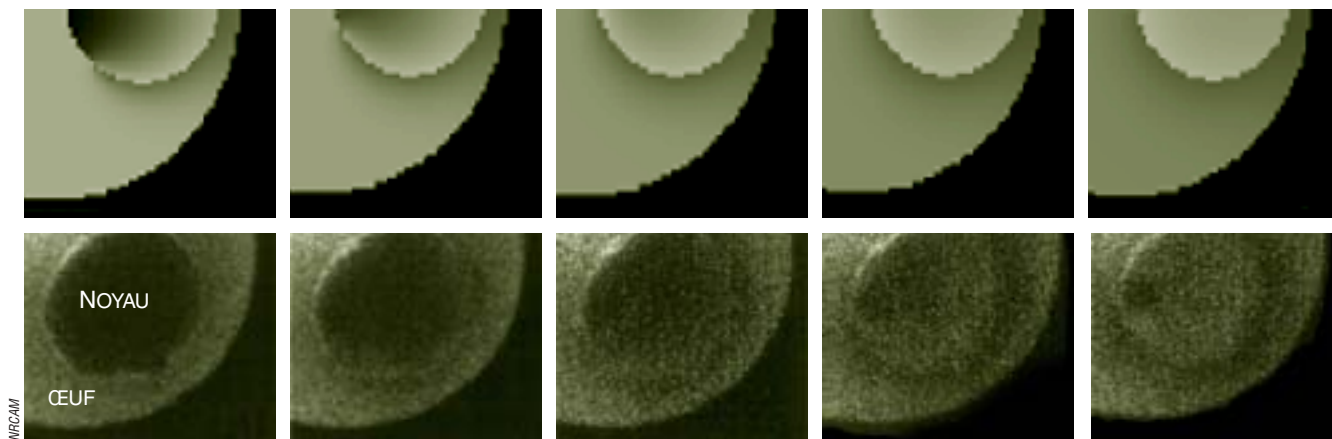




2. DANS UN ŒUF D'ÉTOILE DE MER, les grosses molécules associées à un marqueur fluorescent ne pénètrent pas dans le noyau. Cependant, des simulations informatiques (*en haut*) montrent qu'au moment de la mitose, la division cellulaire, la membrane nucléaire

s'ouvre et laisse passer les molécules fluorescentes : elles entrent dans le noyau, dont la couleur s'éclaircit. Cette modélisation est confirmée par l'observation expérimentale (*en bas*). Le grossissement est égal à 300.

20 000 fois par seconde. Le métabolisme de la bactérie *Escherichia coli* (une bactérie qui colonise les intestins), dont le cycle de 20 minutes requiert entre 10^{14} et 10^{16} réactions, est modélisable sur un ordinateur multiprocesseur.

Jusqu'à présent, les modèles en biologie ont été limités, parce que la puissance des ordinateurs disponibles était insuffisante et que, pour certains systèmes biologiques, les connaissances sont encore trop parcellaires : par exemple, on ne connaît que 66 pour cent des protéines du bactériophage lambda, un virus de bactérie. Par ailleurs, on néglige souvent l'influence du milieu intracellulaire sur les réactions biochimiques des cellules et sur le comportement des molécules. Par exemple, le fonctionnement de la rhodopsine, une protéine qui intervient dans la vision, varie en fonction de la température : des études ont montré que les crapauds voient mieux quand il fait froid.

Outre ces différents facteurs, les interactions moléculaires, innombrables, ne sont pas toutes connues. Par exemple, une même protéine kinase existe sous plus de 25 666 formes différentes selon le nombre de groupes phosphate fixés et selon qu'elle est associée ou non à une autre kinase.

Lorsque le nombre de paramètres requis est atteint, la représentation virtuelle d'un système biologique devient possible et permet d'améliorer notre compréhension du système. Des cycles de modélisations informatiques, de prévisions et de tests expérimentaux accélèrent le rythme des hypothèses formulées et testées.

