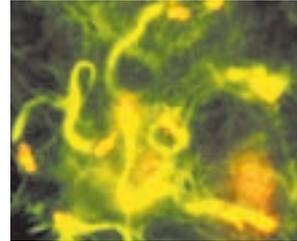


# Saurons-nous construire une bactérie synthétique ?

Antoine Danchin

► La Biologie Synthétique, nouvel avatar des applications de la recherche en biologie, postule qu'il suffira de combiner des modules de base et d'introduire un génome artificiel dans une cellule réceptrice appropriée pour mettre en route une usine cellulaire capable de se reproduire et d'effectuer des opérations programmées par l'Homme. Cette vision repose sur l'idée implicite qu'une cellule se comporte comme un ordinateur qui serait capable de construire des ordinateurs. Cela suppose qu'un objet physique portant un ensemble [données + programme] peut, dans la réalité, être séparé de la machine, et que le programme peut s'exprimer sous la forme d'un « réplicateur » et d'un « constructeur » qui contienne quelque part une image de la machine qu'il va construire. Ce que nous savons des bactéries nous confirme dans cette vue et nous donne une image du plan de construction de ce qui serait une cellule synthétique, si l'on débarrasse les génomes connus de tous leurs appendices « inutiles ». En bref, un ensemble de gènes persistants (le paléome, qui persiste depuis l'origine de la vie) définit les programmes du réplicateur et du constructeur, en y ajoutant un ensemble de gènes importants pour la maintenance et la réparation de l'ensemble. Quelques règles du développement de la vie « en contexte », permettent de comprendre comment nous pourrions un jour faire en sorte que des cellules synthétiques produisent ce que nous attendons d'elles. ◀



Génétique des Génomes  
Bactériens, URA 2171 CNRS,  
Institut Pasteur,  
25, rue du Docteur Roux,  
75724 Paris Cedex 15, France.  
[antoine.danchin@normalesup.org](mailto:antoine.danchin@normalesup.org)

en accord avec elles, mais avec des composants chimiques différents de ceux que nous connaissons chez les organismes vivants [1]. Un caractère propre à la vie nous aide : loin d'aller contre la Nature, la vie en domestique les lois. Au lieu d'être une lutte épuisante contre le deuxième principe de la thermodynamique (comme un temps, bizarrement, certains l'ont cru), la vie en fait un moteur essentiel, la base de toute exploration, concrétisation de cette entropie qui fait que tout système matériel tend à occuper tous les états et tous les lieux qui lui sont accessibles [2]. Saurons-nous donc construire une bactérie synthétique, et quelle bactérie ?

## Ce qu'est la vie

Trois processus, compartimentation, métabolisme et transfert d'information, organisés en ensembles matériels aux propriétés bien différentes, constituent la vie. D'abord, la vie suppose la définition d'un intérieur et d'un extérieur, l'intérieur distinguant la matière vivante de la matière de l'environnement. Il y a donc une enveloppe : la cellule isolée, atome de vie, est entourée d'une enveloppe plus ou moins compliquée comprenant une ou deux membranes permettant les échanges entre l'intérieur et l'extérieur et servant de protection ; dans le cas d'organismes plus complexes, de multiples membranes et « peaux » définissent l'organisme. Cette *compartimentation* permet une analogie avec ces constructions humaines, créées pour effectuer une action, que sont les machines et leurs châssis (ou cadres). Cependant cette analogie ne peut se cantonner à celle des automates du XVIII<sup>e</sup> siècle, car la vie ne possède pas les caractères de prévisibilité, de comportement stéréotypé, qu'ont les machines mécaniques. Si automates il y a, ce seront des automates analogues à des ordinateurs, non à des horloges.

## Préambule

Il y a cinquante ans naissait la Biologie Moléculaire, il y a vingt ans c'était la Génomique, et nous assistons aujourd'hui à la naissance de la Biologie Synthétique. Cette discipline est née de l'idée que la pratique des ingénieurs pouvait s'appliquer à ce que l'on sait de la Biologie au point de permettre non seulement la synthèse de nanomachines se développant selon ses lois, mais encore de reconstruire des cellules vivantes



La vie, ensuite, est caractérisée par une dynamique originale : elle nécessite le flux permanent de transformations de composés en d'autres composés, en parallèle avec la domestication de l'énergie chimique nécessaire à la construction/déconstruction de ces composés. Ce processus est le *métabolisme*. Il décrit la transformation mutuelle de quelques centaines de molécules - réalisables chimiquement - à base d'azote, de carbone, d'hydrogène, d'oxygène, de phosphore et de soufre selon des règles très précises, dont le rôle est l'édification des composants de base de la cellule, la gestion des flux de matière et d'énergie. Au moment de la mort, compartimentation et métabolisme vont disparaître pour ne laisser que des scories. Enfin, le troisième processus spécifique de la vie, bien que construit sur des objets très concrets, est un processus abstrait : il s'agit de l'organisation élaborée d'un permanent *transfert d'information*, qui permet de construire l'individu à partir d'un programme de construction et de fonctionnement. Ce transfert s'opère autour d'objets chimiques originaux, les acides nucléiques, qui permettent l'exécution d'un ensemble de procédures résumées sous la forme d'un texte symbolique écrit dans un alphabet à quatre lettres, et qu'on appelle le programme génétique.

