

Ces travaux expérimentaux se déroulent alors que Jean-Jacques Kupiec, à l'hôpital Saint-Louis, puis à l'institut Cochin, développe une théorie probabiliste de la différenciation cellulaire depuis le début des années 1980. Mais ce sont des techniques utilisant la protéine fluorescente verte, ou GFP, qui vont faire la part des choses, en prouvant que des gènes s'expriment de façon stochastique.

La nature stochastique de l'expression des gènes est confirmée, des bactéries jusqu'aux humains

En 2002, l'équipe de Peter Swain et Michael Elowitz, de l'université Rockefeller, à New York, intègre dans le génome d'*Escherichia coli* deux gènes codant deux variants de la GFP : l'un, une protéine à fluorescence bleu cyan, l'autre, une protéine fluorescente jaune. La construction génétique est identique dans les deux cas. D'après la théorie déterministe, chaque bactérie doit exprimer ces deux gènes dans des proportions quasi semblables ; la fluorescence résultante devrait donc être similaire dans toutes les bactéries, qui auront peu ou prou la même « couleur » : un mélange à parts égales de bleu et de jaune, soit du vert. De petites différences ne représenteraient qu'un « bruit » extrinsèque : les bactéries pourraient être plus ou moins vertes.

Le résultat déjoua ces prédictions : la couleur de fluorescence variait d'une bactérie à l'autre, allant du jaune à l'orange et au vert, avec plusieurs couleurs intermédiaires. Cela signifiait que l'expression des gènes de fluorescence dans une bactérie donnée était aléatoire, indépendamment de la concentration des acteurs de la transcription (voir l'encadré page ci-contre).

La même année, l'équipe de Robert Singer, du collège de médecine Albert-Einstein, à New York, a étendu ce type d'analyse à plusieurs

gènes dans des cellules humaines. Des sondes d'ADN fluorescentes se fixaient sur les molécules d'ARN codées par 11 gènes. À chaque gène correspondait ainsi un « code à barres » coloré spécifique, composé par les sondes fluorescentes. Or, dans une population de cellules identiques, chaque cellule présentait une combinaison de codes à barres spécifique qui changeait au cours du temps ; et toutes les cellules avaient des comportements différents.

Depuis, de nombreuses équipes ont confirmé la nature stochastique de l'expression des gènes, des bactéries jusqu'aux cellules humaines. Mais pourquoi des gènes s'expriment-ils ainsi ? On peut identifier deux types de raisons. D'abord, des causes moléculaires, dues aux interactions des molécules sur les sites de la transcription (enzymes, facteurs de transcription, autres protéines, ADN). Ensuite, des causes topologiques, mettant en jeu la facilité avec laquelle les facteurs de transcription accèdent aux gènes selon la forme que prend la molécule d'ADN.

LA DYNAMIQUE DE L'EXPRESSION

Pour comprendre, rappelons qu'un gène est une séquence d'ADN, une suite de nucléotides (les lettres A, C, T et G), qui contient un message codé. Le code génétique établit une correspondance entre des séries de trois nucléotides et les acides aminés qui composent les protéines. L'expression d'un gène comporte plusieurs étapes qui aboutissent à la synthèse d'une protéine (voir *les Repères*, page 6). Elle commence par l'activation du promoteur, une petite séquence située au début du gène. Puis, une enzyme (l'ARN polymérase II) se fixe sur le promoteur et se déplace le long de l'un des brins de l'ADN : ce faisant, elle transcrit la suite de nucléotides composant le gène en un ARN dit messager ; celui-ci sort du noyau et rejoint des ribosomes qui traduisent le message en protéines.

Ce mécanisme d'expression est régulé par des mécanismes dits épigénétiques, qui ne dépendent pas directement de la séquence d'ADN initiale : de petites molécules se fixent sur l'ADN ou sur les histones, ces protéines autour desquelles l'ADN s'enroule dans les cellules eucaryotes, dotées d'un noyau (voir *L'hérédité sans gène*, par M. Skinner, page 44). Par ailleurs, de petites molécules d'ARN, les micro-ARN, bloquent la traduction de certains ARN messagers. Or, à chacune de ces étapes, les structures moléculaires en jeu interagissent de façon dynamique, c'est-à-dire que leurs interactions évoluent constamment. C'est ce qu'a montré l'utilisation de sondes fluorescentes.

