Qui est le plus gros : l'ARN messager ou la protéine qu'il code ?

Publié le 06.10.26 | Par Pascal Combemorel

Dans les ouvrages de biologie et biochimie, les molécules d'ARN messager (ARNm) et les protéines qu'elles codent sont généralement représentées sur les schémas sans tenir compte de la taille relative de ces deux types de macromolécules. Combien mesurent et pèsent les ARN messagers, ainsi que les protéines qu'ils codent ? Et, entre les deux, qui est le plus gros : l'ARNm ou la protéine codée ?

Le lecteur est invité à consulter au préalable l'article intitulé Les acides aminés et la synthèse peptidique.

L'expression des gènes codant des protéines nécessite un intermédiaire, l'ARN messager (ARNm). Dans les ouvrages traitant de biologie, qu'il s'agisse de manuels scolaires ou universitaires, ainsi que sur internet, les molécules d'ARNm sont souvent représentées par un simple trait ou par les nucléotides qui les composent, tandis que les protéines sont généralement dessinées sous la forme d'un collier d'acides aminés ou sous une forme globulaire (Figure 1). Quoi qu'il en soit, sur de tels schémas, les protéines sont généralement aussi grosses ou plus grosses que les ARNm qui ont servi à leur production. Or, en réalité, les ARN messagers sont plus grands et plus lourds que les protéines qu'ils codent. Pour le montrer, nous allons voir quels sont les ordres de grandeur de masse et de taille des protéines, puis des ARN messagers, pour terminer en comparant ces deux types de macromolécules.

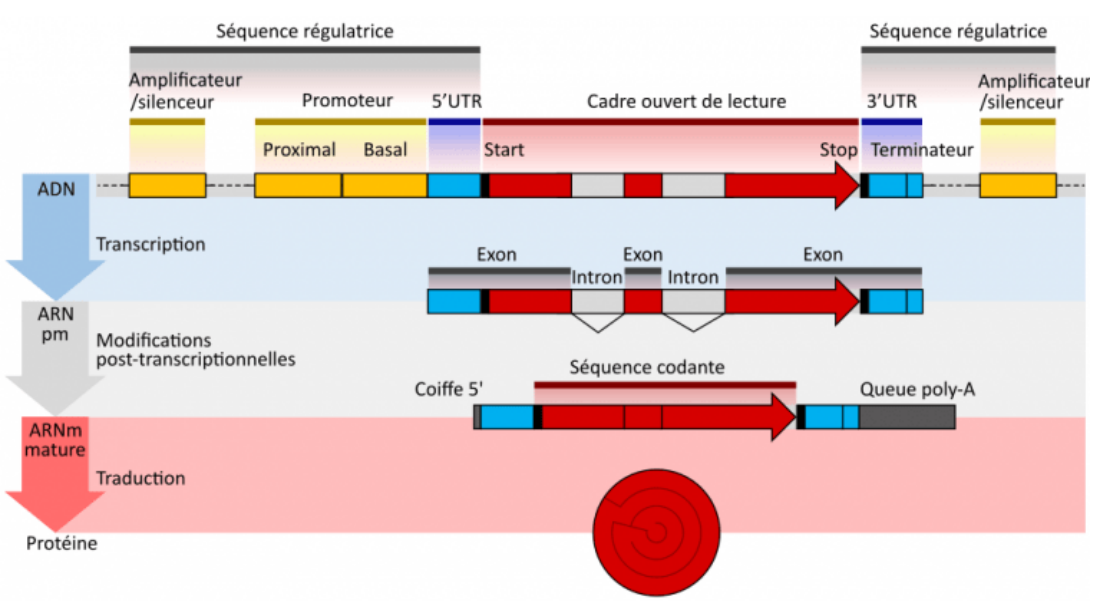


Figure 1 - Le dogme central de la biologie moléculaire : de l'ADN à la protéine en passant par l'ARN messager
Le gène, sous forme d'ADN, est d'abord transcrit en ARN prémessager qui, une fois maturé en ARN messager, est traduit en protéine.

Auteur(s)/Autrice(s) : Thomas Shafee traduit par Pascal Combemorel – Licence : CC-BY – Source : Wikimedia

1. La masse des protéines

1.1. Masses molaires des acides aminés

Tous les acides aminés possèdent un squelette commun, formé d'un carbone α, d'un groupement carboxyle COOH, d'un groupement amine NH_2 et d'un atome d'hydrogène (Figure 2). De formule brute $C_2H_4NO_2$, ce squelette pèse 74 g/mol[1]. Les masses des acides aminés varient alors en fonction de leurs chaînes latérales : la glycine est l'acide aminé le plus léger, tandis que le tryptophane est le plus lourd (Tableau 1).