Le plus frappant dans ce couplage entre compartimentation, métabolisme et transfert d'information, est que tout se passe comme si il y avait séparation entre les deux premiers et le programme génétique. C'est l'hypothèse implicite qui fonde la Biologie Synthétique. L'ADN peut être considéré comme une suite linéaire de symboles constituant l'algorithme alphabétique, et il est séparable de la cellule (= machine) qui met en œuvre le programme qu'il spécifie. Cette séparation est illustrée par le génie génétique - de fait la première phase de la Biologie Synthétique - qui reposait sur la reprogrammation d'une cellule par des fragments de programmes importés d'ailleurs ou même entièrement artificiels. On remarque alors, qu'au moins de façon superficielle, la cellule ressemble à une machine de Turing (*voir encadré 1 et Figure 1*).

1

### LA MACHINE DE TURING, CONCEPT CENTRAL DE LA CONSTRUCTION DES ORDINATEURS

La machine déroule les opérations suivantes, toutes faciles à réaliser (conceptuellement ce sont celles qui se déroulent dans les ordinateurs, bien qu'elles y soient produites sous forme plus condensée et donc plus compliquée) :

- changer un symbole en un nombre fini de places, après lecture des symboles trouvés (noter que changer plus d'un symbole à la fois se résume en un nombre fini de changements uniques successifs)
- mouvoir la lecture d'un point vers un autre, à une distance finie et limitée dans la suite des symboles
- changer l'état de la machine

Tout cela constitue une *Machine de Turing*. Elle est ainsi spécifiée par un ensemble de quintuplet, de l'une des trois formes suivantes :

$$p\alpha\beta Gq \text{ ou } p\alpha\beta Dq \text{ ou } p\alpha\beta Mq$$

Chaque quintuplet indique que la machine est dans la configuration  $p$ , quand elle lit le symbole  $\alpha$ , et le remplace par  $\beta$ , avant de prendre la configuration  $q$  et de se déplacer vers la gauche ( $G$ ), la droite ( $D$ ), ou de rester au point mort ( $M$ ).

Ce qui est remarquable est que la combinaison d'éléments de ce seul ensemble d'actions suffit à réaliser toutes les opérations de la logique et du calcul.

## Vie et calcul

### Les contraintes de la machine

La biologie se place dans le monde de la physique, dont elle comprend les objets, matière, énergie et temps. Il s'y ajoute d'autres caractères comme codage, information, régulation, contrôle... En parallèle, des programmes opèrent dans ces machines. Intervient ici le monde de l'abstraction mathématique, où l'arithmétique et la Théorie des Nombres étudient les suites de nombres entiers autour de concepts comme codage ou récursivité<sup>1</sup>, fondamentalement à l'œuvre dans le calcul, autour du concept d'algorithme réalisé sous forme de programmes opérant dans des machines. Les contraintes pour la définition de la machine sont très limitées, et, pour l'essentiel, demandent que les données de la suite de symboles (souvent appelée « programme » par abus de langage) soient pour l'essentiel séparables de la machine [3].

Il n'est pas question évidemment ici d'entrer dans le détail des propriétés ou des contraintes de la machine. Nous retiendrons seulement l'interaction de la machine et du programme qu'elle déroule. En bref, ces programmes correspondent à des processus algorithmiques hautement parallèles caractérisés par un début, l'expression (qui peut être récursive) de routines, des points de contrôle et une fin. Or tous les processus de l'expression génétique, que ce soient la réplication de l'ADN, sa transcription en ARN messager ou sa traduction en protéines suivent très exactement cette organisation séquentielle<sup>2</sup>. Par ailleurs, nous l'avons dit, l'ADN est séparable de la cellule : de l'ADN étranger s'exprime de fait parfaitement dans la plupart des cellules. Mieux, on sait aujourd'hui que le génome des bactéries comprend une part importante de gènes provenant d'ailleurs (souvent au moins un cinquième du génome) par transfert génétique horizontal. Et les virus ne sont rien d'autre que des portions de programme qui se propagent d'une cellule à l'autre, « habillées » pour se protéger et s'adsorber sur leurs cibles afin de se faire reconnaître par la cellule. Ici, soulignant sans doute la profondeur de son adéquation au Réel,

<sup>1</sup> Rien n'empêche d'imaginer qu'une fonction puisse s'appeler elle-même. C'est ce que l'on appelle une fonction récursive, comme l'illustre le tableau de MC Escher qui représente un visiteur dans une galerie d'art qui représente la galerie où le visiteur se trouve.

<sup>2</sup> Un exemple de récursivité appliquée à l'ADN : l'ADN spécifie les protéines qui répliquent l'ADN, le transcivent, le réparent ; il faut des protéines pour faire les protéines etc ; par exemple pour réaliser une protéine du ribosome, il faut cette protéine.

