

Université de Montréal

**Étude des mécanismes non-conventionnels de traduction  
chez le virus de l'immunodéficience humaine de type 1 et  
le virus de l'hépatite C**

par  
Martin Baril

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

29 avril, 2005

© Martin Baril, 2005



W .

4

U58

2005

v. 136

**Direction des bibliothèques**

**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 des mécanismes non-conventionnels de traduction chez le virus de  
l'immunodéficience humaine de type 1 et le virus de l'hépatite C

présentée par :  
Martin Baril

a été évaluée par un jury composé des personnes suivantes :

Dr Nikolaus Heveker, président-rapporteur  
Dr Léa Brakier-Gingras, directeur de recherche  
Dr Hugo Soudeyns, membre du jury  
Dr Denis Leclerc, examinateur externe



## Résumé

Le virus de l'immunodéficience humaine de type 1 (VIH-1) et le virus de l'hépatite C (VHC) sont responsables de deux des pandémies mondiales les plus importantes, affectant respectivement 40 et 200 millions de personnes. Mes travaux de recherche au cours de mon doctorat ont permis de caractériser des mécanismes non-conventionnels de traduction chez ces deux virus.

Un de ces mécanismes non-conventionnels de traduction est le changement de cadre de lecture programmé (*frameshift*)  $-1$ . Chez le VIH-1, la synthèse de Gag (le précurseur des protéines structurales du virus) et de Pol (le précurseur des enzymes virales) se fait à partir du même ARN viral pleine longueur, mais la séquence codante de Pol est dans un cadre de lecture  $-1$  relativement à la séquence codante de Gag. L'expression de Pol dépend d'un *frameshift*  $-1$  qui permet à une minorité de ribosomes ayant initié la traduction au codon d'initiation de Gag d'éviter le codon d'arrêt de la protéine Gag et de synthétiser la protéine de fusion Gag-Pol. L'efficacité de *frameshift*  $-1$  contrôle donc le rapport Gag-Pol/Gag qui est critique pour l'assemblage des particules rétrovirales, la réplication et l'infectivité du VIH-1. Ce *frameshift*  $-1$  nécessite la présence de deux éléments dans l'ARN viral : une séquence glissante, où les ribosomes glissent d'un nucléotide en direction 5', suivie d'un signal stimulateur formant une structure secondaire spécifique qui contrôle l'efficacité de *frameshift*  $-1$ . Notre groupe a démontré que le signal stimulateur 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 en Amérique et en Europe) adopte une structure en tige-boucle irrégulière où deux tiges sont séparées par une protubérance de trois purines. Mes travaux ont montré que le signal stimulateur du *frameshift*  $-1$  du VIH-1 de tous les autres sous-types du groupe M se replie de la même façon que le sous-type B, malgré des différences dans la séquence du signal stimulateur. Nous avons aussi montré que l'efficacité de *frameshift*  $-1$  des différents sous-types du groupe M est très conservée, la variation maximale étant de l'ordre de 35% par rapport à l'efficacité moyenne correspondant à ces sous-types. Nos résultats confirment

l'importance du rapport Gag-Pol/Gag pour la réplication virale et valident le *frameshift* -1 comme cible potentielle pour le développement de nouveaux agents anti-VIH.

Nous avons également caractérisé la structure du signal stimulateur du *frameshift* -1 chez le groupe O du VIH-1, un autre groupe qui se retrouve principalement en Afrique centrale. La séquence de la région du *frameshift* -1 du groupe O ne permet pas à son signal stimulateur d'adopter la structure en tige-boucle irrégulière retrouvée chez le groupe M. Par des études de mutagenèse et des analyses structurales, nous avons démontré que ce signal stimulateur adopte une structure en pseudonoeud où la boucle coiffant une tige de huit paires de bases forme une deuxième tige de huit paires de bases avec une séquence complémentaire en aval. Nos résultats, en démontrant la différence de structure des signaux stimulateurs de *frameshift* -1 des groupes M et O du VIH-1, appuient l'hypothèse phylogénétique proposant que les groupes M et O proviennent de transmissions indépendantes du chimpanzé à l'homme. Cette différence structurale souligne l'importance de développer des agents anti-*frameshift* spécifiques au groupe O du VIH-1.

Mes recherches ont également porté sur le mécanisme menant à la synthèse d'une protéine du VHC, la protéine F, dont l'existence a récemment été découverte et qui semble jouer un rôle important dans la pathogenèse des patients infectés par le VHC. Les études publiées sur la synthèse de cette protéine rapportaient des résultats contradictoires. La principale étude avait montré, à l'aide d'un système de traduction *in vitro*, que la protéine F était synthétisée par un *frameshift* +1 effectué par une minorité de ribosomes au niveau d'une suite de 10 adénines lors de la traduction de l'ARN viral, la majorité des ribosomes synthétisant la polyprotéine virale. Toutefois, on sait depuis longtemps que les séquences répétitives, comme cette suite de 10 adénines, induisent des *frameshifts* -1 et +1 non-spécifiques dans des systèmes de traduction *in vitro*. Nos études dans des cellules de mammifères en culture ont démontré que la protéine F n'est pas synthétisée par un *frameshift* +1, mais par une initiation de la traduction directement dans le cadre de lecture +1 par rapport à la polyprotéine du VHC. Par des études de mutagenèse, nous avons

localisé le site d'initiation de la traduction de la protéine F au codon non-AUG chevauchant le codon 26 de la polyprotéine (codon 26+1), qui est GUG ou GCG dépendamment des variants viraux. À l'aide d'un inhibiteur spécifique de la traduction, nous avons également démontré que l'initiation de la traduction de la protéine F s'effectue avec l'ARNt initiateur chargé de méthionine (Met-ARNt<sub>i</sub><sup>Met</sup>), utilisé habituellement pour initier la traduction, et ce, que le codon 26+1 soit GUG ou GCG. Des cas d'initiation de la traduction à un codon GUG, qui diffère d'AUG par un nucléotide, ont déjà été observés, mais l'initiation de la traduction de la protéine F au codon GCG, qui diffère d'AUG par deux nucléotides, révèle pour la première fois une initiation de la traduction où un Met-ARNt<sub>i</sub><sup>Met</sup> décode un codon qui diffère d'AUG par plus d'un nucléotide.

À long terme, nos études sur les mécanismes non-conventionnels de traduction chez le VIH-1 et le VHC pourront contribuer au développement de nouveaux agents anti-VHC et anti-VIH dirigés contre des cibles virales encore inexploitées.

**Mots-clés :** Ribosome eucaryote, traduction, *frameshift*, initiation de la traduction à un codon non-AUG, VIH-1, VHC, protéine F

## Abstract

Human immunodeficiency virus type 1 (HIV-1) and Hepatitis C virus (HCV) are responsible for two major viral pandemics and around 40 and 200 million people, respectively, are infected with these viruses worldwide. During the course of my Ph.D. studies, I investigated unconventional translation mechanisms used by these two viruses.

One of these unconventional translation mechanisms is a programmed  $-1$  ribosomal frameshift that is used by HIV-1. In HIV-1, Gag, the precursor of the viral structural proteins, and Pol, the precursor of the viral enzymes, are synthesized from the same RNA, i.e. the full-length viral RNA, but the coding sequence of Pol is in a  $-1$  reading frame relative to the coding sequence of Gag. The majority of ribosomes translate the viral RNA according to conventional rules and synthesize Gag, but a minority of ribosomes that initiated translation at the initiation codon of Gag shift the reading frame at a specific sequence before they encounter the stop codon of Gag and extend the translation to the *pol* gene, synthesizing Gag-Pol. The expression of Pol is thus regulated by a  $-1$  frameshift and the efficiency of this  $-1$  frameshift event controls the Gag-Pol to Gag ratio, which is critical for particle assembly, replication and viral infectivity. This  $-1$  frameshift requires two *cis*-acting elements in the viral mRNA: a slippery sequence, where ribosomes slip by one nucleotide in the 5' direction, and a downstream stimulatory signal, which is a specific structure that controls the  $-1$  frameshift efficiency. Our group had previously demonstrated that the  $-1$  frameshift stimulatory signal of the subtype B of group M of HIV-1 (the group M is predominant worldwide and the subtype B is the most widespread in America and Europe) is a two-stem helix, in which a three-purine bulge interrupts the two stems. Our studies showed that the  $-1$  frameshift stimulatory signal of all the other subtypes of group M can be folded into the same structure as in subtype B, despite sequence variations. We also showed that the efficiency of the  $-1$  frameshift of every subtype of group M falls within a narrow window, the maximal deviation from the mean value of the efficiency of these subtypes being 35%. Our results confirmed the importance of the Gag-Pol to Gag

ratio in viral replication and validate the  $-1$  frameshift event as a potential target for the development of novel anti-HIV agents.

We also characterized the structure of the  $-1$  frameshift stimulatory signal of the group O of HIV-1, a group found in Central Africa. The sequence in the  $-1$  frameshift region of group O RNA does not allow the formation of the two-stem helix structure acting as a frameshift stimulatory signal in group M. Using mutagenesis and structural approaches, we showed that the stimulatory signal of group O adopts a pseudoknot structure. In this structure, eight nucleotides of a ten-nucleotide loop, capping an eight base-pair stem, base-pair with a downstream complementary sequence. Our results, in demonstrating the structural divergence between the  $-1$  frameshift stimulatory signals of group M and O of HIV-1, support the phylogenetic hypothesis that these groups originate from a different chimpanzee to human transmission. This structural divergence stresses the need to develop anti-frameshift agents specific for HIV-1 group O.

We also investigated the translation mechanism leading to the synthesis of a HCV viral protein, the F protein, which was recently discovered and could be implicated in the development of the pathogenesis in patients infected with HCV. Contradictory results that describe the mechanism accounting for the synthesis of the F protein have been published. Among them, a major study showed that the F protein was synthesized *via* a  $+1$  frameshift by a minority of ribosomes that initiated translation at the AUG codon for the viral polyprotein on a stretch of 10A encompassing codons 9 to 11 of the polyprotein. However, these results were obtained with an *in vitro* translation system and repetitive sequences, like this stretch of 10A, are known to promote unspecific  $-1$  and  $+1$  frameshifts *in vitro*. With translation assays in mammalian cultured cells, we demonstrated that the F protein is not synthesized by a  $+1$  frameshift, but results from direct initiation of translation in the  $+1$  reading frame relative to the polyprotein of HCV. Using a mutagenesis approach, we localized the translation initiation site of the F protein at a non-AUG codon overlapping codon 26 of the polyprotein (codon 26+1). This codon is GUG or GCG, depending on the

viral variants. Using a specific inhibitor of translation, we also demonstrated that the initiation of translation of the F protein was made with the standard initiator tRNA charged with methionine (Met-tRNA<sub>i</sub><sup>Met</sup>), whether codon 26+1 is GUG or GCG. Translation initiation at a GUG codon, which differs from AUG by one nucleotide, has already been reported. However, the translation initiation of the F protein at a GCG codon described in our study is the first report of a translation initiation with Met-tRNA<sub>i</sub><sup>Met</sup> at a codon that differs from AUG by more than one nucleotide.

In conclusion, our studies on unconventional translation mechanisms in HIV-1 and HCV provide important information that could contribute to the development of novel anti-HIV and anti-HCV agents directed against these new viral targets.

**Keywords :** Eukaryotic ribosome, translation, frameshift, initiation of translation at non-AUG codon, HIV-1, HCV, F protein

## Remerciements

Je veux remercier tout particulièrement le Dr Léa Brakier-Gingras qui a été une directrice de recherche exemplaire. Son soutien infatigable dans mes projets de recherche, ainsi que ses efforts soutenus dans la correction de mes travaux écrits et oraux m'ont permis de repousser mes limites et de m'améliorer sans cesse durant les cinq merveilleuses années passées en sa compagnie. Avec le recul, je peux affirmer que je n'aurais pas pu trouver d'environnement plus souhaitable pour la réalisation de mon doctorat.

Je tiens également à remercier tous les membres du laboratoire qui ont créé un environnement de travail dynamique et joyeux. Un merci particulier à Dominic Dulude avec qui j'ai eu la chance de travailler sur plusieurs projets de recherche. J'ai également eu la chance de collaborer avec les Dr S. Steinberg, H. Soudeyns et G. Lemay qui ont gentiment accepté de partager leur expertise avec moi.

Je suis également très reconnaissant à mes parents et à ma belle-famille pour leur soutien qui a grandement facilité la poursuite de mes études durant toutes ces années. Un gros merci aussi à Anik qui est à mes côtés à tous les jours et qui donne un sens à ma vie.

Je suis très heureux et très fier d'avoir effectué mon doctorat au Département de biochimie de l'Université de Montréal. La qualité de la recherche et l'entraide entre les différents groupes de recherche retrouvées à l'intérieur du Département de biochimie devraient servir de modèle aux autres unités de recherche.

Merci à tous encore une fois pour ces merveilleuses années.

*À Anik*



## Table des matières

Liste des tableaux .....	xiii
Liste des figures .....	xiv
Liste des abréviations .....	xviii
Chapitre I - Introduction.....	1
1.1 Mise en contexte .....	2
1.2 Les étapes de la traduction .....	3
1.2.1 L'initiation de la traduction.....	6
1.2.1.1 L'initiation coiffe-dépendante.....	6
1.2.1.2 L'initiation coiffe-indépendante .....	10
1.2.2 L'élongation de la traduction .....	17
1.2.2.1 La sélection d'aminoacyl-ARNt .....	17
1.2.2.2 La formation du lien peptidique.....	19
1.2.2.3 La translocation.....	19
1.2.3 La terminaison de la traduction.....	20
1.3 Les mécanismes non-conventionnels de traduction chez les eucaryotes .....	22
1.3.1 Le changement de cadre de lecture programmé ( <i>frameshift</i> ).....	26
1.3.1.1 Le <i>frameshift</i> programmé -1 .....	26
1.3.1.1.1 Mécanisme du <i>frameshift</i> programmé -1.....	27
1.3.1.1.2 Structure des signaux stimulateurs dans le <i>frameshift</i> programmé -1 .....	33
1.3.1.2 Le <i>frameshift</i> programmé +1.....	36
1.3.2 L'initiation de la traduction à un codon non-AUG .....	41
1.3.2.1 L'initiation de la traduction Met-ARNt <sub>i</sub> <sup>Met</sup> -dépendante à un codon non-AUG.....	41
1.3.2.2 L'initiation de la traduction Met-ARNt <sub>i</sub> <sup>Met</sup> -indépendante à un codon non-AUG .....	42
1.4 Le virus de l'immunodéficience humaine de type 1 (VIH-1).....	45
1.4.1 Les différents groupes et sous-types du VIH-1 .....	47

1.4.2	Le génome du VIH-1 .....	51
1.4.3	Cycle de réplication du VIH-1 .....	54
1.4.4	Le <i>frameshift</i> programmé -1 du VIH-1 .....	56
1.5	Le virus de l'hépatite C (VHC).....	59
1.5.1	Le génome du VHC .....	62
1.5.2	Le cycle de réplication du VHC.....	65
1.5.3	La synthèse de la protéine F du VHC .....	67
1.6	Description du projet de recherche .....	71
Chapitre II - Efficiency of a programmed -1 ribosomal <i>frameshift</i> in the different subtypes of the human immunodeficiency virus type 1 group M.....		74
Chapitre III - The <i>frameshift</i> stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot.....		98
Chapitre IV - Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein .....		129
Chapitre V - Discussion .....		166
5.1	Efficacité du <i>frameshift</i> -1 chez les différents sous-types du groupe M du VIH-1 .....	167
5.2	Le signal stimulateur du <i>frameshift</i> -1 du groupe O du VIH-1 est un pseudonoeud... ..	171
5.3	La protéine F du VHC est synthétisée après initiation de la traduction à un codon non-AUG dans un cadre de lecture +1 par rapport à la polyprotéine virale .....	177
5.4	Conclusions .....	186
Bibliographie.....		187

## Liste des tableaux

Tableau 2-1 : Frameshift efficiency of different mutants of subtype B and of all subtypes of HIV-1 group M .....	93
Tableau 3-1 : Frameshift efficiency for different mutants of group O of HIV-1 .....	122

## Liste des figures

Figure 1-1 : Structure des ribosomes 80S et 70S .....	5
Figure 1-2 : Les étapes de l'initiation de la traduction eucaryote .....	9
Figure 1-3 : Exemples de structure secondaire d'ARN pour les IRES viral et cellulaire	13
Figure 1-4 : Exemples de mécanismes d'initiation interne de la traduction utilisant des facteurs d'initiation différents .....	14
Figure 1-5 : Structure de la sous-unité 40S seule (A à C) et de la sous-unité 40S liée à l'IRES du VHC (D à F) obtenue par cryomicroscopie .....	15
Figure 1-6 : Structure secondaire des IRES de la région intergénique de deux virus d'insecte .....	16
Figure 1-7 : Le cycle d'élongation de la traduction chez les eucaryotes .....	18
Figure 1-8 : La terminaison de la traduction chez les eucaryotes .....	21
Figure 1-9 : Exemples de mécanismes non-conventionnels de traduction .....	25
Figure 1-10 : Mécanisme du glissement simultané en 5' du ribosome et des deux ARNt proposé pour le <i>frameshift</i> programmé -1 .....	29
Figure 1-11 : Raffinements proposés pour expliquer le <i>frameshift</i> programmé -1 .....	32
Figure 1-12 : Structure tertiaire de pseudonoeuds stimulant le <i>frameshift</i> programmé -1	35
Figure 1-13 : Différents mécanismes proposés pour le <i>frameshift</i> programmé +1 .....	39
Figure 1-14 : Mécanisme de sélection des aminoacyl-ARNt au site A du ribosome .....	40

Figure 1-15 : Mécanisme d'initiation de la traduction Met-ARNt <sub>i</sub> <sup>Met</sup> -indépendant .....	44
Figure 1-16 : Vue d'ensemble de la diversité génétique et de l'épidémiologie des différents groupes et sous-types du VIH-1 .....	50
Figure 1-17 : Représentation du génome et du virion du VIH-1 .....	53
Figure 1-18 : Cycle de réplication du VIH-1 .....	55
Figure 1-19 : Structures secondaires proposées pour le signal stimulateur du <i>frameshift</i> -1 du VIH-1 .....	58
Figure 1-20 : Mécanisme d'inhibition de l'initiation de la traduction par PKR .....	61
Figure 1-21 : Représentation du génome, du virion et des réplicons du VHC .....	64
Figure 2-1 : Vectors expressing luciferase used to study the programmed -1 ribosomal frameshift of HIV-1 group M <i>in vitro</i> and in cultured cells .....	94
Figure 2-2 : Relative frameshift efficiency of mutants of subtype B of HIV-1 group M	95
Figure 2-3 : Secondary structure of the frameshift stimulatory signal of all the subtypes of HIV-1 group M .....	96
Figure 2-4 : Relative frameshift efficiency of all subtypes of HIV-1 group M .....	97
Figure 3-1 : Predicted structure for the frameshift region of subtype MVP5180 of HIV-1 group O .....	123
Figure 3-2 : Description of the luciferase expression vectors used for the study of the programmed -1 ribosomal frameshift of HIV-1 group O <i>in vitro</i> and in cultured cells .....	124

Figure 3-3 : Effect of different mutations in the frameshift region of subtype MVP5180 of HIV-1 group O on the frameshift efficiency .....	125
Figure 3-4 : Effect on the frameshift efficiency of mutations impairing formation of stem 2 of the pseudoknot of subtype MVP5180 of HIV-1 group O .....	126
Figure 3-5 : Probing of the structure proposed for the frameshift stimulatory signal of subtype MVP5180 of HIV-1 group O .....	127
Figure 3-6 : Stereo view of the computer modeled structure for the pseudoknot of subtype MVP5180 of HIV-1 group O .....	128
Figure 4-1 : Sequence and secondary structure of the HCV IRES and beginning of the HCV coding sequence (nt 1-510 of genotype 1a) .....	156
Figure 4-2 : Description of the luciferase vectors used for the study of the expression of the HCV F protein <i>in vitro</i> and in cultured cells .....	157
Figure 4-3 : Synthesis of the F protein with constructs containing insertions of different length of the HCV coding sequence .....	158
Figure 4-4 : Synthesis of the F protein when a stop codon is introduced in the reading frame of the polyprotéine .....	159
Figure 4-5 : Synthesis of the polyprotein and the F protein when the AUG start codon of the polyprotein or its context is mutated .....	160
Figure 4-6 : Identification of the translation start site of the F protein .....	161
Figure 4-7 : Confirmation of the location of the translation start site of the F protein ....	162
Figure 4-8 : Further details on the initiation of translation of the F protein .....	163

Figure 4-9 : Inhibition of the synthesis of the polyprotein and the F protein by edeine ..	164
Figure 5-1 : Structure secondaire de signaux stimulateurs de <i>frameshift</i> -1 formant un pseudonoeud et une troisième tige-boucle .....	176
Figure 5-2 : Efficacité d'initiation de la traduction de la protéine F du VHC lorsque le codon 26+1 est remplacé par différents codons .....	184
Figure 5-3 : Effet de la surexpression de la protéine <i>core</i> du VHC sur la synthèse de la polyprotéine et de la protéine F .....	185

## Liste des abréviations

4E-BP	Protéine se liant à eIF4E
A	Adénine
ARFP	<i>Alternate reading frame protein</i>
ARN	Acide ribonucléique
ARNi	Interférence de l'ARN
ARNm	ARN messenger
ARNr	ARN ribosomique
ARNt	ARN de transfert
ATP	Adénosine triphosphate
BWYV	Virus de la jaunisse occidentale de la betterave
BYDV	Virus de la jaunisse nanisante de l'orge
C	Cytosine
CMV	Cytomégalovirus
CRF	Forme recombinante circulante
CrPV	Virus de la paralysie du criquet
<i>Edr</i>	<i>Embryonal carcinoma differentiation regulated</i>



eEF	Facteur d'élongation eucaryote
eIF	Facteur d'initiation eucaryote
EMCV	Virus de l'encéphalomyocardite
eRF	Facteur de terminaison eucaryote
FMDV	Virus de la fièvre aphteuse ( <i>foot and mouth disease virus</i> )
<i>Frameshift</i>	Changement de cadre de lecture programmé
G	Guanine
G418	Généticine
GFP	Protéine fluorescente verte
GTP	Guanosine triphosphate
HAART	Trithérapie antirétrovirale
HDL	Lipoprotéine à haute densité
HTLV-II	Virus humain de la leucémie à cellules T de type II
INF- $\alpha$	Interféron-alpha
IRE	Élément de réponse aux ions fer
IRES	Site d'initiation interne de la traduction
IRP	Protéine de contrôle de la concentration en ions fer
ISDR	Région impliquée dans la sensibilité à l'IFN

LDL	Lipoprotéine de faible densité
LTR	<i>Long terminal repeats</i>
Met-ARN <sub>t</sub> <sup>Met</sup>	ARNt-méthionine initiateur
MHCI	Complexe majeur d'histocompatibilité de classe I
MMTV	Virus de la tumeur mammaire de la souris
MLV	Virus de la leucémie murine de Moloney
NRTI	Inhibiteur nucléoside de la transcriptase réverse
NNRTI	Inhibiteur non-nucléoside de la transcriptase réverse
NS	Protéine non-structurale
ODC	Ornithine décarboxylase
PABP	Protéine liant la queue de poly(A) à l'extrémité 3' de l'ARNm
PEG	Polyéthylène glycol
PEMV	Virus de la mosaïque avec énation du pois
PI	Inhibiteur de protéase
PIC	Complexe de préintégration
PKR	Protéine kinase ARN-dépendante
PSIV	Virus intestinal de <i>Plautia stali</i>
PTB	<i>Polypyrimidine tract-binding protein</i>

RE	Réticulum endoplasmique
<i>Readthrough</i>	Suppression programmée d'un codon d'arrêt
RFP	Protéine fluorescente rouge
RMN	Résonance magnétique nucléaire
RRL	Lysat de réticulocytes de lapin
RSV	Virus du sarcome de Rous
SRAS	Syndrome respiratoire aigu sévère
SBP2	Protéine liant SECIS
SECIS	Élément d'insertion de la sélénocystéine
SIV	Virus de l'immunodéficience simienne
SRV-1	Rétrovirus simien de type 1
TMEV	Virus de Theiler
TSV	Virus du syndrome de Taura
U	Uracile
UTR	Région non-traduite
VHC	Virus de l'hépatite C
VIH-1	Virus de l'immunodéficience humaine de type 1

# **Chapitre I**

## **Introduction**

## 1.1 Mise en contexte

L'émergence de pandémies mondiales a marqué les dernières décennies. Deux des principales pandémies ont été causées par le virus de l'immunodéficience humaine de type 1 (VIH-1) et le virus de l'hépatite C (VHC), des virus aux génomes hypervariables qui établissent des persistances en échappant au système immunitaire.

Dans le cas du VIH-1, plus de 20 années de recherches intensives ont permis de développer plusieurs traitements antiviraux, sous forme d'inhibiteurs spécifiques pour les différentes enzymes virales. Ces traitements permettent de réduire le titre viral chez les patients infectés, augmentant ainsi leur espérance de vie, mais aucun de ces traitements ne permet la prévention de l'infection ou encore l'élimination totale du virus qui permettrait une guérison. De plus, des souches virales résistantes à un ou plusieurs des agents anti-VIH présentement utilisés ont fait leur apparition, ce qui souligne la nécessité de développer de nouveaux agents anti-VIH qui agiraient sur de nouvelles cibles du cycle de réplication du VIH-1. Une de ces nouvelles cibles est un mécanisme non-conventionnel de traduction utilisé par le VIH-1 pour exprimer le gène *pol* codant pour ses enzymes. Ce mécanisme consiste en un changement de cadre de lecture programmé (*frameshift*)  $-1$  et sera décrit en détail dans les sections 1.3.1.1 et 1.4.4. Notre contribution à l'étude du mécanisme de *frameshift*  $-1$  chez le VIH-1 sera présentée dans les articles constituant les Chapitres II et III.

Pour ce qui est du VHC, les 15 années de recherche depuis sa découverte n'ont pas permis le développement de traitements spécifiques contre le VHC, principalement en raison de l'absence d'un système de culture cellulaire permettant l'infection de cellules et la relâche de virions infectieux. Ce problème technique empêche non seulement le développement de traitements anti-VHC, mais retarde également la caractérisation du cycle de réplication et des moyens utilisés par le VHC pour déjouer le système immunitaire et permettre la persistance virale. Récemment, une nouvelle protéine, la protéine F, a été identifiée chez le VHC. La particularité de cette protéine F est d'être codée dans le cadre de lecture  $+1$  par rapport au cadre de lecture de la polyprotéine générant toutes les autres protéines du VHC. Quoique le rôle de la protéine F soit encore inconnu, une fonction

importante dans la réplication du virus *in vivo* est envisageable pour cette protéine, puisque les protéines virales dont l'expression est régulée au niveau de la traduction jouent habituellement des rôles essentiels dans la réplication des virus. Le mécanisme régissant l'expression de la protéine F dans le cadre de lecture +1 a été un sujet de controverse dont l'historique sera décrit à la section 1.5.3 et les réponses que nous avons apportées à cette question se trouvent dans l'article constituant le Chapitre IV.

Mon projet de recherche au cours de mon doctorat m'a amené à caractériser des mécanismes non-conventionnels de traduction à l'étape de l'initiation pour le VHC et à l'étape de l'élongation pour le VIH-1. À long terme, ces études fondamentales pourront contribuer au développement de nouveaux agents anti-VHC et anti-VIH dirigés contre des cibles virales encore inexploitées.

Dans cette Introduction, je décrirai les différentes étapes de la traduction ainsi que les mécanismes non-conventionnels de traduction chez les eucaryotes. En raison de leur importance pour la compréhension des articles présentés aux Chapitres II, III et IV, une attention particulière sera portée aux mécanismes de *frameshift* -1 et +1 ainsi qu'à l'initiation de la traduction à un codon non-AUG. Je présenterai également un survol de l'organisation du génome et du cycle de réplication du VIH-1 et du VHC, les virus dans lesquels se produisent les phénomènes non-conventionnels de traduction que nous avons étudiés.

## 1.2 Les étapes de la traduction

La traduction des ARN messagers (ARNm) en protéines est contrôlée par un complexe ribonucléoprotéique nommé ribosome. Pour effectuer cette traduction des ARNm en protéines, le ribosome utilise comme substrat les ARN de transfert (ARNt) chargés d'acide aminé. Le ribosome eucaryote, nommé 80S d'après sa constante de sédimentation, est constitué de deux sous-unités, la petite (40S) et la grande (60S). La petite sous-unité

contient de l'ARN ribosomique (ARNr) 18S et environ 30 protéines ribosomiques, alors que la grande sous-unité est composée des ARNr 5S, 5.8S et 28S ainsi que d'une cinquantaine de protéines ribosomiques. Du point de vue structural, les informations sur le ribosome eucaryote sont limitées comparativement au ribosome bactérien. Des études récentes de cryomicroscopie ont permis de visualiser le ribosome eucaryote à une résolution de 15 Å (Figure 1-1A) (Spahn et al. 2001a). Quoique cette résolution ne permette pas une visualisation à une résolution atomique, les contours du ribosome sont clairement définis. Sur la sous-unité 40S, on distingue le corps qui est séparé de la tête par l'épaule, ainsi que les régions formant le bec et la plate-forme, alors que la protubérance centrale et le bras L1 sont visibles sur la sous-unité 60S (Figure 1-1A). Cette structure obtenue par cryomicroscopie peut ensuite être superposée avec la structure cristallographique (Figure 1-1B) à 5.5 Å de résolution du ribosome bactérien (70S) (Yusupov et al. 2001). L'utilisation conjointe de ces deux structures permet de mieux comprendre le mécanisme de chacune des étapes de la traduction : l'initiation, l'élongation et la terminaison.

Figure 1-1 : Structure des ribosomes 80S et 70S.

A) Structure du ribosome 80S de *Saccharomyces cerevisiae* à une résolution de 15 Å obtenue par microscopie cryo-électronique (adapté de Spahn et al. 2001a).

La petite sous-unité ribosomique eucaryote (40S) est représentée en jaune et la grande sous-unité ribosomique eucaryote (60S) est représentée en bleu. Les régions qui sont présentes chez le ribosome 80S, mais non chez le ribosome 70S sont identifiées en or pour la sous-unité 40S et en mauve pour la sous-unité 60S. On distingue également l'ARNt lié au site P du ribosome (en vert).

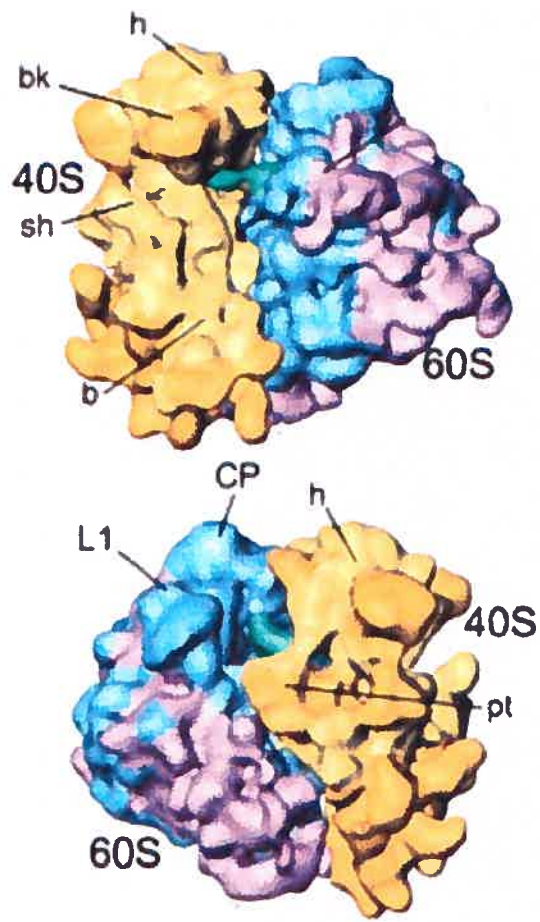
h : tête, bk : bec, sh : épaule, b : corps, pt : plate-forme, CP : protubérance centrale, L1 : bras L1.

B) Structure du ribosome 70S de *Thermus thermophilus* à une résolution de 5.5 Å obtenue par cristallographie (adapté de Yusupov et al. 2001).

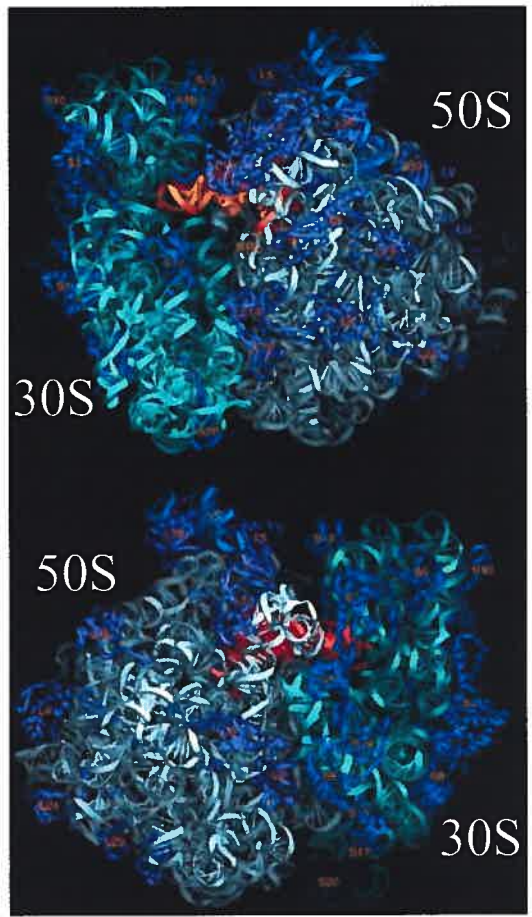
L'ARNr 16S de la petite sous-unité ribosomique procaryote (30S) est représenté en vert. Les ARNr 23S et 5S de la grande sous-unité ribosomique procaryote (50S) sont représentés en gris et en bleu pâle, respectivement. Les protéines ribosomiques des sous-unités 30S et 50S sont représentées en bleu foncé et en mauve, respectivement. L'ARNt au site A est représenté en orange et l'ARNt au site E est représenté en rouge.



**A**



**B**



## 1.2.1 L'initiation de la traduction

### 1.2.1.1 L'initiation coiffe-dépendante

L'initiation de la traduction coiffe-dépendante, qui est utilisée pour la majorité des ARNm eucaryotes, se produit en quatre étapes distinctes faisant intervenir un minimum de 12 facteurs d'initiation (eIF). Ces étapes sont : la formation du complexe de préinitiation 43S, la liaison du complexe de préinitiation 43S à la coiffe de l'extrémité 5' de l'ARNm, le balayage (*scanning*) de la région 5' non-traduite (UTR) de l'ARNm jusqu'au codon d'initiation et l'association des sous-unités ribosomiques 40S et 60S pour former le ribosome complet (revue par Gray et Wickens 1998; Pestova et Hellen 2001; Poulin et Sonenberg 2003; Preiss et Hentze 2003; Merrick 2004).

Pour que le complexe de préinitiation 43S se forme, la première étape fait intervenir eIF3 et eIF1A, qui se lient à la sous-unité 40S, ainsi que eIF6 (Ceci et al. 2003), qui lie la sous-unité 60S, empêchant la réassociation prématurée des sous-unités 40S et 60S. En parallèle, il y a formation d'un complexe ternaire entre eIF2, une molécule de guanosine triphosphate (GTP) et l'ARNt initiateur chargé de méthionine (Met-ARNt<sup>Met</sup>). Le complexe de préinitiation 43S est ensuite formé par la liaison de ce complexe ternaire à la sous-unité 40S, liaison qui est facilitée par eIF3, eIF1 et eIF1A (Figure 1-2A). Dans le complexe de préinitiation 43S, le Met-ARNt<sup>Met</sup> se retrouve au site P de la sous-unité 40S.

Au niveau de l'ARNm à traduire, il y a assemblage du complexe eIF4F (formé de eIF4A, eIF4E et eIF4G). Cet assemblage débute par la liaison de la protéine d'échafaudage eIF4G à eIF4A, une ARN hélicase ATPase, et à eIF4E, qui assure le recrutement du complexe à la coiffe (m<sup>7</sup>GpppN) située à l'extrémité 5' de l'ARNm. Une fois assemblé à la coiffe, eIF4F recrute eIF4B et eIF4H qui vont stimuler l'activité hélicase de eIF4A. Cette activité hélicase utilise l'énergie libérée par l'hydrolyse de l'adénine triphosphate (ATP) pour défaire les structures secondaires de l'ARNm (Rogers et al. 2001). Le complexe eIF4F

recrute ensuite le complexe de préinitiation 43S à l'extrémité 5' de l'ARNm. Ce recrutement se fait principalement via l'interaction entre eIF4G et eIF3 (Figure 1-2B).

En plus de sa liaison avec eIF4E, eIF4G interagit avec la protéine PABP qui lie la queue de poly(A) à l'extrémité 3' de l'ARNm. L'interaction simultanée de eIF4G avec eIF4E (qui lie l'extrémité 5' de l'ARNm) et PABP (qui lie l'extrémité 3' de l'ARNm) circularise l'ARNm. Cette circularisation de l'ARNm permettrait aux ribosomes qui ont terminé un cycle de traduction de réinitier efficacement un nouveau cycle de traduction.

La liaison entre les facteurs eIF4E et eIF4G est contrôlée par le degré de phosphorylation d'une protéine qui peut se lier à eIF4E (4E-BP). Lorsque 4E-BP est hypophosphorylée, elle séquestre eIF4E et empêche la liaison à eIF4G, bloquant ainsi le recrutement du complexe de préinitiation 43S sur l'ARNm et, par le fait même, l'initiation de la traduction. En réponse à certains signaux extracellulaires, 4E-BP hyperphosphorylée permet la relâche de eIF4E et la stimulation de l'initiation de la traduction.

Le complexe de préinitiation 43S glisse ensuite vers l'extrémité 3' de l'ARNm jusqu'à ce qu'il rencontre le premier codon AUG, c'est le mécanisme de balayage (Figure 1-2C). Les facteurs eIF1 et eIF1A sont nécessaires au positionnement du complexe de préinitiation 43S sur un codon AUG (Pestova et al. 1998), alors que eIF3 coordonne l'interaction entre eIF1 et eIF5 pour inhiber l'initiation à un codon non-AUG (Valasek et al. 2004). Lors de la rencontre d'un codon AUG, la principale interaction permettant la reconnaissance de ce codon initiateur provient des paires de bases Watson-Crick formées entre le codon AUG et l'anticodon  $3'UAC^{5'}$  du Met-ARNt<sub>Met</sub> (complexe 48S, Figure 1-2C). En outre, la séquence flanquant cet AUG initiateur, nommée contexte Kozak, a également un impact sur la reconnaissance du codon initiateur. Le contexte Kozak le plus favorable chez les eucaryotes supérieurs est GCC(A/G)CCAAUGG, où l'AUG initiateur est souligné (Kozak 1986). Dans le contexte Kozak, le résidu en position -3 (le résidu A du AUG étant numéroté +1) est celui qui a le plus d'impact sur la reconnaissance du codon AUG, alors que le résidu en position +4 est le deuxième déterminant en importance.

En raison de la direction 5'→3' du balayage, le premier AUG rencontré à partir de l'extrémité 5' est presque toujours utilisé pour initier la traduction. Par contre, si le premier AUG se retrouve trop près de la coiffe (Kozak 1991) ou encore dans un contexte Kozak défavorable, l'initiation à cet AUG est inefficace, ce qui permettra à une majorité des complexes de préinitiation 43S de continuer le balayage de l'ARNm et d'initier à un autre codon initiateur en aval du premier AUG. Cette situation, nommée *leaky scanning*, permet d'initier la traduction à plus d'un codon initiateur avec des efficacités différentes, menant ainsi à la synthèse de plus d'une protéine à partir d'un même ARNm.

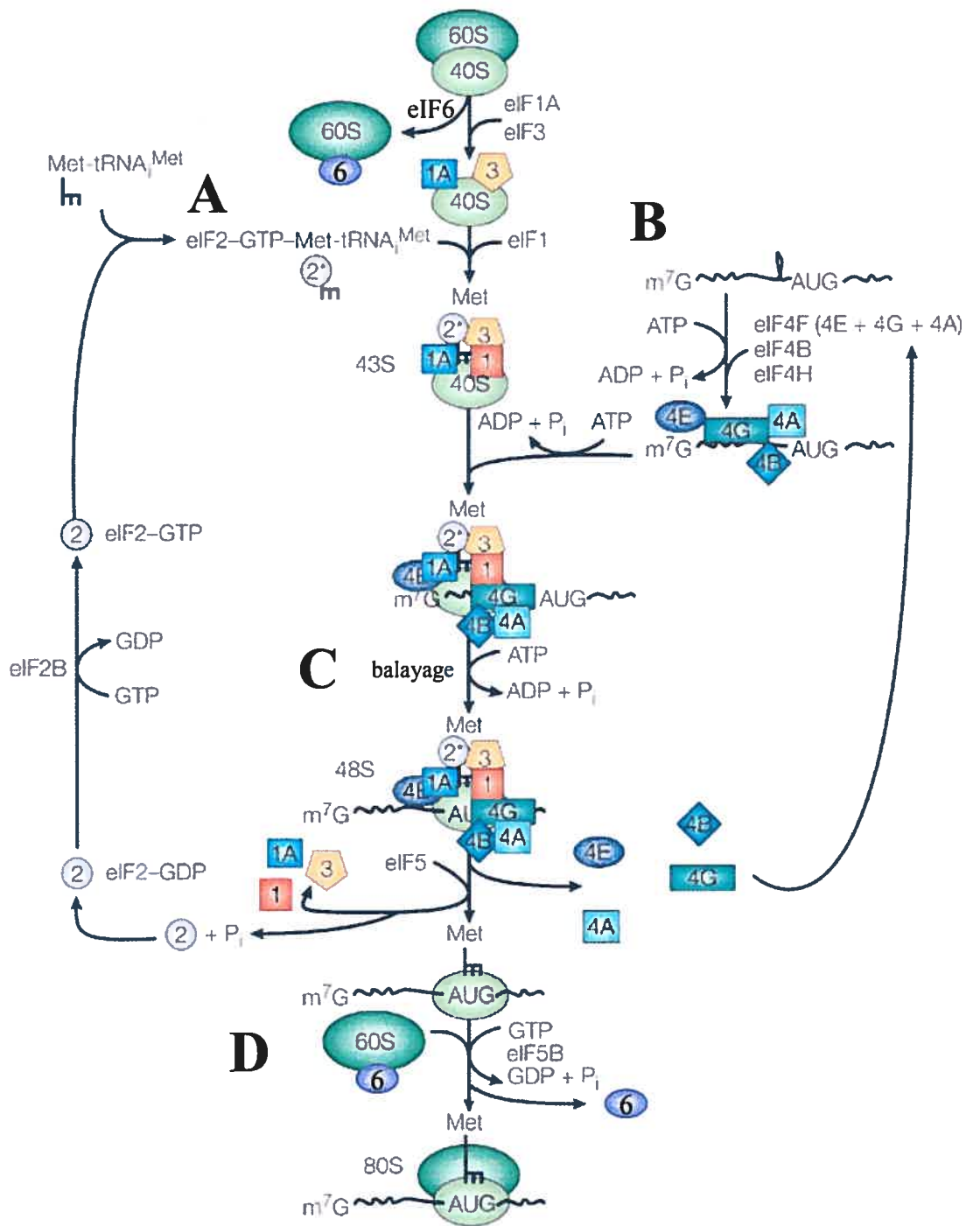
Lors de la reconnaissance du codon initiateur par le complexe de préinitiation 43S, l'interaction entre le Met-ARNt<sup>Met</sup> et le codon initiateur entraînerait un changement de conformation de eIF2 permettant l'hydrolyse du GTP par l'activité GTPase de eIF2, suivie de la relâche de eIF2-GDP et probablement de eIF1, eIF1A et eIF3. L'hydrolyse du GTP est favorisée par la liaison d'un facteur additionnel, eIF5. L'hydrolyse du GTP sert donc de contrôle pour la fidélité de sélection du codon initiateur.

Le facteur eIF2B échange ensuite le GDP lié à eIF2 pour un GTP, permettant à ce facteur de participer à un nouveau cycle d'initiation. En réponse à différents stress cellulaires, eIF2 est phosphorylé, ce qui augmente son affinité pour eIF2B, entraîne la séquestration de ce facteur limitant et empêche le recyclage de eIF2-GDP en eIF2-GTP. La carence en eIF2-GTP va mener à l'inhibition de l'initiation de la traduction.

Ensuite, le facteur d'initiation eIF5B lié au GTP stimule la liaison des sous-unités 60S et 40S et hydrolyse le GTP après la formation du ribosome (Pestova et al. 2000). Cette seconde étape d'hydrolyse du GTP sert donc de contrôle pour l'assemblage du ribosome (Figure 1-2D).

