

Enseigner le numérique à l'école, une nécessité

par Thierry Vieville, sur le blog *Intelligence mécanique*

Dans un article publié récemment sur le blog Internetactu (<http://bit.ly/1n58dj0>), son rédacteur en chef Hubert Guillaud s'insurge contre l'enseignement de l'informatique à l'école et, si l'on suit son raisonnement jusqu'au bout, contre l'intérêt même de faire comprendre à chaque enfant les fondements scientifiques des technologies numériques.

L'argument principal est que « aussi omniprésente que soit une technologie, nous n'avons pas besoin de savoir comment elle fonctionne ». C'est effectivement l'argument des multinationales du secteur numérique, qui nous expliquent que leurs produits sont si faciles à utiliser qu'il suffit de les consommer. Ce raisonnement n'est sans doute pas étranger au fait que la France est aujourd'hui complètement dominée dans le secteur mondialisé du numérique.

Mais cela va plus loin. Ne pas vouloir comprendre comment le numérique fonctionne est une régression intellectuelle. Dès lors qu'il leur serait refusé de s'initier aux fondements scientifiques des technologies numériques, les jeunes ne pourraient en construire que des représentations « mythiques ».

Or que se passerait-il dans une société sans éducation au numérique ? Le savoir numérique ne serait que de la mémorisation. Sans comprendre les lois sous-jacentes, on ne peut rien inférer, mais seulement réciter. Le savoir-faire numérique ne serait que de la répétition : sans comprendre comment cela fonctionne, on ne peut pas innover. Et le savoir-être numérique serait limité à la consommation. Gageons que, comme dans d'autres domaines, le vide laissé par le manque de connaissances sur le numérique serait comblé par des « maîtres à penser », qui sauront construire un discours rassurant.

Sans éducation au numérique, nos enfants n'auront pas la possibilité d'être

acteurs de la construction de leur environnement numérique, et donc de la société qui sera la leur demain. Ils devront se contenter de subir. La liberté numérique finit là où l'ignorance informatique débute.

Mais quels sont au juste les arguments avancés contre l'enseignement du numérique à l'école ? Premièrement, au motif que l'informatique requiert une nouvelle façon d'enseigner – le tableau noir et les livres ne sont pas suffisants ! –, il faudrait surtout s'intéresser aux nouvelles façons d'enseigner, sans s'encombrer d'informatique. Cet argument suppose que le résultat de l'éducation est secondaire et que seule compte la méthode

La liberté numérique commence là où l'ignorance informatique finit.

pédagogique. Pour ma part, peu m'importe la méthode, je fais confiance aux femmes et aux hommes dont c'est le métier d'enseigner pour former les jeunes générations.

Par ailleurs, c'est oublier que l'apprentissage de l'informatique fait l'objet d'études pluridisciplinaires précises depuis plusieurs années, tant en France (citons le laboratoire Sciences, techniques, éducation, formation de l'École normale supérieure de Cachan) qu'à l'international. Reposer sans cesse la question de ce « que serait un enseignement de l'informatique dans le cadre de l'école, et pour apprendre quoi » est en fait obsolète : on le sait, et on est déjà en train de l'enseigner, en terminale S en France ainsi que dans d'autres pays européens.

Second argument : il y aurait « d'autres façons de soutenir l'apprentissage de l'informatique que de passer par la création d'un nouveau contenu disciplinaire avec des enseignants spécialisés ». Certes. Mais pourquoi rejeter d'emblée un outil existant qui garantit à tous un enseignement gratuit et de qualité, l'Éducation nationale ?

Réduire les excellentes initiatives extrascolaires d'éducation au numérique (*Magic Makers*, *École 42*, *Simplon.co*, etc.) à des alternatives à l'éducation nationale, c'est se tromper d'un facteur 1 000. Ces initiatives touchent quelques milliers de jeunes de 13 à 18 ans, alors que plusieurs millions sont concernés !

La question n'est pas de savoir comment éduquer « des jeunes » au numérique, mais comment éduquer des millions de jeunes au numérique. Et refuser la mise en place d'un enseignement généralisé, c'est proposer que seuls des privilégiés bénéficient de cette éducation, par des activités périscolaires ou dans les établissements favorisés.

L'apprentissage pour tous de la science informatique est indispensable. Mais est-il réalisable ? Des arguments tels que « c'est trop cher », « les professeurs n'y arriveront pas » ou « cela se fera au détriment d'autre chose » ne tiennent pas devant les retours d'expérience : l'enseignement de la science informatique a été introduit depuis 2012 comme spécialité en classe de terminale S pour un coût minime et avec succès. Et ce sans empiéter sur les autres spécialités. Il ne reste plus qu'à le généraliser progressivement, au rythme des moyens disponibles et de la formation continue des enseignants. ■



Thierry VIEVILLE est directeur de recherche en neurosciences computationnelles à l'Inria et auteur du blog *Intelligence mécanique* <http://www.scilogs.fr/intelligence-mecanique>

Retrouvez tous nos blogueurs sur www.scilogs.fr et suivez-les sur les réseaux sociaux.



DOSSIER

Réinventer le vivant

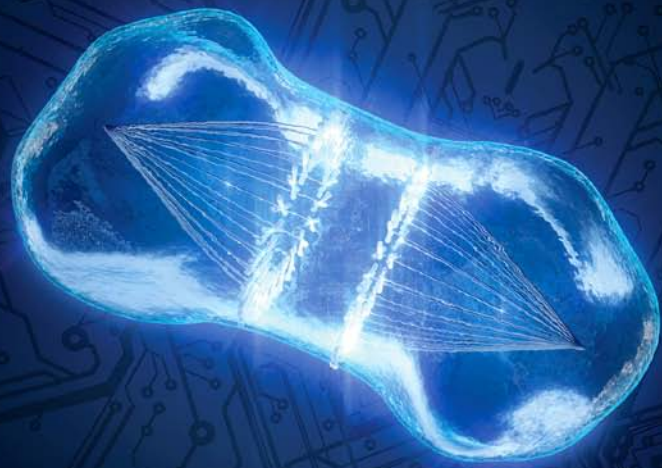
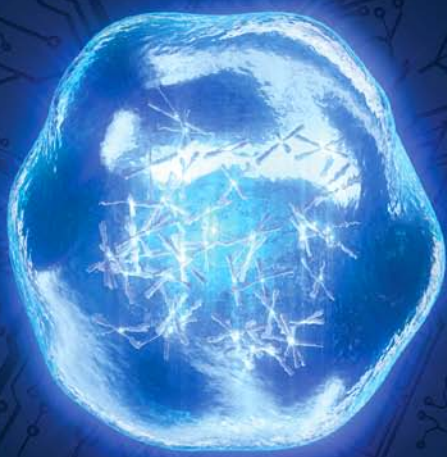
Quels enjeux pour la biologie de synthèse?