Les modèles les plus fréquents portent sur la régulation de circuits génétiques, sur le cycle cellulaire (*voir la*

figure 2), sur la transmission des signaux et sur la séquence des réactions au cours de réactions métaboliques. Par exemple, l'équipe de Bernhard Palsson, de l'Université de Californie, a mis au point des modèles des micro-organismes *Escherichia coli*, *Hemophilus influenzae* et *Helicobacter pylori* (présent dans les ulcères de l'estomac). Les biologistes ont tenté d'imposer diverses conditions de croissance à ces micro-organismes et observé la façon dont ils se comportent. Selon eux, leur modèle fournit, non pas une prévision parfaite, mais une approximation acceptable. Ils ont découvert que *Escherichia coli* a un mécanisme de croissance optimal.

Un virus informatique modèle

Certains biologistes tentent de modéliser des systèmes plus complexes ; ainsi, Drew Endy, de l'Institut des sciences moléculaires de Berkeley, a simulé l'infection de la bactérie *Escherichia coli* par le bactériophage T7.

Le bactériophage T7, l'un des organismes les mieux connus, ressemble à un module d'alunissage. Une enveloppe icosaédrique (à 20 côtés) qui renferme l'information génétique du virus (sous forme d'ADN linéaire double brin) repose sur un tube court, trapu et creux, terminé par de longs appendices articulés (*voir la figure 4*). Grâce à ses appendices, le virus s'agrippe à la paroi externe de la bactérie par l'intermédiaire de récepteurs, puis introduit son ADN dans la bactérie. Ce phénomène, nommé translocation, dure une dizaine de minutes. Le matériel génétique injecté détourne à son profit la machinerie cellulaire de la bactérie qui produit en

série des clones du bactériophage (une centaine) jusqu'à ce qu'elle éclate.

Le génome du bactériophage T7 contient 39 937 paires de bases, lesquelles constituent 56 gènes qui codent 59 protéines. Les gènes sont réunis en trois classes selon leur nature et leur période d'expression pendant le cycle du virus. Ceux de la classe 1 sont des gènes précoces : premiers injectés dans la bactérie, ils sont exprimés dès le début de l'infection. Leurs promoteurs sont reconnus par les enzymes de la bactérie ; les promoteurs sont les séquences géniques situées en amont des gènes, qui indiquent aux enzymes (les ARN polymérases) l'endroit où elles doivent commencer la transcription de l'ADN en ARN messager. Les ARN messagers sont traduits en protéines virales par la machinerie cellulaire de la bactérie. De plus, parmi les premières protéines synthétisées figure l'ARN polymérase spécifique du bactériophage, plus rapide, qui prend la place de l'ARN polymérase bactérienne.

Les gènes de la classe 2 participent à la réplication de l'ADN du bactériophage et ceux de la classe 3 codent des protéines qui fabriqueront l'enveloppe des clones du bactériophage.

L'ordre d'agencement des gènes le long du génome du bactériophage T7 influe sur leur ordre d'entrée dans la bactérie et donc sur le déroulement de l'infection.

Dans la simulation, le génome est divisé en 73 segments numérotés, chacun représentant une partie du génome, par exemple un gène ou un promoteur. L'ordre des nombres que traite l'ordinateur reflète l'ordre des gènes. À partir de paramètres déterminés expérimentalement, telles l'affinité des

promoteurs pour l'ARN polymérase, la force des interactions protéines-protéines, et les constantes cinétiques, le modèle prévoit les concentrations du bactériophage en ARN messagers, en protéines et en diverses macromolécules.

Trois étapes ont été prises en compte : la translocation, la transcription et la traduction. Par exemple, la vitesse de translocation varie de 40 à 200 paires de bases par seconde selon que l'ARN polymérase provient de la bactérie ou du virus.

Les biologistes ont modélisé les effets d'une réorganisation du génome du bactériophage, c'est-à-dire le développement d'un bactériophage où l'ordre des gènes aurait été modifié. Les simulations ont porté, d'une part, sur trois «mutants», où le gène 1, codant l'ARN polymérase, avait été déplacé et, d'autre part, sur 10^5 mutants dont les génomes étaient des permutations aléatoires des 73 gènes.

