

Les surprises du décodage de l'information génétique

Guillaume Stahl
Jean-Pierre Rousset

La conservation du code génétique et de son décryptage par la machinerie traductionnelle est une des universalités du monde vivant. Le génie génétique en fournit une illustration quotidienne en permettant de produire des protéines d'origine humaine chez une bactérie ou chez une plante. Le biologiste moléculaire sait traduire facilement une séquence nucléotidique codante en une séquence protéique et des programmes informatiques simples ont été développés pour automatiser cette tâche. Cependant, pour lire le message génétique, la cellule ne dépend pas d'un logiciel mais d'un mécanisme biologique complexe qui est parfois sujet à des déviations. Il ne s'agit pas ici des exceptions au code universel que l'on observe pour tous les gènes de certains organismes comme les paramécies, mais de modifications qui affectent des gènes spécifiques et qui dépendent de séquences précises de l'ARN messager. L'importance biologique de ces déviations ne doit pas être négligée : ce sont elles qui sont responsables, par exemple, de la synthèse de la transcriptase inverse des rétrovirus. A travers les éléments accumulés ces dernières années, sur les mécanismes moléculaires de ces nouveaux modes de codage chez les eucaryotes, apparaissent les enjeux de ces travaux en recherche fondamentale et appliquée.

C'est au début des années 1970 que l'on a pu montrer pour la première fois l'existence de phénomènes extraordinaires qui impliquent des erreurs de traduction à un haut niveau et de manière reproductible sur certains ARN messagers. La traduction peut ainsi être modifiée au cas par cas : cadre de lecture modifié (décalage de phase), codons stop non reconnus (translecture), ou même « saut de ribosome ». Plutôt

que des erreurs, ces événements impliquent un codage spécifique qui modifie la lecture normale du message génétique par le ribosome. Ils ont été baptisés *recoding* en anglais [1], que l'on pourrait traduire par « recodage »*. Des exemples de plus en plus nombreux de recodage sont

ADRESSES

G. Stahl: Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E. Pratt street, Baltimore, MD 21202, États-Unis. J.P. Rousset: Institut de génétique et microbiologie. Bâtiment 400, Université Paris-Sud, 91405 Orsay Cedex, France.

* La thèse de Guillaume Stahl, qui développe les points évoqués ici, peut être consultée sur le site: <http://www.igmors.u-psud.fr/rousset/tgs/these.html>

apparus chez les virus à partir de 1985, mais ce n'est que très récemment que le phénomène a pu être étendu à des gènes cellulaires impliqués dans des fonctions essentielles telles que le maintien des télomères chez la levure ou le développement de la drosophile. Ce domaine est encore en phase d'exploration et il est clair que : (1) toutes les règles ne sont pas encore connues pour les recodages « classiques » ; et (2) de nouveaux types de recodage peuvent être mis en évidence. Le ribosome n'a certainement pas encore livré tous ses secrets...

Décalage de phase en -1

Le décalage programmé du ribosome d'un nucléotide en amont est observé à la jonction entre les gènes *gag* et *pol*, ou leurs équivalents, chez la plupart des rétrovirus, dans des virus de plantes, de levure, des éléments transposables bactériens ou eucaryotes [2]. Ainsi, chez les rétrovirus, la séquence du gène *pol*, qui code pour les fonctions catalytiques nécessaires au cycle viral, recouvre dans la phase -1 la séquence du gène *gag*, qui code pour une des protéines structurales du virion (figure 1). La majorité des ribosomes va démarrer au premier codon AUG de *gag* et la synthèse va se terminer au premier codon stop, donnant naissance à la protéine GAG. Cependant, quelques ribosomes (de 5% à 20% suivant les virus) vont changer de cadre de lecture au niveau de la jonction *gag-pol* et continuer la traduction dans la phase -1, aboutissant à la synthèse d'une protéine de fusion GAG-POL (figure 1). Cette polyprotéine porte la partie GAG à son extrémité amino-terminale et les domaines protéase, transcriptase inverse, RNase H et intégrase à son extrémité carboxy-terminale. Elle est incorporée dans la nucléocapside, ce qui permet son transfert dans la cellule nouvellement infectée, en même temps que le génome viral. Elle est ensuite mûrie par son domaine protéase, ce qui libère les différents peptides nécessaires à la poursuite du cycle viral. Ce phénomène, d'abord décrit pour le virus du sarcome de Rous, a été ensuite retrouvé dans la majorité des rétrovirus dont le virus de l'immunodéficience humaine de type 1 (VIH-1) [2].