Fabriquer des formes minimales de vie, explorer de nouveaux codes génétiques, modifier génétiquement des organismes pour qu'ils produisent de nouvelles fonctions, concevoir et construire un génome synthétique viable... Ces dernières années, les projets de recherche renouvelant les approches de l'origine de la vie et de la compréhension du vivant se multiplient, suscitant enthousiasme et débats.

Depuis les années 1990, avec l'avènement du séquençage massif et de la bio-informatique, la recherche en biologie a amorcé un tournant. De nouvelles disciplines ont émergé. La biologie des systèmes utilise l'ensemble des données collectées sur un organisme, un organe ou une cellule afin de modéliser le fonctionnement de systèmes vivants dans leurs moindres détails. La biologie de synthèse, avec l'aide de la biologie des systèmes et de l'ingénierie, vise à concevoir des systèmes complexes fondés sur le vivant ou inspirés du vivant, et dotés de fonctions absentes dans la nature.

Que permet aujourd'hui la biologie de synthèse? Quels sont ses objectifs, ses applications à court et moyen termes, ses limites? Quels sont ses impacts sur la société? Telles sont les questions qu'abordent les trois articles de ce dossier, à travers un exemple de biologie des systèmes (voir l'article de Markus Covert, page 20), une présentation de la biologie de synthèse (voir l'article de François Képès, page 28) et une réflexion sur la façon dont les sciences sociales pourraient accompagner cette discipline émergente (voir l'article de Pierre-Benoit Joly et Benjamin Rimbault, page 36). Comprendre ces nouvelles approches de la biologie est une étape indispensable pour mesurer les enjeux et les risques potentiels associés.

– Marie-Neige Cordonnier



Simuler une cellule viv

Markus Covert

Des biologistes ont conçu un modèle numérique d'un organisme unicellulaire, reproduisant son cycle de division. Cet outil éclaire d'un jour nouveau les mécanismes du vivant.

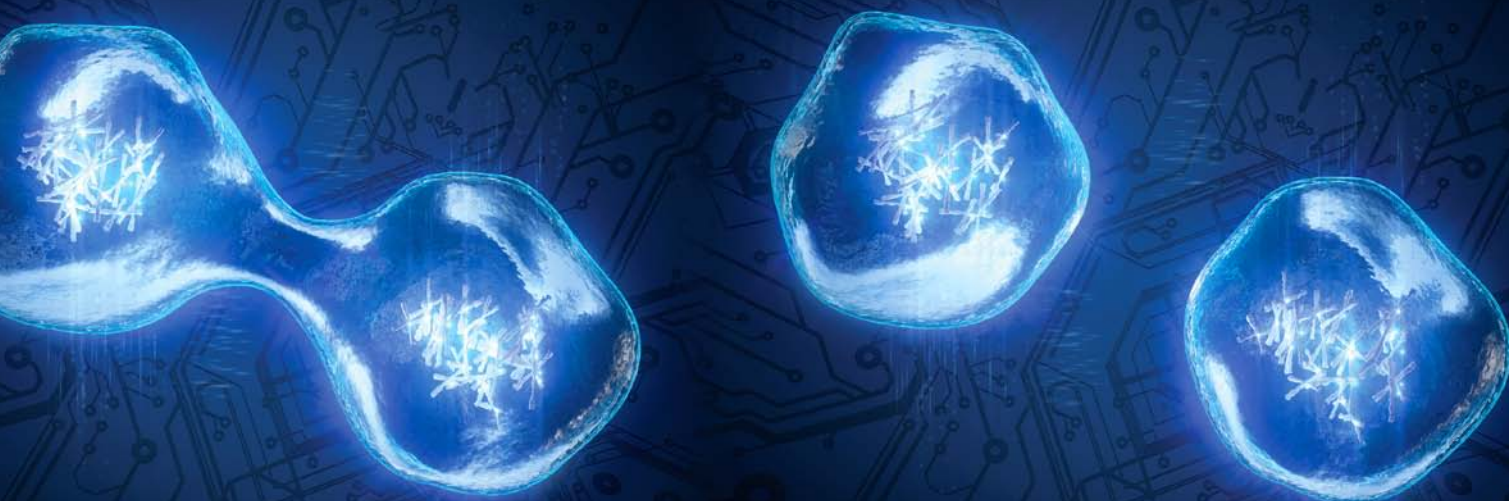
L'idée m'est venue un soir de février 2008, alors que je rentrais du travail à vélo. Je ressassais un problème qui me préoccupait, ainsi que d'autres, depuis plus d'une décennie : existait-il un moyen de simuler la vie – avec toute la complexité de la biochimie qui la rend possible – à l'aide d'un logiciel ?

Un modèle numérique de cellules vivantes, même approximatif et imprécis, serait un outil fort utile. Grâce à lui, les biologistes testeraient des idées d'expériences avant d'engager du temps et de l'argent à les mettre en pratique en laboratoire. Les concepteurs de médicaments accéléreraient la recherche de nouveaux antibiotiques en testant virtuellement leur efficacité. Les bio-ingénieurs transplanteraient et reconnecteraient les gènes de micro-organismes virtuels afin de développer des souches modifiées présentant des caractéristiques données, telles que la capacité d'émettre un rayonnement en cas d'infection par certains virus ou d'extraire

de l'hydrogène gazeux du pétrole. Ces essais virtuels permettraient de limiter les risques associés à la modification de micro-organismes réels. À terme, avec des modèles assez élaborés pour simuler les cellules humaines, ces outils transformeraient la recherche médicale en offrant la possibilité de réaliser des études aujourd'hui inenvisageables parce que nécessitant des types de cellules humaines que l'on ne sait pas cultiver en laboratoire.

Une piste pour modéliser la vie ?

