un livre nommé *Blink* (clignement en français). Sa thèse est simple « the power of thinking without thinking » : dans certains cas, on prend des meilleures décisions si on a accès à moins d'information.

Il fournit un exemple qui m'a marqué. Dans les urgences d'un hôpital américain dans les années 90, les médecins ont dû faire face à une recrudescence de suspicion d'infarctus. Manquant de place, les médecins dirigeaient les patients qu'ils considéraient comme « les plus à risque » vers une unité de soin intensive et ceux « moins à risque » vers des unités standards là où la surveillance est moindre. La question essentielle était donc « comment détecter correctement les personnes à risque ? ». Les médecins, pour évaluer ce risque, demandaient leurs antécédents aux patients en axant sur les facteurs à risque : « *Fumez-vous, Quel âge avez-vous ? Avez-vous déjà eu un infarctus ? Avez-vous du cholestérol ? Souffrez-vous d'hypertension artérielle ? De diabète ? etc...* » puis ils auscultaient le patient *« a-t-il du liquide dans les poumons ? »,*

*« Quel est l'allure de électrocardiogramme ? »* Une fois la consultation terminée, ils décidaient, en fonction de tous ces éléments où placer le patient : unité intensive ou standard. Au fil des mois, les statistiques ont fini par parler : à la grande surprise générale, les choix des médecins étaient quasi-indiscernables du hasard. 50% des infarctus avaient lieu dans la mauvaise unité, l'unité standard moins surveillée. Cela signifiait que les médecins étaient incapables d'évaluer correctement le facteur de risque après l'interrogatoire, l'auscultation et l'examen ECG (noter que l'interprétation des ECG n'est pas triviale : un ECG standard peut cacher un gros infarctus et réciproquement). Le problème fut rapidement démasqué : les médecins avaient trop d'informations. La synthèse de ces informations empêchait de prendre la bonne décision. En effet, comment décider qui va en soin intensif ? Un vieux qui fume et se plaint de violentes douleurs à la poitrine avec un ECG normal ? Ou bien un jeune, en surpoids avec du diabète qui a du liquide dans les poumons et un ECG non standard ?

La solution pour prendre des décisions plus fiables fut de limiter les questions posées lors de la consultation en se focalisant sur uniquement 4 éléments :

⇒ la douleur ressentie ressemble t'elle à celle d'un angor instable ?
⇒ y a-t-il du liquide dans les poumons ?
⇒ la pression sanguine systolique est-elle inférieure à 100 (10) ?
⇒ quelle est l'allure de l'ECG ?

En limitant l'interrogatoire et l'auscultation à ces 4 éléments « les 4 facteurs de risque dans l'urgence », les médecins orientaient le bon patient (à risque) dans la bonne unité (intensive) dans plus de 95% de cas. La raison est simple : le fait de fumer, d'avoir du cholestérol, d'être en surpoids, d'avoir déjà eu un infarctus ou d'être âgé sont des facteurs de risque sur le long terme mais n'apportent aucune information dans l'urgence. Au contraire, ils ajoutent une couche de brouillard dans l'esprit du médecin qui l'empêche de prendre la bonne décision.

Revenons maintenant à notre « reviewer » car je pense qu'on est dans un cas similaire. Sa décision, ses commentaires sont subjectifs car obscurcis par une publication longue et difficile sans moyen de comparaison. Je pense donc qu'il faut *limiter les informations que l'on demande à un « reviewer »*. A l'inverse, il faut lui donner des moyens de comparaison c'est-à-dire lui fournir plusieurs publications. Son rôle est simple : il doit les classer dans l'ordre selon les critères de son choix. Chaque publication est soumise à plusieurs « reviewers » qui ont chacun plusieurs publications à classer (par exemple 4) mais le jeu des publications est différent pour chaque « reviewer ». Il y a plus de publications à lire et donc plus de travail mais pas de rapport à écrire et donc pas d'expériences ou modifications à exiger. Et comme vous le remarquerez, il n'y a pas de critique. Les éditeurs récupèrent les classements, effectuent la somme nécessaire

pour établir le classement final ce qui leur indique quel article publier et quel article rejeter. Ainsi, le « reviewer » donne un sentiment global non justifié sous forme d'un classement. Pour éviter des biais évidents, il n'a ni accès aux noms des auteurs, ni à la liste de références. Le processus est en double aveugle. Le « reviewer » peut sans doute se faire une idée du laboratoire d'où vient la publication mais cela n'est pas un point crucial. L'important c'est de maximiser l'objectivité du processus de « peer-review » et je pense que cette méthode, pas parfaite, est néanmoins plus objective que la méthode des rapports. Remarquez qu'à aucun moment, il n'est fait mention du respect de la méthode scientifique : s'il le souhaite, le « reviewer » classe les publications selon son cœur (ce qu'il a aimé) et le plus étonnant c'est que cela sera, selon moi, plus objectif, plus rationnel et donc plus scientifique.

## 2 Contre l'anonymat

La proposition ci-dessus ne représente qu'une proposition superficielle à portée limitée. En réalité, je pense qu'il faut aller beaucoup plus loin et repenser toutes les fondations de notre système de publication. Sachez, par exemple, que je suis en réalité farouchement opposé à l'anonymat des reviewers (que je défendais pourtant au-dessus). En effet, cet anonymat est une négation affreuse de l'homme scientifique dans toute sa pluralité. Etudions la question de plus prés.

L'anonymat des reviewers est considéré comme une nécessité censée permettre aux reviewers d'être plus « objectifs », de gommer tout facteur humain ou moral, d'empêcher tout conflit d'intérêt. Le reviewer doit pouvoir formuler un avis sans avoir à évaluer les risques pour sa carrière, d'une part, et sans engager sa « moralité » d'autre part. Prenons deux exemples :

⇒ Dire à un ponte, sans la protection de l'anonymat, que son travail est « insignifiant / médiocre », c'est prendre le risque de se faire un ennemi influent.

⇒ Dire à un « chercheur novice », sans la protection de l'anonymat, que son travail est « insignifiant/médiocre », c'est prendre le risque d'être jugé faible/rigide sur le plan psychologique ou pire d'être jugé sur le plan moral.

Car ne nous leurrons pas, dire à quelqu'un que son travail est médiocre revient à dire à quelqu'un qu'il est médiocre, et ce, contrairement à une croyance répandue. Sans l'anonymat, c'est-à-dire dans la vie de tous les jours, les chercheurs « critiquant » sont plus mielleux.

La question essentielle est donc la suivante : tout ce miel ajouté en absence d'anonymat représente-il une intrusion néfaste de la morale dans la sphère scientifique ? Ou bien, au contraire, ce miel doit-il être considéré par delà bien et mal comme un outil, une méthode

scientifique ?

Vous l'aurez compris, je crois à la seconde hypothèse. Les bons professeurs introduisent souvent du doute dans leur jugement en rajoutant le classique « il me semble » : par exemple, « Il me semble que vous devriez vérifier ce résultat avec telle ou telle méthode ». Ils sont donc moins affirmatifs, ils donnent des avis/des conseils à portée consultative car ils ne veulent surtout pas « prétendre détenir la vérité ». Ils savent que leur vérité est elle-même relative et préfèrent donc questionner leur étudiant sur un mode dialectique. Entendons-nous bien : la préférence pour le questionnement n'est pas un acte moral ou « gentil » mais l'unique manière de former un chercheur compétent. A l'inverse, un professeur « mauvais » excelle à « casser » les vérités (les rêves, les fantasmes ?) jugées inferieures de leurs étudiants, jugés eux également inferieurs.

Ainsi, si on est convaincu du bien fondé des « il me semble », pourquoi vouloir absolument les faire disparaître ? Car c'est bien la conséquence de l'anonymat : les « il me semble » disparaissent. Une fois anonyme, le reviewer devient prétentieux (au sens de prétention à la vérité) : il « sait » alors qu'avant il « lui semblait ». On passe de l'interrogation à l'affirmation. Du doute au savoir. Or nous verrons que le phénomène interrogatif (et non la critique) est au fondement de la science. En supprimant l'interrogation, on ne maximise pas la production scientifique.

En pratique, les reviewers ont un pouvoir important sur la décision finale de publication. Or je crois que leur avis devrait être uniquement consultatif. Pourquoi ? A cause du problème du but. Le reviewer ne pourra jamais savoir si son but et celui de l'auteur convergent. Or cette convergence est nécessaire.

Prenons un exemple. Un auteur souhaite publier rapidement un article sur des travaux antérieurs qu'il juge « moyens » pour pouvoir plancher rapidement sur un nouveau problème qu'il croit « plus important ». Le reviewer, lui, identifie une expérience qui permettrait de rendre les résultats de l'article plus robustes. Il impose donc naturellement cette expérience à l'auteur.

Qui a raison, qui a tort ? Le problème réside dans la divergence des buts. Le « reviewer » veut des résultats plus robustes. L'auteur veut pouvoir rapidement travailler sur ce qu'il croit « vraiment important ». Qui maximise le plus la production scientifique ? Impossible à dire. Cependant, une chose est sûre : forcer quelqu'un à faire quelque chose qu'il ne souhaite pas faire n'est jamais optimal. Car l'auteur risque fort de bâcler ou de tricher. Pire : il y aura génération de ressentiment et atteinte potentielle de la motivation. Tout cela parce que la perspective du reviewer ne lui permet pas d'identifier le but recherché par l'auteur.

Le problème c'est que la communauté scientifique ne considère pas qu'il faille minimiser le ressentiment ou la frustration (les forces réactives diraient Gilles Deleuze). Au contraire, elle n'hésiterait pas une seconde à générer du ressentiment si cela lui permettait de préserver à tout

prix « l'examen » qui est sous-jacent à l'acte de publier. En effet, il semblerait qu'aujourd'hui les publications servent avant tout à évaluer et classer les scientifiques. Comme si le fait de pouvoir générer la gaussienne des bons et des mauvais était LA variable importante maximisant le progrès scientifique. Je reviendrai plus tard en détail sur « l'examen », devenu subrepticement une véritable méthode scientifique.

## 3 Faire tomber le mur aristocratique

En réalité, je pense qu'il faut aller beaucoup plus loin. La science est souvent considérée comme fonctionnant sur des bases aristocratiques plutôt que démocratiques. Prenons un exemple : imaginons deux scientifiques qui proposent chacun une solution différente à un théorème mathématique extrêmement complexe. Ces deux scientifiques n'arrivent pas à se mettre d'accord et décident de demander l'avis du peuple par referendum. Vous vous rendez bien compte que le choix du peuple sera bien plus corrélé au charisme des scientifiques qu'à la verité mathématique. Ainsi la démocratie n'est pas l'outil adéquat pour faire avancer la science : la science fonctionne sur une base aristocratique et il n'y a pas de mal à cela.

Le problème c'est que ce système aristocratique a fait émerger une caste, une société qui se juge supérieure par bien des aspects vis-à-vis des autres croyances. Cela a crée une coupure/un mur entre la société (et ses croyances diverses et variées) et la communauté scientifique. Cette coupure est créée par la critique et le scepticisme. Elle est toujours justifiée de la même manière : il faut absolument se protéger des croyances et du marketing (astrologies, sectes, pilules miracle anti-cancer, publicités pour le dentifrice) qui ont une fâcheuse tendance à utiliser la science de manière abusive pour doper leurs ventes et/ou faire du prosélytisme. Pourtant, en créant un filtre critique, on n'empêche pas uniquement les loups de rentrer dans la bergerie mais aussi toute les bonnes volontés qui n'attendent qu'une chose c'est d'être initiées et de participer activement à la science. Un professeur de secondaire qui voudrait publier une idée, une théorie, une découverte sera bloqué non pas parce que son idée, sa théorie sera moins bonne en moyenne que la moyenne des publications qui sortent des laboratoires. Non : il sera bloqué parce qu'il ne connait pas les codes de l'aristocratie scientifique. Un « reviewer » (qu'il soit post-doc ou ponte) a besoin de reconnaitre l'empreinte de sa communauté (ne serait-ce que l'affiliation à un laboratoire qui est toujours indiquée en haut de la publication), les stigmates des codes auxquels il est habitué (structure type : introduction, résultats, discussion, matériel et méthode), la présence des références de ses pairs (démontrant l'appartenance à la communauté). Si ces codes sont absents, le « reviewer » aura énormément de mal à juger le fond. Il

rejettera la publication à cause du non-respect de règles que suivent plus ou moins implicitement sa communauté. Et cela en tout subjectivité. Faites le test, amis sociologues, écrivez ou faites écrire une publication *de bon niveau scientifique* puis débarrassez là de tous les codes de la communauté scientifique. Supprimez toute affiliation à un laboratoire. Supprimez toute référence issue de la communauté scientifique. Réintroduisez l'observateur en insérant beaucoup de « je », faites appel au champ lexical de la culture, de la société, de la métaphysique, de la philosophie, parlez de votre sentiment, de votre ressenti. Puis soumettez l'article et regardez s'il passe le filtre de la critique.

Puis faites exactement l'inverse : écrivez *un article totalement vide de sens* et rajoutez-y tous les codes de la communauté scientifique : article signé par « un ponte » citant d'autres pontes, par exemple, puis soumettez à nouveau. Cela revient à imiter le physicien Alan Sokal et son article–canular *transgresser les frontières : à travers une herméneutique transformative de la gravité quantique*. Cependant, dans notre cas, il ne doit pas s'agir d'un canular mais d'un test sociologique qui devrait avoir, si possible, une valeur statistique.

Etudiez les résultats de ces tests puis dites-moi si mon soupçon de subjectivité est justifié. Si le mur aristocratique que je décris existe réellement. Si tel est le cas, je pense qu'il nous faut tirer des conclusions sur notre système scientifique. Il faut avoir le courage de briser ce mur qui n'a pas de justification. Ce mur qui ne révèle que notre peur inconsciente d'être mélangé à l'extérieur. Notre peur inconsciente que la société se rende compte que la science n'est pas si différente des autres croyances. Et par conséquent, que nous, scientifiques, sommes beaucoup plus proches des croyants que ce que nous imaginons.

Etudions maintenant l'exemple d'un système érigé sans mur : Wikipedia. Il s'agit d'une encyclopédie d'un niveau exceptionnel jamais égalé dans l'histoire autant en termes de nombre d'articles que de consultations d'articles. Sa réussite est due, selon moi, à un double pari étrange :

⇒ ouvrir l'écriture et la modification des articles à l'ensemble des internautes

⇒ mélanger savoir « encyclopédique classique » (un article sur la mécanique quantique par exemple) et savoir « sociétal populaire » (un article sur Madonna par exemple).

Certains critiquent Wikipedia. Les arguments sont : il faut être très prudent car n'importe qui peut y écrire n'importe quoi. En effet, il y a parfois des atteintes (publicité ou prosélytisme caché) au « principe de neutralité ». On enseigne donc aux élèves (et aussi aux grands élèves) à garder une part d'esprit critique vis-à-vis de l'information qu'ils y puisent. Ce qui est une très bonne chose. Ceci dit, déposez-vous votre esprit critique sur le pallier quand vous lisez un article scientifique sur Pubmed ou un article de l'encyclopédie Britannica ?

Prenons un article de science au hasard : par exemple l'article sur la cellule pour les biolo-

| | Wikipedia | Communauté scientifique |
|---|---|---|
| Lecture des articles | gratuite pour toute l'humanité. | payante pour le citoyen (de l'ordre de 35 euro par articles) mais gratuite pour les chercheurs de l'académique (dans le ressenti j'entends car, bien sûr, les instituts payent les abonnements). |
| Langues représentées | + de 250 langues. | principalement l'anglais (devenu l'équivalent de ce qu'était le latin il y a quelques siècles). |
| Ecriture des articles | gratuite et extrêmement simple pour toute l'humanité. Publication en quelques minutes une fois l'article écrit. | payante la plupart du temps (milliers d'euro), processus long et très complexe (des mois, des années). En général, les auteurs ont ou vont avoir un doctorat. Ils appartiennent donc à l'aristocratie scientifique. Les stigmates de cette appartenance doivent être présents dans la publication (nom de son laboratoire, citation d'autres aristocrates etc...). |
| Correction des erreurs | gratuite et extrêmement simple. Les erreurs sont acceptées car statistiquement rapidement corrigées. | réservée à l'auteur. Processus d'erratum peu utilisé. L'erreur est mal vue car elle est en contradiction avec le principe d'excellence. Toute publication devrait tendre vers 100% de vérité. |
| Reviewers | Tout lecteur peut améliorer l'article. | Quelques pairs (scientifiques du domaine) en général 2 ou 3 (en biologie). |
| Règles à respecter | Neutralité de point de vue, Respect des règles de savoir-vivre. | Respect de la méthode scientifique. Le problème c'est, qu'en l'état actuel des connaissances, cette dernière n'a pas encore de statut/définition clair. Dés lors, on baigne en plein subjectivisme. Il en résulte de multiples sous-règles paradigmatiques trop contraignantes et plus ou moins implicites. |
| Vision de la nature humaine | La nature humaine est statistiquement bonne. | La nature humaine est mauvaise. Elle cherche toujours à tricher. Il faut placer des verrous, des contrôles et des filtres pour obtenir l'excellence. |
| Pari | mélanger tous les types de savoir (société, culture, science etc...). Il n'y a pas de sous domaines. | Ne surtout pas mélanger les savoirs car on ne mélange pas les torchons et les serviettes. La science doit être maintenue à l'écart de la culture et de la société. |
| But | Transmettre le savoir au plus grand nombre. | Evaluer, Identifier, filtrer l'excellent. |

gistes et l'article sur le problème P=NP pour les mathématiciens. Au regard de vos connaissances, pensez-vous que le loup s'est immiscé dans la bergerie ? Que des fautes ou des fraudes imputables à « l'attaque des dogmatiques » se retrouvent dans ces articles ? Bien sûr que non. Or pensez-vous vraiment qu'avant la naissance de Wikipedia, la communauté scientifique aurait pu premièrement avoir l'idée et deuxièmement accepter de confier la rédaction de plusieurs centaines de milliers d'articles scientifiques à quelqu'un d'autre qu'à elle-même ? Surtout pas ! Vous n'y pensez pas, l'entrée des dogmatiques !

Wikipedia n'a pas été fondée par un scientifique mais par un financier Jimmy Wales. Ainsi, selon moi, la communauté scientifique n'est pas à l'origine de Wikipedia. Elle ne peut pas afficher Wikipedia sur son tableau de chasse. Elle ne peut pas afficher sur son tableau de chasse ce qui correspond à la plus grande propagation, diffusion du savoir que l'humanité ait jamais connu. Pourquoi ? Parce qu'elle fait le pari inverse de Wikipedia :

⇒ La communauté scientifique aristocratique : L'homme est mauvais et vil par nature c'est pourquoi il est absolument indispensable de critiquer, d'évaluer et de filtrer pour empêcher le loup de rentrer dans la bergerie. L'existence de la caste scientifique est une nécessité. Fermons nous sur nous même.

⇒ Wikipedia : Les contributeurs Wikipedia (sous-ensemble des hommes) sont bons statistiquement. Faisons leur confiance pour créer une pyramide en minimisant le nombre de *règles venant d'en haut* et en laissant les contributeurs se corriger les uns les autres non pas via un processus critique mais dialectique : *Faisons grandir une encyclopédie ensemble en confrontant nos points de vue. Ouvrons-nous au monde en ouvrant les yeux au monde.*

> *« On entendra toujours les défenseurs de l'Internet « civilisé »se plaindre, car la liberté de faire et de dire inspire la méfiance. Ceux-là préféreront la censure au désordre, n'admettront jamais la valeur pédagogique de l'erreur ou de la mise en danger du savoir établi. Ils ne croient pas à l'éducation mais au dressage. »*[5]

Je l'ai dit : je pense que la communauté scientifique, à laquelle j'appartiens, doit opérer une transformation sur elle-même, une révolution : elle doit se décloisonner. Cela n'a jamais été possible à cause de l'absence d'existence de l'expérimentateur prônée par la méthode scientifique. L'expérimentateur ne pouvant avoir une réflexivité scientifique sur lui-même, c'est en fait toute la communauté qui ne peut avoir cette réflexivité sur elle-même. Il est grand temps d'octroyer le droit au scientifique de s'étudier lui-même. Il est grand temps d'octroyer le droit à la communauté scientifique de s'étudier elle-même. Si cette dernière réussit cette métamorphose très profonde et unique au cours de l'histoire des sciences, elle se rendra compte de la nécessité de

[5]Citation de Jean-Noël Lafargue

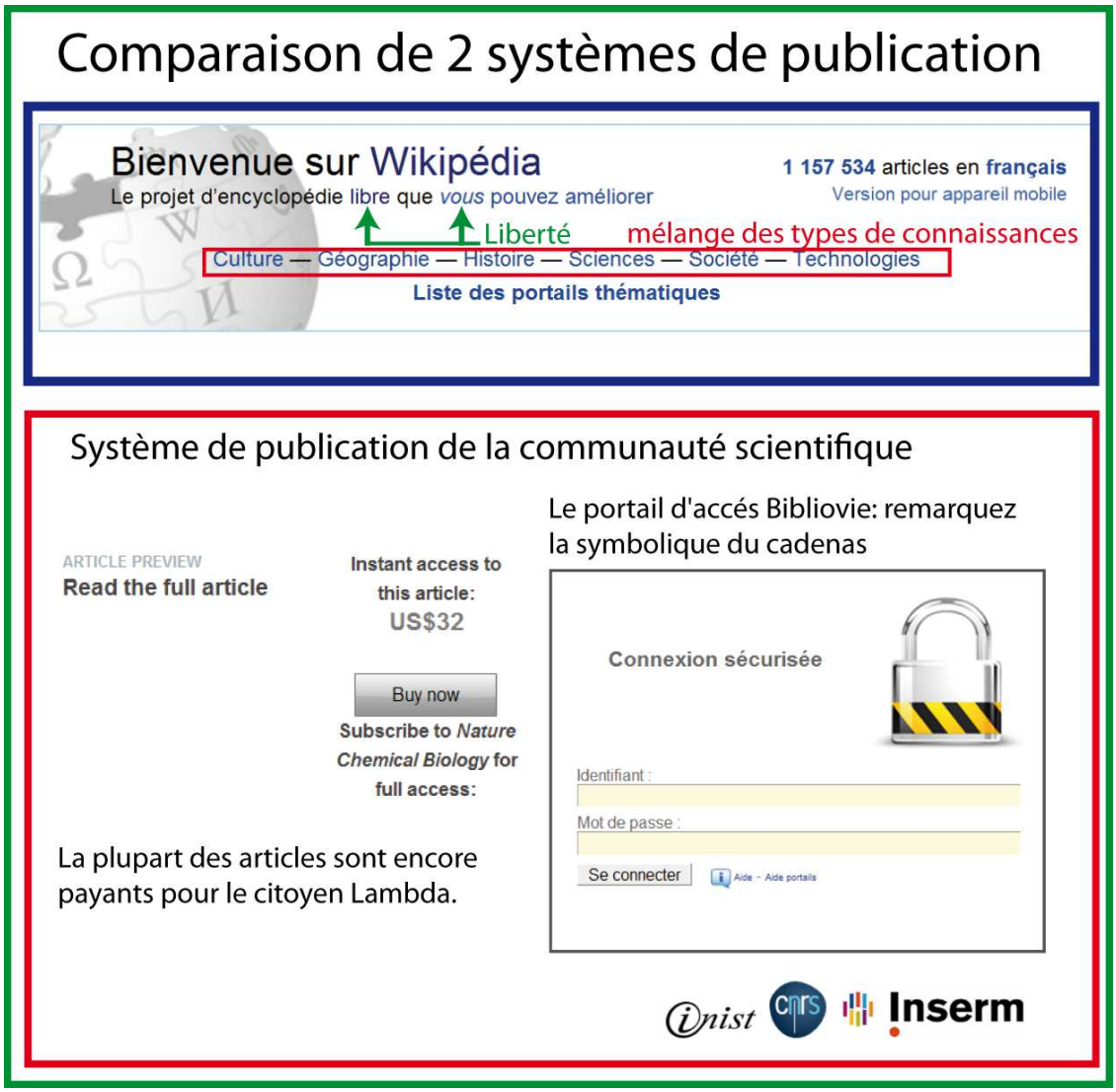



**Figure XVII.1** – L'encyclopédie Wikipedia est libre et ouverte au monde. Le système de publication de la communauté scientifique reste encore trop cloisonné, trop fermé. Je précise que les articles encyclopédiques et les articles de recherche ne sont pas vraiment comparables. D'autre part, il y a des raisons économiques expliquant en partie la fermeture de notre système de publication. Cependant, je pense qu'il faut, petit à petit, ouvrir notre système de publication au monde et ne surtout pas le restreindre à la communauté scientifique.

gommer son complexe de supériorité vis-à-vis des autres modes de pensée. Elle doit s'immerger avec eux et cela ne lui fera pas perdre sa spécificité. Au contraire.

La question c'est comment mettre en place cette métamorphose ?

Je pense qu'il n'y aura jamais assez de publications, d'idées et messages publiés et donc plus il y en aura mieux ce sera y compris si « la matrice » de cette ensemble de publication est relativement « creuse » en qualité. Beaucoup de gens, au contraire, pensent qu'il y a trop de

publications scientifiques. Par conséquent, ils préconisent de renforcer les filtres pour épurer. Leurs règles c'est « moins mais mieux ». Il me semble qu'ils raisonnent selon la mauvaise règle de l'exhaustivité. Le trop grand nombre d'articles les effraie : ils ont peur de ne pas tout contrôler, tout savoir. Ils raisonnent selon des vieux schémas issus de la science du 19$^{\text{ème}}$ siècle où l'homme pouvait encore être un « gentilhomme » capable de suivre et de comprendre l'ensemble des domaines scientifiques et culturels. Mais une métamorphose est en marche dans la science du 21$^{\text{ème}}$ siècle. Celle-ci renie la sacro sainte règle de l'exhaustivité comme seule manière de mener sa pensée. Il nous faudra apprendre à faire naviguer notre pensée dans un réseau gigantesque d'information. Trouver une information le plus vite possible. Fuir l'établissement à un endroit, fuir le stable mais au contraire voyager, pour endurcir sa pensée. Il est étonnant de voir comment tous les scientifiques vantent les mérites des voyages qui ouvrent les yeux : ils voyagent physiquement à travers le monde pour les congrès et autres rituels d'appartenance à la communauté. Pourtant, leurs pensées semblent parfois restées figées au même endroit pendant des décennies à cause de cette règle de l'exhaustivité. Apprendre à danser avec dextérité sur le réseau de la connaissance, voila ce qu'il faut apprendre à nos élèves.

> *« A enseigner, on apprend » [mais] « A toujours enseigner, on n'apprend rien »* [6]

> *« Là où un homme parvient à la conviction fondamentale qu'on doit lui commander, il devient croyant ; à l'inverse, on pourrait penser un plaisir et une force de l'autodétermination, une liberté de la volonté par lesquelles un esprit congédie toute croyance, tout désir de certitude, entrainé qu' il est à se tenir sur des cordes et des possibilités légères et même à danser jusque sur les bords des abîmes. Un tel esprit serait l'esprit libre par excellence »*[7]

> *« Et je ne connais rien que l'esprit d'un philosophe souhaite davantage qu'être un bon danseur. La danse est en effet son idéal, son art également, enfin aussi son unique piété »* [8]

Je rappelle que je ne me place pas à l'extérieur de la communauté scientifique lorsque je la critique. Je revendique totalement mon appartenance à cette communauté et donc je m'autocritique. C'est sans doute pour cette raison que je me permets d'être un peu dur. En secouant la communauté scientifique, je m'auto-secoue. Cependant, il faut savoir aussi reconnaitre les bonnes choses. La communauté scientifique a créé des bases de données gratuites, accessibles à tous et performantes (comme Pubmed par exemple). Un mouvement est initié pour rendre

[6] Schopenhauer, *Les fondements de la morale*, le livre de poche p. 70.
[7] Nietzsche, *Le gai savoir*, Flammarion p. 294.
[8] Nietzsche, *Le gai savoir*, Flammarion p. 350.

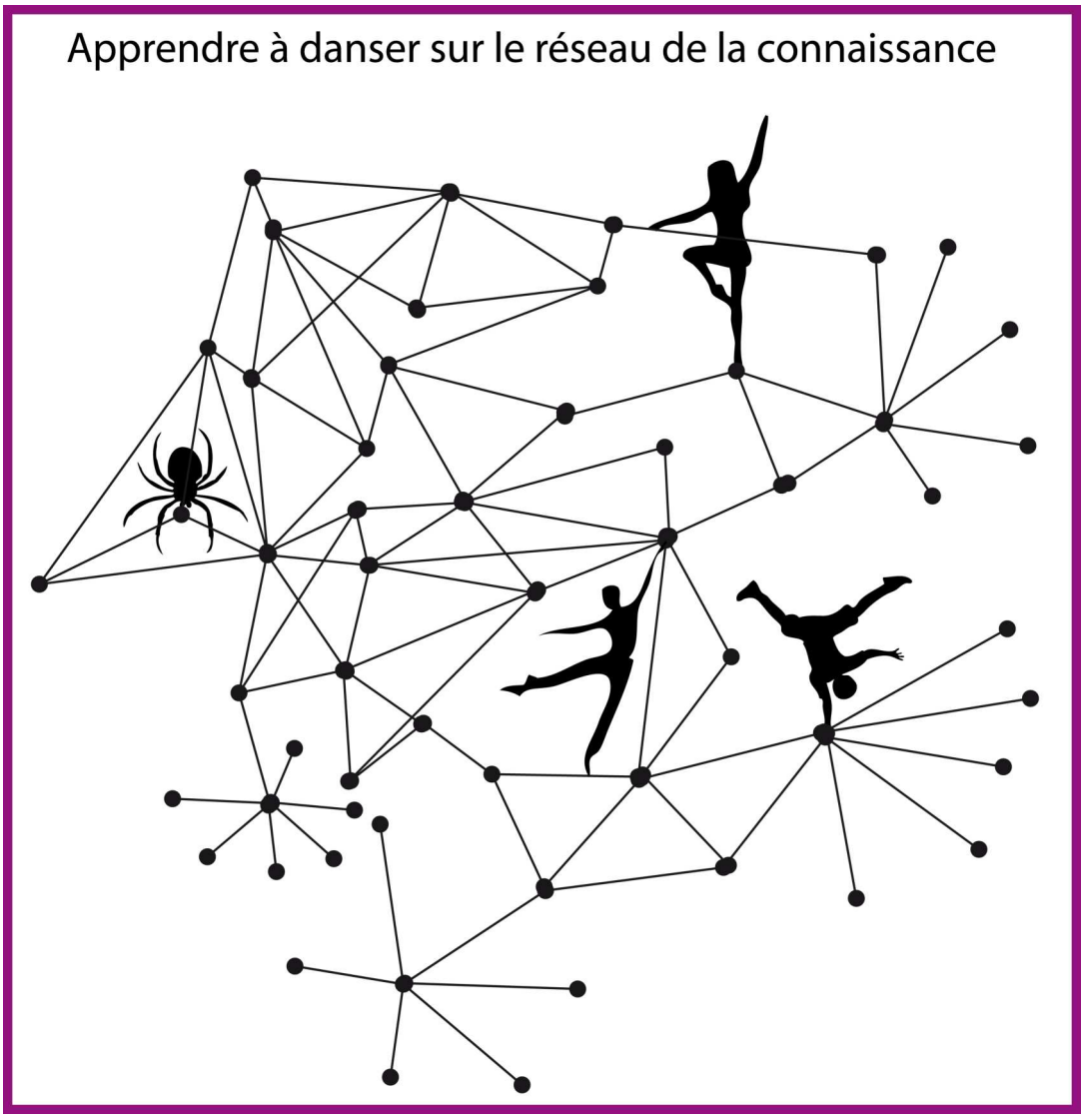



**Figure XVII.2** – Il faut apprendre à danser sur le réseau de la connaissance et ne pas faire comme l'araignée qui reste au chaud au centre de sa toile.

un maximum de publications gratuites à la lecture au grand public (PLOS) et des systèmes de pré-publications existent déja même si ils restent peu utilisés en biologie (arXiv). Le décloisonnement des sciences est peut être déjà en marche. Mais il faut le pousser encore beaucoup plus loin.

Il faut créer un système où toute personne qui a fait une découverte, a une idée originale, une théorie peut la publier facilement de manière à offrir une chance à cette idée de se propager. Ce système unique doit être le même pour tous. L'auteur doit communiquer son nom (ou un pseudonyme si il souhaite rester anonyme) et chaque article doit pouvoir citer et être cité. Les

liens/interactions de ce réseau ne sont pas uniquement les citations. Le lecteur a également la possibilité de cliquer sur le bouton « j'aime » (utilisé par facebook et google). Le poids des « citations » et des « j'aime » pourrait être pondéré par un système de « jus de lien » (utilisé par le google rank) de manière à rendre le système aristocratique. Ainsi, ceux qui ont beaucoup de liens qui pointent vers eux (ceux qui sont très « orientants ») créent des liens vers les autres qui ont beaucoup de poids. Ce poids sert à définir des scores : le score n'indique pas quelle est la meilleur idée mais quelle est l'idée la plus « orientante ». Il sert à définir l'ordre de présentation des résultats aux lecteurs qui effectuent des recherches avec des mots-clés (comme google). Par conséquent, les publications les plus « orientantes » se propageront plus vite. Ainsi, comme il n'y a pas de critique, toute idée est acceptée pour toujours sur le réseau de la connaissance. Aucune publication n'est rejetée/refusée. Toute publication est bonne à prendre. Tout est bon. Le seul paramètre qui varie c'est son accessibilité/sa visibilité. Ainsi il est possible qu'une idée excellente soit ignorée pendant des années puis soit finalement repérée par la personne à qui elle se destinait. La taille moyenne d'une publication sur la toile doit avoisiner 1 Mega-Octet. Avec 1 Tera-Octet (100 euro), on peut donc en stocker 1 millions. Vous voyez donc que le coup de stockage est minime par rapport à « l'or » que représente une idée. Si un humain juge une idée suffisamment intéressante pour faire l'effort de la mettre en forme sur papier dans le but de la transmettre à l'humanité gratuitement, il faut être fou pour la refuser. Le tout c'est de créer un système capable de classer ces idées et de les présenter au lecteur en fonction de leurs notoriétés.

Vous l'aurez compris, le système ne contient presque aucune règle. Des critères de forme à respecter existent néanmoins de manière à augmenter « l'expérience utilisateur » et ce sont les journaux scientifiques actuels qui pourraient se charger d'effectuer ces vérifications. Il faut donc les rémunérer. Mais comme le système ne nécessite ni revue par les pairs (peer-review), ni prise de décision par l'éditeur, ni impression, les coûts de publication devraient chuter. Leur financement pourrait être soit pris en charge par les états soit pris en charge par l'auteur lui-même. Le système peut aussi inclure un système de « post-review » (voir the faculty of 1000) géré par les éditeurs qui pourraient, à la demande d'auteurs et donc contre rémunération, demander à des « reviewers » de lire une publication et de rajouter ou pas le fameux « j'aime » puis de la citer ou pas dans leurs futurs travaux. Cela n'enlève rien à la publication, *ne détruit rien*, cela ne peut que *rajouter, construire*.

Bien sûr, étant donné que le système ne contient presque aucune règle, tout le monde peut publier l'idée qu'il souhaite. Comme sur Wikipedia, inexorablement, les articles scientifiques cohabiteront avec des croyances diverses, dogmes, groupuscules, Madonna, etc...

Mais ce qu'il faut voir/comprendre, c'est que chaque sous-réseau (chaque communauté

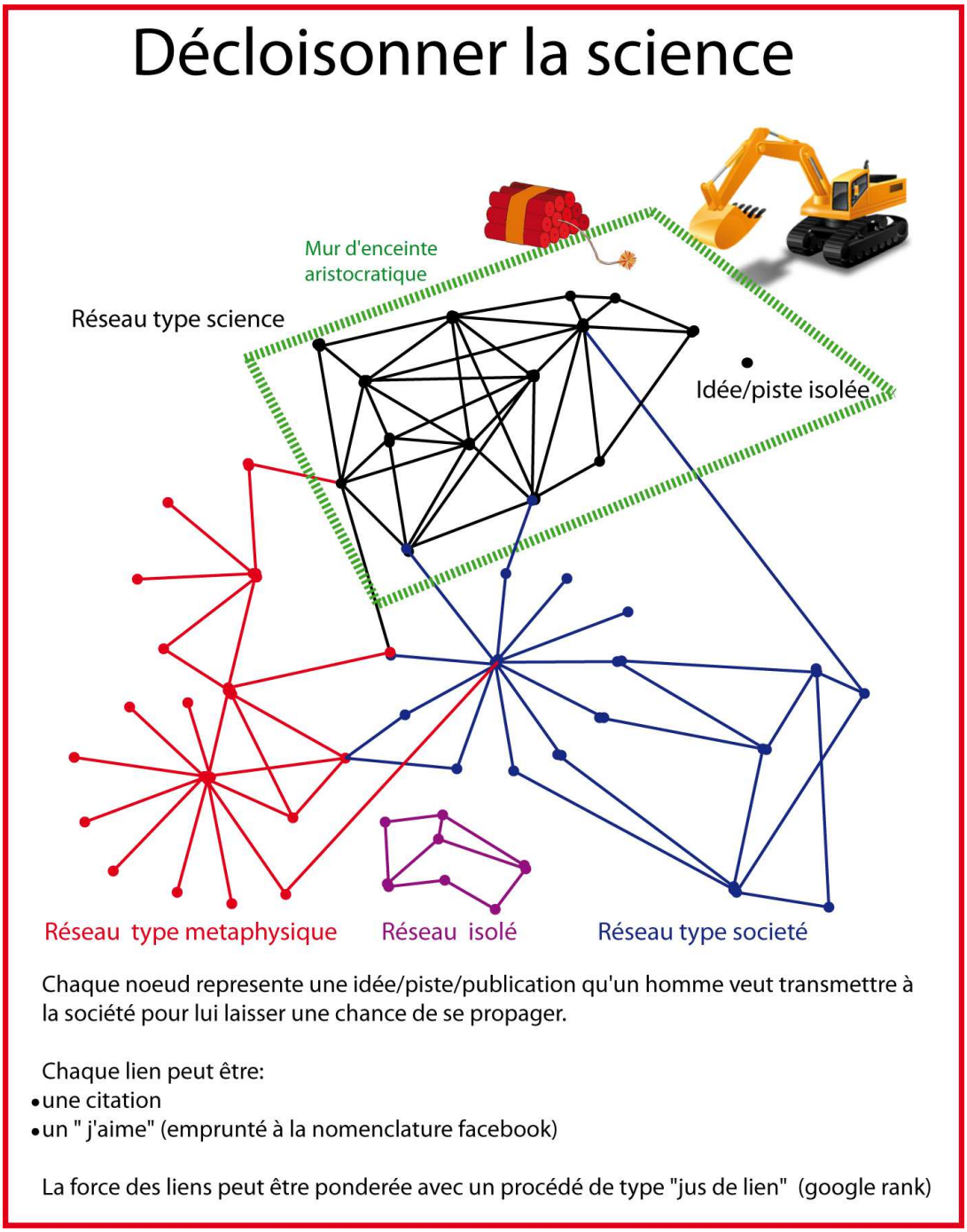



**Figure XVII.3** – Pourquoi la science a-t'elle peur de se mélanger aux autres formes/réseaux de la connaissance ? Nous devons faire sauter le mur d'enceinte aristocratique qui ne protège pas la science mais au contraire la renferme sur elle-même. Cela nous permettra d'identifier la topologie du réseau scientifique et de comprendre ses liens et ses différences avec les autres formes de connaissance.

d'idées) aura une topologie tout à fait particulière. Certaines informations, idées se propa-

geront très vites puis stopperont leurs progressions au bout de quelques jours (idées relatives à l'actualité, aux stars). Les idées relatives aux religions, aux dogmes auront une structure, une expansion différente. Le pari c'est que l'utilisateur sélectionnera le réseau sur lequel il veut se balader sur la base de critères qui représentent fidèlement les caractéristiques de sa communauté de pensée. La science, par exemple, se caractériserait peut être par la robustesse de ses idées au temps qui passe ou par un réseau très densément connecté par des citations créant ainsi une continuité entre les idées passées et futures. Ce qui fait que le scientifique choisirait comme critère de « surfer sur le réseau robuste au temps et densément connecté par les citations ». Il ne choisirait plus de surfer sur un réseau qui aurait subi et résisté à la critique comme c'est le cas actuellement. Il n'y a plus d'idées mauvaises rejetées, de scepticisme, de critiques. Il y a des idées qui se propagent et d'autres qui ne se propagent pas. C'est tout. Il n'y a pas de conclusion à tirer vis-à-vis de ces dernières. L'auteur a eu et conserve sa chance. Peut être que son idée n'a pas eu le succès qu'il escomptait mais il peut toujours se rassurer en se disant que peut être demain elle l'aura. Pas de jugement. Pas de critique. Juste la propagation des idées.

Ainsi en décloisonnant la science et en la mettant à la même portée, soumise aux mêmes règles que n'importe quelle croyance, on pourrait gagner en réflexivité en étudiant la topologie du réseau et en le comparant à la topologie des autres réseaux non scientifiques. Les possibles liens entre science et croyances pourraient aussi nous apparaitre plus clairement.

Enfin, j'insiste sur le fait que supprimer la démarcation actuelle entre science et croyance ne signifie pas que celles-ci sont identiques. Ma proposition n'a rien de relativiste. Elle représente au contraire une véritable méthode scientifique pour étudier, justement, ce qui sépare science et non science. Comprendre d'où provient l'asymétrie ressentie entre les énoncés et déterminer si cet asymétrie est réelle ou supposée passe, je le crois, par le décloisonnement du système de publication.

# Chapitre XVIII

## Falsification, discipline et critique

*« Beaucoup sont des obstinés pour ce qui est de la voie qu'ils ont prise, bien peu le sont quant au but »*[1]

Nietzsche

## 1 Le casse tête du chercheur

Quand ils présentent leurs résultats sous forme d'une publication, les chercheurs n'ont pas forcement conscience que le respect de la méthode scientifique n'est pas l'unique critère qu'ils optimisent. J'identifie 5 critères/variables constituant une fonction et c'est cette fonction que les scientifiques tentent de maximiser (bien sûr le poids relatif de chaque variable est propre à chaque chercheur) :

1. Le respect de la méthode scientifique. Ce critère est optimisé pour satisfaire la communauté scientifique dans son ensemble. Ce critère ne suffit pas car nous verrons que l'on peut publier une découverte à laquelle on ne croit pas tout en respectant la méthode scientifique.
2. La reproductibilité des résultats. Ce critère est optimisé pour satisfaire ses pairs. C'est-à-dire les scientifiques travaillant sur des sujets proches et donc en mesure de comprendre, d'évaluer et de reproduire notre travail. C'est un critère important mais il ne suffit pas car on peut publier un travail tout à fait original et reproductible mais représentant une avancée trop négligeable, un grain de sable trop microscopique pour faire réellement grandir la pyramide.
3. Le rayonnement de la publication. Les pouvoirs publics évaluent l'avancée des recherches en la comparant à celle des autres pays. Un des critères principal est le nombre de pu-

[1] Nietzsche, *Humain trop humain (I)*, Folio Essais p.294

blications (et leurs facteur d'impacts). De manière simpliste, pour le système politique, le rayonnement et les chiffres passent avant le respect de la méthode scientifique et la qualité.

4. La qualité de l'avancée scientifique. Ce critère est optimisé pour satisfaire le peuple qui nous emploie. Peu importe le respect de la méthode scientifique et le facteur d'impact. Ce qui compte c'est que la découverte change/influence leurs vies.
5. Les intérêts personnels et pseudo-personnels (l'individualisme). Ce critère vise à maximiser son intérêt ou celui d'un de ses poulains. En effet, les postes, les promotions et les financements dépendent en grande partie des publications. Il peut être tentant d'accélérer le processus de publication au détriment de la qualité. Ce processus peut également agir indirectement quand un chef aide un thésard ou un post-doc à publier « vite » pour lui permettre, à terme, d'obtenir plus facilement un poste ou un financement par exemple. On voit ici l'intrusion de la morale : aider une personne de son équipe à trouver un emploi peut primer sur le strict respect de la méthode scientifique.

Le problème c'est que ces variables peuvent être antagonistes. Dés lors, le chercheur est très souvent obligé d'opérer des arbitrages. Ces arbitrages sont des falsifications nécessaires.

## 2 La déconstruction du dogme des données inaltérables

> *« Je connais tout cela de trop prés peut être : cette respectable modération philosophique à quoi entraine une telle foi, ce stoïcisme de l'intellect qui finit par s'interdire le non aussi sévèrement que le oui, cette volonté de s'en tenir [aux faits], ce renoncement à toute interprétation, à tout ce qui consiste à faire violence, arranger, abréger, omettre, remplir, amplifier, fausser, et de façon générale à ce qui est le propre de toute interprétation. »*[2]

Je vais donc maintenant aborder un sujet sensible : la falsification des données.

Le système de publication, en biologie, fonctionne de la manière suivante. La publication est un document de quelques pages avec une introduction, les résultats, la discussion et « le matériel et méthode ». Les résultats incluent les figures qui contiennent les données expérimentales ainsi qu'un texte censé extraire et décrire l'information des figures de manière neutre. La discussion est le lieu, quant à elle, où le chercheur peut faire part de ses hypothèses ou de son interprétation.

[2]Nietzsche, *La généalogie de la morale*, Folio p. 181

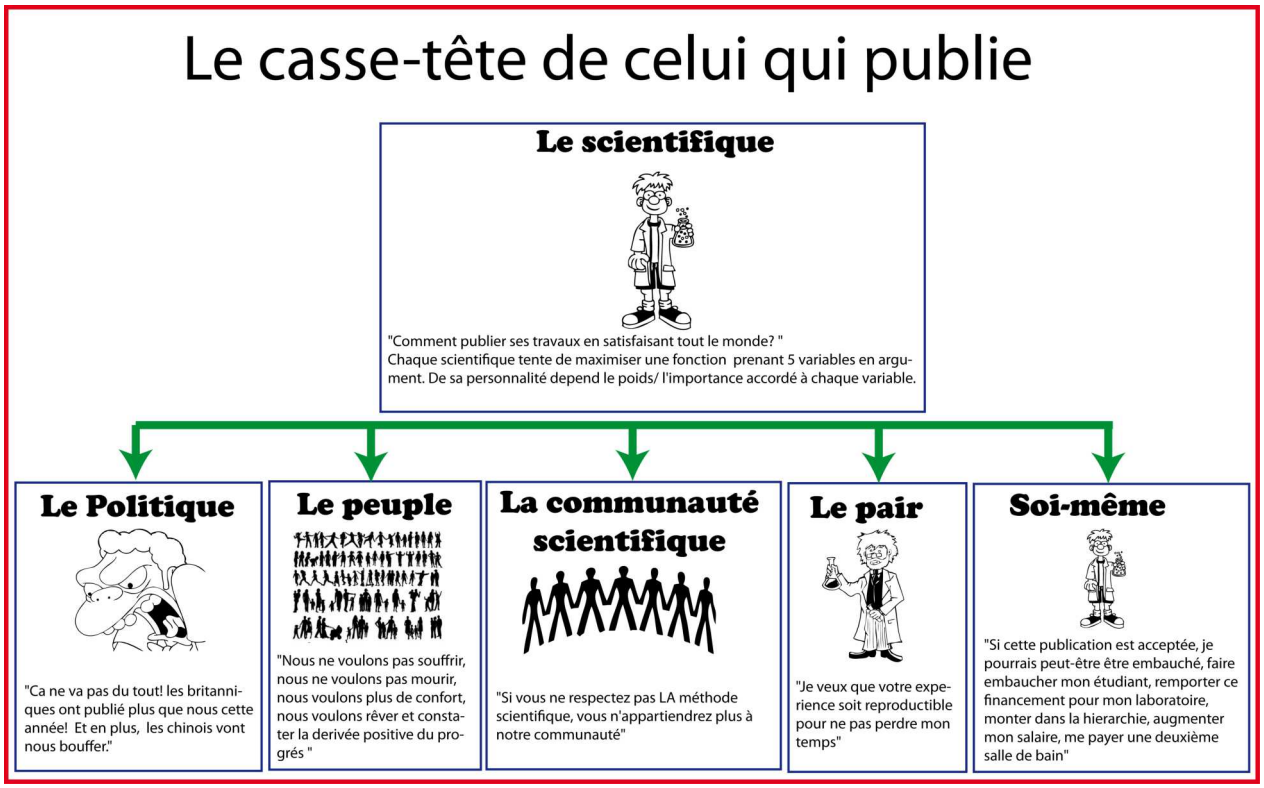



**Figure XVIII.1** – Le casse-tête du chercheur

Il y a donc un découpage entre une partie objective (figures qui contiennent les données + description des résultats) et la partie plus subjective qu'est la discussion.

Le dogme est de croire que l'expérimentateur est absent de la partie résultat. Officiellement, les données brutes sont toujours sacrées, pures : c'est la nature, la vérité, dieu qui a parlé via l'expérience et ce, y compris, si elles mettent en évidence des artefacts. Les données brutes sont les données brutes.

Pourtant, la figure illustrera toujours le message que le scientifique veut faire passer, c'est-à-dire qu'elle sera toujours en accord avec son interprétation favorite. En effet, les chercheurs ne publient que très rarement les résultats négatifs ou les résultats qu'ils n'arrivent pas à interpréter. Cela signifie qu'ils choisissent parmi un grand nombre d'expériences/de données celles qu'ils veulent montrer. Autrement dit, le processus est biaisé : la manière de procéder, pour une figure de publication, ce n'est pas « j'ai des données sacrées, voici mon interprétation » mais plutôt « au fil des expériences, j'ai acquis la certitude d'une interprétation, comment faut-il que je m'y prenne pour que ma figure l'illustre le mieux possible ? » Il arrive aussi parfois que les données brutes soient traitées/corrigées. Pour garder la bénédiction du temple positiviste, ces traitements doivent être justifiables. Il peut s'agir d'enlever des points « aberrants » (outliers),

d' « ajuster » (fitter) les données ou d'effectuer des traitements mathématiques ou statistiques. Selon moi, la méthode de traitement est souvent floue car une couche de brouillard (jargon technique ou absence pure et simple d'information sur le traitement) plus ou moins épaisse existe. Ainsi le processus de traitement de données peut très vite s'apparenter à une boite noire et il y a souvent autant de façons de traiter les données que de chercheurs.

Une figure (et les données qu'elles contiennent) n'est donc pas « sacrée », indépendante de l'expérimentateur. La figure est le reflet de ce que le chercheur a vu, ressenti ou cru voir. Elle est toujours en partie subjective car le chercheur montre toujours *ce qu'il veut montrer* et ce, qu'il s'agisse de données brutes ou de données traités. Et *ce qu'il veut montrer* correspond assez souvent à *ce à quoi il croit* (1) et *ce à quoi il croit* correspond assez souvent à *ce qui est « vrai » ou plutôt disons... ce qui se reproduit* (2) car les chercheurs, pour beaucoup, sont honnêtes (1) et compétents (2).

Ainsi la clairvoyance, la croyance, l'intuition, (totalement ignorées par le positivisme) joueraient un rôle très important dans la méthode scientifique. Retranscrire correctement avec des mots et des figures ce à quoi l'on croit est aussi un critère qu'il convient de maximiser.

Voici une analyse des travaux de Pasteur faite par le sociologue des sciences Harry Collins :

> *« Quoique à l'époque, il [Pasteur] ne connut pas l'origine de la contamination, il n'admettait pas que ces résultats soient considérés comme des présomptions en faveur de l'hypothèse de la génération spontané. Selon ces propres termes : « je ne publiais pas ces expériences ; les conséquences qu'il fallait en déduire étaient trop graves pour que je n'eusse pas la crainte de quelques causes d'erreurs cachées malgré le soin que j'avais mis à les rendre irréprochables ». En d'autre terme, Pasteur était si fermement opposé à la génération spontanée qu'il préférait croire à une erreur inconnue dans son travail plutôt que de publier ses résultats. Il définissait les expériences qui confirmaient la génération spontanée comme ratées et vice versa.[...] Avec le recul nous ne pouvons qu'applaudir à l'intuition de Pasteur. Bien entendu, il avait raison et avait suffisamment le courage de ses convictions pour refuser de s'en laisser détourner par ce qui, à première vue, était une indication expérimentale contraire. Mais c'était de la clairvoyance et non pas l'application neutre de la méthode scientifique »* [3]

Ici vous voyez que l'arbitrage/la falsification de Pasteur a consisté à cacher des résultats d'expériences allant à l'encontre de sa théorie.

J'identifie trois types de falsifications différentes :

[3] Harry Collins, *Tout ce que vous devriez savoir sur la science*, Points P. 119.

1. Modifier activement les données (changer les chiffres, modifier sa courbe avec une logiciel de dessin, modifier la légende...) (falsification active)
2. Omettre de présenter des données « à décharge » (falsification passive)
3. Publier des données « vraies » non falsifiées en livrant une interprétation plausible mais à laquelle on ne croit pas. (falsification ignorée par le positivisme)

Le premier point, dans l'inconscient collectif des chercheurs, est celui qui choque le plus. Celui qui est donc le plus répréhensible car le chercheur (le traitre !) touche « physiquement » au dogme des données inaltérables (dieu). Le deuxième point est toléré, « dépénalisé ». Le troisième point n'a pas d'existence dans l'épistémologie positiviste. En effet, la méthode scientifique type résultats–discussion n'impose pas que l'expérimentateur croie en ce qu'il écrit car l'expérimentateur n'est pas censé exister.

Premièrement, je pense que les falsifications actives ou passives doivent être considérées comme identiques vis-à-vis de la méthode scientifique alors que la première est souvent considérée comme plus grave que la seconde. Deuxièmement (et je sais que cela heurte alors je m'en excuse), selon moi, les falsifications actives ou passives sont non seulement souhaitables mais elles sont surtout nécessaires. Elles sont utilisées depuis toujours plus ou moins inconsciemment par les chercheurs. Vous voyez que mon discours est à la fois « normatif » (il faut falsifier !) et « descriptif » (la science existe et progresse parce qu'heureusement les chercheurs falsifient !). La falsification sans existence épistémologique est la seule qui me semble néfaste. Et encore, je n'en suis plus très sûr [4].

Ainsi, parce que la science est un système complexe, le chercheur falsifie, peut falsifier et doit falsifier. En particulier, il peut falsifier si cela lui permet de faire mieux correspondre sa figure (ses données) avec l'image mentale, la représentation qu'il se fait du phénomène qu'il souhaite montrer. La figure et son message doit être le miroir de sa pensée, c'est-à-dire ce à quoi il croit. Dés lors, je ne fais aucune concession sur *le désintéressement* qui doit animer le chercheur qui falsifie. La falsification peut être utilisée si elle cache une sincérité, une loyauté supérieure pour la vérité.

> *« « Une extrême loyauté à l'égard de tous » est pour lui [Nietzsche] non pas un dogme moral mais une condition tout à fait primaire, élémentaire et indispensable de l'existence « je péris que je suis dans un milieu impur ». L'absence de clarté, la malpropreté morale le dépriment et l'irritent. »*[5]

[4] J'ai changé d'avis depuis la rédaction de ce chapitre. Je ne crois plus que ce type de falsification soit forcement néfaste.

[5] Stephan Zweig, *Nietzsche*, Stock p.65

Enfin, le scientifique peut falsifier si cela ne change pas réellement *la capacité de ses pairs à reproduire ses résultats*. Vous allez me dire : comment cela est possible ? Parce que la science est un système complexe qui revêt de nombreuses facettes irrationnelles *a priori*.

Dés lors qu'on ne pense pas empêcher la reproductibilité par ses pairs, cela signifie que l'on juge que la falsification touchera des éléments insignifiants qui ne mettent pas en péril la qualité du message envisagé dans sa globalité. Evidement c'est un jugement subjectif, c'est pour cela que la méthode scientifique positiviste l'interdit. Il y a un risque d'introduction « d'omniscience » c'est-à-dire l'idée erronée d'un scientifique qui croit détenir la vérité. Pire : la falsification empêche les lecteurs ayant potentiellement plus de recul d'interpréter différemment les données « sacrées » puisque celles-ci sont falsifiées.

> *« Comme nous le savons maintenant, les expériences de Pasteur auraient pu — auraient dû— échouer de bien des façons. A notre humble avis, c'est ce qu'elles ont fait, mais Pasteur savait ce qui devait être considéré comme un résultat et ce qui devait l'être comme une erreur »* [6]

C'est dans le mot « savait » de la phrase ci-dessus qu'il y a introduction d'omniscience. Pasteur a fait un arbitrage et décidé d'accorder un poids supérieur à son intuition qu'aux données issues des expériences. Dans les faits, les données publiées ne sont pas sacrées : elles sont et doivent être malléables. Elles s'adaptent à l'intuition du chercheur —intuition qui provient de la synthèse cognitive de son savoir et des dizaines, centaines, milliers d'expériences qu'il a déjà faites. D'ailleurs, la méthode scientifique a été obligée de s'adapter en proposant de « fitter » les données et d'enlever « les outliers » en demandant tout de même la fameuse justification. Or le chercheur ne peut pas toujours justifier facilement et clairement son intuition ou sa synthèse cognitive. Il ne peut pas justifier *clairement* car ses résultats ne sont pas totalement clairs dans son esprit. Et ses résultats ne sont pas clairs dans son esprit parce qu'ils sont, par définition, à la frontière des connaissances : là où les certitudes, le flou et l'ignorance se côtoient de très prés.

Ainsi, il faut dédramatiser et dépénaliser la falsification c'est-à-dire être pragmatique et oser regarder la verité en face :

1. Car, en l'état actuel des connaissances, il n'existe pas de règles ou méthodes scientifiques clairement définies et immuables à partir desquelles il ne faudrait surtout pas dévier. La falsification n'a donc aucune raison rationnelle d'être rejetée.
2. Car la falsification a toujours existé. Les chercheurs traitent souvent les données de manière arbitraire, choisissent les données qui leurs sont le plus favorables, cachent par

[6] Harry Collins, *Tout ce que vous devriez savoir sur la science*, Points P. 124.

omission des données à « décharge », publient des données vraies avec un message auquel ils ne croient pas...

3. Car, la clarté du message est aussi un élément à maximiser. En effet, qui pourrait reprocher à quelqu'un de falsifier ses données dans le seul but de rendre plus clair son message de manière à laisser la possibilité à son lecteur de le comprendre, de le vérifier, de le critiquer ou de le dépasser ? Cette falsification pour la clarté ne doit pas être mal comprise : il ne s'agit absolument pas « d'exagérer » un phénomène donné car cela éloignerait du « miroir de la pensée » et donc de la capacité des pairs à reproduire.

4. Car le temps est un élément qu'on est obligé, de manière pragmatique, de prendre en compte. Les pairs et la communauté scientifique disent implicitement « prends tout le temps qui est nécessaire avant de publier ton travail pour avoir l'assurance qu'il est parfaitement reproductible ». A l'inverse, le peuple (1), les pouvoirs publics (2) et ses chefs/poulains (3) disent plus ou moins implicitement « dépêche-toi de publier pour augmenter notre confort (1), notre rayonnement (2), notre probabilité d'obtenir un financement /un poste, (3) ». On voit donc qu'il y a des intérêts divergents sur cette question du temps. L'absence de prise de position de la méthode scientifique sur cette question du temps est encore une fois liée au fait que l'expérimentateur n'est pas censé exister. Dés lors, le temps qu'il met à découvrir n'existe pas, ne compte pas. Selon moi, la question du temps est une question de bon sens : il faudrait pouvoir publier facilement ses résultats dans les semaines/mois qui suivent les expériences. Et c'est très rarement le cas : beaucoup de publications mettent des années avant d'être publiées ce qui affecte forcement le moral du scientifique. C'est mon $5^{\text{ème}}$ point.

5. Car le moral des scientifiques doit être préservé à tout prix. Or la raison du décalage temporel énorme entre publication et expérience est liée à l'oppression du système positiviste. le chercheur traine car il vit un conflit intérieur sur la question de la méthode scientifique. La difficulté déontologique que pose la question des arbitrages lui fait toujours reculer à plus tard la publication. Il y a toujours « une dernière expérience » qui l'épuise. Les années passent et inexorablement, le moral du scientifique est attaqué, malmené. Car c'est humain, l'esprit a besoin de « cocher » les travaux finis pour libérer son esprit et passer à la découverte, à la recherche suivante. Une bonne méthode scientifique aiderait « ses soldats » de la manière suivante : dés lors qu'un scientifique fait une découverte qu'il juge capable d'orienter ses pairs dans les bonnes directions, alors il est de son devoir de scientifique de publier sa découverte avec pour règle suprême de minimiser le temps qui le sépare de sa prochaine expérience/découverte y compris si cela nécessite de la falsi-

fication. C'est une règle « anti-épuisement ». Car beaucoup de chercheurs finissent par se détourner de la recherche vers d'autres centres d'intérêt qui les oppriment moins (administratif, éducatif, vulgarisation, philosophique, managérial et autres ; que je respecte totalement évidemment). Ce désintérêt progressif n'est pas uniquement lié à la difficulté de la recherche mais aussi au poids que représentent ces fameux arbitrages et les critiques potentielles qui y sont associées.

Ainsi ce qu'un positiviste appellera « falsification », je l'appelle en fait « arbitrage ». Cet arbitrage est une notion centrale qui compare/soupèse/arbitre/évalue *le risque d'introduction d'omniscience* (penser qu'on sait, alors qu'on ne sait pas) avec *le bénéfice qu'apportera la publication*. Les arbitrages représentent une des difficultés majeures des sciences : c'est un combat intérieur. Une voie crie « avance » et l'autre crie « attends encore un peu, fais encore une expérience pour être encore plus sûr ». Nietzsche sait bien la difficulté qu'il y a à finir *« Il faut plus de courage pour faire une fin qu'un vers nouveau : c'est ce que savent tous les médecins et tous les poètes »*[7]. Un expérimentaliste qui n'a jamais vécu se combat intérieur, qui ne voit pas ce que j'appelle « un arbitrage » ne fait pas la même science que moi. En effet, la méthode scientifique n'offre malheureusement pas la même quiétude d'esprit que celle offerte par l'application stricte d'une recette de cuisine. Le scientifique doit arbitrer. Et le pire c'est qu'il doit le faire en cachette. J'invite donc les sociologues des sciences à s'intéresser de plus prés à cette notion d'arbitrage. Et pour cela, ils devront fouiller là où personne ne veut les voir fouiller.

La falsification que je prône s'effectue en excluant tout intérêt personnel ou pseudo personnel. Cependant, Il existe aussi des falsifications « intéressées ».

> *« La meilleur entité qui puisse à elle seule détourner un scientifique moderne de faire ce que sa conscience scientifique lui dit de poursuivre, c'est encore le dollar »*[8]

Par exemple, en falsifiant, un jeune chercheur peut accélérer la publication de ses résultats ce qui facilitera son embauche. Faut-il reprocher alors à ce chercheur d'avoir triché pour obtenir un emploi stable lui permettant de fonder une famille dans la quiétude ? Et parmi les falsifications « intéressées », il y a celles qui ont un coût sociétal élevé. Par exemple, si un chercheur A publie un résultat qu'il sait « faux » et qu'un chercheur B passe plusieurs mois à tenter de reproduire ce résultat sans succès : A a fait perdre sciemment à B plusieurs mois ce qui a un coût : le salaire de B et les coûts divers de matériels et de fonctionnements. Si A est pris la main dans le sac, faut-il le sanctionner ? Et d'une manière générale faut-il augmenter les contrôles ?

[7] Nietzsche, *Ainsi parlait Zarathoustra/Troisième partie/Des vieilles et des nouvelles tables.*
[8] Paul Feyerabend, *Contre la méthode*, p. 53

> *« Il manque à l'homme supérieur un grand instrument d'éducation : le rire d'hommes supérieurs »* [9]

> *« L'augmentation de la sagesse peut se mesurer avec précision à la diminution du fiel* [c'est à dire l'amertume, le ressentiment] *»* [10]

Je laisse le soin aux juges de toute sorte d'occuper leur temps et d'animer leurs vies avec ce type de question risible. Car je n'ai pas un goût prononcé pour les autodafés. De toute manière, les volontés de puissance du tricheur et du juge sont toutes deux des volontés de puissance de l'homme sur l'homme (voir la fin du dernier chapitre d'épistémologie). Or la science est la volonté de puissance de l'homme sur la nature. Il s'agit d'une guerre. Si vous y combattez, vous n'avez que faire des tricheurs. Et si vous n'y combattez pas, alors vous n'êtes pas très différent du tricheur.

## 2.1 Le système positiviste est un système oppressant

Je m'apprête à critiquer violemment notre système institutionnel positiviste. Pourtant, je ne me revendique (et ne me ressens) ni comme un anarchiste ni comme un révolutionnaire : à l'inverse de Paul Feyerabend, je ne pense pas qu'il faille séparer la science et l'état ! Une partie de mon esprit est positiviste. Quand je fais une expérience, je pense en termes de « vérifiabilité » et de « reproductibilité ». Je m'apprête donc à critiquer une partie de mon esprit ! Accepter de fonctionner/penser avec des théories contradictoires dans son esprit n'est pas pour moi paradoxal. Au contraire, c'est vouloir à tout prix ne disposer que d'une seule théorie (ne penser que d'une seule manière) qui me semble irrationnel.

> *« Il est désavantageux pour le penseur d'être lié à une seule personne. Lorsque on s'est trouvé soi-même, il faut essayer de temps en temps de se perdre et puis de se retrouver »* [11]

J'insiste enfin sur le fait que la critique qui suit n'est pas politique, elle est épistémologique. L'oppression dont je vais parler n'a rien à voir avec les systèmes politiques oppressants/les dictatures barbares qu'a connus l'humanité. J'entrevoie notre système scientifique actuel comme une prison dans laquelle nous nous sommes tous enfermés, nous scientifiques, sans en avoir conscience.

[9] Nietzsche, *Le gai savoir*, p. 201

[10] Nietzsche, *Humain, trop humain*, p. 321 ; Voir aussi les notions antagonistes de forces actives et de forces réactives expliquées par Gilles Deleuze dans son ouvrage *Nietzsche et la philosophie*

[11] Stefan Zweig, *Nietzsche*, Stock p. 82. Voir aussi Nietzsche, *Humain trop humain (II)*, FolioEssais p.309

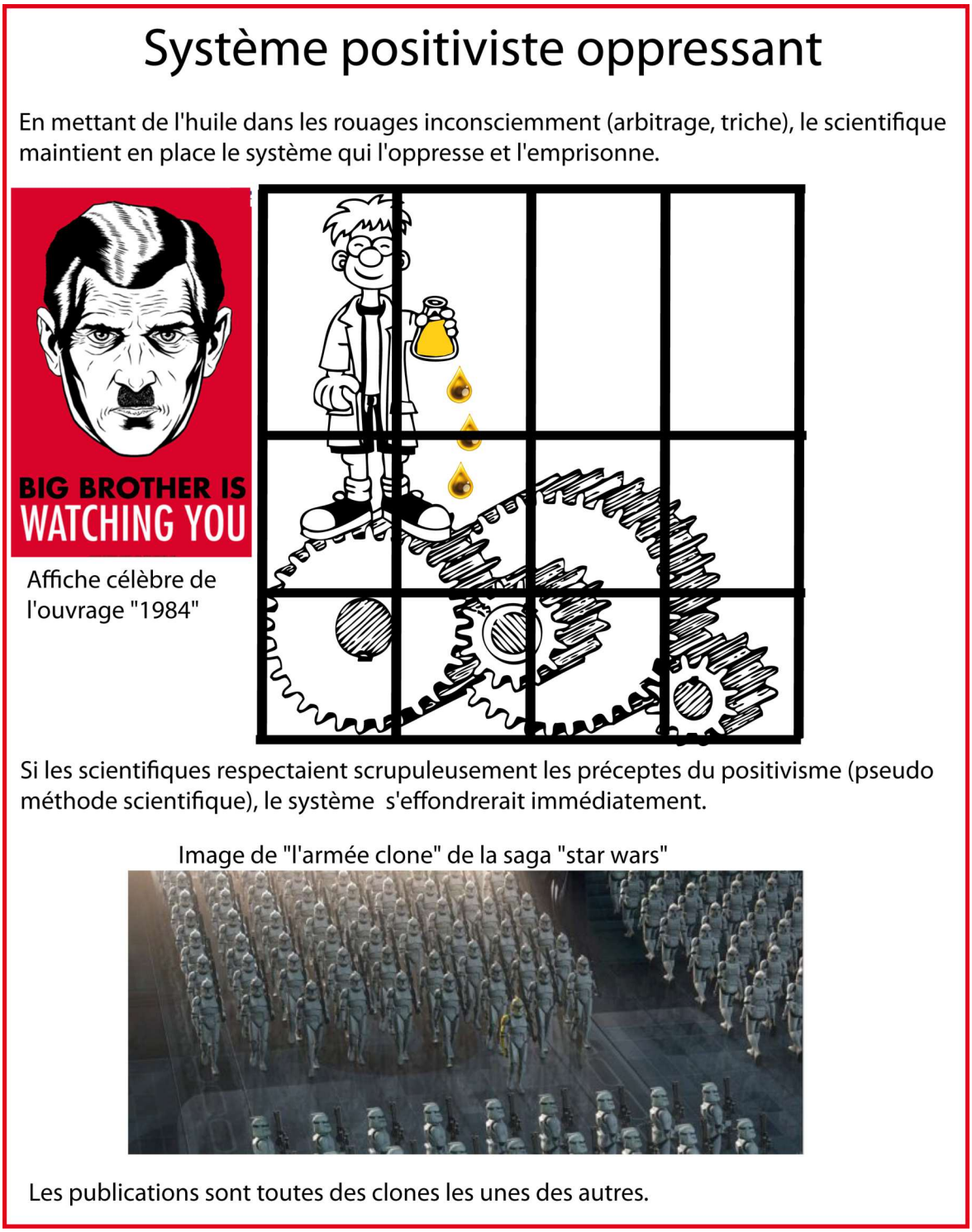



**Figure XVIII.2** – Le système positiviste oppressant.

Notre système positiviste actuel est trop rigide. Il est oppressant. La liberté d'expression y est malmenée. C'est une prison. La science qui en sort est totalement formatée par des standards beaucoup trop liberticides. Nos publications sont toutes des clones les unes des autres. De minuscules grains de sables qui sont sélectionnées sur la base du fait qu'ils sont identiques, standardisés. Toutes les publications qui ont les cheveux longs sont rejetées ou leurs cheveux sont coupés. Une publication doit obéir à un moule. Elle doit respecter les stupides dogmes « hypothèse–expérience–résultats », « données inaltérables » et « expérimentateur absent ». Le schéma type d'une publication actuelle est constitué d'une histoire sexy (inventée, reconstituée) mettant en avant des résultats positifs, originaux et respectant la méthode scientifique. Bien sûr, les angles ont été arrondis (falsification, arbitrage) mais cela correspond à la partie tabou, irrationnelle qui doit être évidemment invisible. Le plus étrange c'est que les chercheurs n'ont pas toujours conscience qu'ils sont obligés d'adapter et d'assouplir le socle positiviste trop rigide pour pouvoir mener à terme le processus scientifique à savoir la communication de leurs résultats. Ils rajoutent de l'huile (de l'irrationalité) dans les rouages pour éviter l'effondrement du système.

> *« Aucune société ne peut vivre avec seulement de l'autorité, des règlements, des normes, des injonctions. Même dans une société comme l'URSS où tout est dirigé, réglementé, totalisé au sommet par l'appareil du Parti qui coiffe l'appareil d'état et qui est omni-compétent, la société vit parce qu'à la base il y a une sorte d'anarchie de fait, où on se débrouille, on triche plus ou moins et l'ordre supérieur ne vit que par le désordre du bas, ce qui est un grand paradoxe mais un paradoxe qu'on retrouve dans tous les domaines puisque à l'usine Renault, les études de Mothé ont montré que si on appliquait à la lettre les instructions de la direction et des ingénieurs tout s'arrêterait [Ce qui fait qu'] Il est évident que pour faire marcher le système qui vous oppresse, il faut tricher avec le système »* [12]

La triche qui maintient le système ne constitue pas une base sérieuse à la science du 21ème siècle. Le positivisme feint d'ignorer la part d'irrationnel nécessaire qui intervient dans la méthode scientifique. Or il est totalement irrationnel de nier l'évidence. Il est irrationnel de nier l'existence et la nécessité de l'irrationnel dans la méthode scientifique. Ce qui est rationnel au contraire c'est d'accepter de voir et de reconnaitre ces éléments irrationnels puis de les intégrer à l'édifice scientifique. Tout le monde camoufle, maquille, truque, falsifie, arbitre mais personne ne le dit (ou ne s'en rend compte). Il est temps que les chercheurs complotent activement pour faire définitivement tomber le temple positiviste. Il est temps que les chercheurs se révoltent

[12] Edgar Morin, *Science avec conscience*, p. 105.

contre un système qui a fait son temps et qui doit maintenant être remplacé par un système beaucoup plus libre. Un système où le scientifique revendique le fait ne pas avoir les idées totalement claires, de ne pas tout savoir. Un système où les scientifiques peuvent faire part de leurs croyances, de leurs intuitions, de leurs sentiments ou de leurs doutes. Un système où l'on peut présenter aussi bien des résultats (positifs) en faveur de sa théorie que des résultats à décharge (négatifs) en faisant part de ses questionnements, de son incompréhension. Un système où le scientifique peut utiliser le « je » sans que cela soit bizarre.

> *« La différence est absolument considérable selon qu'un penseur a un rapport personnel à ses problèmes, de sorte qu'il possède en eux son destin, sa misère et aussi son bonheur le meilleur, ou au contraire un rapport impersonnel : c'est-à-dire s'il ne sait les palper et les saisir qu'avec les antennes d'une pensée froide et curieuse. Dans ce dernier cas, il n'en sortira rien, on peut l'assurer : car les grands problèmes, à supposer même qu'ils se laissent attraper, ne se laissent pas retenir par les grenouilles et les gringalets »*[13]

Dans ce système libre, le scientifique doit pouvoir faire référence à la philosophie, la religion, la société ou la métaphysique c'est-à-dire un système où il n'y a pas de sous-domaines, de sous-références. Dans ce système libre, le scientifique a le droit à l'erreur car l'erreur est considérée comme faisant intrinsèquement partie du processus scientifique puisque par définition nos publications sont des « erreurs ». Un système qui comprend que fondamentalement le scientifique prêche « sa science » tout comme le philosophe prêche « sa philosophie ». Nous devons réintroduire *la liberté* dans le système scientifique. Celui-ci est trop verrouillé. Et il est verouillé par l'obsession de vérification. Il est verouillé par la recherche de l'erreur, de la faute ou de la triche c'est-à-dire par la surveillance généralisée. Il est verrouillé par l'irruption d'un parasite invisible qui s'est greffé sur la science : la discipline. Parasite que nous allons démasquer maintenant.