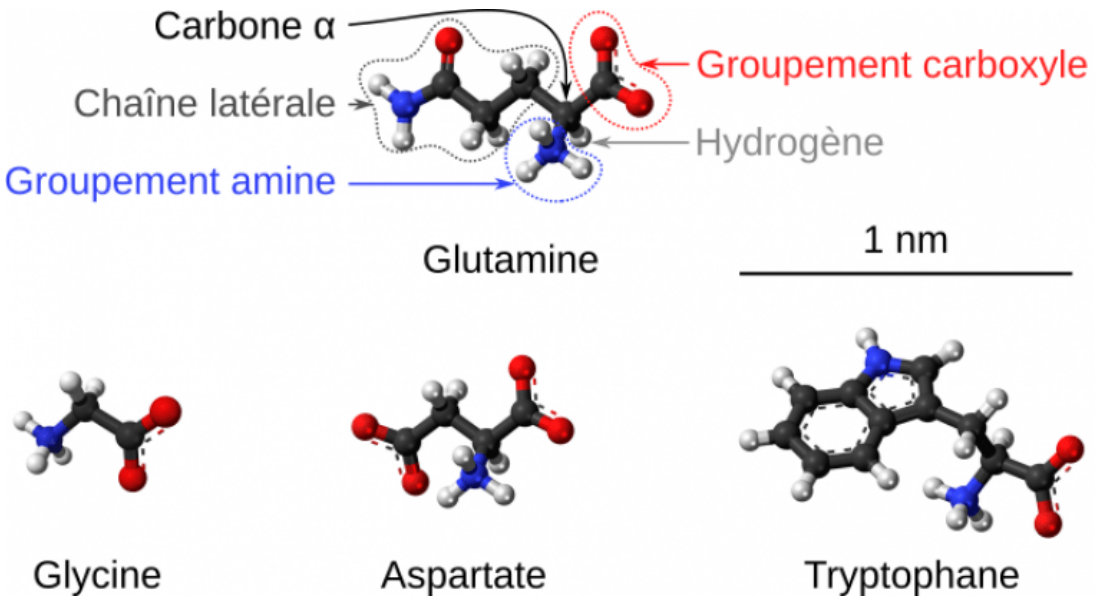


Figure 2 - Structure et comparaison de quelques acides aminés
Quatre acides aminés sont présentés à titre d'exemple. Les atomes sont représentés sous forme de boules (noir : carbone, bleu : azote, rouge : oxygène, blanc : hydrogène) reliés entre eux par des bâtonnets, symbolisant les liaisons chimiques. Sur toutes ces représentations, le groupement carboxyle COOH (ici sous forme carboxylate COO^-, comme c'est le cas au pH cellulaire) est placé à droite. Il est relié au carbone α, lui-même lié au groupement amine (ici sous forme ammonium NH_3^+, comme c'est le cas au pH cellulaire) visible en bas à gauche de celui-ci. La chaîne latérale, variable selon les acides aminés, occupe donc la partie gauche de chaque molécule. Chez la glycine, cette chaîne latérale se limite à un atome d'hydrogène.
Source des images : aspartate, glutamine et tryptophane : Jynto, CC0, Wikimedia. Glycine : Ben Mills, domaine public, Wikimedia.

Auteur(s)/Autrice(s) : Pascal Combemorel à partir des images créditées en légende – Licence : CC0

Masses molaires des différents acides aminés et fréquence dans les protéines

Nom	Masse molaire (g/mol)	Fréquence dans les séquences de la banque de données UniProtKB/Swiss-Prot
Alanine	89,09	8,26
Arginine	174,2	5,53
Asparagine	132,12	4,06
Aspartate	133,1	5,46
Cystéine	121,15	1,39
Glutamate	147,13	6,72
Glutamine	146,15	3,93
Glycine	75,07	7,07
Histidine	155,16	2,28
Isoleucine	131,17	5,91
Leucine	131,17	9,65
Lysine	146,19	5,8
Méthionine	149,21	2,41
Phénylalanine	165,19	3,87
Proline	115,13	4,75
Sérine	105,09	6,66
Thréonine	119,12	5,37
Tryptophane	204,23	1,11
Tyrosine	181,19	2,92
Valine	117,15	6,85

À partir de ce tableau, il est possible de déterminer la masse molaire moyenne d'un acide aminé. Le calcul de la moyenne brute donne un résultat de 136,90 g/mol. Toutefois, la pondération de la moyenne par la fréquence de chaque acide aminé au sein des protéines permet d'aboutir à une valeur de 129,11 g/mol.

1.2. Masses molaires des protéines

Pour calculer la masse molaire d'une protéine, il faut tenir compte des molécules d'eau (M_{H2O} = 18,01 g/mol) éliminées lors de la formation des liaisons peptidiques entre les acides aminés.

La masse molaire moyenne d'une protéine formée par n acides aminés, et contenant donc (n − 1) liaisons peptidiques est :

$M_{protéine} = n \times 129{,}11 - (n - 1) \times 18{,}01$

$M_{protéine} = n \times 111{,}10 + 18{,}01$

$M_{protéine} \sim n \times 111$ (Équation 1)

La masse moyenne d'un acide aminé présent dans une protéine est d'environ 111 g/mol.

1.3. Quelques exemples de masses molaires de protéines

Le site UniProt fournit de nombreuses informations sur les protéines dont, en particulier, leurs nombres d'acides aminés et leurs masses molaires[4]. Les masses molaires estimées d'après l'Équation 1 ainsi que les masses molaires réelles de quatre protéines sont données dans le Tableau 2. La lecture de ce tableau montre que l'approximation $M_{protéine} \sim n \times 111$ donne des estimations très proches des valeurs réelles.

Masses molaires réelles et masses molaires estimées de quatre protéines

Protéine	Référence UniProt	Organisme	Nombre d'acides aminés	Masse molaire estimée d'après l'Équation 1 (g/mol)	Masse molaire réelle (g/mol)
Prépro-insuline	P01308	*Homo sapiens*	110	12 210	11 981
Sous-unité α 1 de l'hémoglobine	P69905	*Homo sapiens*	142	15 762	15 258
Grande sous-unité de la rubisco	O03042	*Arabidopsis thaliana*	479	53 169	52 955
Titine	Q8WZ42	*Homo sapiens*	34 350	3 812 850	3 816 030

2. La taille des protéines

La taille des protéines est intrinsèquement liée à leur forme. Les protéines peuvent très schématiquement être regroupées en deux catégories de formes : globulaires ou fibrillaires. Il existe également des protéines désordonnées, c'est-à-dire sans forme stable. Par ailleurs, les protéines globulaires et fibrillaires peuvent présenter des portions désordonnées. Les protéines globulaires à structure tertiaire mesurent généralement quelques nanomètres de diamètre. Par exemple l'insuline humaine (2HIU[5]) mesure environ 3 nm, tandis que la myoglobine de grand cachalot (1MBD) fait entre 4 et 5 nm de diamètre (Figure 3). Les protéines à structure quaternaire, composées de plusieurs sous-unités, sont logiquement plus grandes. Par exemple la rubisco d'épinard (1RCX) présente un diamètre de 12 nm environ, alors que le ribosome de levure de boulanger (6SNT) a une taille de 25 nm environ[6].

