

Université de Montréal

**Étude du mécanisme de décalage de phase de lecture  
en -1 du virus de l'immunodéficience humaine de type 1  
et de son importance dans la réplication du virus.**

par

Dominic Dulude

Département de Biochimie

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Ph.D.  
en Biochimie

Mai 2005

© Dominic Dulude, 2005



W

4

U58

2005

V.134

## AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

## NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

Étude du mécanisme de décalage de phase de lecture en  $-1$  du virus de l'immunodéficience humaine de type 1 et de son importance dans la réplication du virus.

présentée par :

Dominic Dulude

a été évaluée par un jury composé des personnes suivantes :

Dr Gerardo Ferbeyre, président-rapporteur

Dr Léa Brakier-Gingras, directrice de recherche

Dr Luc Desgroseillers, membre du jury

Dr Sherif Abou Elela, examinateur externe

Dr Gerardo Ferbeyre, représentant du doyen de la FES

## Résumé

Le virus de l'immunodéficience humaine de type 1 (VIH-1), l'agent infectieux responsable du syndrome de l'immunodéficience acquise (SIDA), est l'objet d'intenses recherches depuis plus de 20 ans. Les principaux aspects du processus de réplication du virus sont maintenant mieux compris. Chez ce virus, les gènes *gag* et *pol* sont exprimés par traduction du même ARN messager, l'ARN messager viral pleine longueur. Le gène *gag* code pour le précurseur des protéines structurales, la protéine Gag, tandis que le gène *pol* contient l'information pour la synthèse de la protéine Gag-Pol, le précurseur des enzymes du virus, soit la protéase, la transcriptase inverse et l'intégrase. Lors de mes études de doctorat, je me suis intéressé à la synthèse de ces enzymes virales. Cette synthèse implique une erreur de traduction causant un décalage de phase de lecture en  $-1$  (*frameshift* ou déphasage en  $-1$ ) par les ribosomes des cellules infectées lors de la traduction de l'ARN viral pleine longueur. Ce déphasage en  $-1$  est provoqué par des signaux présents sur cet ARN, soit une séquence glissante où se produit le déphasage, suivie d'un signal stimulateur, une structure secondaire particulière de l'ARN viral, qui contrôle l'efficacité du déphasage.

L'ensemble de nos travaux a eu pour objectif d'approfondir notre connaissance du déphasage en  $-1$  du VIH-1, d'abord en caractérisant la structure du signal stimulateur, et, ensuite, en étudiant l'effet de changements de l'efficacité du déphasage sur la réplication du virus. Enfin, nous avons entamé une étude visant à obtenir des peptides capables de perturber l'efficacité du déphasage, ce qui pourrait conduire au développement d'une nouvelle classe d'agents anti-VIH.

Dans le premier volet de nos travaux, nous avons caractérisé la structure du signal stimulateur du décalage de phase de lecture en  $-1$  du VIH-1. Ce signal avait été initialement défini comme une structure en tige-boucle, séparée de la séquence glissante par une séquence séparatrice de sept nucléotides. Nous avons voulu déterminer si la séquence suivant la structure en tige-boucle pouvait influencer l'efficacité du déphasage en  $-1$  en participant à la formation d'un signal stimulateur plus complexe. En utilisant la

mutagenèse dirigée et la détermination de structure d'ARN par attaque enzymatique, nous avons montré que le signal stimulateur est en effet plus complexe qu'une simple tige-boucle. Il consiste en une hélice irrégulière formée de deux portions séparées par une protubérance de purines, la portion inférieure impliquant un appariement entre la séquence séparatrice et une séquence complémentaire en aval de la tige-boucle. Ces travaux ont conduit à notre premier article et nous ont fourni des informations intéressantes pour comprendre comment le signal stimulateur favorise le déphasage en  $-1$  lors de la traduction.

Dans le second volet de nos études sur le déphasage en  $-1$  du VIH-1, nous avons déterminé comment des variations dans l'efficacité de ce déphasage affectent la réplication du VIH-1. Nous avons utilisé la technique de mutagenèse dirigée pour introduire des mutations dans la séquence glissante ou dans le signal stimulateur du déphasage en  $-1$  du virus. L'efficacité de déphasage résultante était diminuée pour tous les mutants étudiés. Nous avons ensuite déterminé le contenu en transcriptase inverse de pseudo-particules virales (VLP) produites par transfection de cellules en culture avec des vecteurs exprimant Gag et Gag-Pol avec les mutations étudiées dans la séquence glissante et dans le signal stimulateur. Enfin, nous avons introduit ces mutations dans un vecteur contenant l'ADN proviral du VIH-1 et nous avons étudié l'effet du déphasage en  $-1$  sur la réplication virale, en utilisant des essais impliquant un seul cycle de réplication et des essais analysant la réplication à long terme du virus. Nos résultats montrent que la réplication du virus est affectée parallèlement à la diminution de l'efficacité du déphasage et que même des changements modestes de l'efficacité du déphasage, comme une diminution de 30%, réduisent significativement l'infectivité et la réplication virale. Ces résultats, qui ont conduit à un deuxième article, démontrent le potentiel du déphasage en  $-1$  du VIH-1 comme cible thérapeutique.

Un troisième volet de notre recherche a consisté à développer une méthode de sélection de peptides capables de perturber l'efficacité du déphasage du VIH-1. Nous avons construit une librairie combinatoire exprimant de courts peptides, riches en arginine, et utilisé un système d'expression bactérien approprié pour sélectionner les

peptides qui influencent le déphasage en -1 du VIH-1, grâce à un gène rapporteur fluorescent dont l'expression dépend de ce déphasage. Nous avons sélectionné par FACS (*fluorescence-activated cell sorting*) des clones parmi la population de bactéries transformées avec la librairie exprimant des peptides répondant à nos critères de sélection, c'est-à-dire ayant une diminution de l'expression du rapporteur dépendant du déphasage. L'analyse de plusieurs clones individuels montre toutefois que ces peptides influencent plutôt la traduction en général et non seulement le déphasage en -1. Les résultats obtenus dans ce troisième volet de notre recherche, quoique préliminaires, soulignent les limites du système de sélection bactérien et imposent des modifications dans notre approche visant le développement d'agents perturbant le déphasage et, par conséquent, la réplication du VIH-1.

**Mots-clés** : virus de l'immunodéficience humaine de type 1 (VIH-1), décalage du cadre de lecture en -1, structure d'ARN, traduction, ribosome, librairie combinatoire.

## Abstract

The human immunodeficiency virus type 1 (HIV-1), the infectious agent responsible for the acquired immunodeficiency syndrome (AIDS), has been the object of an intense research for more than 20 years. The main aspects of the replication process of the virus are known and our understanding of the HIV infection is substantially advanced. In HIV-1, the *gag* and *pol* genes are translated from the same messenger RNA, the full-length viral messenger RNA. The *gag* gene codes for the precursor of structural proteins, Gag, while the *pol* gene contains the information for the synthesis of Gag-Pol, the precursor of the viral enzymes, i.e. the protease, the reverse transcriptase and the integrase. During my Ph.D. studies, I studied the synthesis of these viral enzymes. This synthesis requires a translational error, causing a shift of the reading frame (frameshift) by the ribosomes translating the full length viral RNA in the infected cells. This frameshift depends upon two *cis*-acting signals present on this RNA, a slippery sequence, where the frameshifting occurs, followed by a specific structural motif, the stimulatory signal that promotes frameshifting and controls its efficiency.

The aim of our work was to increase our understanding of the frameshift event occurring in HIV-1, by characterizing the structure of the stimulatory signal, and by studying how changes in the frameshift efficiency affect the virus replication. Finally, we began a study that aimed at obtaining agents that can perturb the frameshift efficiency of HIV-1 and can constitute a novel class of anti-HIV agents.

In the first part of our work, we characterized the structure of the stimulatory signal controlling the  $-1$  frameshift in HIV-1. This signal was initially defined as a stem-loop structure, separated from the slippery site by a spacer sequence of seven nucleotides. We investigated the impact of the sequence downstream of the stem-loop structure on the frameshift efficiency and its possible contribution to the formation of a more complex stimulatory signal. Using a site-directed mutagenesis and an enzymatic structure probing approach, we demonstrated that the stimulatory signal is indeed more complex than a simple stem-loop. It consists of long irregular helix formed by two stems

separated by a purine bulge: the lower stem is formed by pairing the spacer sequence to a pyrimidine-rich complementary sequence downstream from the stem-loop. The article presenting this work provides interesting information that will help us to understand how the stimulatory signal promotes frameshifting.

In the second part of our studies on HIV-1 frameshifting, we determined how variations in the frameshift efficiency affect the replication of HIV-1. Using site-directed mutagenesis, we introduced mutations into the slippery sequence and in the stimulatory signal responsible for the HIV-1 frameshift. We measured the resulting frameshifting efficiency, which was decreased in all the studied mutants. We then determined the reverse transcriptase content of virus-like particles (VLP) produced by transfection of cultured cells with a vector expressing Gag and Gag-Pol, containing the mutations investigated. We also introduced these mutations into a vector containing the proviral DNA of HIV-1 and we studied how changes in the frameshift efficiency affect the viral replication, using single-round assays and long-term replication kinetics. Our results show that the replication of the virus was affected proportionally to the reduction in the frameshift efficiency and that even modest changes of the frameshift efficiency, such as a reduction of 30%, significantly impaired the viral replication. These results, which generated a second article, show the potential of targeting the HIV-1 frameshift event for the development of therapeutic agents.

The third part of our research consisted in developing a method for the selection of peptides that can perturb the frameshift efficiency of HIV-1. We constructed a combinatorial peptide library expressing short randomized arginine-rich peptides that are prone to interact with RNA motifs and, in order to select efficient peptides, we used a bacterial system with a reporter gene whose expression depends on the HIV-1 frameshifting. We used a bacterial system so as to take advantage of the rapid growth of bacteria, knowing that the HIV-1 frameshifting can be recapitulated in bacteria. With this system, we could detect bacterial clones among the population of bacteria transformed with the library, which express peptides that appear to either increase or decrease the frameshift efficiency. These clones can be selected by fluorescence-

activated cell sorting (FACS), and the sequence of the peptides that perturb frameshifting can be determined by sequencing the plasmid DNA from the selected bacteria. The results obtained so far in this third part of our research, although preliminary, showed that the selected peptides affect translation in general and not specifically the frameshift event. The analysis of individual selected clones showed a decrease of about two to three fold in the expression of the reporter genes monitoring the frameshift event and the conventional translation. This stresses the need to improve this system of selection and, clearly, more work has to be done in order to select peptides that can perturb HIV-1 frameshifting and, consequently, alter the replication of the virus.

**Keywords :** Human immunodeficiency virus type 1 (HIV-1), ribosomal frameshift, combinatorial library, ribosome, RNA structure, translation.

## Table des matières

Liste des tableaux.....	xi
Liste des figures .....	xii
Liste des abréviations.....	xx
Chapitre 1.Introduction .....	1
1.1 Le virus de l'immunodéficience humaine de type 1 (VIH-1).....	2
1.1.1 Description du virus .....	2
1.1.1.1 Organisation du génome viral et synthèse du précurseur Gag-Pol .....	3
1.1.1.2 Rôle des protéines virales.....	5
1.1.2 Le cycle de réplication du VIH-1 .....	11
1.1.2.1 Les étapes précoces .....	11
1.1.2.2 Les étapes tardives .....	13
1.1.3 L'hôte.....	15
1.1.3.1 La réponse immunitaire.....	15
1.1.3.2 Les facteurs cellulaires limitant la réplication virale .....	18
1.1.3.3 Les thérapies contre le VIH-1.....	19
1.2 La traduction .....	20
1.2.1 Le ribosome.....	22
1.2.2 Cycle de traduction.....	26
1.2.2.1 Initiation .....	26
1.2.2.2 Élongation .....	31
1.2.2.3 Terminaison.....	35
1.2.2.4 Recyclage du ribosome .....	37
1.2.3 Contrôle de la qualité des ARNm porteurs de codons non-sens (NMD) .....	37
1.3 Reprogrammation de la traduction ( <i>recoding</i> ).....	40
1.3.1 Saut du ribosome ( <i>hopping</i> ) .....	43
1.3.1.1 Exemples de sauts du ribosome.....	43
1.3.1.2 Signaux <i>cis</i> de l'ARN messager favorisant le saut du ribosome.....	45

1.3.2 Redéfinition d'un codon.....	46
1.3.2.1 Translecture ( <i>readthrough</i> ) .....	47
1.3.2.2 Incorporation de la sélénocystéine.....	50
1.3.2.3 Incorporation de la pyrrolysine.....	51
1.3.3 Décalage de phase de lecture ( <i>frameshift</i> ) .....	53
1.3.3.1 Décalage de phase de lecture en +1 .....	54
1.3.4 Décalage de phase de lecture en -1 .....	69
1.3.4.1 Quelques exemples de décalage de phase de lecture en -1 .....	69
1.3.4.2 Signaux <i>cis</i> de l'ARNm favorisant le décalage de phase de lecture en -1... 73	
1.3.4.3 Fonctions du signal stimulateur.....	82
1.3.4.4 Mécanismes et modèles du décalage de phase de lecture en -1 .....	84
1.3.5 Régulation de la réplication virale par la reprogrammation .....	86
1.3.6 État de la question concernant le déphasage en -1 du VIH-1.....	86
1.4 Description du projet de recherche.....	90
Chapitre 2. Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1 .....	93
Chapitre 3. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1.....	118
Chapitre 4. <i>In vivo</i> selection from a combinatorial library of peptides interfering with the HIV-1 programmed -1 ribosomal frameshift .....	145
Chapitre 5. Discussion .....	165
5.1 Caractérisation du signal stimulateur contrôlant l'efficacité du déphasage en -1 du VIH-1.....	1666
5.2 Des mutations dans les signaux contrôlant l'efficacité du déphasage en -1 interfèrent avec la réplication du VIH-1.....	1777
5.3 Mise au point d'un système <i>in vivo</i> de sélection de peptides interférant avec l'efficacité du déphasage en -1 du VIH-1.....	1855
5.4 Conclusions et perspectives.....	1911

Bibliographie.....	194
--------------------	-----

## Liste des tableaux

Tableau 2-1 : Conservation of the proposed structure for the frameshift stimulatory signal among HIV-1 isolates. ....	113
--	-----

## Liste des figures

- Figure 1-1: Représentation schématique de l'organisation du virion, du génome, des transcrits d'ARNm et des protéines du VIH-1..... 4
- Figure 1-2: Étapes générales du cycle de réplication du VIH-1. La phase précoce (partie supérieure du schéma) s'amorce par la reconnaissance du récepteur cellulaire CD4, et se déroule jusqu'à l'intégration de l'ADN proviral, et la phase tardive (partie inférieure du schéma) implique les événements qui ont lieu à partir de la transcription du provirus jusqu'à la relâche et la maturation du virus (adapté de Turner et Summers, 1999). ..... 12
- Figure 1-3: Relation entre l'activation du système immunitaire par le virus, l'apoptose des cellules-T lymphocytaires et la progression de la maladie avant et après thérapie anti-rétrovirale HAART (adapté de Gougeon, 2003). ..... 17
- Figure 1-4: Structure du ribosome 80S de la levure *S. cerevisiae* obtenue par cryo-EM par le groupe de Frank (Beckmann et al., 2001), à une résolution de 15Å. Les protéines et les ARNr de la sous-unité 40S sont colorés en vert et jaune, et les protéines et les ARNr de la sous-unité 60S sont en bleu et orange, respectivement. Un peptidyl-ARNt occupant le site P est également indiqué dans cette structure (adapté de Doudna et Rath, 2002). ..... 23
- Figure 1-5: Exemple des sites hybrides rencontrés par les ARNt sur le ribosome lors de la traduction. Par exemple, lors de la sélection initiale, le complexe ternaire, contenant un aa-ARNt lié à un facteur d'élongation chargé de GTP, se lie au site hybride A/T, tandis que le peptidyl-ARNt occupe le site P/P classique (schéma A). Par contre, après la formation du lien peptidique, il y a un déplacement de ces deux ARNt, le peptidyl-ARNt

- occupe maintenant le site hybride A/P, et l'ARNt déchargé est au site hybride P/E (schéma B)..... 25
- Figure 1-6: Schéma représentant les étapes du mécanisme de l'initiation coiffe-dépendante chez les eucaryotes. La traduction des ARNm chez les eucaryotes implique le recrutement des ARNm par la machinerie de l'initiation traductionnelle avec la sous-unité ribosomique 40S et la formation du ribosome 80S sur cet ARNm. Ce processus est contrôlé par plusieurs protéines nommées facteurs d'initiation eucaryote (eIF) (voir texte) (Holcik et Sonenberg, 2005). ..... 28
- Figure 1-7: Circularisation des ARNm eucaryotes par la machinerie d'initiation de la traduction (Holcik et Sonenberg, 2005)..... 29
- Figure 1-8: Schéma représentant les étapes du mécanisme de l'élongation de la traduction chez les eucaryotes. L'élongation implique l'incorporation d'acides aminés dans la chaîne protéique en croissance correspondant à la séquence définie par les triplets successifs de l'ARNm. Ce processus est contrôlé par des protéines nommées facteurs d'élongation eucaryote (eEF) (adapté de Kapp et Lorsch, 2004). ..... 33
- Figure 1-9: Schéma représentant les étapes du mécanisme de la terminaison de la traduction chez les eucaryotes. La terminaison permet la relâche de la chaîne protéique en croissance suite au décodage d'un codon de terminaison par le ribosome. Ce processus est assuré par des protéines nommées facteurs de terminaison eucaryote (eRF) (adapté de Kapp et Lorsch, 2004). ..... 36
- Figure 1-10: Modèle de la dégradation des ARNm porteurs d'un codon non-sens (NMD). (a) Mécanisme de contrôle de la qualité des ARNm chez les eucaryotes. (b) La présence d'un codon de terminaison suffisamment en amont de la dernière jonction exon-exon déclenche le NMD et la dégradation du messenger (adapté de Hillman et al., 2004). ..... 39

Figure 1-11: Rôles fonctionnels de la reprogrammation. (A) Contrôle de la synthèse d'une protéine par déphasage ou translecture. La reprogrammation peut se produire au début du processus d'élongation et permettre de compléter la synthèse de la protéine (le bleu indique la portion du messenger avant la reprogrammation). (B) Contrôle de la synthèse de deux protéines dans un rapport précis. La protéine ORFA et la protéine de fusion ORFA-ORFB, sont produites par déphasage ou translecture. Le déphasage permet la production d'une protéine plus longue, ORFA-ORFB (a), ou plus courte, ORFA (b), que la protéine résultant de la traduction conventionnelle. La translecture est nécessaire à la production d'un produit plus long, ORFA-ORFB (c). (C) Incorporation d'un acide aminé particulier. SECIS indique le signal permettant l'incorporation de la sélénocystéine à un codon de terminaison (jaune). PYLIS indique le signal hypothétique permettant l'incorporation de la pyrrolysine à un codon de terminaison (jaune). (adapté de Baranov et al., 2002a)..... 42

Figure 1-12: Modèle du saut du ribosome lors de la traduction du messenger du gène *60* du bactériophage T4. Lors de la traduction, le peptidyl-ARNt<sup>Gly</sup> se détache du codon GGA (décollage), parcourt l'ARNm à la recherche d'une séquence complémentaire (balayage), et se ré-apparie au site GGA pour reprendre l'élongation (atterrissage). Plusieurs signaux sont requis dans l'ARNm pour réaliser ce saut (adapté de Herr et al., 2000).... 44

Figure 1-13: Mécanisme d'incorporation de la sélénocystéine. A) Chez les bactéries, une tige-boucle (séquence bSECIS) situé immédiatement en aval du codon de terminaison UGA interagit avec le facteur SelB qui a lié un ARNt<sup>Sec</sup><sub>UCA</sub> chargé de sélénocystéine. B) Chez les eucaryotes, l'élément SECIS est situé dans la région 3'UTR et permet l'incorporation de la sélénocystéine à plusieurs codon de terminaison UGA d'un messenger. Un facteur additionnel est requis chez les eucaryotes, SBP2, qui se lie

d'une part à la structure SECIS et interagit d'autre part avec eSelB portant l'ARNt<sup>Sec</sup><sub>UCA</sub> chargé de la sélénocystéine (adapté de Namy et al., 2004). ..... 52

Figure 1-14: Description du décalage de phase de lecture en +1 du ribosome sur le messenger codant pour le facteur de terminaison RF2. Le déphasage est causé par une pause du ribosome en élongation lors de la rencontre du codon stop UGA au site A. Le peptidyl-ARNt au site P peut alors se désappairer de l'ARNm, se déplacer d'une base du côté 3' avec le ribosome, puis se ré-appairer à l'ARNm dans le cadre de lecture +1. Le complexe ternaire (EF-Tu\*GTP\*Asp-ARNt<sup>Asp</sup><sub>CUG</sub>) amène ensuite l'aa-ARNt correspondant au nouveau codon au site A. .... 55

Figure 1-15: Modèle de liaison avec occlusion de l'aa-ARNt permettant le décalage de phase de lecture en +1 du ribosome lors de la traduction du rétrotransposon Ty3. Selon le modèle proposé par le groupe de Farabaugh (Farabaugh et al., 1993; Li et al., 2001; Stahl et al., 2001b), le déphasage au site de Ty3 serait causé par l'incorporation de l'aa-ARNt<sup>Val</sup><sub>CAI</sub> directement dans le cadre de lecture +1 au site A du ribosome, sans qu'il y ait glissement du peptidyl-ARNt au site P. L'aa-ARNt<sup>Val</sup><sub>CAI</sub> est introduit au site A, et s'apparie au codon GUU du cadre +1, sans qu'il y ait lecture de l'A précédant GUU. .... 59

Figure 1-16: Exemples de signaux stimulateurs du décalage de phase de lecture en +1. A. Le déphasage en +1 lors de la traduction de l'ARNm du gène *prfB* est stimulé par un appariement entre une séquence de type Shine-Dalgarno (SD), située trois nucléotides en amont du site de glissement, et une séquence complémentaire anti-SD de l'ARNr 16S du ribosome. B. Au site de déphasage des messagers des gènes de l'antizyme, l'efficacité du déphasage en +1 est stimulée par la présence d'une structure secondaire en pseudonoeud, située trois nucléotides en aval du site de glissement. .... 61

- Figure 1-17: Modèles du décalage de phase de lecture en +1. Le modèle de Baranov et al. (2004) suggère que le repositionnement du peptidyl-ARNt au site P se produit durant les étapes du cycle d'élongation où l'aa-ARNt peut être rejeté (c'est-à-dire soit durant la sélection initiale de l'aa-ARNt au site A/T ou durant l'étape de *proofreading*), avant l'occupation du site A. Le déphasage est dicté par le repositionnement du peptidyl-ARNt au site P lorsque le site A est inoccupé si le réappariement dans le nouveau cadre est possible (adapté de Baranov et al., 2004). ..... 65
- Figure 1-18: Mécanisme par lequel les ribosomes empêchent la liaison avec occlusion d'un aa-ARNt (dans le cadre +1). A. Détails des sites A et P du centre de décodage du ribosome. L'ARNt au site A (jaune) et P (bleu). B. Agrandissement de l'interaction codon-anticodon au centre de décodage montré en A. Les nucléotides A1492 et A1493 vérifient l'interaction codon-anticodon du messenger et de l'ARNt. C. Illustration de la liaison d'un ARNt avec occlusion au site A. La liaison avec occlusion (+1) de l'ARNt ne permet pas une reconnaissance normale de l'ARNt par les nucléotides A1492 et A1493. (les section A et B présentent la structure cristallographique correspondante du ribosomes bactérien par diffraction aux rayons X) (adapté de Hansen et al., 2003)... 68
- Figure 1-19: Représentation du décalage de phase de lecture en -1 du ribosome. Le déphasage est causé par une pause du ribosome en élongation. La pause se fait au niveau de la séquence glissante et est causée par la présence d'un signal stimulateur situé quelques nucléotides en aval. À ce moment, le peptidyl-ARNt au site P et l'aa-ARNt au site A se détachent de l'ARNm, le ribosome portant les deux ARNt se déplace d'un nucléotide en direction 5' par rapport au messenger, puis les deux ARNt se ré-apparient avec l'ARNm dans le nouveau cadre de lecture -1. .... 70
- Figure 1-20: Signaux stimulateurs en pseudonoeud présents aux sites de déphasage en -1 de plusieurs virus. A. Le pseudonoeud est une structure

héliçoïdale quasi-continue, où les boucles L1 et L2 en simple brin traversent les sillons majeurs et mineurs des tiges en hélice S2 et S1, respectivement. Pseudonoeuds trouvés dans l'ARN des virus IBV (B), MMTV (C), SRV-1 (D) et BWYV (E).....	77
Figure 1-21: Signaux stimulateurs en tiges-boucles présents aux sites de déphasage en -1 de plusieurs virus tels HAst, HTLV-II, VIH-I, CfMV et RCNMV (adapté de Brierley et Pennell, 2001). .....	80
Figure 1-22: Structure à haute résolution d'une portion du signal stimulateur du VIH-1. A. La conformation d'une portion de la tige-boucle (encadré à gauche) a été déterminé par RMN (à droite). Les nucléotides sont colorés selon : A est rouge, C est bleu, G est vert et U est jaune. B. Conformation de la boucle ACAA coiffant la tige-boucle du signal stimulateur. L'orientation des nucléotides de la boucle suggère qu'ils peuvent interagir avec des partenaires tel le ribosome (adapté de Staple et Butcher, 2003). .....	81
Figure 2-1: Description of pHIV90-luc, a frameshift reporter construct. ....	114
Figure 2-2: Effect on frameshifting of mutations in the HIV-1 frameshift region.....	115
Figure 2-3: Novel structure proposed for the frameshift stimulatory signal of HIV-1. (a) Structure probing of the frameshift stimulatory signal by RNase V1 attack.....	116
Figure 2-4: Effect on frameshifting of mutations in the lower stem of the proposed frameshift stimulatory signal. ....	117
Figure 3-1: Frameshift region of wild-type HIV-1 and its mutants.....	139
Figure 3-2: Measurement of the frameshift efficiency of HIV-1 derivatives with a dual- luciferase system. ....	14040
Figure 3-3: Effect of changes in frameshift efficiency on the release of VLPs, on the RT content of the released VLPs, and on the ratio of Gag-Pol to Gag in these VLPs.. .....	141

Figure 3-4: Impact of changes in frameshift efficiency on the processing of Gag within the released VLPs.....	142
Figure 3-5: Single-round infectivity assays of HIV-1 and its derivatives mutated in the frameshift region..	1433
Figure 3-6: Replication kinetics of HIV-1 and mutants with an altered frameshift efficiency.....	144
Figure 4-1: A) HIV-1 Frameshift region. The underlined slippery site is followed by a frameshift stimulatory signal consisting of a long irregular stem-loop, where a purine bulge separates two stems (Dulude et al., 2002). B) The two-plasmid system used for selecting peptides that perturb the frameshift efficiency.....	162
Figure 4-2: FACS fluorescence patterns obtained with <i>E. coli</i> BL21 cells transformed with different fluorescent reporter constructs, containing either GFP or RFP or both.....	163
Figure 4-3: FACS fluorescence patterns of <i>E. coli</i> BL21 cells transformed with the reporter plasmid and the plasmid expressing the peptide library.....	164
Figure 5-1: Hélices triples. A. Structure secondaire du signal stimulateur du déphasage du VIH-1 tel que proposé par Dinman et al. (2002). B. Caractéristique des hélices triples impliquant un brin riche en pyrimidines. L'orientation antiparallèle des brins riches en pyrimidines est défavorable (Y est une pyrimidine, R une purine et Y' est le brin formant l'hélice triple).....	1744
Figure 5-2: Structures secondaires du signal stimulateur des sous-types du groupe M du VIH-1. Le signal stimulateur est replié selon le modèle identifié chez le sous-type B. Les cercles indiquent les nucléotides qui diffèrent de la séquence du sous-type B, et les flèches indiquent les appariements non-canoniques et les mésappariements (adapté de Baril et al., 2003a)..	175

Figure 5-3: Structure à haute résolution du signal stimulateur du VIH-1. La conformation du signal stimulateur tel que proposé par Dulude et al. (2002) (à gauche) a été confirmée par RMN (à droite) par le groupe de Butcher et le groupe de Fourmy (Gaudin et al., 2005; Staple et Butcher, 2005) (adapté de Gaudin et al., 2005)..... 1766

## Liste des abréviations

A	Adénosine
aa-ARNt	aminoacyl-ARNt
ADN	acide désoxyribonucléique
ARN	acide ribonucléique
ARNc	ARN complémentaire
ARNdb	ARN double-brin
ARNm	ARN messenger
ARNr	ARN ribosomique
ARNsb	ARN simple-brin
ARNt	ARN de transfert
AZ	antizyme
BDYV	virus de la jaunisse de l'orge nain ( <i>barley dwarf yellow virus</i> )
BWYV	virus de la jaunisse de la betterave de l'Ouest ( <i>beet western yellow virus</i> )
C	cytosine
CA	capside
CfMV	<i>cocksfoot mottle virus</i>
CMH-1	complexe majeur d'histocompatibilité de classe 1
CTL	lymphocytes-T cytotoxiques
cyp A	cyclophiline A
cryo-EM	cryo-microscopie électronique
EAV	virus de l'anémie équine ( <i>equine anemia virus</i> )
EIAV	virus de l'anémie infectieuse équine ( <i>equine infectious anemia virus</i> )
EJC	complexe de jonction exon-exon
Env	protéine de l'enveloppe

FACS	tri de cellules basé sur l'émission de fluorescence (Fluorescence-activated cell sorting)
FIV	virus de l'immunodéficience du félin
G	guanosine
GFP	protéine fluorescente verte ( <i>green fluorescent protein</i> )
HAART	thérapie antirétrovirale hautement active
HAstV	astrovirus humain ( <i>human astrovirus</i> )
IS	éléments d'insertion
LTR	<i>long terminal repeats</i>
MA	protéine matrice
MHV	virus de l'hépatite murine
MMTV	virus de tumeur mammaire de la souris ( <i>Mouse mammary tumor virus</i> )
MoMuLV	virus de la leucémie murine Moloney
MuLV	virus de la leucémie murine
NC	protéine de la nucléocapside
NMD	dégradation d'ARNm porteur de codon non-sens
NNRTI	inhibiteur non-nucléosidique de la RT
NRTI	inhibiteur nucléosidique de la RT
ODC	ornithine décarboxylase
ORF	cadre de lecture ( <i>open reading frame</i> )
PABP	protéine liant la queue de poly-A ( <i>polyA binding protein</i> )
PBS	site de liaison de l'amorce ( <i>primer binding site</i> )
PIC	complexe de pré-intégration
PR	protéase
pré-ARNm	précurseur d'ARNm
PTAP	domaine riche en proline
PTC	codon non-sens prématuré
PYLIS	séquence d'insertion de la pyrrolysine

R	purine
RCNMV	virus de la mosaïque nécrotique du trèfle rouge ( <i>red clover necrotic mosaic virus</i> )
RFP	protéine fluorescente rouge
RMN	résonance magnétique nucléaire
RRF	facteur de recyclage du ribosome
RRL	lysate de réticulocytes de lapin
RSV	virus du sarcome de Rous
RT	transcriptase inverse
SD	séquence Shine-Dalgarno
Sec	sélénocystéine
SECIS	séquence d'incorporation de la sélénocystéine
SIDA	syndrome de l'immunodéficience acquise
SIV	virus de l'immunodéficience du singe
SRV	rétrovirus du singe
SU	protéine de la surface
TAR	élément structural de liaison de la protéine Tat ( <i>trans-activator response element</i> )
TFP	protéine <i>transframe</i>
TM	protéine transmembranaire
TMV	virus de la mosaïque du tabac
TRV	<i>tobacco rattle virus</i>
U	uracile
UTR	région non-traduite
VIH-1	virus de l'immunodéficience humaine de type 1
VLP	pseudoparticule virale
Y	pyrimidine

## Remerciements

Je tiens tout particulièrement à remercier ma directrice de recherche, Dr Léa Brakier-Gingras, pour son support constant, son encadrement exemplaire, son ouverture vis-à-vis de mes choix dans l'accomplissement de mes travaux de recherches, ainsi que pour les nombreuses discussions au fil des années. J'ai été très chanceux de faire partie de votre laboratoire. Je veux aussi remercier Nikolaus Heveker pour son excellente collaboration.

Je remercie également les gens que j'ai côtoyés au laboratoire, Martin Baril, François Bélanger, Mélissa Léger, Karine Gendron, Sacha Sidani, Marie-Noëlle Brunelle et Francis Robert. L'ambiance au laboratoire était vraiment très agréable, et j'ai eu beaucoup de plaisir à travailler avec vous. Je tiens aussi à remercier Catherine Spickler et Catherine Payant qui m'ont initié à la recherche lorsque je suis arrivé au laboratoire. Je remercie plus particulièrement mon bon ami Martin Baril, avec qui les discussions tant sur la science que sur l'actualité, les sports ou la Bourse ont toujours été très plaisantes. Je tiens aussi à remercier tous les gens du Département de Biochimie pour leur appui.

Je tiens également à remercier ma famille qui m'a toujours soutenu dans mes études. L'éducation a toujours été une valeur très importante pour ma mère Danielle et mon père Gaétan. Je sais que tu veilles sur moi de là-haut, Maman, et que tu fais partie de ma réussite. Merci aussi à ma sœur Catherine, qui m'a toujours aidé et supporté. Merci à vous, pour votre soutien, votre attention et votre intérêt dans ma carrière.

Merci également à tous mes amis à l'extérieur du laboratoire. Je remercie particulièrement David, Éric et Stéphane, qui se demandent toujours pourquoi un chercheur est si souvent occupé au laboratoire... même les fins de semaine! Merci aussi à Patricia, pour ton support ainsi que pour les belles promenades en plein air. Je suis bien content de vous connaître. Merci à tous!

À ma mère Danielle, mon  
père Gaétan, ma soeur  
Catherine et moi.

# Chapitre 1

## Introduction

Le virus de l'immunodéficience humaine de type 1 (VIH-1), l'agent infectieux responsable du syndrome de l'immunodéficience acquise (SIDA), est l'objet d'intenses recherches depuis plus de 20 ans. En effet, le SIDA est l'une des maladies infectieuses les plus terribles de l'histoire de l'humanité et cette maladie a un impact majeur sur la société et les systèmes de santé. Le VIH-1 a déjà infecté plus de 60 millions de personnes, dont plus de 20 millions sont décédées depuis son apparition au début des années 80 (UNAIDS, 2002). Originaire d'une infection inter-espèce provenant d'un chimpanzé d'Afrique et transmise à l'homme au début du 20<sup>ième</sup> siècle (Hahn et al., 2000), le VIH est aujourd'hui présent partout sur le globe. L'Afrique demeure toutefois la région la plus touchée puisque près de 70% des personnes affectées à travers le monde y résident.

Depuis l'identification du VIH-1 comme la cause du SIDA par le groupe de Montagnier en 1983 (Barre-Sinoussi et al., 1983), plus de vingt années de recherches par la communauté scientifique ont produit plus de 125 000 publications. Nous avons maintenant une très bonne compréhension du processus de réplication du virus et nous savons comment il cause la maladie chez l'homme. Dans sa définition la plus simple, le VIH-1 peut être considéré comme un virus contenant dix gènes qui produisent 16 protéines. L'analyse de la variation génétique des séquences virales a permis de séparer ces séquences en trois groupes, M, N et O. Le groupe M comprend plus de 99% des cas répandus mondialement et est divisé en neuf sous-types (A, B, C, D, F, F, G, H, K et K). Malgré les progrès dans les recherches pour une thérapie contre le SIDA et la mise au point de 17 médicaments actuellement utilisés en clinique, aucun traitement, même prolongé, ne permet d'éliminer le virus chez les personnes infectées car il se dissimule dans des réservoirs où il est inaccessible. L'émergence de variants viraux résistants contrecarre également l'effet des médicaments utilisés (Palella et al., 1998). De plus, les coûts élevés de ces médicaments les rendent inaccessibles aux habitants des pays sous-développés, là où le virus est le plus

présent. La recherche de nouvelles cibles et le développement de nouveaux médicaments sont donc des objectifs prioritaires dans la lutte contre le SIDA.

Lors de mes études de doctorat, je me suis intéressé à une étape cruciale du cycle de réplication du VIH-1, la synthèse des enzymes virales. Cette étape implique une erreur de traduction causant un décalage de phase de lecture en  $-1$  (*frameshift* ou déphasage en  $-1$ ) par les ribosomes lors de la traduction de l'ARN messager (ARNm) viral. Ce déphasage en  $-1$  est provoqué par des signaux présents sur cet ARNm et permet au virus de produire une polyprotéine, Gag-Pol, contenant les enzymes nécessaires à sa réplication.

Dans cette introduction, nous allons explorer les récentes avancées concernant le VIH-1, principalement ses caractéristiques, son mode de réplication et les stratégies qu'il utilise pour court-circuiter les mécanismes de défense de l'hôte. Les thérapies développées actuellement contre ce virus seront aussi abordées. Ensuite, nous examinerons les mécanismes de la traduction chez les eucaryotes, ainsi que les modes non-conventionnels de traduction permettant la reprogrammation de la synthèse protéique de nombreux gènes, parmi lesquels le déphasage en  $-1$  qui nous intéresse particulièrement.

## **1.1 Le virus de l'immunodéficience humaine de type 1 (VIH-1)**

### **1.1.1 Description du virus**

Les images de microscopie électronique nous montrent que le virion du VIH-1 est d'allure sphérique avec un diamètre d'environ 100 nm (voir Figure 1-1A). La partie centrale du virion, la capsid, est formée par les protéines de la capsid (CA) entourant les protéines de la nucléocapsid (NC) liées à deux copies d'ARN d'environ 10 kb qui constituent le génome viral (revue par Ferguson et al., 2002; Frankel et Young, 1998). Il y a également à l'intérieur de la capsid du virion des enzymes virales impliquées dans la réplication du virus, soit la protéase (PR), la transcriptase inverse (RT), l'intégrase (IN), la protéine accessoire Vpr ainsi que l'ARNt<sub>3</sub><sup>Lys</sup>, provenant des cellules infectées et servant d'amorce pour la réaction de transcription inverse (voir section 1.1.2.1). La capsid

d'apparence conique du VIH-1 est entourée d'une membrane lipidique à double couche acquise lors de son bourgeonnement à partir de la cellule infectée. Cette couche membranaire est traversée par les glycoprotéines virales transmembranaires (TM) liées aux protéines de surface (SU). La face interne de cette membrane est tapissée de plusieurs exemplaires de la protéine matrice (MA).

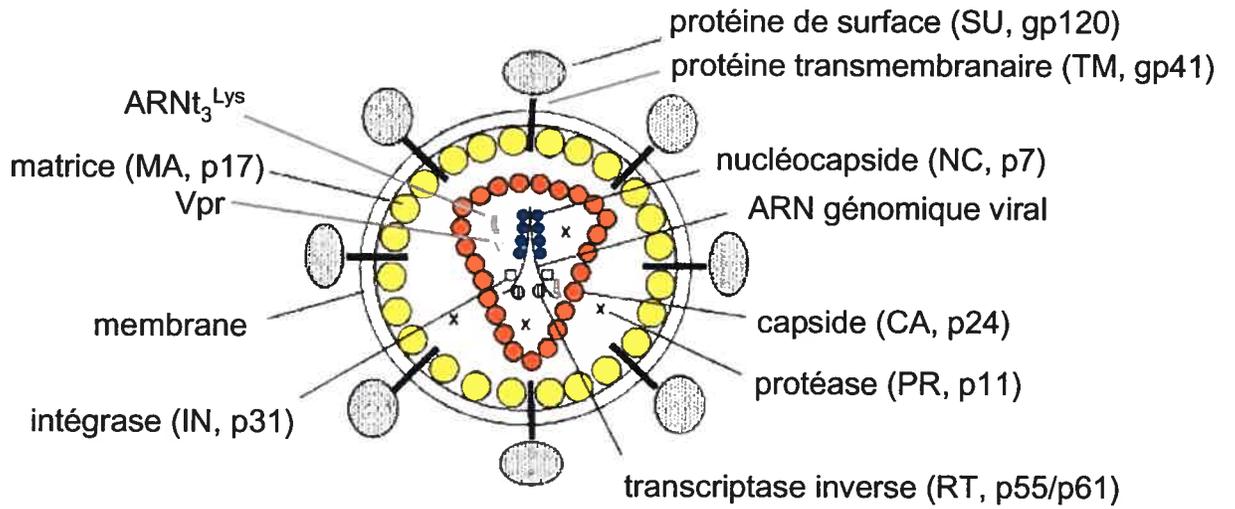
#### 1.1.1.1 Organisation du génome viral et synthèse du précurseur Gag-Pol

Dès l'obtention d'un clone moléculaire contenant le génome du VIH-1 par le groupe de Montagnier (Barre-Sinoussi et al., 1983), les scientifiques ont pu amorcer différentes études permettant la caractérisation de ce virus. Nous savons maintenant que le génome du VIH-1 est une molécule d'ARN d'environ 10 kb contenant dix cadres de lecture (*ORF*) codant pour 16 protéines (Figure 1-1B) (revue par Peterlin et Trono, 2003). Une des particularités du VIH-1, un lentivirus faisant partie de la famille des rétrovirus, est sa capacité à convertir son génome d'ARN simple-brin (ARNsb) en ADN complémentaire (ADNc) grâce à sa transcriptase inverse. Cet ADNc est ensuite inséré dans le génome de l'hôte grâce à l'intégrase du virus, où il forme le provirus et sert de matrice pour générer les dix transcrits en utilisant l'ARN polymérase II de l'hôte (Figure 1-1C). Les promoteurs permettant la transcription des messagers sont localisés dans les séquences LTR (*long terminal repeats*), situés aux extrémités du génome viral. Le VIH-1 contient trois gènes communs à tous les rétrovirus : les gènes *gag*, *pol* et *env*. La séquence des gènes *gag* et *pol* est présente sur le même transcrit, soit l'ARN viral pleine longueur non-épissé. Deux autres classes de transcrits sont également produites à partir du génome viral. Les transcrits mono-épissés contiennent la séquence codante des gènes *vif*, *vpr*, *vpu* et *env*, et les transcrits multi-épissés, celles des gènes *tat*, *rev*, *nef* et *tev*. Une particularité de l'organisation du génome du VIH-1 est le chevauchement du début du gène *pol* avec la fin du gène *gag*. En effet, le gène *pol* est exprimé dans le cadre de lecture -1 par rapport au cadre de lecture de *gag*. La protéine Pol est en fait exprimée sous forme d'une fusion Gag-Pol dans laquelle le début de la protéine est identique à Gag. Sa synthèse requiert un déphasage en -1 des ribosomes lors de la traduction du messenger viral pleine longueur. Les ribosomes changent

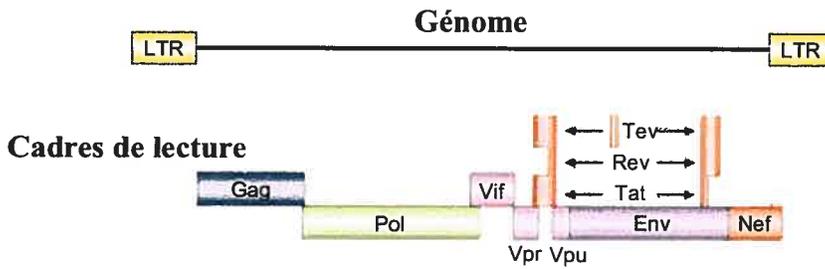
Figure 1-1: Représentation schématique de l'organisation du virion, du génome, des transcrits d'ARNm et des protéines du VIH-1.

- A. Localisation des protéines structurales provenant du gène *gag*, des glycoprotéines du gène *env* et des enzymes virales codées par *pol* dans le virion. Le poids moléculaire en kDa est indiqué à côté du nom de chaque protéine (adapté de Freed, 1998).
- B. Représentation des dix cadres de lectures du génome viral codant pour les protéines virales. Le génome est encadré par deux séquences identiques, les *long terminal repeats* ou LTR (adapté de Peterlin et Trono, 2003).
- C. Représentation des trois classes de transcrits viraux (adapté de Peterlin et Trono, 2003).
- D. Mode de synthèse des protéines Gag et Gag-Pol du VIH-1.
- E. Maturation des précurseurs protéiques Gag, Gag-Pol et Env du VIH-1 (voir détails à la section 1.1.1.2) (adapté de Peterlin et Trono, 2003).

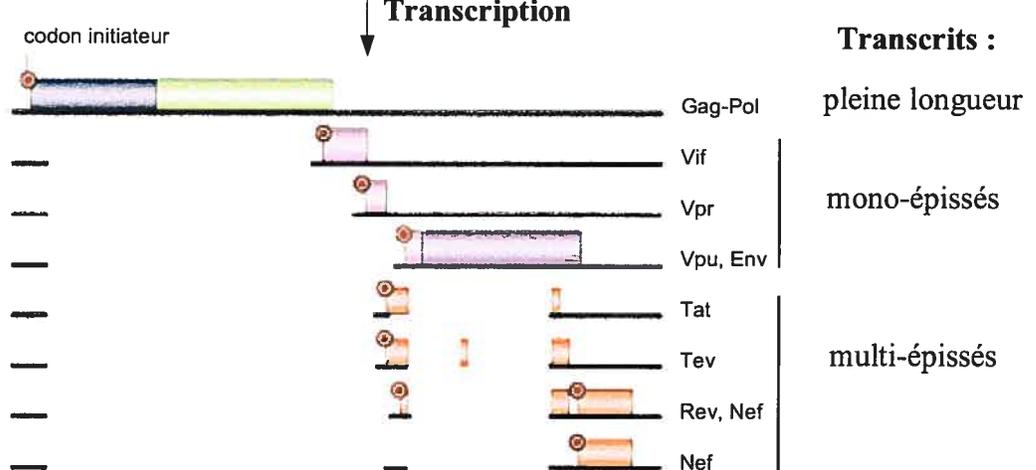
**A**



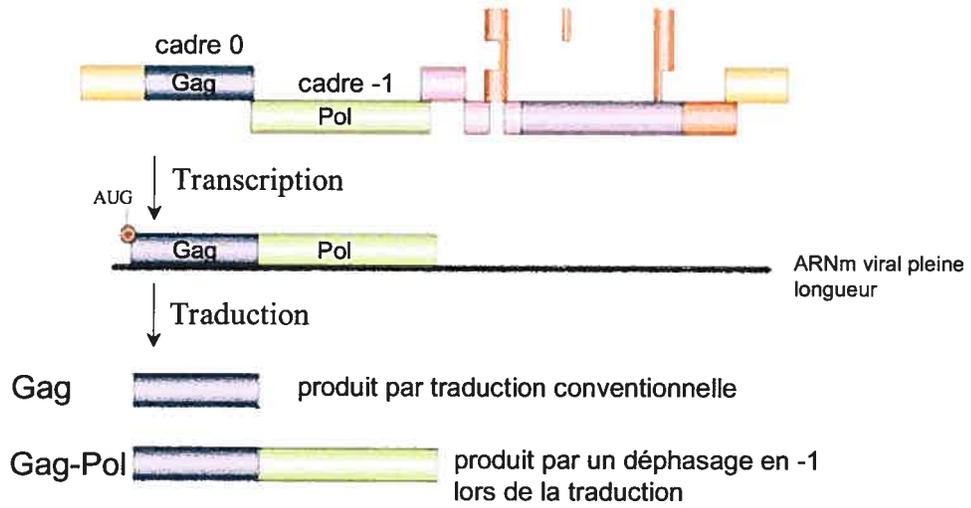
**B**



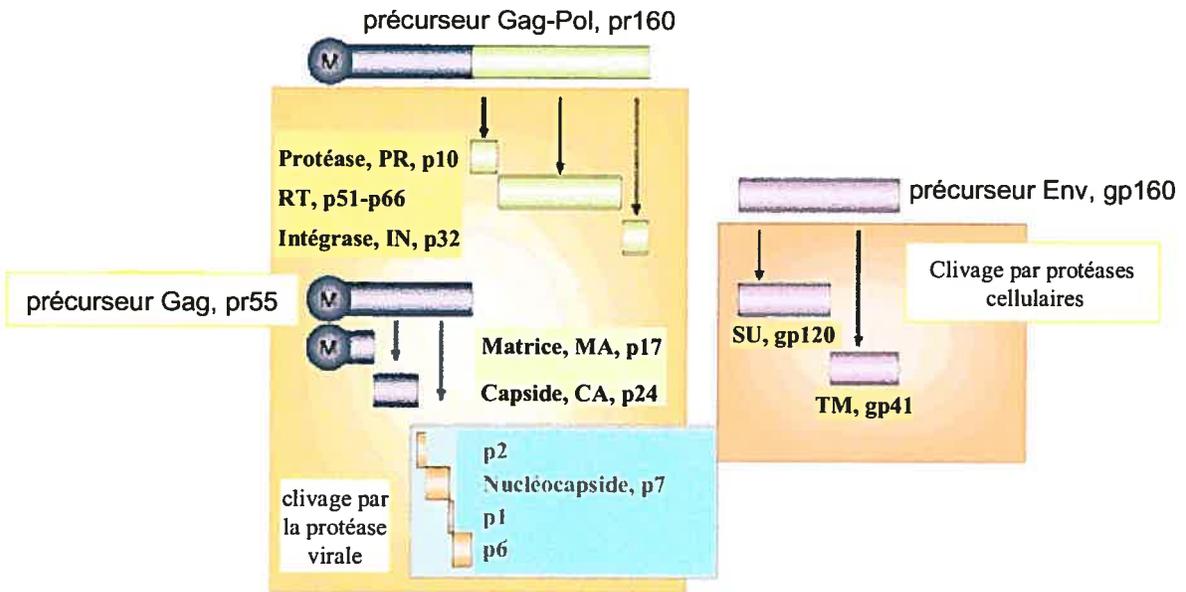
**C**



**D**



**E**



de cadre de lecture et ne reconnaissent pas alors le codon stop de Gag. Le déphasage implique une faible proportion des ribosomes et conduit à la production de Gag-Pol. La majorité des ribosomes traduisant le messenger viral suivent les règles conventionnelles et produisent Gag (Figure 1-1D).

### 1.1.1.2 Rôle des protéines virales

Le produit du gène *gag* est un précurseur polyprotéique de 55 kDa (Gag, pr55<sup>Gag</sup> ou p55<sup>Gag</sup>) codant pour les protéines de structure du virus. De l'extrémité N-terminale à l'extrémité C-terminale le précurseur Gag contient les protéines suivantes : la matrice (MA, p17), la capsid (CA, p24), le peptide p2, la nucléocapsid (NC, p7), le peptide p1 et p6 (Figure 1-1E). La forme mature de ces protéines, c'est-à-dire leur forme libre, est obtenue suite au clivage de Gag par la protéase virale. Les protéines du précurseur Gag ont une fonction dans l'assemblage et la formation de virions. MA, grâce à l'ajout d'un groupement myristoylé à son extrémité N-terminale et la présence de résidus basiques parmi les 50 premiers acides aminés, est responsable du ciblage et de l'ancrage de Gag à la membrane plasmique pour permettre l'assemblage du virion (Gelderblom, 1991). Une étude récente a démontré un équilibre entre deux états de myristoylation de MA : le groupe myristoylé est séquestré si Gag est un monomère, tandis qu'il est exposé et capable de lier la membrane lorsque Gag multimérise (Tang et al., 2004). CA intervient dans la multimérisation de Gag lors de l'assemblage de particules virales (Gamble et al., 1997; von Pöbltzki et al., 1993; Wills et Craven, 1991). CA est également impliquée dans l'incorporation de la cyclophiline A (cyp A), une chaperonne cellulaire, dans le virion (Franke et al., 1994). Cette dernière aurait un rôle important dans le désassemblage (*uncoating*) de la capsid virale lors de l'entrée du virion dans la cellule. Quant à la protéine NC, elle est responsable de l'incorporation spécifique du dimère d'ARN génomique dans la particule virale, grâce à une interaction avec le signal d'encapsidation  $\psi$  (Psi) situé dans la région 5' de l'ARN génomique viral (Aldovini et Young, 1990; Luban et Goff, 1991). La nature basique de NC lui permet également de lier de l'ARN non-spécifiquement, ce qui lui permet de couvrir le génome viral et de le protéger de la dégradation par les nucléases cellulaires. Le peptide p6

permet le détachement des particules virales de la membrane plasmique, facilitant ainsi leur relâche de la cellule. En fait, dès 1991, il a été montré qu'une délétion de p6 résulterait en une diminution de la quantité de particules virales relâchées et une accumulation de ces particules visibles par microscopie électronique à la surface des cellules (Gottlinger et al., 1991). La fonction de relâche de p6 est due à son domaine riche en proline (PTAP) nommé domaine tardif (*late domain*) (Huang et al., 1995). La protéine p6 permet également l'incorporation de la protéine virale Vpr dans les virions lors de l'assemblage (Kondo et Gottlinger, 1996; Lu et al., 1995). Finalement, contrairement aux protéines MA, CA, NC et p6, les peptides p1 et p2 ne jouent pas un rôle direct dans la formation des virions mais semblent avoir évolué de façon à contrôler principalement la vitesse avec laquelle le précurseur Gag est clivé par la protéase virale. En plus, la conformation ou le repliement de p1 semble important pour stabiliser le dimère d'ARN génomique, puisque la mutation de deux résidus proline de ce peptide diminue la stabilité de ce dimère et réduit l'infectivité du virus (Hill et al., 2002).

Le gène *pol* est exprimé uniquement sous la forme d'une fusion Gag-Pol (pr160<sup>Gag-Pol</sup>) suite au déphasage en -1 par une minorité de ribosomes traduisant l'ARN viral. Ce déphasage, qui a lieu à l'extrémité 3' de *gag*, fait en sorte que la séquence MA-CA-p2-NC du précurseur Gag est allongée par les séquences de la protéine *transframe* (TFP), de p6<sup>pol</sup>, de la protéase (PR, p10), de la transcriptase inverse (RT, p66-p51) et de l'intégrase (IN, p32) de Pol. Les séquences TFP et p6<sup>pol</sup> remplacent donc les séquences p1 et p6 de l'extrémité C-terminale de Gag dû au déphasage en -1. Plusieurs évidences suggèrent que TFP et p6<sup>pol</sup> contrôlent l'activité de la protéase (Hill et al., 2002), mais le mécanisme de ce contrôle demeure inconnu. Les recherches effectuées sur les enzymes virales du précurseur Pol sont nombreuses et ont révélé les principales caractéristiques de ces enzymes. D'abord, la protéase virale contrôle le clivage du précurseur Gag et Gag-Pol (Figure 1-1E), un processus critique de l'assemblage et de la maturation lors du cycle de réplication du virus. La protéase fonctionne seulement sous forme de dimère. Lors de l'assemblage deux molécules de Gag-Pol dimérisent de sorte que les deux molécules de protéase viennent en

contact, s'activent et amorcent le clivage des précurseurs (Navia et McKeever, 1990). Des données récentes suggèrent que le clivage initial effectué par la protéase serait intramoléculaire et non intermoléculaire (Pettit et al., 2004). La concentration de Gag-Pol est déterminante pour l'activation de la protéase, et, en fait, une surexpression de PR mène à un clivage prématuré et à une diminution de l'infectivité du virus (Sheng et al., 1997). Comme la production de Gag-Pol dépend d'un déphasage en  $-1$  (voir section 1.3.4 sur le déphasage en  $-1$ ), on voit immédiatement que l'efficacité de ce déphasage est critique. PR clive les précurseurs Gag et Gag-Pol au niveau de neuf sites de clivages dont les séquences sont différentes, mais caractérisées par la présence de résidus hydrophobes (Oroszlan et Copeland, 1985). Ces sites sont reconnus et clivés avec une efficacité distincte, ce qui entraîne une apparition ordonnée des produits de clivages (voir Pettit et al., 2002 et les références citées). Plusieurs études ont permis de déterminer l'ordre d'utilisation des sites de clivage du précurseur Gag. Cet ordre est le suivant :  $p2/NC > p1/p6 \approx MA/CA > CA/p2 > NC/p1$ . La vitesse avec laquelle les peptides  $p1$  et  $p2$  sont clivés aux sites  $CA/p2$  et  $NC/p1$  est de plusieurs centaines de fois plus lente que le premier clivage au site  $p2/NC$ . Le clivage à chacun de ces sites semble nécessaire puisque des substitutions bloquant le clivage aux sites  $p2/NC$ ,  $MA/CA$  et  $CA/p2$  résultent en la formation de particules non-infectieuses (Krausslich et al., 1995; Pettit et al., 1994; Wiegers et al., 1998). L'ordre d'utilisation des sites de clivage du précurseur Gag-Pol est moins connu, mais l'on sait cependant que le clivage initial se produit au niveau des sites  $p2/NC$  ou  $TFP/p6^{pol}$  (Pettit et al., 2004; Pettit et al., 2003). La reconnaissance de plusieurs sites de clivage par PR suggère une évolution de l'enzyme, de façon à ne pas accomplir une seule réaction de façon rapide, mais plutôt plusieurs réactions. Le rôle de la RT est de catalyser la conversion du génome viral, un ARNsb, en ADNc (Goff, 1990). La RT est une ADN polymérase capable d'utiliser soit l'ARN ou l'ADN comme matrice. La RT ne possède pas de mécanisme corrigeant les erreurs d'incorporation de nucléotides et son niveau élevé d'erreur de  $10^{-4}$  entraîne l'introduction d'une mutation dans le génome du virus à chaque cycle de réplication. Elle possède également une activité RNase H, servant à dégrader l'ARN du complexe hybride

ARN-ADN lors de la réaction. La RT est active sous forme d'un hétérodimère avec une sous-unité p66 et une sous-unité p51. Chacune des sous-unité possède l'activité polymérase et p66 contient une activité RNase H additionnelle. Finalement, le rôle de l'intégrase virale est de catalyser une série de réactions menant à l'intégration de l'ADN viral dans le chromosome de l'hôte (Goff, 1992; Vink et Plasterk, 1993). L'intégration peut se produire à plusieurs sites différents à l'intérieur du génome, mais des études *in vitro* ont montré une préférence pour les sites courbés ou distorsionnés (Miller et al., 1995).

Le produit du gène *env*, le précurseur des protéines glycosylées de l'enveloppe virale (Env, gp160) est clivé par des protéases cellulaires, ce qui permet d'obtenir une glycoprotéine extracellulaire (SU, gp120) et une protéine hydrophobe transmembranaire (TM, gp41, Figure 1-1E). Ces deux protéines sont essentielles pour la reconnaissance, la liaison et l'entrée du virus dans les cellules cibles. SU est une glycoprotéine située à la surface membranaire du virus qui initie l'entrée du virus dans la cellule en reconnaissant et en se liant spécifiquement au récepteur de surface cellulaire CD4. La région variable V3 de SU est importante pour le tropisme du virus. Elle devient exposée lors de l'interaction SU-CD4, permettant alors son interaction avec le co-récepteur CCR5 ou CXCR4 (Clapham et Weiss, 1997). Quant à TM, c'est une protéine transmembranaire qui permet la fusion des membranes virale et cellulaire suite à la liaison de SU au CD4 et au co-récepteur CCR5 ou CXCR4. La présence en N-terminal d'un segment hydrophobe riche en glycine est responsable de cette fusion (Hernandez et al., 1996).

Le VIH-1, tout comme certains virus d'immunodéficience des primates, est caractérisé par la présence de nombreux gènes accessoires comprenant : *vif*, *nef*, *vpr*, *vpu* et *tev*, ainsi que deux gènes régulateurs : *tat* et *rev*. Des virus mutants pour les gènes accessoires révèlent un phénotype sans défaut apparent de réplication lorsqu'ils sont testés dans des lignées de cellules communément utilisées en laboratoire. Toutefois, des défauts de réplication sont observés lors d'essais dans les cellules primaires, ce qui est un meilleur reflet de la situation physiologique. Les protéines accessoires semblent avoir évoluées pour

accomplir plusieurs fonctions indépendantes. Aucune observation ne supporte l'idée que les protéines accessoires ou régulatrices aient une activité catalytique. Elles fonctionnent plutôt comme des adaptateurs moléculaires connectant les facteurs viraux ou cellulaires aux différentes voies cellulaires. Le rôle de chacune de ces protéines est résumé ci-dessous (revue par Frankel et Young, 1998; Strebel, 2003) :

Le principal rôle de Tat est l'activation de la transcription des gènes viraux. Tat fonctionne comme un adaptateur moléculaire en dirigeant l'appareil de transcription cellulaire vers l'ARN viral pour augmenter la processivité de l'ARN polymérase II. Tat reconnaît la structure en tige-boucle TAR (*trans-activation-response element*), localisée dans la région 5' LTR de l'ARN viral, et se lie à cette structure en recrutant un facteur cellulaire, la cycline T1, et la kinase Cdk9. Cette dernière phosphoryle l'ARN polymérase II, ce qui augmente sa processivité.

Rev permet l'exportation des ARN viraux non-épissés et mono-épissés du noyau vers le cytoplasme. Cette protéine contient un domaine de liaison à l'ARN riche en arginine, qui lui permet de lier une structure en tige-boucle, RRE (*rev-responsive element*), présente dans la séquence du gène *env*. L'exportation des ARN viraux est accomplie grâce au recrutement du facteur d'exportation cellulaire Crm1 par Rev. La présence de Rev dans le noyau permet à ces ARN d'éviter l'épissage. Des résultats récents de Yedavalli *et al.* (2004) ont montré que DDX3, une hélicase cellulaire, était aussi impliquée dans l'exportation des ARN viraux médiée par Rev.

Vif a été sans contredit la protéine accessoire la plus en vogue durant ces dernières années. Même si l'on reconnaissait son importance comme un facteur d'infectivité virale depuis son identification en 1987, sa fonction est demeurée inconnue jusqu'à récemment. On sait maintenant que Vif permet de contrer l'effet du produit du gène cellulaire CEM15/APOBEC3G, une cytidine désaminase capable de bloquer la réplication virale dans

les cellules non-permissives (Sheehy et al., 2002) (voir aussi section 1.1.3.2 sur les facteurs cellulaires limitant la réplication virale).

Nef réduit le niveau d'expression du récepteur cellulaire CD4 à la surface de la cellule afin d'empêcher son interaction avec SU, ce qui inhiberait la relâche des particules virales. Nef accomplit cette fonction en redirigeant les récepteurs CD4 de la surface cellulaire ou de la surface du Golgi vers le lysosome, où ils sont dégradés. Nef affecte également les voies de signalisation cellulaires et diminue le niveau d'expression des molécules du complexe majeur d'histocompatibilité de classe I (CMH-1), ce qui protège les cellules infectées de la destruction par les lymphocytes T cytotoxiques (CTL).

Vpr permet la localisation nucléaire du complexe contenant l'ADNc, obtenu par copie de l'ARN génomique viral par la RT, associé avec des protéines MA, RT, et IN dans les cellules quiescentes, tels les macrophages, grâce à son signal de localisation nucléaire. Vpr induit aussi l'arrêt du cycle cellulaire en G2 et, lorsqu'elle est surexprimée, provoque l'apoptose des cellules T.

Vpu favorise la dégradation des molécules de CD4 nouvellement produites dans le réticulum endoplasmique. Elle effectue cette tâche par un processus dépendant de l'ubiquitination, empêchant les CD4 de se diriger à la surface des cellules, ce qui favorise la relâche des virions.

Tev est une protéine chimérique synthétisée à partir d'un ARNm contenant, dans l'ordre, la séquence codante du premier exon de *tat*, une portion de *env*, et le second exon de *rev*. Tev possède les activités fonctionnelles de Tat et Rev, mais on ignore si elle joue un rôle dans la réplication du virus.

## 1.1.2 Le cycle de réplication du VIH-1

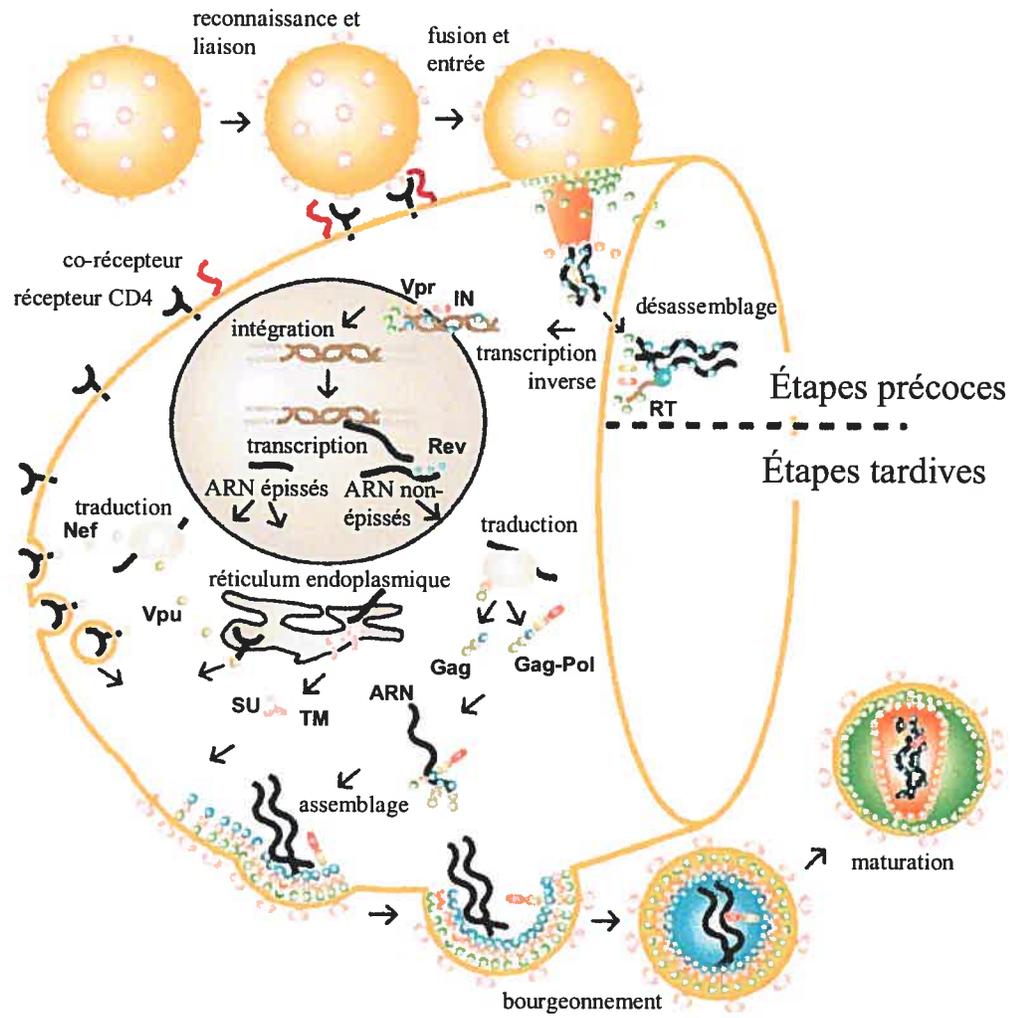
Dès l'identification du VIH-1 comme l'agent causal du SIDA, les chercheurs se sont empressés de tenter de comprendre comment ce virus se répliquait. La connaissance de l'agent infectieux est un atout essentiel si l'on veut relever ses faiblesses. Nous connaissons maintenant de façon détaillée la majeure partie des étapes du cycle de réplication du VIH-1. Ce cycle peut être divisé en une succession d'étapes, représentées à la Figure 1-2, contrôlées par les protéines virales et cellulaires. Ces étapes peuvent toutefois être regroupées en deux classes, soit les étapes précoces et les étapes tardives. Au cours des deux sections suivantes, nous aborderons de façon plus détaillée ces étapes afin de montrer la synergie entre le virus et la cellule infectée (revue par Ferguson et al., 2002; Freed, 1998; Muriaux et al., 2004).

