

DE LA MODÉLISATION CELLULAIRE AUX SYSTÈMES INTÉGRÉS

Une des orientations majeures des biotechnologies concerne la modélisation biologique des systèmes intégrés. La dimension industrielle de cette nouvelle approche est d'exploiter de façon efficace la masse d'informations existantes, aussi complexes que le vivant, pour l'amélioration significative des processus de développement des médicaments. Les besoins sont énormes lorsqu'on sait qu'entre 2002 et 2005, au moins 196 médicaments vont perdre leur protection industrielle, entraînant un manque à gagner estimé à 18 milliards de dollars sur le seul marché américain, et que moins de la moitié sera remplacée par de nouvelles molécules en dépit de dépenses R&D toujours plus grandes (26 milliards de dollars en 1999).

Le continuum physiologique : un système intégratif non-linéaire

Le monde des sciences de la vie est face à une véritable avalanche d'informations. À titre d'exemple, le contenu de la base de données MedLine contenait, en 1984, un peu plus de 6 millions d'articles référencés. Elle en contient aujourd'hui plus de 11 millions (source : NCBI, USA). De plus, une entité biotechnologique telle que Millennium ou Celera produit à elle seule un volume annuel de données supérieur à celui produit par toute la presse mondiale (sources : university of California Berkeley ; Arthur D. Little) et ce volume double tous les 14 mois (sources : NCBI ; Arthur D. Little). De fait, l'industrie croule littéralement sous les cibles potentielles mais difficilement exploitables. Dans le même temps, les concepts qui régissent notre perception du vivant et de ses dysfonctionnements ont eux aussi considérablement évolué.

Il est maintenant largement reconnu que les systèmes vivants sont

constitués de vastes réseaux d'interactions qui, par synergie, donnent naissance à des mécanismes qui sont bien plus que la simple somme des éléments qui les constituent. Ainsi, les réseaux constitués de gènes et de protéines amènent, par intégration, à des réseaux de mécanismes qui, en interaction avec les environnements intra- et extracellulaires, donnent naissance à des réseaux fonctionnels. Les conséquences d'un élément génique donné sur la physiologie d'une cellule ou d'un organisme dépendent donc de la façon dont cet élément s'intègre dans les différentes couches de réseaux interactifs et synergiques. Il en résulte que les corrélations entre gènes et dysfonctionnements sont souvent très indirectes (eg : les très nombreuses maladies dites « multifactorielles ») et leurs thérapies demandent plus qu'une intervention sur le ou les produits des gènes corrélés. Les liens qui relient le génome et le physiome étant de nature non-linéaire, il devient nécessaire de manipuler l'activité biologique en plusieurs points

des réseaux. Ceci requiert une dissection des circuits et réseaux concernés de façon à générer un modèle fonctionnel capable de proposer des modes d'interventions thérapeutiques physiologiquement cohérents.

Du gène aux fonctions physiologiques : une série de démultiplications associées à des discontinuités

De nombreux gènes peuvent encoder, par épissage alternatif, plusieurs ARN messagers distincts (figure 1A). De nouveaux transcrits, émanant souvent de gènes dont la séquence est connue depuis longtemps, sont fréquemment découverts. Ceci démontre qu'il est impossible de déterminer, à partir de la séquence d'un gène, quelles sont les formes alternatives de messagers que ce gène peut encoder. Dans le parcours du gène à l'ARN messager, nous sommes donc face à un phénomène de démultiplication (un gène = plusieurs messagers potentiels) associé