Figure 1-2 : Les étapes de l'initiation de la traduction eucaryote (adapté de Holcik et Sonenberg 2005).

- A) Formation du complexe de préinitiation 43S.
- B) Liaison du complexe de préinitiation 43S à la coiffe de l'extrémité 5' de l'ARNm.
- C) Balayage de la région 5'UTR de l'ARNm jusqu'au codon d'initiation.
- D) Association des sous-unités ribosomiques 40S et 60S pour former le ribosome (80S).



### 1.2.1.2 L'initiation coiffe-indépendante

L'initiation coiffe-indépendante (aussi appelée initiation interne) a été identifiée pour la première fois chez le poliovirus, un membre des picornavirus (Pelletier et Sonenberg 1988). Contrairement aux ARNm cellulaires, l'ARN du poliovirus ne possède pas de coiffe à son extrémité 5'. De plus, sa région 5'UTR est anormalement longue (~750 nucléotides) et contient 6 à 7 codons AUG qui ne sont pas utilisés pour initier la traduction. Dans l'initiation coiffe-indépendante, la région 5'UTR adopte une structure secondaire et tertiaire complexe, appelée site d'initiation interne de la traduction (IRES), qui permet la liaison de la sous-unité 40S sans coiffe (Chen C. Y. et Sarnow 1995). Les IRES sont très utilisés par les virus à ARN, dont le virus de l'hépatite C (Figure 1-3A), mais se retrouvent également chez certains ARN cellulaires impliqués dans la croissance cellulaire, la transcription et l'initiation de la traduction (Figure 1-3B) (revue par Martinez-Salas et al. 2001). Quoique les IRES viraux et cellulaires remplissent les mêmes fonctions, ils ne possèdent pas d'homologie de structure (Figure 1-3).

Pour initier la traduction, la plupart des IRES nécessitent la participation des mêmes eIF que l'initiation coiffe-dépendante, à l'exception de eIF4E (Figure 1-4A). De plus, certains facteurs cellulaires, notamment le facteur de transcription La (Meerovitch et al. 1993; Svitkin et al. 1994) et la *polypyrimidine tract-binding protein* (PTB) (Hellen et al. 1993), augmentent l'efficacité d'initiation interne, probablement en stabilisant la structure secondaire et tertiaire fonctionnelle de l'IRES. Les facteurs liés à l'IRES recrutent alors le complexe de préinitiation 43S, soit directement au niveau de l'AUG initiateur comme chez le virus de la fièvre aphteuse (*foot and mouth disease virus*, FMDV), soit ~150 nucléotides en amont du site d'initiation pour ensuite balayer l'ARN viral jusqu'à l'AUG initiateur comme chez le poliovirus. Le reste de l'initiation se fait de façon identique à l'initiation coiffe-dépendante (Figure 1-2). Le mécanisme d'initiation interne a donc l'avantage d'être indépendant de eIF4E, ce qui permet au virus comme le picornavirus de détourner l'appareil de traduction de la cellule en bloquant spécifiquement la traduction coiffe-

dépendante au profit de la traduction interne, soit en déphosphorylant 4E-BP, soit par clivage de eIF4G ou de PABP (Ohlmann et al. 1996).

D'autres types d'IRES, comme celui du VHC, ne nécessitent aucun des facteurs d'initiation constituant eIF4F (revue par Rijnbrand et Lemon 2000; Kieft et al. 2001). Dans le cas du VHC, le complexe de préinitiation 43S se lie à l'IRES via des interactions directes entre la sous-unité 40S et les domaines II, IIIc, III d et IV de l'IRES (Kolupaeva et al. 2000) et entre eIF3 et le domaine IIIb de l'IRES (Buratti et al. 1998; Sizova et al. 1998) (Figure 1-4B). La structure de l'IRES du VHC lié à la sous-unité 40S a été obtenue par cryomicroscopie et a permis de montrer que l'IRES du VHC interagit avec les régions de la tête et de la plate-forme de la sous-unité 40S (Spahn et al. 2001b). La liaison à l'IRES induit des changements de conformation dans la sous-unité 40S qui se traduisent principalement par le rapprochement de la tête et de l'épaule, permettant la formation du tunnel où passe l'ARNm (Figure 1-5). L'impact de ces changements de conformation dans la sous-unité 40S sur l'initiation de la traduction n'est pas encore connu, mais ils pourraient permettre d'imiter la conformation qu'adopte la sous-unité 40S lorsqu'elle est liée aux facteurs d'initiation classiques. Dans le cas du VHC, la fonction de l'IRES ne se résumerait donc pas seulement à positionner la sous-unité 40S au niveau du site d'initiation de la traduction, mais à lui donner une conformation compétente pour amorcer l'initiation de la traduction.

Finalement, un autre type d'IRES a récemment été identifié chez des virus à ARN(+) simple brin similaires aux picornavirus : le virus de la paralysie du criquet (CrPV) (Wilson et al. 2000), le virus intestinal de *Plautia stali*, également un insecte, (PSIV) (Sasaki et Nakashima 2000) et le virus du syndrome de Taura chez la crevette (TSV) (Cevallos et Sarnow 2005). Ces IRES se retrouvent dans la région intergénique des génomes viraux (~180 nucléotides), soit entre la région codant pour les enzymes et la région codant pour les protéines structurales des virus. L'IRES de la région intergénique permet donc de synthétiser les protéines structurales du virus à partir d'un codon

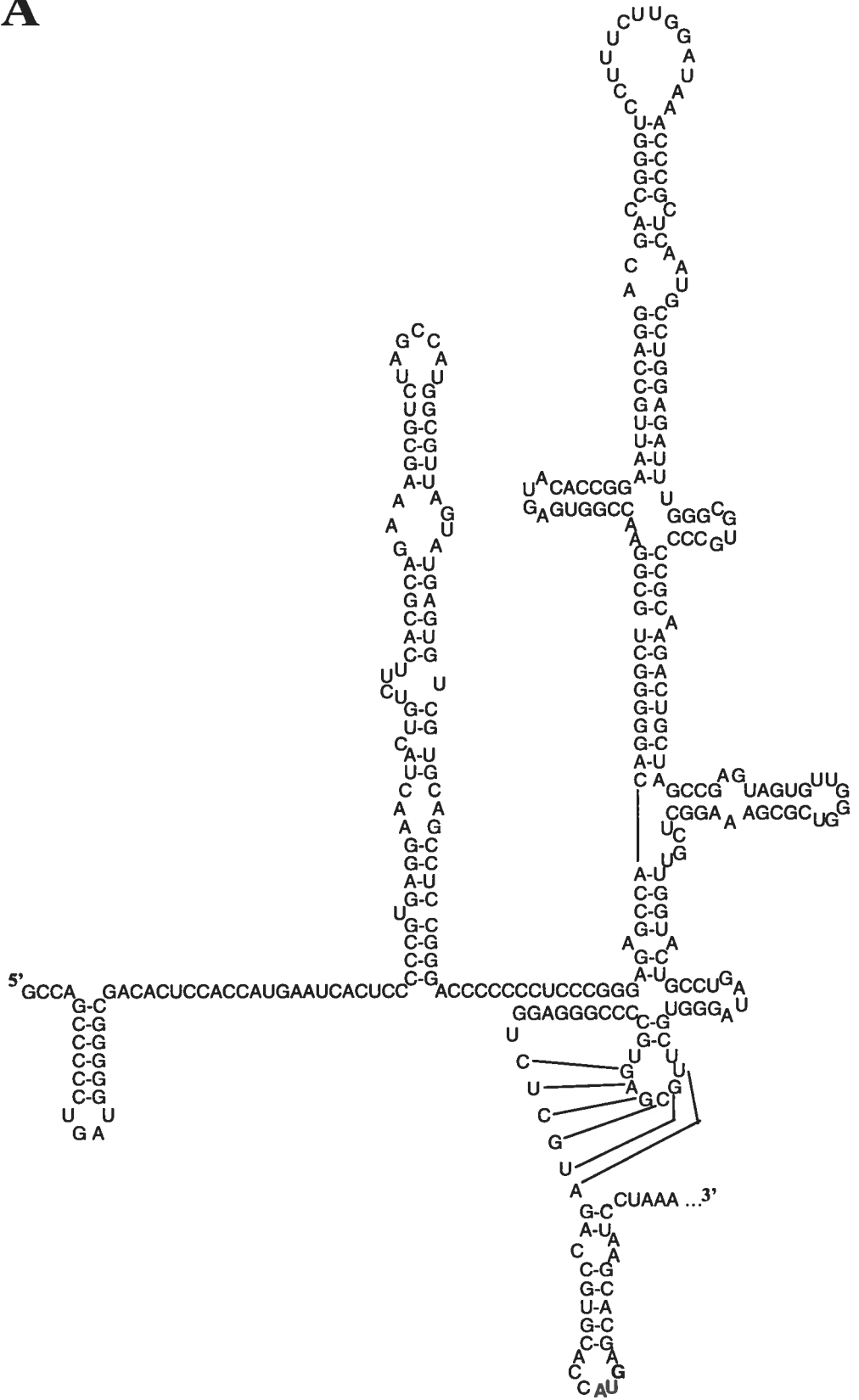


d'initiation non-AUG : GCU pour le CrPV, CAA pour le PSIV et GCU pour le TSV (Figure 1-6). L'initiation de la traduction à partir de ces IRES se produit au site A plutôt qu'au site P du ribosome, puisque le Met-ARN<sub>t</sub><sup>Met</sup> n'est pas utilisé pour initier la traduction. Ce mécanisme inhabituel d'initiation permet donc une initiation indépendante du complexe ternaire (eIF2-GTP-Met-ARN<sub>t</sub><sup>Met</sup>). Ces IRES ont également la particularité de permettre la liaison de la sous-unité 40S, ainsi que son association avec la sous-unité 60S pour former le ribosome sans aucun eIF (Wilson et al. 2000). Le mécanisme permettant cette initiation Met-ARN<sub>t</sub><sup>Met</sup> indépendante sera décrit en détail à la section 1.3.2.2.

Figure 1-3 : Exemples de structure secondaire d'ARN pour les IRES viral et cellulaire.

- A) Structure secondaire de l'IRES de l'ARN génomique du VHC (adapté de Honda et al. 1996). L'AUG initiateur est indiqué en rouge.
- B) Structure secondaire de l'IRES de l'ARNm codant pour *c-myc* (adapté de Le Quesne et al. 2001). L'AUG initiateur est identifié en rouge et les neuf AUG non-utilisés dans la région 5'UTR sont encerclés avec leur nucléotide flanquant en 3'.

A



**B**

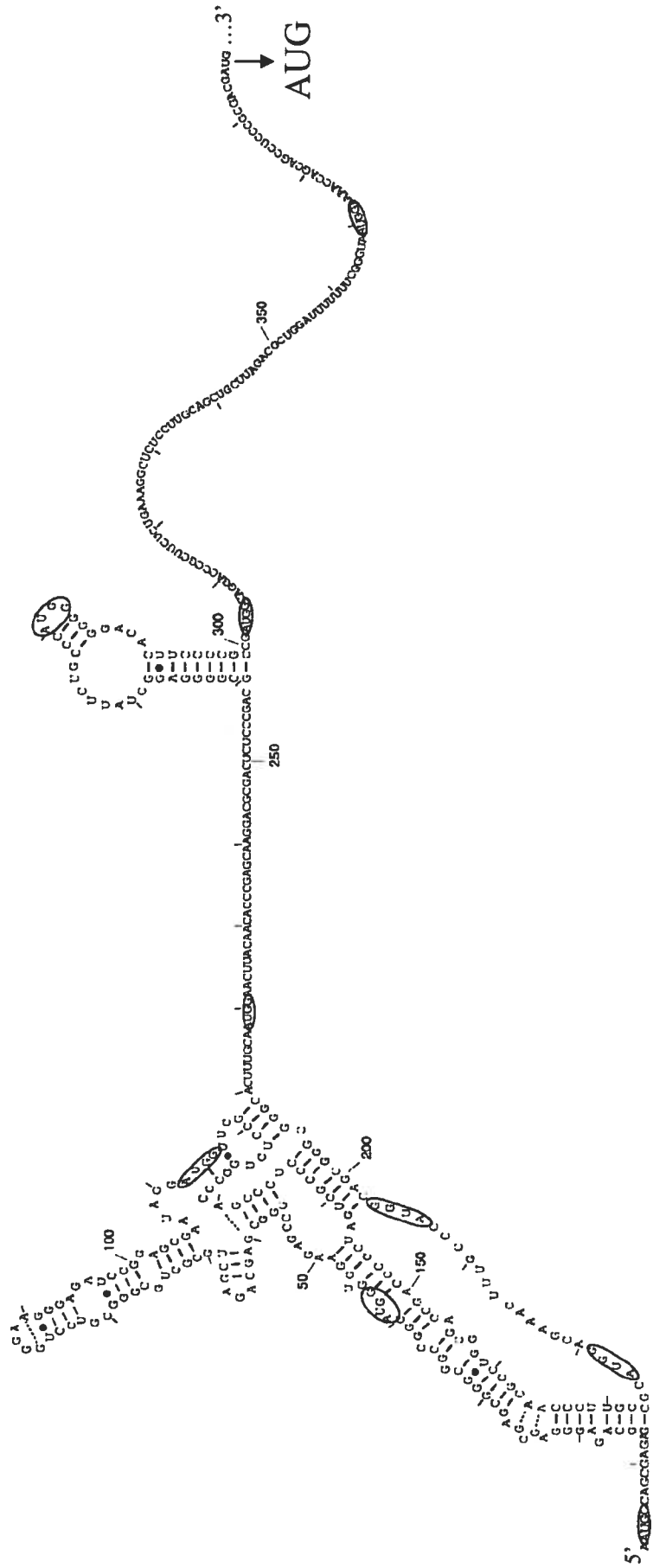
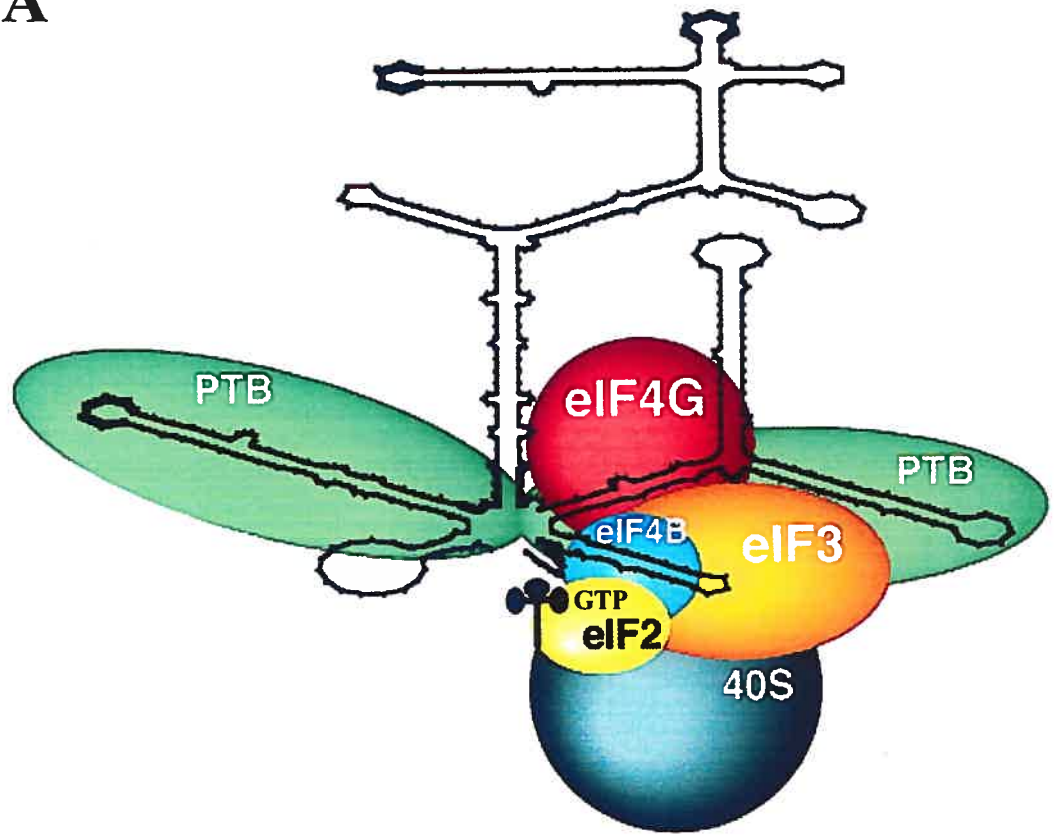


Figure 1-4 : Exemples de mécanismes d'initiation interne de la traduction utilisant des facteurs d'initiation différents (adapté de Martinez-Salas et al. 2001).

- A) Régions d'interactions connues entre l'IRES du FMDV et certaines protéines impliquées dans l'initiation interne de la traduction : PTB, eIF2-GTP-Met-ARN<sub>t</sub><sup>Met</sup>, eIF3, eIF4B et eIF4G.
- B) Interactions entre l'IRES du VHC et les protéines nécessaires à l'initiation interne de la traduction : eIF2-GTP-Met-ARN<sub>t</sub><sup>Met</sup> et eIF3. Les domaines II, IIIa, IIIb, IIIc, IIId, IIIe et IV de l'IRES du VHC sont identifiés en rouge.

**A**



**B**

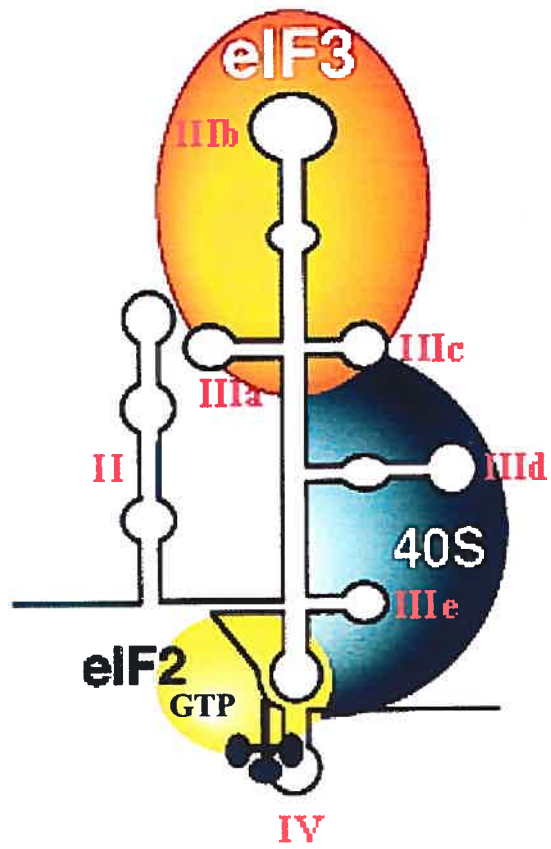


Figure 1-5 : Structure de la sous-unité 40S seule (A à C) et de la sous-unité 40S liée à l'IRES du VHC (D à F) obtenue par cryomicroscopie (adapté de Spahn et al. 2001b).

La région correspondant à l'IRES du VHC est identifiée en mauve. L'entrée (*entry*) et la sortie (*exit*) du tunnel où passe l'ARNm sont indiquées. L'emplacement des hélices 16, 18 et 34 (h16, h18 et h34) de l'ARNr 18S est indiqué. h : tête, bk : bec, sh : épaule, b : corps, pt : plate-forme.

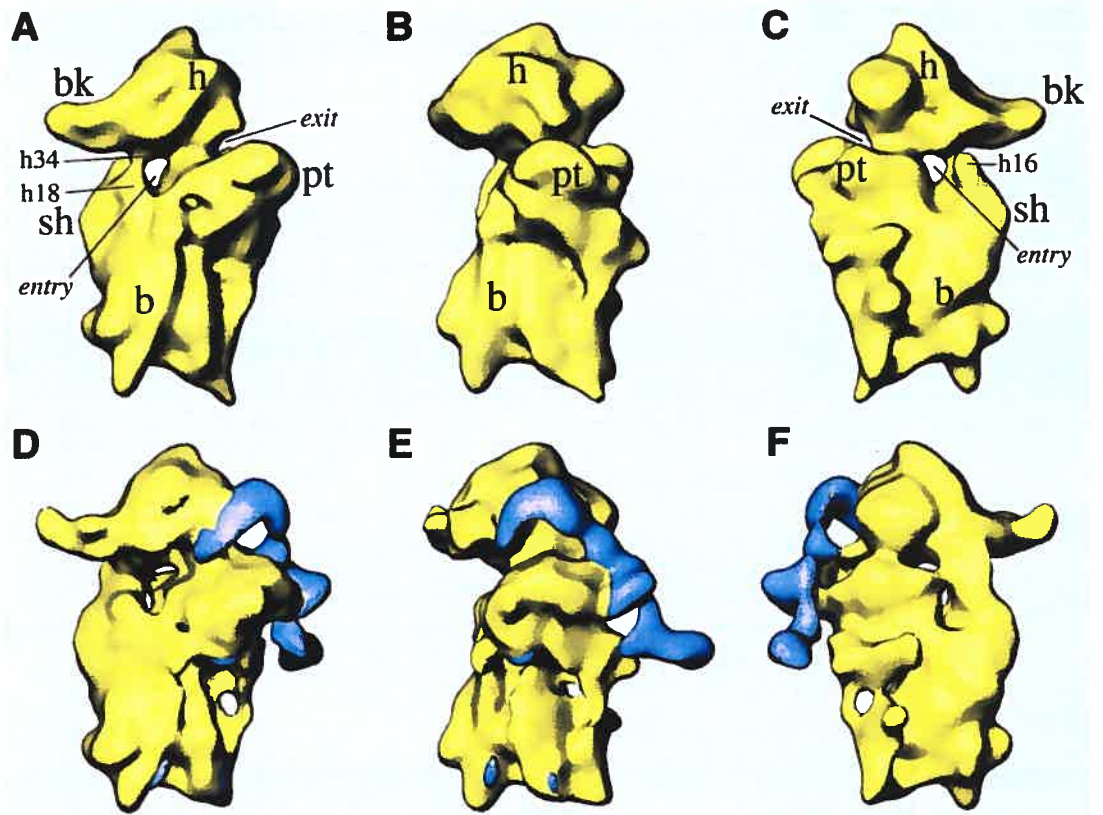
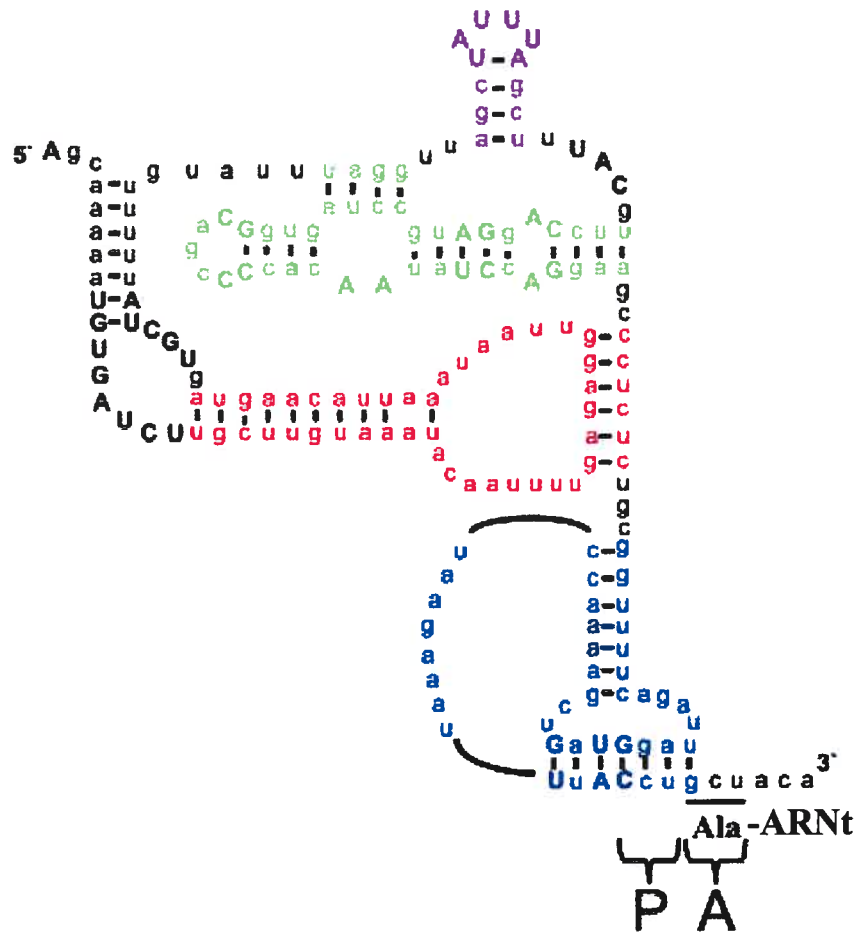




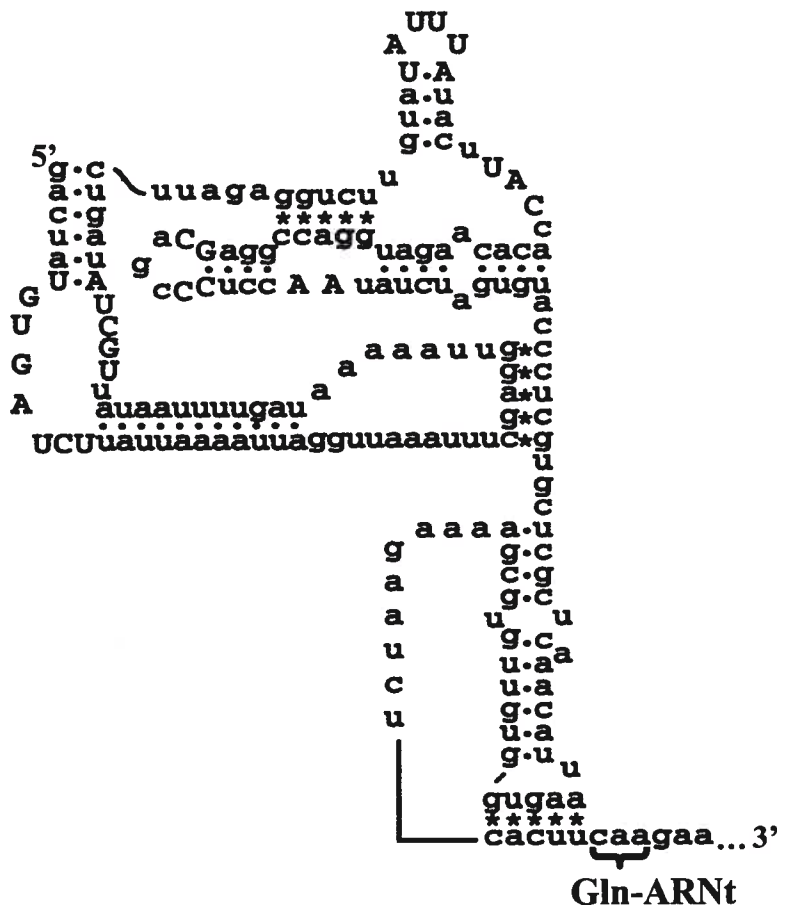
Figure 1-6 : Structure secondaire des IRES de la région intergénique de deux virus d'insecte.

- A) Structure secondaire de l'IRES du CrPV (adapté de Jan et Sarnow 2002).  
Le codon GCU est reconnu par un Ala-ARNt qui y initie la traduction de la capsid.
  
- B) Structure secondaire de l'IRES du PSIV (adapté de Shibuya et al. 2003).  
Le codon CAA est reconnu par un Gln-ARNt qui y initie la traduction de la capsid.

**A**



**B**



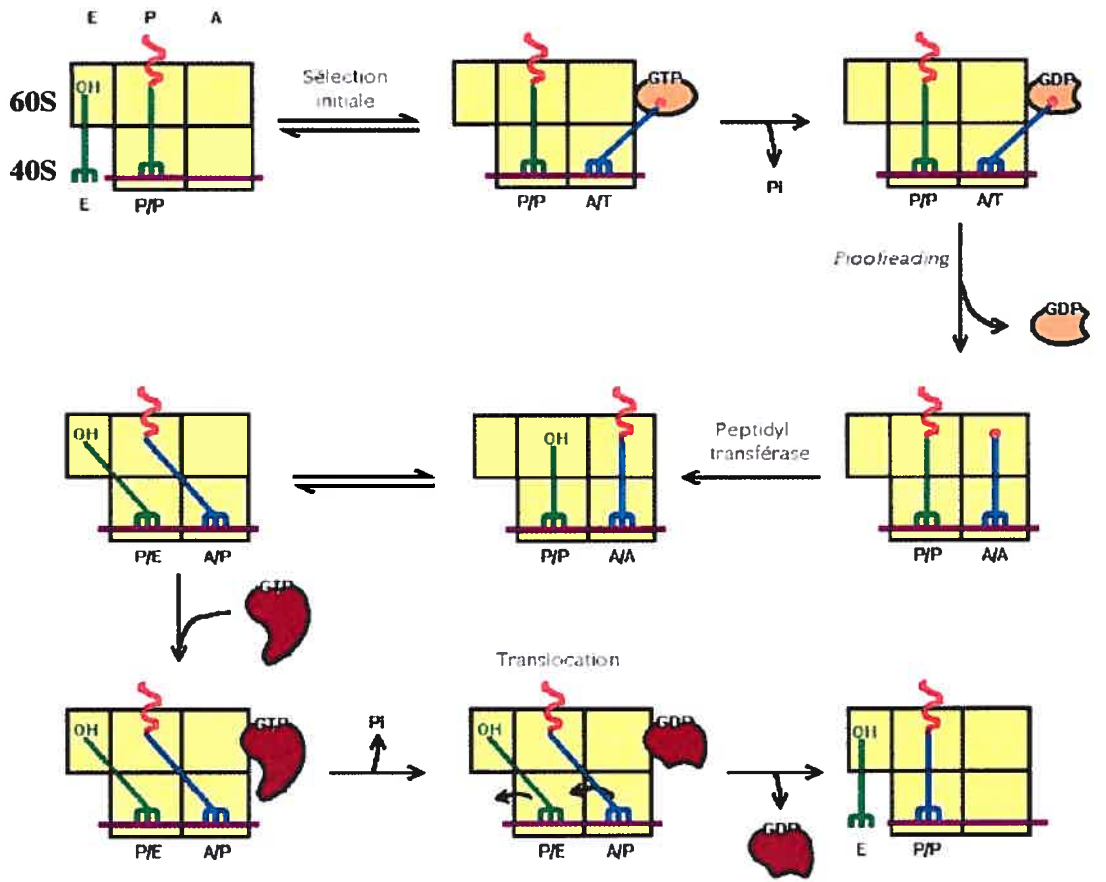
## 1.2.2 L'élongation de la traduction

Une fois le ribosome formé avec le Met-ARNt<sub>i</sub><sup>Met</sup> au site P, le cycle d'élongation de la traduction peut débuter (revue par Merrick et Nyborg 2000; Kapp et Lorsch 2004). Le cycle d'élongation se divise en trois étapes : la sélection d'aminoacyl-ARNt, la formation du lien peptidique et la translocation (Figure 1-7).

### 1.2.2.1 La sélection d'aminoacyl-ARNt

La première étape du cycle d'élongation de la traduction consiste à sélectionner un aminoacyl-ARNt dont l'anticodon est complémentaire au codon de l'ARNm occupant le site A du ribosome. Cette étape s'effectue avec l'aide du facteur d'élongation eEF1A qui est lié au GTP et à un aminoacyl-ARNt, formant un complexe ternaire. L'interaction entre le complexe ternaire et le ribosome permet à l'anticodon de l'aminoacyl-ARNt lié à eEF1A de se positionner au site A/T du ribosome (c'est-à-dire que l'aminoacyl-ARNt occupe le site A sur la sous-unité 40S et le site d'entrée, dit T, sur la sous-unité 60S). Si l'interaction codon-anticodon est incorrecte, le complexe eEF1A•GTP•aminoacyl-ARNt se détache du ribosome et un autre complexe prend sa place. Lors de la formation d'une interaction codon-anticodon correcte, eEF1A subit un changement conformationnel qui active son activité GTPase, permettant l'hydrolyse du GTP en GDP. Après cette hydrolyse, eEF1A•GDP est relâché du ribosome, mais l'aminoacyl-ARNt subit une étape supplémentaire de sélection (*proofreading*). Cette étape permet de rejeter certains aminoacyl-ARNt qui forment un lien codon-anticodon presque parfait (deux paires de bases sur trois) et qui n'ont pas été éliminés à la première étape de sélection. Le rejet de l'aminoacyl-ARNt lors du *proofreading* doit se faire avant son accommodation au site A/T au site A du ribosome. La double vérification de la liaison codon-anticodon permet au ribosome de minimiser le nombre d'erreurs d'incorporation à  $10^{-4}$  par codon.

Figure 1-7 : Le cycle d'élongation de la traduction chez les eucaryotes.  
Le facteur eEF1A est représenté en orange et le facteur eEF2 est représenté en brun.



Le complexe eEF1A•GDP lie ensuite eEF1B qui permet l'échange de la molécule de GDP pour une molécule de GTP, suivie de la relâche de eEF1B. Le complexe eEF1A•GTP peut alors lier un nouvel aminoacyl-ARNt qui sera utilisé dans un autre cycle d'élongation de la traduction. Il est important de noter que le seul aminoacyl-ARNt qui ne peut pas être lié à eEF1A est le Met-ARNt<sub>i</sub><sup>Met</sup> en raison de sa structure tertiaire particulière (Drabkin et al. 1998). Cette particularité de l'ARNt<sub>i</sub><sup>Met</sup> permet l'usage exclusif de la Met-ARNt<sub>i</sub><sup>Met</sup> pour l'initiation de la traduction via sa liaison à eIF2 (voir section 1.2.1.1).

### 1.2.2.2 La formation du lien peptidique

Une fois que l' aminoacyl-ARNt se trouve accommodé au site A du ribosome, la réaction de transpeptidation se produit immédiatement. Cette réaction très rapide est catalysée par le centre peptidyl transférase situé dans l'ARNr 28S de la sous-unité 60S du ribosome. La réaction consiste en une attaque nucléophile par l' aminoacyl-ARNt au site A du ribosome sur le groupement ester qui relie la chaîne peptidique à l'ARNt au site P. Cette réaction de transpeptidation permet le transfert de la chaîne peptidique de l'ARNt au site P sur l' aminoacyl-ARNt au site A. Le mécanisme qui permet au ribosome de stimuler cette réaction est sujet à controverses et il semble actuellement que son rôle serait uniquement de positionner correctement les deux ARNt pour permettre une réaction de transpeptidation spontanée.

### 1.2.2.3 La translocation

L'étape qui suit la formation du lien peptidique est la translocation, qui consiste au déplacement relatif de l'ARNm lié aux ARNt de trois nucléotides en direction 5' par rapport au ribosome. Après la transpeptidation l'ARNt libéré de la chaîne peptidique et le peptidyl-ARNt vont osciller spontanément entre les sites P et A et les sites hybrides P/E (où l'ARNt désacylé occupe le site P sur la sous-unité 40S et le site de sortie, dit E, sur la sous-unité 60S) et A/P (où le peptidyl-ARNt occupe le site A sur la sous-unité 40S et le site P

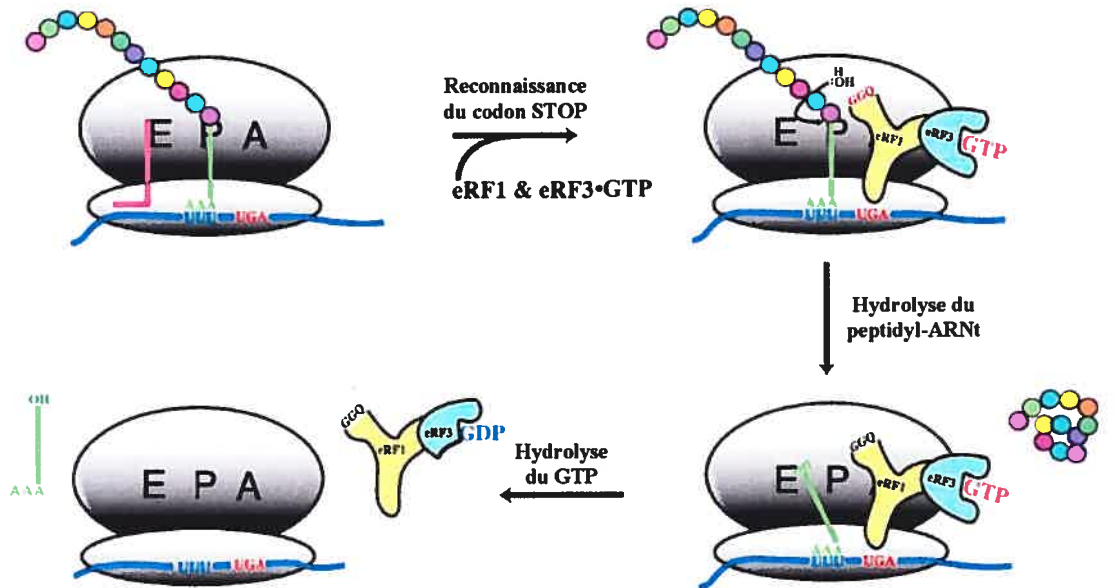
sur la sous-unité 60S). Cet équilibre va être déplacé vers les site P/E et A/P par la liaison de eEF2•GTP au ribosome. La translocation est complétée lorsque eEF2•GTP hydrolyse le GTP en GDP, ce qui entraînerait un changement conformationnel de eEF2, suivie de la relâche de eEF2•GDP. Lors de la translocation, l'ARNt désacylé passe du site P/E au site E et le peptidyl-ARNt passe du site A/P au site P, laissant le site A libre pour le prochain cycle d'élongation. L'ARNt désacylé au site E sera libéré lors de la liaison d'un nouvel aminoacyl-ARNt au site A/T du ribosome.

### 1.2.3 La terminaison de la traduction

La terminaison de la traduction est la dernière étape d'un cycle de traduction et se produit lorsque le site A du ribosome en élongation rencontre un des trois codons d'arrêt (UAA, UAG ou UGA) sur l'ARNm. La nature du nucléotide immédiatement en 3' du codon d'arrêt influence l'efficacité de la reconnaissance du codon d'arrêt, une purine permettant une efficacité de terminaison de 8 à 60 fois supérieure à une pyrimidine (revue par Tate et al. 1995). Il existe deux facteurs de terminaison eucaryote (eRF) : eRF1 et eRF3 qui forment un complexe en interagissant via leur région C-terminale. Le facteur eRF1 reconnaît le codon d'arrêt sur le messager, qu'il soit UAA, UAG ou UGA, alors que eRF3 est une GTPase lié au GTP (Figure 1-8). Le complexe eRF1/eRF3 se lie au ribosome en présence d'un codon d'arrêt et catalyse le transfert du groupement peptidyle du peptidyl-ARNt du site P sur une molécule d'eau, libérant la protéine dont la synthèse est terminée. La libération de la protéine est accompagnée de la relâche de l'ARNt désacylé et de l'hydrolyse du GTP en GDP. Le complexe eRF•GDP est ensuite relâché du ribosome, le ribosome quitte l'ARNm et se dissocie en sous-unités 40S et 60S qui vont initier un autre cycle complet de traduction.

Figure 1-8 : La terminaison de la traduction chez les eucaryotes (adapté de Kapp et Lorsch 2004).





## 1.3 Les mécanismes non-conventionnels de traduction chez les eucaryotes

Les eucaryotes utilisent plusieurs mécanismes non-conventionnels durant l'initiation, l'élongation ou la terminaison de la traduction pour augmenter la diversité ou encore réguler l'expression génétique. Parmi les principaux mécanismes non-conventionnels de traduction, on rencontre le changement de cadre de lecture programmé (*frameshift*), le parcours sans lecture d'un ou de plusieurs codons lors du balayage (*shunt*) ou de l'élongation (*bypassing*) ou la redéfinition d'un codon. Cette section se concentre sur ces mécanismes particuliers qui peuvent se produire durant le balayage, l'élongation et la terminaison de la traduction, alors que les anomalies de l'initiation de la traduction seront traitées à la section 1.3.2.

Le *frameshift* peut se produire soit d'un nucléotide en direction 5' de l'ARNm (*frameshift* -1, décrit en détail à la section 1.3.1.1), soit d'un nucléotide en direction 3' de l'ARNm (*frameshift* +1, décrit en détail à la section 1.3.1.2).

Le mécanisme du *shunt* a été observé pour la première fois chez le virus de la mosaïque du chou-fleur (Futterer et al. 1993). Le *shunt* consiste en un balayage discontinu de l'ARNm par la sous-unité ribosomique 40S qui parcourt des régions de l'ARNm riches en structures secondaires sans les défaire. Les codons d'initiation de la traduction présents ou insérés artificiellement dans ces régions ne sont pas utilisés par la sous-unité ribosomique 40S pour initier la traduction (revue par Ryabova et al. 2002). Le *shunt* peut se produire lors du balayage suivant l'initiation de la traduction à la coiffe de l'ARNm et précédant l'initiation de la traduction à un codon initiateur (Latorre et al. 1998) ou lors du balayage suivant la terminaison de la traduction d'un court cadre de lecture et la réinitiation de la traduction à un deuxième codon initiateur de la traduction en aval (Ryabova et Hohn 2000). Le mécanisme du *shunt* permet donc d'initier la traduction à des codons initiateurs

en aval de codons AUG non utilisés pour initier la traduction, qui ne pourraient pas être reconnus lors d'un mécanisme de balayage traditionnel. Cette caractéristique du *shunt* rappelle l'initiation interne de la traduction à partir d'un IRES, à la différence que le *shunt* nécessite la présence de la coiffe du facteur d'initiation eIF4G intacte (Giorgi et Kolakofsky 1984). Comme la plupart des mécanismes de traduction non-conventionnels, le *shunt* a principalement été caractérisé chez les virus (Futterer et al. 1993; Latorre et al. 1998; Sen et al. 2004), mais un exemple vient d'être décrit pour l'ARNm de la beta-sécrétase chez l'humain (Rogers et al. 2004), où le mécanisme du *shunt* pourrait potentiellement être impliqué dans le développement de l'Alzheimer.

De son côté, le mécanisme du *bypassing* a été caractérisé chez le gène 60 du bactériophage T4, qui code pour une sous-unité de la topoisomérase. Ce mécanisme consiste en ce qu'une région de 50 nucléotides au milieu de l'ARNm correspondant au gène 60 n'est pas traduite (Huang et al. 1988; revue par Herr et al. 2000). La région non-traduite est délimitée par un codon de décollage (GAA), où le peptidyl-ARNt au site P du ribosome va se dissocier de l'ARNm, et un codon synonyme (GAA) qui sert de site d'atterrissage, où la traduction va reprendre. La région non-traduite forme une tige-boucle qui contient un codon d'arrêt dont la présence stimule le *bypassing*. Environ la moitié des ribosomes en élongation parcourent cette région sans la traduire, tandis que les autres quittent l'ARNm (Herr et al. 2001) (Figure 1-9A). La séquence de la protéine synthétisée par la région en 5' du site de décollage stimule également l'efficacité du *bypassing* (Weiss et al. 1990). Contrairement au *bypassing* du gène 60 du bactériophage T4 dont l'efficacité est contrôlée par plusieurs éléments, d'autres exemples de *bypassing* ne semblent requérir qu'une pause de la traduction, qui est induite par la présence d'un codon dont l'aminoacyl-ARNt est rare au site A du ribosome (Gallant J. A. et Lindsley 1998). Le ribosome parcourt alors une région de l'ARNm sans la traduire jusqu'à la rencontre d'un codon correspondant à l'anticodon du peptidyl-ARNt au site P du ribosome (Gallant J. et al. 2003).

