

SUR LES NOUVEAUX RAPPORTS ENTRE DONNÉES EXPÉRIMENTALES ET THÉORIE DANS LA BIOLOGIE CONTEMPORAINE

Pierre-Alain Braillard

Chercheur indépendant

brailla6@hotmail.com

RÉSUMÉ

Cet article discute la manière dont les récents progrès dans les techniques expérimentales en biologie ont amorcé une nouvelle dynamique dans les stratégies de recherche et ont conduit à modifier les rapports entre approches expérimentales et théoriques. Un des enjeux est de comprendre jusqu'où peuvent mener des démarches de type inductif, très présentes depuis quelques années, dans l'étude des systèmes biologiques. Une autre tendance forte de la recherche contemporaine est le développement des cadres théoriques visant à donner un sens à l'avalanche de données empiriques produites par ces techniques et à en tirer une véritable compréhension des propriétés systémiques. Quelques pistes sont proposées pour mieux comprendre cette situation qui soulève un grand nombre de questions épistémologiques.

ABSTRACT

This article discusses how recent progress in experimental biology has transformed research strategies and has modified the relations between experimental and theoretical approaches. One important issue is to understand how far can inductive methods, which have recently been much discussed, lead in the study of biological systems. Another strong trend in contemporary research is the development of theoretical frameworks aiming at making sense of the tsunami of data produced by these techniques and able to gain a real understanding of systemic properties. I offer some general remarks that can clarify this situation which raises many philosophical questions.

MOTS-CLÉS

philosophie de la biologie, biologie des systèmes, biologie théorique.

1 INTRODUCTION

La biologie des vingt dernières années a été profondément transformée par une série de progrès technologiques qui ont complètement modifié la manière de conduire la recherche dans la plupart des domaines. C'est ce qu'on appelle souvent le tournant post-génomique. On peut considérer que cette révolution a été initiée en bonne partie par le programme du génome humain, lancé à la fin des années 1980 (Collins *et al.*, 2003). Ce projet titanesque, qui visait à obtenir la séquence complète de l'ADN d'un individu de notre espèce, a nécessité le développement de nouvelles technologies capables d'accélérer le séquençage de plusieurs ordres de grandeur, ainsi que des moyens informatiques pour l'analyse de ces séquences à la hauteur de la quantité de données générées. Ce projet avait la particularité de n'être fondé sur aucune hypothèse particulière, mais de consister en une sorte de cartographie complète du matériel génétique. Par son succès, il a ouvert la voie à d'autres projets d'acquisition de données à large échelle, comme je le décrirai plus bas.

Ce tournant dans les approches expérimentales soulève de nombreuses questions d'un point de vue épistémologique. S'il est clair que la manière de mener les expériences, de construire des modèles et de chercher des explications en biologie a été bouleversée, il n'est pas évident de déterminer quel est le gain qu'on peut attendre d'un tel déluge de données. Si beaucoup de scientifiques y voient la réponse à presque toutes leurs questions, certains n'hésitent pas à critiquer cette course effrénée aux données. Au-delà des difficultés pratiques liées au traitement, à l'analyse, à l'échange et au stockage de ces données, qui peuvent les rendre difficilement exploitables dans certains cas, il semble illusoire aux yeux de certains d'attendre des progrès fondamentaux dans notre compréhension des systèmes biologiques de ces approches qui sont avant tout descriptives. Ce qui semble leur manquer c'est un cadre théorique solide. C'est pourquoi on assiste en parallèle à d'importants efforts pour développer des modèles théoriques capables de jouer ce rôle.

Le but de cet article est de rapidement présenter ces enjeux en discutant la question des rapports entre données empiriques et modèles théoriques dans la biologie contemporaine. Après avoir décrit ces approches fondées sur la collecte de données (que je désignerai par l'expression anglaise « *data-driven* »), j'examinerai certaines de leurs limites, ce qui m'amènera à discuter la contribution d'approches plus théoriques. Je terminerai par quelques considérations générales sur les influences réciproques qui s'exercent dans ce contexte entre les deux types d'approches.