### 1.1.2.1 Les étapes précoces

*Liaison au récepteur, liaison au co-récepteur, fusion et entrée*

Tel que présenté à la Figure 2, le cycle de réplication s'amorce par l'entrée du virus dans la cellule en une succession de trois étapes menant à la fusion des membranes virale et cellulaire. C'est la protéine SU qui possède l'information nécessaire à la reconnaissance de la cellule cible et la protéine TM qui permet la fusion des membranes. Le tout s'amorce par une liaison de haute affinité de SU avec le récepteur CD4 présent à la surface des cellules cibles, les lymphocytes T *helpers* et les macrophages-monocytes (Wang et al., 2000). Cette liaison entraîne un changement de conformation de SU, exposant les épitopes de la région variable de la boucle V3 qui sont alors reconnus par un co-récepteur. La présence d'un co-récepteur est nécessaire pour permettre l'entrée du virus. Plusieurs récepteurs couplés aux protéines G, peuvent servir de co-récepteurs pour l'infection par le VIH-1 *in vitro*. Toutefois, seuls CCR5 et CXCR4 semblent servir de co-récepteurs *in vivo* (Wang et al., 2000). La liaison aux co-récepteurs entraîne un réarrangement de la protéine TM : la région de fusion et le domaine transmembranaire se rapprochent, ce qui favorise la fusion des

Figure 1-2: Étapes générales du cycle de réplication du VIH-1. La phase précoce (partie supérieure du schéma) s'amorce par la reconnaissance du récepteur cellulaire CD4, et se déroule jusqu'à l'intégration de l'ADN proviral, et la phase tardive (partie inférieure du schéma) implique les événements qui ont lieu à partir de la transcription du provirus jusqu'à la relâche et la maturation du virus (adapté de Turner et Summers, 1999).



membranes virale et cellulaire et permet d'introduire la capsid virale dans le cytoplasme de la cellule infectée.

#### *Désassemblage, transcription inverse et intégration*

Par la suite, la capsid est défaire par un processus moins connu, appelé désassemblage, et quelques molécules de MA, NC, ainsi que des molécules de RT, IN, l'ARN viral génomique et la protéine accessoire Vpr demeurent groupés pour former un complexe de haut poids moléculaire. Durant ce désassemblage, la RT entame la conversion du génome d'ARNsb viral en ADNc par une réaction de transcription inverse. Cette réaction complexe s'amorce par un appariement spécifique de l'ARNt<sub>3</sub><sup>Lys</sup> au site de liaison de l'amorce (PBS ou *primer binding site*), à l'extrémité 5' du génome viral, et implique des sauts de la RT d'une extrémité à l'autre de l'ARN/ADN. Une fois la transcription inverse complétée, ce complexe de haut poids moléculaire est nommé complexe de pré-intégration (PIC) et est transporté aux pores nucléaires pour être introduit de façon active dans le noyau. À cet endroit, l'intégrase virale catalyse l'intégration de l'ADNc dans le chromosome de l'hôte pour former le provirus. L'ADNc proviral contient deux séquences identiques aux extrémités, les LTR, dont celle à l'extrémité 5' sera utilisée comme promoteur de la transcription.

#### **1.1.2.2 Les étapes tardives**

##### *Transcription et synthèse protéique*

Les étapes tardives du cycle de réplication du VIH-1 requièrent la présence de facteurs cellulaires pour accomplir la synthèse, l'épissage et la traduction des ARNm viraux. L'initiation de la transcription est amorcée, de façon peu efficace, à partir des promoteurs LTR. Les premiers transcrits à être produits sont les transcrits multi-épissés de *tat*, *rev* et *nef*, qui, une fois présents dans le cytoplasme, sont traduits par les ribosomes pour produire les protéines Tat, Rev et Nef. L'expression de Tat et Rev permet alors d'augmenter la processivité de l'ARN polymérase II et d'exporter les transcrits non-épissés et mono-

épissés, respectivement. Les précurseurs Gag et Gag-Pol sont alors synthétisés dans le cytoplasme par traduction du transcrit non-épissé, alors que Env est produite au niveau du réticulum endoplasmique à partir d'un transcrit mono-épissé. Env est ensuite glycosylée et clivée dans le Golgi par des enzymes cellulaires. Les protéines Vpr, Vpu et Vif sont également traduites par les ribosomes à partir de messenger mono-épissés.

*Assemblage, bourgeonnement, relâche et maturation*

Plusieurs études sur l'assemblage et la formation de particules virales ont montré que Gag contient toute l'information nécessaire à ce processus. Gag est ciblé à la membrane plasmique, grâce à MA, où il multimérise, par la participation de CA et aussi NC, pour créer une structure ordonnée qui forme la particule virale. Cette structure est relâchée de la cellule par bourgeonnement avec l'intervention de p6. Deux molécules d'ARN génomique pleine longueur sont incorporées dans les virions par l'entremise de NC. Jusqu'à tout récemment, on croyait que l'assemblage des particules virales se produisait exclusivement à la membrane plasmique. Il a récemment été démontré que l'assemblage de virions se produit aussi dans les endosomes dits tardifs chez les macrophages et que ces virions sont amenés à la membrane plasmique par les voies de sécrétion (Pelchen-Matthews et al., 2003). Gag interagit également avec TM, grâce à MA, lors de l'initiation du bourgeonnement afin de permettre l'incorporation des protéines de membrane dans le virion, TM entraînant SU. Le bourgeonnement est causé par la présence d'un domaine protéique conservé chez les rétrovirus. Chez le VIH-1, ce domaine, dit domaine tardif, est situé dans p6, alors que, par exemple, chez le virus de l'anémie infectieuse du cheval (EIAV), ce domaine se retrouve dans la protéine homologue p9. Gag, par l'intermédiaire du domaine tardif de p6, interagit avec la machinerie cellulaire de la voie de formation des corps multivésiculaires (MVB), ce qui favorise la relâche des virions. Finalement, suite au bourgeonnement, les virions relâchés passent de l'état immature à un état mature infectieux grâce à l'action de la protéase virale qui clive les précurseurs polyprotéiques Gag et Gag-Pol.

### 1.1.3 L'hôte

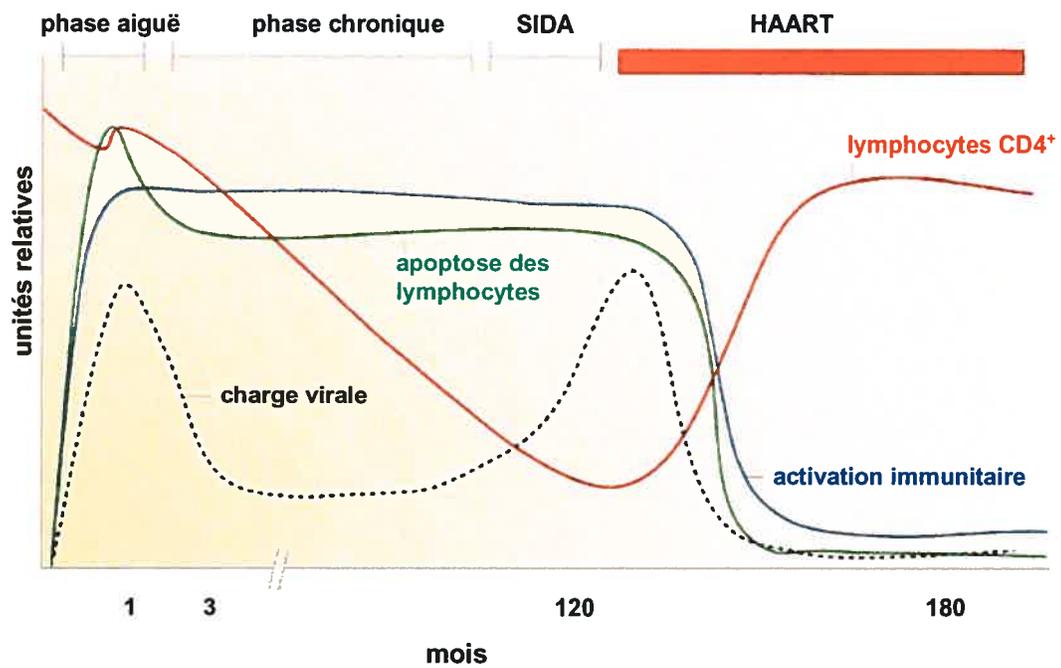
L'infection d'une personne par le VIH-1 mène à une longue bataille entre son système immunitaire et la réplication du virus. Cette bataille n'est presque jamais gagnée par l'hôte malgré l'utilisation d'armes immunitaires de destruction comme les lymphocytes T cytotoxiques (CTL). En effet, l'hôte ne réussit pas à se débarrasser du pathogène qui joue le jeu de se cacher, de se déployer et de contre-attaquer. Des études récentes ont permis d'identifier certains gènes cellulaires participant à la défense de la cellule contre cet envahisseur. Ces gènes codant pour les protéines APOBEC3G et TRIM5 $\alpha$  ont incité la communauté scientifique à rechercher d'autres facteurs de restriction cellulaire et à identifier les moyens permettant à la cellule d'utiliser efficacement ces facteurs. L'hôte n'est cependant pas laissé à lui-même pour lutter contre le VIH-1. De nombreux inhibiteurs de la réplication virale sont disponibles et l'utilisation de la thérapie antirétrovirale hautement active (HAART) permet de réduire considérablement la réplication du virus (voir section 1.1.3.3). Malheureusement, si l'on interrompt ces traitements, le virus recommence à se répliquer à un rythme normal (Cohen et Fauci, 2001). De plus, à la vitesse avec laquelle l'hôte élimine le virus lors du HAART, l'éradication du virus nécessiterait des traitements plus longs que la durée de vie des personnes infectées (Blankson et al., 2002). Cependant, l'apparition de virus résistants à ces inhibiteurs rend la thérapie inefficace bien avant cette éradication. Des informations générales sont fournies plus bas sur la réponse immunitaire de l'hôte, les facteurs cellulaires de restriction et les thérapies contre le VIH-1.

#### 1.1.3.1 La réponse immunitaire

Les CTL du système immunitaire sont importants pour la prévention et le contrôle des infections virales (revue par Collins, 2003). Une étude élégante a permis de montrer

que la déplétion de CTL, plus particulièrement des CTL CD8+, chez le singe empêchait ce dernier de contrôler l'infection par le virus de l'immunodéficience du singe (SIV) (Schmitz et al., 1999). Le rôle de ces CTL est de reconnaître les cellules infectées qui possèdent des peptides étrangers présentés à leur surface par les molécules CMH-I. De cette façon, l'hôte est en mesure de reconnaître les envahisseurs présents dans ses propres cellules. Les CTL détruisent les cellules infectées en induisant leur apoptose et provoquent également leur lyse en sécrétant des perforines, c'est-à-dire des protéines qui entraînent la perforation des membranes. Suite à l'infection de l'hôte par le VIH-1, le virus se réplique et s'accumule à de fortes concentrations durant la phase aiguë d'infection. Un schéma du comportement du système immunitaire et du virus lors de la maladie est présenté à la Figure 1-3. L'accumulation du VIH-1 diminue ensuite lors de l'apparition des cellules CTL anti-VIH (Borrow et al., 1994). Malgré cela, durant la phase chronique, le virus est toujours en mesure de se répliquer à des taux élevés, et se répand dans le sang et les tissus lymphoïdes. On considère donc l'infection par le VIH-1 comme une infection chronique puisque chez les personnes infectées, le virus persiste et est capable d'échapper à la réponse immunitaire par de nombreux moyens. En effet, les protéines virales de surface échappent aux anticorps grâce à un taux élevé de mutations et elles sont aussi protégées par la glycosylation. De plus, le virus à l'intérieur de la cellule diminue l'expression des molécules du CMH-I et la quantité de récepteurs CD4 à la surface, grâce à Nef, qui contre l'effet apoptotique des CTL. Ultimement, le virus entraîne la destruction des cellules infectées par l'activité pro-apoptotique de plusieurs de ses protéines, après avoir préalablement produit une grande quantité de virions et entraîné une destruction des CTL par apoptose. Avec le temps, le VIH mène à une diminution graduelle du nombre de cellules-T lymphocytaires, incluant les cellules CD4+, à la déficience immunitaire et au développement du SIDA. Le VIH avec ses protéines accessoires est donc capable de prolonger la survie des cellules qu'il infecte, ce qui lui permet de se répliquer et de produire de nombreux variants capables d'éviter la reconnaissance par le système immunitaire, grâce à l'accumulation de mutations à chaque cycle de réplication virale.

Figure 1-3: Relation entre l'activation du système immunitaire par le virus, l'apoptose des cellules-T lymphocytaires et la progression de la maladie avant et après thérapie anti-rétrovirale HAART (adapté de Gougeon, 2003).



### 1.1.3.2 Les facteurs cellulaires limitant la réplication virale

En plus de la réponse immunitaire, les organismes supérieurs expriment de façon constitutive une panoplie de gènes qui préviennent les infections virales. Ces gènes, différents de ceux qui sont exprimés dans le système immunitaire, codent pour des protéines nommées facteurs de restriction (selon Bieniasz, 2004). Ces facteurs constituent en quelque sorte la première ligne de défense contre les pathogènes puisque la cellule peut agir suite à l'interaction avec le virus, sans l'activation du système immunitaire. Jusqu'à ce jour, deux groupes majeurs de gènes de défense contre les rétrovirus ont été identifiés : le groupe TRIM5 et le groupe APOBEC. Pendant longtemps, les virologistes ont tenté de comprendre pourquoi les hommes, et non les singes, étaient susceptibles au VIH-1. L'identification par Stremlau et al. (2004) de la protéine TRIM5 $\alpha$ , qui bloque l'infection des singes rhésus macaques par le VIH, a permis de répondre à cette question. La protéine TRIM5 $\alpha$  est le produit du gène *Lv1*, dont on connaissait la capacité d'inhiber la réplication des rétrovirus chez le singe (Besnier et al., 2002; Cowan et al., 2002). Chez le singe, l'activité antivirale de TRIM5 $\alpha$  est due à un blocage de la réplication lors de l'étape du désassemblage de la capsid, empêchant la synthèse de l'ADN viral et son transport au noyau. Ce processus, encore mal défini, serait causé par l'interaction de TRIM5 $\alpha$  avec les protéines CA, causant une désorganisation de la capsid empêchant l'action de la RT. Il est aussi possible que TRIM5 $\alpha$  séquestre le virus dans des compartiments cellulaires où il sera dégradé. L'efficacité de TRIM5 $\alpha$  des macaques contre le VIH-1 n'est pas reproductible chez l'homme. L'homologue humain du gène primate *Lv1*, *Ref1*, possède une activité antivirale efficace contre le rétrovirus de la leucémie murine (MuLV) et EIAV, mais non contre le VIH-1 (Goff, 2004). Il semble en fait que le VIH-1 ait évolué afin d'acquérir une activité de liaison capsid-cyclophiline A. Cette activité lui permettrait d'éviter la restriction par les facteurs liant la capsid tel TRIM5 $\alpha$  quand il infecte l'homme (Towers et al., 2003). À cet effet, on a récemment identifié un variant de TRIM5 $\alpha$  fusionné à la

cyclophiline-A, TRIM5-cypA, chez les singes hiboux (*owl world monkeys*) capable de bloquer l'infection par le VIH-1 (Nisole et al., 2004).

Le second groupe majeur d'activité cellulaire antirétrovirale connue n'empêche pas la cellule d'être infectée, mais a plutôt comme objectif que les virions produits soient non-infectieux. Pendant plusieurs années, les chercheurs ont tenté de comprendre pourquoi le besoin de Vif pour la réplication du VIH-1 dépendait de facteurs cellulaires. En effet, certaines lignées cellulaires nommées *permissives* permettent la réplication du virus dépourvu de Vif, alors que d'autres lignées *non-permissives* ne le permettent pas (Gabuzda et al., 1992). De plus, ce phénotype dépend aussi du type de cellules produisant les virions, c'est-à-dire que des virions produits par des cellules *permissives* peuvent infecter des cellules *non-permissives*. La recherche du gène conférant le phénotype non-infectieux dans les cellules non-permissives a mené à l'identification du gène APOBEC3G, une cytidine désaminase (Sheehy et al., 2002). APOBEC3G est incorporée dans les particules virales et son action cytidine désaminase entre en jeu lors de la réaction de la transcription inverse à l'intérieur de la cellule infectée (Gabuzda et al., 1992; Harris et al., 2003; Lecossier et al., 2003; Zhang et al., 2003). L'action de APOBEC3G génère un génome viral ayant une accumulation de mutations G-A et une stabilité réduite. Dans la bataille entre l'hôte et le VIH-1, ce dernier a réussi à contrecarrer la restriction de APOBEC3G en acquérant la protéine Vif, qui entraîne la dégradation d'APOBEC3G par le protéasome via la voie d'ubiquitination (Marin et al., 2003; Sheehy et al., 2003; Yu et al., 2003).

### 1.1.3.3 Les thérapies contre le VIH-1

En vingt ans, depuis la découverte du VIH-1, la recherche et le développement sur les médicaments ont permis de transformer le SIDA en une maladie traitable et d'augmenter la qualité de vie des patients. Toutefois, lorsque les patients sont traités sur une longue période de temps, les problèmes dus à la toxicité des médicaments s'aggravent

et l'émergence de virus résistants contrecarre l'action de ces médicaments (revue par Pomerantz et Horn, 2003). Les thérapies contre le VIH-1 comprennent habituellement une combinaison de nucléosides inhibiteurs de la RT (NRTI), d'inhibiteurs de la RT non-nucléosidiques (NNRTI) et d'inhibiteurs de la PR. Ces inhibiteurs ciblent l'action de deux protéines virales du cycle de réplication, mais les problèmes surgissent lorsque la population virale acquiert une résistance contre chaque classe d'inhibiteurs. Ceci rend donc nécessaire le développement d'inhibiteurs efficaces et peu toxiques contre de nouvelles cibles virales. Le premier inhibiteur qui a été développé contre une nouvelle cible, l'enfivirtide, est un inhibiteur bloquant l'entrée du virus dans la cellule en empêchant l'étape de la fusion. Il y a d'autres nouvelles cibles, et les plus prometteuses concernent l'intégration, la liaison du virus aux récepteurs CD4 et co-récepteurs CCR5/CXCR4 et la transactivation par Tat. Des médicaments agissant sur ces étapes sont actuellement étudiés en clinique (revue par De Clercq, 2002). Suite à l'identification récente du rôle de facteurs cellulaires de restriction, le développement de médicaments bloquant l'action de Vif sont aussi prometteurs. Nous proposons que des agents ciblant le déphasage en  $-1$  du VIH-1 ont également un potentiel très intéressant car ils perturberaient l'expression des enzymes virales. Lors de mes études de doctorat, j'ai entamé le développement d'un procédé permettant la sélection de courts peptides ciblant le mécanisme du déphasage en  $-1$  du VIH-1. Les résultats préliminaires de ce projet seront présentés au Chapitre 4.

## 1.2 La traduction

La traduction est un processus par lequel l'information génétique contenue dans la séquence des ARN messagers, sous la forme d'une suite de trois bases ou codon, est utilisée pour synthétiser des protéines dont la séquence en acides aminés est déterminée par ces codons. La lecture de la séquence des ARNm nécessite la participation des ARN de transfert chargé d'un acide aminé (aa-ARNt). Le ribosome est la particule contrôlant cette synthèse protéique. La traduction se produit principalement en quatre étapes : l'initiation,

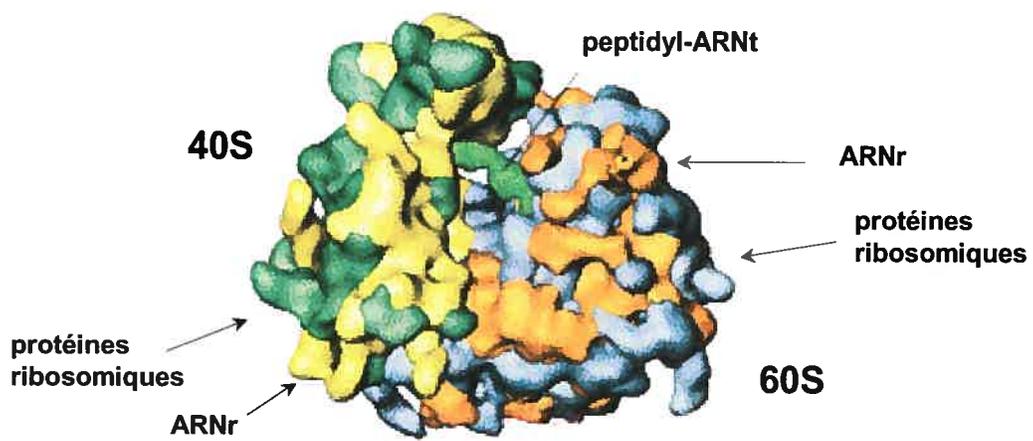
l'élongation, la terminaison et le recyclage des ribosomes. La précision avec laquelle le ribosome traduit le messenger et ajoute à la chaîne protéique en croissance l'acide aminé correct est remarquable. En effet, le ribosome possède des mécanismes contrôlant le choix de l'aa-ARNt, et il peut rapidement corriger une mauvaise sélection, c'est-à-dire la liaison d'un aa-ARNt dont l'anticodon ne correspond pas au codon du messenger. De cette façon, seulement un acide aminé incorrect est incorporé tous les  $5 \times 10^3$  codons (Kurland, 1992). Un autre aspect important de la synthèse protéique est la capacité du ribosome à maintenir le bon cadre de lecture. En effet, le codon initiateur du messenger dicte le cadre de lecture à suivre, mais à partir du moment où le ribosome traduit un gène, il ne possède aucune référence lui confirmant qu'il traduit le bon cadre de lecture. Toute défaillance dans le maintien de ce cadre, soit due au hasard soit voulue (voir section 1.3.3 sur le décalage de phase de lecture), mène à la synthèse d'une autre protéine que la protéine correspondant au cadre de lecture initial. Le ribosome fait une erreur de maintien du cadre de lecture une fois tous les  $3 \times 10^5$  codons lus (Stahl et al., 2002).

De grandes avancées ont été effectuées depuis les 30 dernières années dans la compréhension du fonctionnement du ribosome. C'est au cours des dernières années, principalement grâce aux informations structurales obtenues par cristallographie aux rayons X et par cryo-microscopie électronique (cryo-EM) du ribosome bactérien (revue par Ogle et al., 2003; Ogle et Ramakrishnan, 2004; Puglisi et al., 2000; Ramakrishnan, 2002), que l'on a commencé à comprendre les secrets des mécanismes moléculaires de cette machine complexe. Les connaissances que nous avons du ribosome eucaryote sont moins détaillées, mais beaucoup d'informations peuvent être déduites compte tenu du haut degré de conservation entre le ribosome eucaryote et le ribosome bactérien. Dans cette section, nous allons d'abord parcourir les connaissances que nous avons du ribosome eucaryote et les différentes étapes du cycle de la traduction. Nous examinerons aussi un mécanisme cellulaire de surveillance de la qualité des ARNm.

### 1.2.1 Le ribosome

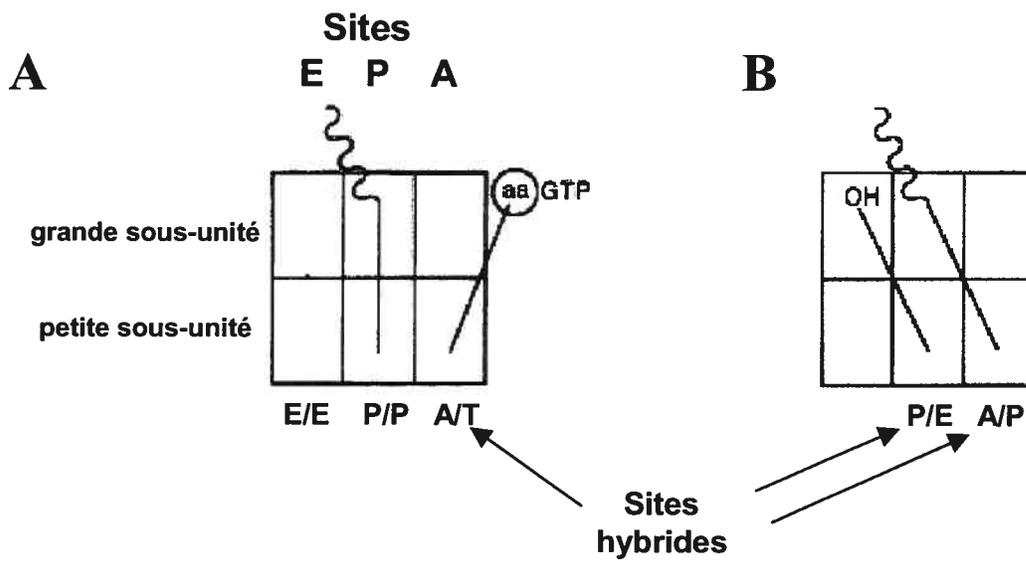
Chez les eucaryotes, le ribosome 80S, est formé de l'association de la petite sous-unité, la 40S, et d'une grande sous-unité, la 60S, contenant à elles deux près de 80 protéines liées à quatre molécules d'ARN ribosomique (ARNr). Les données structurales que nous possédons sur le ribosome 80S eucaryote proviennent surtout du ribosome de la levure *S. cerevisiae* (Spahn et al., 2001) (Figure 1-4). Bien que le ribosome 80S contiennent un ARNr de plus et 20 à 30 protéines de plus que le ribosome bactérien, les similarités entre plusieurs composantes suggèrent que le ribosome possède un noyau conservé chez tous les organismes (Doudna et Rath, 2002; Mears et al., 2002). La petite sous-unité 40S est responsable de la liaison et du décodage des ARNm, et elle vérifie l'appariement correct entre le codon de l'ARNm et l'anticodon de ARNt. Elle est composée chez la levure d'un ARNr 18S de 1798 nucléotides et de 31 protéines différentes. Quant à la grande sous-unité ribosomique 60S, elle contient le centre peptidyle transférase responsable de la catalyse du lien peptidique durant la synthèse protéique. Elle est formée de trois ARNr : le 25S de 3392 nucléotides, le 5.8S de 158 nucléotides et le 5S de 121 nucléotides, et de 41 protéines différentes. Il est maintenant reconnu que les principales fonctions du ribosome sont assurées par l'ARNr. La liaison au messenger, l'assemblage des sous-unités, la sélection des ARNt, et la formation du lien peptidique, pour ne citer que ces étapes, sont catalysées par l'ARNr. En fait, l'ARNr se retrouve principalement à l'interface des deux sous-unités du ribosome, alors que les protéines ribosomique se retrouve en surface, du coté exposé au solvant, et ce, tant chez les bactéries, les eucaryotes et les archéobactéries. Le ribosome initial était seulement composé de molécules d'ARN et l'ajout de protéines serait survenu plus tard au cours de l'évolution (Noller, 2004). Cette addition de protéines aurait permis la stabilisation des structures tertiaire de l'ARNr et aurait favorisé l'activité catalytique du ribosome. La présence de nombreuses protéines, sans équivalents chez les bactéries et les archéobactéries, laisse supposer que ces protéines additionnelles interagissent avec des facteurs ou des régulateurs spécifiques à la traduction eucaryote.

Figure 1-4: Structure du ribosome 80S de la levure *S. cerevisiae* obtenue par cryo-EM par le groupe de Frank (Beckmann et al., 2001), à une résolution de 15Å. Les protéines et les ARNr de la sous-unité 40S sont colorés en vert et jaune, et les protéines et les ARNr de la sous-unité 60S sont en bleu et orange, respectivement. Un peptidyl-ARNt occupant le site P est également indiqué dans cette structure (adapté de Doudna et Rath, 2002).



Lors de la synthèse protéique, les acides aminés portés par les ARNt sont successivement introduits dans la chaîne polypeptidique en croissance. Au cours de ce processus, les ARNt interagissent avec plusieurs sites sur le ribosome, nommés site A (aminoacyl), P (peptidyle) et E (sortie ou *exit*). Dans le modèle classique, le site A du ribosome correspond au site de liaison de l'ARNt chargé de l'acide aminé (aa-ARNt) qui va être introduit dans la chaîne protéique, le site P correspond à l'endroit où se place l'ARNt chargé de la chaîne protéique en croissance (peptidyl-ARNt), et le site E correspond au site de sortie de l'ARNt déchargé après la formation du lien peptidique. À partir de leurs recherches sur le ribosome bactérien, Moazed et Noller (1989) ont proposé le modèle des sites hybrides pour décrire le déplacement de ces ARNt au cours de l'élongation, et ce, en se basant sur des études de protection de l'ARNr contre des agents chimiques lors de la liaison des ARNt au ribosome bactérien. Ce modèle peut être extrapolé au ribosome eucaryote. La particularité de ce modèle est l'indépendance de l'interaction des extrémités des ARNt, soit la tige acceptrice et la boucle anticodon avec la grande et la petite sous-unité ribosomique, respectivement. Une nomenclature plus complète pour les sites d'interaction des ARNt avec le ribosome consiste à les nommer respectivement A/A, P/P et E/E, où la première lettre désigne la portion du site occupé sur la petite sous-unité, et la deuxième lettre indique la portion du site occupé sur la grande sous-unité. Lorsque nous référons aux sites classiques A, P et E, il s'agit en fait du site A/A, P/P et du site E/E, respectivement, mais pour des raisons de simplicité nous continuons à utiliser la nomenclature A, P, E dans ces cas. Durant la synthèse protéique il y aura aussi occupation des sites hybrides comme A/P pour le peptidyl-ARNt et P/E pour l'ARNt déchargé, tel qu'illustré à la Figure 1-5. Par exemple, dans le site A/P, la portion du peptidyl-ARNt interagissant avec la petite sous-unité occupe le site A, tandis que la partie de ce peptidyl-ARNt interagissant avec la grande sous-unité occupe le site P. Le groupe de Noller a également proposé un site initial de liaison du complexe ternaire, c'est-à-dire le complexe formé par un aa-ARNt lié au facteur d'élongation EF-Tu chargé de GTP. Ce site est défini comme le site hybride A/T, où la portion de l'aa-ARNt interagissant avec la petite sous-unité est le site A, et la portion de

Figure 1-5: Exemple des sites hybrides rencontrés par les ARNt sur le ribosome lors de la traduction. Par exemple, lors de la sélection initiale, le complexe ternaire, contenant un aa-ARNt lié à un facteur d'élongation chargé de GTP, se lie au site hybride A/T, tandis que le peptidyl-ARNt occupe le site P/P classique (schéma A). Par contre, après la formation du lien peptidique, il y a un déplacement de ces deux ARNt, le peptidyl-ARNt occupe maintenant le site hybride A/P, et l'ARNt déchargé est au site hybride P/E (schéma B).



l'aa-ARNt interagissant avec le grande sous-unité est le site T, différent du site A de cette sous-unité (Moazed et al., 1988; Noller et al., 2002; Wilson et Noller, 1998) (voir la section 1.2.2.2 décrivant l'étape d'élongation de la traduction).

## 1.2.2 Cycle de traduction

L'initiation, l'élongation, la terminaison et le recyclage des ribosomes constituent les quatre étapes effectuées par le ribosome permettant la synthèse d'une protéine, et ce, dans tous les organismes (revue par Kapp et Lorsch, 2004). Lors de l'initiation, le ribosome avec l'ARNt initiateur chargé d'une méthionine au site P, le Met-ARNt<sub>i</sub><sup>Met</sup>, est formé à partir des deux sous-unités au niveau du codon initiateur AUG de l'ARNm. Par la suite, en élongation, un aa-ARNt est amené au site A pour y être décodé. Si l'aa-ARNt sélectionné est correct, c'est-à-dire qu'il y a complémentarité entre le codon du messager et l'anticodon de l'ARNt, il y a catalyse du lien peptique entre l'acide aminé et la chaîne peptidique portée par l'ARNt au site P. Le résultat est que la chaîne peptidique se retrouve chargée sur l'ARNt au site A. Ceci est suivi par une translocation du peptidyl-ARNt au site P, laissant le site A libre pour l'incorporation de l'aa-ARNt suivant. Cette étape est répétée dépendant de la taille de la protéine à produire. La terminaison correspond à la fin de la synthèse de la chaîne protéique et son relâchement du ribosome. Elle survient lorsque le ribosome atteint un codon stop (ou de terminaison ou d'arrêt ou non-sens). Finalement, le recyclage permet la dissociation des deux sous-unités du ribosome à partir de l'ARNm. Ces sous-unités sont alors prêtes à amorcer un nouveau cycle de traduction. Voici donc ces étapes en détails.

### 1.2.2.1 Initiation

Le but de l'initiation, autant chez les eucaryotes que chez les bactéries, est d'apporter l'ARNt initiateur chargé d'une méthionine et la petite sous-unité ribosomique à l'ARNm à traduire, de trouver le codon AUG initiateur et d'associer ensuite les deux sous-unités ribosomique pour former le ribosome. Cependant, chez les eucaryotes, le nombre

d'étapes et de facteurs requis sont beaucoup plus élevés que chez les bactéries. L'initiation de la traduction chez les eucaryotes est sans doute l'étape la plus complexe du cycle, faisant intervenir un très grand nombre de facteurs d'initiation (revue par Jackson, 2000). Il existe différentes façons d'initier la traduction sur un messenger, mais le modèle de balayage (*scanning*) coiffe-dépendante est presque exclusivement utilisé pour la traduction des gènes cellulaires chez les eucaryotes. Tous les ARN messagers cellulaires (à l'exception de ceux des organelles) ont une structure  $m^7GpppX$ , dite la coiffe, à leur extrémité 5' (où  $m^7G$  est une guanosine méthylée, X est n'importe lequel des nucléotides et où la liaison entre  $m^7G$  et X se fait par un lien 5'-5' triphosphate). L'initiation coiffe-dépendante par balayage (voir Figure 1-6) s'amorce par la formation du complexe ternaire comprenant le facteur d'initiation eIF2 lié à une molécule de GTP et à l'ARNt initiateur chargé de méthionine (eIF2\*GTP\*Met-ARNt<sup>Met</sup>). Contrairement à l'initiation bactérienne où la méthionine utilisée pour l'initiation possède un groupement amine modifié, la méthionine utilisée pour l'initiation eucaryote n'est pas modifiée. L'ARNt initiateur qui lie la méthionine et participe à l'initiation (ARNt<sup>Met</sup>) a une structure différente de celle des ARNt intervenant dans l'élongation de la traduction. Le complexe ternaire s'associe avec la sous-unité 40S où le Met-ARNt<sup>Met</sup> occupe le site P de la 40S pour former le complexe 43S. Les facteurs d'initiation eIF1, eIF1A et eIF3 sont également présents dans ce complexe. La liaison du complexe 43S à l'ARNm est assurée par une interaction entre eIF3 associé à la 43S, et le facteur eIF4F lié au messenger. eIF4F est un complexe composé de trois polypeptides : eIF4A, une ATPase ARN dépendante ayant une fonction ARN hélicase capable de défaire les structures d'ARN double-brin (ARNdb) en hydrolysant l'ATP, eIF4E, une protéine liant la structure en coiffe à l'extrémité 5' du messenger, et eIF4G, un large polypeptide qui possède des sites de liaisons pour eIF4E, eIF4A, eIF3. En plus, eIF4G contient un site de liaison pour PABP, une protéine liant la queue de polyA du messenger. Il est à noter que cette interaction entre PABP et eIF4G est responsable de la circularisation des ARNm eucaryotes (voir Figure 1-7). Il faut aussi ajouter que eIF4A est aidée dans sa fonction hélicase par un facteur supplémentaire eIF4B. Le complexe 43S procède ensuite à un

Figure 1-6: Schéma représentant les étapes du mécanisme de l'initiation coiffe-dépendante chez les eucaryotes. La traduction des ARNm chez les eucaryotes implique le recrutement des ARNm par la machinerie de l'initiation traductionnelle avec la sous-unité ribosomique 40S et la formation du ribosome 80S sur cet ARNm. Ce processus est contrôlé par plusieurs protéines nommées facteurs d'initiation eucaryote (eIF) (voir texte) (Holcik et Sonenberg, 2005).

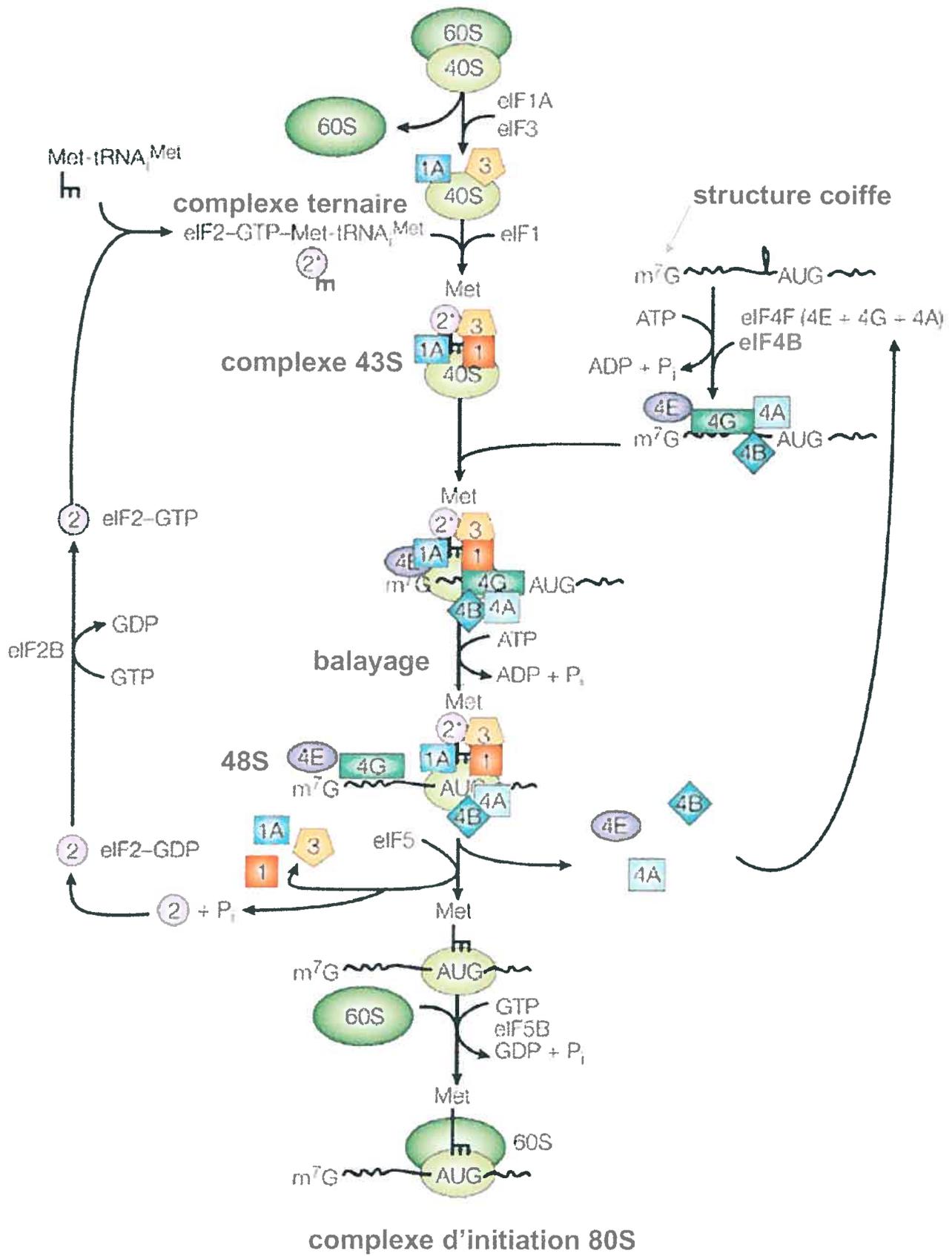
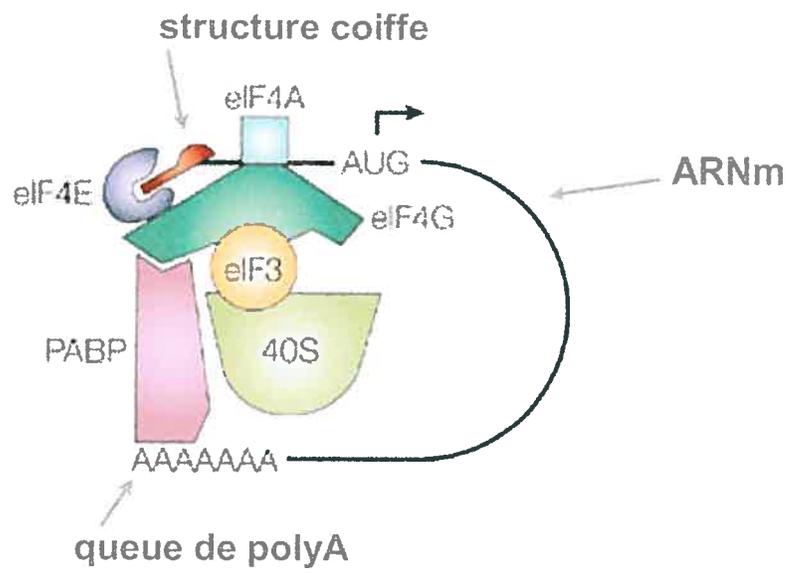


Figure 1-7: Circularisation des ARNm eucaryotes par la machinerie d'initiation de la traduction (Holcik et Sonenberg, 2005).



balayage du messager en direction 3' à la recherche d'un codon initiateur AUG. Ce balayage est un processus qui requiert l'hydrolyse de l'ATP, permettant l'élimination des structures secondaires bloquant ce processus, et est favorisé par les facteurs eIF1 et eIF1A de la 43S. Lorsque le complexe 43S rencontre un codon AUG favorable, c'est-à-dire un codon initiateur avec un environnement correspondant à un site consensus (a/g)ccAUGg (nommé séquence Kozak), l'appariement codon-anticodon entre le Met-ARNt<sub>i</sub><sup>Met</sup> du complexe ternaire et l'AUG du messager active la fonction GTPase de eIF2. Il y a hydrolyse du GTP lié à eIF2, ce qui favorise la relâche des facteurs présents soit eIF1, eIF1A, ainsi que eIF3 et eIF2\*GDP. L'hydrolyse du GTP par eIF2 est facilitée par la liaison du facteur eIF5, une protéine activatrice de GTPase. Si la séquence entourant le premier codon AUG rencontré est différente de la séquence Kozak, l'initiation est soit inefficace soit inexistante et la 43S continue le balayage du messager jusqu'à ce qu'elle rencontre un codon AUG, où aura lieu l'initiation si le contexte est favorable. Lorsque l'initiation n'a pas ou peu lieu au premier AUG rencontré, on parle de *leaky scanning*. Le facteur eIF2\*GDP après éjection, échange la molécule de GDP associé pour du GTP avec l'aide de eIF2B afin de pouvoir participer, de nouveau, à l'initiation. La sous-unité 40S, positionnée au codon AUG, interagit ensuite avec le facteur eIF5B\*GTP qui permet la liaison de la sous-unité 60S. La formation du ribosome 80S déclenche l'hydrolyse du GTP en GDP. Comme le complexe eIF5B\*GDP a moins d'affinité pour le ribosome, il se dissocie, et le ribosome est alors prêt à accueillir un aa-ARNt au site A pour l'étape de l'élongation, le Met-ARNt<sub>i</sub><sup>Met</sup> occupant le site P.

Environ 10% des messagers eucaryotes, principalement ceux codant pour des facteurs de croissance, des suppresseurs de tumeurs, des facteurs de transcription, des proto-oncogènes et plusieurs ARN viraux contiennent une région 5' non-traduite très structurée qui peut contenir un ou plusieurs petits cadres de lecture en amont de la région codante principale. Ces messagers utilisent des modes alternatifs d'initiation de la traduction. Parmi ces modes, on retrouve le *leaky scanning*, où l'on n'initie pas la traduction au premier codon initiateur, mais au codon initiateur ayant un contexte favorable, en aval du premier (revue par Kozak, 1999). L'initiation interne est un autre

mode d'initiation permettant à la sous-unité ribosomique 40S de se lier directement à une région à l'intérieur du messager nommée site d'initiation interne du ribosome (*IRES*) et contenant le codon initiateur du messager (Holcik et al., 2000; revue par Jackson et Kaminski, 1995). La ré-initiation, quant à elle, se produit lorsqu'un ribosome traduit un court cadre de lecture, demeure associé au messager après la terminaison de cette traduction, amorce un balayage du messager et initie à nouveau la traduction lorsqu'il rencontre un codon initiateur (revue par Hinnebusch, 1997; Morris et Geballe, 2000; Ryabova et al., 2002). Le saut du ribosome (*shunt*), est un cas particulier de l'initiation de la traduction, où un ribosome ayant amorcé un processus normal de balayage coiffe-dépendante, évite de balayer une région du messager en se transférant d'un site dit donneur à un site accepteur sur le messager (revue par Ryabova et al., 2002).

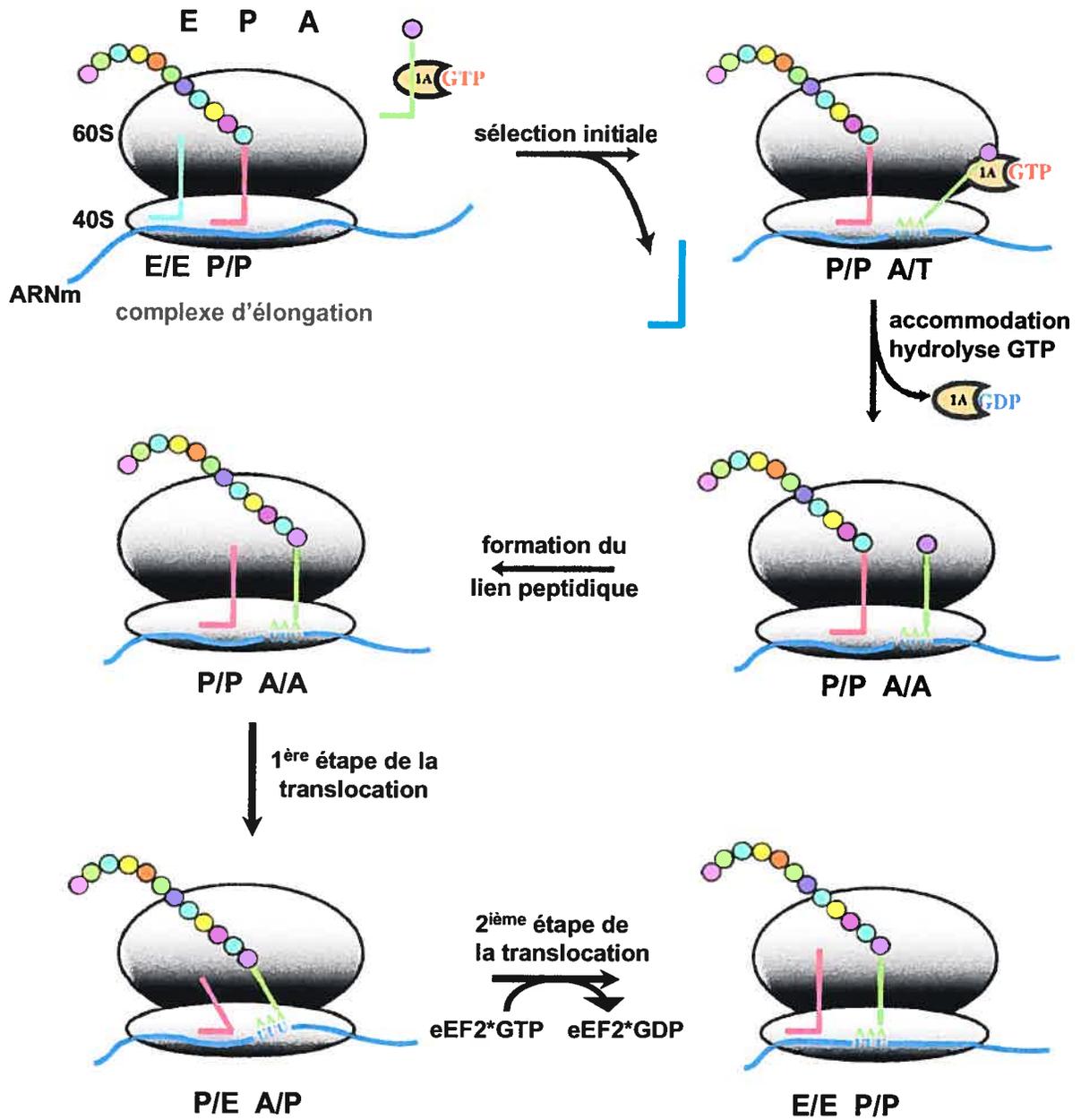
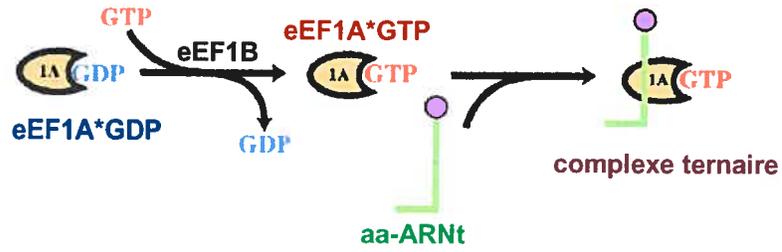
Chez la bactérie, le processus d'initiation requiert la présence de seulement trois facteurs d'initiation, IF1, IF2 et IF3, qui se lient à la petite sous-unité 30S. Brièvement, la petite sous-unité 30S, portant l'ARNt<sup>Met</sup> initiateur chargé d'une méthionine formylée et les différents facteurs d'initiation, dont IF2 associé à une molécule de GTP, se lie directement au niveau d'un codon d'initiation AUG. Cette liaison est contrôlée par un appariement entre une séquence Shine-Dalgarno (SD) riche en purines, située 5 à 13 nucléotides en amont du codon initiateur de l'ARNm, et une région complémentaire riche en pyrimidines, localisée à l'extrémité 3' de l'ARNr 16S de la sous-unité 30S, la séquence anti Shine-Dalgarno (Shine et Dalgarno, 1974). La grande sous-unité, la 50S, vient s'associer ensuite à la petite sous-unité 30S pour former le ribosome 70S. Il y a hydrolyse du GTP et départ des facteurs d'initiation (revue par Laursen et al., 2005).

#### 1.2.2.2 Élongation

Une fois l'initiation complétée, le ribosome est en mesure de répéter de façon quasi imperturbable l'incorporation d'acides aminés pour former une protéine suivant les règles du processus d'élongation. L'étape d'élongation du cycle de traduction emploie une machinerie cellulaire hautement conservée chez les eucaryotes, les bactéries et les

archéobactéries (Ramakrishnan, 2002; Spahn et al., 2001). Tel que représenté à la Figure 1-8, l'étape d'élongation chez les eucaryotes s'amorce avec un ribosome portant le Met-ARNt<sup>Met</sup> au site P, et le site A vacant. Le site P sera occupé par l'ARNt chargé de la chaîne peptidique en croissance durant l'élongation. L'aa-ARNt suivant est amené au ribosome par le complexe ternaire : eEF1A\*GTP\*aa-ARNt, eEF1A étant un facteur d'élongation homologue au facteur bactérien EF-Tu. Tous les complexes ternaires possèdent la capacité de se lier au site A/T du ribosome. Toutefois, par un processus nommé la sélection initiale, le ribosome vérifie la géométrie de l'appariement codon-anticodon entre l'ARNm et l'aa-ARNt. Seul un appariement codon-anticodon correct provoque une série de changements conformationnels dans le ribosome, stabilisant la liaison de cet aa-ARNt au site de décodage de la petite sous-unité. Ces changements sont communiqués à eEF1A dans le complexe ternaire, qui hydrolyse alors le GTP associé, quitte le ribosome libérant l'aa-ARNt qui peut occuper le site A/A par un processus nommé accommodation (Rodnina et Wintermeyer, 2001). Le facteur eEF1A\*GDP, aidé d'un facteur d'échange guanine-nucléotide, eEF1B, remplace son GDP par un GTP, et peut alors participer à nouveau à la formation d'un complexe ternaire. Suite à l'accommodation de l'aa-ARNt, la formation du lien peptidique a alors lieu instantanément entre la chaîne en croissance portée par le peptidyl-ARNt au site P et l'acide aminé qui vient d'être amené au site A. Cette réaction est catalysée par le centre peptidyle transférase faisant partie de la grande sous-unité. Ceci entraîne la désacétylation du peptidyl-ARNt au site P/P, alors que la chaîne peptidique se retrouve sur l'ARNt au site A/A. La translocation est le prochain processus. D'abord, il y a déplacement de la tige acceptrice des ARNt sur la grande sous-unité, l'ARNt déchargé est alors au site hybride P/E et le peptidyl-ARNt au site hybride A/P. Ensuite, un mouvement de la petite sous-unité relativement à la grande sous-unité permet le déplacement des ARNt, d'un triplet en direction 3' de l'ARNm, et positionne le peptidyl-ARNt au site P/P et l'ARNt déchargé du site E/E (Spahn et al., 2004). Cette étape est catalysée par eEF2, et l'énergie nécessaire à ce déplacement est fournie par l'hydrolyse du GTP. Il a été suggéré que la présence de l'ARNt déchargé au site E est importante pour le maintien du cadre de

Figure 1-8: Schéma représentant les étapes du mécanisme de l'élongation de la traduction chez les eucaryotes. L'élongation implique l'incorporation d'acides aminés dans la chaîne protéique en croissance correspondant à la séquence définie par les triplets successifs de l'ARNm. Ce processus est contrôlé par des protéines nommées facteurs d'élongation eucaryote (eEF) (adapté de Kapp et Lorsch, 2004).



lecture par le ribosome et prévient le glissement du peptidyl-ARNt au site P en interagissant avec ce dernier (Nierhaus, 1990; Noller et al., 2002). L'ARNt déchargé quitte probablement le site E lorsque l'aa-ARNt suivant est accommodé au site A/A. Après la translocation, le ribosome possède un site A libre et est prêt à recevoir le prochain complexe ternaire pour amorcer un autre cycle d'élongation.

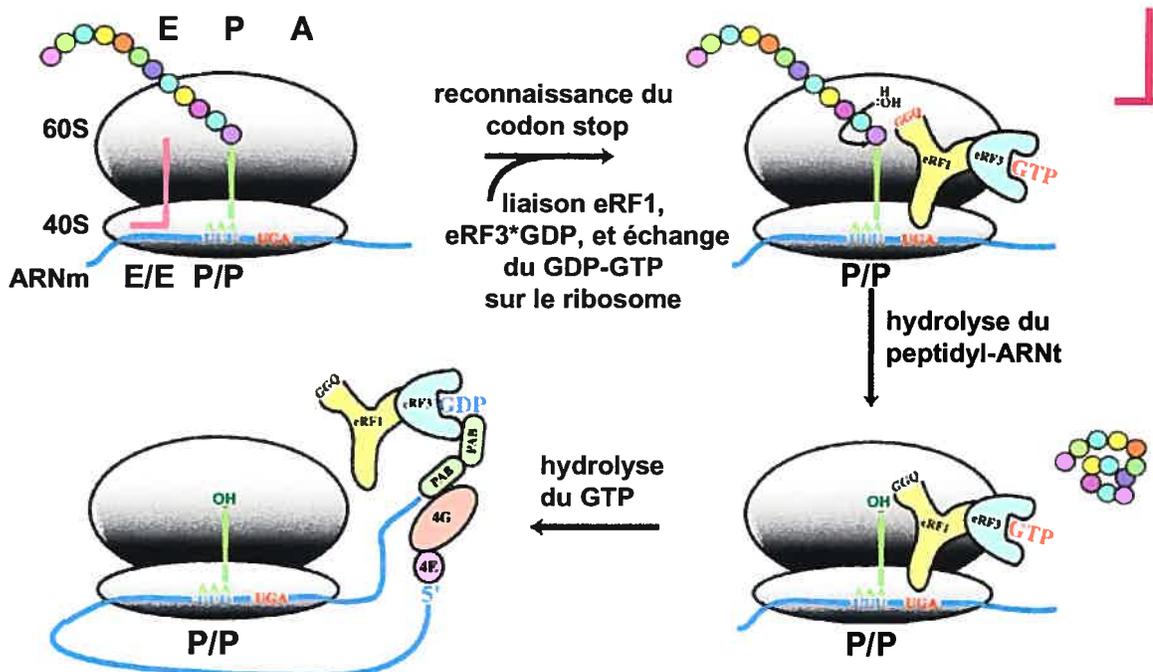
Comment le ribosome fait-il pour assurer la précision de la traduction? C'est principalement dû à la configuration de son site de décodage. Ce site, tel que mentionné ci-haut, permet au ribosome de reconnaître la géométrie de l'appariement codon-anticodon et de discriminer spatialement les appariements incorrects des appariements corrects, dans un processus nommé la sélection initiale. Les bases A1492, A1493 et G530 (selon la nomenclature de l'ARNr 16S de *E. coli*, et correspondant à A1823, A1824 et G861 de l'ARNr 18S de la levure *S. cerevisiae*, respectivement) sont universellement conservées et assurent cette tâche (Ogle et al., 2001). Lors de la liaison de l'aa-ARNt à l'ARNm au site A/T, ces bases changent d'orientation et de position en direction du duplex d'ARN formé par l'interaction codon-anticodon. Dans ces conditions, la base A1493 vérifie le sillon mineur de la première paire de bases du duplex codon-anticodon, alors que les bases A1492 et G530 s'occupent de l'interaction de la seconde paire de bases. Si ces bases détectent un bon appariement codon-anticodon, il s'en suit une série de changements conformationnels, transmis principalement à travers l'aa-ARNt, menant à l'activation de l'activité GTPase du facteur eEF1A qui hydrolyse alors le GTP qui lui est associé. On remarque que l'interaction de la troisième paire de bases du duplex codon-anticodon est moins contrôlée par le ribosome, en accord avec l'hypothèse de Wobble qui avait suggéré qu'un ARNt peut décoder un codon en formant un appariement non-canonique avec la troisième base de ce codon, formant par exemple une paire de bases G-U (Crick, 1966). Le ribosome possède un second mécanisme de correction, nommé *proofreading*, pendant l'accommodation de l'aa-ARNt du site A/T au site A/A. Durant cette transition, l'aa-ARNt a une seconde possibilité d'être rejeté après l'hydrolyse du GTP mais avant d'être accommodé et que l'acide aminé soit ensuite incorporé dans la chaîne peptidique en croissance.

### 1.2.2.3 Terminaison

La terminaison de la traduction se produit en réponse à la présence d'un codon stop, tel UGA, UAG et UAA, au site A du ribosome (revue par Nakamura et Ito, 2003). Tel que représenté à la Figure 1-9, la terminaison entraîne la relâche de la chaîne peptidique suite à l'hydrolyse du lien ester reliant cette chaîne au peptidyl-ARNt au site P. Cette réaction est catalysée par le centre peptidyle transférase de la grande sous-unité. Lors de l'étape de terminaison, le facteur de terminaison eRF1, qui possède la capacité de reconnaître les trois codons de terminaison, se lie au site A du ribosome où il interagit avec un second facteur de terminaison, eRF3 lié au GDP. La fonction de eRF1 est de permettre l'accès d'une molécule d'eau au site peptidyle transférase, ce qui va déclencher l'hydrolyse du lien peptidique entre la protéine et l'ARNt. La fonction de eRF3 demeure inconnue. Toutefois, en se basant sur le modèle de la terminaison de la traduction chez la bactérie, la relâche de la chaîne peptidique favorisée par eRF1 pourrait permettre l'échange du GDP de eRF3 en GTP. L'hydrolyse subséquente du GTP par eRF3, qui possède une activité GTPase, déstabiliserait l'interaction entre eRF1 et eRF3 et le ribosome.

Les codons de terminaison devraient être considérés comme des quadruplets et non des triplets (Bertram et al., 2001; Bonetti et al., 1995; Major et al., 1996; Tate et al., 1995). Des études statistiques ont démontré un fort biais pour le nucléotide immédiatement en 3' du codon stop chez plusieurs organismes. Chez l'homme, le codon UAAG est le plus abondant, alors que UAAA est le moins présent. Chez la levure, UAAG et UGAG sont les plus abondants. Chez *E. coli*, UAAU et UAAG sont hyper présents. L'analyse de ces sites suggère que le biais de la position 3' du codon stop favorise l'ordre suivant : G>U>A>C (G étant le plus efficace) (Poole et al., 1995; Tate et al., 1996). Même si les codons de terminaison les plus fréquemment utilisés diffèrent entre les bactéries et les eucaryotes, on remarque un fort biais pour l'absence du nucléotide C en 3' du codon de terminaison pour tous les organismes (Brown et al., 1990a, b). Le nucléotide en +1 du codon stop influence ainsi la terminaison, probablement en influençant l'efficacité avec laquelle le facteur de terminaison reconnaît le codon stop. Il a d'ailleurs été montré que le facteur de terminaison

Figure 1-9: Schéma représentant les étapes du mécanisme de la terminaison de la traduction chez les eucaryotes. La terminaison permet la relâche de la chaîne protéique en croissance suite au décodage d'un codon de terminaison par le ribosome. Ce processus est assuré par des protéines nommées facteurs de terminaison eucaryote (eRF) (adapté de Kapp et Lorsch, 2004).



bactérien RF2 pouvait être ponté au nucléotide en +1 du codon stop avec une efficacité variant en fonction de la nature de ce nucléotide (Poole et al., 1998).

#### 1.2.2.4 Recyclage du ribosome

La quatrième et dernière étape de la traduction est le recyclage des sous-unités ribosomique qui sont ensuite réutilisées dans un autre cycle de traduction. À la fin de la terminaison, le ribosome est bloqué sur l'ARNm avec un ARNt déchargé au site P. Chez la bactérie, il existe un facteur de recyclage du ribosome (RRF) permettant la dissociation du complexe de terminaison (Karimi et al., 1999). Chez les eucaryotes, les facteurs impliqués ainsi que le processus de dissociation des complexes de terminaison demeurent inconnus. En fait, il ne semble pas exister de facteur homologue au RRF bactérien, et le processus de recyclage pourrait être différent chez les eucaryotes. La nature circulaire des ARNm eucaryotes suggère que la terminaison et le recyclage pourrait ne pas libérer la sous-unité 40S dans le cytoplasme après la terminaison. La sous-unité 40S serait transportée à travers la queue de polyA, de l'extrémité 3' vers l'extrémité 5' du même messenger, via les facteurs associés aux extrémités. Ce modèle favoriserait donc la ré-initiation de la traduction du même messenger. Une interaction entre eRF3 et PABP a d'ailleurs été montrée (Hoshino et al., 1999).

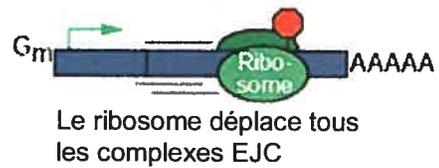
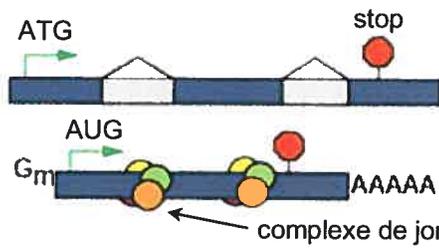
#### 1.2.3 Contrôle de la qualité des ARNm porteurs de codons non-sens (NMD)

Le nombre de précurseurs d'ARNm (pré-ARNm) différents produits par les cellules de mammifères est estimé à plus de  $10^5$ . Ces pré-ARNm doivent subir une maturation permettant l'élimination des introns par épissage avec ligation des exons codants qui vont être traduits par les ribosomes. Ce processus complexe est effectué par le *spliceosome*, un complexe moléculaire formé de cinq petits ARN nucléaires (snARN) et plus d'une centaine

de protéines (revue par Burge, 1999; Krainer, 1997). Les organismes eucaryotes ont développé des mécanismes de contrôle de la qualité des ARNm, ce qui permet de reconnaître et dégrader les ARNm ayant subi une maturation incomplète ou ne codant pas pour la protéine désirée. Ces contrôles empêchent la synthèse de protéines potentiellement délétères. Un de ces mécanismes de contrôle de la qualité est la dégradation des ARNm porteurs de codons non-sens (NMD). Cette voie de surveillance de l'ARNm détecte et dégrade les ARNm contenant un codon de terminaison prématuré (PTC) (revue par Hentze et Kulozik, 1999; Maquat, 2000; Singh et Lykke-Andersen, 2003). Différents groupes d'ARNm sont la cible du NMD, principalement ceux qui proviennent de gènes ayant une mutation créant soit un codon de terminaison ou un déphasage causant une terminaison prématurée de la traduction, et ceux qui proviennent de gènes ayant acquis un PTC à cause d'une erreur d'épissage ou de transcription. Le mécanisme par lequel la cellule distingue un PTC d'un codon normal de terminaison a fait l'objet d'intenses recherches, et nous avons maintenant un modèle détaillé de ce mécanisme chez les mammifères (selon Hillman et al., 2004). Tel que présenté à la Figure 1-10, lors de l'épissage, le *spliceosome* élimine les introns et laisse comme marqueur sur l'ARN messager un complexe protéique de jonction exon-exon (EJC) environ 20 nucléotides en aval de l'endroit où l'intron a été enlevé. Le complexe EJC comprend des facteurs d'épissage, des facteurs d'exportation de l'ARNm, ainsi que des protéines impliquées dans le NMD comme les protéines Upf. Lors de son exportation dans le cytoplasme, l'ARNm subit une première ronde de traduction par le ribosome qui déplace alors les complexes EJC rencontrés durant son parcours. Avec les messagers normaux, qui possèdent un codon de terminaison dans le dernier exon, le ribosome déplace tous les complexes EJC. Cependant, en présence d'un PTC, un complexe EJC demeure en aval du ribosome arrêté au codon de terminaison. Une série d'interactions entre le complexe EJC et le ribosome, impliquant les protéines Upf, entraîne alors la dégradation de cet ARNm. Expérimentalement, il a été observé qu'un codon de terminaison situé à plus de 50 nucléotides en amont du dernier exon est prématuré et cause le NMD.