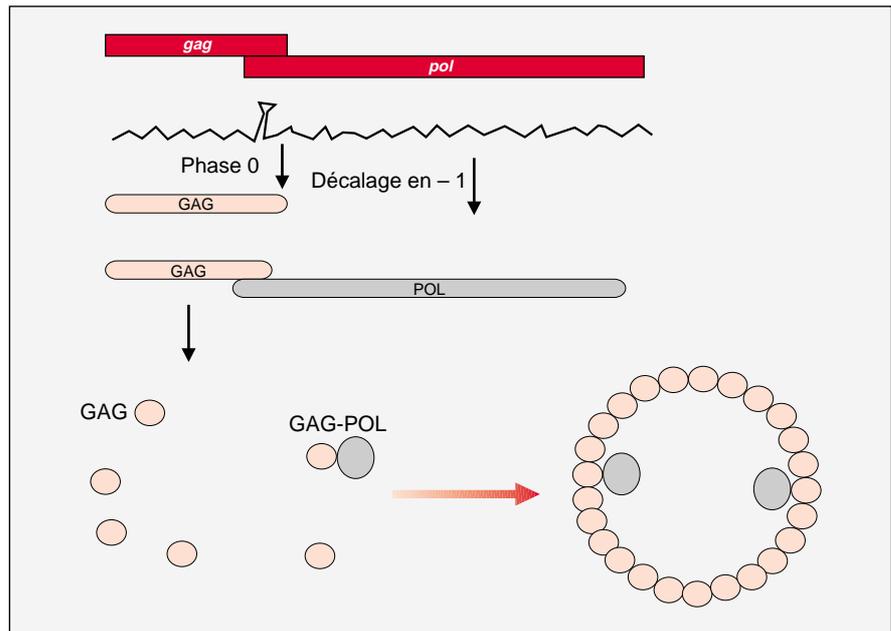


Figure 1. **Expression différentielle des gènes *gag* et *pol* par décalage de phase en -1 chez les rétrovirus.** Un seul ARNm porte les informations pour la synthèse des deux protéines GAG et GAG-POL. Lors de la traduction de cet ARNm, la majorité des ribosomes décode le message de façon « classique », ce qui aboutit à la synthèse de la protéine GAG. Une faible proportion des ribosomes se décale d'un nucléotide en amont, ce qui entraîne la synthèse d'une protéine de fusion GAG-POL, qui est incorporée dans le virion. C'est grâce à ce mécanisme que les activités transcriptase inverse, protéase et intégrase sont introduites dans la cellule nouvellement infectée et qu'un nouveau cycle de réplication virale va pouvoir démarrer.

La jonction *gag-pol* contient une séquence glissante qui est un heptanucléotide de type -X XXY YYZ. Une mutation de la dernière base de l'heptamère modifie deux acides aminés : le dernier acide aminé avant le décalage et le premier acide aminé après le décalage (figure 2A). Ce nucléotide est donc décodé deux fois, par deux ARNt différents [3]. Par ailleurs, une mutation du premier nucléotide de l'heptamère diminue fortement le taux de décalage de phase en -1. Cela suggère que le peptidyl-ARNt et l'aminocyl-ARNt, présents respectivement aux sites P et A du ribosome, se décalent simultanément (*tandem shift*). Grâce à la flexibilité de l'appariement codon-anticodon au niveau de la troisième base (règle du *wobble*), le ribosome peut se réappairer sur les séquences homopolymériques et lire la séquence *pol* dans la nouvelle phase (figure 2B). La quasi-totalité des décalages de phase en -1 étudiés à ce jour chez les eucaryotes correspondent au modèle décrit ci-dessus [2]. Cepen-

nant, tous les nucléotides ne sont pas interchangeables [4, 5]. Au niveau du premier homopolymère, les pyrimidines provoquent un meilleur décalage de phase que les purines, dans l'ordre U > C > A > G. Au niveau du deuxième homopolymère, seules les bases A et U permettent d'obtenir des taux significatifs de décalage de phase. Ces contraintes reflètent probablement la possibilité pour les ARNt de se désappairer efficacement du premier codon pour se réappairer avec le codon en phase -1. On ne peut cependant pas exclure qu'il y ait des ARNt qui soient particulièrement enclins à provoquer le décalage de phase [3]. En aval de l'heptanucléotide, on retrouve presque toujours une séquence pouvant adopter une structure secondaire de type tige-boucle (figure 3). La délétion de cette séquence diminue considérablement le décalage de phase et des mutations ponctuelles ont permis de montrer que l'appariement était effective-

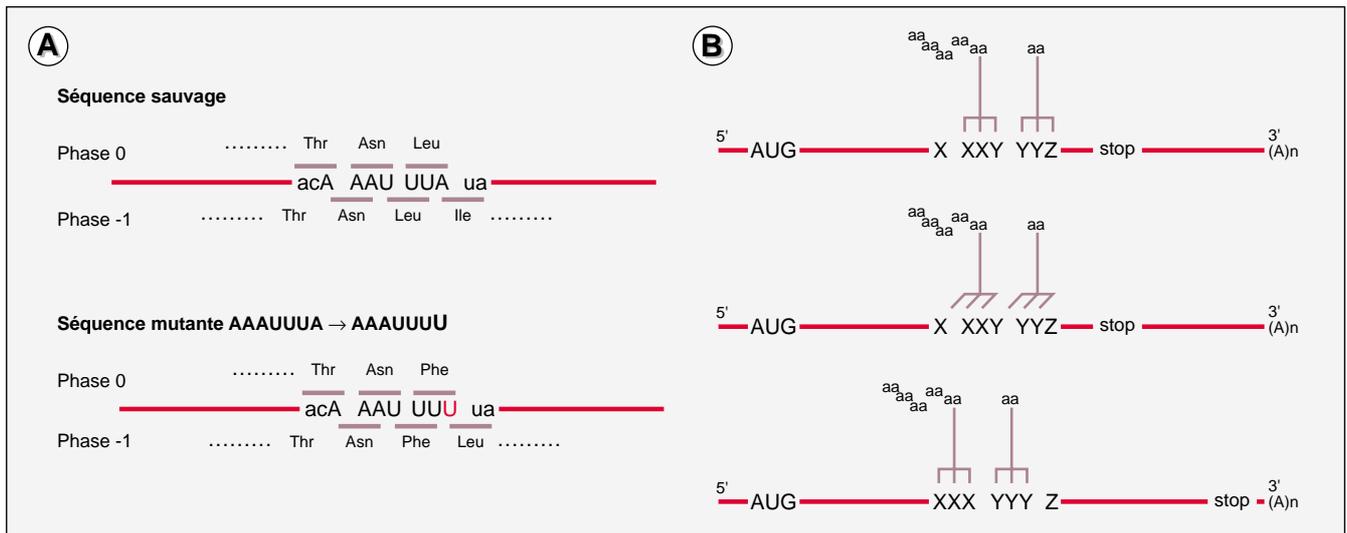


Figure 2. **Le site de glissement du ribosome en -1 est un heptanucléotide.** **A.** La mutation de la dernière base de la séquence heptamérique glissante, ici A vers U, entraîne la modification de deux acides aminés (Leu vers Phe et Ile vers Leu) dans la protéine de fusion. Cela démontre que ce nucléotide est décodé deux fois, par deux ARNt différents. **B.** Dans la grande majorité des cas, le peptidyl-ARNt, situé au site P (pour « peptide ») du ribosome et l'aminoacyl-ARNt, situé au site A (pour « acide aminé »), se décalent simultanément. Cela est possible grâce à l'identité des nucléotides (XXX et YYY) et par la flexibilité de l'appariement codon/anticodon, au niveau de la troisième base wooble.