Toutefois, sans un moyen concret de démêler l'écheveau de réactions chimiques et d'interactions physiques qui se déroulent dans les cellules, ce rêve semblait inaccessible. De nombreuses tentatives, dans mon laboratoire et ailleurs, s'étaient heurtées à divers obstacles. Mais ce soir-là, en pédalant, j'ai repensé à des enregistrements vidéo de cellules vivantes isolées que j'avais



vante

LE CYCLE D'UNE CELLULE (vue d'artiste).

Le contenu de la cellule est dupliqué, puis la cellule se resserre en son milieu jusqu'à former deux cellules filles identiques.

effectués récemment. Et la solution m'est apparue : pour construire un simulateur réaliste fonctionnel, il fallait prendre comme modèle un microbe unicellulaire unique, choisi parmi les plus simples. Limité à une seule cellule, le problème serait en principe assez simplifié pour permettre d'inclure tous les processus biologiques connus de cette cellule : le déroulement de chaque boucle d'ADN, la transcription de chaque message de cet ADN en une copie sous forme d'ARN, la fabrication de chaque protéine obtenue à partir de ces instructions, et les interactions de tous ces acteurs et bien d'autres, tout cela participant à la croissance de la cellule et à sa division en deux cellules filles. En somme, la simulation reproduirait, à partir de considérations générales, toutes les péripéties d'une vie unicellulaire.

Dans les tentatives passées, on avait toujours essayé de simuler une colonie de cellules, car presque toutes les données sur le comportement cellulaire avaient été

L'ESSENTIEL

- Un modèle numérique reproduisant le fonctionnement d'une cellule offrirait une nouvelle compréhension du vivant.
- Une simulation complète d'une bactérie infectieuse commune a été réalisée l'année dernière.
- Les bio-informaticiens élaborent à présent des modèles pour des organismes plus complexes. Leur objectif à long terme : simuler une cellule humaine.

André Kulschauer

recueillies à partir de populations cellulaires. Mais les avancées en biotechnologie et en informatique ont facilité l'étude de cellules isolées. Aussi disposons-nous désormais des outils nécessaires pour une autre approche.

À peine arrivé chez moi, j'ai esquisé les plans d'un simulateur d'une cellule simple, la bactérie *Mycoplasma genitalium*. Le lendemain matin, je me suis attelé à l'écriture du code de quelques-uns des nombreux processus qui ont lieu au sein d'un micro-organisme vivant. En moins d'une semaine, j'avais achevé plusieurs modules prototypes, chacun représentant un processus cellulaire particulier. Les modules produisaient des résultats qui semblaient tout à fait réalistes.

J'ai montré mes travaux à quelques biologistes. La plupart ont pensé que j'avais perdu la tête, mais j'ai persisté dans mon idée. Deux doctorants, Jonathan Karr et Jayodita Sanghvi, se sont joints au projet. Réaliser un tel modèle nécessitait

de concevoir des dizaines de modules à partir des données scientifiques. Concrètement, cela signifiait passer au peigne fin les données biochimiques de près d'un millier d'articles scientifiques, puis d'utiliser les valeurs collectées pour contraindre et ajuster des milliers de paramètres, tels que l'affinité des enzymes avec leurs cibles ou la fréquence de liaison et de décrochage des protéines qui « lisent » l'ADN. Même avec l'aide de collaborateurs, le projet prendrait des années ! Mais j'avais bon espoir.

Les 400 gènes essentiels du mycoplasme

Nous nous sommes inspirés des premières recherches visant à modéliser la vie. En 1984, Harold Morowitz, alors à l'Université Yale, avait ouvert la voie. Pour lui, il fallait commencer par modéliser les bactéries les plus simples que l'on savait cultiver, les mycoplasmes. Minuscules et relativement

simples, ces bactéries ont un intérêt particulier : deux espèces déclenchent des maladies humaines, le parasite à transmission sexuelle *M. genitalium*, qui colonise les voies urinaires et génitales, et le germe *Mycoplasma pneumoniae*, qui peut entraîner une pneumonie atypique. Un modèle de l'une ou l'autre espèce aurait un intérêt tant biologique que médical.

La première étape, selon H. Morowitz, devait être le séquençage du génome du microbe choisi. En 1995, J. Craig Venter et ses collègues de l'Institut pour la recherche génomique ont mené cette tâche à bien pour *M. genitalium*, qui ne compte que 525 gènes (une cellule humaine en a plus de 20000).

Quatre ans plus tard (j'étais alors étudiant en thèse à San Diego), l'équipe de J. C. Venter concluait que seuls environ 400 de ces gènes sont essentiels à la vie de cette espèce (tant que les mycoplasmes sont cultivés dans un milieu riche). J. C. Venter et

ses collègues ont alors synthétisé les gènes essentiels d'une espèce de mycoplasme et montré qu'ils étaient fonctionnels dans une cellule.

Pour moi et d'autres jeunes biologistes de la fin des années 1990, ce groupe était comme Led Zeppelin : des personnalités iconoclastes, remarquables, jouant une musique que personne n'avait entendue jusque-là. Clyde Hutchinson, un biologiste du groupe, a prédit que le test ultime de notre compréhension des cellules simples serait la modélisation informatique de l'une d'elles : on peut construire une cellule fonctionnelle en laboratoire en combinant des éléments sans comprendre dans le détail comment ils s'assemblent. Ce n'est pas le cas en informatique.

H. Morowitz aussi avait appelé de ses vœux la construction d'un simulateur de cellule fondé sur les données du génome de *M. genitalium*. Selon lui, toute expérience réalisable en laboratoire l'est aussi sur un

Les grandes étapes de la modélisation cellulaire

Le long parcours de l'auteur pour aboutir à la première modélisation d'une cellule unique d'une simple bactérie, *Mycoplasma genitalium*, a été nourri des travaux théoriques, des études génétiques et des essais de modélisation d'autres chercheurs, dont voici les principaux. Concevoir un modèle numérique de cellule humaine sera encore plus difficile : une cellule humaine contient presque 40 fois plus de gènes qu'une bactérie, regroupés en chromosomes dont la structure et les informations contenues sont bien plus complexes. Quelques étapes essentielles pour y parvenir sont proposées à la fin de cette chronologie.

1967

Francis Crick et Sydney Brenner proposent le « Projet K : la solution complète d'*Escherichia coli* », une initiative visant à dresser le « patron » de cette bactérie digestive commune, y compris des détails fins de sa génétique, de ses processus énergétiques et de sa reproduction.