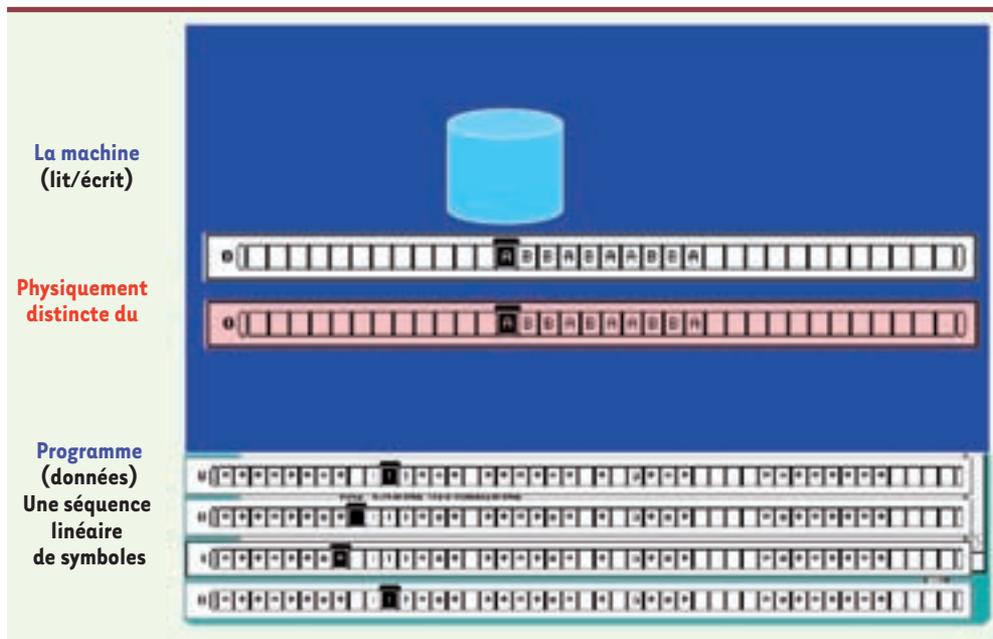


Figure 1. La machine de Turing.

la métaphore est allée en sens inverse, de la biologie au calcul, puisqu'on parle (avec raison) de virus informatiques. Enfin, récemment, Hamilton Smith et ses collègues ont démontré expressément la séparation machine/données-programme en remplaçant entièrement l'ADN d'une bactérie par celui d'une espèce différente [4]. Toute tentative de construction d'une cellule *ab initio* doit donc prendre en compte ces étonnantes contraintes.

### L'image de la machine, ou l'image de l'organisme dans le génome

Mais si l'analogie paraît profonde, les ordinateurs ne font pas (encore) des ordinateurs, alors que les cellules font des cellules. Que faut-il de plus ? Von Neumann, dans une réflexion fondatrice, montre qu'il faut alors la présence simultanée, dans le programme, d'un « réplicateur » et d'un « constructeur », avec, au sein de ce programme de construction, une image de la machine [5]. Peut-on trouver une image de la machine dans les organismes vivants ? Cette question pourrait être le signe d'une imagination débridée, mais il existe justement une propriété énigmatique des organismes supérieurs qui suggère l'idée d'un plan dans le génome, d'un ordre sous-jacent dont on ne comprend pas l'origine et qui montre bien un lien entre l'architecture d'un organisme et l'architecture de certains éléments du programme génétique. Les animaux sont formés de segments successifs définissant une partie antérieure et une partie postérieure avec souvent un dos et un ventre, une tête, un thorax et un abdomen. Or on sait aujourd'hui que ce plan particulier résulte de l'expression de gènes de contrôle, les homéogènes, qui sont répartis dans le génome dans le même ordre que l'ordre des segments. Mieux, on sait - par exemple en comparant insectes et crustacés, comme l'avait fait Geoffroy Saint-Hilaire contre les idées de son temps - que le déplacement de ces gènes dans les chromosomes déplace les segments du corps. Par exemple, il y a une inversion de deux gènes

Hox chez les crustacés par rapport à la situation chez les insectes, et sous le thorax chez les crustacés, le ventre devient le dos [6]. Le déplacement d'autres gènes fait apparaître des pattes à la place des antennes chez la mouche drosophile, ou lui met des yeux un peu partout ! Ces inversions se sont produites pour d'autres gènes au cours de l'évolution, comme de nombreux remaniements chromosomiques dont certains ont eu une descendance, d'autres pas. L'idée, par conséquent, qu'il pourrait exister une image<sup>3</sup> de l'organisme dans le génome

n'est peut-être pas si délirante : on la trouve dans tous les organismes multicellulaires organisés, sans, d'ailleurs en comprendre l'origine ou la raison. Pour faire une cellule synthétique il faudra donc poursuivre cette hypothèse et explorer ce qu'il en est dans la cellule elle-même, y analyser les classes de gènes, et la façon dont ils y sont répartis, et le cas échéant appliquer ces règles à la construction de la cellule artificielle.