ment important pour la stimulation. Cependant, chez certains virus comme le virus de la bronchite infectieuse (IBV), la séquence en aval de l'heptanucléotide est capable de constituer une structure plus complexe: un pseudo-nœud de type H (pour *hairpin-like*), où la boucle est appariée avec une séquence située en aval. Cette structure peut être vue comme deux tiges co-axiales empiilées, reliées par une boucle (figure 3). Les pseudo-nœuds sont les structures les plus fréquemment retrouvés mais certains virus, comme le VIH-1, utilisent une simple tige-boucle pour stimuler le décalage de phase. Nous avons pu démontrer que l'efficacité du décalage de cadre du VIH-1 est directement proportionnelle à la stabilité de la tige-boucle, indépendamment de sa séquence primaire [6]. Dans les autres systèmes analysés, il ne semble pas non plus y avoir d'effet de la séquence primaire de la région stimulatrice. En revanche, une modification de la distance entre la séquence glissante et la structure secondaire affecte fortement le décalage de phase en -1 [7].

Durant l'étude *in vitro* du décalage de phase en -1, nécessaire à l'expression de la sous-unité γ de l'ADN polymérase III de *Escherichia coli*, il était

apparu des intermédiaires transitoires de traduction, dont la taille correspondait à un blocage juste en amont de la tige-boucle. L'auteur avait interprété ce résultat comme reflétant une pause du ribosome au niveau de la tige-boucle [8]. Une bonne corrélation entre la pause du ribosome et le décalage est par ailleurs observée chez l'IBV, bien que quelques structures secondaires provoquent une pause sans induire de décalage [9]. Ainsi, le rôle de la structure secondaire pourrait être de constituer une barrière physique, énergétique, où le ribosome viendrait buter, permettant aux ARNt de se décaler (figure 4). Étant donné la difficulté de prédire la stabilité et la conformation des structures secondaires, des analyses physico-chimiques seront nécessaires pour élucider plus précisément le mécanisme de stimulation du décalage.

Aujourd'hui, aucun élément ne permet de penser qu'un facteur spécifique, cellulaire ou codé par le virus [10], soit capable de moduler le décalage en se fixant sur la structure secondaire. On peut rendre compte de la totalité des résultats expérimentaux par l'intervention des constituants classiques du ribosome ou de l'appareil traductionnel soluble.

Décalage de phase en +1

On connaît moins d'exemples de décalage en +1 que de décalage en -1 mais leurs rôles physiologiques sont plus divers, en particulier par le fait que certains sont retrouvés dans des gènes cellulaires. Le site de décalage est également un heptamère et la stimulation semble aussi dépendre d'une pause du ribosome. Cependant, on observe des variations importantes des mécanismes impliqués. Nous décrivons ici quelques exemples choisis, soit pour leur valeur de modèle, soit pour leur intérêt biologique.

Les éléments Ty constituent une famille de rétrotransposons de *Saccharomyces cerevisiae*, dont l'expression nécessite un décalage de phase en +1. Comme les virus utilisant le décalage de phase en -1, ils possèdent une séquence codant pour une protéine de type GAG suivie, dans la phase +1, d'une séquence homologue à *pol* [11]. Le taux de décalage est de quelques pour cent. Chez le Ty1, le décalage de phase se produit sur le site heptamérique CUU AGG C [12] (figure 5A). Une pause du ribosome permettant le décalage est provoquée par la très faible quantité de l'ARNt_{Arg} qui vient décoder le codon

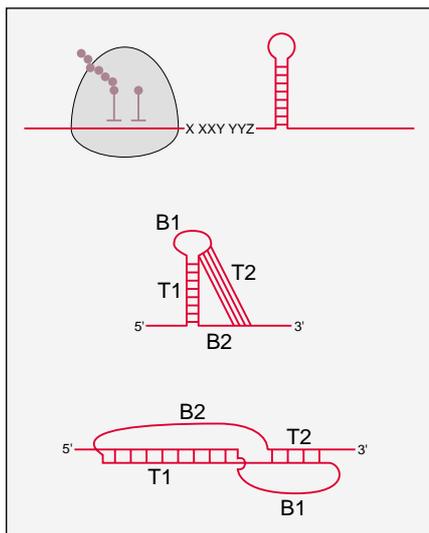


Figure 3. **Élément stimulateur du décalage en -1.** En aval du site de décalage, on trouve une structure secondaire de type tige-boucle (schéma du haut) ou pseudo-nœud (schéma du milieu). Dans le cas du pseudo-nœud, des nucléotides de la boucle sont appariés avec une séquence située en aval. Une autre représentation (schéma du bas) illustre que cette structure correspond à deux tiges empilées (B: boucle et T: tige).

AGG; ainsi, la surexpression de cet ARNt provoque la disparition du décalage [12]. Pendant la pause, l'ARNt_{Leu} situé sur le site P est capable de se réapparier sur le tétranucléotide CUU A en phase +1, grâce à son anticodon ₃GAU₅, qui décode les codons leucine CUU en phase 0 et UUA en phase +1. L'élongation se poursuit par le décodage de GGC en phase +1 par un ARNt_{Gly}. L'abondance de cet ARNt influence également le taux de décalage: sa surexpression augmente le taux de décalage, semblant fixer l'événement mis en route par l'ARNt_{Leu}. Une variante est observée pour le Ty3 où, au vu des appariements possibles, le processus de décalage n'est pas mis en route par le glissement de l'ARNt mais serait dû à sa propension à masquer la base située en 3' (figure 5B). Dans ce cas, seuls certains ARNt sont capables de promouvoir l'arrivée hors de phase du prochain ARNt [13]. Il ne s'agit pas vraiment d'un glissement, puisque des ARNt supposés glissants, tels l'ARNt_{UUU}, ne provo-

quent pas de décalage de phase en +1, alors que des codons que l'on conçoit difficilement glissants le permettent: GCG, CCG, AAU.

