

Vers une théorie probabiliste du vivant¹

Thomas Heams²

Commentaires de Franck Varenne³

Prisme N°12
Octobre 2008

¹ Ce texte est la transcription par Léo Coutellec du séminaire « Les probabilismes » des 21 mars et 5 juillet 2008 au Centre Cournot, Paris.

² Maître de conférences en génomique à AgroParisTech.

³ Maître de conférences en philosophie de la connaissance à l'Université de Rouen.

Résumé

Longtemps, la biologie a été déterministe. L'existence d'un code génétique, voire d'un programme génétique, a fait ressembler les descriptions des processus biologiques à des événements finement régulés, précis et écrits à l'avance dans notre A.D.N. Cette approche a été féconde pour comprendre les grandes lignes des processus à l'œuvre à l'intérieur de chaque cellule. De nombreux arguments expérimentaux remettent en cause les approches déterministes en biologie depuis quelques années. Une découverte récente a mis en évidence le caractère fondamentalement aléatoire de l'expression des gènes : l'enjeu est de la décrire et de la comprendre. Nous évoquons ici les causes moléculaires et topologiques qui l'expliquent, au moins en partie. Nous montrons que c'est un phénomène répandu, contrôlable, qui peut être transmis d'un gène à un autre et même de génération cellulaire en génération. Reste à savoir si cette expression aléatoire des gènes est un bruit de fond ou un paramètre biologique. Nous plaidons pour la deuxième hypothèse, en cherchant à comprendre comment ce désordre élémentaire peut donner de l'ordre. Les phénomènes probabilistes sont ainsi replacés au cœur du fonctionnement du vivant. Nous discutons en conclusion de la possibilité d'un dépassement de l'antagonisme apparent entre déterminisme et probabilisme en biologie.

Introduction⁴

L'objet de mon exposé est de vous parler de la pertinence de l'approche probabiliste pour comprendre l'expression des gènes, un sujet fondamental en biologie. Je voudrais vous montrer qu'une telle approche, que l'on peut aussi qualifier d'aléatoire ou de stochastique, est relativement nouvelle dans nos disciplines, tant en termes de programme de recherche qu'en termes de perspective pour la biologie.

Je souhaite d'abord exposer ce qu'est la notion d'expression aléatoire des gènes et ce en quoi elle est contre-intuitive, voire contradictoire avec ce que l'on a pu lire depuis des années dans la littérature scientifique en biologie. Des citations d'articles scientifiques récents laissent penser que l'idée d'un programme génétique est encore très forte. Dans cette affirmation, le grand nombre d'états cellulaires observable au cours de la vie d'un organisme et leur reproductibilité n'indiquent pas seulement l'existence de programmes mais aussi de mécanismes qui assurent leur reproduction fiable (Martinez-Arias et Hayward, 2006), l'idée classique qu'il existe un programme génétique qui gouverne l'ensemble des phénomènes cellulaires. Cette idée est fréquemment développée dans des publications prestigieuses. Elle atteint son paroxysme lorsqu'elle représente la régulation du développement embryonnaire par un réseau de relations très précises, une sorte de circuit imprimé, entre des gènes et des protéines. Cette représentation donne l'idée d'un programme, similaire à ce que l'on peut trouver dans un ordinateur, lequel permet de réaliser une fonction donnée (Davidson *et al*, 2002). Ces travaux ont tous leurs forces et leurs pertinences. De tels développements s'inscrivent dans une grille de lecture déterministe où chaque gène a un rôle attribué par un programme de manière très précise.

Si je commence par cela, c'est aussi parce que depuis quelques années apparaissent de plus en plus fréquemment des expressions et des contributions qui vont à l'encontre de ces idées. Un premier article théorique paru en 1983 (Kupiec, 1983) présente un schéma probabiliste de la différenciation cellulaire. Une expérience menée six ans plus tard est l'objet d'un article qui porte sur la transcription de n'importe quel gène dans n'importe quel type de cellule et décrit ainsi la

⁴ Note des auteurs : nous avons souhaité conserver le caractère oral des interventions. Certaines expressions, inappropriées à l'écrit ont néanmoins été modifiées dans la transcription. Merci au lecteur de bien vouloir considérer la version écrite comme faisant foi.

« transcription illégitime » (Chelly *et alii*, 1989). Cette expression est pionnière. Il faut lui rendre hommage bien qu'elle sous-entende l'existence d'une « transcription légitime » à laquelle on pourrait parfois échapper. C'est donc une vision encore déterministe, mais nuancée. À partir de 1994, on voit apparaître des articles qui nous parlent de « transcriptions de gènes individuels dans des cellules eucaryotes de manière aléatoire et rare » (Ross *et al.*, 1994). Au début des années 2000, cette conception se diffuse, on n'apprend que l'expression des gènes est un processus stochastique ou bruité (Swain *et al.*, 2002) et que les cellules sont intrinsèquement des moteurs biochimiques générateurs de bruit (Thattai et Oudenaarden, 2001). Par la suite, les mécanismes stochastiques sont présentés comme ubiquitaires dans les systèmes biologiques (Ozbudak *et al.*, 2002), ce qui est en contradiction avec tout ce que l'on a pu lire et apprendre pendant une bonne quarantaine d'années.

Que s'est-il passé pour que ce changement de perspective ait lieu ? Avant de répondre à cette question, je formule deux hypothèses concernant l'expression génétique. Selon une première hypothèse classique, dans des conditions données et dans un environnement homogène, les cellules d'un même organe qui possèdent donc exactement les mêmes gènes ont une utilisation et une expression homogène de ces gènes. Selon une autre hypothèse, les cellules pourraient avoir une utilisation de leur gène imprévisible et aléatoire, alors même qu'elles disposeraient des mêmes gènes et qu'elles seraient dans le même environnement. Ce sont ces deux hypothèses que j'essaie de mettre en balance. La question initiale devient donc « que s'est-il passé en quelques années pour que ce changement de paradigme se produise » ? Le changement de perspective intervient ainsi avec l'adoption de nouvelles techniques. Au cours de la décennie 1990, ce sont les avancées techniques qui ont permis de regarder ce qui se passait dans les cellules une par une et qui permettent de trancher entre les deux hypothèses. Si on regarde les cellules séparément dans un environnement donné, et que l'on se place dans la première hypothèse, l'observation attendue est que chacune des cellules synthétise à peu près la même chose (mêmes gènes exprimés, mêmes protéines). Ce qui s'observe au niveau de la population refléterait donc assez fidèlement ce qui se passe dans chaque cellule. La vision aléatoire, c'est-à-dire la seconde hypothèse, nous prédit que la moyenne des synthèses observée au niveau de la population cellulaire n'est bien qu'une moyenne et qu'elle cache une variabilité. Ceci ne pouvait être expérimenté pendant des années car l'expérience n'était techniquement pas possible. Il a fallu attendre l'amélioration de la microscopie,

permettant les études sur cellules isolées pour découvrir cette variabilité. Néanmoins, nous aurions pu ne jamais l'observer dans le cadre déterministe. La variabilité interindividuelle de cellules qui possèdent les mêmes gènes devient d'ailleurs un objet d'étude de la plus grande importance. Une analyse bibliométrique sommaire, à partir de la requête dans *PubMed* des occurrences conjointes « *stochastique + gène + expression* », révèle l'explosion du nombre de publications sur ces thématiques, dont la moitié remonte à moins de cinq ans.

Une première démonstration de l'expression aléatoire des gènes

Afin de démontrer concrètement qu'il existe une expression aléatoire des gènes et avant d'estimer les conséquences que cela peut avoir, il est bon de s'intéresser à des manipulations qui l'ont clairement mise en évidence. Une expérience menée par l'équipe d'Elowitz et publiée en 2002 dans *Science* pose la question suivante : l'expression des gènes de bactéries est-elle homogène ou aléatoire ? Le protocole a consisté à introduire deux gènes étrangers dans l'Acide désoxyribonucléique (A.D.N.) d'une bactérie. Ces deux gènes permettent de synthétiser des protéines fluorescentes vertes ou rouges, d'observer et de suivre *in vivo* l'expression de ces gènes. Par construction, tout laissait à penser que le niveau d'expression de ces deux gènes, dans un même milieu, n'avait aucune raison d'être différent. C'était la première hypothèse des auteurs : quel que soit le temps, apparaît à peu près la même quantité de protéines vertes et rouges. Ainsi, la fluorescence résultante, une fluorescence moyenne jaune, est toujours la même et permet de démontrer cette première hypothèse. Selon l'autre hypothèse, ces gènes s'expriment de manière aléatoire l'un par rapport à l'autre. Le fonctionnement de ces gènes est imprévisible et il n'y a donc pas de corrélation entre les niveaux d'expression de ces deux gènes, pas de corrélation entre les couleurs. Coexistent donc des bactéries plutôt vertes, d'autres plutôt rouges et d'autres jaunes. Elowitz et son équipe ont mis ces deux hypothèses à l'épreuve. La publication détaille les premiers résultats, qui dévoilent une hétérogénéité dans la fluorescence des bactéries, quand bien même elles possèdent, je le répète, toutes le même génome. Ceci montre une expression imprévisible d'une cellule à une autre de certaines de ces protéines, alors que l'on pouvait penser jusqu'à présent qu'elles fonctionnaient de manière homogène. Cette expérience n'est pas restée longtemps

inaperçue, elle a d'ailleurs fait la couverture de *Science*, ce qui n'est pas en soi un critère de scientificité, mais qui montre une évolution par rapport à une manière de voir très ancienne, la remettant en cause et donnant matière à penser.

