

› Les enjeux des standards dans la recherche biologique

Huit cent mille ! Il existe aujourd'hui près de huit cent mille standards : standards technologiques et scientifiques, d'ingénierie et de physique, standards culinaires et hôteliers, d'informatique et de chimie. De la nano- à la macroéchelle, ils touchent nos cinq continents, de l'Europe à l'Asie, de l'Amérique à l'Afrique. Au-delà de notre planète, les standards ont permis l'exploration de nos voisines, la Lune et Mars. Tous ces standards sont répertoriés, classés, décrits, dessinés et calculés, souvent avec une très grande précision. Ils sont fréquemment acceptés par une majorité d'utilisateurs, mais pas toujours unanimement. Beaucoup sont couramment utilisés, certains plus que d'autres.

L'histoire des standards est ancienne. Ils apparaissent avec la structuration des sociétés humaines et leur industrialisation. Leur utilité s'impose très tôt. On cite souvent comme l'un des exemples les plus anciens le « code d'Hammourabi », qui fixe un étalon monétaire pour régler les échanges économiques. Aujourd'hui, il est clair que notre monde hautement technologique et intimement organisé en réseaux complexes de fabricants et d'utilisateurs, de vendeurs et d'acheteurs, de donneurs et de receveurs, d'éducateurs et d'élèves, ne fonctionnerait pas en leur absence. Mais pourquoi la biologie moléculaire a-t-elle échappé pour l'instant à cette tendance, et quels sont donc les besoins actuels de standards en biologie ? Comment pourrait-on appliquer des standards en recherche biologique ? Dans quel cadre faudrait-il en proposer ? Ce sont ces questions que cet ouvrage se propose d'aborder.

La biologie systémique. Le monde macromoléculaire qui existe à l'intérieur de chacune de nos cellules est lui-même un monde intimement organisé en réseaux complexes d'interactions physiques, chimiques, logiques et dynamiques, qui prennent place entre protéines, lipides, glucides, acides nucléiques et autres métabolites. Ces nombreuses interactions forment des systèmes macromoléculaires dont les propriétés ne correspondent pas exactement à la somme des propriétés de leurs constituants. En effet, ces systèmes cellulaires sont caractérisés par ce que l'on appelle des « propriétés émergentes ». Pour illustrer ce concept, disons par exemple que les propriétés émergentes sont celles qui permettent à un avion de voler : ce ne sont pas des propriétés de ses composants pris individuellement, mais de leur organisation précise. S'agissant à présent de biologie, on pourrait de manière métaphorique considérer les machineries moléculaires complexes que recèle une cellule comme des puzzles qui ne sont, en effet, pas réductibles à leurs pièces. Plus précisément, la simple connaissance des composants d'un organite multiprotéique, ou la description des enzymes impliquées dans un chemin métabolique ou dans une voie de signalisation ne révéleront le plus souvent leur signification fonctionnelle que lorsqu'on sera arrivé à saisir leur agencement spatial ou temporel. Comme c'est le cas pour un puzzle, le rôle biologique joué par une machinerie doit parfois attendre que la mise en place de la dernière pièce soit réalisée. À titre d'illustration, on peut considérer une machinerie de complexité très réduite où l'on reconnaît par exemple la présence de trois composants ou agents en interaction deux à deux. Si l'un des agents a un effet négatif et les deux autres positifs, le motif global constitue une boucle négative. Si un éventuel quatrième

composant a échappé à l'attention et que celui-ci a un effet négatif, la « signification » du système, de négative qu'elle était, devient cette fois inverse, c'est-à-dire que la boucle est positive.

Une compréhension précise de ces propriétés émergentes et de leurs perturbations observées dans les situations pathologiques devrait permettre, à terme, de découvrir de nouvelles stratégies thérapeutiques et de contrôler leurs effets dans les patients soignés. La compréhension et la manipulation des propriétés émergentes des systèmes cellulaires constituent le but central de la « biologie systémique ».

On assiste actuellement à une véritable renaissance de la biologie systémique permise, en tout ou en partie, par :

- l'avènement de l'informatique;
- le séquençage et l'annotation du génome humain et de celui de nos organismes modèles préférés;
- le déploiement d'une nouvelle génération d'outils de manipulation des gènes à l'échelle de tout le génome;
- l'arrivée des jeux de données biologiques fonctionnelles à haut débit;
- le développement de nouvelles techniques de modélisation, qu'elles soient quantitatives ou booléennes, ou qu'elles touchent à la théorie des graphes ou aux méthodes statistiques.

La biologie systémique ajoute ainsi à la biologie moléculaire cette dimension d'interactions complexes entre composants individuels qui, il nous faut bien l'avouer, avait été largement laissée de côté depuis la découverte de la double hélice et la formulation assez réductionniste du « dogme central ADN → ARN → protéine » par Francis Crick et de la théorie « un gène, une fonction » illustrée par les travaux de George Beadle et Edward Tatum.