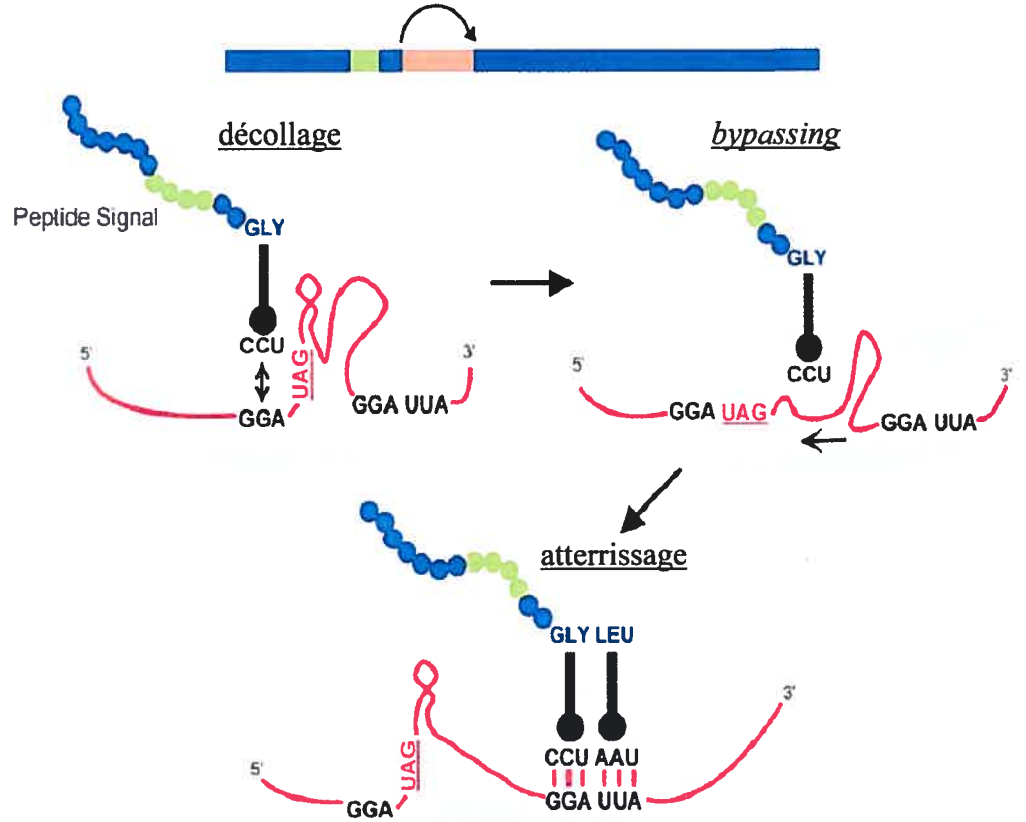
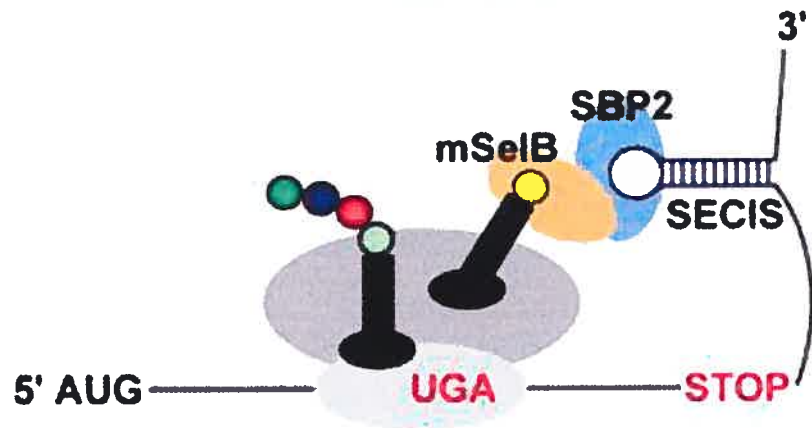
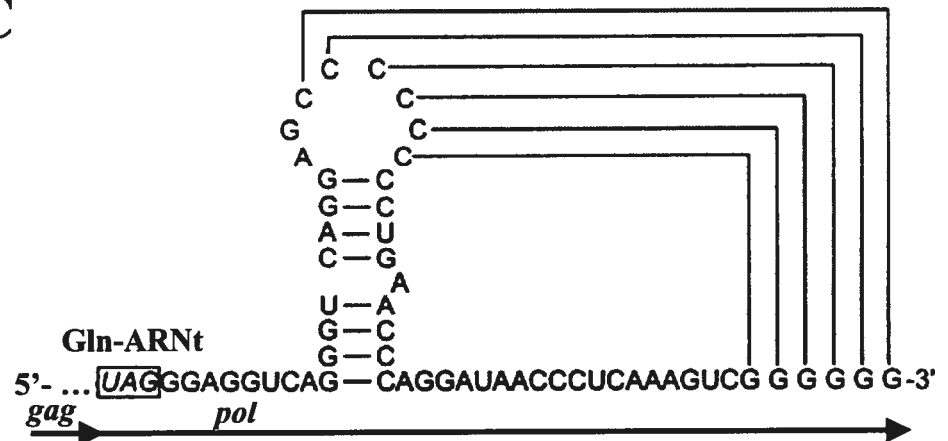
Finalement, plusieurs mécanismes différents existent pour la redéfinition d'un codon, qui permet de lire un codon d'arrêt comme un codon sens. Par exemple, une cinquantaine d'ARNm de mammifères possèdent une structure particulière, nommée élément d'insertion de la sélénocystéine (SECIS), dans leur région 3'UTR. Cette structure permet la liaison d'un facteur d'élongation spécialisé, mSelB (ou eEFsec), par l'entremise d'une protéine adaptatrice, SBP2 (protéine liant SECIS). Le facteur mSelB se lie à un ARNt spécialisé chargé de sélénocystéine, le 21<sup>ième</sup> acide aminé, qui peut être accommodé au site A du ribosome à un codon UGA (Figure 1-9B) (Walczak et al. 1996; Behne et Kyriakopoulos 2001; Namy et al. 2004).

Un codon d'arrêt suivi d'une structure secondaire en pseudonoeud, c'est-à-dire une tige-boucle où la boucle interagit via des liaisons Watson-Crick avec une région en 3' pour former une deuxième tige, peut également être reconnu comme un codon sens. Ce mécanisme est nommé suppression programmée d'un codon d'arrêt (*readthrough*) et permet à une portion des ribosomes de ne pas arrêter la traduction au codon d'arrêt et de poursuivre la traduction de l'extrémité 3' de l'ARNm. Le cas le mieux caractérisé de *readthrough* est celui du virus de la leucémie murine de Moloney (MLV). Lors de la traduction de l'ARNm de ce rétrovirus, 5-10% des ribosomes décodent le codon d'arrêt UAG, marquant la fin du précurseur des protéines structurales du virus (Gag), à l'aide du Gln-ARNt, permettant la synthèse des enzymes virales (Pol) sous forme de protéine d'une fusion Gag-Pol (Figure 1-9C) (Wills et al. 1994; Alam et al. 1999).

Quoique ces mécanismes soient utilisés dans l'expression de certains gènes chez les mammifères, les mécanismes non-conventionnels de traduction sont principalement caractérisés chez les virus eucaryotes qui utilisent plusieurs de ces mécanismes afin de maximiser l'information génétique contenue dans leur génome de taille limitée.

Figure 1-9 : Exemples de mécanismes non-conventionnels de traduction.

- A) *Bypassing* d'une portion de l'ARNm du gène 60 du bactériophage T4 (adapté de Herr et al. 2000).
- B) Incorporation de la sélénocystéine (tiré de Namy et al. 2004).  
SECIS : élément d'insertion de la sélénocystéine, mSelB : facteur d'élongation spécialisé, SBP2 : protéine liant SECIS.
- C) Région de *readthrough* de MLV (adapté de Beier et Grimm 2001).  
*gag* : gène codant pour les protéines structurales du virus, *pol* : gène codant pour les enzymes virales.

**A****By-passing du gène 60 du bactériophage T4****B****C**

### 1.3.1 Le changement de cadre de lecture programmé (*frameshift*)

La synthèse de protéines fonctionnelles est essentielle au fonctionnement et à la survie cellulaire. Il est donc primordial que la traduction des ARNm par les ribosomes se fasse de manière adéquate, en évitant les erreurs d'élongation, soit des erreurs d'incorporation (*missense*), des erreurs de terminaison (*nonsense*) et des changements du cadre de lecture (*frameshift*) (revue par Farabaugh et Bjork 1999). Les erreurs de *frameshift* sont les plus rares avec une probabilité d'erreur inférieure à  $10^{-5}$  par codon (Kurland 1992). Ce taux d'erreurs de *frameshift* peut cependant être grandement augmenté par la présence de structures secondaires particulières dans l'ARNm (revue par Brierley et Pennel 2001; Ivanov et al. 2003). Dans ces cas précis où le *frameshift* se produit à un endroit prédéterminé et non au hasard, on parle de *frameshift* programmé.

#### 1.3.1.1 Le *frameshift* programmé –1

Le premier cas de *frameshift* programmé identifié était de type –1 et se produisait chez le virus du sarcome de Rous (RSV) (Jacks et Varmus 1985). Suite à cette observation, de nombreux sites de *frameshift* programmé –1 ont été identifiés (revue par Brierley 1995; Gesteland et Atkins 1996; Dinman et al. 1998). La majorité des cas se retrouvent chez les rétrovirus, mais on en observe également chez des virus eucaryotes à ARN(+), comme par exemple le virus du syndrome respiratoire aigu sévère (SRAS) (Ramos et al. 2004; Baranov et al. 2005), chez des virus à ARN en double brin comme le virus L-A de la levure (Dinman et al. 1991) et chez des virus à ARN de plantes comme le virus de la jaunisse nanisante de l'orge (BYDV) (Brault et Miller 1992). Des cas de *frameshift* –1 se retrouvent également chez les bactériophages comme T7 (Condrón et al. 1991) et  $\lambda$  (Levin et al. 1993). Chez *Escherichia coli*, le gène *dnaX* codant pour la sous-unité  $\gamma$  de la DNA polymérase III (Tsuchihashi et Brown 1992), ainsi que certains éléments d'insertion (Chandler et Fayet 1993) utilisent également le *frameshift* –1 pour leur expression. Plus récemment, il a été observé que le gène humain *Edr* (*embryonal carcinoma differentiation regulated*),

impliqué dans le développement des fœtus, utilise aussi un *frameshift* -1 (Shigemoto et al. 2001; Manktelow et al. 2005). *Edr* code pour une protéine semblable à Gag, qui est le précurseur des protéines structurales chez les rétrovirus, alors qu'après *frameshift* -1, il y a production d'une protéase semblable à la protéase de Pol, qui est le précurseur des enzymes chez les rétrovirus. Une origine rétrovirale du gène *Edr* a donc été proposée.

### 1.3.1.1.1 Mécanisme du *frameshift* programmé -1

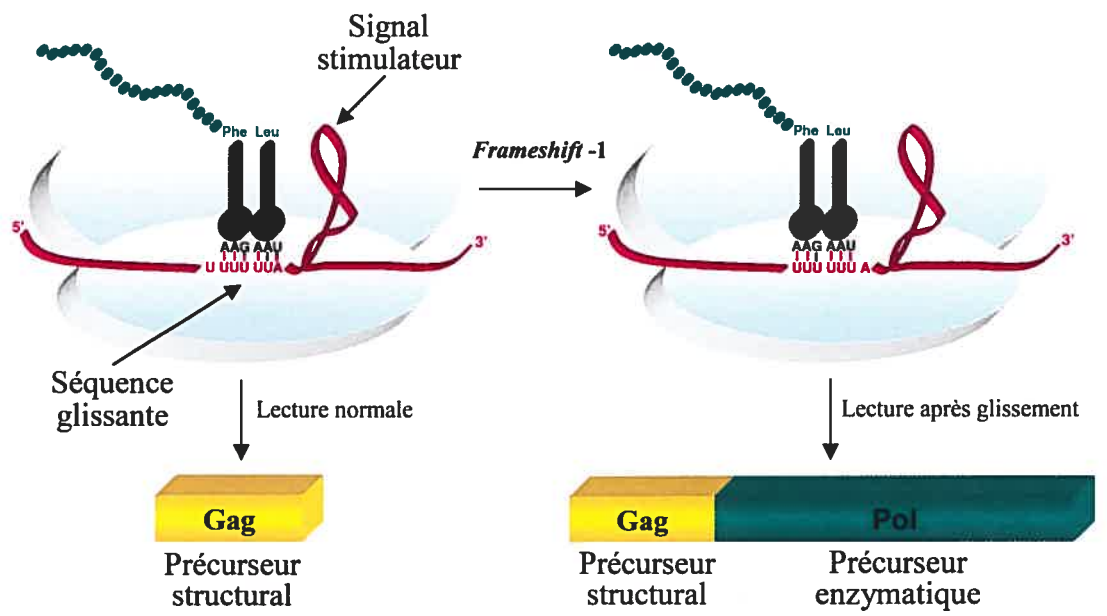
Deux éléments sur l'ARNm sont nécessaires pour produire un *frameshift* programmé -1. Le premier élément est une séquence glissante où se produit le *frameshift* -1. Cette séquence glissante est un heptamère de type X XXY YYZ (les espaces correspondent au cadre de lecture normal) où X peut être n'importe quel nucléotide, Y peut être A ou U, et Z peut être A, C ou U (Brierley et al. 1992). Le nucléotide directement en 3' de l'heptamère a aussi un effet sur l'efficacité du *frameshift* -1, U et C favorisant le *frameshift* à l'inverse de A et G. Cet effet résulterait de la stabilisation (A et G) ou de la déstabilisation (U et C) de l'interaction codon:anticodon au site A du ribosome (Bertrand et al. 2002). Le deuxième élément nécessaire au *frameshift* -1 est un signal stimulateur, une structure particulière de l'ARNm, qui adopte fréquemment une conformation en pseudonœud pour les virus utilisant le *frameshift* -1 (revue par Brierley 1995; Farabaugh 2000). Ce signal stimulateur se retrouve habituellement 5 à 8 nucléotides en aval de la séquence glissante.

Le mécanisme initialement proposé pour le *frameshift* programmé -1 consiste en un glissement d'un nucléotide en direction 5' sur l'ARNm par le ribosome portant le peptidyl-ARNt au site P et l'aminoacyl-ARNt au site A. Dans ce mécanisme, le *frameshift* -1 se produirait après l'accommodation de l'aminoacyl-ARNt au site A et avant la formation du lien peptidique (Jacks et al. 1988a; Dinman et al. 1991; Brierley et al. 1992, Figure 1-10). Le rôle du signal stimulateur serait d'entraîner une pause du ribosome en élongation au niveau de la séquence glissante. La majorité des ribosomes va continuer la traduction dans



le même cadre de lecture, mais pour une minorité de ribosomes, durant la pause, les deux ARNt vont se désappairer de l'ARNm, glisser avec le ribosome d'un nucléotide en direction 5' et se ré-apparier dans le cadre de lecture  $-1$ . Dans le cas des rétrovirus comme le VIH-1 (le *frameshift* programmé  $-1$  du VIH-1 sera décrit en détail à la section 1.4.4), les ribosomes qui ont effectué un *frameshift*  $-1$  continuent la traduction dans le nouveau cadre de lecture, évitant le codon d'arrêt de la protéine Gag pour permettre la production de la protéine de fusion Gag-Pol.

Figure 1-10 : Mécanisme du glissement simultané en 5' du ribosome et des deux ARNt proposé pour le *frameshift* programmé -1.



Suite à la présentation de ce mécanisme initial expliquant le *frameshift* -1, divers raffinements ont été apportés. Tout d'abord, il a été proposé que le *frameshift* -1 se produise quand l'aminoacyl-ARNt se trouve au site A/T, plutôt qu'au site A. Cette proposition découle de l'observation que la formation du lien peptidique après l'accommodation de l'aminoacyl-ARNt au site A est extrêmement rapide, ce qui ne laisserait pas beaucoup de temps au *frameshift* -1 de se produire (Farabaugh 1997a; Brunelle et al. 1999). D'un autre côté, l'étape d'accommodation est beaucoup plus longue, puisqu'il doit y avoir hydrolyse du GTP par eEF1A et relâche du complexe eEF1A-GDP, avant que l'aminoacyl-ARNt puisse occuper le site A (voir section 1.2.1.3). Ceci laisserait donc plus de temps au *frameshift* -1 de se produire. Une étude récente de notre groupe de recherche appuie l'hypothèse suggérant que l'aminoacyl-ARNt se trouve au site A/T lors du *frameshift* -1, car des mutations dans l'ARNr de la petite sous-unité ribosomique qui accélèrent l'accommodation de l'aminoacyl-ARNt au site A diminuent l'efficacité de *frameshift* -1 (Léger et al. 2004).

Un autre raffinement du mécanisme de *frameshift* -1 a été proposé grâce à la modélisation de la structure cristallographique du signal stimulateur en pseudonoed du virus de la jaunisse occidentale de la betterave (BWYV) avec la structure cristallographique du ribosome procaryote (Plant et al. 2003, Figure 1-11A). Dans ce modèle, le ribosome en élongation portant un peptidyl-ARNt au site P et un aminoacyl-ARNt au site A/T est bloqué au niveau de la séquence glissante lorsqu'il rencontre le signal stimulateur. Lors de l'accommodation de l'aminoacyl-ARNt du site A/T au site A, la boucle anticodon de l'aminoacyl-ARNt se déplace d'approximativement 9 Å en direction 5' (Noller et al. 2002). En raison de la liaison codon:anticodon, ce déplacement de l'aminoacyl-ARNt entraînerait l'ARNm, ce qui induirait une tension dans la portion de l'ARNm située entre le site A et le signal stimulateur (espaceur). Cette tension pourrait être supprimée de deux façons. La plus simple consiste à détruire la structure du signal stimulateur grâce à l'activité hélicase du ribosome, ce qui permet de continuer la traduction dans le cadre de lecture normal. Une autre façon de supprimer la tension serait de briser les

liens codon:anticodon entre l'ARNm et les ARNt aux sites A et P afin de permettre à l'ARNm de glisser d'un nucléotide en direction 3'. La traduction continuerait ensuite dans le cadre de lecture -1. Ce mouvement de l'ARNm serait suffisant pour relâcher la tension, puisque la distance entre deux nucléotides dans un ARNm non-structuré est de 5,2 Å. Le même groupe de recherche vient également de proposer que le ribosome qui tente de défaire un signal stimulateur en pseudonoeud serait repoussé en direction 5' de l'ARNm en raison d'une augmentation de la torsion dans le pseudonoeud (Plant et Dinman 2005).

Un dernier raffinement du mécanisme de *frameshift* -1 suggère que le lien codon:anticodon au site P puisse être brisé indépendamment de celui au site A (Baranov et al. 2004, Figure 1-11B). Selon ce modèle, quand le ribosome en élongation se trouve au niveau de la séquence glissante, le lien codon:anticodon du peptidyl-ARNt au site P serait brisé d'abord et pourrait se reformer dans le cadre de lecture -1, alors que le lien codon:anticodon de l'aminoacyl-ARNt au site A/T demeurerait intact. Cette situation anormale laisserait un nucléotide non apparié sur l'ARNm entre les sites P et A du ribosome et serait stimulée par le signal stimulateur. Une telle situation non favorable serait suivie du retour du peptidyl-ARNt au site P normal dans la majorité des ribosomes, mais, dans une minorité de ribosome, l'aminoacyl-ARNt au site A/T se repositionnerait dans le cadre de lecture -1 avant d'être accommodé au site A.

Finalement, il est important de noter que la pause du ribosome en élongation au niveau de la séquence glissante, qui est induite par le signal stimulateur, est nécessaire, mais non suffisante à la stimulation du *frameshift* -1 (Tu et al. 1992; Somogyi et al. 1993; Lopinski et al. 2000; Kontos et al. 2001). Il a été proposé qu'outre la pause, qui peut être induite par n'importe quelle structure possédant une stabilité suffisante, une interaction spécifique entre le signal stimulateur et le ribosome pourrait expliquer la stimulation du *frameshift* -1 (Shen et Tinoco 1995; Kang et Tinoco 1997).

Figure 1-11 : Raffinements proposés pour expliquer le *frameshift* programmé -1.

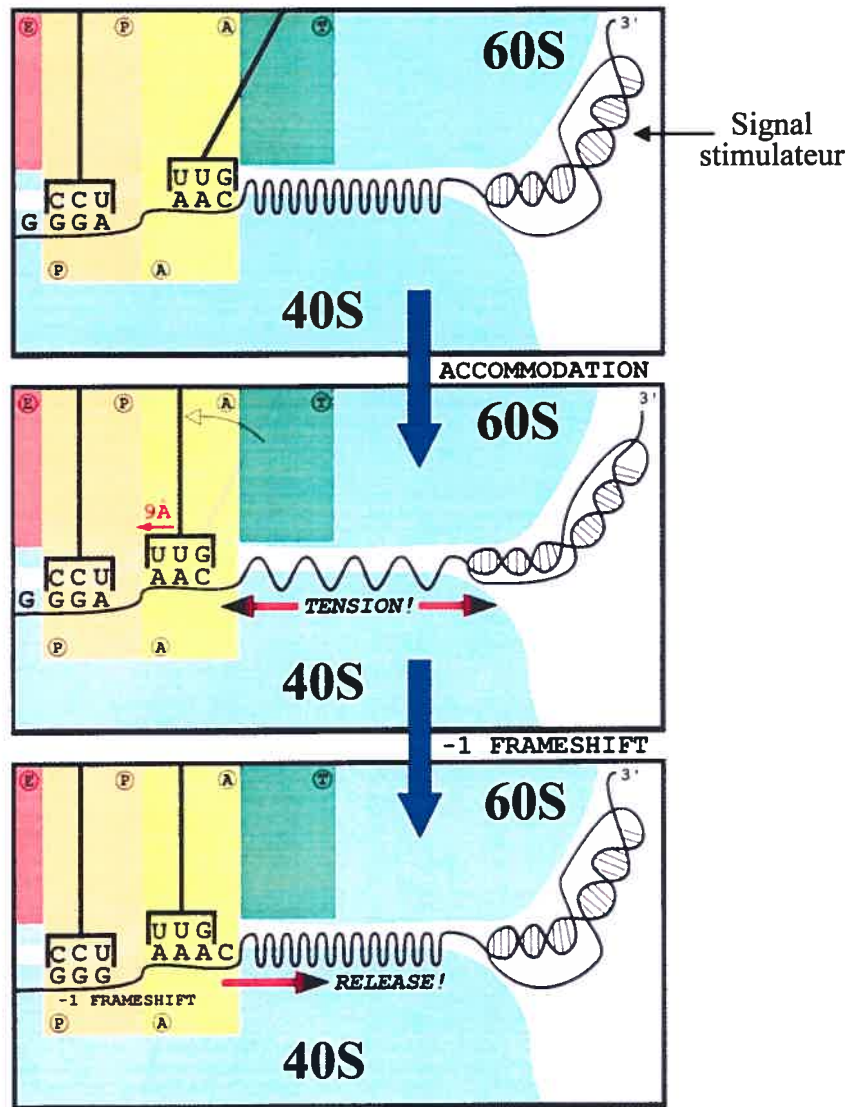
A) Modèle où le signal stimulateur permet le *frameshift* -1 en induisant une tension dans l'ARNm (tiré de Plant et al. 2003).

La séquence glissante (G GGA AAC) et le signal stimulateur en pseudonoëud sont ceux du BWYV.

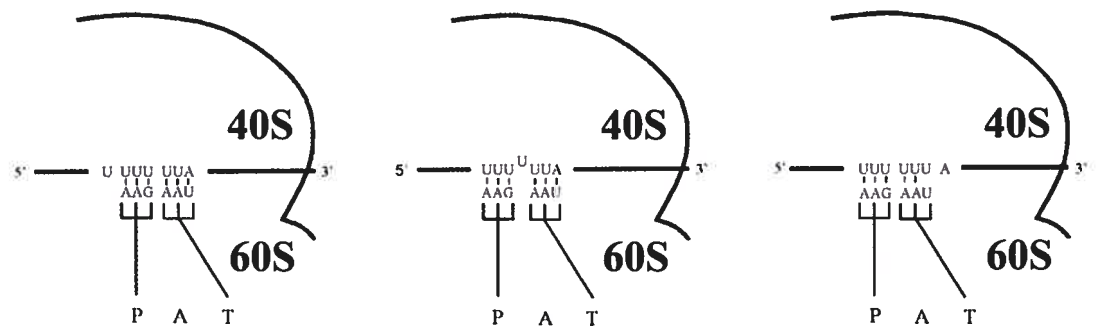
B) Modèle où le *frameshift* -1 se produit séquentiellement, le peptidyl-ARNt au site P glissant avant l'aminoacyl-ARNt au site A/T (adapté de Léger et al. 2004).

La séquence glissante (U UUU UUA) est celle du VIH-1. Le signal stimulateur n'est pas présenté ici.

**A**



**B**



### 1.3.1.1.2 Structure des signaux stimulateurs dans le *frameshift* programmé -1

Le signal stimulateur de *frameshift* -1 chez les virus infectant les cellules eucaryotes est très souvent un pseudonoeud (Brierley et Pennel 2001). Ce n'est toutefois pas le cas chez le groupe M du VIH-1 (voir détails dans la section 1.4.4) (Dulude et al. 2002), chez le virus humain de la leucémie à cellules T de type II (HTLV-II) (Falk et al. 1993) et chez les astrovirus (Marczinke et al. 1994). Cette section décrit quelques pseudonoeuds rencontrés chez les virus eucaryotes. Plusieurs structures de pseudonoeuds stimulateurs de *frameshift* -1 ont été caractérisées par mutagénèse et par analyse de la structure secondaire d'ARN grâce à des attaques avec des agents chimiques et enzymatiques (Brierley et al. 1989; Marczinke et al. 1998). Quoique ces données structurales donnent des renseignements fiables sur la structure secondaire adoptée par le pseudonoeud, elles ne permettent pas l'élucidation de la structure tertiaire. Ce problème a été surmonté durant les dernières années avec la caractérisation de la structure du signal stimulateur en pseudonoeud par RMN pour le rétrovirus simien de type 1 (SRV-1) (Du et al. 1997; Michiels et al. 2001, Figure 1-12A) et le virus de la tumeur mammaire de la souris (MMTV) (Shen et Tinoco 1995, Figure 1-12B), et par diffraction aux rayons X pour le BWYV (Su et al. 1999; Egli et al. 2002, Figure 1-12C) et le virus de la mosaïque avec énation du pois (PEMV) (Nixon et al. 2002).

Ces études structurales ont permis de classer les pseudonoeuds stimulateurs de *frameshift* -1 en deux grands groupes. Les membres du premier groupe possèdent un nucléotide non-apparié (souvent un A) à la jonction des tiges 1 et 2 du pseudonoeud, la tige 2 étant formée par l'appariement de la boucle coiffant la tige 1 avec une séquence complémentaire en aval (Figure 1-12B). Ce nucléotide non-apparié provoque une inclinaison entre les deux tiges du pseudonoeud et il a été proposé que cette particularité du pseudonoeud soit requise pour stimuler le *frameshift* -1 (Chen X. et al. 1996). Les pseudonoeuds du MMTV, du BWYV et du PEMV font partie de ce premier groupe. Les



pseudonoeuds stimulateurs de *frameshift* -1 du deuxième groupe ne possèdent toutefois pas de nucléotide non-apparié à la jonction des tiges 1 et 2. SRV-1 est un membre de ce deuxième groupe et la résolution de sa structure a permis de démontrer que les deux tiges sont empilées de façon coaxiale, montrant que la présence d'un nucléotide non-apparié à la jonction des tiges 1 et 2 n'est pas toujours nécessaire à la stimulation du *frameshift* -1.

Les pseudonoeuds stimulant le *frameshift* -1 possèdent toutefois certaines similarités structurales (revue par Giedroc et al. 2000). Premièrement, les tiges 1 et 2 sont riches en paires de bases G-C, ce qui assurerait une stabilité suffisante pour entraîner la pause du ribosome en élongation (Brierley 1995). De plus, la boucle 1 ne possède jamais plus de deux nucléotides et ces nucléotides interagissent avec le sillon majeur de la tige 2. Enfin, la boucle 2 contient une séquence AACAA très conservée qui forme des interactions sous forme de triplex avec le sillon mineur de la tige 1. Il a été proposé que la formation de ce triplex soit le motif caractéristique des pseudonoeuds qui stimulent le *frameshift* -1. Pour confirmer cette hypothèse, la résolution de plusieurs structures de pseudonoeuds au niveau atomique doit être effectuée.

Figure 1-12 : Structure tertiaire de pseudonoeuds stimulant le *frameshift* programmé -1.

A) Structure du pseudonoeud de MMTV déterminée par RMN (tiré de Shen et Tinoco 1995).

Le nucléotide A14 à la jonction des deux tiges n'est pas apparié.

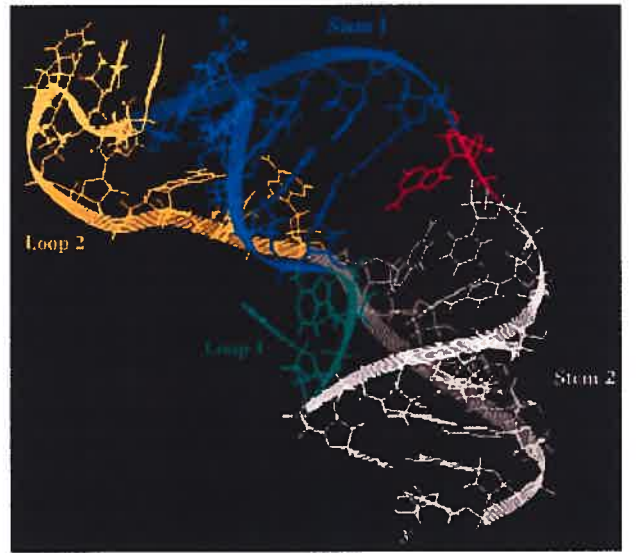
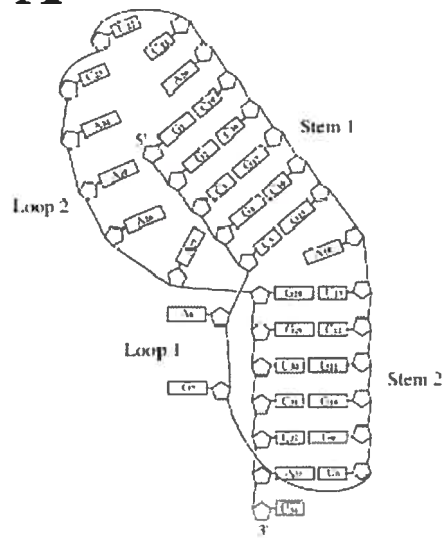
B) Structure du pseudonoeud de BWYV déterminée par diffraction aux rayons X (tiré de Su et al. 1999).

Le résidu U13 à la jonction des deux tiges n'est pas apparié. La séquence AACAA de la boucle 2 qui forme le triplex est encerclée.

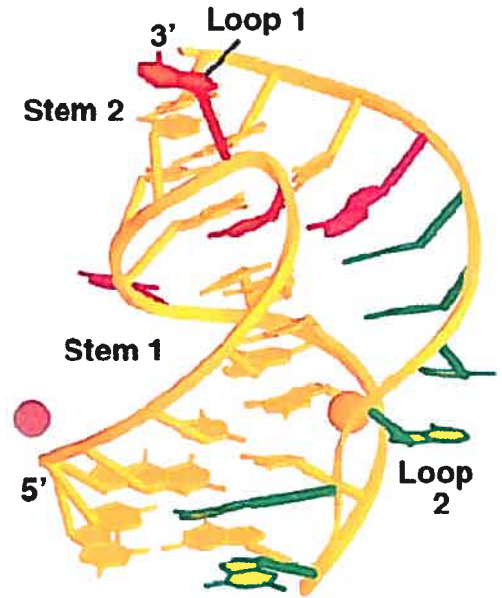
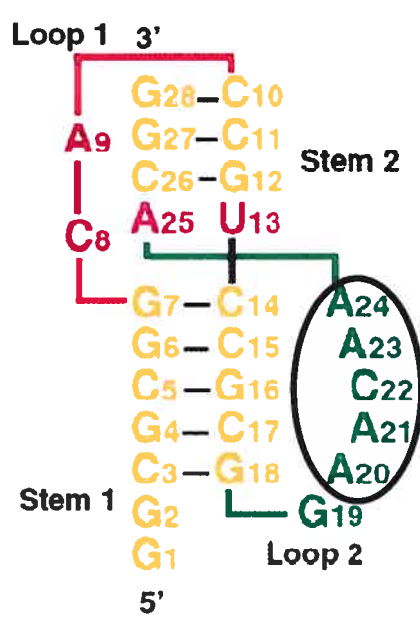
C) Structure du pseudonoeud de SRV-1 déterminée par RMN (tiré de Michiels et al. 2001).

La séquence AACAA de la boucle 2 qui forme le triplex est encerclée.

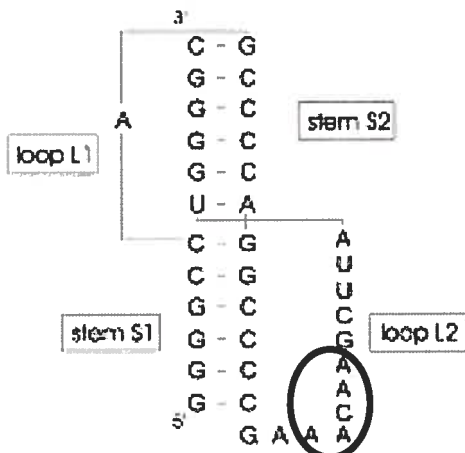
**A**



**B**



**C**



### 1.3.1.2 Le *frameshift* programmé +1

Les événements de *frameshift* programmé +1 sont moins répandus que les événements de *frameshift* programmé -1, mais leurs mécanismes sont bien caractérisés. Des exemples de *frameshift* +1 ont été répertoriés chez *E.coli* dans le gène *prfB*, codant pour le facteur de terminaison RF2 (Craigén et Caskey 1986), chez la levure dans les rétro-transposons Ty1 (Belcourt et Farabaugh 1990) et Ty3 (Farabaugh et al. 1993), dans le gène EST3, qui code pour une sous-unité de la télomérase (Morris et Lundblad 1997), et dans le gène ABP140, qui code pour une protéine liant les filaments d'actine (Asakura et al. 1998). Des exemples de *frameshift* +1 ont aussi été répertoriés chez l'homme dans le gène codant pour l'antizyme responsable de la dégradation de l'ornithine décarboxylase (ODC), l'enzyme qui contrôle la synthèse des polyamines (Matsufuji et al. 1995; Ivanov et al. 1998a). Tous ces *frameshifts* +1 se produisent durant une pause du ribosome induite par un codon d'arrêt peu efficace ou par un codon dont l'aminocyl-ARNt est rare.

Par exemple, pour produire RF2 chez *E.coli*, les ribosomes bactériens doivent changer de cadre de lecture au codon 25 de l'ARNm de RF2 pour éviter le codon 26 qui est un codon d'arrêt (CUU UGA C, Figure 1-13A). Dans ce cas, quand il y a *frameshift* +1, l'ARNt<sup>Leu</sup> dont l'anticodon est <sup>3'</sup>GAG<sup>5'</sup> se désapparie du codon CUU pour se réappairer dans le cadre de lecture +1 avec UUU (Craigén et al. 1985). Il y a donc glissement d'un ribosome portant un seul ARNt, le peptidyl-ARNt au site P. Ce phénomène de glissement se produit pour un nombre limité de ribosomes, les autres terminant la traduction au codon d'arrêt pour produire un peptide non fonctionnel (Craigén et Caskey 1986). L'efficacité de ce glissement dépend donc de l'efficacité de terminaison, qui dépend elle-même de la quantité de RF2 présente. En présence d'une faible quantité de RF2, la terminaison est peu efficace, ce qui favorise le *frameshift* +1 et la production de RF2. Cette augmentation de la concentration de RF2 accroît alors l'efficacité de terminaison et diminue l'efficacité de *frameshift* +1. La production de RF2 est donc autorégulée. Il y a aussi un signal stimulateur du *frameshift* +1, qui est une séquence riche en purines de type Shine-Dalgarno, semblable

à la séquence présente en amont de l'AUG initiateur de l'ARNm pour permettre l'initiation de la traduction procaryote. L'interaction de cette séquence avec la séquence anti-Shine-Dalgarno de l'ARNr 16S stimulerait le *frameshift* +1 en favorisant la pause au niveau de la région contenant le codon d'arrêt (Curran et Yarus 1988; Weiss et al. 1988).

Un autre exemple de *frameshift* +1 est celui de l'antizyme de l'ODC chez les eucaryotes (Figure 1-13B). Le glissement causant le *frameshift* +1 a lieu quand le ribosome rencontre un codon d'arrêt (UCC UGA U) dans l'ARNm de l'antizyme de l'ODC. Dans ce cas, l'ARNt<sup>Ser</sup> dont l'anticodon est 3' AGI<sup>5'</sup> (où I est une inosine, un précurseur des purines) se désapparie du codon UCC pour se réappairer dans le cadre de lecture +1 avec CCU (Farabaugh 1997b). Dans ce cas, le glissement du peptidyl-ARNt dans le cadre de lecture +1 entraînerait un mésappariement A-C et une paire de base *wobble* I•G au site P. La présence de ce mésappariement a mené à l'hypothèse que le *frameshift* +1 de l'antizyme ne se produirait pas par glissement, mais plutôt par l'occlusion du nucléotide U directement en 3' du codon UCC, qui serait ignoré par le ribosome (Matsufuji et al. 1995). Comme dans le cas de RF2, le *frameshift* +1 de l'antizyme est autorégulé. En effet, l'efficacité *frameshift* +1 est stimulée par les polyamines dont la présence augmente la synthèse de l'antizyme, conduisant à la dégradation de l'ODC et conséquemment à une diminution de la concentration de polyamines (Rom et Kahana 1994). Une caractéristique intéressante de ce *frameshift* +1 est la présence d'un signal stimulateur en pseudonoed situé en aval du site de glissement (Matsufuji et al. 1995).

Finalement, dans le cas des rétro-transposons Ty1 et Ty3 chez la levure, les deux événements de *frameshift* +1 ne se produisent pas à un codon d'arrêt, mais à un codon dont l'aminoacyl-ARNt est rare (Farabaugh et al. 1993; Kawakami et al. 1993). Un mécanisme d'occlusion a été proposé pour Ty3 (Matsufuji et al. 1995, Figure 1-13C), alors que le *frameshift* +1 de Ty1 se produit après glissement du peptidyl-ARNt (Belcourt et Farabaugh 1990, Figure 1-13D). L'efficacité de *frameshift* +1 de Ty3 est stimulée par une séquence en aval du site de glissement qui est complémentaire à une portion de l'hélice 18 de l'ARNr

18S (Li Z. et al. 2001). Cette observation est particulièrement intéressante puisque l'hélice 18 est impliquée dans la fidélité de sélection des aminoacyl-ARNt au site A du ribosome (revue par Ogle et Ramakrishnan 2004).

Le groupe de Atkins (Hansen et al. 2003) a récemment démontré que le *frameshift* +1 se produit toujours par glissement du peptidyl-ARNt au site P, même dans les cas où un mécanisme d'occlusion avait été proposé (Figure 1-14). Leurs résultats s'appuient aussi sur les données cristallographiques obtenues récemment pour le ribosome bactérien, qui ont permis de mieux caractériser le mécanisme de sélection de l' aminoacyl-ARNt au site A (revue par Ogle et Ramakrishnan 2004). En effet, lors de la liaison d'un aminoacyl-ARNt au site A, un réarrangement conformationnel du ribosome fait pivoter les nucléotides A1492 et A1493 de l'ARNr 16S de *E.coli* (correspondant à A1823 et A1824 de l'ARNr 18S eucaryote) pour leur permettre d'interagir avec le sillon mineur de l'hélice formée par l'interaction codon:anticodon. Le nucléotide A1493 interagit avec la première paire de bases du lien codon:anticodon, alors que le nucléotide A1492, ainsi que le nucléotide G530 (G626 dans l'ARNr 18S), interagissent avec la deuxième paire de bases du lien codon:anticodon. De son côté, la troisième paire de bases est moins étroitement surveillée par le ribosome, ce qui lui permettrait de tolérer un appariement *wobble* G-U à cette position (Ogle et al. 2001). Ce mécanisme permet au ribosome de sélectionner l' aminoacyl-ARNt correct au site A en reconnaissant la géométrie du pairage codon:anticodon. Selon ce mécanisme de sélection, la liaison de l' aminoacyl-ARNt au site A dans le cadre de lecture +1 ne pourrait pas être acceptée par les nucléotides A1492 et A1493 avec le mécanisme d'occlusion où le peptidyl-ARNt ne glisse pas (*out-of-frame* dans la Figure 1-14C), puisque l'hélice formée par l'interaction codon:anticodon ne serait pas correctement positionnée par rapport au ribosome. Par contre, si le peptidyl-ARNt au site P glisse en même temps que le ribosome d'un nucléotide en direction 3' sur l'ARNm avant la liaison de l' aminoacyl-ARNt au site A, l' aminoacyl-ARNt sera accepté et la traduction se poursuivra dans le cadre de lecture +1 (*slippage* dans la Figure 1-14C).

Figure 1-13 : Différents mécanismes proposés pour le *frameshift* programmé +1.

- A) Mécanisme par glissement proposé pour RF2 (adapté de Marquez et al. 2004)
- B) Mécanisme sans glissement (occlusion) initialement proposé pour l'antizyme (adapté de Ivanov et al. 1998b)
- C) Mécanisme sans glissement (occlusion) initialement proposé pour le rétro-transposon Ty3 (adapté de Li Z. et al. 2001).  
Le codon AGU au site A est un codon rare. Le facteur eEF1A est représenté en gris, l'étoile et le cercle noirs représentent le GTP et le GDP, respectivement.
- D) Mécanisme par glissement proposé pour le rétro-transposon Ty1 (adapté de Li Z. et al. 2001).  
Le codon AGG au site A est un codon rare.

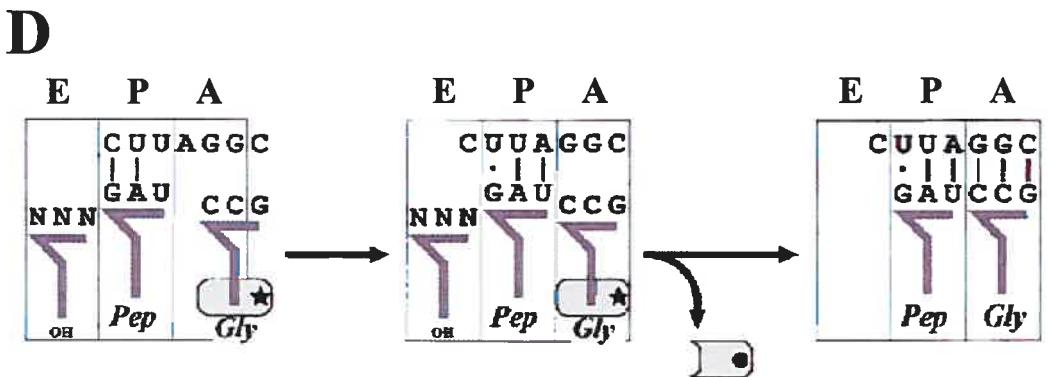
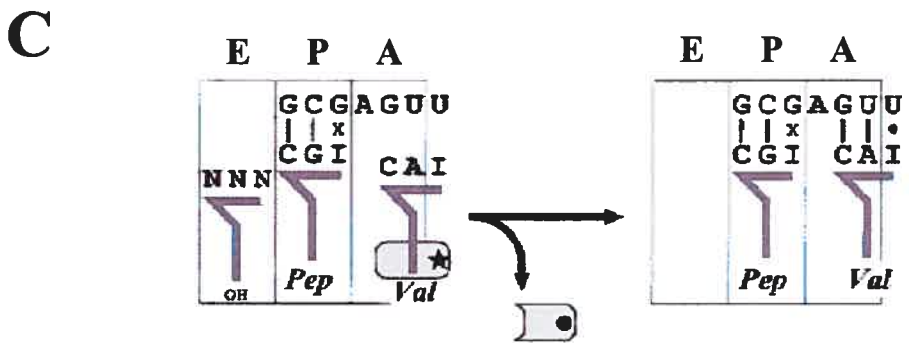
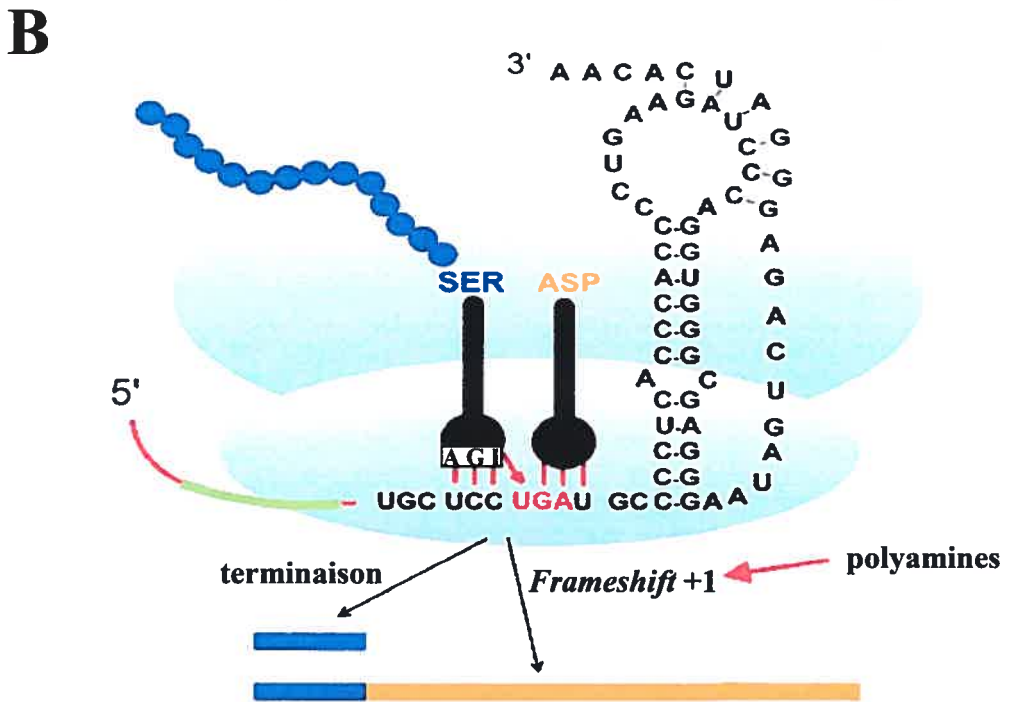
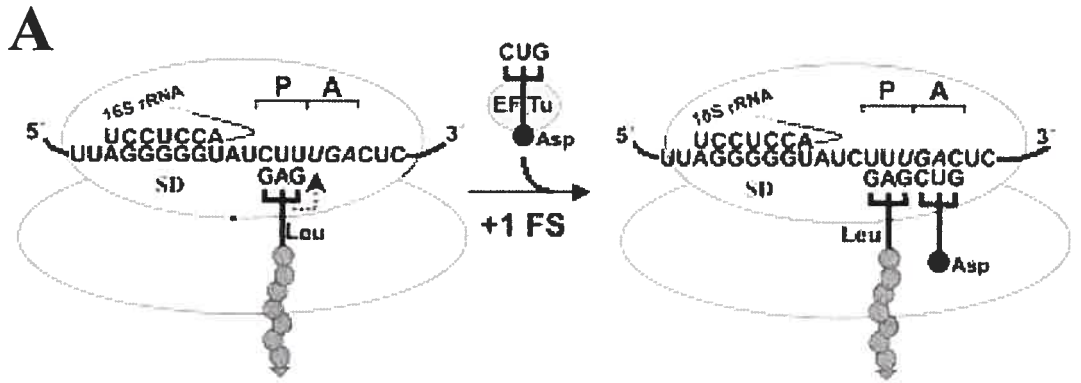




Figure 1-14 : Mécanisme de sélection des aminoacyl-ARNt au site A du ribosome (Hansen et al. 2003).