### Du génome à la vie synthétique : quels gènes ?

Deux approches permettent de se demander comment construire une cellule. On peut extraire les concepts de ce qu'on connaît de la biologie, faire le catalogue minimum de ses objets et tenter de mettre le tout ensemble, comme le font les ingénieurs lorsqu'ils construisent des machines. Cette approche est par exemple celle mise en œuvre par la compétition internationale iGEM (*The international genetically engineered machine competition*), qui suscite l'imagination d'étudiants du monde entier (→ voir l'article de D. Bikard et F. Képès, page 541 de ce numéro (→)).

<sup>3</sup> Il s'agit ici d'un renouvellement de l'image de l'homunculus : au lieu d'une régression de poupées russes, impossible comme Buffon l'a montré, c'est l'algorithme de construction qui est transmis de génération en génération - et là, cela ne conduit pas à une régression à l'infini, c'est donc un renouvellement complet du concept de préformation ; il y a « préformation » mais non pas de la forme finale, mais de la recette pour la construire. L'épigénèse, qui était opposée à la préformation, est alors l'actualisation explicite dans un individu donné, de l'algorithme de construction ; une conséquence fondamentale, curieusement peu comprise est que parler d'inné et d'acquis n'a pas vraiment de sens : tout est à la fois inné et acquis dans un individu donné.

On sait reconstituer *in vivo* toutes les étapes élémentaires du « calcul » des ordinateurs [8, 9]. Toutefois, les organismes vivants s'en distinguent par le niveau très élevé du bruit<sup>4</sup>, une contrainte que les compétitions de biologie synthétique ne savent pas encore bien prendre en compte. Au surplus, il est difficile, sinon impossible, de reconstruire une cellule fonctionnelle sans plan d'ensemble : travailler sur des composants élémentaires est donc très limitant. Il est irréaliste de croire avoir déjà tout compris de ce qu'est la vie, il faut donc procéder en sens inverse, en faisant la synthèse conceptuelle de tout ce que l'on observe et comprend dans un grand nombre d'organismes, que, pour simplifier, on restreindra à un domaine particulier, celui des bactéries. Près de 4 000 programmes de séquençage de génomes sont à ce jour en cours ou terminés, et en utilisant leur comparaison mutuelle à partir d'organismes modèles choisis comme *Pierre de Rosette*, il est possible de se faire une idée des contraintes qui font la vie.

Une première observation remarquable montre que tous les gènes n'ont pas un statut équivalent. Certains (dits gènes persistants, ou formant le paléome) se retrouvent dans un très grand nombre de génomes (ils sont présents dans une « clique ») (Figure 2) ; d'autres, au contraire, ne sont présents que dans un seul. Ces deux ensembles bien distincts ont une propriété commune : les gènes les plus fréquents, qui persistent dans un grand nombre de génomes, tendent à être regroupés en un petit nombre d'endroits du génome et il en est de même des gènes les plus rares. Ce dernier cas est facile à comprendre : ces gènes proviennent de transferts génétiques horizontaux, de l'invasion locale d'un génome par une portion de génome venant d'ailleurs – souvent un virus, un bactériophage. Ces derniers gènes sont spécifiques de

l'occupation par la cellule d'une niche particulière. Ils codent des fonctions d'exploration, le mouvement de la cellule, et la dégradation/récupération de composés de l'environnement – et constituent ce qu'on peut appeler le cénome (de *κοινός*, commun) de l'organisme, ce qui correspond à un milieu commun, partagé par les organismes vivant dans ce même milieu. Il est plus difficile de comprendre comment les gènes du premier ensemble (paléome) sont regroupés : présents un peu partout ils semblent nécessaires à la vie, et sont donc depuis longtemps soumis sans cesse au flux séparateur de l'invasion par les gènes du cénome.

Une explication simple (mais qui demande une justification approfondie) serait la suivante. Chacun de ces gènes (quelques centaines au plus, qui codent les fonctions de base nécessaires aux trois processus qui font la vie, compartimentation, métabolisme et transfert d'information, mais aussi à leur maintenance et à leur protection) contribue à chaque génération à la capacité de la cellule à produire une descendance. Or, comme le génome est sans cesse envahi par de l'ADN extérieur, il code une règle de destruction de gènes, de façon à maintenir sa taille à peu près constante. Ainsi, à chaque génération, de l'ADN entre et de l'ADN disparaît. Mais si l'ADN qui s'en va contient des gènes essentiels pour produire une descendance, alors la cellule affectée n'aura pas d'avenir. On comprend alors que si les gènes persistants sont répartis plus ou moins uniformément dans le chromosome, le risque d'une délétion accidentelle de l'un d'entre eux est considérable (Figure 3). Au contraire, s'ils sont regroupés, ou bien un certain nombre disparaîtra (mais cela n'est pas pire que la disparition d'un seul) alors que la délétion de gènes

du cénome n'aura que des conséquences mineures [10], ou bien la délétion aura lieu dans une région sans gènes persistants, et cela n'aura pas d'effet sur la descendance. Ainsi ce regroupement est shakespearien, c'est la sélection par l'existence !