Si ce phénomène se limitait aux bactéries, ce serait déjà très intéressant puisque les bactéries constituent à peu près 99,9 % des individus sur terre, ce n'est pas qu'un détail. De plus, ces êtres vivants se retrouvent dans tous les environnements, jusqu'aux plus hostiles. Les mouvements d'expressions aléatoires des gènes ont aussi été repérés chez les champignons, chez les animaux et chez les plantes, même si les données à vous présenter pour ce dernier ensemble sont moins nombreuses. À ce stade, l'expression aléatoire des gènes a donc été démontrée expérimentalement et c'est un phénomène répandu de manière majeure dans le vivant.

La causalité de l'expression aléatoire : origines biochimiques et topologiques

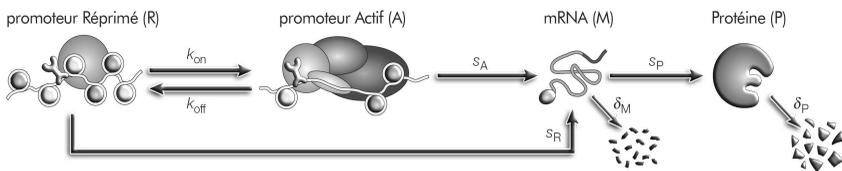


Fig.1 Processus multiétape de l'expression génétique.

Pourquoi cette expression génétique est-elle aléatoire ? Qu'est-ce qui fait qu'elle n'est pas strictement programmée comme on pourrait le penser ? Il faut faire un peu de biologie moléculaire et préciser que l'expression des gènes est un processus à plusieurs étapes (cf. fig. 1). Les gènes se situent sur l'A.D.N., ils sont transcrits en Acide ribonucléique (une sorte d'intermédiaire, une petite molécule qui ressemble à l'A.D.N., mais qui est beaucoup plus courte et qui peut sortir du noyau où elle est synthétisée), cet A.R.N. sert à fabriquer des protéines. Le passage de l'A.D.N. aux protéines n'est pas direct : le processus multiétape intervient. Il se peut d'ailleurs que cela ne marche pas parfaitement à chacune de ces étapes, qu'il y ait une certaine marge de manœuvre. Des différences d'étape en étape entre deux cellules peuvent être observées alors qu'elles sont censées posséder exactement les mêmes gènes. De manière contre-intuitive par rapport à la littérature classique, l'association des structures nucléaires

agrégées autour de l'A.D.N. n'est pas si stable que cela. Cette association est très dynamique, ce qui explique qu'elle peut créer des différences dans l'activité de la chromatine, qui peuvent avoir une influence sur les premières étapes de l'expression des gènes. En outre, les activateurs de la transcription, nécessaires pour que l'A.D.N. soit transcrit en A.R.N., sont parfois en nombre très faible, parfois en moyenne largement inférieur à un par cellule. Concrètement, dans certaines cellules *à un moment donné*, il n'y en a pas, mais dans certaines autres il y en a plus qu'un, ce qui peut créer une différence entre deux cellules. De même, le complexe de l'initiation de la transcription est lui aussi extrêmement dynamique. Tout cela consiste à construire des usines à transcription, des zones à l'intérieur du noyau, où l'A.D.N. est préférentiellement transcrit. Ceci permet localement à un moment donné, l'augmentation de la probabilité de sa transcription et, par contre coup, la diminution de la probabilité d'expression de toutes les autres zones de l'A.D.N. Toutes ces usines à transcriptions, environ 2.500 sites par noyau à un moment donné, ne représentent pas grand-chose lorsque l'on considère que l'on a 30.000 gènes. On commence à percevoir comment, topologiquement, deux cellules peuvent être hétérogènes dans l'expression de certains gènes. Il suffit que les usines à transcription ne soient pas placées exactement au même endroit. Il n'y a d'ailleurs pas de raison de penser qu'elles soient programmées pour cela.

Cette première étape permet de battre en brèche le postulat de disponibilité des molécules régulatrices. Ce postulat sous-entend qu'il y a toujours ce qu'il faut, en termes de nombre de molécules disponibles, pour faire tourner la machine. Il suffirait de donner un signal à des cellules pour qu'elles puissent exprimer les gènes qui correspondent à ce signal. En fait, c'est plus compliqué. Fondamentalement, les protéines, et notamment celles qui servent à réguler l'expression des gènes ou le fonctionnement de l'A.D.N., sont parfois en très petit nombre. Notons que 80% des protéines d'une cellule sont présents à moins de cent copies par cellule, et notamment les protéines qui sont engagées dans des phénomènes très importants comme la division cellulaire, l'initiation de la réplication de l'A.D.N. ou encore la réparation de l'A.D.N. Ces petits nombres peuvent induire un effet d'échantillonnage et des variations entre cellules. Il peut y avoir certaines cellules qui en possèdent plus que d'autres, quand bien même elles ont exactement les mêmes gènes, ce qui induit une activité d'expression des gènes différenciée.

Les origines topologiques de la composante stochastique

J'insiste sur les origines topologiques de l'expression aléatoire des gènes. La cellule est souvent représentée comme un sac, contenant de l'eau avec quelques molécules qui circulent librement à l'intérieur. C'est une représentation qu'il faut

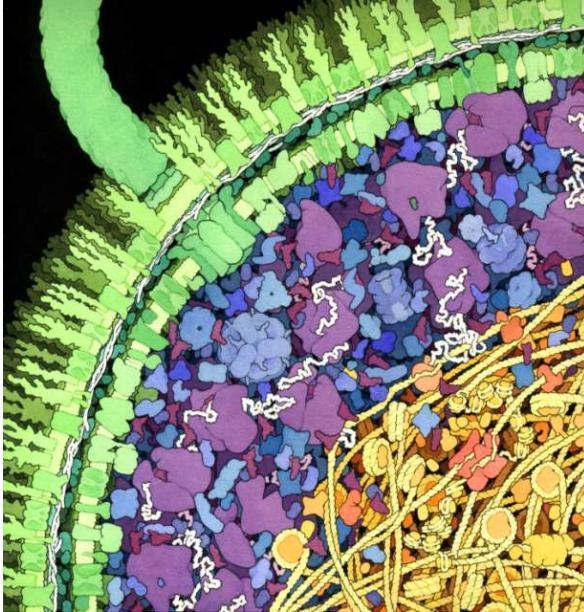


Fig. 2 Représentation de l'encombrement moléculaire d'une cellule (Goodsell, 1993).

à droite, nous voyons l'endroit où les gènes sont exprimés dans l'A.D.N. de la bactérie. L'encombrement moléculaire ne permet pas aux molécules de diffuser librement, ce qui peut provoquer une différence d'accessibilité en fonction du moment et de l'endroit où se trouve la protéine.