Figure 1-10: Modèle de la dégradation des ARNm porteurs d'un codon non-sens (NMD).  
(a) Mécanisme de contrôle de la qualité des ARNm chez les eucaryotes. (b) La présence d'un codon de terminaison suffisamment en amont de la dernière jonction exon-exon déclenche le NMD et la dégradation du messenger (adapté de Hillman et al., 2004).

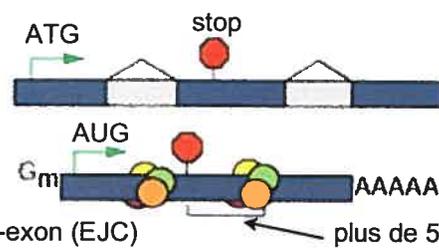
(a) Codon de terminaison dans le dernier exon



Le ribosome déplace tous les complexes EJC

**Plusieurs rondes de traduction normale**

(b) Codon de terminaison prématuré



Le ribosome ne déplace pas tous les complexes EJC



**Dégradation de l'ARNm par la voie du NMD**

Des études controversées ont proposé que les protéines Upf du complexe de surveillance NMD étaient aussi impliquées dans le maintien du cadre de lecture (Cui et al., 1996; Dinman et al., 2000; Ruiz-Echevarria et al., 1998). En effet, ces études ont montré que la délétion de l'une des protéines Upf, Upf3, entraînait une augmentation de l'efficacité du décalage de phase de lecture en  $-1$  d'environ trois fois chez la levure, avec un rapporteur dont l'expression dépend du déphasage (voir section 1.3.4 sur le décalage du cadre de lecture en  $-1$ ). Toutefois, des études du même groupe ont ensuite révélé que l'augmentation du déphasage était plutôt causée par une augmentation de deux à trois fois de la quantité de messagers rapporteurs produits lorsque les protéines Upf étaient délétées (Harger et Dinman, 2004). Ces dernières études confirment une proposition initiale de Stahl et al. (2000), énonçant que les protéines Upf ne sont pas impliquées dans le maintien du cadre de lecture.

Le rôle du signal de décalage de phase de lecture en  $-1$  dans la dégradation des ARNm a d'ailleurs été mis en évidence récemment suite à des études utilisant la levure. Plant et Dinman (2005) ont introduit différents sites de décalage de phase de lecture dans la séquence codante d'un ARNm de façon à ce que les ribosomes effectuant un tel déphasage rencontrent rapidement un codon de terminaison considéré comme prématuré. Ils ont démontré que plus l'efficacité du déphasage en  $-1$  était élevée, plus les messagers étaient dégradés rapidement via le NMD.

### **1.3 Reprogrammation de la traduction (*recoding*)**

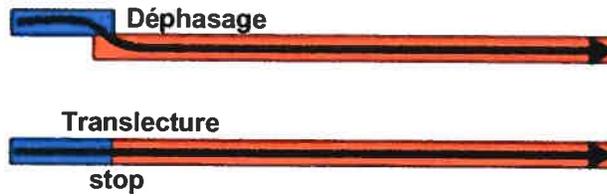
Les règles classiques de la traduction décrites dans tout manuel de biologie moléculaire depuis l'identification du code génétique, il y a plus de 50 ans, sont simples : un ARN messager transcrit à partir d'ADN et contenant un codon initiateur et un codon de terminaison permet la synthèse d'une protéine. Toutefois, la nature sophistiquée de la traduction rend difficile la prédiction de la séquence de toutes les protéines produites à partir de séquences d'ADN. En fait, le code génétique n'est pas universel, et chez certains

organismes ou dans une population de gènes, la signification de codons particuliers a été réassignée. Par exemple, chez la mitochondrie, le codon UGA ne spécifie par un codon de terminaison mais bien un codon tryptophane (Barrell et al., 1979). La présence de signaux à l'intérieur des ARNm peut influencer le décodage du code génétique et fausser les mécanismes régissant la synthèse des protéines. C'est un événement que l'on nomme la reprogrammation (Gesteland et al., 1992). Les virus, dont la réplication dépend de l'appareil de traduction cellulaire, utilisent fréquemment la reprogrammation. Ils sont ainsi passés maîtres dans l'art de pousser le ribosome à l'erreur. Lors de la traduction, la reprogrammation peut se produire durant l'élongation (décalage de phase de lecture en +1 ou -1, saut du ribosome) ou lors de la terminaison (translecture d'un codon de terminaison) (revue par Atkins et al., 2001; Baranov et al., 2002a; Brierley et Pennell, 2001; Gesteland et Atkins, 1996; Stahl et al., 2001a). Cette reprogrammation est souvent en compétition avec le fonctionnement normal de traduction, et seulement une certaine proportion des ribosomes effectue une reprogrammation à un site précis de ces messagers. La reprogrammation a plusieurs rôles fonctionnels importants. Tel que présenté à la Figure 1-11, elle est impliquée dans le contrôle de la synthèse de certaines protéines ou la synthèse de deux protéines dans un rapport précis (par décalage de phase de lecture ou la translecture). Elle permet aussi l'incorporation d'acides aminés particuliers (la sélénocystéine et la pyrrolysine) au cours de la traduction. La majorité des événements de translecture et de déphasage a été identifiée chez les virus et les éléments de transpositions, mais ces événements peuvent également se produire avec le messenger de certains gènes cellulaires.

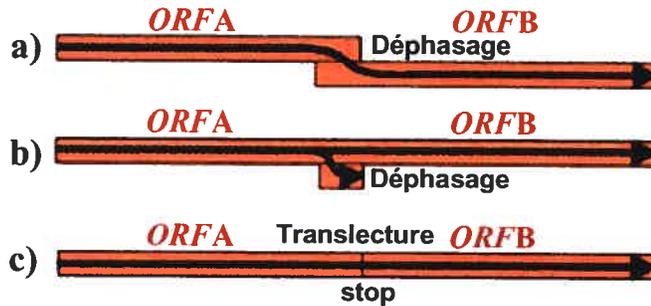
Le VIH-1 utilise le décalage de phase de lecture en -1 pour la production de ses enzymes lors de sa réplication et ce déphasage a fait l'objet de nos recherches. Nous allons donc consacrer une section entière aux cas de déphasage en -1, à l'analyse des signaux *cis* de l'ARNm le favorisant, ainsi qu'à l'examen des modèles proposés pour expliquer le déphasage. Finalement, nous verrons comment la reprogrammation permet la régulation de l'expression de gènes cellulaires et de la réplication virale. Pour le moment, nous examinerons quelques cas de reprogrammation au cours de la traduction.

Figure 1-11: Rôles fonctionnels de la reprogrammation. (A) Contrôle de la synthèse d'une protéine par déphasage ou translecture. La reprogrammation peut se produire au début du processus d'élongation et permettre de compléter la synthèse de la protéine (le bleu indique la portion du messenger avant la reprogrammation). (B) Contrôle de la synthèse de deux protéines dans un rapport précis. La protéine ORFA et la protéine de fusion ORFA-ORFB, sont produites par déphasage ou translecture. Le déphasage permet la production d'une protéine plus longue, ORFA-ORFB (a), ou plus courte, ORFA (b), que la protéine résultant de la traduction conventionnelle. La translecture est nécessaire à la production d'un produit plus long, ORFA-ORFB (c). (C) Incorporation d'un acide aminé particulier. SECIS indique le signal permettant l'incorporation de la sélénocystéine à un codon de terminaison (jaune). PYLIS indique le signal hypothétique permettant l'incorporation de la pyrrolysine à un codon de terminaison (jaune). (adapté de Baranov et al., 2002a).

(A) Contrôle de la synthèse d'une protéine

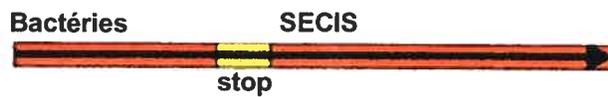


(B) Contrôle de la synthèse de deux protéines dans un rapport précis



(C) Incorporation d'un acide aminé particulier

Sélénocystéine



Pyrrolysine

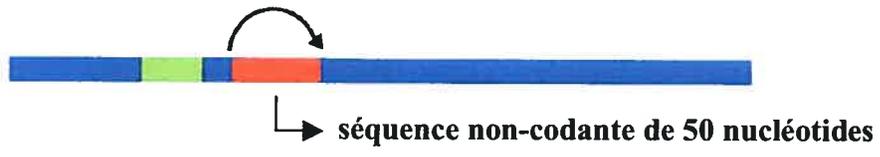


### 1.3.1 Saut du ribosome (*hopping*)

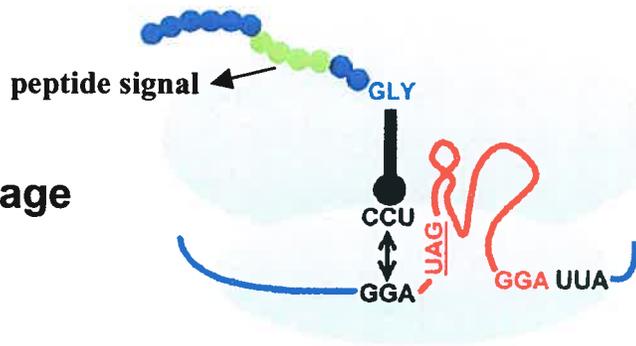
#### 1.3.1.1 Exemples de sauts du ribosome

Le saut du ribosome lui permet d'éviter de décoder une portion de l'ARN messager. Selon une description simplifiée, le saut du ribosome se produit en trois étapes : le décollage, le balayage, et l'atterrissage. Lors de cet événement, le peptidyl-ARNt porté par un ribosome, dont le site A est vacant, se détache de l'ARNm (décollage). Par la suite, le ribosome parcourt le messager à la recherche d'un codon qui pourrait s'apparier au peptidyl-ARNt (balayage). Lorsqu'il rencontre ce codon, le peptidyl-ARNt se ré-apparie à l'ARNm (atterrissage), et un aa-ARNt occupe ensuite le site A du ribosome, qui peut continuer la synthèse protéique selon les règles conventionnelles. Les premiers exemples de saut du ribosome ont montré que le peptidyl-ARNt pouvait balayer le messager sur cinq à six nucléotides et n'avait besoin pour ce saut que d'une complémentarité entre le codon au site de décollage et au site d'atterrissage, d'une part, et l'anticodon du peptidyl-ARNt, d'autre part (Weiss, 1987). Par exemple, le séquençage de protéines révéla qu'une seule asparagine ou leucine était incorporée pour les séquences codantes AAC UCA AU et CUU UAG CUA (où les espaces indique le cadre de lecture zéro et où les codons soulignés correspondent au site de décollage et d'atterrissage). Ces séquences indiquant un saut de cinq ou six nucléotides par le ribosome, respectivement. C'est toutefois l'identification d'un saut du ribosome lors de la traduction du gène 60 du bactériophage T4 qui déclencha le plus d'enthousiasme et d'intérêt. Dans ce cas, les ribosomes font un saut de 50 nucléotides entre le codon 46 et le codon numéroté 47, pour synthétiser une sous-unité de la topoisomérase (Herr et al., 2000; Herr et al., 2001; Huang et al., 1988). L'efficacité du saut est de 50%, mais celle-ci dépend, en plus de la complémentarité entre l'anticodon du peptidyl-ARNt et le codon des sites de décollage et d'atterrissage, de plusieurs signaux agissant en *cis* (décrits plus bas). Le saut du messager du gène 60 est sans contredit le cas le plus étudié, et celui dont on connaît le mécanisme le plus en détail (voir Figure 1-12).

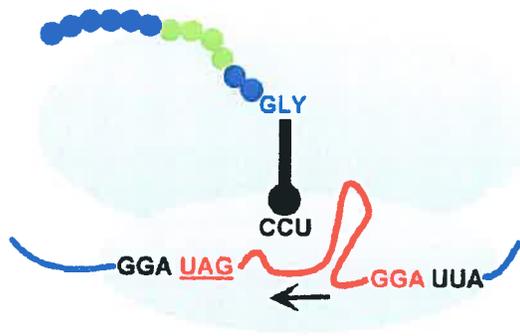
Figure 1-12: Modèle du saut du ribosome lors de la traduction du messenger du gène *60* du bactériophage T4. Lors de la traduction, le peptidyl-ARNt<sup>Gly</sup> se détache du codon GGA (décollage), parcourt l'ARNm à la recherche d'une séquence complémentaire (balayage), et se ré-apparie au site GGA pour reprendre l'élongation (atterrissage). Plusieurs signaux sont requis dans l'ARNm pour réaliser ce saut (adapté de Herr et al., 2000).



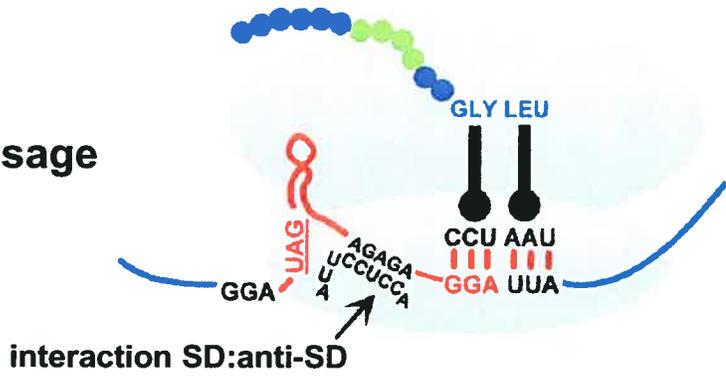
**Décollage**



**Balayage**



**Atterrissage**



Selon le modèle admis, le ribosome, dont le peptidyl-ARNt<sup>Gly</sup> se détache du site de décollage GGA, balaye 50 nucléotides de séquences non-codantes de l'ARNm et reprend l'élongation au site d'atterrissage GGA suite au ré-appariement du peptidyl-ARNt<sup>Gly</sup> au messenger. Ce saut peut être supprimé avec des mutants de l'ARNt<sup>Gly</sup> ou des mutants de la protéine L9 de la grande sous-unité ribosomique (Herr et al., 2001). Le saut du ribosome du gène 60 est le seul saut de cette envergure confirmé expérimentalement. Quelques autres cas non démontrés semblent toutefois probables : la traduction du messenger du gène codant pour l'adhésine de *Prevotella loescheii* pourrait nécessiter un saut de 29 nucléotides (Manch-Citron et al., 1992; Manch-Citron et al., 1999a; Manch-Citron et al., 1999b; Manch-Citron et London, 1994), et il a été suggéré qu'un faible pourcentage de ribosomes sautent durant le décodage du codon de terminaison du messenger du gène codant pour la bêta-hémoglobine du lapin (Chittum et al., 1998).

### 1.3.1.2 Signaux *cis* de l'ARN messenger favorisant le saut du ribosome

De nombreuses études de mutagenèse dirigée portant sur le gène 60 ont mené à l'identification de plusieurs éléments de l'ARNm agissant en *cis* permettant au saut de se produire avec une haute efficacité (revue par Herr et al., 2000). L'un de ces éléments est un codon de terminaison situé immédiatement en 3' du site de décollage GGA. La mutation du codon de terminaison UAG en UGG ou UAC réduit l'efficacité du saut de 10 et 50 fois, respectivement (Weiss et al., 1990). La présence du codon stop UAG servirait à ralentir ou arrêter le ribosome. Des études indépendantes, effectuées avec un système rapporteur ont montré que l'efficacité du saut était beaucoup plus efficace lors de conditions empêchant l'occupation du site A par l'aa-ARNt, ce qui montre le lien entre la pause du ribosome et l'efficacité du saut (Gallant et Lindsley, 1998). Un autre élément *cis* est une structure secondaire en tige-boucle formée en 5' de la région balayée. La présence de cette structure stimule l'efficacité du saut d'environ 30 fois. Son rôle serait d'empêcher l'accès du facteur de terminaison RF1 au codon stop qui suit le site de décollage. Il a également été suggéré qu'elle favorise la dissociation du peptidyl-ARNt de l'ARNm au site P. Bien que ceci soit

purement spéculatif, il est possible que cette structure ne soit pas détruite lors du balayage de la région non-codante. La longueur de la séquence non-codante est importante mais on ignore pour quelle raison. La séquence de la chaîne peptidique en croissance influence également l'efficacité du saut. Des mutations changeant l'identité des acides aminés 17 à 35 de la chaîne protéique en croissance, qui sont majoritairement chargés positivement, polaires et hydrophobes, diminuent de 140 fois l'efficacité du saut. Il a été suggéré que ce signal déstabilise l'interaction codon-anticodon au site de décollage par son interaction avec des protéines ou de l'ARNr dans le tunnel par où chemine la protéine en croissance. Finalement, la séquence AGAGA située cinq nucléotides en amont du site GGA d'atterrissage semble aussi importante. Cette séquence pourrait favoriser l'appariement codon-anticodon au site d'atterrissage par une interaction de type Shine-Dalgarno avec l'ARNr 16S. Mais cette suggestion n'a toutefois pas été démontrée.

### 1.3.2 Redéfinition d'un codon

Le contexte nucléotidique entourant le codon de terminaison peut mener le ribosome à une gymnastique lui permettant de contourner les règles normales de la traduction. Dans certains cas, ces contextes, favorisent l'incorporation d'un aa-ARNt à ce codon stop, ce qui permet la synthèse d'une protéine plus grande au détriment de sa terminaison. La redéfinition du codon de terminaison n'est pas valable pour tous les gènes de l'organisme mais a lieu seulement pour certains gènes. Comme nous le verrons au cours de cette section, il existe deux types d'événements permettant la redéfinition d'un codon de terminaison : la translecture, permettant l'incorporation d'un acide aminé normal faisant partie des 20 acides aminés rencontrés habituellement dans les protéines, comme la glutamine ou le tryptophane, l'incorporation d'acides aminés particuliers comme la sélénocystéine et la pyrrolysine. L'incorporation de la sélénocystéine nécessite la présence de facteurs *trans* particuliers contrairement aux autres cas de translecture. Le mécanisme permettant l'incorporation de la pyrrolysine est encore mystérieux.

### 1.3.2.1 Translecture (*readthrough*)

La translecture est une erreur de terminaison : lorsqu'un ribosome en élongation, portant un peptidyl-ARNt au site P rencontre un codon stop au site A, il arrive qu'il accepte un aa-ARNt au site A et catalyse l'incorporation de l'aa dans la chaîne peptidique en croissance. Il s'agit en fait d'une erreur puisque généralement, l'anticodon de cet aa-ARNt s'apparie au codon stop par seulement deux bases sur trois. Cependant, le ribosome commet à peine une erreur sur 2000 codons lus car, comme nous l'avons vu, il vérifie l'appariement codon-anticodon au site A et possède un mécanisme de confirmation de lecture (*proofreading*). La translecture accidentelle a donc une très faible efficacité. Mais il existe une translecture programmée, qui peut atteindre des efficacités de 5 à 10%. Comme nous le verrons, les processus de translecture programmé sont influencés par des séquences adjacentes au codon stop. Cependant, dans certains cas, les organismes possèdent en faible abondance des ARNt, tel l'ARNt<sup>Gln</sup><sub>UCA</sub>, nommés supprimeurs de codon de terminaison, dont l'anticodon s'apparie parfaitement à un codon de terminaison. Dans ce cas, la translecture ne résulte pas d'une erreur et son efficacité dépend de la compétition entre les facteurs de terminaison et les ARNt supprimeurs.

#### 1.3.2.1.1 Exemples de translecture

Les cas de translecture sont très rares lors de la traduction de gènes cellulaires. Chez les bactéries, la synthèse du pilus CS3 responsable de l'entérotoxicité de la souche *E. coli* CFA/II nécessite la translecture à un codon UGA et la présence de l'ARNt<sup>Gln</sup><sub>UCA</sub> supprimeur (Jalajakumari et al., 1989). Le gène *topo-3* de *Bacillus firmus* codant pour une sous-unité de l'ADN topoisomérase I semble aussi nécessiter un événement de translecture car son cadre de lecture est interrompu par un codon de terminaison UGA (Ivey et al., 1992). L'authenticité de cette translecture demeure toutefois spéculative, dû au manque de données expérimentales et au peu de conservation du codon stop de ce gène chez les bactéries. Des études sur le développement de la drosophile ont montré une translecture

pour au moins trois gènes : *kelch* (Xue et Cooley, 1993), *oaf* (Bergstrom et al., 1995), *hdc* (Steneberg et al., 1998). Il y a translecture à un codon UGA dans le cas de *kelch* et à un codon UAA pour *oaf* et *hdc*. Ces gènes sont impliqués dans le développement, et des mutations dans *kelch* causent la stérilité chez les mouches femelles. Toutefois, des études supplémentaires seront requises afin d'identifier les signaux favorisant la translecture du messager de ces gènes et les acides aminés incorporés aux sites de translecture. La translecture se produit aussi lors de la traduction du gène *PDE2* de la levure *S. cerevisiae*, qui code pour une phosphodiesterase liant et dégradant l'AMP cyclique (Namy et al., 2002). Chez plusieurs rétrovirus, tel le virus de la leucémie murine (MuLV), les gènes *gag* et *pol* sont dans le même cadre de lecture mais séparés par un codon stop et la translecture permet la synthèse du précurseur Gag-Pol, le précurseur des enzymes du virus, alors que la traduction conventionnelle génère Gag, le précurseur des protéines structurales du virus. Le virus de la mosaïque du tabac (TMV) est un des nombreux virus de plantes utilisant aussi la translecture.

#### 1.3.2.1.2 Signaux cis de l'ARNm favorisant la translecture

Plusieurs groupes ont montré que la translecture d'un codon de terminaison était influencée par la séquence des nucléotides entourant ce codon de terminaison chez les ARNm procaryotes ou eucaryotes (Buckingham, 1994; Kopelowitz et al., 1992; Phillips-Jones et al., 1995). Ces contextes peuvent être simplement la séquence des nucléotides un à six du côté 3' du codon stop ou impliquer des signaux plus complexes en 3', comme des structures en pseudonoeud (c'est-à-dire une tige-boucle dans laquelle une portion de la boucle coiffant cette première tige s'apparie à une séquence complémentaire sur l'ARNm, adjacente à cette tige, formant ainsi une seconde tige).

Une comparaison de séquences de plusieurs ARN viraux de plantes et d'animaux ayant un codon stop permettant la translecture a démontré que les séquences

immédiatement en 3' de ce codon stop présentait des homologies (Beier et Grimm, 2001). Certains virus de plantes, tel TMV, possèdent une séquence consensus CAR YYA (où R est une purine et Y est une pyrimidine) du côté 3' des codons stop de translecture UAG ou UAA. Des mutations à chaque position de cette séquence suppriment la translecture (Valle et al., 1992; Zerfass et Beier, 1992). Une autre séquence homologue consensus CGG ou CUA a été identifiée chez plusieurs virus de plantes et d'animaux immédiatement en 3' d'un codon UGA (Beier et Grimm, 2001). Des études chez le virus du *rattle* du tabac (*tobacco rattle virus*, TRV) ont montré que la mutation de deux ou trois nucléotides de la séquence CGG abolissait la translecture (Urban et al., 1996). Ces études de comparaison de séquences de sites de translecture soulignent l'importance de l'efficacité avec laquelle les « quadruplets » de terminaison sont décodés pour le contrôle de l'efficacité de la translecture. L'analyse des contextes montre en fait que la présence d'un C immédiatement en 3' du codon de terminaison favorise la translecture suite à une terminaison peu efficace à ce codon.

Un signal plus complexe est requis pour la translecture du codon stop UAG du virus MuLV. Dans ce cas, une séquence de huit nucléotides riche en purines sépare le codon UAG d'une séquence de 49 nucléotides formant une structure en pseudonoeud (Feng et al., 1992; Honigman et al., 1991). Ce signal se retrouve aussi chez plusieurs gammarétrovirus (Beier et Grimm, 2001) et des mutations altérant la structure du pseudonoeud et de la séquence séparatrice empêchent la translecture (Alam et al., 1999; Felsenstein et Goff, 1992; Feng et al., 1992; Wills et al., 1994).

Les mécanismes moléculaires selon lesquels ces séquences et structures en aval du codon stop influencent la translecture demeurent toutefois obscurs. Comme indiqué ci-dessus, il est généralement admis que tous ces signaux influencent la compétition entre les ARNt et les facteurs de terminaison pour le décodage des codons stop en favorisant la translecture. Ils stabiliseraient l'interaction codon-anticodon au site A ou interagiraient avec l'ARNr (Buckingham, 1994).

### 1.3.2.2 Incorporation de la sélénocystéine

Durant les premiers balbutiements de la biologie moléculaire, dans les années 1960, on croyait que, sur les 64 triplets possibles du code génétique, 61 spécifiaient les 20 acides aminés, et trois la fin de la séquence codante. Toutefois, en 1986, l'identification de séquences montra qu'un des codons stop (UGA) pouvait ne pas agir comme signal d'arrêt et spécifier le 21<sup>ième</sup> acide aminé, la sélénocystéine (Sec). La sélénocystéine est un acide aminé rarement utilisé et aucun codon ne lui avait été assigné. La découverte qu'un codon d'arrêt UGA pouvait correspondre à la sélénocystéine a suscité plusieurs recherches. L'incorporation de la sélénocystéine dans une chaîne protéique en croissance requiert plusieurs processus complexes. D'abord un ARNt peu abondant, ARNt<sup>Sec</sup><sub>UCA</sub>, dont la boucle anticodon UCA est complémentaire du codon stop UGA, se charge de sérine. Une sélénocystéine synthétase convertit ensuite la sérine en sélénocystéine par une réaction en deux étapes impliquant la perte d'une molécule d'eau et l'addition de sélénium, un micronutriment essentiel pour l'homme et pour plusieurs organismes (revue par Driscoll et Copeland, 2003; Hatfield et Gladyshev, 2002). Le nombre de gènes codant pour des sélénoprotéines varie selon les organismes. *E. coli* en possède trois, alors qu'on en dénombre pas moins de 30 à 50 chez les mammifères (Behne et Kyriakopoulos, 2001; Huttenhofer et Böck, 1998). Le mécanisme d'incorporation de la sélénocystéine a d'abord été résolu chez la bactérie par les travaux du groupe de Böck (Böck, 2000; revue par Krol, 2002). Chez les bactéries, tel que présenté à la Figure 1-13A, l'incorporation de la sélénocystéine à un codon UGA particulier dépend de la présence d'une structure secondaire d'ARN en tige-boucle située immédiatement en aval de ce codon. Cette tige-boucle est nommée la séquence d'incorporation de la sélénocystéine (bSECIS, avec b pour bactéries). La boucle de cette tige-boucle lie un facteur d'élongation particulier, SelB, qui lui-même s'associe à l'ARNt<sup>Sec</sup><sub>UCA</sub> chargé de la sélénocystéine. SelB est homologue au facteur d'élongation EF-Tu et son positionnement sur la tige-boucle bSECIS lui permet de placer l'ARNt<sup>Sec</sup><sub>UCA</sub> chargé de sélénocystéine au site A du ribosome. Il y a alors incorporation de la sélénocystéine à la protéine en croissance. Chez les eucaryotes, peut-

être dû au découplage de la transcription et de la traduction, la séquence SECIS (homologue de bSECIS) est localisée dans la région 3' non-traduite du messager (3'UTR). Un ou deux de ces éléments SECIS sont souvent responsables de l'incorporation de sélénocystéines à plusieurs codons UGA d'un même messager. Le facteur d'élongation eSelB (parfois nommé eEFsec) lie l'ARNt<sup>Sec</sup><sub>UCA</sub> chargé de sélénocystéine, comme chez les bactéries, mais ne reconnaît pas directement la séquence SECIS (Lescure et al., 2002). Il y a intervention d'un facteur supplémentaire, SBP2, qui se lie d'une part à la structure SECIS et interagit d'autre part avec eSelB (revue par Copeland, 2003) (voir une représentation schématique à la Figure 1-13B). Ce qui fait qu'un codon UGA est considéré comme un triplet désignant la sélénocystéine plutôt qu'un signal de terminaison demeure mystérieux. Une hypothèse suggère que l'incorporation de la sélénocystéine et l'arrêt de la synthèse protéique soit en compétition directe, et que le contexte entourant le codon stop entraîne ou prolonge une pause du ribosome, tout en rendant la terminaison inefficace. À cet effet, nous avons déjà souligné que la base suivant les codons de terminaison détermine l'efficacité de la terminaison. Plusieurs études ont d'ailleurs révélé que plus ce contexte était favorable à la terminaison, moins l'incorporation de Sec était efficace et vice-versa (Grundner-Culemann et al., 2001; McCaughan et al., 1995; Nasim et al., 2000). L'efficacité avec laquelle se produit l'incorporation de Sec aux codons UGA demeure indéterminée. Elle est probablement liée à la régulation de l'expression des sélénoprotéines, qui dépend elle-même de la disponibilité du sélénium et du recrutement de facteurs limitant tels SBP2 (revue par Driscoll et Copeland, 2003).

### 1.3.2.3 Incorporation de la pyrrolysine

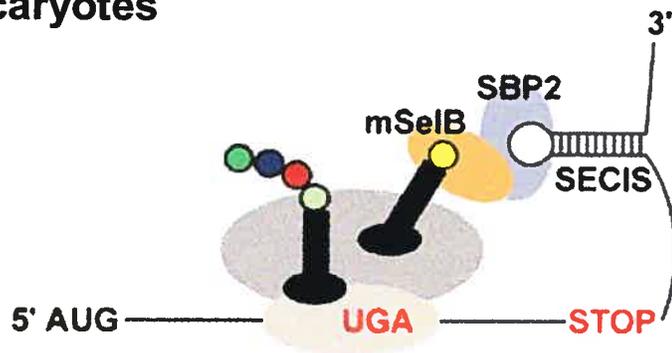
La pyrrolysine, une lysine modifiée et considérée comme le 22<sup>ième</sup> acide aminé, est spécifiée par un codon de terminaison UAG. Elle a été identifiée dans la séquence protéique d'une méthyle transférase de l'archéobactérie *M. barkeri* (Hao et al., 2002; Srinivasan et al., 2002). Une pyrrolysine forme le site actif de cette enzyme dont le rôle est

Figure 1-13: Mécanisme d'incorporation de la sélénocystéine. A) Chez les bactéries, une tige-boucle (séquence bSECIS) situé immédiatement en aval du codon de terminaison UGA interagit avec le facteur SelB qui a lié un  $\text{ARNt}^{\text{Sec}}_{\text{UCA}}$  chargé de sélénocystéine. B) Chez les eucaryotes, l'élément SECIS est situé dans la région 3'UTR et permet l'incorporation de la sélénocystéine à plusieurs codon de terminaison UGA d'un messenger. Un facteur additionnel est requis chez les eucaryotes, SBP2, qui se lie d'une part à la structure SECIS et interagit d'autre part avec eSelB portant l' $\text{ARNt}^{\text{Sec}}_{\text{UCA}}$  chargé de la sélénocystéine (adapté de Namy et al., 2004).

## A) Bacteria



## B) Eucaryotes



de catalyser la conversion de méthylamines en méthane. Le mécanisme par lequel la pyrrolysine est incorporée dans cette protéine demeure incertain (revue dans Cobucci-Ponzano et al., 2005). La présence d'un ARNt<sup>Pyl</sup><sub>CUA</sub>, c'est-à-dire un ARNt qui peut lier la pyrrolysine et dont l'anticodon UCA est complémentaire au codon stop UGA, a été montrée chez cet organisme (Polycarpo et al., 2003; Srinivasan et al., 2002). Des éléments favorisant l'insertion de la pyrrolysine (PYLIS), similaires à l'élément SECIS qui favorise l'incorporation de la sélénocystéine, peuvent être prédits à proximité du codon de terminaison (Namy et al., 2004). De plus, il a récemment été confirmé que l'ARNt<sup>Pyl</sup><sub>CUA</sub> est fonctionnel et pouvait être chargé de la pyrrolysine (Blight et al., 2004). Cependant, la présence de cette pyrrolysine dans plus de 97% des molécules de méthyle transférase est peu compatible avec un mécanisme de compétition entre l'incorporation d'un aa-ARNt et la terminaison du codon stop pour ce genre d'événements (Atkins et Gesteland, 2002). Il y aurait plutôt une réassignation de ce codon chez l'archéobactérie, comme la spécification du codon UGA pour le tryptophane chez la mitochondrie (Barrell et al., 1979). Clairement, d'autres études sont nécessaires pour éclaircir le mécanisme d'incorporation de la pyrrolysine.

### 1.3.3 Décalage de phase de lecture (*frameshift*)

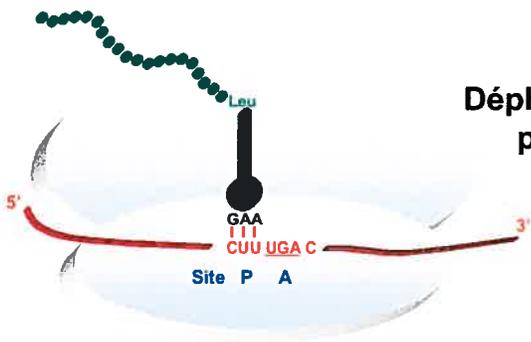
Lors du déphasage du cadre de lecture, le ribosome change le cadre de lecture de l'ARNm, relativement au cadre imposé par le codon initiateur. Le mouvement du ribosome sur l'ARNm définit le type de déphasage : si le ribosome se déplace en direction 5' par rapport au messenger, il s'agit d'un déphasage en -1, alors qu'un déplacement en direction 3', indique un déphasage en +1. Suite au déphasage, le ribosome reprend la traduction de façon conventionnelle. On considère deux groupes de déphasage, spontané ou programmé. Le décalage de phase de lecture spontané résulte d'une erreur de maintien du cadre de lecture lors de l'élongation. Cette erreur se produit à une fréquence de  $3 \times 10^{-5}$  et représente simplement les limites de la fidélité de l'appareil traductionnel (Kurland, 1992). L'erreur

spontanée n'a aucun but fonctionnel. Par contre, si ce déphasage est favorisé par des signaux particuliers du messager, on parle alors de déphasage programmé du cadre de lecture. Ce dernier peut atteindre une efficacité de 50%, soit  $10^4$  fois plus élevée que le déphasage spontané. Les sites de décalage de phase de lecture programmé se retrouvent chez tous les organismes, chez des virus et des phages infectant ces organismes ou sous forme d'éléments mobiles tels les rétrotransposons. Dans cette section nous allons nous consacrer à analyser l'événement de reprogrammation qu'est le décalage de phase de lecture en +1, et nous aborderons ensuite le décalage de phase de lecture en -1, un événement important pour les travaux de recherches que je vais vous présenter. Pour la suite de ce travail, nous allons référer au décalage de phase de lecture programmé, et non spontané, de façon implicite.

### 1.3.3.1 Décalage de phase de lecture en +1

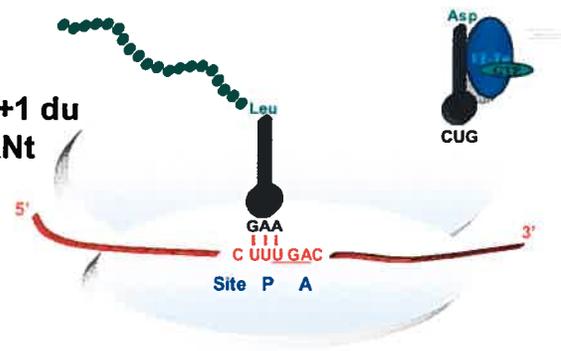
Le décalage de phase de lecture en +1 permet la synthèse d'une protéine à partir de deux cadres de lecture différents se chevauchant. Son importance physiologique a été mise en évidence lors de l'identification du mécanisme de synthèse du facteur de terminaison bactérien RF2, par exemple. L'endroit du messager où se produit le déphasage se nomme la séquence glissante. De façon générale, le décalage de phase de lecture en +1 du ribosome est causé par une pause découlant de la présence d'un codon stop ou rare (c'est-à-dire spécifié par un aa-ARNt peu abondant) au site A. À l'occasion de cette pause, le peptidyl-ARNt au site P se désapparie de l'ARNm, le ribosome portant le peptidyl-ARNt glisse dans le cadre +1, et le peptidyl-ARNt se ré-apparie à l'ARNm. Le ribosome poursuit ensuite la traduction dans le nouveau cadre de lecture +1 de façon conventionnelle (voir la Figure 1-14). La séquence où se produit le déphasage est telle que le peptidyl-ARNt peut former au moins deux appariements corrects sur trois après le déphasage. Plusieurs séquences *cis* peuvent augmenter l'efficacité avec laquelle le déphasage en +1 se produit (Baranov et al., 2002a; Namy et al., 2004; revue par Stahl et al., 2001a).

Figure 1-14: Description du décalage de phase de lecture en +1 du ribosome sur le messenger codant pour le facteur de terminaison RF2. Le déphasage est causé par une pause du ribosome en élongation lors de la rencontre du codon stop UGA au site A. Le peptidyl-ARNt au site P peut alors se désapparier de l'ARNm, se déplacer d'une base du côté 3' avec le ribosome, puis se réapparier à l'ARNm dans le cadre de lecture +1. Le complexe ternaire (EF-Tu\*GTP\*Asp-ARNt<sup>Asp</sup><sub>CUG</sub>) amène ensuite l'aa-ARNt correspondant au nouveau codon au site A.



Appariement avant déphasage

Déphasage en +1 du  
peptidyl-ARNt



Appariement après déphasage

### 1.3.3.1.1 Exemples de décalage de phase de lecture en +1

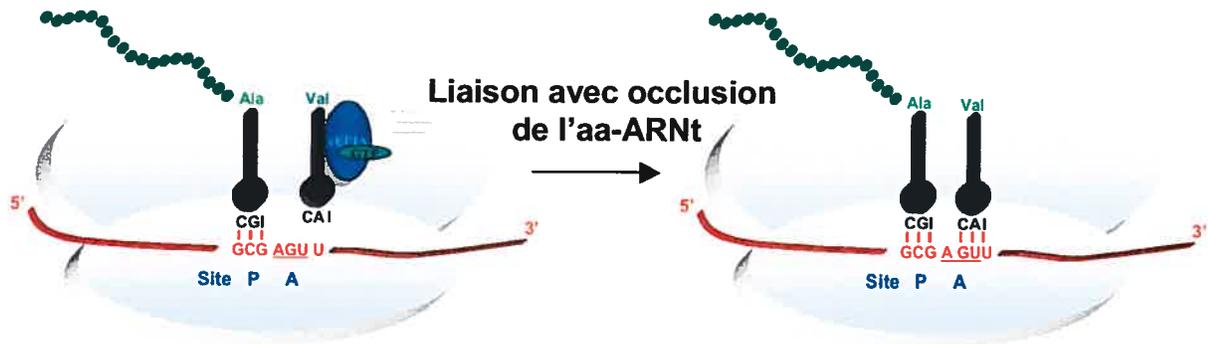
Le gène *prfB* de *E. coli*, codant pour le facteur de terminaison RF2, est sans contredit le modèle des exemples de déphasage en +1 (Craig et Caskey, 1986). Il est le seul exemple connu de déphasage en +1 chez les bactéries. RF2 déclenche la terminaison de la traduction lorsqu'un codon UAA ou UGA occupe le site A du ribosome. La séquence codante de *prfB* est interrompue par un codon stop, UGA, situé 25 codons en aval du codon initiateur et continue alors dans un cadre de lecture +1 pour 340 codons additionnels. Le déphasage en +1 se produit pour 30 à 50% des ribosomes rencontrant la séquence CUU UGA C, où les espaces indiquent le cadre zéro. Lors du déphasage, le peptidyl-ARNt<sup>Leu</sup> se désapparie du codon CUU et s'apparie au codon UUU dans le cadre +1. Ce glissement est favorisé par la présence du quadruplet de terminaison UGAC, qui est reconnu peu efficacement par le facteur de terminaison RF2 (Poole et al., 1995). Il a été montré que l'efficacité du déphasage en +1 à cet endroit dépend de la durée de la pause traductionnelle (Curran et Yarus, 1989; Donly et al., 1990). Le contrôle de la synthèse de RF2 par décalage de phase de lecture +1 constitue un cas élégant d'événement de reprogrammation régulant la synthèse d'une protéine (Adamski et al., 1993). À haute concentration de RF2, la compétition entre le décalage de phase de lecture +1 et la terminaison de la traduction est déplacée en faveur de la terminaison, ce qui mène à une diminution de la synthèse et donc de la concentration de RF2 dans la cellule. À faible de concentration de RF 2, le déphasage en +1 prédomine, augmentant le niveau de RF2 dans la cellule. Ce mécanisme est simple, efficace et sensible.

Plusieurs cas de déphasage en +1 ont été identifiés chez les eucaryotes. Le plus connu est celui qui a lieu lors de la traduction de l'ARNm de l'antizyme (AZ1) de l'ornithine décarboxylase (ODC), qui est l'enzyme limitante de la voie de synthèse des polyamines (Matsufuji et al., 1995; Matsufuji et al., 1990). L'antizyme est responsable de la dégradation de l'ODC par la voie ubiquitine-protéasome (Murakami et al., 1992). Le déphasage en +1 de l'antizyme se produit avec une efficacité d'environ 20% et contrôle

l'expression de l'antizyme chez plusieurs organismes comme les mammifères, la drosophile, le vers *C. elegans*, et les levures (Ivanov et al., 1998a; Ivanov et al., 2000a). Cette efficacité est influencée par la quantité de polyamines dans la cellule. Deux autres gènes AZ2 et AZ3, paralogues d'AZ1, ont également été identifiés chez les mammifères. AZ2 est exprimé de façon ubiquitaire comme AZ1, mais en moindre abondance (Ivanov et al., 1998b), tandis que AZ3 est exprimé spécifiquement dans certains tissus lors de la spermatogenèse (Ivanov et al., 2000b). AZ2 et AZ3, tout comme AZ1, sont exprimés via un déphasage. L'antizyme constitue avec RF2 un autre cas intéressant où il y a régulation de la synthèse d'une protéine par le déphasage en +1. À haute concentration de polyamines, le niveau de déphasage en +1 est augmenté, ce qui favorise la synthèse de l'antizyme qui, en retour, dégrade l'ornithine décarboxylase (ODC). Comme l'ODC catalyse l'étape limitante de la voie de biogenèse des polyamines, sa dégradation mène à la réduction du niveau de polyamines cellulaires, ce qui réduit l'efficacité du déphasage en +1 et la synthèse de l'antizyme (Ivanov et al., 2000a). Le site de déphasage des trois antizymes des mammifères est conservé. Il comprend le codon de terminaison UGA dans le contexte suivant : UCC UGA U, qui permet le désappariement du peptidyl-ARNt<sup>Ser</sup> du codon UCC et son réappariement au codon CCU dans le cadre +1. Chez la levure, les gènes *EST3*, *ABP140*, et les rétrotransposons de la famille Ty utilisent aussi un déphasage en +1 pour leur expression. Le gène *EST3* est responsable de la synthèse d'une des quatre sous-unités de la télomérase, dont la présence est essentielle pour le maintien des télomères, alors que *ABP140* code pour une protéine liant les filaments d'actine. Il a été démontré que le déphasage au site de *EST3* se produit à la séquence CUU AGU U (Morris et Lundblad, 1997) et celui de *ABP140* (Asakura et al., 1998) et du rétrotransposon Ty1 a lieu à la séquence CUU AGG C (Belcourt et Farabaugh, 1990). Le mécanisme par lequel le glissement du ribosome sur le message de *EST3*, *ABP140* et Ty1 se produit est similaire à celui de *prfB* et implique le déphasage en +1 du peptidyle-ARNt<sup>Leu</sup> du codon CUU au codon UUA en +1, ce qui est favorisé par la présence du codon rare arginine (AGG) ou sérine (AGU). Par contre, le déphasage en +1 du rétrotransposon Ty3 est un cas controversé. Il a été suggéré que son site de glissement, GCG AGU U, ne permet pas le

glissement au site P du peptidyl-ARNt<sup>Ala</sup><sub>CGI</sub> du codon GCG au codon CGA en +1. En effet, un seul appariement codon-anticodon sur trois est possible pour le peptidyl-ARNt dans le nouveau cadre +1. Il a alors été proposé que le déphasage en +1 résultait de l'incorporation au site A de l'aa-ARNt<sup>Val</sup>, dont l'anticodon est 3'CAI<sup>5'</sup>, au codon GUU du cadre +1 sans lecture de l'A suivant le codon GCG (Figure 1-15) (Farabaugh et al., 1993; Li et al., 2001; Stahl et al., 2001b). Il s'agirait de l'occlusion d'une base. Comme nous le verrons plus bas, à la section 1.3.3.1.3 traitant des mécanismes et modèles du décalage de phase de lecture en +1, si nous utilisons les récentes données structurales analysant le contrôle de la fidélité de la sélection des aa-ARNt au site A du ribosome bactérien, ce mécanisme d'occlusion d'une base lors de la traduction est peu probable (Hansen et al., 2003).

Figure 1-15: Modèle de liaison avec occlusion de l'aa-ARNt permettant le décalage de phase de lecture en +1 du ribosome lors de la traduction du rétrotransposon Ty3. Selon le modèle proposé par le groupe de Farabaugh (Farabaugh et al., 1993; Li et al., 2001; Stahl et al., 2001b), le déphasage au site de Ty3 serait causé par l'incorporation de l'aa-ARNt<sup>Val</sup><sub>CAI</sub> directement dans le cadre de lecture +1 au site A du ribosome, sans qu'il y ait glissement du peptidyl-ARNt au site P. L'aa-ARNt<sup>Val</sup><sub>CAI</sub> est introduit au site A, et s'apparie au codon GUU du cadre +1, sans qu'il y ait lecture de l'A précédant GUU.



**Appariement avant déphasage**

**Appariement après déphasage**

### 1.3.3.1.2 Signaux cis de l'ARNm favorisant le décalage de phase de lecture en +1

Comme nous venons de le mentionner, le déphasage en +1 nécessite la présence d'un codon stop ou rare ralentissant la traduction et d'une séquence permettant le ré-appariement du peptidyl-ARNt au site P dans le cadre de lecture +1. Toutefois, des signaux stimulateurs additionnels situés en 5' ou 3' du site de changement de cadre influencent l'efficacité du déphasage. Lorsque le stimulateur est situé en 5' du site de glissement, il exerce son effet via une interaction complémentaire entre sa séquence primaire et l'ARNr, tandis que lorsqu'il est situé en 3', il peut soit agir de la même façon ou soit former une structure secondaire qui interagirait avec le ribosome.

Le premier cas identifié de signal stimulateur en 5' est une séquence de type Shine-Dalgarno (SD), riche en purines, située en amont du site de déphasage du gène *prfB* (RF2) (Weiss, 1987; Weiss et al., 1988). Il a été démontré que le déphasage en +1 est stimulé par un appariement entre cette séquence située seulement trois nucléotides en amont du site de glissement, et la séquence complémentaire anti-SD riche en pyrimidines de l'ARNr 16S du ribosome (Figure 1-16A). La position de cette séquence par rapport au site de glissement est critique. Rappelons ici que l'interaction SD:anti-SD, généralement située de cinq à treize nucléotides en amont du codon initiateur AUG, permet à la petite sous-unité ribosomique de localiser ce codon et d'initier la traduction chez la bactérie (Chen et al., 1994; Shine et Dalgarno, 1974). Au site de déphasage de *prfB*, la distance entre la séquence de type SD et la séquence glissante est plus courte que ce qui est observée pour ce type d'interaction lors de l'initiation. Cette interaction SD:anti-SD durant l'élongation créerait une tension entre la région anti-SD et le centre de décodage du ribosome, qui favoriserait un déplacement du ribosome d'un nucléotide en direction 3' du messager. Un tel déplacement allègerait la tension créée par l'interaction SD:anti-SD (Atkins et al., 2001; Baranov et al., 2002b). Une seconde explication, qui n'exclut pas la première, serait que l'interaction SD:anti-SD interfère avec l'éjection de l'ARNt déchargé à partir du site E. Ceci empêcherait la sortie de l'ARNt déchargé, et ce dernier, en interagissant avec le

Figure 1-16: Exemples de signaux stimulateurs du décalage de phase de lecture en +1. A. Le déphasage en +1 lors de la traduction de l'ARNm du gène *prfB* est stimulé par un appariement entre une séquence de type Shine-Dalgarno (SD), située trois nucléotides en amont du site de glissement, et une séquence complémentaire anti-SD de l'ARNr 16S du ribosome. B. Au site de déphasage des messagers des gènes de l'antizyme, l'efficacité du déphasage en +1 est stimulée par la présence d'une structure secondaire en pseudonoeud, située trois nucléotides en aval du site de glissement.



peptidyl-ARNt, pourrait le pousser vers le site A, favorisant ainsi le déphasage de ce dernier dans le cadre de lecture +1 (Baranov et al., 2002b). Une autre interprétation assez différente est que cette interaction, SD:anti-SD, pourrait, au contraire, favoriser le départ prématuré de l'ARNt déchargé du site E, ce qui diminuerait le maintien du cadre de lecture par le ribosome et donnerait plus de mobilité au peptidyl-ARNt (Marquez et al., 2004). Ce type d'interaction SD:anti-SD n'est pas unique au cas de déphasage en +1. Comme nous le verrons plus loin, lorsqu'elle est positionnée différemment par rapport au site de glissement, elle peut aussi favoriser le déphasage en -1. Un deuxième exemple de stimulateur situé en 5' a été identifié chez l'antizyme. Une séquence CCG GGG CCU CGG, située six nucléotides en amont du site de glissement, augmente l'efficacité du déphasage en +1 de l'antizyme d'un facteur d'environ deux et demi (Matsufuji et al., 1995). Cette séquence est conservée chez les trois antizymes des mammifères mais son mécanisme d'action demeure inconnu.

Les deux stimulateurs du déphasage en +1 situés en 3' identifiés à ce jour exercent leurs fonctions de façon très différente. Le premier et le seul cas connu de déphasage en +1 impliquant une structure secondaire en 3' est le cas des messagers de l'antizyme. Ce stimulateur est un pseudonoeud, situé trois nucléotides en aval du site de glissement, qui favorise le déphasage en +1 (Figure 1-16B) (Matsufuji et al., 1995). Sa délétion réduit l'efficacité du déphasage en +1 d'un facteur 30 dans les cellules de mammifères. Comment ce pseudonoeud favorise le déphasage en +1 demeure nébuleux. Il est probable que sa présence à proximité du site A du ribosome bloque l'accès des facteurs de terminaison au codon stop, ce qui augmente la probabilité de glissement du peptidyl-ARNt au détriment de la terminaison. Il a été montré dans les cas de déphasage en -1, que ce type de signal, favorise une pause du ribosome (Somogyi et al., 1993; Tu et al., 1992), et nous pouvons penser que la même situation se rencontre avec ce pseudonoeud. Des études intéressantes de Matsufuji et al. (1996) sur le site de glissement de l'antizyme chez la levure *S. cerevisiae* ont montré que ce pseudonoeud favorise un déphasage de cadre en -2 (à 90% du temps) plutôt qu'un déphasage en +1 (10% du temps), alors qu'on rencontre seulement un

déphasage en +1 chez les mammifères. Cependant, des études similaires chez la levure *S. pombe* avec la même séquence ont montré que 80% du temps, les ribosomes effectuaient un déphasage de cadre en +1, et un déphasage en -2 seulement 20% du temps (Ivanov et al., 1998c). Une augmentation de la distance entre le pseudonoeud et le site de glissement de trois nucléotides favorise le déphasage en +1 au détriment du déphasage -2 chez *S. cerevisiae*. Ce changement dans la direction du déphasage observé chez la levure suggère que le pseudonoeud est bien plus qu'une simple barrière causant une pause du ribosome, et qu'il influence aussi la direction du déphasage.

Un second cas de signal stimulateur situé en 3' a également été identifié au site de glissement du rétrotransposon Ty3 de la levure *S. cerevisiae*. Dans ce cas, une séquence de 14 nucléotides située immédiatement en aval du site de glissement augmente l'efficacité du déphasage en +1 d'un facteur dix (Farabaugh et al., 1993; Li et al., 2001). Ce segment ne forme pas de structure secondaire. Son effet dépend de sa séquence primaire et de sa position vis-à-vis du site de glissement. Il a été suggéré que cette séquence s'apparie à une séquence complémentaire de la boucle de l'hélice 18 de l'ARNr de la petite sous-unité ribosomique. L'analyse des données cristallographiques du ribosome bactérien a révélé une interaction entre les nucléotides C518 et G530 de la boucle coiffant l'hélice 18 et l'ARNm et l'anticodon de l'ARNt au site A, montrant que ces bases ont un rôle dans le contrôle de la fidélité de lecture (Ogle et al., 2001). Farabaugh et son groupe suggèrent que l'interaction entre le stimulateur de Ty3 et la boucle de l'hélice 18 perturbe les interactions entre l'ARNm et l'aa-ARNt, ce qui permet au ribosome d'accepter des aa-ARNt non-correct. Ceci entraîne des erreurs de traduction et cause le déphasage en +1.

#### 1.3.3.1.3 Mécanismes et modèles du décalage de phase de lecture en +1

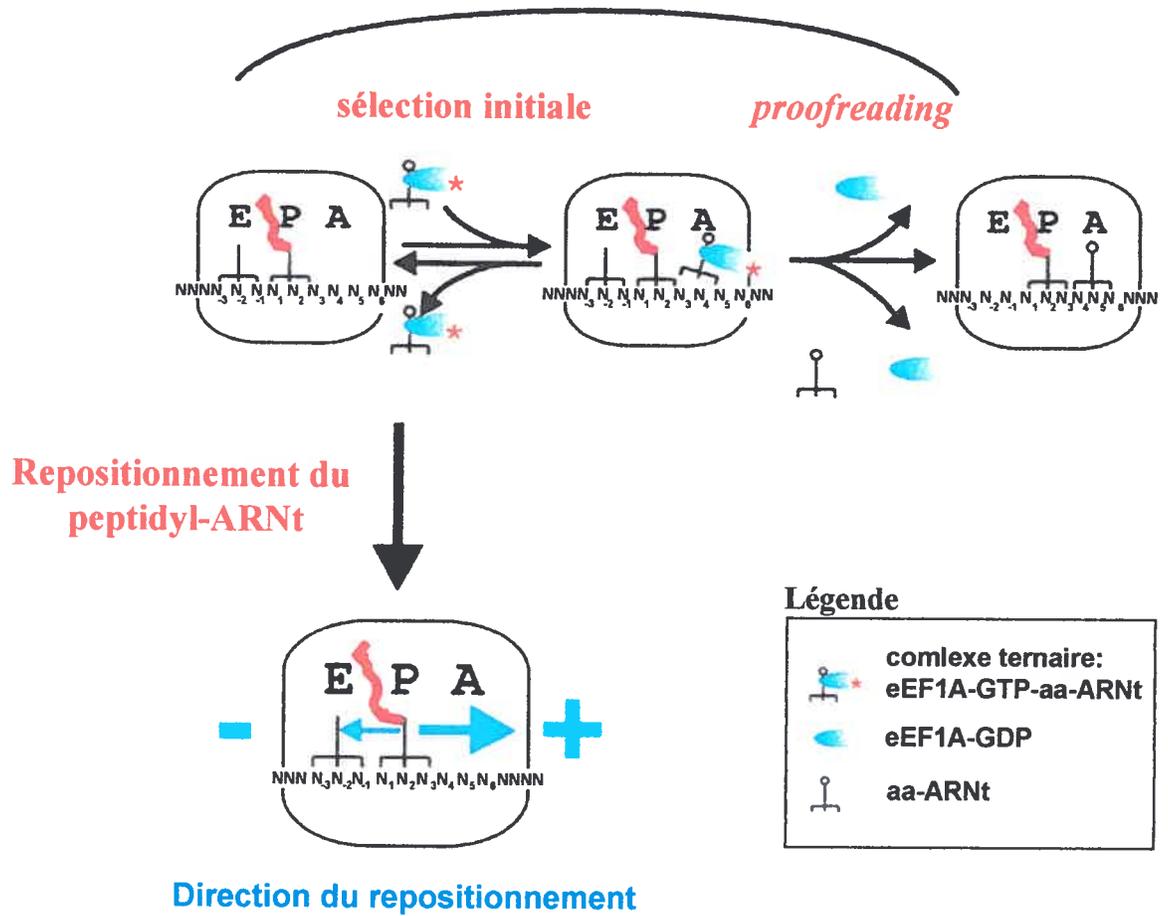
Plusieurs modèles ont été proposés pour expliquer comment et à quelle étape du cycle de l'élongation se produit le déphasage en +1 (Farabaugh et Bjork, 1999; Harger et

al., 2002; Weiss et al., 1989). Le modèle récent de Baranov et al. (2004) suggère que le déphasage est dicté par le repositionnement du peptidyl-ARNt au site P, qui impose le cadre de lecture où se placera l'aa-ARNt suivant au site A lors de la traduction. Ce concept permet également d'expliquer le déphasage en +1 ou -1, mais dans cette section, nous allons discuter du déphasage en +1.

Le modèle, présenté à la Figure 1-17, s'applique à la situation où le codon suivant le dernier codon lu dans le cadre zéro est un codon rare. Selon le modèle, le repositionnement du peptidyl-ARNt au site P se produit avant l'acceptation irréversible de l'aa-ARNt au site A/A. Le déphasage se produit s'il y a rejet à la sélection initiale de l'aa-ARNt au site A/T, et durant l'étape de correction (*proofreading*), précédant l'occupation du site A. En effet, dès que l'aa-ARNt est accommodé au site A/A, le peptidyl-ARNt au site P ne peut plus se déplacer dans la direction +1 puisque l'aa-ARNt au site A occupe cet endroit. Le repositionnement du peptidyl-ARNt implique, nous l'avons vu, une dissociation de l'interaction codon-anticodon, un déplacement du ribosome portant le peptidyl-ARNt d'un nucléotide en direction +1 par rapport au messager, et le ré-appariement du peptidyl-ARNt avec l'anticodon de l'ARNm dans le nouveau cadre de lecture +1. Le prochain aa-ARNt incorporé est donc spécifié par le codon du messager dans le nouveau cadre +1. Logiquement, un repositionnement du ribosome portant le peptidyl-ARNt peut se produire dans les deux directions. Cependant, cette direction dépendra de l'appariement possible entre l'anticodon du peptidyl-ARNt et le codon dans le nouveau cadre. De plus, dans le cas du glissement dans la direction en -1, la présence de l'ARNt déchargé au site E peut interférer avec le glissement (Baranov et al., 2002b; Horsfield et al., 1995). Toutefois, si le peptidyl-ARNt peut se dissocier du site P et se repositionner dans le cadre +1, il peut aussi se dissocier du cadre +1 et revenir dans le cadre initial zéro avant qu'un aa-ARNt soit irréversiblement incorporé au site A. Nous comprenons aisément que le mouvement du peptidyl-ARNt au site P dépend ainsi de la disponibilité des aa-ARNt spécifiés par les différents cadres de lecture. Le modèle peut-être analysé de la même façon lorsque le codon suivant le dernier codon dans le cadre zéro est un codon de terminaison.

Figure 1-17: Modèles du décalage de phase de lecture en +1. Le modèle de Baranov et al. (2004) suggère que le repositionnement du peptidyl-ARNt au site P se produit durant les étapes du cycle d'élongation où l'aa-ARNt peut être rejeté (c'est-à-dire soit durant la sélection initiale de l'aa-ARNt au site A/T ou durant l'étape de *proofreading*), avant l'occupation du site A. Le déphasage est dicté par le repositionnement du peptidyl-ARNt au site P lorsque le site A est inoccupé si le réappariement dans le nouveau cadre est possible (adapté de Baranov et al., 2004).

L'aa-ARNt peut être rejeté avant d'être accommodé au site A/A



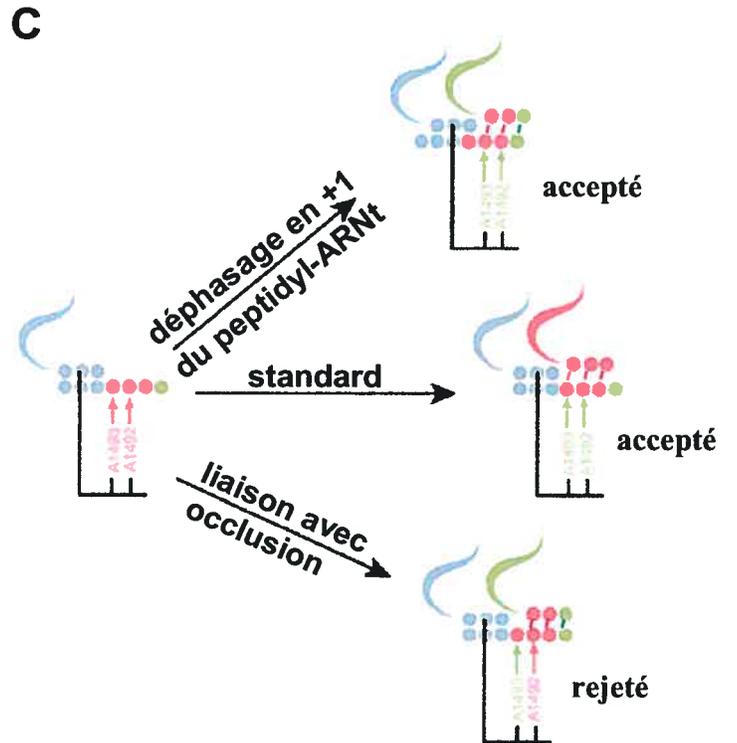
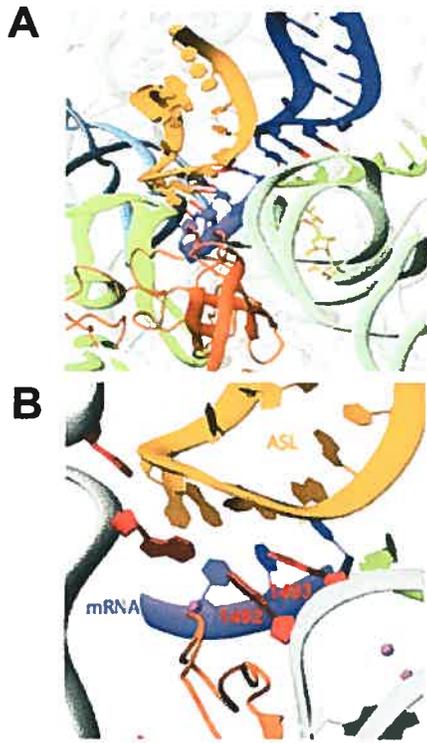
Un aspect important du modèle de Baranov et al. est l'influence de la stabilité de l'interaction codon-anticodon du peptidyl-ARNt au site P sur l'efficacité du déphasage. Comme nous l'avons précédemment remarqué, l'appariement des deux premières positions du duplex codon-anticodon au site A est mieux vérifiée que la géométrie de la troisième position par le ribosome, pour discriminer les aa-ARNt corrects des aa-ARNt non-corrects (Ogle et al., 2003; Ogle et Ramakrishnan, 2004; revue par Ramakrishnan, 2002). Par exemple, même si le codon UUU peut former un appariement de même stabilité avec les anticodons  $3'AAG^{5'}$  ou  $3'GAA^{5'}$ , seul  $3'AAG^{5'}$  est le « bon » anticodon pour le codon UUU. Par contre, on ignore s'il y a un contrôle de la géométrie de l'interaction codon-anticodon au site P. En fait, si l'identité du peptidyl-ARNt au site P était contrôlée, son rejet terminerait la synthèse protéique. Dans le cas de l'initiation de la traduction, où la reconnaissance du codon AUG de l'ARNm se produit au site P, il a été montré que les codons GUG, UUG et AUU, permettent également l'initiation par l'ARNt<sup>Met</sup> chez les bactéries. Ceci suggère donc que le ribosome ne contrôle pas l'appariement au site P aussi rigoureusement qu'il le fait au site A. On peut considérer que l'efficacité du repositionnement du peptidyl-ARNt au site P dépend de la stabilité de l'interaction codon-anticodon dans le cadre de lecture initial versus la stabilité de l'interaction codon-anticodon dans le nouveau cadre de lecture. Moins la différence énergétique est grande, plus il est probable que le peptidyl-ARNt au site P change de cadre de lecture.

L'efficacité du repositionnement du peptidyl-ARNt au site P dépend également des facteurs influençant la présence de l'aa-ARNt correspondant au site +1. Une faible concentration de l'aa-ARNt spécifiant le cadre +1 favorise le décodage normal, tandis qu'une concentration élevée augmente l'efficacité du déphasage. De la même manière, une faible affinité pour le peptidyl-ARNt pour le codon du cadre +1 favorise la traduction normale, tandis qu'une haute affinité augmente le déphasage. En fait, plusieurs évidences expérimentales supportent ces propos. Par exemple, au site de glissement CUU AGG C du rétrotransposon Ty1, la surexpression de l'ARNt<sup>Arg</sup><sub>UCC</sub>, qui reconnaît le codon AGG du cadre de lecture initial au site A et qui est normalement peu abondant, réduit l'efficacité du

déphasage en +1 de 40 fois (Belcourt et Farabaugh, 1990). De plus, la délétion de cet ARNt<sup>Arg</sup><sub>UCC</sub> augmente l'efficacité du déphasage en +1 (Kawakami et al., 1993). De la même façon, la surexpression de l'ARNt<sup>Gly</sup><sub>CCG</sub> reconnaissant le codon GGC spécifié par le cadre +1 augmente l'efficacité du déphasage en +1 (Pande et al., 1995). Ces expériences confirment donc que la concentration des aa-ARNt est critique pour le maintien du cadre de lecture. De nouveau, si le codon décodé au site A par le ribosome est un codon stop, l'occupation de ce site par un facteur de terminaison va influencer l'efficacité du déphasage.

Le modèle de liaison avec occlusion a été proposé par le groupe de Farabaugh pour expliquer le déphasage en +1 au site de glissement du rétrotransposon Ty3, car le peptidyl-ARNt au site P ne peut former un bon appariement dans le cadre +1 (Farabaugh et al., 1993; Li et al., 2001; Stahl et al., 2001b), comme nous l'avons vu plus haut. Dans ce modèle, l'aa-ARNt est directement introduit au site A dans le cadre +1, sans que le peptidyl-ARNt soit repositionné au site P (Figure 1-15). Toutefois, le contrôle de la géométrie du duplex codon-anticodon au site A par le ribosome n'appuie pas ce modèle. Comme nous l'avons vu, les nucléotides A1492 et A1493 (selon la nomenclature de l'ARNr 16S de *E. coli*, et correspondant à A1823 et A1824 de l'ARNr 18S de la levure *S. cerevisiae*) vérifient les deux premières paires de bases du duplex codon-anticodon au site A, et sont positionnés dans une orientation déterminée par rapport au site P. Dans le modèle de liaison avec occlusion, la première base du codon au site A est donc non-appariée, ce qui implique que le nucléotide A1493 ne peut vérifier l'interaction codon-anticodon. Le fait que le ribosome ne détecte pas d'appariement à cet endroit devrait entraîner le rejet de l'aa-ARNt reconnaissant le codon +1 par le ribosome (détails dans la Figure 1-18) (Hansen et al., 2003).

