

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

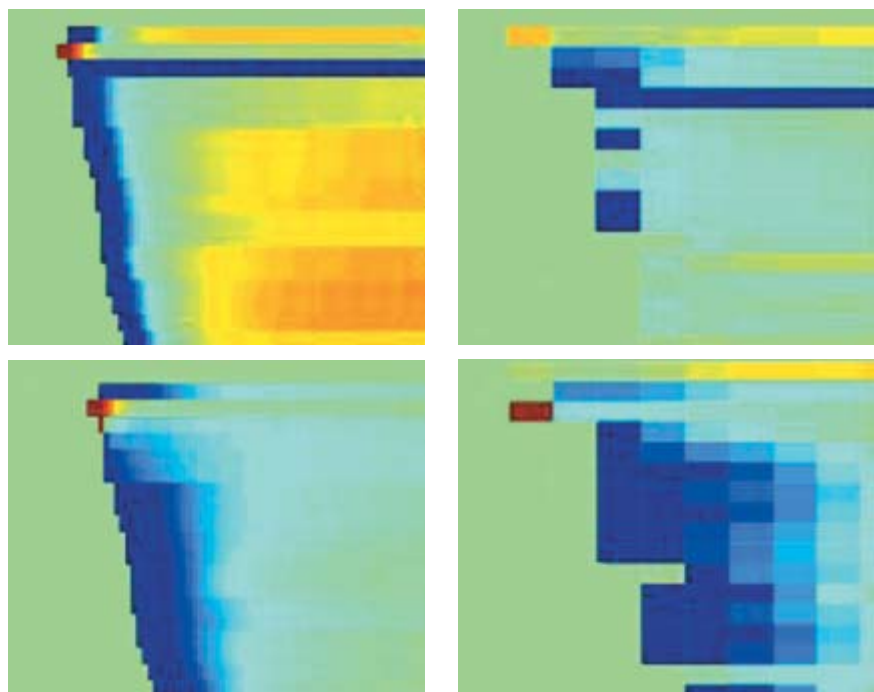
Les simulations cellulaires les plus élaborées doivent refléter la robustesse des véritables cellules : quelle que soit leur forme, elles ont à faire face à des variations notables de température, à des modifications de l'apport d'éléments nutritifs, à des agressions de composés chimiques toxiques et 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

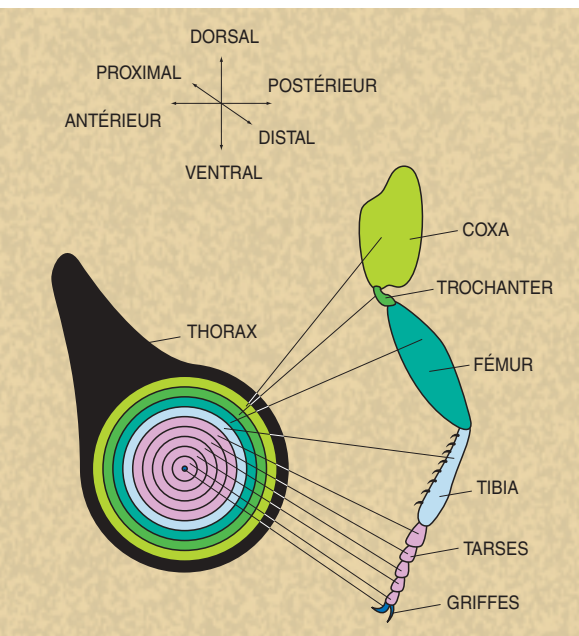


3. SIMULATION DE LA SYNTHÈSE PROTÉIQUE du bactériophage T7 (un virus de bactérie) où le gène 1 (celui de l'ARN polymérase) a été déplacé (dans le gène 1, 7, en haut ; dans le gène 3, 8, au centre ; dans une région intergénique, en bas). Les modélisations (à gauche) sont en accord avec les résultats expérimentaux (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.



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.

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.

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 imaginaux, 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,

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

sont en faible concentration, elles déclenchent l'expression du gène *dachshund*, sans l'intermédiaire de la protéine distal-less. De la même façon, l'expression du gène *aristaless* est sous le contrôle des protéines wingless et decapentaplegic quand celles-ci ont une concentration supérieure à celle que requiert l'expression du gène *distal-less*. Ces données sur les interactions de gènes et de protéines du développement ont été introduites dans le modèle de H. Kitano, qui a révélé comment se forment les anneaux d'expression, c'est-à-dire l'axe proximal-distal, à partir des seules protéines wingless et decapentaplegic.

