

Quelques termes-clef de biologie moléculaire et leur définition

Isabelle Quinkal
INRIA Rhône-Alpes
Septembre 2003

Acide aminé (AA)

Petite molécule dont l'enchaînement compose les *protéines* - on dit qu'une protéine est un *polymère* d'acides aminés (les *monomères*). Il existe 20 acides aminés différents utilisés pour fabriquer les protéines.

Acide nucléique

Polymère formé par l'enchaînement de *nucléotides*. Les acides nucléiques jouent un rôle fondamental dans le stockage, le maintien et le transfert de l'information génétique. Il existe deux types d'acide nucléique : l'acide ribonucléique (ARN) et l'acide désoxyribonucléique (ADN).

Acide désoxyribonucléique (ADN)

Support biochimique de l'information génétique chez tous les êtres vivants (à l'exception de quelques virus qui utilisent l'*ARN*). Principal composant des chromosomes, l'ADN se présente le plus souvent sous forme de deux longs filaments (ou chaînes) torsadés l'un dans l'autre pour former une structure en double hélice. Chacune de ces chaînes est un polymère formé de l'assemblage de quatre nucléotides différents, désignés par l'initiale de la base azotée qui entre dans leur composition : A (Adénine), C (Cytosine), G (Guanine) et T (Thymine).

Acide ribonucléique (ARN)

Dans les cellules, on distingue plusieurs types d'ARN suivant leur fonction. Les trois types principaux sont : les *ARN messagers*, les *ARN de transfert* et les *ARN ribosomiaux*. L'ARN est un *acide nucléique* constitué d'une seule chaîne de nucléotides, de structure analogue à celle de l'ADN. Il existe cependant des différences chimiques entre ces deux acides nucléiques qui donnent à l'ARN certaines propriétés particulières. L'ARN est produit par *transcription* de l'ADN.

ARN messenger (ARNm)

Photocopie du *gène*, il sert à transférer l'information génétique de son lieu de stockage (le *chromosome*) jusqu'au lieu de synthèse des protéines (les *ribosomes*). Les ARNm des cellules *eucaryotes* doivent subir une *maturation*, comprenant souvent un processus d'excision de leurs *introns* et d'*épissage* de leurs *exons* avant leur *traduction* en protéines.

ARN ribosomal (ARNr)

Constituant principal des ribosomes, la machinerie cellulaire où a lieu la *traduction* en protéines de l'information contenue dans les *ARNm*.

ARN de transfert (ARNt)

Petits ARN responsables du transport des acides aminés jusqu'aux ribosomes lors de la *traduction* des *ARNm* : chaque ARNt transporte un acide aminé, de façon spécifique. Sa séquence comporte une série de trois nucléotides, nommée anticodon, qui reconnaît le *codon* (*cf code génétique*) correspondant à l'acide aminé qu'il transporte.

Code génétique

Système de correspondance permettant de traduire une séquence d'acide nucléique en protéine. Dans ce système, un triplet de nucléotides, ou codon, désigne un acide aminé. Comme il existe 4 nucléotides, il y a $4 \times 4 \times 4 = 64$ codons différents. À un codon donné correspond un seul et unique acide aminé. Par contre, il n'existe que 20 acides aminés différents dans les protéines, c'est pourquoi plusieurs codons peuvent désigner un même acide aminé – on dit que le code génétique est redondant.

Certains de ces 64 codons ne désignent aucun acide aminé. Ces triplets « non-sens » indiquent à la machinerie cellulaire la fin de la lecture de l'information contenue dans les gènes, et provoquent l'arrêt de fabrication des protéines. On les appelle *codons STOP*.

Tous les êtres vivants (à quelques variantes près) possèdent le même code génétique : il est universel.

Épissage

S'applique aux *gènes mosaïques* des *eucaryotes* : mécanisme consistant, sur l'*ARN messenger* qui vient d'être transcrit, à éliminer (exciser) les *introns* et réunir (épisser) les *exons* entre eux. Le produit de l'épissage est un ARN messenger *mature*, prêt à être *traduit* en protéine.

Eucaryotes / procaryotes

L'ensemble des organismes vivants peut être classé en trois grands groupes : les *eucaryotes*, les *eubactéries*, les *archaebactéries*. A l'intérieur de chacune de leurs cellules, les eucaryotes possèdent un noyau : petit sac entouré d'une membrane semi-perméable renfermant les chromosomes. L'Homme, ainsi que les animaux, les plantes et les champignons, sont des eucaryotes. Les eubactéries et les archaebactéries ne possèdent pas de vrai noyau, mais une structure beaucoup plus simple, non entourée d'une membrane. C'est de ce « proto-noyau » que vient le nom générique qui les désigne : procaryotes. **Exon / intron / gène mosaïque**

Chez les eucaryotes, les gènes sont le plus souvent constitués de deux types de séquence nucléotidique : l'une est dite codante et l'autre non codante. Les parties codantes, appelées exons, portent l'information qui sera directement utilisée pour fabriquer les protéines. Entre les exons se trouvent les introns, non « lus » lors de la *traduction*. Du fait de cette disposition alternée exon/intron, on emploie l'expression *gène mosaïque*.

