

La biologie virtuelle

WAYT GIBBS

L'informatique vole au secours de la biologie pour traiter la masse considérable des informations, pour modéliser des systèmes biologiques et pour prévoir leur comportement.

Les ingénieurs en aéronautique connaissent les propriétés d'un futur avion avant même la construction du premier prototype, car ils disposent de simulations fiables. Les physiciens des particules prévoient certains résultats d'expériences grâce à la théorie du modèle standard. Les prévisions météorologiques reposent sur des modèles de l'atmosphère terrestre. Les exemples abondent, sauf dans un domaine : la biologie.

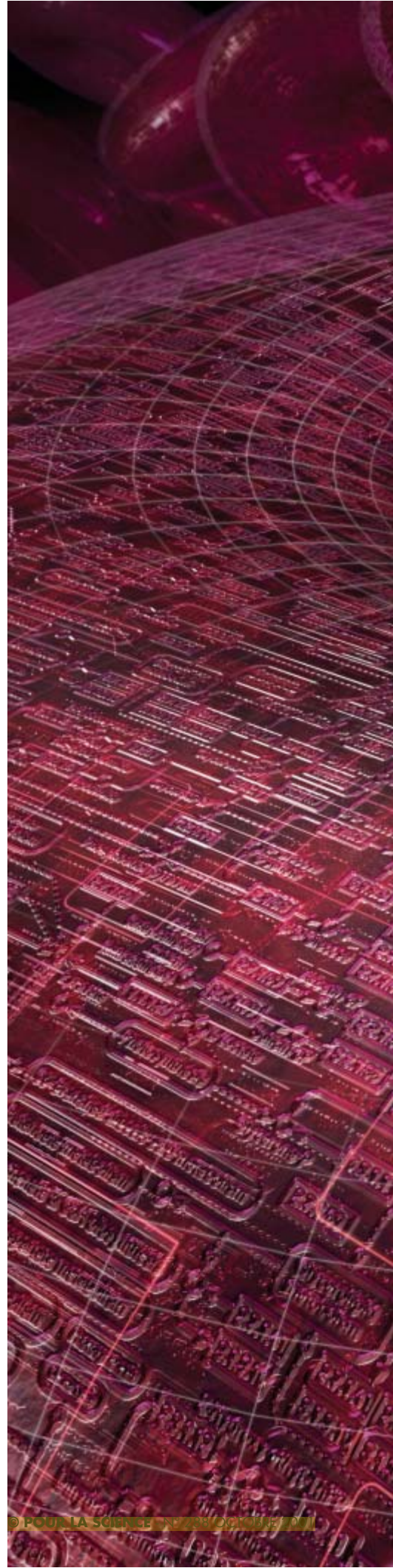
Depuis 50 ans, le réductionnisme prévaut : la plupart des organismes vivants ont été décortiqués, disséqués en éléments toujours plus petits – des organismes aux organes, des tissus aux cellules, des cellules aux molécules, des chromosomes à l'ADN et aux gènes. On connaît tous – ou presque tous – les éléments constitutifs de la vie, mais les données se sont accumulées et elles constituent aujourd'hui une énorme masse d'informations, surtout depuis le séquençage du génome humain, il y a quelques mois. Que faire d'une telle liste de composés et de leurs propriétés spécifiques ? Peut-on comprendre le fonctionnement d'une cellule à partir de celui de ses constituants ? Dans un mouvement inverse d'intégration, l'informatique et les simulations tentent de rassembler toutes les pièces du puzzle biochimique en un ensemble cohérent. Grâce à la biologie virtuelle, nommée aussi biologie *in silico*, on accéderait ainsi à des propriétés émergentes, encore inconnues, et l'on vérifierait l'adage : «Le tout est plus que la somme des parties.»