A) Détails des sites A et P du ribosome.

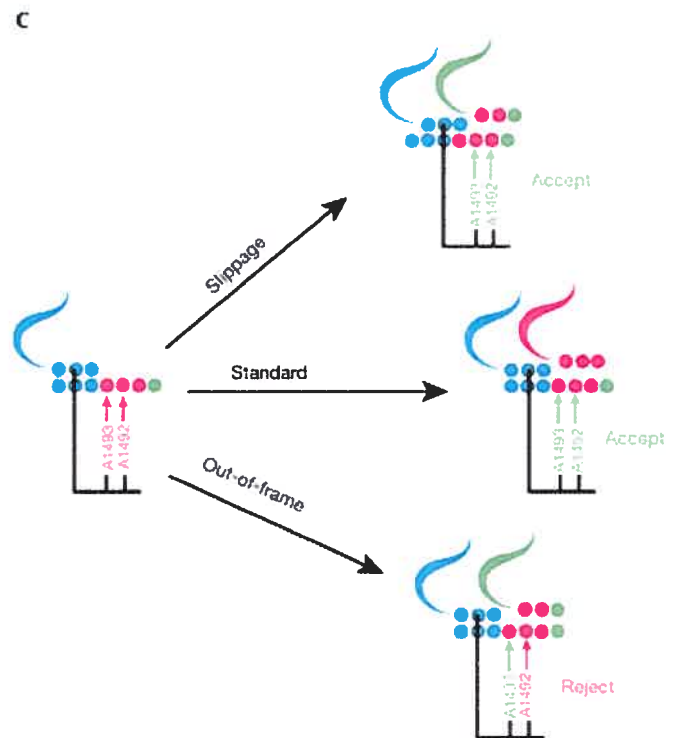
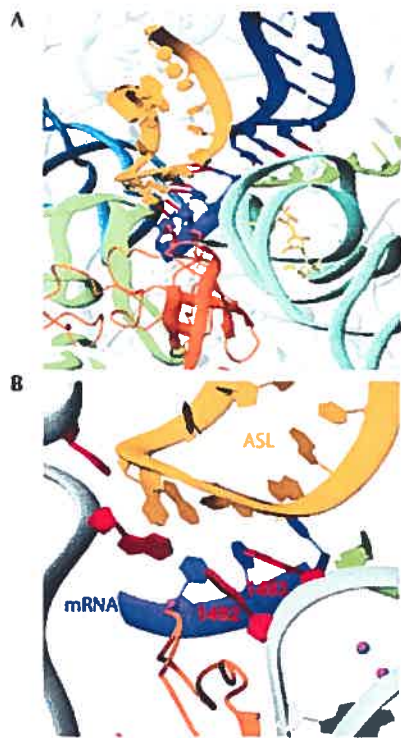
Le peptidyl-ARNt au site P est représenté en bleu et l'aminoacyl-ARNt au site A est représenté en jaune.

B) Détails des composantes du ribosome entourant le site A.

Les nucléotides A1492 et A1493 qui assurent la sélection d'un aminoacyl-ARNt correct au site A sont identifiés en rouge.

C) Représentation schématique de la liaison d'un aminoacyl-ARNt dans le cadre de lecture normal ou dans le cadre de lecture +1.

Les flèches vertes indiquent que les nucléotides A1492 et A1493 reconnaissent un appariement codon:anticodon correct, alors que les flèches mauves indiquent qu'ils ne reconnaissent pas un appariement correct.



### 1.3.2 L'initiation de la traduction à un codon non-AUG

En plus des mécanismes non-conventionnels de traduction lors de l'élongation et de la terminaison, comme ceux présentés précédemment, certains mécanismes non-conventionnels peuvent également se produire lors de l'initiation de la traduction. Ces mécanismes permettent entre autre l'utilisation de plusieurs codons initiateurs de la traduction sur un même ARNm, contribuant à la génération d'une grande diversité de protéines (revue par Touriol et al. 2003). Ces codons initiateurs sont principalement des codons AUG dans le même cadre de lecture et en aval du premier codon AUG initiateur, mais un nombre grandissant de codons initiateurs non-AUG sont maintenant identifiés. La sélection de ces codons non-AUG comme site d'initiation de la traduction semble être gouvernée par plusieurs facteurs, dont la séquence primaire et la structure secondaire de l'ARNm entourant ce codon.

#### 1.3.2.1 L'initiation de la traduction Met-ARNt<sub>i</sub><sup>Met</sup>-dépendante à un codon non-AUG

Le premier exemple d'initiation à un codon non-AUG a été découvert chez le virus Sendai où une initiation en amont et dans un cadre de lecture différent du premier codon AUG, qui code pour l'ARN polymérase ARN-dépendante du virus, mène à la synthèse de la protéine C' (Gupta et Patwardhan 1988). Cette protéine est indispensable à la propagation virale et à la pathogenèse *in vivo* (Kurotani et al. 1998). Plusieurs autres exemples d'initiation non-AUG, en amont ou en aval du codon AUG initiateur traditionnel, ont été caractérisés par la suite chez les virus. Le virus de Theiler (encéphalomyélite de la souris ou TMEV) est un exemple d'initiation à un codon ACG en aval d'un AUG initiateur traditionnel utilisé pour synthétiser la polyprotéine virale. L'initiation dans le cadre de lecture +1 permet la synthèse de la protéine L\* qui est essentielle à la persistance du virus (van Eyll et Michiels 2000, 2002).

Plusieurs exemples d'initiation à des codons non-AUG ont également été caractérisés dans des ARNm cellulaires. Ces exemples se rencontrent principalement dans des classes de protéines régulatrices comme des proto-oncogènes, des facteurs de transcription, des kinases ou des facteurs de croissance (revue par Touriol et al. 2003). Des études de traduction *in vitro* ont montré que, sur les neuf codons qui diffèrent d'AUG par un seul nucléotide, seuls les codons AGG et AAG ne permettaient pas une initiation de la traduction Met-ARNt<sub>i</sub><sup>Met</sup>-dépendante (Peabody 1989). Jusqu'à maintenant, aucune initiation avec un Met-ARNt<sub>i</sub><sup>Met</sup> n'a été observée à des codons qui diffèrent d'AUG par plus d'un nucléotide. Cette observation a mené à l'hypothèse qu'un minimum de deux paires de bases doivent être formées entre le codon et l'anticodon pour pouvoir permettre une initiation Met-ARNt<sub>i</sub><sup>Met</sup>-dépendante. D'autres études ont également montré qu'un contexte Kozak optimal (Kozak 1989), ainsi que la présence d'une tige-boucle stable en aval du codon non-AUG initiateur augmentent l'efficacité d'initiation à ce codon (Kozak 1990; Prats et al. 1992). Le facteur d'initiation eIF1 pourrait également jouer un rôle dans ce mécanisme d'initiation anormale, puisqu'en son absence, le complexe de préinitiation 43S ne différencie pas les codons AUG des codons non-AUG, qu'ils soient dans un contexte Kozak optimal ou non (Pestova et Kolupaeva 2002).

### 1.3.2.2 L'initiation de la traduction Met-ARNt<sub>i</sub><sup>Met</sup>-indépendante à un codon non-AUG

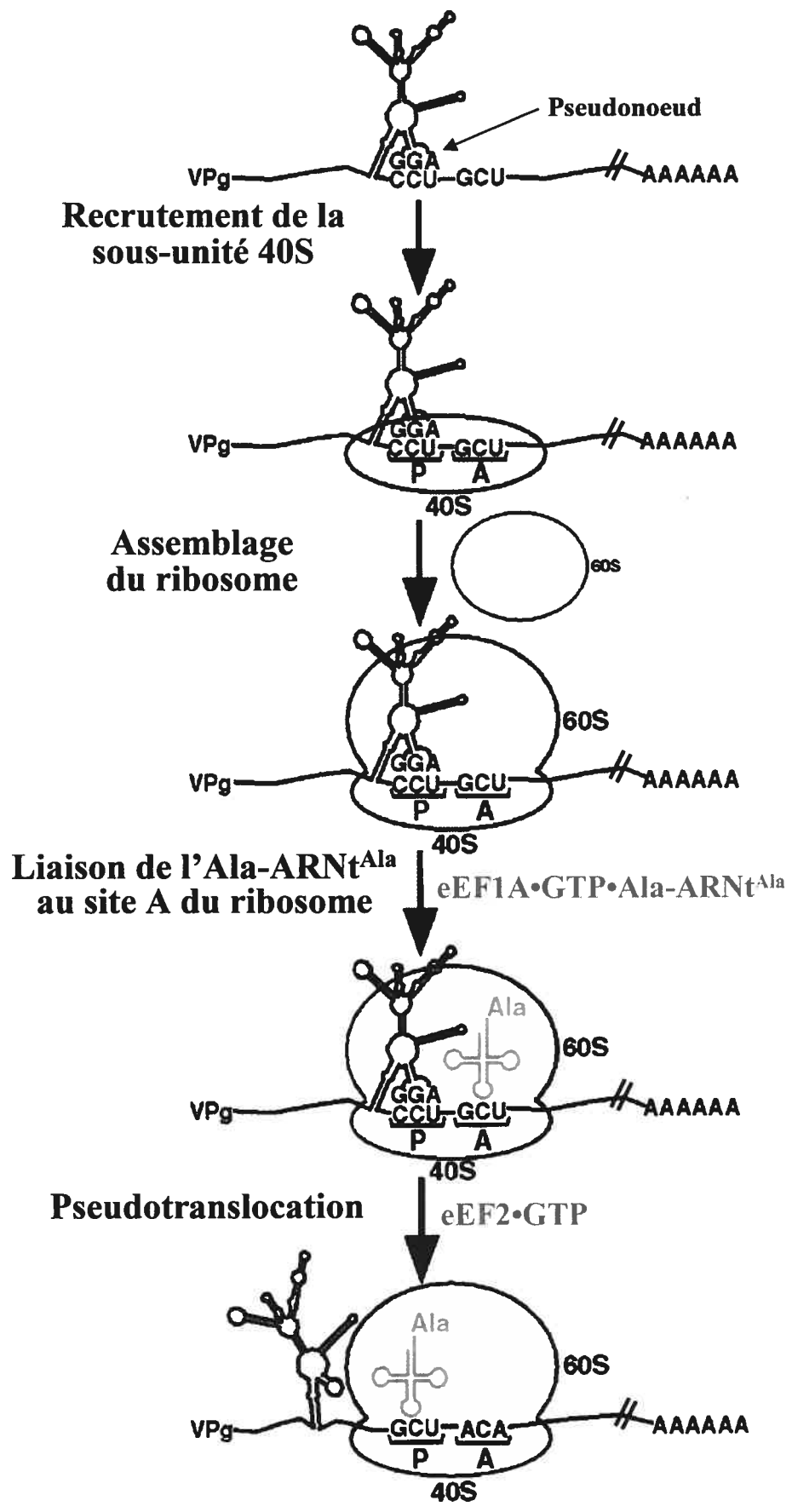
Comme mentionné à la section 1.2.1.2, certains virus d'insecte possèdent un IRES dans leur région intergénique pour permettre la synthèse du précurseur de leurs protéines structurales (revue par Hellen et Sarnow 2001). Ces IRES ont la particularité d'initier la traduction à un codon non-AUG, sans utiliser le Met-ARNt<sub>i</sub><sup>Met</sup>. L'IRES recrute d'abord la sous-unité 40S au niveau du site d'initiation de la traduction (Figure 1-15). Le recrutement subséquent de la sous-unité 60S permet l'assemblage du ribosome. Dans un système d'initiation *in vitro*, ces deux étapes de l'initiation peuvent se faire sans l'aide d'aucun des eIF (Wilson et al. 2000). On ne peut toutefois pas exclure que, *in vivo*, certains eIF

permettent d'accélérer les différentes étapes de l'initiation. Lorsque le ribosome est assemblé, le site P est occupé par une structure en pseudonoeud de l'IRES, bloquant ainsi l'accès au Met-ARNt<sub>i</sub><sup>Met</sup>. Cette particularité de l'IRES permet une initiation sans recourir au complexe ternaire eIF2•GTP•Met-ARNt<sub>i</sub><sup>Met</sup>. Le facteur d'élongation eEF1 lié au GTP positionne un aminoacyl-ARNt (Ala-ARNt<sup>Ala</sup> dans le cas du CrPV) au site A/T du ribosome, comme lors de l'élongation normale (voir section 1.2.2). Il y a ensuite hydrolyse du GTP, relâche de eEF1•GDP et accommodation de l'aminoacyl-ARNt au site A du ribosome. Cette étape est suivie par une pseudotranslocation médiée par eEF2 où l'aminoacyl-ARNt passe du site A au site P sans formation de lien peptidique, puisque le site P n'est pas occupé par un peptidyl-ARNt. L'élongation se poursuit alors suivant le mode normal, avec occupation du site A par un aminoacyl-ARNt, formation du lien peptidique et translocation.

Récemment, un autre exemple d'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-indépendante à un codon non-AUG a été caractérisé chez un peptide cryptique présenté à la surface cellulaire par le complexe majeur d'histocompatibilité de classe 1 (Schwab et al. 2004). Dans cet exemple, un codon CUG est reconnu par un Leu-ARNt<sup>Leu</sup> qui se lie directement au site A, dans un mécanisme indépendant du complexe ternaire eIF2•GTP•Met-ARNt<sub>i</sub><sup>Met</sup>, après balayage de l'ARNm. Le mécanisme précis menant à cette initiation non-traditionnelle de la traduction reste toutefois nébuleux, puisque, contrairement au mécanisme retrouvé chez le CrPV, aucune séquence primaire ou structure secondaire d'ARN ne semble nécessaire à la sélection de ce site d'initiation. De plus, les auteurs suggèrent que l'efficacité d'initiation peut être augmentée par un contexte Kozak optimal, contexte qui est habituellement spécifique à l'initiation traditionnelle avec un Met-ARNt<sub>i</sub><sup>Met</sup> au site P.

Figure 1-15 : Mécanisme d'initiation de la traduction Met-ARNt<sub>i</sub><sup>Met</sup>-indépendant (Wilson et al. 2000).

L'IRES du CrPV recrute la sous-unité 40S au niveau du site d'initiation de la traduction. Le recrutement subséquent de la sous-unité 60S permet l'assemblage du ribosome où un pseudonoeud de l'IRES occupe le site P. Un Ala-ARNt<sup>Ala</sup> se lie alors au site A du ribosome, comme lors de l'élongation normale (voir section 1.2.2). Cette étape est suivie par une pseudotranslocation où l'Ala-ARNt<sup>Ala</sup> passe du site A au site P sans qu'il n'y ait formation de lien peptidique. Le ribosome effectue ensuite les étapes normales de l'élongation.



Pour faire suite aux sections portant sur les mécanismes conventionnels et non-conventionnels de la traduction, nous présenterons maintenant les deux virus, le VIH-1 (section 1.4) et le VHC (section 1.5), chez lesquels nous avons étudié des mécanismes non-conventionnels de traduction. Dans le cas du VIH-1, nous avons étudié le *frameshift* programmé  $-1$  permettant la synthèse de Gag-Pol, alors que pour le VHC, nous avons caractérisé le mécanisme menant à la synthèse d'une protéine virale récemment identifiée, la protéine F. Des études antérieures avaient suggéré qu'un *frameshift*  $+1$  contrôlerait la synthèse de la protéine F, mais nos études ont plutôt démontré qu'elle était synthétisée après initiation de la traduction à un codon non-AUG.

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

Le VIH-1 est un membre de la famille des lentivirus, la sous-famille des rétrovirus les plus complexes, qui cause le syndrome de l'immunodéficience acquise (SIDA) chez l'homme. Cette maladie se développe après l'affaiblissement du système immunitaire par le VIH-1. Malgré de nombreuses années de recherche, aucun traitement n'a permis jusqu'à maintenant d'éradiquer ce virus mortel. Plusieurs agents anti-VIH permettent toutefois de ralentir le développement de la maladie. Les agents anti-VIH les plus utilisés sont répartis en trois classes : les inhibiteurs nucléosides de la transcriptase réverse (NRTI), comme par exemple l'AZT, le 3TC, le ddI et le ddC (Mitsuya et Broder 1987), les inhibiteurs non-nucléosides de la transcriptase réverse (NNRTI), comme la Nevirapine (Merluzzi et al. 1990) et les inhibiteurs de la protéase virale (PI), comme l'Indinavir et le Saquinavir (McQuade et al. 1990). Quoique très efficace à court terme, le traitement à long terme de patients avec ce type d'inhibiteurs entraîne la sélection de mutations dans le génome viral, qui confèrent une résistance à chacun de ces agents anti-VIH : NRTI (Larder et al. 1989), NNRTI (Richman et al. 1991) et PI (Condra et al. 1995). La combinaison de ces traitements, pour mener éventuellement à la trithérapie antirétrovirale (HAART), a permis



de minimiser l'apparition de ces résistances. Dernièrement, des virus résistants à la trithérapie ont toutefois émergé (Gallego et al. 2001), démontrant l'urgence de développer des agents anti-VIH contre de nouvelles cibles de la réplication du virus. Un de ces nouveaux agents, le T20, est un peptide de 36 acides aminés correspondant à une portion de la glycoprotéine transmembranaire de l'enveloppe du VIH-1 (gp41) qui permet d'inhiber l'entrée du virus dans la cellule (Kilby et al. 1998). Des virus résistants au T20 sont rapidement apparus (Baldwin et al. 2004), mais l'utilisation de cet agent en combinaison avec les traitements traditionnels permet le traitement efficace de certains patients résistants aux NRTI, NNRTI ou PI. Plusieurs autres traitements sont également en développement, comme par exemple des antagonistes du co-récepteur CCR5 qui empêchent sa liaison au virus (Strizki et al. 2001) des inhibiteurs de l'intégrase virale (Hazuda et al. 2000) ou une nouvelle génération de NNRTI efficaces contre les virus déjà résistants aux NNRTI traditionnels (Andries et al. 2004). Finalement, la découverte récente du mécanisme naturel de l'interférence de l'ARN (ARNi), qui utilise un court ARN double-brin (duplex de 21 nucléotides) pour dégrader spécifiquement un ARNm complémentaire, a rapidement mené au développement de traitements anti-VIH ciblant le génome d'ARN viral (Gitlin et al. 2002; Jacque et al. 2002; Novina et al. 2002). Malheureusement, la grande spécificité de l'ARNi a permis au virus de développer facilement des résistances en mutant directement la séquence de nucléotides ciblée par l'ARNi (Boden et al. 2003) ou en modifiant la structure secondaire entourant la séquence cible pour bloquer son accessibilité (Westerhout et al. 2005).

Malgré des avancements prometteurs dans le développement de nouveaux agents anti-VIH, l'apparition de nouvelles résistances reste inévitable. Le développement de nouveaux traitements, moins dispendieux et moins toxiques que ceux présentement utilisés, ne règlera pas le problème de la pandémie mondiale, mais pourra donner le temps nécessaire au développement de vaccins qui pourraient éventuellement éradiquer la maladie. En attendant, la propagation du VIH-1, qui a déjà causé 16 millions de décès,

continue de s'accroître avec 6 millions de nouvelles infections qui s'ajoutent annuellement aux 36 millions de personnes déjà infectées mondialement.

### 1.4.1 Les différents groupes et sous-types du VIH-1

Le VIH-1 provient de transmissions inter-espèces chimpanzé-homme du virus de l'immunodéficience simienne (SIV<sub>cpz</sub>) et a été isolé d'un patient pour la première fois en 1983 (Barre-Sinoussi et al. 1983). Les analyses phylogénétiques suggèrent toutefois que la première transmission du chimpanzé à l'homme s'est produite autour des années 1930 et que deux autres transmissions inter-espèces se sont également produites à cette époque ou plus récemment (Korber et al. 2000; Sharp et al. 2000; Salemi et al. 2001). Chacune de ces trois transmissions aurait mené à l'apparition d'un groupe du VIH-1 : M (*main*), N (*new* ou non-M, non-O) et O (*outlier*) (Figure 1-16A). L'accumulation de mutations dans le génome de ces groupes a généré plusieurs sous-types, principalement en raison de l'absence d'activité correctrice 5'→3' exonucléase (*proofreading*) de la transcriptase reverse qui introduit environ une mutation par cycle de rétrotranscription de l'ARN viral en ADN (Roberts et al. 1988) et de la rapidité à laquelle se réplique le génome viral.

Le groupe M est responsable de 99% des infections dans le monde. L'analyse phylogénétique des séquences des gènes viraux *gag*, *pol* ou *env* a permis de subdiviser le groupe M en neuf sous-types : A, B, C, D, F, G, H, J et K (McCutchan 2000; Robertson et al. 2000; Moore et al. 2001). Les divergences de séquence à l'intérieur d'un même sous-type sont de l'ordre de 5 à 20%, alors que les divergences entre les sous-types vont de 20 à 30% (Los Alamos National Laboratory 2002, <http://hiv-web.lanl.gov/content/hiv-db/mainpage.html>). La recombinaison entre les différents sous-types augmente également la diversité virale et certains de ces virus recombinants jouent un rôle important dans l'épidémie du VIH-1 (Peeters et Sharp 2000). Les sous-types E et I ne sont pas inclus dans la liste des sous-types puisqu'ils ont uniquement été identifiés sous formes recombinantes. Pour l'instant, le sous-type B fait l'objet de recherches intensives en raison de sa prévalence

en Amérique du Nord et en Europe de l'Ouest, où la majorité des recherches sur le VIH-1 sont effectuées (Figure 1-16A). Par contre, les formes virales prédominantes dans le monde sont le sous-type A (Afrique de l'Ouest et Afrique centrale), suivie du sous-type C (Afrique australe), du CRF01\_AE (forme recombinante circulante), du CRF02\_AG et du sous-type B (Figure 1-16B).

Le groupe N est encore mal défini, puisque très peu d'isolats sont présentement disponibles. Les membres du groupe N ont la particularité d'avoir un gène *env* proche du SIV<sub>cpz</sub>, alors que les gènes *gag* et *pol* ressemblent au groupe M du VIH-1 (Simon et al. 1998; Corbet et al. 2000). L'analogie de séquence entre les gènes *env* du SIV<sub>cpz</sub> et du groupe N du VIH-1 suggère une transmission récente du chimpanzé à l'homme (Gao et al. 1999).

Le groupe O, dont la séquence diffère à 60% de celle du groupe M, se divise en cinq sous-types (Yamaguchi et al. 2002). Deux de ces cinq sous-types sont principalement étudiés : ANT70, qui a été le premier membre du groupe O isolé en 1990 (De Leys et al. 1990), et MVP5180, isolé pour la première fois en 1994 et qui semble être le sous-type du groupe O le plus répandu. Présentement, les infections avec le groupe O sont concentrées en Afrique centrale, principalement au Cameroun et en Guinée équatoriale (Figure 1-16C) (Peeters et al. 1997). Récemment, des recombinants entre le groupe M et le groupe O ont été identifiés (Peeters et al. 1999; Takehisa et al. 1999).

La diversité génétique du VIH-1, combinée à la recherche systématique sur le sous-type B du groupe M, entraîne des problèmes aux niveaux du dépistage et du traitement. En effet, les systèmes de détection commerciaux disponibles ne détectent pas la charge virale avec la même efficacité chez les patients infectés avec les sous-types non-B du groupe M et ne permettent pas la détection du groupe O. Dans un même ordre d'idée, les membres du groupe O et les sous-types non-B du groupe M possèdent des variations naturelles qui diminuent l'efficacité des inhibiteurs de protéase (Velazquez-Campoy et al. 2001) et conduisent à une résistance plus rapide à ces inhibiteurs (de Baar et al. 2000; Yang et al.

2000; Luk et al. 2001). De plus, le groupe O est naturellement résistant aux NNRTI (Descamps et al. 1997; Vergne et al. 2000). Cette grande diversité génétique du VIH-1, qui provoque également des difficultés dans le développement de vaccins (Moore et al. 2001), pourrait encore se diversifier via la transmission de nouvelles souches de SIV à l'homme. En effet, il existe présentement plus de 20 virus SIV différents, dont certains peuvent être cultivés dans des lymphocytes humains. Jusqu'à aujourd'hui, trois transmissions inter-espèces ont mené à la pandémie mondiale du VIH-1, alors qu'au moins six transmissions du *sooty mangabeys* (SIV<sub>sm</sub>) à l'homme ont mené à l'émergence du VIH-2, un autre type de VIH beaucoup moins virulent. Une surveillance accrue des différentes souches de SIV est donc primordiale (Hahn et al. 2000).

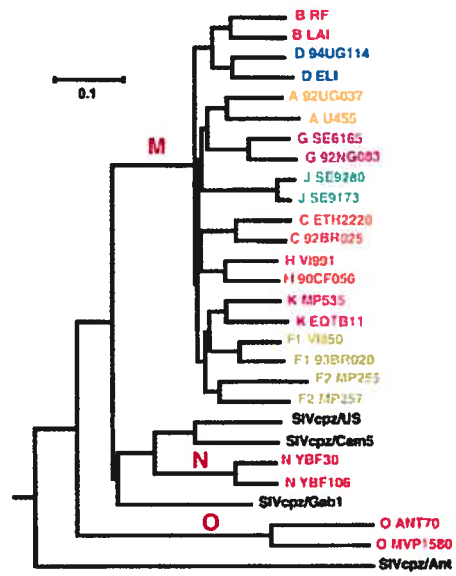
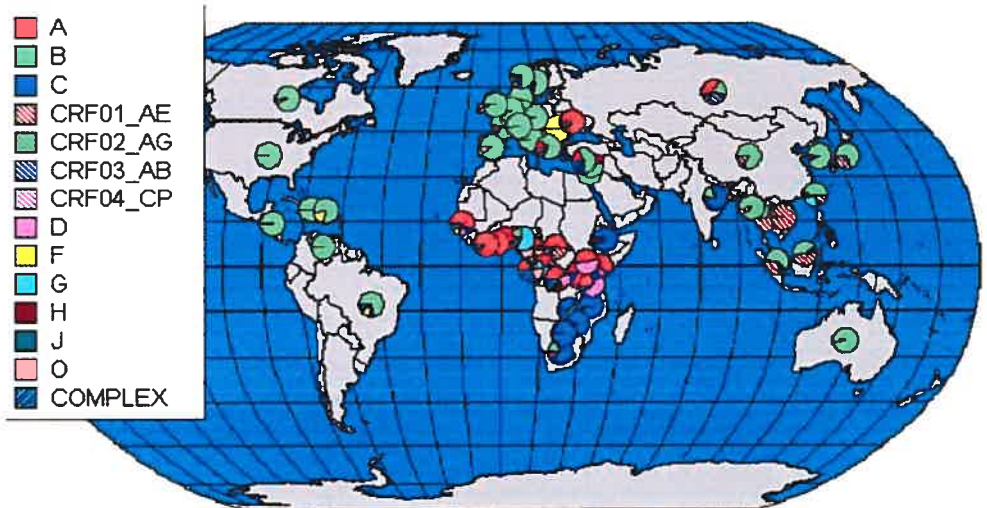
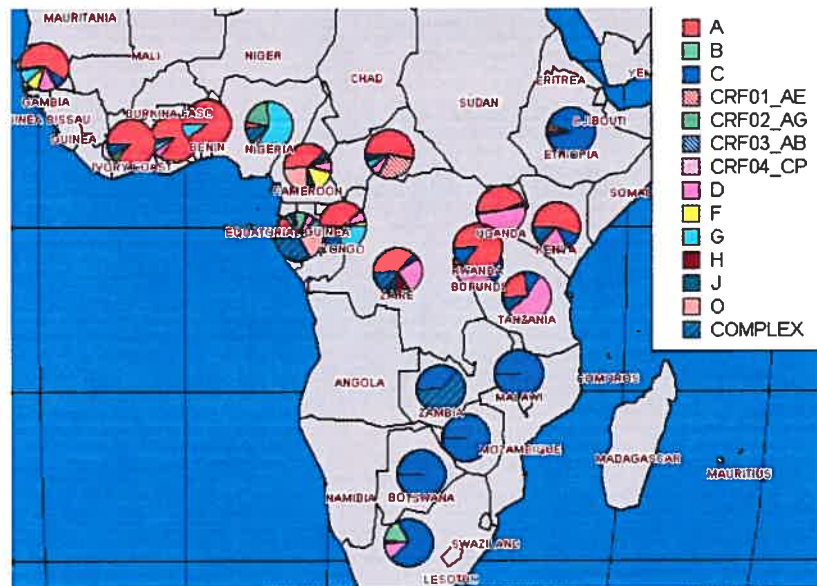
Figure 1-16 : Vue d'ensemble de la diversité génétique et de l'épidémiologie des différents groupes et sous-types du VIH-1.

A) Arbre phylogénétique des différents groupes et sous-types du VIH-1 (tiré de Sharp 2002).

B) Répartition et prévalence des différents groupes et sous-types du VIH-1 mondiale (tiré de Los Alamos National Laboratory 2002).

CRF : forme recombinante circulante, complex : forme recombinante entre plus de deux sous-types.

C) Répartition et prévalence des différents groupes et sous-types du VIH-1 sur le continent africain (tiré de Los Alamos National Laboratory 2002).

**A****B****C**

## 1.4.2 Le génome du VIH-1

Le génome du VIH-1 est composé d'un ARN(+) d'environ 9 kb qui contient neuf cadres de lecture codant pour 15 protéines (revue par Frankel et Young 1998) (Figure 1-17A). Neuf de ces protéines sont synthétisées sous forme de trois précurseurs polyprotéiques : Gag, Pol et Env. Gag est le précurseur des protéines structurales du virus : la matrice (MA), qui cible Gag à la membrane cellulaire via son site myristylé, la capsid (CA), qui forme une carapace autour du complexe ribonucléoprotéique, la nucléocapsid (NC), qui lie l'ARN génomique viral, et un court peptide (p6), qui favorise le relâchement des particules virales. Pol est le précurseur des enzymes virales : la protéase (PR), qui clive les précurseurs polyprotéiques après le relâchement des particules virales immatures par les cellules infectées, la transcriptase réverse (RT), qui copie le génome ARN du VIH-1 en ADN, et l'intégrase (IN), qui intègre l'ADN viral dans les chromosomes de la cellule hôte. Env est le précurseur des protéines de l'enveloppe du virus : la protéine de surface (SU ou gp120) et la protéine transmembranaire (TM ou gp41) qui vont permettre la reconnaissance et l'entrée dans les cellules lymphocytaires T CD4<sup>+</sup>. Les six autres protéines produites sont les protéines régulatrices Tat et Rev et les protéines accessoires Vpr, Vpu, Nef et Vif, spécifiques aux lentivirus (revue par Cullen 1998). Tat permet d'augmenter la processivité de l'ARN polymérase II, Rev permet l'exportation des ARN viraux non-épissés et mono-épissés vers le cytoplasme, Vpr est responsable de l'importation du complexe de préintégration (PIC) dans le noyau de la cellule infectée et de l'arrêt du cycle cellulaire en phase G2, Vpu, qui est unique au VIH-1, cause la dégradation des récepteurs CD4, Nef provoque la dégradation des récepteurs CD4 et des molécules du complexe majeur d'histocompatibilité de classe I (MHCI) en plus d'augmenter l'infectivité des virions et, finalement, Vif permet d'augmenter l'infectivité des virions en entraînant la dégradation par ubiquitination d'APOBEC3G, une désaminase cellulaire conférant une immunité innée contre les infections rétrovirales (revue sur vif par Rose et al. 2004). Ces protéines régulatrices et accessoires assurent une propagation efficace du VIH-1 en favorisant son

efficacité de réplication, sa capacité à déjouer le système immunitaire et sa persistance dans les lymphocytes T mémoires.

La synthèse de Gag et de Pol se fait à partir du même ARN viral pleine longueur, mais la séquence codante de Pol est dans un cadre de lecture  $-1$  relativement à la séquence codante de Gag. En fait, l'expression de Pol dépend d'un *frameshift* programmé  $-1$  (voir section 1.4.4). La majorité des ribosomes qui traduisent l'ARN viral pleine longueur produisent Gag, mais une minorité de ribosomes vont effectuer un *frameshift*  $-1$  au niveau d'une séquence précise de l'ARN viral pour produire la protéine de fusion Gag-Pol. Ce terme est utilisé parce que le début de la protéine est identique à Gag, mais qu'après le *frameshift*  $-1$ , la protéine correspond à l'expression du gène *pol*.



Figure 1-17 : Représentation du génome et du virion du VIH-1 (tiré de Coffin et al. 1997).

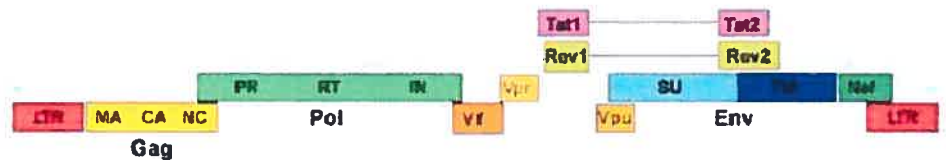
A) Représentation de l'organisation du génome du VIH-1.

Gag : précurseur des protéines de structure, Pol : précurseur des enzymes, Env : précurseur des protéines de l'enveloppe, LTR : *long terminal repeat* (voir section 1.4.3), MA : matrice, CA : capsid, NC : nucléocapsid, PR : protéase, RT : transcriptase réverse, IN : intégrase, SU : protéine de surface de l'enveloppe, TM : protéine transmembranaire de l'enveloppe.

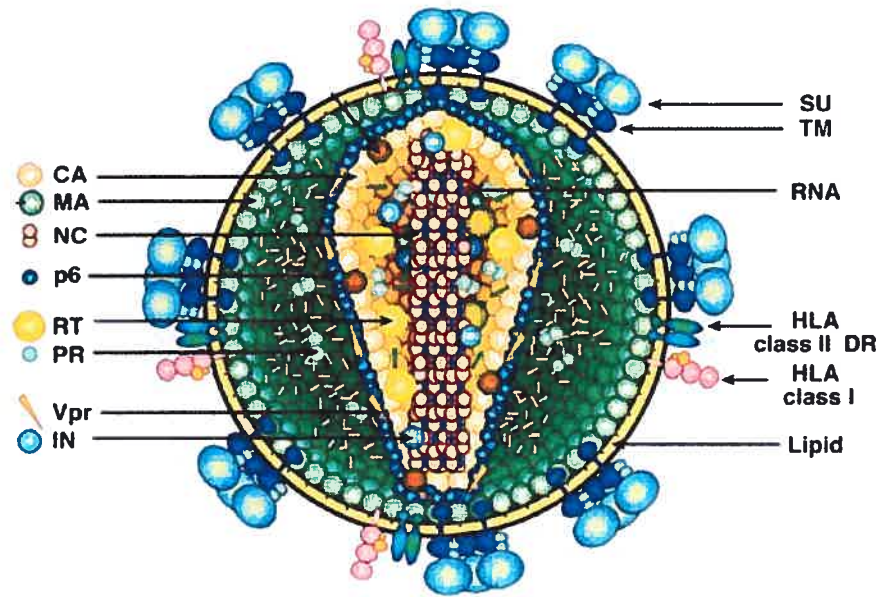
B) Représentation schématique de la structure du virion mature du VIH-1.

HLA : antigène leucocytaire humain.

**A**



**B**



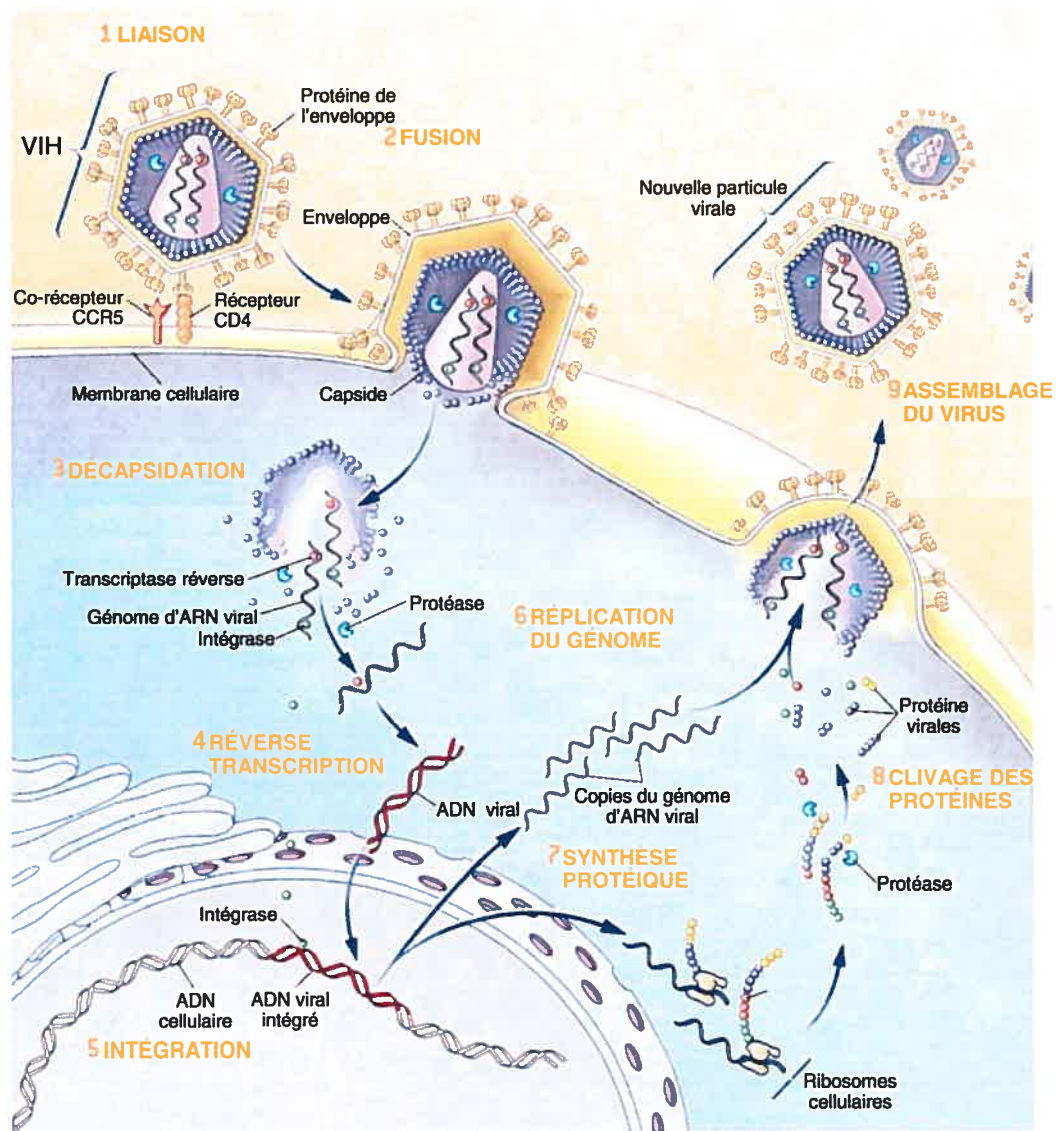
### 1.4.3 Cycle de réplication du VIH-1

Lors de l'infection, la protéine de l'enveloppe SU du VIH-1 se lie au récepteur CD4 et au co-récepteur CCR5 (ou CXCR4 lors de l'infection tardive) des cellules cibles, principalement les cellules dendritiques, les macrophages et les lymphocytes (Figure 1-18). L'enveloppe du virus se fusionne alors avec la membrane grâce aux protéines SU et TM. Une fois entré dans la cellule, le virus est décapsidé pour former le complexe de réverse transcription. La RT virale synthétise un ADN(-) complémentaire au génome d'ARN à partir de l'ARNt<sup>Lys3</sup> qui est encapsidé dans les virions et sert d'amorce à la RT. Le brin d'ARN est ensuite dégradé par l'activité intrinsèque RnaseH de la RT. Un brin d'ADN(+) est ensuite synthétisé par la RT pour former un ADN génomique viral double-brin, formant ainsi le complexe de préintégration (PIC). Le PIC est ensuite importé dans le noyau et intégré dans le génome de l'hôte par l'intégrase virale.

Une fois intégré, le génome viral sera exprimé et produira d'autres virions. Pour débiter, il y a transcription à un niveau basal du génome viral pleine longueur par l'ARN polymérase II cellulaire à partir des promoteurs du virus (ou *long terminal repeats*, LTR). Les ARNm viraux multi-épissés (Tat et Rev) sont exportés vers le cytoplasme par le mécanisme cellulaire d'exportation traditionnel. Une fois synthétisées, les protéines précoces Tat et Rev vont permettre, respectivement, une transcription plus efficace des ARNm viraux et l'exportation vers le cytoplasme des ARNm viraux non-épissés et mono-épissés. Il y a alors traduction des protéines virales tardives, assemblage à la membrane, encapsidation de deux copies de l'ARN génomique viral et relâche des particules virales par bourgeonnement de la membrane cellulaire. Les virions immatures relâchés sont ensuite maturés par la protéase virale, produisant ainsi des virions pouvant infecter d'autres cellules cibles.

Figure 1-18 : Cycle de réplication du VIH-1 (adapté de Haseltine 2001).

- 1) LIAISON : la protéine de l'enveloppe lie le récepteur CD4 et le co-récepteur CCR5.
- 2) FUSION : les membranes virale et cellulaire fusionnent.
- 3) DÉCAPSIDATION : le virus est décapsidé et relâche les enzymes et l'ARN du virus dans le cytoplasme.
- 4) RÉVERSE TRANSCRIPTION : l'ARN génomique viral est réverse transcrit en ADN génomique viral double-brin pour former le complexe de préintégration (PIC) qui est importé dans le noyau de la cellule infectée.
- 5) INTÉGRATION : le PIC est intégré dans le génome de l'hôte par l'intégrase virale.
- 6) RÉPLICATION DU GÉNOME : les ARNm viraux non-épissés sont exportés dans le cytoplasme par Rev et serviront d'ARN génomique viral lors de l'assemblage de nouveaux virions (Étape 9).
- 7) SYNTHÈSE PROTÉIQUE : les ARNm viraux sont traduits en protéines virales.
- 8) CLIVAGE DES PROTÉINES : la protéase virale clive les précurseurs polyprotéiques en protéines virales matures (NOTE : la plupart de ces clivages se font après la relâche du virus).
- 9) ASSEMBLAGE DU VIRUS : assemblage des nouvelles particules virales à la membrane cellulaire et relâche des virions par bourgeonnement.