Figure 1-18: Mécanisme par lequel les ribosomes empêchent la liaison avec occlusion d'un aa-ARNt (dans le cadre +1). A. Détails des sites A et P du centre de décodage du ribosome. L'ARNt au site A (jaune) et P (bleu). B. Agrandissement de l'interaction codon-anticodon au centre de décodage montré en A. Les nucléotides A1492 et A1493 vérifient l'interaction codon-anticodon du messenger et de l'ARNt. C. Illustration de la liaison d'un ARNt avec occlusion au site A. La liaison avec occlusion (+1) de l'ARNt ne permet pas une reconnaissance normale de l'ARNt par les nucléotides A1492 et A1493. (les section A et B présentent la structure cristallographique correspondante du ribosomes bactérien par diffraction aux rayons X) (adapté de Hansen et al., 2003).



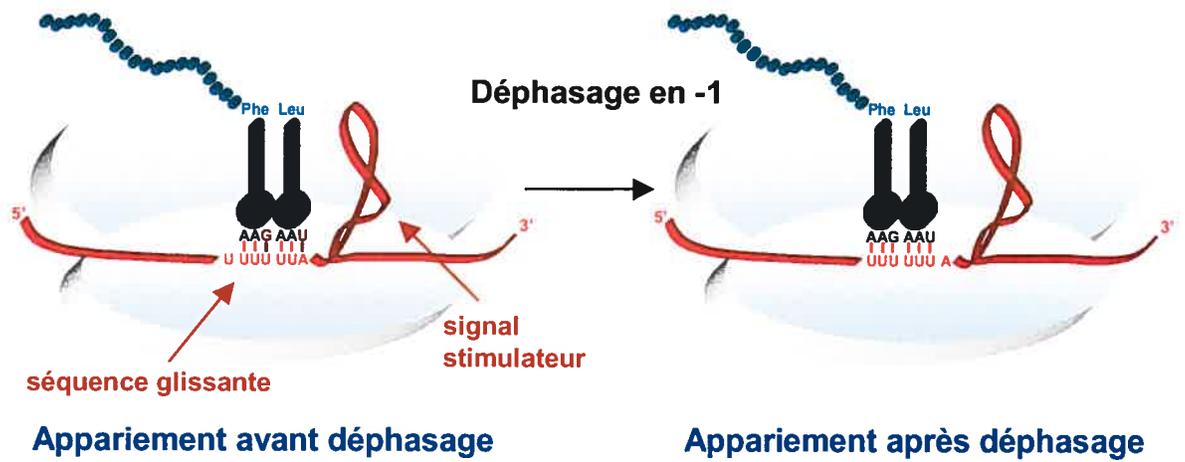
### 1.3.4 Décalage de phase de lecture en $-1$

Le décalage de phase de lecture en  $-1$  permet la synthèse d'une protéine à partir de deux cadres de lecture différents se chevauchant : le premier cadre de lecture situé dans la phase initiale de lecture, dite zéro, et le second cadre de lecture situé en aval du changement de cadre, dite en phase  $-1$ . De façon générale, tel que présenté à la Figure 1-19, le décalage de phase de lecture en  $-1$  est causé par la présence de deux éléments dans l'ARNm : une séquence glissante de type heptamérique, X XXY YYZ (les espaces indiquent le cadre de lecture zéro), où se produit le changement de cadre de lecture et une structure secondaire d'ARN situé quelques nucléotides en aval, dont la présence augmente l'efficacité du déphasage. Lorsque les ribosomes atteignent la séquence glissante X XXY YYZ, une certaine proportion change de cadre et se retrouve dans le nouveau cadre de lecture  $-1$  (XXX YYY Z). La nature répétitive de la séquence glissante permet aux anticodons des ARNt de maintenir au moins deux appariements corrects sur trois avec les codons du messenger après le déphasage en  $-1$ . La synthèse protéique à partir de deux cadres de lecture se chevauchant permet au produit du cadre de lecture  $-1$  d'être fusionné au produit du cadre de lecture zéro pour former une nouvelle protéine. Celle-ci sera soit plus longue soit plus courte que le produit synthétisé par les ribosomes qui ont continué la traduction conventionnelle dans le cadre zéro. La compétition entre le déphasage et la traduction conventionnelle permet en fait la synthèse de deux protéines de taille différente dans un rapport précis.

#### 1.3.4.1 Quelques exemples de décalage de phase de lecture en $-1$

Les exemples de décalage de phase de lecture en  $-1$  sont beaucoup plus abondants que les cas de décalage de type  $+1$ . Le premier exemple de déphasage en  $-1$  a été identifié chez le rétrovirus du sarcome de Rous (RSV). Jacks et Varmus (1985) ont démontré que la

Figure 1-19: Représentation du décalage de phase de lecture en -1 du ribosome. Le déphasage est causé par une pause du ribosome en élongation. La pause se fait au niveau de la séquence glissante et est causée par la présence d'un signal stimulateur situé quelques nucléotides en aval. À ce moment, le peptidyl-ARNt au site P et l'aa-ARNt au site A se détachent de l'ARNm, le ribosome portant les deux ARNt se déplace d'un nucléotide en direction 5' par rapport au messenger, puis les deux ARNt se ré-apparient avec l'ARNm dans le nouveau cadre de lecture -1.



synthèse de la protéine Gag-Pol de RSV nécessitait un déphasage en  $-1$  dans une région de l'ARN viral où se chevauche les gènes *gag* et *pol*. Des sites de déphasage en  $-1$  ont ensuite été identifiés chez plusieurs autres rétrovirus (HTLV-I, HTLV-II, VIH-1, MMTV, CfMV, FIV, EIAV, SIV, SRV), chez des coronavirus (IBV, EAV, MHV, HCV), chez des virus de plantes et d'animaux (HAstV, RCNMV, BWYV, BYDV), chez des rétrotransposons, chez des virus de la levure (dsRNA virus, L-A virus), chez des gènes bactériens (*dnaX*, *ccd*), chez des gènes de bactériophages (gène *10*), des séquences d'insertions (IS) bactériennes (*IS1*, *IS911*, *IS231*, *IS150*, *IS3*), et chez un gène humain (*edr*) (voir abréviations<sup>¶</sup> en bas de page) (revue par Brierley, 1995; Farabaugh, 1996; Gesteland et Atkins, 1996; Namy et al., 2004). Des analyses informatiques de séquences du génome humain suggèrent qu'une vingtaine de gènes, tel CCR5, contiennent des sites de déphasage en  $-1$  (Hammell et al., 1999). Voici en détails quelques exemples de cas démontrés de déphasage en  $-1$ .

Le décalage de phase de lecture en  $-1$  du gène bactérien *dnaX* est probablement le seul exemple de déphasage cellulaire impliqué dans un rôle biologique important. Le gène *dnaX* code pour les sous-unités  $\gamma$  (gamma) et  $\tau$  (tau) de l'holoenzyme ADN polymérase III chez *E. coli*. La sous-unité  $\gamma$  permet à l'holoenzyme de lier les amorces d'ARN avec une haute affinité. Quant à la sous-unité  $\tau$ , elle coordonne la synthèse des brins continus et discontinus de l'ADN en favorisant la dimérisation de l'ADN polymérase III. La synthèse de la sous-unité  $\gamma$  résulte d'un décalage de phase de lecture en  $-1$  à la séquence glissante AAA AAG, et, suite à ce décalage, les ribosomes rencontrent rapidement un codon de terminaison. La sous-unité  $\tau$ , plus longue, est produite par le décodage conventionnel (Blinkowa et Walker, 1990; Flower et McHenry, 1990; Tsuchihashi et Brown, 1992). Environ 50% des ribosomes effectuent un déphasage en  $-1$ , ce qui permet de produire de

---

<sup>¶</sup> abréviations : HTLV-I, II : human T-cell lymphotropic virus type I et II. MMTV : mouse mammary tumor virus, CfMV : coxsfoot mottle virus, FIV : feline immunodeficiency virus, EIAV : equine infectious anemia virus, SIV : simian immunodeficiency virus, SRV : simian retrovirus, IBV : avian infectious bronchitis coronavirus. EAV, equine arteritis virus, MHV, murine hepatitis virus, HCV : human coronavirus, HAstV : human astrovirus, RCNMV : red clover necrotic mosaic dianthovirus, BWYV : beet western yellow virus, BYDV : barley dwarf yellow virus.

façon stœchiométrique une sous-unité  $\gamma$  pour chaque sous-unité  $\tau$  produite de façon conventionnelle. Cette haute efficacité de déphasage en  $-1$  est due à une particularité de l'ARNt<sup>Lys</sup><sub>UUU</sub> qui reconnaît les codons AAA et AAG. Chez *E. coli*, cet ARNt possède un U modifié (5-méthylaminométhyl-2-thiouridine) à la position wobble de l'anticodon et s'apparie plus efficacement au codon AAA qu'au codon AAG, dû à une faible interaction entre le U modifié et le G en troisième position du codon (Sundaram et al., 2000). Deux signaux *cis* permettent également d'atteindre cette haute efficacité de glissement : une séquence de type SD située dix nucléotides en amont de la séquence glissante et une structure secondaire en tige-boucle, localisée cinq nucléotides en aval de cette dernière. Ces signaux augmentent l'efficacité du déphasage en  $-1$  d'un facteur deux et dix, respectivement (Larsen et al., 1994). Contrairement à ce que nous avons vu dans le cas de RF2, la distance entre la séquence de type SD et le début de la séquence glissante est plus grande que ce qui est habituellement observé lors de l'initiation de la traduction. Ceci pourrait créer des tensions dans l'ARNr favorisant un déplacement du ribosome portant les ARNt en direction 5' par rapport au messenger. L'exemple du déphasage du messenger du gène *dnaX* illustre bien comment le déphasage jumelé à la traduction conventionnelle est un moyen efficace pour produire deux protéines ayant des fonctions différentes, dans un rapport précis.

Le décalage de phase de lecture  $-1$ , qui se produit lors de la traduction du rétrovirus VIH-1, permet la synthèse du précurseur polyprotéique Gag-Pol. Comme nous l'avons déjà indiqué, le précurseur des protéines de structure, Gag, et celui des enzymes, Gag-Pol, du VIH-1 sont produits à partir du même ARNm, mais dans une stœchiométrie différente. La séquence de glissement U UUU UUA, située environ 200 nucléotides en amont du codon de terminaison de Gag, assure qu'une proportion des ribosomes en traduction effectue un déphasage en  $-1$  et produise Gag-Pol, tandis que les autres ribosomes achèvent la synthèse de Gag. Une structure secondaire en aval de la séquence glissante favorise le déphasage en  $-1$ , la caractérisation de cette structure a d'ailleurs fait l'objet de ma première publication (Chapitre 2). L'efficacité du déphasage en  $-1$  du VIH-1 dépend du système utilisé pour le

mesurer et les résultats varient de 2 à 10% (Biswas et al., 2004; Cassan et al., 1994; Grentzmann et al., 1998; Harger et Dinman, 2003; Jacks et al., 1988b; Reil et al., 1993). Ce décalage de phase de lecture en  $-1$  permettant la synthèse de Gag-Pol est relativement fréquent pour les rétrovirus : il contrôle le rapport Gag-Pol sur Gag critique pour la réplication et l'infectivité virale (voir section 1.3.5) et il assure l'incorporation à l'intérieur du virions des enzymes virales nécessaires à la réplication.

À présent, le seul cas connu de gènes cellulaires de mammifères contenant une séquence de décalage de phase de lecture en  $-1$  est le gène *edr* (Shigemoto et al., 2001). Le glissement se produit à la séquence G GGA AAC et est stimulé d'un facteur 15 par une structure en pseudonoeud située en aval (Manktelow et al., 2005). Le signal de déphasage du gène *edr* ressemble au signal classique de déphasage en  $-1$  présent chez les rétrovirus. Le premier cadre de lecture possède une homologie de séquence avec un domaine de liaison à doigt de zinc souvent retrouvé dans Gag tandis que le second cadre de lecture  $-1$ , situé en aval, contient un motif apparenté au site catalytique de protéases aspartiques virales. La fonction de ce gène demeure inconnue, mais sa conservation chez la souris et l'homme ainsi son patron d'expression différent durant le développement suggèrent qu'il pourrait avoir une fonction importante.

#### 1.3.4.2 Signaux *cis* de l'ARNm favorisant le décalage de phase de lecture en $-1$

Comme nous avons pu le constater avec les quelques derniers exemples, les sites de décalage de phase de lecture en  $-1$  de l'ARNm contiennent deux éléments essentiels : une séquence glissante, où le ribosome change de phase de lecture, et une structure secondaire stimulatrice située en aval de la séquence glissante. Nous allons maintenant voir au cours de cette section quelles sont les propriétés de ces deux éléments et comment ils permettent le déphasage en  $-1$ .

#### 1.3.4.2.1 Séquence glissante

Des analyses de comparaisons de séquences et des études de mutagenèse dirigée sur plusieurs sites de décalage de phase de lecture en  $-1$  ont permis d'identifier un motif commun au chevauchement des cadres de lecture. Ce site, nommé séquence glissante, consiste en un heptamère de type X XXY YYZ constitué de deux triplets homopolymériques. Comme nous l'avons mentionné, la répétitivité de la séquence glissante permet aux ARNt de maintenir au moins deux appariements corrects sur trois avec le messager après le déphasage. ten Dam et al. (1990) ont identifié ce type de site de glissement chez 38 virus à ARN, mais ont trouvé seulement 11 séquences différentes utilisées : A AAA AAC, A AAU UUA, A AAU UUU, G GGA AAC, G GGC CCC, G GGU UUA, G GGU UUU, U UUA AAC, U UUA AAU, U UUU UUA et G GAU UUA. Seule la dernière séquence n'est pas conforme au motif X XXY YYZ. On constate une restriction au niveau de l'utilisation des séquences X XXY YYZ, puisque le nombre de cas possibles est beaucoup plus élevé que ce qui est rencontré. La nature homopolymérique de ces sites est importante pour l'efficacité du glissement. Des mutations dans les trois premiers nucléotides du site, changeant par exemple la séquence glissante de A AAU UUA à C AAU UUA, A CAU UUA ou A ACU UUA, réduisent l'efficacité du déphasage en  $-1$  d'un facteur cinq environ, tandis que des mutations du même type dans les trois nucléotides suivants abolissent le déphasage en  $-1$  (Brierley et al., 1992; Dinman et al., 1991; Dinman et Wickner, 1992; Jacks et al., 1988a). Les essais de mesure du déphasage en  $-1$  dans plusieurs cas ont révélé que la base X peut être A, C, G ou U, mais que Y peut seulement être U ou A (Brierley et al., 1992; Dinman et al., 1991; Jacks et al., 1988a). De plus, si Z est un G, comme dans les séquences de types X XXU UUG ou X XXA AAG, le déphasage n'est pas favorisé chez les mammifères et la levure (Brierley et al., 1992; Dinman et al., 1991; Marczinke et al., 2000). Le séquençage de la région de glissement des protéines produites par décalage de phase de lecture en  $-1$  chez MMTV, VIH-1 et RSV a aussi permis de confirmer que le changement de cadre se produit après le second triplet de la séquence heptamérique. En d'autres termes, le dernier codon YYZ est le dernier codon

décodé dans le cadre zéro par le ribosome (Hizi et al., 1987; Jacks et al., 1988a; Jacks et al., 1988b; Nam et al., 1993; Weiss et al., 1989). Des études récentes ont cependant suggéré que la séquence glissante est plus complexe qu'un simple heptamère. D'abord, Bertrand et al. (2002) ont montré que les trois nucléotides suivant la séquence glissante influencent l'efficacité du déphasage en  $-1$ . En effet, la présence de purines à cet endroit diminue l'efficacité du déphasage en  $-1$ , probablement par stabilisation de l'interaction codon-anticodon au site A, grâce à un effet d'empilement (*stacking*). D'autre part, Bekaert et Rousset (2005), ont récemment montré que le codon situé immédiatement en 5' de l'heptamère influence le processus de déphasage. La présence d'un codon spécifiant un ARNt modifié à cet endroit augmente l'efficacité du déphasage en  $-1$ , probablement en facilitant l'éjection de cet ARNt déchargé du site E.

#### 1.3.4.2.2 Structures secondaires stimulatrices

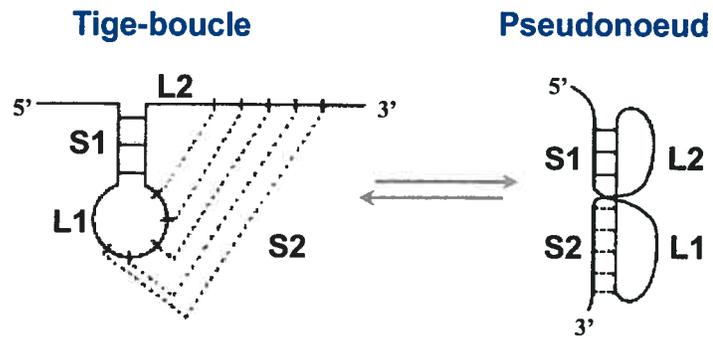
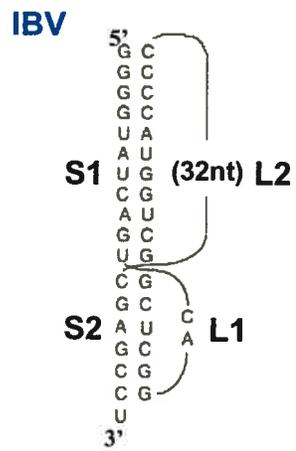
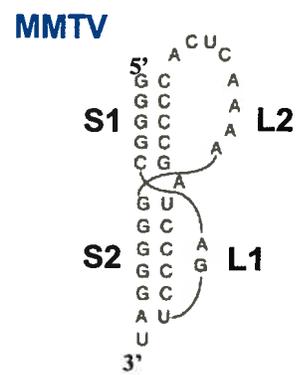
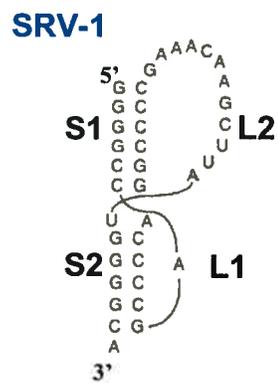
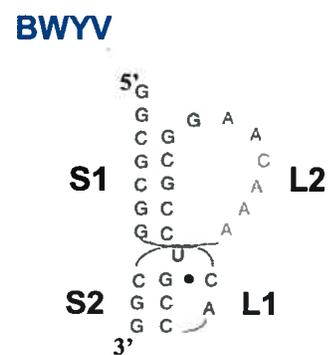
Le décalage de phase en  $-1$  nécessite la présence d'une structure d'ARN stimulatrice située quelques nucléotides en aval de la séquence glissante. Cette structure secondaire est parfois une tige-boucle, mais le plus souvent il s'agit d'un pseudonoeud (revue par Brierley et Pennell, 2001). La distance séparant la séquence glissante et la structure d'ARN stimulatrice est souvent de cinq à neuf nucléotides et son maintien est important pour l'efficacité du déphasage en  $-1$  (Brierley et al., 1989; Brierley et al., 1992; Kollmus et al., 1994).

Le premier type de signal stimulateur de déphasage en  $-1$  que nous allons aborder est le pseudonoeud (revue par Brierley et Pennell, 2001; Giedroc et al., 2000). Les pseudonoeuds forment des motifs structuraux impliqués dans plusieurs fonctions de l'ARN. On les retrouve dans l'ARN ribosomique, les ARN catalytiques et auto-catalytiques, les sites d'initiation interne de la traduction pour le ribosome eucaryote (IRES) et des sites contrôlant la répression de l'initiation de la traduction d'ARNm de protéines ribosomiques

bactériennes (voir Giedroc et al., 2000 pour des références sur ce sujet). Parmi les différents types de pseudonoeuds, le type H (*hairpin*) est le plus souvent rencontré. Ce dernier est formé par l'interaction de la boucle d'une tige-boucle avec une séquence complémentaire adjacente, résultant en une structure ayant deux tiges (S1 et S2) jointes par deux boucles (L1 et L2) (Figure 1-20A). Le pseudonoeud est une structure hélicoïdale quasi-continue, dont les boucles L1 et L2 en simple brin traversent les sillons majeurs et mineurs des tiges S2 et S1, respectivement (Pleij et al., 1985). Le premier cas de pseudonoeud stimulant le déphasage en  $-1$  a été identifié par Brierley et al. (1989) dans leurs études sur le coronavirus aviaire (IBV). La structure tridimensionnelle de ce signal stimulateur n'a toutefois pas encore été définie à une résolution atomique et nos connaissances de cette dernière sont basées sur des expériences de mutagenèse dirigée et de détermination de structure (*probing*) par des attaques avec des agents chimiques ou enzymatique (Figure 1-20B) (Brierley et al., 1989; Brierley et al., 1992; Brierley et al., 1991; Liphardt et al., 1999; Naphthine et al., 1999). Plusieurs pseudonoeuds stimulant le décalage de phase de lecture en  $-1$  ont été caractérisés ensuite par des approches similaires, chez MMTV (Chamorro et al., 1992), SRV-1 (ten Dam et al., 1994; ten Dam et al., 1995), le virus L-A de la levure (Dinman et Wickner, 1992; Tzeng et al., 1992), des éléments d'insertions bactériens (IS) (Sekine et al., 1994), BYDV (Paul et al., 2001), le coronavirus responsable du SARS (Dos Ramos et al., 2004), le groupe O du VIH-1 (dont j'ai participé à la caractérisation, Baril et al., 2003b), FIV (Yu et al., 2005) et le gène humain *edr* (Manktelow et al., 2005).

Au cours des dernières années, des expériences de résonance magnétique nucléaire (RMN) ont permis de résoudre plusieurs structures en pseudonoeud stimulant le décalage de phase de lecture en  $-1$ . Les travaux du groupe de Tinoco portant sur le signal stimulateur du virus MMTV ont révélé qu'un résidu adénosine non-apparié induit une courbure prononcée de  $60^\circ$  entre les deux tiges du pseudonoeud et que la délétion de cette adénosine abolit cette courbure (Figure 1-20C, Chen et al., 1996; Kang et al., 1996; Kang et Tinoco, 1997). La conformation courbée du pseudonoeud est très importante, puisque la délétion de

Figure 1-20: Signaux stimulateurs en pseudonoeud présents aux sites de déphasage en  $-1$  de plusieurs virus. A. Le pseudonoeud est une structure hélicoïdale quasi-continue, où les boucles L1 et L2 en simple brin traversent les sillons majeurs et mineurs des tiges en hélice S2 et S1, respectivement. Pseudonoeuds trouvés dans l'ARN des virus IBV (B), MMTV (C), SRV-1 (D) et BWYV (E).

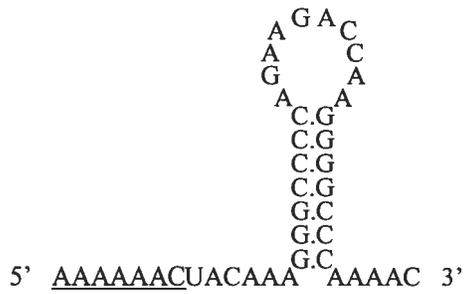
**A****B****C****D****E**

cette adénosine réduit l'efficacité du déphasage en  $-1$  d'un facteur dix. Les données de RMN suggèrent également une interaction entre les bases de la boucle L2 et le sillon mineur de la tige S1. Toutefois, des études de mutagenèse dirigées portant sur la boucle L2 de cette structure n'ont pu montrer son implication dans le déphasage en  $-1$ . La résolution par RMN de la structure du pseudonoeud du rétrovirus SRV-1 en 2001 a montré, contrairement à MMTV, que les deux tiges n'étaient pas courbées, mais bien empilées l'une sur l'autre (Figure 1-20D) (Du et al., 1997; Michiels et al., 2001). Su et al. (1999) ont décrit la structure du pseudonoeud stimulant le déphasage en  $-1$  du lutéovirus BWYV qu'ils ont résolue par diffraction aux rayons X, avec une résolution de 1.6Å (Figure 1-20E). Il s'agit du premier cas de structure cristallographique d'un pseudonoeud stimulant le déphasage en  $-1$  qui a été élucidée par cette approche. Cette haute résolution dans l'analyse de la structure du pseudonoeud de BWYV a permis d'obtenir de nouvelles informations très intéressantes sur les pseudonoeuds favorisant le déphasage en  $-1$ . Par exemple, comme dans le cas de MMTV, les deux tiges ne sont pas coaxiales mais courbées. Cependant, la courbure est due à une rotation de  $48^\circ$  à la jonction des deux tiges, et non à une inclinaison due à la présence d'un nucléotide intercalé entre les deux tiges. La structure révèle un triplex d'ARN formé d'interactions non-canoniques entre les bases AACAA de la boucle L2 et le sillon mineur de la tige S1. Ce triplex stabiliserait la formation de la tige S1. On observe aussi une interaction quadruple entre une cytosine de la boucle L1 et des bases du sillon majeur de la tige S2. Des études de mutagenèse dirigée ont montré que l'identité de cette cytosine de la boucle L1 et des bases du triplex formé entre la boucle L2 et la tige S1 étaient essentiels pour l'efficacité du déphasage en  $-1$  (Kim et al., 1999). Les études de mutagenèse dirigée du même groupe ont révélé que plusieurs bases du pseudonoeud de BWYV, qui ne sont pas impliquées dans des interactions tertiaires stabilisant cette structure, mais qui sont importantes pour l'efficacité du déphasage. Ceci laisse supposer que ces résidus pourraient être impliqués dans des interactions avec le ribosome. Malgré l'avancement des connaissances sur ces structures, il demeure cependant difficile de différencier les particularités des pseudonoeuds stimulant le déphasage en  $-1$  des

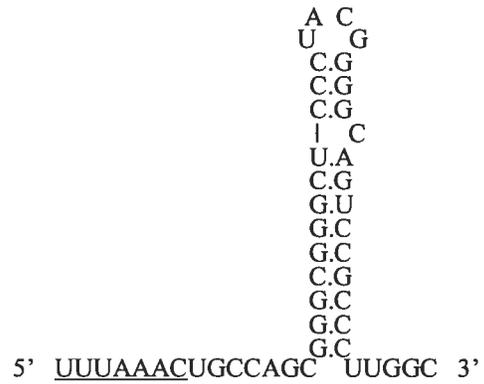
caractéristiques qui n'influencent pas ce processus. On peut suggérer qu'une courbure particulière entre les tiges S1 et S2, ainsi qu'une interaction de la boucle L2 avec la tige S1 sont des caractéristiques favorables au décalage de phase de lecture en  $-1$ .

Le second type de signal stimulateur favorisant le décalage de phase de lecture en  $-1$  est la tige-boucle. On retrouve ce signal chez plusieurs virus d'animaux tel HAst-1 (Jiang et al., 1993), HTLV-II (Mador et al., 1989), VIH-1 (Jacks et al., 1988b) et des virus de plantes tel CfMV (Jacks et al., 1988b) et RCNMV (Xiong et Lommel, 1989). Tel que présenté à la Figure 1-21, on observe de grandes variations entre la taille de ces signaux stimulateurs en tige-boucle. De façon générale, les signaux stimulateurs de virus de plantes possèdent une taille plus grande et une séquence glissante moins homopolymérique. L'analyse de structure des signaux stimulateurs en tige-boucle a reçu moins d'attention que celle des pseudonoeuds. Dans le cas du VIH-1, la structure en tige-boucle de 11 paires de bases, coiffée par une tétraboucle et est séparée de la séquence glissante par sept nucléotides, à d'abord été proposée à partir de programmes informatiques de prédictions de structures (Le et al., 1989), puis démontré par des travaux de mutagenèse dirigée (Parkin et al., 1992) et d'expériences de détermination de structures utilisant des attaques par des agents enzymatiques (Kang, 1998). Peu d'analyses biophysiques ont été effectuées sur ces structures. Une étude récente de Staple et Butcher (2003) a déterminé la structure à haute résolution d'une portion du signal stimulateur en tige-boucle du VIH-1 par RMN (Figure 1-22A). Les informations obtenues par cette approche sont très intéressantes et suggèrent que cette structure possède des déterminants permettant une interaction fonctionnelle avec le ribosome pour le déphasage. En effet, la structure révèle que la boucle ACAA coiffant la tige-boucle, adopte une conformation similaire aux conformations des structures d'ARN reconnues par la RNase III (Lebars et al., 2001; Wu et al., 2001). Selon les données de RMN, la conformation de la boucle ACAA expose la surface des bases au solvant (Figure 1-22B), ce qui pourrait permettre de former un site potentiel de contact avec l'ARNr ou des protéines ribosomiques. Il serait important de déterminer la structure à haute résolution d'autres tiges-boucles stimulant le déphasage.

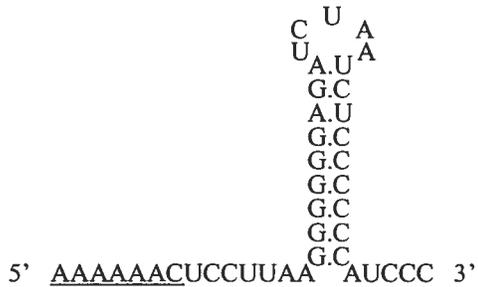
Figure 1-21: Signaux stimulateurs en tiges-boucles présents aux sites de déphasage en  $-1$  de plusieurs virus tels HAst, HTLV-II, VIH-I, CfMV et RCNMV (adapté de Brierley et Pennell, 2001).



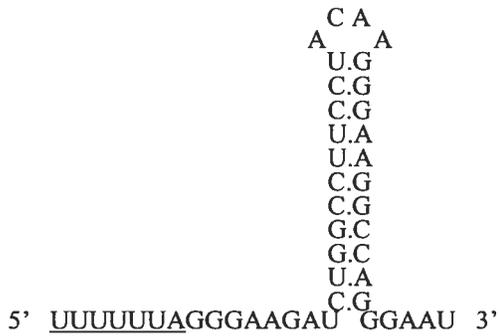
**HAsT**



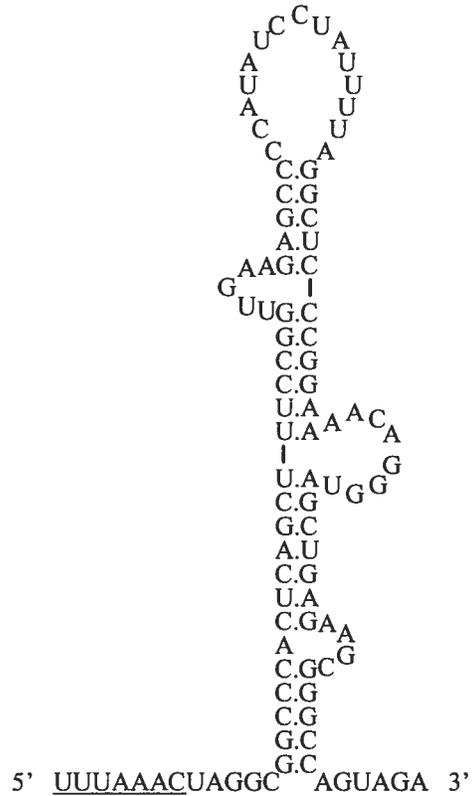
**CfMV**



**HTLV-II**



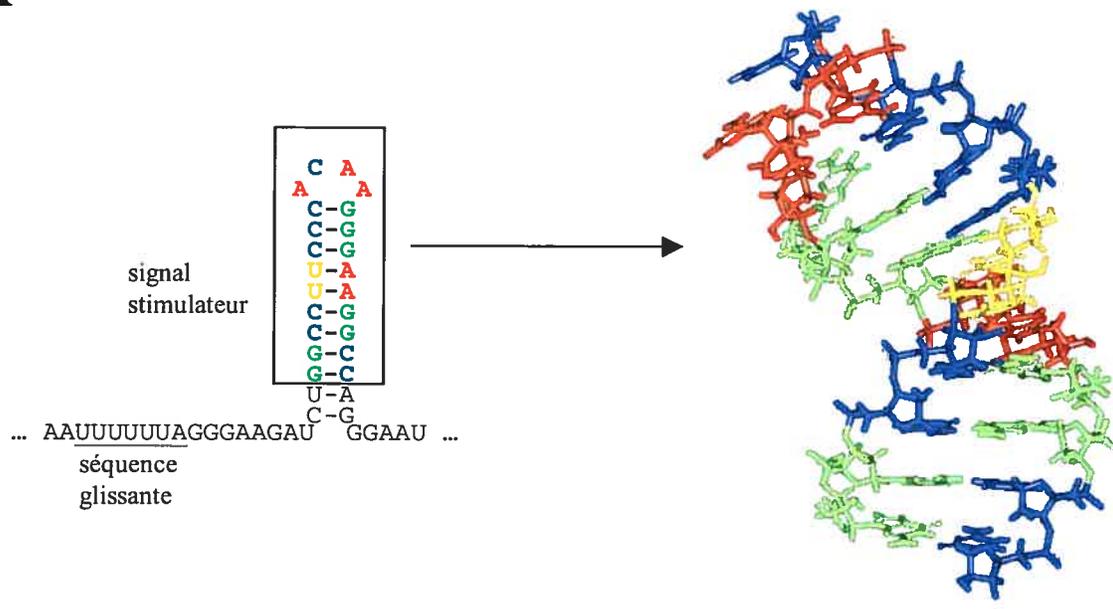
**VIH-1**



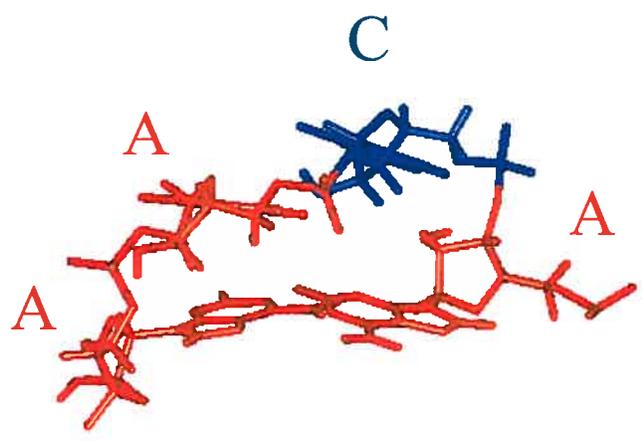
**RCNMV**

Figure 1-22: Structure à haute résolution d'une portion du signal stimulateur du VIH-1. A. La conformation d'une portion de la tige-boucle (encadré à gauche) a été déterminé par RMN (droite). Les nucléotides sont colorés selon : A est rouge, C est bleu, G est vert et U est jaune. B. Conformation de la boucle ACAA coiffant la tige-boucle du signal stimulateur. L'orientation des nucléotides de la boucle suggère qu'ils peuvent interagir avec des partenaires tel le ribosome (adapté de Staple et Butcher, 2003).

**A**

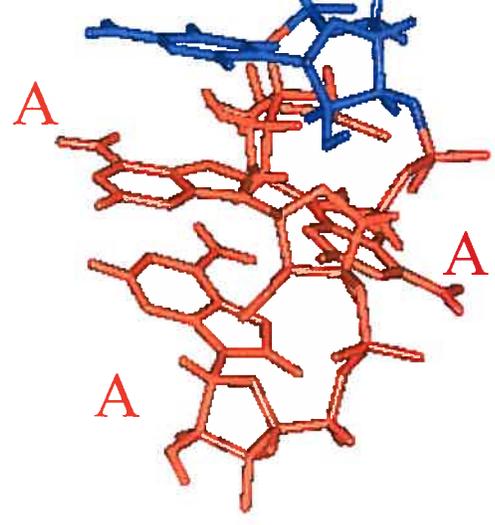


**B**



90°

**C**



### 1.3.4.3 Fonctions du signal stimulateur

La pause du ribosome est au centre de tous les événements de reprogrammation. Celle-ci augmente le temps que les ribosomes passent au site de reprogrammation, ce qui favorise les événements cinétiquement non-favorables (Farabaugh, 2000). Jacks et al. (1988a) ont proposé, dans leur modèle du décalage de phase de lecture en  $-1$ , que le principal rôle du signal stimulateur était d'arrêter le ribosome en traduction à la séquence glissante, augmentant ainsi la durée durant laquelle les deux ARNt peuvent effectuer le déphasage en  $-1$ . La pause du ribosome aux sites de glissement a été démontrée par plusieurs approches expérimentales. D'abord, des polypeptides incomplets correspondant aux produits de traduction par des ribosomes arrêtés sur la séquence glissante au niveau du signal stimulateur en pseudonoeud ont été identifiés sur l'ARN d'IBV (Somogyi et al., 1993) et du virus L-A (Lopinski et al., 2000). Des études par *heelprinting*, c'est-à-dire de positionnement de ribosomes en élongation sur l'ARN ont également identifié des ribosomes arrêtés au site de déphasage des virus L-A, IBV et SRV-1 (Kontos et al., 2001; Lopinski et al., 2000; Somogyi et al., 1993; Tu et al., 1992). On sait que les ribosomes eucaryotes couvrent en moyenne 32 nucléotides de l'ARNm. En utilisant cette information, les analyses des résultats de *heelprinting* confirment que les ribosomes sont arrêtés sur la séquence glissante et qu'ils sont en contact avec le pseudonoeud. Imputer à la pause du ribosome tout le mécanisme du déphasage en  $-1$  serait toutefois simpliste. Plusieurs études ont démontré que produire une pause n'était pas suffisant pour favoriser le déphasage en  $-1$  (Brierley et al., 1991; Chen et al., 1995; Kontos et al., 2001; Liphardt et al., 1999; Lopinski et al., 2000; Naphthine et al., 1999; Somogyi et al., 1993; Tu et al., 1992). Ces expériences n'excluent cependant pas l'implication de la pause dans le mécanisme de déphasage, puisqu'il n'est pas possible d'observer un déphasage sans détecter une pause du ribosome. Nous pouvons en conclure que la pause est nécessaire au déphasage en  $-1$  mais non suffisante, ce qui suggère qu'il y aurait des interactions spécifiques entre le signal stimulateur et le ribosome favorisant le glissement.

Comme nous l'avons remarqué plus haut, les pseudonoeuds possèdent des caractéristiques particulières, comme des interactions triples, stabilisant leurs structures. Ces caractéristiques empêchent peut-être la fonction hélicase associée au ribosome de défaire les structures secondaires d'ARN du messager. Cette fonction hélicase du ribosome n'a toutefois pas encore été caractérisée. Les seules informations connues à ce sujet proviennent de structures du ribosome bactérien résolues par diffraction au rayon X. Yusupova et al. (2001) ont suggéré que trois protéines, S3, S4 et S5 de la petite sous-unité, situées à proximité du tunnel d'entrée par où passe l'ARNm, sont impliquées dans le démantèlement des structures d'ARN. Il est possible qu'une topologie particulière associée aux pseudonoeuds leur permette de résister à la fonction hélicase du ribosome. À ce sujet, Plant et al. (2003) ont proposé que la difficulté du ribosome à défaire le pseudonoeud entraîne un blocage de ce dernier dans le tunnel d'entrée du ribosome. Ceci empêcherait le déplacement de l'ARNm vers l'intérieur du ribosome qui est normalement requis lors de l'accommodation de l'aa-ARNt au site A/A. Ce blocage introduirait alors une tension, un étirement, dans la portion de l'ARNm entre la séquence glissante et le signal stimulateur, et cette tension serait abolie en désappariant les ARNt de l'ARNm permettant alors à l'ARNm de glisser d'un nucléotide en direction 3'. Des études du même groupe ont récemment suggéré que des restrictions dans la torsion d'un pseudonoeud peuvent expliquer comment un tel signal favorise le désappariement des ARNt et le mouvement subséquent de l'ARNm lors du déphasage en  $-1$  (Plant et Dinman, 2005). En effet, lorsque le ribosome défait le pseudonoeud, il déroule d'abord le bas de la tige 1, ce qui entraîne normalement une rotation avec une torsion de cette tige qui est restreinte par la présence de la tige 2. Plant et Dinman suggèrent que l'énergie emmagasinée par la tension au niveau de la tige 2 du pseudonoeud lorsque le ribosome commence à défaire la tige 1 est redistribuée à l'ARNm pour permettre la dissociation des ARNt de l'ARNm et le déphasage en  $-1$ . Ce modèle est intéressant si on considère le signal stimulateur comme un participant actif impliqué dans des interactions spécifiques avec le ribosome lors du déphasage en  $-1$ . On peut même aller jusqu'à considérer la pause comme un événement secondaire résultant de l'interaction entre

le ribosome et le signal stimulateur, mais quoiqu'il en soit, le rôle exact du pseudonoeud demeure encore mystérieux.

Le mécanisme par lequel les tiges-boucles favorisent le déphasage en  $-1$  demeure tout aussi obscur. Étant donné qu'elles forment une structure moins complexe qu'un pseudonoeud, il est possible que les tiges-boucles favorisent un plus faible niveau de déphasage en agissant simplement comme de barrières énergétiques causant une pause du ribosome. Cependant, aucune relation n'existe entre l'efficacité du déphasage en  $-1$  et la présence d'un pseudonoeud ou d'une tige-boucle comme signal stimulateur. Bidou et al. (1997) ont montré une corrélation entre la stabilité et l'efficacité du déphasage en  $-1$  pour un stimulateur en tige-boucle. Toutefois, comme certaines tiges-boucles sont incapables de favoriser le déphasage en  $-1$ , il est possible que les structures stimulatrices en tige-boucle possèdent des caractéristiques encore non identifiées leur permettant entre autres d'interagir de façon spécifique avec le ribosome. Il est aussi possible que les séquences adjacentes aux tiges-boucles influencent la cinétique avec laquelle ces structures secondaires sont défaites par le ribosome (Bertrand et al., 2002). Clairement, d'autres travaux sont nécessaires afin de déterminer comment les tiges-boucles favorisent le glissement.

#### **1.3.4.4 Mécanismes et modèles du décalage de phase de lecture en $-1$**

Comme nous venons de le constater, la structure secondaire stimulatrice fait partie intégrale du mécanisme de décalage de phase de lecture en  $-1$ , soit de façon passive en produisant une pause ou de façon active en interagissant avec le ribosome. Toutefois, en l'absence de séquence glissante, le signal stimulateur ne cause pas de déphasage, ce qui montre l'importance des événements se produisant à la séquence glissante pour le déphasage en  $-1$ . Le premier modèle de déphasage en  $-1$  a été proposé par Jacks et al. (1988a) suite à leurs études portant sur le virus de Rous (RSV). Selon ce modèle, lorsqu'un

ribosome en élongation rencontre le signal stimulateur, il effectue une pause à la séquence glissante (X XXY YYZ), ce qui permet au peptidyl-ARNt au site P et à l'aa-ARNt au site A du ribosome, liés aux codons XXY et YYZ, de se désappairer de l'ARNm et de glisser simultanément d'un nucléotide en direction 5' avec le ribosome pour ensuite se ré-apparier aux codons adjacents XXX et YYY, dans le nouveau cadre de lecture -1. Le modèle suggère que le glissement simultané se produit avant la formation du lien peptidique et que les deux ARNt se déplacent comme une seule entité dont le mouvement est coordonné par rapport à l'ARNm. Même s'il est possible que le ribosome soit responsable de cette coordination et que le signal stimulateur provoque le désappariement simultané des deux ARNt de l'ARNm, aucune observation expérimentale ne supporte ce mouvement simultané. Comme nous le verrons plus bas, un modèle récent de Baranov et al. (2004) suggère que le mouvement des ARNt est plutôt séquentiel. Plusieurs études employant des inhibiteurs de la traduction ou des mutants de l'appareil traductionnel bloquant ou accélérant des étapes précises du cycle d'élongation ont défini l'étape durant laquelle se produit le déphasage. Ces études montrent que le déphasage se produit avant la formation du lien peptidique et probablement lorsque l'aa-ARNt est au site A/T, donc avant qu'il n'occupe le site A/A (Brunelle et al., 1999; Dinman et al., 1997; Dinman et al., 1998; Hudak et al., 2001; Léger et al., 2004; Tumer et al., 1998).

Récemment, Baranov et al. (2004) ont proposé que le décalage de phase de lecture dépend du mouvement du peptidyl-ARNt au site P. Comme dans le cas du déphasage en +1, le mouvement du peptidyl-ARNt au site P dans le nouveau cadre dépend des différences d'énergies entre le complexe codon-anticodon dans le cadre zéro et le nouveau cadre -1. Selon le modèle de Baranov et al., le déphasage du peptidyl-ARNt au site P se produit lorsque l'aa-ARNt est au site A/T, avant qu'il n'occupe le site A/A. Le succès du déphasage dépend du mouvement séquentiel du peptidyl-ARNt puis de l'aa-ARNt, dans le nouveau cadre de lecture -1. Dans le cas du repositionnement de l'aa-ARNt dans le cadre -1 au site A, il dépend du bon appariement codon-anticodon qui est vérifié avec rigueur par le ribosome à cet endroit. À cet effet, on peut maintenant comprendre pourquoi des

mutations dans le triplet décodé par le site A diminuent plus l'efficacité du déphasage en  $-1$  que celles qui ont lieu au site P. Selon le modèle, le déplacement du peptidyl-ARNt au site P en direction  $-1$  peut aussi se produire lorsque le site A est complètement inoccupé, c'est-à-dire avant qu'un aa-ARNt interagisse avec le ribosome ou que l'aa-ARNt au site A/T soit rejeté lors de la sélection initiale ou durant l'étape de *proofreading*. Dans ce cas, l'incorporation de l'aa-ARNt suivant au site A se fera directement dans le nouveau cadre de lecture en  $-1$ . Il est intéressant de noter que pour le VIH-1, la séquence glissante U UUU UUA dans le cadre zéro code pour une phénylalanine (UUU) et une leucine (UUA). Le séquençage de protéines produites par le déphasage en  $-1$  du VIH-1 a montré que 70% d'entre elles contenaient une phénylalanine suivie d'une leucine au site de glissement, consistant avec un glissement séquentiel ou en tandem des deux ARNt, tandis que 30% contenaient deux phénylalanines consécutives (Jacks et al., 1988b; Yelverton et al., 1994), ce qui correspondrait à un glissement de l'ARNt au site P uniquement, suivi de l'incorporation de l'aa-ARNt<sup>Phe</sup> directement dans le cadre de lecture en  $-1$  au site A.

### 1.3.5 Régulation de la réplication virale par la reprogrammation

Dans les sections précédentes, nous avons étudié les mécanismes de reprogrammation utilisés par les organismes et les virus pour produire les protéines désirées. Dans cette section, nous analyserons l'importance du maintien de l'efficacité du décalage de phase de lecture pour la réplication virale.

Dès 1988, Felsenstein et Goff ont étudié l'effet d'une augmentation de l'efficacité de la translecture du codon stop de Gag sur la réplication du rétrovirus MuLV. L'expression d'un provirus mutant de MuLV permettant l'expression constitutive de Gag-Pol, mais non de Gag, en changeant le codon stop en codon glutamine, ne permettait pas l'assemblage et la relâche de particules virales. Des résultats similaires ont d'ailleurs été observés par Karacostas et al. (1993) avec le VIH-1 peu de temps après. Ils ont montré que

l'introduction d'un nucléotide immédiatement en aval de la séquence glissante du signal de décalage de phase de lecture en  $-1$  de façon à produire seulement Gag-Pol était délétère pour la production de particules virales. Il a été suggéré que les défauts d'assemblage étaient probablement causés par un encombrement stérique dû à la présence de Pol à l'extrémité de Gag (Swanstrom, 1997) ou par une activation prématurée de la protéase virale dans la cellule (Karacostas et al., 1993). Ces études intéressantes ont le désavantage de comparer seulement le rapport Gag-Pol sur Gag dans les conditions sauvages à une situation où on ne produit que Gag-Pol. En 1992, une intéressante étude de Dinman et Wickner permettait de confirmer l'importance de l'efficacité du déphasage et du rapport Gag-Pol sur Gag pour la propagation virale. Ils ont introduit des mutations dans la séquence glissante et dans le pseudonoeud, au site de déphasage en  $-1$  du virus L-A de la levure, de façon à modifier l'efficacité de ce déphasage. Ils ont observé qu'une augmentation ou une diminution de l'efficacité du déphasage d'un facteur deux empêchait la propagation virale. Des études récentes du groupe de Mak ont permis d'obtenir plus d'informations sur les causes de la diminution de l'infectivité virale suite au changement du rapport Gag-Pol sur Gag. D'abord, ils ont réussi d'une élégante façon à augmenter le rapport Gag-Pol sur Gag en co-transfectant des cellules en culture avec un premier vecteur proviral contenant le génome du VIH, et différentes quantités d'un second vecteur exprimant le précurseur Gag-Pol. Avec des rapports Gag-Pol sur Gag allant de  $1/20$  pour le virion sauvage à  $21/20$ , et passant par différentes valeurs intermédiaires, ils ont observé une diminution progressive de l'infectivité corrélant avec une diminution de la stabilité du dimère d'ARN génomique dans les virions et une formation aberrante de particules virales dans le cytoplasme (Shehu-Xhilaga et al., 2001; Shehu-Xhilaga et al., 2002b). Pour leur part, Telenti et al. (2002) ont analysé des isolats cliniques du VIH-1 et observé que seuls les isolats où l'efficacité de déphasage en  $-1$  était de 40% ou moins de la valeur sauvage avaient des défauts de réplication. Des études récentes avec un autre virus ont également souligné l'importance du rapport Gag-Pol sur Gag dans la réplication virale. Chen et Montelaro (2003) ont montré que des rétrovirus mutants de EIAV n'étaient pas capables de se répliquer lorsque des

mutations dans le signal stimulateur en pseudonoeud diminuaient l'efficacité du déphasage en  $-1$  à 40% de la valeur sauvage.

Il est évident que les rétrovirus ont besoin que Gag-Pol et Gag soient synthétisés dans un rapport précis. Jusqu'à récemment, on croyait que seuls les signaux *cis* présents dans l'ARN génomique de ces virus étaient responsables de la définition de ce rapport. Une étude du groupe de Goff a récemment mis en évidence le rôle de la transcriptase inverse (RT) du virus de la leucémie murine de Moloney (MoMLV) dans ce processus. Cette enzyme contrôle l'événement de translecture permettant son expression (Orlova et al., 2003). Nous avons mentionné plus haut que les enzymes virales de MoMLV (PR, RT et IN) sont synthétisés sous la forme d'un précurseur Gag-Pol par la translecture d'un codon stop à l'extrémité 3' du gène de *gag*. Le groupe de Goff a démontré que la RT de MoMLV interagit avec le facteur de terminaison eRF1 et le séquestre, augmentant l'efficacité avec laquelle se produit la translecture du codon de terminaison de Gag. Des virus mutants de MoMLV pour lesquels l'interaction RT-eRF1 est abolie sont incapables de produire Gag-Pol et ne produisent que des virions non infectieux. Ceci indique que l'interaction de la RT avec les facteurs de terminaison joue un rôle important dans la détermination du rapport Gag-Pol sur Gag requis pour la réplication de MoMLV.

### **1.3.6 État de la question concernant le déphasage en $-1$ du VIH-1**

Au cours de cette introduction, nous avons souligné que l'utilisation d'un mécanisme de reprogrammation constituait un moyen efficace permettant aux rétrovirus de produire les enzymes, sous forme du précurseur Gag-Pol, en quantité requise et de les emmagasiner dans le virion. Nous avons également vu l'importance de maintenir un rapport Gag-Pol sur Gag précis pour la réplication virale (section 1.3.5). Dans le cas du VIH-1, nous connaissons la séquence glissante et le signal stimulateur en tige-boucle formant la région de déphasage (voir section 1.3.4.1 et 1.3.4.2.2, respectivement). Malgré

l'avancement de nos connaissances sur le déphasage en -1 du VIH-1, plusieurs aspects de ce processus demeurent mystérieux. L'un d'entre-eux concerne la structure du signal stimulateur du déphasage en -1. Pourquoi ce signal est-il une simple tige-boucle pour le VIH-1 alors que, pour la majorité des virus utilisant un déphasage en -1, il s'agit d'une structure en pseudonoeud ? Nous avons donc étudié si la séquence en aval de cette tige-boucle pouvait influencer l'efficacité du déphasage en participant à la formation d'une structure stimulatrice plus complexe. La caractérisation approfondie du signal stimulateur du VIH-1 a constitué le premier objectif de notre recherche (voir Chapitre 2).

Nous avons vu, dans la section précédente, l'importance du maintien de l'efficacité du déphasage pour la réplication virale. Nous savons que le signal stimulateur du déphasage contrôle l'efficacité avec laquelle se produit le déphasage. Toutefois, peu de données sont disponibles sur les conséquences de changements dans ces signaux sur l'efficacité du déphasage et sur la réplication du VIH-1. Les études du groupe de Mak (Shehu-Xhilaga et al., 2001; Shehu-Xhilaga et al., 2002b) se sont concentrées sur les effets d'une augmentation artificielle du rapport Gag-Pol sur Gag (voir plus haut). Telenti et al. (2002), ont analysé des variants naturels ou obtenus au laboratoire du signal stimulateur du VIH-1, qui se limitaient à un très petit nombre de valeurs différentes d'efficacité de déphasage. Dans la seconde partie de notre recherche, nous avons voulu étudier l'effet d'un plus large spectre de variations du rapport Gag-Pol sur Gag suite à des mutations dans la séquence glissante et le signal stimulateur afin de mieux définir la fenêtre de variations du déphasage tolérées par le virus. Les résultats de cette étude sont présentés au Chapitre 3.

Le mécanisme par lequel les signaux stimulateurs, dont celui du VIH-1, favorisent le déphasage n'est pas encore élucidé. Il a été suggéré que ces signaux interagissent spécifiquement avec le ribosome. Nous avons souligné à la section précédente l'importance du maintien du rapport Gag-Pol sur Gag et nous avons observé que de légères variations dans l'efficacité du déphasage handicapent la réplication du virus (résultats du Chapitre 3). Est-il possible de sélectionner des agents influençant l'efficacité du déphasage? Nous avons

développé une approche visant à sélectionner des composés interagissant avec le signal stimulateur. Cette approche se basait sur le principe des bibliothèques combinatoires pour identifier des composés interférant le déphasage du VIH-1. Notre hypothèse est que ces composés pourraient interagir avec le signal stimulateur qui possèdent une structure spécifique et ainsi empêcher l'interaction postulée entre ce signal stimulateur et le ribosome. Notre approche s'inspire des travaux de Harada et al. (1996) qui ont sélectionné des peptides liant des structures secondaires en tige-boucle comme TAR et RRE du VIH-1. Nos résultats préliminaires sont présentés au Chapitre 4.

## 1.4 Description du projet de recherche

Nos travaux de recherches ont été divisés en trois volets. Dans le premier volet, notre but a été de caractériser le signal stimulateur du décalage de phase de lecture en  $-1$  du VIH-1, et comme nous l'avons mentionné plus haut, nous avons voulu déterminer si la séquence suivant cette tige-boucle pouvait influencer l'efficacité du déphasage en  $-1$  et si la structure du signal stimulateur était plus complexe qu'une simple tige-boucle. Nous avons d'abord introduit la région de déphasage du VIH-1, contenant la séquence glissante et la tige-boucle suivie d'environ 30 nucléotides de séquence virale additionnelle, en amont d'un gène rapporteur luciférase, de façon à ce que la synthèse du gène dépende d'un événement de déphasage en  $-1$ . Nous avons observé que l'allongement de la séquence du VIH-1 insérée par rapport à une région se limitant à la séquence glissante suivie de la tige-boucle augmentait l'efficacité du déphasage. Nous avons ensuite procédé à des délétions de la séquence suivant la tige-boucle, vérifié l'effet de ces délétions sur l'efficacité du déphasage dans des cellules en culture. Ceci nous a conduit à l'observation qu'une séquence de neuf nucléotides suivant la tige-boucle participait au déphasage. Par la suite, nous avons utilisé une approche permettant de déterminer la conformation du signal stimulateur du déphasage du VIH-1 avec des sondes enzymatiques. Pour ce faire, de petits fragments d'ARN correspondant à la région de déphasage du VIH-1 ont été produits par transcription *in vitro* et soumis à une attaque par l'endonucléase  $V_1$ , une enzyme qui clive les structures d'ARN

ayant une conformation hélicoïdale. Nous avons également utilisé une approche de mutagenèse dirigée afin de vérifier la structure prédite à partir des attaques par l'endonucléase  $V_1$ . Nos résultats indiquent que le signal stimulateur est plus complexe qu'une simple tige-boucle et consiste en une hélice irrégulière formée de deux portions séparées par une protubérance riche en purines. Ces travaux ont conduit à notre première publication. Dans cette publication, nous avons dû confronter nos résultats avec une structure alternative, qui a été proposée pendant que nous complétions notre étude. Cette analyse est présentée dans la première partie de la discussion (section 5.1).

Dans le deuxième volet de notre étude, nous avons examiné la contribution des différents éléments contrôlant le déphasage en  $-1$  du VIH-1 sur l'efficacité de ce déphasage et sur la réplication du virus. Nous avons utilisé la technique de mutagenèse dirigée pour introduire des mutations dans la séquence glissante ou le signal stimulateur du déphasage en  $-1$ . L'effet de ces mutations sur l'efficacité du déphasage en  $-1$  a d'abord été évalué par un système rapporteur *dual*-luciférase introduit dans des cellules en culture. Ce vecteur permet l'expression de deux gènes rapporteurs codant pour une luciférase à partir d'un même messenger. Il s'agit des gènes de la luciférase de la *Renilla* et de la luciférase de la luciole (*firefly*). La région de déphasage du VIH-1 a été introduite entre ces deux gènes de façon à ce que la synthèse de la luciférase de la *Renilla* dépende d'une traduction conventionnelle alors que la synthèse de la luciférase de la luciole, dont la séquence codante est située en aval du site de déphasage, dépende de ce déphasage et soit synthétisée sous la forme d'une fusion luciférase *Renilla*-luciole. Nous avons ensuite vérifié le contenu en RT de pseudo-particules virales (VLP) produites par transfection dans des cellules en culture de vecteurs exprimant Gag et Gag-Pol, avec les mutations étudiées dans la séquence glissante et le signal stimulateur. Finalement, nous avons introduit ces mutations dans un vecteur proviral du VIH-1, ce qui nous a permis de vérifier directement l'effet du déphasage en  $-1$  sur la réplication d'un cycle et la réplication à long terme du virus. Nos résultats montrent que même des changements modestes de l'efficacité du déphasage,

comme une diminution de 30%, réduisent significativement l'infectivité et la réplication virale. Ces résultats ont été soumis pour publication à *Virology*.

Dans le troisième volet de notre projet, nous avons entamé la mise au point d'un système permettant la sélection de peptides qui perturbent l'efficacité du déphasage, probablement en liant le signal stimulateur du VIH-1. Pour ce faire, nous avons utilisé la stratégie de bibliothèques combinatoires et un système d'expression bactérien sachant que le ribosome bactérien répond de la même façon que le ribosome eucaryote au signal stimulateur du VIH-1. Nous avons construit un vecteur rapporteur codant pour la protéine fluorescente verte (GFP) et la protéine fluorescente rouge (RFP). Dans ce vecteur, l'expression de la GFP dépend du déphasage en  $-1$  du VIH-1 tandis que l'expression de la RFP dépend de la traduction conventionnelle. Nous avons construit un second vecteur permettant l'expression d'une bibliothèque combinatoire de peptides fusionnés à la thioredoxine. Le système a été construit de façon à mesurer les variations de fluorescence de la GFP et de la RFP par un appareil de type FACS (*fluorescence-activated cell sorter*) suite à l'introduction des deux vecteurs dans une souche bactérienne. Nous avons sélectionné par FACS des clones parmi la population de bactéries transformées avec la bibliothèque exprimant des peptides où l'expression du rapporteur dépendant du déphasage était modifiée alors que l'expression du rapporteur dépendant de la traduction conventionnelle semblait normale. Cependant, les clones sélectionnés pour l'augmentation du déphasage se sont tous avérés des faux positifs. Quand aux clones sélectionnés pour une diminution du déphasage, l'analyse de plusieurs d'entre eux a montré que les peptides exprimés influencent plutôt la traduction en général et non spécifiquement le déphasage en  $-1$ . Les résultats obtenus dans ce troisième volet de notre recherche, quoique préliminaires, soulignent les limites du système de sélection bactérien et la nécessité de modifier notre approche visant à sélectionner des agents perturbant le déphasage et, par conséquent, la réplication du VIH-1.

## **Chapitre 2**

Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1.

Dominic Dulude, Martin Baril and Léa Brakier-Gingras\*

Département de Biochimie, Université de Montréal, Montréal, Québec H3T 1J4, Canada.

*Nucleic Acids Research Vol 30 p5094-5102 Dec 2002*

\* Corresponding author:

Léa Brakier-Gingras, Département de Biochimie,  
Université de Montréal, 2900 Blvd Édouard-Montpetit  
Montréal, Québec, H3T 1J4, Canada



**ABSTRACT**

Synthesis of the Gag-Pol protein of the human immunodeficiency virus type 1 (HIV-1) requires a programmed  $-1$  ribosomal frameshifting when ribosomes translate the unspliced viral messenger RNA. This frameshift occurs at a slippery sequence followed by an RNA structure motif that stimulates frameshifting. This motif is commonly assumed to be a simple stem-loop for HIV-1. In this study, we show that the frameshift stimulatory signal is more complex than believed and consists of a two-stem helix. The upper stem-loop corresponds to the classic stem-loop, and the lower stem is formed by pairing the spacer region following the slippery sequence and preceding this classic stem-loop with a segment downstream of this stem-loop. A three-purine bulge interrupts the two stems. This structure was suggested by enzymatic probing with nuclease V1 of an RNA fragment corresponding to the *gag/pol* frameshift region of HIV-1. The involvement of the novel lower stem in frameshifting was supported by site-directed mutagenesis. A fragment encompassing the *gag/pol* frameshift region of HIV-1 was inserted in the beginning of the coding sequence of a reporter gene coding for the firefly luciferase, such that expression of luciferase requires a  $-1$  frameshift. When the reporter was expressed in COS cells, mutations that disrupt the capacity to form the lower stem reduced frameshifting whereas compensatory changes that allow re-formation of this stem restored the frameshift efficiency near wild-type level. The two-stem-structure that we propose for the frameshift stimulatory signal of HIV-1 differs from the RNA triple helix structure recently proposed by Dinman *et al.*, 2002, Proc. Natl. Acad. Sci. USA, 99, 5331-5336.

**Keywords:** ribosomal frameshifting; RNA structure; ribosomes; human immunodeficiency virus type 1.

## INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) uses a programmed  $-1$  ribosomal frameshift to produce the Gag-Pol polyprotein, the precursor of its enzymes, when ribosomes translate the full-length viral messenger RNA. Conventional translation of this RNA by the majority of ribosomes produces the Gag polyprotein, the precursor of the structural proteins of the virus, while the frameshift allows a small number of ribosomes to change the reading frame at a specific sequence and extend their reading over the stop codon of the 0 frame *gag* gene, until the stop codon of the *pol* gene in the  $-1$  frame is encountered (1). The ratio of Gag-Pol to Gag is critical for viral assembly and replication, and increasing or decreasing the frameshifting efficiency interferes with the formation of infectious viral particles (2-4). Programmed  $-1$  ribosomal frameshift has also been reported in several other retroviruses, coronaviruses, plant viruses (5,6), in a yeast virus (7,8), in bacteria (9-12), and, recently, in humans (13).

A programmed  $-1$  ribosomal frameshift requires two elements in the mRNA: (i) a heptanucleotide where the frameshift occurs, called the slippery sequence, with an X XXY YYZ consensus sequence (where the 0 frame is indicated by spaces), and (ii) a secondary structure located downstream of the slippery sequence that stimulates the frameshift. These two elements are separated in most cases by a short spacer sequence (5). When ribosomes bearing two tRNAs, the peptidyl-tRNA and the aminoacyl-tRNA, in the P and A site, respectively, encounter the slippery sequence, a minority of them shift the reading frame. This recoding event can be described as follows: the two tRNAs whose anticodons interact with the codons of the mRNA in the 0 frame (X XXY YYZ) unpair from this mRNA, the ribosome shifts backward by one nucleotide, and the tRNAs re-pair in the  $-1$  frame (XXX YYY) (14). Peptide bond formation then occurs and translation resumes in the new reading frame (reviewed in 15). The stimulatory secondary structure downstream of the slippery sequence makes the ribosome pause at the slippery sequence (16-18). However, this pausing is necessary but not sufficient for promoting the frameshift and little correlation was found between the extent of this pause and the frameshift efficiency (19). It has been

suggested that a specific interaction between the frameshift stimulatory signal and the ribosome is required for the shift to occur (20,21). The slippery sequence is U UUU UUA for HIV-1, and the frameshift stimulatory signal is commonly assumed to be a simple stem-loop, with an 11-base-pair stem, separated from the slippery sequence by an eight-base spacer. This contrasts with the majority of known frameshift events, where the stimulatory signal is a pseudoknot, although there are other well-documented cases of stem-loop-containing frameshift signals (6). The structure of the HIV-1 frameshift stimulatory signal was first proposed from a computer structure analysis program (22) and later demonstrated by structural probing experiments, using a short RNA transcript encompassing the HIV-1 classic frameshift region and ending immediately after the stem-loop (23).

In this study, we investigated whether the sequence following the stem-loop could influence the frameshift efficiency in HIV-1. We inserted the HIV-1 frameshift region in the beginning of the coding sequence of a luciferase reporter gene, such that its expression depends on a  $-1$  frameshift, and assessed the expression of luciferase in COS cells and *in vitro*, in a rabbit reticulocyte lysate (RRL). Based on site-directed mutagenesis experiments, we propose that the frameshift stimulatory signal of HIV-1 is larger than assumed, and consists of a two-stem helix, where the upper stem-loop corresponds to the classic stem-loop and the lower stem is made by base-pairing the spacer sequence to a segment following the classic stem-loop. While this study was underway, a paper was published by Dinman *et al.* (24), supporting our findings that the frameshift region of HIV-1 is longer than believed but suggesting a different structure for the frameshift stimulatory signal.