En conclusion, à ce point de notre raisonnement, le génome des bactéries ressemble formellement au génome de la bactérie synthétique souhaitée, qui devra comprendre l'ensemble des gènes persistants, et – c'est l'équivalent du

<sup>4</sup> Notion de bruit : les réalisations explicites des fonctions biologiques sont extraordinairement fluctuantes au contraire de ce qui se passe dans les machines conçues par l'homme ; la raison en est que à la température de 300 K les énergies mises en jeu à chaque étape des réactions du métabolisme, de la synthèse des macromolécules etc. ne sont pas très différentes de celle de l'agitation thermique, d'où une variation importante dans des actions qui devraient autrement être d'un type unique ; en bref l'agitation thermique conduit à une frange d'incertitude considérable dans le devenir de toutes les actions biologiques élémentaires.

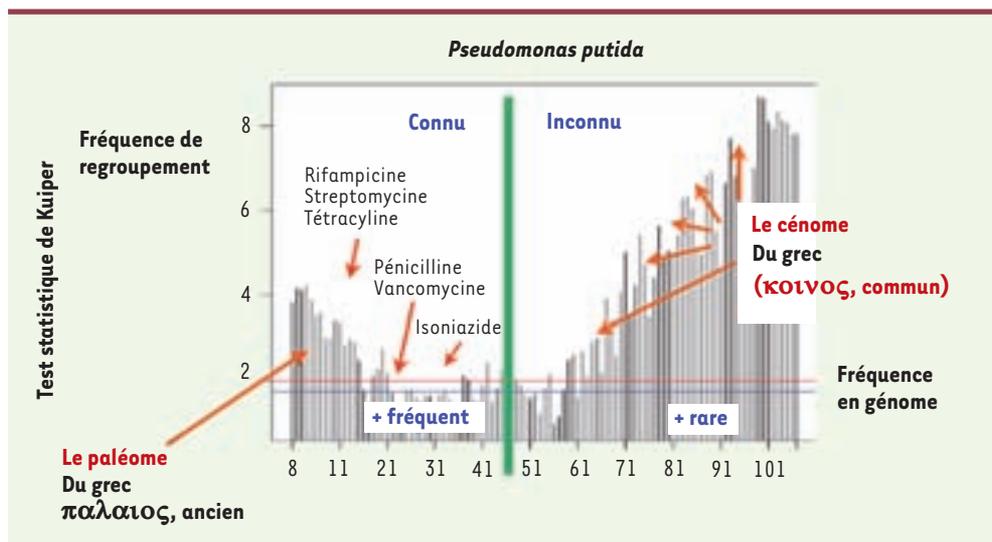


Figure 2. Fréquence des gènes du paléome et du cénome dans le génome (d'après [12]).

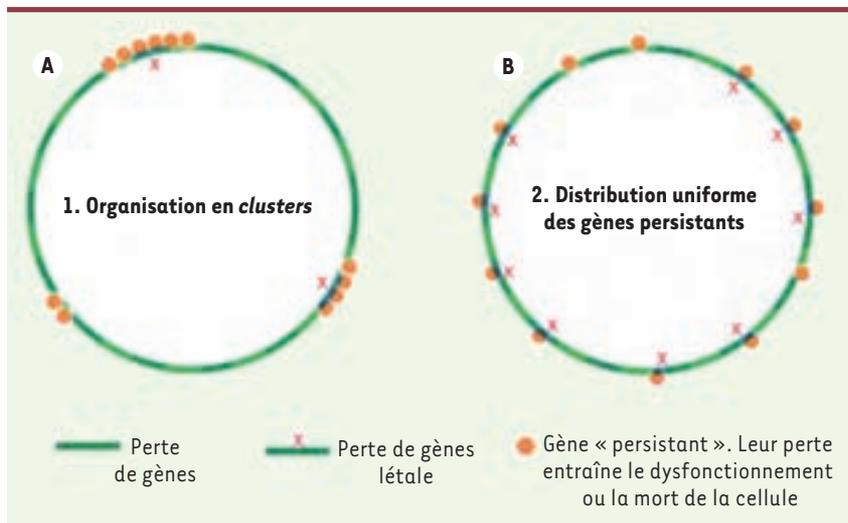


Figure 3. Le groupement des gènes « persistants » en clusters assure la survie de la cellule [10].

cénome - les gènes qui répondront spécifiquement à l'intention de son créateur, pour dépolluer l'environnement, produire de l'hydrogène, fabriquer des molécules importantes pour la chimie fine, etc.

## Le paléome

### Un nombre décroissant de gènes persistants

Que sont donc, dans le détail, ces gènes qui persistent dans la plupart des génomes ? D'abord un préalable important : si l'on cherche les gènes qui se trouvent dans tous les génomes pour en déduire la cellule minimale, on tombe sur une observation remarquable. Au fur et à mesure que de nouveaux génomes sont déchiffrés, ce nombre ne cesse de décroître, au point qu'on peut penser qu'un jour aucun gène ne sera plus commun à tous les génomes. Ce fait, surprenant pour un observateur superficiel, vient de ce que la vie ne constitue pas un ensemble d'objets, mais correspond à un processus physique beaucoup plus abstrait, formé de relations entre objets (et même de relations entre relations) [2]. Il s'en suit que plusieurs objets d'origine différente peuvent avoir la même place et la même fonction dans l'ensemble relationnel qui constitue la vie. Au cours de l'évolution, il arrive donc constamment qu'un gène venant d'ailleurs puisse coder une fonction de façon plus efficace, ou mieux adaptée qu'un gène préexistant. Après quelques générations, au travers du phénomène de délétion décrit plus haut, ce nouveau gène pourra avoir remplacé l'ancien. C'est ce processus de capture de structures pour réaliser des fonctions qui est caractéristique de l'aspect « bricolé » des organismes vivants [11]. Notons que ce fait est d'une importance capitale pour la Biologie Synthétique, car il signifie que le bricolage à l'œuvre (l'évolution par remplacement d'un objet par un autre est purement accidentelle<sup>5</sup>,

