HAL
open science

De la génomique aux modèles computationnels : les outils de la physiologie contemporaine

Laurent Tiret

To cite this version:

Laurent Tiret. De la génomique aux modèles computationnels : les outils de la physiologie contemporaine. Bulletin de l'Académie Vétérinaire de France, 2005, 158 (4), pp.525-536. hal-02681077

HAL Id: hal-02681077

https://hal.inrae.fr/hal-02681077

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

De la génomique aux modèles computationnels : les outils de la physiologie contemporaine

From genomics to computational biology: the tools of contemporary physiology

Par Laurent TIRET[1]

(communication présentée le 20 octobre 2005)

RÉSUMÉ

Les 15 années charnières entre le XXe et le XXIe siècle resteront marquées par les progrès spectaculaires de la biologie moléculaire. En particulier, les techniques et concepts de la génétique moléculaire, appliqués à l'Homme et aux espèces modèles, ont permis d'identifier par l'analyse de mutants spontanés ou induits la fonction individuelle de milliers de gènes. Le balisage fin des génomes a beaucoup contribué à accélérer cette identification. Il est cependant évident que la compréhension des mécanismes de régulation du vivant doit dépasser l'étude de gènes isolés pour intégrer au contraire l'expression combinée et dynamique de l'ensemble des gènes. Dans cette veine, le décryptage des génomes a inauguré la mise en œuvre d'une série de techniques permettant de photographier en temps réel le résultat de l'expression du génome dans différents types cellulaires, éventuellement placés dans des situations physiologiques ou pathologiques différentes. L'analyse et la comparaison des profils obtenus passe par l'utilisation massive de l'outil informatique, omniprésent aujourd'hui dans la production et l'échange des données biologiques. *In fine*, il est attendu de pouvoir expliquer le fonctionnement du vivant à tous les niveaux de son organisation, de la molécule à l'organisme tout entier. Loin de projet futuristes, les premières ébauches computationnelles de modélisation physiologique ont déjà vu le jour et marquent un évident tournant dans notre approche descriptive et prédictive des phénomènes biologiques.

Mots-clés : génome, annotation fonctionnelle, génétique monogénique, physiologie, post-génomique, transcriptomique, modélisation, modèle computationnel, biologie des systèmes, physiome.

SUMMARY

The 15 years straddling the end of the XXth and the beginning of the XXIst centuries saw spectacular progress in molecular biology. The analysis of spontaneous or induced mutants in man and model species, using these new techniques and concepts of molecular genetics, has lead to the identification of the specific function of thousands of genes. Fine mapping of genomes has been a key step in such identification. However, it is now clear that understanding regulatory mechanisms cannot be achieved by studying isolated genes only, but that, on the contrary, it must focus on the combined and dynamic expression of the entire gene network. Genome sequencing projects have inaugurated a series of new techniques, photographing in real time the result of genomic expression in different cell types under various physiological or pathological conditions. The analysis and comparison of these specific patterns requires the extensive use of bio-informatic tools, already omnipresent in the production and exchange of biological data. Such research is expected to reveal the full functioning of living organisms at all levels of its organization, from molecules to the whole body. Far from being futuristic projects, the first computational attempts at physiological modeling are already in existence and constitute a clear turning point in our descriptive and predictive approach of biological phenomena.

Key words: genome, functional annotation, monogenic genetics, physiology, post-genomic, transcriptomic, models, computational models, systems biology, physiome.

(1) Maître de conférences, UP de Physiologie-Thérapeutique & UMR 955 ENVA-INRA de Génétique Moléculaire et Cellulaire. École Nationale Vétérinaire d'Alfort. 7 avenue du Général de Gaulle, 94704 Maisons-Alfort CEDEX. Courriel: ltiret@vet-alfort.fr

• INTRODUCTION

On dit des étudiants d'aujourd'hui qu'ils sont nés avec la télévision en couleur, qu'ils ont fait leurs premiers pas en compagnie des ordinateurs familiaux et qu'ils communiquent en priorité par courriel ou textos. Il faudrait ajouter que les premiers hommes qui ont marché sur la Lune sont de la génération de leurs grands-parents et qu'en pleine adolescence, ils entendaient à la radio que la première version du séquençage du génome humain était achevée. Pour eux, cette nouvelle phase de vie que représente l'entrée dans le monde du travail est donc concomitante de l'apogée des "30 glorieuses" de la biologie moléculaire. Le séquençage du génome humain est un projet scientifique remarquable tant par l'envergure de son ambition que par son assise internationale, sa performance technologique, sa vitesse d'exécution et sa vocation profondément trans-disciplinaire. Pourtant, de l'aveu même de ses auteurs, il faut reconnaître que l'identification à l'unité près de la séquence des 2 milliards 850 millions de paires de bases qui composent notre génome n'a pas permis, à elle seule, de progresser sur le chemin de la connaissance des mécanismes fonctionnels qui sous-tendent l'organisation et le maintien de la vie sur Terre. Comment tirer le meilleur parti de ces données livrées par les biologistes moléculaires, généticiens compris, depuis 30 ans ? Comment intégrer ces données de grande qualité, mais par nature réductionnistes, aux concepts de la Physiologie résolument intégrative ? Quels sont les outils modernes qui permettent de classer, comparer, lier entre elles ces données complexes livrées en quantité telle que les cerveaux humains les plus armés sont dépassés ? Peut-on arriver à quantifier ces données moléculaires et à les modéliser à chaque niveau de l'organisation du vivant au point d'élaborer des modèles computationnels simulés du fonctionnement de nos organismes ? Cet article vise à fournir quelques éléments de réponse à ces questions ainsi que des références d'entrée qui permettront à chacun de se faire une idée sur l'état actuel des outils et projets qui s'y rapportent, communément rassemblés en une discipline virtuelle appelée la post-génomique.

• LA PHYSIOLOGIE À L'ÈRE DE LA GÉNÉTIQUE MOLÉCULAIRE

Les principes fondateurs de la physiologie

Quiconque lirait des articles de synthèse récents sur des questions biologiques serait sans doute surpris d'y déceler en filigrane une certaine incompatibilité qui est soulignée entre la biologie moléculaire d'aujourd'hui et la physiologie. Dans le contexte scientifique actuel, il n'est en effet pas rare de considérer que la physiologie a été oubliée dans ces étapes descriptives des bases moléculaires de l'organisation du vivant. Comme le rappelle Jean-Didier Vincent dans une synthèse collective remise en 2000 à l'Académie des Sciences, cette crise n'est pas la première traversée par la physiologie. Au milieu du XIXe siècle, la physiologie s'est en effet trouvée dans une situation comparable lorsque les premiers chimistes organiciens ont visé par leurs réels exploits à résoudre *in vitro* tous les problèmes de la vie. Il faudra toute la puissance de la pensée bernardienne en action pour énoncer un principe qui nous apparaît aujourd'hui dans toute son évidence : il manque un élément crucial dans un tube à essai pour la compréhension du vivant, celui-là même qu'il nommera le régulateur du *milieu intérieur* (VINCENT *et al.*, 2000). Loin d'effacer la physiologie sous la densité parfois opaque de leur contenu, la chimie et la biologie moléculaire offrent au contraire aux physiologistes de mettre à jour les constantes internes et d'en percer les bases mécanistiques de leur maintien par la compensation des écarts. Ce qui est vécu comme des crises existentielles pour cette discipline correspond donc en réalité à une redéfinition permanente de ses contours, justifiée par le raffinement des outils d'ouverture des boîtes noires dans lesquelles sont enfermés les codes et les lois qui président, au delà de leur structure, à l'organisation et aux mécanismes de régulation des fonctions des corps vivants. En d'autres termes, le défi posé à la physiologie moderne est de saisir, au delà d'une première étape incontournable d'identification de la fonction propre à chaque gène ou protéine, comment ils s'inscrivent dans un modèle intégratif d'ensemble à l'échelle de l'organisme entier. Bien qu'individualisée de la médecine dès le milieu du XVIe siècle, comme discipline propre en charge de lui fournir une référence du fonctionnement du corps humain sain, il est remarquable de constater qu'aujourd'hui plus que jamais, la physiologie demeure indissociable de sa discipline mère. Une très grande proportion des connaissances accumulées sur la fonction normale des gènes est en effet déduite de l'identification de ces gènes et des conséquences de leur dysfonctionnement chez des individus malades (génétique inverse).