## MATERIALS AND METHODS

### Construction of plasmids

Plasmid pHIV90-luc is derived from pcDNA3.1/Hygro(+) (Invitrogen). A portion of the firefly luciferase gene encompassing the coding sequence and part of the 3' untranslated region, contained between the *Bam*HI and *Xho*I sites from pGEM<sup>®</sup>-luc (Promega), was first

introduced between the same restriction sites of pcDNA3.1/Hygro(+), and the AUG initiation codon of luciferase was removed by standard mutagenesis, generating pcDNA3.1-luc. The HIV-1 *gag/pol* frameshift region encompassing bases 2076 to 2158 of HIV-1 HXB2 molecular clone (GenBank K03455) was amplified from plasmid pHC(-1) (25), where it had been inserted in the beginning of the chloramphenicol acetyltransferase gene. The amplification used the following forward and reverse primers, respectively: 5'-TAATACGACTCACTATAGGG-3' and 5'-GCATGCTGG GGATCCTGTTGGC-3'. The resulting PCR fragment was inserted before the second codon of luciferase, between *KpnI* and *BamHI* restriction sites of pcDNA3.1-luc, generating pHIV90-luc(-1) (see details in Figure 2-1). In this construct, the *gag/pol* frameshift region corresponds to bases 8 to 90. A short region containing an AUG initiation codon precedes the HIV portion, and the HIV-luc fusion is under the control of a CMV and a T7 promoter. This construction is such that the luciferase coding sequence is in the -1 frame relative to the initiator codon, so that only ribosomes that make a -1 frameshift produce luciferase. An in-frame plasmid control, pHIV90-luc(0), was created by inserting an additional adenine at position 81 in pHIV90-luc(-1), using PCR amplification. Derivatives of pHIV90-luc (deletion and substitution mutants) were created by PCR, by first amplifying mutated DNA fragments from pHIV90-luc for the (0) and (-1) constructs with two primers for deletion mutants or mutants with substitutions in the 3' strand of the lower stem of the stimulatory signal, and with four primers for mutants with substitutions in the slippery sequence or in the 5' strand of the lower stem, according to the procedure of Ho *et al.* (26). The amplified DNA fragments were then subcloned between the *KpnI* and *BamHI* sites of pcDNA3.1-luc, and all the constructs were verified by sequencing the entire insert.

### **Transient transfections and luciferase assays**

COS1 cells were cultured in Dulbecco's modification of Eagle's medium supplemented with 10% (v/v) fetal bovine serum (BioMedia, Canada). Transfections were carried out in duplicate, using a standard calcium-phosphate precipitation method (27). Cells ( $1.5 \times 10^5$ ) were seeded in 35-mm dishes the day before transfection, 90  $\mu$ l of the calcium-phosphate

transfection precipitate containing 5  $\mu\text{g}$  of a pHIV-luc construct and 2  $\mu\text{g}$  of pcDNA3.1/Hygro/lacZ coding for  $\beta$ -galactosidase were overlaid over the cells. Cells were harvested 48 hours post-transfection, washed twice with 2 ml of PBS, and 600  $\mu\text{l}$  of the Cell Culture Lysis Reagent (Promega) were added. Following an incubation of 15 minutes at room temperature, cells were harvested and the supernatants assayed immediately for luciferase and  $\beta$ -galactosidase activity. For luciferase assays, 1.5  $\mu\text{l}$  of cell extracts was added to 100  $\mu\text{l}$  of the Luciferase Assay Reagent (Promega) and the light output was measured in relative light units with a Berthold Lumat LB 9507 luminometer. The  $\beta$ -galactosidase activity was measured with the chlorophenolred- $\beta$ -galactopyranoside substrate (Calbiochem), as described (28), with aliquots of 10  $\mu\text{l}$  of cell extracts, and used to normalize luciferase activities for variations in transfection efficiency.

### ***In vitro* transcription and translation**

*In vitro* transcriptions were carried out essentially as previously described (29), using *Stu*I-linearized pHIV-luc constructs. The RNA transcript was extracted twice with one volume of phenol/chloroform (50:50, v/v), once with one volume of chloroform/isoamyl alcohol (24:1, v/v), followed by precipitation with 1/10 volume of 3 M sodium acetate (pH 6.0) and 2.5 volumes of 100% ethanol. The RNA pellet was dissolved in water, and unincorporated nucleotide triphosphates removed by G-50 Sephadex chromatography (MicroSpin G50 column, Amersham Pharmacia Biotech). RNA was quantified by spectrophotometry and checked for integrity by electrophoresis on 1.5% (w/v) agarose gels containing 0.1% (w/v) sodium dodecyl sulfate. For translation assays, RNA transcripts were heated at 65°C for ten minutes and then briefly kept on ice prior to use, 0.2  $\mu\text{g}$  of these transcripts were translated in 25  $\mu\text{l}$  of RRL (Promega) at 30°C for 15 minutes, and the reaction was stopped by addition of EDTA at a final concentration of 6 mM. Luciferase activity was monitored as mentioned above, with 2.5  $\mu\text{l}$  of the translation mixture.

### Enzymatic probing of RNA structure

Enzymatic probing of the structure of an RNA fragment encompassing the frameshift region of HIV-1 was performed as described with minor modifications (30). An oligonucleotide cassette containing a T7 promoter followed by the HIV-1 *gag/pol* frameshift region (bases 16 to 70 according to the numbering of the construct depicted in Fig.1) was cloned between the *NaeI* and *Bsp119I* sites of pGEM®-7Zf(-) (Promega), generating the recombinant plasmid pGEM-HIV. The RNA transcript, produced by *in vitro* transcription of the *Bsp119I*-linearized plasmid, was 5'-end-labelled with [ $\gamma$ - $^{32}\text{P}$ ], using a standard dephosphorylation-rephosphorylation method (31), purified from a 10% acrylamide-7M urea gel and dissolved in 500 mM  $\text{NH}_4\text{OAc}$ , 10 mM  $\text{Mg}(\text{OAc})_2$ , 1mM EDTA and 0.1% SDS. Probing with RNase V1 (in a total volume of 10  $\mu\text{l}$  containing  $10^5$  cpm of 5'-end-labelled RNA supplemented with 1  $\mu\text{g}$  of yeast tRNA) was done at 25°C for 15 minutes in 10 mM Tris-HCl (pH 8), 10 mM  $\text{MgCl}_2$ , 100 mM KCl and 0 to 0.01 units of enzyme (Ambion). The reaction was stopped by adding an equal volume of formamide gel loading buffer, the sample heated two minutes at 95°C and immediately loaded for analysis on a 20% polyacrylamide-7 M urea gel.

## RESULTS

### Length of the HIV-1 frameshift region

Several studies on HIV-1 frameshift have been performed where the frameshift region, encompassing the slippery site, the spacer and the classic stem-loop, was inserted in the beginning of the coding sequence of a reporter gene, such that its expression requires a translational frameshift (32,33). We decided to investigate whether the frameshift region that was previously studied contains all the signals required for maximal frameshifting and examined whether the sequence following the classic stem-loop could influence the

frameshift. To this end, we made a HIV-1 frameshift reporter construct, using as a vector the pcDNA3.1/Hygro(+) plasmid, that contains a CMV and a T7 promoter, allowing us to assess the expression of a reporter gene in cultured cells and in a RRL, respectively. The firefly luciferase (*luc*) reporter gene was first introduced into this plasmid, and the HIV-1 frameshift region, longer by 30 bases than the classic frameshift region, was inserted in the beginning of the *luc* coding sequence. This generated plasmid pHIV90-luc(-1) (see Figure 2-1 and Materials and Methods). This HIV-1 insertion is such that expression of the LUC protein requires a -1 frameshift. To quantify the frameshift efficiency, an «in-frame» version (pHIV90-luc(0)) was derived from pHIV90-luc(-1), in which an additional adenine base was inserted upstream of the luciferase coding sequence so that expression of the reporter does not require a frameshift event. Linearization of the plasmid with *Stu*I and transcription with T7 RNA polymerase generate a transcript of about 1.9 kb, which was translated in a RRL. For assays in cultured cells, constructs were transfected into COS cells and luciferase levels were measured 48 hours post-transfection. In both types of experiments, the frameshift efficiency was calculated by determining the ratio of the luciferase activity with the (-1) construct to the luciferase activity of the (0) construct plus that of the (-1) construct, assuming that the same level of frameshift should take place in the in-frame and in the (-1) construct. Different deletions were made in the frameshift region, generating pHIV60-luc(-1), which corresponds to the insertion of the classic frameshift region, ending after the stem-loop, and pHIV69-luc(-1), which contains the classic frameshift region plus ten bases after the stem-loop (see Figure 2-2(a)). For each of these constructs, an in-frame (0) control was also created. Results from experiments made in cultured cells and *in vitro* are presented in Figure 2(b). As a control, mutations that altered the slippery sequence were made in pHIV90-luc to generate pHIV-k/o-luc. In cultured COS cells, these mutations almost abolished the frameshift (0.1%). The level of frameshifting in cultured cells was about 2.5% in the presence of the classic stem-loop (pHIV60-luc), a value in full agreement with previous studies using reporters where the classic frameshift region of HIV-1 was inserted in the beginning of the coding sequence and whose expression depends upon a -1 frameshift (32,33). It was reduced to 0.2% in the

absence of the stem-loop (pHIV-DSL-luc), stressing the importance of the classic stem-loop for the frameshift event. When frameshifting was assessed with the construct containing a longer frameshift region (pHIV90-luc), the frameshift efficiency was increased to 4.8% in cultured cells, two-fold higher than with the classic frameshift region. With construct pHIV69-luc, where the classic frameshift region was extended by a small pyrimidine-rich ten-base segment, the frameshift efficiency (4.6%) was identical to the frameshifting level observed with construct pHIV90-luc. These results indicate that a segment of ten bases adjacent to the classic stem-loop contributes to increase the level of frameshift, showing that the frameshift region is longer than believed. Similar values for frameshift efficiencies were obtained when transfecting another cell line, 293T (data not shown). Moreover, the frameshift efficiencies were independent of the reporter gene, since replacing the luciferase reporter with the green fluorescent protein (GFP) reporter provided identical results (data not shown). Our findings that the frameshift region of HIV-1 is longer than believed are thus in perfect agreement with the recent study of Dinman *et al.* (24). The same constructs that were used in cultured cells were also assayed *in vitro* in a RRL. With the construct containing the conventional frameshift region, it was found that the classic stem-loop influences the frameshift efficiency but to a lesser extent than in cultured cells. Indeed, the frameshift level was 9.5% in its presence and decreased to 3.7% in its absence. With mutations that alter the slippery sequence, the frameshift efficiency was 1%, suggesting the possibility of aberrant initiation events. Lengthening the frameshift region resulted in a frameshift efficiency of 8.8% and 10.4% (constructs pHIV69-luc and pHIV90-luc, respectively), showing no significant variation, compared to the value of 9.5% obtained with the classic frameshift region. It has already been observed that the slippery sequence of HIV-1 is very shifty on its own *in vitro* (4 to 5%) compared to other natural slippery sequences, for which the frameshift efficiency is 1% or less (reviewed in 34). Previous reports have also shown that *in vitro* the frameshift of HIV-1 is more efficient and less sensitive to the presence of the stimulatory signal, compared to the situation in cultured cells (33,35). It has been proposed that the lower *in vitro* rate of translation accounts for the higher efficiency of the frameshift promoted by the slippery sequence of HIV-1 on its own

(35), and this higher frameshift efficiency probably minimizes the effect of the frameshift stimulator. The weak effect of the classic stimulator on the frameshift efficiency, as shown by our results and in agreement with previous reports, led us to conclude that an *in vitro* system is not sensitive enough to investigate the length of the frameshift region and the structure of the stimulatory signal of HIV-1. Therefore, the following experiments were performed only in cultured cells.

### **Proposed structure for the frameshift stimulatory signal of HIV-1**

Our results reveal that, in cultured cells, a pyrimidine-rich segment downstream of the classic stem-loop enhances the frameshift efficiency for constructs containing an extended frameshift region of HIV-1. To account for the contribution of the downstream segment, we investigated the structure of an RNA fragment encompassing the *gag/pol* frameshift region of HIV-1. Enzymatic probing with RNase V1, an enzyme that cleaves RNA in helical conformation, showed that the purine-rich spacer region following the slippery sequence and preceding this classic stem-loop and the pyrimidine-rich downstream segment were attacked by the enzyme (Figure 2-3). Interestingly, a similar observation had been made by Aupeix *et al.* (36). Based on these observations, we suggested that the frameshift region folds into a two-stem helix, made of an upper stem-loop, corresponding to the classic stem-loop, and a lower stem, formed by pairing the purine-rich spacer region to the pyrimidine-rich segment downstream of the classic stem-loop. A three-purine bulge interrupts these two stems (Figure 2-3b). The thermodynamic stability of the novel frameshift stimulatory signal is  $-24.6$  kcal/mol, as predicted by the MFOLD program at Michael Zucker's MFOLD server at <http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi> (37), whereas the classic frameshift signal has a thermodynamic stability of  $-21.3$  kcal/mol. We then reasoned that if the lower stem contributes to stimulate frameshifting, mutations that prevent the formation of this stem should decrease the frameshift efficiency.

## Mutagenesis studies

To further investigate the structure that we propose for the frameshift stimulatory signal of HIV-1, we, therefore, introduced mutations in the two strands of the putative lower stem, that individually disrupt this stem, but when present simultaneously allow its re-formation (Figure 2-4(a)). Frameshift efficiencies with the different constructs are shown in Figure 4(b). We first replaced five bases of the 5' strand of the putative lower stem of the HIV-1 frameshift stimulator with five bases of the 3' strand of this stem, generating construct pHIV1.1-luc. This mutation was found to reduce frameshifting about three-fold in cultured cells (1.6% frameshifting), near the level obtained with the construct containing the classic frameshift region. Similarly, the frameshift efficiency was decreased two-fold when replacing five bases in the 3' strand of the lower stem with five bases of the 5' strand (pHIV1.2-luc; 2.3% frameshifting). However, when both strands were mutated simultaneously, allowing re-formation of the lower stem, the frameshift was enhanced, although still inferior to the wild-type level (pHIV1.12-luc; 3.8% frameshifting). A second series of complementary and compensatory base-pair change mutations were also made in the putative lower stem. These mutations were such that they can create four new G-C pairs in the lower stem, so as to increase its stability (Figure 2-4(a)). In the compensatory mutants, while pHIV1.12-luc has the same thermodynamic stability as pHIV90-luc ( $\Delta G^\circ = -24.6$  kcal/mol), this stability is significantly increased in pHIV2.12-luc ( $\Delta G^\circ = -32.9$  kcal/mol). Assays with this second series of mutations in the lower stem confirmed the results obtained with the first series of mutations, showing that changes in either the 5' strand or the 3' strand of the lower stem reduced frameshifting level almost two-fold (pHIV2.1-luc and pHIV2.2-luc; 2.3% and 2.1%, respectively), and that the frameshift was restored to a wild-type value when both strands were mutated so as to re-form a lower stem (pHIV2.12-luc; 4.8%). It can be observed that the frameshift efficiency was not higher with the more stable lower stem. We also investigated whether the three-purine bulge separating the two stems participates to the frameshift event. To address this question, mutations were

made that replaced the three purines of the bulge with pyrimidines (pHIV-Bulge-luc). We found that this mutation reduced the frameshifting level about two-fold (2.5% compared to 4.8% for pHIV90-luc), bringing it to the level observed when only the classic stem-loop is present and showing that the bulge contributes to the frameshift.

### **Conservation of the two-stem helix among HIV-1 isolates**

Since the frameshift efficiency is critical for maintenance of the Gag-Pol to Gag ratio that allows viral replication, the structure of the frameshift stimulatory signal should be highly conserved in HIV-1 natural isolates. To verify the capacity of the spacer following the slippery sequence to base-pair with the pyrimidine-rich region downstream of the classic stem-loop, we analyzed 139 complete *pol* sequences retrieved from the Los Alamos National Laboratory HIV Sequence Database (38) (Table 1). This analysis shows that the proposed duplex structure can be formed in 96.4% of the sequences, with 58.3% of them conserving the capacity to form the proposed structure with only Watson-Crick or G-U base-pairs and 38.1% containing non-Watson-Crick A-C base-pairs (35.2% and 2.9% with one and two A-C base-pairs, respectively), which are known not to distort significantly the geometry of an RNA helix (39). Only 3.6% of the sequences analyzed contain mismatches able to distort the RNA helix. Therefore, the analysis of sequences strongly supports our suggestion that the capacity to form the lower stem in the structure that we proposed for the frameshift stimulatory signal of HIV-1 is evolutionary conserved. As mentioned in the Introduction, a different structure was recently proposed for the frameshift stimulatory signal of HIV-1 by Dinman *et al.* (24). In their model, the pyrimidine-rich sequence downstream of the classic stem-loop interacts with three bases in the loop capping this stem through Watson-Crick interactions and forms a triple helix with four base-pairs on top of this stem. Analysis of the *pol* sequences retrieved from the Los Alamos National Laboratory HIV Sequence Database does not contradict their model, as shown in Table 1, if one assumes that a U can replace a C in the third strand of the triplex and interact with the G of a C-G pair without destabilizing the triplex. This, however, is not the case (39-41),

especially in a short triplex formed with only four base-pairs where three U\*G-C triple interactions are proposed. If, nevertheless, we accept the assumption of Dinman *et al.*, 92.8% of the sequences could form the triple helix, with 41% of them conserving the capacity to form the triplex structure and 51.8% containing changes that minimally affect the capacity to form this structure. Only 7.2% of the sequences could not form this structure. However, a major criticism that we address to the model of Dinman *et al.* is that the pyrimidine-rich third strand that forms the triplex is in an antiparallel orientation relative to the purine-rich strand of the duplex with which it interacts, an orientation that is not sterically favored (see 40-42), making this structure highly improbable. Moreover, Dinman *et al.* claim that structural probing supports their model but this probing was done with a short fragment of the frameshift region of HIV-1 encompassing only the classic stem-loop and the downstream sequence. When a longer fragment encompassing in addition the slippery sequence and the spacer region was probed, this spacer region and the downstream segment to which we propose that it base-pairs were found to be sensitive to RNase V1, an endonuclease that cleaves RNA in helical conformation (Figure 2-3 and Ref. 36). This probing fully supports the lower stem of the structure that we propose for the frameshift stimulatory signal of HIV-1. Finally, it can be observed that our mutants pHIV1.12-luc and pHIV2.12-luc have frameshift efficiencies near wild-type level although they cannot participate in the structure proposed by Dinman *et al.*

## DISCUSSION

Our results show that the HIV-1 frameshift stimulatory signal folds into a more complex structure than commonly assumed. A two-stem structure was proposed where the sequences flanking the classic stem-loop base-pairs, thus forming the lower stem of an extended helix. This structure was proposed from structure probing of the frameshift stimulatory signal and supported by mutagenesis studies. Indeed, mutations that disrupted base-pairing between the sequences forming the lower stem of our proposed structure (pHIV1.1-luc and

pHIV1.2-luc; pHIV2.1-luc and pHIV2.2-luc) reduced frameshift efficiencies. It could be argued that sequences flanking the stem-loop influence frameshifting by modulating stem-loop unfolding kinetics (see 43). However, compensatory mutations that enable re-formation of the lower stem, by exchanging the sequence of each strand, restored frameshifting near the wild-type level (pHIV1.12-luc and pHIV2.12-luc), and this could not result from a simple effect of mutations in the flanking sequences on the unfolding kinetics of the classic stem-loop. As mentioned in the Introduction, the function of the stimulatory signal is not completely well-understood, and it is thought that specific interactions between the ribosome and the frameshift stimulatory signal are required to achieve efficient frameshifting. The upper stem of the stimulator obviously plays a major role in the interaction with the ribosome since its presence strongly increases frameshifting, compared to what is observed when only the slippery sequence is present, whereas the lower stem only contributes to enhance frameshifting about two-fold. However, as indicated above, the frameshift efficiency controls the Gag-Pol to Gag ratio, and small variations in frameshifting can dramatically affect the virus propagation. Indeed, studies with the yeast L-A double-stranded RNA virus, for which the synthesis of Gag-Pol also requires a programmed  $-1$  frameshift, showed that either increasing or decreasing the frameshift efficiency by more than two-fold, by altering the slippery sequence, disrupted viral propagation (8). Also, a three-fold increase in the Gag-Pol to Gag ratio in HIV-1 producing cells, achieved by cotransfecting HIV-1 proviral DNA with an HIV-1 Gag-Pol expression vector, resulted in a ten-fold decrease in virion infectivity (4). The novel frameshift stimulatory structure is only slightly more stable than the classic signal, and it is unlikely that its enhancing effect results from this increased stability. Indeed, it was even observed that increasing the stability of the classic stem-loop by lengthening the stem could reduce frameshifting (44). Our data show the importance of the three-purine bulge interrupting the two stems in the HIV-1 frameshift stimulatory signal. The formation of this bulge requires the presence of the lower stem and substitution of the purines of the bulge with pyrimidines decreased frameshifting to the level observed in the absence of the lower stem. Although purely speculative, we suggest that this bulge by interacting with the ribosome through

purine-mediated interactions could contribute to enhance the interaction of the upper stem with the ribosome.

A feature that differentiates the frameshift stimulatory signal in HIV-1 from other signals is that it is distant from the slippery site by one base only. In most cases where frameshifting occurs, there is a spacer sequence of five to eight bases between the slippery sequence and the stimulatory signal (5). However, in the Rous sarcoma virus (RSV) RNA, that uses a programmed  $-1$  frameshift, the frameshift stimulator, which is a pseudoknot, is also located one base after the slippery sequence (30). Since frameshift takes place when the slippery sequence occupies the A and P sites of the ribosome, either the RSV frameshift stimulatory signal or the stimulatory signal that we propose for HIV-1 would have to evade the unwinding activity of the translating ribosome or re-fold once inside the ribosome. Interestingly, the possibility that a signal stimulating a recoding event can evade the ribosome unwinding activity or re-fold within the ribosome has also been proposed for the ribosomal hopping that occurs during translation of the T4 gene 60 (45). In this recoding event, a stem-loop in close proximity to the decoding site stimulates tRNA slippage, implying that this structure is present within the ribosome.

In conclusion, we propose a refinement of the structure of the HIV-1 frameshift stimulatory signal, that consists in increasing the size and complexity of this signal. A complete characterization of the signal that promotes frameshifting in HIV and of its interaction with the ribosome will be extremely useful in providing valuable information for the development of antiviral agents that target programmed  $-1$  ribosomal frameshift.

#### **ACKNOWLEDGEMENTS**

This study was supported by a grant from the Canadian Institutes for Health Research. We thank Guy Lemay for his interest in this work. We are most grateful to Sergey Steinberg for stimulating discussions and comments. We are also grateful to Pascal Chartrand, Luc Desgroseillers and Francis Robert for critical reading of this manuscript.

**Abbreviations used:**

HIV-1, human immunodeficiency virus type 1; RRL, rabbit reticulocyte lysate; CMV, cytomegalovirus; RSV, Rous sarcoma virus; PCR, polymerase chain reaction.

**REFERENCES**

1. Jacks,T., Power,M.D., Masiarz,F.R., Luciw,P.A., Barr,P.J. and Varmus,H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature*, **331**, 280-283.
2. Park,J. and Morrow,C.D. (1991) Overexpression of the *gag-pol* precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. *J. Virol.*, **65**, 5111-5117.
3. Karacostas,V., Wolffe,E.J., Nagashima,K., Gonda,M.A. and Moss,B. (1993) Overexpression of the HIV-1 *gag-pol* polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology*, **193**, 661-671.
4. Shehu-Xhilaga,M., Crowe,S.M. and Mak,J. (2001) Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.*, **75**, 1834-1841.
5. Brierley,I. (1995) Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.*, **76**, 1885-1892.
6. Brierley,I. and Pennell,S. (2001) Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. In *The Ribosome, 66<sup>th</sup> Cold Spring Harbor Symposium on Quantitative Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 233-248.
7. Tzeng,T.H., Tu,C.L. and Bruenn,J.A. (1992) Ribosomal frameshifting requires a pseudoknot in the *Saccharomyces cerevisiae* double-stranded RNA virus. *J. Virol.*, **66**, 999-1006.

8. Dinman,J.D. and Wickner,R.B. (1992) Ribosomal frameshifting efficiency and *gag/gag-pol* ratio are critical for yeast M<sub>1</sub> double-stranded RNA virus propagation. *J. Virol.*, **66**, 3669-3676.
9. Tsuchihashi,Z. and Brown,P.O. (1992) Sequence requirements for efficient translational frameshifting in the *Escherichia coli dnaX* gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev.*, **6**, 511-519.
10. Chandler,M. and Fayet,O. (1993) Translational frameshifting in the control of transposition in bacteria. *Mol. Microbiol.*, **7**, 497-503.
11. Larsen,B., Wills,N.M., Gesteland,R.F. and Atkins,J.F. (1994) rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J. Bacteriol.*, **176**, 6842-6851.
12. Gesteland,R.F. and Atkins,J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.*, **65**, 741-768.
13. Shigemoto,K., Brennan,J., Walls,E., Watson,C.J., Stott,D., Rigby,P.W. and Reith,A.D. (2001) Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting. *Nucleic Acids Res.*, **29**, 4079-4088.
14. Jacks,T., Madhani,H.D., Masiarz,F.R. and Varmus,H.E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell*, **55**, 447-458.
15. Farabaugh,P.J. (2000) Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.*, **64**, 131-170.
16. Somogyi,P., Jenner,A.J., Brierley,I. and Inglis,S.C. (1993) Ribosomal pausing during translation of an RNA pseudoknot. *Mol. Cell. Biol.*, **13**, 6931-6940.
17. Tu,C., Tzeng,T.H. and Bruenn,J.A. (1992) Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc. Natl. Acad. Sci. USA*, **89**, 8636-8640.
18. Lopinski,J.D., Dinman,J.D. and Bruenn,J.A. (2000) Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol. Cell. Biol.*, **20**, 1095-1103.
19. Kontos,H., Naphthine,S. and Brierley,I. (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell. Biol.*, **21**, 8657-8670.

20. Shen,L.X. and Tinoco,I., Jr. (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.*, **247**, 963-978.
21. Kang,H. and Tinoco,I.,Jr. (1997) A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. *Nucleic Acids Res.*, **25**, 1943-1949.
22. Le,S.Y., Chen,J.H. and Maizel,J.V. (1989) Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses. *Nucleic Acids Res.*, **17**, 6143-6152.
23. Kang,H. (1998) Direct structural evidence for formation of a stem-loop structure involved in ribosomal frameshifting in human immunodeficiency virus type 1. *Biochim. Biophys. Acta*, **1397**, 73-78.
24. Dinman,J.D., Richter,S., Plant,E.P., Taylor,R.C., Hammell,A.B. and Rana,T.M. (2002) The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure. *Proc. Natl. Acad. Sci. USA*, **99**, 5331-5336.
25. Doyon,L., Payant,C., Brakier-Gingras,L. and Lamarre,D. (1998) Novel Gag-Pol frameshift site in human immunodeficiency virus type 1 variants resistant to protease inhibitors. *J. Virol.*, **72**, 6146-6150.
26. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.K. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.
27. Jordan,M., Schallhorn,A. and Wurm,F.M. (1996) Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res.*, **24**, 596-601.
28. Eustice,D.C., Feldman,P.A., Colberg-Poley,A.M., Buckery,R.M. and Neubauer,R.H. (1991) A sensitive method for the detection of beta-galactosidase in transfected mammalian cells. *Biotechniques*, **11**, 739-743.
29. Dragon,F., Payant,C. and Brakier-Gingras,L. (1994) Mutational and structural analysis of the RNA binding site for *Escherichia coli* ribosomal protein S7. *J. Mol. Biol.*, **244**, 74-85.

30. Marczinke,B., Fisher,R., Vidakovic,M., Bloys,A.J. and Brierley,I. (1998) Secondary structure and mutational analysis of the ribosomal frameshift signal of Rous sarcoma virus. *J. Mol. Biol.*, **284**, 205-225.
31. Knapp,G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.*, **180**, 192-212.
32. Cassan,M., Delaunay,N., Vaquero,C. and Rousset,J.P. (1994) Translational frameshifting at the *gag-pol* junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J. Virol.*, **68**, 1501-1508.
33. Reil,H., Kollmus,H., Weidle,U.H. and Hauser,H. (1993) A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J. Virol.*, **67**, 5579-5584.
34. Brierley,I., Jenner,A.J. and Inglis,S.C. (1992) Mutational analysis of the «slippery-sequence» component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.*, **227**, 463-479.
35. Parkin,N.T., Chamorro,M. and Varmus,H.E. (1992) Human immunodeficiency virus type 1 *gag-pol* frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J. Virol.*, **66**, 5147-5151.
36. Aupeix,K., Le Tinevez,R. and Toulmé,J.J. (1999) Binding of oligopyrimidines to the RNA hairpin responsible for the ribosome *gag-pol* frameshift in HIV-1. *FEBS Lett.*, **449**, 169-174.
37. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911-940.
38. *Human Retroviruses and AIDS* (2002) *A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos National Laboratory, Los Alamos, NM, 2002, <http://hiv-web.lanl.gov>.
39. Nowakowski,J. and Tinoco,I.,Jr. (1999) RNA structure in solution. In Neidle,S., (ed.), *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press, NY, pp. 567-602.
40. Han,H. and Dervan,P.B. (1993) Sequence-specific recognition of double helical RNA and RNA•DNA by triple helix formation. *Proc. Natl. Acad. Sci. USA*, **90**, 3806-3810.

41. Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65-95.
42. Lim, A.C. and Barton, J.K. (1998) Rh(Phen)<sub>2</sub>Phi<sup>3+</sup> as a shape-selective probe of triple helices. *Biochemistry*, **37**, 9138-9146.
43. Kim, Y-G., Mass, S. and Rich, A. (2001) Comparative mutational analysis of *cis*-acting RNA signals for translational frameshifting in HIV-1 and HTLV-2. *Nucleic Acids Res.*, **29**, 1125-1131.
44. Bidou, L., Stahl, G., Grima, B., Liu, H., Cassan, M. and Rousset, J.P. (1997) *In vivo* HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. *RNA*, **3**, 1153-1158.
45. Herr, A.J., Gesteland, R.F. and Atkins, J.F. (2000) One protein from two open reading frames: mechanism of a 50 nt translational bypass. *EMBO J.*, **19**, 2671-2680.

Tableau 2-1 : Conservation of the proposed structure for the frameshift stimulatory signal among HIV-1 isolates.

Description of the sequence	Proposed structure	
	Two-stem helix <sup>a</sup>	Triplex helix <sup>b</sup>
Total of sequences that do not affect the capacity to form the structure	81 (58.3%)	57 (41.0%)
Consensus sequence	19 (13.7%)	13 (9.3%)
Changes that do not affect the capacity to form the structure <sup>c</sup>	62 (44.6%)	44 (31.7%)
Changes that have a minimal impact on the capacity to form the structure <sup>d</sup>	53 (38.1%)	72 (51.8%)
Total changes that do not or minimally affect the capacity to form the structure	134 of 139 (96.4%)	129 of 139 (92.8%)
Mismatches that disrupt the capacity to form the structure <sup>e</sup>	5 of 139 (3.6%)	10 of 139 (7.2%)

One-hundred-thirty-nine complete *pol* sequences from HIV-1 were retrieved from the Los Alamos National Laboratory HIV Sequence Database (38), and analyzed manually for the capacity to form the proposed frameshift stimulatory signals. The 139 sequences include natural isolates of HIV-1 group M (main), the most widespread group in the world (38). Sequences belonging to members of divergent lineages such as group O (outlier) (four sequences) and N (non-M, non-O) (two sequences) have not been included in the analysis. <sup>a</sup>Two-stem helix proposed in this study. <sup>b</sup>Triplex helix proposed by Dinman *et al.* (24). <sup>c</sup>Sequential changes allowing the formation of Watson-Crick and G-U base-pairs and, in the model of Dinman *et al.*, conserving the capacity to form the triplex structure. <sup>d</sup>Sequential changes allowing the formation of A-C base-pairs that minimally distort RNA double helices. <sup>e</sup>Mismatches (U-U, A-A, C-C, G-A, U-C, C-U and C-A base-pairs) that disrupt double-helical structures and, in the model of Dinman *et al.*, impair the capacity to form the triplex structure.

Figure 2-1: Description of pHIV90-luc, a frameshift reporter construct. Plasmid pHIV90-luc is a derivative of pcDNA3.1/Hygro(+) (Invitrogen) in which a *Bam*HI/*Xho*I fragment encoding the firefly luciferase gene sequence has been added, with an insertion at the beginning of the coding sequence corresponding to the HIV-1 *gag/pol* frameshift region (see Materials and Methods). The *Kpn*I and *Bam*HI sites were used for subcloning the HIV-1 sequence between the vector and the luciferase sequences. The AUG initiator codon is italicized. Bases from positions 8 to 90 (in bold) originate from HIV-1. The slippery site is UUUUUUA (underlined). It is followed by a spacer, the classic frameshift stimulatory signal (a stem-loop) and an additional 30 nucleotide sequence. The figure presents construct (-1), which produces luciferase only when ribosomes make a -1 frameshift. Addition of an adenine at position 81 corresponds to construct (0), which produces luciferase when ribosomes do not shift the reading frame.

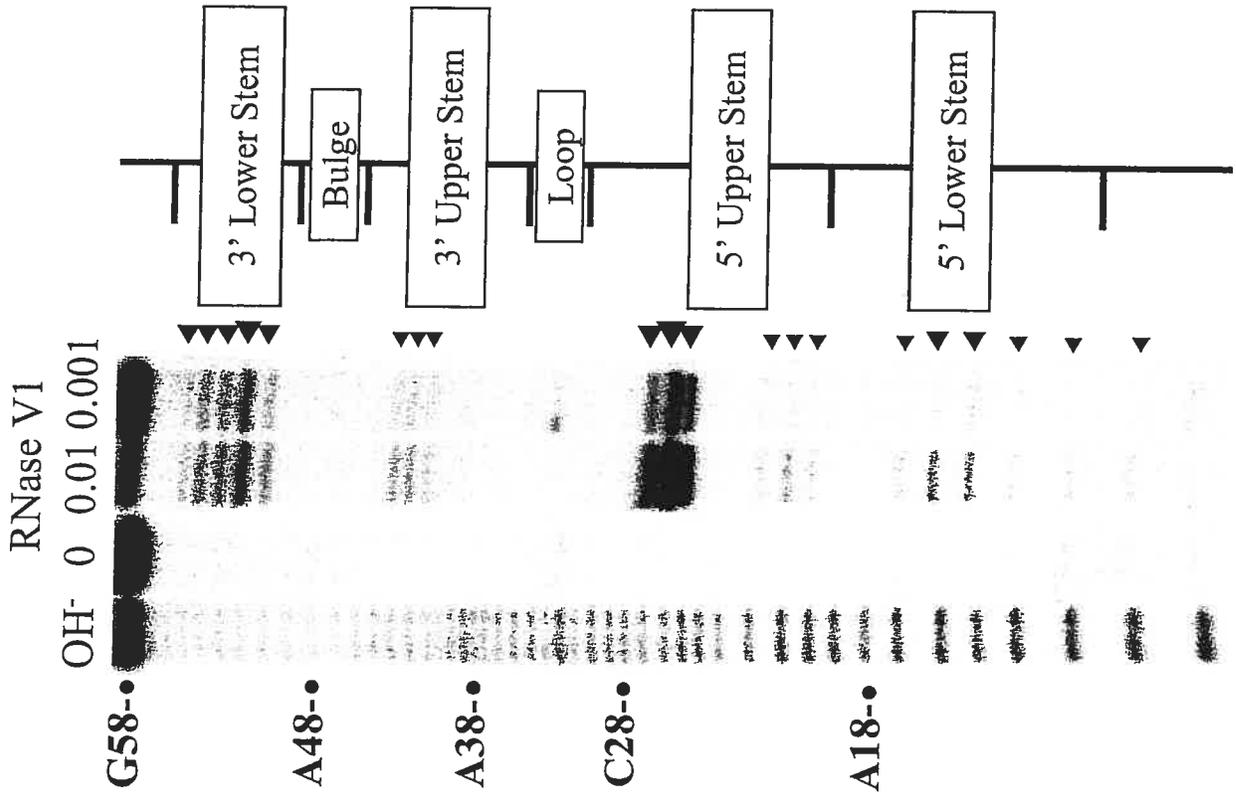


Figure 2-2: Effect on frameshifting of mutations in the HIV-1 frameshift region. (a) A series of mutations were made within the frameshift region of pHIV90-luc (the dots correspond to the *Bam*HI linker connecting the frameshift region to the luciferase coding sequence): deletion mutants pHIV69-luc and pHIV60-luc, where the region 3' to the classic stem-loop is shortened; a slippery site mutant, pHIV-k/o-luc, where the slippery sequence (underlined) is mutated (bases that are changed are in capital letters); a deletion mutant, pHIV-DSL-luc, where the classic stem-loop is eliminated (deletion of bases 31 to 60). (b) Frameshift efficiency *in vitro* and in cultured cells with the pHIV-luc constructs described above. *In vitro* translation experiments were made in 25  $\mu$ l of RRL with 0.2  $\mu$ g of mRNAs transcribed from *Stu*I-digested pHIV90-luc and mutant constructs. Assays in cultured cells were made by cotransfecting COS1 cells with 5  $\mu$ g of pHIV-luc and 2  $\mu$ g of pcDNA3.1/Hygro/lacZ. Frameshift efficiencies were determined as described in the text. Each value represents the mean  $\pm$  standard derivation of five to six independent experiments. The bars indicate the standard deviation on the means.



Figure 2-3: Novel structure proposed for the frameshift stimulatory signal of HIV-1. (a) Structure probing of the frameshift stimulatory signal by RNase V1 attack. An RNA transcript encompassing the HIV-1 *gag/pol* frameshift region was 5'end-labeled with [ $\gamma$ - $^{32}$ P] and digested with RNase V1. Digestion products were analyzed on a 20% acrylamide-7 M urea gel. The sites of cleavage were identified by comparison with a ladder of bands created by limited alkaline hydrolysis of the RNA (OH<sup>-</sup>) and by the position of RNase T1 cuts (not shown). Uniquely cleaved nucleotides were identified by their absence in the untreated control lane (0). The amount of units of enzyme added to each reaction is also indicated. (b) Description of the novel two-stem model for the frameshift stimulatory signal as suggested by structure probing. The upper stem corresponds to the classic stem-loop and the lower stem is formed by pairing the spacer to a segment downstream of this stem-loop. The sensitivity of nucleotides in the HIV-1 frameshift region to RNaseV1 is shown. The size of the arrows is approximately proportional to the intensity of the cleavage at that site. Bases in bold originate from HIV-1.

A



B

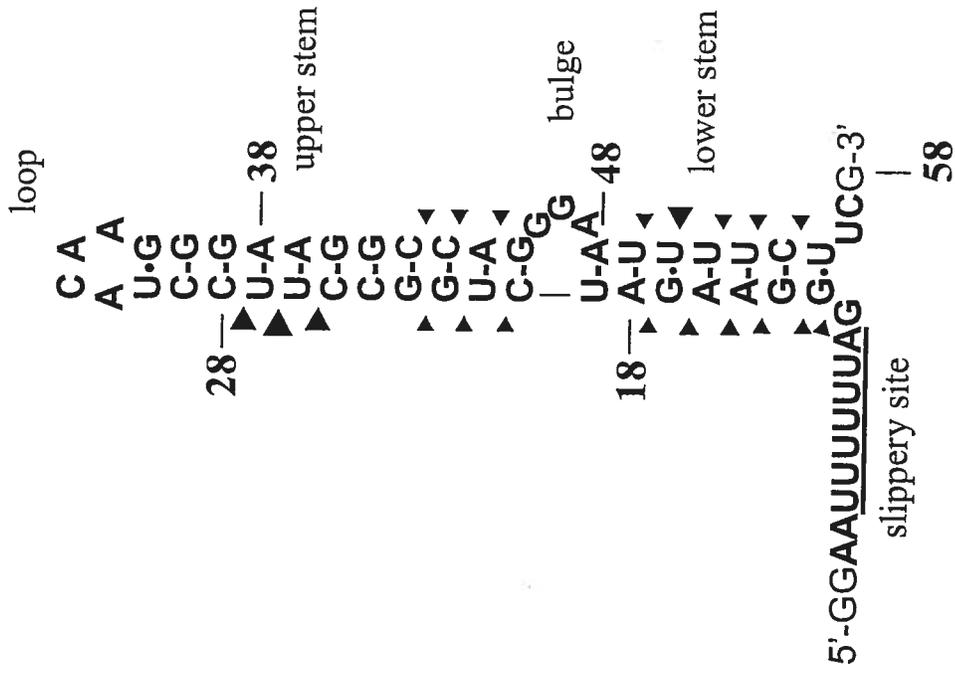
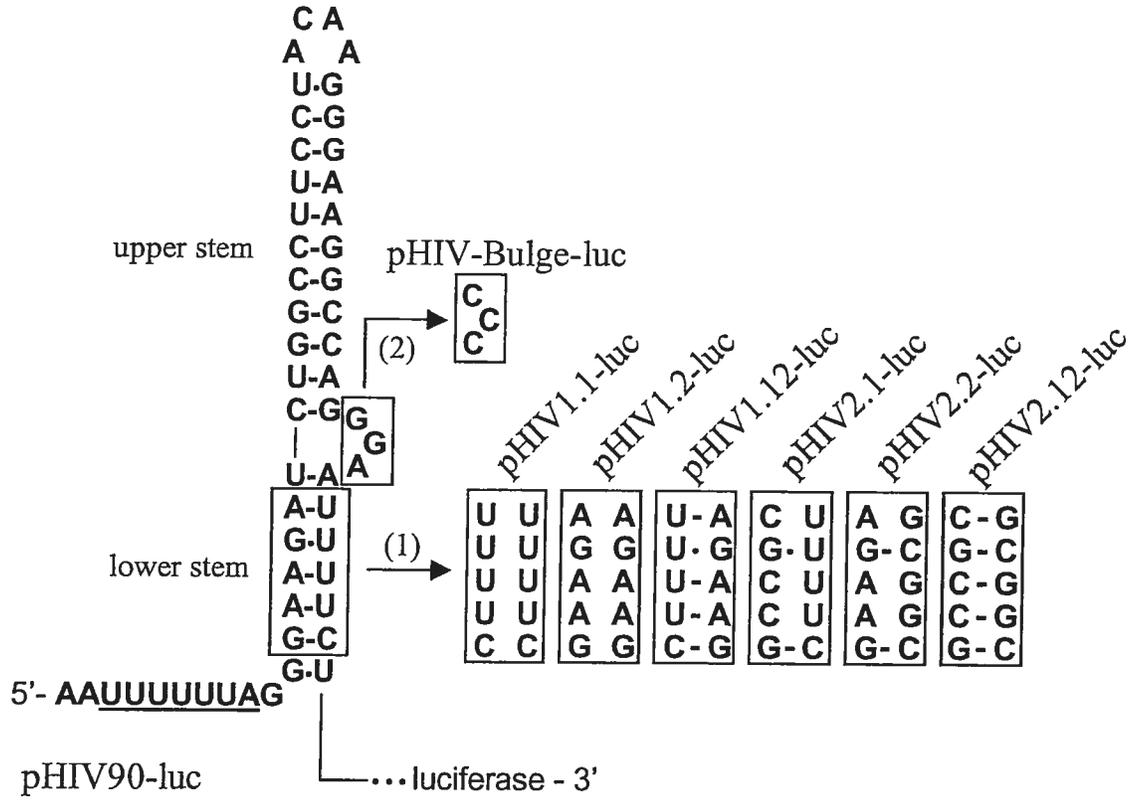
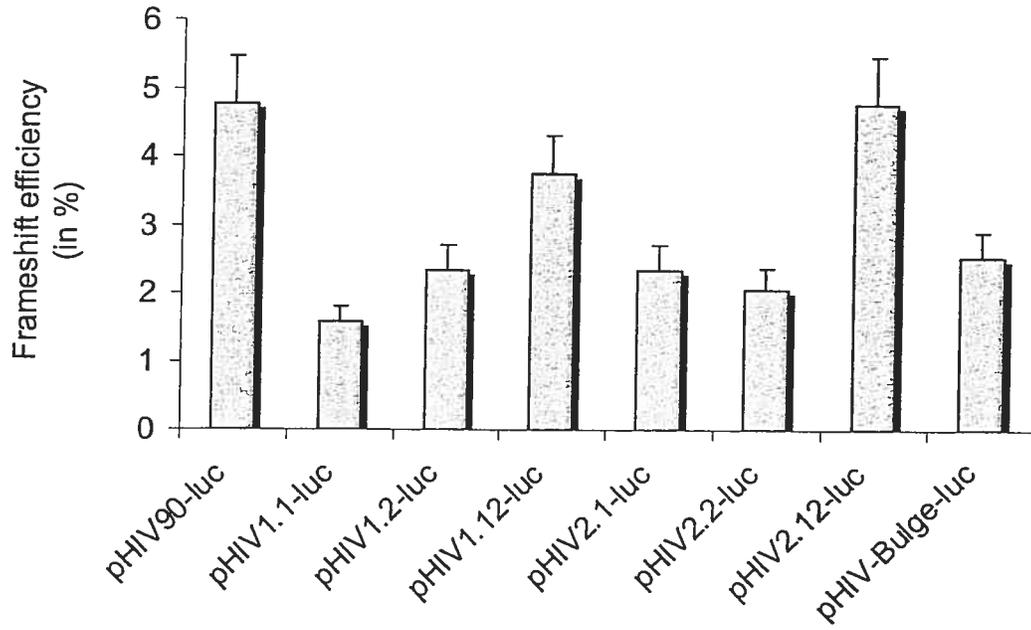


Figure 2-4: Effect on frameshifting of mutations in the lower stem of the proposed frameshift stimulatory signal. (a) Description of mutations made within the *gag/pol* frameshift region of pHIV90-luc. (1) Two series of mutants (the pHIV1 and pHIV2 series) were made in the lower stem. For each series, mutations were introduced either in the 5' or 3' strand of the lower stem, impairing the formation of this stem (pHIV1.1-luc and pHIV1.2-luc; pHIV2.1-luc and pHIV2.2-luc), or allowing re-formation of this stem (pHIV1.12-luc and HIV2.12-luc). (2) The three purines forming the bulge separating the two stems were substituted with pyrimidines (pHIV-Bulge-luc). (b) Frameshift efficiency in cultured cells with the pHIV-luc constructs described above. Assays were as described in the legend to Figure 2.

**A**



**B**



## **Chapitre 3**

Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1.

Dominic Dulude<sup>1</sup>, Yamina A Berchiche<sup>1,2</sup>, Karine Gendron<sup>1</sup>, Léa Brakier-Gingras<sup>1\*</sup>  
& Nikolaus Heveker<sup>1,2\*</sup>.

(*Virology*, acceptable avec revision dans le texte)

<sup>1</sup>Département de Biochimie, Université de Montréal, 2900 Boul. Édouard-Montpetit, Montréal, Québec, H3T1J4 Canada.

<sup>2</sup>Centre de Recherche, Hôpital Sainte-Justine, 3175 Chemin de la Côte Sainte-Catherine, Montréal, Québec, H3T1C5 Canada.

\*correspondent footnote :

Léa Brakier-Gingras : [REDACTED] Phone : 514 343 6316; Fax : 514 343 2210 ;

Nikolaus Heveker : [REDACTED] Phone : 514 345 4931 ext. 4190; Fax : 514 345 4801.

Key-words: HIV-1; -1 ribosomal frameshifting; Gag-Pol

**Abstract**

The Gag-Pol polyprotein of the human immunodeficiency virus type 1 (HIV-1) is the precursor of the virus enzymatic activities and is produced via a programmed  $-1$  translational frameshift. In this study, we altered the frameshift efficiency by introducing mutations within the slippery sequence and the frameshift stimulatory signal, the two elements that control the frameshift. These mutations decreased the frameshift efficiency to different degrees, ranging from  $\sim 0.3$  to 70% of the wild-type efficiency. These values were mirrored by a reduced incorporation of Gag-Pol into virus-like particles, as assessed by a decrease in the reverse transcriptase activity associated to these particles. Analysis of Gag processing in infectious mutant virions revealed processing defects to various extents, with no clear correlation with frameshift decrease. Nevertheless, the observed frameshift reductions translated into equivalently reduced viral infectivity and replication kinetics. Our results show that even moderate variations in frameshift efficiency, as obtained with mutations in the frameshift stimulatory signal, reduce viral replication. Therapeutic targeting of this structure may therefore result in the attenuation of virus replication and in clinical benefit.

## Introduction

The human immunodeficiency virus type 1 (HIV-1) encodes its structural and enzymatic components from the overlapping *gag* and *pol* open reading frames, respectively (reviewed in Frankel and Young, 1998; Freed and Martin, 2001; Tang et al., 1999). When ribosomes of infected cells translate the full-length viral messenger RNA, Gag is synthesized according to conventional rules of decoding, whereas Pol is produced from the same messenger as a Gag-Pol precursor via a programmed  $-1$  ribosomal frameshift event. The frequency of this event was assessed at 2 to 10%, depending upon the system used to measure it (Biswas et al., 2004; Cassan et al., 1994; Dulude et al., 2002; Grentzmann et al., 1998; Harger and Dinman, 2003; Jacks et al., 1988; Reil et al., 1993). The programmed  $-1$  ribosomal frameshift is widely used by retroviruses to produce Gag-Pol (reviewed in Brierley, 1995). It allows a precise control of the Gag-Pol to Gag ratio and also ensures the incorporation of the viral enzymes upon assembly of the viruses due to the fusion of the Pol sequence to Gag. The HIV-1 Gag polyprotein, p55, is the precursor of various endproducts that result from proteolytic breakdown by the HIV-1 protease. These are the matrix (MA or p17), the capsid (CA or p24), the p2 spacer peptide (SP1), the nucleocapsid (NC or p7), the p1 spacer peptide (SP2) and p6 (reviewed in Frankel and Young, 1998; Freed and Martin, 2001; Tang et al., 1999). Gag-Pol shares proteins MA to NC with Gag, but not p1 and p6 because of the frameshift event that overlaps with the p1 reading frame. These two proteins are replaced in Gag-Pol by a transframe (TF) octapeptide and p6\*, the roles of which are still controversial (see Hill et al., 2002; Paulus et al., 2004 and references therein). The Pol portion of Gag-Pol contains the three enzymes that are essential for the replication of the virus: the protease (PR), which cleaves the Gag and Gag-Pol polyproteins, producing mature proteins, the reverse transcriptase (RT), which converts the viral RNA genome into a proviral DNA duplex, and the integrase (IN), which inserts the proviral DNA into the host chromosome (Frankel and Young, 1998; Freed and Martin, 2001; Tang et al., 1999). The Gag polyprotein directs assembly and release of new virions from infected cells (reviewed

in Freed, 1998). During or after the release, the viral protease cleaves the Gag and Gag-Pol polyproteins, producing mature, infectious viral particles.

Previous studies of the  $-1$  programmed frameshift events demonstrated the requirement of two mRNA *cis*-acting elements for efficient frameshifting (reviewed in Brierley and Pennell, 2001) (see Fig. 1A and B). In HIV-1, the first element is a heptamer sequence UUUUUUA, called the slippery sequence, where the  $-1$  frameshift occurs. The second element is a long helix structure located immediately downstream of the slippery sequence, which stimulates the frameshift. This structure, which contains two stems separated by a purine-rich bulge, was first characterized by probing with enzymatic attack and mutagenesis (Dulude et al., 2002) and recently confirmed by NMR (Gaudin et al., 2005; Staple and Butcher, 2005).

Although the importance of the frameshift is clearly established, several questions remain to be answered in order to exploit the frameshift event as a potential drug target. In particular, it remains to determine the extent to which variations in frameshifting efficiently translate into reduced viral replication and, therefore, into potential clinical benefit. Biswas et al. (2004) found that replacing the slippery sequence of HIV-1 with other slippery sequences severely decreases frameshifting and viral infectivity. However, the slippery sequence most likely does not form a defined structure, making specific interactions with candidate compounds difficult. In contrast, targeting of the frameshift stimulator structure by small molecules may be envisioned. A report by Telenti et al. (2002) analyzed natural and laboratory HIV-1 isolates with sequence variations in the upper stem of the stimulatory signal and found reduced viral replication only for two variants with more than 60% reduction in frameshift efficiency, whereas variants with a 16 to 42% reduction did not show any detectable replication deficit. The interpretation of these data suggested the existence of a threshold of tolerance for frameshift variations under which they would have no effect on viral replication. This in turn implied that attempts to modulate frameshift would be of no therapeutic value unless efficiently overcoming this threshold.

The present study was designed to elucidate the impact of mutations in the two *cis*-acting elements of the HIV-1 frameshift region on both frameshifting efficiency and viral

replication. We introduced mutations into the slippery site, the upper stem and the lower stem of the stimulatory signal structure. The effects of these mutants cover a wide range of frameshift efficiencies that were paralleled by a decreased infectivity of mutant virions and a reduced replication in long term cultures. Our results show that moderate reductions of frameshift efficiency are able to slow down viral replication, which suggests that compounds causing such reductions should provide clinical benefit.

## Results

**Description of HIV-1 mutants altered in the frameshift region.** In order to study the influence of changes in frameshift efficiency on HIV-1 replication, we investigated eight mutants containing mutations within the slippery sequence and/or the stimulatory frameshift signal. Fig. 1 shows the HIV-1 wild-type frameshift region and describes the mutants M1 to M8. Mutants M1 to M5 contain mutations that alter the slippery site and influence the frameshift efficiency by changing the dynamics of the codon-anticodon interaction at the A and/or P-site. Mutants M1 (UUUUUUU), M2 (UUUAAAA) and M3 (UUUUUUG) have a mutated A-site codon. M5 is mutated in both the A and P-site codons of the slippery site, the sequence that replaces the HIV-1 slippery sequence being AAAAAAC, which is used as a frameshift slippery site by several viruses (see below in the Discussion). Mutant M4, for which the UUUUUUA slippery sequence was substituted with CUUCCUC, served as a negative frameshift control since it does not allow slippage of the two tRNAs and re-pairing in the  $-1$  frame.

The other mutants are altered in the frameshift stimulatory signal. Mutant M6 is destabilized in the lower stem and M7 in the upper stem of the signal. M8 has mutations that increase the thermodynamic stability of this upper stem by replacing two U-A pairs with two C-G pairs, in addition to having a C-G instead of a U-G pair on top of this stem and having the M1 mutation. This mutant was initially designed to increase frameshifting.

**Frameshift efficiency of the HIV-1 mutants altered in the frameshift region.** The frameshift efficiency of wild-type and mutant constructs was measured with a dual-luciferase system (Fig. 2). The wild-type HIV-1 frameshift region caused a frameshift with an efficiency of  $9 \pm 1\%$  in this system, which is in agreement with previous studies that measured HIV-1 frameshifting with similar reporter systems (Biswas et al., 2004; Harger and Dinman, 2003). The frameshift efficiency of the negative control (M4) was reduced to a background level, as expected. For mutant M1, the frameshift efficiency was 48% of the wild-type whereas for M2 and M3, the frameshift was 8% and 60% of the wild-type, respectively. When the slippery site of HIV-1 was changed to the AAAAAAC sequence (M5) used by other viruses, the frameshift efficiency decreased to 35% of the wild-type construct. Destabilization of the lower stem (M6) led to a frameshift level of 70% and destabilization of the upper stem (M7) to 23%. Finally, with mutant M8, designed to stabilize the upper stem, the frameshift efficiency decreased to 42% of the wild-type value. Therefore, all the eight mutations decreased the frameshift efficiency to variable degrees.

**Influence of frameshift mutations on virus-like particles (VLPs).** We then investigated how changes in frameshift efficiency might affect the release of VLPs and the incorporation of Gag-Pol in these particles. As shown in Fig. 3, no mutation of the frameshift region altered the quantity of released p24. However, we observed changes in the amount of RT activity present in the VLPs, and, consequently, in the ratio of Gag-Pol to Gag incorporated. These variations correlated perfectly with the changes in frameshift efficiency. Those mutations that most dramatically decrease the frameshift efficiency cause the largest decrease in the amount of Gag-Pol incorporated (see the inset of Fig. 3).

**Influence of changes in frameshift efficiency on Gag processing.** The processing of the Gag polyprotein is mediated by the viral protease contained within Gag-Pol. We therefore investigated the effect of frameshift changes on the maturation of Gag within virions produced from cells transfected with the mutant derivatives of HIV-1<sub>LAI</sub> altered in the

frameshift region. The assay consisted in assessing the amount of the capsid protein, p24, produced upon Gag processing. As shown in Fig. 4, the Gag polyprotein was highly processed in the wild-type virions but there was no detectable Gag processing for mutants M2 and M4, which have a very low frameshift efficiency. The Gag processing was decreased about two-fold compared to the wild-type virions for mutants M1, M3, M6, M7 and M8, which have frameshift efficiencies ranging between about 23 and 70% of the wild-type value, and mutant M5, with a frameshift efficiency of 35%, had a decrease of about five-fold in Gag processing. Similar profiles were obtained when analyzing Gag processing within VLPs produced from COS7 cells transfected with pGAGPOL-xm-wt and its mutant derivatives (data not shown). These Gag processing profiles suggest a lack of proportionality between the decrease in frameshift efficiency and processing. Indeed, mutants M6 and M7, with, respectively, a frameshift efficiency of 70 and 23% of the wild-type value, had a reduction of 50% in the production of the p24 breakdown product.

**Influence of changes in frameshift efficiency on viral infectivity.** We also investigated the effect of changes in frameshift efficiency on the infectivity of HIV-1 by performing single-round infectivity assays. Infectivity is the relative number of infectious units per standardized p24 content in the inoculum. As shown in Fig. 5, the infectivity of virions produced from cells transfected with the mutant derivatives of HIV-1<sub>LAI</sub> with an altered frameshift region was inferior to that of the wild-type construct. The decrease in infectivity was proportional to the decrease observed in frameshifting, the effect on viral infectivity being slightly amplified. One apparent exception to this proportionality was mutant M5, which appeared almost non-infectious, despite an appreciable frameshift efficiency and RT incorporation in VLPs.

**Influence of changes in frameshift efficiency on viral replication.** In order to investigate whether the observed decrease in frameshift efficiency and the reduction in infectivity would cause cumulative effects in long term cultures, we analyzed the replication kinetics of mutants M1 to M8 in CEM T-cells (see Fig. 6). With mutants M2, M4, and M5, there

was no detectable virus production during the period of observation and the replication of mutants M1, M7 and M8 was severely attenuated as compared to the wild-type control HIV-1<sub>LAI</sub>. Interestingly, the p24 production of mutant M6 was less than 50% of that of the wild-type virus at all time points, indicating reduced viral replication in spite of a moderate effect on frameshift efficiency. Finally, mutant M3 also produced less p24, and the peak of production was delayed by more than 3 days as compared to the control. These results demonstrate that all mutants with altered frameshift efficiency were appreciably attenuated in long term virus replication.

## Discussion

In the present work, we introduced mutations in the frameshift slippery site and in the downstream frameshift stimulatory signal of HIV-1. Our study shows that even relatively mild effects on frameshift efficiency, such as those resulting from mutations in the frameshift stimulatory signal, lead to a reduced replication of the mutant virions. The frameshift stimulatory signal has a defined structure with which antiviral compounds could interact, unlike the slippery site, which may be a more difficult target. Indeed, even the replication of mutants for which no dramatic changes in Gag processing were detected by Western blot analysis, were affected in comparison to the wild-type virus. Therefore, and in contrast to the view expressed by Telenti et al. (2002), our results suggest that any change in frameshift efficiency results in reduced viral replication and that no threshold of frameshift alteration must be reached to become phenotypically manifest.

In addition to showing a strong correlation between frameshift and virus replication, our results yield new information on the different elements of the frameshift region. Destabilization of the lower stem of the HIV-1 frameshift stimulatory signal (M6) results in a mild but clearly detectable replication phenotype. Destabilization of the upper stem (M7) drastically reduces the replicative capacity of the mutant virus. In contrast to mutants M6 and M7, mutant M8 stabilizes the upper stem but it nevertheless results in a decreased

frameshift efficiency and in a dramatic decrease in the replicative capacity of the corresponding mutant virus. Interestingly, Bidou et al. (1997) also reported that increasing the stability of the upper stem decreases the frameshift efficiency of HIV-1 and concluded that increasing the pause of the ribosome on the slippery site disfavors the frameshift. Our results illustrate the importance of maintaining the structural integrity and stability of the frameshift stimulatory signal for the infectivity of the virus but they contrast with the results of Telenti et al. (2002). Indeed, these researchers failed to detect reduced infectivity or replication for variants containing substitutions that destabilize the upper stem of the frameshift stimulatory signal and reduce frameshift efficiency to a value of about 60% compared to wild-type. However, they observed a reduced infectivity for two mutants with a frameshift efficiency decreased at and below 40% of the wild-type. Telenti et al. concluded that this 40% value constitutes a threshold efficacy, above which variations in frameshifting have no effect on viral replication. The reason for this discrepancy between our results and those of Telenti et al. is unclear and may lie in differences between the respective experimental systems.

We introduced five mutations that alter the slippery sequence, M1 to M5. For mutant M1 (with a stretch of 7U), we detect a 50% reduced frameshift efficiency. With the 7U sequence (M1 and M8), frameshifting occurs with a tRNA<sup>Phe</sup> in the A site rather than a tRNA<sup>Leu</sup> as it is the case for the wild-type. It is possible that tRNA<sup>Phe</sup> is less prone to slip than tRNA<sup>Leu</sup>. Similarly, for UUUAAAA (M2), it is also possible that tRNA<sup>Lys</sup> in the A site is less prone to slip than tRNA<sup>Leu</sup>. The importance of the identity of the tRNA for the frameshift efficiency is also supported by previous observations from Chamorro et al. (1992). A change in the identity of the A-site tRNA<sup>Leu</sup> (from tRNA<sup>Leu</sup> with an IAG anticodon to a tRNA<sup>Leu</sup> with a CAG anticodon) also likely accounts for the decreased frameshift efficiency with the UUUUUUG sequence (M3) (see Brierley et al., 1992). The AAAAAAC sequence (M5) contains a slippery sequence found in astroviruses, in human T-cell leukemia virus type 2, in mouse mammary tumor virus (MMTV) and in equine anemia infectious virus (EIAV) (reviewed in Brierley, 1995). The frameshift efficiency of M5 is 35% of the wild-type, in agreement with Biswas et al. (2004), who also replaced the

slippery sequence of HIV-1 with AAAAAAC. Interestingly, it has been reported that replacing the AAAAAAC sequence of EIAV with the slippery sequence of HIV-1 decreased the frameshift efficiency to 50% in the presence of the pseudoknot that acts as a frameshift stimulator in EIAV (Chen and Montelaro, 2003). Therefore, some degree of functional interdependence seems to exist between the slippery sites and a matching stimulatory signal since both elements cannot be arbitrarily interchanged between viruses. All the mutations that were introduced in the slippery site severely decreased the capacity of the virus to replicate, the severity of this effect correlating with the decrease in the frameshift efficiency, with the exception of M5 (see below).

The mechanisms accounting for the observed effects of the frameshift mutations on HIV-1 replication may be multiple. The frameshift changes result in reduced Gag-Pol production and, consequently, reduced incorporation of the viral enzymes into the virions. The amount of RT activity within VLPs directly mirrors reduced frameshift efficiency, and it is likely that the decreased RT amount largely contributes to the decreased viral infectivity, knowing that viral infectivity is very sensitive to changes in RT activity (see Ambrose et al., 2004; Julias et al., 2001). The reduction in Gag-Pol incorporation can result in altered protease activity leading to altered Gag processing, but it appears that protease activity must be reduced by more than four-fold to significantly impair viral infectivity (Rose et al., 1995). We therefore do not believe that maturation defects play a major part in the reduced infectivity observed for the mutant viruses, except when the amount of protease is dramatically decreased (mutants M2 and M4) and in the case of mutant M5 (see below). It is also possible that the decreased integrase incorporation contributes to reduce the infectivity of mutant viruses. In addition, mechanisms other than just the reduction of incorporated viral enzymes may be at work. It has been shown that over- or underexpressing Gag-Pol impairs the maturation of the virion genomic RNA dimers (Shehu-Xhilaga et al., 2001; 2002), and it can be hypothesized that such an impairment could contribute to the decreased infectivity of mutant viruses. Also, the incorporation of tRNA<sup>Lys</sup> into viruses is a crucial event in the viral replication cycle, which is driven by Gag-Pol (Mak et al., 1994) and could be affected by a decrease in Gag-Pol.

The strong correlation that we observe between the decrease in frameshift efficiency and the decrease in viral infectivity supports our interpretation that the changes in the amount of Gag-Pol produced in the mutants investigated account for the decrease in viral infectivity. However, other changes besides the decrease in frameshift efficiency and the resulting reduction in Gag-Pol production could also contribute to lower the virus infectivity. For example, it could be hypothesized that the mutations that we introduced in the frameshift region influence the secondary and tertiary structure of the viral RNA. These structural changes could perturb long-range interactions such as those involved in the stabilization of the association between the two copies of genomic viral RNA forming a dimer (Hibbert and Rein, 2005; Paillart et al., 2002; 2004). Mutants M1 and M3 have a comparable frameshift efficiency, a comparable level of RT activity associated to the VLPs and do not differ in Gag processing. However, M1 appears to be less infectious than M3 by about two-fold. Although purely speculative, an effect of the mutation in M1 on the viral RNA secondary and tertiary structure could account for its lower infectivity compared to M3. Also, it is difficult to mutate the frameshift signal of HIV-1 without simultaneously changing the amino acid sequence of the corresponding Gag p1 spacer peptide and that of the p7/p1 and p1/p6 cleavage sites, which might also affect viral replication. For example, two proline residues (position 7 and 13) of the 16-amino acid p1 have been shown to be required for efficient HIV-1 infectivity (Hill et al., 2002). Interestingly, the only frameshift mutants found to have reduced replication by Telenti et al. (2002) contain a proline to lysine substitution at position 7. It could therefore be envisioned that the experimental conditions used by Telenti et al. were able to detect defects due to p1 mutations, but were not sensitive enough to identify more subtle effects due to frameshift variations. The impact of some p1 mutations on the viral replication is also illustrated by the drastic phenotype of our mutant M5. This mutant contains three amino acid substitutions at the p7/p1 protease cleavage site, among which the introduction of a lysine residue at the P1 position of this cleavage site, a substitution that was previously shown to abolish the cleavage reaction (Pettit et al., 2002). It could be suggested that these substitutions at the p7/p1 cleavage site affect the overall protease activity. The observed decrease of protease activity of mutant

M5 could indeed contribute to lower its infectivity, which is also disproportional to the effect of the mutation on frameshifting. We should, however, point out that this observation applies only to mutant M5, and it is unlikely that p1 mutations have a major influence on the phenotypes of all our mutants. For example, the mutation of M3 does not affect the amino acid sequence of p1, yet affects viral replication.

Altogether, our study provides evidence that variations in the frameshift stimulatory signal lead to moderate reductions of frameshift efficiency but nevertheless reduce the replicative capacity of the virus. Even a partial reduction of virus replication can translate into clinical benefit. This seems indeed to be the case for many drug-resistant HIV-1 strains that bear mutations reducing their replicative fitness (Berkhout, 1999). Reduction of viral replication by compounds targeting the stimulatory signal may thus be of significant therapeutic interest.