Standards pour la biologie systémique. C'est dans ce contexte de renaissance que s'inscrit le but principal de ce livre de recueils centrés sur les « standards pour la biologie systémique ». Pour le dire simplement, on ne peut imaginer la biologie systémique sans l'utilisation de standards. Les pionniers de la biologie moléculaire opéraient, à juste titre, « localement », capables de mémoriser et de conceptualiser les fonctions de leurs quelques molécules d'intérêt, dont le rôle fonctionnel s'analysait selon des causalités linéaires. Aujourd'hui que cette causalité linéaire est à l'évidence insuffisante pour rendre compte des caractéristiques du vivant (le simple fait que 20000 gènes devraient « expliquer » un nombre de caractères bien plus élevé illustre ce point) les biologistes sont appelés à interagir de plus en plus avec des ingénieurs, des physiciens, des mathématiciens et d'autres experts, qui abordent les problèmes de manière plus globale, en utilisant des outils plus rigoureux et des concepts plus formalisés, plus généraux, adaptables et flexibles. Ces relations interdisciplinaires sont déjà, et seront de plus en plus, basées sur des standards stricts, pour le meilleur et pour le pire.

Histoire sommaire d'un premier standard moderne. L'un des premiers standards de l'ère industrielle à voir le jour, au milieu du XIX^e siècle, concernait un objet en apparence anodin, mais tellement nécessaire : la vis. Pouvons-nous encore imaginer un monde dans lequel le filet d'une vis ne correspondrait à aucun écrou? Comment pourrions-nous construire quoi

que ce soit si les cylindres ne correspondaient aux pistons, les fiches n'étaient compatibles avec aucune prise, si ni le mètre ni le pied n'était défini de manière précise ?

C'est ce genre de monde que deux ingénieurs, Joseph Whitworth, un anglais, et William Sellers, un américain, ont révolutionné en créant, négociant et, à terme, en imposant des standards de dimensions pour la manufacture des vis dans leurs pays d'origine. Ces standards se sont peu à peu imposés, de manière indépendante des deux cotés de l'Atlantique, participant par exemple à l'essor fulgurant de la construction du système ferroviaire aux États-Unis à la fin du XIX^e siècle. Pour la petite histoire, ces standards étant restés différents entre les deux pays jusqu'à la deuxième guerre mondiale, c'est avec surprise que des soldats américains se sont rendus compte que les vis et les écrous dont ils disposaient étaient inutilisables pour réparer les tanks de l'armée alliée anglaise endommagés par l'armée allemande. Dès la fin de la guerre, le Royaume-Uni décida d'opter pour le standard de Sellers, plus pratique à implémenter.

Nous retenons trois leçons importantes de ces événements historiques. D'une part, la production et la consommation de masse sont indéniablement liées à la définition et à l'utilisation de standards. D'autre part, l'imposition de ces standards passe par l'obtention d'un consensus établi entre un grand nombre d'acteurs intéressés qui, à un moment ou à un autre, doivent privilégier l'intérêt d'une communauté par rapport à leurs intérêts personnels. Enfin, une certaine dose de compétition entre différents standards permet la sélection des plus pratiques, moins coûteux, mieux appropriés à un problème particulier.

Classification sommaire des standards à utiliser en biologie systémique. Il est important de distinguer deux classes de standards : les standards de fabrication et ceux de qualité. Les standards de fabrication correspondent à des règles imposées dans un processus de fabrication, comme pour les formats DVD, CD ou VHS, ou à des caractéristiques bien définies comme le mètre et la seconde. Les standards de qualité, eux, correspondent à des mesures de qualité établies en aval de la production. Nous retiendrons en exemple l'étoile du guide Michelin et celles de l'hôtellerie, ou l'appellation contrôlée de type champagne, ou encore les critères de consommation d'énergie imposée par l'Union européenne. Deux exemples illustrent ces différences dans le cadre de la biologie systémique.

Le système de clonage de gènes appelé « clonage Gateway® » fut développé pour la manipulation de phases de lectures ouvertes (*open reading frames* ou ORF) pour l'intégralité des gènes d'un organisme (notion d'ORFéome). Gateway® est principalement basé sur la définition de standards de fabrication, séquences d'ADN et conditions de réactions biochimiques, qui permettent une manipulation de l'ORFéome complètement automatisée et parallèle, hautement contrôlée et systématique, et ceci quelle que soit la séquence des fragments d'ADN à cloner.