Le décalage de phase en +1 n'est pas l'apanage des rétrotransposons, quelques exemples de gènes cellulaires dont l'expression dépend d'un événement de ce type ont été décrits chez la levure et chez les mammifères. Le gène *EST3* avait été identifié génétiquement comme étant impliqué dans le maintien des télomères chez *S. cerevisiae*. D'après les critères du *Yeast Genome Project* [14], ce gène ne contenait pas de phase ouverte de lecture (ORF) de plus de 100 codons et n'était pas référencé comme codant potentiellement pour une protéine. Récemment, Morris *et al.* ont montré qu'*EST3* code en fait pour une protéine de 181 acides aminés, exprimée à partir de deux ORF, séparées par un décalage de phase en +1 [15]. Le mécanisme de ce décalage est très semblable à celui des transposons Ty. Dans le système expérimental utilisé, le décalage se produit avec une efficacité de 80 % à 90 %, mais il n'a pas été évalué dans la situation physiologique. L'élucidation du rôle de ce décalage de phase et de sa régulation nécessite des données complémentaires, mais il pourrait s'agir ici d'un système servant de relais entre l'état physiologique de la cellule, reflété dans l'appareil traductionnel, et le vieillissement cellulaire. On connaît, en revanche, le rôle du recodage dans le contrôle de l'expression de l'ornithine décarboxylase (ODC). L'ODC est une enzyme-clé de la synthèse des polyamines dont la régulation est contrôlée, chez les eucaryotes supérieurs, par un antizyme qui est capable de la cibler vers le protéasome 26S. L'expression de cet antizyme est dépendante d'un décalage de phase en +1 [16]. En l'absence de décalage, un peptide de 68 acides aminés, sans activité biologique connue, est synthétisé. Le taux de décalage est autocontrôlé négativement par le taux de polyamines: plus les polyamines sont abondantes dans la cellule, plus le taux de décalage de phase est élevé, entraînant la synthèse d'antizyme et la dégradation de l'ODCase; la production de polyamines est alors ralentie et le taux de décalage de phase diminue, aboutissant à une diminution de la produc-

tion de l'antizyme (figure 6). Le mécanisme du décalage est de type Ty1, le codon induisant une pause du ribosome est ici le codon stop UGA et c'est un pseudo-nœud, situé 56 nucléotides en aval, qui stimule le décalage [16].

En résumé, pour tous les cas connus où le complexe ribosome/ARNt glisse d'un nucléotide en aval, un ARNt particulier est impliqué sur le site P et le codon présent sur le site A est « affamé » (codon rare ou codon stop). Le site A est, de ce fait, occupé et permet à l'ARNt présent au site P de favoriser l'arrivée d'un ARNt sur le codon non exactement adjacent. Il y a donc compétition entre: (1) l'ARNt censé décoder le codon en phase 0 au site A; et (2) à la fois celui qui est présent au site P et celui qui va décoder le premier codon de la phase +1 au site A.

Translecture d'un codon stop

Un autre type de recodage implique la lecture d'une seule phase interrompue par un codon stop. Le mécanisme est semblable à la suppression génétique classique, sauf qu'il se produit dans un contexte sauvage; on parle alors de translecture (*read-through*) d'un codon stop. Un cas particulier ne sera pas évoqué dans cet article, il s'agit de la redéfinition du codon stop UGA pour spécifier le 21^e acide aminé, la sélénocystéine (pour revue, voir [17]). Nous allons présenter ici quelques exemples choisis pour l'intérêt du mécanisme mis en jeu ou du rôle biologique.

Des études systématiques ont été menées pour comprendre ce qui spécifie exactement un signal de terminaison de traduction et ce qui différencie un stop normal d'un stop soumis à translecture. Une analyse statistique du contexte nucléotidique des codons stop révèle un fort biais sur le nucléotide suivant le stop, variable en fonction des espèces. Par exemple, chez les mammifères, la base +1 est le plus souvent une purine. Ce biais se reflète au niveau fonctionnel, puisque U et C sont associés à une faible efficacité de terminaison [18].

L'efficacité de la translecture étant le résultat de la compétition entre terminaison et poursuite de l'élonga-

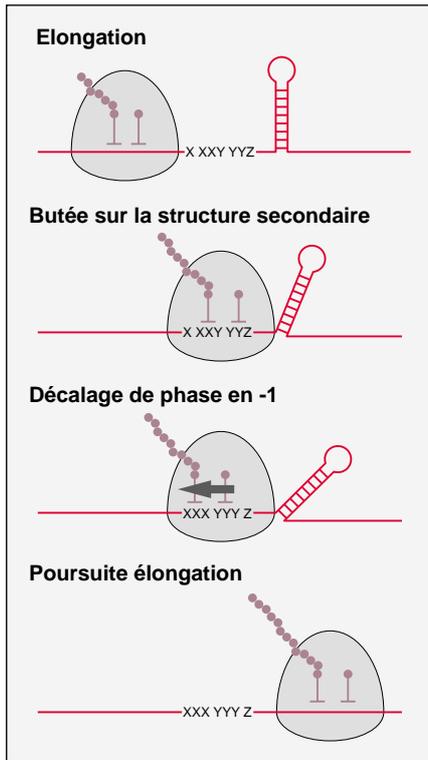


Figure 4. **Mécanisme d'action supposé de la structure secondaire dans la stimulation du décalage en -1.** Le modèle actuel postule que la structure secondaire agirait comme une barrière (physique, énergétique?) à la progression du ribosome. Cette pause dans l'élongation augmenterait la probabilité de décalage sur le site heptamérique glissant.