Par exemple, le groupe de Gordon Hager et Jim McNally, de l'Institut américain du cancer, a ainsi mis en évidence un ballet aléatoire de certains facteurs de transcription autour du promoteur de leurs gènes cibles : chaque molécule ne

reste liée au promoteur que durant quelques secondes ou dizaines de secondes avant d'être remplacée par une autre molécule. De même, on sait désormais que l'association des histones avec l'ADN est beaucoup plus dynamique que ce que l'on imaginait: le temps moyen de cette association varie de quelques secondes à quelques minutes, ce qui modifie sans cesse les zones de l'ADN qui peuvent ou non être transcrites.

On sait aussi, notamment grâce à l'équipe de Peter Fraser, à l'institut Babraham de Cambridge, que l'expression des gènes n'est pas homogène dans le noyau cellulaire; elle se déroule dans quelque 2500 régions particulières, nommées «usines à transcription». Ces régions sont riches en ARN polymérase. Plusieurs gènes partagent la même usine à transcription, sorte de tête de lecture sur laquelle chacun vient tour à tour «s'activer».

Faute de connaître les mécanismes tridimensionnels plausibles qui contrôleraient précisément ces mouvements, on doit privilégier l'hypothèse la plus simple: celle de changements aléatoires de la conformation de l'ADN qui exposeraient les gènes aux usines à transcription. Ce mode dynamique coïncide avec le fait que les ARN et les protéines correspondant à un gène sont produits par bouffées, parfois sous l'impulsion d'une seule molécule.

UNE EXPRESSION EN FONCTION D'UNE CLOCHE

Par ailleurs, les molécules qui régulent l'expression des gènes – facteurs de transcription et répresseurs – sont souvent en petite quantité dans la cellule. Ainsi, 80% des protéines d'une bactérie sont présentes en moins de 100 exemplaires; certaines, impliquées dans la division cellulaire ou dans la réparation de l'ADN, n'ont que quelques dizaines de copies.

Prenons le cas d'une protéine que chaque cellule d'une population contient en 50 exemplaires. En théorie, lors de la division cellulaire, les cellules filles reçoivent, en moyenne, 25 copies chacune. Mais en réalité, cette répartition suit la loi de Gauss; elle est en forme de cloche centrée sur une moyenne. On peut donc s'attendre à ce qu'autour de cette moyenne, surviennent des situations très diverses. Dans environ 6% des cas, la répartition est même très inégale, certaines cellules filles recevant moins de 19 copies, tandis que d'autres en récupèrent plus de 30 (ces estimations sont issues d'une répartition gaussienne standard). De sorte que certaines cellules ont 50% de protéines de plus que les autres. Au final, la transcription d'un gène fonctionne sur un mode instable: son expression n'a lieu que si les molécules de régulation sont suffisamment nombreuses à proximité de son promoteur. Elle peut se produire dans une cellule (et pas dans ses

voisines), puis diminuer avec la dégradation des facteurs de transcription.

Ainsi, des cellules ayant les mêmes gènes produisent des quantités différentes de facteurs de transcription et expriment des gènes distincts. Pour un gène donné, le promoteur de la transcription passe d'un état actif à un état inactif et le rythme de transition d'un état à l'autre détermine à quelle vitesse fluctue l'expression du gène: plus la période séparant deux activations successives du promoteur est courte, plus la variation de l'expression du gène est rapide. Or on a montré que la structure du promoteur influe sur ce rythme. Ainsi, en 2006, l'équipe de James Collins, à l'université de Boston, a découvert chez la levure que les promoteurs qui contiennent un motif de nucléotides, la boîte TATA (riche en thymine et en adénine), réagissent plus vite que les promoteurs qui en sont dépourvus.

Ce sont donc les variations naturelles des interactions moléculaires sur les sites de la transcription qui rendent l'expression génique aléatoire. Cependant, la structure même de l'ADN dans la cellule y contribue également. Cette longue molécule occupe un volume dans

L'expression aléatoire des gènes se traduit par la synthèse de protéines différentes ou dans des proportions différentes dans les cellules génétiquement identiques d'un même tissu. Par exemple, une protéine (ici en rouge) peut être synthétisée en petite quantité tandis que d'autres (en bleu et en jaune) sont produites en plus fortes concentrations. En revanche, la même protéine sera plus abondante dans une cellule voisine, ou absente dans une troisième, tandis que les deux autres protéines seront moins présentes.

