

La nouvelle Physique-Mathématique pour décrypter la biologie cellulaire et booster l'économie de haute technologie

• D. HOLCMAN

Quelques questions en biologie cellulaire comme l'organisation de l'ADN dans le noyau ou le développement du cortex cérébral ont déjà bénéficié de méthodes de calcul, de simulations numériques, ou de l'élaboration de théories mathématiques abstraites. Ce fut aussi le cas à plus grande échelle dans nombre de branches de la physique, chimie, en relativité générale, ou en ingénierie, qui ont motivé des théories sophistiquées. Sur le plan des mathématiques, il a fallu élaborer de nouvelles méthodes pour résoudre de nouvelles équations, et déduire des conclusions solides à partir de leur analyse et de simulations numériques. Cette approche a servi à identifier, trouver de nouveaux motifs cachés dans les données expérimentales et révéler les propriétés secrètes de la matière.

La physique expérimentale a souvent contribué de façon spectaculaire à la biologie avec des grandes révolutions instrumentales telle que la nouvelle microscopie de haute résolution qui a permis de briser la limite de la diffraction et de voir pour la première fois les molécules évoluer dans la cellule avec une résolution de l'ordre de la dizaine de nanomètres. Cette découverte a notamment été récompensée par le Prix Nobel de Chimie en 2014 et fait suite à la découverte et l'utilisation surprenante et efficace de molécules fluorescentes, comme la GFP (Prix Nobel de chimie 2008) pour suivre dans le vivant sous un microscope des molécules d'intérêts.

Pourtant, la contribution de la théorie à la biologie est beaucoup moins spectaculaire. La raison principale étant que la découverte de faits nouveaux et non de preuves (qui occupe une grande partie de l'activité mathématique moderne) repose la plupart du temps en biologie sur l'expérience. Mais quand une multitude de règles empiriques est nécessaire pour rendre compte du comportement aux différentes échelles moléculaires et cellulaires, la modélisation devient nécessaire pour élaborer

des calculs à partir de ces règles. Cette compilation souvent multi-échelle permet ainsi de mieux comprendre les fonctions élaborées du vivant. Dans ces conditions, le bénéfice d'écrire la biologie en langage mathématique, aussi ambiguë que cette entreprise puisse paraître, reviendrait à découvrir, ordonner ces règles, dériver des prédictions et les utiliser dans l'analyse de données expérimentales. Cette approche n'est pas vraiment nouvelle : dans les années 50, Alan Turing a formulé la notion de morphogènes et dérivé les équations de réaction-diffusion pour comprendre comment certains motifs peuvent se former spontanément dans le vivant. De même, l'identification de structures moléculaires par cristallographie ou de diffraction avait permis d'identifier des structures nouvelles comme la double hélice de l'ADN.

L'explosion des connaissances en 30 ans grâce aux nouvelles manipulations génétiques, au criblage systématique de différents gènes impliqués dans les fonctions spécifiques du vivant ou encore les nouvelles approches en électrophysiologie et bien d'autres a propulsé la biologie sur le devant de la scène scientifique : la biologie est ainsi devenue la plus grosse discipline scientifique en termes de ressources humaines et financières, mais aussi la plus représentée dans les revues scientifiques. Cette explosion a un coût financier exorbitant mais nécessaire, alimenté par une vitesse de production et d'apparition des connaissances exponentielle, plus rapide que dans toutes les autres disciplines. Quelle est alors le rôle possible de la physique théorique et des mathématiques ?

Ce que révèle la biologie d'aujourd'hui, c'est une variété de formes et d'interactions qui sont le produit de milliard d'années d'évolution. Ces interactions sont d'une richesse encore indescriptible par les modèles mathématiques actuels et les diverses géométries existantes (Euclidienne ou non). Les re-