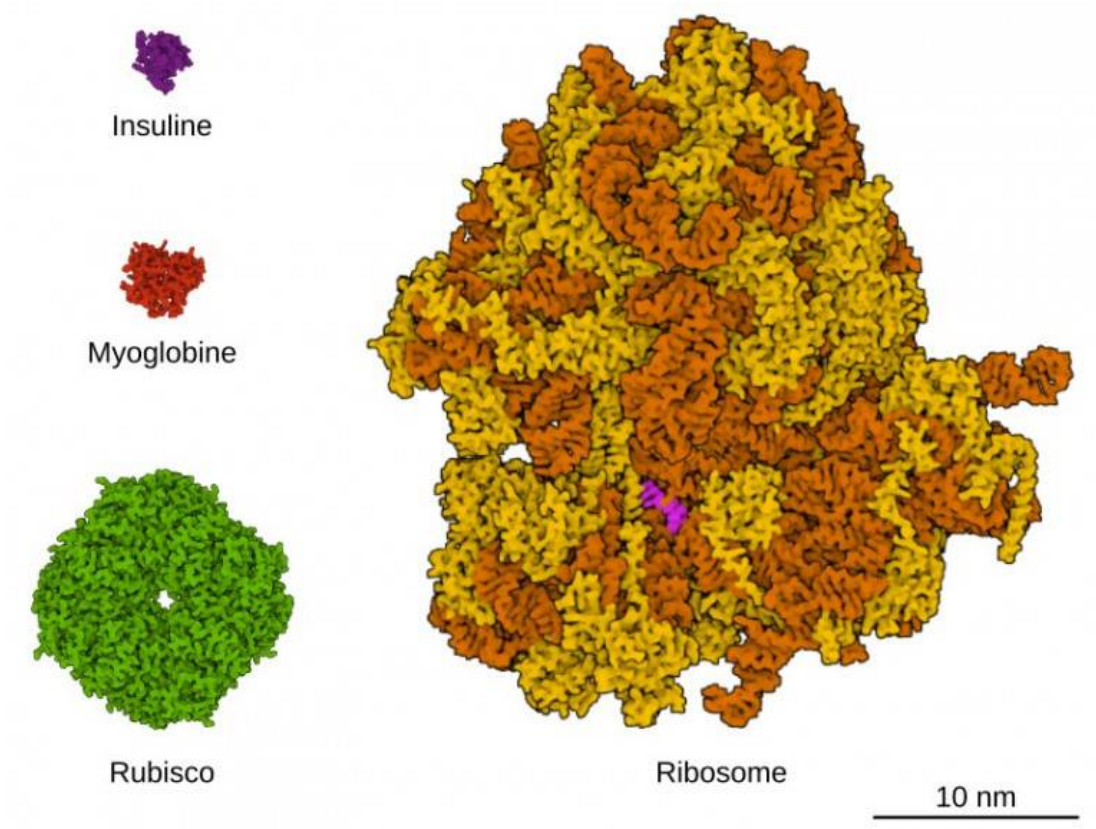


Figure 3 - Comparaison de la taille de quelques protéines globulaires
Cette image a été produite à l'aide du logiciel Mol* grâce à des structures protéiques disponibles sur le site de la banque de données sur les protéines (PDB) : insuline humaine (2HIU), myoglobine de grand cachalot (1MBD), rubisco d'épinard (1RCX) et ribosome de levure de boulanger (6SNT). Sur ce dernier, les protéines sont colorées en jaune, les ARNr en orange et un morceau de l'ARN messager est visible en magenta.

Auteur(s)/Autrice(s) : Pascal Combemorel – Licence : CC0

Les protéines fibrillaires sont formées de sous-unités qui sont soit elles-mêmes fibrillaires, soit globulaires. Le premier cas de figure est représenté par le collagène, formé de trois hélices enroulées. Le tout mesure environ 300 nm de long pour un diamètre de 1,5 nm. Dans la seconde catégorie se trouvent par exemple les microfilaments d'actine et les microtubules. Les microfilaments d'actine se composent d'actine G, globulaire. Assemblées, ces sous-unités produisent des microfilaments d'actine pouvant atteindre plusieurs micromètres de long. Les microtubules sont quant à eux formés de 13 protofilaments eux-mêmes composés de protéines globulaires, les tubulines α et β (le dimère αβ est visible à l'entrée 1TUB de la banque de données sur les protéines). Les microtubules forment des tubes creux d'un diamètre de 25 nm et d'une longueur pouvant atteindre plusieurs micromètres.

Un aperçu de la diversité des protéines du vivant

Deux ressources en ligne permettent d'appréhender la diversité de tailles et de formes des protéines du vivant. La première, en deux dimensions, est hébergée par le site de la banque de données sur les protéines (PDB). La seconde, en trois dimensions, utilise le logiciel Mol* (à prononcer Molstar).

3. La masse des ARN messagers

3.1. Masses molaires des nucléotides

Les ARN messagers sont formés de quatre ribonucléotides monophosphates différents (Figure 4 et Tableau 3).