2 LES APPROCHES INDUCTIVES

Le projet du génome humain a joué un double rôle dans le développement de techniques expérimentales permettant d'effectuer des mesures à l'échelle de la cellule entière. Il a d'une part facilité ce développement en convaincant

la communauté scientifique de l'utilité d'immenses projets collaboratifs, très coûteux et impliquant un grand nombre d'institutions, tout en accélérant les progrès technologiques nécessaires à ce changement d'échelle. D'autre part, il a exigé la mise au point d'autres méthodes permettant de tirer parti du résultat produit. Car, en effet, une fois les séquences d'ADN obtenues, il reste encore à les interpréter, c'est-à-dire à pouvoir découvrir quel est leur rôle dans le fonctionnement de la cellule. Comment déterminer la fonction de ces séquences ?

Traditionnellement en biologie moléculaire, une séquence d'ADN est liée à une fonction biologique au terme d'une longue investigation fondée sur diverses techniques expérimentales (biochimie, génétique, biologie cellulaire, etc.). Ces approches, qui peuvent nécessiter des années pour élucider un petit mécanisme moléculaire, ne sont pas applicables à l'échelle du génome entier. D'où l'idée de développer des méthodes reprenant les ingrédients du succès de la génomique : standardisation et automatisation des procédures, collaborations internationales impliquant des dizaines de laboratoires et des milliers de chercheurs (avec des budgets inconnus jusque là en biologie) et utilisation intensive d'outils informatiques pour analyser les données produites. Parmi ces approches, les principales sont ce qu'on appelle la transcriptomique, qui consiste à mesurer l'activité transcriptionnelle des gènes (c'est-à-dire la quantité d'ARN messenger synthétisés à partir de chaque séquence), et d'autre part la protéomique, dont le but est la caractérisation de toutes les protéines présentes dans une cellule ainsi que de leurs interactions. Dès la fin des années 1990, il était devenu possible de mesurer en parallèle l'activité de milliers de gènes et d'obtenir ainsi une sorte de cartographie fonctionnelle d'une cellule. En somme, le but était que l'étude fonctionnelle du génome suive le rythme donné par les techniques de séquençage. L'ensemble de ces méthodes a pris le nom de génomique fonctionnelle (Hieter & Boguski, 1997 ; Bader *et al.*, 2003). La possibilité d'étudier les processus moléculaires avec cette facilité et cette rapidité a naturellement modifié les stratégies de recherche. Dans ce cadre, les biologistes ne partent pas d'hypothèses particulières ou de modèles qu'ils souhaiteraient tester ou affiner, mais conduisent ces expériences dites à haut-débit de manière systématique et le plus exhaustivement possible, puis tentent de dégager des corrélations dans les vastes ensembles de données produits. Beaucoup d'efforts ont porté sur les méthodes de traitement de données qui relèvent du champ de la bioinformatique. Des outils statistiques combinés à la puissance des ordinateurs doivent permettre de faire ce qui dépasse désormais complètement les capacités humaines et convertir ces données en connaissances biologiques.

Une approche qui a été beaucoup utilisée consiste à déterminer à partir de mesures d'expression des gènes quels gènes sont co-régulés, indication qui suggère qu'ils pourraient être impliqués dans les mêmes processus physiologiques ou pathologiques. Il est ainsi possible d'identifier des groupes de gènes impliqués dans une étape précise du développement embryonnaire ou dans certaines cellules cancéreuses. À une échelle plus large, le but est de

construire des modèles de réseaux d'interactions entre tous les constituants moléculaires d'une cellule, typiquement des réseaux de régulation entre gènes ou d'interactions entre protéines.