L'apport évident de la génétique moléculaire

Le séquençage du génome humain et de celui d'autres espèces, pour important qu'il soit, ne doit pas occulter la grande diversité des apports de la génétique moléculaire. Il serait vain de vouloir citer tous les champs disciplinaires concernés et nous nous limiterons à en rappeler ici quelques uns pris comme exemples, en renvoyant aux autres contributions de ce supplément le lecteur avide d'informations plus détaillées.

La transgénèse et la mutagenèse dans l'espèce murine

En une vingtaine d'années, se sont développées plusieurs techniques dont l'issue est la modification du génome des plantes ou des animaux. La physiologie animale doit beaucoup aux expériences de mutagenèse, prise au sens large, chez la souris. Les cribles phénotypiques de mutants spontanés ou résultant de l'action d'agents chimiques et physiques permettent d'identifier de nombreux gènes impliqués dans la physiologie du développement, la physiologie des grandes fonctions, le métabolisme ou les fonctions d'intégration. Leur grand intérêt réside dans la possibilité qu'ils offrent de découvrir des acteurs fonctionnels sans aucune connaissance structurelle ou fonctionnelle préalable (Revue sur la mutagenèse à la N-éthyl-N-nitrosourée et références dans GUENET, 2004). La transgenèse soit par microinjection dans l'ovocyte fécondé soit par électroporation directe dans un tissu permet quant à elle de forcer l'expression d'un gène dont la séquence est connue

(KOOPMAN *et al.*, 1991). Elle permet ainsi, dans ses applications les plus courantes, de rechercher l'effet résultant du gain de fonction d'un allèle sauvage par surexpression, de suivre l'expression dudit gène par un système rapporteur ou de confirmer, chez la souris, l'effet d'une mutation dominante préalablement identifiée chez l'Homme. La mutagenèse dirigée permise par la recombinaison homologue dans les cellules embryonnaires souches est un raffinement des techniques de transgenèse. Elle est maîtrisée depuis 15 ans. Il devient alors possible de modifier la séquence d'un gène *in situ*, ce qui revient entre autres applications à pouvoir l'invalider (obtenir une lignée KO, pour *Knockout*). Cette technique a permis d'obtenir à ce jour un millier de lignées. Prenons un exemple qui illustre l'apport majeur des mutants murins autour d'une question pourtant déjà bien défrichée, celle de l'action des hormones stéroïdes d'origine gonadique sur les caractères sexuels. Peu à peu, les physiologistes ont approché l'importance fonctionnelles de ces hormones à partir d'expériences convergentes. Ainsi, l'injection d'hormones stéroïdes anabolisantes modifient les masses musculaires, modifie la structure histologique du tractus génital chez la femelle et perturbe le cycle ovarien. Les effets de la gonadectomie peuvent aussi être observés et très finement décrits. Cependant, trois types de problèmes émergent de ce type d'analyses. Le premier provient de ce que l'effet des molécules injectées, souvent à des doses supra-physiologiques, est toujours sujet à polémique. On ne peut nier qu'un androgène à forte dose exerce une action sur la gonade, mais exerce-t-il une action similaire lorsqu'il est sécrété à une dose physiologique ? Le second problème provient de ce que les effets de la gonadectomie sont le résultat de la suppression de l'organisme d'un nombre très important d'hormones : citons rapidement les androgènes et œstrogènes, la progestérone, l'inhibine, l'hormone anti-müllerienne, l'aromatase etc… comment, dans ces conditions expérimentales, conclure sur l'action individuelle de chacune de ces hormones ? Troisième limite non négligeable : le fait que retirer une gonade ne permet plus, c'est une évidence, de voir les conséquences du retrait des hormones gonadiques sur la gonade elle-même. Les mutations moléculaires ciblées ont cet atout de pouvoir permettre l'invalidation d'une protéine et d'observer les effets de ce retrait moléculaire sur le tissu cible, voire même sur la cellule productrice elle-même puisqu'elle n'est pas physiquement retirée de l'organe. Ainsi, des souris femelles après la puberté qui portent des allèles nuls pour les récepteurs nucléaires alpha et bêta aux œstrogènes ont des ovaires hyperhémiques dépourvus de corps jaunes, dans lesquels abondent des follicules au stade antral. En d'autres termes, les follicules n'ovulent pas. Plus tard dans la vie de ces femelles, les ovaires subissent une transdifférenciation partielle avec des cellules de la granulosa qui expriment des marqueurs spécifiques de cellules mâles de Sertoli (COUSE *et al.*, 1999). Au bilan, seule l'analyse de ces mutants aura permis de souligner que chez la femelle, les œstrogènes ne joue aucun rôle sur la différenciation de l'ovaire au cours de l'embryogenèse, jouent en revanche un rôle essentiel dans la différentiation terminale des follicules et l'ovulation, puis dans le maintien du phénotype ovarien chez la femelle adulte. On comprend par de tels exemples la puissance du modèle murin doté d'outils de dissection moléculaire uniques. Cela est d'autant plus vrai qu'une seconde génération de mutants existent depuis quelques années, chez lesquels il est possible d'invalider ou d'augmenter l'expression d'un gène au moment et dans le tissu souhaités par l'expérimentateur (systèmes inductibles par la tétracycline et la combinaison CreloxP, par ex : MOELLER *et al.*, 2005).

La pathologie spontanée des espèces modèles émergentes

Le modèle murin s'impose donc par l'attrait quasi-hypnotique qu'il exerce auprès des biologistes désireux de savoir le plus vite possible comment un embryon se développe et, s'il est viable, comment il vit en l'absence du gène invalidé. La souris ne saurait cependant à elle seule permettre de définir la fonction de tous les gènes, au moins du point de vue de la physiologie humaine dont elle se démarque en partie. Il suffit de rappeler que ce rongeur est un animal qui vit la nuit, moins de 2 ans, possède un cœur qui bat à 600 battements par minutes, donne naissance à des portées de 8 à 10 petits et métabolise les médicaments avec une cinétique différente pour apprécier quelques unes des limites du modèle. C'est pourquoi il importe de maintenir et de développer la culture d'une diversité des modèles d'étude, chacun pouvant être désigné comme le plus approprié pour répondre à une question particulière, qu'il s'agisse d'annotation des gènes ou d'analyses pharmacologiques. Dans le domaine bio-médical, les raisons du choix rationnel d'une espèce modèle plutôt qu'une autre proviennent d'une fiabilité de transposition à l'Homme des données obtenues, à la volonté de réduire le nombre d'animaux utilisés. Les résultats provenant d'une espèce mal choisie doivent être repris dans un second temps, multipliant le nombre total d'animaux utilisés. Il est rassurant de constater que ces évidences énoncées sont aussi travaillées sur le terrain. Par exemple, un choix dichotomique des espèces les plus adaptées aux études pharmacologiques qui mettent en jeu les différentes classes d'enzymes hépatiques dépendantes du cytochrome P450, a été proposé et souligne clairement certaines incompatibilités (ZUBER *et al.*, 2002). Cela est aussi vrai dans le domaine de l'étude des maladies. Là, les modèles émergents (le mouton, la chèvre, la vache, le porc, le chien, le chat etc…) sont intéressants à deux égards. Ils offrent déjà un panel très large de maladies génétiques qui sont autant de chance d'identifier des mutations dans des gènes non encore annotés chez l'Homme ou la souris. Sans aucun doute, le modèle canin est celui dont le potentiel est le plus élevé dans ce domaine. Ensuite, chacun partage un sous-ensemble spécifique de ressemblances physiologiques avec l'Homme. Il en résulte, pour ces ressemblances, des mécanismes pathologiques également proches. C'est dire la puissance de la médecine comparée qui, associée aux analyses de génétique moléculaire, permettra d'accélérer l'identification des acteurs moléculaires impliqués dans les processus de régulation communs à l'animal et à l'Homme.

