

Quelle taille pour l'univers des exons ?

Le nombre de protéines qui pourraient théoriquement être fabriquées à partir des 20 acides aminés formant leur ossature défie l'imagination. Si l'on considère qu'un exon moyen code pour un peptide de 40 acides aminés, le nombre de combinaisons est de 10^{52} , ce qui rendrait illusoire toute tentative de systématisation. Ce chiffre pourrait toutefois être beaucoup plus faible si l'on admet qu'au départ il n'existait qu'un nombre limité d'exons, et qu'une redistribution par brassage d'exons (*exon shuffling*) se trouvait à l'origine de la diversité d'aujourd'hui.

L'idée que les protéines actuelles dérivent d'un fond commun de modules, ou motifs, est dans l'air depuis plusieurs années, mais résultait jusqu'à présent d'une intuition, certes fondée sur des exemples mais ne reposant sur aucune donnée quantitative. C'est donc le mérite d'une équipe de Harvard dirigée par Walter Gilbert d'avoir tenté d'évaluer « l'univers des exons » dans un article de *Science* [1], commenté dans la même revue par Ann Gibbons [2]. Il a fallu aux auteurs adopter des méthodes d'abord théoriques et de calcul qui comportent nécessairement une part d'arbitraire. Ils ont commencé par collecter environ 2 500 séquences connues d'exons conservés dans les banques de données de Genbank et de l'EMBL, en acceptant des exons codant pour au moins 20 acides aminés, avec un pic à 40-50 acides aminés. Ils ont éliminé celles qui proviennent de duplications ou qui sont directement dérivées les unes des

autres, pour éviter les répétitions. Ils ont ensuite repéré, en les examinant par paires, les similarités entre exons par rapport à ce que donnerait une répartition statistique, en utilisant la méthode de calcul appelée de Monte-Carlo. Ils sont ainsi parvenus à conclure à l'existence d'un univers de 1 000 à 7 000 exons originaux, dépourvus de répétitions. Ce résultat peut paraître imprécis ; en réalité, ce qui importe, c'est l'ordre de grandeur et la conclusion que le nombre d'exons initiaux utilisés comme matériaux de base pour la construction des protéines est limité.

Bien entendu ce travail prête à de nombreuses critiques : il se fonde sur les séquences connues, alors que beaucoup d'entre elles restent à découvrir ; il ne tient compte que des gènes eucaryotes, et l'introduction des procaryotes élargirait le nombre des séquences. Le résultat est donc probablement minoré, mais l'ordre de grandeur ne changerait pas. D'autres critiques ont été émises ; elles sont mineures pour la plupart ; la plus grave est que les méthodes de calcul, certes parfaitement standardisées, seraient ici appliquées à tort. Quoi qu'il en soit, le travail de Dorit *et al.* a l'immense qualité de constituer une première tentative pour essayer de comprendre l'origine de la diversité actuelle. Cette origine, à partir de modules préfabriqués, aurait permis d'accélérer l'évolution. Le prix à payer aurait été le sacrifice d'une grande partie de la diversité théoriquement possible. En fait, les études de modèles de structure

des protéines suggèrent que le nombre des formes tridimensionnelles est faible. Unger *et al.* (cité dans [1]) ont examiné les structures de chaque séquence de six acides aminés en trois dimensions, et conclu que l'on trouve seulement 80 types possibles au lieu des 10^8 théoriques. On peut concevoir que des exons homologues adoptent des configurations semblables au sein de protéines différentes. C'est ainsi que, pour les auteurs, une protéine de 200 acides aminés peut résulter de l'arrangement de cinq modules de 40 acides aminés chacun, ce qui, avec 5 000 exons de base, donne 25 000 combinaisons, au lieu des 20^{2000} solutions possibles à partir d'une « recherche au hasard » acide aminé par acide aminé. Il est donc probable que le processus que nous envisageons — assortiment réduit d'exons au départ, puis duplications et divergences — limite la diversité des structures des protéines et n'a pas laissé le temps à la nature de chercher, dans chaque cas, la séquence la plus efficace dont elle aurait pu rêver.

J.-C. D.

1. Dorit RL, Schoenbach L, Gilbert W. How big is the universe of exons? *Science* 1990 ; 250 : 1377-82.
2. Gibbons A. Calculating the original family of exons. *Science* 1990 ; 250 : 1342.