Une mouche modèle

Un disque imaginal de drosophile contient environ 10 000 cellules et a une forme d'ellipse. Le disque virtuel est circulaire et contient environ 1 200 cellules, soit environ une cellule virtuelle pour huit cellules réelles. Le modèle prend en compte les huit gènes cités et leurs interactions (*engrailed*, *hedgehog*, *cubitus interruptus*, *wingless*, *decapentaplegic*, *aristaless*, *distal-less* et *dachshund*). On ignorait si la protéine cubitus interruptus a une action activatrice ou inhibitrice sur le gène *decapentaplegic* : le modèle a permis de tester les deux hypothèses et de trancher en faveur d'une action répressive de la protéine cubitus interruptus sur ce gène (voir la figure 7).

Le développement embryonnaire a été divisé en plusieurs étapes : la transcription, la traduction, la diffusion des protéines et leur dégradation. Pour la transcription, les paramètres consistaient en une quantité seuil d'activateur au-delà de laquelle le gène est transcrit. Chaque protéine, qui est soit un activateur, soit un répresseur, est caractérisée par son affinité pour un site promoteur. Plus l'affinité est élevée, plus la protéine se lie au promoteur et plus son action est importante. Les expressions mathématiques utilisées dans le modèle tiennent également compte de la compétition que peuvent se livrer un activateur et un répresseur face à un même gène.

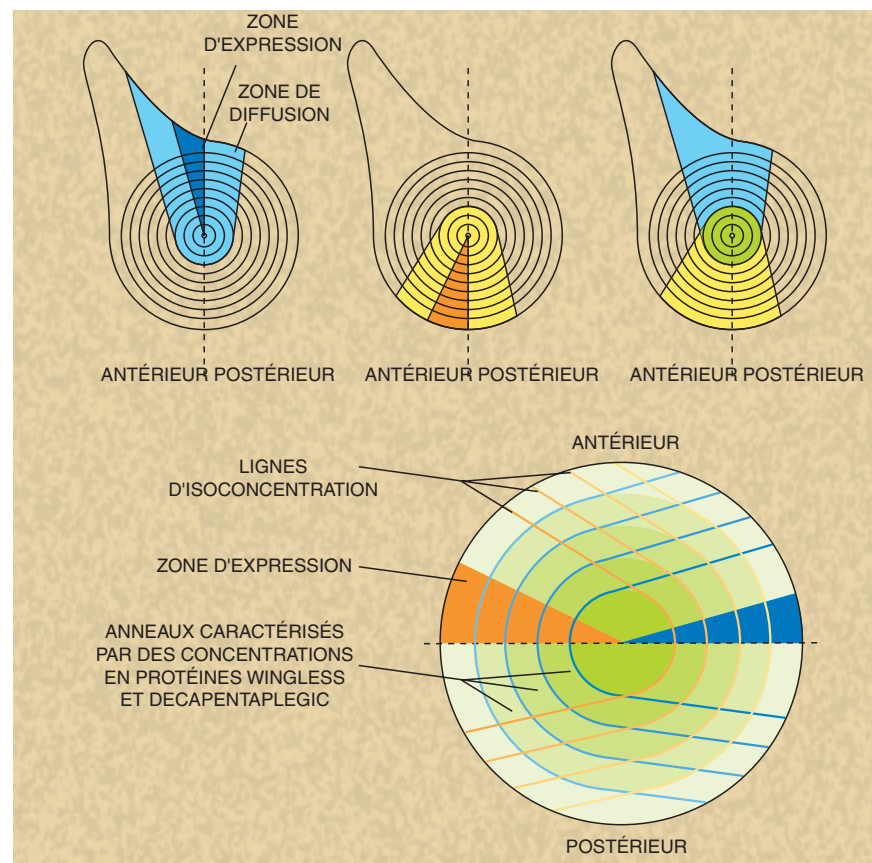
Pour la traduction, la quantité de protéines est identique à celle d'ARN messagers transcrits. Pour chaque protéine (selon qu'elle diffuse ou non), une expression mathématique indiquait sa concentration en chaque endroit du disque imaginal. Enfin, la quantité de protéines diminuait progressivement

pour simuler la dégradation des protéines par méthylation.