Je pourrais compléter cet exemple en mobilisant la thématique de ce que l'on appelle les « territoires chromosomiques », étudiée depuis une dizaine d'années. Sur la figure 3, un noyau de cellule de poulet (*Gallus gallus*) n'est pas présenté dans sa simplification habituelle, avec les chromosomes en forme de petits bâtonnets bien délimités, les chromosomes sont décondensés pour permettre l'accès aux gènes

oublier. Il suffit de regarder la figure de bactéries dessinées par le biochimiste David Goodsell (cf. fig.2) dans laquelle on constate l'encombrement moléculaire. La cellule n'est pas une sorte de piscine dans laquelle diffusent quelques molécules, mais plutôt une grosse concentration de protéines et de molécules très différentes et de grosses

tailles. En couleur claire à l'angle en bas

(« permettre » est un peu finaliste, mais c'est pour les besoins de la démonstration). Bien que décondensé, chacun des chromosomes occupe une place bien particulière à l'intérieur du noyau, un territoire qui peut être représenté comme une sorte de pelote de chromosomes. On s'est aperçu que le niveau d'expression d'un gène était fonction de la distance par rapport au centre du noyau et de la position à l'intérieur d'un même territoire chromosomique. Sur ces territoires, plus ou moins fluctuants, un gène peut, au gré des divisions cellulaires et de l'état de la cellule, se trouver plus ou moins accessible aux facteurs de transcriptions en fonction de sa position à l'intérieur du territoire chromosomique. Il ne suffit donc pas de savoir la composition génétique

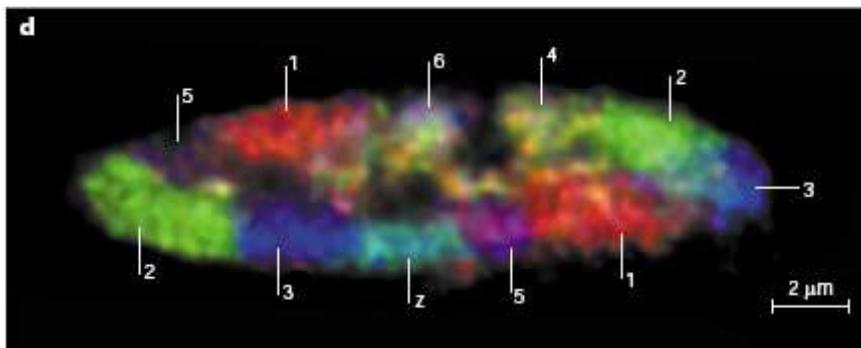


Fig. 3 L'existence de territoires chromosomiques fonctionnels chez *Gallus gallus* (Cremer *et al.* 2001).

d'une cellule donnée, la séquence précise de son génome, pour comprendre où se localise physiquement le gène à l'intérieur du noyau. Ces données semblent pourtant nécessaires à la compréhension du niveau d'expression. Même si on peut dégager de grandes tendances, ceci concourt à montrer qu'il ne suffit pas de connaître le génome d'un individu, qu'il soit bactérien ou animal, pour prédire l'utilisation de leurs gènes par les cellules de cet individu.

Les modalités de l'expression aléatoire des gènes

Nous avons pour l'instant montré que le phénomène d'expression aléatoire existait et qu'il était répandu, quelles sont ses causes biochimiques et topologiques ? Essayons de disséquer cette expression aléatoire. Cette partie aborde la question de « bruit » dans l'expression génétique. La notion de bruit peut sous-entendre une manière non aléatoire de fonctionner, le bruit serait une sorte de parasite. Ce n'est pas l'idée que je développe par la suite. Cependant, le terme de bruit permet dans cette

partie une explication pédagogique. Reprenons nos petits chromosomes bactériens avec leurs deux protéines fluorescentes. En fait, le postulat d'une expression homogène des gènes implique l'homogénéité, quelles que soient les cellules. Autrement dit, on prend le niveau d'expression de la protéine 1, celui de la protéine 2, et à la croisée de ces deux niveaux, on trouve l'endroit où a lieu l'expression génétique attendue. Nous avons vu que ce n'était pas comme cela que cela se passait. La dispersion des points est observée autour de cette valeur attendue et cette dispersion se fait dans deux directions qui permettent de distinguer deux types de bruits. Un premier bruit, que l'on appelle extrinsèque, dépend de la concentration, de l'état de la localisation des molécules intervenant dans la transcription, donc dans la première étape du processus d'expression génétique. Ce bruit est responsable de la configuration différente entre deux cellules, ou dans une même cellule au cours du temps. Un autre bruit, intrinsèque, est dû à la nature fondamentalement aléatoire des éléments microscopiques, qui crée des différences entre deux gènes « rapporteurs » (utilisés pour mettre en évidence un signal) dans une même cellule. Dans le cas qui nous intéresse, c'est plutôt ce bruit-là qui est impliqué ; une combinatoire de ces deux bruits peut néanmoins exister. Les modélisations du bruit dans l'expression génétique ont contribué à les distinguer ; elles ont montré que cela avait un intérêt fonctionnel.

Ces bruits s'accompagnent d'un temps d'autocorrélation, c'est-à-dire d'une période caractéristique de retour à la moyenne. Si vous imaginez l'expression moyenne des gènes et que vous la voyez fluctuer, selon le bruit, le temps caractéristique de retour à la moyenne est ce qu'on appelle le temps d'autocorrélation. Il est indépendant de l'amplitude du bruit. Deux bruits sont perceptibles, qui sont relativement de la même amplitude, vous pouvez avoir un temps d'autocorrélation très court et dans ce cas vous aurez plein d'oscillations. Si l'autocorrélation est plus longue, on observe plutôt des ondulations. Le temps caractéristique du bruit extrinsèque est supérieur à celui du bruit intrinsèque, de l'ordre, par exemple, de quarante minutes contre neuf minutes. Une concurrence avec le cycle cellulaire de certaines bactéries peut être ainsi observée, lorsque l'on est dans des temps de plusieurs dizaines de minutes. Ce sont des ordres de grandeur qui peuvent se chevaucher, ce qui veut dire que les bactéries peuvent se trouver modifiées par ce bruit. Il faut alors tenir compte de ce paramètre.

Le contrôle de l'expression aléatoire des gènes

Essayons maintenant de comprendre comment cette expression aléatoire peut être contrôlée. Il semble nécessaire, en tout cas en première approximation, que les cellules n'expriment pas n'importe quel gène à n'importe quel moment, mais qu'elles puissent, parce qu'elles sont des cellules de foie ou des cellules de peau, suivre le comportement de cellules de foie ou de cellules de peau relativement homogènes permettant le fonctionnement de l'organe. S'il est possible de revenir sur ce point, partons dans un premier temps de l'idée qu'il faut contrôler cette expression aléatoire.

Il y a plusieurs manières de la contrôler. La première est liée à un simple effet de moyenne. Si une cellule de foie ne sert par exemple qu'à synthétiser certaines molécules qui permettent de métaboliser ou de contrôler l'arrivée des nutriments dans le sang, on peut se dire que ce n'est pas grave si certaines cellules de foie soient plus efficaces que d'autres. L'important, c'est que dans la sécrétion globale il y ait une concentration donnée de ces molécules. L'effet fonctionnel n'a pas lieu au niveau de cellules individuelles mais au niveau de la sécrétion de l'organe, c'est donc un effet moyen. Le grand nombre de cellules est perceptible au niveau statistique. Les levures constituent un autre exemple : 75 % de leurs gènes sont à un niveau d'expression inférieur ou égal à un transcrit par cellule. Il y a donc forcément une grande variabilité de la contribution de ces transcrits (et des protéines ensuite) au sein d'une population pourtant homogène et génétiquement identique.

Il existe une deuxième manière de penser le contrôle de l'expression stochastique des gènes, c'est de prendre conscience de l'importance des étapes amont dans l'expression aléatoire. Revenons un peu en arrière, nous avons dit que l'expression des gènes est un processus multi-étape, que ce sont des « machines » qui ne fonctionnent pas à plein, et donc, si elles ne fonctionnent pas à plein, que certaines étapes peuvent être plus efficaces que d'autres. Si l'efficacité des étapes amont est bien supérieure à l'efficacité des étapes aval, on a tendance à minimiser le bruit. Toutes choses égales par ailleurs, quand l'efficacité de la transcription est supérieure à l'efficacité de la traduction, qui est une étape aval, on a une variation intercellulaire de l'expression plus faible, et inversement. On a aussi montré, chez la levure, que le bruit est pareillement limité si l'efficacité de l'activation de la transcription (étape amont) est supérieure à l'efficacité de la transcription (étape aval). Il faut bien comprendre que plus les étapes amont fonctionnent à plein, plus on peut se permettre de faire

fonctionner un peu à vide les étapes en aval et l'on assiste quand même à une diminution du bruit. Cela peut permettre de répondre à une question posée en termes évolutifs : si les étapes aval ne sont pas très efficaces, pourquoi des étapes amont qui le sont au contraire, c'est-à-dire qui produisent beaucoup de molécules, et éventuellement un surplus, ont-elles été sélectionnées si ces surplus ne sont pas utilisés par la machinerie en aval ? On peut considérer qu'ils constituent un « puits énergétique », évolutivement sélectionné pour assurer la fidélité de l'expression. Voilà pour l'importance des étapes amont.