tion, elle peut faire intervenir un contexte nucléotidique défavorable à la terminaison. C'est le cas chez le virus à ARN Sindbis (famille des alphavirus) chez lequel la polymérase est synthétisée *via* un événement de translecture (10%) obtenu par la simple utilisation d'un contexte défavorable à la terminaison, le signal UAGC [19]. Au cours de son cycle, le virus va se multiplier chez deux types d'hôtes – mammifères et insectes – et il a été montré que cette séquence permet une translecture non seulement *in vitro*, mais aussi dans des cellules d'insectes et de mammifères en culture, et même chez la levure [19]. De nombreuses classes de virus de plante à ARN⁺ utilisent la translecture d'un codon stop, soit pour synthétiser une extension carboxy-terminale d'une protéine de surface, soit pour exprimer une activité cataly-

tique de type polymérase. L'environnement de ces stops est très variable en fonction du type viral, mais semble conservé à l'intérieur d'un même groupe [20]. Malgré des études intensives, en particulier sur le système modèle du virus de la mosaïque du tabac, le mécanisme de cette translecture n'est toujours pas élucidé, tout au plus sait-on que le mécanisme implique une suppression par un ARNt, favorisée par le contexte nucléotidique environnant [21].

Chez une minorité de rétrovirus la synthèse des produits du gène *pol* nécessite une translecture et non un décalage comme cela est présenté précédemment. Ainsi, chez le virus de la leucémie murine de Moloney (MoMuLV), un codon UAG est décodé en une glutamine, ce qui permet l'expression de la protéine de fusion GAG-POL à un taux d'environ 5% *in vitro*. Un pseudo-nœud de 57 nucléotides, situé en 3' du stop, est impliqué dans cette translecture, ainsi que la séquence intercalante de 8 nucléotides non appariés entre le stop et le pseudo-nœud [22]. Le rôle du pseudo-nœud n'est probablement pas relayé ici par une pause, le décodage d'un codon stop entraînant déjà à lui seul un ralentissement (site A vide), mais peut-être par une interaction avec un facteur soluble ou constitutif du ribosome.

Plusieurs exemples de gènes cellulaires soumis à translecture ont été relatés ces dernières années, mais, malheureusement, il n'y a pas eu d'étude approfondie. Un cas beaucoup mieux étayé a été décrit récemment: celui du gène *headcase* (*hdc*) de la drosophile. Le gène *hdc* avait été initialement identifié comme étant requis pour le développement des cellules imaginales [23]. Dans ces premiers travaux, un clone sauvage avait été séquencé et montrait l'existence d'une ORF de 650 codons. Dans un travail récent, Steneberg *et al.*, en recherchant des mutants impliqués dans les activités de branchement de la trachée, ont isolé un allèle de ce même gène *hdc* [24]. Ils ont pu montrer que le codon de terminaison initialement défini est suivi d'une ORF d'environ 400 codons. De façon tout à fait inattendue, celle-ci est nécessaire pour obtenir le sauvetage du phénotype par transgénèse,

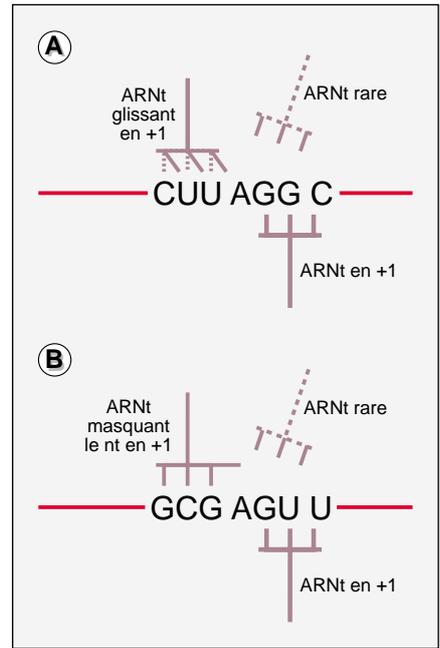


Figure 5. **Mécanismes de décalage en +1.** L'ARNt situé au site P peut soit glisser en +1 (cas A), soit masquer le nucléotide en +1 (cas B). Dans tous les cas, le codon suivant dans la phase 0 correspond à un ARNt rare, ce qui favorise l'arrivée de l'ARNt correspondant à la phase +1.

ce qui suggère très fortement que la lecture du message après le stop est indispensable à la fonction de la protéine sauvage. Le mécanisme du passage de ce stop n'est pas élucidé. Les auteurs ont éliminé l'hypothèse d'un épissage alternatif ou d'une modification (*editing*) de l'ARN, mais, en l'état actuel des données, il pourrait s'agir, soit d'une translecture par suppression, soit d'un saut de ribosome.

Saut de ribosome

Bien que cet article soit consacré aux modèles eucaryotes, on ne peut pas passer sous silence la contorsion la plus étonnante du ribosome, observée dans l'expression du gène *60* du bactériophage T4. Dans ce cas, le ribosome ignore une partie de l'information en se déplaçant par rapport à l'ARNm sans décoder le message correspondant (*ribosome hopping*). Le mécanisme fait intervenir deux sites: un site de décollage (*take off site*) et l'autre d'atterrissage (*landing site*), qui correspondent à deux codons identiques séparés par une