> *« Ils travaillent, semblables à des moulins et à des pilons : qu'on leur jette seulement du grain ! — ils s'entendent à moudre le grain et à le transformer en blanche farine. Avec méfiance, ils se surveillent les doigts les uns aux autres. Inventifs en petites malices, ils épient ceux dont la science est boiteuse — ils guettent comme des araignées.*
>
> *Je les ai toujours vu préparer leurs poisons avec précaution ; et toujours ils couvraient leurs doigts de gants de verre.*
>
> *Ils savent aussi jouer avec des dés pipés ; et je les ai vus jouer avec tant d'ardeur qu'ils en étaient couverts de sueur.*

[13] Nietzsche, *Le gai savoir*, Flammarion p. 288.

*Nous sommes étrangers les uns aux autres* ***et leurs vertus me sont encore plus contraires que leurs faussetés et leurs dés pipés****.*

*Et lorsque je demeurais parmi eux, je demeurais au-dessus d'eux. C'est pour cela qu'ils m'en ont voulu.*

*Ils ne veulent pas qu'on leur dise que quelqu'un marche au-dessus de leurs têtes ; et c'est pourquoi ils ont mis du bois, de la terre et des ordures, entre moi et leurs têtes.*

*Ainsi ils ont étouffé le bruit de mes pas ; et jusqu'à présent ce sont les plus savants qui m'ont le moins bien entendu.*

*Ils ont mis entre eux et moi toutes les faiblesses et toutes les fautes des hommes : — dans leurs demeures ils appellent cela « faux plancher ».*

*Mais malgré tout je marche au-dessus de leur tête avec mes pensées ; et si je voulais même marcher sur mes propres défauts, je marcherais encore au-dessus d'eux et de leur tête.*

*Car les hommes ne sont point égaux : ainsi parle la justice. Et ce que je veux ils n'auraient pas le droit de le vouloir !*

*Ainsi parlait Zarathoustra. »*[14]

# 3 La question disciplinaire

Quelle peut être le rôle/l'impact de la discipline dans la démarche scientifique ? Dans l'élaboration des connaissances ? Doit-on considérer la discipline comme une invention, un outil utilisé en routine par les scientifiques ?

A cette question, il me semble que vous ne trouverez pas aisément de réponses. Il s'agit en effet d'une question délicate et un peu tabou car elle touche directement à l'organisation de notre système de recherche, aux notions de hiérarchie et d'examen. Elle nous renvoie à notre propre société qui est la société de la surveillance.

Quel directeur de thèse (« le chef ») oserait donner comme sujet de thèse à un étudiant la question du rôle/de l'impact de la discipline en science ? Car évidemment l'étudiant ne pourra pas ne pas analyser sa propre situation, ses propres rapports avec sa hiérarchie... La boucle de rétroaction potentielle en cas de conflit pourrait avoir des répercutions comiques... C'est chez Michel Foucault que j'ai trouvé la question ainsi posée :

> *« On fait l'histoire des expériences sur les aveugles-nés, les enfants-loups ou sur l'hypnose. Mais qui fera l'histoire plus générale, plus floue, plus déterminante aussi,*

[14] Friedrich Nietzsche, *Ainsi parlait Zarathoustra/Deuxième partie/Des savants.*

*de l'examen, de ses rituels, de ses méthodes, de ses personnages et de leur rôle, de ses jeux de questions et de réponses, de ses systèmes de notation et de classement ? Car dans cette mince technique se trouve engagés tout un domaine de savoir, tout un type de pouvoir. »* [15]

*« Mais il serait injuste de confronter les procédés disciplinaires avec des inventions comme la machine à vapeur ou le microscope d'Amici. Ils sont beaucoup moins ; et pourtant d'une certaine façon ils sont beaucoup plus. [...] Quel Grand Surveillant fera la méthodologie de l'examen, pour les sciences humaines ? »* [16]

## 3.1 Un système qui nous imprègne

Dés leur plus jeune âge, les chercheurs ont dû subir la discipline du système scolaire. Etant donné qu'ils font de longues études, la discipline est ancrée en profondeur dans leurs connexions neuronales. Les « petits enfants » chercheurs ont souvent fait parti des individus les plus disciplinés. Et si on établit la gaussienne des niveaux, ils se situaient souvent à droite de cette gaussienne, parmi les meilleurs.

*« Les disciplines caractérisent, qualifient, spécialisent ; elles distribuent le long d'une échelle, repartissent autour d'une norme, hiérarchisent les individus les uns par rapport aux autres, et à la limitent disqualifient et invalident. »*

On ne s'étonnera donc pas alors qu'ils finissent par l'aimer cette gaussienne. Ils en ont besoin. Ils ont toujours vécu avec : normaliser, hiérarchiser un groupe humain leur semble nécessaire. C'est, pour eux, la condition sine qua none de toute production. Le fondement de leur agir. « Le stable » qu'on ne peut leur soustraire car ils prennent appuie. Et cela tombe bien car sur la question de la discipline, il y a convergence des intérêts des scientifiques et des politiques. Le politique n'a pas à imposer l'évaluation, la discipline, la hiérarchie aux scientifiques car ceux-ci sont suffisamment disciplinés pour se l'imposer eux-mêmes !

## 3.2 Un système qui nous soumet

La conséquence de cette discipline (que les chercheurs s'auto-imposent) est simple : le scientifique n'est pas un homme libre. C'est un homme soumis. Contrairement au philosophe, le scientifique a une dépendance financière directe vis-à-vis du politique. Et c'est ce fil invisible qui le maintient à l'état d'homme soumis, discipliné, docile.

[15] Michel Foucault, *Surveiller et punir*, Gallimard p. 217
[16] Michel Foucault, *Surveiller et punir*, Gallimard p. 261-263

*« Double effet par conséquent de cette pénalité hiérarchisante : distribuer les élèves selon leur aptitudes et leur conduite, donc selon l'usage qu'on pourra en faire quand ils sortiront de l'école ; exercer sur eux une pression constante pour qu'ils se soumettent tous au même modèle, pour qu'ils soient contraints tous ensemble à la subordination à la docilité, à l'attention dans les études et exercices et à l'exacte pratique des devoirs et de toutes les parties de la discipline. Pour que tous ils se ressemblent. »*

Paul Feyerabend prônait la séparation de l'état et de la science en sous-entendant, relativiste qu'il était, que la science n'est pas différente de la religion. Cette attitude provocante m'a toujours amusé mais je ne l'ai jamais rejoint sur ce terrain (par intérêt personnel sans doute !). Pourtant, quand on voit le lien invisible qui relie science et politique via la discipline, on peut se demander dans quelle mesure, l'indépendance totale de la science vis-à-vis de l'état ne serait pas le chemin nécessaire pour pouvoir gommer librement, si cela s'avérait nécessaire, tous éléments (disciplinaires ou non) néfastes au développement des sciences.

La méthode qu'utilise la discipline pour soumettre est l'examen.

*« L'examen combine les techniques de la hiérarchie qui surveille et celles de la sanction qui normalise. Il est un regard normalisateur, une surveillance qui permet de qualifier, de classer et de punir. Il établit sur les individus une visibilité à travers laquelle on les différencie et on les sanctionne. C'est pourquoi, dans tous les dispositifs de la discipline, l'examen est hautement ritualisé. »*[17]

*« L'examen entouré de toutes ses techniques documentaires fait de chaque individu un cas. Le cas c'est [...] l'individu tel qu'on peut le décrire, le jauger, le mesurer, le comparer à d'autres et cela dans son individualité même ; et c'est aussi l'individu qu'on a à dresser ou redresser, qu'on a à classer, à normaliser, à exclure. »*[18]

Les examens auxquels se plient les scientifiques sont multiples : le doctorat est le premier examen d'entrée dans la communauté scientifique suivi par l'Habilitation à Diriger des Recherche (HDR). Citons aussi les comités d'évaluation des unités (AERES par exemple). La publication en elle-même est devenue un examen. On « *soumet* » (on se soumet ?) sa publication pour évaluation par ses pairs. Si cette dernière est acceptée, cela fait office « de bonne note ». Le processus de « review » est si long (mois, années) que cela crée une attente, une impatience : le désir de se voir décerner « un bon point ». Un écolier, satisfait de sa copie, ressent la même impatience en attendant sa note. Ainsi les rôles dominant-dominé sont inversés. En effet, il pourrait sembler plus logique qu'une publication représente un geste d'apparence d'altruiste :

[17] Michel Foucault, *Surveiller et punir*, Gallimard p. 217
[18] Michel Foucault, *Surveiller et punir*, Gallimard p. 224

un chercheur **fait l'effort** d'écrire une publication pour transmettre son savoir à l'humanité. Il est **le dominant** qui offre « la becquée » aux oisillons (le reste de l'humanité). Pour cela, pas besoin d'années : une fois écrite, une publication pourrait être publiée en quelque heures (comme dans la presse quotidienne). Or les choses se passent dans le sens inverse : l'auteur est l'oisillon, le dominé, l'égoïste, l'enfant qui veut se voir décerner « un bon point ». Pour cela, il passe par un long rituel examinatoire qui peut durer plusieurs mois, années. Officiellement, toute la procédure complexe qui mène à la publication existe pour accroitre la qualité des publications. En réalité, elle est créée pour introduire et ritualiser une forme d'examen. Elle est crée pour normaliser, pour classer.




**Figure XVIII.3** – L'acte de publier revêt l'apparence d'un examen. Nous, chercheurs, nous surveillons les uns les autres pour détecter celui dont la science est boiteuse. Cette surveillance est symbolisée ici par l'architecture d'une prison permettant à un gardien au centre de la tour de surveiller toutes les cellules en même temps grâce au jeu de lumière : on parle de panoptique. Dans le système disciplinaire, ce n'est pas le chef qui est regardé/admiré mais les subordonnées qui sont passés en revue.

## 3.3 Un système qui normalise

La gaussienne obtenue grâce à l'examen a un double rôle. Elle « normalise » tout en distribuant selon le niveau (elle hiérarchise). Comme je vous l'ai déjà dit : les publications sont toutes similaires car elles ont été normalisées.

Le scientifique est habitué à faire en sorte d'être à droite de la gaussienne, *parmi les meilleurs*. Et justement le problème est là. Le scientifique a intériorisé un but crée par la discipline (être à droite de la gaussienne) qui n'est pas le but réel de la science. Le scientifique recherche donc souvent la connaissance dans le but ultime de se maintenir *à droite de la gaussienne*. Derrière ce but ultime se cache la volonté de puissance qu'un homme cherche à avoir sur les autres hommes.

Or justement, la science ne représente pas la volonté de puissance de l'homme sur l'homme. Elle est volonté de puissance de l'homme sur la nature. Le chercheur doit être un guerrier qui, par le biais de la grande santé (concept Nietzschéen qui signifie une santé que l'on sacrifie sans cesse et donc incluant la souffrance), recherche le néant c'est-à-dire qu'il nie sa vie telle qu'elle est. Mais cela nous l'étudierons un peu plus tard.

Le problème de la discipline c'est qu'elle maintient le scientifique attaché à un but non scientifique (obéir, diriger, être à droite de la Gaussienne). Or, pour réellement *attaquer* efficacement la nature, il faut se libérer de cette chaine invisible et aller là où la discipline ne peut intrinsèquement pas nous mener : vers une souffrance consentie mais qui n'est pas investie dans le but de se déplacer à droite de la Gaussienne.

## 3.4 Un système qui récompense

La chasse à la vérité n'a pas, en soi, vocation à une récompense autre que l'atteinte provisoire de la grande santé. Or notre système disciplinaire implémente des récompenses pour ceux qui sont à droite de la Gaussienne : il s'agit de :

⇒ l'argent/le confort

⇒ la gloire/la reconnaissance

⇒ le pouvoir/la responsabilité

Confort, reconnaissance et responsabilité éloignent de la volonté de puissance de l'homme sur la nature. Ces éléments empêchent de ressentir la volonté d'interroger et de souffrir c'est-à-dire d'attaquer la nature. De manière assez étonnante, ces récompenses protègent la discipline : lorsque l'on mène une vie confortable et que l'on est reconnu, il est difficile de juger le système scientifico-politique qui implémente les éléments disciplinaires de manière objective : on est inexorablement tenté de le juger positivement, ou au moins, de le juger nécessaire. Plus étonnant

encore, je soupçonne que la responsabilité soit toujours **la responsabilité de maintenir les éléments disciplinaires**. Jamais la responsabilité de maximiser la production scientifique. Cela est capital car justement ces 2 objectifs sont assez souvent opposés.

## 3.5 Un système qui surveille

> *« Le pouvoir qu'elle [la technologie du panoptisme c'est-à-dire de surveillance généralisée] met en oeuvre et qu'elle permet de majorer est un pouvoir direct et physique que les hommes exercent les uns sur les autres. Pour un point d'arrivée sans gloire, une origine difficile à avouer. »*[19]

> *« Le pouvoir disciplinaire s'exerce en se rendant invisible ; en revanche il impose à ceux qu'il soumet un principe de visibilité obligatoire. Dans la discipline, ce sont les sujets qui ont à être vus. Leur éclairage assure l'emprise du pouvoir qui s'exerce sur eux. C'est le fait d'être vu sans cesse, de pouvoir toujours être vu, qui maintient dans son assujettissement l'individu disciplinaire. »*[20]

> *« La discipline crée entre les individus un lien privé qui est un rapport de contraintes entièrement différent de l'obligation contractuelle ; l'acceptation d'une discipline peut bien être souscrite par voir de contrat, la manière dont elle est imposée, les mécanismes qu'elle fait jouer, la subordination non réversible des uns par rapport aux autres, le « plus de pouvoir » qui est toujours fixé du même coté[...]opposent le lien contractuel et le lien disciplinaire. »*[21]

Un exemple de ce panoptisme (doctrine/technologie pour surveiller les individus) est en train d'émerger dans les laboratoires : le cahier de manipulation.

Dans son objectif initial, le cahier de manipulation est une béquille utilisée par le chercheur pour combler sa mémoire imparfaite. Il s'agit d'un allié précieux et intime, une extension de sa mémoire qui devrait donc avoir un caractère personnel et privé (j'imagine que vous n'autorisez personne à fouiller dans votre mémoire).

Or aujourd'hui, le cahier de manipulation est bien plus qu'une mémoire :

⇒ il n'appartient plus au chercheur mais au laboratoire (c'est l'expropriation).

⇒ il doit être signé chaque semaine par le supérieur : c'est l'examen, la vérification du travail.

[19] Michel Foucault, *Surveiller et punir*, Gallimard p. 261
[20] Michel Foucault, *Surveiller et punir*, Gallimard p. 220
[21] Michel Foucault, *Surveiller et punir*, Gallimard p. 259

⇒ il a valeur juridique en cas de litige sur les notions de propriété intellectuelle. La parole d'un scientifique est plus suspecte que son écrit sur un cahier.

Vous voyez donc que le cahier de manipulation[22] s'est transformé en machine « panoptique » : un outil mis à disposition des chefs pour « examiner » le travail des subordonnées. Mais la machine panoptique est enracinée beaucoup plus profondément dans les sciences. Regardez tout simplement le plaisir presque cruel qu'ont les juges « scientifiques » (que nous sommes tous) à porter le jugement d'irrationalité. Nous nous surveillons tous les uns les autres. Mais on ne s'en rend plus compte. Car la surveillance mutuelle invisible est une des conséquences, un des objectifs de la discipline.

> *« Avec méfiance, ils se surveillent les doigts les uns aux autres. Inventifs et petites malices, ils épient ceux dont la science est boiteuse- ils guettent comme des araignées »* [23]

Jusqu'à maintenant, j'ai décrit la discipline comme un exosquelette, un parasite greffé sur la science mais bien indépendant. En réalité, la réussite la plus forte des disciplines est d'avoir réussi à se fondre dans la définition même de la science : la vérification et la critique sont des examens, des évaluations, des jugements. Ils sont considérés comme au fondement de l'activité scientifique. Mon but sera maintenant de déloger la critique pour montrer que celle-ci est un gène du parasite (la discipline) et surement pas un gène de l'hôte (la science). Or je sais bien que la critique est sacralisée par les scientifiques. Je sais bien qu'il est très peu probable qu'une seule personne me rejoigne sur ce point. C'est pourquoi j'ai prévu une longue démonstration.

# 4 La déconstruction de la critique

## 4.1 Préambule

Je vois un petit problème se profiler avec ma critique de la critique. Mon lecteur (ou rapporteur) pourrait trouver facile (et donc embêtant) que j'écrive une thèse dans laquelle je déconstruis la critique. Comment pourrait-il alors me critiquer ?

Voila ma réponse : ma thèse constructiviste doit être perçue comme un ensemble de constructions cognitives, des spéculations, des pistes ou concepts que je souhaite transmettre. J'espère qu'ils seront « orientants » mais je ne peux le garantir car je ne prétends nullement détenir la vérité. Au contraire, je n'ai aucun doute, ni aucun problème avec le fait que certaines pistes méritent toute simplement d'être fermées. A partir de là, vous, cher lecteur, pouvez soit :

[22] Je précise que je n'ai jamais eu à subir ce genre de pratique dans mon laboratoire.

[23] Nietzsche, Ainsi parlait Zarathoustra/Deuxième partie/Des savants

⇒ suivre ma piste « contre la critique » : contentez vous de ne pas transmettre les idées/pistes que je propose et auxquelles vous ne croyez pas. Soyez certain que c'est violent.

⇒ ne pas suivre ma piste « contre la critique » : sentez-vous libre de me critiquer. En effet, ce n'est pas parce que je déconstruis la critique que je considère que ma piste est « orientante »/ validée. Je le répète, il s'agit également pour moi, de spéculations. Dés lors, je ne peux ni ne souhaite m'opposer à la critique en ce qui concerne mon manuscrit.

Quels sont les arguments des gens qui défendent la critique et le scepticisme en science ? Edgar Morin souligne que la critique dans la communauté scientifique permet l'émergence d'un consensus intersubjectif qui fait office d'objectivité pendant un temps. Ce processus offre une démarcation floue entre science et non science. Ainsi pour pouvoir séparer l'astrophysique de l'astrologie, il faut pouvoir critiquer. Pour séparer un médicament d'une pilule miraculeuse anti-cancer, il faut la critique. Pour séparer science et croyance, il faut la critique. Toucher à la critique, au scepticisme c'est faire rentrer, selon beaucoup, le loup dans la bergerie.

Vous n'aurez pas manqué de constater qu'en critiquant la critique, je crée un paradoxe car si je critique la critique, quelle valeur accorder à ma critique ? C'est pour cela que j'ai astucieusement remplacé le mot critique par le mot déconstruction —petit astuce de sémantique qui ne supprime pas pour autant le paradoxe— . Mais il va falloir faire avec ce dernier et reconnaitre que le match commence mal pour moi. 1–0 pour les tenants de l'importance de la critique.

Tous les scientifiques sont aussi des dogmatiques. Ils peuvent croire en toute sorte de choses, par exemple :

⇒ l'essence des concepts (l'existence de Pi en dehors de toute construction cognitive).

⇒ la réalité essentielle de la réalité existentielle.

⇒ l'existence de méthodes scientifiques pures et donc inattaquables.

⇒ l'idée qu'il existe une démarcation nette entre science et croyance (pourquoi pas la falsifiabilité).

Dés lors que nous sommes nous même, scientifiques, sous l'influence de pensées dogmatiques, est-il prudent de refuser l'entrée des dogmes et croyances à l'intérieur des sciences ? Et donc se refuser sa propre entrée ? Et vous voyez maintenant que le match, du point de vue des paradoxes, est à 1–1

Commençons maintenant la véritable argumentation contre la critique.

## 4.2 La critique mène au principe d'autonomie

> *« La première théorie adéquate a droit de priorité sur les suivantes également adéquates »*[24]

Cela signifie que si une première théorie explique un phénomène, puis qu'une deuxième théorie propose une explication différente mais de qualité semblable, celle-ci ne réussira pas à s'imposer, c'est-à-dire à obtenir une crédibilité de 50%. La première est toujours prioritaire car elle aura ce qu'on appelle aux échecs l'avantage du trait, avantage du joueur qui est le premier à jouer (les blancs) : elle occupe l'espace et empêche la deuxième théorie de se développer. Vous voyez qu'il n'y a rien de rationnel dans ce principe d'autonomie et pourtant c'est comme ça que les choses fonctionnent en science. Mais si par hasard, la deuxième théorie a de meilleurs arguments, explique plus de résultats, il y aura un combat et une gagnante : on rejettera la première théorie. Il ne peut y avoir qu'une théorie majoritaire. Encore une fois, cela est dû à la critique et cela n'a rien de rationnel. Deux théories devraient pouvoir parfaitement cohabiter en parallèle. Mais ce n'est presque jamais le cas. Car nous aimons la compétition, les vainqueurs, et il n'y a qu'une place sur le podium.

## 4.3 La critique mène à l'exhaustivité

La critique rend le chercheur exhaustif ce qui est dangereux. En effet, tout chercheur se doit d'être exhaustif dans son domaine car sinon le risque est trop grand qu'on lui oppose, par la critique, des arguments et des expériences qu'il ne connait pas ce qui pourrait être une atteinte à sa crédibilité. En connaissant tout sur son domaine, le scientifique ne risque pas d'être contredit et mieux, il peut contredire les autres. Pourtant tout le temps qu'il passe à être exhaustif (apprendre des détails le plus souvent), il ne peut l'utiliser pour élargir ses vues aux autres domaines. Il devient alors un chien de garde incapable d'aller sur le territoire des autres et défendant rigoureusement le sien. Et cela Edgar Morin l'a bien vu, il dit humoristiquement :

> *« On arrive à une clôture interdisciplinaire, hyper disciplinaire, ou chacun évidemment est propriétaire d'un maigre territoire et compense son incapacité à réfléchir sur les territoires des autres par l'interdiction rigoureuse faite à autrui de pénétrer sur le sien. Vous savez que les éthologistes ont reconnu cet instinct de propriété territoriale chez les animaux. Dés qu'on entre sur leur territoire : les oiseaux s'égosillent, les chiens aboient etc... Ce comportement a beaucoup diminué dans l'espèce humaine, sauf chez les universitaires et les scientifiques. »*[25]

[24] Paul Feyerabend, *Contre la méthode*, Points p. 34.
[25] Edgar Morin, *Science avec conscience*, Points p. 74.

Nietzche nous fait comprendre les œillères qui poussent quand on se spécialise :

> *« Toute espèce de maitrise se paie cher sur terre, ou tout peut être se paie trop cher ; le prix à payer pour être l'homme de sa spécialité est d'être également la victime de sa spécialité »* [26]

## 4.4 La critique mène à la stabilité des opinions

La science impose d'avoir un esprit très flexible capable de renier l'idée défendue une minute avant. Or les girouettes ne sont pas en vogue dans le monde dans lequel on vit.

> *« La disposition d'esprit de l'homme de connaissance, en tant qu'elle est en contradiction avec la réputation de fermeté est tenue pour déshonorante tandis que la pétrification des opinions s'attire tous les honneurs »*[27]

Et cette phrase si limpide et profonde de Nietzche :

> *« La quantité de croyance dont quelqu'un a besoin pour se développer, la quantité de « stable » auquel il ne veut pas qu'on touche parce qu'il y prend appui — offre une échelle de mesure de sa force (ou pour m'exprimer plus clairement) de sa faiblesse »*[28]

Le stable sur lequel les hommes prennent appuie est à la fois nécessaire et limitant. Tout chercheur devrait essayer de déterminer quel est son stable. Puis essayer d'attaquer ce stable pour déterminer ses points faibles et à l'inverse, ce à quoi il résiste. Tout chercheur devrait rêver d'avoir un stable très dynamique, très changeant, très malléable, très *instable* : le chercheur doit dépenser et consommer des convictions successives. Car on peut déterminer sa profondeur à la dose d'instabilité qu'il est capable de supporter. La critique mène à l'affrontement des stables, au choc des stables. Or ces chocs ne sont pas une nécessité. Ce qui est indispensable, c'est l'interaction entre les stables mais cette dernière ne prend pas nécessairement la forme d'un affrontement.

## 4.5 La critique mène à l'excellence

Il est très difficile, en science, de critiquer sans remettre en cause la compétence du chercheur.

> *« Une des propriétés des controverses scientifiques est de mettre la compétence des protagonistes sur la sellette. En temps normal, les capacités des chercheurs sont*

[26] Nietzsche, *Le gai savoir*, p. 32.
[27] Nietzsche, *Le gai savoir*, p. 234.
[28] Nietzsche, *Le gai savoir*, p. 292.

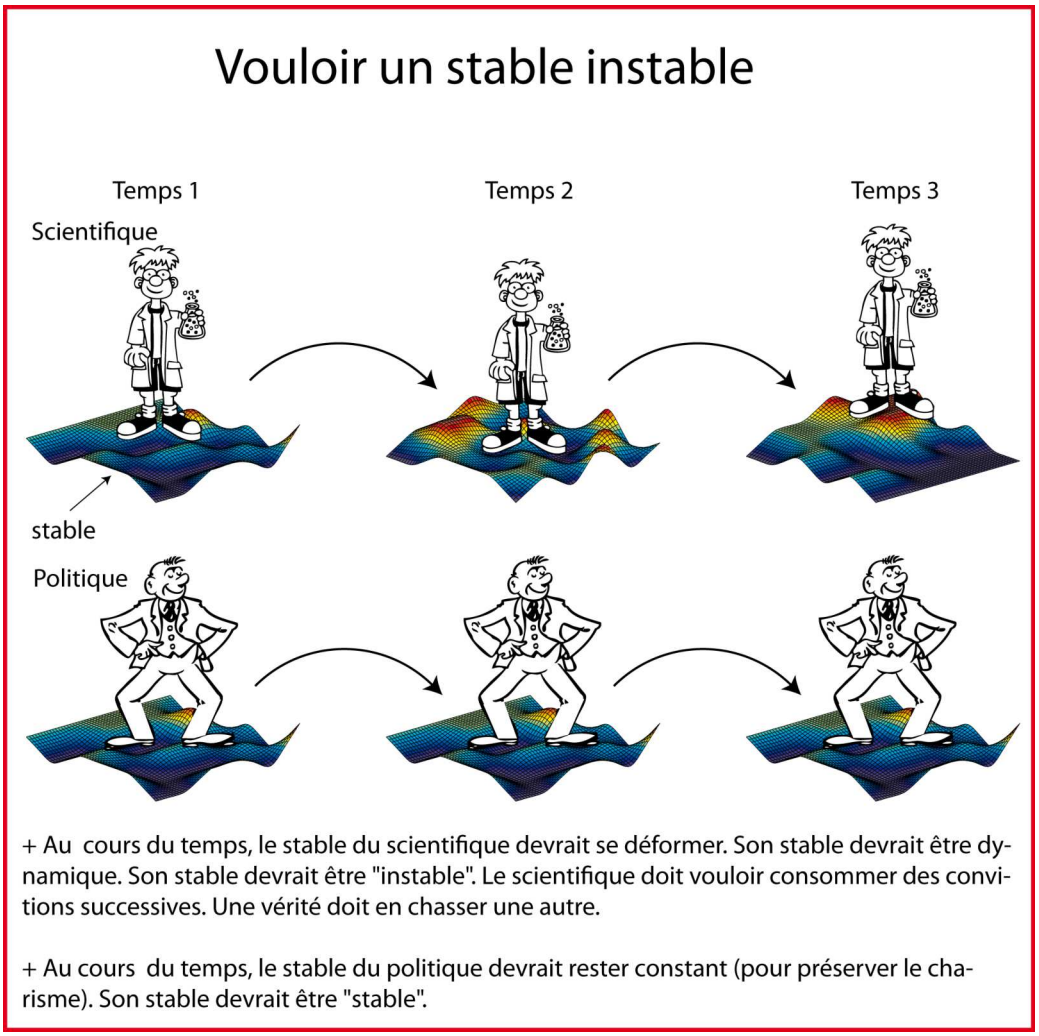



**Figure XVIII.4** – Le rêve de tout scientifique devrait être de posséder un stable instable.

*tenues pour acquises. Dans une controverse, en revanche, il est difficile de dissocier les questions particulières qui sont en jeu de l'aptitude des hommes de science impliqués »* [29]

*« Les détracteurs font avant tout appel à des résultats négatifs pour fonder leur rejet du phénomène controversé et tous les résultats positifs s'expliquent, selon eux, par l'incompétence, l'illusion ou même la fraude. Les défenseurs expliquent quant à eux les résultats négatifs par l'inaptitude à reproduire exactement les conditions de l'expérience qui a permis d'obtenir les résultats positifs. A elles seules, les expériences ne semblent pas suffire à régler la question »*[30]

[29]Harry collins, *Tout ce que vous devriez savoir sur la science*, Points p. 154.
[30]Harry collins, *Tout ce que vous devriez savoir sur la science*, Points p. 101.

Au bout de moment, je crains que l'objectif des détracteurs et des défenseurs ne soit plus tant d'avoir raison mais d'avoir *raison sur l'autre*. La critique n'a pas pour objectif de faire avancer la science mais de montrer la supériorité de son intelligence et de ses raisonnements par rapport à ceux des autres. Or en écrasant l'autre, un des siens, je doute qu'on optimise l'avancée du progrès, de la science. Voici un petit aphorisme de Nietzsche à prendre à la fois au premier et au deuxième degré :

> *« La critique, bornée et injuste aussi bien que compréhensive, vaut tellement de plaisir à qui l'exerce que le monde a une dette de reconnaissance envers toute œuvre, toute action qui incitent beaucoup de gens à beaucoup de critiques ; car derrière elles s'étire une queue étincelante de joie, d'esprit, d'admiration de soi-même, de fierté, de leçons, de ferme propos de mieux faire »* [31]

Derrière la critique se cache cet ersatz de finalité que représente l'excellence. Or l'excellence n'est pas un but en soi et n'a rien à voir avec la méthode scientifique. L'excellence est un concept politique qui repose sur la compétition et l'évaluation. Elle ne sert qu'à faire émerger une aristocratie (aristos, les meilleurs) repliée et fermée sur elle-même. Or ce champ lexical « l'excellence, la compétition, l'évaluation, l'émergence des meilleurs » est ancrée en profondeur dans les connexions neuronales des chercheurs. Ils en sont devenus addicts après avoir subi 25 ans d'imprégnation du système scolaire. Ils ont sanctuarisé/sacralisé la critique pour éviter qu'on ne touche à « leur compétition » devenue « le stable de Nietzsche » dont ils ne peuvent plus se passer. Non ! La critique n'est pas nécessaire, en soi, à l'activité scientifique. Quelque soit la critique (y compris la critique dite constructive), l'homme « critiquant » veut toujours inconsciemment « gagner » quelque chose et il revendiquera souvent l'inverse en disant que c'est l'homme « critiqué » qui gagne quelque chose. En effet, la critique illustre toujours la volonté de puissance de l'homme sur l'homme. Or nous verrons plus tard que la science de haut niveau (au sens informatique : la couche superficielle) est la volonté de puissance de l'homme sur la nature. Cette dernière est donc totalement étrangère à l'idée de gagner quelque chose sur quelqu'un.

Être excellent ne devrait pas être un but en soi. Et je m'applique à ce que mon travail ne le soit pas. Un but possible auquel je crois déjà beaucoup plus c'est le fait d'être « orientant ». Bien que ces deux mots « excellent » et « orientant » vont souvent de pair, ce n'est pas automatique. Si vous me demandez ce qu'est un pomélo, je peux vous répondre qu'il s'agit d'un fruit comestible légèrement sucré à l'écorce mince de couleur jaune et au nom latin *Citrus paradisi* mais je peux aussi vous répondre qu'il s'agit d'une sorte de gros pamplemousse. Quelle est la réponse « excellente » ? Quelle est la réponse « orientante » ? Et quelle est celle que vous préférez ?

[31] Nietzsche, *Humain trop humain (II)*, Folio Essais p.79

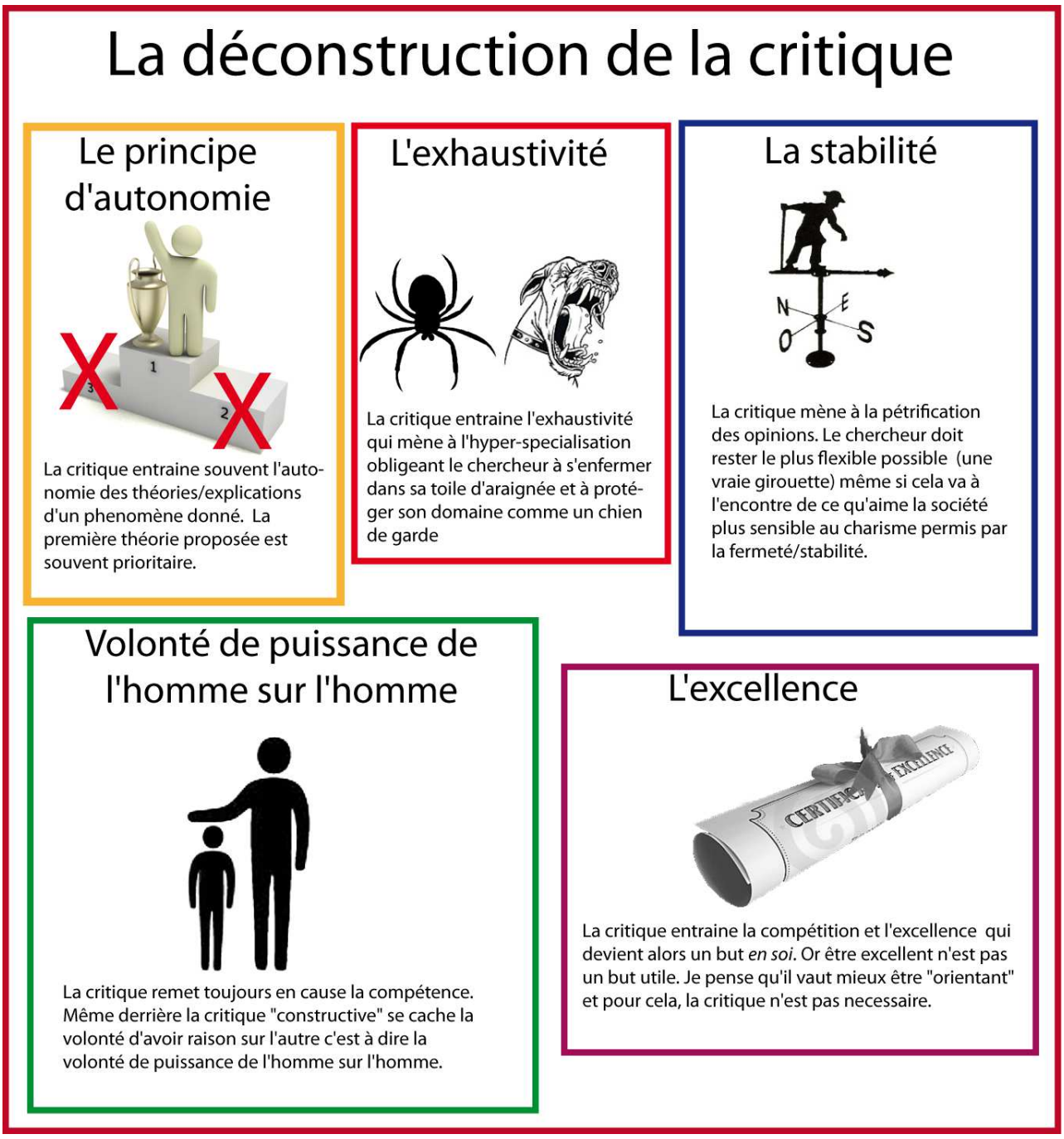



**Figure XVIII.5** – La déconstruction de la critique.

## 4.6 Le scepticisme ne vaut que pour soi

Faisons une expérience de pensée. Je place l'ensemble des résultats scientifiques présents dans les publications de l'année 2011 dans une grande matrice binaire. Puis je reproduis chacune des expériences, une par une, en plaçant un 1 dans la matrice quand le résultat est reproduit et un 0 dans le cas inverse. Cette matrice ne représente pas la vérité ou la fausseté des résultats. Car l'absence de réussite lors de la reproduction ne signifie pas forcement que le résultat est faux. En effet, je peux me tromper ou mal m'y prendre, etc... Cette matrice représente l'ensemble des résultats que je réussis ou pas à reconstruire cognitivement. Ma prédiction, basée sur une synthèse/un ressenti global de mes échecs et réussites personnels en matière de reproduction des résultats des autres, est que cette matrice est globalement creuse (il y a énormément de zéros). Je ne conclus jamais que les autres se trompent car cela serait une analyse trop faible cachant une profonde mécompréhension du fonctionnement des sciences. Mais quand je reproduis le résultat d'un autre, je le vis toujours comme une intense communion d'esprit avec l'auteur de la découverte. Nous voyons un phénomène avec les mêmes yeux. Je le vis toujours comme *un nouveau résultat en moi*, même s'il est déjà publié. J'ai senti et construit cognitivement son phénomène. Et je n'espère rien d'autre en publiant qu'un autre reconstruise cognitivement les phénomènes que j'ai découverts et ressente cette communion d'esprit. Or selon moi, la plupart des chercheurs procèdent, à tort, de manière inverse : tout résultat publié est un résultat acquis : « c'est sûr que cela va marcher ». Si cela ne fonctionne pas, c'est que les « autres » se sont soit trompés (hypothèse d'incompétence) soit ont triché (hypothèse de falsification). Cette tendance/faiblesse/facilité à se laisser tenter par le scepticisme (critique de la prétention à la vérité) est encouragée et validée par l'épistémologie positiviste institutionnelle. Pire, ce scepticisme est considéré par beaucoup comme le moteur essentiel des sciences.

J'ai trouvé chez André Comte-Sponville, l'idée que la morale ne vaut que pour soi.

> *« « Que dois-je faire ? » Et non pas « que doivent faire les autres ? » C'est ce qui distingue la morale du moralisme [...] La morale ne vaut que pour soi ; les devoirs ne valent que pour soi. Pour les autres, la miséricorde et le droit suffisent. »* [32]

J'ai trouvé chez Gilles Deleuze, en substance, l'idée que :

> *« Quand on parle de bêtise, on parle toujours de la sienne. »*

J'ai trouvé chez Nietzsche l'idée que la responsabilité ne vaut que pour soi :

> *« signe de noblesse morale : ne jamais songer à rabaisser ses devoirs pour en faire les devoirs de tout le monde, ne pas abdiquer sa responsabilité propre, ne pas vouloir*

[32] André Compte-Sponville, *Présentation de la philosophie*, le livre de poche p. 20.

*la partager* » [33]

A partir de là, je propose que lorsque l'on parle de scepticisme, on parle toujours du sien. Autrement dit le scientifique doit être le plus grand des sceptiques en ce qui concerne ses résultats et ses découvertes. Il doit douter, reproduire encore et encore jusqu'à ce que sa certitude (qui ne sera jamais totale) l'emporte sur son scepticisme. A l'inverse, il doit à tout prix apprendre (c'est sa responsabilité) à lutter contre sa tendance si naturelle (trop naturelle), si familière et si facile d'être sceptique vis-à-vis des idées des autres. Car le scepticisme c'est la volonté de puissance de l'homme sur l'homme. C'est la volonté d'avoir *raison sur l'autre*. Je ne dis pas que c'est mal ou immoral. Je dis que ce n'est pas de la science de haut niveau (au sens informatique : la couche superficielle) car cette dernière est volonté de puissance de l'homme sur la nature. L'attaqué est toujours, par définition, la nature. Jamais l'homme. Mais cela, je l'aborderai plus en profondeur dans le dernier chapitre d'épistémologie.

## 4.7 L'alternative à la critique : l'interrogation

Comment transmettre et propager les idées les plus « orientantes » sans critiquer ?

Tout d'abord, je vais définir ce que j'entends par critique. Pour moi, la critique contient les notions d'évaluation, de jugement. Il y a également une prétention à détenir une vérité supérieure.

Prenons un exemple. Si un chercheur 1 propose un énoncé A « la terre est plate » et qu'un chercheur 2 lui répond par un énoncé B « non, la terre est ronde ». Le chercheur porte un jugement négatif sur l'énoncé A et juge son énoncé B meilleur, supérieur. C'est cela que j'appelle la critique.

Le critique/le sceptique[34] cherche à détruire/attaquer les représentations cérébrales des interprétations d'expériences (que je nommerai dorénavant *constructions cognitives*) jugées fausses/mauvaises. S'il ne peut les détruire, il fera tout pour empêcher leurs propagations, leurs duplications. Il me semble que cette attitude est profondément néfaste à la science car elle est paradoxale : le sceptique critique la prétention du critiqué à détenir la vérité or en faisant cela, il prétend, lui, détenir une vérité qui est *meilleure*. En se plaçant du point de vue des épistémologies constructivistes, je supprime l'idée de vérité. Il n'y a que des constructions cognitives qui se propagent ou ne se propagent pas.

L'alternative que je propose à la critique est l'interrogation[35]. Si vous n'arrivez pas à re-

[33] Nietzsche, *Par delà bien et mal*, Folio p. 198.

[34] Ces deux termes ne sont en réalité pas tout à fait équivalent. Mais dans notre cas, nous considérerons qu'ils sont très proches.

[35] On pourra bien sûr dresser un pont avec la dialectique Socratique.

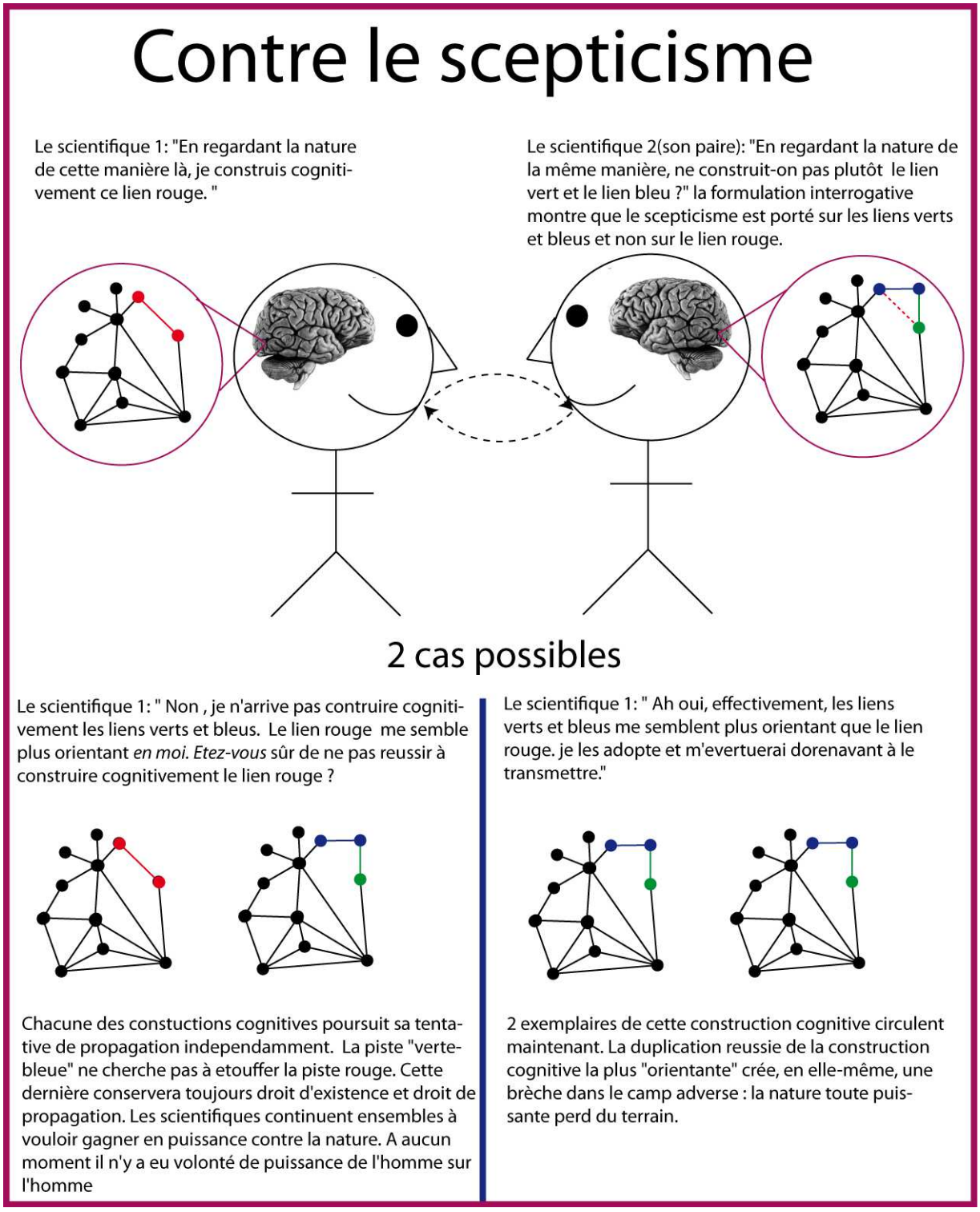



**Figure XVIII.6** – Le scepticisme ne vaut que pour soi.

construire une construction cognitive qui vous est proposée par un pair, vous ne devez pas être sceptique vis-à-vis de sa construction à lui. Vous devez être sceptique vis-à-vis de votre construction cognitive alternative. Et ce scepticisme envers vous même doit ressortir lorsque vous proposerez à votre pair votre alternative sous forme interrogative. Dés lors, soit votre pair adopte votre construction cognitive et celle-ci se duplique, soit il garde sa construction cognitive antérieure en vous répondant, à son tour, avec l'interrogation. Dans ce cas, les deux constructions cognitives continuent leur tentative de duplication chacune de leur coté. A aucun moment n'est mise en cause la pertinence de l'une ou l'autre des constructions cognitives puisque l'idée de verité est abandonnée. De plus, à aucun moment il n'est fait usage du principe d'autonomie qui viserait à ne conserver qu'une construction jugée meilleure en critiquant/attaquant les autres. Toute construction cognitive a et conservera sa chance de propagation sauf si son constructeur choisit de l'abandonner.

Dans la critique constructive, seul le *critiquant* « donne » et le *critiqué* décide s'il reconnait la critique comme constructive (et adopte la nouvelle construction cognitive) ou s'il la rejette. L'absence de réciprocité (« seul le critiquant donne ») me laisse penser qu'il s'agit en réalité de volonté de puissance du critiquant sur le critiqué caché derrière le don en apparence altruiste.

En revanche, avec l'interrogation, *l'interrogeant* reçoit autant qu'il donne quelque soit le comportement de *l'interrogé* (que ce dernier prenne ou rejette la nouvelle construction cognitive proposée). La tentative de duplication est une tentative d'attaque contre la nature, pas une attaque contre l'homme.

Je répète et j'insiste sur le fait qu'il ne s'agit pas de morale. L'interrogation n'a pas pour but « d'adoucir » la critique. L'interrogation rajoute la réflexivité de l'interrogeant sur lui-même ce qui gomme toute volonté de puissance de l'homme sur l'homme. Vous pourriez trouver que je joue avec les mots et qu'interrogation et critique sont en fait très similaires. Je ne le crois pas. En réalité, ce que nous venons de faire c'est de modifier les fondements supposées de la démarche scientifique : ce n'est pas la critique qui est nécessaire pour faire émerger l'objectivité mais le phénomène interrogatif. Et c'est l'interrogation qui est donc au fondement de la démarche scientifique. Nous reviendrons bientôt sur ce point plus en détail.

Dans cette partie, j'ai défendu l'idée que la discipline et la critique sont des éléments parasitaires greffés sur la science mais que ceux-ci ne la constituent pas. Etrangement, l'argument le plus sérieux qui s'oppose à cette vision pourrait être formulé par Paul Feyerabend avec son fameux « tout est bon ». L'argument est simple : selon le but recherché, un scientifique peut vouloir utiliser la critique ou au contraire, ne pas vouloir l'utiliser. Dans certain cas, il aura peut être raison, dans d'autres il aura peut-être tort. Il n'y a pas de règles strictes à suivre (toujours critiquer ou ne jamais critiquer). Idem pour la discipline : pour atteindre un objectif

donné, un chercheur peut décider de mettre en place une discipline stricte et contraignante dans son équipe. Dans d'autres cas, il préférera éviter une dimension trop « verticale » et laisser son équipe libre dans ses mouvements. L'idée est la suivante : il existe des outils (critique, discipline) : parfois ils sont adéquat, parfois ils ne le sont pas. Il n'y a pas de règles. Tout est bon.

Comment contrer cet argument ?

C'est impossible. Par contre, on peut créer un ensemble plus vaste dans lequel s'intègre cet argument. Et pour cela, il faut réfléchir en profondeur à ce qu'est la science. C'est l'objet de la partie suivante.

# Chapitre XIX

# Qu'est-ce que la science ?

Je vais développer ma réponse à la question « qu'est-ce que la science ? » en 4 parties. Il m'a fallu tout d'abord ajouter **la notion de récursivité** pour pouvoir distinguer une science profonde (de bas niveau) intégrant l'ensemble des constructions cognitives possibles (y compris celles jugées irrationnelles) et une science superficielle (de haut niveau) correspondant à la science que le positiviste souhaite « vérifiable ». Ensuite j'insisterai à nouveau sur le fait qu'au fondement de la science, il y a **le phénomène interrogatif**. J'en profiterai pour expliciter en quoi ma pensée est relativiste et en quoi elle ne l'est pas. Puis je montrerai que **la science peut être vu comme une guerre** où l'attaqué est la nature, jamais l'homme. Enfin, je montrerai que le scientifique (le guerrier) qui participe à cette guerre est à la recherche de **la grande santé** (une santé incluant paradoxalement la souffrance). Le but de cette guerre, la victoire telle qu'elle est souhaitée, est à rapprocher de l'idéal ascétique.

## 1 La récursivité

Dans le premier chapitre d'épistémologie, nous avons vu qu'il n'y a pas de frontière nette entre science et non science. Nous avons vu également qu'il est irrationnel de vouloir nier le rôle de l'irrationnel dans la dérivée positive du progrès au cours du temps.

Si un chercheur passe sa vie à chercher avec pour but ultime plus ou moins inconscient de plaire à une femme. Son « but » participe indirectement à la méthode scientifique. La construction cognitive déjà présente « plaire à une femme » est incontestablement reliée à la création de nouvelles constructions cognitives (les découvertes) de ce chercheur. Toute définition de la science incapable d'inclure ces éléments (jugés potentiellement irrationnels) est forcement incomplète. Pour pouvoir les inclure, je vais partir de la définition du langage de bas niveau trouvée sur Wikipedia.

*« Un langage de programmation est dit de bas niveau lorsque le codage de celui-ci*

> *se rapproche du langage machine (dit « binaire »), et donc permet de programmer à un degré très avancé. Les langages de bas niveau sont à opposer aux langages de haut niveau. Ces derniers permettent de créer un programme sans tenir compte de la façon dont fonctionne le matériel de l'ordinateur censé exécuter le programme. La définition reste malgré tout assez floue, et il n'existe pas de méthode particulière qui permettrait de déterminer objectivement si un langage peut être considéré de bas niveau. »*[1]

A partir de là, je propose de définir deux types de science [2] :

⇒ La science de bas niveau (au sens informatique : la couche profonde)

⇒ La science de haut niveau (au sens informatique : la couche superficielle)

Sur un plan rectangulaire, ces deux types de science se placent chacune sur un bord opposé (gauche et droite). Cependant, elles ne sont pas séparées par une barrière/démarcation stricte. Au contraire, elles sont reliées par un continuum/un curseur qui indique le niveau de récursivité (figure XIX.1).

La science de bas niveau est créée, composée par l'ensemble des constructions cognitives qui contribuent ou ont contribué au progrès tel que nous le ressentons (pente globalement positive à travers l'histoire). Ces constructions existent en quantité gigantesque. Ce peuvent être, par exemple, les constructions relatives à la magie, aux religions, aux croyances diverses, les constructions à l'origine de la parole, de l'écriture, les constructions issues des sentiments, des passions, des concepts, l'amour, la morale, l'égoïsme, la liberté, etc... En effet, toutes ces constructions cognitives ont potentiellement participé de manière *indirecte* au progrès. Cet « indirect » est capté par le concept de récursivité (au sens de sa racine latine *recurrere* qui signifie « revenir en arrière »). La science de bas niveau place l'homme, son cerveau et ses constructions cognitives comme étant le socle de base nécessaire au progrès. La science de bas niveau correspond donc à l'ensemble des constructions cognitives qui contribuent ou ont contribué au progrès. L'appel à la récursivité est maximum. Sur la figure XIX.1, le curseur est tout à fait à gauche.

La science de haut niveau se définit de manière opposée : c'est la ou les constructions cognitives qui identifient un phénomène naturel, inconnu et reproductible créant de fait une nouvelle construction. L'appel à la récursivité est minimum car il s'agit des constructions cognitives directement en contact avec la nouvelle construction cognitive créée. Autrement dit, ces

[1] Article Wikipedia : Langage de bas niveau

[2] Avant d'expliciter ma pensée, il me faut m'attarder une seconde sur une question d'ordre sémantique. Ici, les concepts de « bas niveau » et « haut niveau » n'ont strictement rien à voir avec les jugements de valeur. La science de « bas niveau/couche profonde » n'est pas la science médiocre. La science de la couche superficielle (haut-niveau) n'est pas « superficielle » au sens populaire.

constructions génèrent une nouvelle construction par le contact avec la nature si et seulement si la nouvelle construction est reproductible *pour soi, en soi*, c'est-à-dire suffisamment stable.

Lorsque l'on transmet, à un autre scientifique, sa nouvelle construction cognitive (sa découverte), on lui fournit en même temps les constructions cognitives jugées nécessaire et suffisante pour sa reproduction c'est-à-dire *la méthode de construction*. Ces scientifiques *vérifient* que la méthode percutant la nature, permet bien de reconstruire la nouvelle construction cognitive : il faut qu'il y ait reproductibilité en eux. De cette reproductibilité dépend la propagation de la découverte (la nouvelle construction cognitive).

Mais, le processus de transmission de la découverte possède une autre caractéristique particulière mais difficile à saisir : la méthode de construction fournie représente une minimisation de l'appel à la récursivité. Toute nouvelle construction cognitive (la découverte) revendique petit à petit une certaine indépendance vis-à-vis de l'édifice cognitif gigantesque qui a été nécessaire pour la construire la première fois.

Ainsi pour expliquer comment à émerger pour la première fois, la théorie de l'évolution dans la tête de Darwin, l'appel à la récursivité est nécessaire. Voyage, passion, sentiment et tous autres concepts de bas niveau sont indispensables pour comprendre *la construction initiale* des énoncés qui constituent aujourd'hui la théorie de l'évolution. Pourtant, si je souhaite vous enseigner aujourd'hui la théorie de l'évolution, je n'ai plus besoin de vous transmettre les éléments de bas niveau (passion, etc...). Seule *la méthode de construction* (minimisant la récursivité) est nécessaire c'est-à-dire les concepts comme la sélection naturelle ou les mutations.

Ainsi en repartant de la définition des langages de bas niveau :

> *« Les langages de bas niveau sont à opposer aux langages de haut niveau. Ces derniers permettent de créer un programme sans tenir compte de la façon dont fonctionne le matériel de l'ordinateur censé exécuter le programme. »*

On peut effectuer la modification suivante (en remplaçant, entre autre, langage par science)

> *« Les sciences de bas niveau sont à opposer aux sciences de haut niveau. Ces dernières permettent de reconstruire cognitivement un énoncé sans tenir compte de la façon dont fonctionne l'ensemble de l'échafaudage cognitif. »*

Au fondement de la science de bas niveau, je tenterai de montrer que l'on trouve principalement **la recherche de la grande santé** (concept Nietzschéen sur lequel je reviendrai) Au fondement de la science de haut niveau, je tenterai de montrer que l'on trouve **le phénomène interrogatif** que nous allons maintenant étudier en détail.

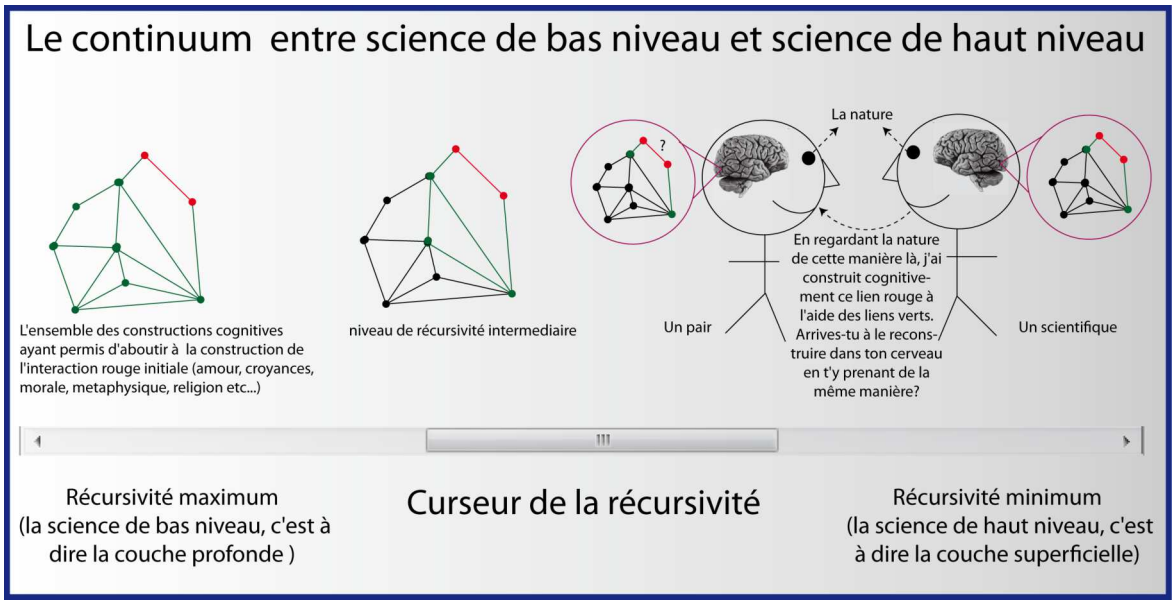



**Figure XIX.1** – Un scientifique utilise des constructions cognitives vertes en observant un phénomène naturel pour construire une construction cognitive rouge. Si il la juge reproductible *en lui*, il tentera de la transmettre à ses pairs en leurs indiquant la méthode c'est-à-dire les liens verts directement en contact du lien rouge. Si ces derniers arrivent à reproduire la construction cognitive rouge grâce aux liens verts, ils la dupliqueront. Bien sûr, il n'est jamais question de verité mais de faisabilité. A l'inverse, la science de bas niveau joue un rôle caché : c'est elle qui, par exemple, a suscité l'envie/ le besoin de construire la construction cognitive rouge. Pour inclure ces mécanismes irrationnels, il suffit de décaler le curseur de la récursivité vers la gauche.

## 2 L'interrogation

Quel sont les éléments qui distinguent la science de haut niveau de la croyance ?

Il me semble que la plupart des épistémologues pensent qu'il existe une asymétrie entre énoncé. Certains énoncés seraient plus « vrais » que d'autres. Certains énoncés seraient plus scientifiques que d'autres. Ceux qui pensent que cette asymétrie n'existe pas sont appelés « relativiste » et sont loin de faire le consensus en épistémologie aujourd'hui.

Le consensus serait plutôt le suivant : il n'existe actuellement aucune démonstration prouvant l'existence d'une asymétrie entre énoncés mais cela ne signifie pas que cette démonstration (et donc l'asymétrie) n'existe pas.

Je pense qu'il n'existe pas d'asymétrie entre énoncés. Pourtant vous verrez que je ne suis pas un relativiste. **L'asymétrie est ailleurs : elle se trouve dans le mode de propagation d'un énoncé.**

Prenons un exemple :

Soit une personne 1 (par exemple un professeur) qui transmet un énoncé A (par exemple la

terre est ronde) à une personne 2 (par exemple un élève). Si l'élève retient la leçon, l'énoncé A s'est dupliqué.

Le professeur fait une affirmation (je rajoute le point d'exclamation pour marquer l'affirmation) :

« A ! » correspond à « La terre est ronde ! »

Et l'élève retient et donc duplique la construction cognitive « A ! »

On passe de 1 « A ! »à 2 « A ! »

**Ce processus de transmission est la croyance.**
Décrivons maintenant le processus scientifique : à ses fondements, on trouve l'interrogation.

Soit une personne 1 (un scientifique) qui transmet un énoncé affirmatif « A ! » (La terre est plate) à une personne 2. Celui-ci répond par un deuxième énoncé « B ? » sous forme interrogative (la terre ne serait-elle pas plutôt ronde ?). A partir de là deux solutions :

Si le scientifique 1 rejoint le scientifique 2, il dira « B ! » (oui effectivement, la terre est ronde). Il y a disparition de A, duplication de B puis transformation de « B ? » en « B ! ».

Si le scientifique 1 ne rejoint pas le scientifique 2, il dira « A ? » (la terre ne serait-elle pas plutôt plate ?). Les deux énoncés perdurent indépendamment tous deux sous forme interrogative c'est-à-dire « A ? »et « B ? ».

**Ce processus de transmission est scientifique.**
La croyance duplique des énoncés affirmatifs :

« ! » engendre « ! ! »

La science duplique des énoncés affirmatifs en insérant l'interrogation entre les deux énoncés :

« ! » engendre « ! ? ! »

Ainsi l'asymétrie ne réside pas *en A ou en B*. Elle réside dans la manière dont A et B se propage. Si la propagation a nécessité le point d'interrogation alors le processus de transmission est scientifique.

L'interrogation nous permet de définir la science. La seule méthode scientifique qui ne peut pas être remise en cause, selon moi, par les relativistes est l'interrogation. Notons que l'interrogation permet également de définir un autre concept : celui d'intelligence. La seule définition incontestable de l'intelligence est l'interrogation. L'intelligence n'est pas définie par la

logique (raisonnement de cause à effet...) ou la sensibilité d'un individu mais par sa capacité à se poser des questions. Il me semble que l'animal ne possède pas l'interrogation et ne possède donc pas l'intelligence. Quand un chien cherche son maître, il ne pose pas la question « ou est mon maître ? » avant de se mettre à chercher. Il remarque une absence et cherche instinctivement[3].

J'invite les anthropologues (s'ils ne l'ont pas déjà fait) à se pencher sur la question suivante de plus prés : *peut-on détecter des signes de l'apparition de l'interrogation au cours de la préhistoire* ?

J'invite les éthologues à ne pas me croire sur parole et à rechercher les signes de l'interrogation chez les animaux. *Des bonobos possèdent-il l'interrogation* ?

En définissant la science avec le processus interrogatif, ce qui est important c'est autant ce que je dis que ce que je ne dis pas : si la science, comme je le crois, n'est pas autre chose que le phénomène interrogatif, alors la science n'a strictement rien à voir avec les notions de vérifiable, de falsifiable ou de critiques.

Je ne suis pas un relativiste car je crois à l'existence d'une asymétrie dans la propagation des énoncés. Pourtant, je crois que :

Si une personne 1 dit à une personne 2 :

« Dieu existe ! »

Et si la personne 2 formule, pour elle-même, l'énoncé sous forme interrogative :

« Dieu existe-il ? »

Puis, si ce scientifique compare l'énoncé avec la nature : par exemple s'il regarde la beauté d'un arc-en-ciel puis en déduit l'énoncé sous forme affirmative :

« Dieu existe ! »

Le processus de transmission est une chaine « ! ? ! ». La présence de l'interrogation entrecoupant les deux affirmations rend le processus de transmission scientifique. La personne 2 est un scientifique pur et dur ! Je sais que cela peut choquer mais c'est la conséquence de ma définition. A l'inverse, si une personne 1 dit à une personne 2 :

« e=mc2 ! »

Et si la personne 2 duplique cet énoncé sans passage par la forme interrogative

« e=mc2 ! »

Le processus de transmission est une chaine « ! ! ». La personne 2 ne s'est pas comportée en scientifique. Le processus de transmission est la croyance.

[3]Je ne suis pas dans la tête d'un chien : il s'agit évidemment d'une pure hypothèse.

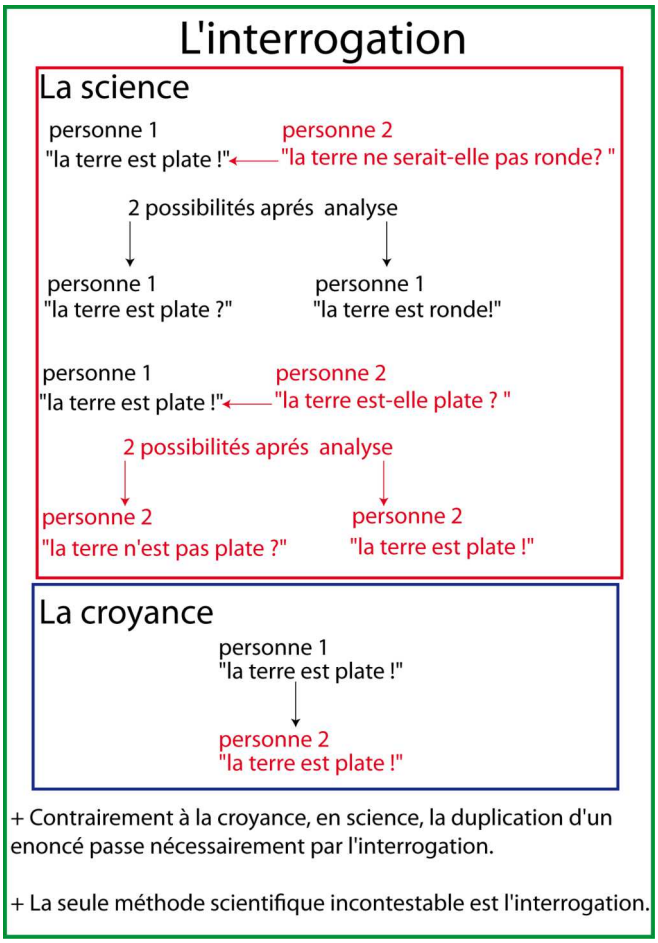



**Figure XIX.2** – Science et Interrogation

# 3 La science est une guerre

La science peut être entrevue comme une guerre. C'est *la volonté de puissance des hommes sur la nature*. Ici, la nature est prise au sens de nature toute puissante que l'on ne comprend pas c'est-à-dire la connaissance future pour l'instant inconnue, le monde qui nous dépasse et nous contrôle.

Dans cette guerre, les hommes sont alliés et regardent vers un but commun. Ce but ne contient aucun élément appartenant à l'autre volonté de puissance : *la volonté de puissance de l'homme sur l'homme*. Violence, critique, compétition, hiérarchie, gloire, pouvoir, confort, reconnaissance, argent, responsabilité ne sont jamais les but ultimes. Au pire, ils peuvent être des moyens ou des instruments. Il existe donc, selon moi, deux volontés de puissance mutuellement exclusive.

Je doute que Nietzsche eu cru à l'existence de deux volontés de puissance distinctes et

mutuellement exclusive inclus dans sa volonté de puissance *à lui*. En d'autres termes, je ne suis pas sûr que « la volonté de puissance de l'homme sur la nature » existe. Et cela pose un vrai problème car si cette volonté de puissance particulière n'existait pas, alors la science ne reflèterai que « la volonté de puissance de l'homme sur l'homme ». Autrement dit, la science ne serait qu'un vulgaire sport, une vulgaire compétition entre les hommes dont la vérifiabilité serait la règle du jeu. Vaincre la nature ne serait plus une fin mais un moyen. Cela signifierait que le scientifique utilise/détourne la science pour maximiser sa puissance personnelle et ce, y compris, si cela est au détriment des autres hommes. La science ne refléterait que la guerre de l'homme sur l'homme avec la nature comme butin bonus. Et cela je me refuse de l'admettre.

Comme la science est une guerre, l'expérience est une attaque qui est lancée pour gagner du territoire sur la nature, sur l'inconnue. Les instruments et méthodes sont des armes. Le scientifique est un guerrier qui effectue une mission (il n'exerce pas une fonction). Quand il réussit une expérience et obtient un résultat significatif, le scientifique-guerrier fait une percée et il blesse la nature car il lui enlève une partie de sa puissance de contrôle. Mais quand il mène des expériences, élabore des théories, il se blesse car il s'acharne. Cette blessure, qui ne le tue pas, le rend évidemment plus fort et vous voyez à qui je fais allusion[4].

Ainsi les percées majeures effectuées par la science sur la nature sont faites par des scientifiques (hommes ou femmes) qui « blessent et sont blessés » et non par des scientifiques (hommes ou femmes) qui « soignent ou sont soignées ». J'ai trouvé chez Nietzsche cette notion de cruauté présente dans la démarche scientifique.

> *« Le simple fait d'étudier sérieusement et à fond est une violence volontaire contre la tendance foncière de l'esprit qui se dirige inlassablement vers l'apparence et la superficie : dans toute volonté de connaitre il entre une goutte de cruauté »*[5]

> *« Il n' y a pas de vérités de grand style qui s'obtiennent par flatterie, il n y' a pas de secrets obtenus par un bavardage familier et séduisant : ce n'est que par violence, par force et par inflexibilité que la nature se laisse arracher ce qu'elle a de plus précieux ; ce n'est que grâce à la brutalité que peuvent s'affirmer dans une morale « de grand style », l'atrocité et la majesté des exigences infinies. Tout ce qui est caché nécessite qu'on ait des mains dures, une intransigeance implacable : sans sincérité il n'y a pas de connaissance [...] ; là où je veux savoir, je veux aussi être sincère, c'est-à-dire dur, sévère, étroit, cruel et inexorable. »* [6]

[4]Nietzsche
[5]Nietzsche, *Par delà bien et mal*, Folio p. 148
[6]Stephan Zweig, *Nietzsche*, Stock p.69

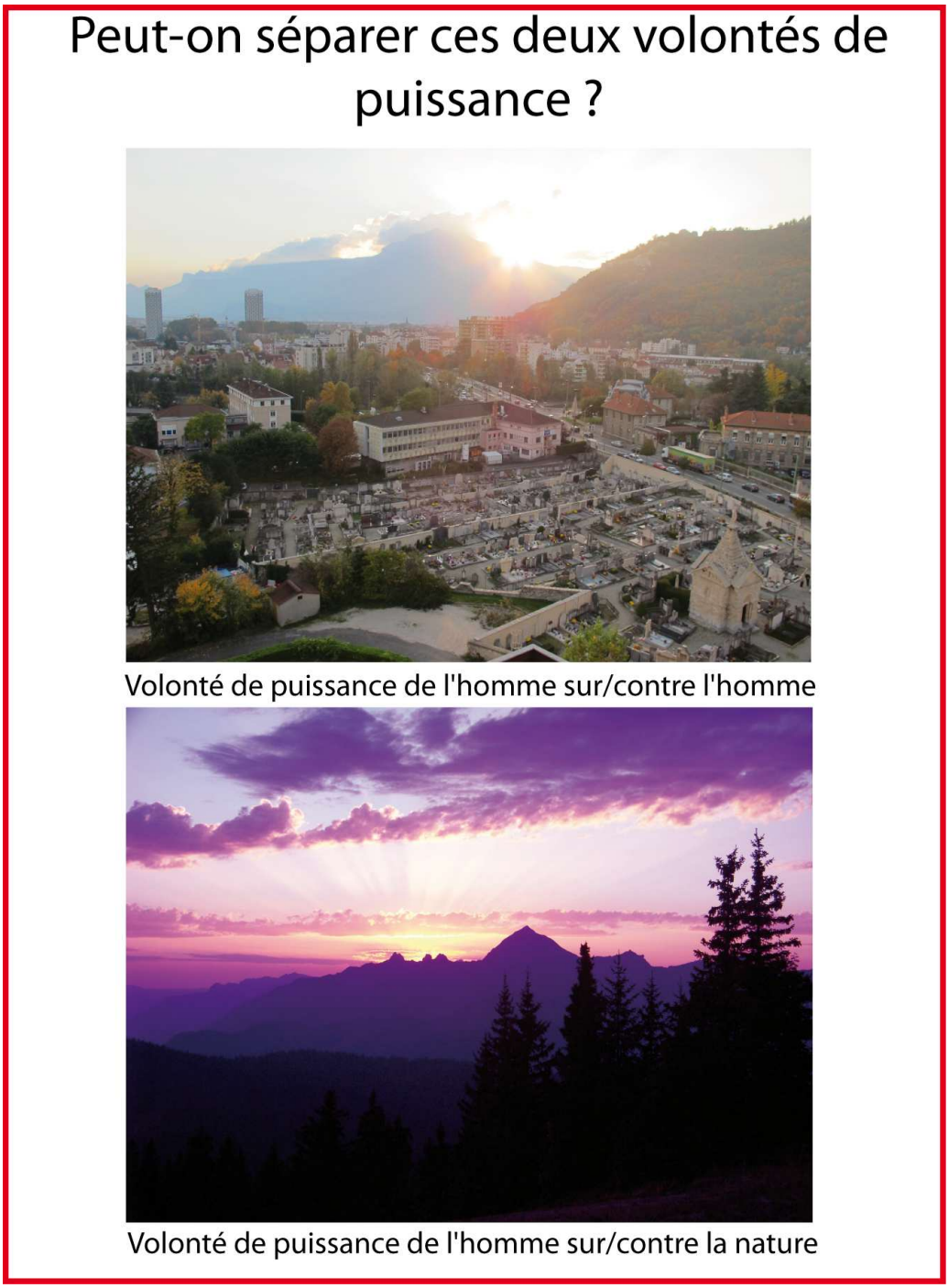



**Figure XIX.3** – La volonté de puissance. Peut-on extraire de la volonté de puissance nietzschéenne, deux entités de volonté de puissance opposées ? L'une serait la volonté de puissance de l'homme sur l'homme symbolisée par la vue de dessus sur la ville : l'homme (le photographe) n'hésiterait pas à dominer et exploiter son prochain pour assouvir sa volonté de puissance. La seconde serait la volonté de puissance de l'homme sur la nature symbolisée par une photo d'un paysage qui révèle notre incompréhension de la nature toute puissante. L'homme cherche à comprendre cette nature pour la contrôler et gagner en puissance.

Bien sûr les scientifique-guerriers qui sont au front ne sont pas seuls. Ils bénéficient de toute la logistique militaire qui se met en place avant la ligne de front : les transmissions, le recrutement, le ravitaillement, le service de santé, la cartographie...Cette logistique nécessaire a été très bien décrite par Bruno Latour :

> *« Les technosciences ont un intérieur parce qu'elles ont un extérieur ; plus la science à l'intérieur est importante, dure et pure, plus les autres chercheurs doivent aller loin vers l'extérieur. L'existence de cet effet retour explique pourquoi on ne voit, quand on se trouve à l'intérieur d'un laboratoire, ni relations publiques, ni politiques, ni problèmes éthiques, ni lutte des classes, ni juristes ; on voit une science isolée de la société. Mais cette isolement n'existe que dans la mesure où d'autres chercheurs consacrent (ou ont consacré) toute leur temps à recruter des investisseurs ou à intéresser et convaincre les gens. Les chercheurs purs ressemblent à des oisillons sans défense qui attendent que les adultes aient fini de construire le nid et leur apportent la becquée [...] La science, à cause du travail de recrutement, a bien un intérieur différent de son extérieur, exactement comme un réacteur nucléaire possède une enceinte de confinement à l'intérieur de laquelle il ne fait pas bon se promener en bras de chemise. Mais de là à dire que l'intérieur de l'enceinte n'est pas dû à la construction à la surveillance à la résistance à l'entretien, à la multiplicité des réglages des concours, des règlements venant de l'extérieur, il y a un pas que franchissent allégrement les épistémologues mais que nous ne hasardons pas. Laissons-les à l'intérieur de leur enceinte et continuons à regarder au contraire ce qui permet à cette muraille de tenir et de confiner »* [7]

Cependant, j'émettrai des réserves sur la métaphore de la centrale nucléaire. Je ne pense pas qu'il soit nécessaire d'opposer un intérieur avec un extérieur et donc de créer une ligne de séparation bien tracée. Je préfère ma métaphore de la guerre avec sa ligne de front constamment alimentée par la logistique militaire des lignes moins avancées. Ici la ligne de front n'est pas opposée au travail logistique : il n'y a pas de séparation nette : on retrouve notre curseur (la récursivité) décrite dans la première partie.