Ces progrès techniques ont fait dire à certains biologistes que le temps était venu de donner une place plus importante à des méthodes non plus fondées avant tout sur la formulation et la mise à l'épreuve d'hypothèses (comme c'est le cas classiquement en biologie moléculaire et cellulaire par exemple), mais sur une sorte d'induction plus ou moins automatisée, qui permettrait d'inférer des modèles à partir de ces vastes ensembles de données empiriques (Kelle & Oliver, 2004). L'expression « *data-driven* » renvoie à cette idée. Il est bien entendu clair que ces méthodes ne sont pas complètement libres d'hypothèses, mais le point important est qu'il n'y a pas d'hypothèses précises qui sont testées lors de ces collectes de données.

Il n'a évidemment jamais été question de devoir choisir entre approches fondées sur les données et celles fondées sur des hypothèses (« *hypothesis-driven* »), chacun reconnaissant que les deux peuvent avoir leur utilité, mais l'arrivée soudaine de ces techniques a suscité des controverses quant aux types de recherche qui devaient être favorisés. La recherche implique toujours des choix stratégiques et visiblement tous les biologistes n'étaient (et ne sont) pas convaincus de l'intérêt d'investir beaucoup de moyens dans la collecte systématique de données, sans réflexion plus approfondie sur ce qu'on pourrait en faire.

3 LIMITES ET PROBLÈMES DES APPROCHES INDUCTIVES

Bien que ces approches aient des avantages évidents et qu'elles aient produit des avancées indéniables, on peut se demander dans quelle mesure elles peuvent contribuer à expliquer le fonctionnement des systèmes biologiques dans toute leur complexité.

Je laisserai de côté un certain nombre de difficultés pratiques, comme le fait que les données sont parfois produites plus rapidement qu'elles ne peuvent être analysées, ou qu'elles puissent ne pas être exploitables par la communauté de chercheurs si les conditions dans lesquelles sont faites les expériences ne sont pas suffisamment bien définies et bien choisies.

Une des limites des ces approches apparaît lorsqu'on réfléchit à la manière dont ces données doivent produire de la connaissance biologique. On l'a dit, un des buts de la génomique fonctionnelle est d'inférer la structure et les mécanismes de régulation des systèmes biologiques à partir des données portant sur leur comportement. C'est ce qu'on appelle parfois l'ingénierie inverse. Idéalement, il suffirait d'accumuler suffisamment de données sur l'expression des gènes, les interactions entre protéines, etc., pour parvenir à construire des modèles représentant les mécanismes moléculaires sous-tendant le fonctionnement des cellules et des organismes. Dans ce cadre, les données constitueraient la matière première traitée par des algorithmes, qui les convertiraient en modèles explicatifs.

Malheureusement, différents obstacles théoriques et pratiques rendent cette stratégie très difficile et souvent impossible à mettre en œuvre. Sans entrer dans les détails techniques, disons que dans le cas de systèmes aussi complexes que des cellules, les méthodes d'ingénierie inverse nécessitent une quantité de données encore beaucoup plus grande que ce que l'on peut produire actuellement pour être capable de choisir entre les différents modèles possibles et déterminer les paramètres des modèles. De manière paradoxale, le récent déluge de données est de ce point de vue insuffisant. D'autre part, la puissance de calcul nécessaire pour mener ces analyses dépasse les capacités actuelles. Enfin, il faut également ajouter que les données produites par les techniques de génomique fonctionnelle sont relativement peu fiables et comportent beaucoup de bruit, ce qui rend leur utilisation délicate dans les procédures d'inférence.

En bref, simplement accumuler des données en espérant en tirer des modèles explicatifs grâce à des outils bioinformatiques n'est pas réaliste (ce qui n'équivaut pas à nier leur intérêt). En d'autres termes, la force brute ne suffit pas. La solution souvent adoptée est de construire des modèles extrêmement simplifiés de ces réseaux, qui ignorent complètement la nature et les détails des interactions moléculaires. On pourra par exemple représenter le fait que deux protéines interagissent sans être capable de dire quel est le sens fonctionnel de cette interaction, ni même si elle en a un.