#### 1.4.4 Le *frameshift* programmé -1 du VIH-1

L'assemblage du VIH-1 est un mécanisme qui est influencé par l'efficacité du *frameshift* programmé -1 qui contrôle le rapport Gag-Pol/Gag. Ce rapport est critique pour l'assemblage des particules rétrovirales, la réplication et l'infectivité du VIH-1 (Park et Morrow 1991; Karacostas et al. 1993). De plus, il a été démontré qu'une augmentation (Shehu-Xhilaga et al. 2001) ou une diminution (Telenti et al. 2002) d'un facteur trois du rapport Gag-Pol/Gag diminuait l'infectivité du VIH-1 de 10 à 100 fois. Le *frameshift* -1 est donc une cible très intéressante pour le développement de nouveaux agents anti-VIH qui affecteraient l'efficacité de ce *frameshift*.

Chez le VIH-1, la séquence glissante est U UUU UUA (Jacks et al. 1988b), suivie d'un espaceur de huit nucléotides et d'un signal stimulateur. Initialement, il avait été montré que la structure secondaire du signal stimulateur du *frameshift* -1 du VIH-1 formait une tige-boucle de 11 paires de bases (tige-boucle classique, Figure 1-19A) (Jacks et al. 1988b; Kang 1998). Toutefois, deux études récentes ont démontré que la présence de la séquence en aval de la tige-boucle classique augmente l'efficacité de *frameshift* -1 du VIH-1 de deux fois en permettant la formation d'un signal stimulateur plus complexe (Dinman et al. 2002; Dulude et al. 2002). Ces deux études proposaient cependant des structures secondaires différentes pour le signal stimulateur complet du VIH-1. Le groupe de Dinman suggérait une structure en pseudonoeud où la séquence riche en pyrimidines en aval de la tige-boucle classique forme trois paires de bases Watson-Crick avec la boucle de cette tige-boucle, en plus de former une triple hélice avec les quatre paires de bases du haut de la tige (Dinman et al. 2002, Figure 1-19C). Une critique majeure de ce modèle est que le brin riche en pyrimidines qui forme la triple hélice est dans une orientation antiparallèle par rapport au brin purine de la tige-boucle avec lequel il interagit, une orientation stériquement défavorable qui rend la formation de cette structure hautement improbable (Figure 1-19C) (Han et Dervan 1993; Frank-Kamenetskii et Mirkin 1995; Lim et Barton 1998). Notre groupe de recherche a par contre démontré que le signal stimulateur du VIH-1 consiste en

fait en une hélice irrégulière interrompue par une protubérance (*bulge*) de trois purines (Dulude et al. 2002, Figure 1-19B). Dans cette structure, démontrée chez le sous-type B du groupe M du VIH-1, la tige-boucle supérieure correspond à la tige-boucle classique et la tige inférieure résulte d'un pairage entre l'espaceur et une séquence complémentaire en aval de la tige-boucle supérieure.

Cette région de *frameshift* -1 a donc la caractéristique de ne pas avoir d'espaceur, ce qui force le ribosome en élongation à défaire la tige inférieure grâce à son activité hélicase avant de se trouver au niveau de la séquence glissante. Un mécanisme où la protubérance de purines interagit spécifiquement avec une région non-identifiée du ribosome en élongation avant que celui-ci ne traduise la séquence glissante a donc été proposé (Figure 1-19D). Cet ancrage de la protubérance sur le ribosome faciliterait une interaction adéquate entre la tige-boucle supérieure et le ribosome, interaction qui serait maintenue jusqu'à ce que la séquence glissante atteigne les sites A et P du ribosome, où elle stimulerait le *frameshift* -1 (Léger et al. 2004).

Figure 1-19 : Structures secondaires proposées pour le signal stimulateur du *frameshift* -1 du VIH-1.

A) Signal stimulateur classique sous forme de simple tige-boucle (inspiré de Jacks et al. 1988b).

La séquence glissante est identifiée en italique.

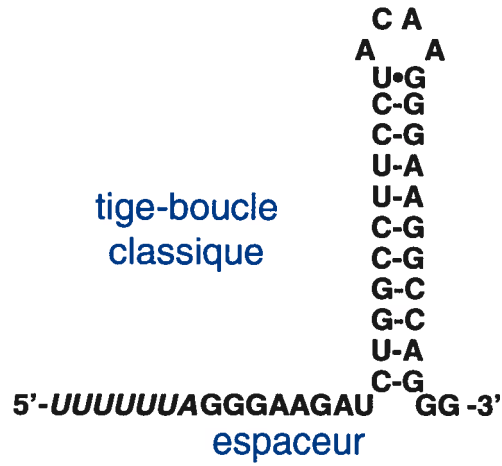
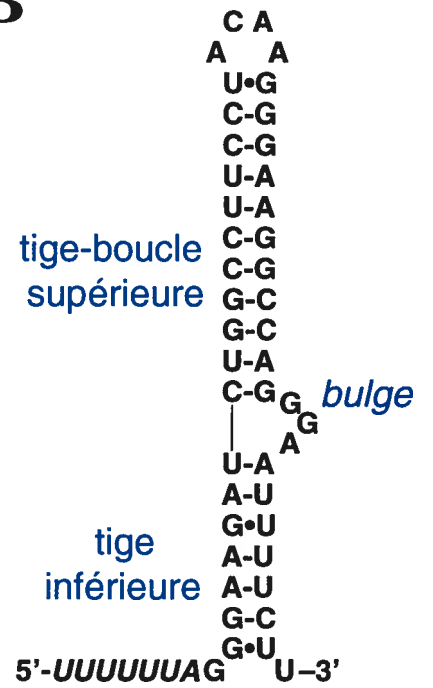
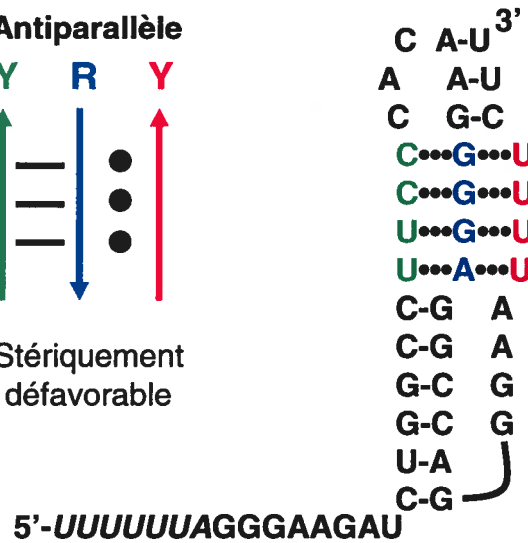
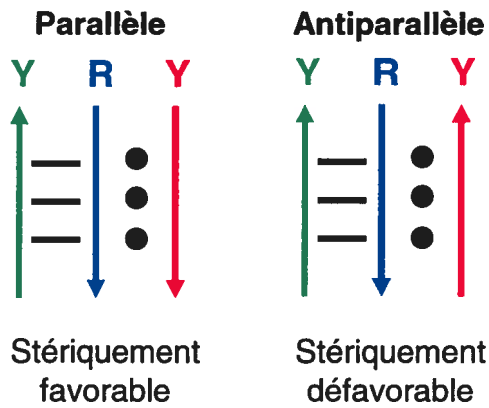
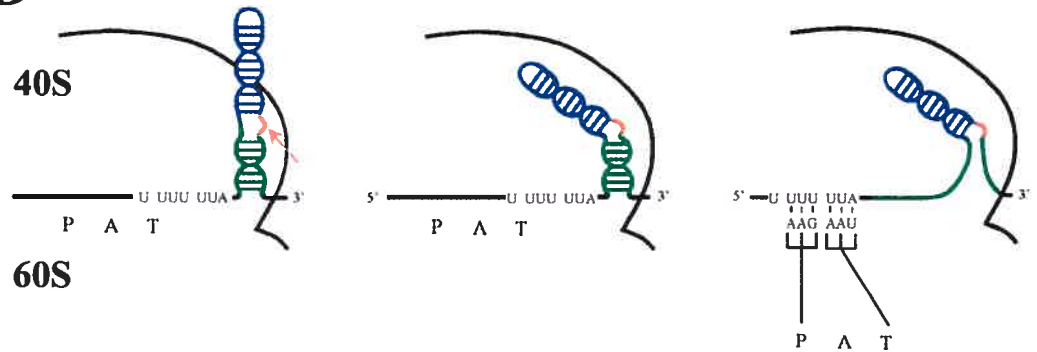
B) Signal stimulateur en tige-boucle irrégulière interrompue par une protubérance de trois purines (adapté de Dulude et al. 2002).

C) Signal stimulateur en pseudonoeud où le brin pyrimidine qui forme une triple hélice d'ARN est dans une orientation antiparallèle par rapport au brin purine de la tige-boucle (inspiré de Dinman et al. 2002).

D) Mécanisme proposé pour l'interaction entre le signal stimulateur du VIH-1 et le ribosome (adapté de Léger et al. 2004).

La protubérance de trois purines est représentée en rouge et identifiée par une flèche. La tige-boucle supérieure et la tige inférieure sont représentées en bleu et en vert, respectivement.



**A****B****C****Triple hélice d'ARN****D**

## 1.5 Le virus de l'hépatite C (VHC)

Le VHC, un membre de la famille des *Flaviviridae*, a été identifié pour la première fois en 1989 (Choo et al. 1989). Le VHC se divise en six génotypes principaux qui sont numérotés de 1 à 6, le génotype 1 étant le plus répandu mondialement. Ces génotypes sont également subdivisés en plusieurs sous-types qui sont identifiés par une lettre minuscule (revue par Howard C. R. 2002). Mondialement, environ 200 millions de personnes sont infectées chroniquement par le VHC. Les patients infectés chroniquement ont un risque élevé de développer une cirrhose ou un cancer du foie (Saito et al. 1990; Scheuer et al. 1992; Lefkowitz et al. 1993), ce qui fait de l'infection par le VHC la cause principale de transplantation du foie dans les pays développés (Willems et al. 2002).

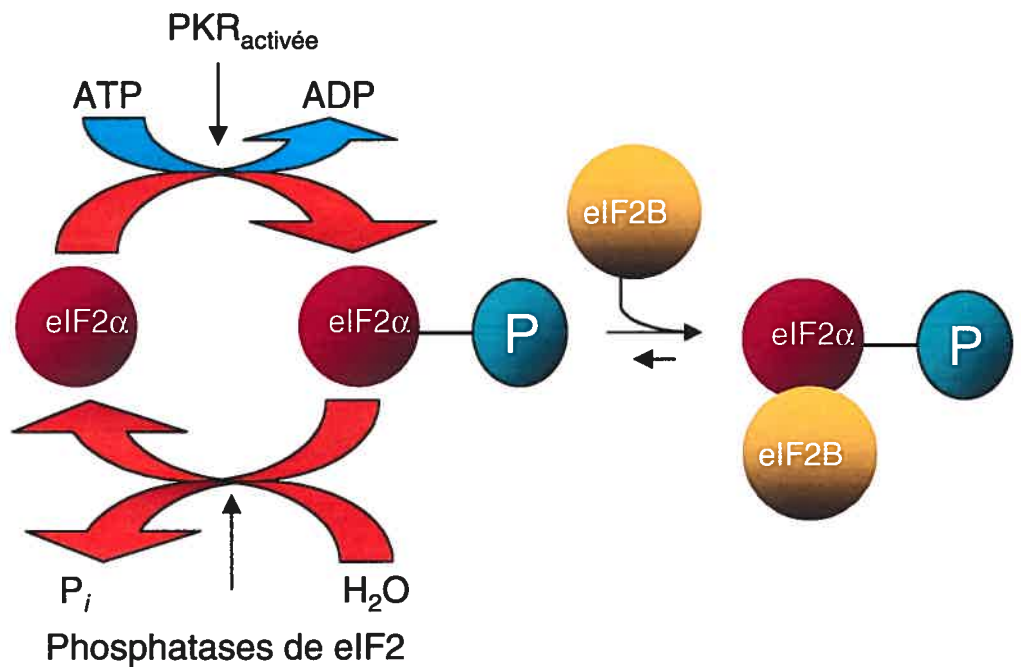
Malheureusement, les traitements disponibles pour traiter l'infection chronique à l'hépatite C sont très limités. Il n'y a présentement aucun agent anti-VHC spécifique qui puisse bloquer la production de virus, ni vaccin pouvant prévenir l'infection, le développement de traitements spécifiques étant ralenti par l'absence d'un système d'infection dans des cellules en culture (voir section 1.5.2). Pour l'instant, seul le traitement à l'interféron-alpha ( $\text{INF-}\alpha$ ), une glycoprotéine endogène sécrétée par les leucocytes en réponse à la présence d'ARN double-brin typiquement associée à des infections virales, est disponible. L'activité antivirale de l' $\text{INF-}\alpha$  provient principalement de l'inhibition de la synthèse protéique en induisant la synthèse de la protéine kinase ARN-dépendante (PKR) et de la (2',5')-oligoadénylate synthétase (revue par Williams 1999). En présence d'ARN double-brin, PKR inhibe l'initiation de la traduction en phosphorylant eIF-2 qui va séquestrer eIF2B et ainsi empêcher le recyclage d'eIF2-GDP en eIF2-GTP (Figure 1-20, voir section 1.2.1.1). PKR active également le facteur de transcription NF- $\kappa$ B qui contrôle l'expression d'une série de gènes reliés à l'apoptose et à l'amplification de la réponse antivirale (revue par Gil et Esteban 2000). De son côté, en présence d'ARN double-brin, la (2',5')-oligoadénylate synthétase cause l'activation de l'ARNase L, une endonucléase qui

dégrade l'ARNm. Initialement, le traitement à l'INF- $\alpha$  permettait d'éliminer complètement le VHC dans seulement 15% des cas. Le taux de guérison a été augmenté à 40% grâce à la combinaison de l'INF- $\alpha$  avec la ribavirine, un analogue nucléosidique de la guanosine qui vide le stock intracellulaire de GTP en inhibant l'inosine monophosphate déshydrogénase qui est impliquée dans la synthèse du GTP (Lau et al. 2002; Dixit et al. 2004). Finalement, la modification de l'INF- $\alpha$  avec du polyéthylène glycol (PEG) a permis de diminuer la dose, la fréquence de l'administration et les effets secondaires du traitement, en plus d'augmenter le taux de guérison à environ 50%. Ce taux de guérison atteint 80% pour les génotypes 2 et 3 du VHC, mais est seulement de 40% chez le génotype 1 (Zein 2000).

En se basant sur les agents anti-VIH présentement disponibles, plusieurs recherches visent actuellement à développer des agents anti-VHC ciblant les enzymes virales (revue par Tan et al. 2002). Pour l'instant, un seul de ces agents anti-VHC, un inhibiteur de la protéase virale (BILN 2061), a obtenu des résultats prometteurs (Lamarre et al. 2003). En plus de cibler les enzymes du VHC, d'autres traitements sont en cours de développement, dont des analogues de la ribavirine (ex : viremagine), des ARN antisens (Witherell 2001), des ribozymes (Macejak et al. 2000) et des ARNi (Randall et al. 2003).

Figure 1-20 : Mécanisme d'inhibition de l'initiation de la traduction par PKR.

La phosphorylation de la sous-unité  $\alpha$  de eIF2 augmente son affinité pour eIF2B, un facteur limitant de l'initiation de la traduction, empêchant ainsi le recyclage de eIF2-GDP en eIF2-GTP. La carence en eIF2-GTP inhibe l'initiation de la traduction.



### 1.5.1 Le génome du VHC

Le VHC, un Hepacivirus avec un génome composé d'ARN(+) d'environ 9,6 kb, contient un grand cadre de lecture codant pour une polyprotéine d'environ 3000 acides aminés (Figure 1-21). La portion 5'UTR du génome viral contient un IRES qui contrôle l'initiation de la traduction de la polyprotéine du VHC (Ji et al. 2004). Durant et après sa synthèse, la polyprotéine est clivée par des protéases cellulaires pour produire quatre protéines structurales (*core*, E1, E2 et p7) et par des protéases virales contenues dans la polyprotéine pour produire six protéines non-structurales (NS2, NS3, NS4A, NS4B, NS5A et NS5B) (revue par Bartenschlager et Lohmann 2000). Enfin, la région 3'UTR du génome contient trois régions indispensables à l'infectivité *in vivo* : une région hypervariable, une région polypyrimidine (poly U-UC) et une région hautement conservée de 98 nucléotides formant trois tiges-boucles (3'X) (Yanagi et al. 1999).

Les protéines structurales du VHC permettent l'entrée du virus lors de l'infection, ainsi que l'assemblage de nouveaux virions pour permettre la propagation de l'infection. En plus de former la nucléocapside du virion (Figure 1-21B), la protéine *core* est impliquée dans la suppression de la réponse immunitaire (Large et al. 1999), la transformation cellulaire (Ray et al. 1996; Jin et al. 2000) et l'induction du cancer du foie (Moriya et al. 1998). La protéine *core* jouerait donc un rôle possible dans la persistance du virus et dans le développement du cancer (revue par Watashi et Shimotohno 2003). Les glycoprotéines de l'enveloppe E1 et E2 forment un hétérodimère qui permet l'entrée dans les cellules cibles, les hépatocytes (Op De Beeck et al. 2000). E1 permettrait la fusion des membranes virale et cellulaire (Ciccaglione et al. 2001), alors que E2 interagirait avec des récepteurs cellulaires : la tétraspanine CD81 (Pileri et al. 1998) et le récepteur des lipoprotéines de faible densité (LDL) (Agnello et al. 1999) ou le récepteur des lipoprotéines à haute densité (HDL) (Scarselli et al. 2002). Il a récemment été démontré que la protéine hydrophobe p7 peut s'oligomériser pour former un canal ionique qui serait responsable d'un transfert

d'ions calcium du réticulum endoplasmique (RE) vers le cytoplasme afin d'assurer un repliement adéquat des glycoprotéines E1 et E2 (Pavlovic et al. 2003).

Figure 1-21 : Représentation du génome, du virion et des réplicons du VHC.

A) Représentation de l'organisation du génome du VHC.

La région 5'UTR contient un IRES (voir Figure 1-3A). Les flèches rouges représentent les sites de clivage par des protéases cellulaires, la flèche bleue représente le site d'autoclivage de la protéase virale NS2-3 et les flèches jaunes représentent les sites de clivages par la protéase virale NS3. C : *core*, F : protéine F encodée dans le cadre de lecture +1 de la polyprotéine (voir section 1.5.3).

B) Représentation schématique de la structure hypothétique du virion du VHC.

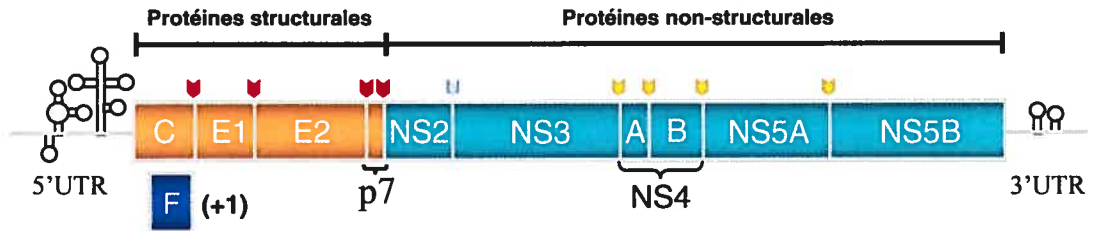
C) Structure d'un réplicon du VHC contenant une portion du génome viral (tiré de Bartenschlager et al. 2003).

Le gène de la néomycine phosphotransférase (*neo*) qui confère la résistance à la généticine (G418) est exprimé sous le contrôle de l'IRES du VHC. L'IRES du virus de l'encéphalomyocardite (EMCV) contrôle l'expression des protéines NS3 à NS5B.

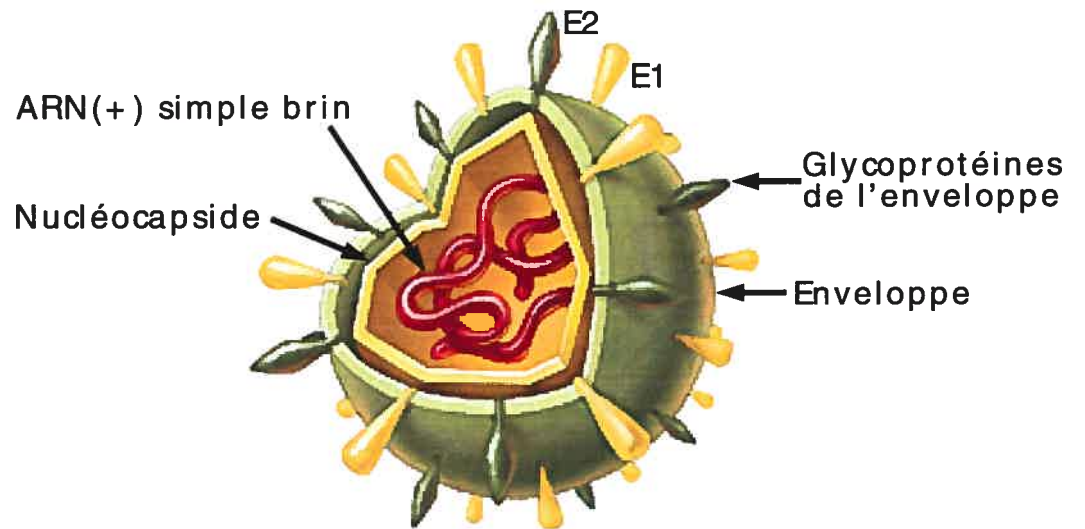
D) Structure du réplicon contenant le génome complet du VHC (tiré de Bartenschlager et al. 2003).



**A**



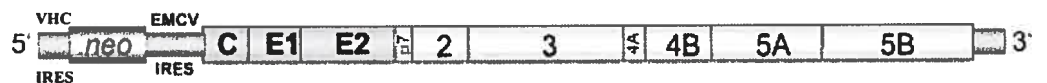
**B**



**C**



**D**



Les protéines non-structurales du VHC contrôlent la réplication du génome viral et la maturation des protéines non-structurales du virus. NS2 et le domaine N-terminal de NS3 forment la cystéine protéase NS2-3, qui catalyse le clivage du site séparant NS2/NS3 (Grakoui et al. 1993; Pallaoro et al. 2001). La portion N-terminale de NS3 possède une activité sérine protéase, qui catalyse le clivage des protéines non-structurales aux sites : NS3/NS4A, NS4A/NS4B, NS4B/NS5A et NS5A/NS5B. La formation d'un hétérodimère entre NS3 et son co-facteur NS4A est requise à l'activité maximale de la protéase (Bartenschlager et al. 1995). La portion C-terminale de NS3 possède une activité ATPase/hélicase qui permet de défaire des hélices d'ARN (Kim D. W. et al. 1995). NS4B est une protéine très hydrophobe qui induirait la formation de vésicules membranaires qui s'accumuleraient dans le cytoplasme et formeraient le site de réplication de l'ARN du VHC (Gosert et al. 2003). NS5A est une métalloprotéine essentielle à la réplication du génome viral qui pourrait agir comme co-facteur de NS5B (Tellinghuisen et al. 2004). Son rôle précis dans le cycle de réplication n'a toutefois pas encore été déterminé. NS5A contient également une région impliquée dans la sensibilité à l'IFN (ISDR), dont la séquence serait un déterminant majeur de la résistance au traitement à l'IFN- $\alpha$ . De plus, NS5A lie et inhibe PKR, contribuant ainsi à la résistance à l'INF- $\alpha$  (Gale et al. 1998). Finalement, NS5B est une ARN polymérase ARN-dépendante qui catalyse la réplication du génome viral (Behrens et al. 1996).

### **1.5.2 Le cycle de réplication du VHC**

La caractérisation du cycle de réplication du VHC a été ralentie par l'absence d'un système adéquat d'infection. En effet, le seul animal pouvant être infecté par le VHC est le chimpanzé, alors qu'un système efficace d'infection dans des cellules en culture Huh7 (hépatocytes cancéreux humain) utilisant un variant du génotype 2a entraînant une hépatite fulminante (souche JFH1) vient tout juste d'être décrit (Lindenbach et al. 2005; Wakita et al. 2005; Zhong et al. 2005). Préalablement, le seul système efficace pour étudier la

réplication du VHC était le réplicon, développé par le groupe de Bartenschlager en 1999 (Lohmann et al. 1999; revue par Bartenschlager et al. 2003). Ce système de réplicon consiste à sélectionner les cellules Huh7 qui expriment efficacement le génome du VHC, en utilisant la généticine (G418) comme moyen de sélection. Le gène de résistance au G418 (*neo*) est exprimé à partir de l'IRES du VHC, alors que les protéines virales sont exprimées à partir de l'IRES du virus de l'encéphalomyocardite (EMCV). Pour permettre une réplication optimale du génome viral, le premier système de réplicon développé ne contenait que les protéines virales NS3 à NS5B (Figure 1-21C), limitant ainsi la taille du génome et éliminant les protéines structurales du VHC en raison de leur potentiel cytotoxique. Le séquençage des clones infectieux a identifié plusieurs mutations adaptatives dans les génomes viraux qui assurent une réplication plus efficace du génome viral. L'insertion de ces mutations dans le système de réplicon a augmenté jusqu'à 10 000 fois le nombre de cellules qui permettent la réplication du VHC (Blight et al. 2000). Ce premier système de réplicon a permis d'étudier les différentes étapes de la réplication du génome du VHC, sans toutefois permettre l'étude des étapes de l'entrée du virus dans la cellule, ni des étapes d'assemblage des particules virales. La caractérisation de ces différentes étapes pourra maintenant être entreprise grâce au système d'infection nouvellement mis au point (Lindenbach et al. 2005; Wakita et al. 2005; Zhong et al. 2005).

La première étape du cycle de réplication est l'entrée du virus dans la cellule. Cette étape est encore mal caractérisée chez le VHC, mais une interaction entre la glycoprotéine virale E2 et la tétraspanine CD81 permettrait l'attachement du virus aux hépatocytes, une interaction subséquente avec un récepteur des lipoprotéines assurerait l'internalisation par endocytose (Bartosch et al. 2003). La glycoprotéine virale E1 induirait ensuite la fusion des membranes virale et cellulaire. Une fois dans le cytoplasme, l'ARN(+) génomique est directement traduit par les ribosomes des cellules infectées à partir de l'IRES et il y a synthèse de la polyprotéine virale. L'efficacité d'initiation de la traduction à partir de l'IRES est augmentée par la présence de la région 3'X de l'ARN du VHC (Ito et al. 1998) et par des facteurs cellulaires, dont PTB (Ali et Siddiqui 1995) et le facteur de transcription

La (Ali et Siddiqui 1997). Une variation dans la concentration intracellulaire de ces facteurs cellulaires durant le cycle de réplication de la cellule pourrait expliquer que l'initiation IRES-dépendante soit favorisée dans les cellules qui se divisent comparativement aux cellules quiescentes (Honda et al. 2000). D'autres facteurs, dont la protéine *core* du VHC (Shimoike et al. 1999) et la vitamine B12 (Lott et al. 2001), ont un effet répresseur sur l'activité de l'IRES du VHC.

La synthèse de la polyprotéine à partir de l'IRES a lieu au niveau du RE où elle est clivée, pendant et après la traduction, par des protéases cellulaires et virales. Les protéines virales matures forment alors un complexe de réplication associé à la membrane du RE. Pour répliquer le génome, NS5B semble initier la synthèse de l'ARN complémentaire *de novo*, c'est-à-dire indépendamment d'une amorce (Luo et al. 2000). En raison de l'interférence potentielle entre la traduction (direction 5'→3') et la transcription (direction 3'→5'), une hypothèse intéressante propose que lors de la synthèse de la polyprotéine, la protéine *core* se lie à l'IRES pour empêcher la traduction de l'ARN viral, permettant ainsi à NS5B de répliquer le génome viral d'ARN(+) en ARN(-) (Zhang et al. 2002; Boni et al. 2005).

En ce qui concerne l'assemblage des virions, peu d'informations sont présentement disponibles. La protéine *core*, qui forme la nucléocapside du virus, joue vraisemblablement un rôle prédominant via sa multimérisation et ses interactions avec l'ARN génomique viral et les glycoprotéines de l'enveloppe : E1 et E2. Les virions seraient ensuite relâchés de la cellule hôte via la voie de sécrétion constitutive, en passant du RE à l'appareil de Golgi (Sato et al. 1993).

### 1.5.3 La synthèse de la protéine F du VHC

Récemment, l'existence d'une nouvelle protéine du VHC a été révélée. L'identification de cette protéine a débuté lorsqu'un cadre de lecture +1 par rapport au

cadre de lecture de la polyprotéine du virus a été identifié (Walewski et al. 2001). Ce cadre de lecture est compris dans la région codant pour la protéine *core* et est très conservé à travers les différents groupes et sous-types du VHC. De plus, des anticorps spécifiques à la protéine hypothétique correspondant à ce cadre de lecture, nommée ARFP (pour *alternate reading frame protein*) ou protéine F, ont été trouvés dans le sérum de patients infectés par le VHC. Cette observation confirmait donc l'expression de la protéine F lors de l'infection par le VHC *in vivo* (Walewski et al. 2001). Peu après, le groupe de Ou a montré que la protéine F était synthétisée suite à un *frameshift* +1 par des ribosomes ayant initié la traduction au codon AUG initiateur de la polyprotéine (Xu et al. 2001). Le radio-séquençage de la région N-terminale du peptide produit lors de la traduction du début de la séquence codante du *core* dans un RRL a permis de préciser que le site de *frameshift* +1 se localise dans la région contenant les codons 9 à 11 de la polyprotéine. L'efficacité de ce *frameshift* +1, qui génère une protéine de 16 kD, était de 30% *in vitro*, mais de seulement 1% dans des cellules Huh7 en culture. Toutefois, ces observations ont été obtenues avec une séquence du VHC qui contient une suite de 10 adénines (10A) au niveau du site de *frameshift* +1, une séquence sous-représentée chez les variants viraux (2 des 721 séquences disponibles dans la banque de génome du VHC : <http://s2as02.genes.nig.ac.jp/>). Une autre étude utilisant un système de traduction *in vitro* a démontré que la suite de 10A était nécessaire à la production de la protéine F, puisque sa production n'était plus observée lorsque la suite de 10A était remplacée par la séquence consensus (AAAGAAAAC) (Varaklioti et al. 2002). Ces études soulevaient donc un doute quant à l'expression de la protéine F *in vivo* chez les patients infectés par des variants du VHC ne contenant pas cette suite de 10A. Cependant, plusieurs études immunologiques ont démontré la présence d'anticorps et de cellules T dirigées contre la protéine F chez les patients infectés par le VHC, indépendamment du génotype et de la présence de la suite de 10A (Yeh et al. 2000; Xu et al. 2001; Boulant et al. 2003; Bain et al. 2004; Komurian-Pradel et al. 2004).

Une autre étude visant à élucider le mécanisme de synthèse de la protéine F a été faite par le groupe de Lavergne (Boulant et al. 2003). La méthode utilisée consistait à

identifier par spectrométrie de masse les peptides exprimés à partir de la séquence codante du *core* du VHC chez *Escherichia coli*. Les résultats de leur étude, qui utilisait une séquence codante du *core* ne possédant pas la suite de 10A, montraient que la protéine F était synthétisée après un *frameshift* +1 au codon 42 (AGG) de la polyprotéine, plutôt qu'au niveau des codons 9 à 11. Ce codon AGG est un codon rare chez la bactérie, mais non chez l'homme (<http://www.kazusa.or.jp/codon/>). La rareté du codon AGG chez *Escherichia coli* pourrait donc favoriser un *frameshift* +1 qui ne serait pas retrouvé chez l'homme (voir section 1.3.1.2).

Finalement, le groupe de Mavromara a caractérisé un mécanisme indépendant menant à la synthèse de la protéine F dans des cellules eucaryotes en culture (Vassilaki et Mavromara 2003). En mesurant l'expression de la protéine F fusionnée à la luciférase, ils ont montré qu'une protéine était produite après initiation de la traduction directement dans le cadre de lecture +1 à l'un ou l'autre de deux codons AUG qui chevauchent les codons 86 et 88 de la polyprotéine. Ce mécanisme mènerait à la synthèse d'une nouvelle protéine de 8 kD, plus courte que la protéine F de 16 kD produite par *frameshift* +1.

Les divergences entre toutes ces observations soulèvent des questions quant au mécanisme réel permettant la synthèse de la protéine F *in vivo* chez les patients infectés par le VHC. Malgré ces incertitudes, plusieurs études ont tenté de caractériser les propriétés de la protéine F et de déterminer son rôle dans l'infection virale. Il a été montré que la protéine F est très basique (pI ~12), qu'elle a une courte demi-vie de moins de 10 minutes et qu'elle est dégradée dans le protéasome (Roussel et al. 2003; Xu et al. 2003). Quant à sa localisation cellulaire, un groupe suggère une association au réticulum endoplasmique, comme les autres protéines du VHC (Xu et al. 2003), alors qu'un autre groupe observe une localisation cytoplasmique avec une concentration périnucléaire (Roussel et al. 2003). Du côté fonctionnel, on sait que la protéine F n'est pas nécessaire à la réplication du génome du VHC, puisque le premier système de réplicon utilisé ne possédait pas la région codant pour la protéine *core*, qui code également pour la protéine F (section 1.5.2) (Lohmann et al. 1999). Une étude a également vérifié si certains rôles associés à la protéine *core* dans la

modulation de l'activité des promoteurs transcriptionnels de *c-myc*, p53 et p21<sup>waf1</sup> du VHC n'étaient pas plutôt dus à la présence de la protéine F. Selon cette étude, les protéines F et *core* du VHC répriment l'expression de p21<sup>waf1</sup> de façon indépendante, alors que la protéine F n'affecte pas l'expression de *c-myc* et de p53 (Basu et al. 2004).

Malgré les renseignements fragmentaires disponibles sur la protéine F, plusieurs indices laissent croire que cette protéine pourrait jouer un rôle important dans l'infection virale. Premièrement, le cadre de lecture +1 est conservé à travers les différents géotypes du VHC, ainsi que chez le GB virus B, un membre des Hepacivirus dont le génome est similaire au VHC (Simons et al. 1995). Ces deux virus font partie des rares virus à ARN(+) qui peuvent causer une infection persistante, suggérant un rôle potentiel pour la protéine F dans l'établissement de la persistance (Martin et al. 2003). Le fait que la protéine F soit la seule protéine virale qui puisse être produite indépendamment des autres protéines virales suggère également qu'elle pourrait jouer un rôle régulateur important. Certaines données préliminaires suggèrent que la synthèse de la protéine F serait augmentée chez les patients souffrant de cirrhose ou de cancer du foie (revue par Branch et al. 2005). Le développement récent d'un système d'infection efficace permettra d'étudier le rôle potentiel des différentes protéines du VHC, dont la protéine F, dans l'établissement de la persistance et le développement de maladies hépatiques.

## 1.6 Description du projet de recherche

Comme mentionné à la section 1.4.4, un article publié par notre groupe, dont je suis coauteur, a démontré que le signal stimulateur du sous-type B du groupe M du VIH-1 est une tige-boucle irrégulière interrompue par une protubérance de trois purines (Dulude et al. 2002, Figure 1-19B).

La première partie de mon projet de recherche a consisté à vérifier si la tige-boucle irrégulière décrite pour le sous-type B pouvait également se former chez les autres sous-types du groupe M. Nous avons décidé de mesurer l'efficacité de *frameshift* -1 de chacun des sous-type du groupe M. L'efficacité de *frameshift* -1 a été mesurée à l'aide d'un gène rapporteur classique, la luciférase de la luciole, en amont duquel nous avons introduit la région de *frameshift* de chacun des sous-types. Pour chacune des insertions, deux types de constructions étaient produites : (0) et (-1). Dans les constructions (0), la séquence codante de la luciférase se retrouve dans le même cadre de lecture que l'AUG initiateur de la traduction placé en amont du site de *frameshift* -1. Dans ces constructions (0), la luciférase est synthétisée par les ribosomes qui n'effectuent pas de *frameshift* -1, ce qui correspond à la synthèse de Gag dans le virus. Des constructions (-1) ont également été produites pour chacun des sous-types. Dans ces constructions (-1), la luciférase est synthétisée uniquement par les ribosomes qui effectuent un *frameshift* -1, ce qui correspond à la synthèse de Gag-Pol dans le virus. Le rapport de la quantité de luciférase produite avec une construction (-1) à celle produite avec une construction (0) permet de calculer l'efficacité de *frameshift* -1, ce qui correspond au rapport Gag-Pol/Gag dans le virus. Nos résultats ont démontré que la structure du signal stimulateur et l'efficacité de *frameshift* -1 sont conservées dans les différents sous-types du groupe M, malgré les variations dans la séquence primaire et dans la stabilité thermodynamique du signal stimulateur. Les résultats de ce premier article (voir Chapitre II) confirment donc l'importance du rapport Gag-



Pol/Gag pour la réplication virale, validant le *frameshift* -1 comme cible potentielle pour le développement de nouveaux agents anti-VIH (Baril et al. 2003a).

La deuxième partie de mon projet de recherche a consisté à analyser la structure du signal stimulateur du *frameshift* -1 du groupe O du VIH-1, qui ne peut pas adopter la conformation en tige-boucle irrégulière retrouvée chez le groupe M. Des expériences de mutagénèse nous ont permis de proposer une structure en pseudonoeud pour le signal stimulateur du *frameshift* -1 du groupe O du VIH-1. Cette structure en pseudonoeud a été confirmée à l'aide d'attaques enzymatiques à la RNase V<sub>1</sub>, qui clive spécifiquement l'ARN double-brin. Finalement, nous avons modélisé la structure du pseudonoeud du groupe O *in silico*, en se basant sur la structure du pseudonoeud de SRV-1 qui avait déjà été résolue par RMN (Michiels et al. 2001, Figure 1-12A). Les résultats de ce deuxième article (voir Chapitre III) mettent en évidence la divergence de structure observée entre les signaux stimulateurs de *frameshift* -1 des groupes M et O du VIH-1 (Baril et al. 2003b). Cette divergence structurale souligne la nécessité de développer des agents anti-*frameshift* différents pour le groupe M et le groupe O du VIH-1.

La troisième partie de mon projet de recherche visait la caractérisation du mécanisme menant à la synthèse de la protéine F du VHC. Comme mentionné à la section 1.5.3, le groupe de Ou avait montré que la protéine F était synthétisée après un *frameshift* +1 au niveau d'une suite de 10A dans un système de traduction *in vitro* (Xu et al. 2001). Toutefois, on sait que les répétitions de nucléotides, comme cette suite de 10A, induisent des *frameshifts* -1 et +1 non-spécifiques dans des systèmes de traduction *in vitro* (Weiss et al. 1987; Farabaugh 1997b). La première partie de notre étude a permis de démontrer que le *frameshift* +1 au niveau de la suite de 10A observé dans un système de traduction *in vitro* ne se produit pas dans des cellules en culture. Des études subséquentes basées sur la mutagénèse nous ont permis de démontrer que la protéine F était en fait produite par initiation de la traduction directement dans le cadre de lecture +1 par rapport à la polyprotéine du VHC, à un codon non-AUG chevauchant le codon 26 de la polyprotéine

(codon 26+1). Des études de traduction *in vitro* en présence d'édéine, un inhibiteur spécifique à la liaison du Met-ARNt<sup>Met</sup> au site P du ribosome, nous ont permis de démontrer que l'initiation de la protéine F dépend du Met-ARNt<sup>Met</sup>, que le codon 26+1 soit GUG ou GCG. Notre étude identifiait donc pour la première fois un mécanisme d'initiation Met-ARNt<sup>Met</sup>-dépendant initiant la traduction à un codon qui diffère d'AUG par plus d'un nucléotide. Les résultats de ce troisième article (voir Chapitre IV) ont permis d'identifier le mode d'initiation menant à la synthèse de la protéine F, ce qui permettra une caractérisation adéquate de ses propriétés et de ses fonctions (Baril et Brakier-Gingras 2005).

## Chapitre II

### **Efficiency of a programmed –1 ribosomal *frameshift* in the different subtypes of the human immunodeficiency virus type 1 group M**

Martin Baril<sup>1</sup>, Dominic Dulude<sup>1</sup>, Karine Gendron<sup>1</sup>, Guy Lemay<sup>2</sup> and Léa Brakier-Gingras<sup>1</sup>

<sup>1</sup>Département de Biochimie et <sup>2</sup>Département de Microbiologie et d'Immunologie  
Université de Montréal  
Montréal, Québec  
H3T 1J4

RNA, 9: 1246-53, 2003.

## ABSTRACT

The synthesis of the Gag-Pol polyprotein, the precursor of the enzymes of the human immunodeficiency virus type 1 (HIV-1), requires a programmed  $-1$  ribosomal frameshift when ribosomes of the infected cells translate the viral RNA. The characteristics of this frameshift have been investigated so far only for subtype B of HIV-1 group M. In this subtype, the frameshift stimulatory signal was found to be a two-stem helix where a three-purine bulge interrupts the two stems. In this study, using a luciferase reporter system, we compare for the first time the frameshift efficiency of all the subtypes of group M. Mutants of subtype B including a natural variant were also investigated. Our results with the mutants of subtype B confirm that the bulge and the lower stem of the frameshift stimulatory signal contribute to the efficiency of the frameshift, in addition to the upper stem-loop previously considered as the sole participant. Our results also show that the frameshift stimulatory signal of all the other subtypes of group M can be folded into the same structure as in subtype B, despite sequence variations. Moreover, the frameshift efficiency of these subtypes, when assessed in cultured cells, falls within a narrow “window” (the maximal deviation from the mean value calculated from the experimental values of all the subtypes being about 35%), although the predicted thermodynamic stability of the frameshift stimulatory signal differs between the subtypes (from  $-17.2$  kcal/mol to  $-26.2$  kcal/mol). The fact that the frameshift efficiencies fall within a narrow range for all the subtypes of HIV-1 group M stresses the potential of the frameshift event as an antiviral target.

## INTRODUCTION

The nucleotide sequence of the human immunodeficiency virus type 1 (HIV-1) is highly variable, due in part to the lack of proofreading activity of the reverse transcriptase that copies the viral RNA into DNA (Roberts et al. 1988). Analysis of this genetic variation resulted into a phylogenetic tree with three main branches for HIV-1: M (main), O (outlier) and N (non-M, non-O or new). Group M accounts for more than 99% of the viral isolates responsible for the worldwide pandemic and is further subdivided into nine subtypes (A, B, C, D, F, G, H, J and K), based on their *env*, *gag* or *pol* gene sequences. Subtypes A and C (in Africa) are prevalent worldwide, followed by subtype B (in North America and Western Europe), AE recombinants (in Southeast Asia) and AG recombinants (in Western Africa) (McCutchan 2000; Robertson et al. 2000; Moore et al. 2001).

Like all retroviruses, HIV-1 contains three major genes that are expressed as precursors of structural proteins (Gag), enzymatic proteins (Pol) and envelope proteins (Env). The *gag* and *pol* products are translated from the full-length viral RNA, but *pol* lacks its own initiation codon and is in a  $-1$  reading frame relative to *gag*. Thus, HIV-1, like many other retroviruses, uses a programmed  $-1$  ribosomal frameshift to produce the precursor of its enzymes, the Gag-Pol polyprotein, while conventional translation of the same RNA produces Gag. The programmed recoding event allows a minority of ribosomes to shift the reading frame at a specific sequence and translate the *pol* coding region (Jacks et al. 1988b). A programmed  $-1$  ribosomal frameshift requires two *cis*-acting elements in the RNA: a heptameric slippery sequence, where the shift occurs, with a X XXY YYZ consensus (the zero frame is indicated by spaces), which is U UUU UUA in all HIV-1 groups, and a downstream stimulator that strongly increases the frameshift efficiency. The slippery sequence is such that when ribosomes bearing the peptidyl-tRNA and the aminoacyl-tRNA shift the reading frame by one base in the 5' direction, the anticodons of both tRNAs can re-pair with the codons of the mRNA after the shift (Jacks et Varmus 1985). The downstream stimulatory signal, which is a pseudoknot in most cases, can also

be a stem-loop (Brierley 1995; Farabaugh 2000; Brierley et Pennel 2001 and references therein). In HIV-1, this signal was previously assumed to be a simple eleven base-pair stem-loop (Jacks et al. 1988b; Le et al. 1991; Chamorro et al. 1992; Kang 1998), and it was suggested that the frameshifting efficiency was directly related to the stability of this stem-loop (Bidou et al. 1997). However, it was recently found by Dinman et al. (2002) and by our group (Dulude et al. 2002) that the frameshift region of HIV-1 is longer than commonly assumed. Using structure probing and mutagenesis studies, we showed that the frameshift stimulatory signal of HIV-1 is a larger helix containing two stems separated by a three-purine bulge. The upper stem-loop corresponds to the classic stem-loop, whereas the lower stem is formed by pairing the purine-rich spacer between the slippery site and the classic stem-loop to a pyrimidine-rich sequence downstream of the classic stem-loop (Dulude et al. 2002). However, Dinman et al. (2002) proposed an alternative structure for the frameshift stimulatory signal, which consists in a pseudoknot, where the pyrimidine-rich downstream segment forms a three base-pair stem with bases in the loop of the classic stem-loop. This short stem is stabilized by the formation of a triple helix between the classic stem and a portion of the pyrimidine-rich segment, but as discussed in Dulude et al. (2002), such triplex structure is sterically unfavorable.