1984

Harold Morowitz, alors à l'Université Yale, présente les grandes lignes d'un plan de séquençage, puis de modélisation d'un mycoplasme.

1984

Une équipe menée par Michael Schuler de l'Université Cornell présente un modèle informatique utilisant des équations différentielles pour traduire la plupart des processus biologiques majeurs intervenant dans la croissance d'une cellule *E. coli*. Le modèle n'est cependant pas en mesure d'inclure l'activité des gènes, car le génome d'*E. coli* n'a pas encore été séquencé.

1989-1990

Bernhard Palsson de l'Université du Michigan publie un modèle complet du métabolisme d'un

globule rouge humain qui inclut les effets des variations de l'acidité et de l'hypoglycémie.

1995

J. Craig Venter, de l'Institut TIGR, et ses collègues achèvent le séquençage du génome de *M. genitalium*.

1999

Masaru Tomita et ses collaborateurs de l'Université Keio au Japon construisent *E-Cell*, un système de modélisation de cellule fondé sur des équations différentielles et incluant

127 gènes, la plupart issus de *M. genitalium*.

2002

L'Alliance pour la signalisation cellulaire, une vaste collaboration d'une cinquantaine de chercheurs, lance un projet ambitieux évalué à dix millions de dollars sur dix ans : modéliser certaines cellules de la souris (les lymphocytes B du système immunitaire et les cellules du muscle cardiaque). Le projet produit des données intéressantes, mais se heurte aux difficultés de culture des lymphocytes B.

2002

B. Palsson, George Church de l'Université Harvard, l'auteur et d'autres conçoivent un modèle à l'échelle génomique du métabolisme d'*Helicobacter pylori*, une bactérie qui infecte l'homme et peut entraîner des ulcères gastriques et un cancer de l'estomac.

ordinateur. Le degré de correspondance entre les résultats expérimentaux et les simulations est une mesure de la validité de la biologie moléculaire actuelle, c'est-à-dire de la théorie expliquant comment l'ADN et les autres biomolécules de la cellule interagissent et donnent naissance à la vie telle que nous la connaissons. En d'autres termes, à mesure que nous assemblons le puzzle de la vie apparaissent de façon plus claire les pièces et les interactions qui manquent à notre théorie.

Les séquenceurs à haut débit et les équipements robotiques des laboratoires ont accéléré la recherche des pièces manquantes. Toutefois, les masses récoltées de séquences d'ADN et d'informations sur l'activité des gènes n'étaient pas (et ne sont toujours pas) accompagnées d'explications sur la façon dont ces activités se combinent. C'est en partie pour cela que, malgré l'annonce régulière de la découverte de nouveaux gènes associés au cancer, à

l'obésité ou au diabète, les traitements curatifs de ces maladies restent hors d'atteinte. Seule la compréhension des nombreuses interactions moléculaires à l'origine de ces maladies conduira à ces traitements.

Les pionniers de la modélisation cellulaire ont compris que les simulations de cellules entières, avec leurs composants et leur réseau d'interactions, seraient de puissants outils pour comprendre ces données incomplètes et désordonnées. Un simulateur de cellule entière transformerait un jeu complet d'hypothèses sur le fonctionnement d'une cellule en un ensemble d'algorithmes mathématiques rigoureux.

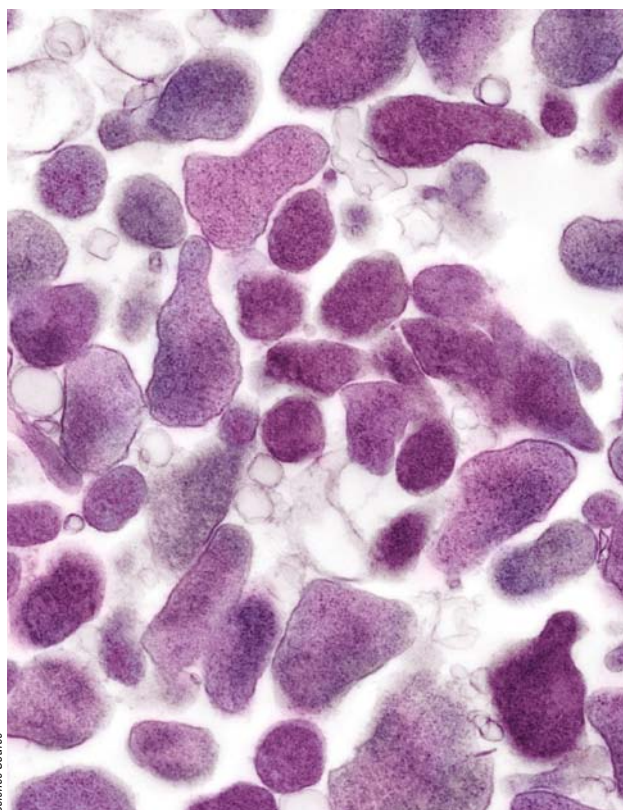
Certes, la route était encore longue : il ne suffisait pas de savoir que tel facteur X régule le gène Y pour traduire ce processus en langage de programmation. Il manquait des précisions. Les programmeurs expriment les processus sous forme d'équations, dont l'une des plus simples est $Y = aX + b$. De telles équations nécessitent d'avancer

■ L'AUTEUR



Markus COVERT est professeur assistant de génie biologique à l'Université Stanford, où

il dirige un laboratoire dédié à la biologie des systèmes.



Science Source

LA BACTÉRIE UNICELLULAIRE *Mycoplasma genitalium* (en violet) est l'un des organismes vivants les plus simples. Cela ne signifie pas pour autant qu'il soit facile de modéliser son cycle de vie...

2004

B. Palsson et l'auteur publient, avec trois autres chercheurs, un modèle informatique des 1010 gènes intervenant dans la régulation du métabolisme et de la transcription de l'ADN d'*E. coli* et montrent que le modèle prédit avec précision les résultats des expériences réalisées sur les bactéries réelles.

2012

L'auteur et ses collègues publient un modèle de cellule entière de *M. genitalium* qui, pour la première fois, simule tous les gènes et processus chimiques connus dans un organisme qui s'autoréplique.

2013

L'auteur et ses collègues montrent que le modèle prédit avec précision l'activité de plusieurs enzymes.

ET APRÈS ?