Ils ont d'abord simulé la croissance des mutants du premier type et mesuré leur taux de croissance, c'est-à-dire la durée pendant laquelle la population de bactériophages double : les simulations informatiques corroboraient les résultats expérimentaux obtenus dans des conditions équivalentes. Dans un deuxième temps, ils ont déterminé les profils de synthèse des protéines de chacun des trois mutants. Il s'agissait de déterminer à quel moment et en quelle quantité les protéines sont fabriquées. Ces profils concordent avec les expériences (voir la figure 3). Enfin, sur les 10^5 mutants au génome aléatoire, les modélisations montrent que l'organisation naturelle du génome viral est quasi optimale : pour 50 pour cent des génomes testés, les bactériophages ne se développaient pas, tandis que dans 2,8 pour cent des simulations, les bactériophages croissaient plus rapidement que le bactériophage initial. Ainsi, plus de 97 pour cent des génomes artificiels testés étaient moins performants que le génome «naturel». Avec de telles simulations, les biologistes obtiennent des informations inaccessibles par l'expérimentation.

Bien sûr, les comportements prédits par ces modèles dépendent beaucoup des valeurs imposées aux paramètres, mais les simulations permettent d'étudier beaucoup plus de situations que ne l'autorise l'expérimentation, et de bouleverser certaines idées reçues. Ainsi, certains biologistes ont récemment remis en cause le dogme selon lequel un gène est transcrit en un ARN, lequel est traduit en une protéine qui

remplit une fonction particulière. Selon eux, ce dogme, sur lequel est fondée toute l'industrie pharmaceutique, n'explique pas que quelque 30 000 gènes (seulement) puissent régir le comportement complexe des milliards de cellules d'un organisme humain. La modélisation informatique devrait faire tomber le dogme et expliquer l'origine de la complexité issue d'un système apparemment simple.

Un métabolisme informatique

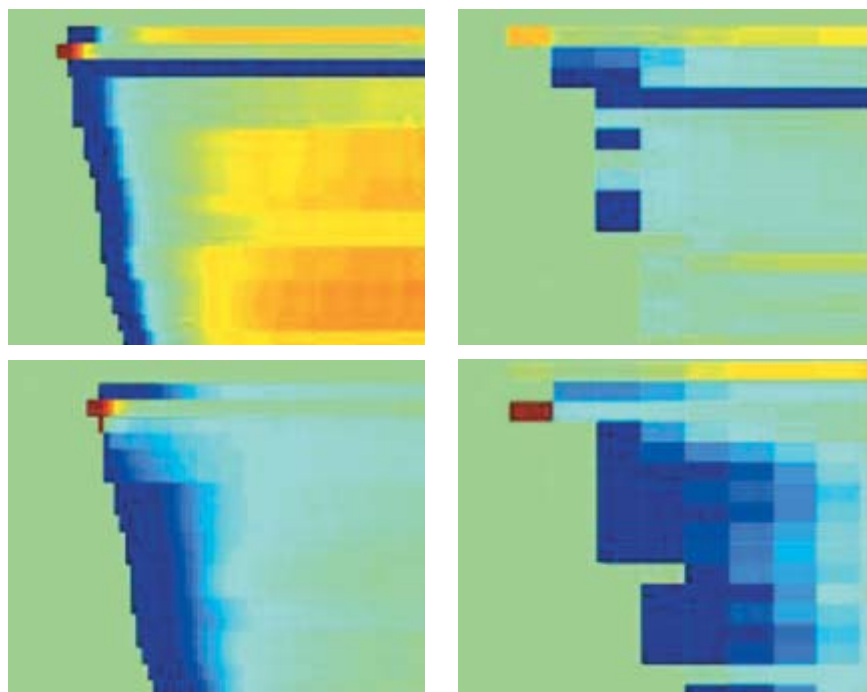
Les simulations cellulaires les plus élaborées doivent refléter la robustesse des véritables cellules : quelle que soit leur forme, elles ont à faire face à des variations notables de température, à des modifications de l'apport d'éléments nutritifs, à des agressions de composés chimiques toxiques et à des attaques venues de l'intérieur comme de l'extérieur. Pour survivre et se multiplier, les cellules doivent résister à ces perturbations.