Fonctions d'une protéine

Rôles remplis par une protéine. Ces fonctions sont très variées et permettent de classer les protéines : les « protéines de structure » sont comparables à des briques cellulaires (ex : le collagène) ; les « protéines de transport » sont chargées du transport d'autres molécules dans la cellule ou entre les cellules d'un organisme (ex : l'hémoglobine transporte l'oxygène) ; les enzymes permettent d'accélérer les réactions chimiques nécessaires à la vie (ex : la glucose 6 phosphatase initie la dégradation du glucose, notre principale source de l'énergie cellulaire) ; les protéines de l'immunité (ou anticorps) contribuent à la défense de notre organisme. La *structure tridimensionnelle* d'une protéine est l'un des éléments qui détermine sa fonction.

Gène

Fragment d'ADN portant les informations nécessaires à la fabrication d'une ou plusieurs protéine(s). Un gène comprend la séquence en nucléotides qui sera transcrite puis traduite en acides aminés, mais aussi des séquences permettant de réguler cette fabrication de protéine en fonction des conditions cellulaires. La longueur d'un gène peut varier de quelques centaines, à plus d'un million de nucléotides.

Génome

Ensemble de l'information génétique d'un organisme. Une copie du génome est présente dans chacune de ses cellules. Le génome est transmis de génération en génération. Par extension, le génome se réfère aussi au support physique de cette information génétique, c'est-à-dire la macromolécule d'ADN.

Génomique

Étude des *génomés*. Son objectif est de séquencer l'ADN d'un organisme et de localiser sur celui-ci tous les gènes qu'il porte, puis de caractériser leurs fonctions.

Génomique fonctionnelle (« post-génomique »)

Étude de la fonction des *gènes* par analyse de leur séquence et de leurs produits d'expression : les *ARNm (transcriptome)* et les *protéines (protéome)*. Elle s'intéresse à leur mode de régulation, et à leurs interactions. L'analyse des protéines peut aller jusqu'à la détermination de leur structure tridimensionnelle.

Hybridation

Propriété caractéristique et essentielle des molécules d'acides nucléiques, qui leur confère leur capacité de transfert de l'information. Deux chaînes (ou brins) ont tendance à s'apparier pour former des doubles brins (ADN/ADN, ADN/ARN ou ARN/ARN), selon un mécanisme rappelant celui d'une fermeture éclair. Il faut pour cela que les nucléotides qui les composent soient complémentaires, c'est-à-dire qu'en face d'un nucléotide (A) se trouve un (T) ou un (U), et en face d'un (C) se trouve un (G), et réciproquement.

Cette propriété intervient lors de la *réplication*, de la *transcription* et de la *traduction*. C'est par ailleurs sur l'hybridation que sont basés les principes de nombreuses techniques de biologie moléculaire (*séquençage, puces à ADN...*)

Macromolécule

Molécule géante dont la masse moléculaire est de plusieurs milliers de daltons (le dalton est l'unité utilisée pour décrire la masse d'une molécule. Il correspond à la masse d'un atome d'hydrogène soit $1,66 \cdot 10^{-24}$ g). Les *protéines* et les *acides nucléiques* (ADN et ARN) sont des macromolécules.

Maturation (des ARN messagers)

Ensemble des modifications des ARNm préalables à leur traduction. L'*épissage* est un des processus de maturation des ARNm.

Nucléotide

Motif structural de base (monomère) des acides nucléiques, formé de l'assemblage de plusieurs molécules : un sucre (ribose pour l'ARN, désoxyribose pour l'ADN), un acide phosphorique et une base azotée (dans le cas de l'ARN cette base peut être l'Adénine - A, la Cytosine - C, la Guanine - G ou l'Uracile - U ; *idem* dans le cas de l'ADN, excepté que l'Uracile est remplacé par la Thymine - T).

Polymère

Macromolécule répétant un même motif structural appelé monomère. Les protéines sont des polymères d'acides aminés, les acides nucléiques sont des polymères de nucléotides.

Promoteur

Courte séquence spécifique d'ADN, située au début des gènes, sur laquelle se fixe l'enzyme qui effectue la transcription (l'ARN polymérase). Etant nécessaire pour que la transcription débute, le promoteur est indispensable au fonctionnement d'un gène.