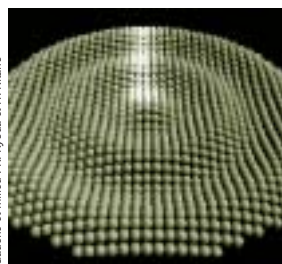
Dans le programme informatique, chaque cellule virtuelle du système contient une liste de protéines et leur quantité. Dans un premier temps, le système vérifie si une protéine peut être sécrétée. Quand la protéine l'est, une autre partie du programme calcule la quantité de protéines qui est synthétisée à partir des gènes selon les concentrations en activateurs et en

répresseurs. La quantité diffusée est calculée par une autre partie du programme qui tient compte de la concentration en cette protéine dans les cellules voisines.

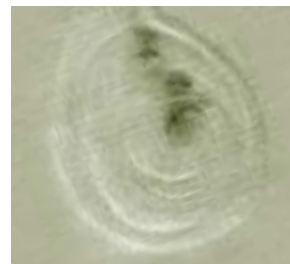
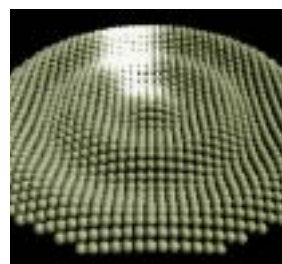
Les résultats des simulations obtenus dans les motifs d'expression des gènes ou de la synthèse des protéines sont en accord avec les résultats expérimentaux (voir la figure 8). Le modèle a aussi montré que l'hypothèse selon laquelle la protéine cubitus interruptus



6. LES ANNEAUX CONCENTRIQUES qui déterminent l'axe proximal-distal de la patte de drosophile sont contrôlés par la diffusion des protéines decapentaplegic et wingless. Chacune diffuse de sa zone d'expression (respectivement en bleu foncé et en orange), et l'on peut tracer des zones d'isococoncentration pour chacune d'elles (respectivement les lignes bleues et orange). Les intersections de ces lignes d'isococoncentration délimitent des zones circulaires (en vert) à l'origine de la formation des anneaux concentriques : selon les concentrations en protéines decapentaplegic et wingless, les gènes activés diffèrent. Deux anneaux distincts deviendront deux parties distinctes de la patte.