## Materials and methods

**Construction of plasmids. (i) pGAGPOL vectors.** The HIV-1 *gag-pol* expression vector used in this study for the synthesis of virus-like particles (VLPs) was derived from pGAGPOL-RRE-r. This vector contains, under the control of the SV40 promoter, the *gag-pol* gene and the *rev*-responsive element (RRE) from the HIV-1 proviral clone BH10-HXB2 (Smith et al., 1990). Cotransfection in COS-7 cells (see details below) of this plasmid together with pRev1 (Smith et al., 1990), which codes for HIV-1 Rev, results in the efficient release of VLPs in the culture supernatant of the transfected cells. To facilitate the cloning of mutants of pGAGPOL-RRE-r altered in the frameshift region, we created a cloning derivative of this plasmid, named pGAGPOL-xm. In this plasmid, the frameshift region was deleted and replaced with a short cassette containing a *Xho*I restriction site at one end and a *Mlu*I restriction site at the other end. This was made by PCR amplification

with a standard overlap extension procedure (Ho et al., 1989). The wild-type construct and the frameshift mutants were made by inserting, between the digested *Xho*I and *Mlu*I sites of pGAGPOL-xm, the appropriate annealed and phosphorylated 74 bp oligonucleotide cassette, purchased from Sigma Genosys. This generated pGAGPOL-xm-wt and pGAGPOL-M1 to pGAGPOL-M8, respectively (Fig. 1C). The insertion of a *Xho*I and a *Mlu*I restriction site in pGAGPOL-xm-wt and its derivatives substituted a tyrosine with a lysine residue and a proline with a valine residue at the amino acid position 427 and 453 of Gag, respectively. These modifications did not change the reverse transcriptase (RT)-associated content of the released VLPs compared to the parental construct (data not shown).

**(ii) Dual-luciferase reporter plasmids.** A dual-luciferase reporter plasmid was created, based on previously published systems (Grentzmann et al., 1998; Harger and Dinman, 2003). Briefly, the HIV-1 frameshift region was inserted between the coding sequence of the *Renilla* luciferase (*Rluc*) and the firefly luciferase (*Fluc*) gene in an appropriate vector. In this construct, only the ribosomes that initiated translation at the initiator codon for the *Renilla* luciferase but changed the reading frame by one base in the 5' direction in the frameshift region synthesize the firefly luciferase, which is thus fused to *Renilla* luciferase (see Fig. 2A). To create this vector, the *Renilla* luciferase coding sequence and part of its 5' untranslated region were amplified by PCR from pRluc-C1, a *Renilla* luciferase fusion protein expression vector (a generous gift from Dr M. Aubry, Université de Montréal). The forward and reverse mutagenic primers: 5'GCTGGTTTAGTGAAGCTTCAGATCCGCTAGAGCCACC-3' and 3'AACGAGCAGGGGGTACCCGGCGAGCTCTC'5 used for this amplification introduced a *Hind*III and a *Kpn*I restriction site at the 5' and 3' end of the gene, respectively. The amplified *Rluc* gene was then inserted upstream of the *Fluc* gene of *Hind*III-*Kpn*I-digested pcDNA3.1-luc (Dulude et al., 2002), generating the parental pDual-luc construct. The wild-type and the mutated frameshift regions (nucleotides 2037 to 2139 according to HXB2 genome nomenclature) were amplified by PCR from pGAGPOL-xm-wt or pGAGPOL-M1 to pGAGPOL-M8 and inserted between the *Rluc* and *Fluc* genes of

pDual-luc, generating pDual-HIV(-1) and the pDual-HIV-M1 to pDual-HIV-M8 constructs, respectively. The amplification used the following primers: Fwd-*KpnI* (5'-GGGCTGTTGGTACCTGGAAAGGAAGG-3') and Rev-*BamHI* (5'-GGGCTGTTGGATCCCCACGCGTGAAGT-3'), and the digested PCR products were then inserted between the *KpnI* and *BamHI* restriction sites of pDual-luc. All these constructs were used to measure the level of firefly luciferase expressed relative to the *Renilla* luciferase. An in-frame plasmid control, pDual-HIV(0), in which the *Rluc* gene is in the same frame as the *Fluc* gene, was created by PCR amplification. This was made by inserting an adenine immediately after the slippery site, which was mutated from UUUUUUA to CUUCCUC to prevent -1 frameshifting, as described by Grentzmann et al. (1998). This plasmid was used to measure the translation level of the in-frame *Fluc* coding sequence relative to that of the *Rluc* gene (see Fig. 2A).

**(iii) Proviral clones.** For infectivity assays, we used the LAI proviral clone, which contains the complete genome of HIV-1<sub>LAI</sub> (Peden et al., 1991). The wild-type construct, pLAI-wt, and the frameshift mutants, pLAI-M1 to pLAI-M8, were LAI-BH10 chimeras made by substituting a 1.2-Kb *AgeI-Bst1107I* fragment from the pGAGPOL-xm-wt and pGAGPOL-M1 to pGAGPOL-M8 plasmids, respectively, into the *AgeI* and *Bst1107I*-digested LAI proviral clone. With the inserted *AgeI-Bst1107I* BH10 fragment, amino acid residues 389, 401, 427, 453 and 473 of Gag and residues 19 to 22 and 53 of Pol, were changed, compared to wild-type HIV-1<sub>LAI</sub>.

**Frameshift assays.** The effect of the mutations on the -1 frameshift efficiency was measured using the dual-luciferase reporter plasmids described above. Frameshifting was monitored by transient transfection of the dual-luciferase vectors into 293FT cells (Invitrogen). The day before transfection,  $2 \times 10^4$  cells/well were seeded in 24-well plates and maintained in Dulbecco's modified Eagle medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Wisent). Briefly, 0.5  $\mu$ g of each dual-luciferase reporter construct was transfected into cells, using a standard calcium phosphate precipitation method (Jordan et al., 1996), and cells were grown for 48 h before being harvested. Cells were washed twice with PBS and lysed with 100  $\mu$ l of the Cell Passive Lysis Buffer

(Promega). The firefly versus the *Renilla* luciferase activities of each construct were measured as relative light units with a Berthold Lumat LB 9507 luminometer, using a Dual-Luciferase<sup>®</sup> Reporter Assay System kit from Promega.

**Production of VLPs.** COS-7 cells ( $3 \times 10^5$  per plate) were plated in 60-mm dishes and cultured in 4 ml of DMEM supplemented with 10% FBS for 24 h prior to transfection. Four  $\mu\text{g}$  of pGAGPOL-xm-wt or its mutant derivatives and 2.5  $\mu\text{g}$  of pRev1 were introduced in cells by transfection, using a standard calcium phosphate precipitation method. The culture medium was removed and replaced with fresh medium 16 h posttransfection. Cells and supernatants were harvested 64 h posttransfection. The culture medium was centrifuged at 3,000 rpm for 20 min at 4°C to remove cellular debris. The clarified medium was centrifuged again for 1 h at 40,000 rpm (Beckman 50Ti rotor) to pellet the VLPs. Pelleted VLPs were resuspended in 200  $\mu\text{l}$  of 10 mM Tris-HCl (pH 7.5) and frozen at -80°C or used immediately. In parallel, cells were washed twice with 2 ml of PBS and lysed with 1 ml of 1% Triton X-100 / PBS. Cell lysates were collected and stored at -80°C or used immediately.

**RT assays.** VLP-associated RT levels were assayed by a standard procedure (Smith et al., 1990). In a 100  $\mu\text{l}$  reaction volume, 65  $\mu\text{l}$  of VLPs resuspended in the Tris buffer were incubated for 90 min at 37°C with 4  $\mu\text{Ci}$  of  $\text{H}^3$ -TTP (Perkin-Elmer) and 2.5  $\mu\text{g}$  of polyA<sub>dT(12-18)</sub> (Pharmacia) in a RT buffer at a final concentration of 50 mM Tris-HCl (pH 7.4), 60 mM KCl, 2 mM DTT, 5 mM MgCl<sub>2</sub>, 0.1% Triton X-100 and 1 mM EDTA. The reaction products were precipitated for 1 h on ice with 1 ml of cold 10% (w/v) trichloroacetic acid containing 10 mM sodium pyrophosphate, collected by filtration on glass-fiber filters and counted for radioactivity.

**p24 ELISA assays.** p24 ELISA assays were performed to measure the amount of Gag inside the cells and in the VLPs released in the supernatant following the transfection of COS-7 cells with the pGAGPOL vectors. Quantification of p24 was also used to normalize the input of viruses in the infectivity assays and to determine the amount of viruses produced in the replication kinetics studies. These assays were performed using a commercially available p24 antigen detection kit (Beckman-Coulter), according to the

instructions provided by the manufacturer. In each experiment, a standard curve was generated, using serial dilutions of a p24 standard. Samples were diluted so that the experimental values were within the linear range of the assay.

**Virus production.** HeLa-P4(CD4<sup>+</sup>/LTR-*lacZ*<sup>+</sup>) cells (Clavel and Charneau, 1994), which contain the *lacZ* gene under the control of the HIV-1 LTR promoter, were used for virus production and single-round infectivity experiments (see below). HeLa-P4 cells ( $8 \times 10^5$  per well) were seeded in 6-well plates and cultured in DMEM supplemented with 10% FBS for 24 h before transfection with 5  $\mu$ g of proviral DNA, using the Polyfect Transfection Reagent (Qiagen). The medium was changed 24 h after transfection. Virus-containing supernatants were collected 48 h posttransfection, clarified of cells debris by a 5 min-centrifugation (13,000 rpm) and stored at  $-80^{\circ}\text{C}$ . The p24 content in the supernatants was determined by a p24 ELISA assay (see above) prior to the single-round infectivity assays and to the long term replication assays.

**Gag processing in virions.** To examine the level of processing of Gag incorporated into virions, 6 ml of supernatant from HeLa.P4 cells transfected with proviral DNA were centrifuged at 25,000 rpm for 3 h at  $4^{\circ}\text{C}$  through a 20% sucrose cushion with a Beckman Coulter Avanti J25 centrifuge, using a JA.25.50 rotor. The viral pellets were resuspended in 200  $\mu$ l of PBS and 15  $\mu$ l were resolved by electrophoresis on a 10% sodium dodecyl sulfate-polyacrylamide gel. Samples were immunoblotted, using a mouse anti-HIV-1 p24 monoclonal antibody (no 24-2), diluted 1/2000 (reagent obtained from Dr. Michael Malim of the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH). Horseradish peroxidase-conjugated sheep anti-mouse was used as the secondary antibody. Antigen-antibody complexes were detected by enhanced chemiluminescence and bands were obtained with films (Biomax, XAR) exposed for a short period. Band were scanned and quantified with Quantity one (Bio-Rad).

Single-round infectivity assays. HeLa-P4 cells were plated in 96-well plates ( $10^4$  cells/well) in DMEM supplemented with 10% FBS and infected in triplicate with a normalized amount of HIV-1 or mutant viruses (180 ng/ml of p24 of virus per well) for 24 h. The supernatant was then removed, cells were lysed with lysis buffer (0.2 % Triton X-

100 / PBS) and  $\beta$ -galactosidase activity was determined, using the chromogenic substrate chlorophenolred- $\beta$ -D-galactopyranoside, as described (Brelot et al., 2000).

**Replication kinetics of mutant virus.** CEM T-cells were used for long term replication experiments. CEM T-cells ( $2 \times 10^6$ ) in 1 ml of RPMI 1640 medium were infected with 360 ng/ml of p24 of wild-type or mutant virus for 2 h, then washed and cultured in T-25 flasks with 5 ml of fresh medium. The cultures were split every second day by replacing 50% of the culture with the same volume of fresh medium. Samples of culture supernatant were withdrawn from the medium at regular intervals and p24 quantified as a measure of ongoing virus replication.

### Acknowledgements

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.B-G. and N.H. Y.B. holds a studentship from the Fondation de l'Hôpital Sainte-Justine. N.H. holds a Chercheur-Boursier award from the Fonds de la Recherche en Santé du Québec (FRSQ). We gratefully acknowledge Gerardo Ferbeyre and Jim Omichinski for helpful discussions and critical reading of this manuscript. We thank Peter Kreuzaler for his participation in this study as an exchange student from the Freie Universität of Berlin, Germany.

### References

- Ambrose, Z., Boltz, V., Palmer, S., Coffin, J. M., Hughes, S. H., and Kewalramani, V. N. (2004). In vitro characterization of a simian immunodeficiency virus-human immunodeficiency virus (HIV) chimera expressing HIV type 1 reverse transcriptase to study antiviral resistance in pigtail macaques. *J Virol* **78**(24), 13553-61.
- Berkhout, B. (1999). HIV-1 evolution under pressure of protease inhibitors: climbing the stairs of viral fitness. *J Biomed Sci* **6**(5), 298-305.

- Bidou, L., Stahl, G., Grima, B., Liu, H., Cassan, M., and Rousset, J. P. (1997). In vivo HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. *RNA* **3**(10), 1153-8.
- Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P., and Peltz, S. W. (2004). The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* **78**(4), 2082-7.
- Brelot, A., Heveker, N., Montes, M., and Alizon, M. (2000). Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* **275**(31), 23736-44.
- Brierley, I. (1995). Ribosomal frameshifting viral RNAs. *J Gen Virol* **76** ( Pt 8), 1885-92.
- Brierley, I., Jenner, A. J., and Inglis, S. C. (1992). Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **227**(2), 463-79.
- Brierley, I., and Pennell, S. (2001). Structure and function of the stimulatory RNAs involved in programmed eukaryotic-1 ribosomal frameshifting. *Cold Spring Harb Symp Quant Biol* **66**, 233-48.
- Cassan, M., Delaunay, N., Vaquero, C., and Rousset, J. P. (1994). Translational frameshifting at the gag-pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J Virol* **68**(3), 1501-8.
- Chamorro, M., Parkin, N., and Varmus, H. E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci U S A* **89**(2), 713-7.
- Chen, C., and Montelaro, R. C. (2003). Characterization of RNA elements that regulate gag-pol ribosomal frameshifting in equine infectious anemia virus. *J Virol* **77**(19), 10280-7.
- Clavel, F., and Charneau, P. (1994). Fusion from without directed by human immunodeficiency virus particles. *J Virol* **68**(2), 1179-85.

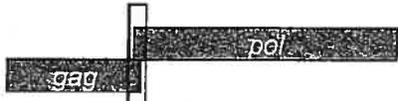
- Dulude, D., Baril, M., and Brakier-Gingras, L. (2002). Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res* **30**(23), 5094-102.
- Frankel, A. D., and Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**, 1-25.
- Freed, E. O. (1998). HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **251**(1), 1-15.
- Freed, E. O., and Martin, M. A. (2001). HIVs and their replication. In "Fields virology, 4th ed." (P. M. H. D. M. Knipe, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus, Ed.), pp. 1971-2041. Lippincott Williams & Wilkins, Philadelphia, Pa.
- Gaudin, C., Mazauric, M. H., Traikia, M., Guittet, E., Yoshizawa, S., and Fourmy, D. (2005). Structure of the RNA signal essential for translational frameshifting in HIV-1. *J Mol Biol* **349**(5), 1024-35.
- Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F., and Atkins, J. F. (1998). A dual-luciferase reporter system for studying recoding signals. *Rna* **4**(4), 479-86.
- Harger, J. W., and Dinman, J. D. (2003). An in vivo dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *RNA* **9**(8), 1019-24.
- Hibbert, C. S., and Rein, A. (2005). Preliminary physical mapping of RNA-RNA linkages in the genomic RNA of Moloney murine leukemia virus. *J Virol* **79**(13), 8142-8.
- Hill, M. K., Shehu-Xhilaga, M., Crowe, S. M., and Mak, J. (2002). Proline residues within spacer peptide p1 are important for human immunodeficiency virus type 1 infectivity, protein processing, and genomic RNA dimer stability. *J Virol* **76**(22), 11245-53.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1), 51-9.

- Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**(6153), 280-3.
- Jordan, M., Schallhorn, A., and Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res* **24**(4), 596-601.
- Julias, J. G., Ferris, A. L., Boyer, P. L., and Hughes, S. H. (2001). Replication of phenotypically mixed human immunodeficiency virus type 1 virions containing catalytically active and catalytically inactive reverse transcriptase. *J Virol* **75**(14), 6537-46.
- Mak, J., Jiang, M., Wainberg, M. A., Hammarskjold, M. L., Rekosh, D., and Kleiman, L. (1994). Role of Pr160gag-pol in mediating the selective incorporation of tRNA(Lys) into human immunodeficiency virus type 1 particles. *J Virol* **68**(4), 2065-72.
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004). Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2**(6), 461-72.
- Paillart, J. C., Skripkin, E., Ehresmann, B., Ehresmann, C., and Marquet, R. (2002). In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA. *J Biol Chem* **277**(8), 5995-6004.
- Paulus, C., Ludwig, C., and Wagner, R. (2004). Contribution of the Gag-Pol transframe domain p6\* and its coding sequence to morphogenesis and replication of human immunodeficiency virus type 1. *Virology* **330**(1), 271-83.
- Peden, K., Emerman, M., and Montagnier, L. (1991). Changes in growth properties on passage in tissue culture of viruses derived from infectious molecular clones of HIV-1LAI, HIV-1MAL, and HIV-1ELI. *Virology* **185**(2), 661-72.
- Pettit, S. C., Henderson, G. J., Schiffer, C. A., and Swanstrom, R. (2002). Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol* **76**(20), 10226-33.

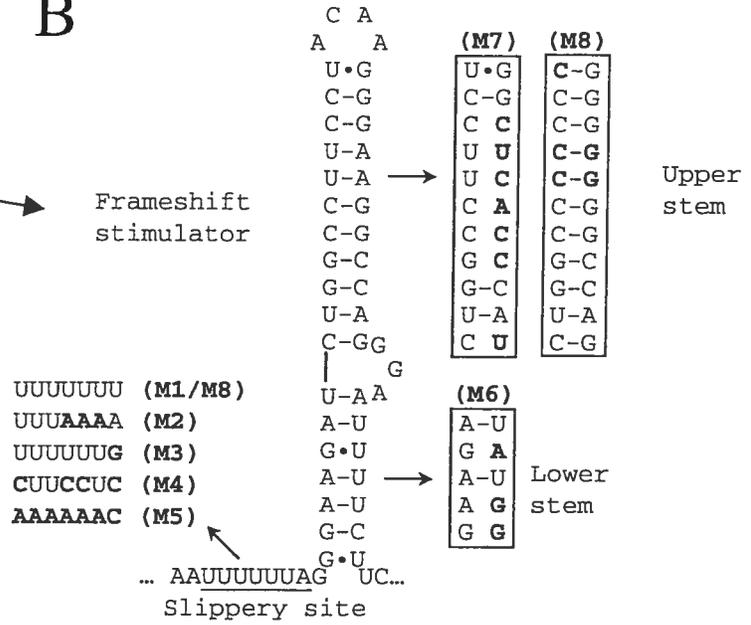
- Reil, H., Kollmus, H., Weidle, U. H., and Hauser, H. (1993). A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J Virol* **67**(9), 5579-84.
- Rose, J. R., Babe, L. M., and Craik, C. S. (1995). Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *J Virol* **69**(5), 2751-8.
- Shehu-Xhilaga, M., Crowe, S. M., and Mak, J. (2001). Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol* **75**(4), 1834-41.
- Shehu-Xhilaga, M., Hill, M., Marshall, J. A., Kappes, J., Crowe, S. M., and Mak, J. (2002). The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol* **76**(9), 4331-40.
- Smith, A. J., Cho, M. I., Hammarskjold, M. L., and Rekosh, D. (1990). Human immunodeficiency virus type 1 Pr55gag and Pr160gag-pol expressed from a simian virus 40 late replacement vector are efficiently processed and assembled into viruslike particles. *J Virol* **64**(6), 2743-50.
- Staple, D. W., and Butcher, S. E. (2005). Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J Mol Biol* **349**(5), 1011-23.
- Tang, H., Kuhen, K. L., and Wong-Staal, F. (1999). Lentivirus replication and regulation. *Annu Rev Genet* **33**, 133-70.
- Telenti, A., Martinez, R., Munoz, M., Bleiber, G., Greub, G., Sanglard, D., and Peters, S. (2002). Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem-loop structure. *J Virol* **76**(15), 7868-73.

Figure 3-1: Frameshift region of wild-type HIV-1 and its mutants. A) HIV-1 *gag* and *pol* open reading frames. The overlapping segment contains the programmed  $-1$  ribosomal frameshift region. The *pol* gene is in the  $-1$  frame relative to *gag* and it is expressed as a Gag-Pol polyprotein via a  $-1$  frameshift. B) Structure of the frameshift stimulatory signal of HIV-1 and summary of the mutations introduced in the frameshift region. The underlined slippery site is followed by a frameshift stimulator consisting of a long irregular stem-loop, where the upper and the lower stems are separated by a three-purine bulge. C) Amino acid and nucleotide sequences of wild-type HIV-1 and its derivatives mutated in the frameshift region. The elements required for the frameshift are located in the spacer p1 protein between the end of the p7 (nucleocapsid) and the beginning of the p6 protein. The slippery sequence is underlined in the wild-type nucleotide sequence (wt). Amino acid and nucleotide residues that are not altered in the Gag mutants (M1 to M8) are represented by dashed lines.

A



B



C

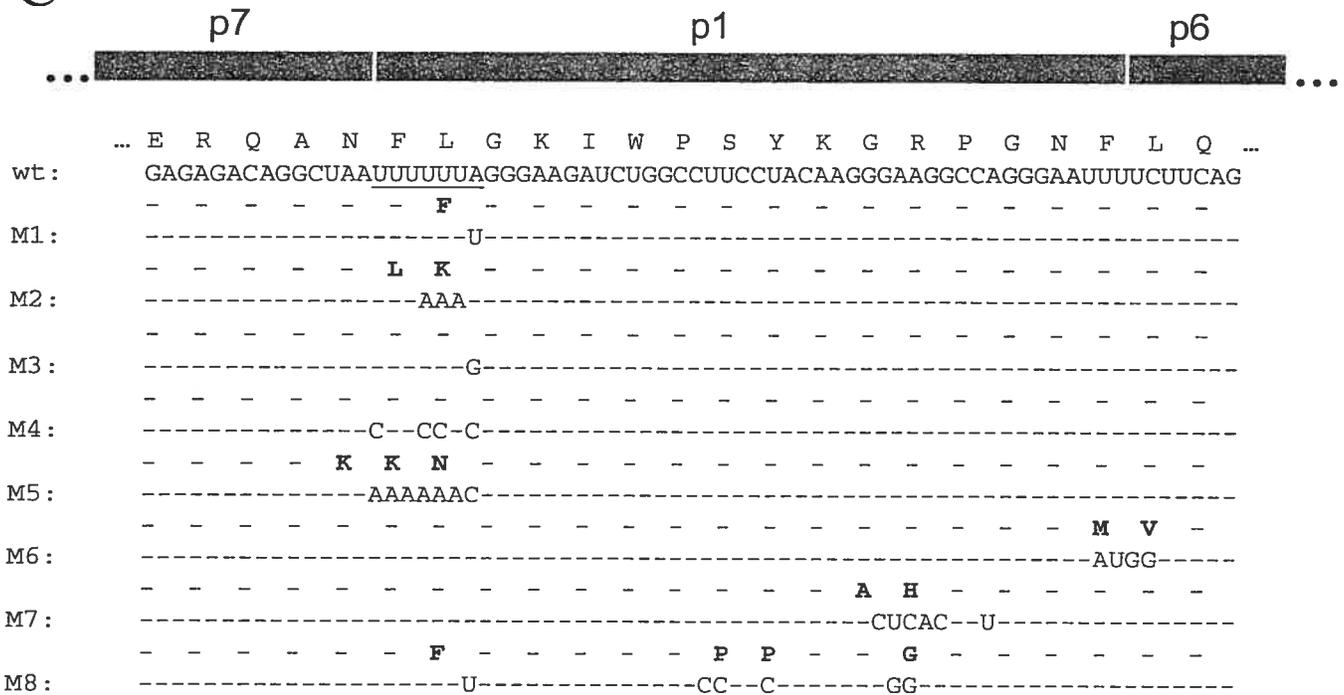
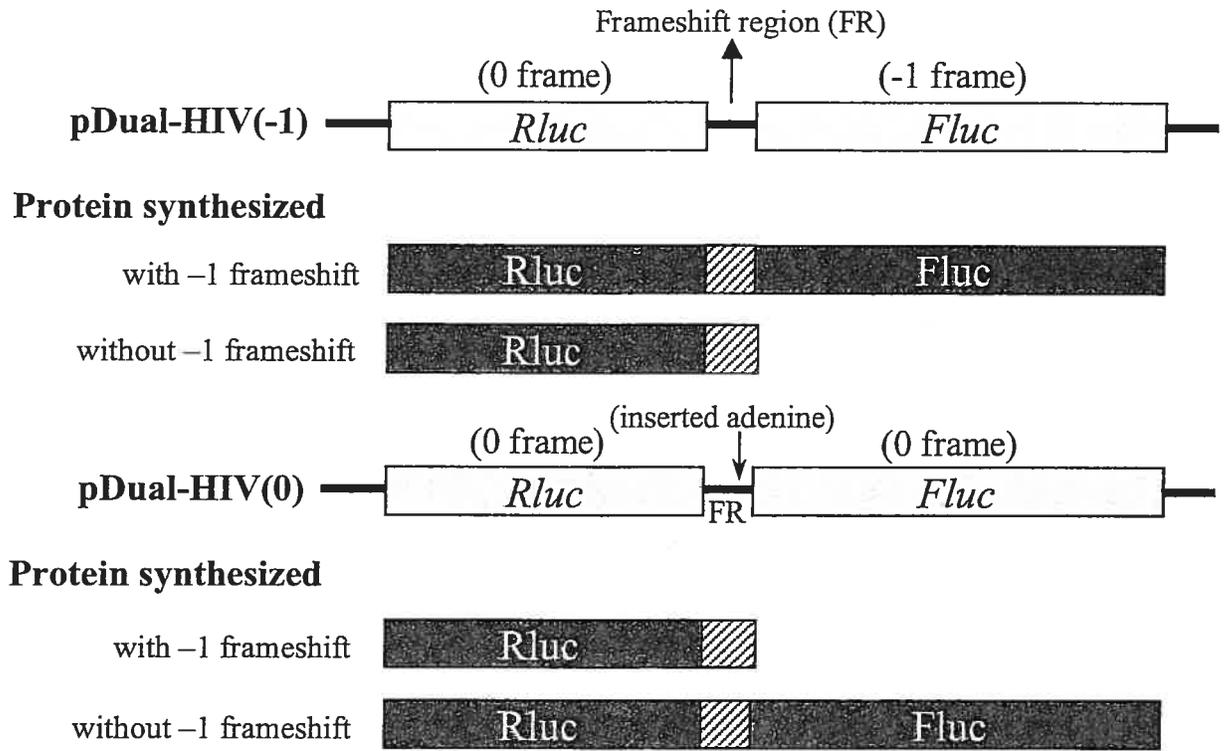


Figure 3-2: Measurement of the frameshift efficiency of HIV-1 derivatives with a dual-luciferase system. A) Scheme of the vectors used to measure the frameshift efficiency. In the pDual-HIV construct, the HIV-1 wild-type or mutant frameshift region (FR) is inserted between the *Renilla* luciferase (*Rluc*) gene and the firefly luciferase (*Fluc*) gene. With the pDual-HIV(-1) construct, the firefly luciferase (*Fluc*) is synthesized as a fusion to the *Renilla* luciferase (*Rluc*) only by ribosomes that make a -1 frameshift in the inserted frameshift region. Plasmid pDual-HIV(0) is an in-frame control, in which an additional adenine is inserted immediately after the slippery site. With this construct, *Fluc* is synthesized as a fusion to *Rluc* only by ribosomes that translate the messenger according to conventional rules. Frameshift efficiencies were determined by dividing the ratio of the firefly to *Renilla* luciferase activities from the wild-type or mutant constructs to the firefly to *Renilla* luciferase ratio from the in-frame pDual-HIV(0) control construct. B) Frameshift efficiency of the frameshift mutants. 293FT cells were transfected with the different dual-luciferase constructs and the activities of *Renilla* and firefly luciferase were measured 48 h later. The frameshift efficiencies of mutants M1 to M8 are indicated relative to the wild-type construct (wt), which is arbitrarily set at 100% (the absolute value is  $9 \pm 1\%$ ). Each value represents the mean  $\pm$  standard deviation of at least four independent experiments.

A



B

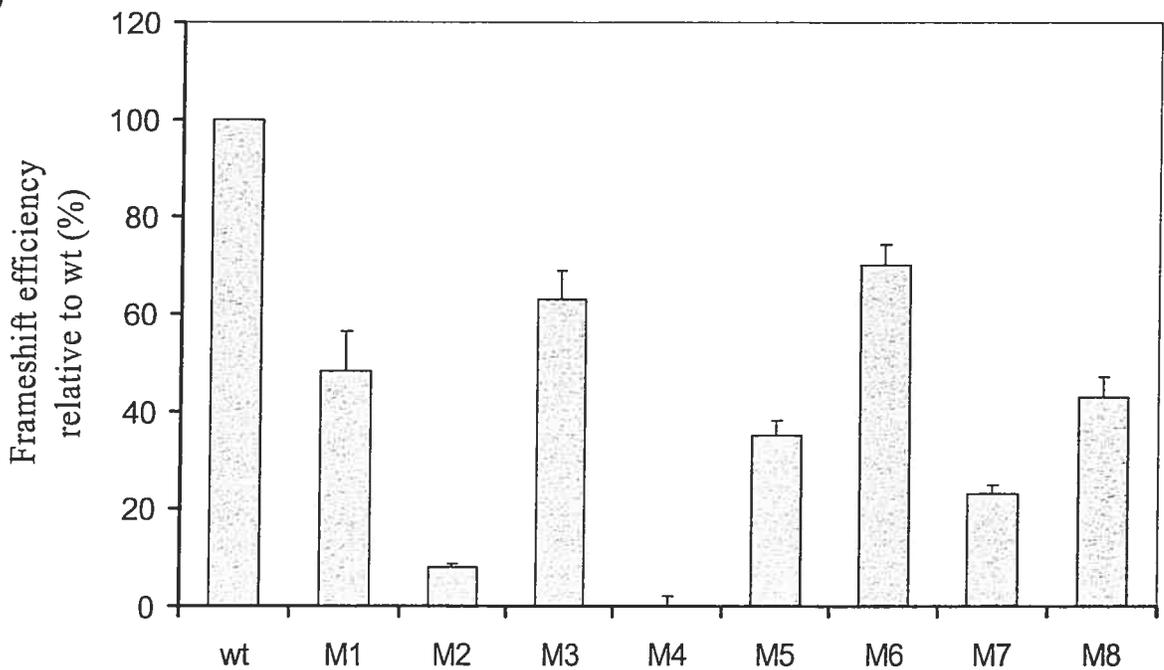


Figure 3-3: Effect of changes in frameshift efficiency on the release of VLPs, on the RT content of the released VLPs, and on the ratio of Gag-Pol to Gag in these VLPs. Four  $\mu\text{g}$  of pGAGPOL-xm-wt or pGAGPOL-M1 to -M8 and 2.5  $\mu\text{g}$  of pRev1 were introduced into COS-7 cells by transfection. The p24 and the RT content of the VLPs in the culture supernatant were measured 64 h posttransfection (see details in Materials and methods). White and grey bars represent, respectively, the p24 and the RT level of the VLPs released in the supernatant. Black bars represent the ratio of RT to p24 for the released VLPs. Changes in the amount of released p24, RT and in the RT to p24 ratio for the mutants are indicated relative to the wild-type construct, which is arbitrarily set at 100%. The inset presents the correlation ( $R^2=0.923$ ) between the RT to p24 ratio and the frameshift efficiency (Fs) of the wild-type construct and frameshift mutants. Each value represents the mean  $\pm$  standard deviation of three to four experiments.

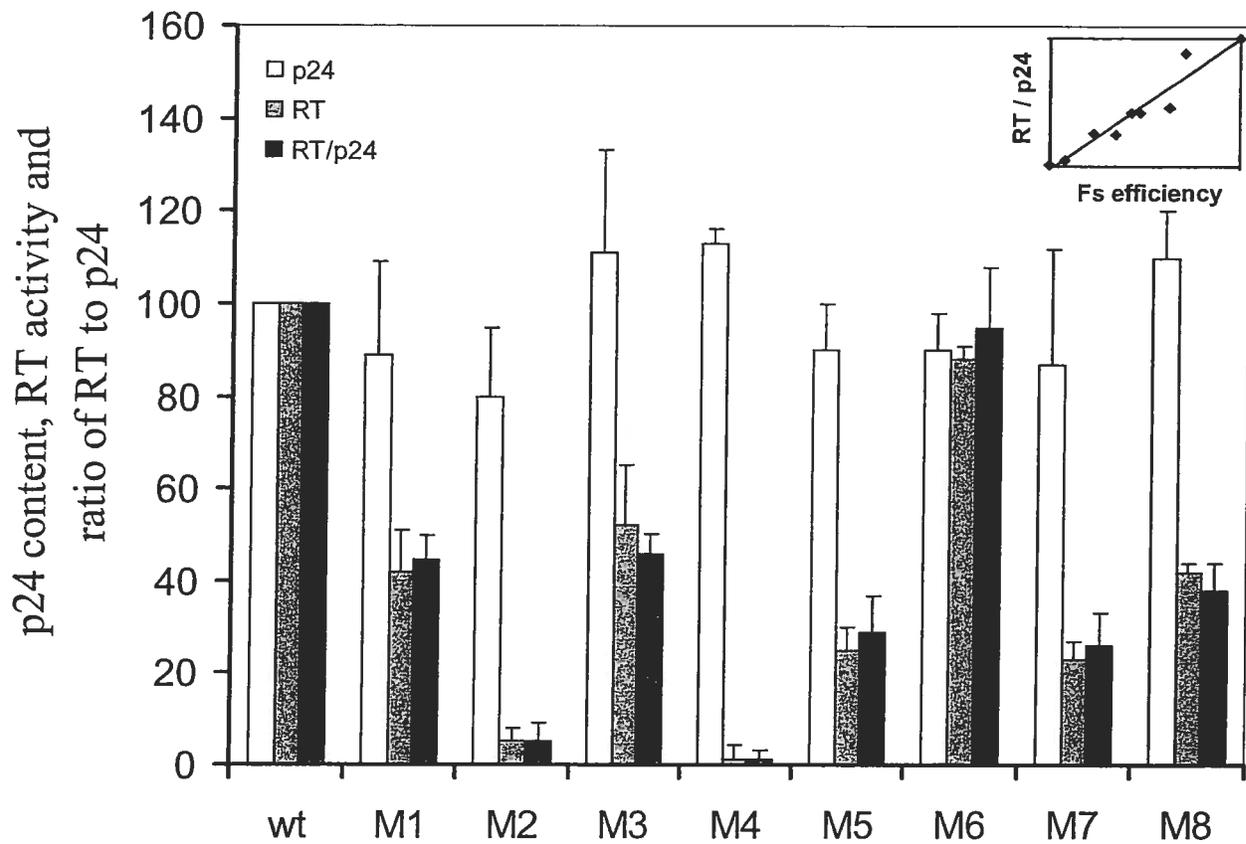


Figure 3-4: Virion samples produced from provirus transfection in HeLa-P4 cells were resolved by a 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and probed by Western blot with an anti-p24 antibody. A representative blot from three separate experiments is shown. P55 (Gag) corresponds to the non-processed polyprotein, whereas p24 (CA) represents a mature component of the processed Gag. The wt lane corresponds to the wild-type protein profile, and lanes M1 to M8 correspond to the frameshift mutants. The ratio of p24 to p55 in the wild-type and mutant virions is indicated under each lane, expressed as the percentage of processing in wild-type virions.

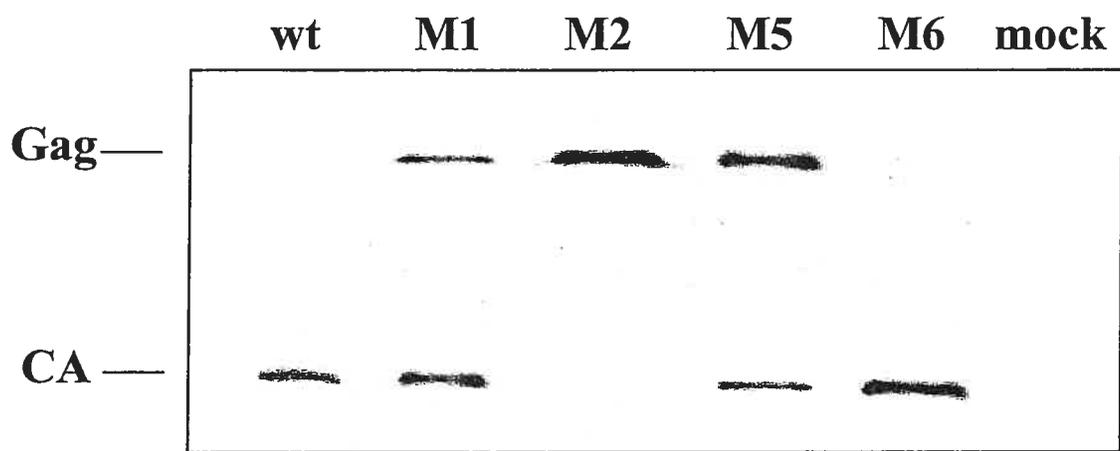


Figure 3-5: Single-round infectivity assays with HIV-1 and its derivatives mutated in the frameshift region. HeLa-P4 cells were infected with p24 normalized amounts of wild-type HIV-1 and mutant viruses. Changes in the infectivity of the mutant viruses (M1 to M8) are indicated relative to the wild-type HIV-1<sub>LAI</sub> (wt), which is arbitrarily set at 100%. The inset presents the correlation ( $R^2=0.752$ ) between the infectivity and the frameshift efficiency (Fs) of the wild-type virus and the frameshift mutants. Each value represents the mean  $\pm$  standard deviation of a triplicate analysis.

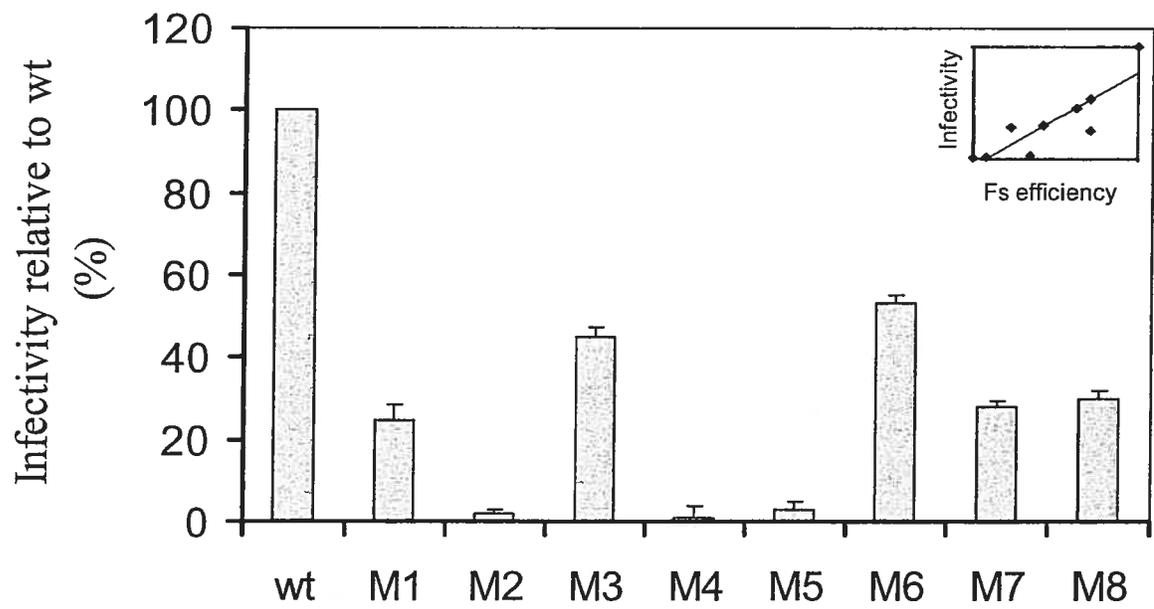
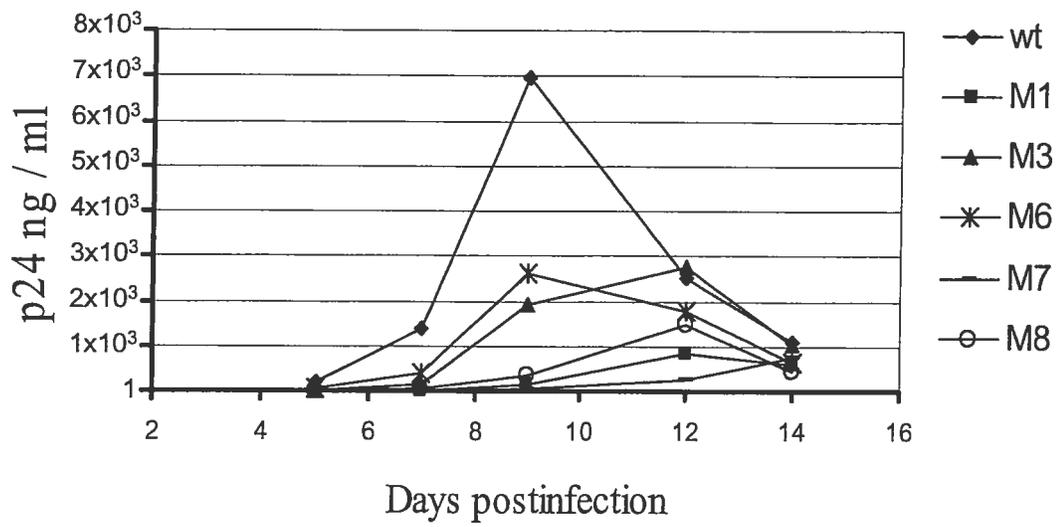


Figure 3-6: Replication kinetics of HIV-1 and mutants with an altered frameshift efficiency. CEM T-cells were infected with p24 normalized amounts of wild-type HIV-1<sub>LAI</sub> (wt) and mutant viruses. Viral replication was monitored in regular intervals by quantification of the p24 concentration in the culture supernatant with a p24 ELISA assay. Mutants M2, M4 and M5 replicate below the level of detection of the assay. A representative analysis from three separate experiments is shown.



## **Chapitre 4**

*IN VIVO* SELECTION FROM A COMBINATORIAL LIBRARY OF PEPTIDES  
INTERFERING WITH THE HIV-1 PROGRAMMED -1 RIBOSOMAL FRAMESHIFT

**Running Title:** Selection of peptides interfering with HIV-1 frameshifting.

Dominic Dulude & Léa Brakier-Gingras\*.

Département de Biochimie, Université de Montréal, Montréal, Québec, Canada.

(in preparation)

\*Corresponding author:

Léa Brakier-Gingras

[REDACTED]

Phone: 514 343 6316

Fax: 514 343 2210

## Abstract

The synthesis of Gag-Pol, the precursor of the enzymatic activities of HIV-1, requires a programmed  $-1$  ribosomal frameshift when ribosomes translate the full-length viral messenger RNA. This frameshift, which occurs with a defined efficiency, is critical for the production of infectious particles. This efficiency is controlled by an RNA structure motif, the stimulatory signal, following the slippery sequence where the frameshift occurs. We previously showed that the bacterial ribosome responds exactly like the eukaryotic ribosome to the frameshift stimulatory signal of HIV-1. In this study, we describe a strategy that we developed for screening a combinatorial peptide library in bacteria, so as to select peptides that perturb HIV-1 frameshifting. A bacterial reporter vector was constructed that codes for the red (RFP) and green (GFP) fluorescent proteins. The frameshift region of HIV-1 was inserted in the beginning of the coding sequence of GFP, such that its synthesis depends upon a  $-1$  frameshift while the RFP expression assesses conventional translation. In parallel, an expression vector was constructed that codes for an arginine-rich 14-residue peptide library fused to thioredoxin. Analysis by fluorescence-activated cell sorting (FACS) of bacteria transformed with the reporter vector and the vector expressing the peptide library detected a population for which the expression of the reporter depending on the  $-1$  frameshifting was decreased. However, the analysis of individual clones from this population showed that the expression of the reporter depending upon conventional translation and that of the reporter depending upon frameshifting were both decreased by two to three fold. This stresses the need to improve this system of selection, and, clearly, more work has to be done in order to select peptides that can perturb HIV-1 frameshifting and, consequently, alter the replication of the virus.

Key words: HIV-1 frameshift, combinatorial peptide library, FACS

## Introduction

The human immunodeficiency virus type 1 (HIV-1) is the causative agent of AIDS, a pandemic that has infected more than 60 million people, over a third of them died (World Health Organization or WHO). The pathogenesis of this virus has been studied intensively during the last decades, and recent reviews on this topic can be found in (Cimarelli et Darlix, 2002; Freed, 2001a). The highly active antiretroviral therapy (HAART), based on the use of a combination of inhibitors of the viral protease and reverse transcriptase, has led to a significant improvement in the treatment of the infected patients (see Fauci, 2003; Gulick, 2003). However, the increasing emergence of drug-resistant virus strains mutated in the reverse transcriptase and protease genes (LaBonte et al., 2003; Little et al., 2002) warrants the identification and validation of new targets for the development of innovative antiretroviral therapy. New classes of antiviral agents are currently being tested in trail and the most promising drugs targets are the viral integration, receptor or co-receptor binding or Tat transactivation, but clearly more viral targets are needed (reviewed in De Clercq, 2002).

The HIV-1 full length viral messenger encodes the structural and enzymatic components from the overlapping *gag* and *pol* open reading frames, respectively (reviewed in Frankel et Young, 1998; Freed, 2001b). Gag is synthesized according to conventional rules of decoding by the ribosomes from the infected cells, whereas Pol shares the same initiator codon as Gag and is produced as a Gag-Pol precursor *via* a programmed  $-1$  ribosomal frameshift that occurs near the 3' end of Gag. The proportion of ribosomes that was found to make a  $-1$  frameshift and synthesize Gag-Pol is estimated to 2 to 10%, depending upon the system of investigation (Biswas et al., 2004; Dulude et al., 2002; Grentzmann et al., 1998; Jacks et al., 1988b). The HIV-1 frameshift event requires two *cis*-acting elements: a heptameric slippery sequence, U UUU UUA, where the shift of frame occurs, and a stimulatory signal located immediately downstream, which is an irregular two-stem helix structure (Dulude et al., 2002) that controls the frameshift efficiency. How the stimulatory signal exerts its function remains unclear. It is known that it makes the ribosomes pause over the slippery sequence, but such pausing is necessary albeit not sufficient for the frameshift to occur (Kontos et al., 2001; Lopinski et al., 2000; Somogyi et al., 1993; Tu et

al., 1992). This suggests that the stimulatory signal is not simply a barrier to the movement of the ribosomes but also interacts specifically with them to control the efficiency of the frameshift. A two-tRNA simultaneous slippage model was initially proposed to account for the frameshift (Jacks et al., 1988a). In this model, for a minority of elongating ribosomes with the peptidyl-tRNA and the aminoacyl-tRNA in the P and A site, respectively, when these ribosomes encounter the slippery sequence, the two tRNA simultaneously un-pair from the mRNA, shift back together with the ribosome that bear them by one nucleotide in the 5' direction and re-pair in the -1 frame (reviewed in Brierley et Pennell, 2001). It was recently proposed that the displacement of the tRNAs is rather sequential, with the movement of the P-site tRNA preceding that of the A-site tRNA (Baranov et al., 2004).

The Gag-Pol to Gag ratio is critical for virus replication and infectivity (Karacostas et al., 1993; Park et Morrow, 1991; Shehu-Xhilaga et al., 2001). It was found that the slippery sequence of HIV-1 is invariant and that the replacement of the HIV-1 slippery sequence by other slippery sequences reduce frameshifting levels and dramatically inhibit infectivity (Biswas et al., 2004). Also, an increase by about three-fold of the frameshift efficiency, obtained when cotransfecting cultured cells with one vector expressing Gag and the other Gag-Pol in a determined proportion, does not produce infectious viral particles (Shehu-Xhilaga et al., 2002b). Telenti et al.(2002) who studied the effect of natural mutations within the stimulatory signal on frameshifting and HIV-1 replication found a severe defect in viral replication when the frameshift efficiency was decreased about two-fold. We recently investigated the effects of altering the frameshift efficiency of HIV-1 by introducing various mutations in the slippery sequence and the frameshift stimulatory signal and found that mutations that decreased this efficiency by as little as 30% also severely decrease viral infectivity and replication of HIV-1 (Dulude et al., 2005).

Kollmus et al. (1996) replaced the HIV upper stem-loop with the iron-responsive element from ferritin mRNA, a stem-loop structure that binds iron regulatory proteins depending upon the iron status of the cells. They found that the frameshift efficiency could be increased about two-fold upon binding of the regulatory proteins, demonstrating that HIV frameshifting can be regulated by RNA-binding proteins. RNA-protein interactions are

known to play a critical role not only in translation but also in transcription, RNA trafficking and packaging. Arginine-rich motifs constitute a particularly versatile framework for recognizing RNA structures (Das et Frankel, 2003; Weiss et Narayana, 1998). Various RNA-protein interactions are needed for HIV-1 replication. Among them, the TAR (trans activation response)-Tat interaction is required for efficient transcription of the HIV genome (Carroll et al., 1992; Wimmer et al., 1999) and the RRE (Rev response element)-Rev interaction controls the cytoplasmic export of intron-containing viral mRNAs of HIV (Battiste et al., 1996).

Harada et al. (1996) showed that a 14-residue peptide library with arginine, serine, asparagine and histidine at each position and a codon bias for arginine produces peptides that bind tightly and selectively to the RRE element of HIV. This approach prompted us to develop a system that enables us to select *in vivo* peptides that perturb the frameshift event in HIV-1. In a previous study, we had shown that the bacterial ribosome responds exactly like the eukaryote ribosome to the frameshift stimulatory signal of HIV-1 (Léger et al., 2004). We therefore decided to use a bacterial system for the selection of peptides, taking advantage of the speed and low cost associated with studies using bacteria, and to exploit the potential of the Fluorescence-Activated Cell Sorting (FACS) approach to sort bacteria containing the wanted peptides.

## Material and Methods

### DNA constructs.

(i) **The reporter plasmid p3RFP-GFPfs.** The p3RFP-GFPfs reporter plasmid was constructed to monitor changes in the frameshift efficiency in the presence of peptides generated from a library. Briefly, this bacterial vector contains the HIV-1 frameshift region (slippery site and the stimulatory signal) inserted in the beginning of the green fluorescent protein (GFP) coding sequence such that its expression requires a programmed  $-1$  frameshift. The vector also contains three copies of the red fluorescent protein (RFP) coding sequence that is used as an internal reporter to assess conventional translation. The

GFP and RFP transcripts are constitutively expressed from a *trp* constitutive promoter. This reporter plasmid was derived from pRNA<sub>luc2</sub> (Belanger et al., 2002), that contains, under control of a *lac* promoter, an IPTG inducible ribosomal RNA operon in which the 16S anti Shine-Dalgarno sequence is mutated. The two reporter RFP and GFP genes have a complementary mutated ribosome-binding site (RBS) so that only ribosomes containing the plasmid-encoded 16S rRNA translate the reporter mRNAs. This specificity of the plasmid is exploited elsewhere to characterize which sites in the 16S rRNA can influence the frameshift efficiency (Léger and Brakier-Gingras, unpublished results).

The p3RFP-GFPfs reporter plasmid was created by first amplifying the GFP coding sequence by PCR from pGFP<sub>emd-N1</sub>, a GFP fusion protein expression vector (a generous gift from Dr M. Bouvier, Université de Montréal). The forward and reverse mutagenic primers: 5'-CCGCAAATGCATGTGAGCAAGGGCGAGGAGCTGT-3' and 3'-AAGGGCTCGAGCGTACGTGGTATGGCTGATTATG-5' used for this amplification introduced an *Nsi*I and an *Xho*I restriction site at the 5' and 3' end of the gene, respectively. The amplified GFP gene was inserted downstream of the *trp* promoter of *Nsi*I-*Xho*I-digested pRNA<sub>luc2</sub>, generating the pGFP construct. The frameshift region of HIV-1 (nucleotides 2075 to 2143 according to HXB2 genome nomenclature) was introduced in the beginning of the GFP coding sequence by inserting an appropriate annealed and phosphorylated 72 bp oligonucleotide cassette, purchased from Sigma Genosys, between the digested *Nsi*I site of pGFP. This generated pGFPfs. An in-frame control plasmid, pGFP(0), was generated from pGFPfs by inserting an adenine immediately after the slippery sequence in the HIV-1 frameshift region, such that GFP is synthesized according to the conventional rules of translation. Next, a pRFP construct derived from pRNA<sub>luc2</sub> was made, using the same approach as that used for creating pGFP, by introducing the RFP gene amplified from pDsRed-T4 (Bevis et Glick, 2002) (a generous gift from Dr B. Glick, University of Chicago). The following forward and reverse mutagenic primers were used: 5'-CGCAAATGCATGCCTCCTCCGAGGACGTCATC-3' and 3'-AAGGGCTCGAGTCGTACGCGCGGCCGCTACAGGAACAGG-5'. These primers also introduced an *Nsi*I and an *Xho*I restriction site at the 5' and 3' end of the gene, respectively.

The pRFP construct was then used as a template to PCR-amplify the RFP coding sequence and its promoter with the following forward and reverse mutagenic primers: 5'-AGCTCGGTACCCAGATCCGGAACATAATGGTGC-3' and 3'-TCTGGGTACCTCTCTTCCGGGCGCTATCATGCC-5' that introduced a *KpnI* restriction site at the 5' and 3' end of the RFP gene. This RFP digested *KpnI*-fragment was inserted in the digested *KpnI* site of pGFPfs and pRFP to generate pRFP-GFPfs and p2RFP, respectively. Finally, the two RFP copies contained in plasmid p2RFP were amplified with the following forward and reverse mutagenic primers: 5'-ATAGCGCTAGCGCCAGTGAATTCGAGCTCGG-3' and 3'-GTGCTGCTAGCTCTCTTCCGGGCGCTATCATGCCATAACC-5' as a fragment with a *NheI* restriction site at the 5' and 3' end. This *NheI* digested fragment was then inserted in the digested *NheI* site of pRFP-GFPfs, generating p3RFP-GFPfs.

**(ii) The peptide library expression plasmid pThio-LIB.** The plasmid used for the expression of the peptide library was derived from the pBAD-Thio-E (Invitrogen), which contains the *E. coli* thioredoxin (*trxA*) gene under control of an arabinose-inducible promoter. This commercial plasmid was modified by replacing its ampicillin resistance gene and its origin of replication with a kanamycin resistance gene and a p15a origin, so as to be compatible with p3RFP-GFPfs. To this end, pBAD-Thio-E was digested with *BsrBI*, the 2.4 Kb fragment containing the *trxA* gene with the upstream arabinose-inducible promoter was isolated and ligated with a 3.2 Kb fragment from *DraI*-digested pACYC177 (New England Biolabs) that contains the p15a origin and the kanamycin resistance gene. This generated pACYC-Thio.

This plasmid was used to express a combinatorial library coding for 14-residue peptides fused to the thioredoxin acting a scaffold protein. The library was made following a procedure adapted from Harada et al. (1996), using a degenerate oligonucleotide, 5'-GACTGACTGGTCCG(MRT)<sub>14</sub>GGTCCTCAGCAGTCAG-3', where M is a A:C mixture in a 1:3 ratio, and R is an A:G mixture in a 3:1 ratio. Each randomized MRT codon encodes arginine, serine, histidine and asparagine in a proportion of 56.25%, 18.75%, 18.75% and 6.25%, respectively. Double-stranded DNAs were synthesized by elongation of an annealed primer (5'-CAGGGGATTT GCTGCATTC-3') using the Sequanase 2.0 (USB) as a DNA

polymerase. These cassettes were digested with *Ava*II, ligated into *Rsr*II-digested pACYC-Thio and used to transform *E. coli* XL-10 Gold (Stratagene). This generated about  $5 \times 10^6$  pThio-LIB transformants.

#### **Expression of plasmids for functional assays.**

*E. coli* BL21-CodonPlus-RP cells (Stratagene), which have extra copies of *argU* and *proL* genes that codes for rare tRNAs, were used for the expression of the reporter and expressor plasmid. This strain was first transformed with the reporter plasmid p3RFP-GFPfs. We then expressed the peptide library in this derived strain. To this end, 300 ng of pThio-LIB DNA were electroporated into these BL21 cells, generating about  $1.5 \times 10^7$  transformants, and grown in 1 ml of LB for 1 h at 37°C without antibiotic. The culture was then diluted with LB to about 0.2 O.D. at 600 nm and the following antibiotics were added: ampicillin (100 µg/ml), kanamycin (50 µg/ml) and chloramphenicol (25 µg/ml). IPTG (1 mM) and arabinose (0,2% w/v) were also added to the culture medium to allow expression of the p3RFP-GFPfs reporter and pThio-LIB expression plasmids. Cells were allowed to grow for an additional 8 h at 37°C before being analysed by FACS.

Plasmids pGFPfs, pGFP(0) and p2RFP were also introduced into the parental BL21 cells, so as to obtain strains that express independently the RFP or the GFP and to calibrate the FACS apparatus. For this purpose, overnight cultures with these different strains were diluted to 0.2 O.D. at 600 nm and grown in 1 ml of LB for 1 h at 37°C with ampicillin (100 µg/ml) and chloramphenicol (25 µg/ml). IPTG (1 mM) was then added and the culture grown for an additional 8h.

#### **FACS analysis.**

For the FACS analysis, bacteria from 1ml of a culture transformed with the reporter and/or expression plasmid were pelleted by a short centrifugation for 1 min at 13,000 rpm. The pellets were washed three times with 1ml of PBS and samples were aliquoted in 2 ml of PBS, with a density of 0.05 O.D.at 600 nm. They were then analyzed by FACS, using the FACSVantage apparatus from BD Biosciences. Fluorescent proteins were excited at 488

nm and the fluorescence emission for the GFP and RFP was recorded at 530 and 586 nm, respectively.

## Results and discussion

### Construction of the reporter plasmid p3RFP-GFPfs.

Fluorescent proteins are used as a tool to study almost every aspect of molecular biology, from subcellular localization of organelles or targeting of fusion proteins to a specific location, transcriptional reporting or bacterial expression, to name just a few examples. Here we used the green and red fluorescent proteins (GFP and RFP, respectively) to develop a two-plasmid system for the screening of a combinatorial peptide library in bacteria, so as to select peptides that perturb frameshifting (Figure 1A). The fluorescent reporters allow us to monitor the fluorescence changes with a FACS apparatus. FACS sorting has emerged as a tool offering significant promise for the high-throughput screening of very large libraries with single cell resolution. The reporter plasmid contains the fluorescent proteins, whereas the expression plasmid generates the peptides (Figure 1B). The reporter plasmid was constructed with the frameshift region of HIV-1 inserted in the beginning of the coding sequence of GFP, such that its synthesis depends upon a -1 frameshift while the RFP expression assesses conventional translation.

A pattern of fluorescence was obtained by FACS for the different intermediary constructs expressed in *E. coli* BL21 cells that lead to the creation of the p3FP-GFPfs reporter plasmid. This pattern is shown in Figure 4-2. Figure 4-2A shows the background fluorescence of BL21 cells: with an excitation at 488 nm, each dot represents the fluorescence at 530 nm on the X-axis and at 586 nm on the Y-axis, which corresponds to the GFP and RFP emission, respectively. In these FACS patterns, two axes arbitrarily determine four quadrants: the lower left one represents background cells without fluorescence, the upper left and lower right quadrant show cells positive for RFP or GFP, respectively, and the upper right quadrant displays cells positive for both RFP and GFP.

Expression of the pGFPfs construct, with the frameshift region of HIV-1 inserted in the beginning of the coding sequence of GFP, showed a level of fluorescence about five fold above the background (Figure 4-2B). The fluorescence from the in-frame GFP control, pGFP(0), allowed us to measure a frameshift efficiency of  $3.7 \pm 0.7\%$  for the pGFPfs construct (Figure 4-2C). This is in perfect agreement with the value of  $4.0 \pm 0.6\%$  that was previously obtained when measuring HIV-1 frameshifting with a luciferase reporter in bacteria (Léger et al., 2004). Also, it can be seen that the fluorescence level of bacteria transformed with pGFPfs is sufficient to allow us to monitor changes in frameshifting efficiency.

The fluorescence of the RFP protein measured with the pRFP construct, that contains one copy of the RFP gene and for which the RFP mRNA is translated according to conventional rules, showed that only 50% of the bacteria express this protein, with an average emission value two fold above the background (Figure 4-2D). This low value might be caused by a slow maturation of the RFP protein, and also the fact that excitation is made at only 40% of the RFP maximal excitation value, since only one wavelength can be used to monitor simultaneously RFP and GFP fluorescence. When the two fluorescent proteins were expressed from the same plasmid pRFP-GFPfs, with the in-frame RFP and the GFPfs under the -1 frameshift control, only 63% of bacteria showed a fluorescence emission corresponding to both the RFP and GFP (Figure 4-2E, upper right quadrant), and this low amount was again caused by a low level of RFP fluorescence. To circumvent this drawback, we increased the number of copies of the RFP gene in the reporter plasmid. As shown in Figure 4-2F, when two copies of the RFP gene are present on the same plasmid, p2RFP, about 90% of bacteria show a level of fluorescence about three and a half fold above the background level. Two additional RFP copies were then inserted in plasmid pRFP-GFPfs, generating the p3RFP-GFPfs reporter plasmid. The fluorescence pattern obtained with this plasmid revealed that about 97% of the bacteria exhibited both GFP and RFP fluorescence (Figure 4-2G, upper right). These results obtained with the p3RFP-GFPfs reporter plasmid demonstrate the capacity of the FACS to resolve bacteria as single events

and to appropriately discriminate the GFP and RFP signals, and we therefore used this plasmid in our subsequent screening.

### **Construction of the pThio-LIB peptide library.**

A combinatorial library of 14-residue peptides fused to *E.coli* thioredoxin, as a carrier protein, was used to express arginine-rich peptides within bacteria. Thioredoxin was chosen as a carrier protein for random peptides because of its stability and the presence of a convenient *RsrII* restriction site in the region encoding the active site (Cys-Gly-Pro-Cys) of this enzyme, that allows to conformationally constrain peptides expressed at this site (LaVallie et al., 1993). The tethering of both ends of the inserted peptide by a disulfide bond provides conformational constraints that may increase the affinity of a peptide for its target (Corey et Corey, 1996; Lu et al., 1998).

To obtain the library, we first constructed a vector that places the expression of thioredoxin under the tight control of the inducible *E. coli* P<sub>BAD</sub> promoter (Guzman et al., 1995), a promoter that is positively regulated the presence of arabinose. The pACYC-Thio vector that we used is derived from a parental plasmid, pBAD-Thio-E, which contains the *E. coli* thioredoxin gene (*trxA*) but differs from this plasmid by its origin of replication and a resistance to kanamycin instead of ampicillin. An oligonucleotide with restriction sites at each end and a central segment encoding 14-residue peptides was cloned into the *RsrII* restriction site that corresponds to the active site of thioredoxin. Similar constructions have been used to make thioredoxin fusions that were selected for various biological activities (Colas et al., 1996; LaVallie et al., 1993; Lu et al., 1998). The 14-residue peptide library was made with a subset of amino acids (arginine, serine, asparagine and histidine, with a bias for arginine), following a procedure used by the group of Frankel (Harada et al., 1996). A 14-residue peptide has approximately  $3 \times 10^8$  possible sequences. Based on the number of bacterial transformants we obtained, our construction resulted in a library of approximately  $5 \times 10^6$  peptides. This represents only a small fraction of all the possible sequences that could be expressed with our library ( $4^{14}$ ) and there should be virtually no duplication of sequences in this library.

**Selection of small peptides: an attempt to the validation of the concept.**

Our goal was to select small arginine-rich peptides that perturb the frameshift efficiency of HIV-1. This is expected to be achieved when selecting bacteria transformed with the plasmid expressing the peptide library (pThio-LIB) and containing the reporter plasmid in which the GFP expression is decreased or increased while the RFP expression is not altered. Figure 4-3 compares the pattern of fluorescence of BL21 cells containing the p3RFP-GFPfs reporter and transformed with the uninduced expression plasmid to that of cells in which the plasmid expressing the library is induced. As seen from the FACS analysis, two regions in the pattern of fluorescence could be reproducibly observed in which potential peptide candidates either decrease or increase the GFP expression, thus increasing or decreasing the efficiency of HIV-1 frameshift. This analysis thus indicates that it is possible to select peptides among the library that are able to alter HIV-1 frameshift. We then sorted  $1 \times 10^7$  bacteria transformed with the induced peptide library by FACS and selected a pool of about 240 and 40 clones that met our selection criteria and corresponded to a decrease or an increase in the GFP expression, respectively, whereas the RFP expression did not appear to be significantly altered (data not shown). These clones were submitted to two successive rounds of selection and then analyzed individually for the expression of GFP and RFP, under conditions where the expression of the selected peptide was induced or not induced. This was done with a microplate reader (Flex station, Molecular Devices). However, all the individual clones selected for an increase in the GFP expression were false positives, whereas, those selected for a decrease in the GFP expression upon induction of the peptide expression showed a decreased of two to three fold in the level of GFP and also RFP, the ratio of GFP to RFP being unchanged (data not shown). This suggests that the peptides expressed in these clones affect translation in general and not specifically the frameshift event. It is possible that these peptides recognize specific RNA structures but their target could be the ribosome. Clearly, more work needs to be done to obtain a system that is appropriate for the selection of peptides that can perturb HIV-1 frameshifting. If we succeed in the selection of bacterial clones that express peptides

that influence frameshifting in further studies, the sequence of these peptides will be obtained by PCR amplification and sequencing of the bacterial DNA. The next steps in this research will be to confirm that these peptides perturb HIV-1 frameshifting in a eukaryote system, and in the affirmative, to investigate the effect of these peptides on viral replication and also to investigate whether these peptides interact with HIV-1 frameshift stimulatory signal.

### **Acknowledgements**

This work was supported by a grant from CIHR. We thank Gerardo Ferbeyre and Nikolaus Heveker for their interest in this study and for stimulating discussion. We are grateful to Serge Sénéchal for his technical help with the FACS analysis.