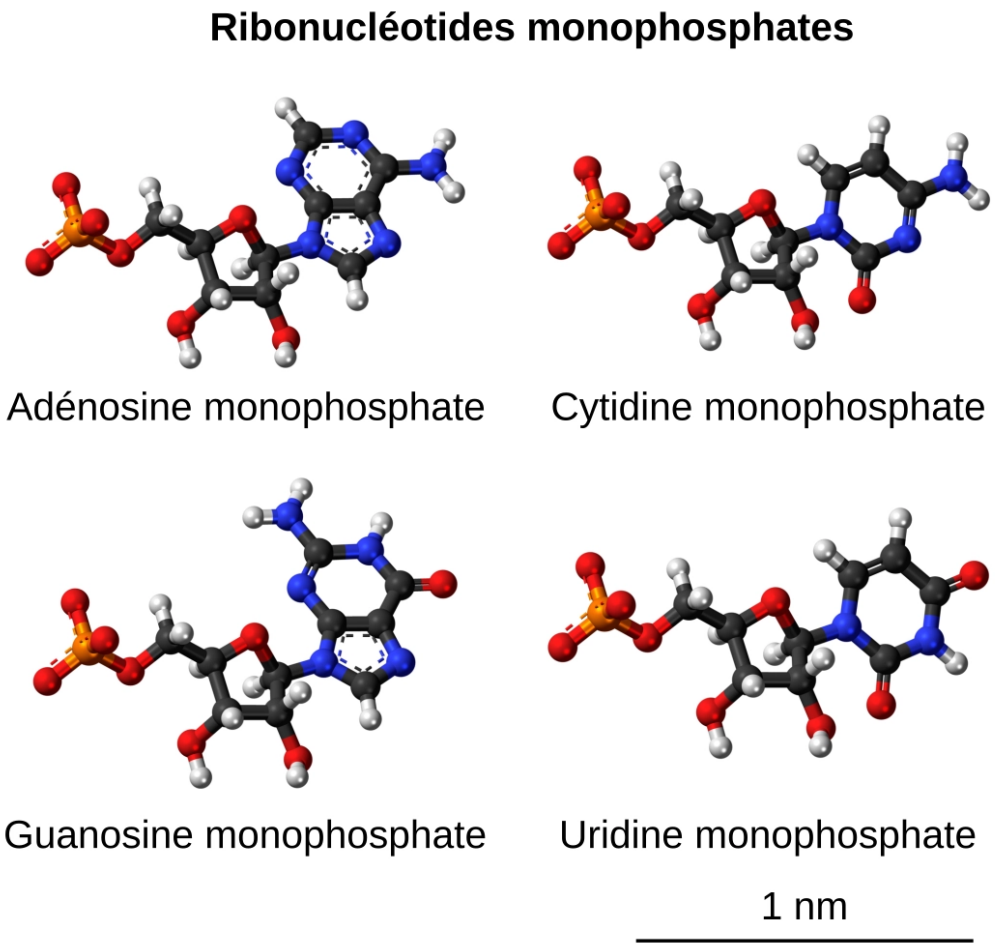


Figure 4 - Les quatre ribonucléotides monophosphates composant l'ARN
Les atomes sont représentés sous forme de boules (noir : carbone, bleu : azote, rouge : oxygène, blanc : hydrogène, orange : phosphore) reliés entre eux par des bâtonnets, symbolisant les liaisons chimiques. Sur toutes ces représentations, le phosphate est à gauche, le ribose au centre et la base azotée à droite.
Source des images : Jynto, CC0, Wikimedia : adénosine monophosphate, cytidine monophosphate, guanosine monophosphate, uridine monophosphate

Auteur(s)/Autrice(s) : Pascal Combemorel à partir des images créditées en légende – Licence : CC0

Formules et masses molaires des quatre ribonucléotides constituant les ARN messagers

Ribonucléotides monophosphates	Formules brutes	Masses molaires (g/mol)
Adénosine monophosphate	$C_{10}H_{12}N_5O_7P^{2-}$	345,21
Uridine monophosphate	$C_9H_{11}N_2O_9P^{2-}$	322,17
Cytidine monophosphate	$C_9H_{12}N_3O_8P^{2-}$	321,18
Guanosine monophosphate	$C_{10}H_{12}N_5O_8P^{2-}$	361,21

Les formules brutes sont données ici pour des nucléotides dont le groupement phosphate porte deux charges négatives, ce qui est le cas au pH physiologique.

La masse molaire moyenne d'un nucléotide monophosphate, calculée à partir des quatre valeurs du tableau, vaut 337,44 g/mol.

3.2. Liaison phosphodiester

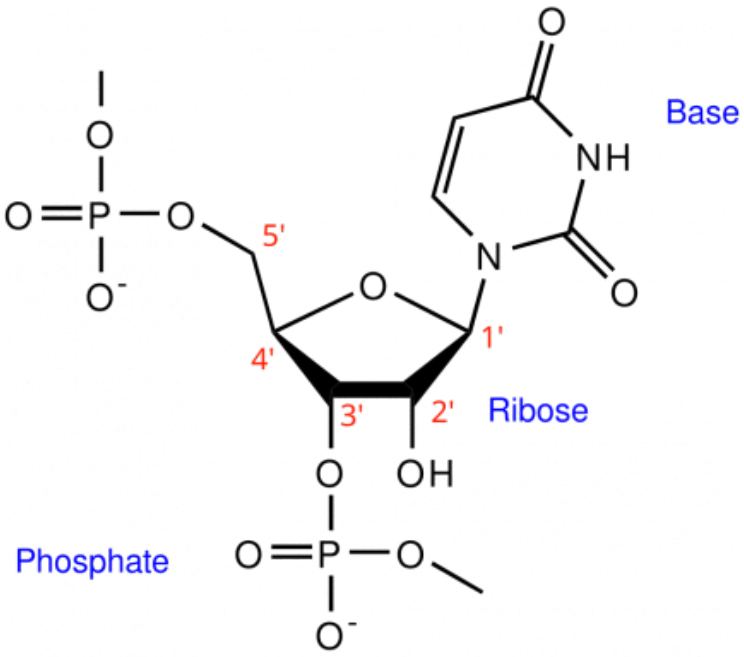


Figure 5 - Liaison phosphodiester au sein de l'ARN
Par convention, les atomes de carbone de la base azotée sont numérotés 1, 2, 3, 4... et les atomes de carbone du ribose sont numérotés 1', 2', 3'..., le numéro 1' étant attribué au carbone impliqué dans la liaison avec la base azotée.