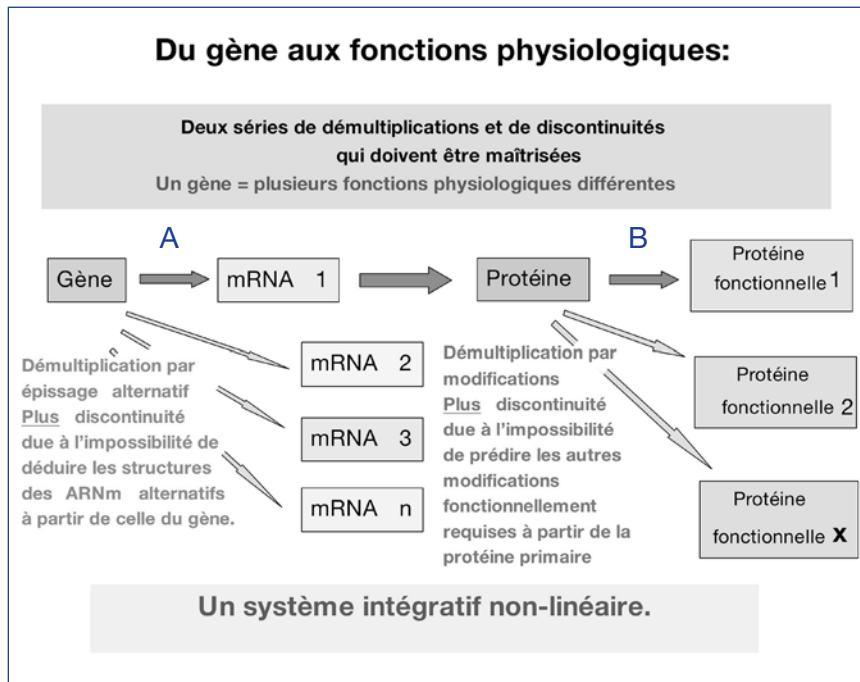


Figure 1.

à une discontinuité. En effet, à partir de la séquence d'un gène, on ne peut prédire que le messager le plus long pouvant être produit et non pas les autres messagers plus courts produits par épissage.

Des phénomènes semblables sont rencontrés dans le parcours allant du messager à la protéine biologiquement active. Chaque messager est traduit en une protéine primaire qui, pour devenir biologiquement active doit subir des altérations (modifications post-traductionnelles). (figure 1B). Par exemple l'insuline, produite sous une forme inactive (la pré-pro-insuline), doit subir, pour devenir l'hormone active, plusieurs modifications, dont par exemple l'amputation d'un segment interne. Il existe au moins six différentes formes de modifications post-traductionnelles, chacune entraînant des conséquences fonctionnelles plus ou moins profondes, et de nombreuses protéines sont sujettes à plus d'une forme de modifications. Comme dans le cas précédent, la conséquence est l'impossibilité

de prédire, à partir de la séquence primaire, les modifications que devra subir une protéine pour une fonction déterminée. On est donc à nouveau en présence d'une démultiplication (une protéine « native » = plusieurs formes fonctionnellement différentes) associée à une discontinuité.

Ce phénomène n'apparaît pas seulement au niveau des protéines, mais affecte aussi les complexes protéiques. De nombreuses protéines, en effet, ne peuvent agir qu'en interaction avec d'autres protéines, formant ainsi des complexes capables de moduler les activités physiologiques. En fonction des partenaires qui constituent ces complexes, ceux-ci peuvent aboutir à des fonctions semblables ou très différentes selon les contextes intracellulaires locaux, ce qui influe directement sur l'expression future des gènes. On se trouve donc, à nouveau en présence d'une démultiplication associée à une discontinuité.

Environnements et physiologie cellulaire

Dans tous les systèmes eucaryotes et procaryotes, les protéines donnent naissance à des groupes d'interactions qui entraînent des événements intracellulaires soit d'ordre structurel (récepteurs, construction de fibres, de parois, de membranes, de tubes...) soit d'ordre biochimique (phosphorylation, myristilation*, transformation du glucose en acides organiques, en alcool, en polysaccharides, etc.) avec consommation d'énergie mais aussi production d'énergie sous forme latente (ATP/GTP, acides gras, amidons/glycogène, etc.). Ces événements entraînent à leur tour des changements dans la nature des environnements locaux (composition et structure du cytosquelette, tonicité ionique, flux membranaires...). Ces changements d'environnement vont influencer de façon différentielle non seulement sur les potentiels d'interactions entre protéines, mais aussi sur les événements en cours et surtout sur la nature et les types d'événements qui pourront apparaître dans le futur immédiat. Ceci implique non seulement la transcription même des gènes, mais aussi l'état physique et la durée de vie des ARN messagers et des protéines. La simple naissance d'un événement d'ordre biochimique induit donc une cascade d'effets différentiels qui vont entraîner des vagues de modifications contextuelles. Ici, les environnements locaux et globaux jouent des rôles fondamentaux et complémentaires. Alors que les environnements locaux définissent les limites du spectre d'événements qui restent possibles dans le futur immédiat, l'environnement global définit les limites du potentiel évolutif à court terme d'un système dans son ensemble. Ce sont donc les contextes, issus des ensembles d'événements,