## References

- Baranov PV, Gesteland RF et Atkins JF. 2004. P-site tRNA is a crucial initiator of ribosomal frameshifting. *Rna* 10:221-230.
- Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD et Williamson JR. 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* 273:1547-1551.
- Belanger F, Leger M, Saraiya AA, Cunningham PR et Brakier-Gingras L. 2002. Functional studies of the 900 tetraloop capping helix 27 of 16S ribosomal RNA. *J Mol Biol* 320:979-989.
- Bevis BJ et Glick BS. 2002. Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat Biotechnol* 20:83-87.
- Biswas P, Jiang X, Pacchia AL, Dougherty JP et Peltz SW. 2004. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* 78:2082-2087.
- Brierley I et Pennell S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic-1 ribosomal frameshifting. *Cold Spring Harb Symp Quant Biol* 66:233-248.
- Carroll R, Peterlin BM et Derse D. 1992. Inhibition of human immunodeficiency virus type 1 Tat activity by coexpression of heterologous trans activators. *J Virol* 66:2000-2007.
- Cimarelli A et Darlix JL. 2002. Assembling the human immunodeficiency virus type 1. *Cell Mol Life Sci* 59:1166-1184.
- Colas P, Cohen B, Jessen T, Grishina I, McCoy J et Brent R. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380:548-550.
- Corey MJ et Corey E. 1996. On the failure of de novo-designed peptides as biocatalysts. *Proc Natl Acad Sci U S A* 93:11428-11434.

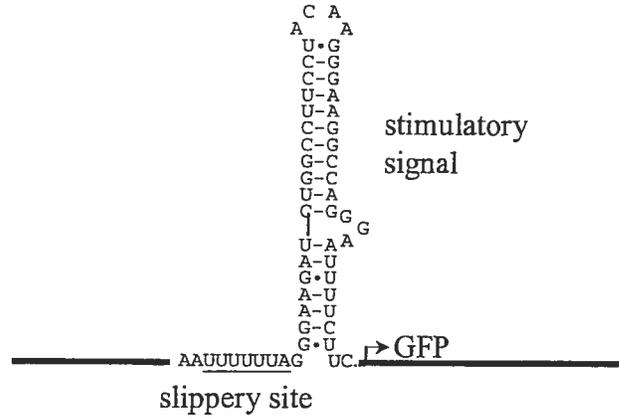
- Das C et Frankel AD. 2003. Sequence and structure space of RNA-binding peptides. *Biopolymers* 70:80-85.
- De Clercq E. 2002. New developments in anti-HIV chemotherapy. *Biochim Biophys Acta* 1587:258-275.
- Dulude D, Baril M et Brakier-Gingras L. 2002. Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res* 30:5094-5102.
- Dulude D, Berchiche YA, Gendron K, Brakier-Gingras L et Heveker N. 2005. Mutations within the signals that control the efficiency of the programmed -1 ribosomal frameshift interfere with the replication of the human immunodeficiency virus type 1. *submitted to Virology*.
- Fauci AS. 2003. HIV and AIDS: 20 years of science. *Nat Med* 9:839-843.
- Frankel AD et Young JA. 1998. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 67:1-25.
- Freed EO. 2001a. HIV-1 replication. *Somat Cell Mol Genet* 26:13-33.
- Freed EO, and M. A. Martin. 2001b. HIVs and their replication. In: D. M. Knipe PMH, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus, ed. *Fields virology, 4th ed*. Philadelphia, Pa.: Lippincott Williams & Wilkins. pp 1971-2041.
- Grentzmann G, Ingram JA, Kelly PJ, Gesteland RF et Atkins JF. 1998. A dual-luciferase reporter system for studying recoding signals. *Rna* 4:479-486.
- Gulick RM. 2003. New antiretroviral drugs. *Clin Microbiol Infect* 9:186-193.
- Guzman LM, Belin D, Carson MJ et Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177:4121-4130.
- Harada K, Martin SS et Frankel AD. 1996. Selection of RNA-binding peptides in vivo. *Nature* 380:175-179.
- Jacks T, Madhani HD, Masiarz FR et Varmus HE. 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55:447-458.

- Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ et Varmus HE. 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 331:280-283.
- Karacostas V, Wolffe EJ, Nagashima K, Gonda MA et Moss B. 1993. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* 193:661-671.
- Kollmus H, Hentze MW et Hauser H. 1996. Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA* 2:316-323.
- Kontos H, Napthine S et Brierley I. 2001. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* 21:8657-8670.
- LaBonte J, Lebbos J et Kirkpatrick P. 2003. Enfuvirtide. *Nat Rev Drug Discov* 2:345-346.
- LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF et McCoy JM. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology (N Y)* 11:187-193.
- Léger M, Sidani S et Brakier-Gingras L. 2004. A reassessment of the response of the bacterial ribosome to the frameshift stimulatory signal of the human immunodeficiency virus type 1. *Rna* 10:1225-1235.
- Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, Koup RA, Mellors JW, Connick E, Conway B, Kilby M, Wang L, Whitcomb JM, Hellmann NS et Richman DD. 2002. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 347:385-394.
- Lopinski JD, Dinman JD et Bruenn JA. 2000. Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol* 20:1095-1103.
- Lu Z, Tripp BC et McCoy JM. 1998. Displaying libraries of conformationally constrained peptides on the surface of Escherichia coli as flagellin fusions. *Methods Mol Biol* 87:265-280.

- Park J et Morrow CD. 1991. Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. *J Virol* 65:5111-5117.
- Shehu-Xhilaga M, Crowe SM et Mak J. 2001. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol* 75:1834-1841.
- Shehu-Xhilaga M, Lee JY, Campbell S, Marshall JA, Crowe SM et Mak J. 2002. Overexpression and incorporation of GagPol precursor does not impede packaging of HIV-1 tRNA(Lys3) but promotes intracellular budding of virus-like particles. *J Biomed Sci* 9:697-705.
- Somogyi P, Jenner AJ, Brierley I et Inglis SC. 1993. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* 13:6931-6940.
- Telenti A, Martinez R, Munoz M, Bleiber G, Greub G, Sanglard D et Peters S. 2002. Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem-loop structure. *J Virol* 76:7868-7873.
- Tu C, Tzeng TH et Bruenn JA. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci U S A* 89:8636-8640.
- Weiss MA et Narayana N. 1998. RNA recognition by arginine-rich peptide motifs. *Biopolymers* 48:167-180.
- Wimmer J, Fujinaga K, Taube R, Cujec TP, Zhu Y, Peng J, Price DH et Peterlin BM. 1999. Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. *Virology* 255:182-189.

Figure 4-1: A) HIV-1 Frameshift region. The underlined slippery site is followed by a frameshift stimulatory signal consisting of a long irregular stem-loop, where a purine bulge separates two stems (Dulude et al., 2002). B) The two-plasmid system used for selecting peptides that perturb the frameshift efficiency. In the expression plasmid, p3RFP-GFPfs, thioredoxin is expressed under the control of an arabinose-inducible  $P_{BAD}$  promoter on a pACYC177-derived plasmid. A 14-residue arginine-rich peptide library is expressed fused to the thioredoxin as a carrier protein. The reporter plasmid contains two reporter genes coding for RFP and GFP. The frameshift region of HIV-1, containing the slippery site and the downstream stimulatory signal, is inserted in the beginning of the coding sequence of GFP, such that GFP synthesis depends upon a -1 frameshift. The RFP expression depends on a conventional translation.

**A**



**B**

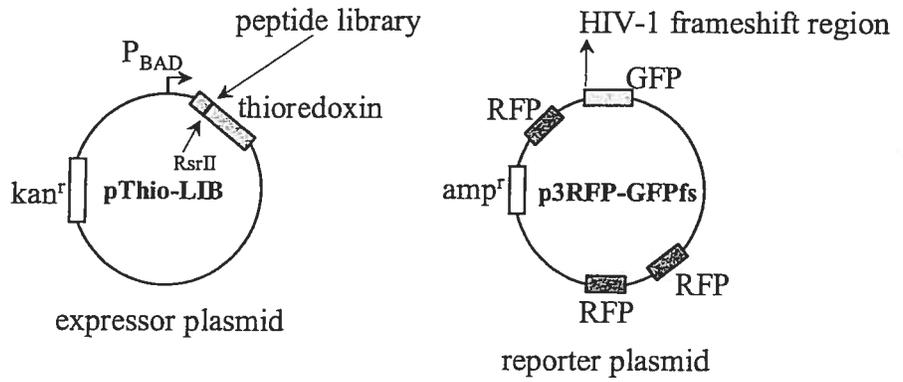


Figure 4-2: FACS fluorescence patterns obtained with *E. coli* BL21 cells transformed with different fluorescent reporter constructs, containing either GFP or RFP or both. RFP is always expressed according to the conventional rules of translation. Fluorescence at 530 nm (X-axis/GFP) and 586 nm (Y-axis/RFP) was measured with  $10^4$  cells excited at 488 nm. A. Background fluorescence with BL21 cells. B. BL21 cells containing pGFPfs, in which GFP expression requires a -1 frameshift. C. BL21 cells containing pGFP(0), in which the in-frame GFP is synthesized according to the conventional rules of translation. D. BL21 cells containing pRFP and expressing RFP. E. BL21 cells containing pRFP-GFPfs, expressing RFP and GFP, the GFP expression requiring a -1 frameshift. F. BL21 cells containing p2RFP, which has two copies of the RFP gene. G. BL21 cells containing p3RFP-GFPfs, with three copies of the RFP gene plus one copy of the GFP gene, the GFP expression requiring a -1 frameshift.

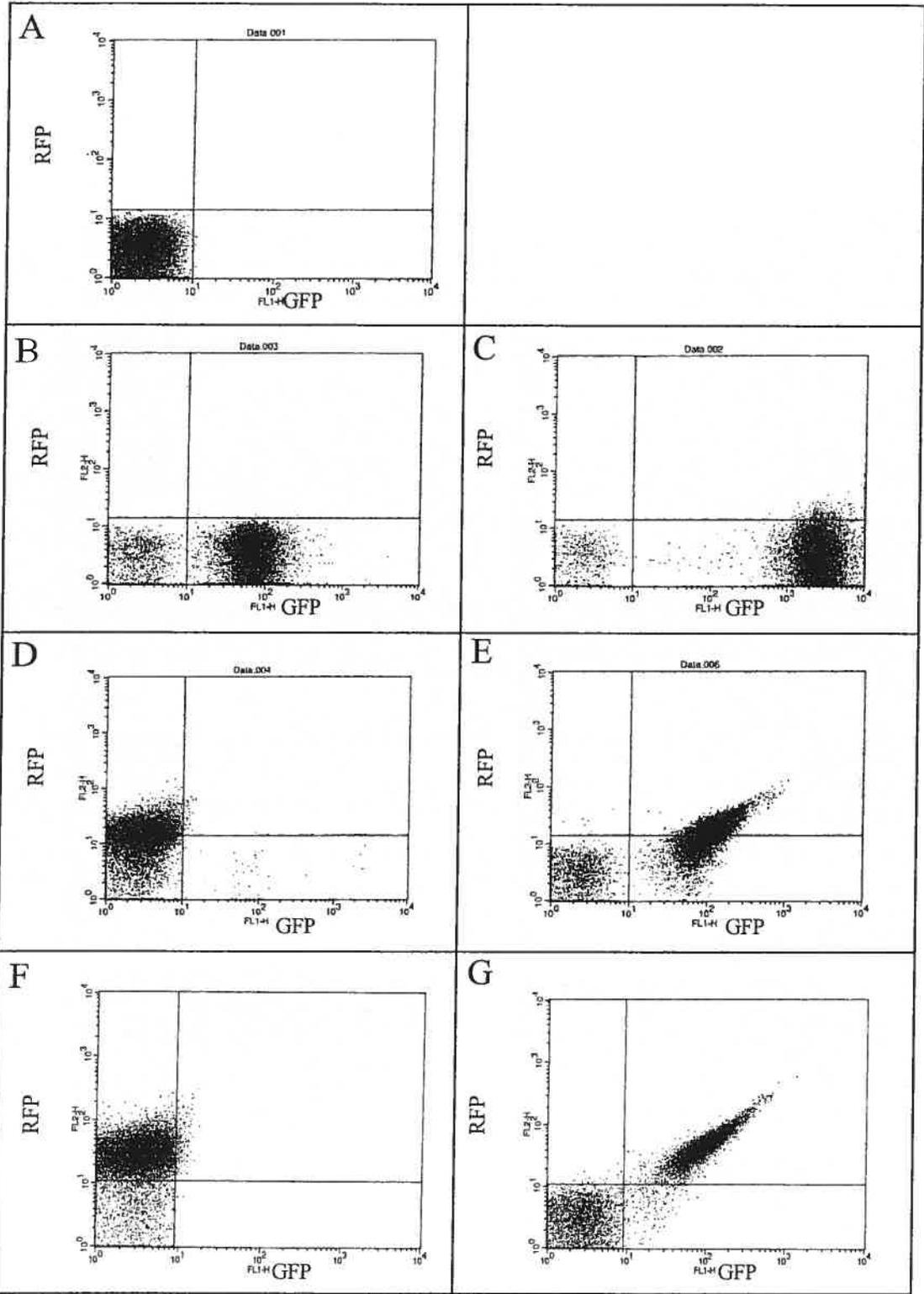
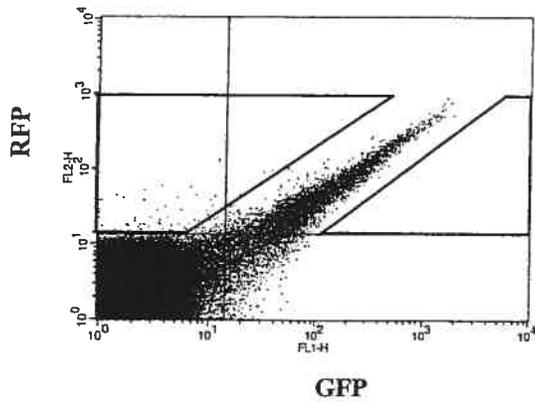
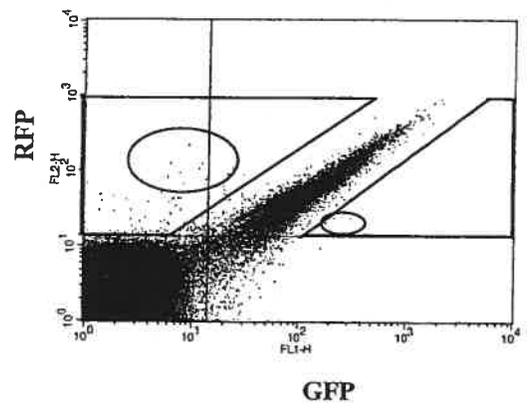


Figure 4-3: FACS fluorescence patterns of *E. coli* BL21 cells transformed with the reporter plasmid and the plasmid expressing the peptide library. Fluorescence at 530 nm (X-axis/GFP) and 586 nm (Y-axis/RFP) was measured with  $10^4$  cells excited at 488 nm. A. BL21 cells containing the p3RFP-GFPfs reporter and the uninduced pThio-LIB expression plasmid. B. BL21 cells containing the p3RFP-GFPfs reporter and the pThio-LIB expression plasmid, with induction of the peptide library. Circles indicate bacterial clones expressing peptides that influence the GFP fluorescence when compared to conditions under which the peptide library is not induced. The parallelogram-like shapes determine the boundaries that will be used to sort out bacteria containing potential peptide candidates that decrease (left) or increase (right) GFP fluorescence but do not affect RFP fluorescence.

**A**



**B**



## Chapitre 5

## Discussion

Depuis l'identification du premier cas de décalage de phase de lecture en  $-1$ , chez le virus du sarcome de Rous, par Jacks et Varmus en 1985, les nombreuses recherches effectuées par la communauté scientifique ont identifié près d'une centaine de cas impliquant ce type d'anomalie traductionnelle (<http://recode.genetics.utah.edu>, Baranov et al., 2003). Le mécanisme contrôlant le déphasage en  $-1$  demeure toutefois encore mal connu. L'ensemble de nos travaux a eu pour objectif de mieux définir le déphasage en  $-1$  du VIH-1, d'abord en caractérisant la structure du signal stimulateur qui contrôle son efficacité, et, ensuite, en étudiant l'effet de changements de l'efficacité du déphasage sur la réplication du virus. Finalement, nous avons entamé une étude visant à obtenir des substances qui perturbent le déphasage du VIH-1 et qui pourront conduire au développement d'une nouvelle classe d'agents anti-VIH.

Dans le premier volet de notre recherche, nous avons caractérisé la structure du signal stimulateur du décalage de phase de lecture en  $-1$  du VIH-1. Rappelons que nos études portent sur la séquence du sous-type B du groupe M du VIH-1, le groupe M étant prépondérant dans le monde et le sous-type B étant le plus répandu des sous-types de ce groupe dans le monde occidental. Le signal stimulateur avait été initialement défini comme une structure en tige-boucle séparée de la séquence glissante, où se produit le déphasage, par une séquence séparatrice de sept nucléotides. Notre objectif était de déterminer si la séquence de l'ARN viral suivant la structure en tige-boucle pouvait aussi influencer l'efficacité du déphasage en  $-1$  et participer à la formation d'une structure stimulatrice plus complexe. Nos résultats ont démontré que le signal stimulateur du déphasage du VIH-1 est effectivement plus complexe qu'une simple tige-boucle et consiste en une hélice irrégulière plus longue que la structure proposée initialement. Cette caractérisation nous fournit des informations intéressantes, à partir desquelles nous pouvons suggérer comment le signal stimulateur peut favoriser le déphasage en  $-1$ .

Dans le second volet de notre étude sur le déphasage en  $-1$  du VIH-1, nous avons examiné comment la variation de l'efficacité de cet événement affecte la réplication du VIH-1. Nos résultats révèlent que même un changement modeste de l'efficacité du déphasage réduit significativement l'infectivité et la réplication virale. Cette observation souligne l'importance potentielle du déphasage du VIH-1 comme cible thérapeutique.

Un troisième volet de notre recherche, suite à cette observation, a consisté à développer une méthode de sélection de peptides capable de perturber l'efficacité du déphasage du VIH-1, probablement en interagissant avec le signal stimulateur de ce déphasage. Nous avons construit un vecteur exprimant une librairie combinatoire de peptides et utilisé un système d'expression bactérien approprié pour sélectionner les peptides ayant l'effet recherché. Les résultats préliminaires obtenus avec cette approche montrent que nous sélectionnons plutôt des clones où la traduction est affectée en général, et non, spécifiquement lors du déphasage en  $-1$ . Ces données nous indiquent qu'il faut modifier notre approche de sélection pour augmenter sa capacité d'obtenir des peptides agissant spécifiquement sur le déphasage.

## **5.1 Caractérisation du signal stimulateur contrôlant l'efficacité du déphasage en $-1$ du VIH-1.**

Le décalage de phase du cadre de lecture en  $-1$  qui se produit lors de la traduction de l'ARNm viral pleine longueur du VIH-1 permet la synthèse du précurseur des enzymes virales, Gag-Pol. La traduction conventionnelle de ce messenger viral par la majorité des ribosomes permet la synthèse du précurseur des protéines de structure, Gag, tandis que, suite au déphasage, une faible proportion de ribosomes changent le cadre de lecture à un endroit spécifique dans la séquence correspondant au gène *gag* et poursuivent la traduction du gène *pol* dans le cadre  $-1$ . Il a d'abord été démontré que le déphasage se produisait à la séquence glissante U UUU UUA du messenger (Jacks et al., 1988b), qui correspond aux

nucléotides 2085 à 2091, selon la nomenclature du génome HXB2 du VIH-1. Des études de mutagenèse dirigée par la même équipe ont montré que le signal stimulateur du déphasage, en aval de la séquence glissante, est une structure en tige-boucle où la tige a 11 paires de bases et est coiffée par une tétraboucle. Ce résultat était appuyé par des analyses informatiques de prédictions de structure (Le et al., 1989). Des expériences de détermination de structures d'ARN utilisant des attaques par des agents enzymatiques ont également montré l'existence de la structure en tige-boucle (Kang, 1998). Cependant, ces études de structure portaient sur une portion d'ARN se limitant à la séquence glissante et la région en aval de cette séquence incluant la tige-boucle. Nous avons donc voulu examiner si la séquence du messenger qui suivait la tige-boucle pouvait aussi influencer le déphasage.

Pour ce faire, nous avons introduit la région de déphasage du VIH-1, contenant la séquence glissante, la séquence séparatrice de sept nucléotides, la tige-boucle, ainsi qu'environ 30 nucléotides de la séquence suivant cette tige-boucle, au début de la séquence codante d'un gène rapporteur dans un vecteur approprié. Le gène rapporteur codait pour la luciférase de la luciole et l'insertion était telle que la synthèse de la luciférase dépendait d'un déphasage en  $-1$ . Nous avons introduit le vecteur dans des cellules de mammifères en culture et observé que le taux de synthèse de la luciférase, qui reflète l'efficacité du déphasage, était supérieur à celui obtenu avec une insertion se limitant à la séquence glissante, la séquence séparatrice et la tige-boucle. Nous avons ensuite procédé à une série de délétions raccourcissant la séquence suivant la tige-boucle et vérifié l'effet de ces délétions sur l'efficacité du déphasage. Nous avons observé qu'une séquence longue de neuf nucléotides, riche en pyrimidines, située juste après la tige-boucle était suffisante pour obtenir une efficacité du déphasage semblable à celle obtenue avec l'insertion plus longue (Figure 2-2). Nous avons par la suite soumis des transcrits d'ARN correspondant à la région de déphasage du VIH-1, englobant ces neuf nucléotides additionnels, à une digestion par l'endonucléase  $V_1$ , qui attaque spécifiquement les structures d'ARN hélicoïdale. Le patron des attaques a montré que le signal stimulateur du déphasage consiste en une hélice irrégulière où deux tiges hélicoïdales sont séparées par une protubérance de trois purines

(Figure 2-3). La tige supérieure de cette structure correspond à la tige-boucle classique et la tige inférieure est formée par un appariement entre la séquence, nommée auparavant séquence séparatrice, en amont de cette tige-boucle, et la séquence complémentaire, riche en pyrimidines, en aval de cette tige-boucle. Nous avons ensuite utilisé la mutagenèse dirigée pour confirmer l'existence de la portion hélicoïdale inférieure de la nouvelle structure que nous proposons pour le signal stimulateur du déphasage VIH-1, en utilisant le vecteur d'expression codant pour la luciférase (Figure 2-4). Nous avons aussi analysé 139 séquences du gène *pol* obtenues d'isolats naturels du groupe M du VIH-1 provenant d'une base de données de séquences du VIH, afin de vérifier si des changements observés dans la séquence de la tige inférieure du signal stimulateur permettaient la formation de cette structure (Los Alamos, <http://hiv-web.lanl.gov>, Kuiken et al., 2003). Parmi ces séquences, 19 étaient de type sauvage et 62 contenaient des mutations qui conservent les appariements Watson-Crick ou permettent des appariements non-canoniques tel G-U. Parmi les autres séquences, 53 avaient des mutations qui conduisent à la formation d'appariements de type A-C, mais de telles paires peuvent participer à la formation d'hélice d'ARN (Nowakowski et Tinoco, 1999). Seulement cinq séquences sur les 139 examinées contenaient des mésappariements de type C-U, G-A, U-U qui déstabilisent les hélices d'ARN (Table 2-1). Cette analyse montre donc que la formation de la tige inférieure du signal stimulateur du déphasage du VIH-1 est conservée dans plus de 96% des séquences virales analysées.

Nos résultats montrent donc que le signal stimulateur du déphasage du VIH-1 est plus complexe que ce qui avait été initialement proposé. Comment les signaux stimulateurs du déphasage favorisent le changement du cadre de lecture est mal compris. Il a été montré que ces signaux provoquent une pause du ribosome et il a été proposé qu'ils interagissent de façon spécifique avec le ribosome. En ce qui concerne la structure du signal stimulateur que nous proposons pour le VIH-1, il est évident que la portion supérieure du signal joue le rôle le plus important. En effet, nos travaux montrent que la destruction de cette tige supérieure diminue d'environ dix fois l'efficacité du déphasage en -1, tandis que la destruction de la portion inférieure du signal ne diminue l'efficacité du déphasage que d'un

facteur deux. L'impact de la tige inférieure ne semble pas majeur, à priori, mais comme nous l'avons souligné au cours de l'Introduction de cette thèse, des changements d'un facteur deux de l'efficacité du déphasage en  $-1$  peuvent être suffisants pour empêcher la propagation d'un virus (section 1.3.5). Le signal stimulateur que nous proposons pour le VIH-1 est thermodynamiquement légèrement plus stable que le signal classique (le  $\Delta G^\circ$  est  $-24.6$  Kcal/mol versus  $-21.3$  Kcal/mol pour le signal classique). Toutefois, il est peu probable que l'effet observé sur l'efficacité du déphasage soit dû à l'accroissement de la stabilité causé par la formation de la portion inférieure du signal. En fait, Bidou et al. (1997) avaient observé qu'une augmentation de la stabilité de la tige classique (variation de  $\Delta G^\circ$  de  $-3.4$  Kcal/mol) diminuait plutôt l'efficacité du déphasage et proposé que, si le signal stimulateur est plus stable, la période durant laquelle les ribosomes demeurent arrêtés sur la séquence glissante augmente, avec le résultat qu'une proportion de ces ribosomes ayant changé de phase peuvent se replacer dans le cadre initial. Lors de nos études de mutagenèse du signal stimulateur du VIH-1, nous avons montré l'importance de la protubérance de trois purines séparant les deux tiges de ce signal. En effet, lorsque nous avons remplacé ces trois purines par des pyrimidines, l'efficacité du déphasage a été réduite d'environ deux fois, en fait à la valeur obtenue avec la région classique de déphasage. Il est probable que ces trois purines forment un point de reconnaissance spécifique entre le signal stimulateur et le ribosome.

Nos résultats montrent que la nouvelle structure du signal stimulateur du VIH-1 est séparée de la séquence glissante par seulement un nucléotide. Une telle séparation diffère de celle de cinq à neuf nucléotides habituellement observée pour des signaux stimulateurs du déphasage (Brierley, 1995). Une distance de cinq à neuf nucléotides semble importante pour que le ribosome s'arrête au niveau de la séquence glissante lorsqu'il rencontre le signal stimulateur. La présence de la tige inférieure dans le signal stimulateur du déphasage du VIH-1 suggère deux possibilités : soit le ribosome déroule cette structure lorsqu'il la rencontre, ce qui se produit avant qu'il ait atteint la séquence glissante soit, au contraire, il engouffre l'hélice irrégulière complète dans le tunnel d'entrée du messager en continuant à

progresser. Chez RSV, qui utilise aussi un déphasage en  $-1$ , le signal stimulateur en pseudonoeud est également situé juste un nucléotide après la séquence glissante, et on peut aussi s'interroger sur la séquence d'évènements lors de la rencontre entre le ribosome et ce signal. Selon les récentes données cristallographiques sur la structure du ribosome bactérien, il semble peu probable qu'une structure en hélice double-brin comme le signal stimulateur du VIH-1 puisse pénétrer dans le tunnel d'entrée du ribosome en restant intact. Les peu de données structurales que nous possédons au sujet du ribosome eucaryote ne nous indiquent si la même situation s'applique au tunnel d'entrée de ce ribosome. Toutefois, étant donné la conservation des mécanismes d'élongation de la traduction chez les bactéries et les eucaryotes, le ribosome eucaryote devrait, lui aussi, restreindre l'accès de structures en hélice dans le tunnel d'entrée du messageur. Il faut cependant noter que certains évènements de reprogrammation suggèrent que des structures en hélice pourraient pénétrer à l'intérieur du ribosome. Dans le cas du saut du ribosome lors de la traduction du messageur du gène *60* du bactériophage T4 (voir section 1.3.1.2), une structure en tige-boucle suit immédiatement le codon où le ribosome entame son saut, et cette structure est importante pour la réalisation du saut. Aussi, lors de l'incorporation de la sélénocystéine durant la synthèse protéique chez les bactéries, une structure en tige-boucle est également située immédiatement après le codon de terminaison où la reprogrammation de la traduction du ribosome se produit. Bien que ce soit purement spéculatif à l'heure actuelle, nous proposons la séquence suivante d'évènements pour décrire le fonctionnement du signal stimulateur du VIH-1. La présence de la tige inférieure stabiliserait la protubérance de trois purines, qui pourrait interagir avec le ribosome lors d'une première rencontre. Cette interaction favoriserait le positionnement adéquat de la partie supérieure du signal stimulateur sur le ribosome. Ce dernier déroulerait ensuite la tige inférieure en continuant à progresser sur le messageur jusqu'à atteindre la séquence glissante. La probabilité du déphasage y serait augmentée suite à l'interaction de la tige supérieure avec le ribosome.

Pendant que nous complétons la caractérisation du signal stimulateur du VIH-1, une publication par Dinman et al. (2002) a également démontré qu'une séquence d'environ

dix nucléotides suivant la structure de la tige-boucle classique augmentait l'efficacité du déphasage du VIH-1. En se basant sur des expériences de mutagenèse dirigée et d'attaque avec des agents enzymatiques, ces chercheurs ont proposé une structure différente de celle que nous avons caractérisée pour le signal stimulateur. Selon leur modèle, la séquence riche en pyrimidines suivant la tige-boucle classique, nommée le troisième brin, interagit avec trois nucléotides de la boucle coiffant cette tige pour former la tige 2 d'un pseudonoed. Cette tige 2 est stabilisée par une hélice triple impliquant quatre paires de bases de la tige classique et ce troisième brin (Figure 5-1A). La formation d'une hélice triple requiert que le troisième brin établisse des ponts hydrogènes avec les paires de bases présentes dans la tige en duplex (revue par Frank-Kamenetskii et Mirkin, 1995; Nowakowski et Tinoco, 1999). Cependant, l'orientation ou la direction du troisième brin est critique pour la formation d'une hélice triple. En effet, si ce brin est riche en pyrimidines, comme dans le cas du VIH-1, il doit être dans une orientation parallèle par rapport au brin riche en purines du duplex (Figure 5-1B) (Frank-Kamenetskii et Mirkin, 1995; Han et Dervan, 1993; Lim et Barton, 1998). Or, dans la structure proposée par le groupe de Dinman, ce brin est dans une orientation antiparallèle, ce qui est défavorable. De plus, l'hélice triple proposée par ce groupe implique trois interactions sur quatre comprenant un uracile. Ce type d'interaction U\*G-C (où U\* correspond au troisième brin) n'est pas habituellement rencontré dans les hélices triples (revue par Frank-Kamenetskii et Mirkin, 1995; Han et Dervan, 1993; Nowakowski et Tinoco, 1999). Une critique additionnelle du modèle de Dinman et ses collègues est qu'ils ont analysé la structure du signal stimulateur par des attaques avec des agents enzymatiques en utilisant un court fragment ARN comprenant seulement la tige-boucle classique suivie d'une dizaine de nucléotides. La présence de la région précédant cette tige-boucle aurait pu modifier leurs résultats.

Le signal stimulateur que nous avons proposé a été déterminé à partir d'une séquence du sous-type B du groupe M du VIH-1. Nous avons donc voulu déterminer si une telle structure pouvait être formée chez les autres sous-types du groupe M du VIH-1 (sous-types A, C, D, F, G, H, J et K), malgré les variations de séquences entre les sous-types. Des

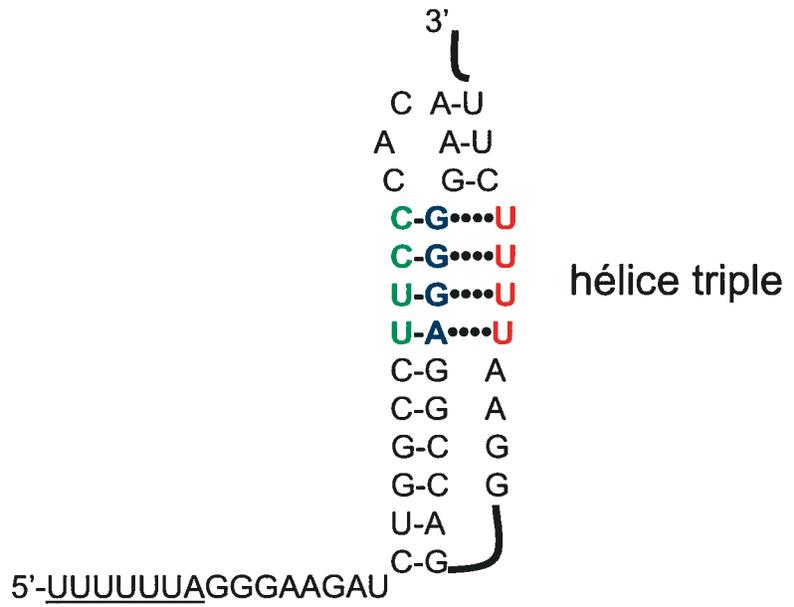
travaux publiés par notre équipe, auxquels j'ai participé en tant que co-auteur, montrent que le signal stimulateur du déphasage du VIH-1 peut adopter dans ces sous-types une structure semblable à celle du sous-type B (Baril et al., 2003a). Comme présenté à la Figure 5-2, la variation des séquences dans la région du déphasage chez ces sous-types n'empêche pas la formation de la tige inférieure. On retrouve dans la tige inférieure un appariement A-C chez les sous-types C et D et un mésappariement U-U chez le sous-type K(a). Cependant, comme indiqué plus haut, on sait que les appariements A-C peuvent être rencontrés dans les hélices d'ARN (Nowakowski et Tinoco, 1999). Nous avons mesuré l'efficacité du déphasage pour les différents sous-types avec un gène rapporteur luciférase dont l'expression dépend du déphasage en  $-1$  du VIH-1. Ces mesures, faites dans les cellules de mammifères en culture, ont montré que l'efficacité du déphasage des sous-types C et D était semblable à celle du sous-type B. Quant au sous-type K(a) qui contient un mésappariement U-U, il s'agit du sous-type où l'efficacité du déphasage a la valeur la plus faible (65%) par rapport au sous-type B. Si on tente, par contre, de replier les séquences des différents sous-types selon le modèle du signal stimulateur en hélice triple proposé par Dinman et al., les sous-types H et J possèderaient deux mésappariements sur trois dans la tige 2 du pseudonoeud proposé, tandis que les sous-types A et K(a) auraient un mésappariement dans cette tige, ce qui devrait se refléter par une diminution correspondante de l'efficacité du déphasage. Nos résultats expérimentaux montrent que ce n'est pas le cas, et l'analyse des différents sous-types appuie donc également notre modèle de structure du signal stimulateur. Il est intéressant d'ajouter ici que le déphasage du VIH-1 peut être reproduit dans un système bactérien où l'expression d'un gène rapporteur luciférase dépend d'un déphasage en  $-1$ , suite à l'insertion de la région de déphasage du VIH-1 en amont de ce gène rapporteur. Des études de mutagenèse dirigée sur le signal stimulateur du déphasage du VIH-1, ont confirmé avec ce système la structure que nous avons observée avec le système eucaryote et indiquent que le ribosome bactérien réagit de la même façon que le ribosome eucaryote au signal stimulateur du VIH-1 (Léger et al., 2004).

Tel que mentionné dans l'Introduction (voir la section 1.3.4.2.2 sur les structures secondaires stimulatrices), Staple et Butcher (2003) ont déterminé la structure à haute résolution d'une portion de la tige supérieure du signal stimulateur du VIH-1 par RMN. Tout récemment, pendant que nous complétons la rédaction de cette thèse, deux études ont décrit la structure du signal stimulateur du déphasage du VIH-1 à haute résolution par la technique de RMN. Les groupes de Butcher et de Fourmy ont déterminé par RMN la structure du signal stimulateur à partir d'un fragment d'ARN de la région de déphasage incluant la région riche en pyrimidines suivant la tige-boucle (Gaudin et al., 2005; Staple et Butcher, 2005). Les résultats montrent que le signal stimulateur du déphasage du VIH-1 adopte une structure à deux hélices interrompues par une protubérance de trois purines (voir Figure 5-3). La présence de la protubérance entraîne une courbure entre les deux tiges, caractéristiques des pseudonoeuds favorisant le déphasage (voir section 1.3.4.2.2). Ces données de RMN sont donc parfaitement en accord avec la structure du signal stimulateur contrôlant le déphasage du VIH-1 que nous avons déterminée.

Comme nous l'avons mentionné au cours de l'Introduction, l'analyse des variations des séquences virales a permis de diviser ces séquences en trois groupes : le groupe M, le groupe N et le groupe O. L'analyse de la séquence de la région de déphasage du groupe O du VIH-1 a montré que ce signal stimulateur ne pouvait adopter la structure en hélice irrégulière trouvée dans le groupe M. Des travaux de notre équipe, auxquels j'ai participé comme co-auteur ont montré que le signal stimulateur du déphasage du groupe O est un pseudonoeud (Baril et al., 2003b). Quant au troisième groupe, le groupe N, qui est très peu représenté, l'analyse des séquences virales suggère que le signal stimulateur du déphasage pour ce groupe adopte une conformation en hélice irrégulière semblable à celle que nous avons trouvée pour le groupe M.

Figure 5-1: Hélices triples. A. Structure secondaire du signal stimulateur du déphasage du VIH-1 tel que proposé par Dinman et al. (2002). B. Caractéristique des hélices triples impliquant un brin riche en pyrimidines. L'orientation antiparallèle des brins riches en pyrimidines est défavorable (Y est une pyrimidine, R une purine et Y' est le brin formant l'hélice triple).

**A**



**B**

Orientation des brins pyrimidines de l'hélice triple

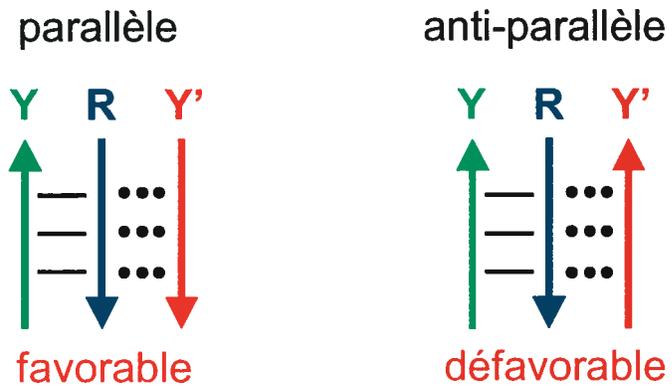


Figure 5-2: Structures secondaires du signal stimulateur des sous-types du groupe M du VIH-1. Le signal stimulateur est replié selon le modèle identifié chez le sous-type B. Les cercles indiquent les nucléotides qui diffèrent de la séquence du sous-type B, et les flèches indiquent les appariements non-canoniques et les mésappariements (adapté de Baril et al., 2003a).

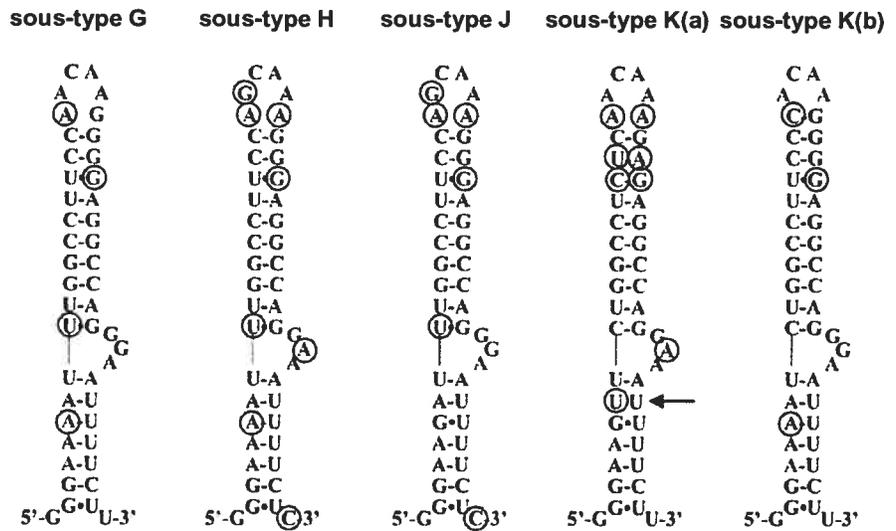
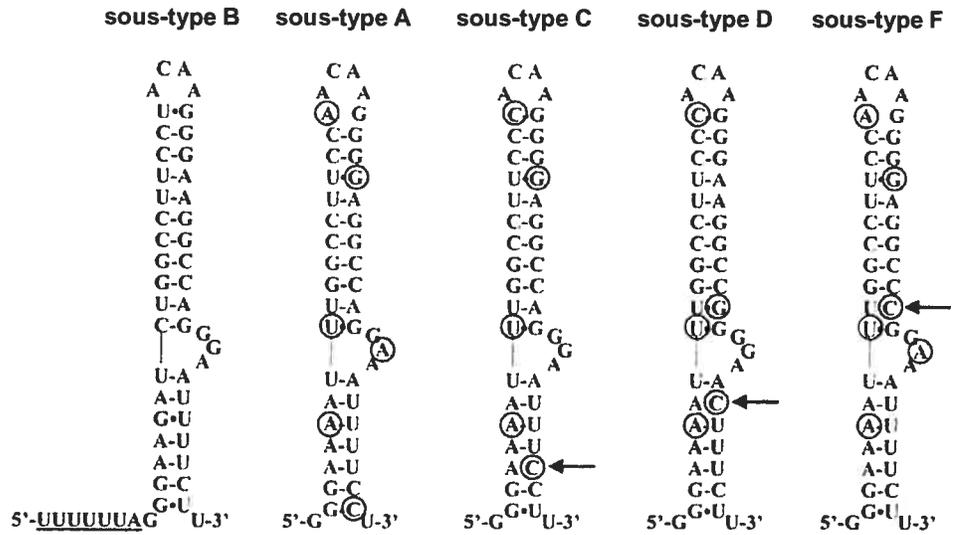
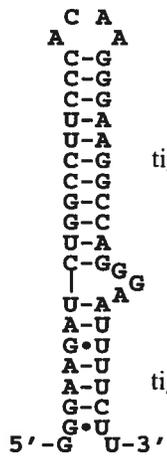


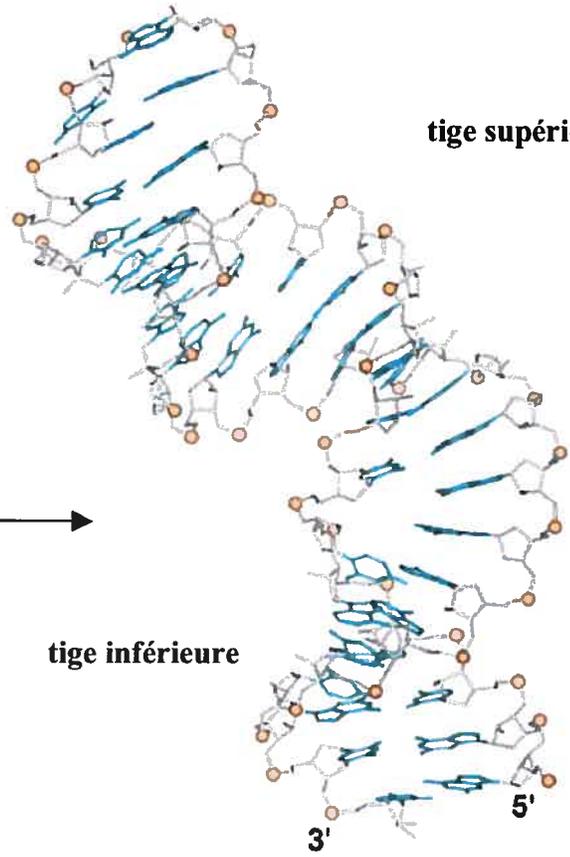
Figure 5-3: Structure à haute résolution du signal stimulateur du VIH-1. La conformation du signal stimulateur tel que proposé par Dulude et al. (2002) (à gauche) a été confirmée par RMN (à droite) par le groupe de Butcher et le groupe de Fourmy (Gaudin et al., 2005; Staple et Butcher, 2005) (adapté de Gaudin et al., 2005).

signal  
stimulateur



tige supérieure

tige inférieure



## 5.2 Des mutations dans les signaux contrôlant l'efficacité du déphasage en $-1$ interfèrent avec la réplication du VIH-1.

Les signaux stimulateurs jouent un rôle de premier plan dans l'efficacité avec laquelle se produit le déphasage en  $-1$ . Il a déjà été observé que le maintien de l'efficacité du déphasage est important pour la réplication virale (section 1.3.5). Par exemple, les recherches du groupe de Mak sur le VIH-1, ont montré qu'une augmentation du rapport Gag-Pol sur Gag d'un facteur trois et plus diminuait drastiquement la réplication virale. Ce défaut de réplication est causé par une déstabilisation du dimère d'ARN génomique dans les virions (Shehu-Xhilaga et al., 2001). Pour ces études, le rapport Gag-Pol sur Gag était défini par une transfection de cellules en culture avec des quantités différentes de deux vecteurs, l'un contenant le génome proviral et l'autre, le précurseur Gag-Pol du VIH-1. Par ailleurs, une étude de Telenti et al. (2002) portant sur l'analyse d'isolats cliniques et de mutants de laboratoire du VIH-1 a conduit à la conclusion que seuls les virus dont l'efficacité de déphasage en  $-1$  était moins de 40% de la valeur sauvage avaient un défaut de réplication. Dans ce cas, le défaut de réplication résultait d'un défaut dans la maturation des précurseurs protéiques du virus.

Dans le deuxième volet de notre étude, nous avons décidé d'examiner l'effet de mutations dans les signaux de déphasage du VIH-1, c'est à dire dans la séquence glissante et le signal stimulateur, sur la réplication du virus. Notre hypothèse est que des mutations introduites dans la séquence glissante peuvent influencer l'efficacité du déphasage en modifiant la dynamique de l'interaction codon-anticodon, tandis que les mutations introduites dans le signal stimulateur perturbent la formation de cette structure ou la déstabilisent, l'empêchant d'exercer son rôle. Toutes les mutations que nous avons introduites (Figure 3-1) ont diminué l'efficacité du déphasage en  $-1$ , les valeurs d'efficacité allant de 0.3 et 70% de la valeur obtenue avec la séquence sauvage (Figure 3-2B). Pour la mesure de l'efficacité du déphasage, nous avons utilisé un système rapporteur *dual-*

luciférase. Dans cette construction, la région de déphasage du VIH-1 est insérée entre la séquence du gène rapporteur codant pour la luciférase de la *Renilla*, en amont, et celle du gène de la luciférase de la luciole (*firefly*), en aval. L'expression de la luciférase de la *Renilla* dépend de la traduction conventionnelle tandis que l'expression de la luciférase de la luciole dépend d'un déphasage en  $-1$ , au site de glissement du VIH-1, par des ribosomes ayant traduit la luciférase de la *Renilla*. Le déphasage permet donc la synthèse d'une protéine de fusion entre la luciférase de la *Renilla* et la luciférase de la luciole. Nous avons ensuite examiné le contenu en RT de pseudo-particules virales (VLP) obtenues par transfection de cellules en culture par un vecteur codant pour Gag et Gag-Pol, contenant les différentes mutations étudiées. Nous avons observé que le contenu en RT des VLP diminue parallèlement à la diminution de l'efficacité du déphasage (Figure 3-3), ce qui était prévisible puisque Gag-Pol, qui code pour les enzymes virales, requiert ce déphasage pour être exprimée. Finalement, l'effet de ces mutations dans la séquence glissante et le signal stimulateur sur l'infectivité du VIH-1 a été évalué. Nous avons utilisé des essais mesurant soit un seul cycle de réplication virale (*single-round assays*) (Figure 3-5) soit la capacité de réplication du virus à long terme (Figure 3-6). Nous avons observé que la réplication du virus dans ces essais varie proportionnellement à la valeur de l'efficacité du déphasage en  $-1$ . Plus l'efficacité du déphasage était diminuée, plus la réplication virale était affectée. Seul un mutant, M5, avait un niveau de réplication beaucoup plus bas que ne le laissait prévoir la diminution de son efficacité de déphasage. Cette étude a montré que même de légères variations (comme une diminution d'environ 30%) dans l'efficacité du déphasage réduisent l'infectivité et la réplication virale.

Une récente étude par Biswas et al. (2004) a suggéré que la séquence glissante du VIH-1 est invariable. Les chercheurs ont en effet remplacé la séquence glissante du virus par d'autres séquences glissantes correspondant à la composition X XXY YYZ (voir la section 1.3.4.2.1), et ils ont observé que ces remplacements réduisaient l'efficacité du déphasage à des valeurs comprises entre 0 et 23% de la valeur obtenue avec le virus sauvage, en utilisant un système de *dual*-luciférase semblable à celui que nous avons

utilisé. Biswas et al. ont aussi montré que les virions possédant ces mutations n'étaient pas infectieux. Les mutations que nous avons introduites dans la séquence glissante du VIH-1 font varier l'efficacité du déphasage dans un plus large spectre de valeurs, allant de 0.3 à 60% de la valeur de la séquence sauvage. L'effet des mutations de la séquence glissante, en accord avec les observations de Biswas et al, montre une proportionnalité directe entre l'infectivité et l'efficacité du déphasage. Quant aux mutations dans le signal stimulateur, l'analyse de nos résultats montre l'importance de l'intégrité de ce signal pour l'infectivité du virus. Une faible diminution de 30% de l'efficacité du déphasage, causée par une mutation de la portion inférieure du signal stimulateur (mutant M6), réduit l'infectivité du virus à environ 50% de l'infectivité du virus sauvage dans les essais mesurant un cycle de réplication et l'infectivité à long terme. Nos résultats diffèrent donc de ceux de Telenti et al. (2002), présentés plus haut, qui ont observé que des mutations dans le signal stimulateur diminuant l'efficacité du déphasage à 60% de la valeur du virus sauvage n'entraînaient aucun effet sur la réplication du virus et qu'il fallait une réduction de l'efficacité au moins à 40% pour que les virus deviennent moins infectieux. Cette contradiction entre nos résultats et ceux de Telenti et al. ne peut être expliquée actuellement. Telenti et al. ont mesuré l'efficacité du déphasage avec un système de rapporteurs exprimés dans la levure alors que nos rapporteurs sont exprimés dans des cellules humaines en culture. Le système avec la levure a peut-être surévalué l'amplitude des changements de l'efficacité du déphasage correspondant aux mutants infectieux étudiés par Telenti et al.

Il y a plusieurs raisons qui peuvent être avancées pour expliquer comment la diminution du rapport Gag-Pol sur Gag peut affecter la réplication virale. D'abord, nos résultats indiquent une diminution de l'incorporation de la RT dans les virions correspondant à la diminution de l'efficacité du déphasage (Figure 3-3). Ceci montre que moins d'enzymes virales ont été produites et incorporées dans les virions, ce qui suggère un défaut possible dans les processus catalysés par ces enzymes. Nous avons observé une faible diminution de l'efficacité du clivage du précurseur Gag par la protéase dans des virions produits par la transfection de cellules en culture avec l'ADN proviral des mutant

dont l'efficacité de déphasage était diminuée jusqu'à 25% de la valeur observée avec la séquence sauvage (Figure 3-4). Cependant, pour des diminutions amenant le déphasage à moins de 25% de la valeur sauvage, la maturation de Gag est fortement affectée. Les mêmes observations ont été faites en analysant la maturation de Gag dans les VLP produites par transfection de cellules en culture avec un vecteur codant pour Gag et Gag-Pol contenant les différentes mutations étudiées (résultats non montrés). Ces résultats sont en accord avec des observations précédentes de Rosé et al. (1995) montrant qu'une diminution de l'activité de la protéase du VIH-1 d'un facteur quatre diminue faiblement l'infectivité du virus, tandis que des réductions d'activité plus grandes interfèrent avec l'infectivité du virus, suggérant qu'il y a un niveau critique d'activité protéasique indispensable pour le virus. Nous pouvons également supposer qu'une réduction d'activité de la RT et d'IN contribue aussi à diminuer l'infectivité des virus. Même si, individuellement, l'effet de la réduction de la quantité de chaque enzyme peut être faible, la combinaison des effets provoqués par la réduction simultanée de la quantité des trois enzymes peut vraisemblablement expliquer en grande partie la diminution d'infectivité des virions mutants. D'autres problèmes peuvent aussi handicaper la réplication virale lorsque la production de Gag-Pol est réduite. Il a été observé que la surexpression ou la sous-expression de Gag-Pol influence le processus de maturation stabilisant le dimère de l'ARN génomique dans les virions (Shehu-Xhilaga et al., 2001; Shehu-Xhilaga et al., 2002a). Il est possible que même de faibles variations du rapport Gag-Pol sur Gag entraînent un défaut de maturation du dimère, ce qui expliquerait la diminution de l'infectivité pour ces mutants, qui produisent probablement suffisamment d'enzymes pour assurer leur réplication. Aussi, Gag-Pol étant impliquée dans l'incorporation de l'ARN<sub>t</sub><sup>Lys</sup> qui sert d'amorce à la RT (Mak et al., 1994), et nous pouvons suggérer qu'une diminution de la quantité de Gag-Pol dans les virions perturbe l'incorporation de cet ARN<sub>t</sub><sup>Lys</sup> et, donc, la transcription inverse.

Il faut ajouter que la séquence glissante du VIH-1 fait partie du site de clivage NC/p1 de la protéase virale. En plus d'influencer le déphasage, des mutations dans la séquence glissante peuvent interférer avec le clivage par la protéase virale à cet endroit. Le

clivage au site NC/p1 se fait avec une cinétique beaucoup plus lente que le premier clivage effectué par la protéase, qui se produit au site p2/NC. Il a récemment été montré que la conformation ou le repliement de p1 est important pour stabiliser le dimère d'ARN génomique, car la mutation de deux résidus proline du peptide p1 déstabilise ce dimère et réduit l'infectivité du virus (Hill et al., 2002). Dans notre étude, un des mutant, M5, est beaucoup moins infectieux que des mutants ayant une efficacité de déphasage semblable. Ce mutant contient une substitution d'une arginine par une lysine au site de clivage NC/p1. Petit et al. (2002), qui ont analysé l'effet *in vitro* de mutations dans les sites de clivages de la protéase virale, ont montré que la présence d'une lysine à ce site empêchait le clivage à cet endroit. Nous pouvons suggérer que, dans le cas de M5, la faible infectivité est causée non seulement par la diminution du déphasage mais aussi par un défaut lors du clivage de Gag libérant p1.

Nous avons indiqué que la structure du signal stimulateur du déphasage du VIH-1 a été récemment résolue à haute résolution par RMN. La caractérisation de cette structure nous permet d'examiner l'impact des mutations introduites dans la région de déphasage sur cette structure. Les mutants M6, M7 et M8 contiennent des mutations dans le signal stimulateur tandis que les autres mutations affectent la séquence glissante. Les mutants M6 et M7 ont pour effet de déstabiliser le signal stimulateur en introduisant des mésappariements dans l'hélice inférieure et supérieure, respectivement. L'analyse du repliement de la structure secondaire du signal stimulateur à l'aide du programme informatique *mfold* (Mathews, et al. 1999; Zucker, 2003) nous indique que seule la tige supérieure peut se former dans le cas du mutant M6, alors que la structure secondaire complète du signal est défaite pour le mutant M7 (résultats non-montrés). Pour M6, des analyses de minimisation d'énergie (par le programme Insight II/Discover package version 2000 Molecular Simulations Inc., San Diego California) ne nous ont révélé aucun changement dans la conformation de la tige supérieure comparativement à la structure sauvage. Dans le mutant M8, les mutations introduites avaient pour but de stabiliser la partie supérieure du signal stimulateur. Deux paires U-A et une paire U-G y sont

remplacées par des paires C-G, dans la tige supérieure. L'analyse de la structure secondaire de M8 par *mfold* montre un repliement semblable à celui observé avec la séquence sauvage, mais avec une augmentation de la stabilité thermodynamique ( $\Delta G^\circ$ ) d'environ 4 Kcal/mol. L'analyse par minimisation d'énergie n'a révélé aucune différence (résultats de prédiction par Insight II non-montrés). Il faut noter ici que le mutant M8 a une efficacité de déphasage diminuée par rapport à la séquence sauvage, mais cet effet résulte probablement du fait que, outre les mutations dans le signal stimulateur, M8 est aussi muté dans la séquence glissante. Nous pouvons aussi mentionner que la structure du signal stimulateur à haute résolution rapportée par le groupe de Fourmy est basée sur l'utilisation d'une séquence contenant une paire U-G terminant la tige supérieure, alors que, pour Stapple et Butcher (2005), c'est une paire C-G qui termine la tige supérieure. La superposition, à l'aide d'Insight II, des deux structures obtenues par RMN montre un léger déplacement de la troisième base de la tétraboucle, une adénine, vers l'extérieur quand U-G est remplacé par C-G (résultat non montré). Toutefois, il faut demeurer critique devant cette différence, car elle peut également provenir de différences dans les données de RMN utilisées par les deux groupes. Il faut ajouter, dans le même ordre d'idée, que d'autres différences mineures sont détectables quand on compare les structures déterminées par les deux groupes, comme, par exemple, une légère différence dans l'angle entre les deux hélices du signal ainsi que dans la conformation de la protubérance de purines qui sépare les deux hélices.

Quand aux changements introduits dans la séquence glissante, les résultats sont à première vue étonnants, puisqu'on aurait pu croire que le virus accepterait que sa séquence glissante soit remplacée par d'autres séquences glissantes. Mais nos résultats montrent que l'efficacité du déphasage en  $-1$  produite par une combinaison de séquence glissante et de signal stimulateur est imprévisible. La combinaison du signal stimulateur du VIH-1 et de la séquence glissante AAAAAAC, une séquence glissante utilisée par MMTV et EIAV, diminue l'efficacité du déphasage en  $-1$  d'environ trois fois comparée à la séquence glissante sauvage du VIH-1. Ceci est en accord avec les résultats de Biswas et al. (2004), qui ont remplacé la séquence glissante du VIH-1 par différentes séquences glissantes

naturelles, et aussi montré que ces séquences, dont AAAAAAC, diminuaient grandement l'efficacité du déphasage. On pourrait en conclure que la séquence glissante du VIH-1 (UUUUUA), est une séquence qui favorise le glissement du ribosome de façon plus efficace que d'autres séquences. Cependant, il a aussi été observé que le remplacement de la séquence AAAAAAC chez EIAV, dont le signal stimulateur est un pseudonoeud, par la séquence glissante du VIH-1, diminuait l'efficacité du déphasage en  $-1$  d'environ deux fois (Chen et Montelaro, 2003). Paradoxalement, des études avec MMTV ont montré que le remplacement de la séquence glissante AAAAAAC par la séquence glissante du VIH-1 en présence du pseudonoeud de MMTV n'affectait pas l'efficacité du déphasage en  $-1$  (Chamorro et al., 1992). L'ensemble de ces résultats suggère que l'efficacité du déphasage pour une région de déphasage donnée est le résultat d'une coopération complexe entre la séquence glissante et le signal stimulateur. Cette coopération pourrait être causée par des facteurs additionnels agissant en *trans*. Il a déjà été suggéré que le déphasage puisse être influencé par des protéines cellulaires ou virales liant les signaux stimulateurs (revue dans Brierley et Pennell, 2001). Cependant, aucune de ces protéines n'a été identifiée à ce jour, que ce soit par des expériences de titration avec un excès de signal stimulateur, par des essais de retard de migration sur gel ou par des tentatives de pontage aux rayons UV. Nous avons aussi mentionné dans l'Introduction de cette thèse que le pseudonoeud de MMTV doit adopter une conformation avec une courbure particulière pour agir efficacement et qu'il y a des bases du signal stimulateur de BWYV, qui sont importantes pour l'efficacité du déphasage, mais qui ne sont pas impliquées dans des interactions tertiaires (section 1.3.4.2.2). Ces observations sont compatibles avec l'existence de facteurs protéiques additionnels qui reconnaîtraient ces structures et influenceraient le phénomène de déphasage.

Dans cette étude sur la contribution de la séquence glissante et du signal stimulateur à la réplication du VIH-1, nous avons donc, comme expliqué plus haut, utilisé un vecteur rapporteur *dual*-luciférase pour mesurer l'efficacité du déphasage. Cette construction contient la région de déphasage du VIH-1 introduite entre les séquences codantes de deux

gènes de luciférase. Le premier rapporteur y sert de contrôle pour évaluer la traduction conventionnelle et le second mesure le déphasage. Avec une telle construction, nous avons mesuré une efficacité de déphasage en  $-1$  d'environ 9% pour le VIH-1. Cette valeur est le double de l'efficacité de déphasage d'environ 5% que nous avons obtenue avec une construction de type monocistronique où l'expression d'un gène rapporteur dépend d'un déphasage en  $-1$ . Nous ne pouvons que spéculer pour expliquer la différence entre le système *dual*-luciférase et le système avec un rapporteur monocistronique. Pour la construction monocistronique, les ribosomes initient la traduction seulement 18 nucléotides avant la séquence glissante tan dis qu'ils doivent traduire la région codante complète du gène rapporteur de la luciférase de la *Renilla* dans le cas de la construction *dual*-luciférase avant d'atteindre la séquence glissante. Il est possible que le nombre de ribosomes décodant la région de déphasage soit plus élevé dans le cas de la construction de type monocistronique. Ceci pourrait interférer avec le repliement du signal stimulateur et diminuer l'efficacité du déphasage, le signal stimulateur étant déroulé lors du passage de chaque ribosome. Le contexte de la construction *dual*-luciférase est au fait plus près du phénomène de la traduction de l'ARN viral, où les ribosomes traduisent une grande partie de la région correspondant au gène *gag* avant d'atteindre la région de déphasage.

Toutes les mutations que nous avons introduites dans la région de déphasage du VIH-1 influencent la réplication du virus. Nous pouvons nous demander si l'introduction de ces mutations pourraient aussi déstabiliser et influencer l'intégrité des ARN traduits, par exemple, en créant des sites d'attaque par des nucléases. Toutefois, nous observons une très bonne corrélation (à l'exception du mutant M5, pour lequel nous avons fourni une explication directe plus haut) entre l'efficacité du déphasage, d'une part, et la quantité de RT incorporé dans les VLP et l'infectivité, d'autre part. Cette corrélation n'appuie donc pas l'hypothèse d'une déstabilisation de l'ARN viral. De plus, l'utilisation d'un système *dual*-luciférase pour la mesure du déphasage nous permet d'évaluer s'il y a des changements dans l'abondance d'ARNm contenant la région de déphasage du VIH-1 mutée, et nous

n'avons observé de tels changements pour aucun des mutants étudiés (résultats non-montrés).

### **5.3 Mise au point d'un système *in vivo* de sélection de peptides interférant avec l'efficacité du déphasage en -1 du VIH-1.**

Les traitements actuels contre la réplication du VIH-1 permettent de transformer le SIDA en une maladie qui peut être traitée et d'augmenter la qualité de vie des patients (section 1.1.3.3). Les médicaments utilisés actuellement ciblent soit la RT, la PR ou l'étape d'entrée du virus dans la cellule, et des composés en développement ciblant l'intégration font l'objet d'essais cliniques. Toutefois, le traitement des patients durant de longues périodes entraîne des complications dues à la toxicité des médicaments, et l'émergence de virus résistants contrecarre l'efficacité des traitements (revue par Pomerantz et Horn, 2003). L'identification de nouvelles cibles virales est donc indispensable. Nous avons démontré que le maintien de l'efficacité du déphasage en -1 est critique pour la réplication du VIH-1 (Chapitre 3). En effet, nous avons observé que même une faible diminution de l'efficacité du déphasage réduit l'infectivité et la réplication virale.

Notre équipe a montré que des oligonucléotides ciblant le ribosome et se liant à l'ARNr 18S de la petite sous-unité peuvent réduire efficacement l'efficacité du déphasage du VIH-1 et la réplication du virus dans des cellules en culture (Brakier-Gingras et al., 1998). L'avantage de cibler un composant cellulaire comme le ribosome permet d'éviter les problèmes de résistance associés aux variations de la cible. L'ARNr du ribosome étant codé par un gène présent en multiples copies, il ne peut y avoir de résistance suite à une mutation dans l'ARNr, l'effet d'une mutation dans une copie étant occulté par la présence des autres copies non mutées. Cependant, cibler une molécule aussi importante que le ribosome présente le risque de déclencher de multiples effets secondaires. Par exemple, Hung et al. (1998) avait identifié un composé interférant avec le déphasage en -1 du VIH-1, mais ils

ont observé que ce composé était toxique dans les cellules de mammifères, probablement parce qu'il interagit avec le ribosome. L'identification d'un gène cellulaire (*edr*) utilisant le déphasage et la possibilité qu'il y ait plusieurs autres cas de déphasage, d'après des analyses informatiques de séquences génomiques, suggèrent que le mécanisme de déphasage en  $-1$  pourrait être requis pour la synthèse de plusieurs gènes cellulaires chez l'homme (Bekaert et al., 2003; Hammell et al., 1999; Namy et al., 2004). Il s'ensuit que des composés affectant la capacité des ribosomes à effectuer un déphasage pourraient également interférer avec l'expression de ces gènes. Nous en concluons qu'il est plus approprié de cibler l'ARN du virus pour empêcher le déphasage traductionnel.

Nous avons souligné que le processus de déphasage est contrôlé par un signal stimulateur, qui interagirait spécifiquement avec le ribosome, et il est donc envisageable de développer des composés interférant avec cette interaction. Une étude intéressante par Kollmus et al. (1996) a montré que le déphasage du VIH-1 pouvait être contrôlé par des protéines liant l'ARN. Ces chercheurs ont remplacé le signal stimulateur du VIH-1 par une structure en tige-boucle, l'élément de réponse au fer du messager de la ferritine. Des protéines régulatrices peuvent se lier à cette tige-boucle dépendamment de la concentration du fer dans le milieu. Kollmus et al. ont montré que l'efficacité du déphasage était augmentée dans les conditions où les protéines régulatrices se lient à la structure en tige-boucle. Dans le troisième volet de notre projet, nous avons entamé la mise au point d'un système de sélection permettant l'identification de peptides perturbant l'efficacité du déphasage en  $-1$ , probablement en se liant au signal stimulateur du déphasage du VIH-1. Pour ce faire, nous avons donc utilisé une librairie combinatoire et un système d'expression bactérien. Nous avons décidé d'utiliser un système bactérien vu sa simplicité d'utilisation, sachant que le ribosome bactérien répond exactement comme le ribosome eucaryote au signal stimulateur du déphasage du VIH-1. Nous avons construit un vecteur rapporteur codant pour la protéine fluorescente verte (GFP) et pour la protéine fluorescente rouge (RFP), et un second vecteur permettant l'expression d'une librairie combinatoire de peptides. Ces peptides contiennent 14 acides aminés, le choix des résidus d'acide aminé

étant restreint à la sérine, l'arginine, l'asparagine et l'histidine, avec une majorité d'arginine. Le choix de cette composition s'inspirait d'un travail par le groupe de Frankel, qui a utilisé ce type de librairie pour sélectionner des peptides se liant à des structures secondaires d'ARN, comme les structures de TAR et RRE du VIH-1 (Harada et al., 1996). La longueur des peptides était choisie d'après la taille du domaine de la protéine Tat du VIH-1 permettant la liaison de cette protéine à la structure TAR de l'ARN viral. Les peptides sont fusionnés au site actif de la thiorédoxine, ce qui leur impose une contrainte favorisant leur capacité d'interagir avec une cible. Avec notre vecteur rapporteur, p3RFP-GFPfs, l'expression de la GFP dépend d'un déphasage en  $-1$  suite à l'introduction de la région de déphasage du VIH-1 au début de la séquence codante de la GFP, tandis que l'expression de la RFP dépend de la traduction conventionnelle et sert à normaliser l'efficacité du déphasage. Des changements de l'efficacité du déphasage se reflètent donc par une variation du niveau de fluorescence émise par la GFP. Le système a été construit de façon à mesurer les variations de fluorescence de la GFP et de la RFP par un appareil de type FACS (*fluorescence-activated cell sorter*) suite à l'introduction des deux vecteurs dans une souche bactérienne. Cet appareil permet de trier les cellules en fonction de la fluorescence qu'elles émettent.