Les outils de cartographie et la séquence des génomes

La génétique inverse est un concept fondé sur le fait que l'on va identifier un gène initialement inconnu à partir de l'analyse d'une maladie. Le gène identifié sert ensuite de base pour

déduire une protéine dont la fonction sera précisée. Cette démarche repose sur l'utilisation de ressources génétiques très particulières: les cartes génétiques, les cartes physiques et la séquence des génomes. Les cartes génétiques sont à ce jour disponibles dans de nombreuses espèces. Elles permettent de localiser un locus de maladie dans un intervalle chromosomique d'abord assez large, puis réduit par l'utilisation de marqueurs polymorphes additionnels. Quand elle est disponible et ordonnée, la séquence des génomes apporte une aide précieuse aux généticiens car elle fournit l'ensemble des polymorphismes de l'ADN dans l'intervalle d'intérêt. Ce polymorphisme est exploité pour identifier entre les individus de générations successives les zones de recombinaison chromosomiques, paramètre étalon pour restreindre de proche en proche la portion de chromosome impliquée dans le développement de la maladie. *In fine*, lorsque l'intervalle sera réduit au point de ne plus englober que quelques gènes dits *candidats*, les données de séquences serviront encore pour amplifier et comparer l'ADN des individus sains avec celui des individus atteints, étape essentielle dans l'identification des mutations causales délétères. C'est ainsi qu'aujourd'hui, plus de 45 000 mutations dans 1 800 gènes sont repertoriées chez l'Homme, concourant à une meilleure annotation fonctionnelle du génome (Human Gene Mutation Database, http://www.hgmd.cf.ac.uk). Bien qu'apparaissant déjà avec beaucoup d'évidence dans la caractérisation des gènes impliqués dans des maladies, le bénéfice du séquençage des génomes ne se limite pas à cette application, comme nous le verrons plus loin.

L'insuffisance du schéma réductionniste "une maladie, un gène, une fonction"

Nous l'avons vu, identifier dans un pedigree un allèle qui ségrège systématiquement avec la maladie étudiée, et de manière cohérente, avec son mode de transmission génétique, c'est attribuer implicitement au gène considéré une action indispensable à l'homéostasie du tissu, de l'organe ou du système touché. En d'autres termes, la mutation dans ce gène conduit au déséquilibre observé qu'est la maladie, et aucune compensation n'a pris de relais fonctionnel. L'analyse des maladies monogéniques, loin d'être achevée par ailleurs, va donc permettre de révéler au mieux les protéines qui jouent un rôle irremplaçable dans l'organisme. Au contraire, les systèmes physiologiques les plus élaborés, et sans doute aussi les plus informatifs pour la médecine, sont articulés autour de dispositifs de secours. En termes moléculaires, ces dispositifs correspondent à la mise en jeu de complexes de régulation multi-protéiques. Le succès de la génétique *monogénique* ne doit donc pas faire oublier que la physiologie attend plus encore de l'analyse moléculaire débutante des maladies monogéniques avec gènes modificateurs et des maladies polygéniques (PETIT, 2000). Dans cette tâche, le généticien ne doit pas rester seul. Le physiologiste moléculaire peut aussi contribuer à l'élucidation de mécanismes plus complexes à partir des données fonctionnelles déjà collectées. Plusieurs approches sont possibles, qui toutes visent pour une protéine donnée à identifier ses partenaires biologiques et la pléïotropie éventuelle des fonctions auxquelles elle participe à l'échelle de l'organisme entier.

L'exemple de la myopathie centronucléaire canine nous permettra de mieux illustrer ce dernier point. Cette myopathie congénitale ségrège chez le Labrador retriever selon le mode autosomique récessif. Elle a été initialement identifiée et baptisée par Stéphane Blot, membre de l'Unité Pédagogique de Médecine et directeur du laboratoire de Neurobiologie à l'ENV d'Alfort. Après plusieurs années de cartographie génétique, l'UMR 955 a finalement identifié le gène responsable de cette maladie (PELÉ *et al.*, 2005). De ce gène, appelé *PTPLA*, on ne savait presque rien, sauf qu'il est normalement exprimé chez l'embryon dans les muscles squelettiques. Nous avons pu montrer que la mutation repérée produit sa quasi-invalidation fonctionnelle, dont les conséquences sont un profond remodelage structural, métabolique et fonctionnel des muscles. Dans ces conditions, ils deviennent incapables de conserver leur force habituelle et les chiens éprouvent très rapidement après la naissance des difficultés locomotrices (un film est visionnable sur le site http://www.labradorcnm.com). Le travail de génétique réalisé a donc permis d'assigner au gène *PTPLA* une fonction unique dans l'homéostasie musculaire. Ce gène est pourtant exprimé dans de nombreux autres organes, parfois à des niveaux plus élevés encore. C'est le cas du cœur par exemple (figure 1). Étonnamment, aucune anomalie cardiaque précoce n'est enregistrée chez les Labradors myopathes. Nous avons vérifié que la mutation identifiée chez le Labrador retriever produit les mêmes anomalies d'expression dans le cœur que dans les muscles squelettiques. Il en ressort que contre l'hypothèse qui pourrait être émise, la mutation dans *PTPLA* exerce des effets plus dramatiques dans les muscles squelettiques où le gène est moins exprimé. Autrement dit, il existe sans doute dans le cœur un mécanisme de compensation (une redondance fonctionnelle) qui permet de pallier la déficience de *PTPLA* et assure à l'organe un fonctionnement suffisamment normal pour ne pas induire d'anomalies détectables par les techniques d'investigation clinique de routine. On peut imaginer que dans le cardiomyocyte, la protéine *PTPLA* est exprimée dans un sous-compartiment cellulaire particulier, ou qu'elle interagit avec d'autres partenaires qui peuvent se "passer d'elle" dans cet organe. Ces hypothèses seront vérifiées par des techniques d'imagerie cellulaire, de biologie moléculaire et de physiologie cellulaire. Cet exemple illustre la complexité, au delà des fonctions indirectement révélées par le processus pathologique, d'obtenir un tableau synthétique complet de la fonction des gènes à l'échelle des différents organes du corps. De nouvelles techniques doivent émerger pour permettre de transformer l'essai du séquençage des génomes en une compréhension intégrée des phénomènes biologiques, et ce dans leur complexité.