<sup>5</sup> Le bricolage aboutit à la capture ; ce qui se passe est que la « nécessité » d'une fonction donnée peut exister dans l'absolu - l'abstrait -, elle crée donc une sorte de béance pour un objet qui remplirait la fonction ; la conséquence de ce fait est que dès qu'un objet peut satisfaire la fonction, même mal, il est sélectionné ; mais le plus souvent il fait mal son office « faute de mieux » ; aussi dès qu'un autre objet, soit venant d'un transfert génétique horizontal, soit résultant de la création de novo d'un gène - question

même si la sélection oriente sans cesse vers une meilleure adaptation à des conditions données) pourra être remplacé par un vrai dessein intelligent, celui décidé par l'Homme ! Les fonctions à comprendre sont donc celles d'un ensemble de gènes « persistants », présents dans une majorité de génomes (clique), et non d'hypothétiques gènes ubiquistes.

### Une organisation en trois classes

L'analyse des gènes persistants dans un grand nombre de génomes bactériens permet une autre observation marquante. On trouve en effet qu'ils tendent à rester regroupés selon une organisation surprenante. En bref, on en trouve trois classes [12]. Une première classe

est faite de gènes qui ne conservent pas vraiment la même position relative par rapport aux autres dans les différents génomes. Tout se passe comme si ces gènes étaient tellement anciens que leur place avait eu le temps de changer considérablement. Pourtant ces gènes ont une particularité fonctionnelle commune : ils codent les voies essentielles du métabolisme de la cellule, constructeur des petites molécules à squelette de carbone à la base des activités catalytiques, de la membrane et des macromolécules, acides nucléiques et protéines. Une deuxième classe est formée de gènes qui s'organisent autour de ceux qui codent la fonction de passage entre le message du programme génétique, famille d'acides nucléiques particuliers, les ARN, et sa réalisation sous la forme de protéines. Ce passage est géré par une famille d'adaptateurs, les ARN de transfert. Ces ARN particuliers font le lien entre le monde des acides nucléiques et le monde des protéines. La troisième classe, elle, est extrêmement connectée, et elle s'organise principalement autour de la nanomachine qui traduit les messagers en protéines (via la lecture par les ARN de transfert), le ribosome.

Cette organisation est frappante. Elle s'interprète le plus simplement en imaginant que la classe la plus ancienne est la moins connectée, alors que la plus récente l'est restée. On se trouve alors en face d'un scénario plausible de l'origine de la vie ! Le parallèle

entièrement ouverte - se présente qui réalise mieux la fonction il aura tendance à remplacer celui qui fait mal son office ; au lieu d'une évolution par amélioration progressive d'un objet initial, on aura soudain le remplacement par un objet ayant une histoire évolutive totalement différente... ; le résultat de ce processus est que le nombre des objets essentiels restant apparentés entre tous les êtres vivants se réduit comme peau de chagrin au fur et à mesure qu'on séquence de nouveaux génomes ; les fonctions elles, sont préservées, mais les structures qui les réalisent ne le sont pas.

## 2

### UN SCÉNARIO D'ORIGINE

À l'origine, la surface de solides chargés (comme la pyrite de fer, des argiles contenant des phosphates et des métaux, etc.) assure la sélection par la charge électrique de molécules qui vont pouvoir réagir entre elles, ainsi que la compartimentation primitive de molécules chargées (retenues par la charge électrique de la surface des pierres), et la formation de membranes primitives. Ce métabolisme de surface synthétise les coenzymes (les molécules au cœur des réactions du métabolisme) et les maillons élémentaires des acides nucléiques (les nucléotides) et des protéines (les acides aminés) et les membranes puis conduit à un métabolisme dont le support était l'ancêtre des ARN de transfert, suivi par un « monde ARN » où apparaissent les transferts d'information avec la loi de complémentarité (celle qui permet la réplication de l'ADN et la synthèse des ARN messagers) et le concept de matrice, puis le code génétique. La polymérisation des acides aminés et des nucléotides avec élimination d'eau est ici favorisée par effet positif de l'augmentation d'entropie (dilution de l'eau produite durant la polymérisation au sein de l'environnement aqueux). Une fois ainsi compartimenté, le métabolisme crée des substituts des surfaces *via* la polymérisation des nucléotides (le monde ARN) [13].

entre ce scénario et la conservation des gènes correspondants au sein d'un grand nombre de génomes, conduit à considérer les gènes persistants comme très anciens : cette portion du génome des bactéries constitue leur paléome (de *παλαιος*, ancien). Pour assurer le succès de l'entreprise, une bactérie synthétique aura donc un génome constitué de gènes codant les fonctions du paléome, et permettant ainsi la construction de la vie, et d'un ensemble de maintenance et de réparation, associés à des gènes spécifiques accomplissant la fonction choisie pour l'organisme synthétique.