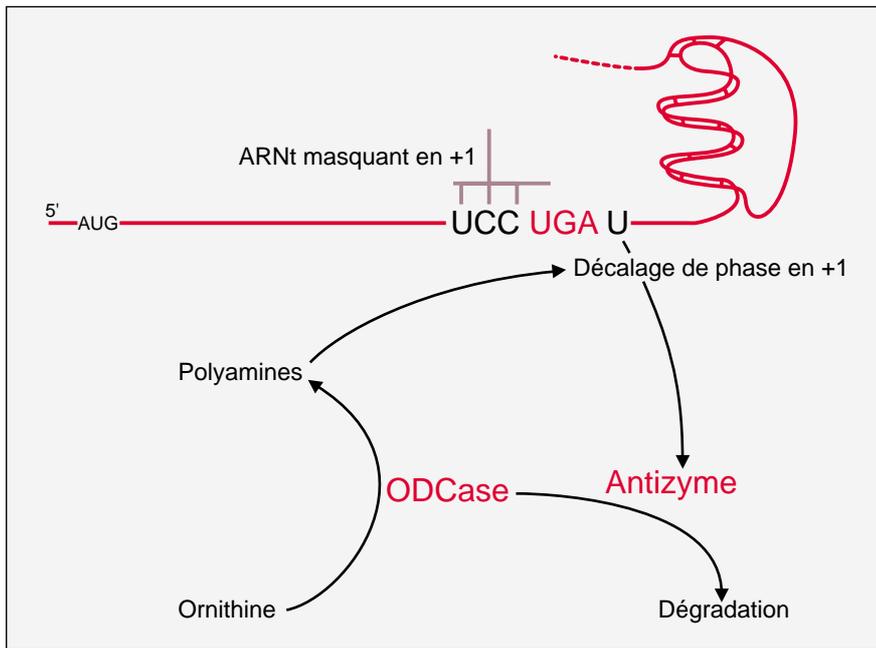


Figure 6. **Boucle de régulation négative de l'antizyme de l'ODC par décalage en +1.** En cas de concentration limitante de polyamines, le décalage n'est pas induit, ce qui entraîne une absence de synthèse d'antizyme. Dans ces conditions, l'ODC n'est pas dégradée et permet la production de polyamines. Lorsque la concentration de polyamines augmente, il y a induction du décalage, production d'antizyme et dégradation de l'ODC, et donc arrêt de la synthèse de polyamines.

distance de 50 nucléotides [25]. Ce saut permet d'éviter un codon stop et de changer de phase avec une efficacité qui peut atteindre 100 %. Quatre

éléments sont critiques dans ce saut de ribosome (figure 7): (1) l'identité des deux codons glycine GGA; (2) la présence d'un stop et d'une tige-

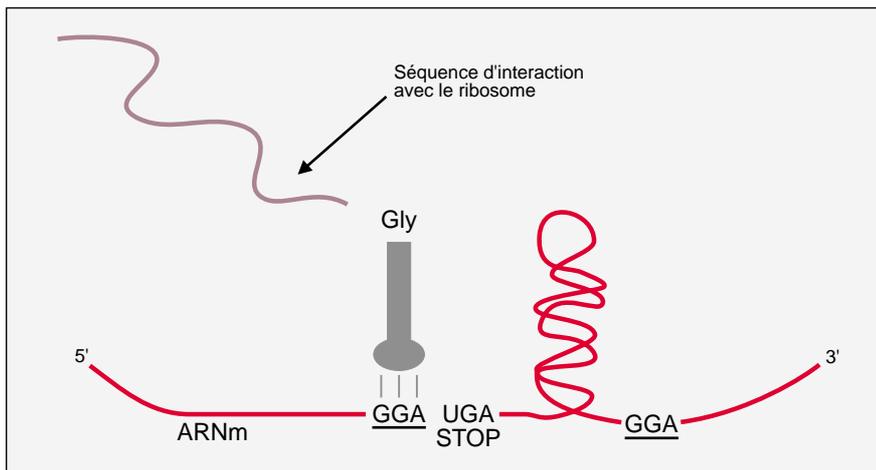


Figure 7. **Saut de ribosome au cours de la traduction du gène 60 du bactériophage T4.** Le ribosome ignore la partie du message comprise entre les deux codons GGA. Quatre éléments stimulateurs interviennent dans ce saut: l'identité des deux codons glycine GGA, la présence du codon stop et d'une tige-boucle dans la région non décodée, la distance optimale de 50 nucléotides et une séquence de 16 acides aminés de la chaîne polypeptidique en croissance.

boucle entre les deux codons; (3) la distance optimale de 50 nucléotides; et (4) la séquence de la chaîne peptidique en croissance (16 acides aminés). Ce dernier point suggère l'existence d'une interaction avec le ribosome, le tunnel de sortie du peptide contenant jusqu'à 30 acides aminés.

Un autre exemple bien documenté d'événements de ce type a été observé chez *Escherichia coli*, aux alentours de certains codons rares [26]. Dans ce cas, les auteurs ont pu montrer qu'après le décollage, le ribosome reste associé à l'ARNm et « glisse » sur celui-ci jusqu'à la rencontre d'un codon d'atterrissage identique à celui de décollage. Plusieurs autres cas de saut de ribosome ont été rapportés chez diverses bactéries, mais n'ont pas fait l'objet d'études approfondies [27].

Conclusions

Même si, par la diversité des sites et des systèmes, le recodage semble peu homogène, deux éléments sont communs à beaucoup de ces événements: (1) un site mécaniquement favorable au recodage; et (2) un élément induisant une pause qui favorise cinétiquement le recodage par rapport au décodage classique. A ce jour, aucun facteur protéique modulant spécifiquement le recodage n'a été isolé, sauf en ce qui concerne le cas particulier de l'incorporation de la sélénocystéine. De plus, on observe une conservation remarquable des mécanismes, de la levure (voire des bactéries pour certains exemples) à l'homme. Ainsi, les événements de recodage pourraient être relayés – outre par les facteurs *cis* – par des caractéristiques intrinsèques très générales de l'appareil traductionnel. Cela permet d'utiliser des eucaryotes simples tels que la levure, comme tube à essai génétique pour étudier des recodages intervenant chez les eucaryotes supérieurs [28]. Quant à la « raison d'être » du recodage, les deux arguments souvent avancés pour son utilisation par les virus et par les rétrotransposons, sont la compacité du signal, critère important pour des organismes à capacité de codage limitée, et le risque de modifier l'ARN lorsque celui-ci sert de génome. En fait, aucun de ces