- Concevoir un modèle de cellule entière pour une bactérie plus classique et mieux étudiée, telle *E. coli*.
- Modéliser un organisme eucaryote unicellulaire tel que la levure *Saccharomyces cerevisiae*. Dans une telle cellule, l'ADN est contenu dans un noyau entouré d'une membrane, et non libre comme dans une bactérie.
- Construire un modèle de cellule animale facile à cultiver en laboratoire, tel un macrophage (un type de cellule immunitaire) de souris.
- Construire un premier modèle de cellule humaine, sans doute, là encore, un macrophage.
- Modéliser d'autres types de cellules humaines, en particulier celles jouant un rôle important dans les maladies communes.

LE PRINCIPE DU SIMULATEUR

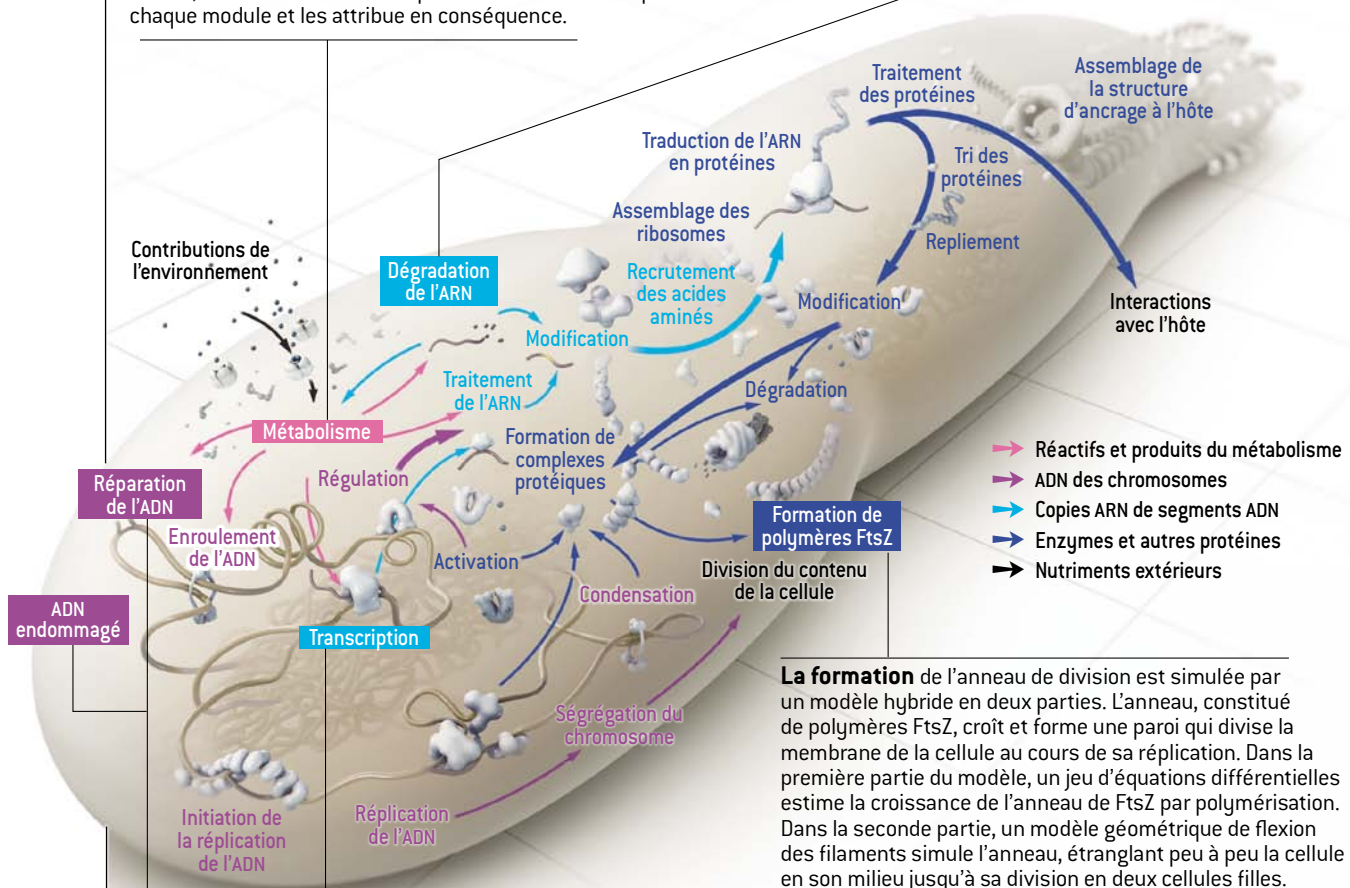
L'auteur a mis au point un modèle numérique de la bactérie infectieuse *Mycoplasma genitalium* qui représente presque tous les aspects de la vie, de la croissance et de la reproduction de ce microbe. Aucune approche mathématique ne permet de simuler toutes les fonctions biologiques de la cellule, si bien que ces fonctions ont été divisées en 28 modules distincts (*indiqués dans la cellule ci-dessous*) qui interviennent dans le traitement de l'ADN (*en violet*), de l'ARN (*en bleu clair*), des protéines (*en bleu foncé*), et de l'énergie, des nutriments et des déchets (*en rose*). Pour chaque module, les chercheurs ont sélectionné la méthode mathématique la plus appropriée (*voir ci-dessous quelques exemples de ces méthodes*).

Au début du programme, tous les modules tournent selon une séquence aléatoire, l'ensemble simulant une seconde de temps réel. Beaucoup des valeurs d'entrée sont tirées d'un grand tableau de variables représentant leurs états initiaux, d'autres sont sélectionnées à partir de fourchettes ou de calculs de probabilités. Les chercheurs peuvent simuler différents scénarios en modifiant la configuration de départ. Après ce premier pas, le programme met à jour le tableau de variables afin de refléter les sorties des modules. La séquence tourne alors pour un nouveau pas d'une seconde, met à jour le tableau, etc. La boucle continue jusqu'à ce que la cellule se divise, meure ou devienne trop vieille pour avoir un sens.

Le métabolisme des nutriments et des déchets est modélisé en utilisant l'analyse du bilan des flux. Cette méthode exploite des techniques de programmation linéaire pour calculer les vitesses de réaction qui optimisent la croissance, la production d'énergie ou toute autre caractéristique choisie par le modélisateur. Elle suppose que les réactions ont lieu assez vite pour obtenir un état stationnaire pendant la seconde correspondant à un pas de simulation. Afin d'empêcher les premiers modules d'utiliser des substances dont d'autres modules ont besoin, le simulateur estime la part de ces ressources qui revient à chaque module et les attribue en conséquence.