Bruno Latour a également utilisé l'analogie martiale pour décrire le processus scientifique (avec cependant un sens différent du mien) :

> *« C'est seulement maintenant que le lecteur peut comprendre pourquoi j'ai utilisé tant d'expression qui ont des connotations martiales : épreuve de force, controverses, lutte, gagner et perdre, stratégie et tactique, équilibre des pouvoirs, forces, nombres,*

[7] Bruno Latour, *La science en action*, La découverte p.382

> *alliés), expressions qui, bien que constamment utilisés par les chercheurs sont rarement employées par les épistémologues pour décrire le monde pacifique de la science pure [...]On s'aperçoit que la plus grande partie des techno sciences se préoccupe de logistique et de force de frappe. »* [8]

Bruno Latour identifie et décrit bien *la logistique, la force de frappe*. Mais n'oublie t'il pas de décrire *la frappe* en elle-même ? Pourquoi passe-il à coté du *scientifique guerrier* ? De *l'attaque* ? A cause d'une des limites inhérente à son parti pris (l'étude de la science sous le prisme de la sociologie). Bruno Latour s'intéresse à ce qui émerge de la relation entre les agents, pas au fondement de l'agir de ces agents. De plus, il doit (de fait) se cantonner à une attitude descriptive (ce qui est) et non normative (ce qui devrait).

On peut *expliquer* les victoires d'une guerre en décrivant les différentes tactiques et les stratégies employées par les officiers. Mais, on ne peut pas *définir* la guerre avec ces seuls concepts. Il manque l'état de « trans » dans lequel doit se mettre le guerrier pour pouvoir aller égorger l'ennemi ou se faire égorger. Si on ne décrit pas cet état second très particulier, on passe à coté d'un élément essentiel et indispensable de toute guerre. Et la définition reste incomplète.

> *« Un énoncé dans le champ scientifique est autant en danger qu'une balle de rugby. Si aucun joueur ne la reprend, elle reste simplement posée sur le gazon. Pour qu'elle puisse bouger à nouveau, il faut qu'il y ait une action : que quelqu'un la saisisse et la lance ; mais le lancer dépend à son tour de l'agressivité, de la rapidité, de la ruse ou de la tactique des autres. [...] La métaphore du jeu de rugby ne peut plus être maintenue car la balle reste identique à elle-même, alors que, dans la partie de science à laquelle nous assistons, l'objet est soumis à des modifications multiples quand il passe de main en main. Il n'est pas simplement transmis d'un acteur à son voisin par un processus collectif, il est composé collectivement par eux »* [9]

Il est de bonne guerre (désolé pour le jeu de mot) que je vienne modifier un peu les propos de Bruno Latour étant donné ses petits pics envoyés aux épistémologues. Bruno Latour remarquera donc peut être que « sa » balle de rugby a subi quelques ajustements$\breve{}$ : et oui, dans mon discours, son *oisillon sans défense qui reçoit la becquée* s'est transformée en un *guerrier qui égorge et se fait égorger*.

Mais pourquoi la science est-elle une guerre et pas un simple combat individuel ? Car la science est un travail d'équipe. Par là, je veux dire quelque chose de fondamental : le scientifique guerrier a besoin des autres guerriers pour *pouvoir vouloir* attaquer la nature. Il doit pouvoir

[8] Bruno Latour, *La science en action*, La découverte p.415
[9] Bruno Latour, *La science en action*, la découverte p. 251

avoir l'assurance qu'il pourra transmettre les bénéfices de son attaque à ses pairs. Ses pairs doivent pouvoir *capitaliser* sur ses découvertes (et donc aussi sur ses blessures). Ce n'est pas une question d'applaudissements car le guerrier n'en a pas besoin (ou pour être plus exacte, il n'a besoin que de ses propres applaudissements [10]). Le scientifique-guerrier a besoin des autres pour attaquer la nature pour la même raison qui pousse Zarathoustra à redescendre vers les hommes.

> *« Tu n'auras pas de sens, toi le soleil, si ceux que tu éclaires n'existent pas »*

Ainsi pour faire de la science, il faut fondamentalement s'interroger. Et pour s'interroger, il faut nécessairement les autres. Une des conditions nécessaire à l'émergence de l'interrogation est donc, je le crois, le fait que l'homme soit un animal social. Ainsi l'interrogation n'a de sens que si les autres existent. Un homme seul sur terre ne chercherait pas.

# 4 La recherche de la grande santé

La science possède une téléologie, un but, une finalité. En terme martial, cela s'appelle la victoire. Victoire sur la nature. Avènement du surhomme. Connaissance absolue. Accès à « la vérité ». Accès à Dieu. Mais Nietzsche arrive encore à voir plus loin :

> *« Il n'y a pas de doute possible, le véridique, dans ce sens audacieux et ultime que présuppose la croyance à la science,* ***affirme en cela un autre monde que celui de la vie****, de la nature et de l'histoire ; et dans la mesure où il affirme cet autre monde, comment ne doit-il pas par la même — nier son opposé, ce monde, notre monde ? »* [11]

Ainsi la science peut-être comparé à la recherche de l'idéal ascétique.

> *« Une vie ascétique est une contradiction de soi : il y règne un ressentiment sans égal celui d'un instinct insatisfait, d'une volonté de puissance qui voudrait dominer non pas quelque chose dans la vie, mais la vie elle-même, ses conditions majeures les plus profondes, les plus fondamentales »* [12]

> *« « Le degré de danger dans lequel un homme vit avec lui-même » est pour lui la seule mesure valable de toute grandeur. Seul celui qui joue sublimement le tout pour*

[10] Nietzsche, *Le gai savoir*, p. 266
[11] Nietzsche, *Le gai savoir*, p. 287
[12] Nietzsche, *La généalogie de la morale*, p. 138

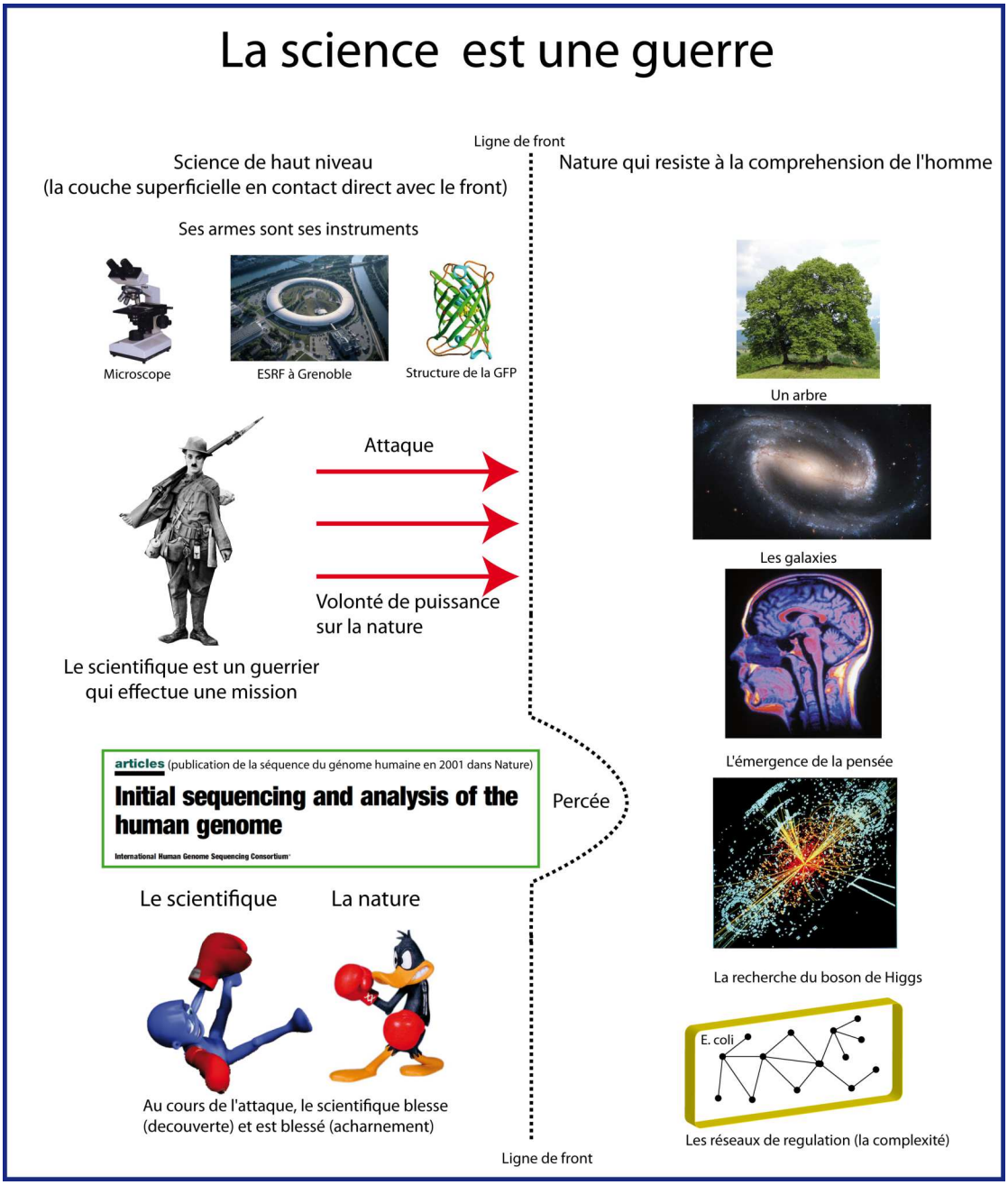



**Figure XIX.4** – La science est une guerre.

*le tout peut gagner l'infini ; seul celui qui risque sa propre vie peut donner à son étroite forme terrestre la valeur de l'infini. »* [13]

Pourtant ce n'est pas le désir de la victoire qui est au fondement de l'agir du scientifique-guerrier. C'est **la recherche** de la vérité, **la traque** de la verité qui génère l'addiction, la passion démoniaque.

*« Ce qui l'excite jusqu'à la souffrance, jusqu'au désespoir ce n'est pas la conquête, ce n'est pas la possession ni la jouissance mais toujours uniquement l'interrogation, la recherche, la chasse. Son amour est incertitude et non pas certitude. [...] Il veut non pas une proie mais simplement l'esprit, le chatouillement et les jouissances de la chasse et des intrigues de la connaissance jusqu'à ses plus hautes et plus lointaines étoiles.[...] La verité n'existe, dans tous les problèmes, que pour un moment et il n'y en a pas où elle existe pour toujours.[...] Il ne conquiert rien pour lui ni pour personne après lui, ni pour un Dieu, ni pour un roi, ni pour une foi mais uniquement pour la joie de la conquête car il ne veut rien posséder rien acquérir rien conquérir »*[14]

Cette traque à la vérité a une allure cyclique. Le scientifique-guerrier consomme des interrogations et des convictions successives selon le schéma suivant (décrit précédemment).

« ! ? ! ! ? ! ! ? ! ! ? ! ! ? ! ! ? ! ! ? ! ! ? ! »

La capacité à endurer un grand nombre de cycles est directement reliée à la tolérance aux blessures et donc à la souffrance.

*« Gai savoir cela veut dire les saturnales d'un esprit qui a résisté patiemment, fermement, froidement, sans s'incliner, mais sans espoir - et qu'envahit soudain l'espoir, l'espoir de la santé, l'ivresse de la guérison [...] et assez souvent je me suis demandé si, somme toute, la philosophie jusqu'à aujourd'hui n'a pas été seulement une interprétation du corps et une mécompréhension du corps [...] Nous ne sommes pas des grenouilles pensantes, des instruments de mesure objective et d'enregistrement aux viscère congelés- nous devons constamment enfanter nos pensées à partir de notre douleur et leur transmettre maternellement tout ce qu'il y a en nous de sang, de coeur, de feu, de plaisir, de passion, de torture, de conscience, de destin, de fatalité[...] Et pour ce qui est de la maladie : ne serions-nous pas presque tentés de demander s'il nous est seulement possible de nous en dispenser ? seule la*

[13]Stephan Zweig, *Nietzsche*, Stock p. 75
[14]Stephan Zweig, *Nietzsche*, Stock p 47-49-50

*grande douleur est l'ultime libératrice de l'esprit, en ce qu'elle est le professeur du grand soupçon [...] seule la grande douleur, cette longue , lente douleur qui prend son temps, dans laquelle nous brulons comme sur du bois vert, nous oblige, nous philosophes, à descendre dans notre ultime profondeur et à nous défaire de toute confiance, de toute bonté d'âme, de tout camouflage, de toute douceur, de tout juste milieu, en quoi nous avons peut être autrefois placé notre humanité. Je doute qu'une telle douleur améliore ; mais je sais qu'elle nous approfondit »*[15]

*« Enfin la grande question demeurerait encore ouverte : celle de savoir si nous pourrions nous passer de la maladie, même pour le développement de notre vertu, et si en particulier notre soif de connaissance et de connaissance de nous même n'aurait pas tout autant besoin de l'âme malade que de l'âme saine : bref si la volonté exclusive de santé ne serait pas un préjugé, une lâcheté et peut être un reste de barbarie et de mentalité arriérée des plus raffinées »* [16]

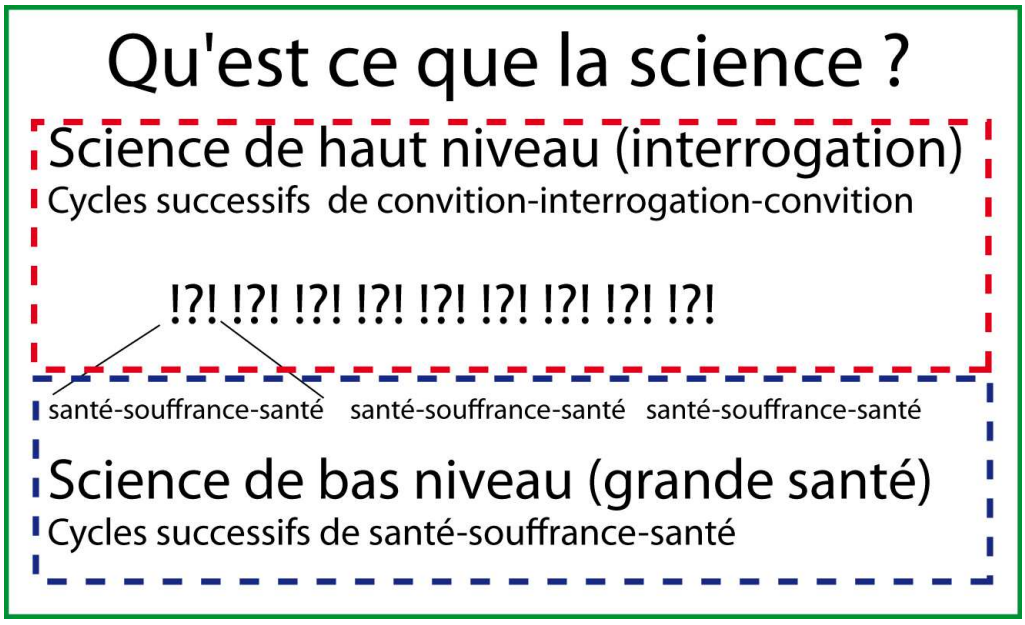



**Figure XIX.5** – Interrogation et grande santé. L'interrogation est la seule méthode scientifique incontestable. La recherche de la grande santé est au fondement de l'agir du scientifique.

Cette souffrance n'est pas constante mais cyclique. Elle est la marque, la cicatrice de la recherche de l**a grande santé** c'est à dire « d'une santé que l'on possède non seulement, mais qu'il faut conquérir sans cesse, puisque sans cesse il faut la sacrifier ! »

[15] Nietzsche, *Le gai savoir*, préface à la seconde édition
[16] Nietzsche, *Le gai savoir*, p. 173

Le scientifique-guerrier enchaine les cycles successifs conviction-interrogation-conviction grâce à sa capacité d'enchaîner les cycles santé-souffrance-santé. Et c'est l'alternance de ces cycles qui est justement la marque de la grande santé. Ainsi les cycles conviction-interrogation-conviction marquent la science de haut niveau : ils sont aux fondements de la méthode scientifique. A l'inverse, les cycles santé-souffrance-santé appartiennent à la science de bas niveau. Ils sont aux fondements de l'agir du scientifique.

# Bibliographie globale des livres


Atlan, Henry. *Entre le cristal et la fumée*. Points edn.

Bachelard, Gaston. *La formation de l'esprit scientifique*. Bibliothèque des textes philosophiques edn.

Bolouri, Hamid. *Computationnal Modeling of Gene Regulatory Networks*. Imperial college press edn.

Canguilhem, Georges. *La connaissance de la vie*. bibliothèque des textes scientifiques edn.

Collins, Harry. *Tout ce que vous devriez savoir sur la science*. Points edn.

Comte-Sponville, André. *Présentation de la philosophie*. le livre de poche edn.

Danchin, Antoine. *La barque de Delphes*. Odile jacob edn.

Dawkins, Richard. *Le gène égoïste*. Odile jacob edn.

Delahaye, Jean-Paul. *Information, complexité et hasard*. Hermes science edn.

Descartes. *Discours de la méthode*. le livre de poche edn.

Einstein, Albert. *Comment je vois le monde*. Flammarion edn.

Ferry, Luc. *Une lecture des trois "Critiques"*. Le livre de poche edn.

Feyerabend, Paul. *Contre la méthode*. Points edn.

Foucault, Michel. *Surveiller et punir*. Gallimard edn.

Gladwell, Malcolm. *Blink*. Back bay books edn.

Gladwell, Malcolm. *Outliers*. Penguin books edn.

Gladwell, Malcolm. *The tipping point*. Abacus edn.

Gleick, James. *La théorie du Chaos*. Champs flammarion edn.

Huxley, Aldous. *Le meilleur des mondes*. Pocket edn.

Jacob, François. *La logique du vivant*. Gallimard edn.

Jonas, Hans. *Pour une éthique du futur*. Payot-rivages edn.

Kauffman, Stuart. *At home in the universe*.

Kierkegaard. *Traité du désespoir*. folio essais edn.

Korn, Henry. *Les menaces biologiques*. Puf edn.

Kuhn, Thomas. *La structure des révolutions scientifiques*. Flammarion edn.

Kupiec, Jean-Jacques, Gandrillon, Olivier, Morange, Michel, & Silberstein, Marc. *Le hasard au cœur de la cellule*. Syllepse edn.

Latour, Bruno. *La science en action*. La découverte edn.

le Moigne, Jean-Louis. *Que sais-je : les épistémologies constructivistes*. Puf edn.

Manent, Pierre. *Cours familier de philosophie politique*. Gallimard edn.

Monod, Jacques. *Le hasard et la nécessité*. Points edn.

Morange, Michel. *La vie expliquée*. Odile jacob edn.

Morange, Michel. *Les secrets du vivant*. La découverte edn.

Morin, Edgar. *La méthode (1. La Nature de la Nature)*. Points edn.

Morin, Edgar. *L'homme et la mort*. Odile jacob edn.

Morin, Edgar. *Science avec conscience*. Point edn.

Nietzsche. *Ainsi parlait Zarathoustra*.

Nietzsche. *Crépuscule des idoles*. Folio edn.

Nietzsche. *Ecce Homo*. Mille est une nuit edn.

Nietzsche. *Humain trop humain*. Folio essais edn.

Nietzsche. *La généalogie de la morale*. Folio edn.

Nietzsche. *Le gai savoir*. Flammarion edn.

Nietzsche. *Par delà le bien et le mal*. folio essais edn.

Ogien, Ruwen. *L'éthique aujourd'hui*. Folio essais edn.

Onfray, Michel, & Roy, Maximilien Le. *Nietzsche (BD)*. Le lombard edn.

Penrose, Roger. *Les deux infinis et l'esprit humain*. Champs flammarion edn.

Pradeu, Thomas. *Chapitre sur la Philosophie de la biologie dans Précis de philosophie des sciences*.

Regis, Ed. *What is life ?* Oxford university press edn.

Rousseau, Jean-Jacques. *Discours sur l'origine et les fondements de l'inégalité parmi les hommes*. Flammarion edn.

Russ, Jacqueline. *La pensée éthique contemporaine*. Puf edn.

Sartre, Jean-Paul. *L'existentialisme est un humanisme*. folio essais edn.

Schopenhauer, Arthur. *Le fondement de la morale*. le livre de poche edn.

Schrödinger, Erwin. *Qu'est ce que la vie ?* Points edn.

Stiegler, Barbara. *Nietzsche et la biologie*. Puf edn.

Thom, René. *Prédire n'est pas expliquer*. Champs edn.

Weber, Max. *Le savant et le politique*. Bibliothèque edn.

Wolfram, Stephen. *A New Kind of Science*. Wolframscience edn.

Zimmer, Carl. *Microcosm*. Pantheon books edn.

Zweig, Stephan. *Nietzsche*. Stock edn.

Zwirn, Hervé. *Les systèmes complexes*. Odile jacob edn.

### Code01ImporteStructureProt.m

```
% Ce petit script permet d'importer la strcuture d'une protéine et de la
% visualiser. Si il ne fonctionne pas, verifier votre connection internet
%ainsi que votre proxy (à parametrer dans matlab)
ubi = getpdb('1ZRC');% recupere la structure 3D
%de Crp en complexe à l'ADN dans Protein Data Bank (PDB) database
h1 = molviewer(ubi);% affiche la molecule
evalrasmolscript(h1,['background white; spin; cartoons on'...
';wireframe off;cpk off;moveto 15 1 0 0 90']);
% paramètre pour un affichage adequate (codé en Rasmol)
```

### Code02SequenceLogo.m

```
% Ce script est un  peu difficile (ne vous cassez pas la tête dessus)
% mais il permet d'identifier comment on stoc des données dans une matrice,
% comment on utilise les fonctions de programmation de base et comment on
% utilise les facilités de plot.Notez qu'il existe une fonction plus
% pratique dans la bioinformatic toolbox pour plotter une sequence-logo


% Fréquence d'apparition des  A
%(sur la première ligne), puis C, G et T sur une
% séquence de 6 pb et un set de 20 sequences
FrequenceApparition= [3 12 2 2 3 0; 5 1 0 11 6 6;...
    2 7 12 6 10 13; 10 0 6 1 1 1]

% passage en probabilité en divisant par 20
ProbabiliteApparition=FrequenceApparition/20

% probabilité de chaque paire de base
%à une position donnée (bruit de fond)
Bruit=[0.1 0.1 0.1 0.1 0.1 0.1 ; 0.4 0.4 0.4 0.4 0.4 0.4;...
    0.3 0.3 0.3 0.3 0.3 0.3 ;0.2 0.2 0.2 0.2 0.2 0.2]

% normalisation des probabilités
%d'apparition en fonction du bruit
ProbabiliteNormalise=ProbabiliteApparition./Bruit

% calcul de l'information et
%donc de la hauteur des lettres
Information=ProbabiliteApparition .*...
    log2(ProbabiliteNormalise)

% on se debarasse des valeurs
%inférieur à 0 et on ordonne les données
[NumBase,Position]=find(Information>0);
HauteurLettre=Information(Information>0);
HauteurLettreOrdonne=[];
BaseOrdonne=[];
for i= 1:6
    [Ordonne,Indice]=sort(HauteurLettre(Position==i));
    Base=NumBase(Position==i);
    HauteurLettreOrdonne=[HauteurLettreOrdonne; Ordonne];
    BaseOrdonne=[BaseOrdonne; Base(Indice)];
end

% parametre de visualisation de la figure
f=figure;
set(f,'Color',[1 1 1])
axis([0 7 -0.1 2])
xlabel('Position','FontSize',20)
ylabel('Contenu en information','FontSize',20)
title('Logo basique','FontSize',30)
hold on

% puis on affiche les lettres
%en fonction de leur contenu en information
ParametrePosition=[0 0 0 0];
for i=1:length(HauteurLettreOrdonne)
    if Position(i)>1 && Position(i)>Position(i-1)
        ParametrePosition(4)=0;
```

```
    end
    switch BaseOrdonne(i)
        case 1
            ValeurBase='A';
        case 2
            ValeurBase='C';
        case 3
            ValeurBase='G';
        case 4
            ValeurBase='T';
    end
    Textlogo=text(Position(i),ParametrePosition(4),...
        ValeurBase, 'HorizontalAlignment'...
        ,'center','VerticalAlignment',...
        'baseline','FontUnits','centimeters'...
        ,'FontSize',7*HauteurLettreOrdonne(i),...
        'Color',[rand(1) rand(1) rand(1)]);
    ParametrePosition=get(Textlogo,'Extent');
end
```

### Code03ActivationSimple.m

```
function ActivationSimple
% tout ce qui est en vert (derrière le pourcentage sont des commentaires
% qui ne sont pas "interprétés".
% on crée une fonction qui contient les equations.
function dydt = equations(t,y)
% crée une matrice de deux ligne et une colonne qu'avec des zeros
% la ligne 1 concentration de A c'est y(1) et ligne 2 celle de B c'est y(2)
    dydt=zeros(2,1);
% on considère la concentration de A constante donc sa derivée est nulle
    dydt(1) = 0;
% la fameuse equation
    dydt(2) = kappa*(y(1)/(kd+y(1)))- gamma*y(2);
end
% on fixe une valeur numerique à nos trois paramètres
kappa=10;
kd=5;
gamma=1;
%y0 c'est le vecteur (=matrice à une ligne) correspondant aux conditions
%initiales. A est à une concentration initiale relative de "6" et B a une
%concentration initiale nulle.
y0=[6, 0];
% fixe les options du solveur numerique. les concentrations ne peuvent pas
% être négative
Options = odeset('NonNegative',[1,2]);
%Vecteur Temps qui va de 0 à 10 par petit pas de 0.1
Time=[0:0.1:10];
% utilise un solveur numerique qui "integre" les equations. Nous
% biologistes nous moquons de comment se solveur fonctionne: il fait le job
% c'est tout. ceux qui veulent plus de detail peuvent taper méthode de
% rugge kutta sur wikipedia. A partir de dydt de la fonction equation, le
% solveur nous renvoie y en fonction de t etant donné la condition initiale
% et les options
[t,y]=ode45(@equations,Time,y0, Options);

% crée une nouvelle figure matlab dans laquel on plot les resultats
figure('Color', 'w')
hold on % permet de superposer les courbes

% trace l'evolution de la concentration de A en fonction du temps
plot(t, y(:,1),'-b','LineWidth',3)
% trace l'evolution de la concentration de B en fonction du temps
plot(t, y(:,2),'-r','LineWidth',3)
% annotation des axes, titre et legende
axis([0 10 0 7])
xlabel('Temps','FontSize',16)
ylabel('Concentrations des protéines','FontSize',16)
title('Une activation simple','FontSize',16)
legend('Facteur de transcription A','Protéine cible B');
end
```

### Code04RegulationAllosterique.m

```
function RegulationAllosterique
function dydt = equations(t,y)
    dydt=zeros(2,1);
    AMPc=kappa*(t>T_AMPc);
    dydt(1) = 0;
    dydt(2) = kappa*(y(1)*AMPc)/(kd2+AMPc)/(kd+(y(1)*AMPc)/(kd2+AMPc))- gamma*y(2);
end
kappa=10;
kd=5;
kd2=5;
gamma=1;
T_AMPc=3;
y0=[6, 0];
Options = odeset('NonNegative',[1,2]);
Time=[0:0.1:10];
[t,y]=ode45(@equations,Time,y0, Options);
figure('Color', 'w')
hold on
plot(t, y(:,1),'-b','LineWidth',3)
plot(t, y(:,2),'-r','LineWidth',3)
plot(t, kappa*(t>T_AMPc),'-g','LineWidth',3);
xlabel('Temps','FontSize',16)
axis([0 10 0 11])
ylabel('Concentrations','FontSize',16)
title('Une regulation allosterique','FontSize',16)
legend('Facteur de transcription Crp','Protéine cible Acs','AMPc(unité relative)');
end
```

### Code05ParametreHill.m

```
function ParametreHill
% l'objectif de cette figure est de montrer coment varie B en fonction de
% l'activateur A quand on fait varier le paramètre de cooperativité n
figure('Color', 'w')
hold on
    function dydt = equations(t,y)
        dydt=zeros(2,1);
        dydt(1) = 0;
        % on ajoute le paramètre de hill n
        dydt(2) = kappa*(y(1).^n/(kd.^n+y(1).^n))- gamma*y(2);
    end
% deux vecteurs qui sont "initialiser"
Y1=[];
Y2=[];
% on fait une boucle for pour tester differents paramètres n
for n=[1,2,4,10,20,300]
    kappa=10;
    kd=5;
    gamma=1;
% on fixe la protéine B et on fait varier la concentration initiale de A
% avec une nouvelle boucle for
    yo(2)=0;
    for i=1:0.2:10
        yo(1)=i;
        Options = odeset('NonNegative',[1,2]);
        Time=[0:0.1:10];
        [t,y]=ode45(@equations,Time,yo, Options);
 % concatenation  dansun vecteur de la valeur finale de B quand on fait
 % varier la condition initale A
        Y1=[Y1 y(end,1)];
        Y2=[Y2 y(end,2)];

        if i==10
            plot(Y1, Y2,'LineWidth',3,'color',[rand(1) rand(1) rand(1)]);
            Y1=[];% delete le vecteur avant le changement de n
            Y2=[];
        end
    end
end
xlabel('Protéine A','FontSize',16);
ylabel('Protéine B','FontSize',16);
title('Variation du paramètre de hill','FontSize',16);
legend('n=1','n=2','n=4','n=10','n=20','n=300');
```

```
end
```

### Code06IdentificationParametre.m

```
% Identification de paramètres à l'aide d'un algorithme genetique.
% ce script necessite la toolbox "genetic algorithm"
function IdentificationParametre

%pour creer le film: decommenter le code film à plusieurs endroits dans le
%script
%crée un objet avi dans lequel on ajoute des frames pour la video
%code film
%mov = avifile('FilmGA','fps',5,'quality',100,'compression','none');

%% "Invention" des données experimentales
kappa1=10;%taux de synthese
kappa2=10;%taux de synthese
kd=7;%constante de dissociation
gamma1=1;%taux de degradation
gamma2=1;%taux de degradation
n=10;%paramètre de hill

y0=[0, 10];%conditions initiales
Options = odeset('NonNegative',[1,2]);
Time=[0:0.3:10];
[t,dydt]=ode45(@equations,Time,y0, Options);%solveur numerique

% on crée une figure dans laquelle on met les données reels qui seront
% recherché par la suite par l'algorithme génétique
f=figure('Color', 'w');
axis([0 8 0 12]);
hold on
plot(t, dydt(:,1),'+b','LineWidth',3);% plot le represseur A
%Notez seul B est "donné" en input à l'algo génétique(données experimtales)
plot(t, dydt(:,2),'+r','LineWidth',3);% plot la concentration de B
text(3.5,9,'Paramètres réels:')
text(4,8,['kappa1: ' num2str(kappa1)])
text(4,7,['kappa2: ' num2str(kappa2)])
text(4,6,['kd: ' num2str(kd)])
text(4,5,['gamma1: ' num2str(gamma1)])
text(4,4,['gamma2: ' num2str(gamma2)])
text(4,3,['n: ' num2str(n)])
xlabel('Temps','FontSize',16)
ylabel('Concentrations des protéines','FontSize',16)
title('Identification des paramètres avec un algorithme génétique',...
    'FontSize',14)
legend('Represseur A','Protéine cible B');
ExpData=dydt; %stoc les données experimentales

%% mise en place de l'optimisation
ParamGene1=[0 0 0 0 0 0];%
LB1=[0 0 0 0 0 0];
UB1=[20 20 20 20 20 20];

GAoptions1 = gaoptimset('Display','off','Generations',150);
it=0; % parametre pour le film (aucune importance)
ParamGeneOptim1 = ga(@optimregGA1,length(ParamGene1),...
    [],[],[],[],LB1,UB1,[],GAoptions1);

kappa1=ParamGeneOptim1(6);
kappa2=ParamGeneOptim1(2);
kd=ParamGeneOptim1(4);
gamma1=ParamGeneOptim1(3);
gamma2=ParamGeneOptim1(5);
n=ParamGeneOptim1(1);

%% plot des resultats finales
y0=[0, 10];
Options = odeset('NonNegative',[1,2]);
Time=[0:0.3:10];
[t,dydt]=ode45(@equations,Time,y0, Options);

plot(t, dydt(:,1),'-b','LineWidth',2)
plot(t, dydt(:,2),'-r','LineWidth',2)
```

```
text(6,9,'Recherche:')
text(6,8,['kappa1: ' num2str(kappa1)])
text(6,7,['kappa2: ' num2str(kappa2)])
text(6,6,['kd: ' num2str(kd)])
text(6,5,['gamma1: ' num2str(gamma1)])
text(6,4,['gamma2: ' num2str(gamma2)])
text(6,3,['n: ' num2str(n)])

%% fonction  utiliser par le GA
    function dydt = equations2(t,y,par)

        kappa1=par(6);
        kappa2=par(2);
        kd=par(4);
        gamma1=par(3);
        gamma2=par(5);
        n=par(1);

        dydt=zeros(2,1);
        dydt(1) = kappa1- gamma1*y(1);
        dydt(2) = kappa2*(kd.^n/(kd.^n+y(1).^n))- gamma2*y(2);

    end
%% fonction utiliser pour générer les données experimentales
    function dydt = equations(t,y)

        dydt=zeros(2,1);
        dydt(1) = kappa1- gamma1*y(1);
        dydt(2) = kappa2*(kd.^n/(kd.^n+y(1).^n))- gamma2*y(2);
    end

%% fonction de la minimisation et de l'affichage
    function Optim=optimregGA1(PG)

        y0=[0, 10];
        Options = odeset('NonNegative',[1,2]);
        Time=[0:0.3:10];
        [t,dydt]=ode45(@(t,y) equations2(t,y,PG),Time,y0,Options);

        p1=plot(t,dydt(:,1),'-b','LineWidth',2);
        p2=plot(t,dydt(:,2),'-r','LineWidth',2);

        t0=text(6,9,'Recherche:');
        t1=text(6,8,['kappa1: ' num2str(kappa1)]);
        t2=text(6,7,['kappa2: ' num2str(kappa2)]);
        t3=text(6,6,['kd: ' num2str(kd)]) ;
        t4=text(6,5,['gamma1: ' num2str(gamma1)]);
        t5=text(6,4,['gamma2: ' num2str(gamma2)]);
        t6=text(6,3,['n: ' num2str(n)]) ;

        it=it+1;
%code film
%         if rem(it,15)==0
%         F=getframe(f);
%         mov = addframe(mov,F);
%         end

        pause(0.001);

        delete(p1);
        delete(p2);
        delete(t0);
        delete(t1);
        delete(t2);
        delete(t3);
        delete(t4);
        delete(t5);
        delete(t6);

        if length(dydt(:,2))==length(ExpData(:,2))
            CLS = dydt(:,2)'-ExpData(:,2)';
            Optim = sum(CLS.^2); % somme des moindre carrés
        else
            Optim=inf;
        end
    end
```

```
%code film
%mov = close(mov);
end
```

### Code07FitElephant.m

```
% algotithme qui montre que l'identification des paramètres peut echouer
%si le nombre de paramètres recherchés est trop grand.
function FitElephant
%pour creer le film: decommenter le code film à plusieurs endroits dans le
%script
%code film
%mov = avifile('FilmElephant','fps',5,'quality',100,'compression','none');

%% "Invention" des données experimentales
kappa1=10;
kappa2=10;
kd=7;
gamma1=1;
gamma2=1;
n=10;

y0=[0, 10];
Options = odeset('NonNegative',[1,2]);
Time=[0:0.3:10];
[t,dydt]=ode45(@equations,Time,y0, Options);

f=figure('Color', 'w');
axis([0 8 0 12]);
hold on

plot(t, dydt(:,1),'+b','LineWidth',3)
plot(t, dydt(:,2),'+r','LineWidth',3)

text(3.5,9,'Paramètres réels:')
text(4,8,['kappa1: ' num2str(kappa1)])
text(4,7,['kappa2: ' num2str(kappa2)])
text(4,6,['kd: ' num2str(kd)])
text(4,5,['gamma1: ' num2str(gamma1)])
text(4,4,['gamma2: ' num2str(gamma2)])
text(4,3,['n: ' num2str(n)])

xlabel('Temps','FontSize',16)
ylabel('Concentrations des protéines','FontSize',16)
title('Identification des paramètres avec un algorithme génétique','FontSize',14)
legend('Represseur A','Protéine cible B');

ExpData=dydt; %stoc les données experimentales

%% mise en place de l'optimisation
ParamGene1=[0 0 0 0 0 0];
LB1=[0 0 0 0 0 0];
UB1=[20 20 20 20 20 20];

GAoptions1 = gaoptimset('Display','off','Generations',150);
it=0; % parametre pour le film (aucune importance)
ParamGeneOptim1 = ga(@optimregGA1,length(ParamGene1),[],[],[],[],LB1,UB1,[],GAoptions1);

kappa1=ParamGeneOptim1(6);
kappa2=ParamGeneOptim1(2);
kd=ParamGeneOptim1(4);
gamma1=ParamGeneOptim1(3);
gamma2=ParamGeneOptim1(5);
n=ParamGeneOptim1(1);

%% plot des resultats finales
y0=[0, 10];
Options = odeset('NonNegative',[1,2]);
Time=[0:0.3:10];
[t,dydt]=ode45(@equations,Time,y0, Options);

plot(t, dydt(:,1),'-b','LineWidth',2)
plot(t, dydt(:,2),'-r','LineWidth',2)
```

```
text(6,9,'Recherche:')
text(6,8,['kappa1: ' num2str(kappa1)])
text(6,7,['kappa2: ' num2str(kappa2)])
text(6,6,['kd: ' num2str(kd)])
text(6,5,['gamma1: ' num2str(gamma1)])
text(6,4,['gamma2: ' num2str(gamma2)])
text(6,3,['n: ' num2str(n)])

%% fonction  utiliser par le GA
    function dydt = equations2(t,y,par)

        kappa1=par(6);
        kappa2=par(2);
        kd=par(4);
        gamma1=par(3);
        gamma2=par(5);
        n=par(1);

        dydt=zeros(2,1);
        dydt(1) = kappa1- gamma1*y(1);
        dydt(2) = kappa2*(kd.^n/(kd.^n+y(1).^n))- gamma2*y(2);

    end
%% fonction utiliser pour générer les données experimentales
    function dydt = equations(t,y)

        dydt=zeros(2,1);
        dydt(1) = kappa1- gamma1*y(1);
        dydt(2) = kappa2*(kd.^n/(kd.^n+y(1).^n))- gamma2*y(2);
    end

%% fonction de la minimisation et de l'affichage
    function Optim=optimregGA1(PG)

        y0=[0, 10];
        Options = odeset('NonNegative',[1,2]);
        Time=[0:0.3:10];
        [t,dydt]=ode45(@(t,y) equations2(t,y,PG),Time,y0,Options);

        p1=plot(t,dydt(:,1),'-b','LineWidth',2);
        p2=plot(t,dydt(:,2),'-r','LineWidth',2);

        t0=text(6,9,'Recherche:');
        t1=text(6,8,['kappa1: ' num2str(kappa1)]);
        t2=text(6,7,['kappa2: ' num2str(kappa2)]);
        t3=text(6,6,['kd: ' num2str(kd)]) ;
        t4=text(6,5,['gamma1: ' num2str(gamma1)]);
        t5=text(6,4,['gamma2: ' num2str(gamma2)]);
        t6=text(6,3,['n: ' num2str(n)]) ;

        it=it+1;
%code film
%         if rem(it,15)==0
%         F=getframe(f);
%         mov = addframe(mov,F);
%         end

        pause(0.001);

        delete(p1);
        delete(p2);
        delete(t0);
        delete(t1);
        delete(t2);
        delete(t3);
        delete(t4);
        delete(t5);
        delete(t6);

        if length(dydt(:,2))==length(ExpData(:,2))
            CLS = dydt(:,2)'-ExpData(:,2)';
            Optim = sum(CLS.^2); % somme des moindre carrés
        else
            Optim=inf;
        end
    end
%code film
```

```
%mov = close(mov);
end
```

### Code08ToggleSwitch.m

```
function ToggleSwitch
clear all
figure('Color', 'w')
hold on
for i=1:4 % genere 4 petit plots et leur attribut un "handle"
    Handle(i)=subplot(2,2,i);
    hold(Handle(i));
end
Vector=[];
VectorPool1=[];
VectorPool2=[];
% système d'équation differentielle avec la repression mutuelle
function dydt = equations(t,y)
    dydt=zeros(2,1);
    dydt(1) = kappa*(kd.^n/(kd.^n+y(2).^n))- gamma*y(1); % A
    dydt(2) = kappa*(kd.^n/(kd.^n+y(1).^n))- gamma*y(2); % B
end
n=5;% paramètre de hill
kappa=10;% taux de synthese
kd=5;% consante de dissociation
gamma=1;% taux de dégradation

y1=5;% conditions initiales de A

it=0; % pour les iterations

for n=5:-0.2:1 % boucle qui fait varier n
     it=it+1;
for gamma=1:0.1:3 % boucle qui fait varier gamma

 for y2=[1 14] % conditions initiales de B
yo=[y1 y2];
    Options = odeset('NonNegative',[1,2]);
Time=[0:0.2:20];
[t,dydt]=ode45(@equations,Time,yo, Options);

%% plot numero 1 (fait varier la condition initiale  de Y et montre que
%% le systeme bascule dans deux états stables differents

if n==5 && gamma==1
    if y2==14 % quand conditions initiales de B plus grandes que celle de A
plot(Handle(1),t(1:25), dydt(1:25,1),'--r','LineWidth',2)
plot(Handle(1),t(1:25), dydt(1:25,2),'--g','LineWidth',2)
    else  % quand conditions initiales de A plus grandes que celle de B
plot(Handle(1),t(1:25), dydt(1:25,1),'r','LineWidth',2)
plot(Handle(1),t(1:25), dydt(1:25,2),'g','LineWidth',2)
    end

xlabel(Handle(1),'Temps','FontSize',16)
ylabel(Handle(1),'Concentrations des protéines','FontSize',16)
title(Handle(1),'Une répression mutuelle','FontSize',16)
legend(Handle(1),'Protéine A (1ère condition initiale)'...
,'Protéine B(1ère condition initiale)'...
,'Protéine A (2ème condition initiale)'...
,'Protéine B(2ème condition initiale)');
end

%% plot numero 2 (une fronce/bifurcation/cusp en anglais)
%on fait varier le parametre de hill n et le taux de degradation gamma et
%on montre que la dynamique du systeme passe de 1 à 2 états stables.
Vector=[Vector dydt(end,1)];

 end
Vector1=max(Vector);
Vector2=min(Vector);
Vector=[];
VectorPool1=[VectorPool1 Vector1];
VectorPool2=[VectorPool2 Vector2];
end
```

```
VectorFin1(it,:)=VectorPool1;
VectorFin2(it,:)=VectorPool2;
VectorPool1=[];
VectorPool2=[];
end
mesh(Handle(2),5:-0.2:1,3:-0.1:1,VectorFin1')
mesh(Handle(2),5:-0.2:1,3:-0.1:1,VectorFin2')
view(Handle(2),-143,17);
zlabel(Handle(2),'concentration de A','FontSize',16)
ylabel(Handle(2),'paramètre de hill','FontSize',16)
xlabel(Handle(2),'taux de dégradation','FontSize',16)
title(Handle(2),'Changement de regime dynamique (bifurcation)','FontSize',16)
%legend(Handle(1),'Protéine A','Protéine B');

%% plot numero 3 et 4: espaces des phases,  nullclines et  champs de
%% vecteurs
for n=[1 3]% compare l'espace des phases pour deux valeurs de n
gamma=1;% refixe gamma;
[a,b]=meshgrid(0:1:12,0:1:12);
dadt = kappa*(kd.^n./(kd.^n+b.^n))- gamma.*a;
dbdt = kappa*(kd.^n./(kd.^n+a.^n))- gamma.*b;
if n==1
quiver(Handle(3),a,b,dadt,dbdt,'LineWidth',1); % trace le champs de vecteur pour n=1
end

if n==3
quiver(Handle(4),a,b,dadt,dbdt,'LineWidth',1);% trace le champs de vecteur pour n=3
end

b=0:1:12;
a=kappa*(kd.^n./(kd.^n+b.^n))./gamma;% nullcline de a (dans ce cas
%précis celui de b est identique); pour me faciliter la tache, je change
%juste l'ordre des absisses et ordonnées dans le plot

if n==1
plot(Handle(3),a,b,'-g','LineWidth',2)
plot(Handle(3),b,a,'-r','LineWidth',2)
xlabel(Handle(3),'A','FontSize',16)
ylabel(Handle(3),'B','FontSize',16)
title(Handle(3),'espace des phases quand n=1','FontSize',16)
legend(Handle(3),'champs de vecteur','nullcline de A','nullcline de B')
end
if n==3
plot(Handle(4),a,b,'-g','LineWidth',2)
plot(Handle(4),b,a,'-r','LineWidth',2)
xlabel(Handle(4),'A','FontSize',16)
ylabel(Handle(4),'B','FontSize',16)
title(Handle(4),'espace des phases quand n=3','FontSize',16)
legend(Handle(4),'champs de vecteur','nullcline de A','nullcline de B')
end
axis(Handle(4),'square')
axis(Handle(4),'tight')
axis(Handle(3),'square')
axis(Handle(3),'tight')
end
end
```

### Code09Hysteresis.m

```
function Hysteresis
%petit fonction qui illustre  l'hysteresis observé avec une boucle de
%retroaction positive

%% gestion de la figure
clear all
figure('Color', 'w')
for i=1:2
    Handle(i)=subplot(2,1,i);
    hold(Handle(i));
end
Condition1=[];
Condition2=[];
%% système d'equations differentielle
% l'IPTG est ajouté par l'experimentateur à t=0 puis enlevé à t=25
```

```
%(centrifugation  des bactéries et resuspension dans du milieu frais).
%Puis on ajoute à nouveau de l'IPTG à t=50.
    function dydt = equations(t,y)
dydt=zeros(4,1);
Iptg_ext=IPTG*(t<25||t>40);% IPTG extracellulaire
dydt(1) = (kappa1*Iptg_ext + kappa2*Iptg_ext*(y(4).^n...
    /(kd3.^n+y(4).^n))- gamma1*y(1));% IPTG intracellulaire
dydt(2) = kappa3-gamma2*y(2);% LacI (non regulé et donc considéré constant)
dydt(3) = kappa4*(kd.^n/(kd.^n+ (y(2)-(y(2)...
    *y(1).^n/(kd2.^n + y(1).^n)))))- gamma3*y(3);%LacZ
dydt(4) = kappa4*(kd.^n/(kd.^n+(y(2)-(y(2)*y(1).^n...
    /(kd2.^n + y(1).^n)))))- gamma4*y(4);% LacY
    end
%% paramètres (totalement arbitraire)
n=10;% paramètre de hill
kappa1=2;% taux d'import de l'IPTG de manière permease independante
kappa2=100;% taux d'import de l'IPTG via LacY
kappa3=30;% taux de synthèse de LacI
kappa4=20;% taux de synthèse de LacZ et LacY

kd=0.87;% constante de dissociation de LacI sur le promoteur de l'operon
%lac
kd2=4;% constante de dissociation de l'IPTG sur LacI
kd3=0.87;% constante de dissociation de l'IPTG sur LacY
gamma1=0.5;% taux d'export de l'IPTG
gamma2=0.5;%taux de degradation de LacI
gamma3=0.5;%taux de degradation de LacZ
gamma4=0.1;%taux de degradation de LacY

%% le code ci dessous montre l'evolution des concentrations en fonction du
%% temps et le graphe d'hysteresis.
%% la gestion des deux graphes en même temps complique un peu le code mais
%% le raccourci

%i=1; la cellule n'a jamais vu l'IPTG ni induit LacY.
%i=2; la cellule "se rappelle" de l'IPTG car elle a encore des LacY
for i=1:2
% on fait varier la concentration d'IPTG extracellulaire
    for IPTG=[0.005:0.005:2]
        yo=[0 60 0 0];% conditions initiales
% lors du deuxieme ajout d'IPTG, les cellules ont la concentration de LacY
%obtenu pour i=1 au temps 40
        if i==2
            yo(4)=y(40,4);
        end
% integration du systeme d'equations differentielles
        Options = odeset('NonNegative',[1 2 3 4]);
        Time=[0:1:60];
        [t,y]=ode45(@equations,Time,yo,Options);
% gestion du graphe des concentrations en fonction du temps
        if IPTG==1 && i==1
            plot(Handle(1),t,y,'LineWidth',3)
            xlabel(Handle(1),'Temps','FontSize',16)
            ylabel(Handle(1),'Concentration','FontSize',16)
            title(Handle(1),'Evolution du système','FontSize',16)
            plot(Handle(1),t,225*IPTG*(t<25|t>40),'color','m',...
                'LineWidth',3)
            legend(Handle(1),'IPTG ic','LacI','LacZ','LacY','IPTG ec');
        end
        if  i==1
            Condition1=[Condition1 y(20,3)];
        end
        if  i==2
            Condition2=[Condition2 y(20,3)];
        end
    end
end
% graphe de l'hysteresis
plot(Handle(2),0.005:0.005:2,Condition1,'-+g','LineWidth',3)
plot(Handle(2),0.005:0.005:2,Condition2,'-+r','LineWidth',3)
xlabel(Handle(2),'[IPTG]','FontSize',16)
ylabel(Handle(2),'[B-galactosidase]','FontSize',16)
title(Handle(2),['Hysteresis induit par une boucle'...
'de retro-action positive'],'FontSize',16);
legend(Handle(2),'Première induction du gène lacZ'...
    ,'Deuxième induction du gène lacZ');
set(Handle(2),'XScale','log')
```

```
axis(Handle(2),[0.005 2 0 40])
end
```

### Code10Oscillation.m

```
function Oscillation

figure('Color', 'w')
hold on
for i=1:4 % genere 4 petit plots et leur attribut un handle
    Handle(i)=subplot(2,2,i);
    hold(Handle(i));
end
% systèmes de 3 equations differentielles
    function dydt=equations(t,y)

        dydt=zeros(3,1);
        dydt(1) = kappa*(kd.^n/(kd.^n+y(3).^n))- gamma*y(1);% LacI
        dydt(2) = kappa*(kd.^n/(kd.^n+y(1).^n))- gamma*y(2);% tetR
        dydt(3) = kappa*(kd.^n/(kd.^n+y(2).^n))- gamma*y(3);%CI

    end
for n=2:3% boucle qui regarde la dynamique du système pour n=2 puis n=3
    kappa=10; %taux de synthèse
    gamma=0.1; %taux de dégradation
    kd=5;% constante de dissociation

    y0=[10, 25, 1];% conditions initiales arbitraires
    Options = odeset('NonNegative',[1,2,3]);
    [t,y]=ode45(@equations,[0:1:400],y0, Options);% integration numerique

    if n==3
        plot(Handle(1),t(1:200), y(1:200,1),'-b','LineWidth',2)
        plot(Handle(1),t(1:200), y(1:200,2),'-g','LineWidth',2)
        plot(Handle(1),t(1:200), y(1:200,3),'-r','LineWidth',2)
        plot(Handle(2),y(:,1),y(:,2),'LineWidth',2)
    end
    if n==2
        plot(Handle(3),t(1:200), y(1:200,1),'-b','LineWidth',2)
        plot(Handle(3),t(1:200), y(1:200,2),'-g','LineWidth',2)
        plot(Handle(3),t(1:200), y(1:200,3),'-r','LineWidth',2)
        plot(Handle(4),y(:,1),y(:,2),'LineWidth',2)
    end
end
xlabel(Handle(1),'Temps','FontSize',16)
ylabel(Handle(1),'Concentrations des protéines','FontSize',16)
title(Handle(1),'Repressilator pour n=3','FontSize',16)
legend(Handle(1),'LacI','TetR','CI');
xlabel(Handle(3),'Temps','FontSize',16)
ylabel(Handle(3),'Concentrations des protéines','FontSize',16)
title(Handle(3),'Repressilator pour n=2','FontSize',16)
legend(Handle(3),'LacI','TetR','CI');

xlabel(Handle(2),'[LacI]','FontSize',16)
ylabel(Handle(2),'[TetR]','FontSize',16)
title(Handle(2),'espace des phases pour n=3','FontSize',16)

xlabel(Handle(4),'[LacI]','FontSize',16)
ylabel(Handle(4),'[TetR]','FontSize',16)
title(Handle(4),'espace des phases pour n=2','FontSize',16)
end
```

### Code11MotifAlon.m

```
function MotifAlon
figure('Color', 'w')
for i=1:3 % genere 3 petit plots et leur attribut un handle
    Handle(i)=subplot(3,1,i);
    hold(Handle(i));
end
```

```
% système d'equation de la figure 1
% boucle de retroaction négative
    function dydt = equations1(t,y)
        dydt=zeros(2,1);
        dydt(1) = kappa-gamma*y(1);
        dydt(2) = kappa2*(kd.^n/(kd.^n+y(2).^n))- gamma2*y(2);
    end
% système d'equation de la figure 2
% feed forward loop coherent
    function dydt = equations2(t,y)
        dydt=zeros(3,1);
        AMPc=AMPC*(t>5&&t<6)||(t>15&&t<25);
        dydt(1) = kappa-gamma*y(1);
        dydt(2) = kappa*(y(1).^n*AMPc.^n)/(kd2.^n+AMPc.^n)/...
(kd.^n+(y(1).^n*AMPc.^n)/(kd2.^n+AMPc.^n))- gamma*y(2);
        dydt(3) = kappa*(y(2).^n/(kd3.^n+y(2).^n))*(y(1).^n*AMPc.^n)...
/(kd2.^n+AMPc.^n)/(kd.^n+(y(1).^n*AMPc.^n)/(kd2.^n+AMPc.^n))- gamma*y(3);
    end
% système d'equation de la figure 3
% feed forward loop incoherent
    function dydt = equations3(t,y)
        dydt=zeros(3,1);
        Signal2=Signal*(t>15);
        dydt(1) = kappa-gamma*y(1);
        dydt(2) = kappa*(y(1).^n*Signal2.^n)/(kd2.^n+Signal2.^n)...
/(kd.^n+(y(1).^n*Signal2.^n)/(kd2.^n+Signal2.^n))- gamma*y(2);
        dydt(3) = kappa*(kd3.^n/(kd3.^n+y(2).^n))*(y(1).^n*Signal2.^n)...
/(kd2.^n+Signal2.^n)/(kd.^n+(y(1).^n*Signal2.^n)/(kd2.^n+Signal2.^n))...
- gamma*y(3);
    end
% 3 boucles pour simuler les 3 systemes et plotter les données
% correspondantes
for i=1:3
    if i==1   % boucle de retroaction négative
        n=2;kappa=10;kappa2=50;kd=5; gamma=1; gamma2=1;
        % vous remarquez que kappa2 est 5 fois plus fort que kappa1
        yo=[0 0];%condition initiale
        Options = odeset('NonNegative',[1,2]);
        Time=[0:0.1:5];
        [t,y]=ode45(@equations1,Time,yo, Options);%integration
        plot(Handle(1),t, y(:,1),'-b','LineWidth',3)
        plot(Handle(1),t, y(:,2),'-r','LineWidth',3)
        xlabel(Handle(1),'Temps','FontSize',14)
        ylabel(Handle(1),'Concentrations','FontSize',14)
        title(Handle(1),'Boucle de retro-action négative','FontSize',16)
        legend(Handle(1),'sans boucle','avec boucle');
        axis(Handle(1),[0 5 0 12])
    end
    if i==2 % feed forward loop coherent
        n=10;kappa=50;kd=5;kd2=5;kd3=50;gamma=1;gamma2=1;AMPC=60;
        yo=[50 0 0];%condition initiale
        Options = odeset('NonNegative',[1,2,3]);
        Time=[0:0.1:32];
        [t,y]=ode45(@equations2,Time,yo, Options);%integration
        plot(Handle(2),t, AMPC*((t>5&t<6)|(t>15&t<25)),'-m','LineWidth',3)
        plot(Handle(2),t, y(:,1),'-b','LineWidth',3)
        plot(Handle(2),t, y(:,2),'-r','LineWidth',3)
        plot(Handle(2),t, y(:,3),'-g','LineWidth',3)
        xlabel(Handle(2),'Temps','FontSize',14)
        ylabel(Handle(2),'Concentrations','FontSize',14)
        title(Handle(2),'Coherent feed forward loop','FontSize',16)
        legend(Handle(2),'AMPc','Crp','araC','araBAD');
    end
    if i==3% feed forward loop incoherent
        n=10; kappa=10;kd=5;kd2=5;kd3=7;gamma=1;gamma2=1;Signal=12;
        yo=[10 0 0];%condition initiale
        Options = odeset('NonNegative',[1,2,3]);
        Time=[0:0.1:32];
        [t,y]=ode45(@equations3,Time,yo, Options);%integration
        plot(Handle(3),t, Signal*(t>15),'-m','LineWidth',5)
        plot(Handle(3),t, y(:,1),'-b','LineWidth',3)
        plot(Handle(3),t, y(:,2),'-r','LineWidth',3)
        plot(Handle(3),t, y(:,3),'-g','LineWidth',3)
        xlabel(Handle(3),'Temps','FontSize',14)
        ylabel(Handle(3),'Concentrations','FontSize',14)
        title(Handle(3),'Incoherent feed forward loop','FontSize',16)
        legend(Handle(3),'Signal','A','B','C');
```

```
        axis(Handle(3),[0 32 0 12])
    end
end
end
```

**Code12AlgorithmeGillespie.m**

```
function GillespieAlgorithm
% Algorithme de gillespie pour un petit modèle simple. ce script se base
% sur des versions de script de van Oudenaarden et Poyatos. Il est
% cependant largement modifié.
% une protéine P est synthétisé à un taux kappa et degradé à un taux gamma
clear all
figure('Color', 'w')
hold on

% cette boucle for reproduit l'algorithme de Gillespie 100 fois
% pour voir un grand de profils de cellules.
for NombreCellule=1:100
k= 25;% taux de synthèse
gamma = 1;% taux de dégradation
P = 0;% nombre de protèines
Tspan{NombreCellule}=[];
Pstochastic{NombreCellule} = P;% evolution de P au cours du temps
tmax = 10;% temps maximum
t = 0;%
tspan = t;

% debut de l'algorithme de Gillespie
while t < tmax
a = [k, gamma*P(1)];
a0 = sum(a);
% determine le temps de la prochaine reaction
r1 = rand;
tau = -log(r1)/a0;
t = t + tau;
% determine le type de reaction
r2 = rand;
acumsum = cumsum(a)/a0;
chosen_reaction = min(find(r2 <= acumsum));
if chosen_reaction == 1;
P(1) = P(1) + 1; %evenement de synthèse
else
P(1) = P(1) - 1;%evenement de degradation
end
tspan = [tspan,t];

Tspan{NombreCellule}=tspan;% stoc les temps
% stoc l'evolution de P dans Pstochastic
Pstochastic{NombreCellule} = [Pstochastic{NombreCellule};P];
end

% Stoc la valeur finale de P à la fin de chaque cinétique
 Distribution(NombreCellule)= Pstochastic{NombreCellule}(end)
end

% equatiion deterministe
P0 = 0;
options = [];
[t P] = ode23(@code1equations,tspan,P0,options,k,gamma);
% plot de la concentration de P en fonction du temps avec les deux modèles
handle1=subplot(3,1,1)
plot(t,P,Tspan{1},Pstochastic{1},'r')
legend('deterministe','stochastique')
xlabel('Temps','Fontsize',12)
ylabel('Nombre de protéines P','Fontsize',12)
title ('Evolution de P dans une cellule','Fontsize',12)

% plot de 100 profils (100 cellules)
handle2=subplot(3,1,2)
hold(handle2)
for i=1:NombreCellule
plot(Tspan{i},Pstochastic{i},'color',[rand  rand rand])
end
```

```
xlabel('Temps','Fontsize',12)
ylabel('Nombre de protéines P','Fontsize',12)
title ('100 Profils de P (100 cellules)','Fontsize',12)

% histogramme de la distribution de P dans les 100 cellules.
handle3=subplot(3,1,3)
hist(Distribution)
xlabel('Nombre de protéines','Fontsize',12)
ylabel('Nombre de cellules','Fontsize',12)
title ('Distribution de P dans 100 cellules','Fontsize',12)

function dPdt = code1equations(t,P,kappa,gamma)
dPdt = [kappa-gamma*P(1)];
end
end
```