The Gag-Pol to Gag ratio in HIV-1 is critical for the RNA dimerization, particle assembly, replication and viral infectivity (Park et Morrow 1991; Karacostas et al. 1993; Shehu-Xhilaga et al. 2001), which suggests that altering the frameshift efficiency could affect viral replication. So far, the frameshift efficiency has only been studied for subtype B of HIV-1 group M. However, the development of an antiviral strategy directed against frameshifting requires the characterization of the frameshift efficiency of the other subtypes. All the other subtypes of group M (A, C, D, F, G, H, J, K) possess the same slippery sequence as subtype B (U UUU UUA), but exhibit variations in the sequence of the stimulatory region (Los Alamos National Laboratory 2002, <http://hiv-web.lanl.gov>). In this study, we assessed the frameshift efficiency of all subtypes of HIV-1 group M. The sequence variations in the frameshift stimulatory signal of these subtypes compared to

subtype B affect the predicted stability of the stimulator (see below) and consequently could alter the frameshift efficiency. Using a luciferase reporter system, the frameshift efficiency of the different subtypes was assessed *in vitro* and in cultured cells. While the *in vitro* system is not sensitive enough to analyze the efficiency of the frameshift stimulatory signal of HIV-1 group M, we found that the frameshift efficiency in cultured cells did not significantly differ between the different subtypes, despite the predicted variations in the stability of the stimulator.

## RESULTS

### Frameshift efficiency of different mutants of subtype B of HIV-1 group M

We first examined the frameshift efficiency in 293T cultured cells and *in vitro*, in a rabbit reticulocyte lysate (RRL), of a construct containing the complete frameshift region of subtype B of HIV-1 group M, the so-called 3' extended frameshift region (Figure 1A). The frameshift efficiency of subtype B was 1.9% in 293T cells and 7.0% *in vitro* in a RRL (Table 1). These values are comparable to those previously obtained for the frameshift efficiency of subtype B, using an appropriate reporter gene, which were found to vary from 2.1 to 5.5% in cultured cells and 5.6 to 11.6% *in vitro* (Reil et Hauser 1990; Cassan et al. 1994; Kim Y. G. et al. 2001; Dulude et al. 2002). It can be noted that the values that we obtained are inferior to those obtained by Dulude et al. (2002) with a similar construct. This difference can be ascribed to a difference in the sequence upstream the slippery site, since it is known that this sequence affects the frameshift efficiency by influencing the rate with which the ribosomes encounter the slippery site (Kim Y. G. et al. 2001). The frameshift efficiency of a construct containing only the slippery sequence was 0.2% and 2.8% in 293T cells and *in vitro*, respectively, showing that the corresponding stimulation of the frameshift efficiency with the 3' extended region was about ten-fold and three-fold (data not shown).

Prior to the analysis of the frameshift efficiency in the different subtypes of HIV-1 group M, we examined two mutants of subtype B: subtype B short, where the frameshift stimulatory signal contains only the classic stem-loop, and subtype B bulge subs, where the purine bulge is substituted with pyrimidines (Figure 1B). The structure of the stimulator of each mutant and its thermodynamic stability, calculated using *mfold* (Mathews et al. 1999, <http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi>), are presented in Figure 2A. This stability was calculated either for the extended bulged stem-loop structure (upper + lower stems) found by Dulude et al. (2002) or for the classic structure of the stimulator (only the upper stem-loop). In cultured cells, the frameshift efficiency compared to subtype B decreased by about 60% and 40% for subtype B short and subtype B bulge subs, respectively (Figure



2B). Dulude et al. (2002) had also investigated the effect on the frameshift efficiency of mutations preventing the formation of the lower stem and substituting the purine bulge with pyrimidines. Although the experimental values we obtained for the frameshift efficiency of subtype B short and subtype B bulge subs are inferior to the values they obtained for the corresponding mutations, the decrease relative to the subtype B is the same, confirming that the lower stem and the bulge contribute to the frameshift efficiency. No decrease in the frameshift efficiency was seen *in vitro*, which suggests the frameshift efficiency under these conditions depends only upon the classic stem-loop and that the *in vitro* system is not sensitive enough to characterize the structure of the frameshift region of HIV-1, as previously discussed (Dulude et al. 2002).

The frameshift efficiency of a third mutant called subtype B\* was also investigated. The sequence of the frameshift stimulatory signal of this mutant originates from a natural isolate of subtype B and has a three-nucleotide deletion (Figure 1B). This deletion changes the structure of the stimulator, which can be predicted to be a variant of the two-stem structure, with a shorter upper stem separated from the lower stem by a symmetric internal loop. The lower stem has the same size as in subtype B but is slightly more stable due to the replacement of a G-U pair with a G-C pair. The predicted stability for this structure was lower than that of the stimulator of subtype B, whether one considers only the upper stem-loop, or the extended bulged stem-loop structure (Figure 2A). The frameshift efficiency of subtype B\* compared to subtype B was decreased by about 25% and 60% *in vitro* and in cultured cells, respectively (Figure 2B). The lower *in vitro* frameshift efficiency can be directly related to the lower stability of the upper stem of this construct. In cultured cells, the frameshift efficiency of subtype B\* is the same as that of subtype B short, showing that the presence of the lower stem counteracts the destabilization of the upper stem. The effect of the symmetrical internal loop between the two stem in subtype B\* will be analyzed below (see Table 1 for a summary of the results).

### **Structure and stability of the frameshift stimulatory signal of all the subtypes of HIV-1 group M**

Next, the frameshift efficiency of all the subtypes of group M was investigated. The sequences used are the consensus sequences of each subtype, except for subtype K, and were obtained from the Los Alamos HIV Sequence Database (Los Alamos National Laboratory 2002, <http://hiv-web.lanl.gov>). For subtype K, only two highly divergent sequences (named arbitrarily K(a) and K(b)) are available, making the identification of a consensus sequence impossible. Figure 3 shows that although the frameshift stimulatory signal sequence displays variations in the different subtypes, it can be folded into the extended bulged stem-loop structure previously demonstrated for subtype B. Indeed, the majority of the sequence variations allow the conservation of the base-pairing in the two stems. As to the bulge, it contains purines in every subtype (GGA or GAA), which fully agrees with the demonstrated requirement of purines in the bulge to obtain maximal frameshift efficiency in subtype B (see above).

The thermodynamic stability of the stimulator of every subtype was calculated for the classic structure (upper stem-loop only) and for the extended bulged stem-loop structure (upper + lower stems), as it has been done for subtype B and its mutants. The hierarchy, by order of decreasing stability, is: subtype K(b) > B > K(a)  $\cong$  C > D > H = J  $\cong$  G = A > F, when considering only the upper stem-loop (see also Chang et al. 1999). Subtypes A, F, G, H, J and K(a) have a loop of six nucleotides instead of four in subtypes B, C, D and K(b), which decreases their upper stem length by one base-pair. Subtype F is further destabilized by an U-C mismatch in the lower part of the upper stem. When considering the extended bulged stem-loop, the hierarchy of stability changes to K(b) > B > A > K(a) > J  $\cong$  H  $\cong$  G > D > C > F. The predicted stability of subtypes C and D is reduced due to an A-C mismatch in their lower stem. However, we cannot exclude the possibility that the A-C pair forms a non-Watson-Crick base-pair geometrically similar to a G-U pair (Nowakowski et Tinoco Jr 1999) and that the thermodynamic stability calculated with *mfold* does not correspond to

the actual stability of the structure. In subtype K(a), a U-U mismatch destabilizes the lower stem.

### **Frameshift efficiency of all the subtypes of HIV-1 group M**

The frameshift efficiency of all the subtypes of HIV-1 group M was measured with the luciferase reporter system and is presented in Figure 4 (summarized in Table 1). It can be seen that the *in vitro* frameshift efficiency varies between 4.5% (subtype F) and 8.4% (subtypes K(b)), with a mean value of 6.0%. As discussed above for subtype B and its mutants, the stimulation of the frameshift efficiency *in vitro* appears to depend only upon the upper stem-loop. Our results with the different subtypes of group M confirm this observation, since the frameshift efficiency varies in parallel to the predicted stability of the upper stem *in vitro*. For example, subtype F has the lowest frameshift efficiency *in vitro* and is the only subtype with a mismatch in its upper stem, drastically reducing its thermodynamic stability. In cultured cells, however, the frameshift efficiency of the different subtypes ranges from 1.2% (subtype K(a)) to 2.4% (subtype H), with a mean value of 1.9%, clearly showing that the complete extended bulged stem-loop and not just the upper stem contributes to the frameshift efficiency. This is best seen with subtypes F and K(a). The upper stem of subtype F has a low stability, but its lower stem is very stable and its frameshift efficiency relative to the others subtypes is higher in cultured cells than *in vitro*. In contrast, the upper stem of subtype K(a) has a high stability and its frameshift efficiency *in vitro* is proximal to the mean value, but a U-U mismatch in its lower stem, which compromises the formation of this stem, likely accounts for the low frameshift efficiency of this subtype in cultured cells, compared to the other subtypes.

It can be observed that the frameshift efficiency of the different subtypes varies within a narrow range in cultured cells, the maximal deviation from the mean value calculated from the experimental values of all the subtypes being about 35%. This does not reflect the range of variations of the thermodynamic stability of the frameshift stimulator (from  $-17.2$  kcal/mol to  $-26.2$  kcal/mol). However, as suggested by Kim et al. (2001), it is

not the stability of the isolated stimulatory signal alone that must be considered but its stability when it interacts with the ribosome, an interaction which remains to be defined. It is worth recalling that a stimulator makes the ribosome pause at a slippery sequence and that the duration of the pause likely correlates with the stability of the signal, but, on the other hand, pausing is not sufficient to promote frameshifting (Tu et al. 1992; Somogyi et al. 1993; Kontos et al. 2001). We found that the thermodynamic stability of subtype B\* is higher than that of subtype F, but its frameshift efficiency in cultured cells is two-fold less than subtype F. A direct explanation for this observation is that this low frameshift efficiency of subtype B\* results from the absence of the bulge, which is replaced by a symmetric internal loop. It is likely that this internal loop does not interact with the ribosome in a way promoting a productive interaction of the frameshift stimulatory signal.

## DISCUSSION

Our analysis presents for the first time an assessment of the frameshift efficiency of the different subtypes of HIV-1 group M. It shows that the frameshift efficiencies of the different subtypes of HIV-1 group M investigated fall within a narrow “window”, despite variations in the stimulatory signal sequence and in its predicted thermodynamic stability. This likely reflects the existence of a selective pressure to maintain the Gag-Pol to Gag ratio at a critical value for optimal virus assembly and infectivity. Indeed, it was observed that a three-fold increase of the Gag-Pol to Gag ratio, caused by cotransfecting HIV-1 proviral DNA with an HIV-1 Gag-Pol expression vector in 293T cells, produces a ten-fold decrease in virion infectivity (Shehu-Xhilaga et al. 2001). In this study, we show that the frameshift efficiencies of the different subtypes of group M are found within 35% of the mean frameshift efficiency, calculated from the experimental values obtained for all subtypes. Telenti et al. (2002) recently investigated how mutations destabilizing the classic stem-loop of natural variants and laboratory mutants of HIV-1 affect the frameshift efficiency and the capacity of the mutant virions to replicate in peripheral blood lymphocytes. Their results show that a decrease of the frameshift efficiency greater than 60% compared to subtype B severely impairs viral replication, while a decrease up to 40% has no effect. With the natural subtype B\* variant we investigated here, the frameshift efficiency decreased by about 60% compared to either subtype B or to the mean value obtained from all subtypes, but whether subtype B\* is infectious has yet to be determined. Our results coupled to those of Telenti et al. suggest that agents targeted against the frameshift event should affect the frameshift efficiency by more than 60% to efficiently impair viral replication.

We can also use the natural sequence variations of the different subtypes of group M to address the possible formation of the alternate frameshift stimulatory signal structure proposed by Dinman et al. (2002). This structure is derived from the analysis of sequences from GenBank, which include mostly subtype B sequences. In this structure, which is a

pseudoknot, the pyrimidine-rich segment downstream of the classic stem-loop forms an additional three base-pair stem with bases in the loop of the classic stem-loop. This stem is further stabilized by a triple helix, where a portion of the pyrimidine-rich segment interacts with the classic stem. If the frameshift region of all the subtypes investigated here was folded according to the model of Dinman et al. (2002), subtypes H and J would have mismatches in two out of the three base-pairs of the additional stem, while subtypes A and K(a) would have one mismatch in this stem. However, we found that the frameshift efficiency of these subtypes was not significantly lower than the mean value calculated for all the subtypes, except for subtype K(a). In this latter case, as discussed above, the presence of a mismatch in the lower stem of the extended bulged stem-loop structure can readily account for the decrease of the frameshift efficiency of this subtype. These observations raise an additional criticism against the model of Dinman et al. (2002), in addition to the fact that the pyrimidine-rich strand of the triplex is in an antiparallel orientation relative to the purine-rich strand of the duplex with which it interacts, an orientation shown to be sterically unfavorable (Han et Dervan 1993; Frank-Kamenetskii et Mirkin 1995; Lim et Barton 1998).

Agents that decrease the frameshift efficiency in HIV-1 have already been developed (Hung et al. 1998), and, although the toxicity of these agents precludes their use for clinical treatments, they illustrate the potential of an approach targeting the frameshift event. The absence of treatments able to eradicate HIV-1 and the emergence of viruses resistant to the protease and reverse transcriptase inhibitors presently in use (Condra 1998; O'Brien 2000) make it imperative to develop alternative strategies against novel viral targets (Dinman et al. 1998; Gallego et al. 2001). Our results identify the frameshift event as such a potential target since the frameshift efficiency is very similar in the different subtypes of HIV-1 group M. Any agent perturbing the frameshift event, by bringing the Gag-Pol to Gag ratio outside a narrow "window", could impair the viral propagation in a broad population of HIV-1 subtypes.

## ACKNOWLEDGMENTS

This study was supported by a grant from the Canadian Institute for Health Research to L. B.-G. and G.L. We are most grateful to Pascal Chartrand and Sergey Steinberg for stimulating discussions and for critical reading of this manuscript.

## MATERIALS AND METHODS

### Construction of plasmids

A plasmid construct was derived from pcDNA3.1/Hygro(+) (Invitrogen), in which a reporter gene coding for the firefly luciferase was inserted under the control of the T7 and CMV promoter, allowing the assessment of its expression *in vitro* and in cultured cells, respectively. The *luc* sequence originated from pGEM<sup>®</sup>-*luc* (Promega) and was bracketed between a *Bam*HI and a *Xho*I site. The AUG initiation codon of *luc* was deleted (so as to avoid aberrant initiations) and an oligonucleotide fragment containing an AUG codon followed by a short linker encompassing an *Eco*47III site and the frameshift region of subtype B of the HIV-1 group M (GenBank K02007) was inserted upstream of the *luc* sequence, between the *Kpn*I site of the vector and the *Bam*HI site preceding the *luc* sequence. This generated the pLUC/HIV/B (-1) construct (Figure 1A). The frameshift region in this construct, coined the 3'extended frameshift region, encompasses the slippery sequence, the spacer, the classic stem-loop plus ten nucleotides downstream of this stem, i.e. the stimulator defined in a previous study (Figure 2A) (Dulude et al. 2002). The slippery site was preceded by an *Eco*47III site, so that constructs corresponding to mutants or other subtypes of HIV-1 could be readily generated by replacing the *Eco*47III-*Bam*HI fragment of subtype B with an appropriate oligonucleotide cassette, corresponding to either subtype A (GenBank AF069673), C (GenBank AF110962), D (GenBank U88822), F (GenBank AJ249238), G (GenBank AF061640), H (GenBank AF005496), J (GenBank AF082394), K(a) (GenBank AJ249235) and K(b) (GenBank AJ249239) (see Figure 1B). Different mutants of subtype B, i.e. subtype B short, subtype B bulge subs and subtype B\*,

a natural variant of subtype B (GenBank M17449), were also derived from the pLUC/HIV/B (-1) construct (Figure 1B). In all (-1) constructs, a -1 ribosomal frameshift is required for expression of the *luc* reporter gene, otherwise translation terminates at a stop codon overlapping codon six of the *luc* sequence. For each (-1) construct, a pLUC/HIV (0) in-frame control was made, in which an additional adenine was introduced before the *luc* coding sequence (see Figure 1A) so that luciferase is synthesized by ribosomes that do not make a frameshift. All constructs were verified by sequencing the entire insert.

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

For assays in cultured cells, transient transfections were made in 293T cells, using the standard calcium phosphate precipitation method (Jordan et al. 1996). To this end, precipitates were formed with 3  $\mu$ g of pLUC/HIV and 1.25  $\mu$ g of pcDNA3.1/Hygro(+)/lacZ (Invitrogen), which codes for  $\beta$ -galactosidase. Precipitates were overlaid over 40% confluent cells grown into six-well plates and maintained in Dulbecco's modification of Eagle's medium with 10% (v/v) fetal bovine serum (BioMedia). Cells were harvested with 600  $\mu$ l of the Cell Culture Lysis Reagent (Promega) 48 hours post-transfection, and the luciferase activity was measured in cell lysates. Luminescence produced by adding 2  $\mu$ l of the lysate to the Luciferase Assay Reagent (Promega) was measured in relative light units in a Berthold Lumat LB 9507 luminometer. The  $\beta$ -galactosidase activity of the cell lysates was determined using chlorophenolred- $\beta$ -D-galactopyranoside substrate (Calbiochem) (Eustice et al. 1991) and was used to normalize the transfection efficiency. The frameshift efficiency of the different pLUC/HIV constructs was calculated by dividing the level of luciferase activity produced with the (-1) construct by the sum of the luciferase activity of the (0) and (-1) constructs.



***In vitro* transcription and translation**

For *in vitro* studies, plasmids were linearized at the *StuI* site located 3' of the *luc* coding sequence. HIV-LUC mRNA was generated by run-off transcription with T7 RNA polymerase, and 0.2  $\mu\text{g}$  of this mRNA was translated at 30°C for 15 min in 25  $\mu\text{l}$  of RRL (Promega). The reaction was stopped by addition of EDTA to a final concentration of 6 mM. Luciferase activity was monitored with 2.5  $\mu\text{l}$  of the translation mixture, as described in the “transient transfections and luciferase assays” section. Frameshift efficiencies were calculated as described above.

## REFERENCES

- 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.
- Brierley, I. 1995. Ribosomal frameshifting viral RNAs. *J. Gen. Virol.* **76**: 1885-1892.
- Brierley, I. and Pennel, S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. *The Ribosome*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press. pp 233-248.
- 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.
- 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.* **89**: 713-717.
- Chang, S.Y., Sutthent, R., Auewarakul, P., Apichartpiyakul, C., Essex, M. and Lee, T.H. 1999. Differential stability of the mRNA secondary structures in the frameshift site of various HIV type 1 viruses. *AIDS Res. Hum. Retroviruses* **15**: 1591-1596.
- Condra, J.H. 1998. Resistance to HIV protease inhibitors. *Haemophilia* **4**: 610-615.
- 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.* **99**: 5331-5336.
- Dinman, J.D., Ruiz-Echevarria, M.J. and Peltz, S.W. 1998. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.* **16**: 190-196.
- 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**: 5094-5102.

- 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**: 739-740, 742-733.
- Farabaugh, P.J. 2000. Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.* **64**: 131-170.
- Frank-Kamenetskii, M.D. and Mirkin, S.M. 1995. Triplex DNA structures. *Annu. Rev. Biochem.* **64**: 65-95.
- Gallego, O., de Mendoza, C., Perez-Elias, M.J., Guardiola, J.M., Pedreira, J., Dalmau, D., Gonzalez, J., Moreno, A., Arribas, J.R., Rubio, A., Garcia-Arata, I., Leal, M., Domingo, P. and Soriano, V. 2001. Drug resistance in patients experiencing early virological failure under a triple combination including indinavir. *AIDS* **15**: 1701-1706.
- 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.* **90**: 3806-3810.
- Hung, M., Patel, P., Davis, S. and Green, S.R. 1998. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J. Virol.* **72**: 4819-4824.
- 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.
- Jacks, T. and Varmus, H.E. 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**: 1237-1242.
- 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.
- 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.

- 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.
- Kim, Y.G., Maas, 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.
- Kontos, H., Naphine, 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.
- Le, S.Y., Shapiro, B.A., Chen, J.H., Nussinov, R. and Maizel, J.V. 1991. RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet. Anal. Tech. Appl.* **8**: 191-205.
- 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.
- Los Alamos National Laboratory. 2002. A compilation and analysis of nucleic acid and amino acid sequences.: Los Alamos National Laboratory.
- 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.
- McCutchan, F.E. 2000. Understanding the genetic diversity of HIV-1. *AIDS* **14 Suppl 3**: S31-44.
- Moore, J.P., Parren, P.W. and Burton, D.R. 2001. Genetic subtypes, humoral immunity, and human immunodeficiency virus type 1 vaccine development. *J. Virol.* **75**: 5721-5729.
- Nowakowski, J. and Tinoco Jr, I. 1999. RNA structure in solution. In: Press OU, ed. *Oxford Handbook of Nucleic Acid Structure*. New-York: Oxford University Press. pp 567-602.

- O'Brien, W.A. 2000. Resistance against reverse transcriptase inhibitors. *Clin. Infect. Dis.* **30** **Suppl 2**: S185-192.
- 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.
- Reil, H. and Hauser, H. 1990. Test system for determination of HIV-1 frameshifting efficiency in animal cells. *Biochim. Biophys. Acta* **1050**: 288-292.
- Roberts, J.D., Bebenek, K. and Kunkel, T.A. 1988. The accuracy of reverse transcriptase from HIV-1. *Science* **242**: 1171-1173.
- Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S. and Korber, B. 2000. HIV-1 nomenclature proposal. *Science* **288**: 55-56.
- 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.
- 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.
- 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**: 7868-7873.
- Tu, C., Tzeng, T.H. and Bruenn, J.A. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc. Natl. Acad. Sci.* **89**: 8636-8640.

Tableau 2-1 : Frameshift efficiency of different mutants of subtype B and of all subtypes of HIV-1 group M.

Construct	Description	Frameshift efficiency (%)	
		<i>In vitro</i>	In cultured cells
<b>Frameshift efficiency of mutants of subtype B</b>			
pLUC/HIV/B	Frameshift region of subtype B	7.0	1.9
pLUC/HIV/B short	Classic HIV-1 frameshift region	7.3	0.9
pLUC/HIV/B bulge subs	Substitution in the bulge	8.5	1.4
pLUC/HIV/B*	Natural variant of subtype B	5.4	0.8
<b>Frameshift efficiency of subtypes A, B, C, D, F, G, H, J, K(a) and K(b)</b>			
pLUC/HIV/A	Frameshift region of subtype A	5.9	1.8
pLUC/HIV/B	Frameshift region of subtype B	7.0	1.9
pLUC/HIV/C	Frameshift region of subtype C	5.3	2.0
pLUC/HIV/D	Frameshift region of subtype D	6.4	2.0
pLUC/HIV/F	Frameshift region of subtype F	4.6	1.8
pLUC/HIV/G	Frameshift region of subtype G	5.4	1.6
pLUC/HIV/H	Frameshift region of subtype H	5.6	2.4
pLUC/HIV/J	Frameshift region of subtype J	5.9	1.8
pLUC/HIV/K(a)	Frameshift region of subtype K(a)	6.4	1.2
pLUC/HIV/K(b)	Frameshift region of subtype K(b)	8.4	2.0

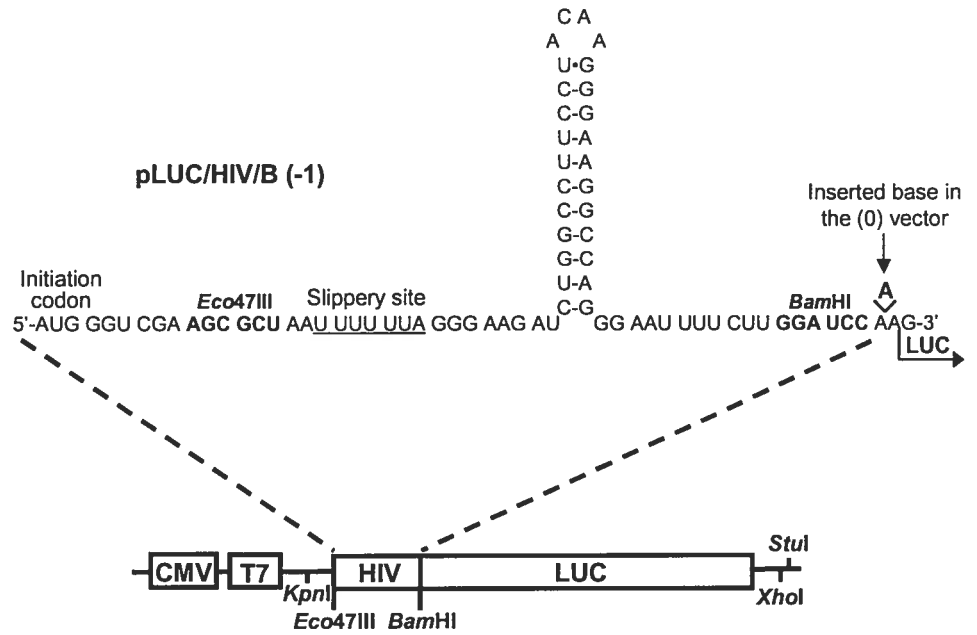
The pLUC/HIV constructs contain the HIV-1 *gag/pol* frameshift region. The frameshift region of each subtype is identified by the corresponding letter (A, B, C, D, F, G, H, J, K(a) and K(b)) and corresponds to the 3' extended HIV-1 frameshift region. Mutants of the subtype B frameshift region are further identified by either a short description (B short, B bulge subs) or by an asterisk for the natural variant (see details in Figure 1B). Results are the means of five independent experiments. The standard error on the mean was inferior or equal to 10%.

Figure 2-1 : Vectors expressing luciferase used to study the programmed -1 ribosomal frameshift of HIV-1 group M *in vitro* and in cultured cells.

- A) The frameshift region of subtype B was inserted upstream of the coding sequence of the luciferase reporter gene, which is under control of a CMV and a T7 promoter, generating pLUC/HIV/B (-1). The classic structure of the frameshift stimulatory signal is represented here. For the study of all subtypes of group M and mutants of subtype B, the corresponding vectors were constructed by exchanging the *Eco47III-BamHI* fragment of pLUC/HIV/B (-1) with an appropriate oligonucleotide cassette. For the (-1) constructs, the luciferase sequence is in the -1 reading frame relative to the AUG initiation codon, so that a -1 frameshift is required to produce luciferase. An adenine nucleotide was added between the frameshift region and luciferase sequence for the (0) constructs, so that luciferase is expressed by ribosomes that do not shift the reading frame.
- B) Sequences of the frameshift region of all the constructs used in this study. Nucleotides modified or deleted relative to subtype B are underlined or represented by dashed lines, respectively.



A



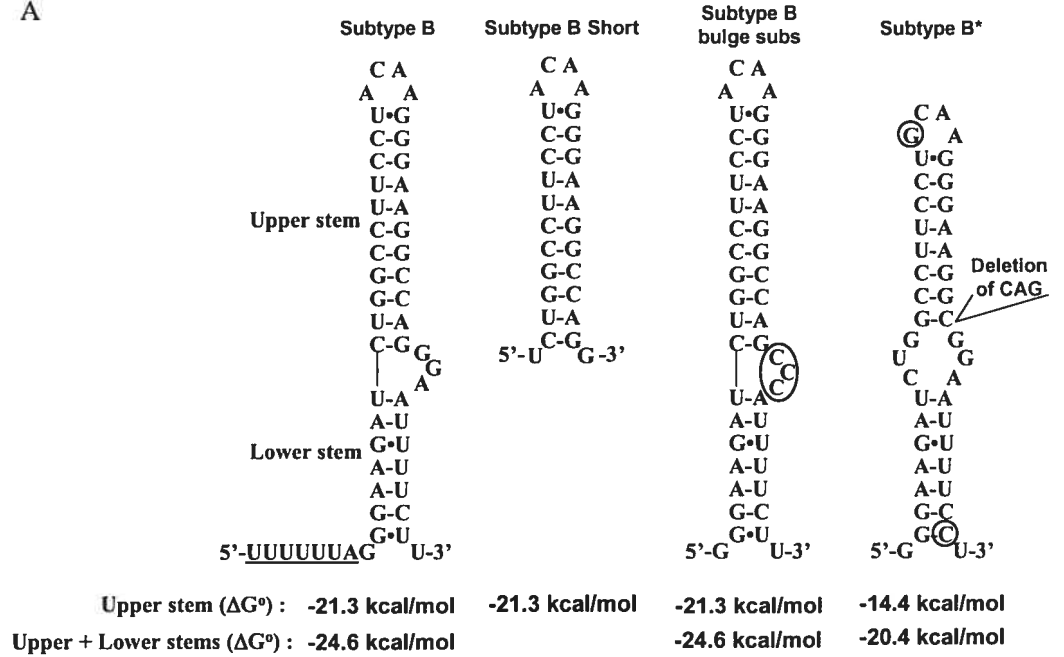
B

Subtype B	<b>GC</b> UAAUUUUUUAGGGAAGAUCUGGCCUCCUACAAGGGAAGGCCAGGGAAUUUUCUUG
Subtype B short	<b>GC</b> UAAUUUUUUAGGGAAGAUCUGGCCUCCUACAAGGGAAGGCCAGGG-----G
Subtype B bulge subs	<b>GC</b> UAAUUUUUUAGGGAAGAUCUGGCCUCCUACAAGGGAAGGCCAGCC <u>CA</u> UUUUCUUG
Subtype B*	<b>GC</b> UAAUUUUUUAGGGAAGAUCUGGCCUCCUG <u>CA</u> AGGGAAGGC---GGAAUUUUC <u>CU</u> G
Subtype A	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCCAGCAAGGGAGGCCAGGGAAUUUUC <u>CU</u> G
Subtype C	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCC <u>CA</u> AGGGAGGCCAGGGAAUUUCC <u>U</u> G
Subtype D	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCC <u>CA</u> AGGGAGGCC <u>GG</u> GAA <u>CU</u> UUCUUG
Subtype F	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCC <u>CA</u> AGGGAGGCC <u>CG</u> GAAUUUUCUUG
Subtype G	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCC <u>CA</u> AGGGAGGCCAGGGAAUUUUCUUG
Subtype H	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUGGCCUCCAGCAAGGGAGGCCAGGAAUUUUC <u>CG</u>
Subtype J	<b>GC</b> UAAUUUUUUAGGGAAGAUUUGGCCUCCAGCAAGGGAGGCCAGGGAAUUUUC <u>CG</u>
Subtype K(a)	<b>GC</b> UAAUUUUUUAGGGAAGUUCUGGCCU <u>CU</u> CAACAAGAGAGGCCAGGAAUUUUCUUG
Subtype K(b)	<b>GC</b> UAAUUUUUUAGGGAA <u>AA</u> UUCUGGCCUCC <u>CA</u> AGGGAGGCCAGGGAAUUUUCUUG

Figure 2-2 : Relative frameshift efficiency of mutants of subtype B of HIV-1 group M.

- A) Secondary structure of the frameshift stimulatory signal for subtype B, subtype B short, subtype B bulge subs and subtype B\*, a natural variant that has a deletion of three nucleotides in the frameshift stimulatory signal. The stimulatory signal is represented with the extended bulged stem-loop structure found for HIV-1 subtype B (Dulude et al. 2002). Circles indicate nucleotides that differ from subtype B. The thermodynamic stability,  $\Delta G^\circ$  (kcal/mol), of the frameshift stimulatory signal of the different mutants of subtype B is indicated. The predicted  $\Delta G^\circ$  were calculated with the *mfold* program of M. Zuker (Mathews et al. 1999) for the classic stem-loop structure (upper stem) and the extended bulged stem-loop structure (upper stem + lower stems). The complete frameshift region with the slippery sequence underlined is represented for subtype B, while only the frameshift stimulatory signal is shown for the mutants.
- B) Relative frameshift efficiency of mutants of subtype B. *In vitro* studies in RRL and assays in cultured cells (293T) were performed as described in the text. For assays in cultured cells, the experimental values were normalized for variations in transfection efficiency, by co-transfecting the various constructs used for measuring the frameshift efficiency with pcDNA3.1/Hygro(+)/LacZ and monitoring the  $\beta$ -galactosidase activity. The frameshift efficiency of subtype B is arbitrarily set at 100%. The values of the frameshift efficiency are the means of at least five independent experiments, with bars representing the standard error on the means.

A



B

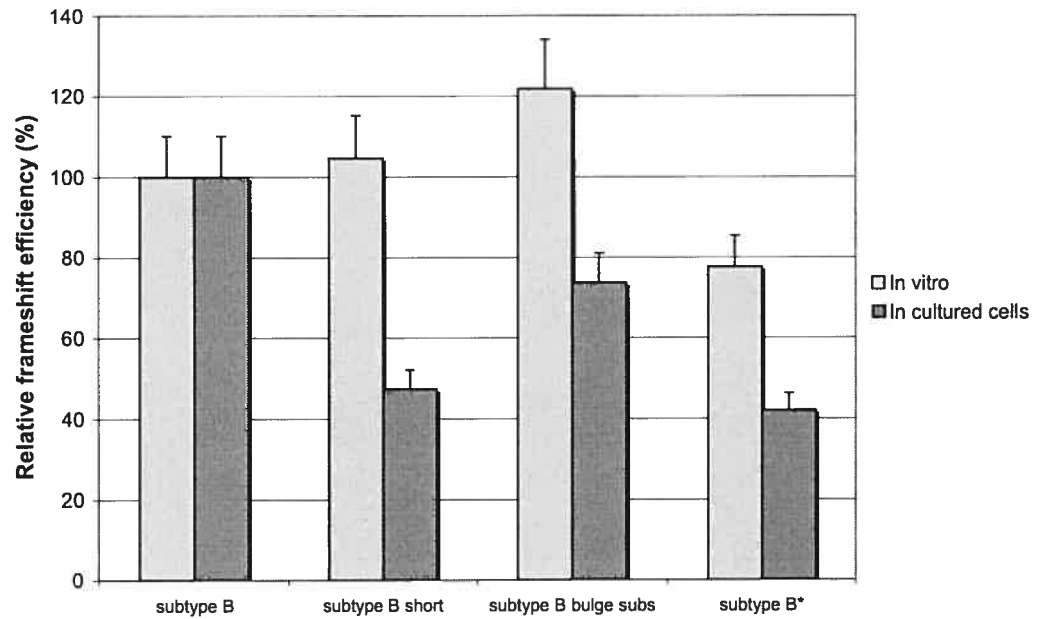


Figure 2-3 : Secondary structure of the frameshift stimulatory signal of all the subtypes of HIV-1 group M.

The frameshift stimulatory region is folded according to the extended bulged stem-loop structure found for HIV-1 subtype B stimulator (Dulude et al. 2002). Circles indicate nucleotides that differ from subtype B. The arrows point to non-canonical base-pairs and mismatches. The thermodynamic stability,  $\Delta G^\circ$  (kcal/mol), of the frameshift stimulatory signal of the different subtypes is indicated. The predicted  $\Delta G^\circ$  were calculated as described in the legend to Figure 2.

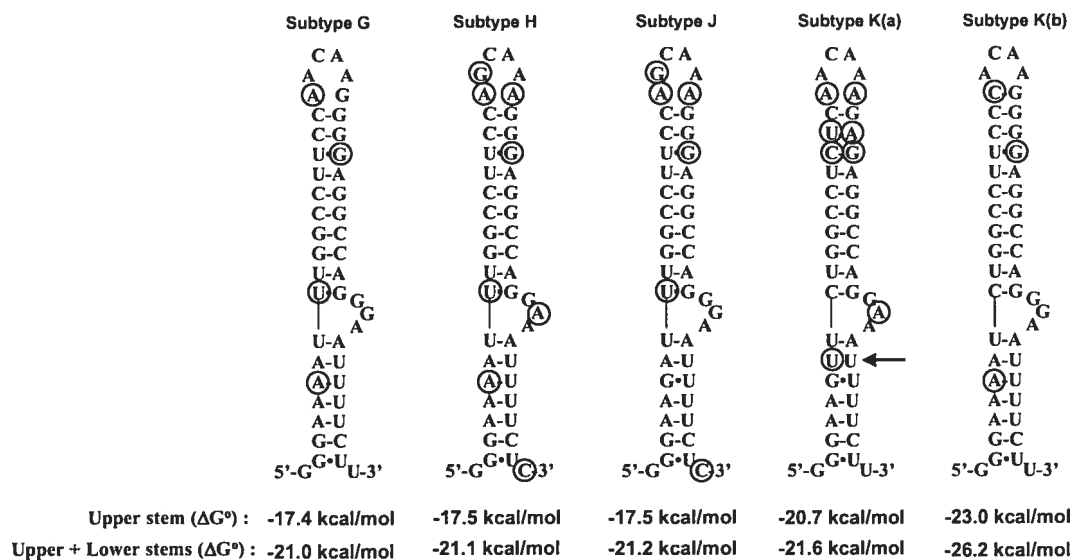
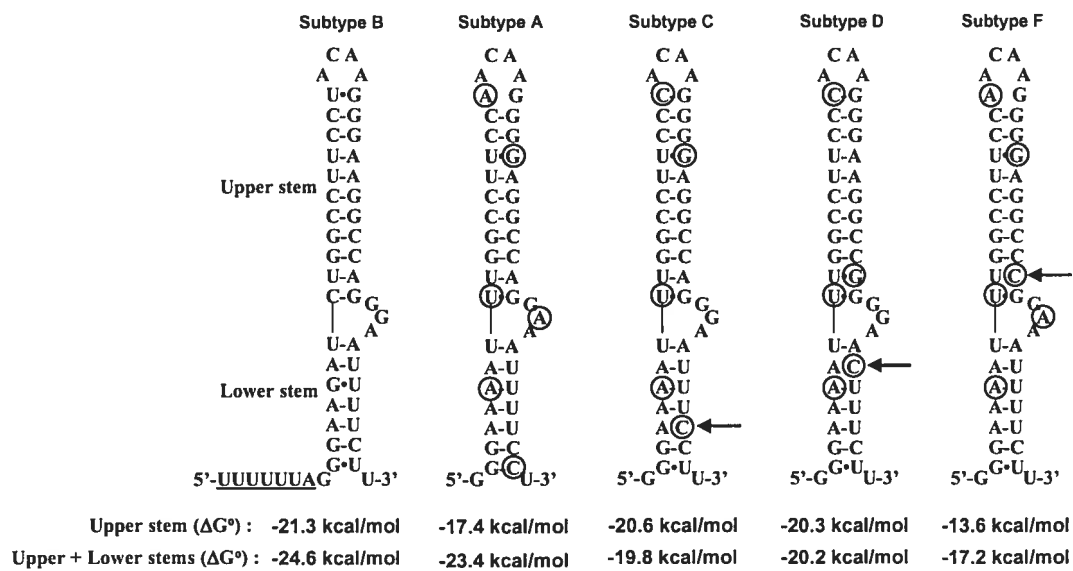
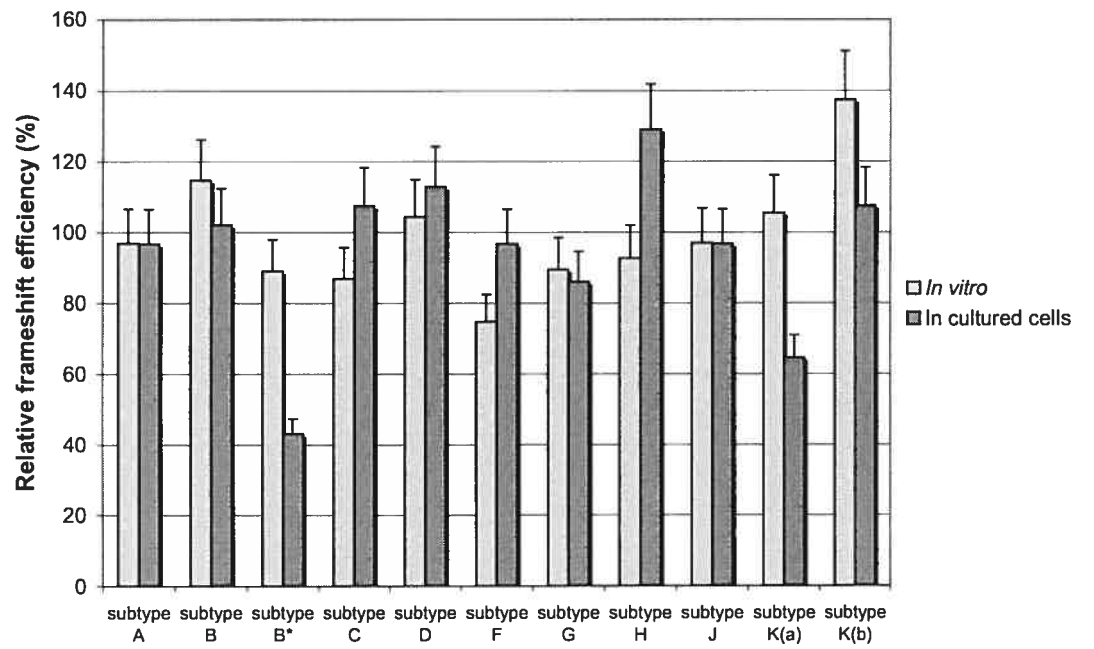


Figure 2-4 : Relative frameshift efficiency of all subtypes of HIV-1 group M.

*In vitro* studies in RRL and assays in cultured cells (293T) were performed as indicated in Figure 2. The mean value of the frameshift efficiency, calculated from the values obtained for all the subtypes (1.9% in cultured cells and 6.0% *in vitro*), is arbitrarily set at 100%. The frameshift efficiency of subtype B\* has been added to the figure to facilitate comparison with the different subtypes. The values of the frameshift efficiency are the means of at least five independent experiments, with bars representing the standard error on the means.



## Chapitre III

### **The *frameshift* stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot.**

Martin Baril, Dominic Dulude, Sergey V. Steinberg and Léa Brakier-Gingras

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

J. Mol. Biol., 331: 571-583, 2003.



## SUMMARY

Human immunodeficiency virus type 1 (HIV-1) requires a programmed  $-1$  ribosomal frameshift to produce Gag-Pol, the precursor of its enzymatic activities. This frameshift occurs at a slippery sequence on the viral messenger RNA and is stimulated by a specific structure, downstream of the shift site. While in group M, the most abundant HIV-1 group, the frameshift stimulatory signal is an extended bulged stem-loop, we show here, using a combination of mutagenesis and probing studies, that it is a pseudoknot in group O. The mutagenesis and probing studies coupled to an *in silico* analysis show that group O pseudoknot is a hairpin-type pseudoknot with two coaxially stacked stems of eight bp (stem 1 and stem 2), connected by single-stranded loops of two nt (loop 1) and 20 nt (loop 2). Mutations impairing formation of stem 1 or stem 2 of the pseudoknot reduce frameshift efficiency, whereas compensatory changes that allow re-formation of these stems restore the frameshift efficiency near wild-type level. The difference between the frameshift stimulatory signal of group O and group M supports the hypothesis that these groups originate from a different monkey to human transmission.

**Keywords:** Human immunodeficiency virus type 1; HIV-1 group O; ribosomal frameshifting; RNA structure; RNA pseudoknot.

## INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) is classified into three groups: M (main), that accounts for more than 99% of the worldwide pandemic, N (new or non-M/non-O), that is not clearly defined yet, and O (outlier), that has a genomic sequence differing by 60% from group M. Each group is furthermore divided into subtypes based on differences in their *env*, *gag* or *pol* gene sequences.(Yamaguchi et al. 2002) Although much less abundant than group M, group O viruses are spread widely in Cameroon and Equatorial Guinea.(Peeters et al. 1997) Their resistance to non-nucleoside reverse transcriptase inhibitors active against group M(Descamps et al. 1997) as well as the rapid rate at which they develop resistance to protease inhibitors indicate that they could become an important threat.(de Baar et al. 2000; Yang et al. 2000; Luk et al. 2001)

HIV-1 uses a programmed  $-1$  ribosomal frameshift during translation of its full-length messenger RNA (mRNA) to produce the Gag-Pol polyprotein, the precursor of the viral enzymes. This programmed event allows a minority of ribosomes to shift the reading frame at a specific sequence before they encounter the stop codon of *gag*, and to extend the translation to the *pol* gene, while conventional translation of the same mRNA by the majority of ribosomes produces the Gag polyprotein, the precursor of the viral structural proteins.(Jacks et al. 1988b; Brierley et Pennel 2001) The efficiency of the frameshift event controls the Gag-Pol to Gag ratio, which is critical for RNA dimerization, particle assembly, replication and viral infectivity.(Park et Morrow 1991; Karacostas et al. 1993; Shehu-Xhilaga et al. 2001) Therefore, increasing or decreasing the frameshifting efficiency can interfere with the formation of infectious viral particles.