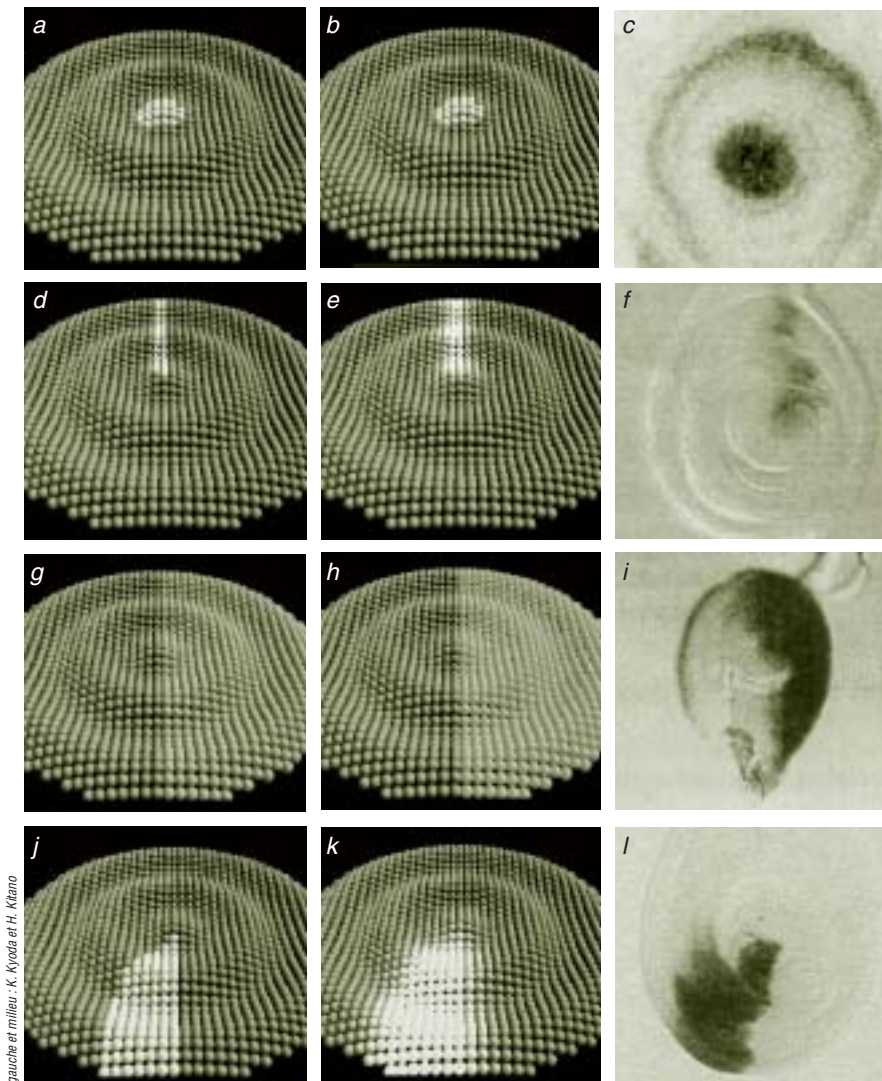


Gauche et milieu : K. Kyôda et H. Kitano



D. Brook et S. Cohen

7. L'ACTION DE LA PROTÉINE CUBITUS INTERRUPTUS sur le gène *decapentaplegic* n'a pas été mise en évidence expérimentalement. Un modèle du développement de la patte de drosophile a permis de tester les deux hypothèses (à gauche, la protéine active le gène, au centre, elle le réprime) et ainsi de trancher en comparant le profil d'expression du gène *decapentaplegic* avec des données expérimentales (à droite) : la protéine réprime l'expression du gène *decapentaplegic*.



gauche et milieu : K. Kyoda et H. Kitano

A droite, de haut en bas : G. Campbell et al., W. Brook et S. Cohen, S. Eaton et T. Kornberg et M. Zecca et al.

8. LA MODÉLISATION DU DISQUE IMAGINAL d'une patte de drosophile permet de caractériser la zone d'expression de divers gènes (à gauche, en blanc) et la zone où les protéines sont synthétisées (au centre, en blanc). Les simulations concordent avec les résultats obtenus expérimentalement (à droite ; les zones noires sont celles où la protéine est présente). Les gènes étudiés sont, de haut en bas, *aristaless* (a, b et c), *decapentaplegic* (d, e et f), *engrailed* (g, h et i) et *wingless* (j, k et l).

est un répresseur du gène *decapentaplegic* concorde avec les résultats expérimentaux.

Des séries de simulations ont montré que l'axe proximal-distal résulte directement de la diffusion des produits des gènes *wingless* et *decapentaplegic*. L'équipe de H. Kitano a ainsi échafaudé un nouveau modèle, dit bipolaire, dans lequel la diffusion des protéines *wingless* et *decapentaplegic* agit sur la segmentation du disque en anneaux : ces protéines diffusent à partir de leur zone de synthèse et sont de moins en moins concentrées à mesure qu'on s'éloigne. À partir des zones de synthèse, on peut tracer des lignes d'isoconcentration. Les intersections des lignes d'isoconcentration en protéine *wingless* et des lignes d'isocon-

centration en protéine *decapentaplegic* forment une série de cercles concentriques : l'ensemble des lignes d'isoconcentration découpe le disque imaginal en anneaux ; chacun donnera naissance à une partie de patte.

Wayt GIBBS est journaliste à la revue *Scientific American*.

Koji KYODA et Hiroaki KITANO, *Simulation of genetic interaction for drosophila leg formation*, in *Proceedings of Pacific Symposium on Biocomputing*, 1999.

Drew ENDY, Lingchong YOU, John YIN et Ian MOLINEUX, *Computation, prediction and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes*, in *Proceedings of National Academy of Science*, vol. 97, n° 10, pp. 5375-5380, 2000.

Ainsi, le modèle, fondé sur des résultats expérimentaux, a permis de trancher parmi plusieurs hypothèses, d'une part, sur le rôle de la protéine *cubitus interruptus* et, d'autre part, sur la formation de l'axe proximal-distal. Rappelons que le premier objectif de ces modèles biologiques n'est pas d'obtenir une prévision parfaite, mais plutôt une approximation fiable de la réalité.

La biologie virtuelle sera de plus en plus utilisée, notamment pour la mise au point, dans un proche avenir, des modèles de fonctionnement de différents types cellulaires, par exemple les cellules du cœur ou encore les lymphocytes B, des acteurs essentiels du système immunitaire.

D'ici une dizaine d'années, ces modèles constitueront un incroyable moteur pour la découverte de nouveaux médicaments. On pourra modéliser les conséquences cellulaires d'une maladie et, ensuite, observer comment agit un médicament ou une molécule en cours d'évaluation. Nous disposerons de modèles quantitatifs du fonctionnement d'une cellule, d'un organe, voire d'un organisme entier.