### Code13Oscillation3D.m

```
%% E.Coli virtuelle
% Cette bacterie virtuelle a été crée pour augmenter l'experience matlab:
% l'objectif était de faire generer des oscillations basées sur la pulbi de
% elowitz à l'aide du mouvement brownien, le  tout visualisable en 3D.
% pour fabriquer le film, decommenter le code film

%% petit necessité de visualisation
function Oscillation3D
F=figure;
set(F,'color','w')
graph2D =axes('Fontsize', 18,'Units','pixels', 'Position',...
    [100 500-450 400 100]);
axis([0 2000 0 100]);
graph3D =axes( 'Fontsize', 8,'Units','pixels', 'Position',...
    [100 500-330 400 300]);
axis([-1.3 1.8 -1.0 1 -1 1]);
axis off
set([graph2D,graph3D],'Units','normalized')
hold on
grid off
box off

xlabel(graph2D,'Time (minutes)','fontsize',18)
ylabel(graph2D,'Proteins/cell','fontsize',18)

%code film
%mov=avifile('testfilmavrilN11.avi','compression','Cinepak','quality',100);

%% definition des paramètres
temps=2000; % temps

n=1;% nombre de proteine A
m=1;% nombre proteine B
q=1;% nombre de protéine C

vitesse=0.06; % vitesse des molécules dans la bacterie
v1=vitesse*ones(n,1);% vitesse
v2=vitesse*ones(m,1);% vitesse
v3=vitesse*ones(q,1);% vitesse

KappaA=100; % taux de synthèse
KappaB=100;% taux de synthèse
KappaC=100;% taux de synthèse

KappaA_init=100; % taux de synthèse
KappaB_init=100;% taux de synthèse
KappaC_init=100;% taux de synthèse

GammaA=1;% taux de dégradation
GammaB=1; % taux de dégradation
GammaC=1;% taux de dégradation

kaffAsurb=1.98;% compris entre 1 et 2
kaffBsurc=1.98;% compris entre 1 et 2
kaffCsura=1.98;% compris entre 1 et 2
```

```
SyntheseA=0; %alloue de l'espace
DegradationA=0; %alloue de l'espace

SyntheseB=0; %alloue de l'espace
DegradationB=0; %alloue de l'espace

SyntheseC=0; %alloue de l'espace
DegradationC=0; %alloue de l'espace

N=[];% ensemble des valeurs de N
M=[];
Q=[];
I=[];

xc=0;% sert à générer l'ellipsoid (forme coli)
yc=0;
zc=0;
xr=2;
yr=1;
zr=1;
mt=20;

% génére alleatoirement les coordonées de départ des molcules
A=(rand(n,1)-0.5)*2; % nombre aléatoire entre -1 et 1 (je crois)
% voir equation ellipsoide
A=[A(:,1) sqrt(yr.^2*((1-A(:,1).^2/xr.^2))).*(rand(n,1)-0.5)*2   ];
A=[A(:,1) A(:,2) sqrt(zr.^2*((1-(A(:,1).^2/xr.^2)-(A(:,2).^2/yr.^2)))).... 
    *(rand(n,1)-0.5)*2 zeros(n,1) zeros(n,1) zeros(n,1)];

B=(rand(m,1)-0.5)*2;
B=[B(:,1) sqrt(yr.^2*((1-B(:,1).^2/xr.^2))).*(rand(m,1)-0.5)*2   ];
B=[B(:,1) B(:,2) sqrt(zr.^2*((1-(B(:,1).^2/xr.^2)-(B(:,2).^2/yr.^2)))).... 
    *(rand(m,1)-0.5)*2 zeros(m,1) zeros(m,1) zeros(m,1)];

C=(rand(q,1)-0.5)*2;
C=[C(:,1) sqrt(yr.^2*((1-C(:,1).^2/xr.^2))).*(rand(q,1)-0.5)*2   ];
C=[C(:,1) C(:,2) sqrt(zr.^2*((1-(C(:,1).^2/xr.^2)-(C(:,2).^2/yr.^2)))).... 
    *(rand(q,1)-0.5)*2 zeros(q,1) zeros(q,1) zeros(q,1)];

%% simple hélice d'ADN

t=0:pi/50000:2*pi;% nombre de point de l'hélice d'ADN
% plot3((cos(t*1000)/150)+cos(t), sin(t*1000)/150,sin(t),'color','r');%
% plot l'hélice d'ADN mais sans le double superenroulement

plot3((cos(t*100)/50)+cos(t), sin(t*100)/50+cos(t*5000)/500,...
    sin(t)+sin(t*5000)/500,'color','r');

% definit les zones ou seront plotées les bases atcg
promoteurA=0.33*pi:pi/500:0.53*pi;
% affiche sequence codante
text((cos(promoteurA*100)/50)+cos(promoteurA),sin(promoteurA*100)...
    /50+cos(promoteurA*5000)/500,sin(promoteurA)+sin(promoteurA*5000)...
    /500,randsample('ATCG',101,true,[0.5 0.5 0.5 0.5])','Color','r');

promoteurB=1.00*pi:pi/500:1.20*pi;
% affiche sequence codante
text((cos(promoteurB*100)/50)+cos(promoteurB),sin(promoteurB*100)...
    /50+cos(promoteurB*5000)/500,sin(promoteurB)+sin(promoteurB*5000)...
    /500,randsample('ATCG',101,true,[0.5 0.5 0.5 0.5])','Color','g');

promoteurC=1.66*pi:pi/500:1.86*pi;
% affiche sequence codante

text((cos(promoteurC*100)/50)+cos(promoteurC),sin(promoteurC*100)...
    /50+cos(promoteurC*5000)/500,sin(promoteurC)+sin(promoteurC*5000)...
    /500,randsample('ATCG',101,true,[0.5 0.5 0.5 0.5])','Color','b');
%% membrane coli

[x, y, z] = ellipsoid(xc,yc,zc,xr,yr,zr,mt);
surface(x, y, z,'EdgeColor','y','FaceColor','y')
alpha(0.1)%transparence

%% fabrication des cubes
for chaqueN=1:n % boucle de la protéine cubique A

vert1 = [A(chaqueN,1) A(chaqueN,2) A(chaqueN,3);...
```

```
    A(chaqueN,1) A(chaqueN,2)+0.02 A(chaqueN,3);...
    A(chaqueN,1)+0.02 A(chaqueN,2)+0.02 A(chaqueN,3);...
    A(chaqueN,1)+0.02 A(chaqueN,2) A(chaqueN,3) ; ...
        A(chaqueN,1) A(chaqueN,2) A(chaqueN,3)+0.02;...
        A(chaqueN,1) A(chaqueN,2)+0.02 A(chaqueN,3)+0.02;...
        A(chaqueN,1)+0.02 A(chaqueN,2)+0.02 A(chaqueN,3)+0.02;...
        A(chaqueN,1)+0.02 A(chaqueN,2)...
        A(chaqueN,3)+0.02];

fac = [1 2 3 4; ...
    2 6 7 3; ...
    4 3 7 8; ...
    1 5 8 4; ...
    1 2 6 5; ...
    5 6 7 8];
graph1(chaqueN)=patch('Faces',fac,'Vertices',vert1,'FaceColor','r');
chaqueN=chaqueN+1;

end

for chaqueM=1:m % boucle de la protéine cubique B

vert2 = [B(chaqueM,1) B(chaqueM,2) B(chaqueM,3);...
    B(chaqueM,1) B(chaqueM,2)+0.02 B(chaqueM,3);...
    B(chaqueM,1)+0.02 B(chaqueM,2)+0.02 B(chaqueM,3);...
    B(chaqueM,1)+0.02 B(chaqueM,2) B(chaqueM,3) ; ...
        B(chaqueM,1) B(chaqueM,2) B(chaqueM,3)+0.02;...
        B(chaqueM,1) B(chaqueM,2)+0.02 B(chaqueM,3)+0.02;...
        B(chaqueM,1)+0.02 B(chaqueM,2)+0.02 B(chaqueM,3)+0.02;...
        B(chaqueM,1)+0.02 B(chaqueM,2)...
        B(chaqueM,3)+0.02];

fac = [1 2 3 4; ...
    2 6 7 3; ...
    4 3 7 8; ...
    1 5 8 4; ...
    1 2 6 5; ...
    5 6 7 8];

graph2(chaqueM)=patch('Faces',fac,'Vertices',vert2,'FaceColor','g');
chaqueM=chaqueM+1;

end

for chaqueQ=1:q % boucle de la protéine cubique C


vert3 = [C(chaqueQ,1) C(chaqueQ,2) C(chaqueQ,3);...
    C(chaqueQ,1) C(chaqueQ,2)+0.02 C(chaqueQ,3);...
    C(chaqueQ,1)+0.02 C(chaqueQ,2)+0.02 C(chaqueQ,3);...
    C(chaqueQ,1)+0.02 C(chaqueQ,2) C(chaqueQ,3) ; ...
        C(chaqueQ,1) C(chaqueQ,2) C(chaqueQ,3)+0.02;...
        C(chaqueQ,1) C(chaqueQ,2)+0.02 C(chaqueQ,3)+0.02;...
        C(chaqueQ,1)+0.02 C(chaqueQ,2)+0.02 C(chaqueQ,3)+0.02;...
        C(chaqueQ,1)+0.02 C(chaqueQ,2)...
        C(chaqueQ,3)+0.02];

fac = [1 2 3 4; ...
    2 6 7 3; ...
    4 3 7 8; ...
    1 5 8 4; ...
    1 2 6 5; ...
    5 6 7 8];

graph3(chaqueQ)=patch('Faces',fac,'Vertices',vert3,'FaceColor','b');
chaqueQ=chaqueQ+1;

end

%%  creation de la vie bacterienne !!
for i=1:temps % definit le temps

    A(:,4)=(rand(n,1)-0.5)*2;% choisit une variation aléatoire sur x
    A(:,5)=(rand(n,1)-0.5)*2;% choisit une variation aléatoire sur y
    A(:,6)=(rand(n,1)-0.5)*2;% choisit une variation aléatoire sur z

    A(:,1)=A(:,1)+v1(:,1).*A(:,4);% equation du mouvement brownien
```

```
    A(:,2)=A(:,2)+v1(:,1).*A(:,5);
    A(:,3)=A(:,3)+v1(:,1).*A(:,6);

    B(:,4)=(rand(m,1)-0.5)*2;
    B(:,5)=(rand(m,1)-0.5)*2;
    B(:,6)=(rand(m,1)-0.5)*2;

    B(:,1)=B(:,1)+v2(:,1).*B(:,4);
    B(:,2)=B(:,2)+v2(:,1).*B(:,5);
    B(:,3)=B(:,3)+v2(:,1).*B(:,6);

    C(:,4)=(rand(q,1)-0.5)*2;
    C(:,5)=(rand(q,1)-0.5)*2;
    C(:,6)=(rand(q,1)-0.5)*2;

    C(:,1)=C(:,1)+v3(:,1).*C(:,4);
    C(:,2)=C(:,2)+v3(:,1).*C(:,5);
    C(:,3)=C(:,3)+v3(:,1).*C(:,6);

%% rester dans le cocon familial elipsoidal
chaqueN=1;
for chaqueN=1:n

if (A(chaqueN,1)-xc).^2/xr.^2+(A(chaqueN,2)-yc).^2/yr.^2+...
        (A(chaqueN,3)-zc).^2/zr.^2>1 % voir equation ellipsoid

    A(chaqueN,:)=[A(chaqueN,1)-v1(chaqueN,1)...
        *A(chaqueN,4) A(chaqueN,2)-v1(chaqueN,1)...
        *A(chaqueN,5) A(chaqueN,3)-v1(chaqueN,1)*A(chaqueN,6) 0 0 0];
end
chaqueN=chaqueN+1;
end

chaqueM=1;
for chaqueM=1:m

if (B(chaqueM,1)-xc).^2/xr.^2+(B(chaqueM,2)-yc).^2/yr.^2+...
        (B(chaqueM,3)-zc).^2/zr.^2>1

    B(chaqueM,:)=[B(chaqueM,1)-v2(chaqueM,1)...
        *B(chaqueM,4) B(chaqueM,2)-v2(chaqueM,1)...
        *B(chaqueM,5) B(chaqueM,3)-v2(chaqueM,1)*B(chaqueM,6) 0 0 0];
end
chaqueM=chaqueM+1;
end

chaqueQ=1;
for chaqueQ=1:q

if (C(chaqueQ,1)-xc).^2/xr.^2+(C(chaqueQ,2)-yc).^2/yr.^2+...
        (C(chaqueQ,3)-zc).^2/zr.^2>1

    C(chaqueQ,:)=[C(chaqueQ,1)-v3(chaqueQ,1)...
        *C(chaqueQ,4) C(chaqueQ,2)-v3(chaqueQ,1)...
        *C(chaqueQ,5) C(chaqueQ,3)-v3(chaqueQ,1)*C(chaqueQ,6) 0 0 0];
end
chaqueQ=chaqueQ+1;
end

%% synthese et degradatation

    SyntheseA=SyntheseA+KappaA;
    if SyntheseA>100 % definit la manière dont se passe la synthèse
        n=n+1;
% cree une nouvelle protéine à l'interieur de l'ellipsoid
    A(n,1)=(rand(1)-0.5)*2;
    A(n,2)=sqrt(yr.^2*((1-A(n,1).^2/xr.^2))).*(rand(1)-0.5)*2 ;
    A(n,:)=[A(n,1) A(n,2) sqrt(zr.^2*((1-(A(n,1).^2/xr.^2)-...
        (A(n,2).^2/yr.^2)))).*(rand(1)-0.5)*2 0 0 0];


        vert1 = [A(n,1) A(n,2) A(n,3); A(n,1) A(n,2)+0.02 A(n,3);...
            A(n,1)+0.02 A(n,2)+0.02 A(n,3); A(n,1)+0.02 A(n,2) A(n,3) ; ...
        A(n,1) A(n,2) A(n,3)+0.02;A(n,1) A(n,2)+0.02 A(n,3)+0.02;...
        A(n,1)+0.02 A(n,2)+0.02 A(n,3)+0.02;A(n,1)+0.02 A(n,2)...
        A(n,3)+0.02];
```

```
    graph1(n)=patch('Faces',fac,'Vertices',vert1,'FaceColor','r');
        SyntheseA=0;
        v1=[v1; vitesse] ;
    end
% equation correcte pour la degradation
    DegradationA=DegradationA+GammaA*n;
    if DegradationA>100 % definit la manière dont se passe la dégradation
        if v1(n,:)==0
            KappaB=KappaB_init;
        v1(n,:)=[];
        A(n,:)=[];
        delete (graph1(n))
        n=n-1;
        DegradationA=0;
        else
        v1(n,:)=[];
        A(n,:)=[];
        delete (graph1(n))
        n=n-1;
        DegradationA=0;
        end
    end

    SyntheseB=SyntheseB+KappaB;
    if SyntheseB>100 % definit la manière dont se passe la synthèse
        m=m+1;

        B(m,1)=(rand(1)-0.5)*2;
        B(m,2)=sqrt(yr.^2*((1-B(m,1).^2/xr.^2))).*(rand(1)-0.5)*2 ;
        B(m,:)=[B(m,1) B(m,2) sqrt(zr.^2*((1-(B(m,1).^2/xr.^2)-...
            (B(m,2).^2/yr.^2)))).*(rand(1)-0.5)*2 0 0 0];

        vert2 = [B(m,1) B(m,2) B(m,3); B(m,1) B(m,2)+0.02 B(m,3);...
            B(m,1)+0.02 B(m,2)+0.02 B(m,3); B(m,1)+0.02 B(m,2) B(m,3) ; ...
        B(m,1) B(m,2) B(m,3)+0.02;B(m,1) B(m,2)+0.02 B(m,3)+0.02;...
        B(m,1)+0.02 B(m,2)+0.02 B(m,3)+0.02;B(m,1)+0.02 B(m,2)...
        B(m,3)+0.02];

    graph2(m)=patch('Faces',fac,'Vertices',vert2,'FaceColor','g');
        SyntheseB=0;
        v2=[v2; vitesse] ;
    end

    DegradationB=DegradationB+GammaB*m;
    if DegradationB>100 % definit la manière dont se passe la dégradation
        if v2(m,:)==0
             KappaC=KappaC_init;
            v2(m,:)=[];
        B(m,:)=[];
        delete (graph2(m))
        m=m-1;
        DegradationB=0;
        else
        v2(m,:)=[];
        B(m,:)=[];
        delete (graph2(m))
        m=m-1;
        DegradationB=0;
        end
    end

    SyntheseC=SyntheseC+KappaC;
    if SyntheseC>100 % definit la manière dont se passe la synthèse
        q=q+1;

        C(q,1)=(rand(1)-0.5)*2;
        C(q,2)=sqrt(yr.^2*((1-C(q,1).^2/xr.^2))).*(rand(1)-0.5)*2 ;
        C(q,:)=[C(q,1) C(q,2) sqrt(zr.^2*((1-(C(q,1).^2/xr.^2)...
            -(C(q,2).^2/yr.^2)))).*(rand(1)-0.5)*2 0 0 0];


        vert3 = [C(q,1) C(q,2) C(q,3); C(q,1) C(q,2)+0.02 C(q,3);...
            C(q,1)+0.02 C(q,2)+0.02 C(q,3); C(q,1)+0.02 C(q,2) C(q,3) ; ...
        C(q,1) C(q,2) C(q,3)+0.02;C(q,1) C(q,2)+0.02 C(q,3)+0.02;...
        C(q,1)+0.02 C(q,2)+0.02 C(q,3)+0.02;C(q,1)+0.02 C(q,2)...
        C(q,3)+0.02];
```

```
    graph3(q)=patch('Faces',fac,'Vertices',vert3,'FaceColor','b');
        SyntheseC=0;
        v3=[v3; vitesse] ;
    end

    DegradationC=DegradationC+GammaC*q;
    if DegradationC>100 % definit la manière dont se passe la dégradation
       if  v3(q,:)==0
           KappaA=KappaA_init;
           v3(q,:)=[];
        C(q,:)=[];
        delete (graph3(q))
        q=q-1;
        DegradationC=0;
       else
        v3(q,:)=[];
        C(q,:)=[];
        delete (graph3(q))
        q=q-1;
        DegradationC=0;
       end
    end

%% fixation à son promoteur préféré: cette fixation n'est pas parfaite

    chaqueN=1;
        for chaqueN=1:n

    if A(chaqueN,1)<cos(pi*1000)/150+cos(pi)+0.025 &&...
            A(chaqueN,1)>cos(pi*1000)/150+cos(pi)-0.2 &&...
       A(chaqueN,2)<sin(pi*1000)/150+0.025 &&...
       A(chaqueN,2)>sin(pi*1000)/150-0.2 &&...
       A(chaqueN,3)<sin(pi) && A(chaqueN,3)>sin(pi)-0.2 ;

     v1(chaqueN,1)=0;
    KappaB=GammaB;

       LienAsurb=rand(1)/kaffAsurb;
       if LienAsurb>0.5
           v1(chaqueN,1)=vitesse;
           KappaB=KappaB_init;
       end

    end
     chaqueN=chaqueN+1;
        end

       chaqueM=1 ;
      for chaqueM=1:m

    if B(chaqueM,1)<cos(1.66*pi*1000)/150+cos(1.66*pi)+0.025 &&...
            B(chaqueM,1)>cos(1.66*pi*1000)/150+cos(1.66*pi)-0.2 &&...
       B(chaqueM,2)<sin(1.66*pi*1000)/150+0.025 &&...
       B(chaqueM,2)>sin(1.66*pi*1000)/150-0.2 &&...
       B(chaqueM,3)<sin(1.66*pi) && B(chaqueM,3)>sin(1.66*pi)-0.2 ;

     v2(chaqueM,1)=0;
    KappaC=GammaC;

     LienBsurc=rand(1)/kaffBsurc;
       if LienBsurc>0.5
           v2(chaqueM,1)=vitesse;
           KappaC=KappaC_init;
       end

    end
     chaqueM=chaqueM+1;
      end

      chaqueQ=1 ;
        for chaqueQ=1:q

    if C(chaqueQ,1)<cos(0.33*pi*1000)/150+cos(0.33*pi)+0.025 &&...
            C(chaqueQ,1)>cos(0.33*pi*1000)/150+cos(0.33*pi)-0.2 &&...
       C(chaqueQ,2)<sin(0.33*pi*1000)/150+0.025 &&...
       C(chaqueQ,2)>sin(0.33*pi*1000)/150-0.2 &&...
       C(chaqueQ,3)<sin(0.33*pi) && C(chaqueQ,3)>sin(0.33*pi)-0.2 ;
```

```
     v3(chaqueQ,1)=0;
    KappaA=GammaA;

     LienCsura=rand(1)/kaffCsura;
       if LienCsura>0.5
           v3(chaqueQ,1)=vitesse;
           KappaA=KappaA_init;
       end

    end
     chaqueQ=chaqueQ+1;
        end

%% visualisation à chaque itération

chaqueN=1;
for chaqueN=1:n
set(graph1(chaqueN),'XData',[A(chaqueN,1) A(chaqueN,1) ...
    A(chaqueN,1)+0.02 A(chaqueN,1)  A(chaqueN,1)  A(chaqueN,1);...
   A(chaqueN,1) A(chaqueN,1) A(chaqueN,1)+0.02 A(chaqueN,1)...
   A(chaqueN,1)  A(chaqueN,1)  ;...
   A(chaqueN,1)+0.02    A(chaqueN,1)+0.02...
   A(chaqueN,1)+0.02 A(chaqueN,1)+0.02...
   A(chaqueN,1) A(chaqueN,1)+0.02 ;...
   A(chaqueN,1)+0.02    A(chaqueN,1)+0.02...
   A(chaqueN,1)+0.02 A(chaqueN,1)+0.02...
   A(chaqueN,1) A(chaqueN,1)+0.02],...
'YData',[A(chaqueN,2) A(chaqueN,2)+0.02...
A(chaqueN,2) A(chaqueN,2)  A(chaqueN,2)  A(chaqueN,2)   ;...
   A(chaqueN,2)+0.02 A(chaqueN,2)+0.02...
   A(chaqueN,2)+0.02 A(chaqueN,2)  A(chaqueN,2)+0.02...
   A(chaqueN,2)+0.02  ;...
   A(chaqueN,2)+0.02    A(chaqueN,2)+0.02 A(chaqueN,2)+0.02...
   A(chaqueN,2) A(chaqueN,2)+0.02 A(chaqueN,2)+0.02;...
   A(chaqueN,2)   A(chaqueN,2)+0.02...
   A(chaqueN,2) A(chaqueN,2) A(chaqueN,2) A(chaqueN,2)],...
'ZData',[A(chaqueN,3) A(chaqueN,3) A(chaqueN,3)...
A(chaqueN,3)  A(chaqueN,3)  A(chaqueN,3)+0.02   ;...
   A(chaqueN,3) A(chaqueN,3)+0.02...
   A(chaqueN,3) A(chaqueN,3)+0.02  A(chaqueN,3)  A(chaqueN,3)+0.02  ;...
   A(chaqueN,3)    A(chaqueN,3)+0.02...
   A(chaqueN,3)+0.02 A(chaqueN,3)+0.02...
   A(chaqueN,3)+0.02 A(chaqueN,3)+0.02;...
   A(chaqueN,3)   A(chaqueN,3) A(chaqueN,3)+0.02...
   A(chaqueN,3) A(chaqueN,3)+0.02 A(chaqueN,3)+0.02])
chaqueN=chaqueN+1;
end



for chaqueM=1:m
set(graph2(chaqueM),'XData',[B(chaqueM,1) B(chaqueM,1) ...
    B(chaqueM,1)+0.02 B(chaqueM,1)  B(chaqueM,1)  B(chaqueM,1);...
   B(chaqueM,1) B(chaqueM,1) B(chaqueM,1)+0.02 B(chaqueM,1)...
   B(chaqueM,1)  B(chaqueM,1)  ;...
   B(chaqueM,1)+0.02    B(chaqueM,1)+0.02...
   B(chaqueM,1)+0.02 B(chaqueM,1)+0.02...
   B(chaqueM,1) B(chaqueM,1)+0.02 ;...
   B(chaqueM,1)+0.02    B(chaqueM,1)+0.02...
   B(chaqueM,1)+0.02 B(chaqueM,1)+0.02...
   B(chaqueM,1) B(chaqueM,1)+0.02],...
'YData',[B(chaqueM,2) B(chaqueM,2)+0.02...
B(chaqueM,2) B(chaqueM,2)  B(chaqueM,2)  B(chaqueM,2)   ;...
   B(chaqueM,2)+0.02 B(chaqueM,2)+0.02...
   B(chaqueM,2)+0.02 B(chaqueM,2)  B(chaqueM,2)+0.02...
   B(chaqueM,2)+0.02  ;...
   B(chaqueM,2)+0.02    B(chaqueM,2)+0.02 B(chaqueM,2)+0.02...
   B(chaqueM,2) B(chaqueM,2)+0.02 B(chaqueM,2)+0.02;...
   B(chaqueM,2)   B(chaqueM,2)+0.02...
   B(chaqueM,2) B(chaqueM,2) B(chaqueM,2) B(chaqueM,2)],...
'ZData',[B(chaqueM,3) B(chaqueM,3) B(chaqueM,3)...
B(chaqueM,3)  B(chaqueM,3)  B(chaqueM,3)+0.02   ;...
   B(chaqueM,3) B(chaqueM,3)+0.02...
   B(chaqueM,3) B(chaqueM,3)+0.02  B(chaqueM,3)  B(chaqueM,3)+0.02  ;...
   B(chaqueM,3)    B(chaqueM,3)+0.02...
   B(chaqueM,3)+0.02 B(chaqueM,3)+0.02...
```

```
   B(chaqueM,3)+0.02 B(chaqueM,3)+0.02;...
   B(chaqueM,3)   B(chaqueM,3) B(chaqueM,3)+0.02...
   B(chaqueM,3) B(chaqueM,3)+0.02 B(chaqueM,3)+0.02])
chaqueM=chaqueM+1;
end

for chaqueQ=1:q
set(graph3(chaqueQ),'XData',[C(chaqueQ,1) C(chaqueQ,1) ...
    C(chaqueQ,1)+0.02 C(chaqueQ,1)  C(chaqueQ,1)  C(chaqueQ,1);...
   C(chaqueQ,1) C(chaqueQ,1) C(chaqueQ,1)+0.02 C(chaqueQ,1)...
   C(chaqueQ,1)  C(chaqueQ,1)  ;...
   C(chaqueQ,1)+0.02    C(chaqueQ,1)+0.02...
   C(chaqueQ,1)+0.02 C(chaqueQ,1)+0.02...
   C(chaqueQ,1) C(chaqueQ,1)+0.02 ;...
   C(chaqueQ,1)+0.02    C(chaqueQ,1)+0.02...
   C(chaqueQ,1)+0.02 C(chaqueQ,1)+0.02...
   C(chaqueQ,1) C(chaqueQ,1)+0.02],...
'YData',[C(chaqueQ,2) C(chaqueQ,2)+0.02...
C(chaqueQ,2) C(chaqueQ,2)  C(chaqueQ,2)  C(chaqueQ,2)   ;...
   C(chaqueQ,2)+0.02 C(chaqueQ,2)+0.02...
   C(chaqueQ,2)+0.02 C(chaqueQ,2)  C(chaqueQ,2)+0.02...
   C(chaqueQ,2)+0.02  ;...
   C(chaqueQ,2)+0.02    C(chaqueQ,2)+0.02 C(chaqueQ,2)+0.02...
   C(chaqueQ,2) C(chaqueQ,2)+0.02 C(chaqueQ,2)+0.02;...
   C(chaqueQ,2)   C(chaqueQ,2)+0.02...
   C(chaqueQ,2) C(chaqueQ,2) C(chaqueQ,2) C(chaqueQ,2)],...
'ZData',[C(chaqueQ,3) C(chaqueQ,3) C(chaqueQ,3)...
C(chaqueQ,3)  C(chaqueQ,3)  C(chaqueQ,3)+0.02   ;...
   C(chaqueQ,3) C(chaqueQ,3)+0.02...
   C(chaqueQ,3) C(chaqueQ,3)+0.02  C(chaqueQ,3)  C(chaqueQ,3)+0.02  ;...
   C(chaqueQ,3)    C(chaqueQ,3)+0.02...
   C(chaqueQ,3)+0.02 C(chaqueQ,3)+0.02...
   C(chaqueQ,3)+0.02 C(chaqueQ,3)+0.02;...
   C(chaqueQ,3)   C(chaqueQ,3) C(chaqueQ,3)+0.02...
   C(chaqueQ,3) C(chaqueQ,3)+0.02 C(chaqueQ,3)+0.02])
chaqueQ=chaqueQ+1;
end
drawnow
% code film
%mov = addframe(mov,F);
    i=i+1;

%% petit graphe des concentrations

    N=[N n];
    M=[M m];
    Q=[Q q];
    I=[I i];

hold (graph2D,'on' )
graph4=plot(I,N,'-r','parent',graph2D);
graph5=plot(I,M,'-g','parent',graph2D);
graph6=plot(I,Q,'-b','parent',graph2D);

end
% code film
% mov = close(mov);
end
```

**Code14VirusOndeParticule.m**

```
% script qui gènère la petit animation des virus projetés sur les fentes
% d'young et generant un paterne d'interferences sur l'ecran detecteur

%Film = avifile('FilmYoung','fps',15,'quality',100,'compression','none');
f=figure('Color', 'w');
box off
axis([0 1000 0 1000 0 1000]);
axis off

% paramètre de la première plaque rouge
X=[500 510 510 500 500 500; 510 510 500 500 510 510;...
    510 510 500 500 510 510; 500 510 510 500 500 500 ];
Y=[0 0 400 400 0 0;0 400 400 0 0 0;0 400 400 0 400 400;...
```

```
    0 0 400 400 400 400];
Z=[0 0 0 0 0 1000;0 0 0 0 0 1000; 1000 1000 1000 1000 0 1000;...
    1000 1000 1000 1000 0 1000];

% paramètre de la deuxième plaque rouge
X2=[500 510 510 500 500 500; 510 510 500 500 510 510;...
    510 510 500 500 510 510; 500 510 510 500 500 500 ];
Y2=[450 450 550 550 450 450 ;450 550 550 450 450  450;...
    450 550 550 450 550 550;450 450 550 550 550 550];
Z2=[0 0 0 0 0 1000;0 0 0 0 0 1000; 1000 1000 1000 1000 0 1000;...
    1000 1000 1000 1000 0 1000];

% paramètre de la deuxième troisième plaque rouge
X3=[500 510 510 500 500 500; 510 510 500 500 510 510;...
    510 510 500 500 510 510; 500 510 510 500 500 500 ];
Y3=[600 600 1000 1000 600 600 ;600 1000 1000 600 600  600;...
    600 1000 1000 600 1000 1000;600 600 1000 1000 1000 1000];
Z3=[0 0 0 0 0 1000;0 0 0 0 0 1000; 1000 1000 1000 1000 0 1000;...
    1000 1000 1000 1000 0 1000];

% montage des trois plaques (cubes) rouges correspondant aux fentes
patch(X,Y,Z,'r');
patch(X2,Y2,Z2,'r');
patch(X3,Y3,Z3,'r');
text(500,1200,1000,'Les deux fentes','FontSize',18)
text(1000,1000,1000,'Ecran détecteur','FontSize',18)
text(0,1000,1000,'Le canon à virus','FontSize',18)

hold on

%sert à generer les points bleus avec une certaine periode.
A=rand(10,10);
B2=[];
for i=1:200:1000
    B=i+35*randn(500,1);
    B2=[B2 ; B];
end

t=15;% taille de la particule virale.t=15: petite particule.
for It=1:20 % 20 particules virales apparaitront

    k=1000*rand(1);% coordonnée de depart aleatoire (reste fixe)
    j=500+100*rand(1);% coordonnée de depart aleatoire (reste fixe)

    % i represente l'avancé de la particule vers la fente puis le detecteur
    for i=0:50:1000
        if i==500||i==550 % si passage au niveau des fentes
            t=100;%taille de la particule augmente pour symboliser une onde
            % montagne de la grosse particule ovoide
            [x, y, z] = ellipsoid(i,j,k,t,t,t,20);
            a=surface(x,y,z,'FaceColor',[0 1 0]','edgecolor','none',...
                'faceAlpha',0.5);
        else
            t=15;
            % montagne de la petite particule virale
            [x, y, z] = ellipsoid(i,j,k,t,t,t,4);
            a=surface(x,y,z,'FaceColor',[0 1 0]) ;
        end
        pause(0.05)
%         F=getframe(f);
%         Film = addframe(Film,F);
        delete(a);

 % Accelère l'apparition des boules bleues pour faire apparaitre les
 % interferences à partir de l'iteration 10.
        if It>=10
            plot3(1000*ones(10,1)',B2(randi(2500,[1,10])),...
                randi(1000,[1,10]),'.','MarkerSize',15);
            pause(0.00005);
        end
    end
    plot3(i,j,k,'.','MarkerSize',15);
end
%Film=close(Film)
```

### Code15PaysageAdaptatif.m

```
% petit script pour générer des paysages adaptatifs (fitness landscape)
figure('Color','w')
for i=1:3
    Handle(i)=subplot(1,3,i);
    axis(Handle(i),'off');
    hold(Handle(i));
    view(-42,26);
end

% plot 1
[x,y]=meshgrid([0:1:5]);
A=[0 0 0 0 0 0; 0 1 2 2 1 0; 0 3 4 4 3 0;  0 3 4 4 3 0;...
    0 1 2 2 1 0; 0 0 0 0 0 0 ];
[XI,YI] = meshgrid(0:.1:5);
ZI = interp2(x,y,A,XI,YI);
x = {linspace(0,6,51),linspace(0,6,51)};
[xx,yy] = ndgrid(x{1},x{2}); y = ZI;
[smooth,p] = csaps(x,y,[],x);
smoother = csaps(x,y,.996,x);
surf(Handle(1),x{1},x{2},smoother.'), axis off

% plot 2
[x,y]=meshgrid([0:1:5]);
A=rand(6,6)*5;
[XI,YI] = meshgrid(0:.1:5);
ZI=rand(51,51)*5;
x = {linspace(0,6,51),linspace(0,6,51)};
[xx,yy] = ndgrid(x{1},x{2}); y = ZI;
[smooth,p] = csaps(x,y,[],x);
smoother = csaps(x,y,.996,x);
surf(Handle(2),x{1},x{2},smoother.'), axis off

% plot 3
[x,y]=meshgrid([0:1:5]);
A=[0 1 0 1 0 0; 0 4 0 3 1 1; 0 0 0 0 1 1;  0 0 1 3 1 5;...
    0 0 0 2 2 3; 0 0 3 0 2 3 ];
[XI,YI] = meshgrid(0:.1:5);
ZI = interp2(x,y,A,XI,YI);
x = {linspace(0,6,51),linspace(0,6,51)};
[xx,yy] = ndgrid(x{1},x{2}); y = ZI;
[smooth,p] = csaps(x,y,[],x);
smoother = csaps(x,y,.996,x);
surf(Handle(3),x{1},x{2},smoother.'), axis off
```

### Code16TuneRugositePaysageAdaptatif.m

```
% petit script pour tuner "faussement" la rugosité d'un paysage adaptatif.
% en effet, ce script ne se base evidemment pas sur le modèle NK car on ne
% peut pas representer un paysage adaptatif de plus de 2 nucléotides en 3
% dimensions.
figure('Color','w')
for i=1:9
    Handle(i)=subplot(3,3,i);
    axis(Handle(i),'off')
    hold(Handle(i));
    view(44,60);
end
for i=1:9
Fit=[ 0.1 0.7 0.9 0.99 0.995 0.999 0.9995 0.9999 1];
[x,y]=meshgrid([0:1:5]);
[XI,YI] = meshgrid(0:.1:5);
ZI=rand(51,51)*5;
x = {linspace(0,6,51),linspace(0,6,51)};
[xx,yy] = ndgrid(x{1},x{2}); y = ZI;
[smooth,p] = csaps(x,y,[],x);
smoother = csaps(x,y,Fit(i),x);
surf(Handle(i),x{1},x{2},smoother.'), axis off
end
```

### Code17RegleWolfram.m

```
%code pour générer 6 petites figures d'automates cellulaires élementaires
figure('Color', 'w');
Regles=[254 90 30 30 110 110];% regles des automates cellulaires
NombreLignes=[50 50 50 200 50 500];
NombreColonnes=[100 100 100 400 100 1000];
for j=1:6
Regle=Regles(j);
NombreLigne=NombreLignes(j);% nombre total de ligne
NombreColonne=NombreColonnes(j);% nombre total de colonne
% J'utilise le booleen 1(off) et 2(on) et non 0 et 1
Ligne=ones(1, NombreColonne);% Ligne en cour d'analyse
Ligne(floor((NombreColonne+1)/2))=2;% cellule du milieu est "on" (2)
RegleCorrespondante = (bitget(Regle, 1:8) + 1);% decompresse les règles
Grille= ones(NombreLigne, NombreColonne);% allocation pour la grille
% Iteration pour generer le reste de la grille
for i=1:NombreLigne
    Grille(i, :) = Ligne;% Ecrit l'état courant dans la ligne en cour
    % étapes centrales : applique les règles des automates cellulaires à
    %à propager le long du patterne 1D
    % J'utilise la technique de David young: les deux lignes ci-dessous ne
    % sont pas simples à comprendre...prendre son temps...
    indice = sub2ind([2 2 2], ...
    [Ligne(2:end) Ligne(1)], Ligne, [Ligne(end) Ligne(1:end-1)]);
    Ligne = RegleCorrespondante(indice);
end
subplot(3,2,j);
% repasse en binaire classique (0et1)puis inverse les couleurs
imshow(~(Grille-1));
xlabel(['règle: ' int2str(Regle)])
end
```

### Code18JeuxVieConway.m

```
%Jeux de la vie de conway
% si le script génére un erreur, commenter le code du film
clear all
F=figure('color','w');
%mov = avifile('conway4.avi','quality',100,'compression','none');% code film
n=200; % taille de la grille (carré de longueur n)
%taille de la matrice
cellule=(rand(n,n)>0.7);% matrice generée avec quelques cellules "on" (1)
somme=zeros(n,n);% alloue l'espace à la variable somme
x = 2:n-1;
y = 2:n-1;
for i=1:200
        %somme des voisins les plus proches
        somme(x,y) = cellule(x,y-1) + cellule(x,y+1) + ...
            cellule(x-1, y) + cellule(x+1,y) + ...
            cellule(x-1,y-1) + cellule(x-1,y+1) + ...
            cellule(3:n,y-1) + cellule(x+1,y+1);
        % regle du jeux de la vie
        cellule = (somme==3) | (somme==2 & cellule);
%montre les images à chaque iteration et inverse les couleurs
imshow(~cellule)
%mov = addframe(mov,F);% code film
pause(0.1)
end
%mov = close(mov);% code film
```

### Code19GrammaireWolfram.m

```
% système de substitution sequentielle
% ce script tente de reproduire le systeme de la figure (h) page 91 de
% l'ouvrage de Stephen Wolfram "A new kind of science".
% Ce systeme fonctionne comme la fonction "find and replace" des editeurs
% de texte. Dans la Sequence de base S, l'algorithme trouve la sequence
% 011, il la remplace par 1000. Idem avec 0100 =>101 et 1=>1100
%le script fait evoluer ce systeme au cours du temps et chaque ligne
%represente une itération successive.
```

```
%evidemment, les 0 et 1 sont remplacés dans l'image par noir et blanc.

clear all
% grammaire (règle de substitution)
Sequence{1}='0 1 1';% si trouve 1, remplace par 6

Sequence{2}='0 1 0 0';% si trouve 2, remplace par 7

Sequence{3}='1';% si trouve 3, remplace par 8

Sequence{6}='1 0 0 0';

Sequence{7}='1 0 1';

Sequence{8}='1 1 0 0';

% Sequence initial qui evolue au cour du temps en fonction de la grammaire
S='0 1 0 1';

Matrice=zeros(500,50);% alloue la memoire pour l'image correspondante

for iteration=1:500
% écrit S dans la ligne numero "iteration".
    Matrice(iteration,1:length(str2num(S)))=str2num(S);

    for i=1:3 % cherche la sequence 1, puis 2 puis 3 dans S
        Match=findstr(Sequence{i},S);

% si il trouve la sequence, il opère la substitution
        if ~isempty(Match)
            S1=S(1:Match(1)-1);
            S2=S(Match(1)+length(Sequence{i}):length(S));
            S3=Sequence{i+5};
            S=[S1 S3 S2];
            break
        end
    end
end
% montre l'image du résultat
h=figure
imshow(Matrice)
saveas(h,'Grammaire500','eps')% sauve l'image en format  vectoriel
```

**Code20GrammaireSoupeOriginelle.m**

```
% petit script pour tester l' analogie  entre grammaire et soupe originelle
clear all
F=figure('color','white');
NombreIteration=1500;
LongueurChaine=50;
NombreRegle=150;
Diviseur=6;%sert à pooler les iterations dans une images
NombreImages=50;% nombre de fois que la simulation tournera.
TailleMaxPerm=15;
TailleMinPerm=3;
% test l'algorithme un certain nombre de fois avec des grammaires
% differentes
for t=1:NombreImages
    for i=1:NombreRegle
% pour chaque regle, decide de la taille du string à remplacer
       Taille=randi(TailleMaxPerm-TailleMinPerm)+TailleMinPerm;%

% choisit aleatoirement un string subtrat, puis enzyme, puis produit
        Sequence{i}.substrat=num2str(round(rand(1,Taille)),'%1d ');
        Sequence{i}.enzyme=num2str(round(rand(1,Taille)),'%1d ');
        Sequence{i}.produit=num2str(round(rand(1,Taille)),'%1d ');
    end

%String de depart
S=num2str(round(rand(1,LongueurChaine)),'%1d ');
%alloue la taille de la grille image
Grille=false(NombreIteration/Diviseur,LongueurChaine*Diviseur);

    for j=1:NombreIteration
```

```
%cette ligne un peu complexe sert à pooler les iterations en colonne de la
%largeur de LongueurChaine et de la longueur de NombreIteration/diviseur
% cela permet de visualiser un grand nombre d'iteration dans une seule
% grande image.
Grille((mod((j-1),NombreIteration/Diviseur)+1),(LongueurChaine*...
ceil(j/(NombreIteration/Diviseur))-LongueurChaine+1)...
:(LongueurChaine*ceil(j/(NombreIteration/Diviseur))))=str2num(S);

%applique l'ensemble des regles à chaque iteration mais dans un ordre
%aleatoire
        for i=randperm(NombreRegle)
   % cherche la sequence du substrat dans la sequence S
            MatchSubstrat=findstr(Sequence{i}.substrat,S);
   % cherche la sequence de l'enzyme dans la sequence S
            MatchEnzyme=findstr(Sequence{i}.enzyme,S);
   % si l'enzyme et le substrat sont présent alors...
            if ~isempty(MatchSubstrat) && ~isempty(MatchEnzyme)...
                    && length(S)>length(Sequence{i}.substrat)+1
                % fait autant de remplacement que l'element (enzyme ou
                % substrat) limitant
                for z=1:min([length(MatchSubstrat) length(MatchEnzyme)])
                   %remplacement aleatoire
                    Var=randperm(length(MatchSubstrat));
                   % change le substrat par le produit
                    S1=S(1:MatchSubstrat(Var(z))-1);
                    S2=S(MatchSubstrat(Var(z))+...
                    length(Sequence{i}.substrat):length(S));
                    S3=Sequence{i}.produit;
                    S=[S1 S3 S2];
                end
            end
        end
    end
    imshow(Grille)
    saveas(F,num2str(t),'png')%sauve chaque image avec son numero
    %pause(0.1)
end
```

**Code21CompareInformationTroisImages.m**

```
F=figure('Color', 'w');
for i=1:3  % genere 3 petit plots
    Handle(i)=subplot(1,3,i);
    hold(Handle(i));
end
% Image 1 uniforme noir
axes(Handle(1));
cdata=uint8(zeros(420,560));
imshow(cdata);
figure;
cdata=uint8(zeros(420,560));
imshow(cdata);
saveas(gcf,'imageUniforme','bmp');
clear cdata;
% Image 2 aleatoire
axes(Handle(2));
cdata=uint8(rand(420,560)*256);
imshow(cdata);
figure;
imshow(cdata);
saveas(gcf,'imageAleatoire','bmp');
clear cdata;

% Image 3 : ensemble de Mandelbrot
x = linspace(-2.1,0.6,1000);
y = linspace(-1.1,1.1,1000);
[X Y] = meshgrid(x,y);
C = complex(X,Y);
Zmax= 2;
kmax= 50;
Z = C;
for k=1:kmax
Z = Z.^2 + C;
```

```
end
contourf(x,y,Z);
axis off;
colormap gray;
saveas(gcf,'imageMandelbrot','bmp');
imread('imageMandelbrot.bmp');
axes(Handle(3));
imshow(ans);
imshow(F);
saveas(gcf,'Information3Images2','eps');
```

### Code22ProfondeurLogiqueNucleotide.m

```
% petit script didactique pour montrer que la profondeur logique de
% chaque nucleotide augmente au cours de l'evolution
scrsz = get(0,'ScreenSize');
F=figure('Color','w','Position',[0 0 scrsz(3) scrsz(4) ]);
%mov = avifile('ProfondeurLogique3.avi','quality',100,'compression',...
%    'none','fps',5);% code film
g1=subplot(1,2,1);
xlabel('Temps (millions d années)','FontSize',18);
ylabel('Profondeur logique/Complexité organisée/Information','FontSize',18)
hold(g1);
set(g1,'XTick', []);
set(g1,'YTick', []);
axis square;
axis on;
g2=subplot(1,2,2);
axis([1 3 1 3]);
axis square;
axis off;

Sequence(1:90)='A';
% cree un matrice de place aleatoire dans le plasmide (entre 1 à 90pb). sa
% taille depend de nombre de bacteries qui mutent
Place= randi(90,[1,1500]);
% cree matrice de base string de la même taille que place (même proba pour
% chaque base d'apparaitre)
Base=randsample('ACGT', 1500,true,[0.5 0.5 0.5 0.5]);

Cercle=0:pi/45:2*pi;
Fitness=1;
hold on
for Indice=1:90
    h{Indice}=text(2+cos(Cercle(Indice)),2+sin(Cercle(Indice)),...
        Sequence(Indice),'FontSize',16);% sinon base noir
end
for i =1:300
    if rand(1)>0.75
        delete(h{Place(i)});
        h{Place(i)}=text(2+cos(Cercle(Place(i))),2+...
            sin(Cercle(Place(i))),Base(i),'FontSize',16,'Color','g');
        if i>1
            Fitness(i)=Fitness(i-1)+1;
        end
    else
        Green=get(h{Place(i)},'Color');

        if ~strcmp(num2str(Green),'0  1  0')
            delete(h{Place(i)});
            h{Place(i)}=text(2+cos(Cercle(Place(i))),2+...
                sin(Cercle(Place(i))),Base(i),'FontSize',20,'Color','r');

            pause(0.08);
%             frame = getframe(F);
%             mov = addframe(mov,frame);% code film
            delete(h{Place(i)});
            h{Place(i)}=text(2+cos(Cercle(Place(i))),2+...
                sin(Cercle(Place(i))),Sequence(Place(i)),...
                'FontSize',16,'Color','k');
        end
        if i>1
            Fitness(i)=Fitness(i-1);
        end
    end
```

```
    plot(g1,1:i,Fitness(1:i))  ;
end
%mov = close(mov);% code film
```

### Code23ExemplaireLog.m

```
% Demonstration que le nombre d exemplaires d une sequence P appartenant
% à une sequence aleatoire S est proportionnelle au logarithme

% Ce script gènere une Sequence S aleatoire (symbolisant Mars) et cherche
% quelle est la taille maximal d'une sequence P appartenant à S tel que P
% est au moins en deux exemplaires dans S.
clear all
iteration=0;
% boucle qui test differentes tailles de S
for TailleS=[10:10:150 200:100:1000]

    % choisit un Sequence binaire aleatoire et la convertit en "string"
    S=num2str(round(rand(1,TailleS)));
    S=S(isspace(S)==0);% Supprime les espaces dans la sequence

    for    TailleP=1:TailleS % parcourt toute les tailles de P
        for DepartP=1:TailleS-TailleP %parcourt tout les departs de P

            P=S(DepartP:DepartP+TailleP);% Sequence P qui appartient à S
            % cherche le nombre de fois que P est dans S
            NombreExemplaire=findstr(P,S);

            % stoc les informations quand P est présent en deux exemplaires
            if length(NombreExemplaire)==2
                Exemplaire{iteration+1}.S=S;
                Exemplaire{iteration+1}.Length=length(P);
                Exemplaire{iteration+1}.P=P;
                Exemplaire{iteration+1}.Place=NombreExemplaire;
                break
            end
        end
    end
    iteration=iteration+1;
end

for j=1:iteration
    % recupere les longueurs max de P (fonction de la taille de S)
    MatriceExemplaire(j)=Exemplaire{j}.Length(end);
end
% plot des résultats
figure('Color', 'w');
axes('FontSize', 15)
hold on
plot([10:10:150 200:100:1000],MatriceExemplaire,'-bs','LineWidth',3)
plot([10:10:150 200:100:1000],3.4*log([10:10:150 200:100:1000])...
,'--g','LineWidth',3)
xlabel('Taille de S','FontSize', 15)
ylabel('Taille de P','FontSize', 15)
legend('Données de simulation','Ligne de vie (3.4 log)')
text(500,24,'vie','FontSize', 20)
text(500,18,'non-vie','FontSize', 20)
text(500,9,'non-vie','FontSize', 20)
set(legend,'FontSize', 15)

% plot S (en noir) et les deux exemplaires de P en rouge
figure('Color', 'w');
axis off
for j=1:15
text(-0.1,1-j/20,[Exemplaire{j}.S(1:Exemplaire{j}.Place(1)) ...
    '\color{red}'...
Exemplaire{j}.S(Exemplaire{j}.Place(1):Exemplaire{j}.Place(1)...
+ Exemplaire{j}.Length(end)-1) '\color{black}'...
Exemplaire{j}.S(Exemplaire{j}.Place(1)+...
Exemplaire{j}.Length(end):Exemplaire{j}.Place(2)) '\color{red}'...
Exemplaire{j}.S(Exemplaire{j}.Place(2):...
Exemplaire{j}.Place(2)+ Exemplaire{j}.Length(end)-1) '\color{black}'...
Exemplaire{j}.S(Exemplaire{j}.Place(2)+...
Exemplaire{j}.Length(end):end)],'FontSize', 12) ;
end
```

UNIVERSITÉ DE GRENOBLE

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Virologie Microbiologie Immunologie**

Arrêté ministériel : 7 août 2006

Présentée par

## [2023 : Viafx24]

Thèse dirigée par **[2023 : identité protégée]**

préparée au sein du
**Laboratoire Adaptation et Pathogénie des Microorganismes**
dans **l'École Doctorale Chimie et Sciences du Vivant**

# Thèse positiviste portant sur les réseaux de régulation chez *Escherichia coli*

Thèse soutenue publiquement le **29 aout 2012**,
devant le jury composé de :

[2023 : identités protégées]

# Table des matières

**Table des figures** **iii**

**I Introduction** **1**

1 La biologie des systèmes . . . . . 1
2 Le système d'étude . . . . . 2
3 Logique de régulation de l'expression d'*acs* . . . . . 4
4 Un criblage à haut-débit pour détecter de nouvelles interactions géniques . . . . 5
5 Recherche du mécanisme répresseur . . . . . 8
6 Un phénomène général . . . . . 10
7 Le flux glycolytique . . . . . 11
8 Expériences à l'échelle de la cellule individuelle . . . . . 11
9 Utiliser la répression catabolique en biologie synthétique . . . . . 13
10 Système synthétique de communication intercellulaire par l'AMPc . . . . . 15
11 Utiliser l'activité luciferase pour inférer l'état métabolique de la cellule . . . . . 17

**Références bibliographiques** **23**

**II A genome-wide screen for identifying all regulators of a target gene** **25**

**III Study of carbon catabolite repression in *Escherichia coli* by monitoring the activity of Crp-cAMP dependent promoters** **51**

**IV A synthetic *Escherichia coli* communication system mediated by extracellular cyclic AMP** **83**

**V Using luciferase activity to infer the metabolic state of *Escherichia coli*** **105**

**VI Conclusion et perspectives** **127**

**Références bibliographiques** **133**

# Table des figures


I.1 Fabrication de la fusion transcriptionnelle p-acs-lux. . . . . 3
I.2 Mesure de l'expression du gène *acs* dans une souche sauvage. . . . . 4
I.3 Expérience phare montrant une incohérence dans le modèle de la répression catabolique. . . . . 5
I.4 Un nouveau criblage à haut-débit pour détecter de nouvelles interactions géniques. 7
I.5 Schéma récapitulatif des deux modèles principaux de la répression catabolique. . 9
I.6 Souche *crp** et autres promoteurs Crp-AMPc dépendants. . . . . 10
I.7 Rôle du flux glycolytique. . . . . 11
I.8 Mesure de l'expression du gène *acs* à l'échelle de la microcolonie. . . . . 12
I.9 Un interrupteur synthétique contrôlé par le flux glycolytique. . . . . 14
I.10 Un système de communication intercellulaire via l'AMPc. . . . . 16
I.11 La dépendance énergétique de la luciferase. . . . . 18
I.12 Capacité à maintenir l'activité luciferase durant la croissance sur glucose et acétate dans différent mutants. . . . . 19
I.13 Un outil pour mesurer la concentration extérieure de glucose, d'acétate et le débordement du métabolisme. . . . . 21

VI.1 Expérience pour tenter de faire communiquer deux royaumes distincts. . . . . 129
VI.2 Induction de l'operon *luxCDABE* directement dans le tractus gastro-intestinal d'une souris grâce à l'AMPc. . . . . 130

# Chapitre I

# Introduction

## 1 La biologie des systèmes

La biologie des systèmes cherche à intégrer les différents niveaux d'information qui structurent un organisme. Ces niveaux d'information s'établissent à différentes échelles : macroscopique comme les mécanismes physiologiques, microscopique comme la communication cellulaire ou nanométrique comme la régulation génique. Interconnecter ces différents niveaux permettra, à terme, de comprendre, de prédire, voire de contrôler la dynamique de fonctionnement global d'un système vivant. La difficulté d'une telle approche vient du fait que chaque niveau est régi par des lois qui lui sont propre. Ainsi, avant de pouvoir construire un modèle global multi-échelle d'un système vivant complexe, il est nécessaire de diviser les difficultés en se focalisant à un niveau d'information donné sur un organisme suffisamment simple. Dans cette étude, nous nous focaliserons sur la dynamique de fonctionnement du réseau de régulation d'un organisme très étudié et facilement manipulable : la bactérie *Escherichia coli*.

L'adaptation d'une bactérie aux changements dans son environnement est contrôlée au niveau moléculaire par un réseau de régulation large et complexe impliquant des gènes, des ARN, des protéines et des métabolites. Transduction du signal, régulations transcriptionnelles, flux métaboliques sont étroitement interconnectés même si ils fonctionnent à des échelles de temps très différentes (de la milliseconde au jour). Connaître la topologie de ce réseau de régulation est le premier objectif de la biologie des systèmes. La base de données Ecocyc liste plus de 5345 interactions régulatrices chez *E. coli* et ce nombre ne cesse d'augmenter (Keseler *et al.*, 2010). Etudier le fonctionnement dynamique de ce réseau est le second objectif de la

biologie des systèmes. Cette dynamique est non linéaire et comprend de nombreuses boucles de rétroaction. De plus, la nature intrinsèquement stochastique des mécanismes de régulation (mouvement brownien) rend le réseau de régulation idiosyncratique : pour un même génotype, une topologie du réseau identique et une même perturbation, deux bactéries peuvent adopter deux phénotypes distincts lors de l'adaptation (Veening *et al.*, 2008).

En l'état actuel des connaissances, il n'est pas encore possible de comprendre et de contrôler finement la dynamique de fonctionnement de cet énorme réseau de régulation. Nous nous sommes donc focalisés sur un module de régulation plus restreint mais faisant intervenir à la fois une réorganisation de la transcription et une réorganisation des flux métaboliques : nous avons choisi, comme point de départ, d'étudier l'adaptation de la bactérie *E. coli* à un changement de sources de carbone : du glucose à l'acétate (Wolfe, 2005).

## 2 Le système d'étude

Lorsque les bactéries *E. coli* sont placées dans un milieu contenant du glucose, elles consomment ce glucose tout en secrétant de l'acétate dans le milieu extérieur. Une fois le glucose épuisé, les bactéries réimportent cet acétate et l'utilisent comme source d'énergie. Cette transition glucose-acétate entraîne une inversion des flux métaboliques (de la glycolyse à la néoglucogenèse) et une profonde réorganisation de l'expression des gènes (Oh *et al.*, 2002).

Pour pouvoir utiliser l'acétate comme source d'énergie principale, il est nécessaire de le convertir en acétyl-coenzyme A. Cette première étape est réalisée par l'acétyl-coenzyme A synthétase encodée par le gène *acs* (Kumari *et al.*, 1995). Ce gène a un rôle précoce et prédominant dans l'établissement de la transition glucose-acétate c'est pourquoi nous nous sommes focalisés sur la régulation de son expression. En effet, sa transcription est connue pour être activée par le facteur de transcription Crp qui n'est actif que lorsque le glucose est épuisé (Kumari *et al.*, 2000). Ce mécanisme, appelé répression catabolique, fait intervenir une petite molécule : l'adénosine monophosphate cyclique (AMPc). Selon un modèle classique, lorsque le glucose est épuisé, une cascade de réactions aboutit à l'activation d'une enzyme, l'adenylate cyclase, qui produit alors une forte concentration d'AMPc. Cette molécule se fixe à la protéine Crp et change sa conformation. Le complexe ainsi formé a une plus grande affinité pour son site de liaison à l'ADN, au niveau de la séquence promotrice du gène *acs*. La fixation de ce complexe Crp-AMPc entraine l'activation de la transcription de ce gène.

Pour pouvoir mesurer l'expression du gène *acs* avec une haute résolution temporelle, nous utilisons la technologie des fusions transcriptionnelles. Nous fusionnons, sur un plasmide multi-

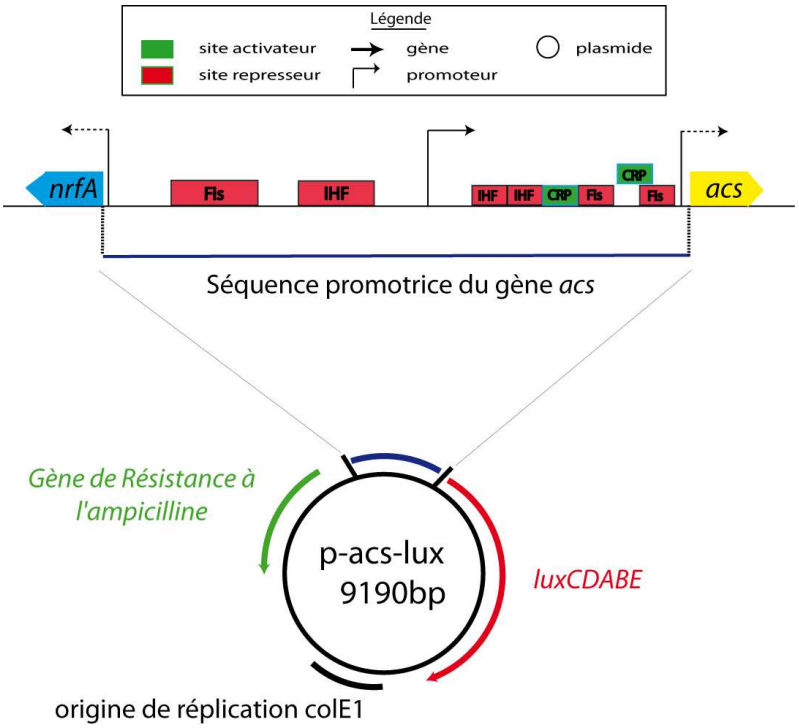



**Figure I.1** – Fabrication de la fusion transcriptionnelle p-acs-lux. La première étape consiste à amplifier la région promotrice du gène *acs* à partir du chromosome bactérien. Puis cette région est clonée en amont du gène rapporteur (*luxCDABE*) sur un plasmide multicopie. Cet « instrument » permet d'inférer l'activité du promoteur *acs* à partir du signal de luminescence généré.

copie, la séquence promotrice du gène *acs* en amont d'un gène rapporteur (l'operon *luxCDABE*) codant pour la luciferase : cet outil serait désormais appelé p-acs-lux (Figure I.1). Ainsi l'activité du promoteur *acs* est directement reliée à la concentration de luciferase qui est, elle-même, proportionnelle à la production de lumière, facilement mesurable Une fois notre outil construit, nous l'avons placé, par transformation, dans des bactéries *E. coli* sauvages et nous avons fait pousser ces dernières dans un milieu minimum contenant du glucose et de l'acétate. La croissance a lieu dans un lecteur de microplaque ce qui nous permet de mesurer, à intervalle de temps régulier, l'absorbance qui reflète le nombre de bactéries et la luminescence qui reflète l'activité du promoteur *acs*. Avec cette méthode, nous avons pu observer que l'activité du promoteur *acs* est fortement réprimée durant la croissance sur glucose, puis augmente brutalement lorsque ce dernier est épuisé (Figure I.2). Pour s'assurer que ce profil d'expression est correct, nous l'avons contrôlé à l'aide d'une autre fusion transcriptionnelle composée d'un autre rapporteur : la *gfp*. Le profil obtenu en fluorescence est similaire à celui obtenu en luminescence. Enfin, pour définitivement exclure tout artefact, nous avons également confirmé ce profil à l'aide d'une méthode totalement indépendante : la qRT-PCR.

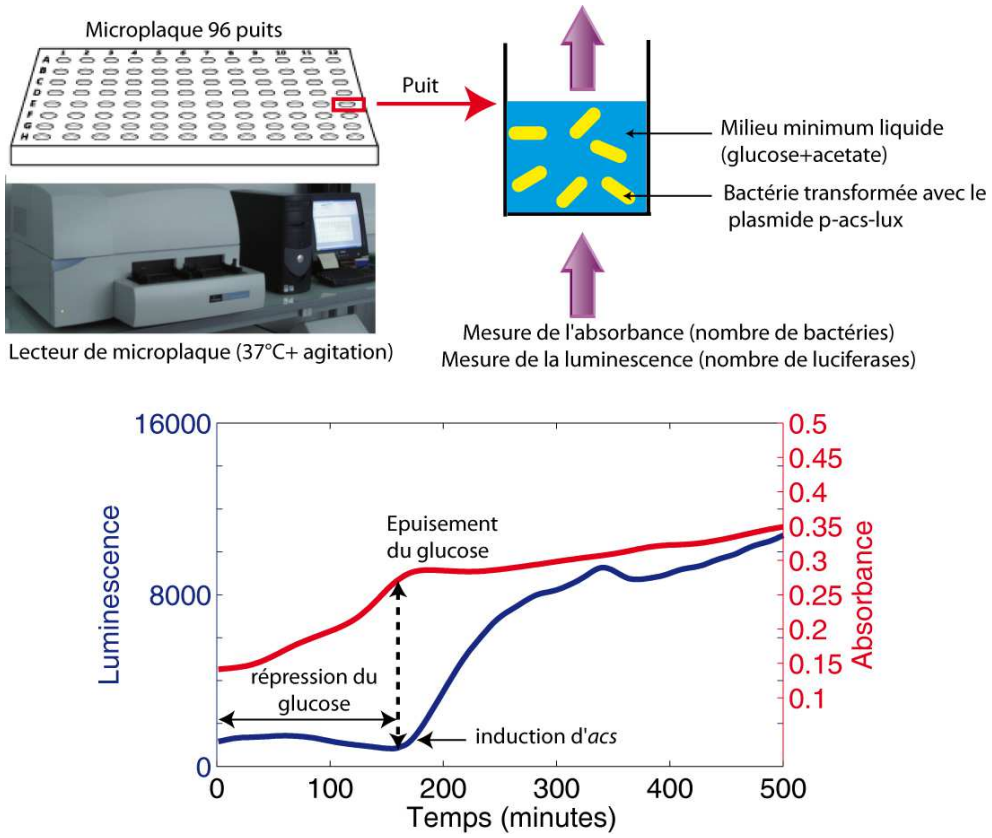



**Figure I.2** – Mesure de l'expression du gène *acs* dans une souche sauvage. Les mesures d'expression génique sont effectuées dans une microplaque incubées à 37°C sous agitation. Des bactéries sauvages, contenant le plasmide p-acs-lux, sont placées dans un milieu minimum liquide avec du glucose et de l'acétate. Le lecteur de microplaque mesure l'absorbance et la luminescence toutes les 5 minutes pendant plusieurs heures. La courbe rouge correspond à la courbe de croissance. On remarque un point de cassure qui signale l'épuisement du glucose. C'est à ce moment là qu'a lieu une forte induction de luminescence correspondant à une forte expression du gène *acs* (courbe bleue).

# 3 Logique de régulation de l'expression d'*acs*

Dans les paragraphes qui suivent, nous allons nous intéresser à la cascade d'événements qui a lieu entre l'épuisement du glucose et l'activation du promoteur *acs*. C'est en effet cette cascade d'événements qui est en mesure d'expliquer l'allure du profil d'expression décrit précédemment. La littérature décrit l'activation de l'expression du gène *acs* par le facteur de transcription Crp. Nous avons vérifié cette information en mesurant son expression dans un mutant $\Delta crp$. Comme prévu, dans une telle souche, il n'y a aucune expression du gène *acs*. Ceci entraîne un phénotype facile à observer : l'absence de croissance sur acétate. En réintroduisant, dans ce mutant, une copie du gène *crp*, on retrouve l'expression du gène *acs* et la croissance sur acétate. Les choses se passent de manière très similaire dans un mutant $\Delta cyaA$, c'est à dire

dans un mutant sans adenylate cyclase. En absence d'AMPc, le facteur de transcription Crp ne peut pas se lier à la séquence promotrice du gène *acs* et ce dernier n'est pas exprimé. A l'inverse, l'ajout d'AMPc exogène dans le milieu rétablit l'expression du gène *acs*. Le résultat de cette dernière expérience a mis en évidence une divergence vis-à-vis du modèle classique de la répression catabolique. L'ajout d'AMPc dans le milieu aurait dû contourner la répression du glucose et induire directement l'expression du gène *acs* malgré la présence de glucose. Or ce n'est pas ce que nous observons : l'expression d'*acs* dans le mutant $\Delta cyaA$ est complémentée par l'ajout d'AMPc exogène, mais seulement après épuisement du glucose (Figure I.3). Nous avons donc déduit qu'il existait un autre mécanisme régulateur empêchant l'expression d'*acs* tant qu'il y a du glucose dans le milieu.

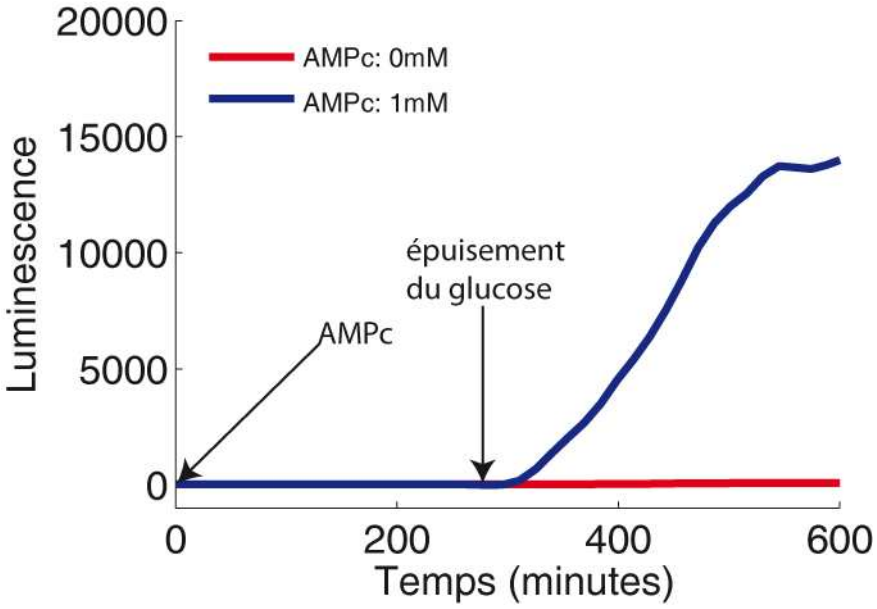



**Figure I.3** – Expérience phare montrant une incohérence dans le modèle de la répression catabolique. nous mesurons l'expression du gène *acs* dans une souche $\Delta cyaA$ poussant dans un milieu liquide avec du glucose et de l'acétate. Sans AMPc (courbe rouge), il n'y a aucune expression du gène *acs*. La présence d'AMPc exogène depuis le début de la cinétique permet de rétablir l'expression d'*acs*. Pourtant ce rétablissement a lieu seulement une fois que le glucose est épuisé.

# 4 Un criblage à haut-débit pour détecter de nouvelles interactions géniques

Pour rechercher ce mécanisme de régulation inconnu, nous avons développé un nouveau type de criblage à haut-débit. Nous avons transformé les 4000 simples mutants d'*Escherichia coli*,

provenant d'une collection (Baba et al. 2006), avec notre fusion transcriptionnelle p-acs-lux. Le criblage consiste à comparer la luminescence des colonies mutantes avec celle de la colonie sauvage : une différence significative pouvant suggérer que le gène déleté contrôle, directement ou indirectement, l'expression du gène *acs* (Figure I.4). Parmi les 4000 mutants, nous avons identifié deux groupes de gènes bien distincts. Le premier groupe est composé de gènes impliqués dans la fabrication de l'énergie. La luciferase, pour émettre des photons, a besoin de pouvoir réducteur ce qui implique la présence d'un métabolisme central fonctionnel (Meighen, 1991). Certaines colonies mutantes, émettant significativement moins de lumière, révèlent un épuisement énergétique plutôt qu'une régulation spécifique sur l'activité du promoteur *acs* : il s'agit donc de faux positifs. Nous reviendrons plus tard en détail sur cet artefact énergétique pour montrer qu'il peut également être utile.

Un deuxième groupe de gènes a été identifié par notre criblage : les gènes du système de transfert de phosphate (PTS). Les enzymes correspondantes sont impliquées dans l'import et la phosphorylation du glucose pour donner du glucose-6-phosphate. Le phosphate ajouté a subi une cascade de transfert partant de phospho-enol-pyruvate (PEP) et se déplaçant successivement sur l'enzyme E1 (*ptsI*), puis HPR (*ptsH*) puis $\text{EIIA}_{glc}$ (*crr*), puis sur le transporteur $\text{EIIB/C}_{glc}$ (*ptsG*) pour se retrouver enfin sur le $6^{ème}$ carbone du glucose formant le glucose-6-phosphate (Postma et al. 1993).

La détection par notre criblage des gènes appartenant au PTS n'est pas étonnante en soi. En effet, l'implication de ce système dans la répression catabolique est connue depuis longtemps. Ce système fait le lien entre l'épuisement du glucose et l'augmentation de la concentration d'AMPc dans la cellule. Le modèle actuel postule que seule la forme phosphorylée de l'enzyme $\text{EIIA}_{glc}$ soit capable d'activer l'adenylate cyclase et donc d'augmenter la synthèse d'AMPc. Or la forme phosphorylée d'$\text{EIIA}_{glc}$ est la version prédominante en absence de glucose (schéma récapitulatif de la figure I.5). Notre criblage n'a pas seulement réussi à faire ressortir les mutants du PTS, il a également révélé de nouvelles subtilités dans le fonctionnement de la répression catabolique. En effet, les colonies $\Delta crr$ (sans $\text{EIIA}_{glc}$) et $\Delta ptsI$ (sans E1) ont été sélectionnées pour leur faible luminescence alors que la colonie $\Delta ptsG$ (sans $\text{EIIB/C}_{glc}$) a été sélectionnée sur la base de sa forte luminescence. Ce résultat n'est ni prédictible ni explicable par le modèle ci-dessus.

Un des avantages de notre méthode à haut-débit est d'offrir la possibilité de mener des investigations plus poussées en testant les phénotypes intéressants dans différentes conditions et en dynamique. Par exemple, nous avons pu vérifier que, dans mutant $\Delta crr$ (sans $\text{EIIA}_{glc}$), l'ajout d'AMPc exogène dans le milieu rétablit l'expression du gène *acs*. Ce résultat suggère (sans le démontrer) le mécanisme d'activation de l'adenylate cyclase par la version phosphorylée de $\text{EIIA}_{glc}$. Pourtant (et comme observé avec le mutant $\Delta cyaA$) dans le mutant $\Delta crr$, l'AMPc

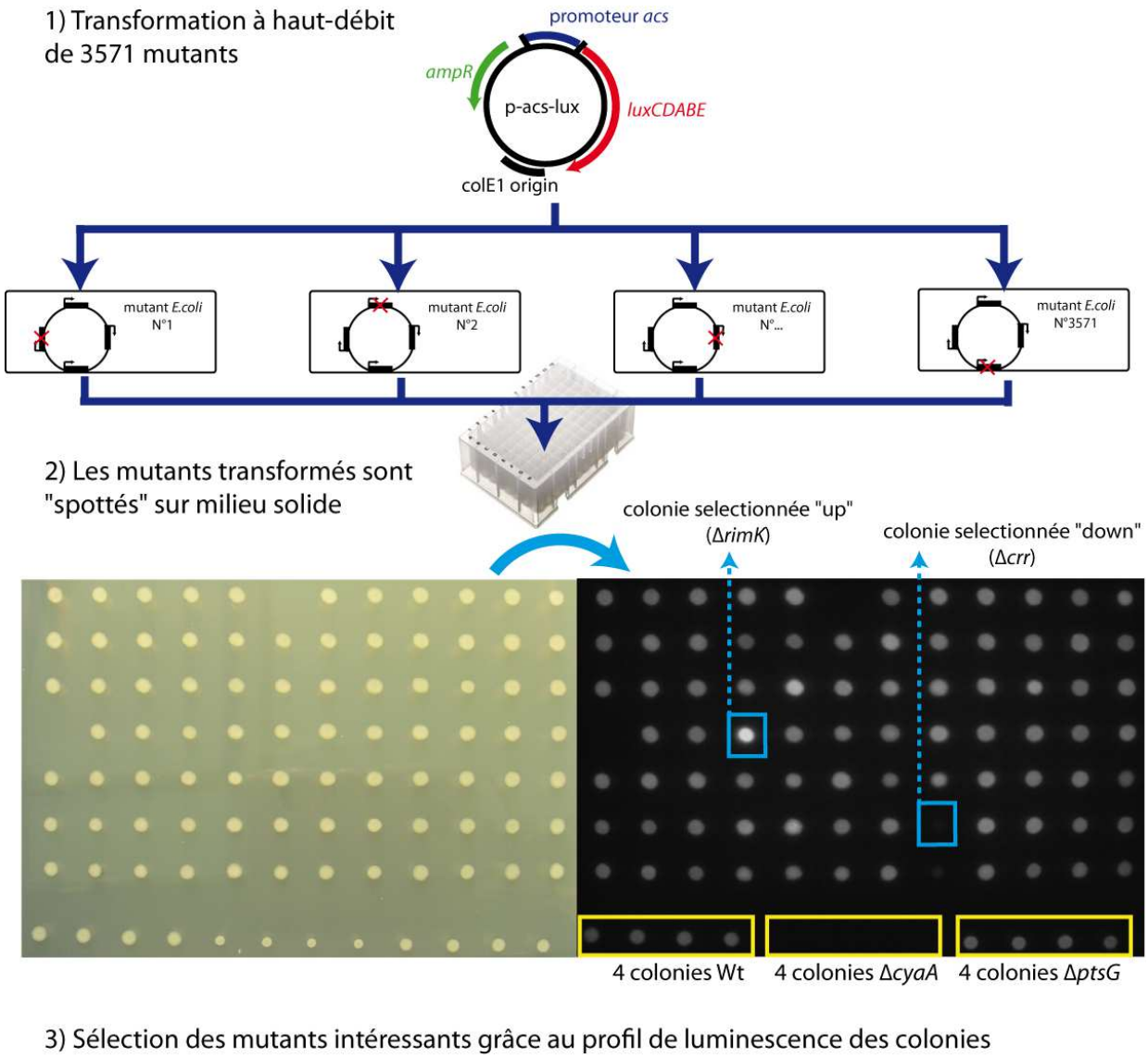



**Figure I.4** – Un nouveau criblage à haut-débit pour détecter de nouvelles interactions géniques. Nous transformons la fusion transcriptionnelle p-acs-lux dans l'ensemble des simples mutants d'*E. coli*. Puis nous « spottons » sur boite de petri (il est possible de tester différents types de milieu) les transformants. Après une nuit à 37°C, les colonies apparaissent. Le profil de luminescence de ces colonies est mesuré à l'aide d'une caméra sensible. Les colonies très luminescentes (par rapport à la souche sauvage) peuvent révéler l'absence d'un répresseur transcriptionnelle. A l'inverse, les colonies peu luminescentes peuvent révéler l'absence d'un activateur.

rétablit l'activité du promoteur *acs* seulement une fois que le glucose est épuisé : il subsiste donc un mécanisme de répression indépendant de l'enzyme $EIIA_{glc}$. Nous allons voir que ce résultat d'expérience contredit un deuxième modèle de la répression catabolique.

## 5 Recherche du mécanisme répresseur

Il existe une deuxième explication au phénomène de répression catabolique : le mécanisme d'exclusion de l'inducteur (Saier & Roseman, 1972). Beaucoup de gènes chez *E. coli* ont leur expression contrôlée par un couple facteur de transcription-inducteur. Citons le couple classique LacI-lactose[1]. Dans le modèle de l'exclusion de l'inducteur, c'est, cette fois, la forme déphosphorylée de $EIIA_{glc}$ qui inhibe l'import du lactose au niveau du transporteur. Or la forme déphosphorylée d'$EIIA_{glc}$ est prédominante en présence de glucose (schéma récapitulatif de la figure I.5). L'inhibition de l'import du lactose empêche ce dernier de jouer son rôle d'inducteur.

Dans le cas d'*acs*, il n'existe pas de facteur de transcription connu qui, en présence d'acétate, activerait ou de-réprimerait l'expression du gène *acs*. Notre criblage n'a d'ailleurs pas détecté l'existence d'un tel facteur. Il est difficile d'exclure expérimentalement l'hypothèse du rôle « inducteur » de l'acétate car ce dernier est produit et sécrété par la bactérie elle-même. Cependant, si la version déphosphorylée de $EIIA_{glc}$ inhibait l'entrée de l'inducteur potentiel « acétate », alors l'ajout d'AMPc exogène dans le milieu de croissance d'un mutant $\Delta crr$ (sans $EIIA_{glc}$) devrait contourner la répression exercée par le glucose via $EIIA_{glc}$. Or ce n'est pas ce que nous observons. Ainsi ni le modèle d'exclusion de l'inducteur ni le modèle passant par l'activation de l'adenylate cyclase ne sont en mesure d'expliquer nos données.

A partir de là, il restait une dernière hypothèse à exclure : l'AMPc exogène que nous ajoutons pourrait ne pas réussir à rentrer dans la bactérie en présence de glucose ce qui pourrait expliquer l'absence de levée de répression. D'autant que la bactérie *E. coli* secrète des quantités non négligeables d'AMPc dans le milieu extérieur sans que l'on connaisse la fonction de cet AMPc secrété. Ce pool extérieur d'AMPc pourrait être la résultante d'un mécanisme d'import/export servant à contrôler finement la concentration d'AMPc intracellulaire. Pour exclure cette hypothèse, nous avons utilisé une protéine dite Crp*, qui est une copie de Crp légèrement modifiée (Khankal *et al.*, 2009). Cette protéine a la particularité d'être « AMPc indépendante » : elle peut activer les promoteurs Crp-AMPc dépendants sans nécessité la formation du complexe Crp-AMPc. Nous avons donc construit une souche $\Delta crr\ \Delta cyaA\ crp^*$. Quand cette souche pousse dans un milieu avec du glucose et de l'acétate (et sans AMPc exo-

[1] En réalité l'allolactose

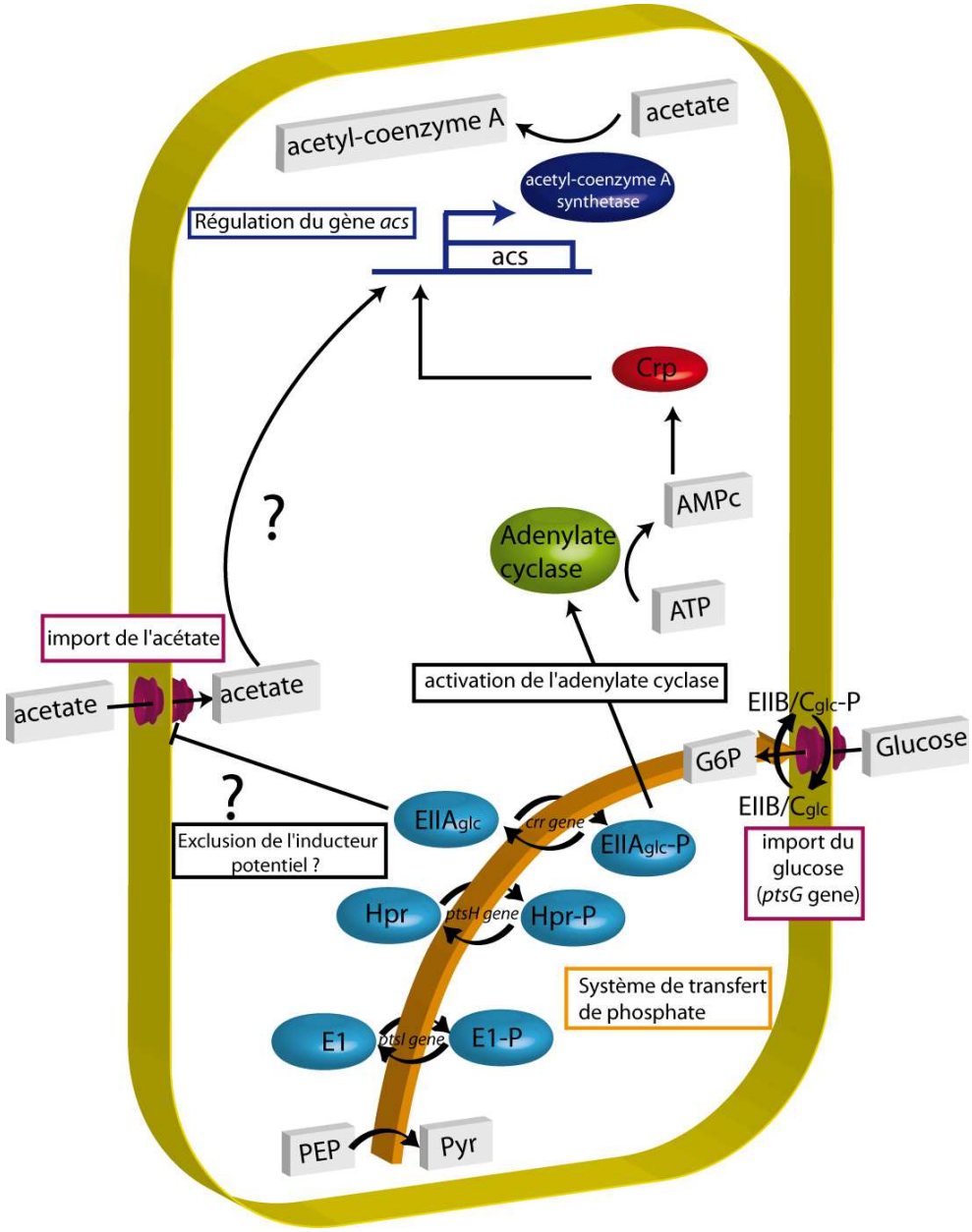



**Figure I.5** – Schéma récapitulatif des deux modèles principaux de la répression catabolique. Dans le modèle « AMPc », l'arrêt de l'import de glucose, entraine la phosphorylation des enzymes du système de transfert de phosphate. L'enzyme $EIIA_{glc}$, dans sa version phosphorylée, active l'adenylate cyclase ce qui augmente la concentration d'AMPc. Cette molécule se fixe ensuite au facteur de transcription Crp qui active alors l'expression du gène *acs*. Bien que le rôle de l'acétate comme inducteur du gène *acs* n'ait jamais été montré, il est possible d'imaginer un deuxième mécanisme de répression catabolique : l'exclusion de l'inducteur. Dans ce modèle, durant la croissance sur glucose, l'enzyme $EIIA_{glc}$ est principalement sous forme déphosophorylée. Or, cette forme pourrait inhiber le système d'import de l'acétate et empêcher son rôle hypothétique d'inducteur. La suite de notre exposé montrera que ni l'un ni l'autre de ces modèles ne peuvent expliquer la répression catabolique que nous observons.

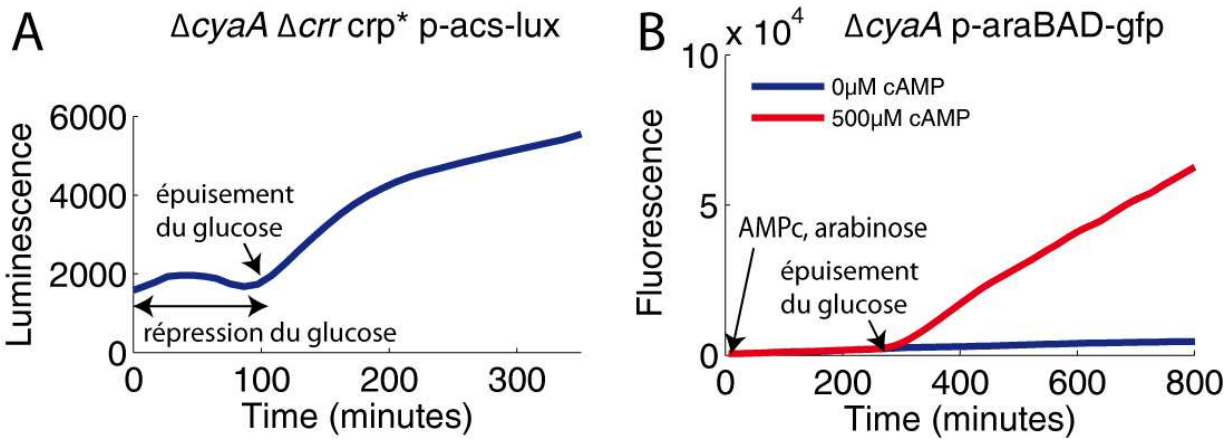



**Figure I.6** – Souche *crp** et autres promoteurs Crp-AMPc dépendants. A. Expression du gène *acs* dans une souche Δ*crr* Δ*cyaA* *crp** poussant dans un milieu minimum avec du glucose et de l'acétate. Comme la courbe de croissance n'est pas montrée, nous indiquons l'épuisement du glucose par une flèche. On constate que le glucose réprime l'expression du gène *acs* malgré l'absence d'AMPc et de l'enzyme $EIIA_{glc}$. B. Expression de l'operon *araBAD* dans une souche Δ*cyaA* poussant dans un milieu minimum avec du glucose, de l'acétate, de l'arabinose (l'inducteur) avec ou sans AMPc. Ici le gène rapporteur est la *gfp*, c'est pourquoi nous mesurons la fluorescence et non la luminescence. On constate que l'AMPc ajouté permet bien de rétablir l'expression de l'operon *araBAD* mais seulement une fois que le glucose est épuisé.

gène), le profil d'expression du gène *acs* est très similaire à celui de la souche sauvage : la répression du glucose est toujours visible malgré le fait que les deux pierres angulaires des modèles de la répression catabolique, $EIIA_{glc}$ et l'adenylate cyclase, ont été retirées (Figure I.6A). L'hypothèse du contrôle de l'import/export de l'AMPc, comme explication de nos résultats, devient caduque car la répression du glucose perdure dans cette souche en absence d'AMPc.

## 6 Un phénomène général

Ces résultats obtenus sont-ils spécifiques à la régulation du gène *acs* ou sont-ils généralisables à l'ensemble des promoteurs Crp-AMPc dépendants ? Pour répondre à cette question, nous avons sélectionné plusieurs fusions transcriptionnelles Crp-AMPc dependantes dans une collection (Zaslaver *et al.*, 2006). Il s'agit des promoteurs *araBAD*, *glpA* et *sdhC*. Bien qu'il existe des subtilités quantitatives, nous avons pu montrer que les résultats obtenues avec *acs* sont généralisables à ces 3 promoteurs (Figure I.6B). Nous pensons donc que nos recherches portent sur le mécanisme général de la répression catabolique.