Le problème est toutefois plus fondamental, car même en admettant que l'on dispose de bons modèles de réseaux représentant toutes les interactions entre composants d'un système, qu'a-t-on expliqué et compris sur le fonctionnement de ce système ?

4 UN RÉEL BESOIN DE THÉORIE

De nombreux scientifiques ont souligné que les approches de génomiques fonctionnelles ne peuvent en elles-mêmes, même en conjonction avec de puissants outils d'analyse bioinformatique, mener à une véritable compréhension des processus dynamiques qui produisent les phénomènes observés. Une cellule est un système complexe caractérisé par des processus hautement non linéaires se déroulant à différentes échelles spatio-temporelles. De la dynamique de ces processus émergent les fonctions biologiques. Il est nécessaire de disposer d'un cadre permettant de décrire, analyser et expliquer ces phénomènes et ces propriétés. Pour cela, les approches « *data-driven* » doivent être complétées par le développement de modèles théoriques.

L'idée qu'il est nécessaire d'aller au-delà d'une description mécaniste des systèmes biologiques est ancienne et a été souvent discutée pendant les années 1960 et 1970, mais sans grandes conséquences au niveau de la recherche dominante. Cela s'explique en bonne partie par le manque de données expérimentales (j'y reviendrai plus bas). De manière intéressante, ces idées ont commencé à refaire surface avec le tournant post-génomique. Il est devenu de plus en plus évident qu'il ne suffit pas d'accumuler des données

sur les composants et leurs interactions pour avoir une véritable compréhension systémique (par cette expression, il faut entendre une compréhension de la manière dont les propriétés à l'échelle d'un système émergent des interactions entre ses composants, ou en d'autres termes de son organisation).

Un bon exemple de cadre théorique qui semble pouvoir éclairer de nombreuses propriétés essentielles des systèmes biologiques est la théorie des systèmes dynamiques. Très utilisée dans des domaines comme la physique et l'ingénierie, elle offre de puissants outils pour analyser les systèmes complexes à dynamiques non linéaires. Elle se révèle en particulier très utile pour comprendre les boucles de rétroactions qui contrôlent la plupart des fonctions biologiques. Comme le souligne un article encourageant l'utilisation de ce cadre en biologie :

Les mécanismes de rétroaction sont la base de toutes les formes de régulation, de contrôle et de coordination à tous les niveaux dans les organismes vivants. Ils ne peuvent être compris par l'analyse conventionnelle de donnée, le traitement des données, ou l'apprentissage automatique [une méthode relevant de l'intelligence artificielle] - ils nécessitent plutôt la théorie des systèmes dynamiques. (Wolkenhauer *et al.*, 2007)

Ces méthodes ont par exemple commencé à révéler comment des circuits biologiques peuvent osciller de manière stable et robuste. Cette propriété est cruciale pour des processus comme le cycle cellulaire ou les cycle circadiens. Un autre type de comportement qui a bénéficié de ces approches est la bistabilité de certains circuits de régulation, que l'on observe par exemple quand une cellule bascule soudainement vers un nouvel état dans un processus de différenciation cellulaire. Il peut être important que le mécanisme contrôlant ces « décisions » agisse comme une sorte d'interrupteur, afin d'assurer la rapidité et l'efficacité du processus. Ce sont là des cas relativement simples, mais qu'il faut clarifier pour comprendre la plupart des fonctions biologiques.

Une question fondamentale en biologie concerne la robustesse des systèmes biologiques, c'est-à-dire leur étonnante capacité à continuer de fonctionner malgré des perturbations internes et externes (Wagner, 2005). On ne peut vraiment progresser dans la compréhension de cette propriété en amassant des données sur les composants moléculaires et leurs interactions. La robustesse peut être produite par de la redondance (par exemple nous avons deux reins et il est possible de vivre avec un seul), mais il semble que dans la plupart des cas ce qui confère à un mécanisme ou un système sa robustesse tient à son mode d'organisation. Ce peut être l'existence de boucles de rétroaction particulières dans les mécanismes de régulation, ou à l'échelle du système entier à sa modularité. Dans tous les cas, une bonne compréhension du rapport entre structures et propriétés dynamiques est fondamentale.