* **myristilation** : liaison covalente d'une molécule de myristil (alcool gras constitué d'une chaîne de 14 carbones) à une protéine, permettant son insertion dans une membrane cellulaire.

et leurs interactions, qui définissent les relations entre éléments et leurs fonctions, et non pas l'inverse.

Modélisation biologique, intégration des systèmes et validation expérimentale

Le problème majeur aujourd'hui n'est plus la production d'informations fiables mais bien l'analyse, l'interprétation et l'intégration des données en des ensembles cohérents amenant à la production de connaissances. Il existe un besoin criant d'outils d'analyses intégratives qui, par le biais d'une exploitation rationnelle des informations, entraîne une réduction sensible du temps de développement du médicament, un meilleur ciblage thérapeutique et un meilleur taux de succès en phases cliniques.

Une telle approche intégrative et systémique des phénomènes biologiques requiert la collecte et l'organisation dynamique de l'ensemble des informations, quelles qu'en soient les sources, concernant la totalité des composantes biologiques. Elle doit en outre, sans *a priori*, mettre en relief les interactions entre les composantes et leurs interdépendances contextuelles. Il s'agit de générer, à partir de la littérature scientifique, un ensemble de connaissances pertinentes permettant la construction de modèles biologiques. Ceux-ci doivent amener à une description détaillée des mécanismes physiopathologiques, et suggérer des cibles et des modes d'interventions thérapeutiques physiologiquement cohérents. C'est un changement majeur de paradigme permettant, par un processus de modélisation des fonctionnements physiologiques, d'accéder à un niveau supérieur de compréhension du vivant.

Un prototype analytique semi-automatisé, permettant ce processus de modélisation (figure 2), a été développé par des biologistes

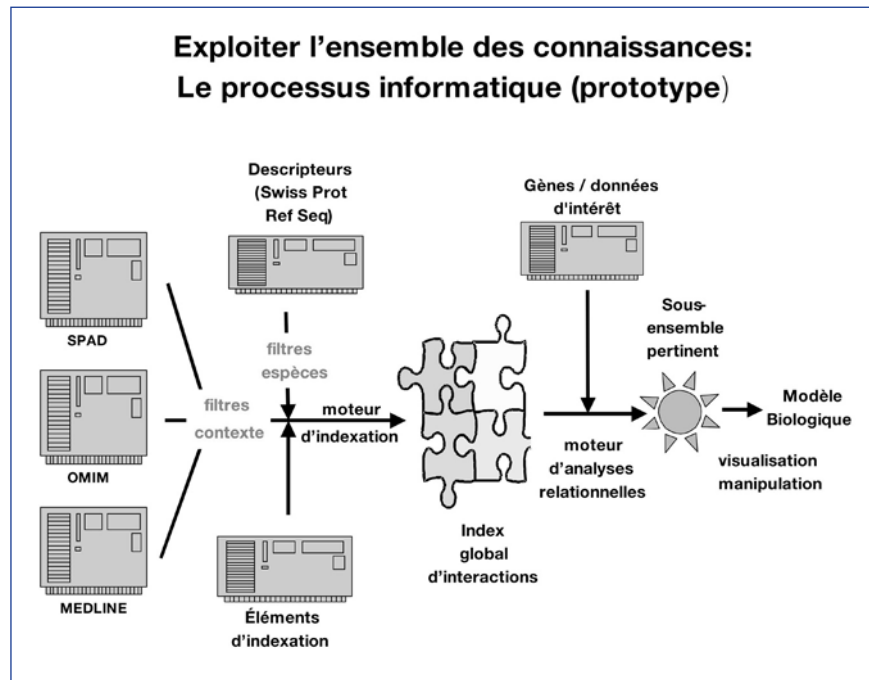


Figure 2.

