

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.