Le contrôle de l'expression aléatoire des gènes peut également se faire par le nombre de copies d'un gène. Il a en effet été établi que l'importance du bruit était fonction décroissante du nombre de copies de ce gène (Raser et O'Shea, 2004). Plus un gène est présent en grande quantité (par exemple, cinquante fois dans un génome), plus il a une probabilité forte d'être exprimé, même s'il y a des mécanismes de canalisation qui permettent aussi d'aller utiliser des gènes qui ne sont présents qu'à une seule copie. On peut toutefois imaginer qu'avoir plusieurs copies d'un même gène peut garantir un niveau d'expression raisonnable s'il y a de l'aléa topologique dans cette expression (souvenez-vous, plus haut, de l'encombrement à l'intérieur de la cellule). C'est donc une hypothèse évolutive possible du maintien de plusieurs copies de certains gènes après leurs duplications. Ce phénomène, que l'on observe dans différents génomes, peut être aussi une explication partielle de ce qu'on appelle la polypléidie, c'est-à-dire le fait de posséder plusieurs exemplaires de chromosomes et pas seulement une paire. Ce contrôle peut aussi avoir lieu au niveau de l'emplacement des copies du gène, car la position par rapport au centre du noyau, à la périphérie d'un territoire chromosomique, peut aussi jouer sur le niveau d'expression. À supposer que ces territoires, dont on découvre la structure en ce moment, ne soient pas eux-mêmes complètement formés aléatoirement, un avantage évolutif peut ainsi placer certains gènes à certains niveaux de ces territoires. Là encore, posséder la séquence de ce génome ne permet pas, tant que l'on ne connaît pas les positions relatives des gènes les uns par rapport aux autres dans le volume du noyau, de déterminer cet avantage évolutif potentiel que peuvent représenter certaines positions topologiques par rapport à d'autres. Ensuite, chose plus classique en biologie moléculaire, nous pouvons parler de ce qu'on appelle les effets de rétrocontrôle. On peut imaginer que l'expression aléatoire des gènes est aussi l'objet de différents rétrocontrôles, notamment négatifs, très classiques en biologie moléculaire. Qu'est-ce qu'un rétrocontrôle négatif ?

Considérant un gène α qui est activé donc exprimable, on observe assez classiquement en biologie que le produit aval de la chaîne d'expression de ce gène a un effet négatif, c'est-à-dire qu'il a tendance à diminuer l'expression de ce gène α . En toute logique, quand le même gène α est moins exprimé, le produit aval est lui aussi moins présent et donc réprime moins α : l'augmentation de l'expression de α peut être constatée. Ainsi, l'expression du gène se maintient à peu près constamment. C'est ce qu'on appelle un **rétrocontrôle en constance**.

Ce **rétrocontrôle** est cependant paramétré par un certain nombre de constantes, notamment la constante d'association au gène et la constante de destruction de ce produit en plusieurs sous-produits. Ces détails importent, car le bruit est bien plus faible quand la constante d'association au gène est nettement supérieure à la constante de destruction du produit de ce gène. Ceci montre que même les mécanismes classiques en biologie peuvent contribuer à expliquer comment une expression aléatoire peut être contrôlée.

La transmissibilité d'une expression stochastique dans un réseau de gènes

Je résume ce qui précède : l'expression stochastique ou aléatoire des gènes est un phénomène répandu dans le monde du vivant, j'en ai présenté les causes et montré qu'il était contrôlé. Je voudrais maintenant expliquer, grâce à des résultats récents, que c'est un phénomène transmissible. C'est d'abord un phénomène transmissible le long d'une cascade d'expression génétique. Qu'est-ce qu'une telle cascade ? Quand on dit que les gènes fonctionnent en cascade, cela signifie que, souvent, un gène exprime une protéine et que cette protéine elle-même influe sur l'expression d'un autre gène, et ainsi de suite. Il y a ainsi plusieurs gènes intermédiaires entre le premier qui est activé et le produit final de cette cascade, une protéine. En utilisant des méthodes de fluorescence, on observe que la variabilité intercellulaire d'expression, ainsi que le temps de réponse de l'expression de cette protéine, augmentent avec la longueur de cette cascade. Ceci permet de montrer une chose toute simple, à savoir que l'expression aléatoire d'un gène peut se transmettre d'un gène à l'autre. Le bruit peut être transmis le long de cette cascade (Hooshangi *et al.*, 2005), et ce, d'autant que la cascade est longue. Nous dirons donc que le bruit est transmissible. Ce phénomène, sur lequel nous disposons d'encore peu de résultats,

est toutefois utilisable fonctionnellement.

Le bruit est transmissible au sein d'une cascade d'expression ; montrons maintenant qu'il est transmissible de générations cellulaires en générations. Le bruit est dit « héritable », l'adjectif est en soi assez surprenant. Prenons, dans le cas des levures, une cellule-mère et une cellule-fille. On s'attend à ce qui se passe chez une cellule-mère soit indépendant de ce qui peut se passer chez sa cellule-fille, devenue une levure autonome à la suite de la division asymétrique de la première. Les individus autonomes sont donc indépendants. En fait, des études récentes ont fourni des résultats contre-intuitifs. Pour les comprendre, il faut faire appel à deux informations : la première est que les levures peuvent passer d'un état « marche » à un état « arrêt ». Autrement dit, elles peuvent être dans deux états globaux d'expression différents. Dans chacun de ces états s'expriment des paquets de gènes différents. Ce sont, en sorte, deux développements différents. Par des techniques de coloration, nous sommes capables de révéler ces états : les cellules passent d'un état à l'autre de manière aléatoire. La deuxième information est que nous sommes capables, dans des conditions particulières, de suivre les généalogies des levures : on place une caméra et l'on observe le devenir des cellules, de sorte que l'on identifie quelle cellule provient de quelle autre.

Lorsque l'on cumule ces deux informations, qu'on voit des cellules qui passent de l'état « marche » à l'état « arrêt » et qu'on est capable de savoir si ce sont des cellules apparentées mères/filles ou non, on vérifie que la probabilité pour une cellule-fille de passer de l'état *marche* à *arrêt* dépend du fait que la cellule-mère est passée de l'état *marche* à *arrêt*. Autrement dit, il y a une probabilité augmentée pour la cellule-fille de passer d'un état à un autre, quand la cellule-mère y est passée (et ce, de manière imprévisible). Il y a en quelque sorte une « héritabilité » d'un comportement aléatoire. Cette chose assez curieuse et pas encore très bien comprise montre en tout cas qu'un comportement aléatoire peut être héritable d'une cellule-mère à une cellule-fille. Le fait qu'il soit transmissible nous amène tranquillement vers l'idée que ce sont des phénomènes qui ont une certaine fonctionnalité et qui peuvent être utilisables par l'organisme.

Une signification biologique de l'expression stochastique des gènes

Quelle est la signification biologique de l'expression aléatoire des gènes ? Comment montrer que la version des bruits de fond, conforme à la vision classique, n'est pas la bonne ? Selon cette vision, il est en effet envisageable de penser que les cellules ne sont pas des robots et que toutes les conditions ne sont peut-être pas rigoureusement identiques. Il est donc légitime d'imaginer, qu'à la marge, elles puissent avoir des comportements un peu différents les unes aux autres. Il y aurait un programme génétique qui impose précisément aux cellules ce qu'elles doivent faire, et puis de temps en temps elles ne font pas exactement ce qu'on attendait. C'est la première hypothèse qui permet de rester dans le paradigme classique. Comme nous avons vu que le phénomène aléatoire était répandu, contrôlé (c'est-à-dire que l'évolution semble avoir mené à des mécanismes pour le contrôler), et plus ou moins transmissible (avec toute la prudence qu'il convient d'avoir sur ce point-là), nous mettons en doute ce statut de simple « bruit ». Pourquoi n'est-il pas un paramètre biologique ? Pourquoi la cellule, voire les organismes, ne pourraient-ils pas avoir intégré ce fonctionnement complètement aléatoire comme un paramètre biologique qui leur permet d'avoir une certaine souplesse face à des variations environnementales ?

Prenons un nouvel exemple chez les bactéries. Imaginons deux réseaux de gènes, r_1 et r_2 , fonctionnant indépendamment l'un de l'autre chez la bactérie. Chaque réseau de gènes est commandé par un promoteur qui lui est propre. Au centre du réseau r_1 , nous avons un gène crucial (si ce gène n'est pas exprimé la cellule ne peut fonctionner). La manipulation consiste à faire passer ce gène crucial de r_1 dans le réseau r_2 en modifiant son promoteur, en le remplaçant par le promoteur du réseau r_2 . Bloquons ensuite le fonctionnement du réseau r_2 (c'est techniquement possible). Celui-ci, qui inclut désormais le gène crucial initialement membre de r_1 , s'éteint : tous les gènes cessent de s'exprimer. La bactérie se trouve donc privée de ce gène crucial, elle est en bien mauvaise posture.