# 7 Le flux glycolytique

Dans toutes nos expériences, le paramètre qui semble crucial pour contrôler l'état de la répression catabolique (et donc l'état d'activation du régulon Crp) est l'état du flux glycolytique. En effet, dans le mutant Δ*ptsG* (sans le transporteur du glucose), le niveau basal d'expression d'*acs* est surélevé par rapport à la souche sauvage car le flux glycolytique y est, à l'inverse, réduit (Figure I.7A). Une autre manière pour diminuer le flux glycolytique consiste à utiliser un sucre non métabolisable (l'$\alpha$-methyl-glucose) agissant en compétition avec le glucose. La littérature décrit ces sucres comme des répresseurs cataboliques alors que nous ne pensons pas qu'ils le soient. Au contraire, lorsque le ratio glucose/$\alpha$-methyl-glucose génère la réduction adéquate du flux glycolytique, nous avons pu activer l'expression de plusieurs promoteurs Crp-AMPc dépendants (Figure I.7B).

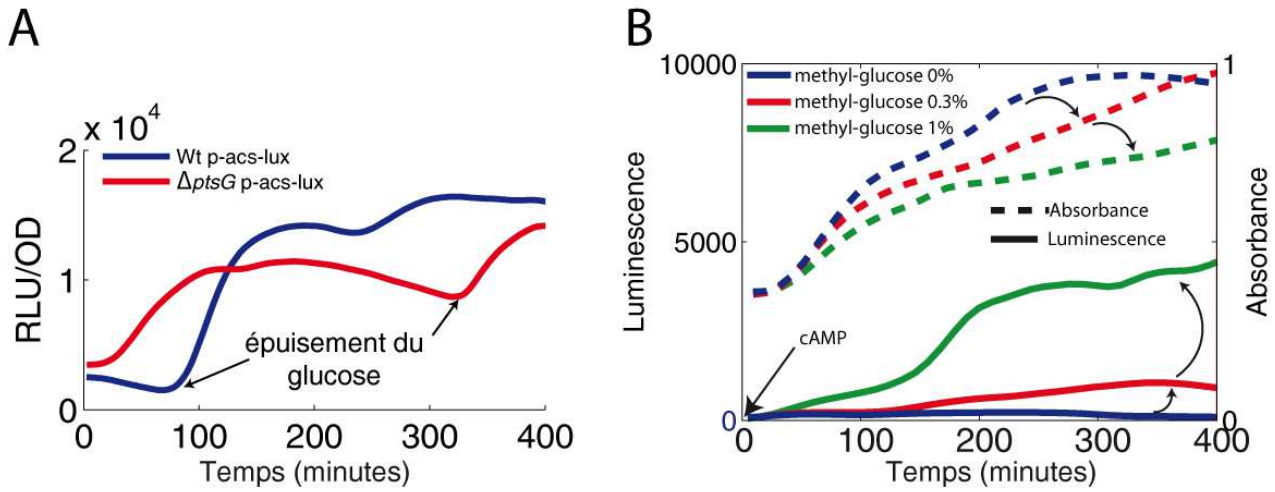



**Figure I.7** – Rôle du flux glycolytique. A. Dans un mutant Δ*ptsG* (courbe rouge), déleté du transporteur principal du glucose, poussant sur un milieu minimum avec du glucose et de l'acétate, la consommation du glucose (et donc le flux glycolytique) est plus lente ce qui, d'une part, retarde l'induction d'*acs* et, d'autre part, rehausse son expression par rapport à une souche sauvage. B. Dans une souche Δ*cyaA* rapportant l'expression du promoteur *sdh* (Crp-AMPc dépendant) et poussant sur un milieu avec du glucose et de l'AMPc, il est possible de rehausser l'expression du gène *sdh* de manière dose-dépendante en diminuant la consommation du glucose (ce qui diminue le taux de croissance) à l'aide du methyl-glucose, non métabolisable.

# 8 Expériences à l'échelle de la cellule individuelle

Les données présentées jusqu'à maintenant sont principalement issues d'expériences effectuées à l'échelle de la population dans les puits d'une microplaque. Nous avons également vérifié

certains de nos résultats à l'échelle de la cellule isolée. Par exemple, nous avons pu mettre en évidence l'induction du gène *acs* lorsque le glucose est épuisé (Figure I.8). De même, nous avons pu observer, sous le microscope, la répression catabolique exercée sur l'expression de l'operon *araBAD* malgré la présence de l'inducteur (arabinose) et de l'AMPc. L'expression de l'operon *araBAD* est connue pour amplifier les effets du bruit moléculaire à cause de la présence d'une boucle de rétroaction positive (Megerle *et al.*, 2008). Nous avons également observé ces variations stochastiques. De plus, nous pensons avoir été en mesure de contrôler finement la distribution de l'expression de cet opéron dans la population de cellules en faisant varier le flux glycolytique à l'aide de l'$\alpha$-methyl-glucose.

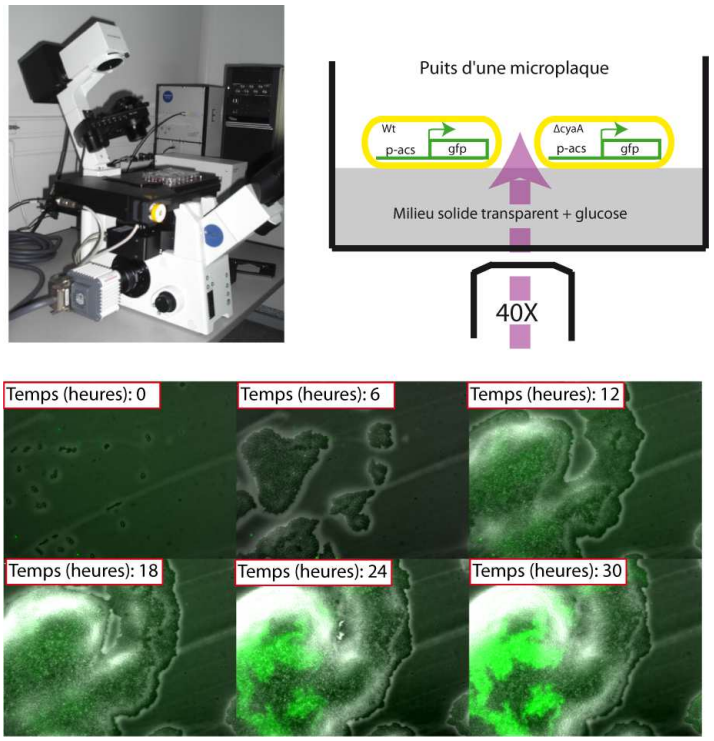



**Figure I.8** – Mesure de l'expression du gène *acs* à l'échelle de la microcolonie. Un puits d'une microplaque 6 puits contient un milieu minimum solide (agarose) avec du glucose. Deux types de bactéries sont présentes sur ce milieu : la souche sauvage et la souche $\Delta cyaA$, toutes deux transformées avec un plasmide rapportant l'expression du gène *acs* via la *gfp*. Les deux types de bactéries commencent à former des micro-colonies et ces dernières rentrent en contact et se mélangent. Une fois le glucose épuisé (arrêt de la croissance de la microcolonie), seul le tapis bactérien composé des bactéries sauvages voit sa fluorescence augmenter brusquement, signe de l'expression du gène *acs*.

Notre système expérimental sous le microscope (milieu solide, formation de microcolonies) est très différent de celui en milieu liquide. Pourtant, les résultats obtenues dans ces deux types de conditions sont similaires ce qui démontrent leur certaine robustesse. Cela peut sembler anodin mais le mécanisme responsable de la répression catabolique fait encore l'objet de nombreux

débats (Inada *et al.*, 1996; Crasnier-Mednansky, 2008; Gorke & Stulke, 2008; Narang, 2009). Les désaccords sont souvent liés à l'obtention de résultats différents entre les équipes pour une même expérience : par exemple, l'effet de l'ajout d'AMPc exogène sur le regulon Crp en présence de glucose. Tout le monde s'accorde cependant pour dire que les modèles actuels sont incomplets et doivent être révisés. Sans certitude cependant, nous pensons que le mécanisme responsable de la répression catabolique est encore inconnu. Il ne s'agit, pour nous, ni des variations de concentration de l'AMPc intracellulaire ni des variations de l'état de phosphorylation d' $EIIA_{glc}$. Paradoxalement, et c'est ce qui rend les choses si complexes, les rôles du PTS et de l'AMPc sont incontestables.

# 9 Utiliser la répression catabolique en biologie synthétique

Ainsi nous pensons que le fonctionnement de l'interrupteur, contrôlé par le flux glycolytique et contrôlant l'activité des promoteurs Crp-AMPc dépendants, reste inconnu. Nous avons créé un petit système synthétique qui illustre la fiabilité de cet interrupteur (Figure I.9). Notre système est constitué de deux fusions transcriptionnelles distinctes. La première rapporte l'activité du promoteur *uhpT* via le rapporteur GFP. Ce promoteur est activé par l'inducteur glucose-6-phosphate (G6P). La deuxième fusion rapporte l'activité du promoteur *acs* via le rapporteur luciferase. Le mécanisme d'interrupteur fonctionne de la manière suivante : les bactéries poussent sur un milieu minimum contenant du G6P et de l'acétate. La présence du G6P dans le milieu active le promoteur *uhpT* (ON), ce qui permet l'utilisation de ce sucre comme source de carbone. La mise en place du flux glycolytique empêche l'activité du promoteur *acs* (OFF). Lorsque le G6P est épuisé, le promoteur *uhpT* cesse d'être actif (OFF), le flux glycolytique s'effondre ce qui active le promoteur *acs* (ON). Ce petit système synthétique, bien que basé sur le mécanisme inconnu de la répression catabolique, est robuste, performant et rapide. Nous avons également implémenté une version un peu différente à l'échelle de la cellule isolée. Ce petit détour par la biologie synthétique nous offre une excellente transition pour amorcer la prochaine partie : la description d'un système synthétique de communication intercellulaire par l'AMPc.

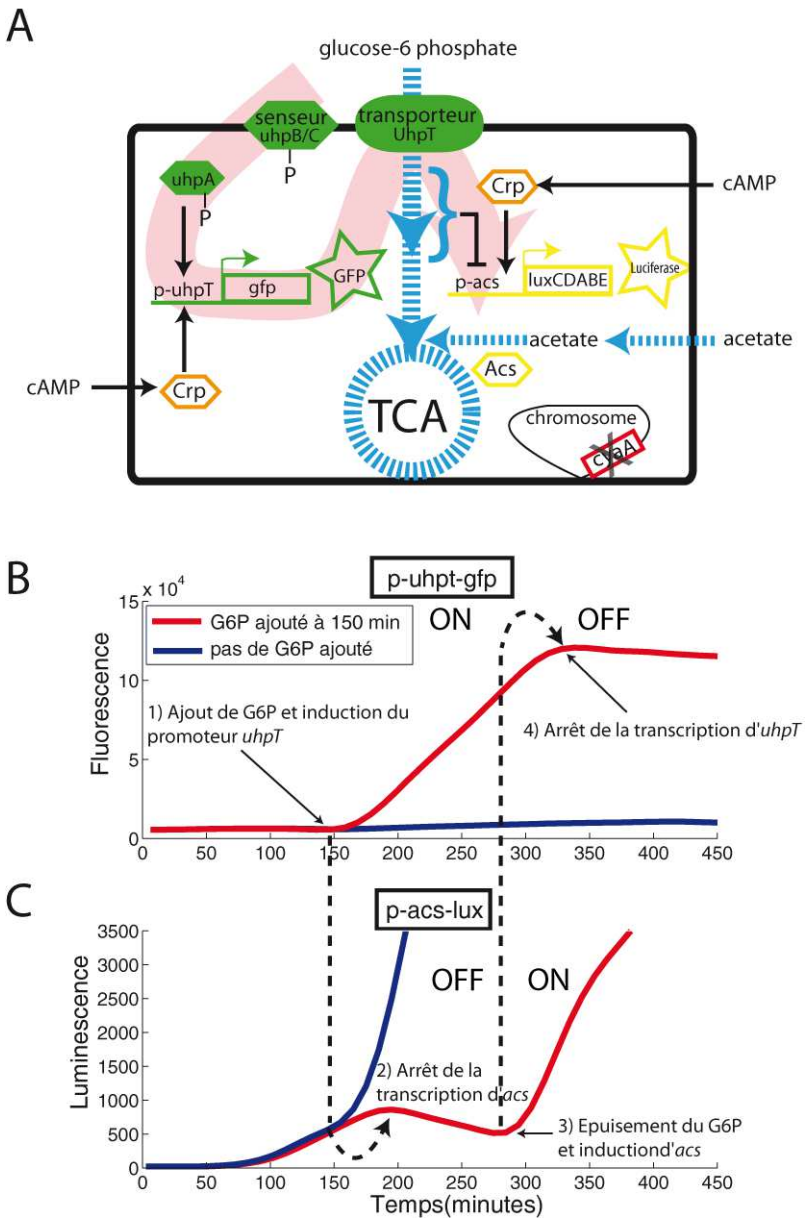



**Figure I.9** – Un interrupteur synthétique contrôlé par le flux glycolytique. A. Schéma d'un mutant $\Delta cyaA$ transformé avec deux fusions transcriptionelles (p-acs-lux et p-uhpT-gfp). Le glucose-6-phosphate (G6P) est détecté par un mécanisme à deux composantes (UhpB/UhpA) qui active l'expression du gène *uhpT*. Ce gène code pour le transporteur du G6P. Ce dernier, par le flux glycolytique qu'il génère, stoppe l'expression d'*acs* par le mécanisme inconnu de la répression catabolique. B et C : les bactéries poussent sur un milieu minimum avec du glucose, de l'acétate et de l'AMPc. Au bout de 100 minutes, il y a la première induction du gène *acs* dû à l'épuisement du glucose (C). L'ajout de G6P après 150 minutes déclenche l'expression d'*uhpT* (B). Ceci aboutit à la mise en place d'un fort flux glycolytique qui stoppe l'expression d'*acs* (C). Une fois épuisé, le G6P ne peut plus activer l'expression d'*uhpT* (B). Le flux glycolytique chute et une nouvelle induction d'*acs* se produit (C).

# 10 Système synthétique de communication intercellulaire par l'AMPc

Pour comprendre le mécanisme responsable de la répression catabolique, nous avons développé de nombreux outils (fusions transcriptionnelles, vecteurs de sur-expression inductibles, construction de souches mutantes). Cependant, ces efforts ne nous ont pas permis de percer à jour le mécanisme responsable de la répression catabolique. En échange, la nature nous a accordé la « serendipity » c'est-à-dire la découverte au hasard qui rend heureux. En plaçant en contact un mutant $\Delta crp$ et un mutant $\Delta cyaA$, nous nous sommes rendu compte que ces deux souches réussissaient à communiquer grâce à l'AMPc extracellulaire. L'expérience est relativement simple : nous avons placé une gélose contenant des bactéries $\Delta crp$ (dite « sender ») en contact d'une gélose composée des bactéries $\Delta cyaA$ (dite « receiver ») contenant un bio-senseur luminescent très sensible capable de détecter la présence d'AMPc extracellulaire. Ce biosenseur correspond au promoteur *sdh* (Crp-AMPc dépendant) fusionné à l'operon *luxCDABE* (plasmide p-sdh-lux). Après quelques heures, la gélose contenant les bactéries $\Delta cyaA$ s'est mise à émettre de la lumière. L'AMPc produit par la souche « sender » a diffusé dans la seconde gélose puis a pénétré dans les bactéries $\Delta cyaA$, s'est fixé à Crp ce qui a permis l'activation du promoteur *sdh* puis l'émission de lumière (Figure I.10A). Pour parfaire la démonstration, nous avons montré que la gélose « receiver » ne s'allume jamais si l'adenylate cyclase (production de l'AMPc) est absente dans la gélose « sender ». De même, nous avons montré que la présence de Crp dans la gélose « receiver » est indispensable au processus de communication. Ces jeux avec le complexe Crp-AMPc nous ont permis d'illustrer « artistiquement », via la communication, le fonctionnement de cette jolie porte AND que constitue le complexe Crp-AMPc (Figure I.10B).

Pour caractériser plus précisément ce nouveau module de communication extracellulaire, nous avons également fait des expériences en milieu liquide. Ces expériences montrent que l'on peut contrôler très finement la communication en faisant varier la concentration d'adenylate cyclase dans les « senders » ou la concentration de Crp dans les « receivers ». Un des avantages du complexe Crp-AMPc c'est qu'il est déjà connecté à de nombreuses autres fonctions cellulaires. Par exemple, un mutant $\Delta cyaA$ souffre d'un défaut de motilité lié à son incapacité à fabriquer des flagelles. Nous avons pu rétablir, par communication intercellulaire, la fabrication de flagelles et donc la motilité de bactéries $\Delta cyaA$ créant de fait une sorte de système de guidage synthétique. Cela montre qu'il est possible de connecter aisément notre système de communication à d'autres modules (physiologiques ou synthétiques) présents dans la cellule. Ce processus de communication est-il physiologique ? L'AMPc secrété par la bactérie a-t-il

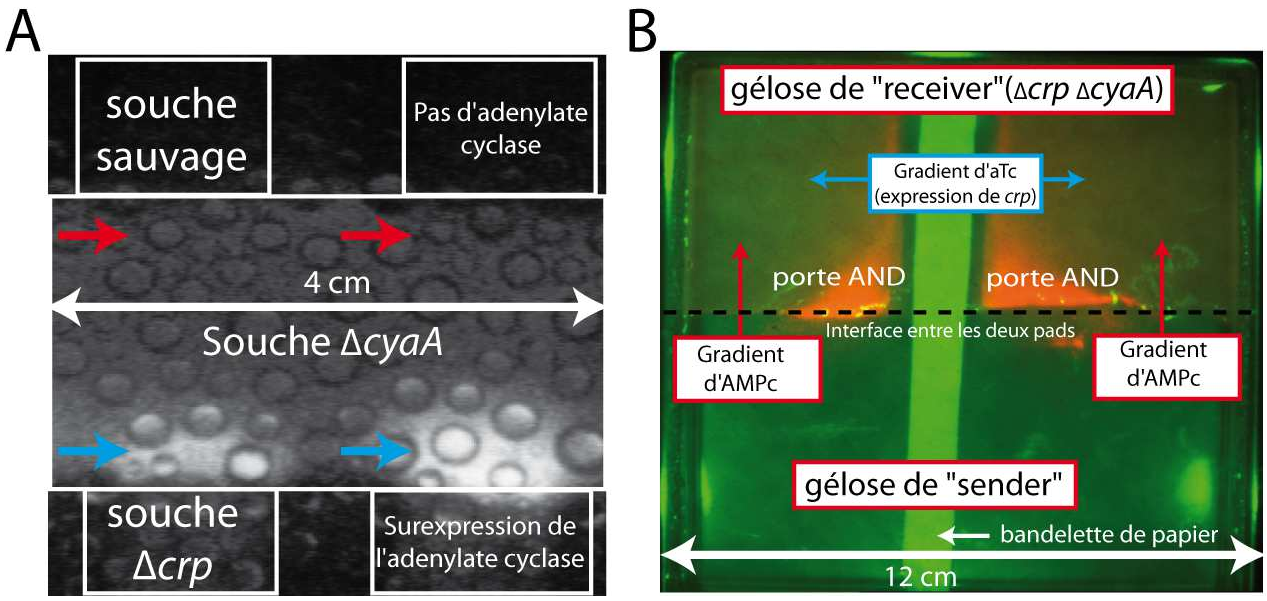



**Figure I.10** – Un système de communication intercellulaire via l'AMPc. A. Image de luminescence vue de dessus d'une gélose solide. La photo a été prise dans une chambre noire à l'aide d'une caméra sensible (les ronds sont des gouttes de condensation sur le couvercle). Au milieu de la photo, il y a une longue gélose contenant les cellules « receiver » produisant un niveau basal de lumière. En haut à gauche (flèche rouge), l'interface créée entre la gélose « receiver » et une gélose contenant des bactéries sauvages ne produit pas d'augmentation de lumière. A l'inverse, en bas à droite (flèche bleue), un halo de lumière est observé à l'interface entre la gélose « sender » contenant les bactéries $\Delta crp$ et la gélose « receiver ». En bas à droite (flèche bleue), des cellules « sender », sur-exprimant l'adenylate cyclase, déclenchent l'augmentation de lumière dans la gélose « receiver ». A l'inverse, sans expression d'adenylate cyclase, il n'y a pas de halo lumineux visible (en haut à droite, flèche rouge). B. Une photo d'une boite de petri et des données de luminescence (fausse couleur rouge) sont superposées. Une moitié de la boite de petri contient une gélose « sender » produisant de l'AMPc qui diffuse dans la gélose « receiver » (seconde moitié). Au milieu et perpendiculairement à l'interface, une mince bandelette de cellulose saturée avec un inducteur (aTc) permet l'expression de la protéine Crp dans les cellules « receiver ». Deux zones lumineuses (rouges) de part et d'autre de la bandelette apparaissent là où les concentrations d'AMPc et de Crp sont suffisamment élevées pour permettre la formation du complexe Crp-AMPc dans les cellules « receiver » et donc la transcription de la luciferase.

pour fonction la communication ? Nous ne sommes pas encore en mesure de répondre à cette question. Un résultat nous a particulièrement troublé : la limite basse de détection de l'AMPc extracellulaire de notre biosenseur ( 5-10$\mu$M) correspond exactement à la limite haute de production d'AMPc extracellulaire par une souche sauvage ( 1-10$\mu$M). C'est comme si les bactéries sauvages produisaient une concentration d'AMPc extracellulaire juste en dessous du seuil qui commencerait à modifier significativement l'activité du régulon Crp des autres bactéries. Nous n'avons pas d'explications à ce résultat surprenant.

# 11 Utiliser l'activité luciferase pour inférer l'état métabolique de la cellule

Le rapporteur luciferase a besoin de sources d'énergie cellulaires pour émettre des photons. Cette dépendance est souvent considérée comme un risque d'artefact majeur. Nous allons montrer qu'elle peut aussi être utilisée pour inférer de nombreuses informations sur l'état des cellules. Pour émettre des photons, la luciferase que nous utilisons (*Photorhabdus luminescens*) ne nécessite pas l'ajout de substrat dans le milieu de culture. En effet, ce dernier est produit directement dans la cellule à l'aide de trois enzymes (encodées par *luxC*, *luxD*, *luxE*). La luciferase (encodée par *luxA* et *luxB*) oxydes simultanément ce substrat et le $FMNH_2$ en présence d'oxygène pour émettre un photon à 490 nm. La production de $FMNH_2$ requiert le fonctionnement des réactions centrales du métabolisme du carbone. Ainsi l'émission de lumière n'est pas seulement corrélée à la concentration de la luciferase mais également à l'activité du métabolisme central. Toute perturbation dans cette activité affecte la production de lumière. Cet effet peut être observé lorsque les bactéries épuisent la source de carbone de leur milieu et rentrent en phase stationnaire. A ce moment précis, on observe une chute de l'activité luciferase (que nous appellerons dorénavant ADLA, Abrupt Decline of Luciferase Activity) due à une chute du pouvoir réducteur (Koga *et al.*, 2005). Il est facile de prouver que cette chute est liée à l'épuisement de la source d'énergie car l'ajout dans le milieu, après cet ADLA, de n'importe quelle source de carbone utilisable par les bactéries, rétablit quasi immédiatement l'émission de lumière (Figure I.11).

Pour pouvoir étudier cette dépendance énergétique de la luciferase, nous nous sommes tout d'abord affranchi de toutes régulations transcriptionnelles. Nous avons fusionné un promoteur synthétique indépendant (tet) à l'operon *luxCDABE* (plasmide p-tet-lux). Avec cette construction, l'expression de la luciferase est sous notre contrôle : toutes les variations mesurées dans le signal de luminescence proviennent de changements d'états métaboliques. En faisant pousser des bactéries *E. coli* sauvages transformées avec ce senseur métabolique dans un milieu minimum contenant du glucose et de l'acétate, on observe deux phénomènes intéressants (figure I.12A) :

⇒ il y a une légère diminution transitoire de luminescence lors de la transition glucose-acétate. Cet événement suggère une diminution transitoire du fonctionnement du métabolisme central.

⇒ il y a une chute brutale de l'activité luciferase lorsque l'acétate est épuisé, c'est-à-dire lorsqu'il n'y a plus aucune source de carbone dans le milieu.

Cette méthode de mesure de l'état métabolique des cellules à l'aide d'un « biosenseur » est

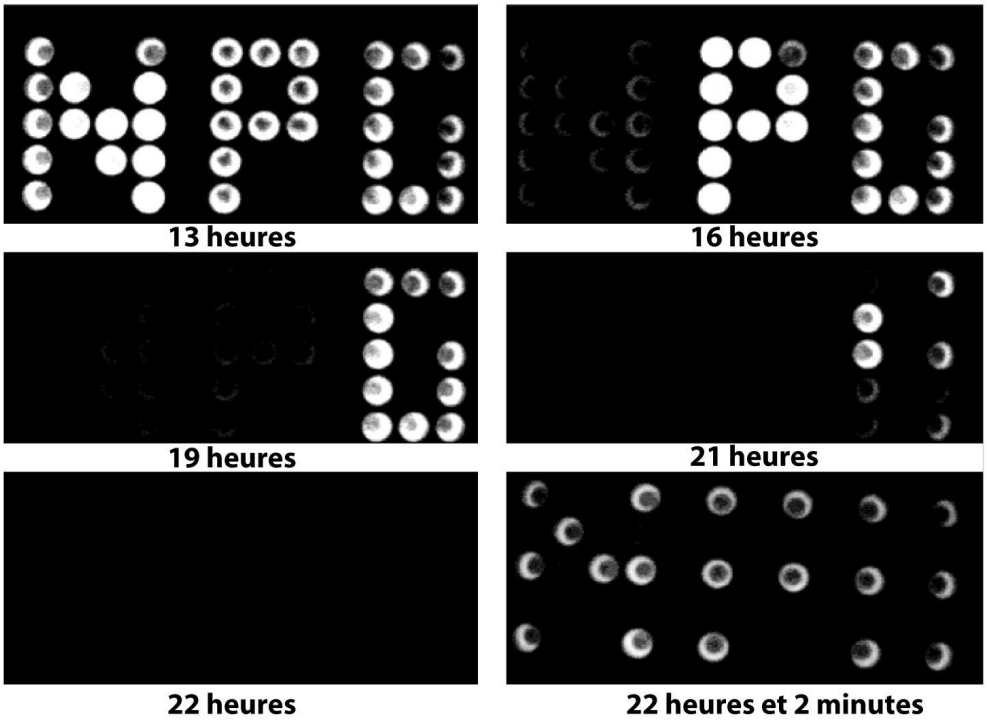



**Figure I.11** – La dépendance énergétique de la luciferase. Images de luminescence d'une microplaque 96-puits vue de dessus. Les lettres N, P, G sont formées par le regroupement de plusieurs puits. Chaque puits contient un milieu minimum avec du glucose et des bactéries *E. coli* exprimant l'operon *luxCDABE*. La concentration du glucose augmente de la gauche vers la droite. Par conséquent, sa consommation prend de plus en plus de temps. Quand la source de carbone externe est épuisée, la production d'énergie s'arrête et les lettres s'éteignent les unes après les autres (fonction de la concentration de glucose). L'ajout de glucose dans le milieu (ici dans un puits sur deux) permet le retour immédiat (∼minutes) du signal de luminescence ce qui démontre que les enzymes « luciferases » sont toujours présentes dans la cellule.

intéressante mais elle ne dévoile toute sa puissance que lorsqu'elle est couplée à une banque de simples mutants. Nous avons donc transformé des centaines de mutants avec notre biosenseur à la recherche de signaux qui diffèrent de la souche sauvage.

A titre d'exemple, le profil de luminescence d'un mutant *sucB*, déleté d'un gène codant pour une enzyme du cycle de Krebs, montre une chute brutale de l'activité luciferase non pas lors de l'épuisement de l'acétate (comme c'est le cas pour la souche sauvage) mais lors de l'épuisement du glucose (figure I.12B). En effet, ce mutant ne peut pas utiliser l'acétate comme source de carbone car la néoglucogenèse est coupée. Une souche *sdhB* est aussi déletée d'un gène codant pour une enzyme du cycle de Krebs : on observe aussi une chute de l'activité luciferase lorsque le glucose est épuisé. Cependant l'ampleur de cette chute est bien moins prononcée que celle observée avec le mutant *sucB*. Il est très probable que cette différence de phénotype soit significative et mette en exergue l'existence d'une petite fuite métabolique permettant de

maintenir un fonctionnement minimal du cycle de Krebs et donc une petite activité luciferase dans le mutant *sdhB*. Ainsi nous pensons que la capacité à maintenir l'activité luciferase associée à la forte sensibilité de détection de la lumière permettrait peut-être de détecter et de quantifier des variations fines dans les flux métaboliques. Menée de manière systématique (grand nombre de mutants), cette manière de procéder pourrait peut-être permettre de redessiner les flux métaboliques de manière plus quantitative dans une condition donnée.

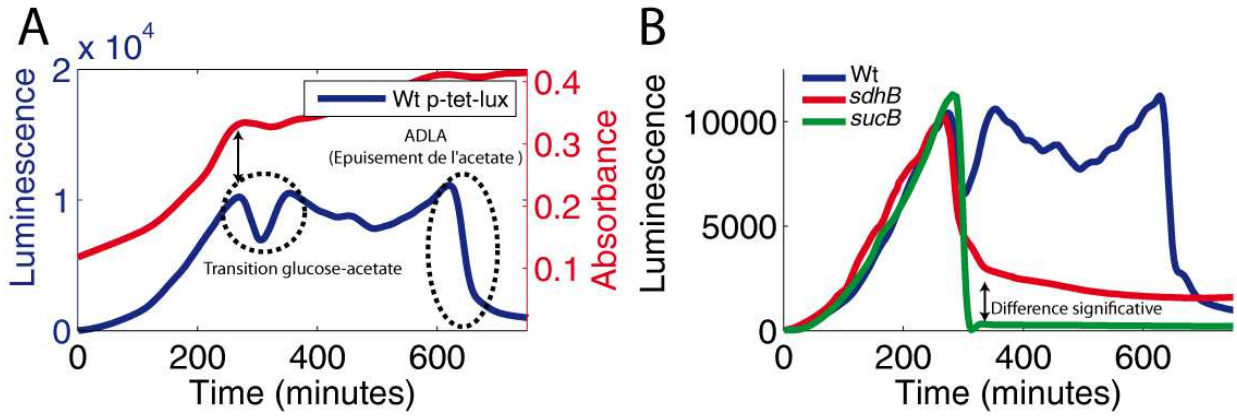



**Figure I.12** – Capacité à maintenir l'activité luciferase durant la croissance sur glucose et acétate dans différent mutants. A. Une souche sauvage transformée avec le plasmide p-tet-lux pousse dans un milieu minimum avec du glucose, de l'acétate et l'anhydrotetracycline (l'inducteur du promoteur tet). Quand le glucose est épuisé, il y a une légère diminution suivi du rétablissement de l'activité luciferase (courbe bleue). En effet, cet événement survient au même moment que la cassure visible sur la courbe de croissance (courbe rouge). Plus tard, lorsque l'acétate est épuisé, il y a une chute brutale de l'activité luciferase. B. Dans un mutant *sucB*, cette chute brutale de lumière n'a pas lieu lors de l'épuisement de l'acétate mais lors de l'épuisement du glucose. En effet, la délétion du gène *sucB* empêche la néoglucogenèse et donc la possibilité d'utiliser l'acétate comme source d'énergie. Dans un mutant *sdhB*, la capacité à maintenir l'activité luciferase est sévèrement diminuée par rapport à la souche sauvage mais significativement supérieure au mutant *sucB* : il subsiste sans doute une petite fuite métabolique permettant quand même l'utilisation de l'acétate comme source d'énergie.

La capacité à maintenir l'activité luciferase permet de se focaliser sur une enzyme, un flux spécifique. Ainsi nous avons pu mesurer, en temps réel, l'activité relative d'une enzyme : l'acétyl-coenzyme A synthétase encodée par le gène *acs*. L'idée consiste à créer et contrôler un goulot d'étranglement métabolique dans des bactéries saturées en luciferase. Nous avons donc supprimé toute trace d'acétyl-coenzyme A synthétase dans des bactéries placées dans un milieu minimum avec de l'acétate. Dans ces conditions, ces bactéries ne peuvent pas utiliser l'acétate : le métabolisme central est inactif. La luciferase, pourtant présente en forte concentration, ne peut pas émettre de photons. Puis nous rouvrons « la vanne » lentement : l'arrivée des premières enzymes d'acetyl-coenzyme A synthétase réamorce le métabolisme central qui peut alors utiliser l'acétate comme source d'énergie : la brusque remontée du pouvoir réducteur entraine

le rétablissement de l'activité luciferase et le niveau de ce rétablissement est directement corrélé à la concentration de l'acetyl-coenzyme A synthétase. Avec cette méthode, il est donc possible de connaître son activité relative, directement *in vivo* et en temps réel.

Enfin, l'utilisation du promoteur *acs* fusionné à l'operon *luxCDABE* (plasmide p-acs-lux) nous a permis d'inférer certains paramètres extérieurs, encore une fois directement et en temps réel. Nous avons vu que l'épuisement du glucose lève la répression exercée sur ce promoteur. Dans un milieu contenant du glucose et de l'acétate, l'induction de luminescence indique donc l'épuisement du glucose (Figure I.13A). A l'inverse, la chute brutale de l'activité luciferase signale l'épuisement de l'acétate (Figure I.13B). En associant ces deux événements, nous pensons qu'il est possible de mesurer la quantité relative d'acétate secrété par les bactéries dans le milieu c'est-à-dire le débordement du métabolisme (overflow metabolism). Lorsque les bactéries poussent sur glucose, le métabolisme central est saturé et une partie de l'énergie est évacuée vers l'extérieur sous forme d'acétate en vu d'une utilisation future. Pour mesurer ce phénomène, nous avons fait pousser des bactéries dans un milieu contenant uniquement du glucose. Nous avons remarqué que le temps entre l'induction de la luciferase (épuisement du glucose) et la chute de l'activité luciferase (épuisement de l'acétate) augmente proportionnellement avec la concentration de glucose utilisée (Figure I.13C). L'explication la plus raisonnable pour expliquer l'accroissement de ce délai est l'augmentation de la concentration extérieure d'acétate. Pour étayer cette hypothèse, nous avons testé des mutants (*ptsG*, *pgi*) qui ralentissent la consommation du glucose (et donc la glycolyse) ce qui limite le débordement du métabolisme. Dans ces mutants et pour une même concentration de glucose, le délai entre l'induction de la luciferase et la chute de l'activité luciferase est beaucoup plus court comparé à une souche sauvage (Figure I.13D). Ce résultat suggère, comme escompté, que la sécrétion d'acétate dans ces mutants est d'une ampleur beaucoup plus limitée.

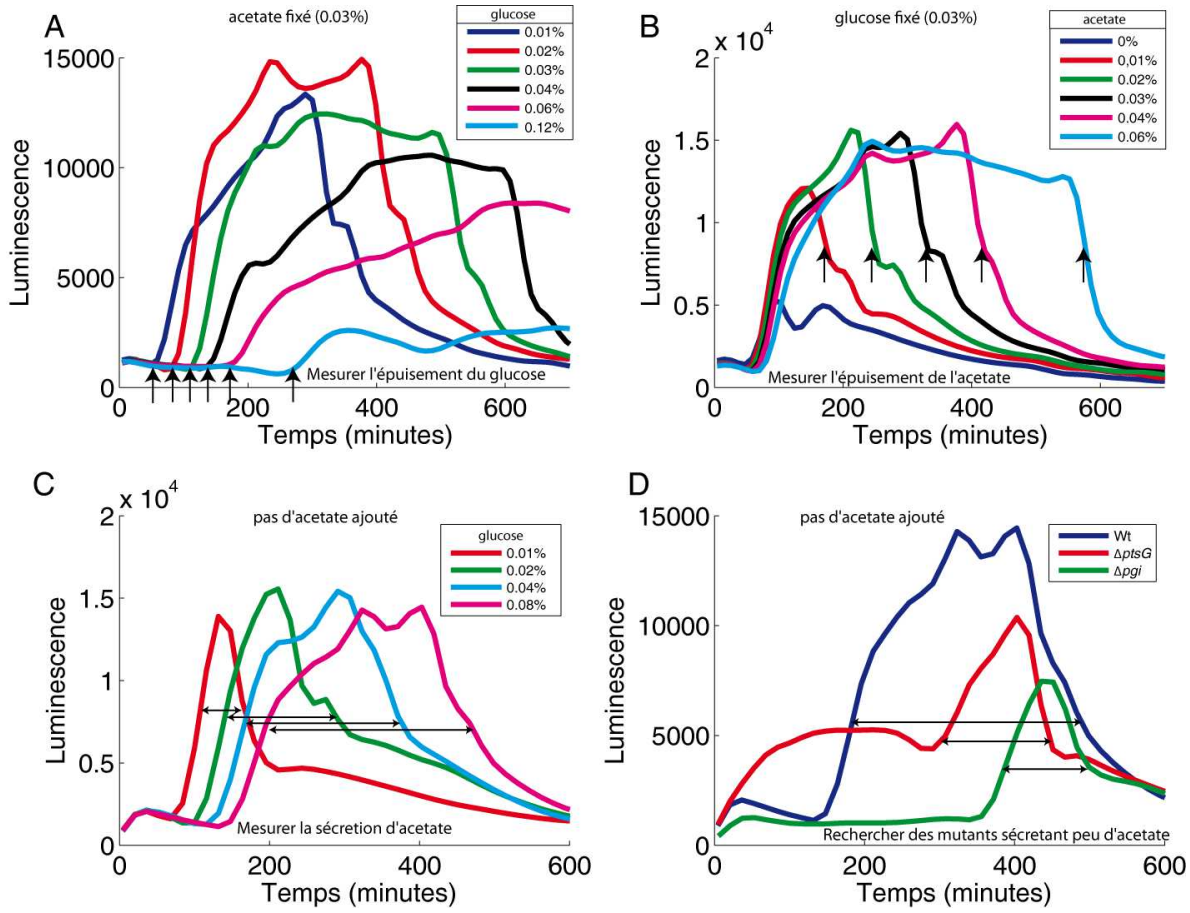



**Figure I.13** – Un outil pour mesurer la concentration extérieure de glucose, d'acétate et le débordement du métabolisme. A. Le plasmide p-acs-lux peut être utilisé pour mesurer l'épuisement du glucose (la concentration d'acétate est fixée). En effet, une forte augmentation de luminescence se produit lorsque le glucose est épuisé (de manière dose-dépendante). B. Le plasmide p-acs-lux peut être utilisé pour mesurer l'épuisement de l'acétate (la concentration du glucose est fixée). En effet, le temps auquel survient la chute de lumière est proportionnel à la concentration d'acétate. C. Des bactéries poussent dans un milieu avec du glucose mais sans acétate. Quand la concentration de glucose augmente, le délai entre l'induction de la luminescence (épuisement du glucose) et la chute brutale de la luminescence (épuisement de l'acétate) augmente proportionnellement. Ce délai pourrait mettre en évidence l'acétate sécrété dans le milieu dû au débordement du métabolisme durant la croissance sur glucose. D. Pour une même concentration de glucose, des mutants avec un flux glycolytique réduit (*ptsG*, *pgi*) ont un délai entre l'induction de la luminescence et sa chute brutale qui est réduit par rapport à celui observé avec la souche sauvage. La réduction de ce délai pourrait mettre en exergue une diminution de la sécrétion d'acétate et donc un débordement du métabolisme plus faible.

# Références bibliographiques


Crasnier-Mednansky, M. 2008. Is there any role for cAMP-CRP in carbon catabolite repression of the Escherichia coli lac operon ? *Nat Rev Microbiol*, **6**(12), 954 ; author reply 954. 13

Gorke, B., & Stulke, J. 2008. Is there any role for cAMP-CRP in carbon catabolite repression of the Escherichia coli lac operon ? Reply from Gorke and Stulke. *Nat Rev Microbiol*, **6**(12), 954. 13

Inada, T., Kimata, K., & Aiba, H. 1996. Mechanism responsible for glucose-lactose diauxie in Escherichia coli : challenge to the cAMP model. *Genes Cells*, **1**(3), 293–301. 13

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P., & Karp, P. D. 2010. EcoCyc : a comprehensive database of Escherichia coli biology. *Nucleic Acids Res*, **39**(Database issue), D583–90. 1

Khankal, R., Chin, J. W., Ghosh, D., & Cirino, P. C. 2009. Transcriptional effects of CRP* expression in Escherichia coli. *J Biol Eng*, **3**(1), 13. 8

Koga, K., Harada, T., Shimizu, H., & Tanaka, K. 2005. Bacterial luciferase activity and the intracellular redox pool in Escherichia coli. *Mol Genet Genomics*, **274**(2), 180–8. 17

Kumari, S., Tishel, R., Eisenbach, M., & Wolfe, A. J. 1995. Cloning, characterization, and functional expression of acs, the gene which encodes acetyl coenzyme A synthetase in Escherichia coli. *J Bacteriol*, **177**(10), 2878–86. 2

Kumari, S., Beatty, C. M., Browning, D. F., Busby, S. J., Simel, E. J., Hovel-Miner, G., & Wolfe, A. J. 2000. Regulation of acetyl coenzyme A synthetase in Escherichia coli. *J Bacteriol*, **182**(15), 4173–9. 2

Megerle, J. A., Fritz, G., Gerland, U., Jung, K., & Radler, J. O. 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys J*, **95**(4), 2103–15. 12

Meighen, E. A. 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev*, **55**(1), 123–42. 6

Narang, A. 2009. cAMP does not have an important role in carbon catabolite repression of the Escherichia coli lac operon. *Nat Rev Microbiol*, **7**(3), 250. 13

Oh, M. K., Rohlin, L., Kao, K. C., & Liao, J. C. 2002. Global expression profiling of acetate-grown Escherichia coli. *J Biol Chem*, **277**(15), 13175–83. 2

Saier, M. H., & Roseman, S. 1972. Inducer exclusion and repression of enzyme synthesis in mutants of Salmonella typhimurium defective in enzyme I of the phosphoenolpyruvate : sugar phosphotransferase system. *J Biol Chem*, **247**(3), 972–5. 8

Veening, J. W., Smits, W. K., & Kuipers, O. P. 2008. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol*, **62**, 193–210. 2

Wolfe, A. J. 2005. The acetate switch. *Microbiol Mol Biol Rev*, **69**(1), 12–50. 2

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods*, **3**(8), 623–8. 10

# Chapitre II

A genome-wide screen for identifying all regulators of a target gene

# A genome-wide screen for identifying all regulators of a target gene

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

* E-mail: Corresponding author@institute.edu

## Abstract

We have developed a new screening methodology for identifying all genes that control the expression of a target gene through genetic or metabolic interactions. The screen combines mutant libraries with luciferase reporter constructs. Instead of a static picture of gene expression, this method allows dynamical monitoring in different environmental conditions. Mutants with interesting phenotypes can thus be selected based on multiple criteria, and the expression dynamics of the target gene can be extensively characterized. We apply the method to identify the genes that control the expression of the *acs* gene of *Escherichia coli*. We confirm most of the known genetic regulators and identify new regulatory influences, many of which involve metabolic intermediates or metabolic sensing. An analysis of mutants involved in glycolysis and glucose transport demonstrates that the classical model of catabolite repression in *E. coli* needs to be amended.

## Introduction

The adaptation of bacteria to changes in their environment is controlled on the molecular level by a large and complex regulatory network involving genes, mRNAs, proteins and metabolites (Jacob & Monod, 1961). Understanding, and therefore eventually predicting, the dynamics of this regulatory network is a general goal of biology, and in particular of the field of systems biology. The basic ingredient of this type of research is knowing the topology of the underlying regulatory network. An increasing number of such regulatory connections are documented in bioinformatics databases such as EcoCyc (`http://EcoCyc.org`), which lists more than 5345 regulatory interactions for *E. coli* (Keseler *et al.*,

2010).

Different methods are used to identify and characterize these interactions. Evidence for specific interactions can come from in vitro measurement of physical DNA-protein interactions (footprinting, gel retardation assay). Computational tools can predict interactions based on consensus sequence. Transcriptomics (DNA microarrays) is a popular approach to study transcription on a global scale (Brown & Botstein, 1999). Typically, this technique allows detecting the targets of a particular regulator by studying the effect of the deletion of this regulator on the expression of all genes of a genome. ChIP-Chip is another global approach, identifying direct interactions of a specific transcription factor through the detection of chromosome-wide DNA binding (Grainger & Busby, 2008). All these tools aim at identifying regulons, i.e., all the targets of a given regulator (Neidhardt & Savageau, 1996). However, to our knowledge, no global method exists to determine the regulators that control, directly or indirectly, the expression of a particular target gene. We present such a method in this manuscript.

We use the GFP and luciferase reporters to measure promoter activities in living cells (Van Dyk *et al.*, 2001; Hakkila *et al.*, 2002). These constructs provide not only a static picture of expression levels, but allow measuring the dynamics of gene expression with a high temporal resolution (Zaslaver *et al.*, 2006). The luciferase system is particularly useful since it allows measuring gene expression in a colony growing on solid medium. We have developed a technique for efficiently transforming 3571 *E. coli* single-gene knockout mutants (Baba *et al.*, 2006) with the appropriate reporter plasmid. The screening strategy consists in comparing the luminescence of mutant colonies with the luminescence of a wild-type colony: a significant difference suggests that the deleted gene controls, directly or indirectly, the expression of the target gene.

We have applied this method to identify regulators that control the expression of the *acs* gene, coding for the acetyl coenzyme A synthetase (Kumari *et al.*, 1995). This enzyme converts acetate to acetyl coenzyme A. We have chosen this gene for three reasons. First, this enzyme plays a key role in the "acetate switch": when bacteria grow rapidly on glucose, they excrete acetate which is subsequently used as a carbon source when glucose is exhausted (Wolfe, 2005). Acetyl coenzyme A synthetase is thus at the center of an inversion of metabolic flow from glycolysis to neoglucogenesis. Second, the accumulation of acetate in the culture medium is an important problem in industrial fermentations since this organic acid inhibits cell growth and recombinant protein production. Different approaches are currently developed to reduce acetate accumulation by modifying central metabolic pathways (Gosset, 2005). Lastly, the *acs* gene

was also selected because its complex regulation has been extensively described and thus allows a thorough validation of the method (Kumari *et al.*, 2000; Sclavi *et al.*, 2007; Shin *et al.*, 2009). Briefly, transcription of this gene is repressed by the nucleoproteins Fis and IHF and positively regulated by the CRP-cAMP complex. The latter activation implies the well-studied carbon catabolite repression mechanism where the glucose concentration controls the formation of the CRP-cAMP complex by adjusting the cAMP production via the phosphotransferase system (PTS) (Gorke & Stulke, 2008a). The *acs* gene is thus at the end of a well-known, complex regulatory cascade, an ideal property for our purpose. Our screen confirms known regulatory mechanisms of *acs* expression. We also find new candidates for regulatory input that highlight the tight links between metabolism and gene expression.

# Results

## A screening strategy for identifying the regulators of *acs*

Our screening strategy is outlined in Fig. 1. To measure *acs* promoter activity we have constructed a reporter plasmid containing a transcriptional fusion of the *acs* promoter region (394 bp) to the *luxCDABE* operon. This operon codes for the heterodimeric luciferase protein (*luxAB*) and the enzymes (*luxCDE*) necessary for producing the luciferase substrate, a long-chain aldehyde (Meighen, 1991). We have chosen this reporter system because of the highly sensitive luminescence signal, the absence of background, the continuous production of bioluminescence without added substrate, and the possibility to measure luminescence at the colony level. The reporter plasmid was transformed into the 3911 single-gene knockouts of *Escherichia coli* (Keio collection) using our high-throughput transformation procedure (see Methods). Transformation efficiency was 91%: we have thus obtained 3571 transformants. We spotted the transformed bacteria on solid media and acquired the luminescence images of the colonies that appeared on the plates.

The luminescence of the colonies changes as a function of culture time and growth condition. The powerful features of our screen are that we can monitor the dynamics of gene expression at the colony level (see an example in Supplementary Information 2: Movie) and adapt growth conditions to be most relevant to the gene under investigation. The *acs* promoter is activated after exhaustion of the preferred carbon source. We therefore measured the luminosity of colonies after 24h and 48h of growth on two different media: rich LB medium and minimal glucose medium. We thus obtained four complete sets

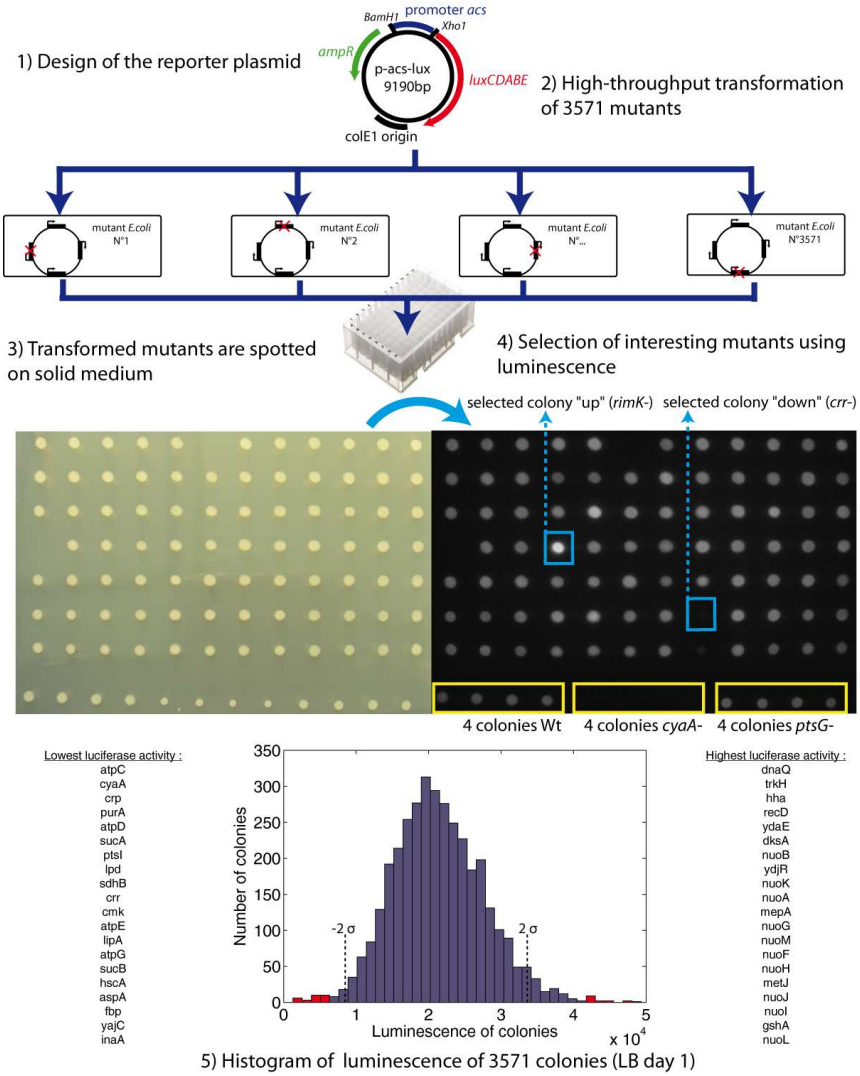



**Figure 1. A screening strategy for identifying new regulators.** The *acs* promoter region was cloned upstream of the *luxCDABE* operon on a low-copy number plasmid. This plasmid is transformed into 3571 single knock-out mutants of *E. coli* (top). The transformed bacteria were spotted onto solid medium and the luminescence of the colonies was quantified (LB-agar day 1) (middle). The histogram (bottom) shows the distribution of luminescence emitted by all colonies. The names of the 20 mutants with the highest/lowest luciferase activity (in red on the histogram) are listed on the sides. In these mutants, light emission is significantly different from the wild-type (0.95 confidence), as determined by fitting a normal distribution to the histogram and noting that the mean of the distribution corresponds to wild-type levels. The broken vertical lines indicate two standard deviations above and below the mean. Among the mutants with the lowest luciferase activity, we find 7 genes (*atpC, atpD, atpE, atpG, sucA, sucB*, *sdhB, lpd*A) involved in energy supply and four genes involved in catabolite repression (*cyaA*, *crp*, *crr*, *ptsI*). Among the mutants with the highest luciferase activity, our screen detects a transcriptional regulator (*hha*) and 10 genes of the *nuo* operon.

of luminosity images of the plates. We quantified the luminosities of the 3571 colonies after 24h of growth on LB plates (Supplementary Data). Fig. 1 shows the corresponding distribution of luminosities. Automatic quantification was not possible for the other growth conditions, but the complete dataset is available in Supplementary Data. From these four batches of luminescence pictures, we have selected 517 (15%) colonies for a detailed kinetic analysis. We also include data about the growth phenotype: among the 3571 transformants, eighty auxotrophic mutants did not grow on minimal medium-agar-glucose (Supplementary Information 4).

## The dynamics of *acs* expression

For the 517 pre-selected genes, we measured the dynamics of *acs* expression in liquid cultures grown in a minimal medium containing glucose and acetate in order to focus on the physiologically important transition between those two carbons sources. The expression patterns were measured in an automated microplate reader and analyzed as previously described (de Jong *et al.*, 2010). Fig. 2A shows the expression profile of *acs* in the wild-type strain. We observe a sharp increase of *acs* transcription when glucose is exhausted. This expression pattern is very robust and independent of the reporter system since we observe an identical pattern with the *gfp* reporter carried on a plasmid with a completely different replication mechanism (Fig. 2B). Note that luminescence is largely preferable for screening because bacteria have a significant auto-fluorescence background, probably due to flavin mononucleotide or riboflavin (Billinton & Knight, 2001).

## Known regulators

An important criterion for the validity of a screen is that known interactions should be confirmed. The *acs* gene is positively regulated by the CRP-cAMP complex (*crp* gene) and thus by the well-studied carbon catabolite repression mechanism. The influx of glucose through the phosphotransferase system (PTS) controls adenylate cyclase (*cyaA* gene) activity and thus cAMP production. The PTS is composed of enzymes EI (*ptsI* gene), HPR (*ptsH* gene), $EIIA_{glc}$ (*crr* gene) and $EIIB/C_{glc}$ (the glucose transporter encoded by the *ptsG* gene). When glucose is exhausted, the four PTS enzymes are phosphorylated and $EIIA_{glc}$ binds to and activates adenylate cyclase. The cAMP concentration quickly rises and the Crp-cAMP complex activates target promoters such as the *acs* promoter (Gorke & Stulke, 2008a) (Fig. 5).

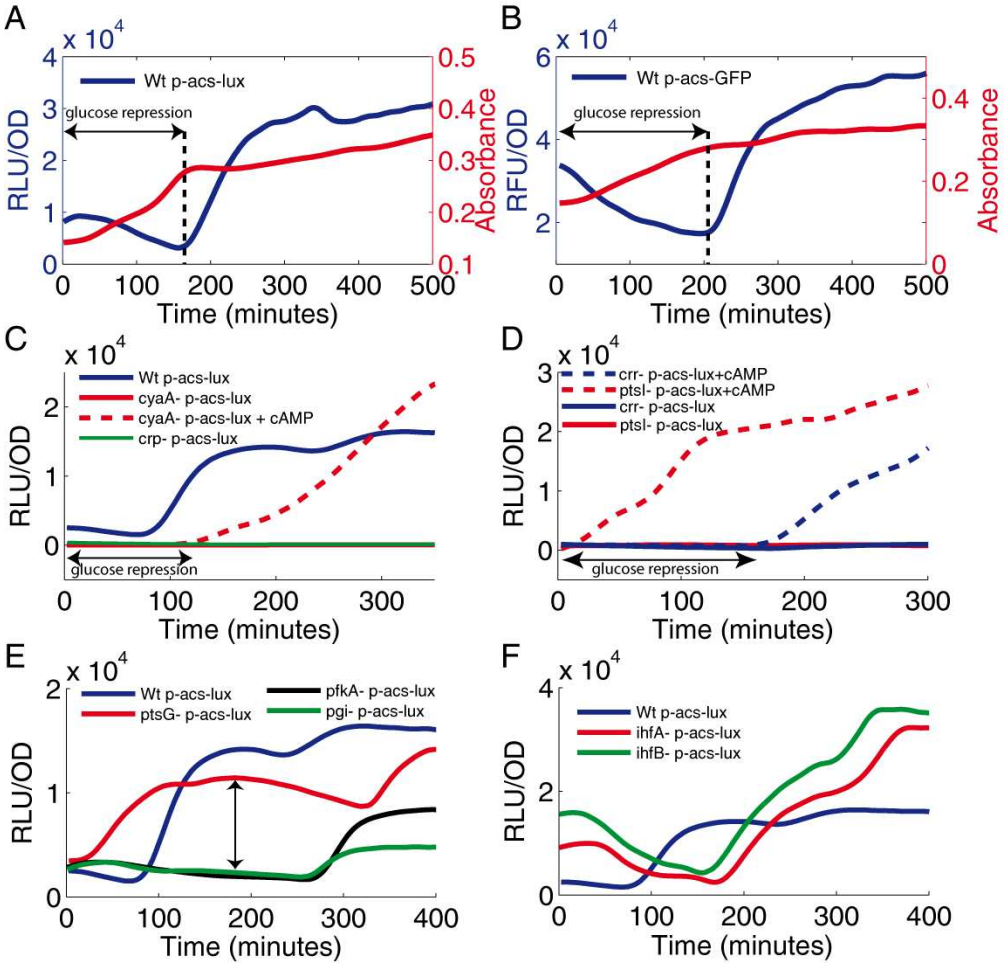



**Figure 2. Dynamics of gene expression of known mutants in a liquid minimal medium containing glucose and acetate.** A. The expression profile of the *acs* luciferase reporter plasmid shows that *acs* expression is induced when glucose is exhausted. B. The *acs* expression profile using GFP as a reporter is identical to the profile observed with luciferase (A). C. *acs* expression is completely absent in *crp* and *cyaA* strains. cAMP added to the medium (5mM) at the beginning of the experiment complements the *cyaA* strain but glucose repression is maintained. D. Very low expression of *acs* in the *crr* and *ptsI* strains. cAMP added to the medium at the beginning of the experiment directly complements the *ptsI* strain during growth on glucose, whereas the *crr* strain only responds once glucose is exhausted. E. In the *ptsG* mutant, *acs* induction is delayed because glucose is consumed more slowly. Moreover, *acs* expression during glucose consumption is much stronger than in the wild-type strain. The glycolysis mutants *pgi* and *pfkA* also grow more slowly, but *acs* expression remains very low in exponential phase. F. The expression profiles of the *ihfA* and *ihfB* strains show stronger *acs* expression than the wild-type during growth on acetate. The highly similar expression kinetics of the two mutants (with genes located at different sites on the chromosome) shows the high internal consistency of the method.

The analysis of the genes involved in glucose repression confirms that our screen reliably identifies dramatic, but also subtle changes in the regulation of the target gene. In *cyaA* and *crp* strains, there is no transcription of *acs* at all, demonstrating the strict dependence of *acs* transcription on the CRP-cAMP complex. The activity of the *acs* promoter can be restored in the *cyaA* strain by adding cAMP to the medium but, interestingly, only after glucose exhaustion (Fig. 2C). The *crr* and *ptsI* strains show the expected phenotype: a very low *acs* expression, although significantly higher than in the *cyaA* and *crp* strains. cAMP added to the medium clearly restores *acs* expression in these two strains confirming that these mutants cannot activate adenylate cyclase. As in the *cyaA* strain, cAMP fails to complement the *crr* strain while glucose is still present in the growth medium. In contrast, the *ptsI* strain is directly complemented by cAMP even in presence of glucose (Fig. 2D). This difference of reactivity is not explained by the current model of catabolite repression: adding cAMP to the medium should bypass the glucose repression. Consequently, the *cyaA* and *crr* mutants should behave as the *ptsI* strain, with a direct activation of *acs* even in presence of glucose. The screen has thus revealed new subtleties of a well characterized control system.

Expression of *acs* is reported to be negatively regulated by two histone-like proteins Fis and IHF (Browning *et al.*, 2004). The Fis protein is abundant in exponential phase (Ali Azam *et al.*, 1999). Since *acs* expression is already very low in this growth phase, the observed effect of the *fis* deletion is rather subtle: the *acs* expression profile is rather similar to that of the wild-type strain (data not shown). IHF acts mainly in stationary phase (Wolfe, 2005) and we accordingly observe a higher *acs* expression in late stationary phase (Fig. 2F). The perfect superposition of *acs* expression profiles observed in the *ihfA* and *ihfB* strains (each coding for a subunit of IHF and transcribed from different chromosomal locations) confirms the high accuracy and internal consistency of the method.

## Metabolic influences on the reporter system

For the purpose of our screen, we want luciferase to report only the changes in gene expression of the target gene. However, the emission of light produced by the colonies does not only depend on *luxCDABE* transcription, but is also influenced by the metabolic state of the cell, such as concentrations of oxygen, $FMNH_2$, ATP, and NADPH. These metabolites are involved in the enzymatic reactions that lead to light emission (Koga *et al.*, 2005). The *nuo* operon (13 genes) is a striking example of these metabolic effects on luciferase activity: all corresponding colonies show a higher luminescence on LB plates (Fig. 3A).

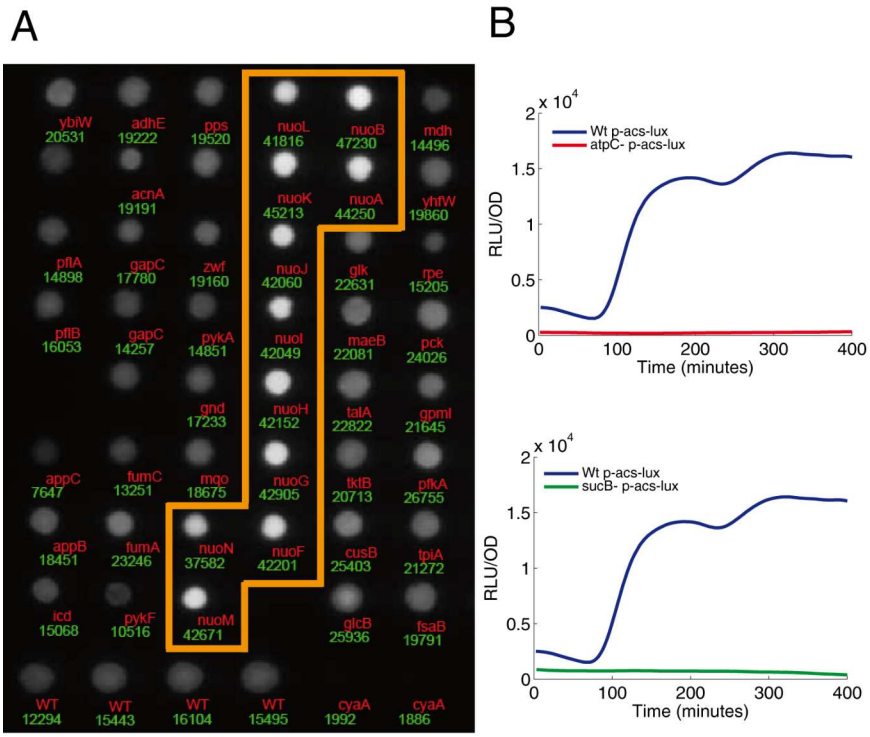



**Figure 3. Luciferase and the energy supply.** A. Higher level of luminescence of different mutants of the *nuo* operon. This effect is probably due to the dependence of the luciferase reaction on reducing power. The *nuo* operon codes for the NADH dehydrogenase involved in the respiratory chain. In the absence of a functional enzyme, reducing power accumulates and thus stimulates the luciferase reaction. The uniform increase in luminescence of all genes in the operon illustrates the high reproducibility of the screen. B. The *atpC* and *sucB* mutants do not induce *acs* (producing light) when glucose is exhausted.

The *nuo* operon codes for the NADH dehydrogenase involved in the respiratory chain. In the absence of a functional enzyme, reducing power accumulates (Pruss *et al.*, 1994), thus stimulating the luciferase reaction. This phenotype of the *nuo* mutants was not observed during the kinetics in liquid medium. In fact, it is the combination of a specific deletion with a specific medium that leads to this metabolic artifact. Consequently, it is not possible to generate a universal list of false positive mutants.

However, a systematic way to detect and eliminate false positives due to the reporter system is to validate the results with a different reporter system that is independent of the metabolic state of the cell. We therefore transformed the 517 selected mutants with the *gfp* reporter plasmid, also carrying a different origin of replication (thus controlling for mutants that would affect the plasmid copy number). We acquired the expression profile of *acs* for all selected mutants using the two reporter systems. Fig. 4A shows, for each reporter, the strength of *acs* induction, computed from the area under the induction curve. Mutants for which *acs* activation is very different in the two reporter systems (points far off the diagonal) likely represent artifacts of at least one of the reporter systems. We report, for example, 14 false positive mutants (at the left border of the scatterplot), probably affecting light production rather than diminished *acs* expression.

As a further control for the validity of the reporter assay, we measured transcription for several mutants using quantitative RT-PCR. Mutants in genes coding for the PTS (*ptsG*, *crr*, *ptsI*), cAMP production (*cyaA*), regulatory proteins (*ihfB*) and metabolic enzymes (*lpdA*, *sucB*) precisely reproduce the effects observed with the reporter genes (Fig. 4B).

## Metabolic regulation of *acs* expression

Even though part of the PTS, the *ptsG* mutant (which lacks the glucose transporter) behaves very differently from the *ptsI* and *crr* mutants: in the *ptsG* strain, colonies on M9 glucose are much more luminescent than those of the wild-type strain (Supplementary data) and the dynamical analysis shows a higher *acs* expression level in exponential phase as compared to the wild-type strain (Fig. 2E). According to the classical model of carbon catabolite repression, this phenotype reflects an increased intracellular concentration of cAMP in exponential phase as a consequence of the lower rate of glucose influx (Hogema *et al.*, 1998). The deletion of metabolic enzymes (*pgi*, *pfkA*) interrupts glycolysis and forces the metabolic flux to pass though the pentose phosphate pathway. Like the *ptsG* strain, these mutants grow more slowly on glucose (delayed induction of *acs* expression), but in this case, *acs* expression remains completely shut-

off in exponential phase (Fig. 2E).

Several of the mutants with the lowest colony luminescence carry deletions of genes involved in the energy supply of the cell (Fig. 1), such as the genes coding for ATP synthetase (*atpC, atpD, atpE, atpF*) and the genes coding for enzymes of the TCA cycle: the oxoglutarate dehydrogenase complex (*sucA*, *sucB* and *lpdA*) and, to a lesser extent, succinyl-CoA synthetase (*sucCD*) and succinate dehydrogenase (*sdhCDAB*). We found that theses mutants fail to grow on acetate (Supplementary Information 4). In these mutants, the luminescence profile shows an almost total absence of *acs* expression (Fig. 3B). This result may either represent (i) an artifact of the luminescence reporter system or (ii) a genuine regulation of the *acs* promoter. Our current data do not allow us to definitively decide between the two possibilities: the nature of the deleted genes (enzymes involved in the energy supply) strongly suggest a mebabolic artifact but the following observations argue for a genuine regulation: the effect is reproduced with the fluorescence reporter and qRT-PCR (Fig. 4) and expression is absent even in rich medium where energy supply should not play a role (Supplementary Data).

# Discussion

## A new method for identifying the regulators of *acs*

The classical way to dissect a regulatory network, closely linked to classical approaches in genetics, consists in disrupting a gene of interest and observing the effects on the organism. This approach identifies the downstream targets of the gene of interest. Here, we have developed a method to do the opposite: given a gene of interest, we can identify all factors, genetic and metabolic, that affect the expression of this gene. Our method relies on a high-throughput transformation of a luciferase reporter plasmid into an extensive mutant collection. Even though we have focused on transcriptional regulation, the method can certainly be extended to include translational fusions. We use the Keio-mutant collection, but the scope of the screen could be extended to include, e.g., mutants of non-coding RNAs and intergenic regions that may contain unknown ORFs. We have chosen luciferase as a reporter because it allows easy screening at the colony level without any significant background signal. Furthermore, the luciferase signal of a colony is highly reproducible, with a standard deviation of about 5% for independent colonies of the same mutant. The enzymatic reaction of luciferase depends on cellular factors such as ATP concentration or redox potential. For a given medium, deletions that seriously affect the concentration of these factors may

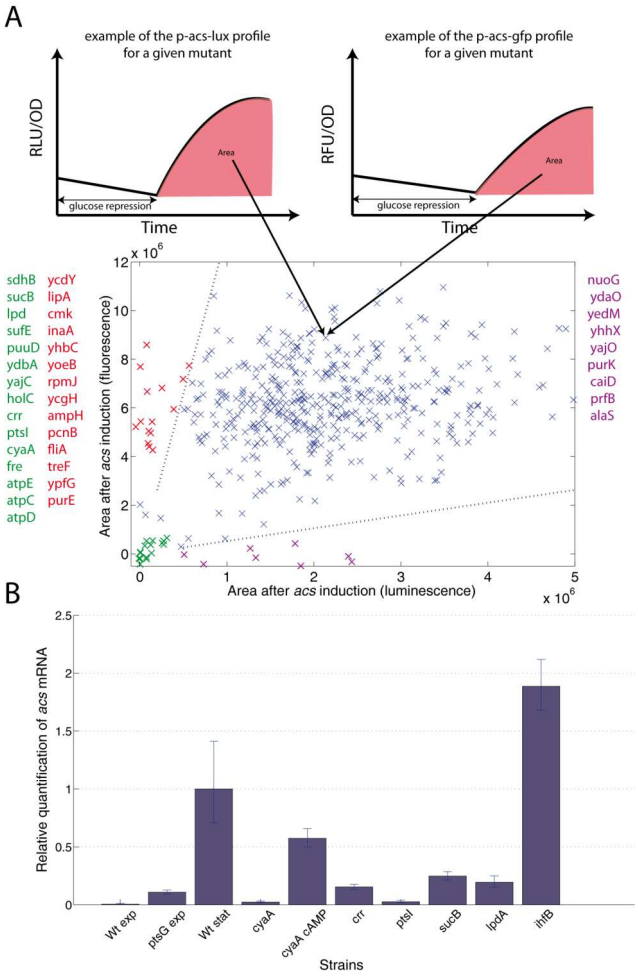



**Figure 4. Comparing luminescence, fluorescence and qRT-PCR.** A. The area under the expression curve after glucose exhaustion is a good measure of the induction of *acs* expression. The areas were computed for all measured luminescence and fluorescence profiles. The scatterplot shows the correspondence between induction measured by the luminescence and fluorescence reporters. Mutants beyond the dotted lines (four fold deviation from the correlation line) show a significant discrepancy between the two reporter systems. Mutants for which the fluorescence induction is normal, but luminescence low (red), probably affect the luminescence reporter system. Mutants that affect catabolite repression, the TCA cycle, or ATP synthetase show very low fluorescence and luminescence (green). B. Direct measurement of relative mRNA levels of selected mutants using qRT-PCR. All mRNA levels were measured 30 min after glucose exhaustion except for the first two samples, marked Wt exp and *ptsG* exp, that were measured during growth on glucose. All concentrations were normalized to the expression level of the induced wild-type strain. The *acs* mRNA is almost undetectable in the wild-type strain in exponential phase, whereas the *ptsG* mutant shows a clear signal. No *acs* expression is observed in the *cyaA* mutant (*cyaA*), but expression is restored when the medium is complemented with cAMP (*cyaA* cAMP). As expected, the *crr*, *ptsI*, *sucB* and *lpdA* strains show a reduced *acs* expression in comparison with the wild-type strain, and the *ihfB* mutant has a stronger *acs* expression.

generate false positives. However, these can be easily excluded by using an alternative reporter system or verification by qRT-PCR.

An important novel feature of our screen is the possibility to directly measure the dynamics of gene expression during the screen or by high-throughput measurements of selected mutants in an automated plate reader. Furthermore, the original screen, as well as these additional analyses, can be carried out in many different environmental and growth conditions.

## The regulation of *acs* transcription

The effects of some interesting mutants identified by the screen are summarized in Fig. 5. We have confirmed known interactions that regulate the transcription of *acs* and we have discovered interesting new directions to explore. The *acs* promoter activity is mainly controlled by the CRP-cAMP complex: transcription is completely abolished without CRP or cAMP. We also confirm the involvement of the phospho-transferase system which sets the intracellular concentration of cAMP: the *crr* and *ptsI* mutants have a very limited *acs* expression except when cAMP is added to the medium. Even though part of the same glucose uptake machinery, the *ptsG* mutant behaves differently from the other PTS mutants: it does not require the addition of exogenous cAMP to activate the *acs* expression.