Masaru Tomita, de l'Université Keio, à Fujisawa, a constaté, lors des simulations qu'il a effectuées sur son modèle nommé *e-Cell*, que ce dernier acquiert une remarquable capacité de résistance aux modifications de son environnement. Son équipe a construit une

cellule virtuelle contenant 127 «gènes», dont on connaît le comportement et les protéines qu'ils codent. La plupart étaient des gènes de *Mycoplasma genitalium*, un micro-organisme unicellulaire qui a le plus petit génome découvert à ce jour. Cette simulation devait indiquer le nombre minimal de gènes nécessaires à un organisme autosuffisant. Toutefois, l'équipe japonaise n'a pas obtenu le résultat recherché, car la modification drastique de l'expression de divers gènes du modèle perturbait à peine le comportement de *e-Cell*.

On en déduit que le comportement d'une cellule, par exemple en réponse à un médicament, est peu modifié par l'augmentation ou par la diminution de l'activité d'un gène particulier, ou par le blocage de telle ou telle protéine ; en revanche, ce comportement dépend de la façon dont tous les gènes et toutes les protéines interagissent.

John Koza, de l'Université Stanford, a utilisé la programmation génétique pour recréer une petite partie seulement du modèle *e-Cell*. Cette méthode crée des logiciels performants à partir de logiciels modifiés («mutés») aléatoirement ; on sélectionne ensuite les logiciels ainsi engendrés qui accomplissent le mieux la tâche désirée. J. Koza modélise ainsi un ensemble de réactions enzymatiques



3. SIMULATION DE LA SYNTHÈSE PROTÉIQUE du bactériophage T7 (un virus de bactérie) où le gène 1 (celui de l'ARN polymérase) a été déplacé (dans le gène 1, 7, en haut ; dans le gène 3, 8, au centre ; dans une région intergénique, en bas). Les modélisations (à gauche) sont en accord avec les résultats expérimentaux (à droite). À chaque ligne correspond une protéine. La couleur reflète le rapport de la quantité de protéine fabriquée par le mutant et par le bactériophage naturel (du bleu, où la synthèse est 100 fois inférieure, au rouge, où elle est 10 fois supérieure).



Slim Films

4. LES BACTÉRIOPHAGES (des virus à bactéries) sont les organismes dont on connaît le mieux les molécules qui les composent : ils sont idéaux pour des modélisations.

connues qui transforme deux molécules d'acide gras et une de glycérol en une molécule de diacylglycérol. Parmi les multiples «circuits» biologiques créés par le programme informatique, ceux qui correspondaient le plus précisément au «fonctionnement» du modèle *e-Cell* étaient retenus pour qu'ils poursuivent leur évolution, tandis que les autres étaient éliminés.

Au bout d'une journée, l'ordinateur a mis au point un programme qui correspondait à l'ensemble des réactions chimiques réelles mises en jeu lors de la synthèse du diacylglycérol. Il fait intervenir quatre enzymes, cinq molécules intermédiaires et des boucles de rétroaction. Par ailleurs, les vitesses de réaction des enzymes «virtuelles» sont les mêmes que celles des enzymes réelles. Aucune autre configuration ne fonctionnait aussi bien.

La programmation génétique aidera peut-être les biologistes à simuler des métabolismes plus complexes, par exemple la transformation de la nourriture en énergie, la croissance et la production de déchets. Cela ne sera possible que dans les cas où les biochimistes connaissent le déroulement, au cours du temps, des réactions chimiques dans les cellules.

le gène *engrailed* code une protéine qui influe sur l'expression d'un autre gène. Cette protéine engrailed entraîne l'expression du gène *hedgehog*, et la protéine hedgehog synthétisée diffuse alors vers la partie antérieure où les cellules expriment le gène *cubitus interruptus*. Dans la partie postérieure, ce gène est réprimé par la protéine engrailed. Ainsi, la protéine cubitus interruptus, qui réprime l'expression du gène *hedgehog* dans la partie antérieure, n'est fabriquée que dans la partie antérieure. Au total, la partie antérieure contient la protéine cubitus interruptus, tandis que la partie postérieure contient la protéine engrailed. Ces deux protéines, en limitant deux territoires, marquent l'axe antéro-postérieur.

Dans le disque imaginal, près de la frontière entre la partie antérieure et la partie postérieure, la protéine hedgehog entraîne l'expression du gène *wingless* dans la partie ventrale de la partie antérieure et du gène *decapentaplegic* dans la partie dorsale de la partie antérieure. La protéine wingless réprime l'expression du gène *decapentaplegic*, tandis que la protéine decapentaplegic réprime le gène *wingless*. Ainsi, deux territoires marquent l'axe dorso-ventral.