L'expérience montre alors une chute drastique du nombre de cellules, puis, sans que les mécanismes ne soient aujourd'hui compris clairement, un certain nombre de solutions est trouvé par une sous-population de cette population initiale. Une solution émerge pour les cellules, peut-être grâce au recrutement d'un gène qui

ressemble à celui que l'on a déplacé mais qui n'était pas utilisé jusqu'à présent et permet de repartir jusqu'à atteindre un niveau d'expression tout à fait correct. Evidemment, on vérifie dans cette population d'arrivée que le gène crucial d'origine r1 était toujours bien sous commande du réseau r2 dans lequel on l'avait fait passer. Pour vérifier qu'il n'y avait pas eu une inversion, ce qui aurait expliqué le rétablissement de la population, une deuxième expérience a consisté à faire cette manipulation de façon encore plus drastique, en affaiblissant de façon plus importante la structure de ce gène crucial. On observe alors un comportement exploratoire des cellules qui permet de déboucher sur une solution. Cela semble montrer que les cellules sont capables de piocher dans leurs gènes comme dans une boîte à outils, quand bien même on en empêche certains de fonctionner. Une sous-fraction d'entre-elles (pas toutes, sinon on serait toujours dans le cadre d'un programme), retrouve alors une manière de fonctionner face à un problème d'environnement. Voici une des manières d'expliquer que le comportement exploratoire des cellules, et le fait d'aller piocher des gènes éventuellement imprévus dans un ensemble de gène que comptait un génome d'un organisme, peut être une manière de s'adapter à des variations environnementales imprévues.

L'autoamplification et l'équilibre bistable : un aléa de bifurcation

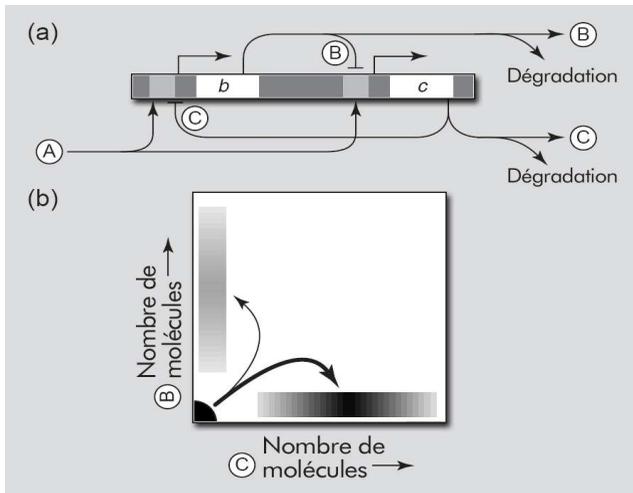


Fig. 4 Equilibre bistable

Une autre déclinaison de l'aléa des gènes apparaît dans le phénomène des équilibres bistables. Prenons un cas de figure qui peut être complètement compris dans le schéma déterministe classique de l'expression des gènes (cf. fig. 4). Soit une molécule A

capable de favoriser l'expression de deux gènes différents, un gène b ou un gène c . Le gène c , s'il est exprimé a tendance à inhiber le b , et le b s'il est exprimé a tendance à inhiber le c . C'est donc soit l'un soit l'autre. Une chose imprévisible se produit, la cellule qui possède ce réseau d'équilibre bistable, va soit vers l'expression du gène b , soit vers l'expression du gène c , de manière complètement imprévue. Deux cellules qui seraient très voisines, qui posséderaient exactement ce réseau et pour lesquelles la molécule A serait activée peuvent partir chacune dans des directions opposées en ce qui concerne l'expression soit de b , soit de c . C'est donc imprévu, aléatoire, même si le réseau est complètement déterministe. Alors que des molécules influencent précisément l'expression d'un gène, toutes ne partent pas dans la même direction. Cet équilibre bistable permet d'utiliser de manière finalement assez astucieuse une dimension imprévisible et aléatoire de l'expression des gènes.

Prenons l'exemple d'un œil de drosophile, composés de facettes que l'on appelle en termes techniques des ommatidies (fig. 5). Chacune de ces ommatidies est composée de huit cellules et parmi ces cellules deux sont responsables de la captation de la lumière. Pour qu'un œil de mouche fonctionne correctement, il faut que 30 % de ses

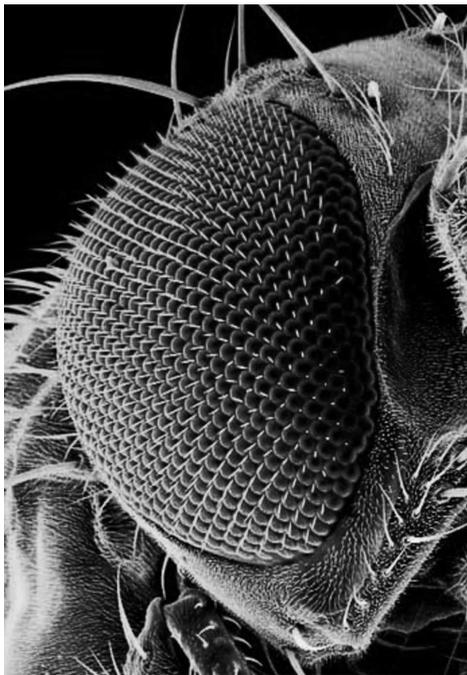


Fig. 5 Œil de drosophile

cellules soient capables de capturer les rayonnements ultraviolets et 70 % le visible. Imaginons un programme génétique pour faire ceci : il faudrait qu'il indique à chacune de ses cellules ce qu'il faut qu'elles captent, il faudrait aussi qu'il sache qu'il indique une chose à l'une et pas à l'autre pour maintenir les bonnes proportions. Avec les équilibres bistables, nul besoin de programme précis pour imposer à chacune des cellules un devenir précis. Les cellules ont 70 % de chance d'aller dans une voie d'expression qui permet de capter dans le visible et 30 % dans l'invisible.

Au final, les probabilités de devenir pour chacune des cellules, se transforment en proportion dans la population de cellules. Le phénomène d'équilibre bistable se produit selon des proportions évolutivement sélectionnées. Nous avons ici une économie, il suffit de voir que les cellules font les choses un peu au hasard, selon des proportions évolutivement sélectionnées, et non selon un programme génétique omniscient. Cela veut dire que l'on peut utiliser de manière fonctionnelle une dimension complètement aléatoire. Lorsque l'on raisonne en termes d'individus à travers une population, ce fonctionnement est très utile.

D'autres cas de différenciation observés sur un mode stochastique

Je voudrais maintenant montrer, sans trop de détail, qu'il existe des cas de fonctionnement aléatoire de différenciations cellulaires (ce que l'on a vu est une sorte de différenciation cellulaire lorsque passent des cellules d'un état à un autre) dans tout le règne animal (en utilisant en première approximation ce terme de règne, un peu daté). Cela s'observe notamment chez un petit nématode très utilisé en laboratoire (*C. elegans*), qui a la particularité de posséder un nombre de cellules fixes, environ 950. En biologie du développement, elles sont typiquement utilisées dans une optique déterministe, c'est-à-dire qu'on est capable de prévoir le devenir de chacune de ces cellules par l'acquisition d'un programme d'instructions précises pour chacune d'entre elles à chaque stade de développement. Ce qui est intéressant dans cet exemple, paradigmatique du développement déterministe, c'est que certaines de ces cellules ont un fonctionnement aléatoire, une différenciation stochastique. C'est notamment le cas des cellules Z1 et Z4 qui peuvent se différencier dans deux lignées différentes, et quand l'une se différencie d'une manière, l'autre se différencie de l'autre. C'est un bon exemple car même au sein d'un organisme qui est très connu pour fonctionner, au moins en apparence, de manière déterministe, il semble y avoir des endroits où l'on voit un peu de probabilisme surgir.

Chez le xénope (*Xenopus laevis*), un batracien, on observe même dans la phase de développement embryonnaire (moment où il n'y a pas beaucoup de cellules) une différenciation du mésoderme plus ou moins aléatoire. Nous observons qu'à chaque étape de la différenciation, seulement une petite proportion des cellules se met en route, et que progressivement tout devient homogène. Autrement dit, au cours de la

différenciation du mésoderme toutes les cellules ne réagissent pas de la même manière, quand bien même elles sont à un endroit précis, soumises au même microenvironnement. Dans le même ordre d'idée, elles ne réagissent pas toutes au même moment. Il existe donc certaines réponses aléatoires dans leurs comportements.