De l'avis d'un nombre croissant de scientifiques, la biologie contemporaine a un besoin essentiel de modèles qui font plus qu'intégrer et synthétiser les données produites par les techniques expérimentales.

5 DISCUSSION SUR LES RAPPORTS ENTRE DONNÉES ET THÉORIE

La situation rapidement esquissée ci-dessus renvoie à une vieille tension qui parcourt l'histoire de la biologie. D'une part, les approches empiriques qui ont souvent été favorisée ont tendance à être descriptives et à cantonner la biologie à collectionner les études de cas particuliers (« *stamp collecting* » selon l'expression célèbre du physicien Rutherford). Mais d'un autre côté, les approches théoriques ont très souvent suscité la méfiance et le rejet dans la communauté des biologistes, comme si cette science devait ne pas s'aventurer trop loin dans l'abstraction et la spéculation.

Depuis que la biologie existe en tant que science, on trouve des mises en garde à propos des limites des approches purement inductives. Darwin écrivait par exemple ceci :

Il y a environ trente ans, on parlait beaucoup du fait que les géologues devraient uniquement faire des observations et non théoriser ; et je me rappelle très bien quelqu'un disant qu'à ce compte, un homme pourrait aussi bien descendre dans une carrière et compter les pierres et décrire leurs couleurs... qu'il est étrange qu'on ne puisse voir que toute observation doit être en faveur ou contre une certaine opinion, pour être d'une quelconque utilité.

(Darwin, lettre à Fawcett, 18 septembre 1861).

Plus récemment, lorsqu'ils ont présenté leur interprétation non gradualiste des fossiles au début des années 1970 (la théorie des équilibres ponctués), Stephen Jay Gould et Niels Eldredge ont défendu l'idée que très souvent les progrès en science ne viennent pas de nouvelles observations mais d'une nouvelle hypothèse ou théorie permettant d'interpréter différemment des données déjà connues (Eldredge & Gould, 1972). Ils ajoutaient qu'une théorie ne peut émerger d'une patiente accumulation de données. Ils accusaient le credo inductiviste de freiner les progrès en paléontologie en détournant les scientifiques de modèles théoriques alternatifs. Si l'on regarde du côté de la biologie fonctionnelle, la révolution commencée par la biologie moléculaire dans les années 1940 doit certainement plus aux nouvelles perspectives introduites par des physiciens (entre autres) – nouvelles questions posées, mais aussi volonté de trouver d'autres principes explicatifs – qu'aux simples progrès expérimentaux en biochimie ou en bactériologie (Morange, 2003).

Si l'exemple de la théorie darwinienne montre bien que des hypothèses audacieuses peuvent permettre de donner un sens complètement nouveau à un vaste ensemble de données qui avaient été accumulées auparavant sans grande visée théorique, il faut reconnaître qu'on trouve dans l'histoire de la biologie plus de cas de développements théoriques à peu près complètement coupés de tout socle empirique et presque totalement inutiles aux expérimentalistes. C'est notamment en grande partie le cas des théories systémiques proposées vers le milieu du siècle dernier, qui n'ont pas réussi à infléchir la tendance réductionniste imposée par la biologie moléculaire, alors en plein essor.

Si une part importante de la biologie post-génomique semble, comme je l'ai décrit, plutôt appartenir à la tendance descriptive que théorique, les conditions sont néanmoins réunies pour un rapprochement entre les deux, avec tous les bénéfices qui peuvent en découler. Ces deux types d'approches ne s'opposent donc pas, mais entretiennent des relations étroites et riches, avec des influences mutuelles complexes.