### Contraintes physico-chimiques

Une fois identifiés les gènes nécessaires, il convient de les mettre ensemble au sein d'un chromosome. Une première hypothèse sera de conserver l'organisation qu'ils ont pour un type architectural de cellule donné (typiquement un bacille). Mais cela suppose d'abord intégrer, dans un petit volume, une longue molécule d'ADN. La question n'est pas si simple si l'on considère les paramètres physiques de cette molécule : en bref, la longueur du chromosome est de l'ordre de mille fois celle de la cellule. Bien sûr il s'agit d'une molécule flexible et l'on sait calculer en fonction de sa rigidité comment elle peut se replier en une pelote statistique : le résultat est sans appel, la pelote en question aurait un diamètre dix fois supérieur à celui de la cellule. On doit donc imaginer que des contraintes supplémentaires forcent la molécule à se tenir dans un petit volume. Cette molécule d'ADN, lorsqu'elle se réplique, ne risque-t-elle pas de ne pouvoir séparer le chromosome parental de sa réplique ? Il s'agit en effet d'un très long fil, passablement embrouillé. C'est là qu'apparaît un caractère particulièrement positif de l'augmentation d'entropie dans la dynamique de la propagation de la vie. Le physicien coréen Suckjoon Jun en a fait la démonstration éclatante. Il a eu l'idée de reprendre le modèle des gaz parfaits de Boltzmann, mais de connecter par un fil continu les molécules de chacun des deux gaz (ce qui est l'équivalent de deux chromosomes) et de regarder ce à quoi conduit alors le deuxième principe de la thermodynamique. Le résultat est clair : si les deux fils occupent le même compartiment cylindrique,

ils vont tendre peu à peu à se séparer et à occuper deux espaces disjoints, si l'espace libre est suffisant [14, 15] (Figure 4). On observe donc très exactement avec un fil continu l'opposé de ce qui arrive au mélange des deux gaz parfait. L'augmentation d'entropie conduit d'un ensemble homogène à un ensemble hétérogène. Mieux, cette contrainte physique - l'augmentation nécessaire de l'entropie - fournit la raison de la séparation des chromosomes, caractère essentiel de la réplication de l'ADN, sans recours à la pré-existence d'une machinerie compliquée. Notons que ce processus demande cependant que ces molécules se trouvent dans un espace non symétrique (dans le modèle exploré, un cylindre). On n'aura donc, dans la construction d'une bactérie synthétique,

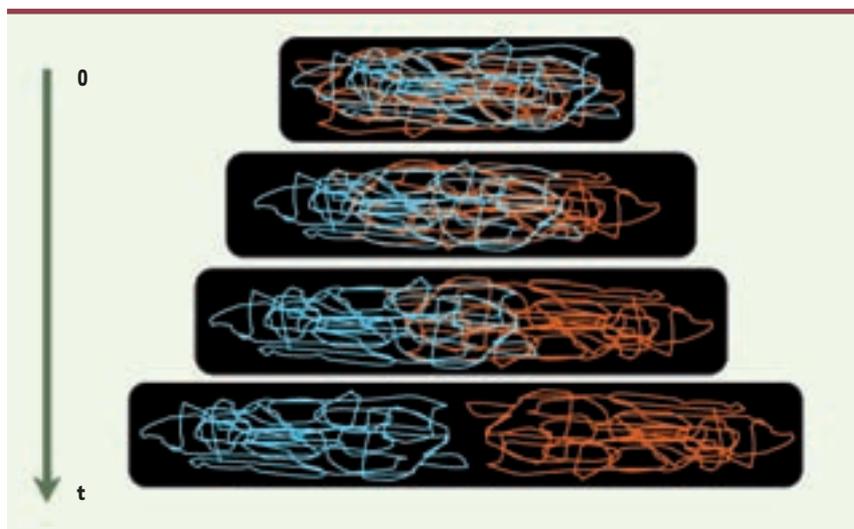


Figure 4. Un accroissement de l'entropie est suffisant pour entraîner la ségrégation des chromosomes.