La dégradation et le recyclage de l'ARN

et des protéines sont modélisés en utilisant des processus de comptage dits de Poisson. Ces processus utilisent un générateur de nombres aléatoires et des fonctions de probabilité pour décider si un fragment donné d'ARN ou de protéine se dégrade ou persiste jusqu'au pas suivant.



La transcription et la traduction des gènes (étapes au cours desquelles sont fabriquées de nombreuses protéines nécessaires à la croissance et la réplication des cellules) sont simulées par des algorithmes en plusieurs parties incluant des modèles de Markov (ces modèles suivent dans le temps l'état des enzymes qui copient les gènes en ARN), un algorithme probabiliste décrivant la fixation de ces enzymes sur l'ADN et une programmation linéaire pour distribuer l'énergie et autres ressources.

Les dommages subis par l'ADN et les réparations sont modélisés selon cette même méthode non déterministe.

des hypothèses sur les paramètres a et b . Cette exigence de précision dicte les expériences de laboratoire à mener pour combler les lacunes sur les vitesses de réaction et autres quantités caractéristiques du processus simulé.

Mais il paraissait clair aussi qu'une fois les modèles vérifiés, ils se substituaient à certaines expériences, permettant de réserver les travaux coûteux aux questions insolubles par les simulations seules. Les expériences simulées aideraient les biologistes à définir leurs priorités, ce qui accélérerait le rythme des découvertes. Les modèles laissaient entrevoir des outils si séduisants que certains voyaient dans la simulation de cellules entières le grand défi du XXI^e siècle.

Étudiant en thèse, j'ai été impressionné par les premiers résultats de ces pionniers de la modélisation cellulaire (voir l'encadré page 22), et ce défi est devenu mon obsession. Plus tard, quand j'ai monté mon laboratoire et développé des techniques d'imagerie de cellules isolées, ce défi est resté dans mes pensées. Et en ce soir de février, j'ai enfin vu un moyen de le relever.

Comment contraindre et vérifier ses hypothèses ?

Afin de simuler le cycle de vie d'une espèce microbienne avec une précision suffisante pour imiter ses comportements complexes, il fallait résoudre trois problèmes. Tout d'abord, il s'agissait d'encoder sous forme de formules mathématiques et d'algorithmes toutes les fonctions importantes : le flux d'énergie, de nutriments et de produits de réactions dans la cellule (en d'autres termes, le métabolisme), la synthèse et la destruction de l'ADN, de l'ARN et des protéines, et l'activité d'une multitude d'enzymes. Ensuite, il fallait trouver un cadre global pour intégrer toutes ces fonctions. Le dernier problème, enfin, était le plus compliqué à bien des égards : fixer judicieusement les limites haute et basse de chacun des quelque 1700 paramètres du modèle.

J'ai vite compris que nous aurions beau éplucher la littérature sur *M. genitalium* et ses proches parents en quête de ces paramètres (nous avons passé deux ans à extraire les

données d'environ 900 articles), il faudrait, dans certains cas, se contenter de faire des hypothèses ou d'utiliser des résultats d'expériences portant sur d'autres types de bactéries, telle *Escherichia coli*, afin d'obtenir certaines valeurs (par exemple, le temps de séjour moyen des molécules d'ARN dans la cellule avant qu'elles ne soient détruites et recyclées). Or sans un moyen de contraindre et de vérifier ces hypothèses, l'entreprise était vouée à l'échec.

C'est précisément un tel moyen qui m'est apparu ce soir d'hiver : j'ai réalisé que modéliser une cellule unique (et non un groupe de cellules, comme l'avaient

LES MODÈLES LAISSAIENT ENTREVOIR des outils si séduisants que certains voyaient dans la simulation de cellules entières le grand défi du XXI^e siècle.

fait presque toutes les études antérieures) offrait peut-être la situation recherchée. Considérons la croissance et la multiplication des cellules. Une population de cellules croît graduellement ; la naissance ou la mort d'une cellule perturbe peu cette croissance. En revanche, pour une cellule donnée, la division est un événement d'une importance capitale. Avant de se scinder en deux, l'organisme double sa masse, plus précisément les quantités d'ADN, de membrane et de toutes les protéines nécessaires à sa survie. Ainsi, si le modèle se limite à une cellule unique, l'ordinateur pourra compter et suivre chaque molécule sur tout le cycle de vie. Et il pourra vérifier les quantités lors du passage de une à deux cellules.

De plus, une cellule se reproduit à un rythme donné. *M. genitalium*, par exemple, se divise toutes les neuf à dix heures dans un environnement normal de laboratoire. Elle met rarement moins de six heures ou plus de 15. Le fait que le contenu entier de la cellule doive se dupliquer dans ce délai permet de choisir des fourchettes raisonnables pour de nombreuses variables qui, sinon, auraient été indéterminées, telles celles qui contrôlent le moment où commence la réplication de l'ADN.

J'ai réuni une équipe de physiciens, de biologistes, de modélisateurs et même un ancien ingénieur informaticien de Google, et nous avons discuté des approches mathématiques à adopter. Michael Schuler, ingénieur biomédical de l'Université Cornell,

était un pionnier des simulations cellulaires fondées sur les équations différentielles. Bernhard Palsson, qui m'avait encadré lors de mes études à San Diego, avait développé une technique qui modélisait bien le métabolisme. D'autres avaient montré que le hasard est un élément important dans la transcription des gènes, ou que la division des cellules implique un changement géométrique de la membrane cellulaire.

Aucune des méthodes développées ne serait suffisante à elle seule pour modéliser toutes les fonctions d'une cellule. Nous avons donc décidé de créer un modèle de cellule entière constitué de 28 modules distincts, dont chacun utilise l'algorithme qui correspond le mieux au processus biologique concerné et à la connaissance que nous en avons (voir l'encadré page ci-contre). Cette stratégie a conduit à un patchwork de procédures mathématiques. Restait à assembler les morceaux.

J'ai repensé à un cours de conception d'usine chimique que j'avais suivi à l'université. Pour le projet de fin d'année, nous avions utilisé un outil de simulation puissant nommé HYSYS pour ébaucher les plans d'une grande raffinerie. HYSYS permettait d'effectuer chaque réaction chimique principale dans un module séparé et de connecter ces modules. En d'autres termes, il offrait un cadre permettant de réunir de nombreux types d'opérations chimiques en un système ordonné et prévisible.

