

Le nouveau visage des ARN

Brice Felden

Les ARN sont des molécules connues surtout comme intermédiaires dans la fabrication des protéines. Pourtant, ils exercent nombre d'autres tâches. Certains se révèlent même multifonctionnels.

L'ESSENTIEL

- Les ARN, les molécules produites à partir de l'ADN, sont bien plus que des intermédiaires dans la fabrication des protéines.
- Nombre d'entre eux ont à la fois une fonction codante et une ou plusieurs fonctions régulatrices.
- Ces ARN mixtes sont présents chez tous les organismes vivants, et certains sont impliqués dans des maladies.
- Ils interrogent aussi sur l'origine du vivant : sont-ils les précurseurs de la vie ou un produit de l'évolution ?

© Pour la Science

En biologie moléculaire, l'un des dogmes fondateurs indique que l'information génétique, stockée chez la majorité des organismes vivants sous forme de molécules d'ADN, est transformée en protéines *via* un intermédiaire transitoire, les ARN. Cependant, depuis 50 ans, un nombre croissant de travaux ont montré que les ARN ont des rôles plus riches et diversifiés. Si tous les génomes des organismes vivants, unicellulaires ou pluricellulaires, présentent une proportion élevée de gènes qui sont transcrits en ARN, cet ARN n'est pas nécessairement traduit en protéines, et c'est en tant qu'ARN qu'il remplit une fonction au sein de la cellule.

À la suite de cette découverte, les ARN traduits en protéines ont été nommés ARN messagers, et ceux qui assurent d'autres fonctions ont reçu le nom d'ARN non codants ou ARN régulateurs. Ainsi les ARN messagers ont-ils été distingués des autres ARN qui ne présentaient pas de séquence codante.

Jusqu'à récemment, la séparation codant/non codant semblait claire : les ARN messagers, qui encodent des protéines, étaient distincts des ARN régulateurs, n'exprimant pas de protéines. Mais aujourd'hui,

la séparation se révèle beaucoup plus floue. Les premiers indices sont apparus lorsque l'on s'est rendu compte que le catalogue des ARN encodés par les génomes – le transcriptome – est bien plus étendu et complexe qu'on ne le pensait. De nombreux exemples d'ARN messagers présentant des régions régulatrices en amont et en aval de leurs séquences traduites en protéines ont été décrits. Inversement, ces dernières années, on a aussi mis en évidence, au sein d'ARN régulateurs de l'expression de gènes cibles, de petites séquences internes capables d'être traduites en peptides ou en protéines.

Cela suggère que de nombreux ARN, que nous dénommerons ici ARN mixtes, présenteraient à la fois des fonctions régulatrices et codantes. Des travaux récents incluant des études approfondies de mutagenèse (on mute une région de l'ARN pour étudier sa fonction) couplées à des tests biologiques ont permis de définir avec précision les différents domaines des ARN mixtes impliqués dans leurs diverses fonctions. Ces dernières sont activées ou réprimées selon les conditions extérieures : le stade de développement de l'organisme, les variations physico-chimiques

dans lesquelles il évolue, les perturbations causées par un micro-organisme pathogène... Après avoir décrit, à travers quelques exemples, les ARN mixtes, leurs fonctions et leur importance dans certaines pathologies, nous envisagerons les mécanismes évolutifs ayant permis l'éclosion de ces ARN qui, tels des couteaux suisses, permettent aux organismes les exprimant de faire face à une multitude de situations dans un environnement en perpétuel changement.

ARN messagers et régulateurs, une frontière floue

Le premier ARN non codant identifié a été l'ARN de transfert alanine, purifié à partir de levures, dont la structure a été publiée en 1965 par le biochimiste américain Robert Holley – travaux récompensés par un prix Nobel en 1968. Cet ARN est impliqué dans la synthèse des protéines : il transporte, jusqu'à la protéine en construction, un constituant des protéines (l'acide aminé alanine) et l'insère dans la chaîne protéique aux endroits codés par l'information génétique. Depuis cette découverte, grâce aux progrès des techniques de séquençage, des centaines d'ARN non codants ont été identifiés dans les génomes de procaryotes (cellules sans noyau) et des milliers chez les eucaryotes (avec noyau). Leurs fonctions sont peu à peu élucidées. La majorité d'entre eux semblent réguler l'expression de gènes cibles en intervenant à diverses étapes du processus conduisant des gènes à la synthèse des protéines. Leur taille varie de quelques nucléotides (les constituants de l'ADN et de l'ARN) à plusieurs dizaines de milliers.

Au cours de la dernière décennie, des progrès techniques spectaculaires ont permis de séquencer le patrimoine génétique de nombreux organismes vivants, des virus à l'homme. Désormais, du fait de la baisse des coûts de séquençage, le nombre de génomes séquencés d'individus distincts au sein d'une même espèce augmente à grande vitesse. Cela permet d'analyser et de comparer les génomes d'entités biologiques similaires provenant de zones géographiques différentes. Une fois les

séquences génomiques déterminées, des approches bio-informatiques automatisées différencient le matériel génétique contenant de l'information codante de celui non traduit en protéines, à l'aide de signatures spécifiques présentes au sein de la séquence même des génomes.

Chez les organismes procaryotes, l'information codante est majoritaire dans le génome. Chez les eucaryotes, dont nous faisons partie, c'est l'inverse : la majeure partie du génome n'encode pas de protéines, mais contient une information colossale sous forme d'une panoplie de gènes transcrits en ARN fonctionnels de différentes

tailles et fonctions. Seul un pour cent environ des génomes de mammifères contient de l'information traduite en protéines, alors que 70 à 90 pour cent sont transcrits au cours de la vie en un transcriptome vaste et complexe d'ARN non codants. Le catalogue de ces transcrits fait l'objet d'une recherche internationale active.

Il est encore incomplet, mais permet déjà de dresser un nouveau portrait des ARN.

Les transcriptomes des organismes vivants contiennent un mélange d'ARN de tailles et de fonctions variées, dont les ARN messagers ne sont qu'un sous-ensemble. Un ARN messager se définit par la séquence codante qu'il contient, construite par une succession de triplets de nucléotides – des codons –, dont

chacun spécifie un acide aminé, avec un triplet marquant le début et un autre la fin de la séquence codante. Ces triplets sont traduits par les ribosomes, les usines de transformation de l'information génétique en protéines, avec l'aide des ARN de transfert qui les approvisionnent en acides aminés (voir l'encadré pages 54 et 55).

Généralement, en amont et en aval de leurs séquences codantes, les ARN messagers contiennent des portions transcrites, mais non traduites, très structurées, dont

1. LES ARN SONT LES MOLÉCULES TRANSCRITES à partir de l'ADN. On découvre aujourd'hui que nombre d'entre eux remplissent plusieurs fonctions. Un même ARN présente plusieurs domaines plus ou moins structurés ayant chacun une fonction différente, symbolisée par une couleur sur cet ARN fictif.