La différenciation stochastique et les cellules-souches

Il est temps d'en venir à la modélisation de la différenciation cellulaire. Dès 1983, Jean-Jacques Kupiec avait émis l'idée générale que l'expression aléatoire des gènes n'était pas forcément un problème à canaliser. Il l'a présentée comme un paramètre biologique utilisable par les organismes. Selon Kupiec, l'apparition de l'« ordre » (des cellules qui se ressemblent toutes, par exemple d'un même tissu) peut être expliquée à partir d'un désordre préalable. Des cellules possédant initialement la capacité de passer dans différents états de manière imprévisible, donnant lieu à plusieurs types cellulaires, pourraient ainsi avoir été évolutivement sélectionnées puis stabilisées par le biais d'interactions, par exemple membranaires, entre molécules de surfaces propres à chacun de ces types. Les populations de cellules d'un seul type ne se stabiliseraient ainsi pas tant que le deuxième ne serait apparu. Chaque type apparaîtrait, avec une certaine probabilité, au niveau cellulaire ; et donc, au niveau de la population, il apparaîtrait donc dans une certaine proportion. Kupiec avait ainsi formulé l'idée que le comportement stochastique des cellules pouvait être sélectionné (ici par stabilisation), une sorte de « hasard puis sélection ». L'expression d'endodarwinisme est ainsi apparue (Heams, 2004), pour qualifier le darwinisme à l'intérieur d'un organisme. Original en cela qu'il est un darwinisme appliqué aux cellules qui possèdent exactement les mêmes gènes, alors que généralement les processus darwiniens sont des processus de hasard-sélection entre individus ou entités qui possèdent des patrimoines génétiques différents, sélectionnés grâce à cette différence. Selon le postulat de Kupiec, ce n'est pas tant la composition en gènes que l'utilisation qui en est faite par les cellules qui permet à certaines d'entre elles de se stabiliser. Il a récemment, par le biais de travaux de modélisation, un peu modifié ces propositions en faisant passer cela, non pas par des potentiels récepteurs membranaires, mais par des échanges de nutriments entre cellules de types complémentaires stochastiquement apparus. C'est-à-dire que les cellules n'émettraient pas forcément en surface des récepteurs mais plutôt sécrèteraient des molécules qui

permettraient à l'une et à l'autre de se nourrir réciproquement, d'être en dépendance les unes par rapport aux autres. Pour ses propositions théoriques pionnières qu'il a cherché à valider par l'expérimentation et la modélisation, Jean-Jacques Kupiec doit être considéré comme le père de la théorie contemporaine probabiliste du vivant.

Conclusion

Je voudrais conclure cette présentation par une réflexion sur la possible compatibilité entre phénomènes déterministes et probabilistes. Est-ce que le vivant a sélectionné certaines niches dans lesquelles des phénomènes aléatoires peuvent prendre leur place au sein de fonctionnements en général déterministes ? Non, je défends simplement l'idée qu'au contraire, tous ces phénomènes ont lieu dans un grand ensemble probabiliste, le déterminisme n'étant qu'un cas limite du probabilisme. Ce que je veux dire par là, c'est que des phénomènes dont la probabilité est proche de 1 semblent déterministes. Une probabilité de 99,9 % reste une probabilité et les aspects fonctionnellement déterministes de certains phénomènes biologiques sont peut-être des apparences évolutives d'un fonctionnement plus général fondamentalement probabiliste. Celui-ci s'observe différemment, selon les contraintes qui pèsent sur ces phénomènes biologiques. Nous pourrions représenter les choses de cette manière : imaginons qu'un phénomène se produise avec une probabilité donnée, par exemple d'un pour mille. Dans une population de cent cellules où il a une chance sur mille de se reproduire, si vous constatez que ce phénomène se reproduit dans toute population (même celle qui a moins de mille cellules), vous pouvez penser qu'il y a des phénomènes de canalisation qui permettent aux phénomènes de se produire. Il semble déterministe. Dans une population d'un million de cellules avec toujours la même probabilité d'un sur mille du phénomène, vous pouvez être certains que ce phénomène a lieu. Il suffit de laisser le hasard faire les choses : environ un pour mille de ces cellules (donc un millier) suit la probabilité et ce millier peut se multiplier à son tour et envahir la population de la génération précédente. Il n'est nul besoin d'un contrôle précis, coûteux en énergie et en stockage d'information.

Finalement, il y a un équilibre à trouver entre les tailles de populations que l'on considère et les probabilités, par cellule, d'occurrence d'un phénomène. Si la taille de la population est suffisamment grande au regard de la probabilité de l'événement

élémentaire, le fait qu'il survienne effectivement n'est pas un miracle... Il faut cependant nuancer ces affirmations. L'événement stochastique de la sous-population qui arrive à réaliser cet événement doit pouvoir se répandre. Il faut trouver les mécanismes qui rendent possible l'homogénéité de cette population à ce moment-là ; il faut aussi moduler par la durée constatée du passage entre deux états. Toutes ces petites modulations doivent permettre, si on s'en donne les moyens, de calculer une sorte d'index synthétique déterminant la pertinence d'invoquer soit la stochasticité soit des mécanismes de canalisations plus déterministes. Cela nous permettra dans les années à venir de définir un domaine de validité réfutable de la stochasticité dans l'expression des gènes et dans les phénomènes biologiques cellulaires.

Commentaires et questions de Franck Varenne, réponses de Thomas Heams

Franck Varenne

Je voudrais d'abord revenir sur deux séquences de ton exposé. Tu as d'abord rappelé qu'un certain nombre de postulats classiques sont battus en brèche par de nombreuses publications récentes. Tu en as montré l'évolution récente : l'idée de stochasticité de l'expression des gènes est vraiment en train de se diffuser. Ce qui m'a marqué, c'est l'idée que le postulat de l'homogénéité était lié à un outil. Par le passé, les techniques ne permettaient pas d'avoir accès aux composants des cellules individuelles mais à des populations. Comment vois-tu ce lien entre histoire des instruments et ce déplacement de paradigme qui semble être en cours actuellement ? Deuxièmement, ce qui me paraît aussi décisif, c'est ce postulat de disponibilité qui nous emmène au niveau de l'individu pour nous apprendre que la cellule n'est pas une piscine où les composants flottent mais qu'il existe une espèce de rareté et une sorte de gestion de la rareté, qui est liée également à l'apparition d'outils de visualisation. L'approche de la cellule individualisée a-t-elle pris le pas sur l'approche de la population de cellules ?

Thomas Heams

Sur le postulat de l'homogénéité, je suis passé un peu vite, mais il est vrai que j'ai attribué cela d'abord aux évolutions des techniques. Certains de mes collègues ont nuancé le rôle des outils : depuis les années 1990 au moins, il était techniquement possible de déceler une hétérogénéité, certes pas de manière aussi fine que ce que j'ai pu montrer. Il y avait notamment les techniques de cytométrie de flux, qui permet de classer les cellules par taille, notamment par la taille de leur noyau ou par leur densité. Depuis des années, des nuages de points sont construits avec cette technique, qui montrent que des cellules placées toutes dans un même environnement ne répondent pas toutes de la même manière mais, ont au contraire, des marges de manœuvre. Cela n'avait pas alerté particulièrement les biologistes au sens qu'ils étaient restés dans l'idée que les cellules, certes n'étaient pas des robots, ne faisaient pas toutes exactement la même chose, mais globalement si, et donc qu'il fallait s'intéresser au centre du nuage de points, là où se trouvait la vraie valeur. Finalement ce rendez-vous manqué nous amène à nous demander si ce n'était qu'un problème technique. Pour ma part, je ne le pense pas. La biologie moléculaire a explosé au sortir

de la seconde guerre mondiale, notamment dans les années 50 quand la structure moléculaire d'A.D.N. a été trouvée par James Watson et Francis Crick, dans un contexte d'essor de l'informatique, de la théorie de l'information, et de l'idée de programmes, de déterminations cybernétiques. Tous les premiers modèles, et y compris les modèles les plus simples fonctionnaient par transposition de concepts déterministes à la biologie.

Franck Varenne

Dans un contexte donc très conceptuel où l'on se concentre sur la moyenne, la notion de bruit venant par exemple de la théorie du signal.

Thomas Heams

Exactement, je pense qu'il y a eu une sorte de percolation, en tout cas d'infusion, de ce concept qui était très fort, la théorie de l'information et de l'informatique donnant des résultats très concrets et puissants. Il y a eu des limitations techniques mais il y a eu aussi des vrais présupposés dont on ne s'est pas aperçu. Notamment le présupposé qui consiste à dire que toutes les cellules que l'on étudie sont égales à la moyenne d'une récupération en masse de toutes les protéines d'un tissu. C'est un vrai présupposé, *a priori* on aurait pu dès le départ dire qu'il y avait une distribution gaussienne et une vraie variabilité entre cellules. On a préféré considérer qu'elles étaient homogènes.

Franck Varenne

En même temps, dans les systèmes biologiques, quand on commence à aborder le niveau microscopique, comme Schrödinger le note, on ne peut appliquer la loi des grands nombres. C'est peut-être un autre problème de l'approche statistique ?