28 modules interconnectés

Avec quelques modifications, cette approche pouvait fonctionner pour notre simulateur de cellule, à condition de faire une hypothèse simplificatrice importante : supposer que, bien que tous les processus biologiques aient lieu simultanément dans une cellule vivante, leurs actions sont indépendantes sur des périodes inférieures à la seconde. Si cette hypothèse était viable, nous pourrions diviser la vie de la cellule en tranches d'une seconde et faire tourner chacun des 28 modules, dans l'ordre, pendant une seconde avant de mettre à jour l'ensemble des variables de la cellule. Le modèle rendrait compte de toutes les interdépendances biochimiques (par exemple, le fait que la transcription des gènes et la synthèse de l'ADN dépendent de l'énergie et des nucléotides

Un tournant de la biologie des systèmes

Le modèle « pleine-cellule » de *Mycoplasma genitalium* proposé par Markus Covert et ses collègues marque un tournant dans la jeune histoire de la biologie des systèmes.

Si l'idée de modéliser un organisme monocellulaire n'est pas nouvelle, le modèle de M. Covert est le premier à s'appuyer sur un inventaire exhaustif de gènes et à donner une représentation aussi complète que possible de leurs fonctions et des processus cellulaires dans lesquels ils sont impliqués.

Pour construire ce modèle, qui agrège une quantité considérable de connaissances, M. Covert et son équipe ont dû vaincre plusieurs difficultés. Ils ont dû créer un cadre original pour combiner des sous-modèles fondés sur des techniques de modélisation très différentes. Par ailleurs, ils ont su trouver des jeux de paramètres pour lesquels le modèle produit

les bonnes proportions des différents ingrédients qu'une cellule doit accumuler pour compléter son cycle et faire une copie d'elle-même. Leur modèle produit des temps réalistes de division cellulaire et prédit correctement la viabilité d'un certain nombre de mutants.

Ce modèle est en rupture avec la tradition en ce qu'il n'est pas destiné à évaluer une hypothèse préalable spécifique. L'enjeu est d'organiser la totalité des interactions qui composent le système de sorte qu'on puisse en observer le comportement collectif. Ce n'est que dans un second temps que l'on cherchera à exploiter cet outil « panoptique » pour tester de nouvelles hypothèses.

Abandonner ainsi une modélisation d'explication pour une modélisation de navigation ne va pas de soi. On passe des principes clairs et concis d'un système à quelques variables à un système plus complexe difficile à appréhender intuitivement. Le modèle est une carte qui permet d'explorer cette complexité. De ce point de vue, il se manipule comme on le ferait du système réel, et les « faits » qu'il produit réclament une explication au même titre que des données expérimentales. Mais comme le modèle ne reflète que la connaissance imparfaite que nous avons de cet organisme, si simple soit-il, et qu'il ne la reflète qu'imparfaitement, le passage du système réel, avec ses difficultés propres, à son substitut numérique vient sans garantie de qualité. Le passage « au numérique » peut ainsi sembler une opération au bénéfice incertain dans l'état de nos connaissances.

Quoi qu'il en soit, la possibilité d'interroger le système dans sa totalité permet d'aborder de nouvelles questions. On peut par exemple étudier l'impact du niveau des ressources énergétiques sur les composants du système, ou comment offre et demande se répondent sur le « marché » de la traduction des gènes, et si des gènes répondent différemment aux conditions de cette compétition ; ou encore examiner l'influence de la densité des molécules associées à l'ADN sur l'activité transcriptionnelle du système. Le modèle pleine-cellule de M. Covert ouvre un nouveau champ d'enquête en biologie des systèmes, où l'on étudie l'interdépendance des différents processus qui composent une cellule et leur intégration fonctionnelle à l'échelle cellulaire.

– Vincent Danos
CNRS et Université
d'Édimbourg (Écosse)

produits par le métabolisme), mais seulement aux échelles temporelles supérieures à la seconde.

Tout en construisant notre cellule virtuelle, nous avons intégré au logiciel des capteurs pour mesurer l'état de la cellule. Chaque lancement du simulateur, couvrant la totalité d'un cycle de vie d'une cellule, produisait 500 mégaoctets de données. En sortie, on obtenait des dizaines de graphiques et de représentations visuelles.

Au début, les résultats furent frustrants. Pendant des mois, tandis que nous corrigeons le code, précisons les équations mathématiques, ajoutons de nouvelles contraintes sur les paramètres et ajustons certaines (grâce à de nouvelles données de laboratoire), la cellule refusait toujours de se diviser et se comportait de façon imprévisible. Pendant un moment, par exemple, elle a produit d'énormes quantités d'un certain acide aminé (l'alanine) et presque rien d'autre.

Et puis un jour, notre microbe cybernétique est arrivé au terme de son cycle cellulaire et s'est enfin divisé. De plus, son

temps de doublement était proche de neuf heures – celui du mycoplasme réel. De nombreuses mesures étaient encore loin de la réalité, mais le succès était enfin à notre portée.

Quelques mois plus tard, la majorité des données obtenues par simulation correspondait exactement à la croissance observée d'une vraie cellule. Et le reste était fascinant : inattendu, mais biologiquement plausible. Le premier modèle numérique d'un organisme vivant entier était fonctionnel. Qu'allait-il nous apprendre ?

Modéliser la vie d'une cellule humaine ?

Aujourd'hui, après environ un an d'utilisation de ce nouvel outil, nous découvrons encore des informations passionnantes à chaque fois que nous scrutons le fonctionnement de notre micro-organisme virtuel, tandis qu'il gère des millions de détails participant à sa vie et à sa réplication. Par exemple, la fréquence à laquelle les protéines participant au décodage des gènes se

décrochent de l'ADN est impressionnante : environ 30000 fois par cycle de vie de neuf heures. Nous avons aussi découvert que la période de doublement du matériel biologique lors de la division cellulaire, très stable, est en fait une propriété émergente qui découle des interactions de deux phases distinctes et indépendantes de la réplication, toutes deux de durée très variable. Et les enregistrements seconde par seconde du comportement de la cellule nous ont permis d'expliquer pourquoi sa division s'arrête net quand on inactive certains gènes, mais se poursuit encore une dizaine de fois si l'inactivation porte sur d'autres gènes essentiels : ces divisions supplémentaires se produiraient à chaque fois que la cellule a en réserve davantage de copies d'une protéine essentielle qu'elle n'en a besoin durant un cycle de vie : l'excédent est légué à ses descendantes, qui ne meurent que quand le stock est épuisé. Il faudra sans doute encore des années pour comprendre tout ce que ces simulations révèlent sur le fonctionnement des mycoplasmes et des cellules en général.