We failed to induce *acs* expression in *cyaA* or wild-type strains by adding cAMP while glucose was still present in the growth medium. Those results could possibly be explained by another well-known mechanism of glucose repression: inducer exclusion. In this control system, the dephosphorylated form of $EIIA_{glc}$, the dominant form in the presence of glucose, binds and inhibits different sugar transport systems, thus preventing the entry of the inducing sugar (Roseman & Meadow, 1990) into the cell. However we do not believe that inducer exclusion can account for the glucose repression of the *acs* transcription in the presence of exogenous cAMP. To date, no inducer of *acs* expression has been described and acetate can enter the cell by passive diffusion (Gimenez *et al.*, 2003). Furthermore, the central component of inducer exclusion is $EIIA_{glc}$, encoded by the *crr* gene. Yet, we still observe glucose repression in a *crr* mutant in the presence of exogenous cAMP (Fig. 2D). This experiment demonstrates that neither variations of cAMP concentration nor variations of the phosphorylated state of $EIIA_{glc}$ are sufficient to explain the glucose repression that we observe. Consequently, and paradoxically, our results contradict the current two main models of carbon catabolite repression while confirming the involvement of the CRP-cAMP complex and the PTS in the transcription of the *acs* gene. Despite more than four decades of research,

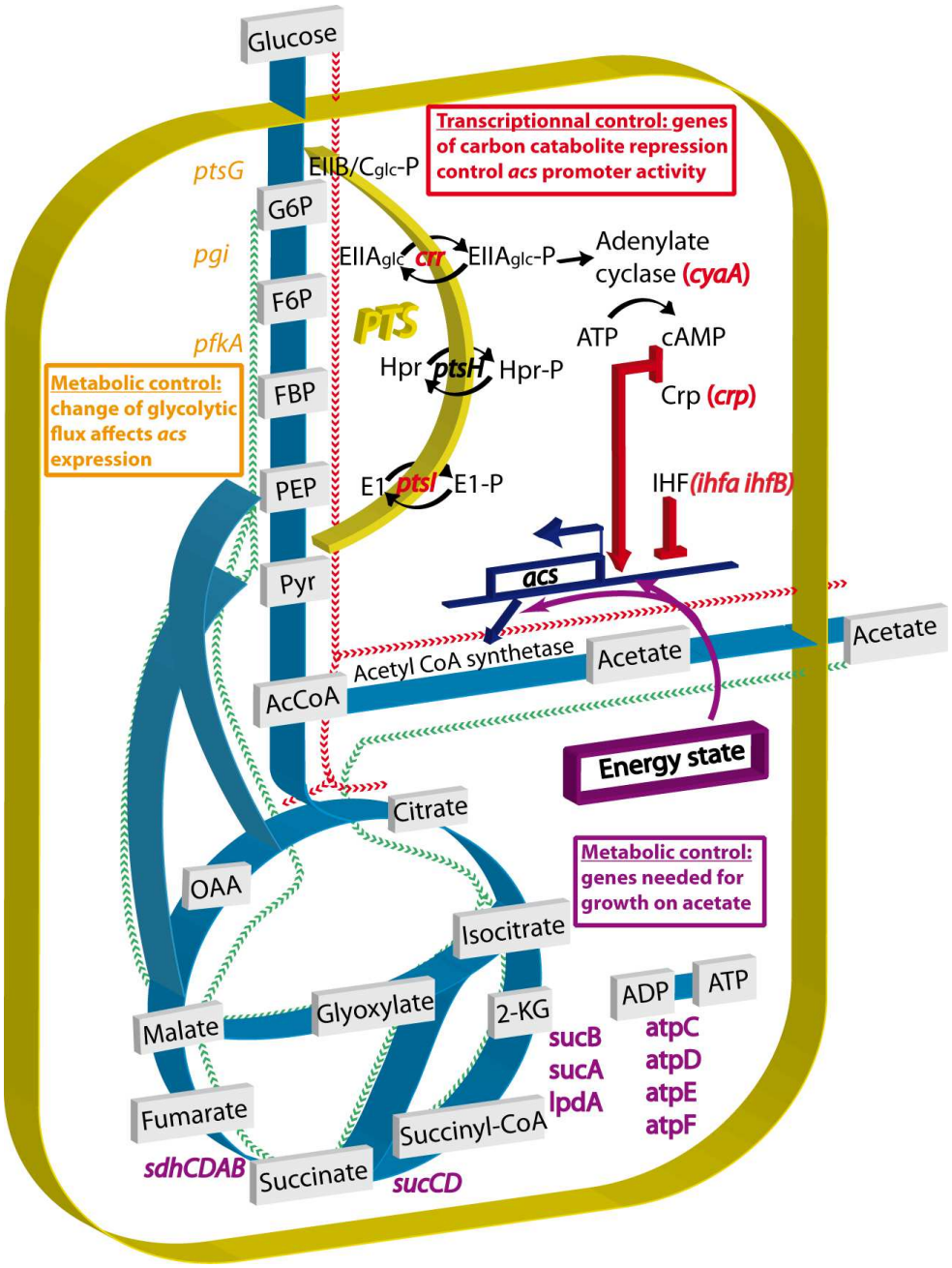



**Figure 5. The regulation of *acs* expression.** The expression of *acs* is regulated by genes involved in glucose repression (*cyaA*, *crp*, *crr* and *ptsI*). A modification of the glycolytic flux has indirect consequences on *acs* expression (*ptsG*, *pgi*, *pfkA*). Genes of the TCA cycle are needed for acetate utilization and mutants in these enzymes shut down *acs* transcription by a yet unknown mechanism. The red and green routes are schematic representations of the flux direction when cells grow on glucose and acetate, respectively.

our understanding of the precise mechanism that underlie the glucose effect (and more generally the carbon catabolite repression) is still far from complete and remains the subjects of debate (Inada *et al.*, 1996; Crasnier-Mednansky, 2008; Gorke & Stulke, 2008b; Narang, 2009). We are currently exploring theses paradoxical results in more detail.

### Metabolism and gene expression

An important strength of our screen is to uncover connections between metabolism and gene expression. In the case of *acs*, we find several novel connections between the metabolic state of the cell and the transcription of *acs*. Although still to be confirmed, it appears that the enzymes involved in the growth on acetate (*suc*, *atp* mutants etc) have a genuine impact on *acs* expression. Even though such a regulation makes sense biologically, the mechanism of this signal transduction remains to be discovered. Another example involves the glycolytic flux. Reducing the glucose inflow by deleting the glucose transporter (*ptsG* gene) increases the basal *acs* expression level during growth on glucose, probably by increasing the intracellular concentration of cAMP. However, the reduction of the glycolytic flux observed during growth of the *pgi* and *pfkA* mutants does not lead to an enhancement of *acs* expression in exponential phase (Fig. 2E). This result shows that neither *acs* promoter activity nor cAMP concentration are a simple function of the glycolytic flux, which is roughly the same for *ptsG*, *pgi*, and *pfkA* strains, but depends on the exact location where glycolysis is interrupted. We do not yet know the metabolic indicator nor the sensing mechanism that transmits the state of glycolysis to the activation of the *acs* promoter.

These different observations clearly show a strong connection between the metabolic state of the cell and *acs* transcription and confirm our prediction that the genetic regulatory network of *Escherichia coli* is densely connected and that a majority of these connections pass through metabolism (Baldazzi *et al.*, 2010).

## Materials and Methods

### Strains

The screen uses the Keio collection, comprising 3985 single gene deletions of the *E. coli* K-12 strain BW25113, called wild-type strain (Baba *et al.*, 2006).

## Plasmids

The nrfA-*acs* intergenic region was amplified by PCR using primers 5'-TCCTCTCGAGAGGGGCTTCA TCCGAAT-3' and 5'-TTTGGGATCCGCTTTTGTTCTCCTTGTAGG-3' and the BW25113 strain as a template. The PCR product, digested with Xho1 and BamHI, was inserted into a *luxCDABE* plasmid backbone without promoter to give p-acs-lux. A strong ribosome binding site drives efficient expression of the reporter gene. The *bla* gene (conferring ampicillin resistance) and the *acs* promoter region of the previous plasmid p-acs-lux were amplified by PCR using primers 5'-TTTGGGATCCGCTTTTGTTCTCCTT GTAGG-3' and 5'-TCTAAAGTGAGCTCGAGTAAACTTGGTCTGACAG-3'. The PCR product, digested with SacI and BamHI, was inserted into pUA66 (Zaslaver *et al.*, 2006) to give p-acs-gfp. The complete sequences of these plasmids are available upon request.

## High-throughput transformation

All steps were performed in 96-well plates. 135 plates, 6 days and 4 persons were needed to transform 3571 mutants. All volumes below are indicated for one well. Mutants were inoculated from frozen stock in sterile microplates containing 200 $\mu$l per well of Luria-Bertani broth (LB) with kanamycin (50 $\mu$g/ml) and placed at 37°C under shaking (200 rpm) overnight. The next day, 4 $\mu$l of the pre-culture were added to new plates containing 400 $\mu$l of LB with kanamycin (50 $\mu$g/ml) and placed at 37°C under shaking for 3 hours. We then transferred 50 $\mu$l of these exponentially growing cells to new plates containing 50 $\mu$l of ice cold TSS 2X (LB containing 10% PEG, 5% DMSO, and 20-50 mM $Mg_{2+}$ at a final pH of 6.5) (Chung *et al.*, 1989) and the mixture was kept on ice for 45 min before adding 4 $\mu$l of plasmid DNA (50 ng/ml). The plates were left on ice for another 10 minutes before applying a heat shock (90 seconds at 42°C) and returned on ice for an additional 10 minutes of incubation. We finally added 1 ml of LB containing ampicillin (100 $\mu$g/ml) and the plates were placed at 37°C under shaking overnight. The following day, we spotted the transformed mutants onto two different petri dishes containing either LB-agar-ampicillin (100 $\mu$g/ml) or minimal medium M9-agar-0.3% glucose (p/v)-ampicillin (100 $\mu$g/ml). Petri dishes were incubated at 37°C overnight and the luminescence of the colonies was measured the following days using an intensified camera (Photonic Science). The quantification of the luminescence of the 3500 colonies of the condition "LB day 1" was performed semi-automatically by means of a custom-made Matlab® scripts (Image processing toolbox). Briefly, we used functions that correct side effects

and compute an appropriate threshold on the images allowing the detection of the colonies. Then, we computed the median of luminescence pixels belonging to a specific colony.

## Measuring the dynamics of gene expression

The dynamics of gene expression was measured for cells growing for 10 hours in a 96-well microplate. The microplate reader (Perkin Elmer, Fusion alpha FP-HT) allows growth with shaking and temperature control (37°C). We measured the cell density (OD at 600 nm) every 30 min and luminescence or fluorescence (480 nm/520 nm) every 3 minutes. Unless otherwise specified, the wells in the microplate contain 180 $\mu$l of minimal medium (M9) supplemented with 0.03% glucose (w/v), 0.03% sodium acetate (w/v), and ampicillin (100 $\mu$g/ml). When indicated, we added cyclic AMP (5 mM) at the beginning of the experiment. The wells were inoculated with 20 $\mu$l of stationary phase cultures (10x dilution) grown in LB-ampicillin (100 $\mu$g/ml). The fluorescence background of the bacterial cells was assessed by measuring the fluorescence of mutant bacteria transformed with a promoterless reporter plasmid.

## Quantification of the expression profiles

In order to compare the strength of *acs* induction at the entry into stationary phase as measured by the luminescence and fluorescence reporter systems, we have calculated the area under the expression profile for four hours following entry into stationary phase (see Figure 4B for a schematic representation). This quantification was carried out using custom-made Matlab® scripts that perform the following tasks: (i) Remove outliers from the data. (ii) Determine the time of entry into stationary phase by classifying the expression profiles into different classes and fitting segment splines to the signal (fluorescence or luminescence) and the absorbance measurements. The entry into stationary phase is, e.g., indicated by a break in the absorbance curve. (iii) Correct the fluorescence expression profiles for the autofluorescence of the cells by subtracting corresponding profiles of control strains. No background correction is necessary for luminescence data. (iv) Numerically integrate the expression curve for 240 min, starting at the entry into stationary phase. The expression value was set to zero at the entry into stationary phase and the integral thus gives a measure of the induction strength of the *acs* promoter in the particular mutant strain. (v) Normalize the integral by the absorbance of the sample at entry into stationary phase.

### qRT-PCR

The strains were grown in a microplate containing M9 0.03% glucose and 0.03% acetate (preculture LB). The cells ( $10^8$) were harvested either before glucose exhaustion (exponential growth on glucose) or after glucose exhaustion (growth on acetate). Total mRNA was protected using the RNAprotect® Bacteria Reagent (Quiagen) and then isolated using the RNeasy mini kit (Quiagen) according to the recommendations of the manufacturer. Potential trace quantities of DNA were removed using the turbo DNAse (Ambion). Samples were protected from RNase using RNaseOut ribonuclease inhibitor (Invitrogen). 1 $\mu$g of RNA was reverse transcribed using SuperScript$^{TM}$ II Reverse Transcriptase (Invitrogen) according to the protocol of the manufacturer. Briefly, a 25 $\mu$l reaction mixture was incubated in a T3000 thermocycler (Biometra) for 10 min at 25°C, 50 min at 42°C and 15 min at 70°C. The primers (Eurogentec) used in the present study were as follows: *acs*: 5' ACAGTTCTGGTGACGGTTCC 3' and 5' ACAGTTCTGGTGACGGTTCC 3'. We have used the *rrsD* gene as internal control for relative quantification: *rrsD*: 5' GCTACAATGGCGCATACAAA-3' and 5'-TTCATGGAGTCGAGTTGCAG 3'. Quantitative PCR was performed in a StepOnePlus Real-Time PCR System (Applied Biosystems) using MESA Green qPCR Master Mix (Eurogentec) according to the instructions of the manufacturer. Briefly, 25 $\mu$l reactions mixtures were incubated for 10 min at 95°C and 40 PCR cycles (15 s at 95°C and 1 min at 60°C). PCRs were run in triplicate. Raw data were transformed into threshold cycle (Ct) values. The relative strength of *acs* expression for each mutant, compared to wild-type, was calculated by the comparative Ct Method ($\Delta\Delta$Ct).

## Acknowledgments

We would like to thank Omaya Dudin, Stephan Lacour and Yves Markowicz for many useful discussion and comments on the manuscript. Laure Giannone helped with the analysis of the data. This work was funded by the EU FP6 grant EC-MOAN, the ANR BioSys grant MetaGenoReg and the Joseph Fourier University.

## References

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., & Ishihama, A. 1999. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol*, **181**(20), 6361–70.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, **2**, 2006 0008.

Baldazzi, V., Ropers, D., Markowicz, Y., Kahn, D., Geiselmann, J., & de Jong, H. 2010. The carbon assimilation network in Escherichia coli is densely connected and largely sign-determined by directions of metabolic fluxes. *PLoS Comput Biol*, **6**(6), e1000812.

Billinton, N., & Knight, A. W. 2001. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Anal Biochem*, **291**(2), 175–97.

Brown, P. O., & Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nat Genet*, **21**(1 Suppl), 33–7.

Browning, D. F., Beatty, C. M., Sanstad, E. A., Gunn, K. E., Busby, S. J., & Wolfe, A. J. 2004. Modulation of CRP-dependent transcription at the Escherichia coli acsP2 promoter by nucleoprotein complexes: anti-activation by the nucleoid proteins FIS and IHF. *Mol Microbiol*, **51**(1), 241–54.

Chung, C. T., Niemela, S. L., & Miller, R. H. 1989. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*, **86**(7), 2172–5.

Crasnier-Mednansky, M. 2008. Is there any role for cAMP-CRP in carbon catabolite repression of the Escherichia coli lac operon? *Nat Rev Microbiol*, **6**(12), 954; author reply 954.

de Jong, H., Ranquet, C., Ropers, D., Pinel, C., & Geiselmann, J. 2010. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol*, **4**, 55.

Gimenez, R., Nunez, M. F., Badia, J., Aguilar, J., & Baldoma, L. 2003. The gene yjcG, cotranscribed with the gene acs, encodes an acetate permease in Escherichia coli. *J Bacteriol*, **185**(21), 6448–55.

Gorke, B., & Stulke, J. 2008a. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, **6**(8), 613–24.

Gorke, B., & Stulke, J. 2008b. Is there any role for cAMP-CRP in carbon catabolite repression of the Escherichia coli lac operon? Reply from Gorke and Stulke. *Nat Rev Microbiol*, **6**(12), 954.

Gosset, G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact*, **4**(1), 14.

Grainger, D. C., & Busby, S. J. 2008. Methods for studying global patterns of DNA binding by bacterial transcription factors and RNA polymerase. *Biochem Soc Trans*, **36**(Pt 4), 754–7.

Hakkila, K., Maksimow, M., Karp, M., & Virta, M. 2002. Reporter genes lucFF, luxCDABE, gfp, and dsred have different characteristics in whole-cell bacterial sensors. *Anal Biochem*, **301**(2), 235–42.

Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Inada, T., Aiba, H., & Postma, P. W. 1998. Inducer exclusion by glucose 6-phosphate in Escherichia coli. *Mol Microbiol*, **28**(4), 755–65.

Inada, T., Kimata, K., & Aiba, H. 1996. Mechanism responsible for glucose-lactose diauxie in Escherichia coli: challenge to the cAMP model. *Genes Cells*, **1**(3), 293–301.

Jacob, F., & Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, **3**, 318–56.

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P., & Karp, P. D. 2010. EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res*, **39**(Database issue), D583–90.

Koga, K., Harada, T., Shimizu, H., & Tanaka, K. 2005. Bacterial luciferase activity and the intracellular redox pool in Escherichia coli. *Mol Genet Genomics*, **274**(2), 180–8.

Kumari, S., Tishel, R., Eisenbach, M., & Wolfe, A. J. 1995. Cloning, characterization, and functional expression of acs, the gene which encodes acetyl coenzyme A synthetase in Escherichia coli. *J Bacteriol*, **177**(10), 2878–86.

Kumari, S., Beatty, C. M., Browning, D. F., Busby, S. J., Simel, E. J., Hovel-Miner, G., & Wolfe, A. J. 2000. Regulation of acetyl coenzyme A synthetase in Escherichia coli. *J Bacteriol*, **182**(15), 4173–9.

Meighen, E. A. 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev*, **55**(1), 123–42.

Narang, A. 2009. cAMP does not have an important role in carbon catabolite repression of the Escherichia coli lac operon. *Nat Rev Microbiol*, **7**(3), 250.

Neidhardt, F.C., & Savageau, M. A. 1996. Regulation beyond the operon. *Pages 1310–1324 of: Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1. Washington, DC: ASM Press.

Pruss, B. M., Nelms, J. M., Park, C., & Wolfe, A. J. 1994. Mutations in NADH:ubiquinone oxidoreductase of Escherichia coli affect growth on mixed amino acids. *J Bacteriol*, **176**(8), 2143–50.

Roseman, S., & Meadow, N. D. 1990. Signal transduction by the bacterial phosphotransferase system. Diauxie and the crr gene (J. Monod revisited). *J Biol Chem*, **265**(6), 2993–6.

Sclavi, B., Beatty, C. M., Thach, D. S., Fredericks, C. E., Buckle, M., & Wolfe, A. J. 2007. The multiple roles of CRP at the complex acs promoter depend on activation region 2 and IHF. *Mol Microbiol*, **65**(2), 425–40.

Shin, S., Chang, D. E., & Pan, J. G. 2009. Acetate consumption activity directly determines the level of acetate accumulation during Escherichia coli W3110 growth. *J Microbiol Biotechnol*, **19**(10), 1127–34.

Van Dyk, T. K., Wei, Y., Hanafey, M. K., Dolan, M., Reeve, M. J., Rafalski, J. A., Rothman-Denes, L. B., & LaRossa, R. A. 2001. A genomic approach to gene fusion technology. *Proc Natl Acad Sci U S A*, **98**(5), 2555–60.

Wolfe, A. J. 2005. The acetate switch. *Microbiol Mol Biol Rev*, **69**(1), 12–50.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods*, **3**(8), 623–8.

# Supplementary information

Author1[1], Author2[2], Author3[3,*]

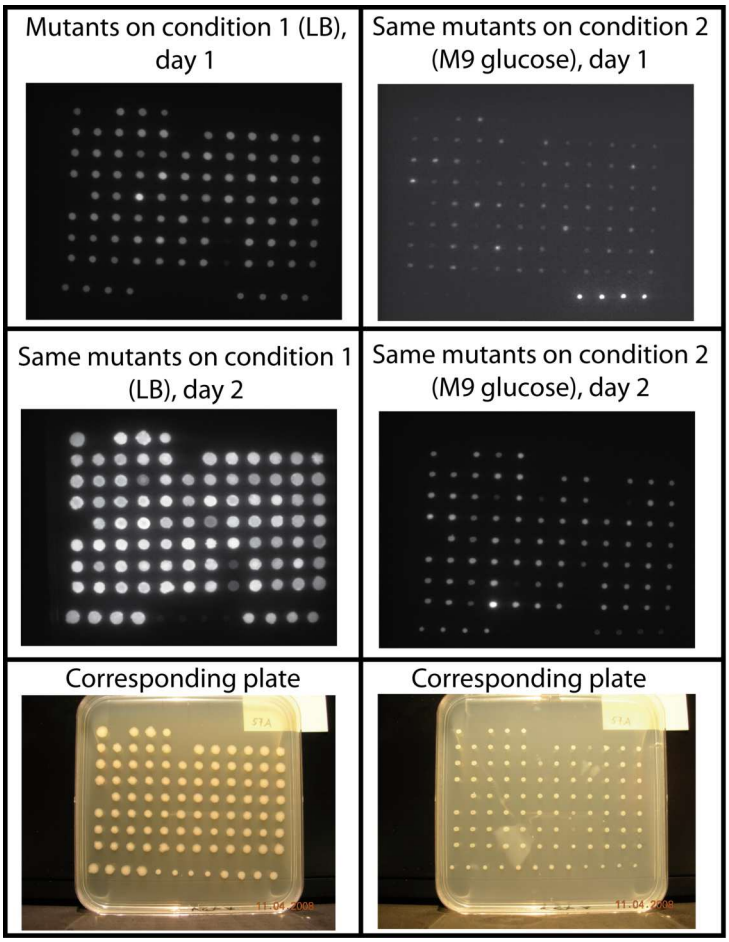
1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

* E-mail: Corresponding author@institute.edu

# 1 The screen can be performed in different conditions




The pattern of luminescence of colonies can significantly change over time and as a function of the growth medium. Our screen allows to easily test different conditions and to perform kinetics at the colony level. Here we show a typical example of a screen where the cells were grown on LB (left column) or M9-glucose minimal medium (right column). The luminescence images were taken after one or two days of incubation. The growth of the colonies is shown in the bottom row. The supplementary data show the luminescence images of colonies grown on LB plates. Luminescence images of M9 plates (and/or colony images) are available upon request.

## 2 The screen can be performed dynamically

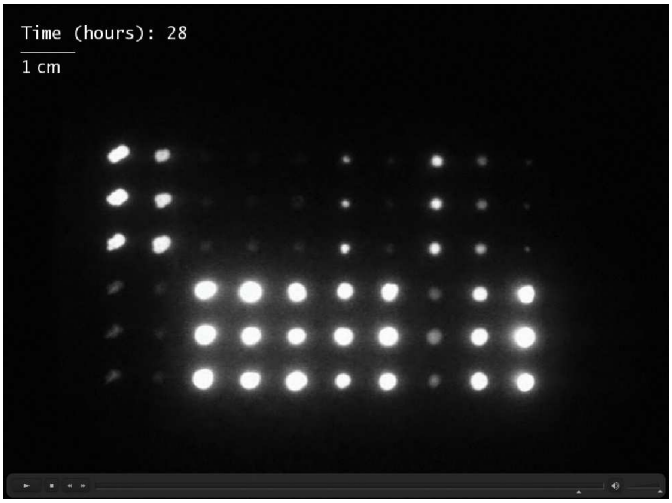



**Figure 1. Video of the development of colonies over time**

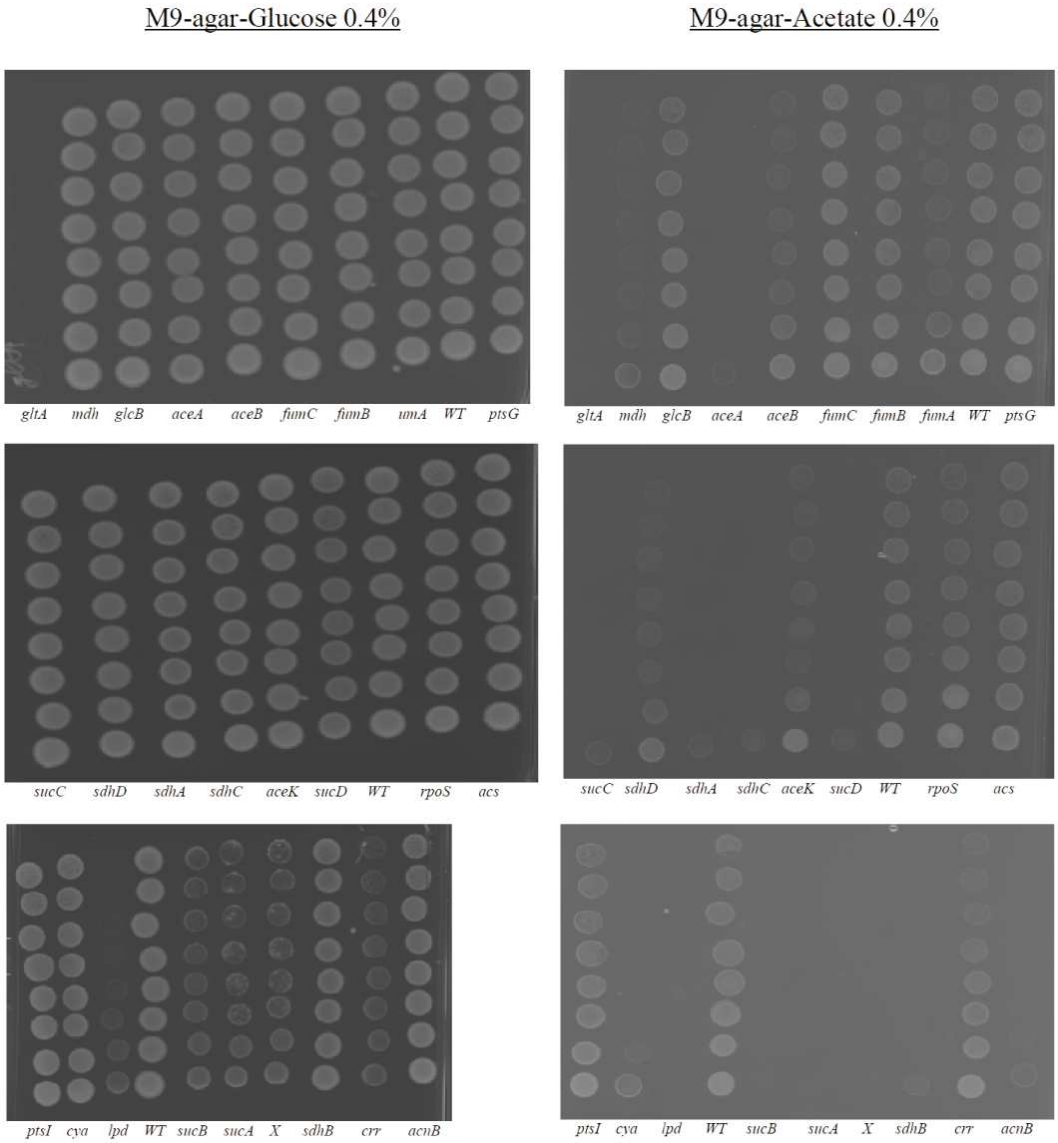
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Wt | Wt 2 | cyaA- | crp- | ptsI- | ptsH- | crr- | ptsG- | pgi- | pfkA- |
| Wt | Wt 2 | cyaA- | crp- | ptsI- | ptsH- | crr- | ptsG- | pgi- | pfkA- |
| Wt | Wt 2 | cyaA- | crp- | ptsI- | ptsH- | crr- | ptsG- | pgi- | pfkA- |
| sucB- | lpdA- | ihfA- | ihfB- | fis- | iclR- | fnr- | cpdA- | rpos- | hha- |
| sucB- | lpdA- | ihfA- | ihfB- | fis- | iclR- | fnr- | cpdA- | rpos- | hha- |
| sucB- | lpdA- | ihfA- | ihfB- | fis- | iclR- | fnr- | cpdA- | rpos- | hha- |

Mutants (see the table) containing the p-acs-lux plasmid were "spotted" on solid M9 medium 0.06% glucose 0.06% acetate 10% LB 1% agar 100 $\mu$g/ml ampicillin. We have filmed the development of colonies (luminescence) for 30 hours using an intensified camera.

### Results

- The expression profiles of triplicate colonies show identical and highly reproducible kinetics.
- The initial Keio *crp* strain we used in the screen was incorrectly constructed. A correct copy of this strain shows the same phenotype than the *cyaA* strain, as expected.
- Wt 2 is a strain with a shorter *acs* promoter sequence that does not include the *acsP1* promoter (maintaining only the *acsP2* promoter). Similar expression patterns are obtained, except that the intensity of *acs* expression is somewhat smaller.

# 3 Mutants that do not grow on minimal medium acetate




Certain mutants that grow on M9+Agar+glucose (0.4%) do not grow on M9+agar+acetate (0.4%). The y axis (upward) represents serial dilutions (factor 10) of strains pre-cultured in LB medium.

# 4 Mutants that do not grow on minimal medium glucose

Mutants that grow on Luria Broth + Agar + ampicillin but do not grow on M9 + agar + 0.3% glucose + ampicillin

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ppc | icd | rpoN | gltA | hisH | argC | ycaL | argG | cysG | trpB | metB | metR |
| proB | trpD | glyA | serA | argE | thrB | proA | serC | trpE | argH | aroA | hisG |
| pheA | metA | leuD | aroD | hisD | aroC | tyrA | aroE | ilvE | leuC | proC | hisC |
| ilvA | leuA | hisF | argA | metE | leuL | serB | nadB | pdxA | nadC | nadA | pdxJ |
| metF | cysN | cysC | cysD | cysH | purK | carB | purE | pyrD | pyrC | aroB | pyrE |
| pyrB | pyrF | purD | purA | purM | fepD | fepB | purL | hisB | hisA | trpC | argB |
| thrC | ilvC | ilvD | thrA | purH | purC | cysB | guaB | | | | |

Mutants with strong growth defects on M9+agar+0.3% glucose+ampicillin

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| sucA | sucB | lysR | Rpe | pfkA | Pgi | Rnt | aceE | cydB | Lpd | ybaS | fepC |
| metC | bioF | pabB | pabA | bioC | bioD | panC | lipA | panB | Cmk | pgpA | cysQ |
| ptsH | ptsI | gcvR | metL | fepG | carA | | | | | | |

# Chapitre III

Study of carbon catabolite repression in *Escherichia coli* by monitoring the activity of Crp-cAMP dependent promoters

# Study of carbon catabolite repression in *Escherichia coli* by monitoring the activity of Crp-cAMP dependent promoters

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## Abstract

Despite decades of intensive study, the understanding of the precise mechanism underlying carbon catabolite repression (CCR) in *Escherichia coli* is still far from complete. The central regulator of CCR is Crp, which becomes active in the presence of cAMP. However, addition of exogenous cAMP often fails to elicit the expected response. In this study, we have investigated some of the reasons for this unexpected result. We measured the activity of numerous CRP-cAMP dependent promoters in different conditions and genetic backgrounds. We used two different reporter genes (GFP and luciferase), measuring expression kinetics at the population level and at the single cell level. Our results indicate that the mechanism of carbon catabolite repression involves a metabolic control unrelated to the concentration of cAMP or to the phosphate transfer system.

## Introduction

In microorganisms, certain carbon sources are preferred to others. For example, when the medium contains glucose and lactose, *Escherichia coli* preferentially uses glucose and starts to metabolize lactose only when glucose is exhausted (Monod, 1942). This phenotype is explained by the inhibitory effect of glucose (the so-called glucose effect) on the expression of genes involved in lactose catabolism: the extensively studied *lac* operon. The phenomenon has been observed with other carbohydrates in many different prokaryotes and has been called carbon catabolite repression (CCR) (Magasanik, 1961). It refers to the general phenomenon whereby the presence of certain carbon sources in the medium represses the expression of certain genes or operons (Gorke & Stulke, 2008a).

Two complementary mechanisms have been proposed to explain the glucose mediated catabolite repression in *E. coli*: the "inducer exclusion model" (Saier & Roseman, 1972) and the "cAMP model" (Ullmann & Monod, 1968; Perlman *et al.*, 1969). The uptake of glucose is mediated by the phosphopenolpyruvate:glucose phosphotransferase system (PTS). This system sequentially transfers a phosphoryl group originating from phospoenolpyruvate to glucose via several enzymes including $EIIA_{glc}$ (Postma *et al.*, 1993). The cornerstone common to both models is the phosphorylation state of the enzyme $EIIA_{glc}$.

In the inducer exclusion model, the dephosphorylated form of $EIIA_{glc}$, the dominant form in the presence of glucose, binds and thereby inhibits different sugar transport systems, thus preventing the entry of the inducing sugar into the cell.

In the cAMP model (Kolb *et al.*, 1993), the phosphorylated form of $EIIA_{glc}$, the dominant form in the absence of glucose, binds and thereby activates adenylate cyclase, the enzyme that converts ATP into cAMP. The concentration of cAMP quickly increases, binds to Crp (cAMP receptor protein) and converts this global transcription factor to the active form, Crp-cAMP. Figure 1 summarizes these two models, taking as an example the well-studied regulation of *lacZ* expression.

Yet, despite more than four decades of research, our understanding of the precise mechanism undelying carbon catabolite repression remains incomplete and subject to debate (Inada *et al.*, 1996; Crasnier-Mednansky, 2008; Gorke & Stulke, 2008b; Narang, 2009a). Although the two models have a certain predictive power, each one has to struggle to account for experiments contradicting particular model predictions. For instance, glucose repression have still been observed in a *crp** strain which harbors a modified Crp protein that binds to the Crp-cAMP binding site even in absence of cAMP (Inada *et al.*, 1996; Khankal *et al.*, 2009). Moreover, both models fail to explain why PTS-independent carbon sources, such as glucose-6-phosphate, induce a strong catabolite repression even in PTS-deficient strains (Hogema *et al.*, 1997).

In a previous study (article not yet published), we studied the expression of *acs*, a gene whose transcription is strongly repressed by glucose and positively regulated by the CRP-cAMP complex (Kumari *et al.*, 2000). We described several experiments whose results did not fit well with the current models of carbon catabolite repression. In particular, we failed to relieve glucose repression by adding exogenous cAMP to the growth medium. Here, we investigate in detail the reasons for these unexpected results. We first measured the *acs* promoter activity in different conditions and in different genetic backgrounds by using two different reporter genes (GFP and luciferase). We then extended our investigation to other

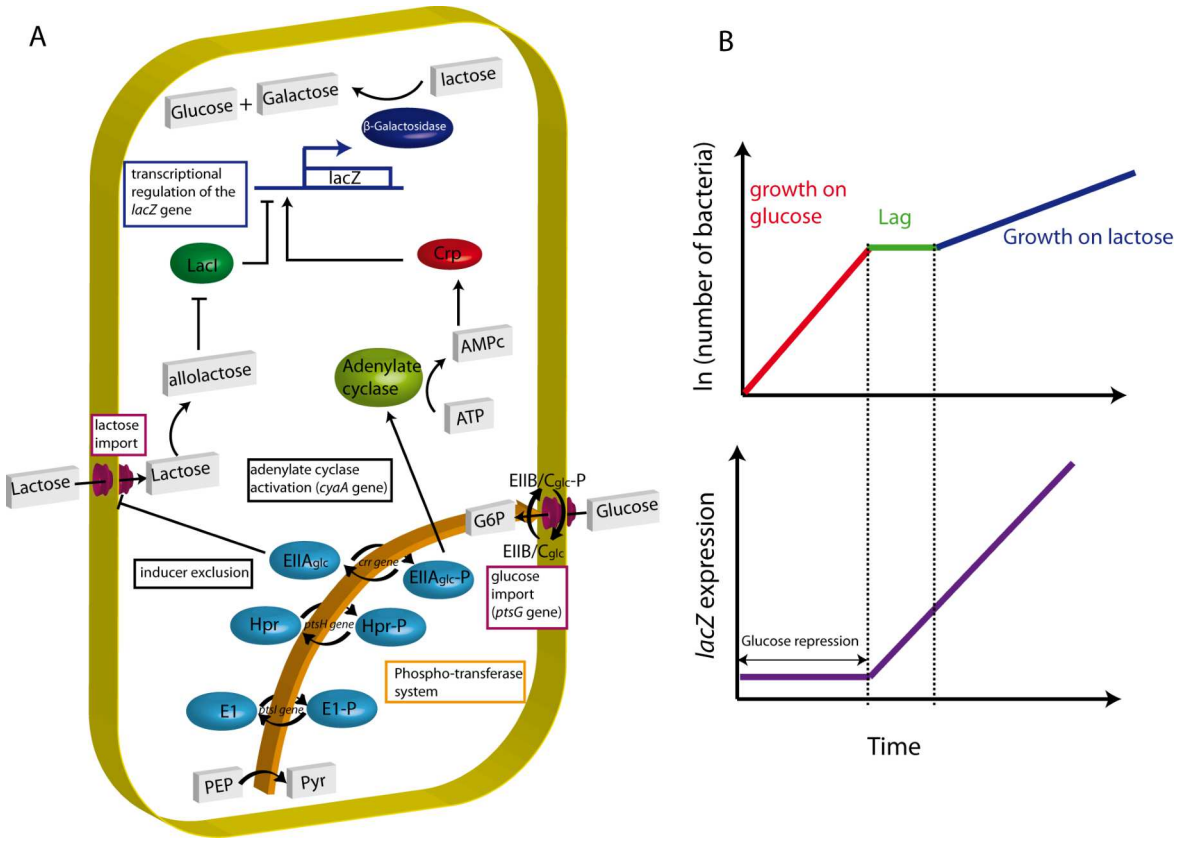



**Figure 1. The two models explaining glucose repression of the *lac* operon.** (A) Glucose import is mediated by the PTS system, which transfers a phosphoryl group originating from PEP to glucose. The inducer exclusion model: during glucose import, the dephosphorylated form of $EIIA_{glc}$ inactivates the import of lactose. This prevents the inactivation of the LacI repressor. Consequently, *lacZ* transcription is repressed. The cAMP model: once glucose is exhausted, the phosphorylated form of $EIIA_{glc}$ becomes predominant. It activates adenylate cyclase, which produces cAMP. This signalling molecule binds to the Crp protein, and the resulting complex activates *lacZ* transcription. (B) Schematic representation: the absence of $\beta$-galactosidase during growth on glucose leads to diauxic growth where lactose is consumed only after glucose exhaustion.

Crp-cAMP dependent promoters. Our results show that the mechanism of carbon catabolite repression cannot be totally explained neither by low levels of intracellular cAMP, nor by the phosphorylation state of $EIIA_{glc}$. We conclude that a reduction of the glycolytic flux, unrelated to the PTS or cAMP level and detected by an as yet unidentified metabolic sensor, is needed to fully induce Crp-cAMP dependent promoters.

# Results

## The *acs* promoter is strongly controlled by carbon catabolite repression

To measure *acs* expression, we used the previously described transcriptional fusions p-acs-lux and p-acs-gfp (article not yet published). Briefly, the *acs* promoter is cloned either upstream *luxCDABE* or *gfpmut2* reporter genes on a plasmid carrying either the colE1 (p-acs-lux) or the pSC101 (p-acs-gfp) origin of replication. To measure gene expression, we grow cells on minimal medium with the indicated carbon sources for several hours within a thermostated micro plate reader with periodic agitation. We measured absorbance, luminescence and fluorescence at intervals of about 5 minutes.

In a wild-type *Escherichia coli* strain growing on minimal medium (M9) supplemented with 0.03% glucose and 0.03% acetate, we observe a strong induction of *acs* expression when growth slows dramatically (glucose exhaustion) (Figure 2A). For reasons of clarity, we will not show the growth curves in the following profiles. However, to assist readers, we generally indicate the time of glucose exhaustion by an arrow. Identical profiles are obtained using the two reporter systems and plasmids carrying two different origins of replication. In addition, we have verified the expression profile by an independent method, qRT-PCR (article not yet published). We can therefore ascertain that the signal we measure reflects transcriptional regulation of the *acs* gene.

We observe an essentially identical expression profile when growing the wild-type strain on the non-PTS sugar glucose-6-phosphate (Figure 2B). This experiment demonstrates that the *acs* promoter is also strongly repressed by glucose-6-phosphate. Adding exogenous cAMP (5 mM) at the beginning of the experiment fails to directly induce the expression of the *acs* gene (Figure 2C). Adding glucose to the well during the experiment after glucose exhaustion, i.e., during the induction of *acs*, immediately stops the *acs* promoter activity. A *acs* induction occurs again when the newly added glucose is exhausted. The more glucose was added, the longer the duration of the shutoff of *acs* (Figure 2D). The time it takes for

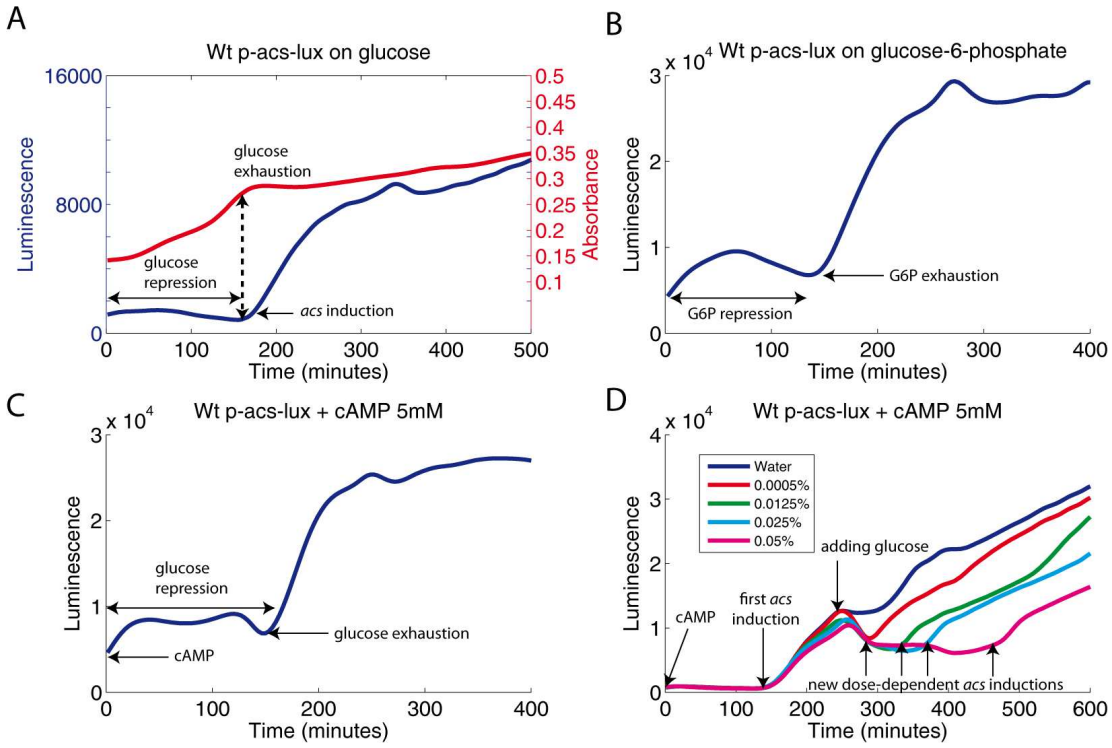



**Figure 2. *acs* expression in the wild-type strain.** (A) *acs* expression measured by the luciferase reporter in the wild-type strain growing on minimal medium, supplemented with 0.03% glucose and 0.03% acetate. The expression of *acs* is inhibited by a strong glucose repression and is induced once glucose is exhausted. (B) A strong repression also occurs in the presence of a non-PTS sugar (glucose-6-phosphate). (C) Glucose repression persists even when cAMP (5mM) is added to the medium at the beginning of the experiment. (D) *acs* induction is interrupted by adding glucose to the medium. A new induction occurs when this newly added glucose is exhausted (in a dose-dependent manner). This switch, controlled by glucose, occurs despite the presence of exogenous cAMP.

the expression of *acs* to stop when adding glucose (well below 10 minutes) is not compatible with the transcription and translation of a new hypothetic regulator. Thus, the repression occurs at the time-scale of signal transduction, e.g., the sensing of a biochemical signal. This switching behavior triggered by glucose is independent of the variation of the cAMP concentration, since cAMP is added in excess to the medium at the beginning of the experiment.

## Role of Crp-cAMP in carbon catabolite repression of the *acs* promoter

To clearly define the roles of Crp and cAMP in the control of *acs* expression , we measured the *acs* promoter activity in an adenylate cyclase deficient strain (*cyaA* mutant) and in a *crp* mutant. In both strains, no transcription of *acs* occurs, demonstrating the total dependence on the Crp-cAMP complex (controls on Figure 3A and 3B). In the *cyaA* strain, adding exogenous cAMP at the beginning of the kinetics restores the same *acs* transcription profile as the one observed in a wild-type strain (Figure 3A), showing strong *acs* induction only when glucose is exhausted. The lack of *acs* expression during growth on glucose could be due to the absence of Crp during this growth phase. To exclude this possibility, we constructed a plasmid that over-express the Crp protein when induced by anhydrotetracycline (aTc). This plasmid complements the *crp* mutant in a dose dependent manner (aTc), but only when glucose is exhausted (Figure 3B).

The sudden induction of *acs* in the wild-type strain and its absence in a *cyaA* strain have also been observed at the single cell level using time-lapse microscopy. We grew these two strains (transformed with the p-acs-gfp plasmid) together on a solid and translucent medium and acquired fluorescence and phase contrast images for several hours. As expected, only part of the bacterial layer (corresponding to the wild-type strain) showed a sudden increase of fluorescence once the micro colony has stopped growing (Figure 3C; supplementary movie 1). Furthermore, in this experiment, the growth conditions (oxygenation, growth rate, contact with neighboring bacteria) are quite different from the conditions of growth in a homogeneous, liquid culture. Yet, the observed expression profile is remarkably resilent.

## Role of the PTS in carbon catabolite repression of the *acs* promoter

To assess the role of the PTS in glucose-mediated repression of the *acs* promoter, we had previously measured its activity in two PTS mutants: the *ptsI* and *crr* strains. The E1 deficient strain (*ptsI* mutant) has a much lower growth rate on glucose. Since this mutant completely inactivates the PTS by blocking

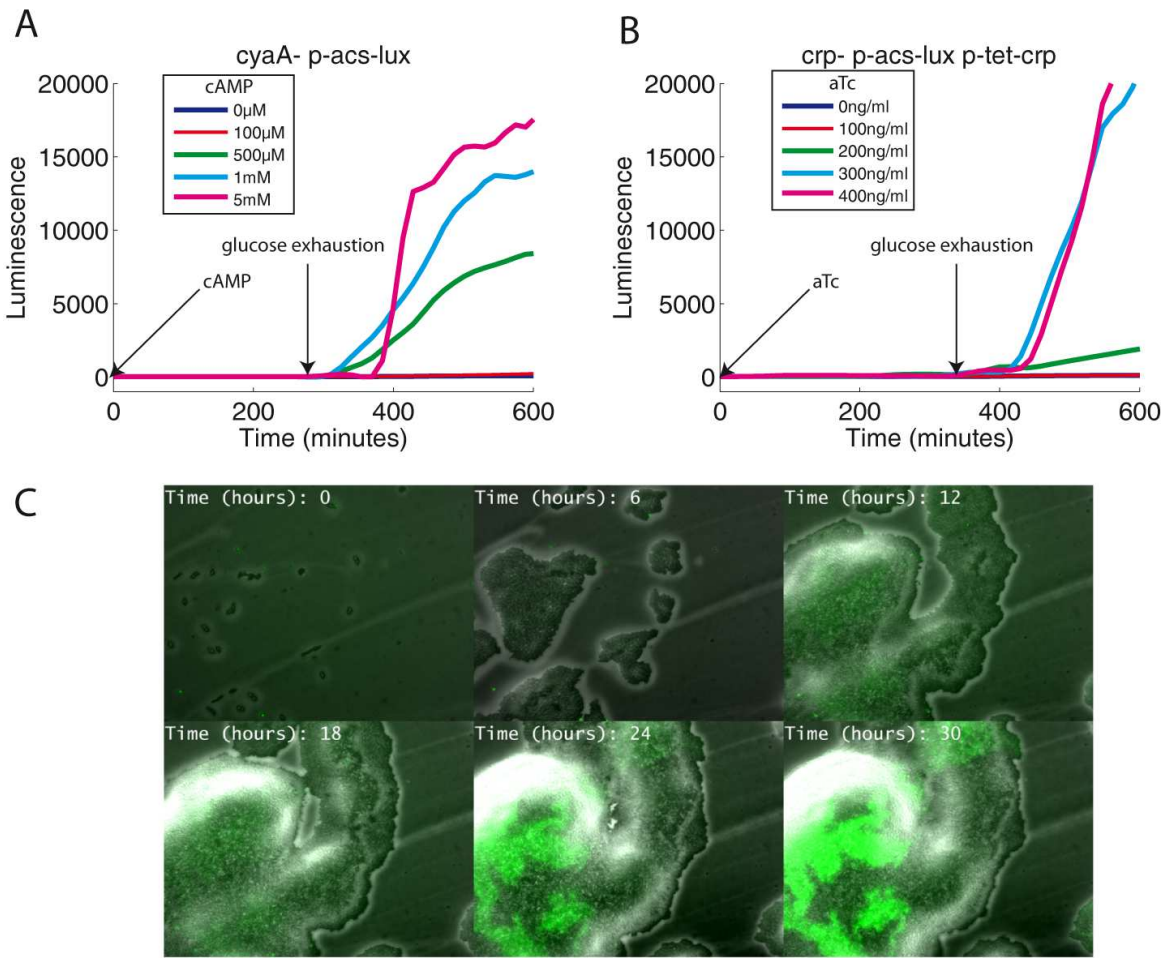



**Figure 3. *acs* expression in *cyaA* and *crp* mutants.** (A and B) Bacteria were grown on a minimal medium supplemented with glucose and acetate. The controls (dark blue) show that the absence of *crp* or *cyaA* prevents the expression of *acs*. (A) The *cyaA* strain is complemented in a dose-dependent manner by adding cAMP to the medium. (B) Similar dose-dependent complementation is observed with the *crp* mutant, by using a plasmid (p-tet-crp) allowing anhydrotetracycline (aTc) dependent expression of Crp. (A and B) In both cases, the repression remains while glucose is still present. (C) Time-lapse microscopy. A wild-type strain and a *cyaA* mutant mixed together were grown on the surface of a solid medium. After growth arrest (glucose exhaustion), only a part of the bacterial layer (corresponding to the wild-type strain) induces *acs* expression, as shown by the rise in fluorescence.

the first step of phosphate transfer, glucose has to enter the cell via PTS-independent transporters (MglC and GalP) (Gosset, 2005). *acs* expression in this strain is very weak, suggesting the role of the PTS in the production of cAMP. Adding exogenous cAMP restores *acs* transcription, even in presence of glucose (Figure 4A).

Restoring a fast growth rate by providing the favorable, non-PTS sugar glucose-6-phosphate re-established the repression of *acs* transcription in presence of exogenous cAMP (Figure 4B). First, this experiment shows that glucose-6-phosphate maintains a strong repression of the *acs* promoter, even in a PTS-deficient strain. Therefore, the repression of the *acs* promoter cannot be due to a modification of the PTS phosphorylation state by glucose-6-phosphate. Second, the major difference between growth on glucose and growth on glucose-6-phosphate in a *ptsI* mutant, and in the presence of exogenous cAMP, is the growth rate. The *ptsI* strain has a strong growth defect on glucose whereas it grows normally on glucose-6-phospate which is transported by a PTS independent hexose-phosphate system. Once in the cell, both substrates are metabolized by glycolysis. The *acs* promoter thus seems to somehow sense the rate of glucose metabolism and is activated when glycolysis is slow.

We also measured *acs* expression in an $EIIA_{glc}$ deficient strain (*crr* mutant), i.e. a strain that lacks the component of the PTS at the crossroad between inducer exclusion and activation of adenylate cyclase. The *crr* mutant grows at an almost normal rate on glucose, i.e., much faster than the *ptsI* strain. Since there are several alternative EIIA proteins that can participate in the PTS, the absence of $EIIA_{glc}$ may probably be compensated by another EIIA protein. The *acs* promoter activity in this strain is very weak (Figure 4C). Exactly as observed with the *cyaA* mutant, exogenous cAMP restores *acs* expression with a profile very similar to the wild-type pattern. These observations suggests that $EIIA_{glc}$ is important for stimulating adenylate cyclase when glucose is absent from the growth medium. However, this experiment also points out an important shortcoming of current models of CCR: the simultaneous absence of a $EIIA_{glc}$ mediated hypothetical (acetate has never been described as an inducer of *acs* expression) inducer exclusion and the presence of exogenous cAMP are not sufficient to remove glucose repression. This result is further confirmed in the *crr cyaA* double mutant (Figure 4D). Glucose repression is clearly maintained until all glucose is exhausted. Thus, deleting the two main components of carbon catabolite repression has failed to remove the glucose mediated *acs* repression. Neither the "cAMP model" nor the "inducer exclusion model" are able to explain this result. Potential explanations involving a major role of Fis, IHF, IclR, FNR, RpoS, FruR or Mlc were discarded by using the corresponding mutants (data not shown).

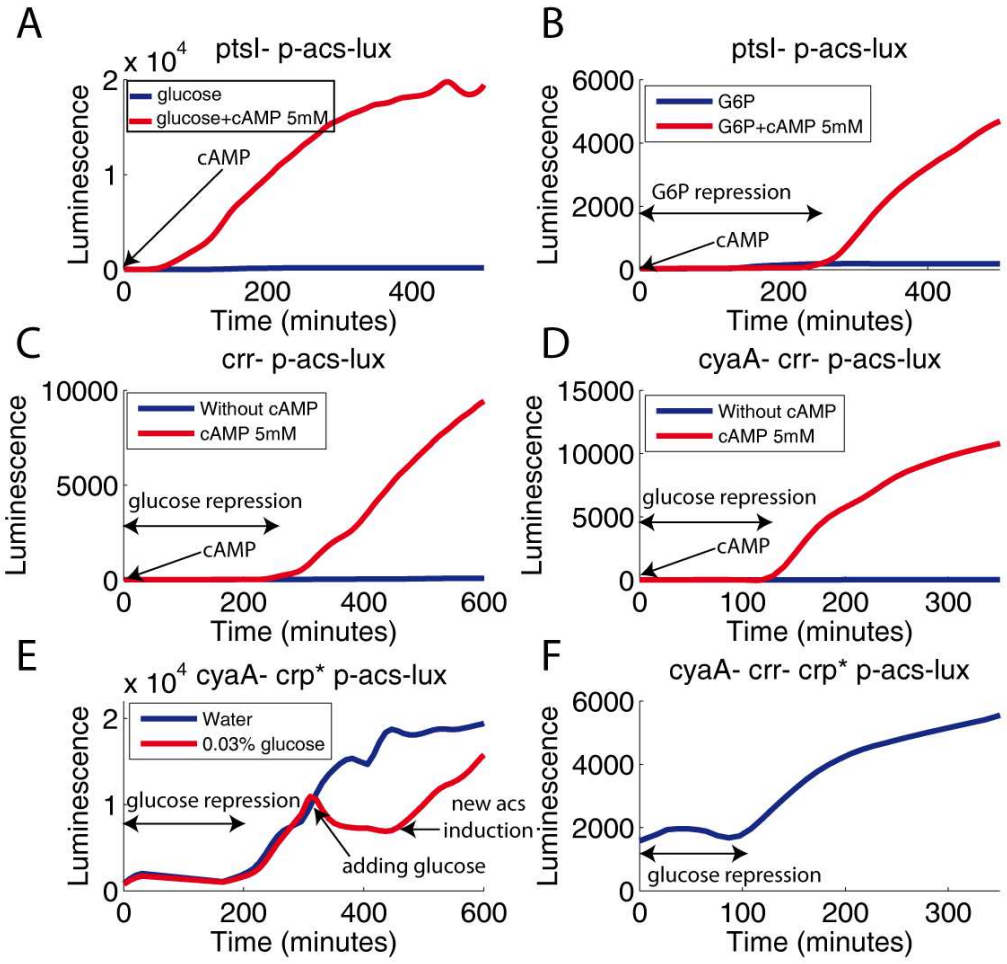



**Figure 4. *acs* expression in PTS deficient strains and *crp** strains.** Bacteria were grown on a minimal medium supplemented with 0.03% glucose and 0.03% acetate. (A and C) *ptsI* and *crr* strains have an extremely low *acs* expression (control in dark blue). (A) cAMP (5mM) complements the *ptsI* strain even during growth on glucose. (B) cAMP complements the *ptsI* strain, but only when G6P (0.03%) is exhausted. (C and D) Glucose repression remains in *crr* and *crr cyaA* strains, even when cAMP is added to the medium. (E) The *acs* expression is recovered in a *cyaA crp** strain without the need of adding cAMP. Yet, the glucose repression is still observed. Adding glucose just after *acs* induction stops the *acs* promoter activity and a new induction occurs when the added glucose is exhausted. (F) The glucose repression is still observed in a *cyaA crr crp** double mutant.

## Role of cAMP import/export in carbon catabolite repression of the *acs* promoter

One possible explanation of the previous results is that glucose controls the import/export of cAMP and thus prevents the entrance of exogenous cAMP (Buettner *et al.*, 1973). Indeed, the mechanism of import/export of cAMP remains largely unexplained, even though recent results suggest that TolC is partly responsible for the active export of cAMP (Hantke *et al.*, 2011). To assess a possible control of cAMP import/export as an explanation of our results, we constructed a *cyaA crp** strain harboring a modified Crp* protein carrying three amino acid substitutions (I112L, T127I, and A144T) (Khankal *et al.*, 2009). This Crp* protein binds to the Crp-cAMP DNA binding site even in the absence of cAMP. We confirm that the three amino acid difference between the *cyaA* mutant and the *cyaA crp** strain is sufficient to restore *acs* transcription without the need of adding exogenous cAMP to the medium. However, again, induction of the *acs* promoter occurs only after glucose exhaustion. Furthermore, adding glucose just after the first *acs* induction immediately stops *acs* promoter activity. A new induction occurs when the newly added glucose is exhausted (Figure 4E). This experiment excludes the possibility that the control of cAMP export or import be responsible for the glucose repression. We obtained similar results using a *crr cyaA crp** triple mutant (Figure 4F). In other words, we observe a strong glucose repression in a strain without cAMP nor $EIIA_{glc}$.

## Carbon catabolite repression of other Crp-cAMP dependent promoters

Are our results a particular quirk of the *acs* promoter or a general characteristic of CCR? In order to answer this question, we studied the expression dynamics of five other Crp-cAMP dependent genes: *araBAD*, *glpA*, *sdhC*, *gntT*, *uhpT*. All corresponding transcriptional fusions to *gfp* (Zaslaver *et al.*, 2006) are clearly sensitive to exogenous cAMP. With the exception of *sdh*, these promoters also need inducers to be activated: arabinose (*araBAD*), glycerol (*glpA*), glucose-6-phosphate (*uhpT*) and gluconate (*gntT*) were therefore added as needed.

For the *sdh* promoter, we also constructed the corresponding transcriptional fusion to the luciferase operon. To minimize the probability of any artifact due to the plasmids or reporter systems, we transformed the *cyaA* mutant with both plasmids (p-sdh-lux and p-sdh-gfp) in order to simultaneously monitor luminescence and fluorescence signals (Figure 5A). Figures 5B and 5C confirm that the expression profile

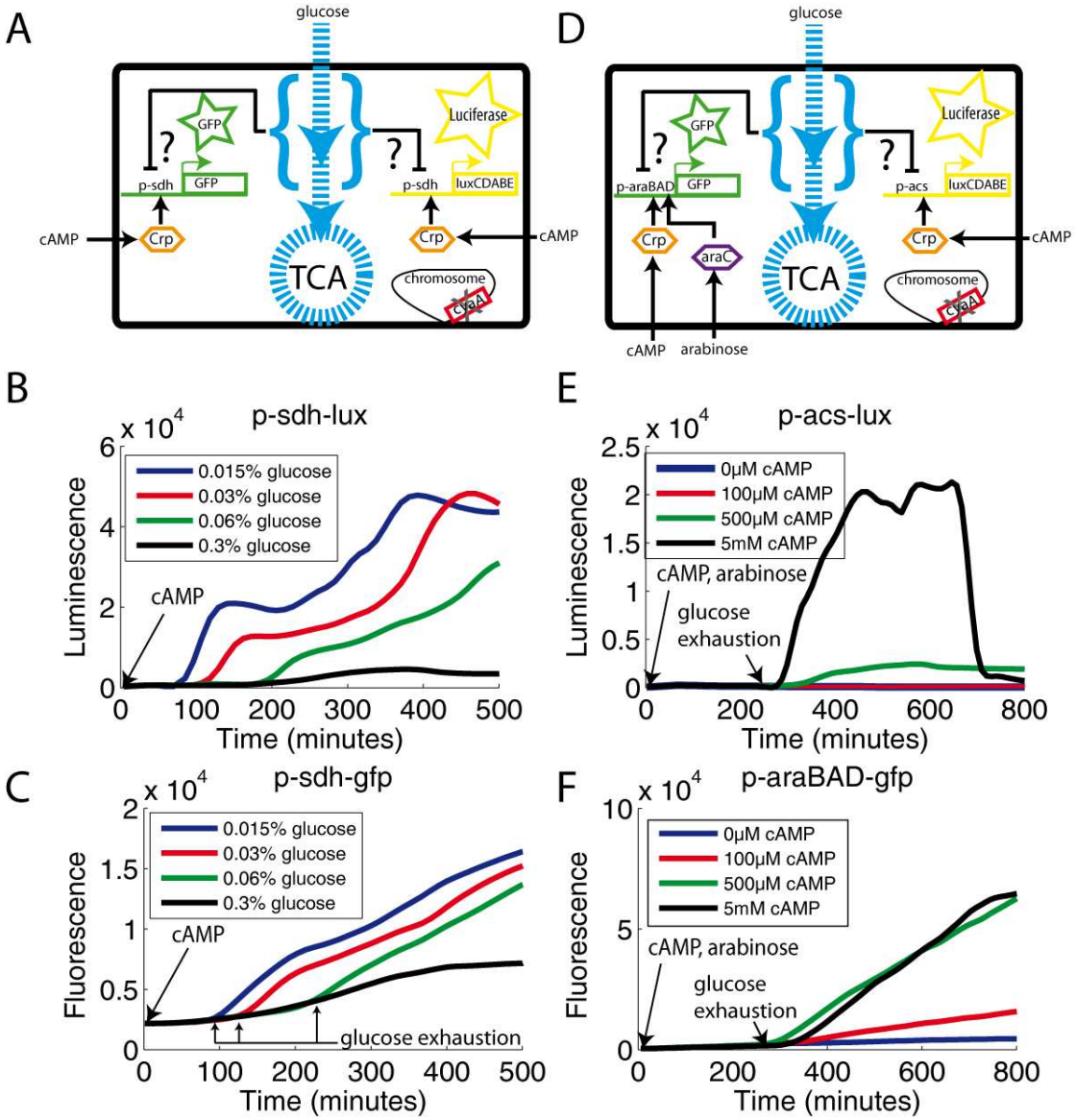



**Figure 5. Other cAMP-Crp promoters.** Bacteria were grown on a minimal medium supplemented with glucose (as indicated or 0.03%) and 0.03% acetate. (A) The *cyaA* mutant used to obtain the data of figures 5B and 5C. Both GFP and luciferase report *sdh* expression. (B and C) The glucose repression remains, despite the presence of exogenous cAMP (500μM). Increasing concentrations of glucose lengthens the period of growth and thus the period of corresponding repression. (D) The *cyaA* mutant used to obtain the data of figures 5E and 5F. GFP reports *araBAD* expression and luciferase reports *acs* expression. (E and F) The glucose repression remains, despite the presence of cAMP and arabinose (0.01%).

is independent of the reporter system, even though GFP is somewhat less reactive than luciferase. This is expected, since GFP has to maturated the fluorophore before emitting a signal, whereas luciferase is active as soon as the protein is synthesized. In this experiment, we added 500µM cAMP to the medium at the beginning of the experiment. As observed with the *acs* promoter, the *sdh* promoter is fully activated only after glucose exhaustion. Increasing concentrations of glucose extend the duration of the growth phase and correspondingly glucose repression (Figure 5B and C).

We also used the double reporter technique to simultaneously monitor the activity of two different promoters (*araBAD*) and (*acs*) in the same cell (Figure 5D). In this experiment, we grew the *cyaA* mutant on glucose and acetate (the carbon sources), in the presence of arabinose and cAMP (the inducers). Both luminescence (*acs*) and fluorescence (*araBAD*) strongly increase after glucose exhaustion (Figures 5E and 5F). As expected, the promoter strength after glucose exhaustion is a function of the cAMP concentration, *araBAD* reaching maximal activity at a lower cAMP concentration than *acs*. Thus, the *sdh* and *araBAD* promoters behave as the *acs* promoter: full induction occurs only after glucose exhaustion, even when cAMP is present in the medium. The same holds true for the *glpA* promoter (see supplementary information). The behavior of the *acs* promoter is therefore representative of several CCR-controlled promoters. Furthermore, for these promoters, the induction behavior mirrors the one of *acs* also in the "cAMP and $EIIA_{glc}$ free" strain (*cyaA crr crp**) (supplementary information). However, we failed to observe glucose repression for the *uhpT* and *gntT* promoters even though the strength of their promoter activity was easily controlled by ranging cAMP (see discussion and supplementary information).

## Role of the glycolytic flux

The experiments shown in figure 4 show that neither the functioning of the PTS nor the variation of cAMP concentration explain carbon catabolite repression. The major, systematic difference between situations where the *acs* promoter is induced or not is the growth rate, and more specifically the glycolytic flux (the rate of glycolysis, going from glucose-6-phosphate to pyruvate). For example, glucose and glucose-6-phosphate, after having entered the cell are both in the form of glucose-6-phosphate. The difference in our experiments is the rate of import of these molecules in different situations and mutant backgrounds. In agreement with this hypothesis, we had previously observed, that *acs* expression in a *ptsG* strain (deleted for the main glucose transporter) was clearly upregulated compared to the wild-type strain during growth on glucose (article not yet published). The *ptsG* strain is able to grow on glucose by using

other transporters, but its growth rate, and thus its glycolytic flux, is strongly reduced.

We therefore set out to verify a prediction of this hypothesis: we should be able to control carbon catabolite repression by modifying the glycolytic flux. This can be done by supplementing the growth medium with a non metabolizable analogue of glucose, $\alpha$-methyl glucose. We could indeed induce several Crp-cAMP operons by adding the appropriate concentration of $\alpha$-methyl glucose into the growth medium. An example is shown in figure 6A, where we control the activity of the *sdh* promoter in a *cyaA* mutant (supplemented with exogenous cAMP) by increasing the concentration of $\alpha$-methyl glucose. Thus, we do not observe catabolite repression induced by non metabolizable sugars (Ambrose & MacPhee, 1998).

We performed a similar experiment at the single cell level. We filmed 4 different wells in parallel (6-wells microplate). Each well contains a solid medium (M9 agarose with glucose and arabinose). The parameter that varies between these 4 wells is the concentration of $\alpha$-methyl glucose. Figure 6B shows the difference of *araBAD* expression between those wells (supplementary movie 3). In well 1, there is no $\alpha$-methyl glucose: cells grow fast while there is glucose and when this latter is exhausted, they induce synchronously the expression of *araBAD*. In well 2, the $\alpha$-methyl glucose has slightly reduced the glycolytic flux and some cells induce the *araBAD* expression before glucose exhaustion. In well 3, the glycolytic flux is strongly reduced, growth is slow and most cells induce the *araBAD* expression before glucose exhaustion. In the last well, the concentration of $\alpha$-methyl glucose is so high that cells appear to be poisoned, but all induce *araBAD* expression before glucose exhaustion. This experiment demonstrates that we can control carbon catabolite repression by tuning the glycolytic flux. The heterogeneity of *araBAD* expression between the cells is a well-known phenomenon (Megerle *et al.*, 2008). This experiment also points out a way to tune the distribution of *araBAD* expression in the population of cells.

## Discussion

In this study, we have explored the mechanism of carbon catabolite repression by studying the regulation of transcription of Crp-cAMP dependent operons in different genetic contexts and in different conditions. We monitored *in vivo*, in real time and with a high temporal resolution, the activity of several Crp-cAMP dependent promoters. Different strategies of controls have been implemented to assess the robustness of our results: we used two reporters systems (GFP/luciferase) and we performed the experiments both at the population level (liquid medium) and at the single cell level (solid medium).

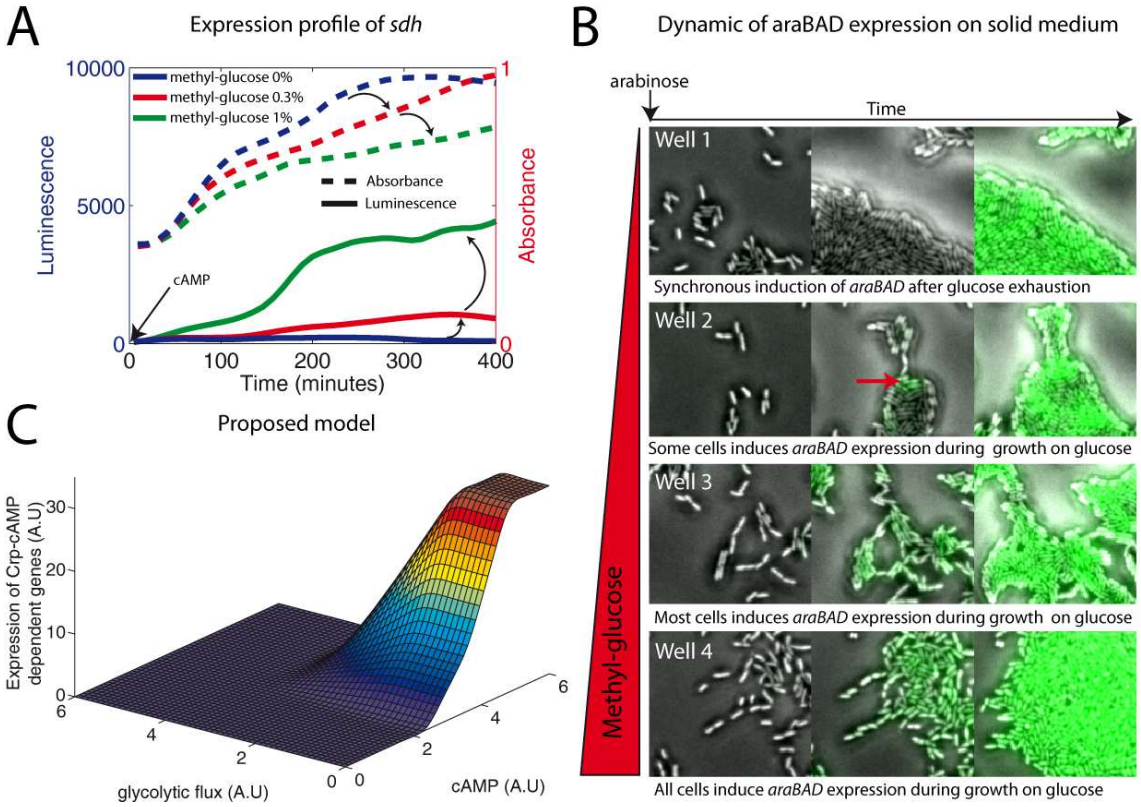



**Figure 6. Role of the glycolytic flux in carbon catabolite repression.** (A) Bacteria (*cyaA* strain) were grown on a minimal medium supplemented with 0.03% glucose, 0.03% acetate and 600µM cAMP. Non-metabolizable $\alpha$-methyl-glucose controls the growth of cells by competing for glucose uptake and catabolism. The resulting reduction of the glycolytic flux induces the *sdh* promoter activity in a dose dependent manner. (B) *araBAD* expression of cells growing on a solid medium composed of glucose (0.06%) and arabinose (0.005%). 4 wells are filmed in parallel. Only the concentration of $\alpha$-methyl-glucose varied between the wells (0%, 0.3%, 1% and 3%). This parameter tunes the level of *araBAD* expression, but also the proportion of cells that express *araBAD*. (C) A model of carbon catabolite repression: the combination of a sufficient concentration of cAMP and a sufficient reduction of the glycolytic flux is needed to synergistically induce CCR dependent genes.

## Is there any role for cAMP in carbon catabolite repression of the *Escherichia coli* Crp-cAMP operons?

This question has been addressed in 2008 for the case of the *lac* operon (Gorke & Stulke, 2008b; Crasnier-Mednansky, 2008; Narang, 2009a). In 1998, Inada et al. have demonstrated than glucose repression is maintained both in a *cyaA* strain complemented with cAMP and in the double mutant *cyaA crp** (Inada *et al.*, 1996). The authors concluded that inducer exclusion, blocking the entry of lactose, was the main contributor to carbon catabolite repression when *E. coli* grows in the presence of both glucose and lactose. Later, Narang (Narang, 2009b), by an accurate and exhaustive review of four decades of results, has pointed out that the small difference of levels of cAMP during growth on glucose compared to growth on lactose cannot account for the large repression of $\beta$-galactosidase on glucose. These results clearly challenge the cAMP model.

In this context, our results demonstrate: (i) that cAMP is necessary in the wild-type strain to increase transcription of Crp-cAMP dependent promoters. Adding cAMP in the medium complements the *cyaA* mutant and thus increases the expression of all operons studied in this work ; (ii) that, in the wild-type strain, cAMP is not sufficient for explaining carbon catabolite repression: adding cAMP to the medium fails to retrieve full gene expression while glucose is still present ; (iii) that cAMP is not necessary in a *cyaA crr crp** strain. This strain, devoid of all supposedly essential components of CCR, still shows strong glucose repression.

## Is there a role for the PTS in carbon catabolite repression of the *Escherichia coli* Crp-cAMP operons?

Dephosphorylation of enzyme $IIA_{glc}$ is thought to be the main cause of glucose repression. However, non-PTS carbohydrates, such as glucose-6-phosphate, also induce a strong catabolite repression. Hogema et al. have shown that glucose-6-phosphate added in the medium induces the dephosphorylation of $EIIA_{glc}$. Yet, they have also observed that glucose-6-phosphate is still able to repress catabolic operons in a *crr* strain (Hogema *et al.*, 1997). This last result clearly challenges the main role of enzyme $EIIA_{glc}$ in carbon catabolite repression.

In this context, our results: (i) demonstrate that, in the wild-type strain, the PTS is necessary for increasing transcription of the *acs* promoter: the *ptsI* and *crr* strains show a very low level of *acs*

transcription, except when exogenous cAMP is added to the medium ; (ii) demonstrate that the PTS is involved in the control of cAMP production: the activation of adenylate cyclase by the $EIIA_{glc}$ enzyme remains poorly understood as it seems to require additional components not yet discovered (Park *et al.*, 2006); (iii) demonstrate that the phosphate transfer system (PTS) is not sufficient to account for carbon catabolite repression: a strong carbon catabolite repression still occurs in a *crr* strain complemented with exogenous cAMP ; (iv) demonstrate that the PTS is not necessary in a *crr cyaA crp** strain for carbon catabolite repression: this strain still shows glucose repression.

## Mechanism of carbon catabolite repression

How can we reconcile these apparently contradictory results? The rise of the intracellular concentration of cAMP when glucose is exhausted is *correlated* with the induction of Crp-cAMP dependent operons. The simplest and most intuitive explanation is that this increase in cAMP concentration is *responsible* for the induction of Crp-cAMP dependent operons (Epstein *et al.*, 1975). However, there still remains the counter-intuitive possibility that those events, although correlated, may not be linked by a cause and effect relationship. Our experiments show that the sequence "glucose exhaustion - rise in cAMP concentration - transcription of Crp-cAMP operons" is not correct, or at least incomplete. The model that we propose in figure 6C assumes that the transcription of CCR-controlled Crp-cAMP operons requires two factors: a sufficient concentration of cAMP and a reduction of the glycolytic flux.

The glycolytic flux would have to be detected by an as yet unknown sensor, detecting the concentration of an as yet unidentified metabolite. Different CCR-controlled promoters differ by their sensitivity to this signal and their sensitivity to the concentration of cAMP. Only when both conditions are met will the corresponding promoter be fully induced. The *acs* promoter, for example, requires a high concentration of cAMP, whereas the *uhpT* promoter is already induced at a much lower concentration. Therefore, in the presence of glucose-6-phosphate, growth rate and glycolytic flux are high, and the cAMP concentration is low. With a qualitatively identical regulation mechanism (Figure 6C), but different parameters, the *acs* promoter is turned off in these conditions, whereas the *uhpT* promoter is fully active. When glucose-6-phosphate is exhausted, growth rate decreases and the concentration of cAMP increases. *Acs* is now fully expressed. According to the model, *uhpT* should still be maximally expressed. However, this promoter also requires the inducer glucose-6-phosphate for full activity, and this inducer is no longer present. *UhpT* transcription is thus shut off (see supplementary information: the metabolic toggle switch). Even though

the underlying regulatory mechanism is the same for both promoters, the observed expression pattern is very different. The same is true for the *gntT* promoter, responding to gluconate. This dual antagonistic effect of inducers has previously been described in a study that also clearly points out problems with the cAMP model (Peekhaus & Conway, 1998).