• LES OUTILS COMTEMPORAINS D'UNE PHYSILOLOGIE INTÉGRATIVE

La transformation des schémas de pensée centrés autour de l'étude des fonctions "gène par gène" vers l'étude de multiples interactions entre les composants moléculaires et cellulaires de l'organisme s'accompagne d'une mise en œuvre de nouvelles méthodes expérimentales. Depuis quelques années, on assiste à l'émergence dans de nombreux pays d'instituts intégralement dédiés à ce type d'analyse de questions bio-

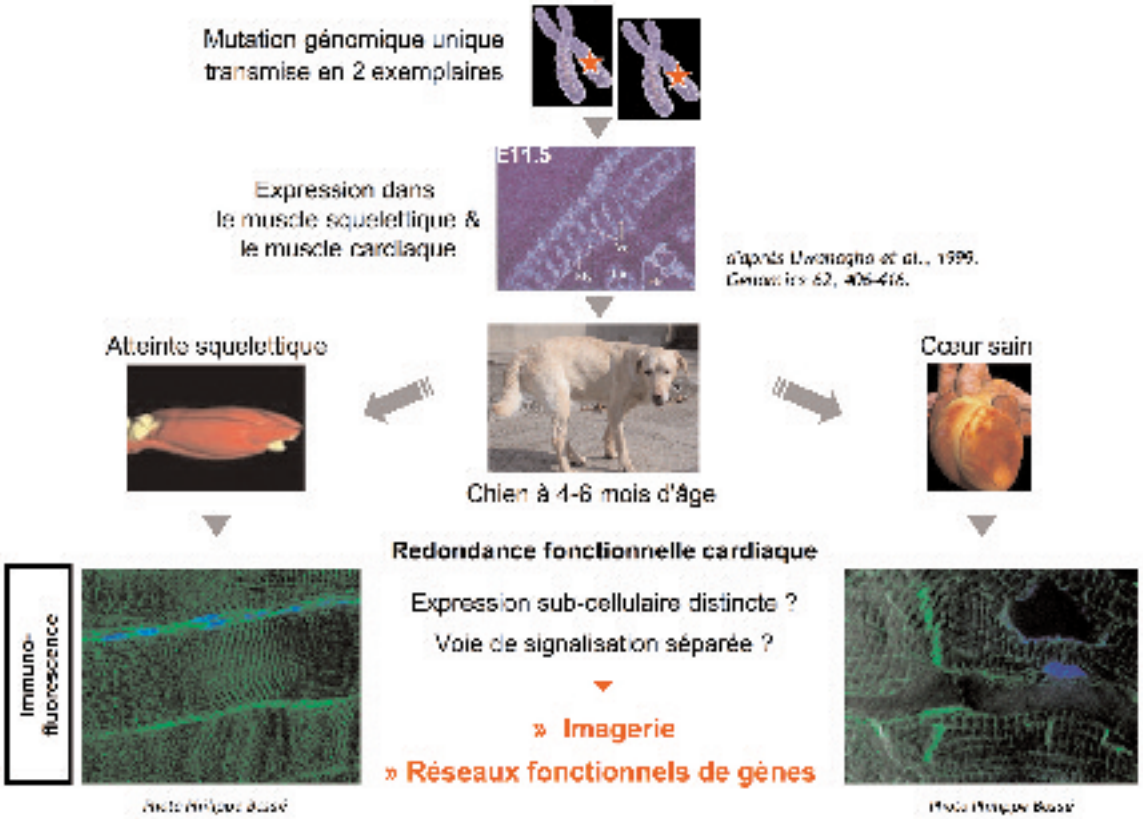

Figure 1: L'insuffisance de la génétique monogénique.
La myopathie centronucléaire du Labrador retriever résulte d'une mutation autosomique récessive qui induit une perte de fonction dans le gène PTPLA. Transmise par les deux parents (en haut), la mutation génomique est transmise à l'ensemble des cellules de l'organisme et peut donc potentiellement produire des anomalies d'expression du gène dans tous les tissus où PTPLA est exprimé. Au 12e jour du développement embryonnaire (E11.5), PTPLA est exprimé assez fortement dans les précurseurs des muscles squelettiques (My) et très fortement dans les précurseurs du muscle cardiaque (He). Chez le jeune adulte myopathe pourtant, seuls les muscles squelettiques sont touchés. La redondance fonctionnelle spécifiquement active dans le cœur désigne donc un tissu dans lequel la perte de fonction de PTPLA n'a pas de conséquence fonctionnelle précoce. La génétique inverse a donc permis d'identifier la fonction de PTPLA dans le muscle squelettique, mais ne permet pas de se prononcer sur celle qu'il exerce, sans doute de concert avec des partenaires-relais, dans le cœur.

logiques complexes telles que la biologie des cancers, la biologie du développement, la réponse aux maladies infectieuses ou la biologie des cellules souches. Les outils développés sont des techniques de biologie moléculaire à haut débit, des logiciels de valorisation du séquençage (décryptage, mise en ordre et diffusion des données par internet à l'ensemble de la communauté scientifique) et des langages informatiques spécifiques visant à modéliser le fonctionnement du vivant.

La génomique comparée et l'approche phylogénétique

Le nombre de génomes séquencés, même partiellement, permet de mener aujourd'hui des analyses de comparaison des séquences entre espèces. Ainsi, il devient possible d'interpréter la séquence d'un génome particulier dans le contexte des génomes d'espèces apparentées puis, de proche en proche, d'espèces plus distantes. Cette phylogénie moléculaire, comparée à la phylogénie fonctionnelle, permet de distinguer les domaines protéiques soumis à de fortes pressions de sélection de ceux qui tolèrent une plus grande variabilité. Schématiquement, les premiers sont les acteurs majeurs des fonctions conservées entre les espèces, tandis que les seconds permettent d'appréhender les mécanismes du bricolage évolutif ayant conduit à la séparation des phylums. La comparaison des séquences permet aussi d'observer les duplications et les pertes géniques qui ont accompagné l'évolution. Leur étude est d'un grand intérêt dans les approches ontogénétiques des mécanismes tels que le contrôle de la croissance, de la reproduction, des rythmes biologiques ou des mécanismes adaptatifs. Les duplications géniques offrent en particulier un mécanisme explicatif compatible avec l'apparent paradoxe de maintenir certaines fonctions essentielles tout en permettant des adaptations évolutives. Lorsqu'un gène est dupliqué, il s'ensuit une redondance qui peut autoriser la modification d'une des deux séquences, par exemple exprimée dans un nouveau compartiment cellulaire qui s'en trouvera modifié. En résumé, l'étude des bases moléculaires de la spécialisation des animaux adaptés à des milieux différents est un élément clé de la compréhension des principes généraux de la physiologie.

Les données qui finissent par -omes : analyse fonctionnelle à grande échelle

On sait que toute cellule sait adapter son fonctionnement, en bien ou parfois en mal, selon des éléments fournis par sa propre activité et par des stimulus extérieurs. Ce processus adaptatif est en partie contrôlé par la plasticité de l'expression du génome. Cela revient à dire qu'une cellule ne contient pas un lot de transcrits, de protéines et de métabolites figé mais au contraire remanié en permanence. Ainsi, la caractérisation de l'ensemble (indiqué par le suffixe –ome) des transcrits (transcriptome), protéines (protéome) ou métabolites (métabolome) d'une cellule à un moment et dans une situation donnés, correspond à une empreinte moléculaire spécifique de cet état particulier. En théorie, la comparaison des empreintes de deux cellules différentes, ou d'une même cellule soumise à divers stimulus, doit donc amener à repérer les seuls transcrits et protéines dont l'expression diffère entre les 2 états. Les gènes codant ces produits-là sont ceux qui accompagnent, voire induisent, les mécanismes adaptatifs des cellules. Ils sont donc les acteurs désignés pour mieux comprendre les mécanismes de transition physiologique ou pathologique des cellules.