Nos travaux sur *M. genitalium* ne sont que la première des nombreuses étapes à franchir pour modéliser les cellules ou tissus humains à l'échelle des gènes et des molécules. Notre modèle est loin d'être parfait, et les mycoplasmes sont l'expression la plus simple de la vie autosuffisante. Nos simulations, le code source, les données expérimentales, le code de visualisation, la base de connaissances, tout est disponible gratuitement sur Internet. Avec d'autres chercheurs, nous travaillons déjà à améliorer le simulateur et à l'étendre à divers organismes, tels que la bactérie *E. coli* et la levure *Saccharomyces cerevisiae*, toutes deux des organismes utilisés fréquemment tant dans les laboratoires que dans l'industrie.

Chez ces espèces, la régulation des gènes est beaucoup plus complexe, et la localisation des événements dans la cellule joue un rôle bien plus important (la levure, par exemple, présente un noyau). Autant de problèmes à résoudre avant de s'attaquer à des cellules plus complexes encore, tels des macrophages de souris ou

humains. Faciles à cultiver, ces cellules du système immunitaire seront sans doute les premières utilisées pour ajuster et valider le modèle.

Il est trop tôt pour estimer le temps que ces améliorations prendront. Comparées aux bactéries, les cellules humaines comportent beaucoup plus de compartiments. De plus, leur contrôle génétique, plus complexe que celui des mycoplasmes, demeure en grande partie mystérieux. Enfin, de par leurs rôles complémentaires dans les tissus multicellulaires, les cellules humaines interagissent plus avec d'autres types cellulaires que les bactéries.

Début 2008, modéliser la plus simple des cellules ne semblait pas pouvoir se concrétiser avant une décennie, et modéliser une forme de vie plus complexe paraissait inenvisageable. Aujourd'hui, l'idée de simuler une cellule humaine ne paraît plus si saugrenue. Et si le logiciel se révèle inapproprié, au moins cet échec mettra-t-il en lumière les nombreuses inconnues à expliciter encore sur nos propres cellules. ■

■ BIBLIOGRAPHIE

E. Gonçalves *et al.*, **Bridging the layers: toward the integration of signal transduction, regulation and metabolism into mathematical models**, *Molecular Biosystems*, vol. 9, pp. 1576-1583, 2013.

J. R. Karr *et al.*, **A whole-cell computational model predicts phenotype from genotype**, *Cell*, vol. 150, pp. 389-401, 2012.

P. L. Freddolino et S. Tavazoie, **The dawn of virtual cell biology**, *Cell*, vol. 150, pp. 248-250, 2012.

W. Gibbs, **La biologie virtuelle**, *Pour la Science* n° 288, pp. 38-44, 2001.

france culture

LA MARCHÉ DES SCIENCES

Découvertes, inventions, aventures savantes au fil de l'Histoire

Aurélie Luneau
14h-15h / chaque jeudi

franceculture.fr

POUR LA SCIENCE



La biologie de vers une ingénierie

François KÉPÈS

Concevoir un génome *ab initio*, inventer une cellule minimale, produire de nouvelles fonctions biologiques... Telles sont quelques-unes des ambitions de la biologie de synthèse. Son objectif : ouvrir de nouveaux horizons en biologie en s'inspirant des méthodes de l'ingénierie.

Au début des années 1960, l'invention des circuits intégrés a métamorphosé l'ingénierie électronique : un schéma imprimé sur un masque suffisait à déposer sélectivement des couches métalliques et chimiques sur un support, et à produire ainsi le circuit imaginé. Bientôt, les ingénieurs ont puisé dans les circuits déjà conçus pour les combiner en des montages toujours plus complexes. La modélisation informatique aidant, les progrès furent considérables, conduisant aux nombreux circuits intégrés que comportent les téléphones portables, ordinateurs et autres dispositifs électroniques.

Et si une telle combinaison de technique et de méthode était transposable en biologie ? S'il était possible, en biologie aussi, de disposer d'une banque d'outils et de

règles rationnelles qui, combinés selon les besoins, aideraient à comprendre le vivant en le fabriquant, ou à concevoir et produire à la demande de nouvelles fonctions biologiques n'existant pas dans la nature ? Depuis quelques années, cette idée fait son chemin parmi les biologistes. Rassemblées sous le nom de biologie de synthèse, leurs approches se construisent aux interfaces de la biologie et de l'ingénierie, avec des apports de la chimie, de la physique, des mathématiques et de l'informatique. Elles ont pour objectif d'injecter en biologie les principes fondant toute ingénierie. De quoi s'agit-il ? Quelles sont les ambitions de la biologie de synthèse ? Que permet-elle aujourd'hui ? Quelles sont ses perspectives et limites ? Telles sont les questions que nous examinerons ici.

La biologie de synthèse est à la biologie des systèmes – l'analyse des interactions des biomolécules et des comportements des systèmes biologiques – ce que la chimie de synthèse est à la chimie analytique (l'analyse des substances chimiques). Après le triomphe de la chimie analytique dans les années 1850, les chimistes se sont tournés vers la synthèse pour obtenir plus de produits naturels à analyser, puis des composés n'existant pas dans la nature. Cette recherche a donné naissance à l'industrie chimique et à la pétrochimie du XX^e siècle. De même, dans les années 1990, les progrès de la biologie analytique ont été considérables grâce au développement des techniques dites « -omiques » : le séquençage du génome (génomique), la cartographie des ARN (transcriptomique) ou des protéines (protéomique) d'un système biologique, ou des modifications non génétiques du génome (épigénomique)... Stimulée par un volume croissant de données, la biologie des systèmes s'est développée, attirant physiciens et informaticiens. Quelques années plus tard est apparue la biologie de synthèse, qui s'appuie sur la compréhension ainsi acquise.

La biologie de synthèse est l'ingénierie rationnelle de la biologie. En d'autres termes,