The programmed  $-1$  ribosomal frameshift requires two *cis*-acting elements in the viral mRNA: a heptameric slippery sequence where the frameshift takes place (U UUU UUA in HIV-1) and a downstream stimulatory signal. In group M, the stimulatory signal was previously assumed to be a simple 11-bp stem capped by a four-nt loop,(Jacks et al. 1988b; Kang 1998) although an alternative structure was proposed but without any

experimental support.(Du et al. 1996) It was recently shown independently by our group(Dulude et al. 2002) and by Dinman *et al.*(Dinman et al. 2002) that the frameshift stimulatory signal is longer than believed and that the sequence downstream of the classic stem-loop contributes to the frameshifting event. Using a combination of site-directed mutagenesis and probing experiments, we demonstrated that this signal is an extended bulged stem-loop for subtype B of group M,(Dulude et al. 2002) the subtype that prevails in North America and Europe, and also that the same extended bulged stem-loop structure is found in all subtypes of group M (our unpublished results). In this signal, the upper portion of the stem corresponds to the classic stem-loop previously considered as the frameshift stimulatory signal, while the lower portion is formed by a base-pairing between the spacer separating the slippery sequence from the classic stem-loop and a complementary pyrimidine-rich sequence downstream of this stem-loop (Figure 1). This structure thus differs from most frameshift stimulatory signals, which were shown to be pseudoknots.(Giedroc et al. 2000; Brierley et Pennel 2001) Dinman *et al.*(Dinman et al. 2002) proposed an alternative structure for the frameshift signal of HIV-1 group M. 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 triplex with 4 bp on top of this stem. However, as discussed by Dulude *et al.*,(Dulude et al. 2002) a careful analysis of this model shows that it is highly improbable, a major criticism being that the pyrimidine-rich 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 sterically not favored.

In this study, the structure of the frameshift stimulatory signal of HIV-1 group O was investigated with a combination of mutagenesis and probing studies. The HIV-1 group O frameshift region or its mutants was inserted at the beginning of the coding sequence of a luciferase reporter gene, such that luciferase expression depends on a -1 frameshift. The luciferase activity was then assessed in 293T cultured cells and *in vitro*, in a rabbit reticulocyte lysate (RRL). The structure of the frameshift stimulatory signal of HIV-1

group O was also analyzed by probing the RNA fragment encompassing this region with RNase V<sub>1</sub>. The proposed structure was then assessed by computer modeling. Group O is subdivided into two subtypes: MVP5180 and ANT70, for which there are seven and one complete *pol* sequences, respectively. Our results show that the frameshift stimulatory signal of HIV-1 group O MVP5180 consists in a pseudoknot, and also, that this pseudoknot promotes a higher frameshift efficiency (and thus a higher Gag-Pol to Gag ratio) than the group M frameshift stimulatory signal. However, sequence differences in subtype ANT70 prevent the frameshift stimulatory signal from adopting the same pseudoknot structure as in subtype MVP5180.

## RESULTS

### Sequence analysis of HIV-1 group O frameshift region

While the totality of studies on the HIV-1 frameshift have been carried out using group M, the frameshift region of group O has not been examined so far. We decided to investigate whether the frameshift region of HIV-1 group O encompasses the same slippery site and stimulatory signal as group M. As mentioned above, there are seven complete *pol* sequences for subtype MVP5180 and one complete *pol* sequence for subtype ANT70 (Los Alamos National Laboratory HIV Sequence Database).(2002) We started the analysis of the frameshift region with MVP5180, the most represented subtype of group O. Examination of MVP5180 sequences suggests that the slippery sequence in this group O subtype is the same as in group M, while the frameshift stimulatory signal could be predicted to be an eight-bp stem (the thermodynamic stability calculated with *mfold*(Mathews et al. 1999) is – 14.9 kcal/mol), capped by a 10-nt loop and separated from the slippery sequence by an eight-nt spacer. This stem-loop is different from that initially proposed as a frameshift stimulatory signal for group M. Moreover, for group O, because of the absence of base complementarity between the spacer region preceding the stem-loop and the sequence downstream of this stem-loop, the extended bulged stem-loop structure demonstrated for group M(Dulude et al. 2002) cannot form (Figure 1). However, examination of the frameshift region sequence in MVP5180 suggests that its frameshift stimulatory signal could form a pseudoknot structure by base-pairing between eight nt of the loop capping the eight-bp stem and a complementary downstream sequence (Figure 1).

### Characterization of the frameshift region of subtype MVP5180 of HIV-1 group O

To investigate whether the frameshift stimulatory signal of HIV-1 group O could adopt the proposed pseudoknot structure, we made a reporter construct for the frameshift region of subtype MVP5180 of HIV-1 group O, using a modified pcDNA3.1/Hygro(+) plasmid, in which the firefly luciferase (*luc*) reporter gene had been introduced under

control of a CMV and a T7 promoter. The frameshift region of subtype MVP5180 or its derivatives was inserted between the first and second codon of *luc*, generating a series of pHIV/O-LUC (-1) constructs (Figure 2). These insertions were such that a -1 frameshift was required to express the LUC reporter protein. An in-frame (0) control construct, in which ribosomes synthesize LUC through conventional translation, was also made for each (-1) frameshift construct examined, by adding an adenine immediately after the slippery sequence. The frameshift efficiency for each construct was measured by dividing the luciferase activity of the (-1) construct by the sum of that of the (0) and (-1) constructs, assuming that the same level of frameshift occurs in both (0) and (-1) constructs. The frameshift efficiency of these constructs was assessed in cultured cells and *in vitro*. For assays in cultured cells, the different pHIV/O-LUC (0) and (-1) constructs were transfected into 293T cells, and luciferase activity was measured in cell extracts 48 hours later. For *in vitro* assays, luciferase activity was measured after translation in RRL of the HIV/O-LUC mRNAs generated by transcription with T7 RNA polymerase of the *StuI*-linearized plasmids.

Figure 3 shows the characteristics of the structure of the different constructs and their frameshift efficiency, and the results are summarized in Table 1. Consistent with what was systematically observed in previous reports, the frameshift efficiency was higher *in vitro* than in cultured cells (see ref. (Dulude et al. 2002) and references therein), and was less sensitive *in vitro* to the presence of the stimulatory signal, compared to the situation in cultured cells, (Parkin et al. 1992; Reil et al. 1993) which was proposed to result from differences between the rate of translation *in vitro* and in cultured cells. As a reference construct for the study of the frameshift region of HIV-1 group O, we used pHIV/O-87-LUC. This construct bears the frameshift region of subtype MVP5180 of group O, encompassing the putative slippery site, the predicted stem-loop and the downstream region, with the sequence proposed to interact with the loop, thus forming a pseudoknot. The frameshift efficiency of pHIV/O-87-LUC was 4.7% in cultured cells and 10.2% *in vitro*. Using construct pHIV/O-k/o-LUC, in which the predicted slippery sequence

UUUUUUA of pHIV/O-87-LUC was replaced by CUUCCUC, we first verified that the slippery sequence is the same in group O as in group M. The frameshift in pHIV/O-k/o-LUC was abolished or dramatically decreased in cultured cells and *in vitro*, respectively, confirming that group O viruses use the same slippery sequence as group M viruses. The influence on frameshift efficiency of the region downstream of the predicted eight-bp stem-loop was assessed with construct pHIV/O-60-LUC, bearing the slippery site and the eight-bp stem-loop, but lacking this downstream region. The frameshift efficiency of pHIV/O-60-LUC decreased two to three-fold in cultured cells and *in vitro*, compared to pHIV/O-87-LUC. This indicates the presence of an element in the region downstream of the putative eight-bp stem-loop of group O that stimulates the frameshift efficiency, although it does not prove the existence of a pseudoknot structure.

Next, the importance of the proposed eight-bp stem-loop for frameshift was demonstrated using three constructs: pHIV/O-1.2-LUC, pHIV/O-1.12-LUC and pHIV/O-DSL-LUC, all derivatives of pHIV-87-LUC. In pHIV/O-1.2-LUC, the eight nt of the 3'-strand of the stem-loop were replaced with the eight nt of the 5'-strand, thus impairing the formation of the stem. In pHIV-1.12-LUC, the 3' and the 5'-strands of the stem were exchanged, thus restoring the capacity to form the stem. Finally, in pHIV/O-DSL-LUC, the stem-loop was deleted. For the two constructs where the stem-loop was either destabilized or deleted, the frameshift efficiency was reduced by about ten-fold in cultured cells and three to five-fold *in vitro*, compared to pHIV/O-87-LUC. However, in pHIV/O-1.12-LUC, where the re-formation of the stem-loop was possible, the frameshift efficiency was restored to the wild-type level *in vitro* and in cultured cells. These results confirm the existence of the predicted eight-bp stem-loop structure and its involvement in the frameshift process. Since the region downstream the stem-loop increases frameshifting in the presence of this stem-loop, as shown in Figure 3, these results also support our suggestion that the frameshift stimulatory signal in MVP5180 is more complex than a simple stem-loop and could correspond to a pseudoknot structure.

### **Characterization of the pseudoknot in the frameshift region of subtype MVP5180 of HIV-1 group O**

Figure 4a shows a standard representation of the hypothetical pseudoknot that we proposed to act as a frameshift stimulatory signal in HIV-1 group O. To demonstrate the existence of this pseudoknot, we made a series of mutations in the loop capping the eight-bp stem, named stem 1, as well as in the downstream complementary sequence predicted to interact with this capping loop and to form stem 2 of the pseudoknot. Individually, these mutations impair the formation of stem 2, but when combined, should allow it to re-form (Figure 4). The frameshift efficiencies of the constructs containing these mutations were assessed *in vitro* and in cultured cells and are presented in Figure 4b (see also Table 1). The mutations consisted in replacing eight nt of the loop capping stem 1, with the downstream set of complementary nt (pHIV/O-2.1-LUC). The frameshift efficiency of this construct was reduced by about two to three-fold *in vitro* and in cultured cells, compared to the wild-type construct pHIV/O-87-LUC. This level of frameshift efficiency is similar to that obtained with the eight-bp stem-loop alone (pHIV/O-60-LUC). Similarly, the frameshift efficiency of pHIV/O-2.2-LUC, in which the eight bases downstream of stem 1 were replaced with the eight complementary bases of the loop, was decreased by about two to three-fold *in vitro* and in cultured cells. However, combining both mutations (pHIV/O-2.12-LUC), so as to allow re-formation of stem 2 of the pseudoknot, restored the frameshift efficiency to the wild-type level *in vitro* and in cultured cells. Therefore, mutagenesis studies fully support our hypothesis that the stimulatory signal of subtype MVP5180 of HIV-1 group O adopts a pseudoknot structure.

### **RNA structure probing of the pseudoknot of subtype MVP5180 of HIV-1 group O**

To provide an independent support for the pseudoknot structure in subtype MVP5180 of HIV-1 group O frameshift stimulatory signal, we made a structural probing analysis of an RNA fragment encompassing the *gag/pol* frameshift region of this subtype. Enzymatic probing with RNase V<sub>1</sub>, an enzyme that cleaves RNA in helical conformation,



showed that the regions corresponding to stems 1 and 2 were attacked by the enzyme, with the strongest cleavage sites in stem 1 (Figure 5). Stem 2 was also sensitive to RNase V<sub>1</sub>, contrasting with loop 2, which was only weakly attacked. The stacking of the two nt in loop 1 likely accounts for its sensitivity to RNase V<sub>1</sub>. Altogether, probing experiments are in perfect agreement with the mutagenesis studies, showing that the frameshift stimulatory signal of HIV-1 group O MVP5180 is a pseudoknot.

### **Computer modeling of the pseudoknot of subtype MVP5180 of HIV-1 group O**

Structural analysis of frameshifting pseudoknots and other frameshift stimulatory signals can lead to the identification of conserved motifs and also to a prediction of the mechanism through which frameshift is stimulated.(Wang Y. et al. 2002) Based on the characteristics of their structure, two major classes of frameshifting pseudoknots can be proposed. In the first class, pseudoknots have an unpaired residue (usually an adenine) at the junction of the two stems, allowing a characteristic bent conformation that was proposed to be the conserved motif promoting frameshift stimulation.(Chen X. et al. 1996) The pseudoknot of the mouse mammary tumor virus (MMTV), resolved by NMR,(Shen et Tinoco 1995) and that of the beet western yellow virus (BWYV), resolved by X-ray crystallography,(Su et al. 1999) also belong to this class. The frameshifting pseudoknots of the second class do not have an unpaired residue at the junction of the coaxially stacked stems. Among pseudoknots of this class are those of the simian retrovirus-1 (SRV-1), resolved by NMR,(Michiels et al. 2001) and of the infectious bronchitis virus (IBV), which has been extensively studied by mutagenesis and probing studies.(Brierley et al. 1989; Liphardt et al. 1999; Napthine et al. 1999) No additional structural data on frameshifting pseudoknots with stacked stems are available, making identification of a conserved motif difficult. In addition to these two classes of pseudoknots, the Rous sarcoma virus (RSV) has an unusual pseudoknot, where the loop capping stem 1 is partly involved in the formation of stem 2 but contains two additional small sub-stem elements.(Marczinke et al. 1998)

Combining our mutagenesis and probing results with the structural information on – 1 frameshifting pseudoknots recently provided by crystallographic and NMR studies, we made an *in silico* modeling of the HIV-1 group O pseudoknot (Figure 6). As shown above, this pseudoknot was predicted to be a classic hairpin-type pseudoknot (e.g. with base-pairing in a hairpin loop). It has two eight-bp stems (stem 1 and stem 2) connected by loops of two nt (loop 1) and 20 nt (loop 2). Although the lengths of the stems and loops make the RNA pseudoknot of group O different from other pseudoknots promoting frameshift, these lengths still fit to the general steric requirements that allow the coaxial arrangement of the two stems.(Pearlman 1995) Among the pseudoknots whose tertiary structure has been resolved, that of the SRV-1 RNA(Michiels et al. 2001) is the best to be used as a starting conformation to model the structure of the group O RNA pseudoknot. In both pseudoknots, there is no intervening nt between the two stems, so that one can presume that, like in the SRV-1 RNA pseudoknot, there is no kink between the two stems of the group O pseudoknot. Also, the lengths of stems 1 and 2 in both molecules differ by only two bp while the length of loop 1 differs by only one nt, which necessitates only minimal rearrangements. The only region that is notably different in both pseudoknots is loop 2, which is eight-nt longer in the group O RNA than in the SRV-1 RNA, and is much longer than needed for connecting the two stems, which suggests the existence of additional elements in the secondary structure of the pseudoknot. Indeed, visual analysis of the nt sequence of loop 2 revealed the possibility to form an additional three-bp stem in this loop, immediately after the junction with stem 1 (stem SS in Figure 6). If this sub-stem is positioned coaxially to stem 1, it would effectively reduce the number of nt in loop 2 down to only eight. Modeling such a structure showed that formation of this additional substructure and its coaxial position with respect to stem 1 does not create any problem for the connection of stem 1 and stem 2 with loop 2. Formation of this sub-stem would provide an additional stabilizing effect for the whole structure, where most nt of the connector region become involved in H-bonding with the minor groove of stem 1, forming a triplex interaction. Although disruption of the putative sub-stem by mutagenesis did not affect

frameshift efficiency in our assays (data not shown), it is still possible that it could affect frameshifting with the full-length viral RNA, and not with a reporter system. A similar four-bp stem in loop 2 can also be proposed for the IBV pseudoknot, although a deletion in loop 2, impairing the formation of this putative stem, was found not to affect the frameshift efficiency.(Brierley et al. 1991) The triplex interaction between loop 2 and stem 1 in group O pseudoknot involves an AACAA sequence. This is reminiscent of the situation observed in BWYV(Su et al. 1999) and SRV-1,(Michiels et al. 2001) where an AACAA sequence present in loop 2 is implicated in loop-helix triplex interactions. Furthermore, as in SRV-1, the cytosine of the AACAA sequence bulges out in our modeling of group O pseudoknot. Su *et al.*(Su et al. 1999) and Michiels *et al.*,(Michiels et al. 2001) from their studies with BWYV and SRV-1, respectively, suggested that the formation of a triplex structure in a pseudoknot that promotes frameshift could be a signal to frameshift recognized by the translating ribosome. Our results with HIV-1 group O support this suggestion.

#### **Subtype ANT70 of HIV-1 group O does not contain the same frameshift stimulatory signal as subtype MVP5180**

As mentioned above, there is a second subtype for HIV-1 group O, ANT70,(Yamaguchi et al. 2002) for which only one complete *pol* sequence is available.(Los Alamos National Laboratory 2002) Sequence analysis suggests that subtype ANT70 uses the same slippery sequence, followed by the same eight-bp stem-loop as in subtype MVP5180. However, sequence differences in the region downstream of this stem-loop prevent formation in ANT70 of stem 2 of the pseudoknot present in MVP5180 (see Figure 4a). The frameshift efficiency of subtype ANT70 was assessed with construct pHIV/O-ANT70-LUC containing the frameshift region from subtype ANT70. It was found to be about half of that of pHIV/O-87-LUC both *in vitro* and in cultured cells (Figure 4b), which is consistent with the incapacity for this frameshift region to form the pseudoknot found in subtype MVP5180. This does not exclude however the possibility that the frameshift region of subtype ANT70 can form another pseudoknot (see Discussion).

## DISCUSSION

Our results show that the frameshift stimulatory signal of subtype MVP5180 of HIV-1 group O is a pseudoknot, where eight nt of a 10-nt loop capping an eight-bp stem base-pair with a downstream complementary sequence. The presence of this pseudoknot was demonstrated by site-directed mutagenesis and enzymatic probing of the frameshift stimulatory region. Among the two distinct classes of  $-1$  frameshifting pseudoknots described above, the group O pseudoknot falls in the same class as those of SRV-1 (Michiels et al. 2001) and IBV (Brierley et al. 1989) characterized by the absence of any intervening nt between the coaxial stems. Interestingly, the frameshift efficiency of the group O pseudoknot is two and four times lower than that of the SRV-1 and IBV pseudoknots, respectively. This can result from the difference in the length of stem 1, since shortening this stem from 11 to 10 bp caused a 85% loss of the frameshift efficiency for the IBV pseudoknot (Naphthine et al. 1999). The lower frameshift efficiency observed for the HIV-1 group O pseudoknot can also relate to the absence of a guanosine-rich region in the 5' arm of stem 1, which was found to contribute significantly to stimulation of frameshift in IBV (Naphthine et al. 1999), SRV-1 (ten Dam E. B. et al. 1995) and MMTV (Chamorro et al. 1992). Formation of stem 2 in group O stimulates frameshift by two to three-fold compared to the effect of a simple stem-loop (pHIV/O-87-LUC vs pHIV/O-60-LUC). A similar situation is encountered in the IBV (Naphthine et al. 1999) and RSV pseudoknots (Marczinke et al. 1998) where destabilization of stem 2 has a less drastic effect than in SRV-1 (ten Dam E. et al. 1994) and MMTV (Chamorro et al. 1992) for which the frameshift efficiency is reduced by eight to ten-fold under these conditions.

The use of a pseudoknot as a frameshift stimulatory signal makes HIV-1 group O distinct from group M, where this signal is an extended bulged stem-loop (Dulude et al. 2002). This observation supports previous phylogenetic analyses according to which the group O and group M clusters of HIV-1 originated from different monkey to human transmissions (Simon et al. 1998; Gao et al. 1999). Since evolution led to the selection of a

stimulatory signal in group O different from group M, it likely corresponds to a need specific to group O. In this study, we found that the frameshift efficiency in cultured cells was nearly 5% for group O MVP5180. When the frameshift region of group O was replaced with the frameshift region of group M subtypes, under conditions ensuring that any difference observed between the frameshift efficiency of group O and that of group M results from the difference in the structure of the stimulatory signal, we found that the frameshift efficiency of group M subtypes was about 2%, half of that of group O (our unpublished results). We can thus propose that the protease of HIV-1 group O viruses could be less active than group M protease, and that group O viruses require a Gag-Pol to Gag ratio higher than that of group M members, so as to incorporate more Gag-Pol in their virions for an optimal fitness during infection. Although there are no data available on the protease activity of group O, this protease is known to contain several secondary mutations that are found together with primary mutations causing resistance to protease inhibitors in variants of group M.(de Baar et al. 2000; Yang et al. 2000; Luk et al. 2001) Moreover, mutations implicated in resistance to protease inhibitors usually reduce the activity of the enzyme in the absence of inhibitors. Another possibility to account for the higher frameshift efficiency of group O is that the sequence variations in the frameshift region of group O, compared to group M, reduce the activity of the three peptides encoded by this region, the spacer peptide, p1, the transframe peptide, TFP, and p6. These peptides are important for the formation of infectious viral particles,(Louis et al. 1999; Demirov et al. 2002; Hill et al. 2002) and a higher production of Gag-Pol in group O viruses could compensate for a possible decrease in the activity of these peptides.

Surprisingly, the frameshift region of the other subtype of group O, ANT70, also contains an 8-bp stem capped by a 10-nt loop, but cannot form the same pseudoknot as MVP5180, since variations in the sequence downstream this stem-loop impair the formation of stem 2. The level of frameshift efficiency in this subtype, as measured in our reporter system, is about two-fold less than in MVP5180. We cannot, however, exclude the possibility that the loop capping stem 1 interacts with another region further downstream to

form a pseudoknot that would promote frameshift to a level close to that in MVP5180, and such complementary regions can indeed be found in the viral RNA sequence. Loop 2 in these putative pseudoknots would be very large, but other viruses, such as the human coronavirus, are known to use a frameshift stimulatory pseudoknot where loop 2 contains more than 160 nt.(Brierley 1995)

Finally, ribosomal frameshifting in HIV-1 represents an interesting novel target for an antiviral treatment. It is particularly important to develop new anti-HIV agents directed against group O viruses, since these viruses are naturally resistant to non-nucleoside reverse transcriptase inhibitors(Descamps et al. 1997) and bear mutations that could lead to faster resistance against protease inhibitors.(de Baar et al. 2000; Yang et al. 2000; Luk et al. 2001) The presence of a pseudoknot as a frameshift stimulatory signal in HIV-1 group O suggests that future anti-frameshift agents targeted against this signal must be developed independently for group M and group O members.

## MATERIALS AND METHODS

### Construction of plasmids

All the plasmids used in this study were derived from pcDNA3.1-LUC, originating from pcDNA3.1/Hygro(+) (Invitrogen) where the luciferase gene was inserted under control of a CMV and a T7 promoter.(Dulude et al. 2002) An oligonucleotide cassette containing a 5'UTR region of 55 nt, an initiator AUG and, two codons downstream, an *Eco47III* restriction site was inserted between restriction sites *KpnI* and *BamHI* of pcDNA3.1-LUC, generating pcDNA3.1-5'UTR-LUC, where the insertion precedes the second codon of luciferase. The frameshift region of HIV-1 group O subtypes MVP5180 and ANT70 (GenBank accession nos. L20571 and L20587, respectively) was inserted between the *Eco47III* and *BamHI* sites, creating pHIV/O-87-LUC (-1) and pHIV/O-ANT70-LUC (-1) constructs (Figure 2). These constructions are such that the luciferase coding sequence is in the -1 reading frame relative to the initiator codon, so that only ribosomes that make a -1 frameshift produce luciferase, while ribosomes translating in the 0 reading frame terminate translation at a stop codon overlapping codon six of the *luc* sequence. For each construct, an in-frame control plasmid, the (0) construct, was created by inserting an additional adenine just after the slippery sequence of the (-1) construct. Derivatives of pHIV/O-87-LUC (deletion and substitution mutants) were created by PCR, by first amplifying mutated DNA fragments from pHIV/O-87-LUC construct with two primers for deletion mutants or mutants with substitutions 3' of the stem-loop of the stimulatory signal, and with four primers for mutants with substitutions in the slippery sequence and in the stem-loop, according to the procedure of Ho *et al.*(Ho et al. 1989) The amplified DNA fragments were then subcloned between the *KpnI* and *BamHI* sites of pcDNA3.1-LUC, and all constructs were verified by sequencing the entire insert.

### Transient transfections and luciferase assays

Transfections of HIV-1 group O plasmids in 293T cells were carried using a standard calcium-phosphate precipitation method,(Jordan et al. 1996) with 3  $\mu\text{g}$  of a pHIV/O-LUC construct and 1.25  $\mu\text{g}$  of pcDNA3.1/Hygro(+)/lacZ coding for  $\beta$ -galactosidase, as previously described.<sup>(Dulude et al. 2002)</sup> For luciferase assays, 1.5  $\mu\text{l}$  of cell extract (600  $\mu\text{l}$ ) was added to 50  $\mu\text{l}$  of Luciferase Assay Reagent (Promega) and the light emitted was measured with a Berthold Lumat LB 9507 luminometer. Frameshift efficiencies were calculated by dividing the luciferase activity of the (-1) construct by the sum of the luciferase activity of the (0) and (-1) constructs. The  $\beta$ -galactosidase activity was measured with the chlorophenolred- $\beta$ -galactopyranoside substrate (Calbiochem), as described,(Eustice et al. 1991) 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,(Dulude et al. 2002) using *Stu*I-linearized pHIV/O-LUC constructs. These RNA transcripts (0.2  $\mu\text{g}$ ) were translated in 25  $\mu\text{l}$  of RRL (Promega) at 30°C for 15 minutes, a reaction time for which the translation system functions at its maximal rate. The reaction was stopped by addition of EDTA to a final concentration of 6 mM. Luciferase activity was monitored as mentioned above, with 2.5  $\mu\text{l}$  of the translation mixture. Frameshift efficiencies were calculated as described above.

### **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.(Marczinke et al. 1998; Dulude et al. 2002) An oligonucleotide cassette containing a T7 promoter followed by the HIV-1 group O *gag/pol* frameshift region (bases 26 to 87, Figure 2) was cloned between the *Nae*I and *Bsp*119I sites of pGEM®-7Zf(-) (Promega), generating the



recombinant plasmid pGEM-HIV/O. The RNA transcript, produced by *in vitro* transcription of the *Bsp*119I-linearized plasmid, was 5'-end-labeled with [ $\gamma$ - $^{32}$ P] using a standard dephosphorylation-rephosphorylation method,(Knapp 1989) purified from a 10% acrylamide-7M urea gel and dissolved in 500 mM NH<sub>4</sub>OAc, 10 mM Mg(OAc)<sub>2</sub>, 1mM EDTA and 0.1% SDS. Probing with RNase V<sub>1</sub> (in a total volume of 10  $\mu$ l containing 10<sup>5</sup> cpm of 5'-end-labeled RNA supplemented with 1  $\mu$ g of yeast tRNA) was done at 25°C for 15 minutes in 10 mM Tris-HCl (pH 8), 10 mM MgCl<sub>2</sub>, 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 was heated two minutes at 95°C and immediately loaded for analysis on a 10% and a 20% polyacrylamide-7 M urea gel.

### **Computer modeling**

Preliminary modeling was done interactively, using InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The solution structure of the SRV-1 RNA pseudoknot(Michiels et al. 2001) (PDB identifier 1E95), having six bp in each of the two stems and one and 12 nt in loops 1 and 2, respectively, was used as a starting conformation. The nt sequence of the molecule was changed and additional bp were appended to both stems according to the suggested model of the secondary structure for the RNA pseudoknot of group O subtype MVP5180. The unpaired nt of loops 1 and 2 were arranged as in the structure of these regions in the SRV-1 RNA pseudoknot, providing a reasonable system of H-bonds and base-base stacking interactions. The model was submitted to unrestrained energy minimization using the AMBER forcefield(Pearlman 1995) until an energy minimum was reached. Visualizations were done in a Silicon Graphics O2 computer. The PDB file is available on request.

## **ACKNOWLEDGMENTS**

This study was supported by a grant from the Canadian Institutes for Health Research to L. B.-G. and Guy Lemay and a grant from the Canadian Institutes for Health Research to S.V.S. S.V.S. acknowledges fellowships from the Canadian Institutes of Health Research and from the Fonds de la Recherche en Santé du Québec. We thank Guy Lemay for his interest in this work. We are grateful to Pascal Chartrand, Gerardo Ferbeyre and Hugo Soudeyns for stimulating discussions and critical reading of this manuscript.

### **Abbreviations used:**

BWYV, beet western yellow virus; CMV, cytomegalovirus; HIV-1, human immunodeficiency virus type 1; IBV, infectious bronchitis virus; LUC, firefly luciferase; MMTV, mouse mammary tumor virus; PCR, polymerase chain reaction; RRL, rabbit reticulocyte lysate; RSV, Rous sarcoma virus; SRV-1, simian retrovirus-1; TFP, Gag-Pol transframe protein.

**REFERENCES**

- Brierley, I. 1995. Ribosomal frameshifting viral RNAs. *J. Gen. Virol.* **76**: 1885-1892.
- Brierley, I., Digard, P. and Inglis, S.C. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**: 537-547.
- Brierley, I. and Pennel, S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. *The Ribosome*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press. pp 233-248.
- Brierley, I., Rolley, N.J., Jenner, A.J. and Inglis, S.C. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **220**: 889-902.
- 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. USA* **89**: 713-717.
- Chen, X., Kang, H., Shen, L.X., Chamorro, M., Varmus, H.E. and 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.
- de Baar, M.P., Janssens, W., de Ronde, A., Fransen, K., Colebunders, R., Kestens, L., van der Groen, G. and Goudsmit, J. 2000. Natural residues versus antiretroviral drug-selected mutations in HIV type 1 group O reverse transcriptase and protease related to virological drug failure in vivo. *AIDS Res. Hum. Retroviruses* **16**: 1385-1394.
- Demirov, D.G., Orenstein, J.M. and Freed, E.O. 2002. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *J. Virol.* **76**: 105-117.
- Descamps, D., Collin, G., Letourneur, F., Apetrei, C., Damond, F., Loussert-Ajaka, I., Simon, F., Saragosti, S. and Brun-Vezinet, F. 1997. Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. *J. Virol.* **71**: 8893-8898.

- 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.
- Du, Z., Giedroc, D.P. and Hoffman, D.W. 1996. Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. *Biochemistry* **35**: 4187-4198.
- 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**: 5094-5102.
- 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**: 739-740, 742-733.
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. and Hahn, B.H. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**: 436-441.
- Giedroc, D.P., Theimer, C.A. and Nixon, P.L. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.* **298**: 167-185.
- 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**: 11245-11253.
- 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**: 51-59.

- 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.
- 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.
- 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.
- 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.
- Knapp, G. 1989. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* **180**: 192-212.
- Liphardt, J., Naphine, S., Kontos, H. and Brierley, I. 1999. Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. *J. Mol. Biol.* **288**: 321-335.
- Los Alamos National Laboratory. 2002. A compilation and analysis of nucleic acid and amino acid sequences.: Los Alamos National Laboratory.
- Louis, J.M., Clore, G.M. and Gronenborn, A.M. 1999. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nat. Struct. Biol.* **6**: 868-875.
- Luk, K.C., Kaptue, L., Zekeng, L., Soriano, V., Gurtler, L., Devare, S.G., Schochetman, G. and Hackett, J., Jr. 2001. Naturally occurring sequence polymorphisms within HIV type 1 group O protease. *AIDS Res. Hum. Retroviruses* **17**: 1555-1561.
- 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.

- 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.
- Michiels, P.J., Versleijen, A.A., Verlaan, P.W., Pleij, C.W., Hilbers, C.W. and Heus, H.A. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**: 1109-1123.
- Napthine, S., Liphardt, J., Bloys, A., Routledge, S. and 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.
- 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.
- 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.
- Pearlman, D.A. 1995. AMBER 4.1. San-Francisco: University of California.
- Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J.P., Badombena, W., Luo, N., Vanden Haesevelde, M. and Delaporte, E. 1997. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**: 493-498.
- 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.
- 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.
- 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.

- Simon, F., Maucclere, P., Roques, P., Lousert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F. and Brun-Vezinet, F. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**: 1032-1037.
- Su, L., Chen, L., Egli, M., Berger, J.M. and Rich, A. 1999. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol.* **6**: 285-292.
- ten Dam, E., Brierley, I., Inglis, S. and 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, E.B., Verlaan, P.W. and Pleij, C.W. 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.
- Wang, Y., Wills, N.M., Du, Z., Rangan, A., Atkins, J.F., Gesteland, R.F. and Hoffman, D.W. 2002. Comparative studies of frameshifting and nonframeshifting RNA pseudoknots: a mutational and NMR investigation of pseudoknots derived from the bacteriophage T2 gene 32 mRNA and the retroviral gag-pro frameshift site. *RNA* **8**: 981-996.
- Yamaguchi, J., Vallari, A.S., Swanson, P., Bodelle, P., Kaptue, L., Ngansop, C., Zekeng, L., Gurtler, L.G., Devare, S.G. and Brennan, C.A. 2002. Evaluation of HIV type 1 group O isolates: identification of five phylogenetic clusters. *AIDS Res. Hum. Retroviruses* **18**: 269-282.
- Yang, C., Gao, F., Funjungo, P.N., Zekeng, L., van der Groen, G., Pieniazek, D., Schable, C. and Lal, R.B. 2000. Phylogenetic analysis of protease and transmembrane region of HIV type 1 group O. *AIDS Res. Hum. Retroviruses* **16**: 1075-1081.

Tableau 3-1 : Frameshift efficiency for different mutants of group O of HIV-1.



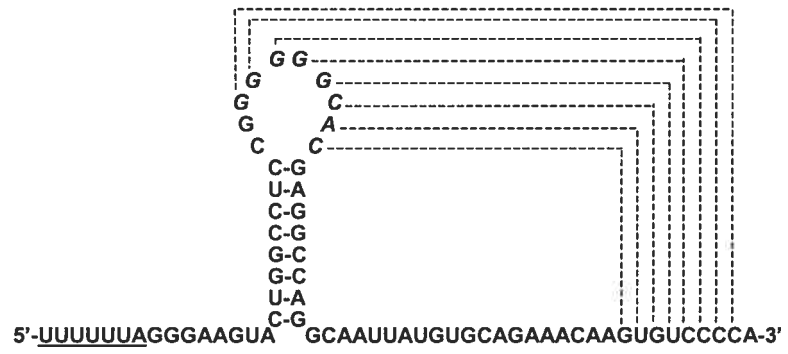
Construct	Description	Frameshift efficiency (%)	
		<i>In vitro</i>	In cultured cells
<b>Frameshift efficiency for regions of different length of group O</b>			
pHIV/O-87-LUC	Long frameshift region	10.2 (100)	4.7 (100)
pHIV/O-60-LUC	Short frameshift region	3.3 (32)	1.9 (40)
<b>Frameshift efficiency for mutants of group O</b>			
pHIV/O-k/o-LUC	Altered slippery sequence	1.0 (10)	0.2 (4)
pHIV/O-1.2-LUC	Substitution (3' strand of stem 1)	3.4 (33)	0.6 (13)
pHIV/O-1.12-LUC	Substitution (stem 1 re-formed)	10.0 (98)	4.5 (96)
pHIV/O-DSL-LUC	Deletion (minus stem 1 and its capping loop)	2.3 (23)	0.4 (9)
<b>Frameshift efficiency for mutations impairing or re-forming the pseudoknot</b>			
pHIV/O-2.1-LUC	Substitution (loop capping stem 1)	4.4 (43)	1.6 (34)
pHIV/O-2.2-LUC	Substitution (downstream segment)	3.9 (38)	2.1 (45)
pHIV/O-2.12-LUC	Compensatory (pseudoknot re-formed)	10.3 (101)	3.8 (81)
pHIV/O-ANT70-LUC	Long frameshift region for subtype ANT70 of group O	4.0 (39)	2.2 (47)

All pHIV/O-LUC constructs contain the HIV-1 Gag-Pol frameshift region of group O MVP5180 (except for construct pHIV/O-ANT70-LUC). Mutants of the group O frameshift region are further identified by a short description recalling their characteristics (see details in Figure 3a and 4a). For each (-1) construct, an in-frame (0) control was made to monitor the frameshift efficiency. The numbers between brackets represent the frameshift efficiency of each construct relative to pHIV/O-87-LUC, which is arbitrarily set at 100%. Results are the means of five independent experiments. Standard deviation was inferior or equal to 15%.

Figure 3-1 : Predicted structure for the frameshift region of subtype MVP5180 of HIV-1 group O.

- A) The slippery sequence (underlined) is followed by an eight-nt spacer and an eight-bp stem, capped by a 10-nt loop. Eight nt of this loop could base-pair with a complementary region downstream of the stem-loop (proposed pairings are represented by dashed lines). Such an interaction results in a pseudoknot with an eight-nt stem 1, a two-nt loop 1, an eight-nt stem 2 and a 20-nt loop 2.
- B) Structure of the frameshift region of subtype B of HIV-1 group M, as determined by Dulude *et al.*(Dulude et al. 2002)

a)



b)

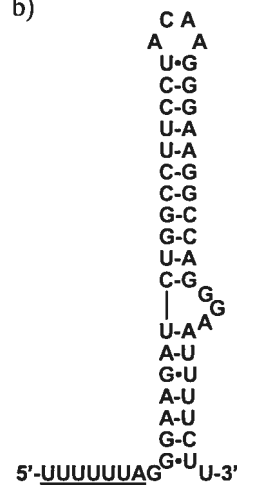


Figure 3-2 : Description of the luciferase expression vectors used for the study of the programmed -1 ribosomal frameshift of HIV-1 group O *in vitro* and in cultured cells.

- A) The HIV-1 frameshift region of subtype MVP5180 of group O was inserted upstream the coding sequence of the luciferase reporter gene, generating construct pHIV/O-87-LUC. The slippery site is UUUUUA (underlined). All mutants of subtype MVP5180 were cloned by inserting between the *KpnI*-*Bam*HI sites of the vector the PCR product bearing the mutation investigated. For subtype ANT70 of group O, the corresponding vector was constructed by exchanging the *Eco*47III-*Bam*HI fragment with an appropriate oligonucleotide cassette. For the (-1) constructs, the luciferase sequence is in the -1 reading frame relative to the AUG initiation codon, so that a -1 frameshift is required to produce luciferase. An adenine was added immediately after the slippery sequence (at position 25) for the (0) constructs, so that luciferase is expressed by ribosomes that do not shift the reading frame.
- B) Sequences of the frameshift region of all constructs used in this study. Nt substituted or deleted compared to subtype MVP5180 of group O are underlined or represented by dashed lines, respectively.



Figure 3-3 : Effect of different mutations in the frameshift region of subtype MVP5180 of HIV-1 group O on the frameshift efficiency.

- A) A series of mutations were made within the frameshift region of pHIV/O-87-LUC (the dots correspond to the *Bam*HI linker connecting the frameshift region to the luciferase coding sequence): a slippery site mutant, pHIV/O-k/o-LUC, where the slippery sequence (underlined) is mutated (bases that are changed are in capital letters); deletion mutant pHIV/O-DSL-LUC, where the region encompassing stem 1 and its capping loop is deleted (deletion of bases 33 to 60); deletion mutant pHIV/O-60-LUC, where the region 3' to stem 1 is deleted; substitution mutants where the 3' strand of stem 1 is altered, impairing formation of stem 1 (pHIV/O-1.2-LUC), and where the 3' and 5' strands of stem 1 are simultaneously altered, allowing re-formation of stem 1 (pHIV/O-1.12-LUC).
- B) Frameshift efficiency *in vitro* and in cultured cells for the pHIV/O-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 the *Stu*I-digested pHIV/O-LUC constructs. Assays in cultured cells were made by co-transfecting 293T cells with 3  $\mu$ g of pHIV/O-LUC and 1.25  $\mu$ g of pcDNA3.1/Hygro(+)/lacZ. Frameshift efficiencies were calculated as described in the text. Each value represents the mean  $\pm$  standard error of five to six independent experiments. The bars indicate the standard error on the mean.

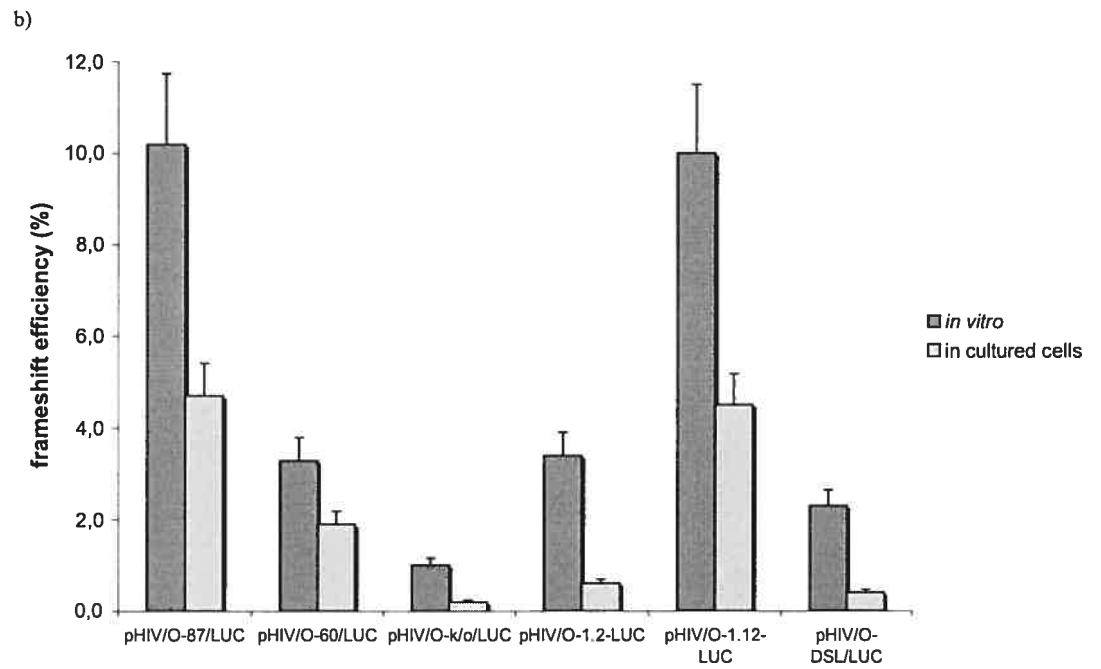
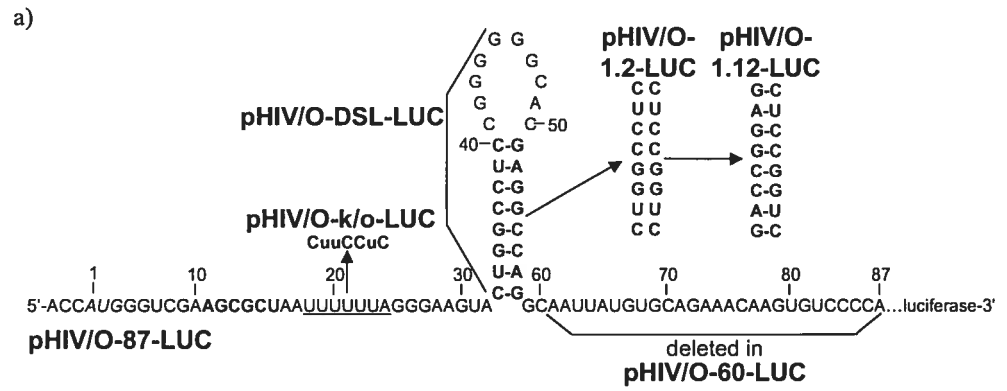


Figure 3-4 : Effect on the frameshift efficiency of mutations impairing formation of stem 2 of the pseudoknot of subtype MVP5180 of HIV-1 group O.

- A) Description of mutations made within the *gag/pol* frameshift region of pHIV/O-87-LUC (the dots correspond to the *Bam*HI linker connecting the frameshift region to the luciferase coding sequence). Mutations were introduced either in the loop capping stem 1 or in the complementary downstream region, impairing formation of stem 2 (pHIV/O-2.1-LUC and pHIV/O-2.2-LUC, respectively), or allowing re-formation of this stem (pHIV/O-2.12-LUC). Subtype ANT70 of group O, where sequence differences compared to subtype MVP5180 impair formation of stem 2, was also assessed.
- B) Frameshift efficiency *in vitro* and in cultured cells with the HIV/O-LUC constructs described above. Assays were as described in the legend to Figure 3.