Thomas Heams

« Qu'est ce que la vie ? », le livre de Erwin Schrödinger de 1944 est passionnant. Cet auteur est un des grands penseurs du probabilisme en physique, il s'approche de la biologie avec intelligence et avec une sorte d'anticipation puisqu'il prévoit, lorsqu'il parle de cristal aperiodique, ce que va être la découverte de la molécule d'A.D.N. huit ans après. Il s'approche de la biologie et fait des estimations formellement exactes, mais il ne calcule pas exactement les bonnes choses. Il se focalise sur le nombre d'atomes qu'il pourrait y avoir dans une molécule et se demande, si ces atomes-là

pourraient fonctionner selon la loi des grands nombres et donc permettre un fonctionnement probabiliste. Il n'imaginait pas, et il n'aurait pas pu le faire à l'époque, comment tous les gènes pouvaient interagir entre eux. Avec 30.000 gènes dans un organisme qui peuvent chacun interagir les uns avec les autres, les probabilités d'interaction dépassent le nombre de molécules dans l'univers. Schrödinger s'est donc approché de très près d'un mariage entre le probabilisme et la biologie, il en a entrouvert la porte et l'a claquée trop vite.

Franck Varenne

... parce qu'il n'a pas intégré la combinatoire ?

Thomas Heams

En tout cas, pas à la bonne échelle. C'est une occasion manquée assez incroyable. Il est resté sur le fonctionnement globalement déterministe, avec des arguments que l'on peut comprendre. Avant de découvrir comment fonctionne un millier de gènes, il n'est pas inutile de comprendre comment fonctionne l'un par rapport à l'autre un petit réseau de gènes avec une protéine qui imprime son expression ou pas. Des résultats très simples sont obtenus au départ, comme tous ceux des travaux de l'école de François Jacob et Jacques Monod, qui ont déblayé un terrain immense. Il ne s'agit pas de remettre en cause l'ampleur de ces travaux. Il s'agit juste de montrer qu'ils ont été effectués dans un cadre conceptuel déterministe, qui a pu ne pas poser de problèmes dans les premières années mais qui a fini par en poser. En effet, lorsque s'est posée la question des gènes qui causent le cancer (programme des années 70 doté de milliards de dollars), les réponses n'ont pas été trouvées, malgré la force de frappe de la biologie moléculaire nord-américaine. Tous les gènes qui sont identifiés depuis trente ans sont impliqués dans autre chose que le seul cancer. On n'a pas la spécificité que l'on attendait, cela c'est pour l'homogénéité. Concernant la disponibilité, nous constatons que les cellules sont hétérogènes et qu'apparaissent ces problèmes de disponibilité pour certaines molécules. Alors que certaines sont en effet présentes à une ou deux copies en moyenne par cellule, avec une distribution gaussienne, certaines cellules n'en ont pas et d'autres en ont trois ou quatre. Ceci peut déclencher des effets de seuil et des fonctionnements cellulaires différents. Je voudrais ajouter que des recherches récentes montrent des mouvements moléculaires très rapides à l'intérieur des cellules, quand

bien même elles ne sont pas des piscines d'eau libre dans lesquelles les molécules peuvent librement diffuser. Même si cet encombrement topologique est constaté, les molécules se déplacent beaucoup plus vite que l'on imagine. Ce point contrebalance une partie de mon propos.

Franck Varenne

La prise en compte des dissymétries de l'espace et de la variété des topologies permet de constater que le niveau d'expression ne peut pas être uniquement connu par la seule séquence. C'est un point très important et il faut bien souligner ce que tu nous as dit : il ne suffit pas de connaître la séquence du génome pour savoir ce qu'il en est.

Thomas Heams

Ça n'est pas inintéressant de connaître la séquence. Je suis impliqué dans des travaux de séquençage moi-même et les travaux sur des animaux séquencés sont précieux. Cependant, cela est tout aussi vrai dans tous les travaux sur les territoires chromosomiques, la topologie nucléaire, et je pense notamment aux études de l'équipe des Cremer en Allemagne, pionnière en la matière. La connaissance de la topologie influence beaucoup notre compréhension du génome et la topologie est typiquement hors de la séquence. Cela rejoint une thématique générale que l'on appelle l'épigénétique, à la mode en ce moment, pour de bonnes ou de mauvaises raisons. D'ailleurs le terme épigénétique lui-même est à étudier. Il fait penser aux vieux temps de Ptolémée. Lorsque son système d'explication des mécanismes célestes ne marchait pas, il rajoutait des épicycles, des explications *ad hoc* de tous les dysfonctionnements. Le terme épigénétique donne l'impression que l'on rajoute un peu tout ce que la génétique ne permet pas d'expliquer, alors qu'elle était censée le faire il y a quelques années. On rajoute les territoires chromosomiques, on rajoute la méthylation des gènes, les modifications sur l'A.D.N. qui ne sont pas visibles dans la séquence mais qui peuvent jouer sur la fonction de chaque gène (les mêmes gènes ne vont pas forcément fonctionner exactement de la même manière). En tout cas, la génomique n'est pas, à elle seule, un outil suffisant pour comprendre le vivant.

Franck Varenne

Je voulais revenir justement sur ce point. L'ambition théorique est manifeste lorsque que tu dis qu'on peut expliquer que des formes de déterminisme sont dues à des

comportements stochastiques fortement canalisés d'une part ; et d'autre part, lorsque tu dis que ce type de modélisation, en se référant au travail de Jean-Jacques Kupiec, permet une économie. Cela parle à un épistémologue, puisqu'il y a une tradition de recherche théorique sous conditions d'économie des postulats explicatifs. Troisième point, cela pose la question de l'explication causale en biologie. Enfin, tu as parlé à un moment de comportement exploratoire des cellules comme s'il y avait une forme d'intelligence : on pourrait penser aux programmes de l'intelligence artificielle avec des essais/erreurs ou encore aux algorithmes génétiques, cette question renvoie à la notion de finalisme.

Thomas Heams

Je dois être plus clair en effet sur cette dernière question. Lorsque j'ai parlé de comportement exploratoire des gènes, je ne prêtai à chacune des cellules aucune intelligence particulière, la question de l'intelligence me dépasse assez largement. Concrètement, je parle d'une population cellulaire dans laquelle les cellules font un peu n'importe quoi dans des conditions de *stress* par exemple, et que certaines d'entre elles trouvent les bonnes combinaisons, par hasard. Le comportement est aléatoire dans une période donnée. Si, au bout d'un moment, parce que la probabilité de trouver la solution est finalement trop faible, aucune de ces cellules n'accède à la bonne combinaison de gènes exprimables, on peut considérer que la population meurt. Je ne leur prête donc aucune intention particulière. Le principe d'économie est quelque chose auquel je suis attaché. Il faudrait d'ailleurs le démontrer. Ce n'est pas aussi simple mais je raisonne en terme évolutif, je pars du principe que fabriquer de l'A.D.N., contrôler la fiabilité de cette séquence d'A.D.N., a un coût énergétique. Je considère comme économe tout ce qui peut permettre à des cellules et des organismes de fonctionner de manière acceptable dans un environnement donné, avec une certaine souplesse sans forcément avoir besoin de tout coder dans l'A.D.N. J'ai aussi conscience que cela se discute.

Franck Varenne

Je pensais à l'épistémologie des théories économes dans laquelle on prône un petit nombre de principes pour des raisons de commodité épistémique. Or, ton approche se prévaut d'un ancrage bien réel car on y voit un vivant qui est en lui-même économe.

Pour beaucoup de physiciens et même d'épistémologues, la science doit chercher à être économe en principes : elle doit en limiter le plus possible le nombre. Expliquer, en un sens, c'est toujours condenser, ramener le divers à l'unité, ou à l'identité ou, au moins, à un petit nombre d'énoncés. Pour quelqu'un comme Ernst Mach, par exemple, la science doit être considérée avant tout comme une stratégie d'« économie de pensée », *La Mécanique* (1904). Pour lui, c'est parce que les lois scientifiques ne sont que des résumés commodes des faits. Ton approche rejoint ce souci d'économie mais en semblant lui donner une justification théorique, fondée dans l'objet même que tu étudies, et non seulement dans une stratégie générale de connaissance (une stratégie qui ne serait qu'épistémique en ce sens). Une telle conception stochastique permet en effet que le vivant ait des comportements complexes et adaptatifs sans que tout soit programmé ou sans que les lois aient à rendre compte de tous les détails.