lations aussi complexes qu'elles soient entre les espaces abstraits provenant de l'analyse semblent en effet bien loin de capturer la richesse des relations entre les formes du vivant et leur fonction. Alors que nous héritons de 60 ans d'effort d'imagination et de rigueur sur des objets mathématiques modernes et des méthodes abstraites pour étudier l'existence et le comportement de solutions d'équations, il semblerait que ce cadre ne soit pas suffisant pour aborder les questions même élémentaires posées par le vivant, comme l'organisation tridimensionnelle des chromosomes, les réponses de la cellule à un seul événement moléculaire ou encore la construction et l'organisation de réseaux cellulaires pour définir des grandes structures fonctionnelles du cerveau. Devant cette grande complexité, pouvoir explorer un espace de paramètres en grande dimension est une étape décisive dans notre compréhension des fonctions du vivant. Cependant, à cause de leur complexité, ces questions demandent des calculs souvent sophistiqués où le besoin de simplification est essentiel. Une partie des nouvelles demandes se trouve ainsi dans la résolution numérique de ces calculs : comment réaliser des simulations en temps raisonnable tout en capturant la complexité du phénomène biologique ? Comment résoudre numériquement des solutions d'équations quand les temps de calculs divergent vers l'infini pour des petites valeurs des paramètres ? Tels sont quelques défis proposés par la biologie aux mathématiciens. L'exigence classique des mathématiques de prouver l'existence et l'unicité des solutions d'équations, de trouver des mesures invariantes, de pousser les inégalités ou autre manipulation élégante est loin de répondre au besoin de comprendre, de calculer et de prédire en biologie.

Cette nouvelle approche nécessite souvent une connaissance fine de la biologie et biophysique, mais aussi de la physique statistique, des mathématiques appliquées et de la physique théorique, ainsi que de la curiosité pour continuer d'apprendre si nécessaire n'importe quelle autre discipline qui deviendrait nécessaire en chemin. Les premières et grandes avancées dans ce domaine proviennent d'environnements scientifiques et académiques propices qui ont su créer les conditions favorables, pour faire tomber les barrières réelles bien qu'arbitraires entre les disciplines, concentrées dans certains endroits par exemple, aux Royaume-Uni, États-Unis (UCSF, Berkeley, Salk, etc.), Israël (Tel Aviv University) ou encore EMBL en Allemagne.

Malgré quelques efforts pour implanter ce type

d'environnement en France à coup de chaire d'excellence pour rapatrier des jeunes talentueux qui ont été formés ou reformés à l'étranger, ce nouvel effort en modélisation mathématique de la biologie est encore largement sous-représenté. De plus, cet effort ne peut être fructueux que s'il y a une véritable volonté de financement sur le long terme afin de favoriser le développement d'équipes et assurer leur compétitivité à l'international. Les tentatives domestiques qui se sont construites en évitant une formation de 3 à 5 ans minimum à l'étranger sont loin d'être concluantes, malgré le potentiel incontestable de chaque génération de chercheurs durement sélectionnés. Il serait temps que les commissions interdisciplinaires arrivent à maturité : elles répondent à un vrai besoin et combler un manque. Elles sont une très bonne initiative dans un système si cloisonné, mais devraient permettre de recruter des chercheurs d'une certaine stature, celle-ci étant nécessaire dans les disciplines interdisciplinaires où une formation postdoctorale longue et intensive permettra aux jeunes chercheurs théoriques de pouvoir évoluer à l'interface de plusieurs disciplines. Enfin, quand les financements deviennent rarissimes, ce sont les tentatives nouvelles qui sont sacrifiées les premières, devant les disciplines traditionnelles mieux établies. Les critères élémentaires pour caractériser cette recherche sont certainement la capacité à produire des résultats qui se retrouvent dans les journaux de mathématiques, de physique et de biologie ainsi que les journaux à l'interface, par le même auteur en position principale. Il est important de prouver qu'une réflexion sur les processus biologiques mène à de nouvelles analyses mathématiques et statistiques, et à de nouveaux modèles qui sont souvent bien reconnus par la communauté de physique.

De plus, il est illusoire de penser que des chercheurs de premier plan vivant à l'étranger et travaillant à ces interfaces vont venir s'installer en France sans pouvoir compenser la lourdeur, la rigidité de la gestion, les petits salaires non-compétitifs et une loi Sauvadet, appliquée avec zèle par l'administration (INSERM et Centre national de la recherche scientifique (CNRS) entre autres) et qui plombe la recherche et la discrédite sur le plan internationale. Alternativement, une recherche efficace se prépare en formant une génération, en offrant des cours spécialisées, des enseignements innovants et créatifs, des stages dans des laboratoires de différentes disciplines : les thèses à l'interface doivent pouvoir conduire à des publications pour l'étudiant en premier auteur dans les journaux

de différentes disciplines, bien qu'il soit aussi envisageable qu'une formation débute par une thèse théorique ce qui permet de bien ancrer les fondamentaux, ~~mais alors~~ un post-doc de 2 à 5 ans est nécessaire pour apprendre le reste. Il est peu fructueux de modéliser un phénomène biologique sans avoir vu et compris les expériences sur le sujet.