On saisit ici tout l'intérêt de disposer d'informations les plus complètes possibles sur l'expression du génome. Le décryptage de sa séquence et le séquençage en masse de portions de transcrits ont permis de définir avec une très grande précision le nombre de gènes contenus dans le génome ainsi que les variants résultant de leur transcription. De ces données, il est possible de déduire par application des règles du décodage génétique l'ensemble des protéines potentiellement traduites dans les cellules. Technologiquement, l'analyse de l'expression simultanée de milliers de gènes est possible depuis la création des puces à ADN, également appelées DNA chips, et explique le succès de la transcriptomique. Avec ces possibilités offertes d'acquérir des données complètes sur l'expression des gènes, la résolution des questions physiologiques connaît avant tout une évolution méthodologique importante. Au contraire des expériences classiques conduites à l'échelle d'un gène unique et fondées sur le concept de la vérification d'une hypothèse initiale, ces approches globales évaluent en aveugle les variations d'expression de plusieurs gènes, sans *a priori*. Par exemple, une comparaison récente des transcriptomes de l'atrium *versus* le ventricule cardiaque a permis de révéler que chacun des deux tissus se démarquait de l'autre par la surexpression d'environ 3 000 gènes. La classification de ces gènes a conduit à confirmer que les myocytes ventriculaires expriment préférentiellement des gènes impliqués dans les mécanismes contractiles et énergétiques tandis que les myocytes atriaux expriment préférentiellement des gènes impliqués dans la réponse neuro-hormonale. De plus, ces myocytes atriaux expriment fortement des gènes impliqués dans les voies pro-fibrotiques et apoptotiques, en conformité avec les phénomènes de mort cellulaire et de fibrose matricielle observés au niveau des atriums du cœur insuffisant (BARTH *et al.*, 2005). Ces gènes, dont certains n'étaient pas encore bien annotés fonctionnellement, sont donc des cibles désignées pour identifier des futurs agents pharmacologiques efficaces dans le traitement de l'insuffisance cardiaque.

L'étude en masse des protéines et des métabolites, plus complexe, explique le moindre nombre des analyses protéomiques et métabolomiques décrites à ce jour, mais les techniques évoluent très rapidement et permettront sans aucun doute de démocratiser ces approches dans la prochaine décennie.

La surfomique : diffusion des données par internet

Pendant longtemps, les données s'échangeaient entre les biologistes exclusivement à l'occasion de discussions ou à travers la lecture des articles. Bien entendu, ces médias existent toujours mais la nature des informations et des échanges a évolué. Pour une part importante des données biologiques, la publication de résultats mis en forme, rédigés, interprétés n'est qu'une facette de ce qu'un groupe d'auteurs peut offrir à la communauté des chercheurs. Par exemple, il n'est plus rare aujourd'hui de lire un article de 5 pages qui est accompagné de 10 à 100 pages d'informations complémentaires publiées sous forme de documents électroniques, les "Supplemental data". Parfois aussi, un article n'est que l'écrin de présentation d'un site web, ce dernier renfermant la totalité des résultats et la véritable source d'informations pour les chercheurs. Un changement de méthode dans ces échanges de données est intervenu dans les années 1980, avec l'arrivée du courrier électronique, mais la véritable révolution remonte au milieu des années 1990 avec l'apparition des premières versions des navigateurs internet. Cette utilisation conviviale du réseau internet par des utilisateurs non spécialistes encourage le développement d'interfaces. Ces logiciels rendent accessibles en toute facilité des données jusque là réservées à des spécialistes. Le développement de sites de ressources biologiques a suivi une courbe exponentielle et il serait impossible aujourd'hui d'en faire une liste exhaustive. Une sélection de sites couvrant une gamme représentative de bases de données, classées par le type d'information qui y figure est fournie (tableau 1). Le grand absent thématique de cette liste serait le rayon des logiciels d'analyse de séquences, jugé trop spécialisé, mais accessible par les portails d'entrée. De plus, les sites de bases de données japonaises ne sont pas indiqués dans ce tableau. Moins utilisés quotidiennement, sans doute pour des questions de mauvaises habitudes, ils contiennent néanmoins bon nombre d'informations extrêmement précieuses.

Les deux plus importants portails d'entrée sont ceux du National Center for Biotechnology Information (NCBI), développé par le National Institute of Health (USA) et l'European Bioinformatics Institute (EBI), développé par la communauté des chercheurs européens. À partir de ces deux portails majeurs, on peut accéder aux données bibliographiques, dont la référence internationale est PubMed. Les sites OMIM et OMIA sont réservés aux maladies héréditaires humaines et animales. Ils reprennent l'ensemble des informations cliniques, génétiques et moléculaires se rapportant à une affection ou à un groupe d'affections. Des liens avec Pub Med et les autres de bases de données sont proposés, rendant la navigation très efficace. Ces deux premières séries de sites sont consultables par un public non chercheur et il est possible de télécharger certains articles au format PDF, lorsque les éditeurs ont accepté de rendre le contenu de leurs journaux accessibles à tous. Parfois, le contenu n'est librement accessible que 3, 6 ou 12 mois après publication. Les