### A unifying view of carbon catabolite repression

We therefore propose that carbon catabolite repression comprises the "classical" mechanisms involving the PTS and the concentration of cAMP. In addition, a yet unknown sensor detects the "metabolic state" of the cell, closely related to the glycolytic flux. The unknown mechanism (i) is independent of the concentration of cAMP, the phosphorylation state of $EIIA_{glc}$ and the concentration of Crp; (ii) is correlated with the glycolytic flux, but unrelated to PTS functioning, and thus unrelated to the PEP/pyruvate ratio (Hogema *et al.*, 1998); (iii) functions as a rapid switch, at the scale of signal transduction events. The activity of the CCR-controlled promoters results then from the synergistic activation of the target promoter by these two signals. We are currently working on the identification of this unknown signal and the way this signal is ultimately transmitted to the activator, Crp.

## Materials and Methods

### Strains

The mutant strains come from the Keio collection (Baba *et al.*, 2006). The wild-type strain (*E. coli* K-12 strain BW25113) has a well-defined pedigree and has not been subjected to mutagens (Datsenko & Wanner, 2000). This wild-type strain contains a deletion of the *araBAD* operon. Consequently, arabinose, that we add as inducer, can't be used as a carbon source and thus its concentration remains constant. We removed several kanamycin cassettes from mutants using flp recombinase to allow antibiotics compatibility with the plasmids derived from the Zaslaver collection. The *cyaA crr* double mutant was constructed by removing the kanamycin resistance cassette of the *cyaA* strain. Then we transduced (P1 phage transduction) the kanamycin resistance cassette of the *crr* strain in the *cyaA* strain and then remove again the new kanamycin cassette. The *crp** strain comes from a gift of Pr. Patrick Cirino (Khankal *et al.*, 2009). The *crp** gene was introduced into the *cyaA* mutant and the *cyaA crr* double mutant by P1 phage transduction using a tetracycline resistance cassette.

## Plasmids

We have used home-made p-acs-lux, p-acs-gfp and p-sdh-lux plasmids whose construction has already been described. We have also used the p-uhpT-gfp, p-araBAD-gfp, p-sdh-gfp, p-gntT-gfp, p-glpA-gfp coming from the Zaslaver collection (Zaslaver *et al.*, 2006). Note that *luxCDABE* genes come from Xenorhabdus luminescens and the *gfp* variant used (*gfpmut2*) is highly stable and non toxic. The plasmid carries either a colE1 replication origin and an ampicillin resistance cassette (*lux*) or a pSC101 origin and a kanamycin resistance cassette (*gfp*). The two types of plasmid are fully compatible and they include a strong ribosome binding site just upstream the reporter gene.

The *crp* gene was amplified by PCR using primers 5'-ATAAGGTACCATGGTGCTTGGCAAAC CGC-3'(kpnI site underlined) and 5'-TCCGGGATCCTTAACGAGTGCCGTAAAC-3'(bamH1 site underlined) using BW25113 strain as template. The PCR product, digested with kpn1 and bamH1, was inserted into the previously described p-tet-lux plasmid to make p-tet-crp.

## Measurement of gene expression at the population level

Dynamic measures of gene expression were performed by growing cells in a minimal medium (M9) with the indicated carbon sources during several hours within a microplate reader (fusion alpha FP-HT). This micro plate reader allows growth with shaking and temperature control (37°C). It measure cell density (OD at 600nm), luminescence and fluorescence (485nm/520nm) at a high temporal resolution ( 5 minutes). The microplate contains 180$\mu$l of medium and 20$\mu$l (dilution 10) of cells in stationary phase (pre-culture LB+corresponding antibiotics). The main parameters that differ between experiments are the cAMP, inducers, anhydrotetracycline (aTc), glucose and acetate concentrations (indicated in the legend of figures).

## Measurement of gene expression at the single cell level

Briefly (details of each experiment are provided in the supplementary section), we add 1 ml of a minimal medium M9 agarose 0.8% (w/v) (with the indicated concentration of carbon sources, inducers, cAMP and antibiotics) into a 6 wells micro plate. After polymerization of agarose, the previous medium becomes solid and we add over the indicated volume of liquid M9 containing cells in stationary or exponential phase (pre-culture LB + antibiotics). We let some cells fall on the solid medium during 10 minutes

and then remove this liquid medium. Cells, which are placed on the top of the agar pad, get a perfect access to oxygen and begin to form micro-colonies. Image acquisition is performed using a 40X objective of the Olympus IX81 motorized inverted microscope. Note that the solid medium is placed between the objective (down) and the cells (up). Fluorescence and phase contrast images are collected each ten minutes during several hours at room temperature. Image processing is limited to the superposition of phase contrast and fluorescence images and to the normalization between minimum and maximum pixel intensity of the images. We have used home-made scripts and the image processing toolbox of Matlab.

## Acknowledgments

## References

Ambrose, M., & MacPhee, D. G. 1998. Catabolite repressors are potent antimutagens in *Escherichia coli* plate incorporation assays: experiments with glucose, glucose-6-phosphate and methyl-alpha-D-glucopyranoside. *Mutat Res*, **398**(1-2), 175–82.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, **2**, 2006 0008. Gm62662/gm/nigms Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England.

Buettner, M. J., Spitz, E., & Rickenberg, H. V. 1973. Cyclic adenosine 3',5'-monophosphate in *Escherichia coli*. *J Bacteriol*, **114**(3), 1068–73.

Crasnier-Mednansky, M. 2008. Is there any role for cAMP-CRP in carbon catabolite repression of the *Escherichia coli* lac operon? *Nat Rev Microbiol*, **6**(12), 954; author reply 954.

Datsenko, K. A., & Wanner, B. L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**(12), 6640–5.

Epstein, W., Rothman-Denes, L. B., & Hesse, J. 1975. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in Escherichia coli. *Proc Natl Acad Sci U S A*, **72**(6), 2300–4.

Gorke, B., & Stulke, J. 2008a. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, **6**(8), 613–24.

Gorke, B., & Stulke, J. 2008b. Is there any role for cAMP-CRP in carbon catabolite repression of the *Escherichia coli* lac operon? Reply from Gorke and Stulke. *Nat Rev Microbiol*, **6**(12), 954.

Gosset, G. 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact*, **4**(1), 14.

Hantke, K., Winkler, K., & Schultz, J. E. 2011. *Escherichia coli* exports cAMP via TolC. *J Bacteriol*. Journal article.

Hogema, B. M., Arents, J. C., Inada, T., Aiba, H., van Dam, K., & Postma, P. W. 1997. Catabolite repression by glucose 6-phosphate, gluconate and lactose in *Escherichia coli*. *Mol Microbiol*, **24**(4), 857–67.

Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Yoshida, H., Takahashi, H., Aiba, H., & Postma, P. W. 1998. Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIAGlc. *Mol Microbiol*, **30**(3), 487–98.

Inada, T., Kimata, K., & Aiba, H. 1996. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells*, **1**(3), 293–301.

Khankal, R., Chin, J. W., Ghosh, D., & Cirino, P. C. 2009. Transcriptional effects of CRP* expression in *Escherichia coli*. *J Biol Eng*, **3**(1), 13.

Kolb, A., Busby, S., Buc, H., Garges, S., & Adhya, S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem*, **62**, 749–95.

Kumari, S., Beatty, C. M., Browning, D. F., Busby, S. J., Simel, E. J., Hovel-Miner, G., & Wolfe, A. J. 2000. Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. *J Bacteriol*, **182**(15), 4173–9.

Magasanik, B. 1961. Catabolite repression. *Cold Spring Harb Symp Quant Biol*, **26**, 249–56.

Megerle, J. A., Fritz, G., Gerland, U., Jung, K., & Radler, J. O. 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys J*, **95**(4), 2103–15.

Monod, J. 1942. Recherches sur la croissance des cultures bactériennes. *2nd edn. Paris:Hermann.*

Narang, A. 2009a. cAMP does not have an important role in carbon catabolite repression of the *Escherichia coli* lac operon. *Nat Rev Microbiol*, **7**(3), 250.

Narang, A. 2009b. Quantitative effect and regulatory function of cyclic adenosine 5'-phosphate in Escherichia coli. *J Biosci*, **34**(3), 445–63.

Park, Y. H., Lee, B. R., Seok, Y. J., & Peterkofsky, A. 2006. In vitro reconstitution of catabolite repression in *Escherichia coli*. *J Biol Chem*, **281**(10), 6448–54.

Peekhaus, N., & Conway, T. 1998. Positive and negative transcriptional regulation of the *Escherichia coli* gluconate regulon gene gntT by GntR and the cyclic AMP (cAMP)-cAMP receptor protein complex. *J Bacteriol*, **180**(7), 1777–85.

Perlman, R. L., De Crombrugghe, B., & Pastan, I. 1969. Cyclic AMP regulates catabolite and transient repression in E. coli. *Nature*, **223**(5208), 810–2.

Postma, P. W., Lengeler, J. W., & Jacobson, G. R. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev*, **57**(3), 543–94.

Saier, M. H., & Roseman, S. 1972. Inducer exclusion and repression of enzyme synthesis in mutants of Salmonella typhimurium defective in enzyme I of the phosphoenolpyruvate: sugar phosphotransferase system. *J Biol Chem*, **247**(3), 972–5.

Ullmann, A., & Monod, J. 1968. Cyclic AMP as an antagonist of catabolite repression in *Escherichia coli*. *FEBS Lett*, **2**(1), 57–60.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods*, **3**(8), 623–8.

# Supplementary information

Author1[1], Author2[2], Author3[3,*]

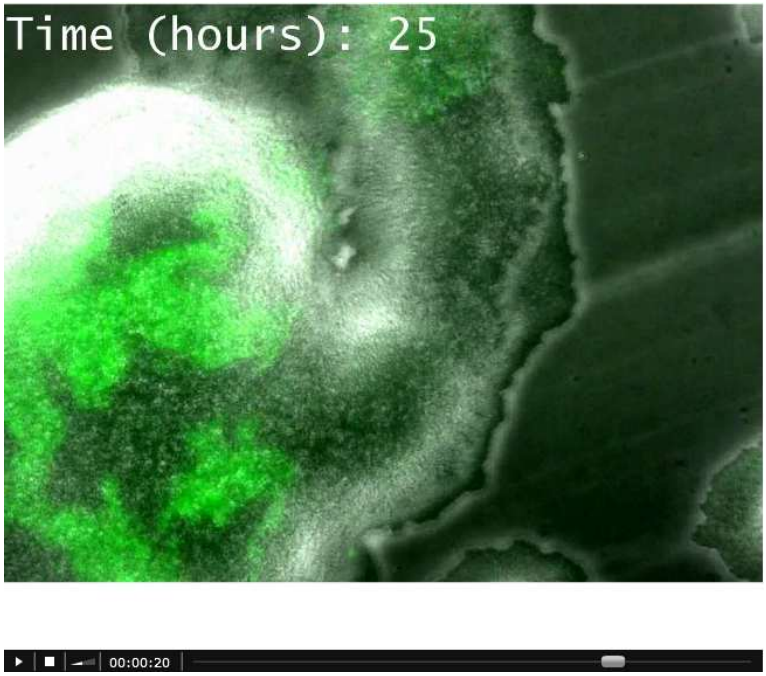
1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## 1 Video 1: Wt p-acs-gfp vs *cyaA*- p-acs-gfp




### Materials and methods

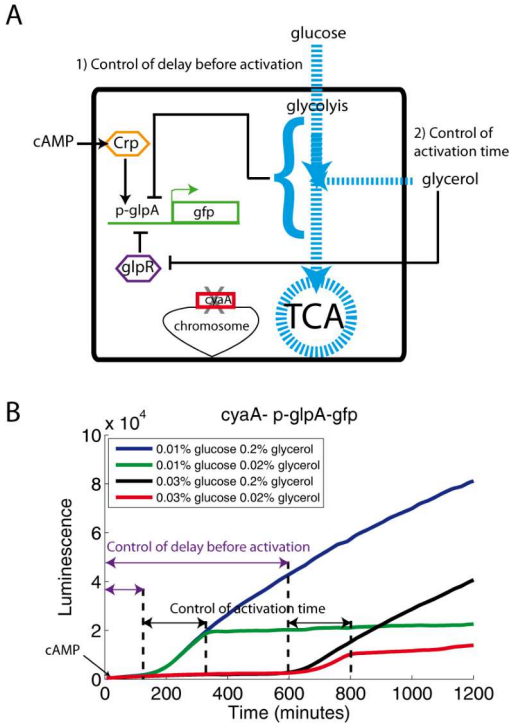
We added 1 ml of a minimal medium M9, containing 0.8% (w/v) agarose, 0.3% glucose, and 50$\mu$g/ml kanamycin into a well of a 6 wells-microplate. After polymerization of the agarose, we overlaid the now solid medium with 400$\mu$l of M9 medium and a mixture of two bacterial strains precultured in LB: 200$\mu$l of the wild-type strain containing the p-acs-gfp plasmid and 200$\mu$l of the *cyaA*- strain containing the p-acs-gfp plasmid. The precultured cells were still in exponential phase (OD>0.5) and their growth

medium, LB, was thus still "fresh". Part of the cells sedimented onto the solid medium during a 10 minute incubation. We then removed the liquid medium. Note that the remaining traces of fresh LB medium may "help" the growth. The remaining cells are thus located on the top of the agar pad; they get a perfect access to oxygen and begin to form microcolonies. Image acquisition was performed using 40X objective of the Olympus IX81 motorized inverted microscope. The solid medium is located between the objective (down) and the cells (up). Fluorescence and phase contrast images were collected every ten minutes for several hours at room temperature. Image processing was limited to the superposition of phase contrast and fluorescence images and to the normalization between minimum and maximum pixel intensity of the entire batch of images. All image treatments were done with home-made scripts and the image processing toolbox of Matlab.

## Results

Towards the end of the video we observe a strong increase of fluorescence in parts of the microcolony, corresponding to the induction of *acs* transcription. The time of fluorescence increase coincides with the growth arrest of the microcolony, probably due to glucose exhaustion. Other parts of the microcolony, corresponding to the *cyaA*- strain, remain dark (no fluorescence increase), as expected. In this video, we also observed six cells that were fluorescent already at the beginning of the experiment. Our interpretation of this observation is that these cells have a strongly reduced glycolytic flux (probably due to an unfavorable local environment, spontaneaous mutations, the particular history of these cells, or a bet-hedging strategy of the cell), and consequently express *acs*.

## 2 The *glpA* promoter



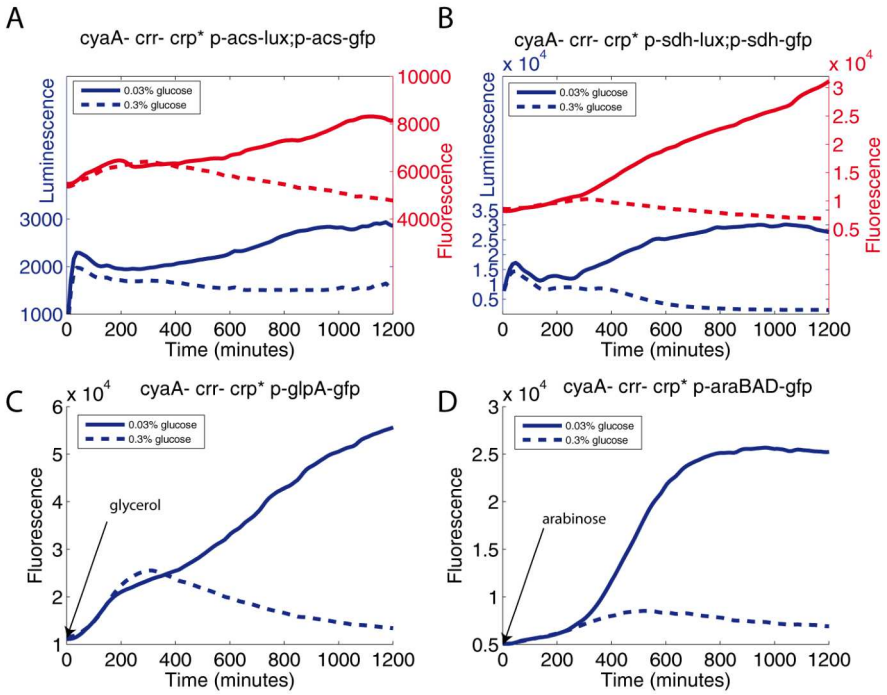

### Materials and methods

A *cyaA*- strain transformed with the *p-glpA-gfp* plasmid (A) was grown in M9 medium containing glucose, glycerol and cAMP (500$\mu$M) (B). In this experiment, we have centrifuged the LB pre-culture and replaced the LB by an equivalent volume of fresh M9 before the 10-fold dilution of the cells.

### Results

This promoter behaves as the *acs*, *sdh* and *araBAD* promoters. Glycerol and cAMP, added from the beginning of the kinetics, are not sufficient to induce the expression of *glpA* while glucose is still present. This behavior may be interesting for synthetic biology: glucose represses the activity of the *glpA* promoter and thus allows controlling the delay before the activation. Once glucose is exhausted, glycerol activates the promoter by inactivating the GlpR repressor: we observe gene expression. When glycerol is also exhausted, the activity of the *glpA* promoter drops again to zero.

# 3 Carbon catabolite repression in the *cyaA crr crp** strain



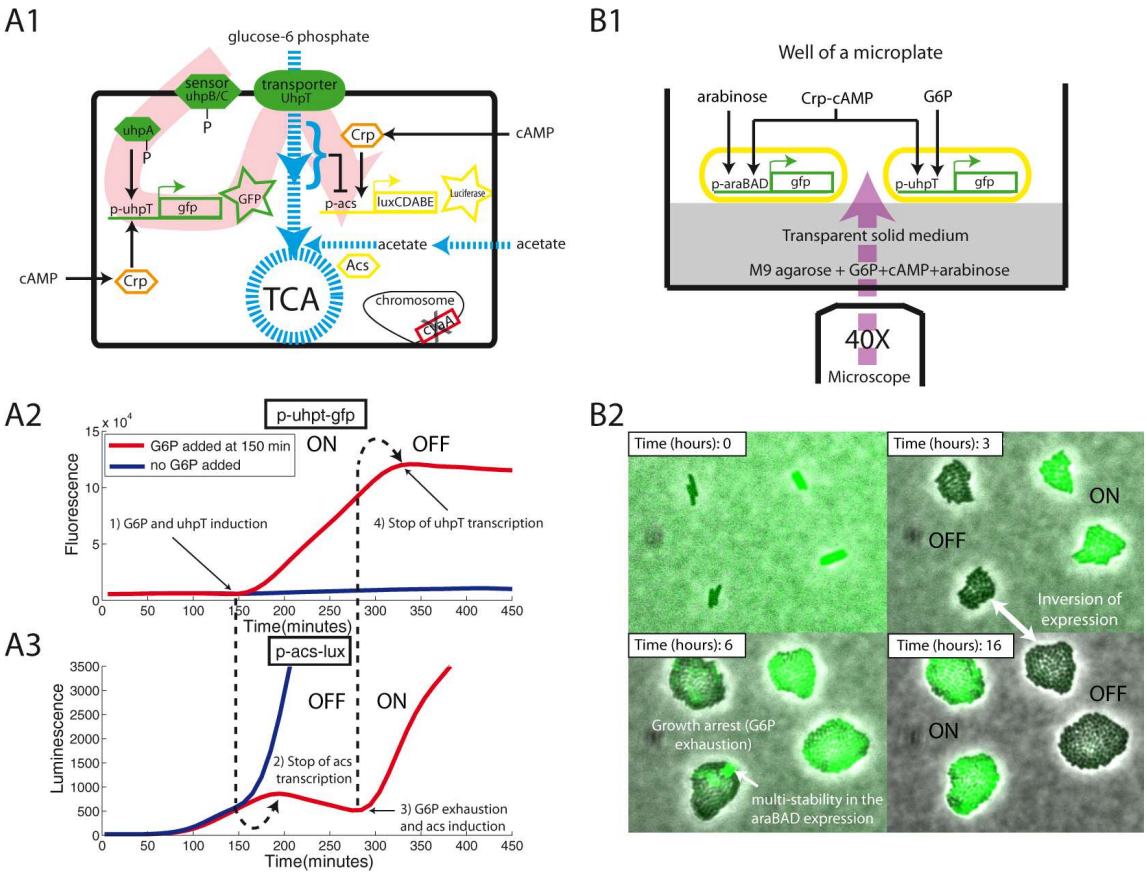

## Materials and methods

The *cyaA crr crp** strain, transformed with the indicated plasmids (transcriptional fusions), were grown (10-fold dilution of the preculture) either in an LB medium supplemented with 0.03% glucose or in a LB medium supplemented with 0.3% glucose. Inducers were added when needed: Glycerol (0.08%) or arabinose (0.001%).

## Results

In this experiment, we measure the activity of 4 Crp-cAMP dependent promoters (*acs*, *sdh*, *glpA*, and *araBAD*) in the "cAMP and EIIA$_{glc}$ free" strain (*cyaA crr crp**). We compare the level of fluorescence of theses strains growing in LB either with a low glucose concentration or a high glucose concentration. In all cases (A, B, C and D) we observe that the level of fluorescence (or luminescence) increases in the low glucose condition compared to the high glucose condition.

# 4 A synthetic toggle switch controlled by the glycolytic flux




## Description of the figure

(A1) Schematic of a *cyaA* mutant transformed with two transcriptional fusions: p-acs-lux and p-uhpT-gfp. (A2 and A3) Bacteria were grown on a minimal medium supplemented with 0.03% glucose and 0.03% acetate. At 100 minutes, we observed the first *acs* induction. Adding G6P (0.03%) to the medium triggers *uhpT* expression, leading to a strong glycolytic flux that stops *acs* expression. Once exhausted, G6P cannot longer activate *uhpT* transcription and a new phase of *acs* induction begins. (B1) Schematic of a well containing a solid minimal medium containing the following carbon sources and complements: G6P (0.06%), cAMP (750$\mu$M), arabinose (0.01%), and acetate (0.03%). Two types of *cyaA* cells monitor either (i) the expression of *araBAD* or (ii) the expression of *uhpT*. (B2) Frames of the time lapse microscopy:

the two microcolonies on the left report *araBAD* expression whereas those on the right report *uhpT* expression. Note the inversion of expression after growth arrest (presence/absence of the glycolytic flux). Note also the heterogeneity of fluorescence among cells reporting the *araBAD* expression. This phenomenon of multi-stability is well-known : it is due to a positive feedback loop involving the arabinose transporter.

## Results

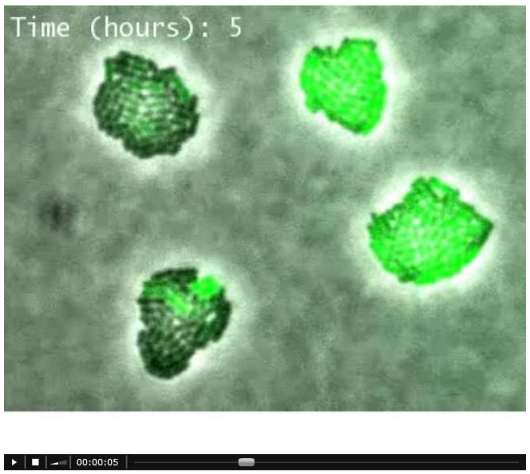
Our experiments suggest that CCR-controlled promoters are activated as a function of a metabolic parameter related to growth rate, e.g., the glycolytic flux. The different CCR-controlled promoters differ by their cofactor requirements (inducers such as arabinose or glucose-6-phosphate) and their sensitivity to cAMP. At the two extremes are the *acs* and *uhpT* promoters. To visualize the complementary behavior of these two promoters, we have constructed a synthetic toggle switch.

Both promoters are dependent on Crp-cAMP, but to different extents. Whereas *acs* requires a high concentration of cAMP for maximal activity, *uhpT* already reaches maximal activity at a much lower concentration of cAMP, but in the presence of the inducer, glucose-6-phosphate. The *acs* promoter is repressed by glucose-6-phosphate, an excellent carbon source allowing fast growth of *E. coli,* whereas this compound is the indispensable inducer of the *uhpT* promoter. In our system (figure A1), two transciptional fusions (p-acs-lux and p-uhpT-gfp) monitor the activity of their respective promoters in a *cyaA* strain. As before, this strain induces *acs* expresssion when glucose is exhausted (luminescence; figure A3). The *uhpT* promoter remains inactive because the inducer is absent. We then added glucose-6-phosphate during the induction phase of the *acs* promoter, resulting in the activation of the *uhpT* promoter (increase of fluorescence; figure A2). At the same time, glucose-6-phosphate now serves as an excellent carbon source, increasing the glycolytic flux and shutting down the *acs* promoter. Once glucose-6-phosphate is exhausted, *acs* transcription is again induced and *uhpT* transcription is stopped. The control experiment, without addition of glucose-6-phosphate, shows no transcription of *uhpT* and no arrest of *acs* transcription, as expected.

A similar switch can be observed at the single-cell level (figure B1; supplementary movie 2). We filmed two types of *cyaA* cells growing on the same agar patch, using glucose-6-phosphate as a carbon source: the first cell type reported the *araBAD* expression (p-araBAD-gfp) that occurs after glucose-6-phosphate exhaustion, whereas the second cell type reported *uhpT* expression (p-uhpT-gfp) that is induced in

the presence of glucose-6-phosphate. All inducers are present from the beginning of the experiment (cAMP, glucose-6-phosphate and arabinose). In figure B2, we show the inversion of expression of GFP as glucose-6-phosphate is being consumed. Cells reporting *uhpT* expression (on the right) are induced during the growth phase, whereas cells reporting *araBAD* expression are induced after exhaustion of glucose-6-phosphate.

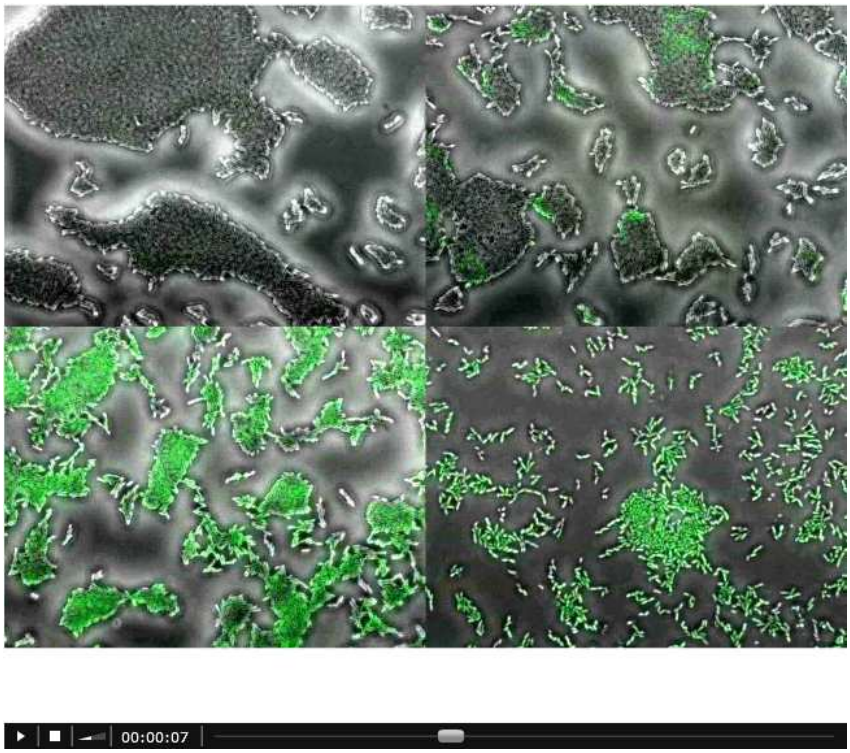
### Video 2: *cyaA*- p-araBAD-gfp vs *cyaA*- p-uhpT-gfp




### Materials and methods

This experiment used the same protocol as that used for Video 1, with the following experimental conditions. The medium used was M9 agarose 0.8%, sodium acetate 0.03%, glucose-6-phosphate 0.06%, cAMP 750$\mu$M, arabinose 0.01%, and kanamycin 50$\mu$g/ml. We used *cyaA*- strains either with the p-araBAD-gfp plasmid or the p-uhpT-gfp plasmid. After polymerization of the agarose, we add over 1 ml of M9 and 20$\mu$l (dilution 50-fold) of each strain, pre-cultured in LB (stationary phase). After having removed the liquid medium, the micro-plate was placed at 37°C for 30 minutes before beginning of the image acquisition (this step facilitates automatic detection of the focus). The fluorescence intensity was not normalized between minimum and maximum pixel intensity of the entire batch of images, but for each image independently. This allows a simpler comparison of the level of fluorescence of the two strains. However, the resulting pictures suggest that GFP is degraded during the experiment, which is not the case.

# 5 Video 3: glycolytic flux and catabolite repression




## Materials and methods

We used the same method as the one of Video 1 with the following modifications. The medium used is M9 agarose 0.8%, glucose 0.06%, and arabinose 0.005%. We used a *sucB* mutant transformed with the *p-araBAD-gfp* plasmid. This strain is used because it can not grow on the secreted acetate. This leads to a very sudden growth arrest when glucose is exhausted (see the video). In this experiment, we filmed four wells in parallel. The only difference between the wells is the concentration of $\alpha$-methyl-glucose (top left: 0%; top right: 0.3%; bottom left: 1%; bottom right: 3%). We assembled the four videos with Simulink (Video and Image Processing Blockset).

## Results

On the top left of the video (Well 1), there is no $\alpha$-methyl glucose: cells grew fast in the presence of glucose. After glucose exhaustion, they activate synchronously the *araBAD* promoter (rise of fluorescence). On

the top right of the video (Well 2), the $\alpha$-methyl glucose has slightly reduced the glycolytic flux and some cells induce the *araBAD* expression before glucose exhaustion. On the bottom left of the video (Well 3), the glycolytic flux is strongly reduced, the growth is slow and most cells induce *araBAD* expression before glucose exhaustion. In the last well (bottom right), the $\alpha$-methyl glucose concentration is so high that cells seem poisoned (very little growth), but all induce *araBAD* expression before the glucose exhaustion.

# 6 Critical parameters to reproduce our results

Given the contradictory results found in the literature about glucose repression, we provide some additional guidelines. Our results are most clearcut for the following experimental conditions:

- using a *cyaA*- strain with exogenous cAMP (or a *crp** strain; and in this case, do not add cAMP).
- using either minimal medium or LB (begin with minimal medium, and switch to LB in case of problem). In our experiments, the transition from pre-culture (LB) to culture is done with a low dilution factor (factor of 10: 180$\mu$l of medium + 20$\mu$l of pre-culture). This ensures a rapid exhaustion of the carbon sources.
- performing the kinetics in micro-plate in order to easily measure many different concentrations of:
  - cAMP
  - inducers (if needed)
  - Glucose: best results are obtained by comparing 0.03% glucose with 0.3% glucose. These conditions focus on what happen when glucose is exhausted (first concentration), compared to continued growth on glucose (second concentration).
- adding an alternative carbon source to ensure correct transcription/translation capabilities when glucose is exhausted. This carbon source must permit very slow growth (best results with sodium acetate 0.03% w/v).

# Chapitre IV

A synthetic *Escherichia coli* communication system mediated by extracellular cyclic AMP

# A synthetic *Escherichia coli* communication system mediated by extracellular cyclic AMP

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## Abstract

In *Escherichia coli*, intracellular cAMP is involved in carbon catabolite repression. It binds to the cAMP receptor protein (CRP) to control the transcriptional of more than 220 operons. *Escherichia coli* also actively exports cAMP into the growth medium where this molecule reaches non-negligible concentrations. However its role is not understood. In this study, we have engineered strains capable of communicating with each other by way of extracellular cAMP. We propose this new communication module to the synthetic biology community and we discuss its potential physiologic role in natural environment.

## Introduction

3'-5'-cyclic adenosine monophosphate (cAMP) is a second messenger performing intracellular signaling in many organisms. Discovered in the 60s, the importance attributed to this compound has never stopped growing (Makman & Sutherland, 1965). The number of publications involving cAMP now exceeds 100000. In humans, the intracellular cAMP-dependent pathway is involved in the regulation of many biological processes ranging from glycogen metabolism to synaptic transmission and the control of heart rate. The deregulation of concentrations of intracellular cAMP is observed in many diseases such as cholera, anthrax, cancers or neurological disorders.

In addition to its role as an intracellular second messenger, cAMP also acts as first messenger (Konijn *et al.*, 1967). This function of cAMP was first discovered in the amoebae *Dictyostelium discoideum*, where extracellular cAMP mediates communication between cells by acting as a chemo-attractant. Periodic waves of extracellular cAMP allow these amoebae to convert from solitary predators into gregarious

community members (Tomchik & Devreotes, 1981). In the 1970s, some reports mentioned that cAMP was actively exported in the blood of humans (and then urine) where it could attain significant concentrations (in the range 10-100 nM) (Broadus *et al.*, 1970). Thirty years later, a wide body of experimental observations suggests that extracellular cAMP exerts endocrine and paracrine actions in several tissues (Bankir *et al.*, 2002).

In *Escherichia coli*, cAMP is involved in carbon catabolite repression (Gorke & Stulke, 2008; Narang, 2009) and binds to the cAMP receptor protein (CRP). The corresponding complex is a transcriptional factor controlling the expression of more than 220 operons (Keseler *et al.*, 2009). It has been known for a long time that *E. coli* actively exports cAMP into the growth medium (Goldenbaum & Hall, 1979; Makman & Sutherland, 1965). However, the role of this extracellular cAMP is not understood. In this study, we have engineered strains capable of communicating with each other by way of extracellular cAMP. Even though we have engineered the strains to clearly show communication via cAMP, the same mechanism could operate in a natural setting and provide a mean for communication between species that are phylogenetically only very distantly related.

# Results

## Communication on solid medium

To build an inter-cellular communication system, we engineered a sender strain capable of producing more than the usual amount of cAMP and a receiver strain able to sense this exported cAMP and to report its presence by a convenient signal. The simplest sender strain is the *crp* mutant, known to produce high extracellular cAMP level compared to a wild-type strain (Potter *et al.*, 1974). The receiver strain harbors a deletion of the *cyaA* gene (encoding the cAMP biosynthetic enzyme adenylate cyclase) in order to prevent any cAMP production. This *cyaA* mutant also contains a "bio-sensor" that converts the cAMP concentration into a luminescence signal. Placed on a plasmid, this biosensor is a transcriptional fusion, i.e, a promoter fused to a reporter gene. We chose the Crp-cAMP dependent *sdhCDAB* promoter (Nam *et al.*, 2005) and fused it upstream of the *luxCDABE* operon coding for bacterial luciferase (Van Dyk *et al.*, 2001).

To demonstrate inter-cellular communication, we used an interface assay. We embedded the sender cells into agarose gel containing minimal glucose medium on a Petri dish. The solid agarose gel containing

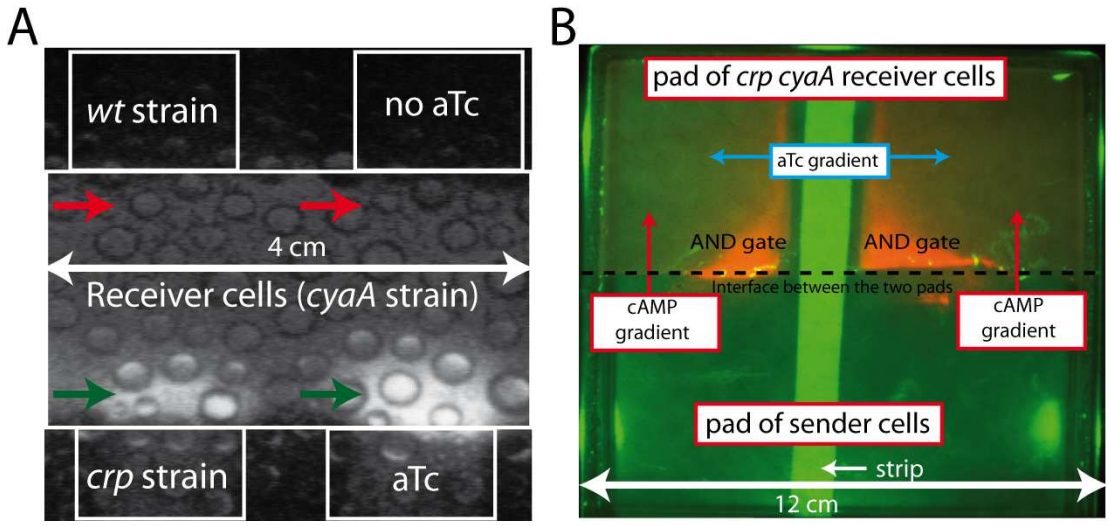



**Figure 1. Demonstration of communication in solid medium.** (A) Luminescence picture (seen from above) of a solid medium in a Petri dish acquired in a dark chamber with a sensitive camera (the round droplets are due to condensation on the lid). In the middle, a long rectangular pad of *cyaA* receiver cells containing the luminescence biosensor produces a basal level of light. On the top left, the interface between the receiver pad and the wild-type pad does not produce any rise of light (red arrow). Conversely, on the bottom left, a halo of light is observed at the interface between the "*crp* cells" pad and the "receiver cells" pad (green arrow). On the bottom right, sender cells overexpress adenylate cyclase (by the way of aTc) and a halo of light is observed at the interface (green arrow). On the top right, no halo appears because adenylate cyclase is not expressed (absence of aTc) (red arrow). (B) Picture of a Petri dish seen from above. Luminescence data (red false color) and photo of a Petri dish (green false color) are overlaid. One half of the Petri dish contains an agar pad of sender cells producing cAMP which diffuses in the agar pad of receiver cells (second half). At the middle, and perpendicular to the interface, a thin strip of cellulose soaked with anhydrotetracycline (aTc) is deposited on the agar and allows expression of Crp. Two bright areas (on either side of the strip) appear where the presence of both optimal aTc and cAMP levels allow formation of the Crp-cAMP complex in the receiver cells, and thus luciferase transcription.

the sender bacteria is brought into contact with a similar pad containing receiver cells. The Petri dish is placed in a heated, dark chamber and filmed for 40 hours using a sensitive CCD camera (see supplementary movie 1). An intense "halo" of luminescence appears in the area of contact between the "receiver cells" and the *crp* strain, called "sender cells" (Figure 1A; bottom left side). cAMP produced by the "sender area" has diffused into the "receiver area" and elicited the expected response: entering the receiver cells, binding to Crp and activating the *sdh* promoter, thereby inducing the expression of luciferase. By contrast, no "halo" appears at the interface between receiver cells and a wild-type strain (Figure 1A; top left side). This observation shows that a wild-type strain does not produce a sufficient amount of extracellular cAMP to allow the detection of communication.

Communication could be due to the exchange of any molecule, not necessarily cAMP. In order to demonstrate that cAMP is the signaling molecule, we remove the gene coding for adenylate cyclase (*cyaA*) in the sender strain. We put into the *cyaA crp* double mutant a plasmid that expresses *cyaA* from an anhydrotectracycline (aTc) inducible promoter. This strain produces cAMP only when the inducer aTc is present in the medium. As expected, a halo of luminescence appears only at the interface of the pad containing aTc (Figure 1A; right bottom side versus right top side). This experiment demonstrates that the expression of adenylate cyclase in the sender strain is necessary for inducing transcription of luciferase in the receiver strain.

A second way to demonstrate that cAMP acts as an intercellular signaling molecule is to show that the presence of the *crp* gene in the receiver strain is needed for the appearance of the luminescence halo. We transformed a *cyaA crp* double mutant with the plasmid PR100, carrying a construct that over-expresses *cyaA*. This latter is under the control of the strong lambda pL promoter and the gene carries the canonical AUG initiation codon, instead of the less efficient TTG found in the chromosomal *cyaA* gene. With this expression system, adenylate cyclase constitutes approximately 30% of the cellular proteins. As a receiver strain we also use the *cyaA crp* double mutant containing two plasmids. The first one carries the luminescence biosensor and the second plasmid allows to tune the expression of Crp by the aTc inducible system. We juxtaposed two agar pads containing these two strains and we added, overlaid in the middle of the Petri dish, and perpendicular to the generated interface, a thin strip of paper soaked with a solution of aTc. Figure 1B shows that in the area where the aTc gradient (expression of Crp) and the cAMP gradient generated by the sender cells overlap, we observe a bright luminescence of the reporter plasmid. Activation of the reporter gene corresponds to the AND logic gate behavior of the

Crp-cAMP complex. Communication passes through the first messenger cAMP produced by sender cells and requires the Crp protein as a second component in the receiver cells.

## Tuning the communication in liquid medium

We have shown that communication via cAMP functions on solid medium. Is the amount of excreted cAMP sufficient to produce communication in liquid medium? In order to obtain quantitative information, we decided to duplicate the reporter system. In addition to the luciferase reporter, we added to the receiver strain an equivalent gfp reporter, also driven by the *sdh* promoter (Zaslaver *et al.*, 2006), but carried on a plasmid with a different origin of replication (pSC101). We measure the *sdh* promoter activity independently with each reporter system in the same cell (Figure 2A1). This double reporter system eliminates all potential artifacts due to the reporter system or the plasmid.

We first tuned the strength of communication by controlling the expression of adenylate cyclase in the sender (Figure 2A1; see also the tridimensional simulation in supplementary movie 2). We grow sender cells (allowing the control of *cyaA* expression by aTc) and receiver cells in a minimal glucose medium for 24 hours. The culture grows in a microplate and we continuously monitor luminescence, fluorescence and optical density. At specific points of the kinetics, we collected samples to measure extracellular cAMP by an immunoassay kit. Figure 2A2 shows that the extracellular cAMP concentration is proportional to the aTc level controlling adenylate cyclase expression in the sender cells. The increase of extracellular cAMP over time is gradual and reaches 80 $\mu$M after about 20h of growth.

Figures 2A3 and 2A4 show the corresponding induction of the *sdh* promoter after glucose exhaustion (at about 550 min). Both reporter systems, fluorescence and luminescence, show the same signal. This experiment shows that we can control the transcription of luciferase and GFP in receiver cells by tuning the adenylate cyclase level in the sender cells. Last, contrary to the gradual increase of extracellular cAMP, the strong increase of fluorescence and luminescence occurs only after glucose exhaustion (growth arrest indicated by an arrow on the figure 2A3). This result does not fit well with the classic model of carbon catabolite repression. Indeed, adenylate cyclase in sender cells should produce cAMP only after glucose exhaustion because its activator (the phosphorylated form of $EIIA_{glc}$.) is supposed to be present only in absence of glucose (Park *et al.*, 2006). Yet, the raise of extracellular cAMP is not linked to the glucose exhaustion and cAMP is already present at a high concentration in the medium well before *sdh* activation at 600 minutes (Figure 2A2 versus Figures 2A3 and 2A4). This result means that the

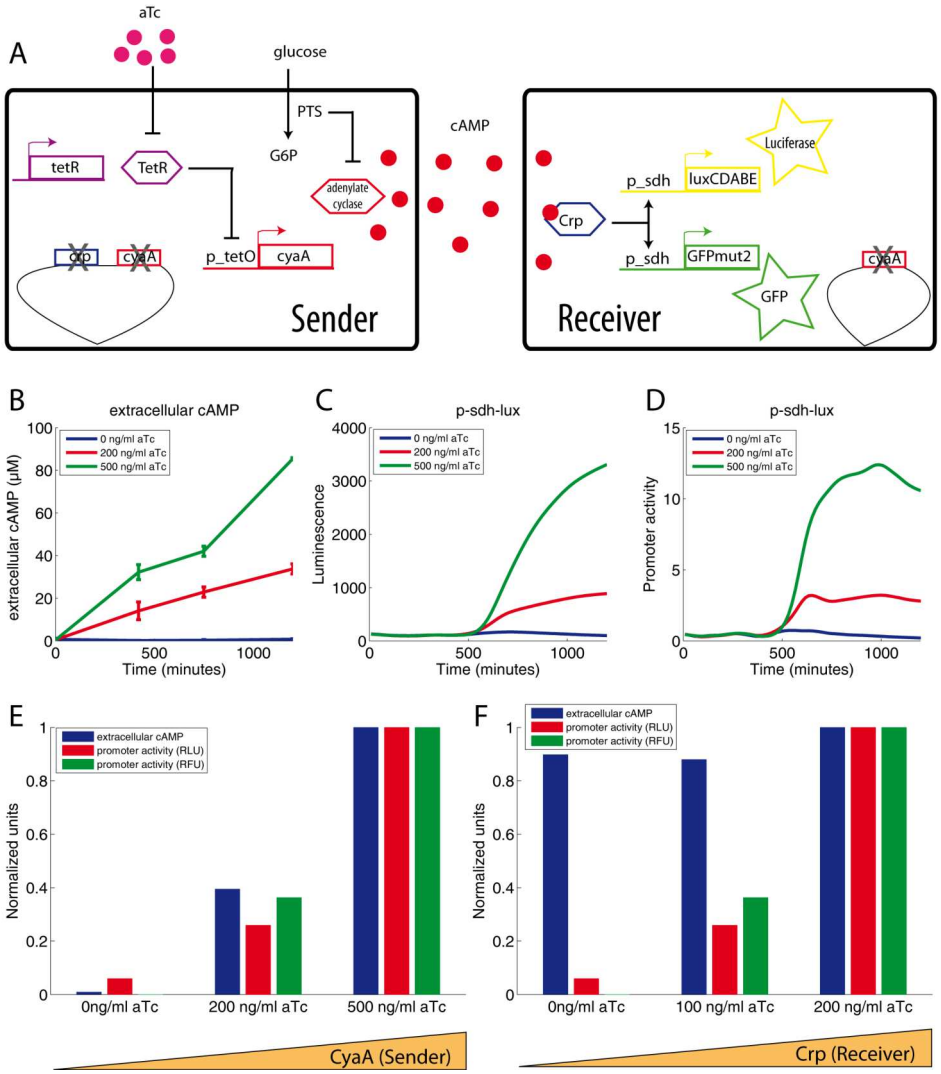



**Figure 2. Tuning the communication in liquid medium.** (A) Tuning the communication by controlling adenylate cyclase production in sender cells. (B) Extracellular cAMP is correlated with aTc levels. (C and D) The luminescence and fluorescence in the receiver cells after glucose exhaustion, when the transcription of *sdh* is induced, correlates with the aTc levels, and therefore the expression of adenylate cyclase in the sender cells and the concentration of cAMP in the medium. (E) Summary of the correlation between the concentration of cAMP in the medium and the response of the receiver cells at different concentrations of aTc; (F) Response of receiver cells with Crp expression under the control of the aTc-inducible promoter and in the presence of sender cells that produce constitutively high levels of cAMP (concentration measured in the medium: 150$\mu$M). Fluorescence is similar to the luminescence profile and is correlated to aTc levels after glucose exhaustion.

repression of *sdh* promoter activity by glucose is not due to the inactivation of adenylate cyclase in sender cells but rather due to an unknown mechanism that occurs in receiver cells.

Another way to tune the strength of communication consists in controlling the Crp concentration in the receiver cells (figure 2F). In this experiment, we used sender cells carrying the PR100 plasmid, which produces large amounts of cAMP. This microbial factory generates an extracellular cAMP concentration close to 150$\mu$M (data not shown). The fluorescence and luminescence levels follow the Crp concentration in the receiver cells (range of aTc) (Figure 2F) but only after glucose exhaustion. Once again, *sdh* promoter activity is repressed by a strong glucose repression. This latter is neither related to variation of extracellular cAMP level nor related to variation of Crp level. Another explanation must be sought.

## Recovery of motility through communication

How could communication be used to elicit a physiological response? The *cyaA* mutant has many characteristic phenotypes. A *cyaA* mutant cannot utilize carbon sources such as maltose, lactose or acetate and it cannot produce flagella, resulting in a non-motile bacterium. This lack of motility is due to the absence of FlhDC, the principal regulator of bacterial flagellum biogenesis. The transcription of the *flhDC* operon is positively regulated by the cAMP-CRP complex (Soutourina *et al.*, 1999). If communication works, cAMP produced and exported by sender cells should complement the non-motile phenotype of *cyaA* receiver cells. We designed an assay based on the faculty of *Escherichia coli* to swim though a semi-solid medium (low concentration of agarose). We filled 3/4 of a well (6-wells microplate) with this semi-solid gel. The bottom left quarter also contains sender cells. Two drops of receiver cells where spotted onto this agarose gel: one at the center of the well close to sender cells, a second spot at the top left, far from sender cells (Figure 3A). We filmed the microplate for 50 hours (supplementary movie 3) and observed that the "center" spot spreads to the bottom left quarter containing sender cells as shown by luminescence (Figure 3B). By contrast, the "top left" spot does not move. We conclude that cAMP produced by sender cells complements the *cyaA* strain and allows the recovery of motility probably by allowing *flhDC* expression followed by flagella biosynthesis. Too far from cAMP, cells of the second spot do not recover motility. It has not escaped our attention that this system can be viewed as a synthetic guidance system.

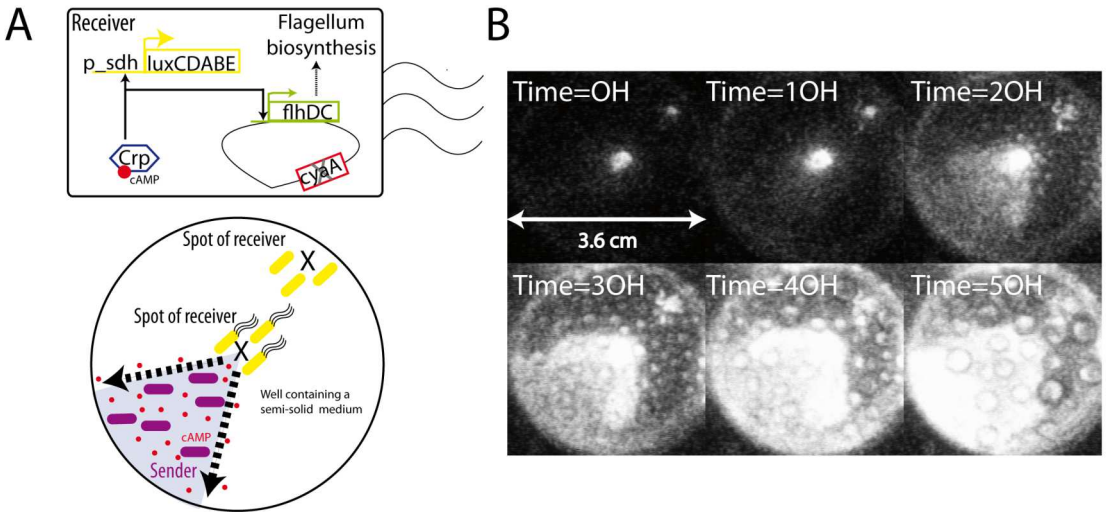



**Figure 3. Recovery of motility allowed by the communication.** (A) Schematics of a receiver cell and the geometry of the motility assay. The Crp-cAMP complex allows both luciferase expression and *flhDC* expression. The presence of cAMP produced by sender cells allows *cyaA* cells to produce flagella and to swim though the semi-solid medium. (B) Luminescence picture of the well showing the evolution of the luminescence pattern during two days (Round droplets are due to condensation on the lid). The spot of receiver cells in the middle of the well spreads to the left bottom quarter of the well because cAMP is present. In contrast, the "control" spot at the top right is too far from cAMP and cannot recover motility.

# Discussion

## A synthetic system

In this study, we have developed an easily tunable synthetic system allowing inter-cellular communication of *Escherichia coli* cells by the way of extracellular cAMP. We propose this system to the synthetic biology community and discuss these advantages and disadvantages. Contrary to other intercellular communication systems such as LuxI/LuxR from Vibrio fischeri or LasI/LasR from Pseudomonas aeruginosa (Balagadde *et al.*, 2008), the main default of our system is its absence of orthogonality. Indeed, it is not independent from the regulatory network of *E. coli*. On the contrary, the cAMP-CRP complex is considered as a "hub". The lack of orthogonality places our system at the edge between synthetic and system biology. Because our system is not uncoupled from natural regulatory network, it can be used as a tool to understand, disturb and control it. The huge knowledge on the cAMP-CRP complex can be used and embezzle to develop new synthetic functions: The high number of promoters controlled by Crp-cAMP offers a wide choice of different affinity binding site (Keseler *et al.*, 2009). The "hub" Crp-cAMP is already connected to a lot of different cellular modules (motility, chemotaxis, virulence, biofilm, stress resistance, carbon utilization) and composes a wide number of known network motifs such as coherent and incoherent feed forward loops (Alon, 2006). The connection of the Crp-cAMP "hub" to metabolism, via the mechanism of carbon catabolite repression, is a case study for system biology and different already existing models could be re-used (Bettenbrock *et al.*, 2006; Hardiman *et al.*, 2007; Kotte *et al.*, 2010; Nishio *et al.*, 2008). Last, cAMP is strongly ubiquitous: it's a messenger shared by most of the living species. Extracellular cAMP is thus a good candidate to perform synthetic inter-kingdom communication systems.

## A microbial factory and a biosensor as new tools

In this study, we engineered a strain able to produce up to 150$\mu$M extracellular cAMP i.e well above the concentration reached by a wild-type strain (100nM-10$\mu$M) (Matin & Matin, 1982). It's still far from the concentration (30mM) produced by Microbacterium sp. no. 205 (Ishiyama, 1990). It is therefore unlikely that our strain is competitive on an industrial scale. However the current yield can probably be optimized by adapting the growth conditions or by strain improvement, for instance, by removing the phosphodiesterases (*cpda*) which has a reported Km of 500$\mu$M (Imamura *et al.*, 1996) or by modifying the

phosphate transfer system to improve adenylate cyclase activation (Gosset, 2005). In this study, we also described a "biosensor", a *cyaA* strain that convert exogenous cAMP concentration into luminescence and fluorescence signals. The lower limit of detectable concentrations is closed to 10$\mu$M. The active export of cAMP is probably the main reason explaining why the biosensor fails to easily report extracellular cAMP concentration yet equivalent to physiologic intracellular concentration in a wild-type strain. Removing this mechanism of export may increase the sensitivity of our biosensor to reach the Crp km. Last, the unknown phenomenon of glucose repression observed in the figure 2 but unrelated to cAMP concentration is a parameter that influences the output signals generated by our reporters. Understanding and removing this phenomenon may also increase the performance of the biosensor.

### A potential physiologic system

In this study, we do not demonstrate that communication by cAMP is a physiologic system in *Escherichia coli*. Indeed wild-type cells produce a maximum extracellular cAMP level (5-10$\mu$M) which roughly corresponds to the minimum threshold of detection of our biosensor. Is that mean that the reachable extracellular cAMP concentration must not change drastically the gene expression of the population? But then why is cAMP exported in the medium? It has already be demonstrated that the cost of the cAMP export in term of ATP molecules is too high to only serve as a mechanism used to regulate intracellular cAMP level (Matin & Matin, 1982). Could we imagine that a tiny fraction of the population begins to sense the extracellular cAMP and expresses Crp-cAMP dependant operon as a bet hedging strategy? is *E. coli* uses its secreted extracellular cAMP to communicate with other species or less "anthropocentrically" to change their phenotypes? One track to search for physiologic communication process may be to look towards the natural environment of *E. coli*: the intestinal ecosystem. By way of conclusion, we invite the reader to watch the movie 4 (and the explanations in supplementary information) where we made an attempt in this direction.

## Materials and Methods

### Strains and plasmids

We used the *E. coli* K-12 strain BW25113 as "wild-type strain". This strain has a well-defined pedigree and has not been subjected to mutagens (Datsenko & Wanner, 2000). We used the *cyaA*- strain and the

*crp*- strain of the Keio collection (Baba *et al.*, 2006). We removed the kanamycin resistance cassette of the *cyaA*- strain using flp recombinase in order to allow utilization of the p-sdh-gfp plasmid which also harbors a kanamycin resistance cassette (Zaslaver *et al.*, 2006). The *cyaA*- *crp*- double mutant was constructed by P1 transduction of the spectinomycin resistance cassette of another *crp*- mutant (Soutourina *et al.*, 1999).

Precultures were always performed in LB medium with the appropriate antibiotics (ampicillin 100 $\mu$g/ml; kanamycin 50 $\mu$g/ml; chloramphenicol 30 $\mu$g/ml; spectinomycin 10 $\mu$g/ml) at 37°C under shaking with the exception of the strain harboring the PR100 plasmid (with the thermosensitive lambda promoter) which has grown at 30°C to prevent too strong expression of *cyaA* resulting in insoluble adenylate cyclase (Reddy *et al.*, 1989).

The gltA-sdhC intergenic region was amplified by PCR using primers 5'-TATC**CTCGAG**TTAAGGT CTCCTTAGCGCCTT-3' (xho1 site in bold) and 5'- TTCT**GAATTC**GAATAACGCCCACATGCTG TTC-3' (EcoRI in bold) using BW25113 strain as template and inserted into *luxCDABE* plasmid backbone without promoter to make p-sdh-lux.

The *crp* gene was amplified by PCR using primers 5'-ATAA**GGTACC**ATGGTGCTTGGCAAACC GC-3'(kpnI site in bold) and 5'-TCCG**GGATCC**TTAACGAGTGCCGTAAAC-3'(bamH1 site in bold) using BW25113 strain as template. The PCR product, digested with kpn1 and bamH1, was inserted into the previously described p-tet-lux plasmid to make p-tet-*crp*.

The *cyaA* gene was amplified by PCR using primers The 5'-GCGA**GGTACC**TTGTACCTCTATATT GAGACT-3' (kpnI site in bold) and 5'-ACAGTC**GCTAGC**TTAACTAGCATCAGCATAATTTTCACT GTAGTTTTCGTCGTTTGCTGCCGAAAAATATTGCTGTAATAGC-3' (nheI site in bold) using BW 25113 strain as template. The PCR product, digested with kpn1 and nheI, was inserted into the previously described p-tet-lux plasmid to make p-tet-*cyaA*. Note that the oligo introduces a degradation tag das+4 in 3' (AANDENYSENYADAS). This tag doesn't play any role in this study but it may significantly reduce the half-life of the corresponding adenylate cyclase (McGinness *et al.*, 2006). The PR100 plasmid was a gift from Pr. Reddy.

## Solid media assay

We used a minimal medium with 0.3% (w/v) glucose and 0.8% (w/v) low melt agarose. After a washing step to remove antibiotics, we added, in the unsolidified medium and just before polymerization, the

receiver cells (dilution factor 1/10) or the sender cells (dilution factor 1/4) coming from LB pre-cultures. Interfaces were created step by step by pouring the unsolidified medium mix in the Petri dish (then sometimes, cutting to adjust final shape) and waiting for solidification. Then the next unsolidified medium mix was added so as to touch the previous pad to create the interface. After polymerization of the entire pad, the Petri dish was placed in a heated chamber and filmed during the appropriate time (days) by a sensitive CCD camera (photonic science). The thin strip of paper (whatman) in Figure 2 is saturated with 100 $\mu$l of a solution of anhydrotetracycline (1mg/ml).

## Liquid media assay

Dynamic analyses were performed by growing cells during 24 hours in a 96-well microplate. The microplate reader (fusion alpha FP-HT) allows growth with shaking and temperature control (37°C). It measure cell density (OD at 600nm), luminescence or fluorescence (480nm/520nm) every 15 minutes. Each well contains 180 $\mu$l of minimal medium (M9) with 0.3% glucose (w/v) and with a range of anhydrotetracycline (0, 100, 200, 500 ng/ml). We added 10 $\mu$l of appropriate receiver cells and 40 $\mu$l of appropriate sender cells after a washing step to remove the antibiotics of the preculture (LB). Contrary to luminescence data which don't produce background, we subtracted a background from fluorescence data: this background is the fluorescence generated by the control (0 ng/ml aTc). Sample for extracellular cAMP measurement were collected at 0, 420, 750 and 1200 minutes. We used the chemiluminescent cAMP HTS immunoassay (Millipore). All measures were performed in triplicate except one outlier removed (at time 1200, 500 ng/ml, Figure 2).

## Semi-solid media assay

The minimal medium (M9) contains 0.3% galactose and 0.3% low melt agarose. We added 4ml in a well of a 6 wells-micro plates. Once "semi-solidified", we were still able to sculpt the agar and we removed, by using a pipetman, one quarter of the semi-solid medium (0.8 ml) that we replaced by 0.8ml of the same medium but containing the sender cells (*crp-cyaA*- double mutant with the PR100 plasmid) (a pullet resuspended coming from the centrifugation of 2ml of preculture). Then we added, on the top of the agar, two spots (center and top left) of 1$\mu$l of receiver cells (concentrated 200X).

### Image processing

All movies and images were processed using Matlab software. No specific image enhancement has been performed.

## Acknowledgments

## References

Alon, Uri. 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/CRC Mathematical & Computational Biology)*. 1 edn. Chapman and Hall/CRC.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, **2**, 2006 0008. Gm62662/gm/nigms Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England.

Balagadde, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., Quake, S. R., & You, L. 2008. A synthetic Escherichia coli predator-prey ecosystem. *Mol Syst Biol*, **4**, 187.

Bankir, L., Ahloulay, M., Devreotes, P. N., & Parent, C. A. 2002. Extracellular cAMP inhibits proximal reabsorption: are plasma membrane cAMP receptors involved? *Am J Physiol Renal Physiol*, **282**(3), F376–92.

Bettenbrock, K., Fischer, S., Kremling, A., Jahreis, K., Sauter, T., & Gilles, E. D. 2006. A quantitative approach to catabolite repression in Escherichia coli. *J Biol Chem*, **281**(5), 2578–84.

Broadus, A. E., Kaminsky, N. I., Northcutt, R. C., Hardman, J. G., Sutherland, E. W., & Liddle, G. W. 1970. Effects of glucagon on adenosine 3',5'-monophosphate and guanosine 3',5'-monophosphate in human plasma and urine. *J Clin Invest*, **49**(12), 2237–45.

Datsenko, K. A., & Wanner, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**(12), 6640–5. Journal Article Research Support, U.S. Gov't, Non-P.H.S. United states.

Goldenbaum, P. E., & Hall, G. A. 1979. Transport of cyclic adenosine 3',5'-monophosphate across Escherichia coli vesicle membranes. *J Bacteriol*, **140**(2), 459–67.

Gorke, B., & Stulke, J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, **6**(8), 613–24.

Gosset, G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact*, **4**(1), 14. Journal article.

Hardiman, T., Lemuth, K., Keller, M. A., Reuss, M., & Siemann-Herzberg, M. 2007. Topology of the global regulatory network of carbon limitation in Escherichia coli. *J Biotechnol*.

Imamura, R., Yamanaka, K., Ogura, T., Hiraga, S., Fujita, N., Ishihama, A., & Niki, H. 1996. Identification of the cpdA gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in Escherichia coli. *J Biol Chem*, **271**(41), 25423–9.

Ishiyama, J. 1990. Isolation of mutants with improved production of cAMP from Microbacterium sp. no. 205 (ATCC 21376). *Appl Microbiol Biotechnol*, **34**(3), 359–63.

Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., & Karp, P. D. 2009. EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic Acids Res*, **37**(Database issue), D464–70.

Konijn, T. M., Van De Meene, J. G., Bonner, J. T., & Barkley, D. S. 1967. The acrasin activity of adenosine-3',5'-cyclic phosphate. *Proc Natl Acad Sci U S A*, **58**(3), 1152–4.

Kotte, O., Zaugg, J. B., & Heinemann, M. 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol Syst Biol*, **6**, 355.

Makman, R. S., & Sutherland, E. W. 1965. Adenosine 3',5'-Phosphate in Escherichia Coli. *J Biol Chem*, **240**, 1309–14.

Matin, A., & Matin, M. K. 1982. Cellular levels, excretion, and synthesis rates of cyclic AMP in Escherichia coli grown in continuous culture. *J Bacteriol*, **149**(3), 801–7.

McGinness, K. E., Baker, T. A., & Sauer, R. T. 2006. Engineering controllable protein degradation. *Mol Cell*, **22**(5), 701–7. AI-16892/AI/United States NIAID Journal Article Research Support, N.I.H., Extramural United States.

Nam, T. W., Park, Y. H., Jeong, H. J., Ryu, S., & Seok, Y. J. 2005. Glucose repression of the Escherichia coli sdhCDAB operon, revisited: regulation by the CRP*cAMP complex. *Nucleic Acids Res*, **33**(21), 6712–22.

Narang, A. 2009. Quantitative effect and regulatory function of cyclic adenosine 5'-phosphate in Escherichia coli. *J Biosci*, **34**(3), 445–63.

Nishio, Y., Usuda, Y., Matsui, K., & Kurata, H. 2008. Computer-aided rational design of the phosphotransferase system for enhanced glucose uptake in Escherichia coli. *Mol Syst Biol*, **4**, 160.

Park, Y. H., Lee, B. R., Seok, Y. J., & Peterkofsky, A. 2006. In vitro reconstitution of catabolite repression in Escherichia coli. *J Biol Chem*, **281**(10), 6448–54.

Potter, K., Chaloner-Larsson, G., & Yamazaki, H. 1974. Abnormally high rate of cyclic AMP excretion from an Escherichia coli mutant deficient in cyclic AMP receptor protein. *Biochem Biophys Res Commun*, **57**(2), 379–85.

Reddy, P., Peterkofsky, A., & McKenney, K. 1989. Hyperexpression and purification of Escherichia coli adenylate cyclase using a vector designed for expression of lethal gene products. *Nucleic Acids Res*, **17**(24), 10473–88. Journal Article England.

Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A., & Bertin, P. 1999. Multiple control of flagellum biosynthesis in Escherichia coli: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the flhDC master operon. *J Bacteriol*, **181**(24), 7500–8.

Tomchik, K. J., & Devreotes, P. N. 1981. Adenosine 3',5'-monophosphate waves in Dictyostelium discoideum: a demonstration by isotope dilution–fluorography. *Science*, **212**(4493), 443–6.

Van Dyk, T. K., Wei, Y., Hanafey, M. K., Dolan, M., Reeve, M. J., Rafalski, J. A., Rothman-Denes, L. B., & LaRossa, R. A. 2001. A genomic approach to gene fusion technology. *Proc Natl Acad Sci U S A*, **98**(5), 2555–60.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods*, **3**(8), 623–8.

# Supplementary information

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## 1 Calibration of the p-sdh-lux biosensor

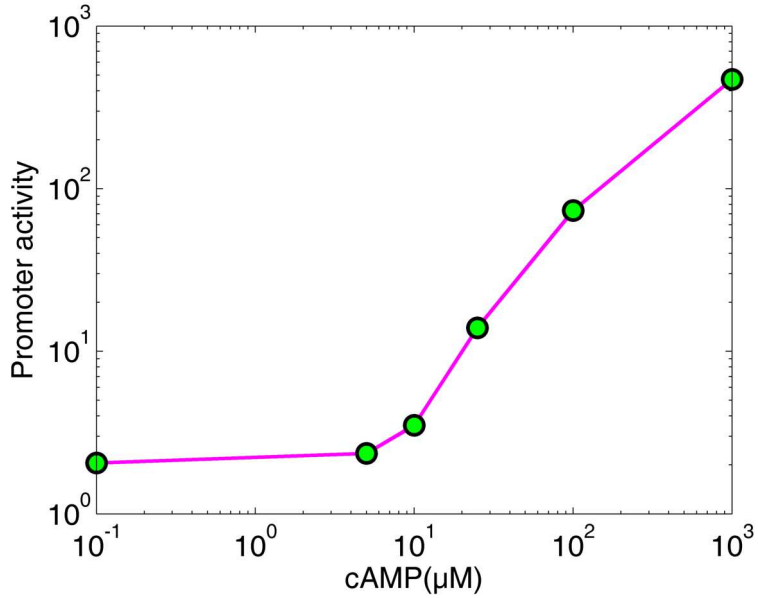



**Figure 1. Calibration of the biosensor**

We measured the luciferase activity of the p-sdh-lux plasmid during growth in a microplate reader (37°C with shaking). Each well contained 180 $\mu$l of minimal medium (M9) supplemented with glucose (0.3%) and the indicated concentration (0, 5, 10, 25, 100, 1000 $\mu$M) of cAMP. We added 20 $\mu$l of a LB pre-culture (*cyaA*- strain transformed with the p-sdh-lux luminescence biosensor) to each well at time zero. The promoter activity, computed from luminescence data, was calculated from the expression profile after the glucose exhaustion. The minimal detectable cAMP concentration of this biosensor, corresponding to a measurable difference in promoter activity, is around 5-10 $\mu M$.

## 2 Stochastic simulation of the communication

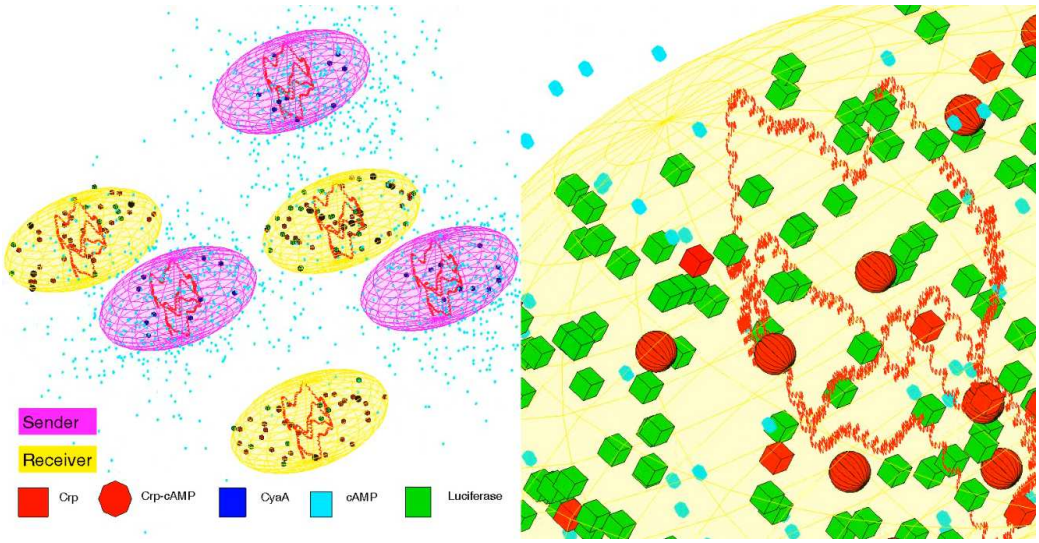



**Figure 2. Supplementary movie 2.**

We have implemented a stochastic, three-dimensional model to simulate the communication process. The code has been written in Matlab and is available upon request. Although this model reproduces the observed patterns, it should rather be considered as didactic. We do not derive any parameters from this simulation.

We added the video in flash format in this pdf document. Movie 2 shows the same simulation at a higher resolution. The screen of the video is divided in two parts. On the left, the video schematizes the different steps of the communication process. On the right, camera rotations and zooms highlight the tridimensional space and the events that we describe below. Each screen corresponds to a distinct simulation. Consequently, due to stochasticity in gene expression, the two simulations are not completely synchronous. At the beginning of the video, one can see three yellow ellipsoids that represent three *cyaA*-receiver cells. The three pink ellipsoids represent the *crp*- senders. All cells harbor a red, super-coiled plasmid. The sender cells produce adenylate cyclases (dark blue cube) that move randomly inside the cell and synthesize cAMP (light blue cube). Contrary to the proteins, cAMP progressively diffuses into the external medium, enters the receiver cells and binds to Crp proteins to form a cAMP-CRP complex. The molecular interactions, too computationally expensive, are not modeled. When cAMP molecules exceed a certain threshold in the receiver cell, there is a certain probability that CRP changes its conformation

(cube to sphere), increasing its affinity for binding site on the plasmid. Once bound to the DNA (watch the event on the right screen), the synthesis rate of luciferase increases, allowing transcription and traduction of *luxCDABE* and subsequent appearance of luciferase molecules (green cube) in the receiver cells. This event signals the success of the communication procedure. The receiver cells become green at different time due to stochasticity in gene expression. Obviously, those temporal differences can not be observed in our experimental data (presented in figure 2), obtained at the population level.

# 3 Investigation of the mouse gastro-intestinal ecosystem

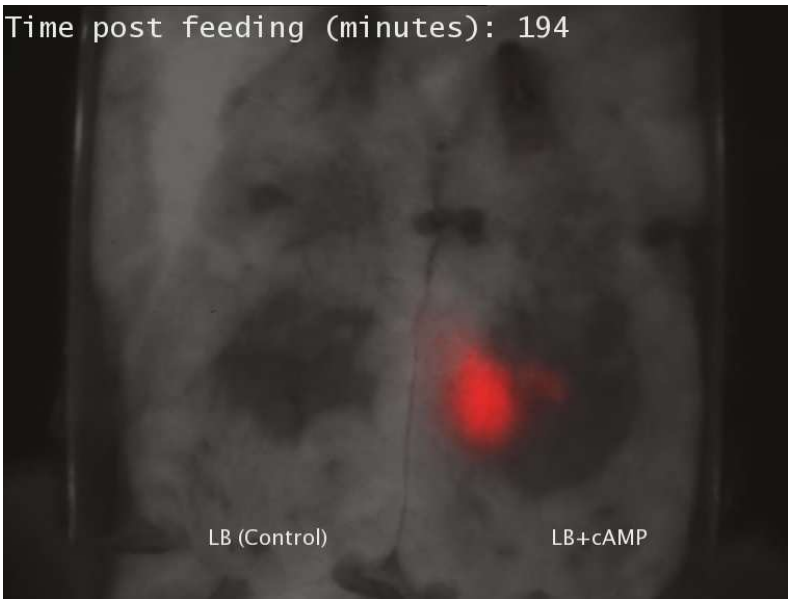



**Figure 3. Supplementary movie 4.**

We wanted to test whether communication by way of extracellular cAMP could occur between *E. coli* and other species in their natural environment: the gastro-intestinal tract. Unfortunately, we could not answer this particular question. The experiment shown here nevertheless demonstrates that external cAMP can induce gene expression in *E. coli* within the gastro-intestinal tract. It has previously been shown that the light produced by an *E. coli* population that has invaded the intestine of a mouse can be detected "transcutaneously" by a sensitive CCD camera. In our experiment, we administered orogastrically 200 $\mu$l of a solution containing approximately 25 billion *cyaA*- bacteria (containing the luminescence biosensor) to two mice. After 30 minutes, we administered orogastrically 250 $\mu$l of LB containing 100 mM cAMP to one mouse and 250 $\mu$l of LB to the second mouse (control). We then anesthetized the mice and placed them in a dark chamber under a sensitive camera to acquire luminescence images. Movie 4 shows that the expression of a gene (*luxCDABE*) is switched on in the gastro-intestinal tract of the mouse that has received cAMP. The mouse on the right, that drank the LB + cAMP solution, emits a strong luminescence detected transcutaneously and thus in a non invasive way. The mouse on the left, that drank LB as control, produces no light. We have reproduced this experiment twice.