Auteur(s)/Autrice(s) : Fdardel – Licence : CC-BY-SA – Source : Wikimedia

De même que les protéines sont des polymères d'acides aminés, les molécules d'ARN sont des polymères de ribonucléotides. La liaison mise en jeu est une *liaison phosphodiester*, impliquant un groupe -OH de la partie "ribose" d'un premier ribonucléotide et le groupe phosphate d'un second ribonucléotide.

Dans l'ARN, la liaison phosphodiester se forme entre l'extrémité 3' OH sur le ribose du premier nucléotide vers l'extrémité 5' phosphate du second nucléotide, par le biais d'une ARN polymérase. La réaction mise en jeu s'apparente à une réaction d'estérification.

3.3. Masses molaires des ARNm

Pour calculer la masse molaire d'une molécule d'ARN, il faut tenir compte de l'élimination d'un groupement OH (M_{OH} = 17,01 g/mol) lors de la formation des liaisons phosphodiesters.

La masse molaire d'une molécule d'ARN formée par n ribonucléotides, et contenant donc (n – 1) liaisons phosphodiesters est :

$M_{ARN} = n \times 337{,}44 - (n - 1) \times 17{,}01$

$M_{ARN} = n \times 320{,}43 + 17{,}01$

$M_{ARN} \sim n \times 320$ (Équation 2)

La masse moyenne d'un ribonucléotide présent dans un ARN est d'environ 320 g/mol.

3.4. Quelques exemples de masses molaires d'ARNm

L'outil *Nucleotide* du site du *National Center for Biotechnology Information* fournit de nombreuses informations sur les gènes et leurs transcrits dont, en particulier, la longueur (en nombre de nucléotides) des ARNm. Les masses molaires estimées d'après l'Équation 2 ainsi que les masses molaires réelles de quatre ARNm sont données dans le Tableau 4. L'approximation $M_{ARN} \sim n \times 320$ donne des estimations très proches des valeurs réelles.

Masses molaires réelles et masses molaires estimées de quatre ARN messagers

ARNm	Référence	Organisme	Nombre de nucléotides	Masse molaire estimée d'après l'Équation 2	Masse molaire réelle, d'après la séquence
Insuline (variant 1)	NM_000207.3	*Homo sapiens*	465	148 800	148 796
Sous-unité α 1 de l'hémoglobine	NM_000558.5	*Homo sapiens*	577	184 640	183 553
Grande sous-unité de la rubisco	L24073.1	*Oryza sativa*	1355	433 600	432 851
Titine, isoforme N2-A	NM_133378	*Homo sapiens*	101 520	32 486 400	32 510 844

Notez que les banques de données fournissent le nombre total de nucléotides de l'ARNm (sans la queue polyA toutefois, puisque celle-ci n'est pas issue de la transcription du génome, mais ajoutée après la transcription) et non pas uniquement de la séquence codante. Autrement dit, la longueur des extrémités non traduites en 5′ (5′UTR) et en 3′ (3′UTR) sont également prises en compte (Tableau 5).

Tailles de l'extrémité 5'UTR, de la séquence codante, de l'extrémité 3'UTR et tailles totales de quatre ARN messagers

ARNm	Référence	Nombre de nucléotides de l'extrémité 5'UTR	Nombre de nucléotides de la séquence codante (y compris codon stop)	Nombre de nucléotides de l'extrémité 3'UTR	Nombre de nucléotides de l'ARNm mature (sans la queue polyA)
Insuline (variant 1)	NM_000207.3	59	333	73	465
Sous-unité α 1 de l'hémoglobine	NM_000558.5	37	429	111	577
Grande sous-unité de la rubisco	L24073.1	34	1 206	115	1 355
Titine, isoforme N2-A	NM_133378	225	100 272	1023	101 520

Chez les êtres vivants pluricellulaires, la longueur moyenne de l'extrémité 5' non traduite est d'environ 300 nucléotides, tandis qu'elle est d'environ 700 nucléotides pour l'extrémité 3' non traduite (calculs réalisés grâce au site de la banque de données sur les extrémités non traduites, UTRdb).

De plus, les ARNm peuvent porter des modifications post-transcriptionnelles, dont certaines sont réversibles, qui ne sont pas prises en compte ici. La plus étudiée d'entre elles est l'addition d'un groupement méthyle en position 6 sur les adénines, modification qui régule la traduction et le métabolisme des ARNm.

4. La taille des ARN messagers

Les ARN sont composés d'un enchaînement de ribonucléotides monophosphates. Chaque nucléotide occupe une « hauteur » de 0,5 nm environ. Cependant, il n'est pas possible de déterminer simplement la longueur totale d'un ARN messager en multipliant la place occupée par un nucléotide par le nombre de nucléotides. En effet, les ARN messagers, comme les autres ARN (ARNt, microARN…) présentent des structures secondaires par appariement de nucléotides (structure tige-boucle par exemple). Dans les ARNm, ces structures secondaires sont séparées par des régions non structurées et la flexibilité qui en résulte rend difficile la détermination expérimentale de la structure tridimensionnelle des ARNm[7].