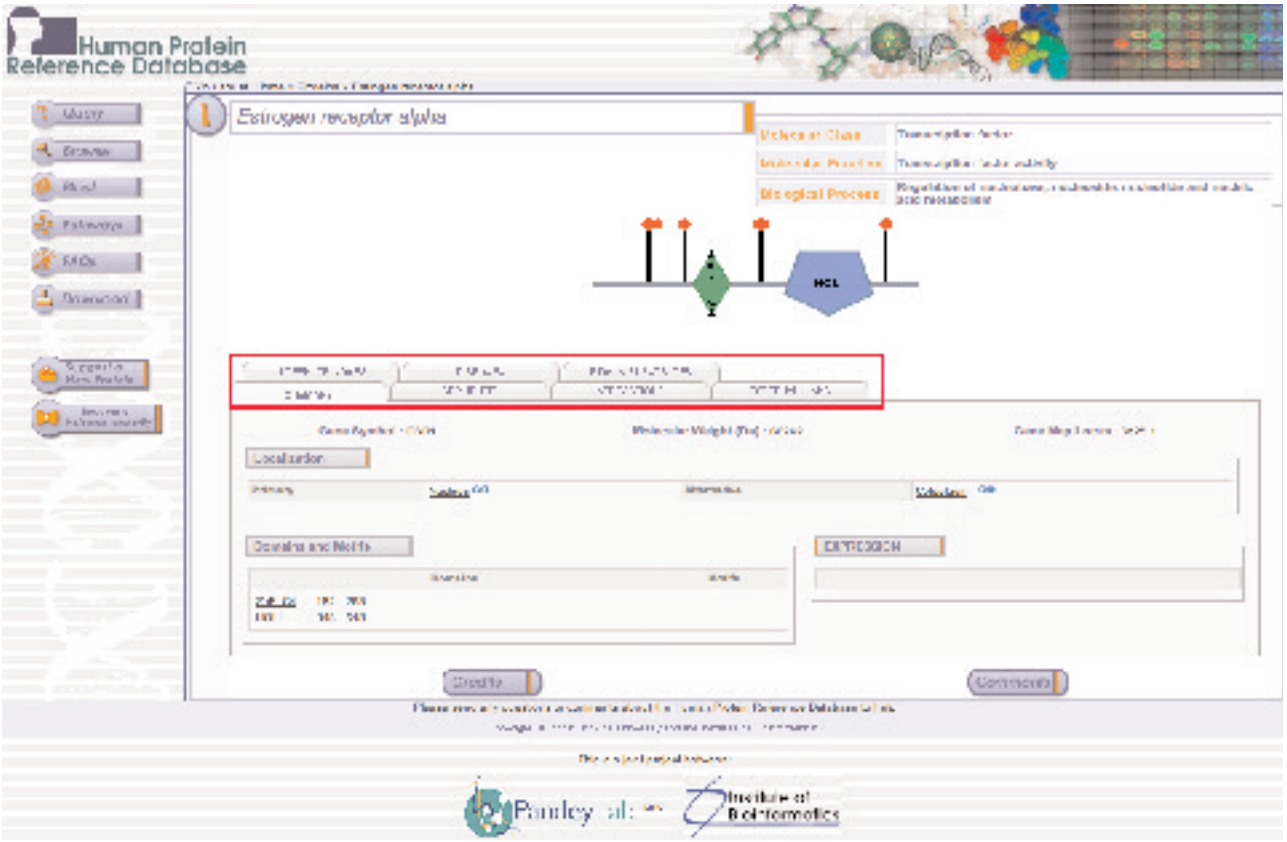

Figure 2: Human Protein Reference Database.
Cette base de données dispose d'un outil de recherche complet. L'exemple décrit ici correspond à une requête sur le récepteur alpha aux œstrogènes. La page est découpée en plusieurs zones informatives. À gauche, un menu classique d'accès aux ressources proposées par le site. En haut à droite, sont indiquées la classe moléculaire (ici un facteur de transcription), les fonctions moléculaires et les fonctions biologiques dans lesquelles la protéine agit. En dessous, un schéma de la protéine avec les motifs fonctionnels connus : ici, un motif de liaison des hormones stéroïdes (HOLI) et un motif en "doigt de zinc" permettant de lier l'ADN (ZnF_C4). En dessous, encadrée en rouge, une série d'onglets interactifs permettant de faire apparaître dans la fenêtre centrale les données connues se rapportant à la protéine, sa séquence, ses interactions, les maladies associées avec des mutations décrites, la bibliographie etc… Cette figure provient d'une page modifiée, consultable à l'adresse suivante: http://www.hprd.org/protein/00589.

informations sur les génomes et leur annotation sont destinées aux chercheurs en priorité. Des outils de recherche de séquences, de leur comparaison entre elles, de description de la structure des gènes, transcrits, et protéines sont combinés à des interfaces permettant une restitution graphique des résultats. Cette visualisation interactive est une des clés du succès des bases de données actuelles. On peut ainsi démarrer une recherche de séquence à partir d'un chromosome, cliquer sur une bande cytogénétique d'intérêt, puis sur un intervalle génétique, un gène, une paire de base et ainsi, de proche en proche, accéder à des informations sur des polymorphismes de séquence, eux mêmes connectés à des maladies recensées par exemple. Dans cette rubrique, une mention particulière sera faite pour le site Entrez – Gene, qui permet d'obtenir l'ensemble des informations connues pour un gène donné. D'un grand intérêt, la rubrique GeneRIF (pour Reference Into Function) permet d'obtenir immédiatement l'ensemble des fonctions connues pour ce gène. En face de la fonction énumérée, un lien dirige automatiquement vers l'article ayant permis de la caractériser. Cela évite un tri personnel parfois très fastidieux. Par exemple, une personne désireuse de rechercher la fonction du récepteur alpha aux œstrogènes devrait taper ESR1 dans le champ de recherche du site "Entrez Gene". Elle rencontrerait alors 80 preuves directes de son implication dans des processus biologiques. Les sites de données fonctionnelles sont également très nombreux. Un site majeur est le Mouse Genome Informatics (MGI), centré sur les modèles murins. Il collectionne l'ensemble des données sur les mutants spontanés et induits, leur caractérisation moléculaire, les gènes, les données d'expression de ces gènes, les différents allèles, leur connexion avec les phénotypes etc… Dans cette rubrique, le site de la Human Protein Resource Database est intéressant à noter. Riche de 20 000 entrées, il permet en effet, au delà des recherches moléculaires directes (par nom de gène ou de séquence), d'effectuer des recherches plus larges. Par exemple, on peut réclamer d'obtenir les récepteurs nucléaires exprimés par l'hypothalamus. Les résultats obtenus pour chaque protéine sont rendus sous forme d'une fiche synthétique assortie de liens sur la séquence de la protéine, les interactions avec d'autres protéines, les tissus dans lesquelles la protéine est exprimée etc… (figure 2).

Chaque année, la quantité de données disponibles sur le canal de la toile croît de manière très importante. Cet échange permanent est une richesse pour les chercheurs qui doivent en conséquence repenser leurs méthodes de travail.

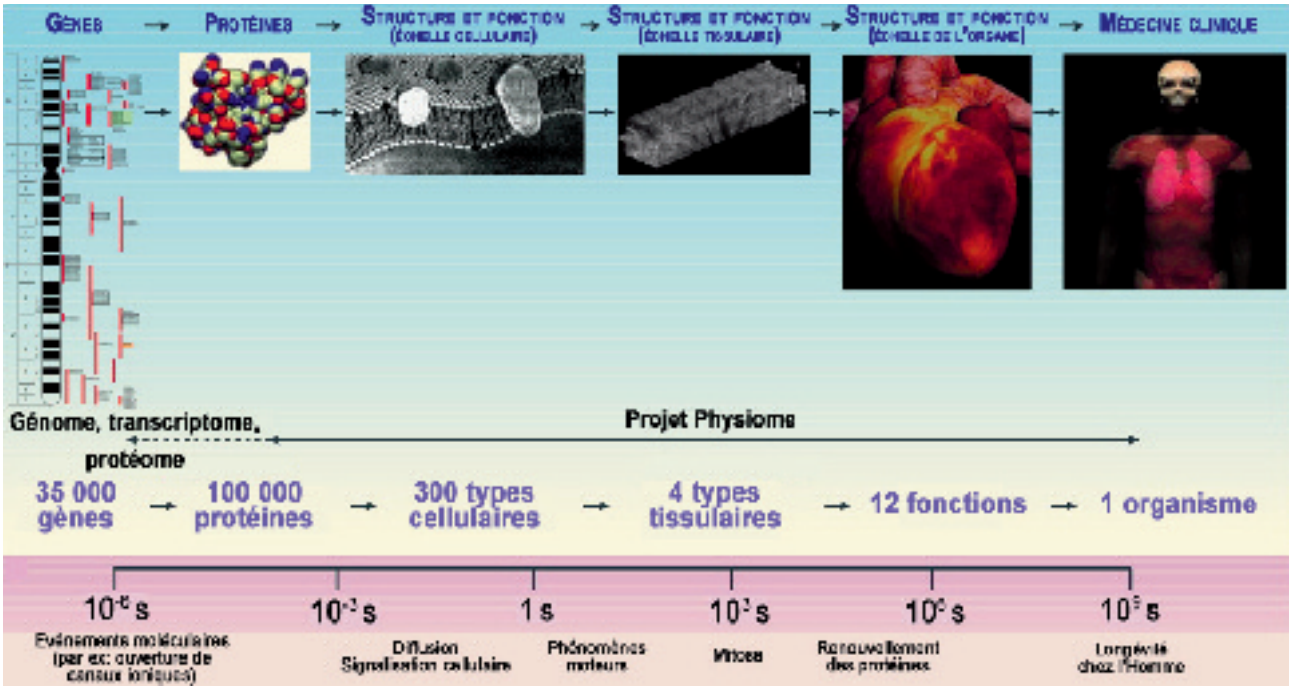

Figure 3 : Les différentes échelles du Projet Physiome.
Lier des informations moléculaires à des fonctions physiologiques nécessite de traiter les données selon plusieurs échelles spatiales et temporelles. À l'échelon moléculaire, correspond des longueurs physiques de l'ordre du nanomètre et des phénomènes d'une durée de quelques microsecondes. Les phénomènes moteurs contractiles se produisent à l'intérieur de cellules dont la taille est de l'ordre du micron, à des fréquences avoisinant les unités d'Hertz (durées ~0.1 à 1s). À chaque niveau d'organisation, correspond donc une matrice spatio-temporelle déclinée en un langage informatique spécifique. Figure modifiée avec permission d'après HUNTER et al. *(2002), Pflugers Arch –* Eur J Physiol.*, vol 445, pp 1-9. © Springer-Verlag (2002).*

Pas question par exemple de refaire une expérience dont la réponse peut être trouvée dans les archives du Net. Ainsi, le quotidien de nombreux chercheurs passe par une veille renforcée de la bibliographie et des données déjà contenues dans les bases, afin de valoriser au mieux leur temps de manipulation à la paillasse par des expériences réellement innovantes.