et informaticiens, travaillant en collaboration depuis 1994, et qui en sont maintenant devenus les utilisateurs. Le processus d'intégration commence par la définition du problème bio-médical devant être traité (cancer du sein ras-dépendant, hypercholestérolémie familiale, détoxification d'une molécule déterminée, etc.) ce qui permet de définir des contextes physio-pathologiques (figure 3A). Toutes les informations, stockées sur les bases de données publiques, ayant un lien quelconque avec ces contextes sont rapatriées et structurées sous forme de base de données indexée (en termes de composantes biologiques, d'interactions, d'interdépendances, etc.). Si des données expérimentales sont disponibles, elles sont injectées dans cette base de données et sont prises en compte dans une nouvelle indexation (figure 3B).

Les arborescences sont ensuite visualisées et les éléments donnant naissance aux embranchements sont utilisés pour générer des hypothèses fonctionnelles (figure 3C). Chaque hypothèse ainsi générée est alors testée contre l'ensemble des bases de données publiques, le but

recherché ici étant la destruction de l'hypothèse (figure 3D). Les éléments d'information permettant de la détruire sont alors utilisés, conjointement avec les éléments d'arborescence, pour générer une nouvelle hypothèse qui est à son tour soumise au processus de destruction et ainsi de suite, de façon itérative, jusqu'à obtention d'une hypothèse ne pouvant être détruite (figure 3E). Les hypothèses non détruites accumulées sont alors intégrées en métahypothèses qui sont à leur tour soumises au même processus de destruction. Ce procédé, qui est étonnamment rapide (moins d'un an pour un désordre complexe tel que le cancer du sein), amène à la construction d'un modèle biologique, souvent très détaillé (figure 3F), ayant trois caractéristiques majeures :

- Il est entièrement documenté par un ensemble important d'informations publiées et bien identifiées,
- Il suggère des mécanismes cohérents pour de nombreuses observations qui restaient jusqu'alors sans explication.
- Il fait des prédictions fonctionnelles et thérapeutiques très précises qui