Aujourd'hui, la biologie semble être dans la situation de l'astronomie du XVI^e siècle. Les astronomes disposaient alors d'innombrables données sur les positions des objets célestes, mais ils ne pouvaient pas prédire avec précision le mouvement des planètes. Ils n'avaient pas encore compris que les orbites sont elliptiques et décrites par une équation simple... jusqu'à ce que Kepler le démontre.

Il n'existe sans doute pas une unique équation simple qui expliquerait toute la biologie cellulaire, mais les biologistes recherchent des principes unificateurs afin d'agencer les données et de parvenir à une compréhension globale des systèmes. La prochaine révolution biologique sera-t-elle informatique ?

Jeremy EDUAROS, Rafael IBARRA et Bernhard PALSSON, *In silico prediction of Escherichia coli metabolic capabilities are consistent with experimental data*, in *Nature Biotechnology*, vol. 19, pp. 125-130, 2001.

Jean-Arcady MEYER et Agnès GUILLOT, *La robotique évolutionniste*, in *Pour la Science*, n° 284, juin 2001.

Drew ENDY et Roger BRENT, *Modeling Cellular Behaviour*, in *Nature*, vol. 409, pp. 391-395, 18 janvier 2001.

Le 15 janvier 2009,
découvrez le **nouveau site**
de **Pour la Science !**
www.pourlascience.fr

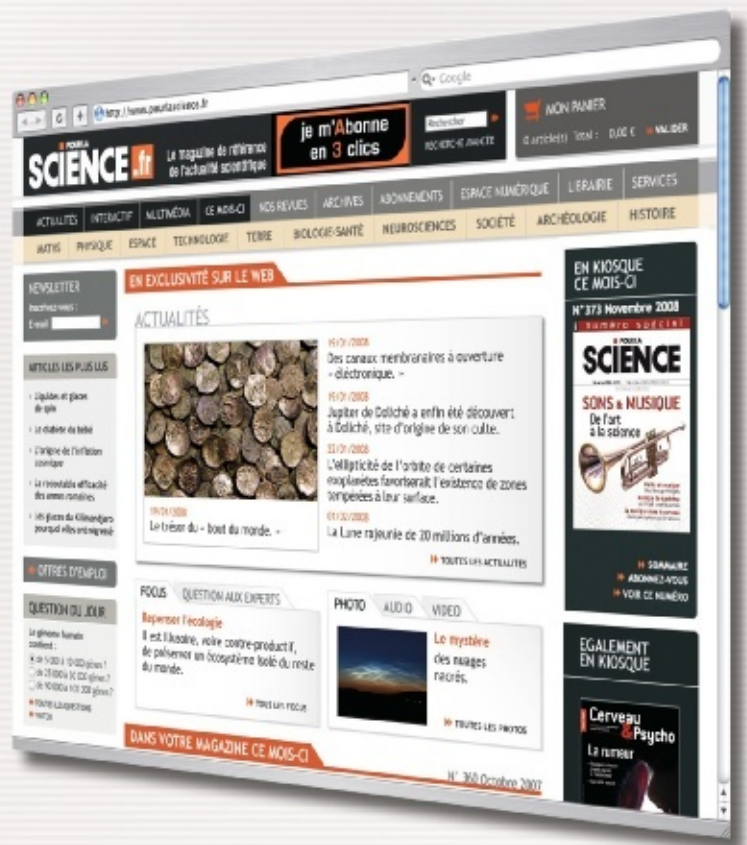
Avec son nouveau site Internet, Pour la Science vous offre plus de contenu, plus de services, plus d'interactivité.

➤ Vous êtes abonné à Pour la Science ?
Retrouvez l'intégralité de votre magazine en ligne et téléchargez les articles de votre choix. Réagissez aux articles et posez vos questions à des experts. Découvrez chaque jour des contenus inédits.

➤ Vous n'êtes pas encore abonné à Pour la Science ?
Découvrez chaque jour des actualités scientifiques, des articles originaux, des photos, des vidéos...
Achetez les articles de votre choix.

➤ Abonnez-vous en ligne en trois clics !
L'abonnement « papier » vous donne accès au magazine imprimé et à tous les contenus web.
L'abonnement « web » vous donne accès à votre magazine au format numérique.

➤ Et bientôt l'accès en intégralité aux archives !



En cadeau de bienvenue pour le lancement de www.pourlascience.fr,
et jusqu'au 25 janvier 2009, abonnez-vous en ligne au tarif exceptionnel
de 41 € pour l'abonnement web, et de 48 € pour l'abonnement
au magazine imprimé + web.

À bientôt sur **POUR LA SCIENCE.fr**