Un point sur lequel j'aimerais insister, c'est que le récent intérêt pour les approches théoriques vient en bonne partie du déluge de données. En effet, les biologistes se sentent souvent un peu perdus face à cette surabondance et voient mal comment en tirer parti et leur donner du sens. Il est devenu de plus en plus évident que de nouvelles stratégies explicatives sont nécessaires pour faire face à la complexité du vivant. D'autre part, un certain nombre de cadres théoriques qui avaient été imaginés alors que les connaissances sur l'organisation des systèmes biologiques étaient encore extrêmement lacunaires, peuvent désormais établir des liens plus étroits avec les connaissances empiriques. Un bon exemple se trouve dans l'étude des propriétés de larges réseaux de régulation (principalement génétiques). Des chercheurs comme Stuart Kauffman (1993) ont ouvert la voie à de passionnants travaux théoriques dans les années 1960 et 1970, mais ceux-ci sont largement restés à l'état de spéculation, puisqu'on ne disposait d'aucun moyen pour déterminer la structure des véritables réseaux moléculaires. Depuis le tournant du siècle, les approches de génomique fonctionnelle ont commencé à révéler la structure de ces réseaux, ce qui a relancé l'intérêt de travaux théoriques sur la manière dont certaines structures (topologies) peuvent favoriser des propriétés biologiquement importantes comme la robustesse ou la facilité à évoluer selon un processus darwinien (l'évolvabilité) (Barabasi & Oltvai, 2004).

De la même manière, on trouve des modèles inspirés de la théorie des systèmes dynamiques dès les années 1960 et 1970, mais la connaissance des détails moléculaires était insuffisante pour qu'ils soient vraiment utiles aux expérimentalistes. Les choses ont bien changé et dans de nombreux cas, il est difficile voire impossible d'interpréter les données et d'expliquer le comportement de mécanismes moléculaires sans recours à ces modèles.

Ces modèles théoriques deviennent donc non seulement de plus en plus testables, mais surtout ils commencent à faire partie intégrante du cadre explicatif mobilisé par les expérimentalistes. Ils ne sont plus considérés comme de purs « jouets théoriques » par la majorité des biologistes, mais comme une contribution essentielle à leurs recherches.

Si, comme je viens de le dire, les progrès expérimentaux ont stimulé l'intérêt pour des approches théoriques, il faut également reconnaître que la nature des données produites par ces méthodes a influencé les types de modèle développés et de manière générale le cadre théorique choisi. Ici encore, le cas des réseaux est parlant. La biologie contemporaine est extrêmement influencée par la perspective des réseaux. La structure des systèmes biologiques est souvent conçue et décrite comme des réseaux complexes. Ce type de modèle permet l'application de puissants outils mathématiques et com-

putationnels, mais il est clair que l'adoption de modèles de réseaux découle en bonne partie de la nature des données issues de la génomique fonctionnelle. La protéomique décrit des réseaux d'interactions entre protéines et la conjonction de diverses données conduit à inférer des réseaux de régulation entre gènes. Le cadre des réseaux fournit ainsi un moyen naturel d'intégrer toutes ces données en des modèles cohérents et manipulables. Une de leurs fonctions est de synthétiser l'ensemble des connaissances produites par ces techniques expérimentales.

Ces modèles ont évidemment de nombreux avantages, mais le récent engouement pour l'étude des réseaux biologiques ne doit pas faire oublier qu'il s'agit d'une manière extrêmement simplifiée de représenter les systèmes biologiques, qui ignore de nombreux aspects fondamentaux de ces systèmes. Ils sont notamment très souvent statiques et représentent très mal les propriétés spatiales de ces systèmes (une cellule est extrêmement structurée et cela a des conséquences majeures pour son fonctionnement). Des modèles alternatifs nécessitent d'autres types de données, qui sont en général plus difficiles à produire à cette échelle.