Thomas Heams

L'idée de *programme* (en grec « écrit en avance ») est une projection anthropomorphique, or rien n'est écrit à l'avance. L'A.D.N. est certes fascinant mais il n'est pas central. Il n'y a pas de centre dans une cellule, il y a des protéines et de l'A.D.N. qui sert à fabriquer ces protéines, mais sans protéines pas d'A.D.N. Tout cela est complètement distribué, décentralisé. Donc je pense que faire appel à la notion de programme c'est faire une hypothèse en terme épistémologique qui demande d'être vérifiée. Il ne faut pas inverser la charge de la preuve. Je pense qu'affirmer l'existence d'un programme est très lourd en termes scientifiques. Partir de l'idée qu'il pourrait ne pas y en avoir, et qu'il y a des phénomènes de canalisations qui donnent l'apparence qu'il en existe un de temps en temps, me paraît intellectuellement plus satisfaisant. Je voudrais juste compléter en disant que j'ai conscience, et c'est un problème qui est posé à la biologie en général, qu'à la suite d'un exposé comme celui-là on pourrait me dire que j'ai choisi les dix exemples dans la littérature qui illustrent l'aléatoire et que je raisonne donc par choix d'hypothèses et par exemples bien choisis. Ceci est vrai dans toutes les démonstrations en biologie. On fonctionne par exemple, nous n'avons pas la puissance théorique de la physique, et c'est un défi pour la biologie de faire des démonstrations. La proposition qui me paraît être honnête à ce stade c'est de comprendre, non pas comment on confronte le déterminisme et le probabilisme, mais comment on peut les faire dialoguer et les rassembler dans un même ensemble. Je ne me contente pas de confronter des exemples contre d'autres exemples.

Franck Varenne

Le déterminisme serait donc une variante, voire une dégénérescence du probabilisme. Cela m'amène à une autre question. Tu as eu des accents un peu poppériens dans ton propos... On pourrait te renvoyer cela en te disant : est-ce que l'on ne va pas se retrouver avec une théorie irréfutable si on dit que le déterminisme c'est seulement quand la probabilité tend vers un ? Deuxième question, n'est-ce pas là retrouver aussi une thématique des changements de niveau ? Ainsi, ce ne serait pas vraiment au niveau de la molécule qu'il faudrait chercher l'explication causale mais peut-être à un autre niveau, qui pour moi reste assez flou, celui de la population ?

Thomas Heams

C'est ce que j'essaie de montrer à la fin de ma présentation. C'est évident que lorsqu'un phénomène à une probabilité très rare de subvenir et qu'il intervient à toutes les générations, dans tous les embryons « au stade 4 cellules », tu ne peux pas invoquer que la probabilité. Il y a des mécanismes qui canalisent les choses. Quand l'ovule est fécondé, des A.R.N. interviennent et permettent à la cellule de se diviser, de faire des protéines pendant quelques divisions avant d'utiliser elle-même son propre génome. Là, on peut dire qu'il y a une sorte de déterminisme qui se met en place, je n'ai pas de problème avec cela, je ne cherche pas à falsifier le déterminisme. Il me semble en revanche raisonnable de penser que le fonctionnement est globalement probabiliste et qu'il peut exister des phénomènes de canalisations.

Franck Varenne

Il est vrai qu'il existe des techniques mathématiques pour rejeter un modèle probabiliste. J'en viens pour finir à une question plus générale, une interrogation qui s'inscrit davantage dans une perspective d'histoire des sciences. Il y a peut-être une tradition, en biologie, de retour récurrent au probabilisme. La biométrie, par exemple, a lutté contre l'approche physiologique dès le début du XX^e siècle. Née dans le cadre d'une interrogation systématique sur la diffusion des caractères génétiques, la biométrie a forgé des instruments mathématiques d'observation et de discrimination nouveaux, montrant qu'il y a des comportements stochastiques à presque toutes les échelles du vivant, y compris au niveau métabolique ou biomécanique, par exemple. Vivez-vous en ce moment, dans la biologie moléculaire, quelque chose de cet ordre-là ?

Vois-tu des similitudes avec l'approche des biométriciens des années 1920-30 qui mettaient les biologistes en garde contre les excès des approches causales naïves et prétendant valoir au seul niveau des comportements moyens ? Entre l'approche physiologique traditionnelle illustrée d'abord par la notion de milieu intérieur, liée ensuite étroitement à celle de comportement moyen, et l'approche biométrique (la biologie étudie quelque chose de mouvant livré aux aléas et aux fluctuations (conformément à la théorie mathématique des probabilités avec Francis Galton puis Ronald Aymler Fisher) une tension perdure. Elle est encore perceptible dans un petit ouvrage-manifeste comme celui d'Eugène Schreider (1967), par exemple. Selon Schreider, il n'y a pas de liaison fonctionnelle en biologie, l'organisme n'est qu'un processus à grand nombre de variables aléatoires, un « processus stochastique multivarié ». Fisher proposa d'ailleurs, dès 1934, une vision indéterministe et probabiliste pour l'ensemble des systèmes causaux, en particulier pour les systèmes biologiques.

Thomas Heams

Mon propos n'est pas de discréditer les approches antérieures. Je les inclus dans un ensemble historique et je sais ce que je leur dois. Il n'y a pas d'ambiguïté là-dessus, la biologie ne serait rien sans les travaux déterministes de la biologie moléculaire de l'après seconde guerre mondiale. Néanmoins, j'ai l'impression que les choses évoluent de façon radicale. Le mouvement probabiliste en génétique a aussi l'ambition de faire réagir face aux excès du déterminisme.

Références

- Chelly, J., J.-P. Concordet, J.-C. Kaplan et A. Kahn (1989), « Illegitimate Transcription: Transcription of Any Gene in Any Cell Type », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, n°8, pp. 2617-2621.
- Cremer, M., J. von Hase, T. Volm, *et al.* (2001), « Non-Random Radial Higher-Order Chromatin Arrangements in Nuclei of Diploid Human Cells », *Chromosome Research*, vol. 9, n°7, p. 541-567.
- Davidson, E.H., J.P. Rast, P. Oliveri, *et al.*, (2002), « A Genomic Regulatory Network for Development », *Science*, vol. 295, n°5560, pp. 1669-78.
- Elowitz, M., A. Levine, E.D. Siggia, P.S. Swain (2002), « Stochastic Gene Expression in a Single Cell », *Science*, vol. 297, n°5584, pp. 1183-1186.
- Felix, M. et P. Sternberg, « Symmetry Breakage in the Development of One-Armed Gonads in Nematodes », *Development*, vol. 122, n°7, p. 2129-2142.
- Fisher R.A., (1934), « Two New Properties of Mathematical Likelihood », *Proceedings of Royal Society, A*, 144: 285-307.
- Goodsell, D. (1993), *The Machinery of Life*, New York: Springer-Verlag.
- Heams, T. (2004), *Approche endodarwinienne de la variabilité de l'expression génétique*, thèse de doctorat, INA P-G.
- Hooshangi, S., S. Thiberge et R. Weiss (2005), « Ultrasensitivity and Noise Propagation in a Synthetic Transcriptional Cascade », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, n° 10, pp. 3581-6.
- Kupiec, J.-J. (1983), « A Probabilistic Theory for Cell Differentiation, Embryonic Mortality and DNA C-Value Paradox », *Speculations in Science and Technology*, vol. 6, n°5, pp. 471-478.
- Kupiec J.-J. (2008), *L'origine des individus*, Paris, Arthème Fayard.
- Mach, E. (1904), *La Mécanique : Exposé historique et critique de son développement*, Paris, A. Hermann.
- Martinez Arias, A. et P. Hayward (2006), « Filtering Transcriptional Noise During Development: Concepts and Mechanisms », *Nature Reviews Genetics*, vol. 7, n°1, pp. 34-44.
- Ozbudak, E., M. Thattai, I. Kurtser *et al.* (2002), « Regulation of Noise in the Expression of a Single Gene », *Nature Genetics*, vol. 31, n°1, pp. 69-73.
- Raser, J.M. et E. O'Shea, (2004), « Control of Stochasticity in Eukaryotic Gene Expression », *Science*, vol. 304, n° 5678, pp. 1811-4.
- Raser, J.M et E. O'Shea (2005), « Noise in Gene Expression: Origins, Consequences, and Control », *Science*, vol. 309, n°5743, p. 2010-2013.
- Ross, I., C. Browne, D. Hume, (1994), « Transcription of Individual Genes in Eukaryotic Cells

Occurs Randomly and Infrequently », *Immunology & Cell Biology*, vol. 72, n°2, pp. 177-185.

Schreider, E. (1967), *La biométrie*, Paris, PUF.

Schrödinger, E. (1993) [1944], *Qu'est-ce que la vie ? [What is Life?]*, Paris, Le Seuil.

Swain, P., M. Elowitz, E. Siggia (2002), « Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, n°20, pp. 12795-12800.

Thattai, M. et A. van Oudenaarden (2001), « Intrinsic Noise in Gene Regulatory Networks », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, n°15, pp. 8614-8619.