Protéine

L'un des quatre matériaux de base de tout organisme, avec les glucides, les lipides et les acides nucléiques. Les protéines sont formées d'un enchaînement spécifique d'acides aminés (de quelques dizaines à plusieurs centaines).

Protéome / protéomique

Le protéome est l'ensemble des protéines produites à partir du *génome* d'un organisme. Comme le transcriptome, le protéome n'est pas identique dans toutes les cellules d'un organisme donné.

La protéomique est l'étude du protéome, dans le but de déterminer l'activité, la fonction et les interactions des protéines, et cela dans diverses conditions.

Puce à ADN

Technologie employée dans l'étude du *transcriptome* et basée sur la capacité des molécules d'ADN et d'ARN à s'*hybrider* entre elles. De courtes séquences d'ADN connues sont fixées sur des supports d'une surface de l'ordre du centimètre carré : les puces. Elles sont mises en présence d'un mélange d'ARN de séquences inconnues. Le système est conçu de sorte à ne détecter que les paires d'ADN/ARN qui se sont hybridées. On en déduit les séquences des ARN présents dans le mélange étudié.

Réplication

Mécanisme de synthèse de l'ADN permettant de transmettre l'information génétique d'une cellule ou d'un organisme à sa descendance. Chaque molécule-fille d'ADN est constituée d'un brin de la molécule-mère, qui sert de modèle à un nouveau brin. Ceci conduit à la duplication des molécules d'ADN de tout le génome.

Réseau de régulation

Interactions complexes entre les gènes et leurs produits (ARN et protéines) régissant l'activité de la cellule afin de lui permettre de s'adapter en permanence aux variations de son environnement. Il existe de véritables cascades d'interactions, faisant souvent intervenir des boucles de rétroaction positive et négative.

Exemple : les gènes bactériens *lacZ lacY* et *lacA*, localisés côte à côte sur le génome, codent pour 3 enzymes impliquées dans le transport et la digestion du lactose. En absence de lactose, une protéine répresseur empêche la synthèse de ces 3 enzymes. En présence de lactose, une molécule dérivée du lactose se fixe sur le répresseur, activant la synthèse des 3 gènes. Le lactose peut alors être utilisé par la cellule.

Ribosomes

Machinerie cellulaire, constituée de protéines et d'ARN (les ARNr), responsable de la traduction des ARNm.

Séquence / séquençage

La séquence est l'ordre d'enchaînement des éléments qui composent les acides nucléiques ou les protéines : respectivement nucléotides ou acides aminés. La séquence d'une protéine est dictée par celle de son gène.

Le séquençage consiste, par des méthodes chimiques ou de biologie moléculaire, à déterminer l'ordre des nucléotides de l'ADN ou des acides aminés des protéines.

Signaux de régulation

Séquences nucléotidiques spécifiques sur lesquelles viennent se fixer des protéines qui provoquent ainsi l'activation ou la répression de l'expression des gènes.

Structure d'une protéine

Toute protéine existe sous une conformation définie par différents niveaux structuraux. La séquence des acides aminés liés dans la chaîne polypeptidique constitue la *structure primaire* de la protéine. Cette séquence est pleinement définie par l'ADN de la cellule.

La *structure secondaire* est un premier niveau de repliement, adopté par des portions de la protéine, résultant d'interactions entre des acides aminés voisins sur la chaîne. Deux motifs de repliement caractéristiques peuvent ainsi se former : des hélices (dites hélices α) ou des feuillettes (dits feuillettes β), réunis par des boucles ou des demi-tours.

La structure finale, c'est-à-dire l'organisation des éléments de la structure secondaire entre eux est appelée *structure tertiaire*. C'est la forme que prend la protéine dans l'espace. Elle est due à des interactions entre acides aminés éloignés sur la chaîne peptidique, qui se retrouvent proches dans cette structure tridimensionnelle.

Certaines protéines, complexes, sont constituées de plusieurs sous-unités (plusieurs chaînes protéiques). La structure quaternaire est l'arrangement spatial de ces différentes unités, nécessaire à l'activité de l'ensemble.

Traduction

Étape de la synthèse (fabrication) des protéines au cours de laquelle le brin d'ARN messenger obtenu lors de la *transcription* est converti en une chaîne d'acides aminés qui donnera une protéine. Les ARNm procaryotes sont traduits tels quels, les ARNm eucaryotes subissent une *maturation* préalable.

Transcription

Processus de fabrication d'un ARNm à partir de la séquence codante d'un gène qui sert de modèle. L'enzyme responsable de cette réaction est appelée ARN polymérase.

Transcriptome

Ensemble des ARN messagers transcrits à partir du *génome*. Comme le protéome, il varie au cours du temps et d'une cellule à l'autre d'un organisme.