Pour terminer, j'aimerais préciser le rôle que peuvent jouer les approches théoriques en biologie expérimentale. On l'a vu, inférer des modèles des systèmes biologiques à partir de données que l'on peut obtenir expérimentalement est une tâche ardue. En général, un grand nombre de modèles peuvent produire le même comportement et il faut pouvoir restreindre cet ensemble. Or, certaines expériences sont plus informatives que d'autres lorsqu'il s'agit de déterminer quel modèle décrit mieux le système. Les techniques de modélisation issues de la physique et de l'ingénierie peuvent être d'une aide précieuse lorsqu'il s'agit de mettre au point une procédure expérimentale. Elles permettent à l'expérimentaliste de décider quelles variables il va perturber et de quelles manières, dans quelles conditions les expériences seront faites, et enfin choisir quelles variables devront être mesurées et selon quel profil temporel (Wolkenhauer *et al.*, 2007).

On est donc ici bien loin de l'idée d'expérience exploratoire, menée de manière systématique sans idée précise de ce que les résultats pourront nous apprendre. Bien entendu, ces deux types d'approches ne sont pas mutuellement exclusifs, mais il paraît douteux qu'on puisse progresser de manière décisive dans la modélisation des systèmes biologiques en accumulant des données sans qu'elles soient intégrées à un cadre théorique solide.

6 CONCLUSION

Les remarques faites dans cet article ont été trop brèves et schématiques, mais elles auront indiqué dans quelle mesure la biologie contemporaine connaît des transformations tant sur le plan expérimental que théorique. La vieille question de la contribution respective des méthodes inductives et des travaux théoriques se repose dans ce nouveau contexte. De manière intéressante, on voit à la fois une généralisation de projets de collectes de données se

réclamant d'une démarche inductive et libre d'hypothèses et d'autre part une multiplication des travaux théoriques qui ambitionnent de dépasser le simple catalogage des composants moléculaires pour atteindre une compréhension générale des principes d'organisation fonctionnelle des systèmes biologiques. J'ai suggéré que ces approches empiriques et théoriques ne s'opposaient pas, mais entretenaient des relations complexes qui méritent l'attention de l'épistémologue.

BIBLIOGRAPHIE

Bader, G.D. *et al.* (2003), « Functional genomics and proteomics : charting a multidimensional map of the yeast cell », *Trends in Cell Biology*, 13, 344-356.

Barabasi, A.-L. & Oltvai, Z.N. (2004), « Network biology : Understanding the cell's functional organization », *Nature Reviews Genetics*, 5, 101-113.

Collins, F.S. *et al.* (2003), « The Human Genome Project : Lessons from large-scale biology », *Science*, 300, 286-290.

Darwin, C. (1861), lettre à Fawcett, 18 septembre 1861, *Darwin Correspondence Database*, [URL](#).

Eldredge, N. & Gould, S. J. (1972), « Punctuated Equilibria : An Alternative to Phyletic Gradualism », dans T. Schopf (eds.), *Models in Paleobiology*, San Francisco, Freeman, Cooper and Co.

Hieter, P. & Boguski, M. (1997), « Functional genomics : It's all how you read it », *Science*, 278, 601-602.

Kauffman, S. (1993), *The origins of order*, Oxford University Press.

Kell, D. & Oliver, S.G. (2004), « Here is the evidence, now what is the hypothesis ? The complementary roles of inductive and hypothesis-driven science in the post-genomic era », *Bioessays*, 26, 99-105.

Morange, M. (2003), *Histoire de la biologie moléculaire*, Éditions La Découverte.

Wagner, A. (2005), *Robustness and Evolvability in Living Systems*, Princeton, NJ, Princeton University Press.

Wolkenhauer, O. *et al.* (2007), « A plea for more theory in molecular biology », dans Bringmann, P., Butcher, E.C., Parry, G., & Weiss, B. (eds.), *Systems Biology - Applications and Perspectives*, Berlin