Pour se représenter convenablement les ARN messagers dans un contexte cellulaire il faut également considérer que ces molécules ne sont pas nues, mais associés à de nombreuses protéines. Ainsi, chez les Eucaryotes, durant la transcription, des protéines se lient au niveau des jonctions introns-exons et contrôlent l'épissage. Les ARN messagers matures sont également liés à de nombreuses protéines : protéines s'associant à la coiffe 5' et à la queue polyA en 3' et circularisant les ARN messagers, protéines du complexe de la jonction exonique et, bien entendu, ribosomes.

5. Bilan

5.1. Taille relative des nucléotides et des acides aminés

Les nucléotides, qui mesurent environ 1,1 nm, sont à peu près 1,5 fois plus grands que la taille moyenne d'un acide aminé, de 0,7 nm (Figure 5). Par ailleurs, leur masse molaire est environ trois fois plus élevée (320 g/mol en moyenne pour un ribonucléotide, contre 111 g/mol pour un acide aminé).

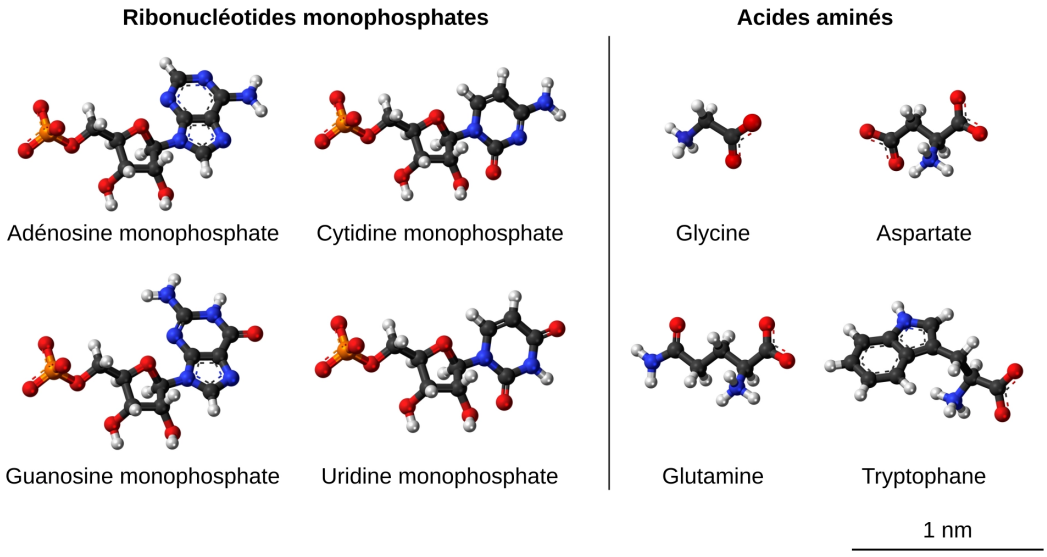


Figure 6 - Comparaison de la taille des ribonucléotides monophosphates et de quelques acides aminés
Les différents nucléotides sont représentés sous leur forme ionisée. Ils mesurent environ 1,1 nm. Le plus petit des acides aminés, la glycine, a une longueur d'à peu près 0,4 nm tandis que le plus grand, le tryptophane, mesure environ 1 nm.
Source des images : Jynto, CC0, Wikimedia : adénosine monophosphate, cytidine monophosphate, guanosine monophosphate, uridine monophosphate, acide aspartique, glutamine, tryptophane. Ben Mills, domaine public, Wikimedia : glycine

Auteur(s)/Autrice(s) : Pascal Combemorel à partir des images créditées en légende – Licence : CC0

5.2. Taille relative des ARNm et des protéines qu'ils codent

Une fois polymérisés, la masse molaire d'un nucléotide ou d'un acide aminé est un peu plus faible que celle de sa version libre, à cause de l'élimination d'un groupement hydroxyle ou d'une molécule d'eau, respectivement. Cependant, la masse molaire d'un nucléotide reste toujours trois fois supérieure à celle d'un acide aminé ($M_{nucléotide}$ = 320 g/mol ; $M_{acideaminé}$ = 111 g/mol).

Par ailleurs, comme il faut un triplet de nucléotides pour coder un acide aminé, cela signifie que la masse d'un ARNm est au moins neuf fois supérieure à celle de la protéine qu'il code. En réalité, ce rapport est souvent encore plus élevé, car il faut également tenir compte de la masse des séquences non traduites (5' UTR et 3'UTR ; Tableau 6). Ainsi, dans les cellules, la notice de fabrication (l'ARNm) est bien plus lourde que le produit lui-même (la protéine) !

Comparaison de la masse de l'ARN messager et de la protéine dans quatre cas de figure

	Masse ARNm	Masse protéine	Rapport
Insuline (variant 1)	148 796	11 981	12
Sous-unité α 1 de l'hémoglobine	183 553	15 258	12
Grande sous-unité de la rubisco	432 851	52 955	8
Titine, isoforme N2-A	32 510 844	3 816 030	9

L'observation d'une unité de traduction en microscopie électronique à transmission permet de se rendre compte de la taille relative d'un ARN messager et de la protéine qu'il code (Figures 6 et 7). Il est également possible de transmettre cette idée sous forme de schéma (Figure 8).



Figure 7 - Unité de traduction d'un ARN messager observée en microscopie électronique à transmission
Cette image, obtenue à partir d'une glande salivaire de chironome (*Chironomus sp.*) permet d'apprécier la taille relative d'un ARN messager (long filament occupant toute la largeur de l'image), des ribosomes qui le décodent (boules jaunes le long de l'ARNm) et des protéines ainsi produites (courts filaments émergeant des ribosomes).