Nos résultats préliminaires ont montré que l'introduction du vecteur rapporteur p3RFP-GFPfs dans les bactéries permet aisément de détecter par FACS les bactéries avec une émission de fluorescence (Figure 3-2G). Ces observations indiquaient que des changements dans l'efficacité du déphasage en  $-1$  causés par la présence de peptides sont détectables. Nous avons construit une librairie combinatoire de peptides ayant une complexité d'environ  $5 \times 10^6$  peptides, ce qui représente environ 2% du total possible ( $2.6 \times 10^8$ ). Il est donc peu probable que des peptides ayant une séquence identique se retrouvent dans cette librairie. Lorsque nous avons introduit la librairie de peptides dans les bactéries transformées avec le vecteur rapporteur p3RFP-GFPfs, l'analyse par FACS nous a révélé la présence de clones bactériens contenant des candidats potentiels. En effet, dans certaines bactéries, il y avait des variations dans l'expression de la GFP, par rapport à la

RFP, qui reflètent une augmentation ou une diminution de l'efficacité du déphasage (Figure 3-3). Nous avons procédé à la sélection de ces bactéries en délimitant des paramètres de fluorescence établissant des frontières, qui correspondent à une diminution ou une augmentation de la fluorescence de la GFP pour lesquelles le niveau de fluorescence de la RFP semblait normal. Nous avons ensuite procédé la sélection par FACS de ces clones correspondant à ces critères, et nous avons récupéré par triage deux populations de clones comprenant environ 240 clones dont l'expression de la GFP était diminuée et environ 40 clones où elle était augmentée. Après avoir remis ces populations de clones en cultures et effectué deux nouvelles rondes de sélections pour enrichir ces populations, nous avons observé que les clones sélectionnés pour l'augmentation du déphasage étaient de faux positifs. Quant à l'examen de plusieurs clones individuels parmi ceux diminuant le déphasage, il a révélé que plusieurs des clones sélectionnés étaient aussi des faux positifs. Toutefois, il y avait aussi plusieurs cas où l'expression des peptides avait un effet mais cet effet consistaient en une diminution de l'efficacité de la traduction en général, se manifestant par une diminution de la fluorescence émise et par la GFP et par la RFP de deux à trois fois. Le déphasage en -1 n'était donc pas affecté spécifiquement et il est fort probable que cet effet résultait d'une interaction des peptides avec le ribosome. Les résultats obtenus dans ce troisième volet de notre recherche, quoique préliminaires, soulignent les limites du système de sélection que nous avons utilisé et indiquent qu'il nous faut introduire des modifications dans notre approche visant le développement d'agents perturbant le déphasage et, par conséquent, la réplication du VIH-1. Nous proposons d'appliquer notre méthode de sélection à un système eucaryote tel que décrit plus bas. Nous pourrions ainsi tirer avantage du fait que le FACS a une meilleure capacité de discriminer des cellules eucaryotes que des bactéries. Nous envisageons aussi la possibilité de modifier la longueur et le contenu en acides aminés de la librairie de peptides afin d'augmenter la probabilité d'identifier des candidats potentiels. Quoiqu'il en soit, la possibilité qu'un grand nombre de ces peptides puissent interagir avec le ribosome ne peut être ignorée et il faudrait trouver un moyen d'appauvrir le contenu de ces peptides dans la

librairie, ce qui augmentera la probabilité de sélectionner les peptides agissant sur le signal stimulateur.

Pour la suite de ce projet, si nous parvenons à optimiser notre approche et sélectionnons des peptides capables de perturber spécifiquement le déphasage, le séquençage de l'ADN des plasmides exprimant les peptides de ces bactéries permettra de connaître leur séquence en acide aminé. Une fois la séquence du peptide connue, nous vérifierons l'effet des peptides sur l'efficacité du déphasage en  $-1$  dans un système eucaryote en exprimant les peptides fusionnés à la thiorédoxine dans des cellules en cultures contenant un rapporteur *dual-luciférase*, tel qu'utilisé dans nos études précédentes. Cette analyse pourra confirmer si les peptides sélectionnés agissent sur le déphasage dans un système eucaryote. Les peptides qui influencent le déphasage par un facteur deux à trois environ seront considérés des candidats intéressants. En effet, les travaux de Shehu-Xhilaga et al. (2001) ont montré qu'une augmentation de l'efficacité du déphasage d'un facteur trois réduisait fortement l'infectivité virale (voir section 1.3.5). D'autre part, nos travaux ont montré qu'une diminution de l'efficacité du déphasage de 30% ralentissait significativement la réplication virale et une diminution d'environ 75% conduisait à la production de virions non-infectieux (Chapitre 3). L'effet des peptides qui perturbent le déphasage sera ensuite étudié sur la réplication du VIH-1, en examinant la réplication du virus dans des cellules en culture en leur présence ou en leur absence. En parallèle, nous allons vérifier la cible des peptides, notre hypothèse étant qu'ils interagissent avec le signal stimulateur du déphasage. Ceci sera examiné par des études de migration sur gel d'un court transcrit d'ARN correspondant au signal stimulateur en présence de chacun de ces peptides. S'il y a interaction entre un peptide et le transcrit, elle devrait causer un retard dans la migration du transcrit. Si les peptides sélectionnés influencent donc le déphasage en interagissant avec le signal stimulateur, nous pouvons nous interroger sur leur mécanisme d'action. Différentes situations peuvent être rencontrées. Si l'interaction avec le peptide diminue l'efficacité du déphasage, nous pouvons supposer qu'elle bloque un site reconnu par le ribosome. Si par contre, elle augmente l'efficacité du déphasage, nous pouvons

supposer qu'elle pourrait stabiliser la structure du signal, avec peut-être une augmentation de la pause et une augmentation de la probabilité de déphasage. Nous pouvons aussi nous demander si cette interaction entre un peptide et le signal stimulateur pourrait interférer avec la traduction conventionnelle, augmentant par exemple la possibilité de terminaison prématurée. Intuitivement, cela nous semble peu probable mais cette possibilité se prête facilement à une vérification expérimentale. Il suffirait en effet d'examiner si la présence d'un des peptides sélectionnés influence l'expression d'un gène rapporteur où la région de déphasage du VIH-1 a été introduite au début de la région codante mais de façon à ce que l'expression de ce gène ne nécessite pas de déphasage. Un cas connu de protéine liant une structure en tige-boucle d'un ARNm afin de contrôler la traduction est le cas du messenger de la ferritine, une protéine de stockage du fer (revue par Rouault et Klausner, 1996). L'interaction à haute affinité entre une tige-boucle particulière, l'élément de réponse au fer (IRE), situé près de la coiffe du messenger, et la protéine cytoplasmique IRE-BP, permet de bloquer la liaison du complexe d'initiation 43S à ce messenger. Toutefois, lorsque la tige-boucle est éloignée de la coiffe, elle n'affecte nullement le scanning du messenger par le complexe 43S. Nous supposons que des peptides interagissant avec le signal stimulateur du déphasage n'empêcheraient pas non plus le mouvement normal des ribosomes en élongation. Par contre, le déphasage du cadre de lecture serait plus sensible à leur présence.

Si nous parvenons à sélectionner des peptides prometteurs, toute une avenue de recherche s'ouvre alors à nous. Ces peptides pourraient être soumis à des études de mutagenèse dirigée afin d'identifier quelles sont les positions-clés qui contrôlent l'effet de ces peptides. Ces peptides ont été obtenus à partir d'un screening de librairie combinatoire de peptides. Si leur cible est bien le signal stimulateur du déphasage du VIH-1, nous pouvons envisager maintenant de recourir aux techniques du design rationnel de médicaments pour obtenir des dérivés peptidiques plus performants dont l'affinité avec la cible est augmentée. Le design rationnel de médicaments est encore à ces débuts et il y a peu de drogues qui ont été développées directement *in silico* uniquement à partir de la connaissance de la structure tridimensionnelle d'une cible

(<http://www.wellcome.ac.uk/en/genome/tacklingdisease/hg09b002.html>). Par contre, la construction de dérivés plus puissants qu'un composé sélectionné initialement est une entreprise réalisable, en tenant compte de l'amplitude des moyens bioinformatiques mis actuellement à la disposition des chercheurs.

Parallèlement à notre approche utilisant un système bactérien, nous envisageons d'appliquer l'approche de sélection de peptides perturbant l'efficacité du déphasage du VIH-1 à un système de sélection eucaryote afin de multiplier le nombre de candidats possibles. Une approche développée par le groupe de Nolan, dans laquelle une librairie combinatoire de courts peptides est exprimée dans des cellules eucaryotes grâce à un vecteur de réplication rétroviral, pourrait être adaptée à notre étude, en conservant le biais utilisé pour la composition de la librairie (Xu et al., 2001). Il sera important de déterminer si les peptides agissant efficacement contre le sous-type B groupe M, que nous avons utilisé pour nos études, sont aussi efficaces contre les autres sous-types du groupe M. Il sera aussi intéressant plus tard d'effectuer la sélection de peptides affectant l'efficacité du déphasage du groupe O du VIH-1. Dans ce cas, les travaux de notre équipe auxquels j'ai participé ont montré que le signal stimulateur du déphasage, comme nous l'avons mentionné (voir section 1.3.4.2.2), est un pseudonoeud (Baril et al., 2003b) et nous prévoyons que les peptides sélectionnés différeront des peptides obtenus contre le groupe M.

## 5.4 Conclusions et perspectives

Les récents progrès dans les recherches sur le déphasage du ribosome, et surtout, la détermination de la structure du ribosome et celle de signaux stimulateurs ont permis d'avancer dans la compréhension des mécanismes contrôlant cet événement. L'ensemble de nos travaux a eu pour objectif de mieux caractériser le déphasage en -1 du VIH-1, d'abord en déterminant la structure du signal stimulateur responsable de son efficacité, et ensuite en examinant l'effet de changements de l'efficacité du déphasage sur la réplication du virus.

De plus, nous avons entamé une procédure visant à valider le signal stimulateur comme cible thérapeutique contre le VIH-1. Si nous parvenons à obtenir des peptides empêchant la réplication du virus, il faudra évidemment d'abord évaluer la toxicité de ces composés pour des cellules en culture, ensuite, la toxicité *in vivo*. Les peptides non-toxiques et perturbant efficacement le déphasage se qualifieront pour des études ultérieures. Une première étape de ces études consistera à examiner l'effet de ces peptides sur le déphasage avec un modèle animal approprié. Nous disposons déjà d'un tel modèle. En effet, une publication de notre équipe, à laquelle j'ai participé comme co-auteur, décrit la construction et la caractérisation d'un rétrovirus chimérique murin, MoMuLV, pour lequel le mécanisme de synthèse du précurseur Gag-Pol a été modifié pour dépendre du signal de déphasage en -1 du VIH-1 (Gendron et al., 2005). Il sera possible de vérifier si les peptides agissant sur le déphasage du VIH-1 interfèrent avec la réplication du rétrovirus MoMuLV chimérique dans des souris athymiques. Ultimement, l'identification de tels agents pourra compléter les régimes de drogues disponibles pour le traitement des personnes infectées par le VIH-1.

Beaucoup de recherches sont encore nécessaires pour comprendre exactement comment le déphasage se produit. La résolution de cette question passe nécessairement l'approfondissement de nos connaissances tant sur le ribosome que sur les signaux de déphasage qui perturbent sa fonction. De grands pas pourront être faits grâce à l'avancée des techniques de cristallographie développées pour l'étude du ribosome. Une récente étude du groupe de Yusupov a permis d'observer la structure cristallographique à haute résolution d'un ARNm, contenant une structure secondaire en tige-boucle, en complexe avec le ribosome bactérien (Jenner et al., 2005). Des images cristallographiques de ribosomes en élongation interagissant avec un signal stimulateur de déphasage pourraient nous fournir des informations détaillées sur les interactions du ribosome avec cette structure. L'approche cristallographique s'applique actuellement au ribosome bactérien, mais vu le haut degré de conservation entre le ribosome bactérien et le ribosome eucaryote, l'information obtenue pourra être extrapolée au ribosome eucaryote. De plus, le progrès de ces techniques permet d'espérer qu'il sera bientôt possible d'obtenir la structure cristallographique d'un ribosome

eucaryote. La caractérisation de sites d'interactions du ribosome avec le signal stimulateur pourra nous préciser le mécanisme d'action du déphasage.

## Bibliographie

- Adamski FM, Donly BC et Tate WP. 1993. Competition between frameshifting, termination and suppression at the frameshift site in the Escherichia coli release factor-2 mRNA. *Nucleic Acids Res* 21:5074-5078.
- Alam SL, Wills NM, Ingram JA, Atkins JF et Gesteland RF. 1999. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. *J Mol Biol* 288:837-852.
- Aldovini A et Young RA. 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* 64:1920-1926.
- Ambrose Z, Boltz V, Palmer S, Coffin JM, Hughes SH et Kewalramani VN. 2004. In vitro characterization of a simian immunodeficiency virus-human immunodeficiency virus (HIV) chimera expressing HIV type 1 reverse transcriptase to study antiviral resistance in pigtail macaques. *J Virol* 78:13553-13561.
- Asakura T, Sasaki T, Nagano F, Satoh A, Obaishi H, Nishioka H, Imamura H, Hotta K, Tanaka K, Nakanishi H et Takai Y. 1998. Isolation and characterization of a novel actin filament-binding protein from Saccharomyces cerevisiae. *Oncogene* 16:121-130.
- Atkins JF, Baranov PV, Fayet O, Herr AJ, Howard MT, Ivanov IP, Matsufuji S, Miller WA, Moore B, Prere MF, Wills NM, Zhou J et Gesteland RF. 2001. Overriding Standard Decoding: Implications of Recoding for Ribosome Function and Enrichment of Gene Expression. *Cold Spring Harbor Symposia on Quantitative Biology* 66:217-232.

- Atkins JF et Gesteland R. 2002. Biochemistry. The 22nd amino acid. *Science* 296:1409-1410.
- Baranov PV, Gesteland RF et Atkins JF. 2002a. Recoding: translational bifurcations in gene expression. *Gene* 286:187-201.
- Baranov PV, Gesteland RF et Atkins JF. 2002b. Release factor 2 frameshifting sites in different bacteria. *EMBO Rep* 3:373-377.
- Baranov PV, Gesteland RF et Atkins JF. 2004. P-site tRNA is a crucial initiator of ribosomal frameshifting. *Rna* 10:221-230.
- Baranov PV, Gurvich OL, Hammer AW, Gesteland RF et Atkins JF. 2003. Recode 2003. *Nucleic Acids Res* 31:87-89.
- Baril M, Dulude D, Gendron K, Lemay G et Brakier-Gingras L. 2003a. Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. *Rna* 9:1246-1253.
- Baril M, Dulude D, Steinberg SV et Brakier-Gingras L. 2003b. The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot. *J Mol Biol* 331:571-583.
- Barrell BG, Bankier AT et Drouin J. 1979. A different genetic code in human mitochondria. *Nature* 282:189-194.
- Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W et Montagnier L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220:868-871.

- Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD et Williamson JR. 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* 273:1547-1551.
- Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J et Blobel G. 2001. Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107:361-372.
- Behne D et Kyriakopoulos A. 2001. Mammalian selenium-containing proteins. *Annu Rev Nutr* 21:453-473.
- Beier H et Grimm M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res* 29:4767-4782.
- Bekaert M, Bidou L, Denise A, Duchateau-Nguyen G, Forest JP, Froidevaux C, Hatin I, Rousset JP et Termier M. 2003. Towards a computational model for -1 eukaryotic frameshifting sites. *Bioinformatics* 19:327-335.
- Bekaert M et Rousset JP. 2005. An Extended Signal Involved in Eukaryotic -1 Frameshifting Operates through Modification of the E Site tRNA. *Mol Cell* 17:61-68.
- Belanger F, Leger M, Saraiya AA, Cunningham PR et Brakier-Gingras L. 2002. Functional studies of the 900 tetraloop capping helix 27 of 16S ribosomal RNA. *J Mol Biol* 320:979-989.
- Belcourt MF et Farabaugh PJ. 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* 62:339-352.
- Bergstrom DE, Merli CA, Cygan JA, Shelby R et Blackman RK. 1995. Regulatory autonomy and molecular characterization of the *Drosophila* out at first gene. *Genetics* 139:1331-1346.

- Berkhout B. 1999. HIV-1 evolution under pressure of protease inhibitors: climbing the stairs of viral fitness. *J Biomed Sci* 6:298-305.
- Bertram G, Innes S, Minella O, Richardson J et Stansfield I. 2001. Endless possibilities: translation termination and stop codon recognition. *Microbiology* 147:255-269.
- Bertrand C, Prere MF, Gesteland RF, Atkins JF et Fayet O. 2002. Influence of the stacking potential of the base 3' of tandem shift codons on -1 ribosomal frameshifting used for gene expression. *Rna* 8:16-28.
- Besnier C, Takeuchi Y et Towers G. 2002. Restriction of lentivirus in monkeys. *Proc Natl Acad Sci U S A* 99:11920-11925.
- Bevis BJ et Glick BS. 2002. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nat Biotechnol* 20:83-87.
- Bidou L, Stahl G, Grima B, Liu H, Cassan M et Rousset JP. 1997. In vivo HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. *Rna* 3:1153-1158.
- Bieniasz PD. 2004. Intrinsic immunity: a front-line defense against viral attack. *Nat Immunol* 5:1109-1115.
- Biswas P, Jiang X, Pacchia AL, Dougherty JP et Peltz SW. 2004. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* 78:2082-2087.
- Blankson JN, Persaud D et Siliciano RF. 2002. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* 53:557-593.

- Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK et Krzycki JA. 2004. Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* 431:333-335.
- Blinkowa AL et Walker JR. 1990. Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res* 18:1725-1729.
- Böck A. 2000. Biosynthesis of selenoproteins--an overview. *Biofactors* 11:77-78.
- Bonetti B, Fu L, Moon J et Bedwell DM. 1995. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. *J Mol Biol* 251:334-345.
- Borrow P, Lewicki H, Hahn BH, Shaw GM et Oldstone MB. 1994. Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol* 68:6103-6110.
- Brakier-Gingras L, Melançon P, Côté M et Payant C. 1998. Use of antisense DNA oligomers for inhibition of HIV by decreasing ribosomal frameshifting. *no* 5707866.
- Brelot A, Heveker N, Montes M et Alizon M. 2000. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* 275:23736-23744.
- Brierley I. 1995. Ribosomal frameshifting viral RNAs. *J Gen Virol* 76 (Pt 8):1885-1892.
- Brierley I, Digard P et Inglis SC. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* 57:537-547.

- Brierley I, Jenner AJ et Inglis SC. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 227:463-479.
- Brierley I et Pennell S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic-1 ribosomal frameshifting. *Cold Spring Harb Symp Quant Biol* 66:233-248.
- Brierley I, Rolley NJ, Jenner AJ et Inglis SC. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 220:889-902.
- Brown CM, Stockwell PA, Trotman CN et Tate WP. 1990a. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res* 18:6339-6345.
- Brown CM, Stockwell PA, Trotman CN et Tate WP. 1990b. The signal for the termination of protein synthesis in procaryotes. *Nucleic Acids Res* 18:2079-2086.
- Brunelle MN, Payant C, Lemay G et Brakier-Gingras L. 1999. Expression of the human immunodeficiency virus frameshift signal in a bacterial cell-free system: influence of an interaction between the ribosome and a stem-loop structure downstream from the slippery site. *Nucleic Acids Res* 27:4783-4791.
- Buckingham RH. 1994. Codon context and protein synthesis: enhancements of the genetic code. *Biochimie* 76:351-354.
- Burge CB, T. Tuschl, and P. A. Sharp. 1999. *Splicing of precursors to mRNAs by the spliceosome*. Cold Spring Harbor, N.Y: Cold Spring Laboratory Harbor Press.

- Carroll R, Peterlin BM et Derse D. 1992. Inhibition of human immunodeficiency virus type 1 Tat activity by coexpression of heterologous trans activators. *J Virol* 66:2000-2007.
- Cassan M, Delaunay N, Vaquero C et Rousset JP. 1994. Translational frameshifting at the gag-pol junction of human immunodeficiency virus type 1 is not increased in infected T-lymphoid cells. *J Virol* 68:1501-1508.
- Chamorro M, Parkin N et Varmus HE. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci U S A* 89:713-717.
- Chen C et Montelaro RC. 2003. Characterization of RNA elements that regulate gag-pol ribosomal frameshifting in equine infectious anemia virus. *J Virol* 77:10280-10287.
- Chen H, Bjerknes M, Kumar R et Jay E. 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* 22:4953-4957.
- Chen X, Chamorro M, Lee SI, Shen LX, Hines JV, Tinoco I, Jr. et Varmus HE. 1995. Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. *Embo J* 14:842-852.
- Chen X, Kang H, Shen LX, Chamorro M, Varmus HE et Tinoco I, Jr. 1996. A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA. *J Mol Biol* 260:479-483.
- Chittum HS, Lane WS, Carlson BA, Roller PP, Lung FD, Lee BJ et Hatfield DL. 1998. Rabbit beta-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps. *Biochemistry* 37:10866-10870.

- Cimarelli A et Darlix JL. 2002. Assembling the human immunodeficiency virus type 1. *Cell Mol Life Sci* 59:1166-1184.
- Clapham PR et Weiss RA. 1997. Immunodeficiency viruses. Spoilt for choice of co-receptors. *Nature* 388:230-231.
- Clavel F et Charneau P. 1994. Fusion from without directed by human immunodeficiency virus particles. *J Virol* 68:1179-1185.
- Cobucci-Ponzano B, Rossi M et Moracci M. 2005. Recoding in archaea. *Mol Microbiol* 55:339-348.
- Cohen OJ et Fauci AS. 2001. Current strategies in the treatment of HIV infection. *Adv Intern Med* 46:207-246.
- Colas P, Cohen B, Jessen T, Grishina I, McCoy J et Brent R. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380:548-550.
- Collins KL. 2003. How HIV evades CTL recognition. *Curr HIV Res* 1:31-40.
- Copeland PR. 2003. Regulation of gene expression by stop codon recoding: selenocysteine. *Gene* 312:17-25.
- Corey MJ et Corey E. 1996. On the failure of de novo-designed peptides as biocatalysts. *Proc Natl Acad Sci U S A* 93:11428-11434.
- Cowan S, Hatzioannou T, Cunningham T, Muesing MA, Gottlinger HG et Bieniasz PD. 2002. Cellular inhibitors with Fv1-like activity restrict human and simian immunodeficiency virus tropism. *Proc Natl Acad Sci U S A* 99:11914-11919.
- Craig WJ et Caskey CT. 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* 322:273-275.

- Crick FH. 1966. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* 19:548-555.
- Cui Y, Dinman JD et Peltz SW. 1996. Mof4-1 is an allele of the UPF1/IFS2 gene which affects both mRNA turnover and -1 ribosomal frameshifting efficiency. *Embo J* 15:5726-5736.
- Curran JF et Yarus M. 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J Mol Biol* 209:65-77.
- Das C et Frankel AD. 2003. Sequence and structure space of RNA-binding peptides. *Biopolymers* 70:80-85.
- De Clercq E. 2002. New developments in anti-HIV chemotherapy. *Biochim Biophys Acta* 1587:258-275.
- Dinman J, Ruiz-Echevarria M, Wang W et Peltz S. 2000. The case for the involvement of the Upf3p in programmed -1 ribosomal frameshifting. *Rna* 6:1685-1686.
- Dinman JD, Icho T et Wickner RB. 1991. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci U S A* 88:174-178.
- Dinman JD, Richter S, Plant EP, Taylor RC, Hammell AB et Rana TM. 2002. The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure. *Proc Natl Acad Sci U S A* 99:5331-5336.
- Dinman JD, Ruiz-Echevarria MJ, Czaplinski K et Peltz SW. 1997. Peptidyl-transferase inhibitors have antiviral properties by altering programmed -1 ribosomal frameshifting efficiencies: development of model systems. *Proc Natl Acad Sci U S A* 94:6606-6611.

- Dinman JD, Ruiz-Echevarria MJ et Peltz SW. 1998. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol* 16:190-196.
- Dinman JD et Wickner RB. 1992. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J Virol* 66:3669-3676.
- Donly BC, Edgar CD, Adamski FM et Tate WP. 1990. Frameshift autoregulation in the gene for Escherichia coli release factor 2: partly functional mutants result in frameshift enhancement. *Nucleic Acids Res* 18:6517-6522.
- Dos Ramos F, Carrasco M, Doyle T et Brierley I. 2004. Programmed -1 ribosomal frameshifting in the SARS coronavirus. *Biochem Soc Trans* 32:1081-1083.
- Doudna JA et Rath VL. 2002. Structure and function of the eukaryotic ribosome: the next frontier. *Cell* 109:153-156.
- Driscoll DM et Copeland PR. 2003. Mechanism and regulation of selenoprotein synthesis. *Annu Rev Nutr* 23:17-40.
- Du Z, Holland JA, Hansen MR, Giedroc DP et Hoffman DW. 1997. Base-pairings within the RNA pseudoknot associated with the simian retrovirus-1 gag-pro frameshift site. *J Mol Biol* 270:464-470.
- Dulude D, Baril M et Brakier-Gingras L. 2002. Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res* 30:5094-5102.
- Dulude D, Berchiche YA, Gendron K, Brakier-Gingras L et Heveker N. 2005. Mutations within the signals that control the efficiency of the programmed -1 ribosomal

frameshift interfere with the replication of the human immunodeficiency virus type 1. *submitted to Virology*.

Farabaugh PJ. 1996. Programmed translational frameshifting. *Microbiol Rev* 60:103-134.

Farabaugh PJ. 2000. Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog Nucleic Acid Res Mol Biol* 64:131-170.

Farabaugh PJ et Bjork GR. 1999. How translational accuracy influences reading frame maintenance. *Embo J* 18:1427-1434.

Farabaugh PJ, Zhao H et Vimaladithan A. 1993. A novel programmed frameshift expresses the POL3 gene of retrotransposon Ty3 of yeast: frameshifting without tRNA slippage. *Cell* 74:93-103.

Fauci AS. 2003. HIV and AIDS: 20 years of science. *Nat Med* 9:839-843.

Felsenstein KM et Goff SP. 1988. Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *J Virol* 62:2179-2182.

Felsenstein KM et Goff SP. 1992. Mutational analysis of the gag-pol junction of Moloney murine leukemia virus: requirements for expression of the gag-pol fusion protein. *J Virol* 66:6601-6608.

Feng YX, Yuan H, Rein A et Levin JG. 1992. Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the gag termination codon followed by an RNA pseudoknot. *J Virol* 66:5127-5132.

Ferguson MR, Rojo DR, von Lindern JJ et O'Brien WA. 2002. HIV-1 replication cycle. *Clin Lab Med* 22:611-635.

- Flower AM et McHenry CS. 1990. The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc Natl Acad Sci U S A* 87:3713-3717.
- Franke EK, Yuan HE et Luban J. 1994. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 372:359-362.
- Frankel AD et Young JA. 1998. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 67:1-25.
- Frank-Kamenetskii MD et Mirkin SM. 1995. Triplex DNA structures. *Annu Rev Biochem* 64:65-95.
- Freed EO. 1998. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* 251:1-15.
- Freed EO. 2001a. HIV-1 replication. *Somat Cell Mol Genet* 26:13-33.
- Freed EO, and M. A. Martin. 2001b. HIVs and their replication. In: D. M. Knipe PMH, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus, ed. *Fields virology, 4th ed.* Philadelphia, Pa.: Lippincott Williams & Wilkins. pp 1971-2041.
- Gabuzda DH, Lawrence K, Langhoff E, Terwilliger E, Dorfman T, Haseltine WA et Sodroski J. 1992. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *J Virol* 66:6489-6495.
- Gallant JA et Lindsley D. 1998. Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream. *Proc Natl Acad Sci U S A* 95:13771-13776.

- Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI et Hill CP. 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* 278:849-853.
- Gaudin C, Mazauric MH, Traikia M, Guittet E, Yoshizawa S et Fourmy D. 2005. Structure of the RNA Signal Essential for Translational Frameshifting in HIV-1. *J Mol Biol* 348.
- Gelderblom HR. 1991. Assembly and morphology of HIV: potential effect of structure on viral function. *Aids* 5:617-637.
- Gendron K, Dulude D, Lemay G, Ferbeyre G et Brakier-Gingras L. 2005. The virion-associated Gag-Pol is decreased in chimeric Moloney murine leukemia viruses in which the readthrough region is replaced by the frameshift region of the human immunodeficiency virus type 1. *Virology* 334:342-352.
- Gesteland RF et Atkins JF. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* 65:741-768.
- Gesteland RF, Weiss RB et Atkins JF. 1992. Recoding: reprogrammed genetic decoding. *Science* 257:1640-1641.
- Giedroc DP, Theimer CA et Nixon PL. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* 298:167-185.
- Goff SP. 1990. Retroviral reverse transcriptase: synthesis, structure, and function. *J Acquir Immune Defic Syndr* 3:817-831.
- Goff SP. 1992. Genetics of retroviral integration. *Annu Rev Genet* 26:527-544.
- Goff SP. 2004. Retrovirus restriction factors. *Mol Cell* 16:849-859.

- Gottlinger HG, Dorfman T, Sodroski JG et Haseltine WA. 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* 88:3195-3199.
- Gougeon ML. 2003. Apoptosis as an HIV strategy to escape immune attack. *Nat Rev Immunol* 3:392-404.
- Grentzmann G, Ingram JA, Kelly PJ, Gesteland RF et Atkins JF. 1998. A dual-luciferase reporter system for studying recoding signals. *Rna* 4:479-486.
- Grundner-Culemann E, Martin GW, 3rd, Tujebajeva R, Harney JW et Berry MJ. 2001. Interplay between termination and translation machinery in eukaryotic selenoprotein synthesis. *J Mol Biol* 310:699-707.
- Gulick RM. 2003. New antiretroviral drugs. *Clin Microbiol Infect* 9:186-193.
- Guzman LM, Belin D, Carson MJ et Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177:4121-4130.
- Hahn BH, Shaw GM, De Cock KM et Sharp PM. 2000. AIDS as a zoonosis: scientific and public health implications. *Science* 287:607-614.
- Hammell AB, Taylor RC, Peltz SW et Dinman JD. 1999. Identification of putative programmed -1 ribosomal frameshift signals in large DNA databases. *Genome Res* 9:417-427.
- Han H et Dervan PB. 1993. Sequence-specific recognition of double helical RNA and RNA.DNA by triple helix formation. *Proc Natl Acad Sci U S A* 90:3806-3810.
- Hansen TM, Baranov PV, Ivanov IP, Gesteland RF et Atkins JF. 2003. Maintenance of the correct open reading frame by the ribosome. *EMBO Rep* 4:499-504.

- Hao B, Gong W, Ferguson TK, James CM, Krzycki JA et Chan MK. 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296:1462-1466.
- Harada K, Martin SS et Frankel AD. 1996. Selection of RNA-binding peptides in vivo. *Nature* 380:175-179.
- Harger JW et Dinman JD. 2003. An in vivo dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *Rna* 9:1019-1024.
- Harger JW et Dinman JD. 2004. Evidence against a direct role for the Upf proteins in frameshifting or nonsense codon readthrough. *Rna* 10:1721-1729.
- Harger JW, Meskauskas A et Dinman JD. 2002. An "integrated model" of programmed ribosomal frameshifting. *Trends Biochem Sci* 27:448-454.
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS et Malim MH. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* 113:803-809.
- Hatfield DL et Gladyshev VN. 2002. How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22:3565-3576.
- Hentze MW et Kulozik AE. 1999. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 96:307-310.
- Hernandez LD, Hoffman LR, Wolfsberg TG et White JM. 1996. Virus-cell and cell-cell fusion. *Annu Rev Cell Dev Biol* 12:627-661.
- Herr AJ, Atkins JF et Gesteland RF. 2000. Coupling of open reading frames by translational bypassing. *Annu Rev Biochem* 69:343-372.

- Herr AJ, Nelson CC, Wills NM, Gesteland RF et Atkins JF. 2001. Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA. *J Mol Biol* 309:1029-1048.
- Hibbert CS et Rein A. 2005. Preliminary physical mapping of RNA-RNA linkages in the genomic RNA of Moloney murine leukemia virus. *J Virol* 79:8142-8148.
- Hill MK, Shehu-Xhilaga M, Crowe SM et Mak J. 2002. Proline residues within spacer peptide p1 are important for human immunodeficiency virus type 1 infectivity, protein processing, and genomic RNA dimer stability. *J Virol* 76:11245-11253.
- Hillman RT, Green RE et Brenner SE. 2004. An unappreciated role for RNA surveillance. *Genome Biol* 5:R8.
- Hinnebusch AG. 1997. Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *J Biol Chem* 272:21661-21664.
- Hizi A, Henderson LE, Copeland TD, Sowder RC, Hixson CV et Oroszlan S. 1987. Characterization of mouse mammary tumor virus gag-pro gene products and the ribosomal frameshift site by protein sequencing. *Proc Natl Acad Sci U S A* 84:7041-7045.
- Ho SN, Hunt HD, Horton RM, Pullen JK et Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59.
- Holcik M et Sonenberg N. 2005. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol* 6:318-327.
- Holcik M, Sonenberg N et Korneluk RG. 2000. Internal ribosome initiation of translation and the control of cell death. *Trends Genet* 16:469-473.

- Honigman A, Wolf D, Yaish S, Falk H et Panet A. 1991. cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology* 183:313-319.
- Horsfield JA, Wilson DN, Mannering SA, Adamski FM et Tate WP. 1995. Prokaryotic ribosomes recode the HIV-1 gag-pol-1 frameshift sequence by an E/P site post-translocation simultaneous slippage mechanism. *Nucleic Acids Res* 23:1487-1494.
- Hoshino S, Imai M, Kobayashi T, Uchida N et Katada T. 1999. The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J Biol Chem* 274:16677-16680.
- Huang M, Orenstein JM, Martin MA et Freed EO. 1995. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* 69:6810-6818.
- Huang WM, Ao SZ, Casjens S, Orlandi R, Zeikus R, Weiss R, Winge D et Fang M. 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* 239:1005-1012.
- Hudak KA, Hammell AB, Yasenchak J, Tumer NE et Dinman JD. 2001. A C-terminal deletion mutant of pokeweed antiviral protein inhibits programmed +1 ribosomal frameshifting and Ty1 retrotransposition without depurinating the sarcin/ricin loop of rRNA. *Virology* 279:292-301.
- Hung M, Patel P, Davis S et Green SR. 1998. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J Virol* 72:4819-4824.
- Huttenhofer A et Böck A. 1998. Selenocysteine inserting RNA elements modulate GTP hydrolysis of elongation factor SelB. *Biochemistry* 37:885-890.

- Ivanov IP, Gesteland RF et Atkins JF. 1998a. Does antizyme exist in *Escherichia coli*? *Mol Microbiol* 29:1521-1522.
- Ivanov IP, Gesteland RF et Atkins JF. 1998b. A second mammalian antizyme: conservation of programmed ribosomal frameshifting. *Genomics* 52:119-129.
- Ivanov IP, Gesteland RF, Matsufuji S et Atkins JF. 1998c. Programmed frameshifting in the synthesis of mammalian antizyme is +1 in mammals, predominantly +1 in fission yeast, but -2 in budding yeast. *Rna* 4:1230-1238.
- Ivanov IP, Matsufuji S, Murakami Y, Gesteland RF et Atkins JF. 2000a. Conservation of polyamine regulation by translational frameshifting from yeast to mammals. *Embo J* 19:1907-1917.
- Ivanov IP, Rohrwasser A, Terreros DA, Gesteland RF et Atkins JF. 2000b. Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: antizyme 3. *Proc Natl Acad Sci U S A* 97:4808-4813.
- Ivey DM, Cheng J et Krulwich TA. 1992. A 1.6 kb region of *Bacillus firmus* OF4 DNA encodes a homolog of *Escherichia coli* and yeast DNA topoisomerases and may contain a translational readthrough of UGA. *Nucleic Acids Res* 20:4928.
- Jacks T, Madhani HD, Masiarz FR et Varmus HE. 1988a. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55:447-458.
- Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ et Varmus HE. 1988b. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 331:280-283.
- Jacks T et Varmus HE. 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230:1237-1242.

- Jackson RJ. 2000. A comparative view of initiation site selection mechanisms. In: Sonenberg N, Hershey, J.W.B. and Mathews, M.B., ed. *Translational Control of Gene Expression*. Cold Spring Harbor, New York,: Cold Spring Harbor Press,. pp 127-183.
- Jackson RJ et Kaminski A. 1995. Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *Rna* 1:985-1000.
- Jalajakumari MB, Thomas CJ, Halter R et Manning PA. 1989. Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic Escherichia coli: novel regulation of pilus production by bypassing an amber codon. *Mol Microbiol* 3:1685-1695.
- Jenner L, Romby P, Rees B, Schulze-Briese C, Springer M, Ehresmann C, Ehresmann B, Moras D, Yusupova G et Yusupov M. 2005. Translational operator of mRNA on the ribosome: how repressor proteins exclude ribosome binding. *Science* 308:120-123.
- Jiang B, Monroe SS, Koonin EV, Stine SE et Glass RI. 1993. RNA sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. *Proc Natl Acad Sci U S A* 90:10539-10543.
- Jordan M, Schallhorn A et Wurm FM. 1996. Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res* 24:596-601.
- Julias JG, Ferris AL, Boyer PL et Hughes SH. 2001. Replication of phenotypically mixed human immunodeficiency virus type 1 virions containing catalytically active and catalytically inactive reverse transcriptase. *J Virol* 75:6537-6546.

- Kang H. 1998. Direct structural evidence for formation of a stem-loop structure involved in ribosomal frameshifting in human immunodeficiency virus type 1. *Biochim Biophys Acta* 1397:73-78.
- Kang H, Hines JV et Tinoco I, Jr. 1996. Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J Mol Biol* 259:135-147.
- Kang H et Tinoco I, Jr. 1997. A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. *Nucleic Acids Res* 25:1943-1949.
- Kapp LD et Lorsch JR. 2004. The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* 73:657-704.
- Karacostas V, Wolffe EJ, Nagashima K, Gonda MA et Moss B. 1993. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* 193:661-671.
- Karimi R, Pavlov MY, Buckingham RH et Ehrenberg M. 1999. Novel roles for classical factors at the interface between translation termination and initiation. *Mol Cell* 3:601-609.
- Kawakami K, Pande S, Faiola B, Moore DP, Boeke JD, Farabaugh PJ, Strathern JN, Nakamura Y et Garfinkel DJ. 1993. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in *Saccharomyces cerevisiae*. *Genetics* 135:309-320.
- Kim YG, Su L, Maas S, O'Neill A et Rich A. 1999. Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc Natl Acad Sci U S A* 96:14234-14239.

- Kollmus H, Hentze MW et Hauser H. 1996. Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA* 2:316-323.
- Kollmus H, Honigman A, Panet A et Hauser H. 1994. 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* 68:6087-6091.
- Kondo E et Gottlinger HG. 1996. A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *J Virol* 70:159-164.
- Kontos H, Naphine S et Brierley I. 2001. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* 21:8657-8670.
- Kopelowitz J, Hampe C, Goldman R, Reches M et Engelberg-Kulka H. 1992. Influence of codon context on UGA suppression and readthrough. *J Mol Biol* 225:261-269.
- Kozak M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234:187-208.
- Krainer AR. 1997. *Eukaryotic mRNA processing*. New York, N.Y: Oxford University Press Inc.
- Krausslich HG, Facke M, Heuser AM, Konvalinka J et Zentgraf H. 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol* 69:3407-3419.
- Krol A. 2002. Evolutionarily different RNA motifs and RNA-protein complexes to achieve selenoprotein synthesis. *Biochimie* 84:765-774.
- Kuiken C, Korber B et Shafer RW. 2003. HIV sequence databases. *AIDS Rev* 5:52-61.

- Kurland CG. 1992. Translational accuracy and the fitness of bacteria. *Annu Rev Genet* 26:29-50.
- LaBonte J, Lebbos J et Kirkpatrick P. 2003. Enfuvirtide. *Nat Rev Drug Discov* 2:345-346.
- Larsen B, Wills NM, Gesteland RF et Atkins JF. 1994. rRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol* 176:6842-6851.
- Laursen BS, Sorensen HP, Mortensen KK et Sperling-Petersen HU. 2005. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 69:101-123.
- LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF et McCoy JM. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology (N Y)* 11:187-193.
- Le SY, Chen JH et Maizel JV. 1989. Thermodynamic stability and statistical significance of potential stem-loop structures situated at the frameshift sites of retroviruses. *Nucleic Acids Res* 17:6143-6152.
- Lebars I, Lamontagne B, Yoshizawa S, Aboul-Elela S et Fourmy D. 2001. Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. *Embo J* 20:7250-7258.
- Lecossier D, Bouchonnet F, Clavel F et Hance AJ. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* 300:1112.
- Léger M, Sidani S et Brakier-Gingras L. 2004. A reassessment of the response of the bacterial ribosome to the frameshift stimulatory signal of the human immunodeficiency virus type 1. *Rna* 10:1225-1235.
- Lescure A, Fagegaltier D, Carbon P et Krol A. 2002. Protein factors mediating selenoprotein synthesis. *Curr Protein Pept Sci* 3:143-151.

- Li Z, Stahl G et Farabaugh PJ. 2001. Programmed +1 frameshifting stimulated by complementarity between a downstream mRNA sequence and an error-correcting region of rRNA. *Rna* 7:275-284.
- Lim AC et Barton JK. 1998. Rh(phen)<sub>2</sub>phi<sub>3</sub><sup>+</sup> as a shape-selective probe of triple helices. *Biochemistry* 37:9138-9146.
- Liphardt J, Naphine S, Kontos H et Brierley I. 1999. Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. *J Mol Biol* 288:321-335.
- Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC, Koup RA, Mellors JW, Connick E, Conway B, Kilby M, Wang L, Whitcomb JM, Hellmann NS et Richman DD. 2002. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 347:385-394.
- Lopinski JD, Dinman JD et Bruenn JA. 2000. Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol Cell Biol* 20:1095-1103.
- Lu YL, Bennett RP, Wills JW, Gorelick R et Ratner L. 1995. A leucine triplet repeat sequence (LXX)<sub>4</sub> in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. *J Virol* 69:6873-6879.
- Lu Z, Tripp BC et McCoy JM. 1998. Displaying libraries of conformationally constrained peptides on the surface of Escherichia coli as flagellin fusions. *Methods Mol Biol* 87:265-280.
- Luban J et Goff SP. 1991. Binding of human immunodeficiency virus type 1 (HIV-1) RNA to recombinant HIV-1 gag polyprotein. *J Virol* 65:3203-3212.
- Mador N, Panet A et Honigman A. 1989. Translation of gag, pro, and pol gene products of human T-cell leukemia virus type 2. *J Virol* 63:2400-2404.

- Major LL, Poole ES, Dalphin ME, Mannering SA et Tate WP. 1996. Is the in-frame termination signal of the Escherichia coli release factor-2 frameshift site weakened by a particularly poor context? *Nucleic Acids Res* 24:2673-2678.
- Mak J, Jiang M, Wainberg MA, Hammarskjold ML, Rekosh D et Kleiman L. 1994. Role of Pr160gag-pol in mediating the selective incorporation of tRNA(Lys) into human immunodeficiency virus type 1 particles. *J Virol* 68:2065-2072.
- Manch-Citron JN, Allen J, Moos M, Jr. et London J. 1992. The gene encoding a Prevotella loescheii lectin-like adhesin contains an interrupted sequence which causes a frameshift. *J Bacteriol* 174:7328-7336.
- Manch-Citron JN, Dey A, Ewell JB et Nguyen NY. 1999a. Mutant analysis of Prevotella sp. plaA-lacZ fusion protein expression in Escherichia coli: support for an essential role of the stem-loop. *Can J Microbiol* 45:153-161.
- Manch-Citron JN, Dey A, Schneider R et Nguyen NY. 1999b. The translational hop junction and the 5' transcriptional start site for the Prevotella loescheii adhesin encoded by plaA. *Curr Microbiol* 38:22-26.
- Manch-Citron JN et London J. 1994. Expression of the Prevotella loescheii adhesin gene (plaA) is mediated by a programmed frameshifting hop. *J Bacteriol* 176:1944-1948.
- Manktelow E, Shigemoto K et Brierley I. 2005. Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Res* 33:1553-1563.
- Maquat LE. 2000. *in* Translational Control of Gene Expression. In: (eds. Sonenberg N, Hershey, J.W.B. & Mathews, M.B.), ed. Cold Spring Harbor: Cold Spring Harbor Press,.

- Marczinke B, Hagervall T et Brierley I. 2000. The Q-base of asparaginyl-tRNA is dispensable for efficient -1 ribosomal frameshifting in eukaryotes. *J Mol Biol* 295:179-191.
- Marin M, Rose KM, Kozak SL et Kabat D. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* 9:1398-1403.
- Marquez V, Wilson DN, Tate WP, Triana-Alonso F et Nierhaus KH. 2004. Maintaining the ribosomal reading frame: the influence of the E site during translational regulation of release factor 2. *Cell* 118:45-55.
- Matsufuji S, Matsufuji T, Miyazaki Y, Murakami Y, Atkins JF, Gesteland RF et Hayashi S. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* 80:51-60.
- Matsufuji S, Matsufuji T, Wills NM, Gesteland RF et Atkins JF. 1996. Reading two bases twice: mammalian antizyme frameshifting in yeast. *Embo J* 15:1360-1370.
- Matsufuji S, Miyazaki Y, Kanamoto R, Kameji T, Murakami Y, Baby TG, Fujita K, Ohno T et Hayashi S. 1990. Analyses of ornithine decarboxylase antizyme mRNA with a cDNA cloned from rat liver. *J Biochem (Tokyo)* 108:365-371.
- McCaughan KK, Brown CM, Dalphin ME, Berry MJ et Tate WP. 1995. Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proc Natl Acad Sci U S A* 92:5431-5435.
- Mears JA, Cannone JJ, Stagg SM, Gutell RR, Agrawal RK et Harvey SC. 2002. Modeling a minimal ribosome based on comparative sequence analysis. *J Mol Biol* 321:215-234.

- Michiels PJ, Versleijen AA, Verlaan PW, Pleij CW, Hilbers CW et Heus HA. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol* 310:1109-1123.
- Miller MD, Bor YC et Bushman F. 1995. Target DNA capture by HIV-1 integration complexes. *Curr Biol* 5:1047-1056.
- Moazed D et Noller HF. 1989. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342:142-148.
- Moazed D, Robertson JM et Noller HF. 1988. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. *Nature* 334:362-364.
- Morris DK et Lundblad V. 1997. Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol* 7:969-976.
- Morris DR et Geballe AP. 2000. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* 20:8635-8642.
- Murakami Y, Tanaka K, Matsufuji S, Miyazaki Y et Hayashi S. 1992. Antizyme, a protein induced by polyamines, accelerates the degradation of ornithine decarboxylase in Chinese-hamster ovary-cell extracts. *Biochem J* 283 ( Pt 3):661-664.
- Muriaux D, Darlix JL et Cimorelli A. 2004. Targeting the assembly of the human immunodeficiency virus type I. *Curr Pharm Des* 10:3725-3739.
- Nakamura Y et Ito K. 2003. Making sense of mimic in translation termination. *Trends Biochem Sci* 28:99-105.
- Nam SH, Copeland TD, Hatanaka M et Oroszlan S. 1993. Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *J Virol* 67:196-203.

- Namy O, Duchateau-Nguyen G et Rousset JP. 2002. Translational readthrough of the PDE2 stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. *Mol Microbiol* 43:641-652.
- Namy O, Rousset JP, Napthine S et Brierley I. 2004. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* 13:157-168.
- Napthine S, Liphardt J, Bloys A, Routledge S et Brierley I. 1999. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J Mol Biol* 288:305-320.
- Nasim MT, Jaenecke S, Belduz A, Kollmus H, Flohe L et McCarthy JE. 2000. Eukaryotic selenocysteine incorporation follows a nonprocessive mechanism that competes with translational termination. *J Biol Chem* 275:14846-14852.
- Navia MA et McKeever BM. 1990. A role for the aspartyl protease from the human immunodeficiency virus type 1 (HIV-1) in the orchestration of virus assembly. *Ann NY Acad Sci* 616:73-85.
- Nierhaus KH. 1990. The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry* 29:4997-5008.
- Nisole S, Lynch C, Stoye JP et Yap MW. 2004. A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *Proc Natl Acad Sci U S A* 101:13324-13328.
- Noller HF. 2004. The driving force for molecular evolution of translation. *Rna* 10:1833-1837.
- Noller HF, Yusupov MM, Yusupova GZ, Baucom A et Cate JH. 2002. Translocation of tRNA during protein synthesis. *FEBS Lett* 514:11-16.

- Nowakowski J et Tinoco IJ. 1999. RNA structure in solution. In: Neidle S, ed. *Oxford Handbook of Nucleic Acid Structure*. New York, NY: Oxford University Press. pp 567–602.
- Ogle JM, Brodersen DE, Clemons WM, Jr., Tarry MJ, Carter AP et Ramakrishnan V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292:897-902.
- Ogle JM, Carter AP et Ramakrishnan V. 2003. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci* 28:259-266.
- Ogle JM et Ramakrishnan V. 2004. Structural Insights into Translational Fidelity. *Annu Rev Biochem*.
- Orlova M, Yueh A, Leung J et Goff SP. 2003. Reverse transcriptase of Moloney murine leukemia virus binds to eukaryotic release factor 1 to modulate suppression of translational termination. *Cell* 115:319-331.
- Oroszlan S et Copeland TD. 1985. Primary structure and processing of gag and env gene products of human T-cell leukemia viruses HTLV-ICR and HTLV-IATK. *Curr Top Microbiol Immunol* 115:221-233.
- Paillart JC, Shehu-Xhilaga M, Marquet R et Mak J. 2004. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* 2:461-472.
- Paillart JC, Skripkin E, Ehresmann B, Ehresmann C et Marquet R. 2002. In vitro evidence for a long range pseudoknot in the 5'-untranslated and matrix coding regions of HIV-1 genomic RNA. *J Biol Chem* 277:5995-6004.
- Palella FJ, Jr., Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Aschman DJ et Holmberg SD. 1998. Declining morbidity and mortality among patients with

- advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* 338:853-860.
- Pande S, Vimaladithan A, Zhao H et Farabaugh PJ. 1995. Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. *Mol Cell Biol* 15:298-304.
- Park J et Morrow CD. 1991. Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. *J Virol* 65:5111-5117.
- Parkin NT, Chamorro M et Varmus HE. 1992. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol* 66:5147-5151.
- Paul CP, Barry JK, Dinesh-Kumar SP, Brault V et Miller WA. 2001. A sequence required for -1 ribosomal frameshifting located four kilobases downstream of the frameshift site. *J Mol Biol* 310:987-999.
- Paulus C, Ludwig C et Wagner R. 2004. Contribution of the Gag-Pol transframe domain p6\* and its coding sequence to morphogenesis and replication of human immunodeficiency virus type 1. *Virology* 330:271-283.
- Peden K, Emerman M et Montagnier L. 1991. Changes in growth properties on passage in tissue culture of viruses derived from infectious molecular clones of HIV-1LAI, HIV-1MAL, and HIV-1ELI. *Virology* 185:661-672.
- Pelchen-Matthews A, Kramer B et Marsh M. 2003. Infectious HIV-1 assembles in late endosomes in primary macrophages. *J Cell Biol* 162:443-455.
- Peterlin BM et Trono D. 2003. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat Rev Immunol* 3:97-107.

- Pettit SC, Everitt LE, Choudhury S, Dunn BM et Kaplan AH. 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J Virol* 78:8477-8485.
- Pettit SC, Gulnik S, Everitt L et Kaplan AH. 2003. The dimer interfaces of protease and extra-protease domains influence the activation of protease and the specificity of GagPol cleavage. *J Virol* 77:366-374.
- Pettit SC, Henderson GJ, Schiffer CA et Swanstrom R. 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol* 76:10226-10233.
- Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, Klein CA et Swanstrom R. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 68:8017-8027.
- Phillips-Jones MK, Hill LS, Atkinson J et Martin R. 1995. Context effects on misreading and suppression at UAG codons in human cells. *Mol Cell Biol* 15:6593-6600.
- Plant EP et Dinman JD. 2005. Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res* 33:1825-1833.
- Plant EP, Jacobs KL, Harger JW, Meskauskas A, Jacobs JL, Baxter JL, Petrov AN et Dinman JD. 2003. The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *Rna* 9:168-174.
- Pleij CW, Rietveld K et Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* 13:1717-1731.
- Polycarpo C, Ambrogelly A, Ruan B, Tumbula-Hansen D, Ataide SF, Ishitani R, Yokoyama S, Nureki O, Ibba M et Soll D. 2003. Activation of the pyrrolysine

suppressor tRNA requires formation of a ternary complex with class I and class II lysyl-tRNA synthetases. *Mol Cell* 12:287-294.

Pomerantz RJ et Horn DL. 2003. Twenty years of therapy for HIV-1 infection. *Nat Med* 9:867-873.

Poole ES, Brown CM et Tate WP. 1995. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *Embo J* 14:151-158.

Poole ES, Major LL, Mannering SA et Tate WP. 1998. Translational termination in *Escherichia coli*: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals. *Nucleic Acids Res* 26:954-960.

Puglisi JD, Blanchard SC et Green R. 2000. Approaching translation at atomic resolution. *Nat Struct Biol* 7:855-861.

Ramakrishnan V. 2002. Ribosome structure and the mechanism of translation. *Cell* 108:557-572.

Reil H, Kollmus H, Weidle UH et Hauser H. 1993. A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J Virol* 67:5579-5584.

Rodnina MV et Wintermeyer W. 2001. Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu Rev Biochem* 70:415-435.

Ruiz-Echevarria MJ, Yasenachak JM, Han X, Dinman JD et Peltz SW. 1998. The upf3 protein is a component of the surveillance complex that monitors both translation and mRNA turnover and affects viral propagation. *Proc Natl Acad Sci U S A* 95:8721-8726.

- Ryabova LA, Pooggin MM et Hohn T. 2002. Viral strategies of translation initiation: ribosomal shunt and reinitiation. *Prog Nucleic Acid Res Mol Biol* 72:1-39.
- Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, Racz P, Tenner-Racz K, Dalesandro M, Scallon BJ, Ghayeb J, Forman MA, Montefiori DC, Rieber EP, Letvin NL et Reimann KA. 1999. Control of viremia in simian immunodeficiency virus infection by CD8<sup>+</sup> lymphocytes. *Science* 283:857-860.
- Sekine Y, Eisaki N et Ohtsubo E. 1994. Translational control in production of transposase and in transposition of insertion sequence IS3. *J Mol Biol* 235:1406-1420.
- Sheehy AM, Gaddis NC, Choi JD et Malim MH. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418:646-650.
- Sheehy AM, Gaddis NC et Malim MH. 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 9:1404-1407.
- Shehu-Xhilaga M, Crowe SM et Mak J. 2001. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol* 75:1834-1841.
- Shehu-Xhilaga M, Hill M, Marshall JA, Kappes J, Crowe SM et Mak J. 2002a. The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol* 76:4331-4340.
- Shehu-Xhilaga M, Lee JY, Campbell S, Marshall JA, Crowe SM et Mak J. 2002b. Overexpression and incorporation of GagPol precursor does not impede packaging of HIV-1 tRNA(Lys3) but promotes intracellular budding of virus-like particles. *J Biomed Sci* 9:697-705.

- Sheng N, Pettit SC, Tritch RJ, Ozturk DH, Rayner MM, Swanstrom R et Erickson-Viitanen S. 1997. Determinants of the human immunodeficiency virus type 1 p15NC-RNA interaction that affect enhanced cleavage by the viral protease. *J Virol* 71:5723-5732.
- Shigemoto K, Brennan J, Walls E, Watson CJ, Stott D, Rigby PW et Reith AD. 2001. Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting. *Nucleic Acids Res* 29:4079-4088.
- Shine J et Dalgarno L. 1974. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* 71:1342-1346.
- Singh G et Lykke-Andersen J. 2003. New insights into the formation of active nonsense-mediated decay complexes. *Trends Biochem Sci* 28:464-466.
- Smith AJ, Cho MI, Hammarskjold ML et Rekosh D. 1990. Human immunodeficiency virus type 1 Pr55gag and Pr160gag-pol expressed from a simian virus 40 late replacement vector are efficiently processed and assembled into viruslike particles. *J Virol* 64:2743-2750.
- Somogyi P, Jenner AJ, Brierley I et Inglis SC. 1993. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* 13:6931-6940.
- Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G et Frank J. 2001. Structure of the 80S ribosome from *Saccharomyces cerevisiae*--tRNA-ribosome and subunit-subunit interactions. *Cell* 107:373-386.
- Spahn CM, Gomez-Lorenzo MG, Grassucci RA, Jorgensen R, Andersen GR, Beckmann R, Penczek PA, Ballesta JP et Frank J. 2004. Domain movements of elongation factor

- eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *Embo J* 23:1008-1019.
- Srinivasan G, James CM et Krzycki JA. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296:1459-1462.
- Stahl G, Ben Salem S, Li Z, McCarty G, Raman A, Shah M et Farabaugh PJ. 2001a. Programmed +1 Translational Frameshifting in the Yeast *Saccharomyces cerevisiae* Results from Disruption of Translational Error Correction. *Cold Spring Harbor Symposia on Quantitative Biology* 66:249-258.
- Stahl G, Ben Salem S, Li Z, McCarty G, Raman A, Shah M et Farabaugh PJ. 2001b. Programmed +1 translational frameshifting in the yeast *Saccharomyces cerevisiae* results from disruption of translational error correction. *Cold Spring Harb Symp Quant Biol* 66:249-258.
- Stahl G, Bidou L, Hatin I, Namy O, Rousset JP et Farabaugh P. 2000. The case against the involvement of the NMD proteins in programmed frameshifting. *Rna* 6:1687-1688.
- Stahl G, McCarty GP et Farabaugh PJ. 2002. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci* 27:178-183.
- Staple DW et Butcher SE. 2003. Solution structure of the HIV-1 frameshift inducing stem-loop RNA. *Nucleic Acids Res* 31:4326-4331.
- Staple DW et Butcher SE. 2005. Solution Structure and Thermodynamic Investigation of the HIV-1 Frameshift Inducing Element. *J Mol Biol* 348.
- Steneberg P, Englund C, Kronhamn J, Weaver TA et Samakovlis C. 1998. Translational readthrough in the *hdc* mRNA generates a novel branching inhibitor in the *drosophila* trachea. *Genes Dev* 12:956-967.

- Strebel K. 2003. Virus-host interactions: role of HIV proteins Vif, Tat, and Rev. *Aids* 17 Suppl 4:S25-34.
- Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P et Sodroski J. 2004. The cytoplasmic body component TRIM5 $\alpha$  restricts HIV-1 infection in Old World monkeys. *Nature* 427:848-853.
- Su L, Chen L, Egli M, Berger JM et Rich A. 1999. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat Struct Biol* 6:285-292.
- Sundaram M, Durant PC et Davis DR. 2000. Hypermodified nucleosides in the anticodon of tRNA(Lys) stabilize a canonical U-turn structure. *Biochemistry* 39:15652.
- Swanstrom R, and Wills, J.R. 1997. Synthesis, assembly and processing of viral proteins. in *Retroviruses*. J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 263-334.
- Tang C, Loeliger E, Luncsford P, Kinde I, Beckett D et Summers MF. 2004. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proc Natl Acad Sci USA* 101:517-522.
- Tang H, Kuhlen KL et Wong-Staal F. 1999. Lentivirus replication and regulation. *Annu Rev Genet* 33:133-170.
- Tate WP, Poole ES, Dalphin ME, Major LL, Crawford DJ et Mannering SA. 1996. The translational stop signal: codon with a context, or extended factor recognition element? *Biochimie* 78:945-952.
- Tate WP, Poole ES, Horsfield JA, Mannering SA, Brown CM, Moffat JG, Dalphin ME, McCaughan KK, Major LL et Wilson DN. 1995. Translational termination efficiency in both bacteria and mammals is regulated by the base following the stop codon. *Biochem Cell Biol* 73:1095-1103.

- Telenti A, Martinez R, Munoz M, Bleiber G, Greub G, Sanglard D et Peters S. 2002. Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem-loop structure. *J Virol* 76:7868-7873.
- ten Dam E, Brierley I, Inglis S et Pleij C. 1994. Identification and analysis of the pseudoknot-containing gag-pro ribosomal frameshift signal of simian retrovirus-1. *Nucleic Acids Res* 22:2304-2310.
- ten Dam EB, Pleij CW et Bosch L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* 4:121-136.
- ten Dam EB, Verlaan PW et Pleij CW. 1995. Analysis of the role of the pseudoknot component in the SRV-1 gag-pro ribosomal frameshift signal: loop lengths and stability of the stem regions. *Rna* 1:146-154.
- Towers GJ, Hatzioannou T, Cowan S, Goff SP, Luban J et Bieniasz PD. 2003. Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. *Nat Med* 9:1138-1143.
- Tsuchihashi Z et Brown PO. 1992. Sequence requirements for efficient translational frameshifting in the Escherichia coli dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev* 6:511-519.
- Tu C, Tzeng TH et Bruenn JA. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci U S A* 89:8636-8640.
- Tumer NE, Parikh BA, Li P et Dinman JD. 1998. The pokeweed antiviral protein specifically inhibits Ty1-directed +1 ribosomal frameshifting and retrotransposition in Saccharomyces cerevisiae. *J Virol* 72:1036-1042.
- Turner BG et Summers MF. 1999. Structural biology of HIV. *J Mol Biol* 285:1-32.

- Tzeng TH, Tu CL et Bruenn JA. 1992. Ribosomal frameshifting requires a pseudoknot in the *Saccharomyces cerevisiae* double-stranded RNA virus. *J Virol* 66:999-1006.
- UNAIDS. 2002. AIDS epidemic update. Genève: United Nations Program on HIV/AIDS (UNAIDS).
- Urban C, Zerfass K, Fingerhut C et Beier H. 1996. UGA suppression by tRNACmCATrp occurs in diverse virus RNAs due to a limited influence of the codon context. *Nucleic Acids Res* 24:3424-3430.
- Valle RP, Drugeon G, Devignes-Morch MD, Legocki AB et Haenni AL. 1992. Codon context effect in virus translational readthrough. A study in vitro of the determinants of TMV and Mo-MuLV amber suppression. *FEBS Lett* 306:133-139.
- Vink C et Plasterk RH. 1993. The human immunodeficiency virus integrase protein. *Trends Genet* 9:433-438.
- von Poblitzki A, Wagner R, Niedrig M, Wanner G, Wolf H et Modrow S. 1993. Identification of a region in the Pr55gag-polyprotein essential for HIV-1 particle formation. *Virology* 193:981-985.
- Wang WK, Chen MY, Chuang CY, Jeang KT et Huang LM. 2000. Molecular biology of human immunodeficiency virus type 1. *J Microbiol Immunol Infect* 33:131-140.
- Weiss MA et Narayana N. 1998. RNA recognition by arginine-rich peptide motifs. *Biopolymers* 48:167-180.
- Weiss R, Dunn DM, Atkins JF, Gesteland RF. 1987. Slippery runs, shifty stops, backward steps, and forward hops: 2, 1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spring Harb Symp Quant Biol.* pp 687-693.

- Weiss RB, Dunn DM, Dahlberg AE, Atkins JF et Gesteland RF. 1988. Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *Embo J* 7:1503-1507.
- Weiss RB, Dunn DM, Shuh M, Atkins JF et Gesteland RF. 1989. *E. coli* ribosomes rephase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol* 1:159-169.
- Weiss RB, Huang WM et Dunn DM. 1990. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* 62:117-126.
- Wieggers K, Rutter G, Kottler H, Tessmer U, Hohenberg H et Krausslich HG. 1998. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* 72:2846-2854.
- Wills JW et Craven RC. 1991. Form, function, and use of retroviral gag proteins. *Aids* 5:639-654.
- Wills NM, Gesteland RF et Atkins JF. 1994. Pseudoknot-dependent read-through of retroviral gag termination codons: importance of sequences in the spacer and loop 2. *Embo J* 13:4137-4144.
- Wilson KS et Noller HF. 1998. Molecular movement inside the translational engine. *Cell* 92:337-349.
- Wimmer J, Fujinaga K, Taube R, Cujec TP, Zhu Y, Peng J, Price DH et Peterlin BM. 1999. Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. *Virology* 255:182-189.

- Wu H, Yang PK, Butcher SE, Kang S, Chanfreau G et Feigon J. 2001. A novel family of RNA tetraloop structure forms the recognition site for *Saccharomyces cerevisiae* RNase III. *Embo J* 20:7240-7249.
- Xiong Z et Lommel SA. 1989. The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1. *Virology* 171:543-554.
- Xu X, Leo C, Jang Y, Chan E, Padilla D, Huang BC, Lin T, Gururaja T, Hitoshi Y, Lorens JB, Anderson DC, Sikic B, Luo Y, Payan DG et Nolan GP. 2001. Dominant effector genetics in mammalian cells. *Nat Genet* 27:23-29.
- Xue F et Cooley L. 1993. kelch encodes a component of intercellular bridges in *Drosophila* egg chambers. *Cell* 72:681-693.
- Yedavalli VS, Neuveut C, Chi YH, Kleiman L et Jeang KT. 2004. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* 119:381-392.
- Yelverton E, Lindsley D, Yamauchi P et Gallant JA. 1994. The function of a ribosomal frameshifting signal from human immunodeficiency virus-1 in *Escherichia coli*. *Mol Microbiol* 11:303-313.
- Yu ET, Zhang Q et Fabris D. 2005. Untying the FIV frameshifting pseudoknot structure by MS3D. *J Mol Biol* 345:69-80.
- Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P et Yu XF. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* 302:1056-1060.
- Yusupova GZ, Yusupov MM, Cate JH et Noller HF. 2001. The path of messenger RNA through the ribosome. *Cell* 106:233-241.

- Zerfass K et Beier H. 1992. Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. *Nucleic Acids Res* 20:5911-5918.
- Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC et Gao L. 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424:94-98.