Les modèles physiologiques computationnels

Les premiers modèles mathématiques en physiologie datent des années 1950. À cette époque, Hodgkin et Huxley réussissent à prédire la vitesse de propagation d'un potentiel d'action dans une fibre nerveuse à partir de modèles basés sur la cinétique d'ouverture des canaux ioniques. L'existence de données dynamiques d'expression des gènes se prête aujourd'hui à la modélisation des processus sub-cellulaires et l'enjeu réside dans l'exploitation de cette mine d'informations pour décrire le fonctionnement de l'organisme à toutes les échelles de l'organisation. Il existe deux grands types de projets computationnels : l'un, retrouvé dans la littérature sous l'appelation "biologie des systèmes" et l'autre, sous la dénomination unique "Physiome". Tandis que le premier vise essentiellement à caractériser les réseaux fonctionnels de gènes à l'échelle cellulaire (ex: voies de signalisation, métabolisme cellulaire), le second ambitionne d'intégrer également des données morphologiques et cliniques à l'échelle des tissus, des organes, des fonctions et finalement du corps entier.

La biologie des systèmes repose sur des projets à très haut débit. Les trois axes fondateurs de cette approche intégrée sont les données biologiques, les outils de calcul et l'organisation de la recherche. Les données biologiques qui alimentent les bases proviennent de la transcriptomique dynamique et des interactions protéiques, c'est-à-dire de la quantification en temps réel des transcrits présents dans une cellule. Alternativement, on peut aussi évaluer l'activité transcriptionnelle du génome en repérant les facteurs de transcription fixés sur l'ADN. Ce type d'analyse a été conduit, par exemple, en vue de caractériser le réseau de gènes qui contrôle la régulation de la différenciation cellulaire lors de l'embryogenèse. Le séquençage en masse de millions de transcrits de cellules embryonnaires souches humaines et murines a permis d'identifier une classe limitée de gènes conservés dans les 2 espèces, soulevant une spécificité inattendue de ce type cellulaire (WEI *et al.*, 2005). Par la suite, la comparaison du transcriptome des cellules embryonnaires souches avec celui des mêmes cellules induites à se différencier sera une clé pour diriger la production *ex vivo*, à partir de ces cellules souches, de types cellulaires utilisables dans le traitement de maladies dégénératives telles que le diabète, les affections hépatiques, l'infarctus cardiaque, les maladies neuro-dégénératives ou l'ostéoporose. Les outils de calcul de cette Biologie des systèmes sont centrés sur la recherche et l'analyse des données moléculaires massivement produites par les étapes de génomique, transcriptomique ou protéomique. Les logiciels doivent également intégrer des données provenant de bases très hétérogènes, et générer des systèmes de présentation des données à grande échelle accessibles aux biologistes de terrain. La qualité de ces logiciels peut être évaluée sur la valeur des prédictions qu'ils établissent pour analyser de nouveaux systèmes biologiques.

Enfin, l'organisation de la recherche dans les instituts dédiés apparaît cruciale, tant elle oblige à abolir la distinction habituelle entre biologistes et informaticiens (LIU, 2005). De même, l'évaluation des carrières dans ce type de projets est considérablement modifiée. Dans ces programmes, il n'existe plus vraiment de chercheur "investigateur principal", chef d'une petite équipe bien identifiée. Au contraire, chacun devient collaborateur quasi-anonyme, noyé mais néanmoins indispensable dans une grosse équipe en charge d'un programme nécessairement collectif par son ampleur.

La première annonce du projet Physiome a été faite à Glasgow en 1993 lors du 32e Congrès de l'Union Internationale des Sciences Physiologiques (IUPS). En 2000, il a été désigné comme l'un des projets majeurs à soutenir par l'IUPS dans la décennie suivante. Le projet Physiome est à la fois la continuation des modèles cellulaires intégratifs mais ne se limite pas à l'utilisation des données moléculaires de l'expression des gènes. Il intègre également des données morphologiques (ex : histologie, anatomie), fonctionnelles (ex : biomécanique, électrophysiologie, mécanique des fluides) et cliniques (ex : imagerie médicale). Par des méthodes computationnelles, il s'agit de modéliser le corps humain en intégrant à toutes les échelles spatio-temporelles les données biochimiques, biophysiques et anatomiques (figure 3). À terme, cette modélisation doit permettre d'obtenir un système de modélisation aussi élaboré que les simulateurs de vol, poursuivant les mêmes buts d'entraînement pour les utilisateurs (chercheurs et médecins), de mise à l'épreuve de l'organisme en situation extrêmes, de validation de protocoles thérapeutiques. Si le choix computationnel s'impose, par l'évidence qu'il est particulièrement adapté à reproduire les phénomènes physiques et chimiques, il incombe aux auteurs de ce projet de remédier à une difficulté d'échelles majeure. En effet, de la molécule à l'organisme ou de l'ouverture du canal ionique à la longévité d'un homme, il y a un facteur d'échelle spatiale de 109 et un facteur d'échelle temporelle de 1015 secondes (figure 3). Un seul modèle ne peut couvrir cette étendue et l'une des premières tâches a été de développer des modèles adaptés à des sous-ensembles d'échelles (HUNTER et BORG, 2003). De plus, un organisme vivant se distingue nettement d'un avion par sa capacité unique de croissance et de remodelage en réponse aux facteurs modificateurs de l'environnement. Cela implique que les algorithmes développés en biologie computationnelle ne peuvent être simplement choisis parmi la collection existante des modèles mathématiques. Ils doivent au contraire être profondément réécrits. L'un des organes dont la modélisation est la plus aboutie est le cœur, fruit d'une collaboration entre trois laboratoires [Université d'Auckland (NZ) ; Université d'Oxford (UK) et Université de San Diego (USA)]. À titre d'exemple, pour obtenir une simulation du fonctionnement cardiaque intégré, il a fallu réaliser autant de modèles que d'éléments à prendre en considération: géométrie et structure des feuillets fibreux ; potentiel de membrane et potentiels d'action à travers canaux ioniques, transporteurs et échangeurs ; propagation de l'onde de dépolarisation; liaison du calcium à la troponine C ; cinétique des filaments fins ; cinétique de liaison des têtes de myosine; déformation élastique ; débit sanguin coronaire; compliance vasculaire ; stress compressif musculaire sur les vaisseaux. On le comprend, un tel projet repose sur une hiérarchie de modèles en liaison avec les bases de données particulières à chaque niveau d'intégration. Bien que nous soyons loin d'une modélisation complète du corps humain, les progrès enregistrés ces dernières années demeurent impressionnants et donneront lieu, de toute évidence, à des applications médicales concrètes dans les prochaines années.