L'un des premiers exemples de développement de standards de qualité correspond au « PHRED score » utilisé en séquençage d'ADN à grande échelle. Le PHRED score indique la chance qu'une base d'ADN ne soit pas correctement assignée à partir des données expérimentales sortant des séquenceurs automatiques. Le PHRED score fut une réelle révolution pour le projet international de séquençage du génome humain, il représente en effet un standard de qualité inégalé jusqu'ici en biologie expérimentale à haut débit.

Peu ou pas de standards en biologie conventionnelle. L'émergence de standards en biologie est intimement liée à la représentation graphique et à la nomenclature des systèmes biologiques étudiés. Par exemple, ils sont utilisés dans les schémas qui décrivent les systèmes vivants contenant cellules, organites, flux de matière et réactions métaboliques. Par ailleurs, ils se sont parfois imposés dans la dénomination des entités : gènes, mutants, protéines, etc. Les chercheurs travaillant sur la mouche *Drosophile* donnent des noms complets à leurs mutants et donc aux gènes correspondants ; les généticiens de la levure, eux, ont adopté un code à trois lettres. Ces standards se sont imposés de façon consensuelle au sein de chaque communauté scientifique. Certaines communautés de chercheurs se sont gardé une liberté totale : les descriptions de protéines ou de gènes humains sont totalement libres — par exemple, un nom d'invention (insuline), un sigle (EGF), un nom descriptif (Tubuline), une valeur d'un paramètre biochimique (p53), etc. Cette complexité a d'ailleurs conduit les biologistes à s'organiser (voir par exemple l'initiative « Human Gene Nomenclature Committee », <<http://www.genenames.org/>>).

Cette situation n'a pas été préjudiciable aux développements des recherches ni de leurs applications dans le domaine industriel ou médical, tant que les connaissances nouvelles étaient échangées entre experts d'un même domaine et que seules les connaissances intégrées, synthétisées et simplifiées étaient communiquées à l'extérieur du champ primaire d'acquisition. Cette situation a complètement changé depuis une dizaine d'années, tant en raison des modifications profondes du mode de recherche en biologie, avec l'apparition de protocoles expérimentaux à grande échelle, que de l'apparition de modes de stockage et de traitement de l'information peu connus auparavant dans cette communauté. Ces outils, largement répandus dans de nombreux domaines de la recherche (physique, chimie, médecine, etc.) ont finalement envahi le champ de la recherche expérimentale en biologie. Les standards ont donc fait leur apparition, par le biais des nouvelles technologies, principalement en génomique (séquençage, puces à ADN) et en protéomique (spectroscopie de masse), mais aussi dans des domaines plus anciens comme la biologie cellulaire (imagerie confocale).

Conclusions : des standards, assez, mais pas trop ! Il est à prévoir que les standards en biologie vont continuer à s'imposer dans la mesure où la biologie des systèmes est par nature holistique avec une accumulation inégalée de données, et que l'interprétation fonctionnelle de celles-ci nécessite des approches quantitatives de type modélisation et simulation exigeant précisément la mise en place de normes standards. La principale difficulté à anticiper est leur harmonisation entre disciplines distinctes, soutenues par des technologies indépendantes. Il est à craindre également que ces standards apparaissent comme un mal nécessaire pour les biologistes pratiquant la biologie à grande échelle, mais restent des outils inutilisés et obscurs pour le reste de la communauté scientifique.

Cet ouvrage vise donc à aplanir ces obstacles de communication en expliquant leur mise en œuvre et en illustrant leur usage à des fins d'échanges interdisciplinaires. Mais nous devons aussi garder à l'esprit que les règles de fonctionnement du vivant nous échappent encore en grande partie. L'appréhension des problématiques biologiques bénéficiera utilement de la mise

en place de standards, notamment sans doute pour représenter des modèles explicatifs des phénomènes étudiés, ce qui est indiscutablement une difficulté actuelle rencontrée par les recherches à haut débit de la biologie systémique. Mais il serait inconsidéré de tenir toutes ces représentations actuelles — qui sont des simplifications nécessaires — comme couvrant la complexité biologique sous-jacente. Il serait en particulier illusoire de vouloir standardiser et figer prématurément, en quelque sorte, des représentations du vivant dont les bases conceptuelles restent incertaines et floues. Les standards ne doivent pas être des éléments contraignants mais, au contraire, trouver leur justification dans l'aide au formalisme qu'ils apportent.

C'est par l'organisation des connaissances toujours plus vastes et complexes que nous réussons le défi de mieux comprendre le vivant, de dégager de nouveaux concepts et finalement de proposer des solutions à des problèmes médicaux ou environnementaux.

Pierre Legrain

(Institut de biologie et de technologies de Saclay, CEA)

Marc Vidal

(Center for Cancer Systems Biology, CCSB,
Department of Cancer Biology, Dana-Farber Cancer Institute,
Department of Genetics, Harvard Medical School, Boston, États-Unis)