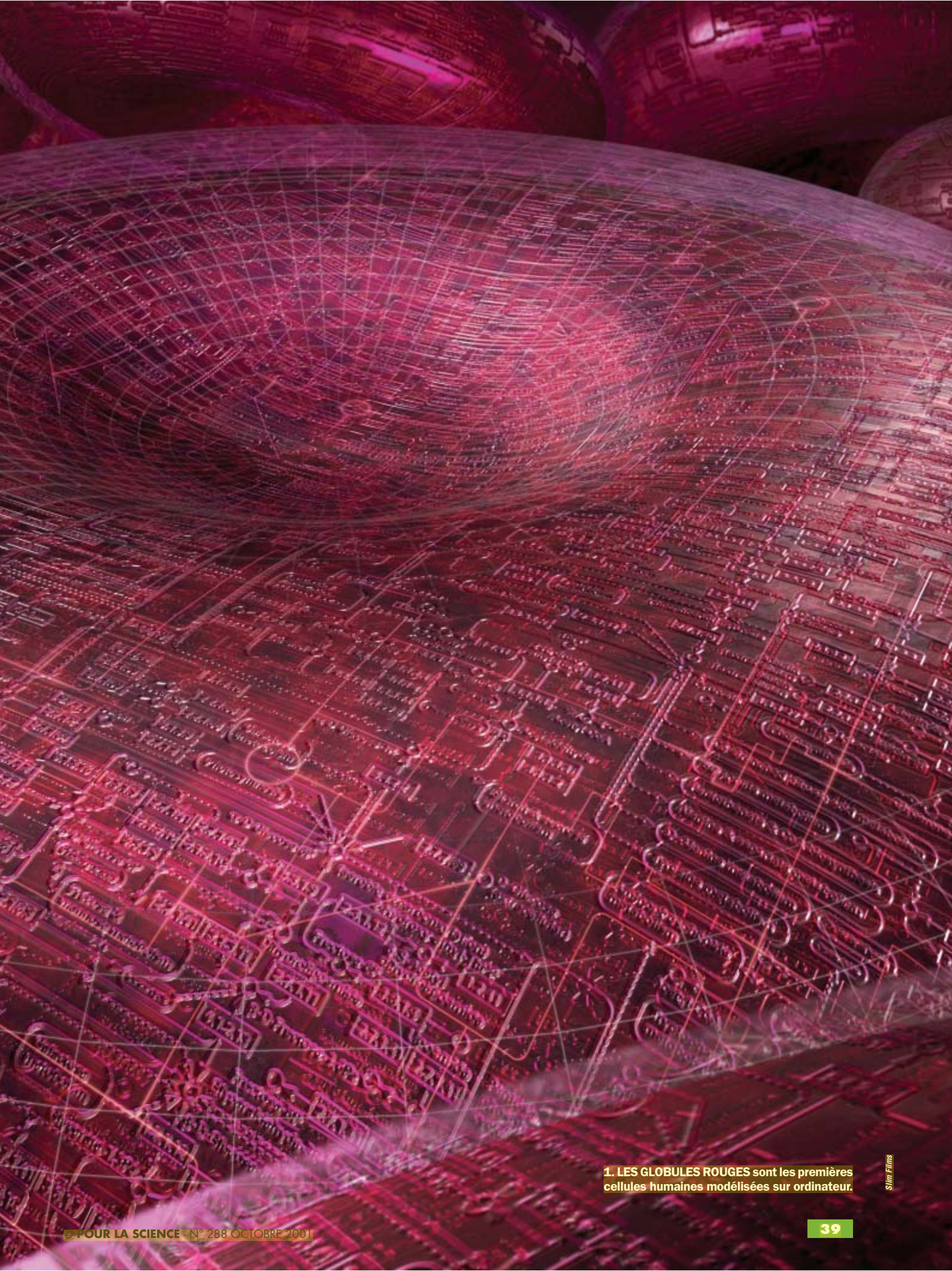
Toutefois, bien que prometteuse, la modélisation en biologie n'en est qu'à ses balbutiements : il ne suffit pas d'introduire dans un ordinateur toutes les informations disponibles sur un être humain pour dénouer les mystères de la médecine. Cependant, des simulations ont déjà fait la preuve de leur utilité. Aussi, après avoir examiné les contraintes que rencontrent les biologistes dans leurs modélisations, nous exposerons quelques-uns de leurs succès, tels le décodage des mécanismes d'infection d'une bactérie par un virus et la génétique du développement d'une patte de drosophile.

Un terrain miné

Par manque de données, les premiers modèles biologiques étaient éloignés de la réalité. Par ailleurs, ils étaient seulement qualitatifs, c'est-à-dire fondés sur les relations qui unissent les molécules entre elles, par exemple, comment telle molécule interagit avec telle autre pour en donner une troisième.

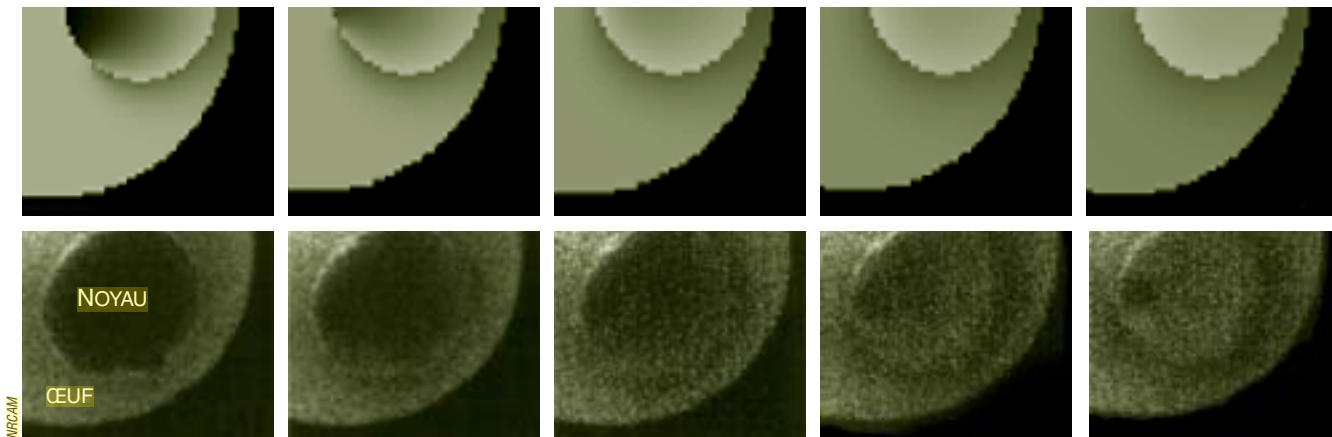
Dans les années 1990, des méthodes mathématiques nouvelles ont été mises au point. Elles ont permis de prendre en compte des paramètres quantitatifs, tels la cinétique des réactions ou les coefficients de diffusion. Parallèlement, la puissance des ordinateurs a augmenté et, aujourd'hui, on peut exécuter des modélisations complexes sur un ordinateur de bureau. Par exemple, avec un microprocesseur courant de 800 mégahertz, un algorithme peut simuler 10^{10} réactions par jour, soit 100 minutes de la vie d'un système biologique, où 100 réactions ont lieu environ





1. LES GLOBULES ROUGES sont les premières cellules humaines modélisées sur ordinateur.

Slim Films



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