Auteur(s)/Autrice(s) : Science Photo Library / Kiseleva, Dr. Elena – Licence : Reproduit avec autorisation – Source : Science Photo Library

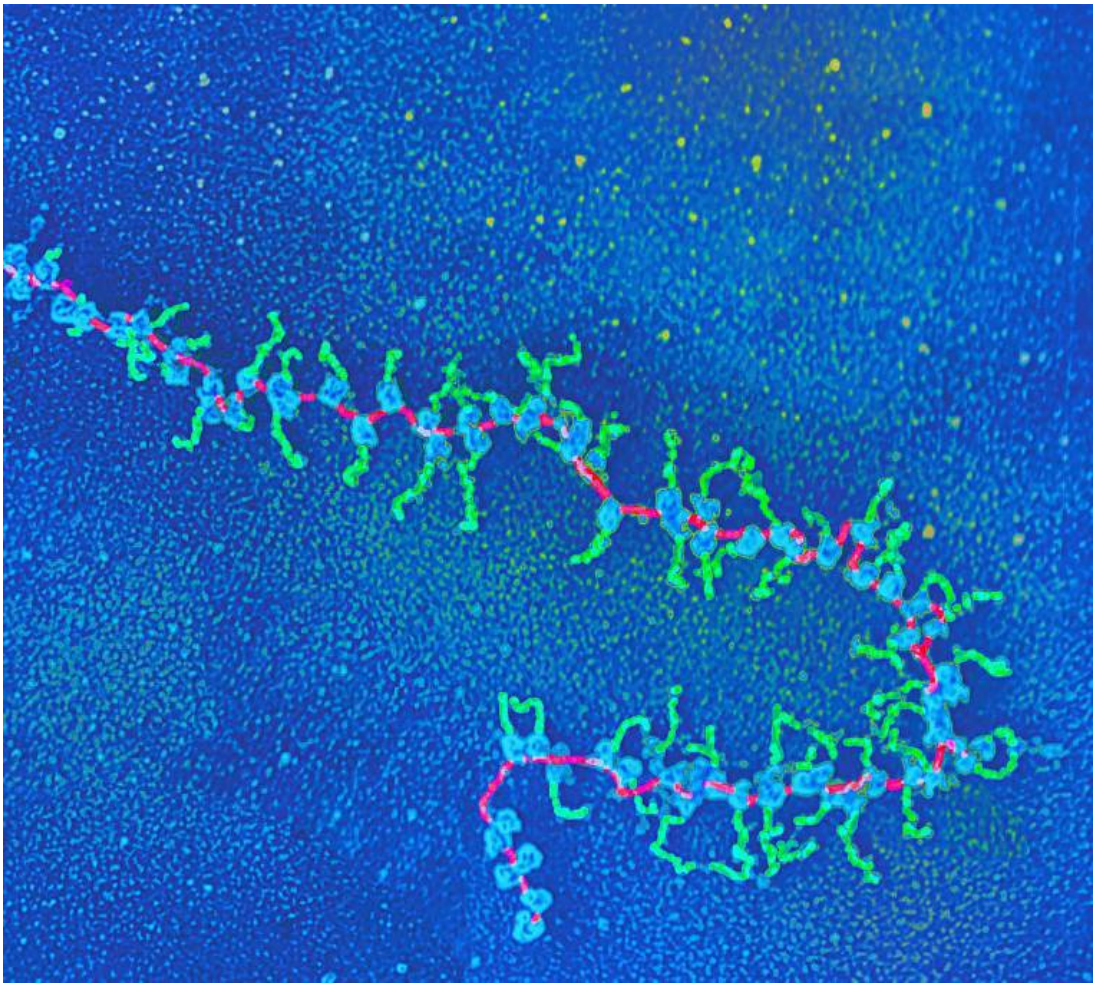


Figure 8 - Détail d'une unité de traduction d'un ARN messager observée en microscopie électronique à transmission

Cette image, obtenue à partir d'une glande salivaire de chironome (Chironomus sp.) permet d'apprécier la taille relative d'un ARN messager (en rose), des ribosomes qui le décodent (en bleu) et des protéines ainsi produites (en vert).

Auteur(s)/Autrice(s) : Science Photo Library / Kiseleva, Dr. Elena – Licence : Reproduit avec autorisation – Source : Science Photo Library

