

Les gènes chevauchants dans les virus et l'origine *de novo* des protéines

Cet exposé sera focalisé sur un aspect du génome qui peut paraître très spécialisé dans un premier temps, mais qui en fait, pourrait être, pour des raisons différentes, d'intérêt biologique générale : les gènes chevauchants, qui constituent une énigme évolutive fascinante et en même temps éclairent la question pas encore résolue de l'origine des nouveautés, c'est-à-dire des nouvelles protéines. Les gènes chevauchants sont des régions génomiques de l'ADN ou l'ARN (dans le cas de beaucoup de virus) qui sont traduites en deux (parfois même trois) cadres de lecture en donnant deux (trois) protéines dont les séquences des acides aminés, donc en général, les fonctions, n'ont aucun rapport. On a raison de penser qu'ils sont nés par « overprinting » (*surcharge*), lorsque des mutations ponctuelles, dans un cadre de lecture qui préexiste, permettent l'expression d'une protéine complètement nouvelle à partir d'un autre cadre de lecture. Chez les virus, dans lesquels les gènes chevauchants sont nombreux, ces nouvelles protéines jouent souvent un rôle critique dans l'infection, bien que fréquemment elles soient sous estimées. Pourtant, leur détection correcte est très importante pas seulement lors de l'étude des virus, puisque en fait, ils se révèlent de plus en plus nombreux même dans les organismes « supérieurs », homme compris.

Comme précédemment expliqué, les gènes chevauchants laissent perplexes, parce que chacune des deux protéines chevauchantes contraint la liberté de changer de l'autre. Pour conduire des études évolutives systématiques et pour développer des méthodes de détection meilleures, on a rassemblé un « dataset » de haute qualité de 80 gènes chevauchants viraux non homologues, dont l'expression a été prouvée expérimentalement. A partir de ça, on a donc avant tout déterminé les caractéristiques compositionnelles, qu'on a montré être partagées aussi par les gènes chevauchants caractérisés dans les mammifères, qui sont significativement différentes de celles des régions génomiques non chevauchantes. Nous croyons que ces caractéristiques sont importantes pour comprendre l'évolution des gènes chevauchants et leur implication dans l'origine *de novo* des protéines pas seulement dans les virus. En fait, il est possible que les nouvelles données sur les gènes chevauchants dans les eucaryotes dont on a fait mention, aient finalement des conséquences profondes sur la vision courante des gènes. Il me semble donc que, du côté historique, il est intéressant de revenir à l'époque de la découverte des gènes chevauchants dans les virus, la deuxième moitié des années '70, en se posant la question de savoir si ce fut perçu comme ébranlant l'idée du gène qui était accepté à l'époque. Pourquoi donc autour de 1980 n'en parlait-on presque plus jusqu'aux premières années du XXI siècle? Si le temps m'était donné, j'esquisserais quelques hypothèses de travail pour répondre à ces questions, qui sont mises à l'épreuve dans mon activité de recherche actuelle.