• CONCLUSION

Dans un récent article, Edison Liu, Directeur Général de l'Institut de Génomique de Singapour, rappelle que la fascination humaine pour la Science provient d'un cheminement intellectuel qui part de l'observation pour élaborer une théorie de la connaissance, support d'une ouverture sur la possibilité de prédire le destin des êtres, ou plus généralement le cours de la matière (LIU, 2005). Pour nous, biologistes, la prédiction est mère d'une médecine de qualité et explique le succès de la recherche biomédicale durant les dernières décennies. Le XXe siècle restera sans doute dans l'histoire comme la période de transition entre une biologie empirique essentiellement descriptive et une biologie assise sur des principes fondamentaux, au premier rang desquels se place la structure de l'ADN, son mode d'expression et depuis peu, la conaissance de sa séquence complète à l'échelle du génome. Depuis 50 ans, la génétique moléculaire a ainsi permis de progresser considérablement sur la nature du lien qui unit l'expression des gènes aux fonctions qu'ils exercent. Aujourd'hui, on sait décrire leur fonction essentielle pour environ 10 % des 20 000 gènes que contient notre génome. La biologie du XXIe siècle aura pour mission de comprendre comment les 35 000 transcrits résultant de l'expression de ces 20 000 gènes organisent collectivement le vivant dans sa complexité et ce, en interaction permanente avec l'environnement. Cette ambition est elle-même complexe car elle est fondée sur une approche nécessairement multidisciplinaire. En particulier, l'outil informatique, par ses capacités quasi-illimitées d'analyses itératives, semble particulièrement adapté à la résolution, à l'échelle de l'organisme entier, des équations traduisant quantitativement le jeu des combinatoires moléculaires des mécanismes de régulation. Ce que, non sans une pensée visionnaire jamais encore démentie, François Jacob a nommé "La logique du Vivant".

Nom de la base	Description	URL (adresse www)
Portails d'entrée		
NCBI	Le portail du National Center for Biotechnology Information (USA)	http://www.ncbi.nlm.nih.gov/
Entrez	Moteur de recherche des sites du NCBI	http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi
EBI	Le portail de l'European Bioinformatics Institute (Europe)	http://www.ebi.ac.uk/
Infobiogen	Centre de ressources génomiques (français)	http://www.infobiogen.fr/
Données bibliographiques		
PubMed	Le site de la National Library of Medicine	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed
PMC	PubMed Central, base unique d'articles gratuits mis à disposition de la communauté par certains éditeurs.	http://www.pubmedcentral.nih.gov/
HighWire Press	Collection non redondante de journaux en accès gratuit	http://highwire.stanford.edu/
Informations sur les maladies héréditaires		
OMIM	Online Mendelian Inheritance in Man Données cliniques et moléculaires sur les maladies héréditaires de l'Homme	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=omim
OMIA	Online Mendelian Inheritance in Animal Données cliniques et moléculaires sur les maladies héréditaires animales	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=omia
Annotation des génomes		
Entrez - Genomes	Le point sur les projets abouts et en cours de séquençage des génomes eucaryotes	http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi
Entrez - Gene	Information centrée sur les gènes (NCBI)	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Gene
Ensembl	Annotation des génomes (projet européen)	http://www.ensembl.org/index.html
Analyses fonctionnelles		
MGI	Données génétiques complètes sur la souris. Séquences, mutations, expression, lignées etc..	http://www.informatics.jax.org/
HPRD	Base de données des protéines humaines : séquences, motifs, fonction, interactions, signalisation.	http://www.hprd.org/
SymAtlas	Atlas d'expression tissulaire des gènes humains et murins (Genomics Institute of the Novartis Research Foundation)	http://symatlas.gnf.org/SymAtlas/
Interactome	Banque des interactions fonctionnelles duales entre plus de 6000 protéines	http://bcfdb.dfci.harvard.edu/
GIS	Les programmes intégratifs de réseaux de gènes du Genome Institute of Singapoure	http://www.gis.a-star.edu.sg/InternetSite/
Physiome	Modèles intégratifs computationnels (systèmes)	http://www.bioeng.auckland.ac.nz/physiome/physiome_project.php
La Biobar de Mozilla	Barre d'accès direct aux bases de données, à installer dans Mozilla ou Firefox	http://biobar.mozdev.org/installation.html

Tableau 1: Sélection de sites internet de ressources biologiques, classés par thèmes.
Tous ces sites sont accessibles librement à tout public. Les adresses pour y accéder (URL, pour Uniform Resource Locators) sont indiquées. La dernière ligne donne un accès direct pour télécharger un logiciel gratuit permettant d'installer, pour les utilisateurs du navigateur libre Firefox (http://www.mozilla-europe.org/fr/), une barre de navigation spécifique – la Biobar -contenant un accès direct aux sites majeurs de bioinformatique.

BIBLIOGRAPHIE

• BARTH AS, MERK S, ARNOLDI E, ZWERMANN L, KLOOS P, GEBAUER M, STEINMEYER K, BLEICH M, KAAB S, PFEUFER A, UBERFUHR P, DUGAS M, STEINBECK G NABAUER M. (2005) Functional profiling of human atrial and ventricular gene expression. *Pflugers Arch.*, **450**, 201-208.

• COUSE JF, HEWITT SC, BUNCH DO, SAR M, WALKER VR, DAVIS BJ KORACH KS. (1999) Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. *Science*, **286**, 2328-2331.

• GUENET JL (2004) Chemical mutagenesis of the mouse genome: an overview. *Genetica*, **122**, 9-24.

• HUNTER PJ BORG TK. (2003) Integration from proteins to organs: the Physiome Project. *Nat. Rev. Mol. Cell.Biol.*, **4**, 237-243.

• KOOPMAN P, GUBBAY J, VIVIAN N, GOODFELLOW P LOVELL-BADGE R. (1991) Male development of chromosomally female mice transgenic for Sry. *Nature*, **351**, 117-121.

• LIU ET. (2005) Systems biology, integrative biology, predictive biology. *Cell*, **121**, 505-506.

• MOELLER MJ, SOOFI A, SANDEN S, FLOEGE J, KRIZ W HOLZMAN LB. (2005) An efficient system for tissue-specific overexpression of transgenes in podocytes *in vivo*. *Am. J. Physiol. Renal Physiol.*, **289**, F481-488.

• PELÉ M, TIRET L, KESSLER JL, BLOT S PANTHIER JJ (2005) SINE exonic insertion in the PTPLA gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. *Hum. Mol. Genet.*, **14**, 1417-1427.

• PETIT C (2000) De la génétique à la physiologie. *In*: GROS, F, editor. *Physiologie animale et humaine : vers une physiologie intégrative*. Académie des Sciences. RST n°2., Tec & Doc., 48-54.

• VINCENT JD, ASSENMACHER I, BUSER P, MOREL F, LAPORTE Y CORVOL P. (2000) De la physiologie. *In*: GROS, F, editor. *Physiologie animale et humaine : vers une physiologie intégrative.* Académie des Sciences. RST n°2. *Tec & Doc*. 48-54.

• WEI CL, MIURA T, ROBSON P, LIM SK, XU XQ, LEE MY, GUPTA S, STANTON L, LUO Y, SCHMITT J, THIES S, WANG W, KHREBTUKOVA I, ZHOU D, LIU ET, RUAN YJ, RAO M LIM B. (2005) Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cell*, **23**, 166-185.

• ZUBER R, ANZENBACHEROVA E ANZENBACHER P. (2002) Cytochromes P450 and experimental models of drug metabolism. *J. Cell. Mol. Med.*, **6**, 189-198.