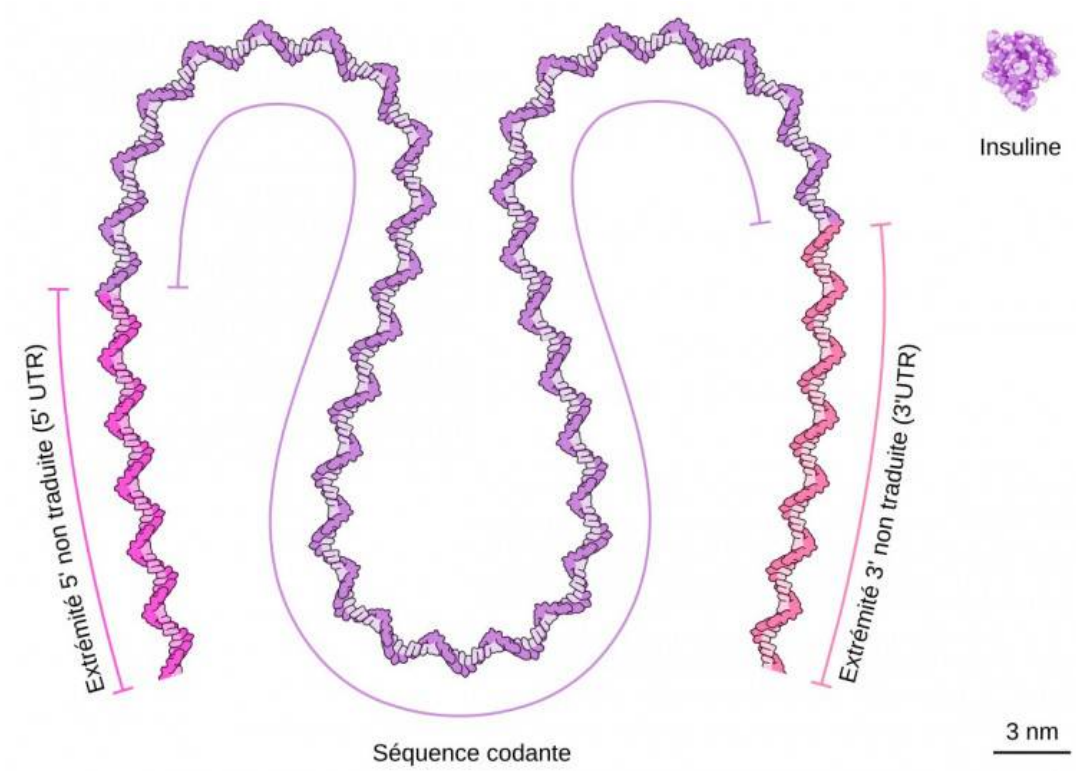


Figure 9 - Taille relative de l'ARN messager codant l'insuline et de l'insuline
L'insuline est une petite protéine (51 acides aminés) codée par un ARNm relativement court (465 nucléotides se décomposant en 59 nucléotides à l'extrémité 5' UTR, 333 nucléotides au niveau de la séquence codante et 73 nucléotides à l'extrémité 3'UTR). La queue polyA n'est pas représentée. La traduction de l'ARNm donne la prépro-insuline, longue de 110 acides aminés, qui est ensuite maturée en insuline (51 acides aminés). NB : L'ARN est ici représenté sous forme simple brin, sans former de structure secondaire, ce qui n'est probablement pas le cas physiologiquement, où les ARN messagers ne sont pas nus, mais associés à de nombreuses protéines. Le schéma présenté ici ne vise donc pas à rendre compte de la structure réelle de l'ARNm, mais uniquement de sa différence de taille par rapport à la protéine qu'il code.

Auteur(s)/Autrice(s) : Pascal Combemorel – Licence : CC-BY

6. Conclusion

Il ne s'agit bien sûr pas ici de remettre en cause la pertinence de schémas fonctionnels comme celui présenté Figure 1. Cependant, de tels schémas gagneraient à être complétés par des images obtenues en microscopie électronique à transmission (Figures 6 et 7), ou encore par des schémas dans lesquels les tailles relatives des différents éléments sont respectées (Figure 8, qui présente cependant comme limite de ne pas donner la structure physiologique de l'ARN messager, et vise uniquement à rendre compte de la différence de taille entre ARNm et protéine). De telles représentations traduisent en effet mieux le volume réellement occupé par les différentes molécules.

CRÉDITS

AUTEUR(S)/AUTRICE(S)

Pascal Combemorel

Agrégé de SVT, il est le responsable éditorial du site Planet-Vie depuis septembre 2016.

RELECTURE SCIENTIFIQUE

Roland Marquet

Après des études de chimie et un doctorat en biophysique, il est directeur de recherche à l'Institut de biologie moléculaire et cellulaire de Strasbourg et étudie les ARN viraux.

LICENCE DU TEXTE DE L'ARTICLE

Creative Commons - Attribution

PARTENAIRE(S)

Cet article est co-publié sur Planet-Vie, site de ressources pour enseigner la biologie. La section sur la liaison phosphodiester a été ajoutée dans la version publiée sur CultureSciences-Chimie.
Planet-Vie

NOTES

1. Les masses des molécules biologiques sont parfois données en daltons (Da), une unité ne faisant pas partie du système international d'unités, à la différence de l'unité de masse atomique unifiée (u). 1 Da = 1 u ≈ masse d'un atome d'hydrogène.
2. D'après la page *Acide aminé protéinogène* de Wikipédia, consultée le 6 mai 2026.
3. D'après les 575 503 séquences protéiques recensées par la banque de données UniProtKB/Swiss-Prot, consultée le 30 juin 2026. Il faut noter que les données alimentant cette banque proviennent d'environ 15 000 espèces « seulement ».
4. Sur la page d'une protéine (par exemple celle consacrée à la prépro-insuline), le nombre d'acides aminés est donné en haut de page, au niveau du champ *Amino acids*, tandis que la masse molaire est indiquée dans la section *Sequence*, au niveau du champ *Mass (Da)*.
5. Les différents liens donnés dans ce paragraphe pointent vers la banque de données sur les protéines (PDB). Pour explorer les structures tridimensionnelles des protéines évoquées, il est également possible d'utiliser un outil plus simple à prendre en main, comme LibMol.

6 On considère ici la taille totale de l'ensemble grande sous-unité + petite sous-unité. Par ailleurs, il faut noter que le ribosome n'est pas exclusivement protéique : s'il est constitué de 79 protéines totalisant environ 13 000 acides aminés, il comporte également quatre molécules d'ARN, appelées ARN ribosomiques (ARNr), comprenant un total de 5 475 nucléotides.

7 La structure tridimensionnelle de certains ARN structurés, seuls ou en complexe avec des protéines, a cependant été déterminée. Ainsi, les ARNt dont la longueur est d'environ 75 nucléotides et la masse molaire d'environ 25 300 g/mol adoptent une structure tridimensionnelle en forme de L dont la distance entre les extrémités est de 7 à 8 nm.