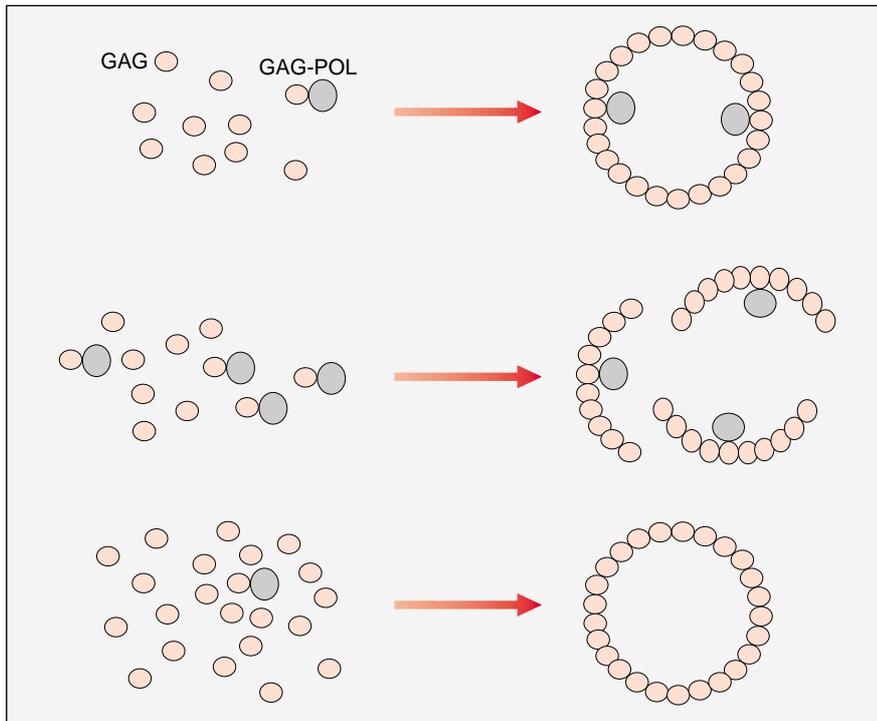


Figure 8. **Effet d'un déséquilibre du rapport GAG-POL/GAG sur la maturation virale.** Une variation du taux de décalage modifie la proportion de protéine de fusion synthétisée. S'il y a trop de protéine GAG-POL par rapport à GAG, il y aura beaucoup moins de virions assemblés correctement. S'il y a trop peu de protéine GAG-POL, il y aura production de virions dépourvus du domaine POL qui seront incapables de déclencher un nouveau cycle de répllication dans les cellules qu'ils auront infectées.

arguments ne semble très convaincant, d'autres solutions éprouvées permettant de surmonter ces écueils. Peut-être sommes-nous ici simplement en face d'un «bricolage» efficace qui, à ce titre, a été conservé au cours de l'évolution. Cette conservation spectaculaire dans les virus à ARN et les éléments transposables pourrait signer un mécanisme ancestral datant du monde à ARN.

Perspectives : recodage et analyse des génomes

L'accumulation de séquences pose le problème du traitement automatisé de leur signification. Ainsi, une séquence soumise à un recodage sera souvent répertoriée dans les banques de données comme séquence non codante ou pseudogène (voir le cas d'EST3 présenté ci-dessus). Des problèmes plus graves risquent même de se poser du fait de la mise en œuvre de programmes qui détectent des décalages de phases entre deux par-

ties codantes d'un même gène [29] et corrigent alors automatiquement ces erreurs supposées. Or une séquence soumise à recodage et une insertion/délétion due à une erreur de séquençage sont des objets informatiques identiques. Il est donc crucial d'élucider les contextes nucléotidiques et les mécanismes mis en jeu dans ces événements, afin de pouvoir mieux prédire leur existence à partir des données de séquences.

Perspectives : recodage et médecine

Le décalage de phase et la translecture étant utilisés par de nombreux virus animaux ou végétaux, le développement de stratégies antivirales agronomiques ou thérapeutiques fondées sur ces mécanismes est envisageable. Ainsi, plusieurs laboratoires sont en train de tenter d'isoler et de caractériser des composés affectant le décalage de cadre en -1 du VIH-1 [30, 31]. Le principe de l'effet de telles molécules est fondé sur des

études – menées sur le virus M1 de la levure – qui ont démontré qu'une faible modification du taux de décalage perturbe gravement la maturation et la propagation du virus [5] (figure 8). L'intérêt potentiel de tels produits est que l'apparition de virus résistants serait probablement difficile, la cible du médicament étant cellulaire. En contrepartie, leur toxicité risquerait d'être élevée. Un composé «antidécage», le RG501, a déjà été identifié par un criblage à grande échelle mené par la compagnie américaine *RiboGene Inc* [31]. Un effet important du RG501 sur la propagation *ex vivo* du VIH-1 a été obtenu. Bien que ce composé lui-même ne soit pas utilisable, pour des raisons non précisées par les auteurs, il est probable que cette nouvelle cible thérapeutique potentielle sera exploitée dans les années qui viennent [32] ■

RÉFÉRENCES

1. Gesteland RF, Weiss RB, Atkins JF. Recoding: reprogrammed genetic decoding. *Science* 1992; 257: 1640-3.
2. Brierley I. Ribosomal frameshifting on viral RNAs. *J Gen Virol* 1995; 76: 1885-92.
3. Jacks T, Madhani HD, Masiarz HD, Varmus HE. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell* 1988; 55: 447-58.
4. Brierley I, Rolley NJ, Jenner AJ, Inglis SC. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 1991; 220: 889-902.
5. Dinman JD, Wickner RB. Ribosomal frameshifting efficiency and *gag/gag-pol* ratio are critical for yeast M1 double-stranded RNA virus propagation. *J Virol* 1992; 66: 3669-76.
6. Bidou L, Stahl G, Grima B, Liu H, Cassan M, Rousset JP. *In vivo* HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. *RNA* 1997; 3: 1153-8.
7. Kollmus H, Honigman A, Panet A, Hauser H. The sequences of and distance between two *cis*-acting signals determine the efficiency of ribosomal frameshifting in human immunodeficiency virus type 1 and human T-cell leukemia virus type II *in vivo*. *J Virol* 1994; 68: 6087-91.
8. Tsuchihashi Z. Translational frameshifting in the *Escherichia coli* dnaX gene *in vitro*. *Nucleic Acids Res* 1991; 19: 2457-62.
9. Somogyi P, Jenner AJ, Brierley I, Inglis SC. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* 1993; 13: 6931-40.