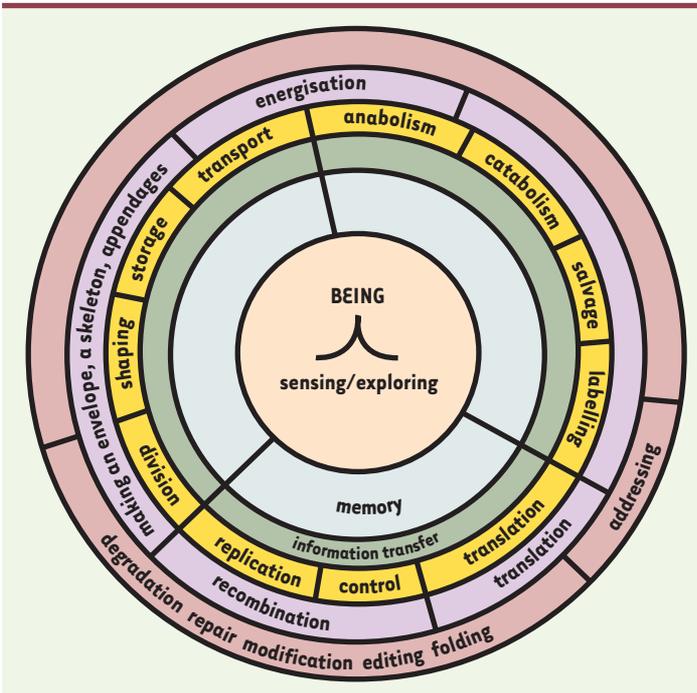


Figure 5. Les fonctions essentielles à la base de la vie.

qu'à se soucier de ne pas commencer par la faire sphérique, mais plutôt, comme le sont d'ailleurs la plupart des bactéries, en forme de bacille cylindrique.

À ce stade nous voyons donc qu'il est possible de construire un programme de recherche visant à la construction d'une bactérie synthétique, à partir de l'ensemble des fonctions du paléome, du regroupement des gènes correspondants selon un ordre connu, et en plaçant l'ADN qui porte ces gènes dans une cellule dont on peut imaginer la construction indépendante, à partir de réalisations chimiques *in vitro*.

### Conclusion provisoire : construire une bactérie synthétique

Une quantité d'autres contraintes, en particulier pour la réalisation de la machinerie cellulaire, sont à considérer, et nous sommes certainement encore assez loin du compte - c'est ce qui explique pour l'instant que le génie génétique se soit contenté de la reprogrammation de certaines cellules pour en diriger la production (production de métabolites, de protéines humaines, ou simplement de biomasse utilisable comme nourriture). Pourtant il semble clair que nous comprenons assez de la structure et de la dynamique du vivant pour envisager la construction d'une bactérie synthétique. La métaphore alphabétique et l'hypothèse que les cellules se comportent comme des machines de Turing (très bruitées il est vrai) semblent tenables. L'organisation du génome en l'association d'un paléome héritier de l'évolution depuis l'origine de la vie, et un cénome spécifique de l'occupation d'une niche, semble bien se prêter à la distinction entre

un couple répliqueur/constructeur et producteur, ce dernier devant jouer le rôle finalisé qu'ont les artefacts humains. On remarque que dans cette approche il est raisonnable de « nettoyer » les séquences génétiques du bruit introduit par l'évolution, en caractérisant les régions du programme qui définissent les processus algorithmiques de l'expression des gènes, un peu à la manière de ce qui est fait en génie logiciel. Cela a déjà été réalisé, avec succès, dans la synthèse de virus « simplifiés » [16]. Tout à l'opposé, on cherche à partir de cellules dont le génome est déjà très petit, à le diminuer encore [17]. Cet effort permet de ne retenir du paléome qu'une partie des gènes, en prévoyant de fournir à l'environnement de la cellule synthétique un certain nombre de composés essentiels (qui seront synthétisés chimiquement) et de coder dans le génome des transporteurs versatiles qui seront capables de les faire entrer dans la cellule. Ainsi la combinaison de l'imitation de la nature, et de l'ingénierie va très probablement permettre, d'ici deux décennies, de réaliser une première bactérie synthétique.

Il convient aussi de se rendre compte que nous n'avons parlé que de la construction d'une cellule synthétique, pas de la pérennité de cette construction. Or, la vie nous apprend que les cellules vieillissent irrémédiablement et disparaissent, tout comme les espèces. Ce que nous avons décrit ici correspond au fait de vivre, pas à l'évitement de la mort. Ce deuxième point, passionnant, demanderait toute une autre réflexion. On sait déjà que les constructions de biologie synthétique sont instables : le bactériophage T7 modifié pour être compris par l'homme, quand on le laisse se reproduire, efface toute la construction humaine et retourne à l'état ancestral, comme le font les espèces domestiquées retournées à l'état sauvage... Et si nous souhaitons réaliser des usines cellulaires il nous faudra pourtant le prendre en compte, ou alors reconstruire l'usine après un petit temps d'usage.

Si cette approche est un succès, nous pourrions envisager une stade ultérieur, où les lois abstraites de la biologie (qui ne spécifient pas explicitement la nature chimique des objets en cause) seront exprimées avec des composés différents de ceux qui constituent la vie. On pourra ainsi probablement modifier la nature du code génétique, et introduire dans les protéines d'autres acides aminés (c'est déjà une réalité, à partir de la sélection traditionnelle de mutants bactériens). On peut même imaginer, pour un futur plus lointain, que le support même de l'hérédité, l'ADN, soit lui-même modifié (c'est déjà le cas chez certains virus). La mise à l'épreuve de ces hypothèses et des modèles correspondants est essentielle, car c'est leur succès