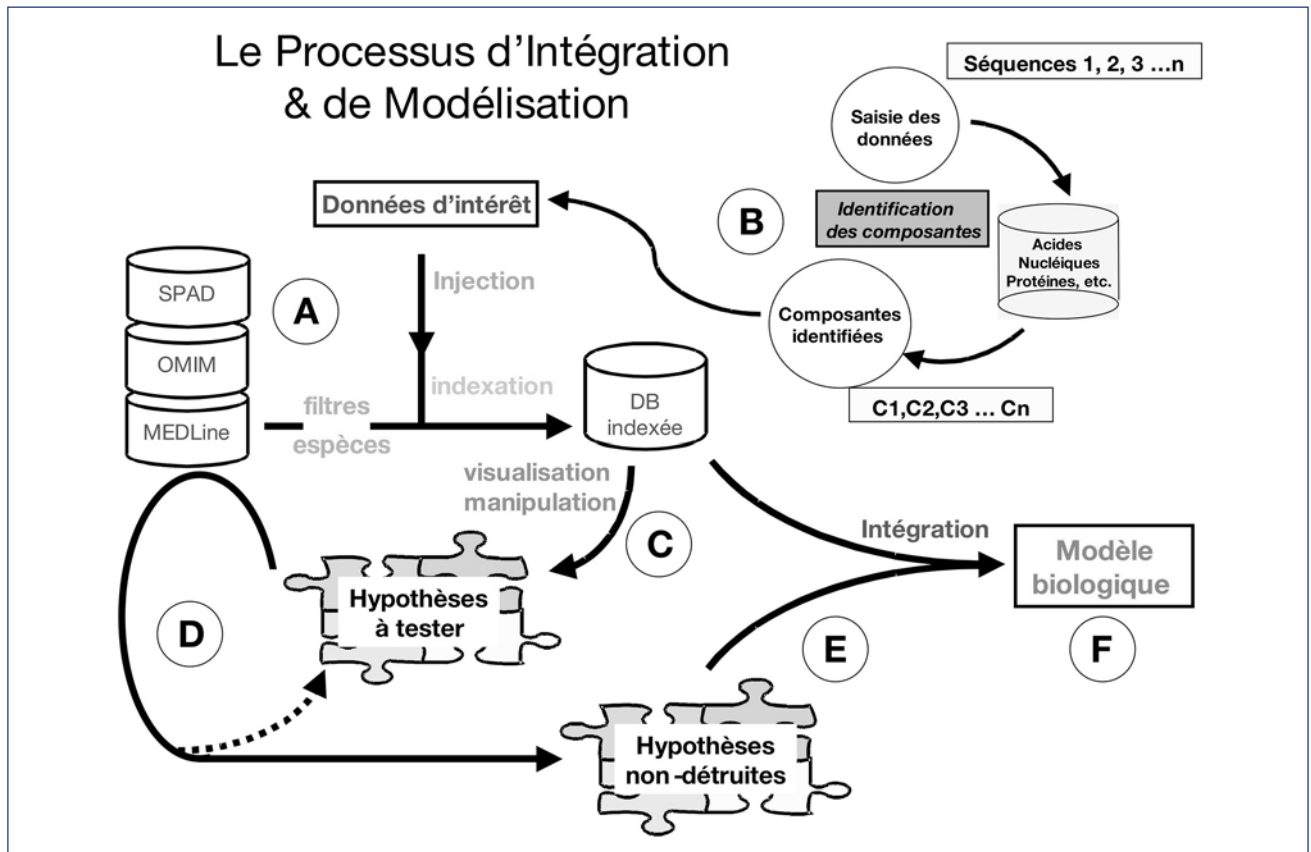


Figure 3.

peuvent facilement être testées *in vitro* et/ou *in vivo* sur modèle animal. En effet, le modèle suggère clairement ce qui doit être vérifié, où, quand, comment et pourquoi, ce qui laisse très peu de place à l'incertitude.

Modèles biologiques intégrés et applications

Ce type de modèle intégré n'est en aucune façon l'équivalent d'une « cellule virtuelle ». Bien que décrivant des enchaînements de mécanismes intracellulaires associés à un état physiologique défini, les modèles produits intègrent aussi des événements appartenant à des types cellulaires ou tissulaires différents, affectés par ce changement d'état (eg : hypercholesterolémie

familiale, maladie de Creutzfeld-Jakob...).

Ce modèle est donc à la fois :

- une représentation dynamique dans le temps des enchaînements de mécanismes intracellulaires associés au développement et à l'instauration d'un changement d'état,
- une image des modifications physiologiques entraînés par ces changements.

Il est important toutefois de souligner qu'il reste très théorique et doit être vérifié expérimentalement.

Ce type de modèle peut offrir cinq débouchés principaux pour l'industrie pharmaceutique :

- Le sauvetage de molécules rencontrant de sérieuses difficultés en phase d'essais cliniques.

- La production de modèles physiopathologiques permettant d'identifier de nouvelles cibles et de nouveaux modes de traitement.
- La découverte de nouvelles applications pour des molécules existantes.
- L'identification des programmes de développement devant impérativement être abandonnés.
- L'identification des critères d'exclusion clés entraînant une amélioration sensible des taux de succès en phase clinique.

François Iris (Ph.D)
Laboratoire d'ingénierie des biotechnologies et santé
École Centrale de Paris
francois.iris@centrale-santé.net

Françoise Xavier
Responsable scientifique du club
Biotechnologies
xavier@ecrin.asso.fr



En complément de ce dossier vous trouverez une synthèse documentaire réalisée par l'Inist à cette adresse : <http://www.inist.fr/ecrin>