## Materials and methods

Briefly, two Swiss CD-1 mice (Janvier, Le Genest-Saint-Isle, France) were given drinking water containing 500 $\mu$g/ml of kanamycin and ampicillin for 16 h to kill most commensal bacteria. The *cyaA* strain containing luminescence and fluorescence biosensors was grown in 100 ml of LB medium (with 100 $\mu$g/ml ampicillin and 50 $\mu$g/ml kanamcyin) overnight at 37°C under shaking. We resuspended the cells in 2 ml of LB and we administered orogastrically 200 $\mu$l of this solution with a feeding needle. 30 minutes later, we administered 250 $\mu$l of LB 100 mM cAMP to one mouse and 250 $\mu$l of LB to the second mouse (control). The mice were then anesthetized with a mixture of ketamine hydrochloride (Imalgen 500, Merial, 25 mg/kg body weight) and xylasine (Rompun, Bayer Healthcare, 12.5 mg/kg body weight), injected intra-peritoneally . The two mice were placed in a heated dark chamber and filmed for several hours by a sensitive CCD camera. Luminescence images were acquired every 30 seconds. We had to anesthetize the mice three times during the kinetics. This explains why the luminescence spot moves in the film just before the next anesthesia: a mouse begins to awake. The images of mice do not detect the awakening because there are static images taken at the beginning of each anesthesia. These images were subsequently overlaid onto the luminescence images (red false color). The movie was created using Matlab. No specific image enhancement has been performed.

## Ethical statement

Animals were housed according with the French Ethical Committee (Decree 87-848) and European Community Directive 86/609/EEC. Experiments were carried out under the supervision of CL (agreement nu 38 08 08) in the animal care facilities approved by the Direction des Services Vétérinaires de l'Isère (Nu A 38 516 10006).

# Chapitre V

## Using luciferase activity to infer the metabolic state of *Escherichia coli*

# Using luciferase activity to infer the metabolic state of *Escherichia coli*

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## Abstract

The terrestrial luciferase of *Photorhabdus luminescens*, encoded by the *luxCDABE* operon, is a classic reporter system used for monitoring promoter activity. In order to emit photons, the luciferase requires an energy source provided by a functioning central metabolism. When the cell lacks energy, there is an Abrupt Decline of Luciferase Activity (ADLA). In this manuscript we show that this dependence of luciferase activity on the availability of cellular energy sources can be used to infer metabolic states of the cell *in vivo* and in real time. The system can also be used to measure more specific parameters such as the relative activity of a specific enzyme, the speed of acetate consumption, the overflow metabolism or the level of catabolite repression. Combined with a library of mutants, the measurement of luciferase activity may also help to increase the accuracy of our metabolic maps or provided a tool for similar applications.

## Introduction

Characterization of the response of *Escherichia coli* cells to genetic and environmental perturbations is often performed *ex situ* using fast sampling techniques with or without a quenching step. Analyses of the transcriptome, proteome and/or metabolome are carried out *a posteriori*, after cell disruption, using DNA microarray or mass spectrometry (Ishii *et al.*, 2007). Those powerful, high-throughput technologies provide large amounts of data at the genome scale, but they do not give access to the behavior of the organism in real time.

Conversely, GFP and luciferase reporters allow measuring promoter activities in living cells (Hakkila

*et al.*, 2002; Van Dyk *et al.*, 2001) with a high temporal resolution (Zaslaver *et al.*, 2006). Contrary to GFP, which needs an external input (excitation light) to produce the output signal, bacterial luciferase emits photons without external dependencies. The ease and sensitivity of light detection, associated with the absence of background, has led to numerous applications using luciferase, such as enzymatic assays, biosensors and *in vivo* imaging (Waidmann *et al.*, 2011).

The terrestrial luciferase of *Photorhabdus luminescens* is part of a five genes operon (*luxCDABE*). The genes of the operon code for luciferase as well as the enzymes necessary for producing the substrates of the light generating reaction. The key substrate, a long-chain aldehyde, is produced within the cell by the reduction of fatty acid catalyzed by the gene products of *luxC, luxD,* and *luxE*. The hetero-dimeric luciferase, encoded by *luxA* and *luxB* genes, oxidizes this substrate and $FMNH_2$ in presence of oxygen, to emit a photon at 490 nm (Meighen, 1991). The reaction is as follow:

$$FMNH_2 + RCHO + 0_2 \rightarrow FMN + H_20 + RCOOH + h\nu \text{ (490 nm)}$$

The production of $FMNH_2$ requires an electron transport system, driven by the central reactions of carbon and energy metabolism (Sunya *et al.*, 2011). Consequently, light production does not only reflect the quantity of luciferase, but also the concentration of the metabolites participating in the light-generating reaction, and thus the activity of the metabolic reactions that produce these metabolites. For example, it has been shown that an Abrupt Decline of Luciferase Activity (ADLA) occurs upon entry in stationary phase of *Escherichia coli* cells due to a drop of the intracellular redox pool (Koga *et al.*, 2005).

In this study, we present several ways in which ADLA, and, more generally the variation of luciferase activity, can be exploited as a measure of the metabolic state of the cell. The variation of luciferase activity, combined with the appropriate genetic tools (mutants), allows to assess specific metabolic fluxes and thus indirectly monitor the activity of the corresponding enzymes.

# Results

## Detection and characterization of the abrupt decline of luciferase activity

We first wanted to show that ADLA can be used as a measure for the availability of a carbon or energy source. In order to verify this basic premise, we grew *Escherichia coli* expressing the *luxCDABE* operon from *Photorhabdus luminescens* on a minimal medium with glucose as a carbon and energy source. We

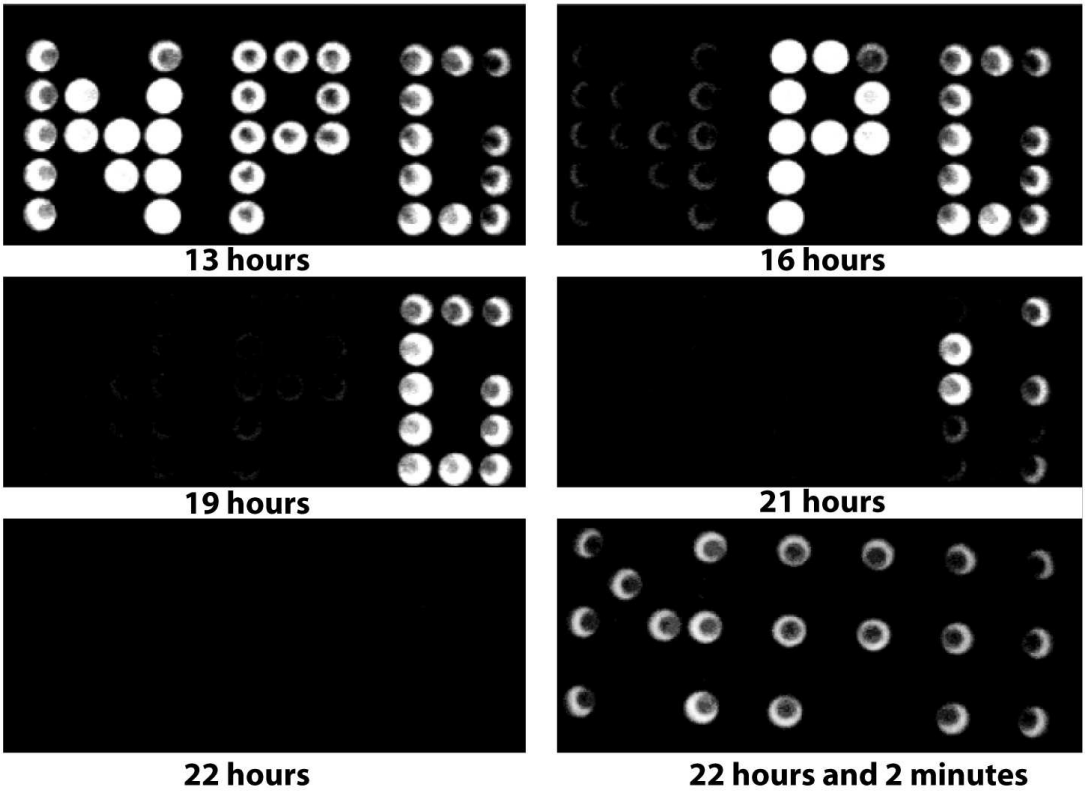



**Figure 1. The abrupt decline of luciferase activity depends on the energy supply of the cell.** Luminescence images extracted from supplementary movie (see details in supplementary information). Specific wells of a 96-well microplate (forming the letters N, P, G) contain a minimal glucose medium and *Escherichia coli* cells expressing the *luxCDABE* operon. Glucose concentration of each letter increases from left to right on the microplate (0.01% to 0.06%). Glucose consumption thus takes more time for cells further to the right on the plate. When the external energy source is exhausted, the energy production of cells drops and letters of this "bacterial billboard" switch off one after the other. Adding glucose to the medium (here in one well out of two) immediately turns on the luminescence signal again, demonstrating that luciferase enzymes are still present within the cells.

observed light production by the luciferase during the growth phase and an abrupt extinction of light emission when the energy source was exhausted. To illustrate the phenomenon, we created a bacterial billboard by growing the bacteria in a 96-well microplate with different amounts of glucose in the different wells (Figure 1 and supplementary information: movie). The concentration of glucose increases from left to right in the microplate. As soon as energy production stops (the duration of growth is a function of the glucose concentration), luciferase activity drops sharply. The luciferase enzymes are still present after the cessation of light production, since adding glucose after the ADLA allows immediate recovery of the original luminescence signal.

## Measuring the metabolic state of cells

Not surprisingly, this experiments shows that light emission is completely shut off when cellular energy is exhausted. Could we obtain a more graded response and use the intensity of light emission as a measure of the intracellular energy level ? In order to separate signals controlling light production, i.e., the regulation of *luxCDABE* transcription, and the energy requirement of luciferase activity, we use the synthetic *tet* promoter to drive the expression of the *luxCDABE* operon. The activity of this promoter is repressed by the TetR repressor, and repression is lifted by adding anhydrotetracycline (aTc) to the medium. We grew a wild-type strain containing this construction on a minimal medium containing glucose, acetate and the inducer (aTc). The temperature was maintained at 37°C and we measured the absorbance and luminescence of the culture grown directly in a microplate reader. The presence of the inducer from the beginning of the kinetics directly triggers the transcription of *luxCDABE* (Figure 2A). When glucose is exhausted (break on the red curve), there is a sharp decrease immediately followed by a rapid recovery of the luminescence signal. This event marks the transient fall of energy during the glucose-acetate transition. When acetate is exhausted, there is no remaining energy source in the medium and we observe an abrupt decline of luciferase activity (ADLA). Thus, measuring luciferase activity driven by a constitutive promoter during the growth of the bacterial culture can be used to detect, in real time, small and large variations of energy supply.

Since the luminescence profile gives an indication of energy availability, we can use this signal to assess the effect of different mutants on the energy metabolism of *E. coli*. We therefore transformed hundreds of single-gene knock-out mutants (Baba *et al.*, 2006) with the biosensor plasmid and measured their corresponding luminescence profile in the same conditions as in Figure 2A. Most of the mutants

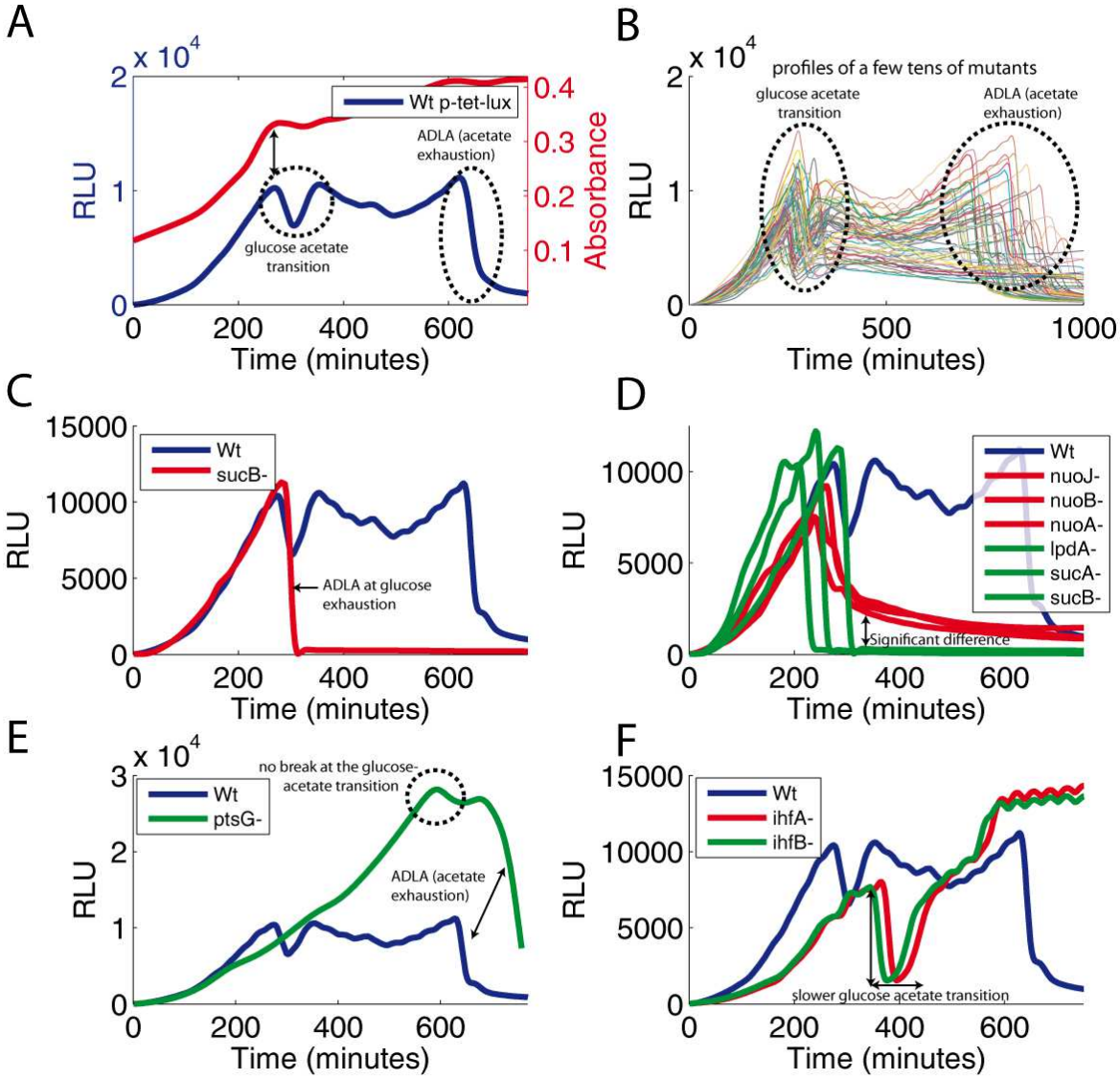



**Figure 2. Luciferase activity during growth on glucose and acetate in different genetic backgrounds.** (A) An *E. coli* wild-type strain, transformed with the *p-tet-lux* plasmid, grows on a minimal medium with glucose (0.03%), acetate (0.03%) and aTc (500ng/ml). When glucose is exhausted, there is a sharp decrease followed by a rapid recovery of luciferase activity. This occurs exactly at the time when growth slows, i.e., when glucose is exhausted (arrow). The Abrupt Decline of Luciferase Activity (ADLA) signals acetate exhaustion. (B) Example of few dozens of mutants showing a profile similar to the one of the wild-type strain. Difference in the time delay of ADLA may therefore reflect difference in the speed of acetate consumption. (C) The profile of a *sucB* mutant shows an ADLA when glucose is exhausted because this strain can not utilize acetate as an energy source. (D) During growth on acetate, *nuo* mutants maintain an intermediate level of luciferase activity. (E) In a *ptsG* mutant, there is no transient decrease of light emission at the point of glucose exhaustion. (F) The more pronounced dip in luciferase activity in *ihfA* and *ihfB* profiles may indicate a slower glucose-acetate transition.

have a luminescence profile similar to that of the wild-type strain (Figure 2B).

However, a subset of the mutants (*sucB, sucA, lpdA, atpC, atpD*) fails to maintain luciferase activity after glucose exhaustion (Figure 2C; exemplified by the *sucB* mutant). In fact, the oxoglutarate dehydrogenase complex and ATP synthetase are needed for growth on acetate. In these mutants, the impossibility to use acetate as an energy source after glucose exhaustion directly leads to the abrupt decline of luciferase activity. For mutants that are impaired in using acetate as an energy source, we observe intermediate phenotypes. The *nuo* operon shown in figure 2D is an example. We interpret the different levels of light production by these mutants as a measure of metabolic flux. The small differences of the growth phenotype would be impossible to detect and a full blown metabolic analysis would be very time consuming and costly. Consequently, we propose to use the capability to maintain luciferase activity as a very sensitive measure of the metabolic state of the cell.

Certain mutants show interesting phenotypes, but are more difficult to interpret. The *ptsG* mutant lacking the main glucose transporter takes much more time to consume glucose, as expected. However, we did not observe the transient decrease of the luminescence signal when glucose is exhausted (Figure 2E; locate the circle). The glucose-acetate transition may occur without transient drop of energy or the cells may consume glucose and acetate at the same time during growth. On the contrary, the transient drop of light emission at the glucose-acetate transition is much more pronounced in the *ihfA* and *ihfB* mutants, compared to the wild-type strain (Figure 2F). This phenotype may reflect a slower glucose-acetate transition in the IHF deficient strain. IHF is a transcriptional activator of the *aceBAK* operon, which codes for enzymes of the glyoxylate by-pass (Resnik *et al.*, 1996), the crucial pathway for gluconeogenesis and thus for acetate utilization. The absence of IHF may slow the establishment of this pathway, an event reported by the almost complete, transient loss of luciferase activity after glucose exhaustion.

## Tuning and measuring acetyl CoA synthetase activity *in vivo*

The transient decrease in luciferase activity after glucose exhaustion probably represents the time needed by the cell to switch from glucose metabolism to the utilization of acetate, the only remaining carbon source. The key enzyme for acetate utilization is acetyl-CoA synthetase, coded by the *acs* gene. This enzyme converts acetate to acetyl coenzyme A, which is the first enzymatic reaction allowing acetate utilization. By controlling the expression of *acs*, we should therefore be able to control the luciferase

activity of a "metabolic biosensor".

*acs* transcription is activated by the Crp-cAMP complex after glucose exhaustion and there is no promoter activity at all in a mutant lacking adenylate cyclase, the enzyme producing cAMP coded by the *cyaA* gene. However, this lack of cAMP can be complemented by adding exogenous cAMP to the growth medium. Different concentrations of added cAMP activate the *acs* promoter to different extents. If our hypothesis is correct, this should lead to different levels of acetyl-CoA synthetase leading to different levels of energy in the cell, which could be detected by the appropriate biosensor. Figure 3A shows the expression of *acs* in a *cyaA* strain, detected by a fusion of luciferase to the *acs* promoter. The transcriptional response is clearly dependent on the concentration of cAMP in the growth medium. The promoter is shut-off during growth on glucose. The *p-acs-lux* fusion thus reports essentially *acs* transcription and not energy availability of the cell.

In order to measure luciferase activity as a function of energy supply, we used the *aceBAK* promoter, again fused to the luciferase operon (*p-aceBAK-lux*). The very strong *aceBAK* promoter leads to a very high luciferase concentration. At those levels of luciferase activity, variation of light emission mainly reports changes in the energy supply rather than the regulation of the *aceBAK* promoter. Consequently, in our experiment, the *aceBAK* biosensor mainly reports the acetyl-coenzyme A synthetase activity, i.e., the rate of acetate metabolism.

We grew the *cyaA* cells as before on minimal medium in the presence of glucose and acetate, but without exogenous cAMP. The *aceBAK* promoter is highly transcribed during growth on glucose (Figure 3B, times before the red arrow). Luciferase activity drops sharply when glucose is exhausted (red arrow). In this state, the *cyaA* mutants no longer grow because acetyl-CoA synthetase is not produced. We kept the cells in this state for about an hour, then added different concentrations of cAMP to the medium. This addition triggers the activation of *acs* transcription (Figure 3A). The arrival of acetyl coenzyme A synthetase enzymes triggers utilization of acetate as a carbon source and thus allows the recovery of luciferase activity of the *aceBAK* biosensor (Figure 3B). The luciferase activity reflects energy production and thus the rate of acetate consumption which is function of the concentration of acetyl coenzyme A synthetase. Indeed, the luciferase activity of the *aceBAK* biosensor during growth on acetate (Figure 3B) correlates well with the activity of the *acs* promoter (Figure 3A). Furthermore, faster acetate utilization (at higher cAMP concentrations) also leads to an earlier onset of ADLA (Figure 3B, blue arrows), because the fixed amount of acetate in the medium is consumed more rapidly.

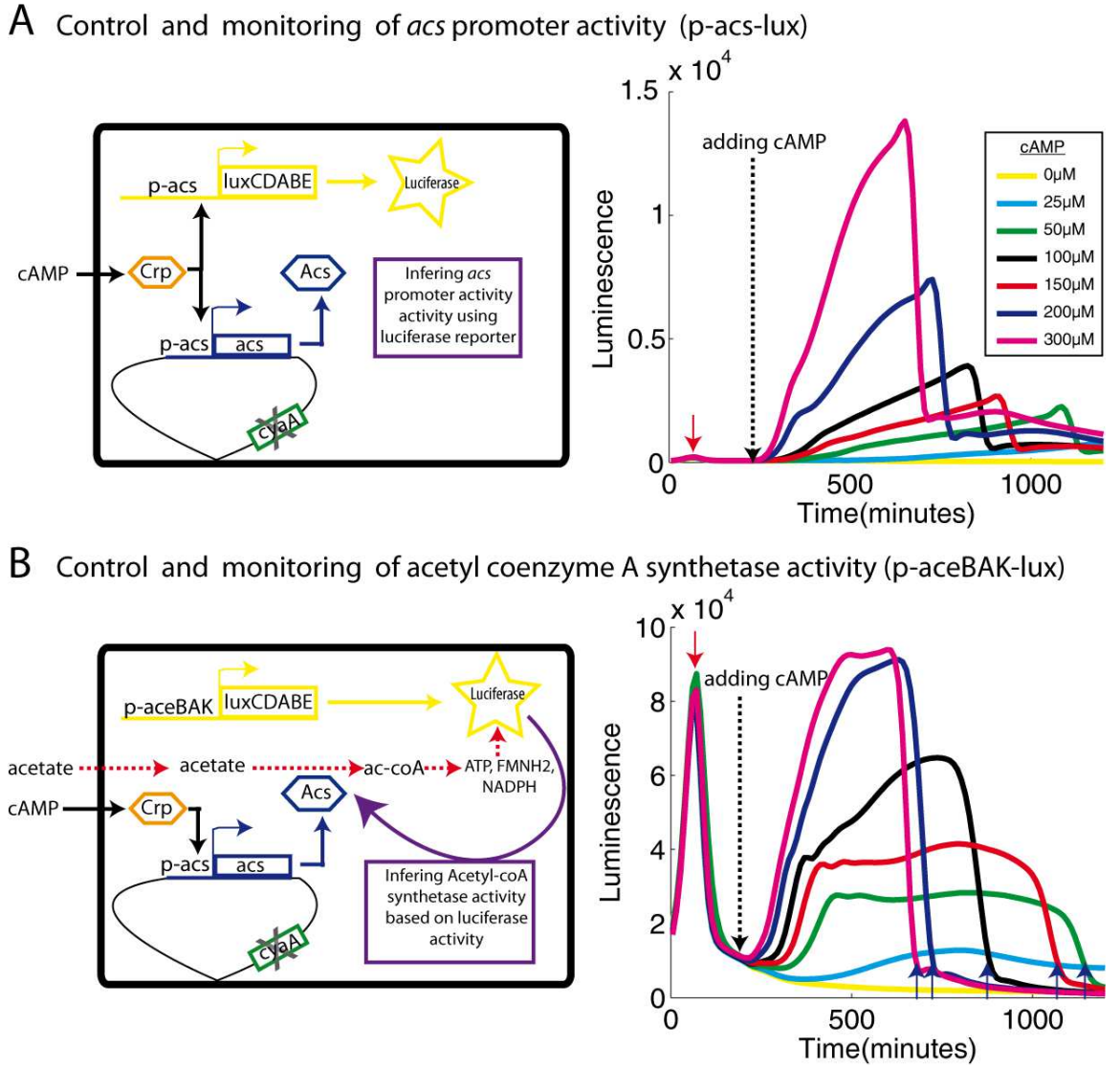



**Figure 3. Inferring the activity of a specific enzyme by measuring luciferase activity.** The *cyaA* strains were grown on a minimal medium supplemented with glucose (0.015%) and acetate (0.03%). We added different concentrations of cAMP to the medium at 200 minutes. (A) The *cyaA* mutant transformed with the *p-acs-lux* plasmid. Luciferase activity mainly reports the transcription of the *acs* gene, which is activated by the added exogenous cAMP in a dose dependent manner. (B) The *cyaA* mutant transformed with the *p-aceBAK-lux* plasmid. The light emission during growth on glucose reaches a high level (red arrow). After glucose exhaustion, there is an abrupt decline of luciferase activity since the *cyaA* mutant can not grow on acetate. Adding cAMP allows *acs* transcription, production of acetyl coenzyme A synthetase, and recovery of luciferase activity in a dose dependent manner. The ADLA occurs more quickly when the acetyl coenzyme A synthetase concentration is high because acetate is consumed more rapidly (blue arrows).

To conclude, in this experiment, we monitored, *in vivo* and in real time, the relative level of acetyl coenzyme A synthetase and thus the rate of acetate consumption by using the information provided by the recovery of luciferase activity. We checked our interpretation by using a *cyaA acs* double mutant that fails to recover luciferase activity despite the presence of exogenous cAMP, as expected (see supplementary information).

## Correlation between ADLA and the end of glucose repression

We have shown in a previous work that carbon catabolite repression (CCR) senses the metabolic state of the cell, probably energy availability or glycolytic flux. Using our biosensor that responds to energy availability, we can assess the connection between this parameter and the expresssion of CCR-controlled promoters.

To monitor CCR, we use the *araBAD* promoter. This promoter is activated both by the AraC-arabinose complex and by the Crp-cAMP complex. Since we wanted to study only CCR, we added the inducer (arabinose) to all media in these experiments. As expected, in these conditions, we observe that the expression of *araBAD* is strongly induced when glucose is exhausted.

To assess energy availability and CCR in the same cell, we transformed a *sucB* mutant with two biosensors (Figure 4A): the previously described p-tet-lux plasmid that reports the energy level of the cells (luminescence), and the p-araBAD-gfp plasmid, that measures the activity of the *araBAD* promoter (fluorescence). We grew the strain in a minimal medium containing glucose and the inducers (arabinose and aTc). As expected for a *sucB* mutant, we observe an abrupt decline of luciferase activity once glucose exhaustion (Figure 4B). Simultaneously with ADLA (luminescence), we observe the induction of the *araBAD* promoter (fluorescence) (Figure 4B vs 4C; black dashed arrow). The drop of luciferase activity indicates the decrease in energy availability and this event seems to trigger the transcription of cAMP-CRP controlled promoters. We confirmed this result with another Crp-cAMP dependent promoter (*acs* promoter; data not shown). When all energy sources are used up, no residual luciferase activity remains (Figure 4BC, red dashed arrow) and the transcription of *gfp* also stops. All cellular functions seem to be shut down at this point.

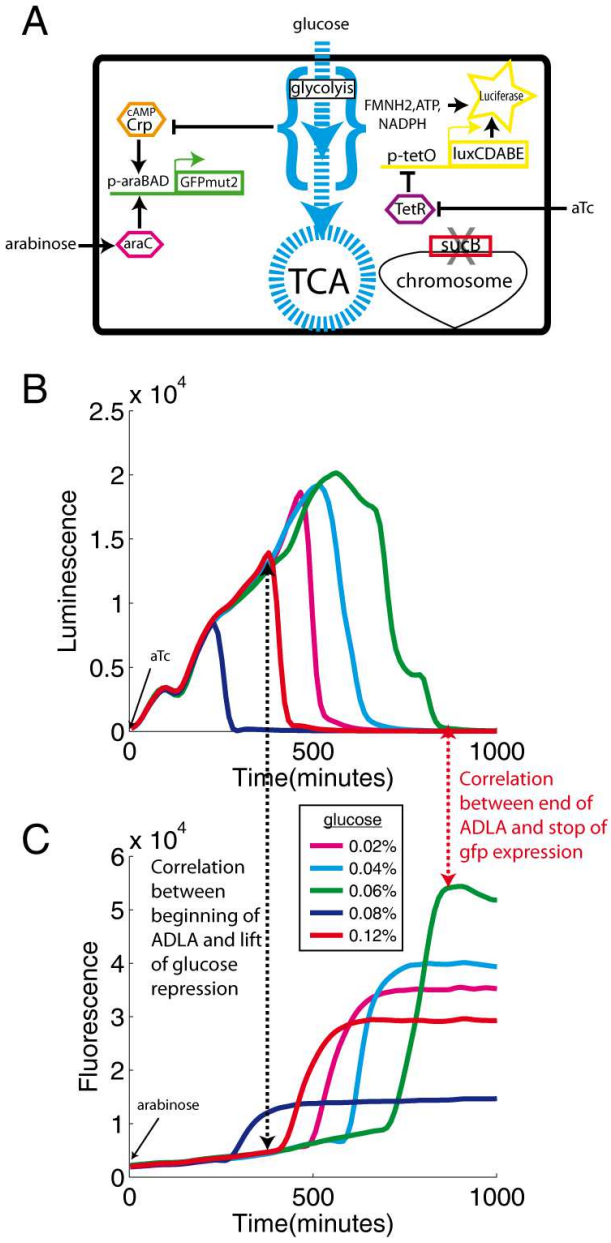



**Figure 4. Measuring carbon catabolite repression.** (A) A *sucB* mutant was transformed with two plasmids. The first plasmid, *p-araBAD-gfp*, reports *araBAD* promoter activity (fluorescence, C) and the second plasmid, *p-tet-lux*, monitors the energy level of the cell (luminescence, B). The strain grew on a minimal medium supplemented with glucose, arabinose (0.005%), and aTc (500ng/ml). (B) Profile of luminescence. ADLA occurs once glucose is exhausted (in a dose-dependent manner) (C) Induction of *araBAD* occurs simultaneously with the beginning of the ADLA (C vs B). The arrest of gfp expression (fluorescence) occurs simultaneously with the end of the ADLA.

### Quantifying the overflow metabolism

In addition to measuring intracellular physiological states, we can use the reporter system also to assess the media composition. As described above, *acs* expression is strongly induced when glucose is exhausted. The *p-acs-lux* plasmid can therefore be used to detect glucose exhaustion. By varying the concentration of glucose, in the presence of a fixed acetate concentration, we observe a strong increase of luminescence correlated with the glucose concentration (Figure 5A). Conversely, by fixing the glucose concentration and varying the acetate concentration, we can control the time of the abrupt decline of luciferase activity that marks acetate exhaustion (Figure 5B). Combining these two phenotypes, we can use the *p-acs-lux* plasmid to detect both glucose and acetate exhaustion in real-time.

The transition from growth on glucose to growth on acetate is a natural phenomenon for *Escherichia coli*. When these bacteria grow rapidly on glucose, they excrete acetate in the culture medium (Wolfe, 2005). This phenomenon have been called "overflow metabolism". Once glucose is exhausted, bacteria import the secreted acetate and used it as a carbon source. Using our reporter plasmid, we were able to quantify the overflow metabolism. We grew a wild type strain carrying the *p-acs-lux* plasmid in varying concentrations of glucose, but without adding acetate to the medium. With increasing glucose concentration, we observe a proportional increase of the delay between *acs* induction (reporting glucose exhaustion) and ADLA (reporting acetate exhaustion) (Figure 5C). We believe that this delay is indicative of the amount of acetate secreted due to overflow metabolism.

An easy way to test this hypothesis consists in testing mutants known to limit overflow metabolism and acetate secretion due to low glycolytic flux (Lin *et al.*, 2005). Compared to a wild-type strain, both the *ptsG* mutant (that lacks the main glucose transporter) and the *pgi* mutant (that interrupts glycolysis) show a delayed *acs* induction. This is expected for strains that consume glucose more slowly (Figure 5D). However, for the same glucose concentration, the delay between *acs* induction and the ADLA is much shorter in these two strains compared to the wild type strain. Our *p-acs-lux* biosensor is therefore a simple tool for estimating acetate secretion and overflow metabolism.

## Discussion

In this study, we have shown that bacterial luciferase can be used, both as a reporter of promoter activity and of the metabolic state of the cell. When there is excess energy, the reporter system measures

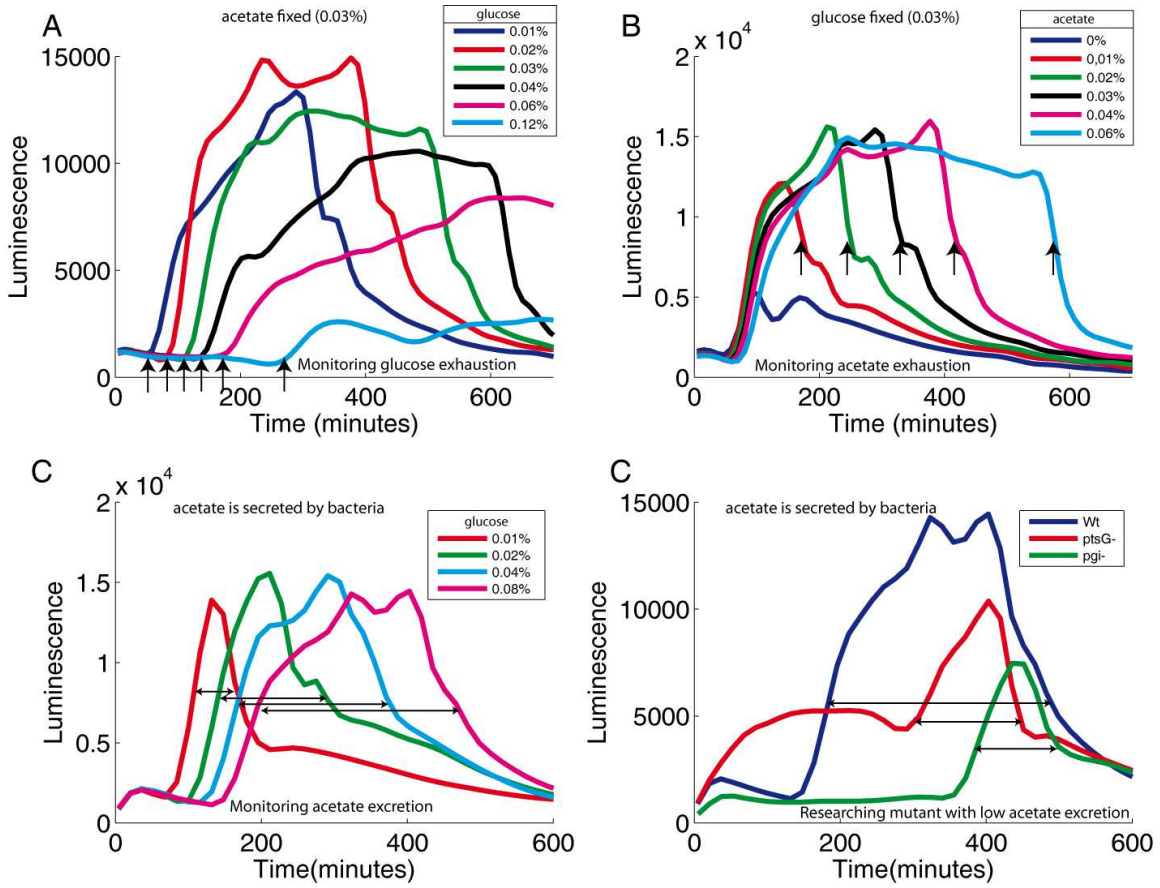



**Figure 5. Measuring glucose, acetate, and overflow metabolism.** (A) The *p-acs-lux* plasmid can be used to monitor glucose exhaustion (acetate concentration is fixed). A strong increase of luminescence occurs upon glucose exhaustion (in a dose dependent manner). (B) The *p-acs-lux* plasmid can also be used to monitor acetate exhaustion (glucose concentration is fixed). The time of Abrupt Decline of Luciferase Activity (ADLA) is proportional to the acetate concentration. (C). Bacteria were grown on a minimal medium with glucose, but without acetate. When the glucose concentration increases, the delay between *acs* induction (glucose exhaustion) and ADLA (acetate exhaustion) increases proportionally. This delay indicates the amount of secreted acetate during growth on glucose due to overflow metabolism. (D) For the same amount of glucose (0.08%), mutants with a reduced glycolytic flux (*ptsG, pgi*) show lower acetate secretion compare to the wild-type strain.

transcription, when cellular energy is limiting, the luciferase activity is determined by the metabolic state of the cell and the reporter system measures metabolic activity. The schematic of figure 6A and the following paragraphs summarize the main results.

## Global information

The main parameter to consider in order to measure the global metabolic activity of the cell is the ratio luciferase-activity / luciferase (ratio LA/L), especially when luciferase concentration (L) is high. When the activity of metabolism, or cellular energy, drops, this ratio drops. The accuracy of light detection, and its large dynamical range, allow detecting small variations of energy supply.

The ratio LA/L also allows to estimate the functioning of the global expression machinery of the cell in a rather obvious way: when no energy is available at all, the luciferase activity, and therefore the ratio LA/L, drops to zero. In synthetic biology, regulatory networks are sometimes considered independently of the activity of metabolism. Yet, variations of the metabolic activity have a profound effect on gene expression and the regulation of gene expression. The LA/L ratio may provide a useful parameter to assess the capacity of the cell to perform the function of a designed, or natural, genetic regulatory network. Last, at intermediate levels of the LA/L ratio, we can assess the activity of certain Crp-cAMP dependent promoters.

## Specific information

Depending on the experimental condition, the LA/L can also be used to obtain more specific information. In this study, we have used this parameter to quantify the activity of acetyl coenzyme A synthetase (and the corresponding flux of acetate utilization). By controlling the expression of acetyl coenzyme synthetase, our *aceBAK* biosensor correctly indicated the rate of acetate utilization and thus the degree of energy availability of the cell. This method of detection/relative quantification of enzyme activity is not limited to acetyl coenzyme A synthetase. The activity of any enzyme could be quantified by setting the conditions such that this enzyme constitutes the bottleneck for luciferase activity. For instance, controlling *lacZ* expression in cells growing on a minimal medium with lactose as the only carbon source will limit the rate of metabolism and a luciferase-based biosensor of metabolic activity would thus indirectly measure $\beta$-galactosidase activity.

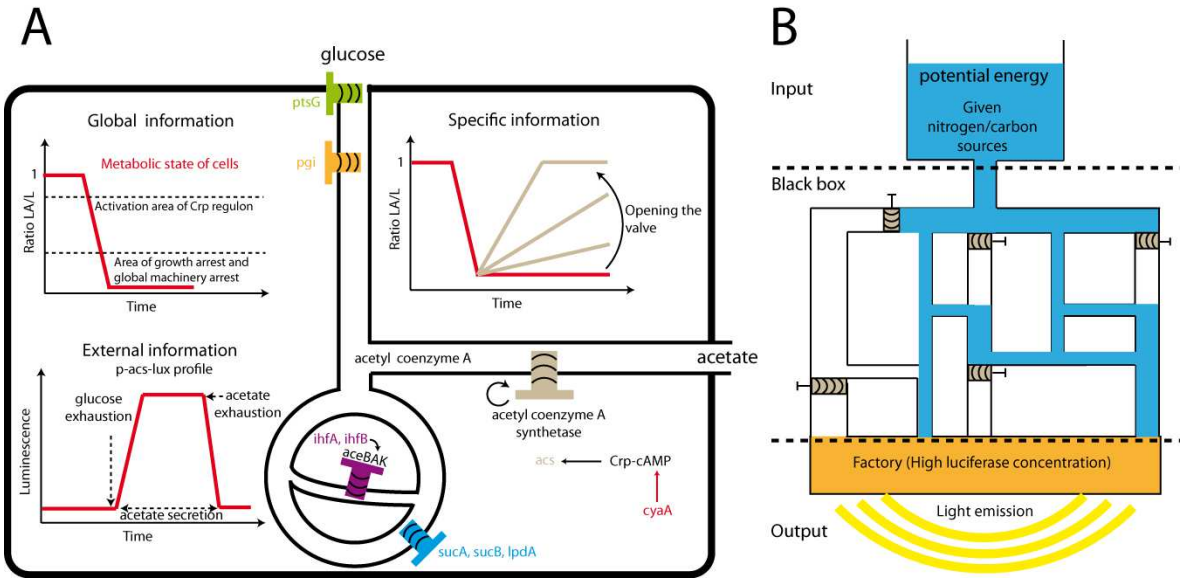



**Figure 6. Schematic of the biosensors.** (A) Representation of a cell. Metabolism (glycolysis and TCA) is represented by pipes, the opening of which can be controlled by valves (mutants). Valves that are shown on the pipes have been used in this work: *ptsG, pgi, ihfA, ihfB, sucA, sucB, lpdA*. The valve "acetyl-coenzyme A synthetase" is more complicated: the degree of opening can be adjusted by the level of cAMP, since cAMP-CRP controls *acs* transcription. The three schematics placed inside the cell show the different kinds of information that can be assessed using luminescence reporters. Specific information (top right) can be obtained by measuring the ratio "luciferase activity/luciferase" (LA/L) in controlled conditions (adjustment of the valve opening). Global information (top left) can also be obtained: a very low LA/L ratio indicates a global shut-down of cellular functions. Intermediate LA/L ratios indicate a regime where targets of the Crp regulon are activated. External information (bottom left) can be inferred using the *p-acs-lux* transcriptional fusion: glucose exhaustion is detected by the induction of *acs* expression. Acetate exhaustion is detected by ADLA. The combination of these two pieces of information allows determining the amount of secreted acetate. (B) The energy is seen as a reservoir flowing in a network of pipes and operating an electric power station (luciferase), placed at the bottom and producing light. To determine the network topology and the flow, it is possible to adjust valves in the network (mutation) while watching the consequences on light production.

## Information about media composition

In this study, we also present a biosensor, *p-acs-lux*, that responds to glucose exhaustion (via a transcriptional activation) and on acetate exhaustion (via the drop of metabolic activity). Taken together, those two signals can be used to measure acetate secretion during overflow metabolism. This biosensor could thus be useful, for example, in a screening procedure that aims at identifying mutants with low acetate secretion. The accumulation of acetate in the culture medium is an important problem in industrial fermentations since this organic acid inhibits cell growth and recombinant protein production. Different approaches are currently developed to reduce acetate accumulation by modifying central metabolic pathways (Gosset, 2005).

## Redrawing metabolic flux

By combining the complete mutant library of *E. coli* with the current knowledge of metabolism, the ratio LA/L may help to re-draw, with a high resolution, the map of metabolic flux for a given medium. Indeed luciferase activity is particularly well-suited for detecting small metabolic leaks. We propose an analogy with a Dam (Figure 6B). Potential energy (water) is stocked as carbon or nitrogen sources. The water flow passes through a complex circuit with valves placed on each channel. At the downstream end, a factory (high luciferase concentration) transforms the potential energy in light, provided sufficient flux arrives at the factory. To reconstruct an unknown circuit of channels, the experimenter would close certain valves (knock out mutant) or enlarges other (over expression) and analyses the effect on light production. A serious limitation of this approach is the fact that luciferase activity is affected by "global energy availability" : ATP, NADPH, NADH, $FMNH_2$. A mutant affecting specifically the concentration of one of those metabolites without affecting the other could generate artifacts: the luminescence signal may drop even though the cell still has a lot of energy. However, among the dozens of mutants involved in energy production that we have tested, we did not find a single one (including *nuo*, *atp*, *fre*, *suc* mutants) growing normally without being able to maintain luciferase activity. In this sense, luciferase activity seems to be a robust, global measure of energy availability in the cell.

### Conclusion

In this study we have shown that reporter systems based on bacterial luciferase may be used for inferring different parameters about the functioning of the cell. The ease and sensitivity of light detection, associated with its small cost, make our method/tool easy to implement/to use. The biosensors can potentially be used in many applications in biotechnology, system and synthetic biology.

## Materials and Methods

### Strains

The mutant strains come from the Keio collection (Baba *et al.*, 2006). The wild type strain (*E. coli* K-12 strain BW25113) has a well-defined pedigree and has not been subjected to mutagens (Datsenko & Wanner, 2000). We removed the kanamycin cassette of the *sucB* mutant using the flp recombinase to allow antibiotics compatibility. The *cyaA*- *acs*- double mutant was constructed by removing the kanamycin resistance cassette of the *cyaA*- strain. Then, we transduced (P1 phage transduction) the kanamycin resistance cassette of the *acs*- strain in the *cyaA*- strain. High-throughput transformation of hundreds of mutants was performed by using a previously described protocol (results not yet published).

### Plasmids

We have used the p-acs-lux and p-tet-lux plasmids whose construction has already been described (results not yet published). The p-araBAD-gfp comes from the Uri Alon collection (Zaslaver *et al.*, 2006). The metA-aceB intergenic region was amplified by PCR using primers 5'-GCTGCTCGAGTCTTCTGTGAT AGTCGATC-3' (xho1 site underlined) and 5'-GTTCGAATTCCGTGCAGCTCCTCGTCAT-3' (EcoRI site underlined) using BW25113 strain as template. The PCR product, digested with Xho1 and EcoRI, was inserted into a *luxCDABE* plasmid backbone without promoter to make p-aceB-lux.

### Measurement of light emission

Images of the figure 1 have been obtained by filming a microplate placed in a heated chamber using a sensitive CCD camera (photonic science). Histogram equalization has been performed to remove the luminescence background.

Dynamic measures of luminescence and/or fluorescence were performed by growing cells during several hours within a microplate reader (fusion alpha FP-HT). This micro plate reader allows growth with shaking and temperature control (37°C). It measure cell density (OD at 600nm), luminescence and/or fluorescence (485nm/520nm) at a high temporal resolution ( 5 minutes). Wells of the microplate contain 180$\mu$l of medium and 20$\mu$l (dilution 10) of cells in stationary phase. Cells in stationary phase (Figure 1, 2, 5) come directly from a LB pre-culture. Cells in stationary phase (Figure 3, 4) come from a LB pre-culture which have been centrifuged and the pellet have been re-suspended in a equivalent volume of clean M9 medium. The main parameters that differ between experiments are the strains, the concentrations of inducers (cAMP, arabinose and aTc) and the concentrations of the carbon sources (glucose and acetate). They are indicated in the legend of the figures.

## Acknowledgments

## References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, **2**, 2006 0008.

Datsenko, K. A., & Wanner, B. L. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, **97**(12), 6640–5.

Gosset, G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact*, **4**(1), 14.

Hakkila, K., Maksimow, M., Karp, M., & Virta, M. 2002. Reporter genes lucFF, luxCDABE, gfp, and dsred have different characteristics in whole-cell bacterial sensors. *Anal Biochem*, **301**(2), 235–42.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., & Tomita, M. 2007. Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. *Science*.

Koga, K., Harada, T., Shimizu, H., & Tanaka, K. 2005. Bacterial luciferase activity and the intracellular redox pool in Escherichia coli. *Mol Genet Genomics*, **274**(2), 180–8.

Lin, H., Bennett, G. N., & San, K. Y. 2005. Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield. *Metab Eng*, **7**(2), 116–27.

Meighen, E. A. 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev*, **55**(1), 123–42.

Resnik, E., Pan, B., Ramani, N., Freundlich, M., & LaPorte, D. C. 1996. Integration host factor amplifies the induction of the aceBAK operon of Escherichia coli by relieving IclR repression. *J Bacteriol*, **178**(9), 2715–7.

Sunya, S., Gorret, N., Delvigne, F., Uribelarrea, J. L., & Molina-Jouve, C. 2011. Real-time monitoring of metabolic shift and transcriptional induction of yciG::luxCDABE E. coli reporter strain to a glucose pulse of different concentrations. *J Biotechnol*.

Van Dyk, T. K., Wei, Y., Hanafey, M. K., Dolan, M., Reeve, M. J., Rafalski, J. A., Rothman-Denes, L. B., & LaRossa, R. A. 2001. A genomic approach to gene fusion technology. *Proc Natl Acad Sci U S A*, **98**(5), 2555–60.

Waidmann, M. S., Bleichrodt, F. S., Laslo, T., & Riedel, C. U. 2011. Bacterial luciferase reporters: The Swiss army knife of molecular biology. *Bioeng Bugs*, **2**(1), 8–16.

Wolfe, A. J. 2005. The acetate switch. *Microbiol Mol Biol Rev*, **69**(1), 12–50.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods*, **3**(8), 623–8.

# Supplementary information

Author1[1], Author2[2], Author3[3,*]

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

∗ E-mail: Corresponding author@institute.edu

## Video of ADLA

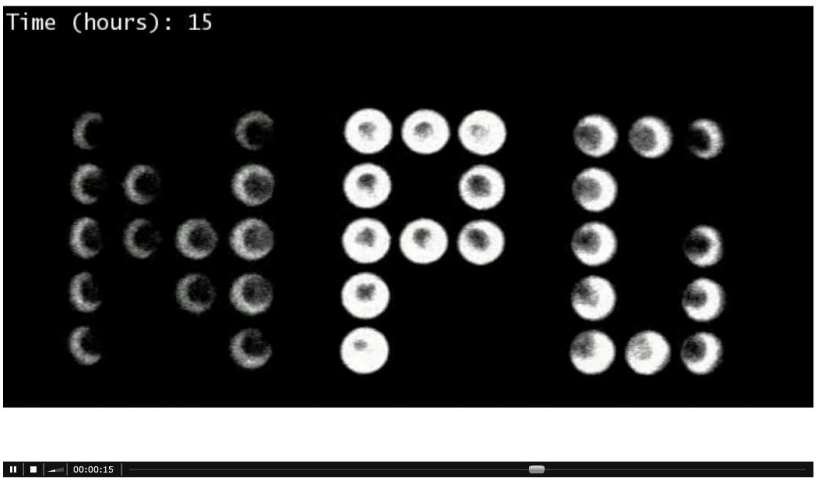



**Figure 1. Supplementary movie.**

Video from which were extracted the images of Figure 1. This bacterial billboard is obtained by filming for 25 hours the luminescence emitted by a 96-well microplate. The results obtained from this visualization of the cellular energy metabolism are somewhat more complete and complex than the description given in the results section. This video shows two different phenomena:

- carbon catabolite repression (CCR).
- Abrupt Decline of Luciferase Activity (ADLA).

The letters N, P, G were formed by appropriately grouping the wells of a 96-microplate. The wells forming the letters contained a minimal medium supplemented with glucose and acetate (0.03% w/v). The wells also contained the wild-type *E. coli* strain transformed with a reporter plasmid carrying the luciferase

operon (*luxCDABE*) under control of the *acs* promoter. The glucose concentration increases from left to right on the microplate (0.01% w/v to 0.06% w/v). Cells further to the right therefore grow for a longer time before exhausting glucose. When glucose is exhausted, the global regulator CRP turns on the transcription of the *luxCDABE* operon and letters switch on, one after the others. This effect is called catabolite repression.

At this stage, bacteria use acetate as a carbon source. When this nutrient is also exhausted, there is no remaining carbon source in the medium and the letters of the "bacterial billboard" switch off, one after the other. This effect is called Abrupt Decline of Luciferase Activity (ADLA).

## Control of luciferase activity with a *cyaA- acs-* double mutant

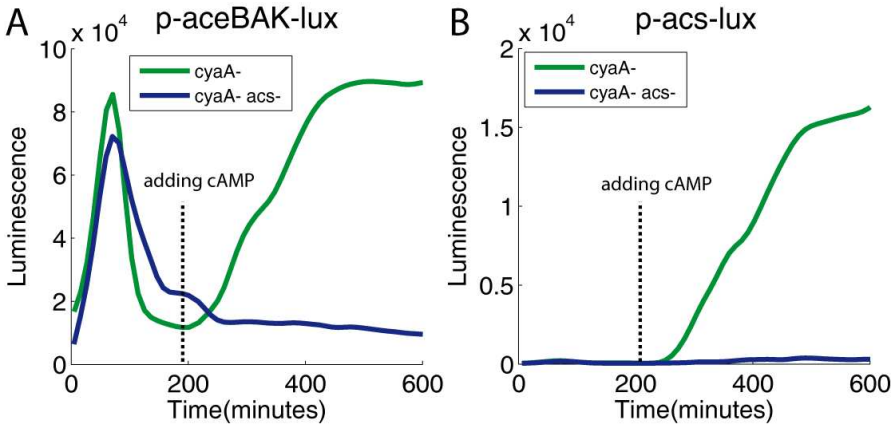



**Figure 2.** Control of luciferase activity with a *cyaA- acs-* double mutant. A *cyaA-* mutant and a *cyaA- acs-* double mutant were grown on minimal medium supplemented with glucose (0.015%) and acetate (0.03%). We added 5 mM cAMP to the medium at 200 minutes. (A) strains were transformed with the p-aceBAK-lux plasmid to monitor the energy level of the cell. There is a peak of luciferase during the phase of glucose consumption and the luminescence decreases when glucose is exhausted. In the *cyaA-* mutant, luciferase activity is restored by the addition of cAMP, which stimulates the transcription of *acs*, thereby allowing acetate to be used as a carbon source. In the *cyaA- acs-* double mutant, there is no recovery of luciferase activity. (B) The bacteria were transformed with the p-acs-lux plasmid (to monitor *acs* transcription): the transcription of *luxCDABE* operon from the *acs* promoter after adding cAMP only occurs if the chromosomal copy of *acs* is present.

In the experiment of the Figure 3B (see the section *Tuning and measuring acetyl CoA synthetase activity in vivo*), we have shown that adding exogenous cAMP in the growth medium leads to a recovery of luciferase activity. This phenomenon is probably due to the transcription of the *acs* gene, known to be induced by the Crp-cAMP complex. To definitively prove the essential role of this gene, we repeated this experiment in a *cyaA- acs-* background. In the supplementary figure 2A, we show that adding cAMP triggers the recovery of luciferase activity in the *cyaA-* mutant, but not in the *cyaA- acs-* double mutant, as expected. Moreover, in the *cyaA- acs-* double mutant, despite cAMP being added to the medium, we failed to induce *acs* expression (Figure 2B). This result strongly suggests that the production of acetyl-coA synthetase after the induction of the *acs* transcription is critical for allowing the consumption of acetate. This alternative carbon source then provides the energy source needed to continue transcription and/or translation of its own promoter. This positive metabolic feedback loop provides a physiologically reasonable regulation in adaptation to the environment: the production of the enzyme needed for utilizing acetate as a carbon source is stimulated by the action of this same enzyme.

# Chapitre VI

## Conclusion et perspectives

La plus grosse partie expérimentale de ma thèse aura consisté à essayer de comprendre pourquoi l'ajout d'AMPc exogène ne déclenche pas l'expression du gène *acs* quand il y a encore du glucose dans le milieu. Pourtant, après avoir testé des centaines d'hypothèses et criblé l'ensemble des simples mutants d'*E. coli*, nous n'avons pas pu déterminer le mécanisme à l'origine de ce résultat. Notre apport a donc principalement consisté à déconstruire les deux modèles principaux expliquant le phénomène de répression catabolique. Nous avons ainsi montré que la variation intracellulaire d' [AMPc] *ne peut pas* être à l'origine de la répression catabolique. Ce constat est d'autant plus paradoxal que l'AMPc joue neamoins un rôle de premier plan dans le processus. De même, nous avons montré que l'état de phosphorylation du PTS *ne peut pas* expliquer la répression catabolique malgré le fait que, encore une fois, le PTS y joue indéniablement un rôle. Nous proposons, humblement et sans certitude, de recentrer les recherches sur la protéine Crp et les variations de flux glycolytique (indépendamment du ratio PEP/pyruvate) : ce sont les deux seuls éléments qui, dans nos conditions, se sont avérés indispensables au fonctionnement de l'interrupteur. Une hypothèse potentielle proposée par Magasanik il y a maintenant 50ans (Magasanik, 1961) est celle du métabolite répresseur : peut-on imaginer un deuxième petit métabolite qui contrôlerait l'expression de Crp sans que nous nous en soyons rendu compte jusqu'à maintenant ? La solution réside-elle dans un mécanisme plus complexe, encore inconnu, et difficile à détecter avec les méthodes actuelles ? Ou, la solution nous est-elle tout simplement passée sous les yeux ? L'avenir nous le dira.

Ainsi, c'est sur ces recherches restées bloquées dans une impasse que se sont greffés les résultats qui suivent. Ces derniers sont plus sûrs, plus facilement reproductibles, plus intéressants

et débouchent *a priori* sur bien plus d'applications. Ils n'ont cependant nécessité qu'une petite fraction de mon temps de travail expérimental.

L'étude de l'opéron *araBAD* est un cas d'école pour décrire les variations stochastiques dans l'expression des gènes. Dans nos expériences, nous avons pu montrer qu'un flux glycolytique fort a tendance à synchroniser l'expression de l'opéron *araBAD* alors que la diminution de ce flux à l'aide d'un sucre non métabolisable module la distribution d'expression de cet operon dans la population de cellules. Les modélisateurs pourraient tenter de comprendre ce phénomène et inclure le flux glycolytique dans leur modèle stochastique.

Notre interrupteur métabolique, basé sur les activités inversées des promoteurs *acs* et *uhpT* en fonction du flux glycolytique, est très robuste, facile à utiliser et très rapide : l'inversion des expressions est visible en une dizaine de minutes. Il possède cependant deux défauts majeurs : on ne sait pas comment il fonctionne, car il est basé sur le mécanisme inconnu de la répression catabolique. De plus, il manque d'orthogonalité c'est-à-dire qu'il n'est pas du tout indépendant/découplé du réseau de régulation de la bactérie. Nous espérons cependant que notre système séduira des biologistes synthétiques ou des modélisateurs.

Un autre système synthétique développé dans cette thèse est le système de communication intercellulaire par l'AMPc. Ce système est robuste et facilement modulable, mais il souffre également du manque d'orthogonalité : il est directement connecté au « hub » que représente le complexe Crp-AMPc. Or, ce défaut *a priori* peut être transformé en avantage car ce complexe, très documenté dans la littérature, est déjà connecté à de nombreux modules (motilité, chimiotactisme, métabolisme, virulence, résistance au stress) et compose un grand nombre de motifs récurrents comme les boucles de type « feed forward ». Il existe, en outre, un très grand nombre de promoteurs contrôlés par le complexe Crp-AMPc et donc un large choix d'affinité de sites de liaison. Celui qui voudrait s'emparer de la difficile question de « la modularité » des réseaux de régulation génique pourrait donc trouver ici un terreau expérimental favorable.

Dans notre système de communication synthétique, il a fallu construire et optimiser deux souches. Une usine microbienne qui fabrique de fortes concentrations d'AMPc extracellulaire et un bio-senseur qui en détecte de très faibles. L'usine bactérienne est capable de produire plus de 150$\mu$M d'AMPc dans le milieu extérieur soit $\sim$10-100 fois plus qu'une souche sauvage. Cette concentration reste très éloignée des 30mM produit par Microbacterium sp. no. 205. Il est donc peu probable que notre souche soit compétitive à l'échelle industrielle. Cependant, ce rendement pourrait être amélioré en supprimant la phosphodiesterase (le gène *cpdA*) qui a justement un $K_m$ de l'ordre de 500$\mu$M. Il n'est donc pas totalement exclu que l'amélioration de cette souche la rende compétitive. Le bio-senseur, lui, a sa limite basse de détection de l'AMPc extracellulaire vers 5-10$\mu$M. Or le $K_m$ de l'AMPc pour Crp est dix fois plus faible. On sait

que l'export de l'AMPc est actif, mais le gène impliqué dans cet export n'a pas encore été découvert. Pour espérer « récupérer » ce facteur 10, il faudrait donc déterminer quels sont les gènes impliqués dans l'import / export de l'AMPc, puis il faudrait bloquer l'export et facilité l'import.

Il reste également à déterminer si le processus de communication intercellulaire par l'AMPc n'existe que dans les mains du biologiste synthétique ou s'il s'agit d'un mécanisme physiologique. Pourquoi la cellule gaspillerait-elle 9% de ses réserves en ATP à secréter de l'AMPc (Matin & Matin, 1982) ? Ce processus de communication, s'il il existe, n'est pas forcement « inter-bactérien » : il pourrait potentiellement être « inter-royaume », car l'AMPc est une molécule de signalisation utilisée également par les cellules eucaryotes. Nous avons entrepris des expériences visant à montrer un processus de communication entre nos cellules et l'amibe sociale *Dictyostelium discoideum*, sans succès néanmoins, faute de temps (Figure VI.1). Cependant, c'est aussi l'énorme avantage de notre petit système synthétique : il peut être implémenté entre des royaumes différents. Nous espérons donc que d'autres réussiront là où nous avons dû renoncer.

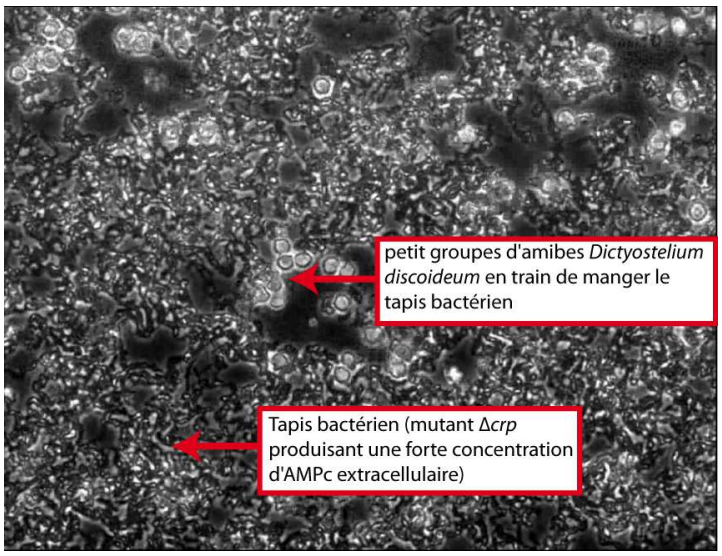



**Figure VI.1** – Expérience pour tenter de faire communiquer deux royaumes distincts. Nous avons mélangé, sur un milieu solide, des bactéries Δ*crp* produisant de fortes concentrations d'AMPc extracellulaire avec des amibes sociales *Dictyostelium Discoideum* connues pour se regrouper grâce à l'AMPc. Cependant, il n'a pas été possible de tirer de conclusions concluantes sur ce type d'expériences.

Enfin, le milieu le plus logique où pourrait prendre place cette communication via l'AMPc chez *Escherichia coli* est l'écosystème gastro-intestinal. Les conditions y sont très différentes des conditions de laboratoire et la densité de cellules beaucoup plus forte pourrait faciliter un processus physiologique de communication. Pour tester cette hypothèse, nous avons inoculé à

des souris, par voie orale, la souche bio-senseur. L'idée était que le bio-senseur s'allumerait dans les intestins de la souris grâce à un processus de communication avec une souche « sender » physiologique. L'expérience n'a pas pu démontrer de processus de communication. Cependant, nous avons ensuite administré, par voie orale, de l'AMPc à la souris. Puis, nous avons observé un accroissement de lumière dans la zone du tractus gastro-intestinal et ce, de manière non-invasive, c'est-à-dire transcutané (Figure VI.2). Dans cette expérience, nous avons donc induit l'expression d'un gène (l'opéron *luxCDABE*) de bactéries placées dans le tractus gastro-intestinal d'une souris et nous avons contrôlé visuellement la réussite de l'induction. Ce type d'expériences pointe du doigt des technologies plus futuristes où nos intestins seront inspectées par des cellules médicaments, véritables micro-ordinateurs synthétiques, chargées de surveiller et de rétablir l'homéostasie de l'hôte ou de signaler la présence d'un pathogène.

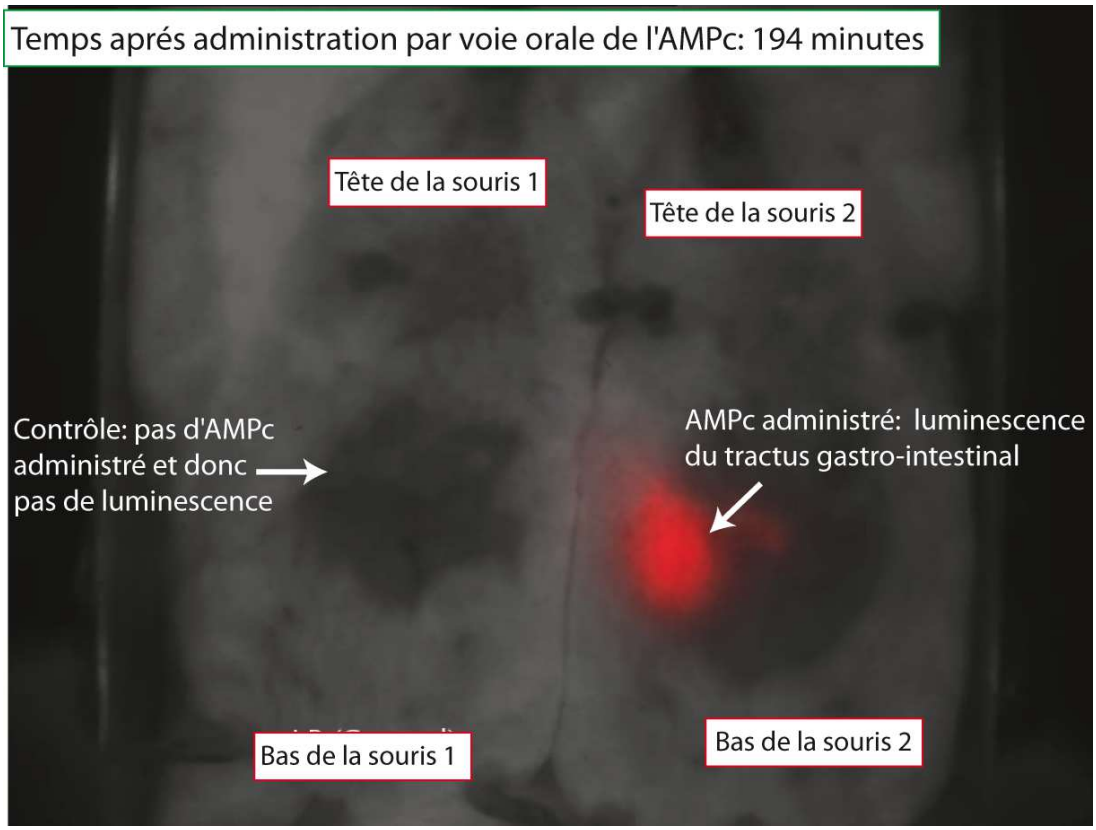



**Figure VI.2** – Induction de l'operon *luxCDABE* directement dans le tractus gastro-intestinal d'une souris grâce à l'AMPc. On a administré à deux souris (placées côte à côte sur l'image) des milliards de bactéries $\Delta cyaA$ contenant la fusion transcriptionnelle p-sdh-lux très sensible à l'AMPc extracellulaire. Puis nous avons administré oralement à la souris de droite une forte dose d'AMPc alors que la souris de gauche n'a eu qu'une solution sans AMPc en guise de contrôle. Après quelques dizaines de minutes, on peut observer, de manière transcutanée grâce à une camera sensible, la luminescence émise, depuis le tractus gastro-intestinal, par ces milliards de bactéries dont l'operon *luxCDABE* a été transcrit grâce à la présence d'AMPc.

Dans ces expériences de communication, il perdure une sorte de regret « artistique » : malgré des dizaines de tentatives, nous n'avons jamais pu filmer le processus de communication par l'AMPc, sous le microscope, à l'échelle de la cellule. Nous attendons donc que quelqu'un relève ce défi et nous envoie les vidéos.

Il reste maintenant à décrire les perspectives liées à l'utilisation de l'activité luciferase pour inférer de nombreux paramètres *in vivo* et en temps réel sur le fonctionnement de la cellule. Il s'agit, sans doute, des résultats qui auront le plus d'impact, même si paradoxalement ce sont aussi ceux qui ont été les plus faciles à obtenir. La combinaison de la mesure de l'activité luciferase, des banques de mutants et du savoir actuel sur le métabolisme pourrait aider à générer des données quantitatives plus précises sur l'état des flux métaboliques dans une condition donnée. L'avantage de cette technique est qu'elle ne nécessite pas d'étapes d'échantillonnage et de « quenching » comme c'est le cas pour les mesures de metabolome effectuées *a posteriori* de l'éclatement des cellules.

La capacité à maintenir l'activité luciferase apporte de nombreuses informations *in vivo* et en temps réel : citons l'état énergétique global, l'état du regulon Crp, l'état de la machinerie globale, l'état du milieu, mais aussi des informations plus précises sur l'état d'un flux ou d'une enzyme particulière. La facilité, la sensibilité et le faible coût de détection de la lumière rend notre méthode relativement facile à implémenter. Elle pourrait avoir de nombreuses applications en biotechnologie, en biologie des systèmes ou en biologie synthétique. L'utilisation de notre méthode en fermenteur industriel semble possible *a priori* : il n'est même pas nécessaire que toutes les cellules disposent de la luciferase (ce qui risquerait d'avoir un coût énergétique trop élevé) : une petite proportion de cellules, présentes dans le fermenteur et saturées en luciferase, collecteraient les informations nécessaires, informations elles-mêmes récupérées en temps réel par un simple photomultiplicateur.

A l'opposé, et même si cela nécessite sans doute un matériel plus coûteux, il n'est pas impossible que notre méthode soit implémentable à l'échelle de la cellule isolée. En effet, des cameras sensibles, placées sur un microscope, arrivent à détecter la luminescence de cellule unique (Mihalcescu *et al.*, 2004). Nous pourrions alors suivre, en temps réel, l'état du métabolisme central de cellules isolées ainsi que tous les autres paramètres cités précédemment.

Enfin nous terminerons par vanter les mérites de la fusion transcriptionnelle p-acs-lux rapportant l'activité du promoteur *acs* grâce à la luciferase. Il s'agit d'un bio-senseur robuste et facile à utiliser. Placé dans un milieu minimum avec du glucose et de l'acétate, il rapporte l'épuisement du glucose (induction de la luciferase) et l'épuisement de l'acétate (chute brutale de la luminescence). Placé dans un milieu minimum avec du glucose uniquement, il rapporte la concentration d'acétate sécrétée dans le milieu extérieur c'est-à-dire le débordement du mé-

tabolisme (overflow metabolism). L'accumulation de l'acétate dans le milieu de culture est un problème important dans la fermentation industrielle car cet acide organique inhibe la croissance cellulaire et donc la production des protéines recombinantes. Différentes approches sont actuellement développées pour tenter de réduire l'accumulation d'acétate en modifiant les flux du métabolisme central (Gosset, 2005). Notre bio-senseur pourrait potentiellement aider à mesurer, en temps réel, l'acétate secrété voire même à cribler des banques de mutants à la recherche de candidats intéressants secrétant peu d'acétate.

# Références bibliographiques

Gosset, G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate :sugar phosphotransferase system. *Microb Cell Fact*, **4**(1), 14. 132

Magasanik, B. 1961. Catabolite repression. *Cold Spring Harb Symp Quant Biol*, **26**, 249–56. 127

Matin, A., & Matin, M. K. 1982. Cellular levels, excretion, and synthesis rates of cyclic AMP in Escherichia coli grown in continuous culture. *J Bacteriol*, **149**(3), 801–7. 129

Mihalcescu, I., Hsing, W., & Leibler, S. 2004. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*, **430**(6995), 81–5. 131

L'adaptation d'une bactérie aux changements de son environnement est contrôlée par un réseau de régulation large et complexe, faisant intervenir de nombreux acteurs et modules différents. Dans ce travail, nous avons étudiés un module de régulation spécifique, contrôlant l'adaptation de la bactérie *Escherichia coli* à un changement de sources de carbone. Dans un milieu contenant du glucose et de l'acétate, la croissance est divisée en deux phases : les bactéries utilisent préférentiellement le glucose et commencent à métaboliser l'acétate qu'après l'épuisement du glucose. En effet, la présence du glucose réprime la transcription d'un gène nécessaire à la croissance sur acétate, le gène *acs* (codant pour l'acétyl-CoA synthétase). Le mécanisme régulateur fait intervenir le facteur de transcription Crp-AMPc et le système de transfert de phosphate (PTS), qui permet l'import du glucose.
Plusieurs modèles décrivent en détail la cascade de réactions moléculaires à l'origine de cette « répression catabolique ». Cependant, certaines de nos observations expérimentales ne sont pas correctement prédites par les modèles actuels. Ces modèles doivent être révisés ou complétés. L'outil majeur que nous employons pour les expériences est la fusion transcriptionnelle : une région promotrice fusionnée en amont d'un gène rapporteur (GFP, luciferase). Avec ces constructions, nous mesurons la dynamique de l'expression génique dans différentes souches (mutants) et différentes conditions environnementales. Les observations à l'échelle de la population sont corroborées par des mesures similaires à l'échelle de la cellule unique. Nous utilisons cette même technologie pour construire de petits systèmes synthétiques qui sondent davantage le phénomène de répression catabolique. Nous avons ainsi créé un interrupteur génétique dont le fonctionnement est contrôlé par le flux glycolytique et nous avons construit un petit système de communication intercellulaire basé sur la molécule AMPc. Enfin, nous proposons une manière originale de mesurer l'état métabolique des cellules en utilisant la dépendance énergétique de la luciferase.

Mots clés : Réseaux de régulation, biologie synthétique, cinétique d'expression génique, répression catabolique, gènes rapporteurs

The adaptation of bacteria to changes in their environment is controlled by a large and complex regulatory network involving many different actors and modules. In this work, we have studied a specific module controlling the adaptation of *Escherichia coli* to a change in carbon sources. In a medium containing glucose and acetate, growth is divided into two phases : the bacteria preferentially use glucose and start to metabolize acetate only after glucose exhaustion. Indeed, the presence of glucose represses the transcription of a gene needed for growth on acetate : the *acs* gene (coding for acetyl-CoA synthetase). The regulatory mechanism involves the Crp-cAMP regulator and the phosphate transfer system (PTS), which is responsible for glucose import.
Several models describe the cascade of molecular reactions responsible for this « catabolite repression ». However, our work shows that many of our experimental observations are incorrectly predicted by current models. These models have to be amended. We use transcriptional fusion, i.e., the fusion of a promoter region upstream of a reporter gene (GFP, luciferase), to measure the dynamics of gene expression in different genetic backgrounds and environmental conditions. Observations at the population level are corroborated by similar measurements at the single cell level. We use this same technology to construct small synthetic systems that probe further aspects of the phenomenon of catabolite repression. We have thus created a genetic toggle switch controlled by the glycolytic flux and we have built an inter-cellular communication system mediated by cAMP. Finally, we propose a novel way to measure the metabolic state of cells by using the energy dependence of the luciferase enzyme.

Keywords : Regulatory network, synthetic biology, kinetics of gene expression, catabolite repression, reporter genes