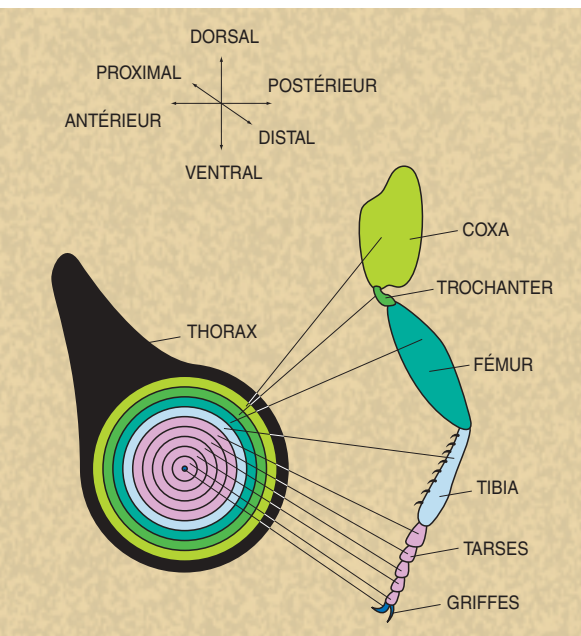
L'axe proximal-distal indique quelle partie du disque imaginal deviendra telle partie de la patte. Il est déterminé par des gènes, tels *aristaless*, *distal-less* et *dachshund*, exprimés sous forme d'anneaux concentriques. Comment ces anneaux se forment-ils? Plusieurs modèles ont été proposés. En 1976, V. French a proposé que les protéines wingless et decapentaplegic entraînent la fabrication d'un facteur dans le disque où elles sont toutes deux présentes. Selon ce modèle, dit polaire, ce facteur diffuse à partir du centre du disque imaginal vers l'extérieur : la différenciation d'un disque en tel ou tel organe dépend de la concentration en ce facteur. D'autres scénarios ont été proposés, mais n'ont pas été validés.

En 1997, Thomas Lecuit et S. Cohen ont montré que les deux protéines wingless et decapentaplegic, les deux protéines qui sont déjà responsables de l'axe dorso-ventral, diffusent dans les cellules autour de la zone où elles sont fabriquées. Dans la région circulaire centrale, les protéines wingless et decapentaplegic sont toutes deux présentes en forte concentration, et activent l'expression du gène *distal-less*. Par ailleurs, aux endroits où ces protéines

Des pattes de mouche

Les systèmes pluricellulaires n'échappent pas aux modélisations. Ainsi, l'équipe de Hiroaki Kitano, de la Corporation japonaise pour la science et la technologie, a modélisé le développement embryonnaire d'une patte de drosophile. Les mécanismes de développement de cet insecte et de l'homme se ressemblent étonnamment.

Pendant l'embryogenèse, l'épithélium se replie et forme plusieurs structures, nommées disques imaginaires, qui donneront naissance aux appendices, telles les ailes ou les pattes. Une patte se développe selon trois axes (voir la figure 5) : l'axe antéro-postérieur (de la tête à l'abdomen), l'axe dorso-ventral et l'axe proximal-distal (de la jonction avec le thorax à l'extrémité de l'appendice). Contrairement à l'axe proximal-distal, les axes antéro-postérieur et dorso-ventral sont déterminés précocement. La mise en place de ces trois axes résulte de l'expression de plusieurs gènes et de la diffusion – ou non – des protéines qu'ils codent. Ces protéines activent ou répriment d'autres gènes. Dans la partie postérieure du disque,



5. LE DISQUE IMAGINAL d'une patte de drosophile se développe selon trois axes : l'axe antéro-postérieur, l'axe dorso-ventral et l'axe proximal-distal. Ce dernier axe résulte de la fragmentation du disque imaginal en anneaux concentriques caractérisés par les gènes qui s'y expriment. Chaque anneau donne naissance à une partie distincte de la patte. Les trois axes sont déterminés par l'expression de huit gènes principaux dont les produits activent ou répriment l'expression d'autres gènes.