En résumé, former des étudiants qui deviendront des chercheurs en sciences théoriques appliquées aux sciences du vivant demande une révision de l'enseignement qui doit largement aboutir à regrouper pour ces cursus, physique, mathématiques, biophysique théorique et l'informatique appliquée (basé sur la programmation pratique). De plus un post-doc long (~~de 3 à 5 ans~~) est nécessaire afin d'apporter une autonomie au jeune chercheur et la capacité de publier dans différentes disciplines en premier auteur. Enfin, le recrutement dans les universités et instances de recherche doit être bien orchestré pour protéger ces disciplines minoritaires. Il va sans dire que les chercheurs participant à ces recrutements doivent eux même avoir démontré leur autorité scientifique à ces interfaces. Une évaluation adaptée et nouvelle doit alors être mise en place pour que ces chercheurs puissent être jugés et promus convenablement, sans quoi il ne faudra pas s'étonner de les voir partir, ou repartir.

Sans promouvoir et subventionner cette nouvelle culture interdisciplinaire, cette richesse scientifique ne pourra émerger. Dans ce cas, comme jadis l'informatique, la cryptographie, la robotique, les drones, ... et bien d'autres, l'économie domestique ne pourra bénéficier de cet apport. Les universitaires comme le monde industriel, en bataille depuis tant d'années ont une grande part de responsabilité pour n'avoir pas su vraiment interagir ou voulu apporter de la diversité, en favorisant par exemple des chercheurs capable de mener en même temps et en symbiose une recherche théorique et appliquée.

Une large part de l'économie de haute technologie de demain est dans l'analyse mathématique fine de la biologie et la recherche médicale translationnelle d'aujourd'hui. Sans un effort, une vision et les moyens de la poussée, il faudra alors des sommes considérables pour acheter à l'étranger au prix fort, ce qui aurait pu être produit en France. La course a déjà bien recommencé avec l'intelligence artificielle et le machine learning dans les entreprises web (Google, Facebook...) et devant la sous-représentation française des biotechs à grand succès installés pour la plupart aux USA, Suisse et Israël notamment. Google fait de la biologie computationnelle un domaine d'investissement majeur via GoogleX. La fondation du mathématicien J. Simons concentre une grande partie de ses investissements à encourager les recherches à l'interface pour étudier l'autisme, l'évolution et bien d'autres. Le développement de centres interdisciplinaires dans les grandes villes est déjà très bénéfique comme à Columbia¹, Berkeley-UCSF-UCSC² et bien d'autres, qui n'ont aucun équivalent en France. La prochaine révolution scientifique passera sans doute par les méthodes d'analyse pour extraire des motifs dans des données de grande dimension acquises en continu, qui pourront être appliqués à prévenir les accidents de santé. Une nouvelle économie va émerger. Ce fut sans doute l'état d'esprit de Pasteur et ses lieutenants, qui ont en une génération fait augmenter l'espérance de vie de 20 ans. Ils ont créé l'institut Pasteur, qui a révolutionné plusieurs fois l'économie en sauvant dès son début les secteurs du vin et du ver à soie en France. À l'aube d'une nouvelle révolution, les mathématiques, la biologie, la physique et l'informatique sont nécessaires. La France saura-t-elle rattraper son retard ou allons nous le regarder s'accumuler, comme ce fut le cas plusieurs fois ces 20 dernières années, à moins que se réveille un certain esprit, comme pendant la Révolution française ?

D. HOLCMAN

Directeur du Groupe de Mathématiques appliquées et Biologie Computationnelle, École Normale Supérieure, Paris, France, Professeur invité au Department of Applied Mathematics and Theoretical Physics Cambridge University et Fellow of the Churchill college, Cambridge University, CB3 0DS UK

Je remercie Thibault Lagache (Modélisation et analyse d'image, Institut Pasteur), Khanh Dao Duc (Statistique, Berkeley) et Nathanaël Hozé (Modélisation, ETH Zurich) pour leurs remarques et critiques.

1. <http://systemsbiology.columbia.edu/center-for-computational-biology-and-bioinformatics-c2b2>
2. <http://qb3.berkeley.edu/ccb/>