RÉFÉRENCES

10. Cassan M, Delaunay N, Vaquero C, Rousset JP. Translational frameshifting at the gag-pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J Virol* 1994; 68: 1501-8.
11. Farabaugh PJ. Programmed translational frameshifting. *Microbiol Rev* 1996; 60: 103-34.
12. Belcourt MF, Farabaugh PJ. Ribosomal frameshifting in the yeast transposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* 1990; 62: 339-52.
13. Vimaladithan A, Farabaugh PJ. Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol Cell Biol* 1994; 14: 8107-16.
14. Oliver SG, Van der Aart QJ, Agostini-Carbone ML, et al. The complete DNA sequence of yeast chromosome III. *Nature* 1992; 357: 38-46.
15. Morris DK, Lundblad V. Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol* 1997; 7: 969-76.
16. Matsufuji S, manquent 2 auteurs, et al. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase anti-zyme. *Cell* 1995; 80: 51-60.
17. Berry MJ, Larsen PR. Recognition of UGA as a selenocysteine codon in eukaryotes: a review of recent progress. *Biochem Soc Trans* 1993; 21: 827-32.
18. Tate WP, Poole ES, Mannering SA. Hidden infidelities of the translational stop signal. *Prog Nucleic Acid Res Mol Biol* 1996; 52: 293-335.
19. Li G, Rice CM. The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *J Virol* 1993; 67: 5062-7.
20. Miller WA, Dinesh-Kumar SP, Paul CP. Luteovirus Gene Expression. *Crit Rev Plant Sci* 1995; 14: 179-211.
21. Skuzeski JM, Nichols LM, Gesteland RF, Atkins JF. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *J Mol Biol* 1991; 218: 365-73.
22. Wills NM, Gesteland RF, Atkins JF. Pseudoknot-dependent read-through of retroviral gag termination codons: importance of sequences in the spacer and loop 2. *EMBO J* 1994; 13: 4137-44.
23. Weaver TA, White RA. Headcase, an imaginal specific gene required for adult morphogenesis in *Drosophila melanogaster*. *Development* 1995; 121: 4149-60.
24. Steneberg P, Englund C, Kronhamn J, Weaver TA, Samakovlis C. Translational readthrough in the *hdC* mRNA generates a novel branching inhibitor in the *drosophila* trachea. *Genes Dev* 1998; 12: 956-67.
25. Weiss RB, Huang WM, Dunn DM. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* 1990; 62: 117-26.
26. Gallant JA, Lindsley D. Ribosomes can slide over and beyond «hungry» codons, resuming protein chain elongation many nucleotides downstream. *Proc Natl Acad Sci USA* 1998; 95: 13771-6.
27. Gesteland RF, Atkins JF. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* 1996; 65: 741-68.
28. Stahl G, Bidou L, Rousset JP, Cassan M. Versatile vectors to study recoding: conservation of rules between yeast and mammalian cells. *Nucleic Acids Res* 1995; 23: 1557-60.
29. Fichant GA, Quentin Y. A frameshift error detection algorithm for DNA sequencing projects. *Nucleic Acids Res* 1995; 23: 2900-8.
30. Dinman JD, Ruiz-Echevarria MJ, Czaplinski K, Peltz SW. Peptidyl-transferase inhibitors have antiviral properties by altering programmed-1 ribosomal frameshifting efficiencies: development of model systems. *Proc Natl Acad Sci USA* 1997; 94: 6606-11.
31. Hung M, Patel P, Davis S, Green SR. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J Virol* 1998; 72: 4819-24.
32. Dinman JD, Ruiz Echevarria MJ, Peltz SW. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol* 1998; 16: 190-6.

TIRÉS À PART

J.P. Rousset.

Summary

Surprising ways of decoding genetic information

Recoding refers to various translational phenomena in which expression of viral and cellular functions uses programmed modifications of canonical decoding rules. It includes ribosome slipping, either forward or backward (frameshifting), incorporation of amino acids instead of terminating translation (readthrough), and even skipping long segments of a message (ribosome hopping). The role of recoding is crucial for the expression of some important biological functions, like the synthesis of the protease, reverse transcriptase, RNase H and integrase activities encoded by the *pol* gene of retroviruses. In this case, *pol* translation is dependent on a frameshifting or a readthrough event, during upstream *gag* translation. Specific sequence contexts are involved in recoding: slippery sequences and stem-loop/pseudoknot RNA secondary structures for -1 frameshifting, special tRNAs for +1 frameshifting, poor termination context for readthrough. In many cases, a translational pause is needed to stimulate recoding efficiency. Recoding is widely conserved, especially in retrotransposons, suggesting that it appeared early during evolution. Since similar mechanisms are often involved for the same event in different species, recoding should lie on very general properties of the translational apparatus. A better understanding of recoding has two important goals. First, in the genomic field, it will help to identify recoded genes which would otherwise escape the routine procedures developed to identify putative protein coding regions. Second, recoding can potentially be used as a therapeutic target since it is essential for the life cycle of numerous pathogenic viruses, including HIV.

HERACLITEAN BIOLOGY GROUP

Nouveau groupe international de biologistes (France, États-Unis, Canada, Japon) s'intéressant à l'évolution et au développement.

Le site Internet www.heraclitean.com propose un forum de discussion, une liste de sites portant sur des thèmes voisins, une base bibliographique et très bientôt des articles *on line*.