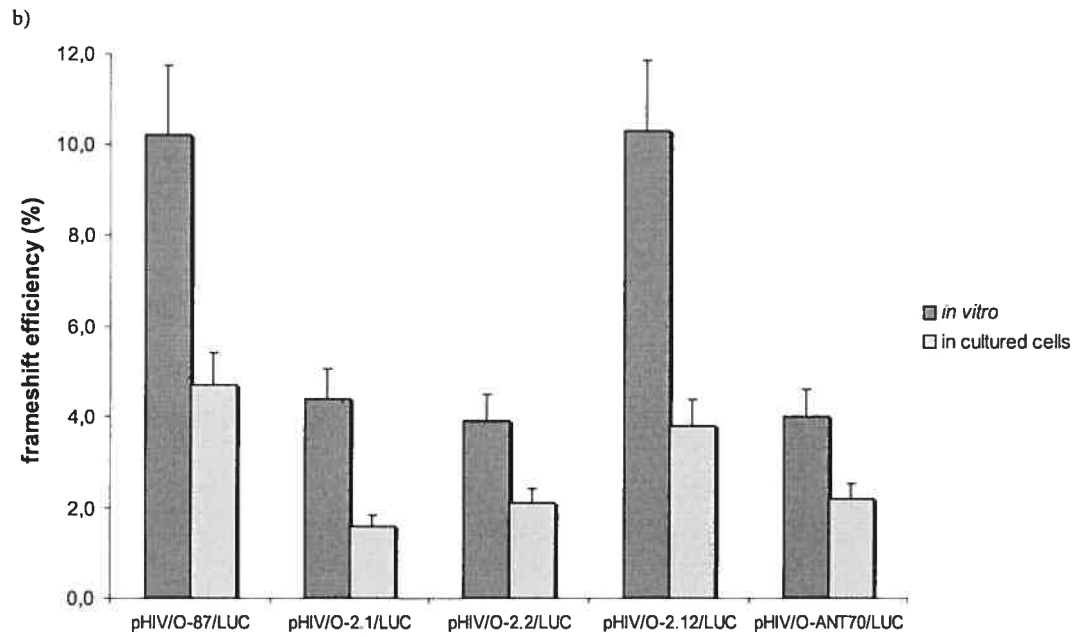
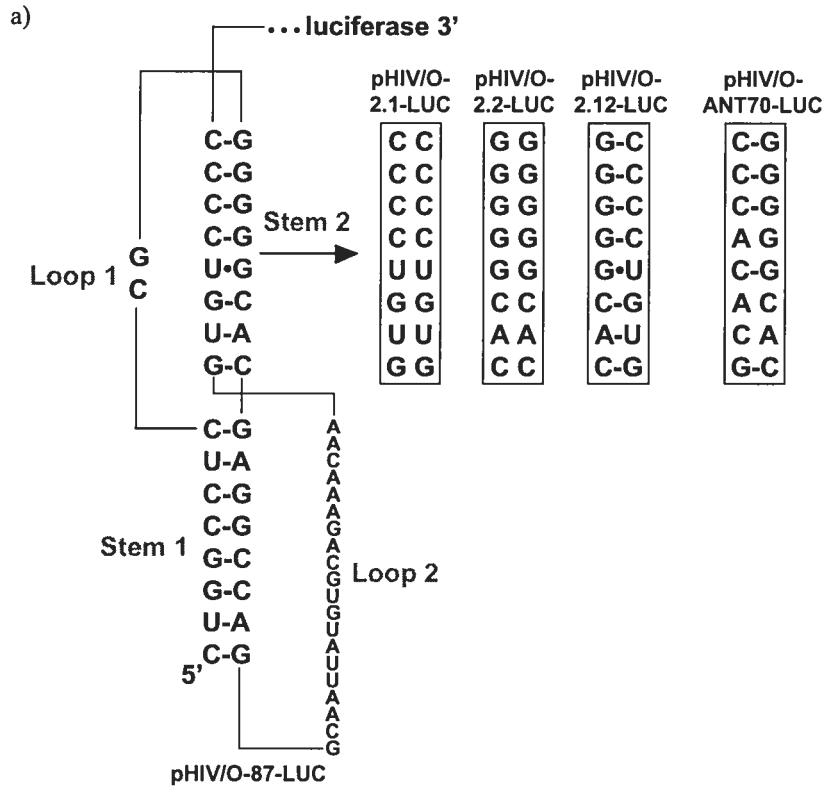


Figure 3-5 : Probing of the structure proposed for the frameshift stimulatory signal of subtype MVP5180 of HIV-1 group O.

- A) Structure probing of the frameshift stimulatory signal by RNase V<sub>1</sub> attack. An RNA transcript encompassing the *gag/pol* frameshift region was 5' end-labeled with [ $\gamma$ -<sup>32</sup>P] and digested with RNase V<sub>1</sub>. Digestion products were analyzed on a 20% (left) and a 10% (right) 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). The nt that were cleaved were identified by the absence of cleavage in the untreated control lane (0). The amount of units of enzyme added to each reaction is also indicated.
- B) Summary of the RNase V<sub>1</sub> attacks in the pseudoknot structure of the frameshift stimulatory signal of subtype MVP5180 of HIV-1 group O. The sensitivity of nt to RNase V<sub>1</sub> is shown by arrows of different size, where the size is approximately proportional to the intensity of the cleavage at that site. Bases 1 to 62 originate from subtype MVP5180, while bases 63 to 66 (in gray) originate from the vector.

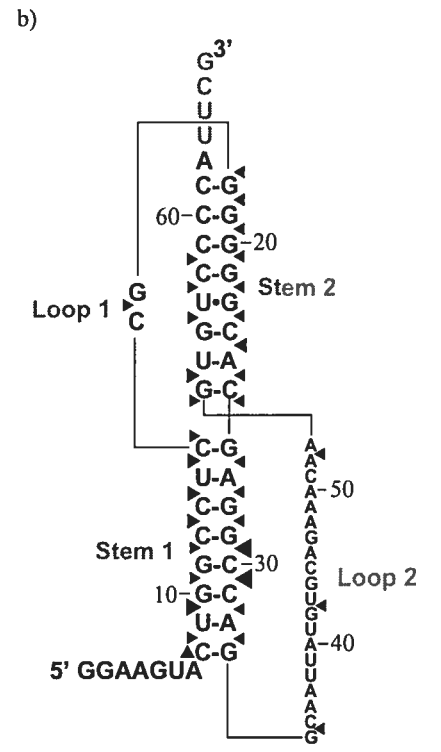
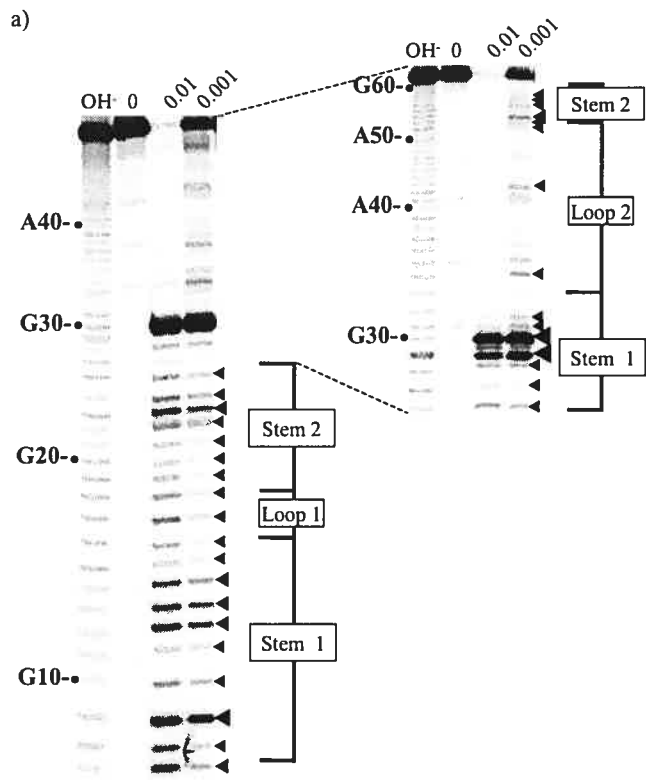
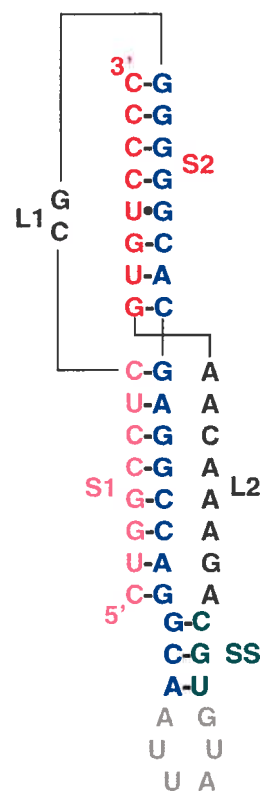
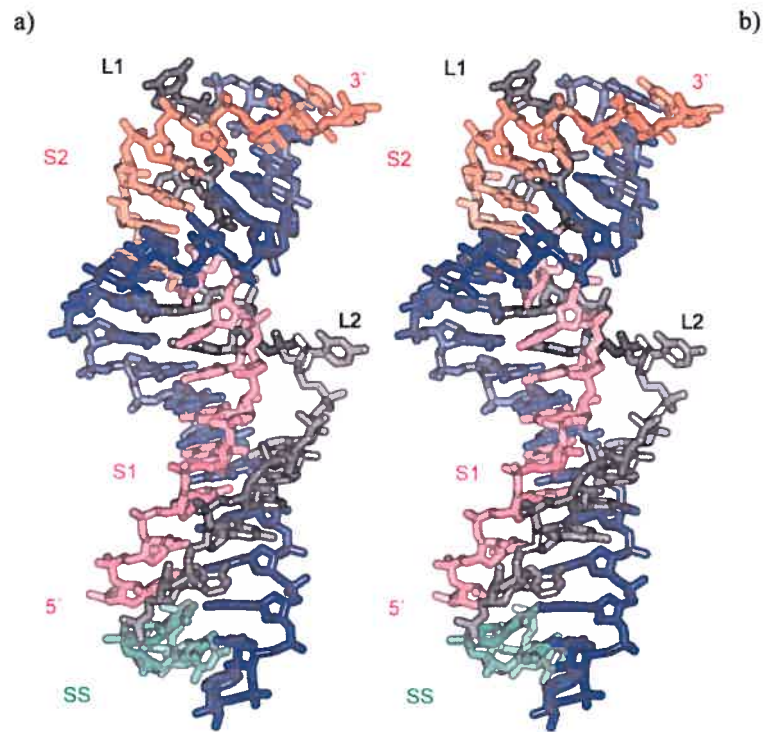


Figure 3-6 : Stereo view of the computer modeled structure for the pseudoknot of subtype MVP5180 of HIV-1 group O.

Stems (S1 and S2) and loops (L1 and L2) are represented in different colors.

A) Structure of the MVP5180 pseudoknot, with a putative sub-stem (SS) in loop 2 (see the text). Nt in the loop of this sub-stem were not included in our modeling.

B) Schematic representation of the pseudoknot.



## **Chapitre IV**

**Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein.**

Martin Baril and Léa Brakier-Gingras

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

NAR, 33: 1474-86, 2005.

## ABSTRACT

The hepatitis C virus (HCV) genome contains an internal ribosome entry site (IRES) followed by a large open reading frame coding for a polyprotein that is cleaved into ten proteins. An additional HCV protein, the F protein, was recently suggested to result from a +1 frameshift by a minority of ribosomes that initiated translation at the HCV initiator AUG of the polyprotein. In the present study, we reassessed the mechanism accounting for the synthesis of the F protein by measuring the expression in cultured cells of a luciferase reporter gene with an insertion encompassing the IRES plus the beginning of the HCV coding region preceding the luciferase coding sequence. The insertion was such that luciferase expression was either in the +1 reading frame relative to the HCV initiator AUG, mimicking the expression of the F protein, or in-frame with this AUG, mimicking the expression of the polyprotein. Introduction of a stop codon at various positions in-frame with the AUG initiator codon and substitution of this AUG with UAC inhibited luciferase expression in the 0 reading frame but not in the +1 reading frame, ruling out that the synthesis of the F protein results from a +1 frameshift. Introduction of a stop codon at various positions in the +1 reading frame identified the codon overlapping codon 26 of the polyprotein in the +1 reading frame as the translation start site for the F protein. This codon 26(+1) is either GUG or GCG in the viral variants. Expression of the F protein strongly increased when codon 26(+1) was replaced with AUG, or when its context was mutated into an optimal Kozak context, but was severely decreased in the presence of low concentrations of edeine. These observations are consistent with a Met-tRNA<sub>i</sub>-dependent initiation of translation at a non-AUG codon for the synthesis of the F protein.

## INTRODUCTION

Hepatitis C virus (HCV) chronically infects around 200 million people worldwide and frequently causes liver cirrhosis and hepatocellular carcinoma (Saito et al. 1990; Scheuer et al. 1992; Lefkowitz et al. 1993). HCV is a member of the Flaviviridae family (Choo et al. 1991) and has a positive-stranded RNA genome of ~9.6 kb. This RNA contains an internal ribosome entry site (IRES) (Wang C. et al. 1993; Reynolds et al. 1996; Rijnbrand et Lemon 2000), which controls the initiation of translation of a large open reading frame encoding a polyprotein of ~3000 amino acids. Proteolytic cleavage of this polyprotein by host signal peptidases and two viral proteases produces four structural proteins (core, E1, E2 and P7) and six nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B) (reviewed in Kato 2000; Reed et Rice 2000; Rosenberg 2001).

It was also observed that the core-encoding region of HCV expresses at a low level an additional protein of about 16-17 kDa (Lo et al. 1994; Lo et al. 1995; Ray et al. 1996). A conserved open reading frame in the +1 reading frame relative to the polyprotein was subsequently observed in the core-encoding region and it was shown that the sera of HCV-infected patients reacted against peptides corresponding to this +1 reading frame (Walewski et al. 2001). Xu *et al.* (Xu et al. 2001) showed that this protein, which was named the F protein, was produced by a +1 translational frameshift, by ribosomes that initiated translation at the AUG codon used for the synthesis of the HCV polyprotein and shifted the reading frame between codons 9 to 11 of this polyprotein. This mechanism was deduced from the N-terminal sequencing of the translation product generated when translating the beginning of the HCV core region in a rabbit reticulocyte lysate (RRL). This +1 frameshift was observed with a HCV sequence that contains a stretch of ten adenines (10A) encompassing the proposed frameshift site, a sequence that is underrepresented among the viral variants (2 out of 721 sequences available at the hepatitis virus database: <http://s2as02.genes.nig.ac.jp/>). However, a stretch of A in a messenger RNA is known to promote frameshifting at a high efficiency *in vitro* (Weiss et al. 1987), but much less in



cultured mammalian cells where the control of the translation accuracy and of the maintenance of the reading frame is much more stringent (see Farabaugh 1997b). Also, antibodies against the F protein can be detected in patients infected with any HCV genotype, whether the 10A stretch is present or not (Yeh et al. 2000; Bain et al. 2004). In an independent study, Boulant *et al.* (Boulant et al. 2003) sequenced translational products generated by expressing a HCV fragment in *E. coli* coding for the core protein and they also concluded that the F protein results from a +1 ribosomal frameshift, but at codon 42 of the polyprotein, an AGG codon. This codon is a rare codon in bacteria but not in humans (codon usage database: <http://www.kazusa.or.jp/codon/>). It could therefore promote a +1 frameshift by slowing down the elongating ribosomes in bacteria (Farabaugh 1997b) but this does not hold for humans. Lastly, Vassilaki and Mavromara (Vassilaki et Mavromara 2003), when expressing a fusion of the HCV core-coding sequence to a luciferase reporter gene in cultured cells, detected another HCV protein, smaller than the protein generated by a +1 frameshift and showed that this protein results from the initiation of translation in a +1 reading frame relative to the polyprotein, at an AUG codon overlapping codon 86 or 88 of this polyprotein. All these observations raise questions on the mechanism that account for the synthesis of the F protein and led us to reassess this mechanism by investigating the expression of a luciferase reporter with an N-terminal insertion encompassing the beginning of the HCV coding sequence *in vitro* and in cultured cells. In the site that was proposed by Xu *et al.* to promote a +1 ribosomal frameshift, we introduced either the consensus sequence (AAAGAAAAC) found in most HCV variants or the previously used stretch of 10A. We found that the previously reported +1 frameshift was artificially caused by the presence of the 10A stretch and that the synthesis of the F protein results from a direct initiation in the +1 reading frame, at a non-AUG codon (GUG or GCG) overlapping codon 26 of the polyprotein.

## MATERIALS AND METHODS

### Construction of plasmids

All the plasmids used in this study were derived from pcDNA3.1-LUC, originating from pcDNA3.1/Hygro(+) (Invitrogen), where the firefly luciferase gene is inserted under control of a CMV and a T7 promoter (Dulude et al. 2002). Plasmid pCore1a-33, which contains the 5'UTR plus the beginning of the coding sequence of HCV, was a gift from Dr. H. Soudeyns (Hôpital Ste-Justine, Montréal) and was created from a viral sequence obtained from the plasma of a patient infected with HCV genotype 1a. This plasmid provided the HCV sequence used in this study. Portions of different length were amplified by PCR from this plasmid and inserted between the *KpnI* and *BamHI* sites of pcDNA3.1-LUC, creating the pHCV-LUC (0) series (Fig. 2). These (0) constructs are such that the luciferase coding sequence is in-frame with the HCV AUG initiation codon, so that only the ribosomes that translate in the reading frame of this AUG produce luciferase. For each pHCV-LUC (0) construct, a (+1) construct was created by inserting an adenine just after the *BamHI* site of the (0) construct. These (+1) constructs are such that the luciferase coding sequence is in the +1 reading frame relative to the HCV AUG initiation codon, so that only ribosomes that translate in this +1 reading frame produce luciferase. Derivatives of pHCV-447-LUC (0) and (+1) constructs (the parental constructs that were used to generate deletion and substitution mutants) were created by PCR, by amplifying mutated DNA fragments with two primers for deletion mutants and with four primers for substitution mutants, according to the procedure of Ho *et al.* (Ho et al. 1989). The amplified DNA fragments were 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**

Transfections of the plasmids of the pHCV-LUC series into HEK 293FT cells (Human Embryonic Kidney fibroblast-like cells, Invitrogen) were carried out as previously described (Dulude et al. 2002), using a standard calcium phosphate precipitation method (Jordan et al. 1996), with 3  $\mu\text{g}$  of pHCV-LUC (0) or (+1) and 1  $\mu\text{g}$  of pcDNA3.1/Hygro(+)/lacZ, which codes for  $\beta$ -galactosidase. For luciferase assays, 1  $\mu\text{l}$  of a 600  $\mu\text{l}$  cell extract was added to 25  $\mu\text{l}$  of the Luciferase Assay Reagent (Promega) and the amount of light emitted was measured with a Berthold Lumat LB 9507 luminometer. The synthesis of the F protein relative to the synthesis of the polyprotein was calculated by dividing the luciferase activity of the (+1) construct by the luciferase activity of the corresponding (0) construct, unless otherwise indicated. The  $\beta$ -galactosidase activity was measured with the chlorophenolred- $\beta$ -galactopyranoside substrate (Calbiochem) (Eustice et al. 1991), with aliquots of 10  $\mu\text{l}$  of cell extracts and it was 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 (Dulude et al. 2002), using *FbaI*-linearized pHCV-LUC (0) and (+1) constructs. The RNA transcripts (0.2  $\mu\text{g}$ ) were translated in 25  $\mu\text{l}$  of RRL (Promega) at 30°C for 20 minutes, a reaction time for which the translation system functions at its maximal rate. When the translation assays were made in the presence of edeine, the RRL was preincubated with the indicated concentration of this compound for 5 min at 37°C before the addition of the RNA transcripts. The reaction was stopped by addition of EDTA to a final concentration of 6 mM. For luciferase assays, 2.5  $\mu\text{l}$  of the translation mixture was added to 25  $\mu\text{l}$  of the Luciferase Assay Reagent (Promega) and the amount of light emitted was measured with a Berthold Lumat LB 9507 luminometer. The synthesis of the F protein relative to the synthesis of the polyprotein was calculated as described above.

## RESULTS

### Mapping of the HCV coding region required to produce the F protein

To reassess the mechanism controlling the synthesis of the F protein from HCV RNA, we introduced fragments of different length of the HCV genotype 1a genome (Fig. 1) before the coding sequence for the firefly luciferase gene (*luc*) harbored by a pcDNA3.1/Hygro(+) plasmid, under control of a CMV and a T7 promoter. This generated a series of pHCV-LUC (0) constructs, an example of which, pHCV-447-LUC (0), is shown in Figure 2A. In these constructs, the insertions are such that ribosomes initiate translation at the HCV AUG initiation codon (nt 342-344 of the HCV RNA) and translate the portion of HCV viral RNA and the fused in-frame *luc* sequence. Synthesis of luciferase in these (0) constructs mimicks the expression of the HCV polyprotein. These constructs contained either the complete IRES, starting at nt 1 of HCV RNA (thereafter referred to as IRES constructs) or a small fragment of the 5'UTR of HCV RNA, starting at nt 290 (thereafter referred to as cap constructs). In the cap constructs, initiation of translation is mediated through recruitment of the 40S ribosomal subunits to the 5' end of the mRNA and scanning of the 5'UTR until the 40S subunits encounter the polyprotein AUG initiation codon (see Hershey et al. 1996; Pestova et Hellen 2001; Dever 2002; Poulin et Sonenberg 2003 for reviews of cap-dependent initiation). In the IRES constructs, translation initiation occurs through the HCV IRES-mediated binding of the 40S subunits directly to the polyprotein AUG initiation codon (see Hellen et Pestova 1999 for reviews of HCV IRES-dependent initiation; Rijnbrand et Lemon 2000). For each (0) construct, a (+1) construct was made by adding an adenine before the *luc* coding sequence (Fig. 2A), so that the luciferase coding sequence is in the +1 reading frame relative to the HCV AUG initiation codon and that only ribosomes that translate the HCV viral RNA in the +1 reading frame synthesize the luciferase. Synthesis of luciferase in these (+1) constructs mimicks the expression of the HCV F protein. The synthesis of luciferase in the pHCV-LUC (0) and (+1) constructs was assessed in cultured cells and *in vitro*, in a RRL.

We started by defining the HCV coding region required to synthesize the maximal amount of the F protein. To this end, four series of (0) and (+1) cap and IRES-constructs of different length were produced: pHCV-387-LUC, pHCV-426-LUC, pHCV-447-LUC and pHCV-510-LUC, where the number in the name of each construct indicates the nt at the 3' end of the HCV insertion (see Fig. 1). In all these constructs, the sequence encompassing codons 9 to 11 of the polyprotein, corresponding to nt 366-374 of the HCV viral RNA, is the consensus HCV sequence. Derivatives of all these constructs were made by mutagenesis, which contained the stretch of 10A (thereafter referred to as 10A constructs) that was previously proposed by Xu *et al.* (Xu *et al.* 2001) to mediate a +1 ribosomal frameshift (see Fig. 2A).

Figure 3 shows the amount of the F protein (assessed by expression of luciferase in the +1 frame) synthesized in cultured cells (Fig. 3A) or *in vitro* (Fig. 3B), with the four series of constructs described above and with the corresponding 10A constructs. The synthesis of the F protein is expressed relative to the amount of polyprotein. The background of expression, which is about 0.4% in cultured cells and about 1% *in vitro*, was subtracted from the results presented in the figures. This background was determined with a control construct where a UGA stop codon was inserted immediately preceding the *luc* coding sequence. In cultured cells, with pHCV-387-LUC, the F protein was produced at 0.1% to 0.2% of the polyprotein for all members of this series (IRES or cap constructs with the consensus or 10A sequence), which is not significantly different from the background. Lengthening the HCV insertion, in pHCV-426-LUC, increased the relative synthesis of the F protein to 0.6% to 0.9% and an additional lengthening of the HCV insertion with nt 427-447, in pHCV-447-LUC, further increased the synthesis of the F protein to 1.5% to 2.6% of the polyprotein. However, a longer HCV insertion, in pHCV-510-LUC, decreased the relative synthesis of the F protein to a level similar to that obtained with pHCV-426-LUC.

Our results with cultured cells show that the synthesis of the F protein occurs whether initiation of translation is cap-dependent or IRES-dependent and that the efficiency of this synthesis is maximal for pHCV-447-LUC, which contains the first 35 codons from

the HCV coding sequence. The level of expression was always low, even for pHCV-447-LUC, for which it is about five-fold the background value, but it was reproducibly observed with independent constructs and DNA preparations. Our results also show that the presence of the 10A sequence used by Xu *et al.* (Xu et al. 2001) does not increase the amount of the F protein synthesized, compared to the consensus sequence. This observation is in agreement with the fact that, as mentioned above, antibodies against the F protein were detected in patients infected with any HCV genotype, regardless of the presence or the absence of the 10A stretch (Yeh et al. 2000; Bain et al. 2004). We also assayed our constructs *in vitro* by translating in an RRL the various corresponding HCV-luciferase mRNAs obtained by transcription of the linearized plasmids. It can be seen (Fig. 3B) that, with the consensus HCV sequence, the pattern of synthesis of the F protein is similar to that obtained in cultured cells, i.e. it does not differ significantly from the background for pHCV-387-LUC, increases with the pHCV-426-LUC and pHCV-447-LUC constructs, reaching a maximum level of 1.7% and 7.0% for the cap and IRES constructs, respectively, but decreases with pHCV-510-LUC to the levels observed with pHCV-426-LUC. However, with the constructs containing the 10A stretch, synthesis of the F protein is around 15% to 20%, independent of the length of the HCV insertion. *In vitro* translational systems are less accurate than cultured cells (see Farabaugh 1997b), as recalled in the Introduction. From comparing the results in cultured cells and *in vitro*, we conclude that the presence of the 10A stretch, a slippery sequence that is known to promote frameshifting *in vitro*, favors a +1 frameshift allowing the synthesis of the F protein, in the +1 reading frame relative to the polyprotein. This efficient frameshift likely masks any other event occurring at a lower efficiency, which could contribute to the synthesis of the F protein. The observation that the F protein was produced by a +1 ribosomal frameshift (Xu et al. 2001) appears to be linked to the use of a peculiar viral sequence and this prompted us to reassess the mechanism accounting for the synthesis of the F protein, using the consensus viral sequence and not the 10A stretch at the site previously proposed for frameshifting.

### **HCV F protein is not produced by a +1 ribosomal frameshift**

To investigate the mechanism responsible for the synthesis of HCV F protein, we used pHCV-447-LUC, the construct with which the amount of the F protein synthesized is maximal. We first introduced mutations in pHCV-447-LUC (cap or IRES construct, with the consensus sequence for codons 9 to 11), by inserting a UGA stop codon in the HCV polyprotein reading frame. This stop codon was introduced at codon 3, 9, 10, 11 or 12, generating pHCV-UGA3-LUC, pHCV-UGA9-LUC, pHCV-UGA10-LUC, pHCV-UGA11-LUC and pHCV-UGA12-LUC (see Fig. 2B for the description of these mutants and the other mutants used in the rest of this study). We reasoned that if the F protein was produced by a +1 ribosomal frameshift, inserting a stop codon in the HCV polyprotein reading frame should abolish translation in the +1 reading frame when it is inserted before the frameshift site, since no elongating ribosomes would then reach this frameshift site. However, assays in cultured cells showed that none of the inserted stop codon decreased the synthesis of the F protein (Fig. 4), whereas all the inserted stop codons severely impaired the synthesis of the polyprotein (data not shown). The same observation was made with translation assays in a RRL (data not shown). These results therefore indicate that the F protein is not produced by elongating ribosomes that initiated translation at the polyprotein AUG initiation codon and shifted the reading frame at a specific site.

### **A decrease in the efficiency of the initiation of translation of HCV polyprotein favors the synthesis of the F protein**

If the F protein is not synthesized by a +1 ribosomal frameshift, a likely hypothesis is that it results from initiation of translation in the +1 reading frame relative to the reading frame of the HCV polyprotein. If this is the case, the translational start site for the polyprotein and the one for the F protein must compete for the amount of available 40S ribosomal subunits. To get further insight into the mechanism of synthesis of the F protein, we made two derivatives of pHCV-447-LUC. The first one is pHCV-uuuAUG-LUC, for

which the three nt upstream of the initiator AUG codon were mutated so as to change its context from an optimal context for initiation of translation, according to Kozak (Kozak 1986), to a weak context. The second one is pHCV-accUAC-LUC, for which the initiator codon for the polyprotein was mutated to UAC, so as to impair recognition by Met-tRNA<sub>i</sub>. It can be seen in Figure 5 that the translation efficiency of the polyprotein was decreased by five to ten-fold with pHCV-uuuAUG-LUC and by 15 to 50-fold with pHCV-accUAC-LUC, in the IRES and cap constructs, respectively. However, the translation of the F protein increased by two to three-fold in pHCV-uuuAUG-LUC and pHCV-accUAC-LUC, whether the initiation of translation was cap or IRES-dependent. These results confirm that the synthesis of the F protein does not result from a +1 frameshift by elongating ribosomes, since reducing the synthesis of the polyprotein did not decrease the amount of the F protein produced but rather increased it. They support the hypothesis that the synthesis of the F protein results from an initiation in the +1 reading frame relative to the AUG initiator used for the polyprotein.

#### **Location of the F protein translation start site**

The next step consists in determining where is located the translation initiation site for the F protein. There is no AUG codon in the +1 reading frame in the HCV coding sequence, inserted in pHCV-447-LUC, which suggests that the translation initiation for the F protein takes place at a non-AUG initiation codon. Initiation at a non-AUG codon is now well-documented in mammalian cells and in viruses (reviewed in Peabody 1989; Kozak 1999; Touriol et al. 2003). To locate the initiation codon for the F protein, we introduced a stop codon at various positions in the +1 reading frame of the HCV coding sequence in pHCV-447-LUC, knowing that insertion of a stop codon in the +1 reading frame should impair translation of the F protein if it is positioned at the translational start site or downstream of it. We first narrowed the region encompassing the translation initiation site for the F protein by inserting a stop codon every tenth codon. We found that the synthesis



of the F protein was not impaired when a stop codon was inserted at positions overlapping codon 10 or 20 of the polyprotein sequence in the +1 reading frame (codon 10(+1) or 20(+1)), but this synthesis was strongly decreased when the stop codon was inserted at codon 30(+1) (data not shown). We then introduced a stop codon in the +1 reading frame overlapping codon 23, 24, 25 or 26 of the polyprotein, generating, respectively, pHCV-UAA23+1-LUC, pHCV-UGA24+1-LUC, pHCV-UGA25+1-LUC and pHCV-UGA26+1-LUC. In pHCV-UAA23+1-LUC and pHCV-UGA24+1-LUC, the nt downstream of the inserted stop codon was mutated to an adenine to ensure efficient termination of translation at this codon (Tate et al. 1995). It can be seen (Fig. 6) that, in cultured cells, the synthesis of the F protein is not decreased with pHCV-UAA23+1-LUC and pHCV-UGA24+1-LUC, but, with pHCV-UGA25+1-LUC and pHCV-UGA26+1-LUC, it is reduced to about one-fourth or one-fifth, to a value near the background. This decrease was also observed with the same two constructs in a RRL (data not shown). An immediate conclusion would be that initiation of translation for the F protein occurs at the CGG codon overlapping codon 25 of the polyprotein. However, we considered the possibility that the GUG codon 26(+1) could also be the translation start site and that mutating codon 25(+1) could affect the F protein synthesis by altering the context of codon 26(+1).

To discriminate between codons 25(+1) and 26(+1) as candidates for the initiation of translation of the F protein, we made another series of mutants derived from pHCV-447-LUC (Fig. 7A): an additional nt (U) was added immediately after codon 25(+1) (pHCV+U418-LUC) or after codon 26(+1) (pHCV+U421-LUC). These insertions change the reading frame of the coding sequence of the reporter gene and should impair luciferase expression, which monitors the synthesis of the F protein, if translation initiation of the F protein takes place before the inserted nt. The synthesis of the F protein was not impaired with pHCV+U418-LUC, but was strongly reduced with pHCV+U421-LUC (Fig. 7B), which indicates that the translational start site for the F protein is located at codon 26(+1). We also deleted nt 447 (U) in pHCV+U421-LUC, generating pHCV+U421 $\Delta$ U447-LUC, in which the luciferase sequence was replaced in the reading frame corresponding to codon

26(+1) (Fig. 7B). In this mutant, the synthesis of the F protein was restored to its value in pHCV-447-LUC, confirming that the decrease in the synthesis of the F protein observed with pHCV+U421-LUC resulted from a change in the reading frame and that codon 26(+1) is used to initiate translation of the F protein.

### **Further information on the initiation of translation of the F protein**

Codon 26(+1), which we identified as the start codon for the synthesis of the F protein, is GUG in pHCV-447-LUC. However, this codon is GUG in only 18% of the HCV sequences available at the hepatitis virus database (<http://s2as02.genes.nig.ac.jp/>), but it is GCG in 81% of the sequences. We therefore decided to mutate this GUG codon to GCG, generating pHCV-GCG26+1-LUC, and to investigate the synthesis of the F protein in the corresponding cap and IRES constructs. The synthesis of the F protein was not decreased in the IRES construct (Fig. 8), revealing that translation initiation of the F protein can take place at codon 26(+1) of the polyprotein, whether it is GUG or GCG. However, with the corresponding cap construct, the synthesis of the F protein was strongly decreased.

It was previously shown that the 40S subunits that initiate translation with the HCV IRES cannot use a leaky scanning to initiate downstream from the polyprotein AUG initiator codon (Rijnbrand et al. 1996). This suggests that initiation of translation at codon 26(+1) should result from direct positioning of the 40S subunits at this codon when the initiation is IRES-mediated. To verify this suggestion, we introduced, in the 0 reading frame, an AUG with an optimal Kozak context (accAUGg) at codon 23 of the HCV polyprotein coding sequence for both the cap and IRES constructs of pHCV-447-LUC. With the pHCV-AUG23-LUC cap construct, the 40S subunits that scan the viral RNA after skipping the first initiator start site should initiate translation at the inserted AUG at codon 23, reducing the amount of 40S subunits available for an initiation at codon 26(+1). This is indeed the case for this cap construct (Fig. 8), with which synthesis of the F protein was strongly reduced. However, the synthesis of the F protein remained unchanged in the

pHCV-AUG23-LUC IRES construct, demonstrating that the initiation of translation occurs after direct positioning of the 40S subunits around codon 26(+1), as expected.

Initiation of translation with the initiator tRNA, Met-tRNA<sub>i</sub>, at a non-AUG codon in the ribosomal P site is well documented with codons that differ from AUG by only one nt (Peabody 1989; Kozak 1999; Touriol et al. 2003), such as the GUG codon identified here as the initiation site of translation of the F protein. However, to our knowledge, this is the first time that a Met-tRNA<sub>i</sub>-mediated initiation is observed with an initiator codon that differs from AUG by two nt, such as the GCG codon found in this study. An alternative possibility could be that the synthesis of the F protein results from an initiation of translation in the A site of the ribosome, without requiring the initiator Met-tRNA<sub>i</sub>. With this mode of initiation, which was demonstrated in insect viruses (Sasaki et Nakashima 2000; Wilson et al. 2000), translation is initiated by the binding of an aminoacyl-tRNA cognate to the codon in the A site, whereas the P site is occupied by a structure involving an RNA-RNA interaction between the codon in the P site and a complementary sequence in the viral RNA. In this Met-tRNA<sub>i</sub>-independent mechanism, the efficiency of the initiation of translation should be independent of the nature of the start codon. However, in our study, when codon 26(+1) was mutated to AUG, generating pHCV-AUG26+1-LUC, the synthesis of the F protein dramatically increased with the cap construct as well as with the IRES construct (Fig. 8). This suggests that the initiation of translation of the F protein does not occur in the A site, but takes place in the P site, with Met-tRNA<sub>i</sub> recognizing a non-AUG codon at position 26(+1), although less efficiently than an AUG codon. Furthermore, when the three nt upstream codon 26(+1) were mutated to ACC (pHCV-accGUG26+1-LUC), so as to create an optimal Kozak context for this codon, the synthesis of the F protein increased by four and 15-fold in the IRES and cap constructs, respectively (Fig. 8). The effect of an optimal Kozak context was also observed when codon 26(+1) is GCG (pHCV-accGCG26+1-LUC), although to a much lesser extent. Although it was observed in one case that a Kozak context could also enhance initiation at the A site (Schwab et al. 2004), this context generally enhances the initiation with Met-tRNA<sub>i</sub> at the ribosomal P site, in

cap- and IRES-mediated translation (Martinez-Salas 1999). This effect of the Kozak context on the synthesis of the F protein thus also supports a translation initiation with Met-tRNA<sub>i</sub> in the P site.

### **Effect of edeine on translation initiation of the F protein**

To confirm that translation initiation of the F protein is carried out with Met-tRNA<sub>i</sub> in the P site, we performed *in vitro* translation experiments in the presence of edeine with transcripts generated from IRES constructs. This compound is a peptide antibiotic, which interferes with initiation, by impairing AUG codon recognition by Met-tRNA<sub>i</sub> in the P site at concentrations below 1  $\mu$ M (Kozak et Shatkin 1978; Odon et al. 1978). At concentrations higher than 1  $\mu$ M, it also interferes with binding of tRNA to the A site (Szer et Kurylo-Borowska 1970; Carrasco et al. 1974). A previous report by Sarnow and his collaborators (Wilson et al. 2000) demonstrated that the presence of 0.25  $\mu$ M edeine reduced IRES-mediated initiation at the P site by 90%, while initiation of translation at the A site for the cricket paralysis virus was only inhibited by 20%. We found here that translation of the polyprotein is inhibited up to 95% when the concentration in edeine increases from 0.1 to 0.3  $\mu$ M (Fig. 9). A similar response to edeine is observed for the synthesis of the F protein, whether the 26(+1) start codon is GUG, GCG or AUG. These results fully support a translation initiation for the F protein at the P site.

## DISCUSSION

In this study, we reassessed the mechanism accounting for the synthesis of the F protein of HCV, a protein whose role is still unknown and which is translated from the viral RNA in a +1 reading frame relative to the HCV polyprotein. It was previously proposed that the F protein results from a +1 ribosomal frameshift which takes place in the beginning of the core-coding region of the polyprotein and is made by a minority of ribosomes that initiated translation at the AUG start codon of the polyprotein (Xu et al. 2001). Our results demonstrate that this conclusion was inferred from *in vitro* translation assays using an underrepresented shifty sequence, encompassing a stretch of 10A, which promotes this +1 ribosomal frameshift *in vitro*. Using the consensus HCV sequence (AAAGAAAAC) instead of this shifty sequence, we showed that the synthesis of the F protein does not result from a ribosomal frameshift. We rather found that the initiation of the synthesis of the F protein takes place at a codon overlapping codon 26 of the polyprotein coding sequence in the +1 reading frame, codon 26(+1). Codon 26(+1) is GUG in 18% of the HCV sequences available and GCG in 81% of the sequences. We found that a GUG codon at position 26(+1) is recognized as a translation start site by the ribosomes whether the initiation of translation is mediated by the HCV IRES or is cap-dependent, whereas initiation of translation at a GCG codon requires the presence of the IRES. When the initiation of translation is IRES-mediated, we showed that the 40S subunits that initiate translation at the start codon of the F protein do not reach this codon by scanning, but are directly positioned at this codon. We also observed that the synthesis of the F protein strongly increases when codon 26(+1) is mutated to AUG or when the context of this codon is mutated into an optimal Kozak context. Furthermore, at low concentrations of edeine that interfere with codon recognition by Met-tRNA<sub>i</sub> in the P site but do not impair binding of a tRNA to the A site, the synthesis of the F protein was severely inhibited. These observations are consistent with a Met-tRNA<sub>i</sub>-dependent initiation of translation at the P site for the F protein.

When inserting segments of different length of the HCV coding region before the coding sequence of the reporter gene, we found that the synthesis of the F protein requires the first 35 codons of the HCV polyprotein to reach its maximal efficiency. Under these conditions, the F protein is synthesized in cultured cells at an efficiency of about 2% relative to the synthesis of the polyprotein. Although the efficiency of translation of the F protein in the context of the complete viral genome is not known, we can assume that it is comparable to the value we observed. A low level of synthesis such as that we observe is not uncommon for viral regulatory proteins whose expression is regulated at the translational level. For example, the barley yellow dwarf virus (Barry et Miller 2002) and the yeast L-A double-stranded RNA virus (Dinman et Wickner 1992) use a  $-1$  ribosomal frameshift with an efficiency of about 1% and 2%, respectively, to produce a polymerase required for their replication.

It is known that a Met-tRNA<sub>i</sub>-dependent initiation of translation can occur at codons that differ from AUG by one nt (Peabody 1989; Kozak 1999; Touriol et al. 2003), like a GUG codon. However, the initiation of translation with Met-tRNA<sub>i</sub> at a GCG codon, that differs from AUG by two nt, is unprecedented. Contrasting with the fidelity of ribosomal decoding in the A site, information about the stringency of codon-anticodon pairing in the P site is rather limited. This stringency was analyzed thoroughly in translational bypassing, an event where ribosomes bearing a peptidyl-tRNA in the P site suspend translation at a given site, by disrupting codon-anticodon interaction, scan the messenger and resume translation downstream. Most frequently, the codon where translation resumes is perfectly complementary to the anticodon of the peptidyl-tRNA, but it was observed that this complementarity is not always perfect (reviewed in Herr et al. 2000). Depending on the local context of the codon where translation resumes, codon-anticodon complementarity can also involve three bp with a wobble bp at any position or two out of three bp with a wobble bp at one or both positions. It is interesting to observe that when initiation occurs at a GUG codon, pairing of this codon with the anticodon of Met-tRNA<sub>i</sub> (3'-UAC) implies

three bp, with a G-U wobble at the first position. With GCG, the pairing forms two out of three bp with also a G-U wobble at the first position.

We showed in this study that, when codon 26(+1) is GUG, it is selected as a translation start site whether the initiation is IRES or cap-dependent. This indicates that there must be some peculiar features in the sequence proximal to this codon, which favors translation initiation at this position. With pHCV-426-LUC, which encompasses the first 28 codons of the polyprotein sequence, the F protein was synthesized at a very low efficiency. Lengthening the HCV insertion so as to encompass the first 35 codons of the polyprotein (pHCV-447-LUC) stimulated synthesis of the F protein by about three-fold. It can be suggested that the short segment encompassing nt 427 to 447 contains a signal that enhances the selection of codon 26(+1) for initiation of translation. We can propose that this signal could base-pair to a complementary sequence within the 18S ribosomal RNA of the 40S subunit or participate in a specific secondary structure, causing in either case a distortion of the P site, which would facilitate the decoding of a non-AUG codon by the Met-tRNA<sub>i</sub>. This selection signal would be trapped in a long irregular helix in pHCV-510-LUC (see on the right of Fig. 1), accounting for the decrease in the synthesis of the F protein observed with this construct. When the complete viral genome is present, a conformational switch between alternative RNA structures could expose codon 26(+1) and its selection signal, controlling the efficiency of the initiation of the synthesis of the F protein. When codon 26(+1) is GCG, the additional presence of the HCV IRES is required for the initiation of translation of the F protein. Upon binding to the HCV IRES, the 40S subunit undergoes drastic conformational changes (Spahn et al. 2001b). We hypothesize that the simultaneous interaction with the HCV IRES and with a signal proximal to codon 26(+1) induces changes in the P site of the 40S subunit that stabilize the pairing between the anticodon of Met-tRNA<sub>i</sub> and the GCG codon, although this pairing is not perfect. Such an effect of the HCV IRES is supported by the fact that initiation of translation of the polyprotein is marginally altered when the AUG initiation codon is mutated to AUU or CUG (Reynolds et al. 1995).

Using the translational start site defined by our study, the molecular weight of the F protein can be calculated to be 14.4 kDa. As mentioned in the Introduction, Vassilaki and Mavromara (Vassilaki et Mavromara 2003), using a fusion of the beginning of the HCV coding sequence to the luciferase gene, observed the synthesis of a novel HCV protein in cultured cells. This protein of ~8 kDa resulted from initiation at an AUG codon overlapping codon 86 or 88 of the polyprotein, but in a +1 reading frame. The constructs used in the present study do not encompass the region containing these two AUG codons, but, when we lengthened the inserted segment of the HCV coding sequence so as to include these two AUGs, synthesis of luciferase in the +1 frame increased four to five-fold (data not shown). This result is consistent with the observation of Vassilaki and Mavromara that an additional protein, shorter than the F protein, could be synthesized from the +1 reading frame of the HCV polyprotein. Whether this short protein is actually synthesized in HCV-infected cells remains to be determined.

Finally, although the function of the F protein is unknown, it is well documented that viral proteins whose expression is regulated at the translational level, whether by frameshift, readthrough of a stop codon or alternative initiation, usually play an important role in viral replication (Farabaugh 1997c, 2000; Brierley et Pennel 2001; Sarnow 2003). The F protein is not required for HCV RNA replication since HCV subgenomic RNA replicons can replicate in its absence (Lohmann et al. 1999; Blight et al. 2000; Guo et al. 2001; Ikeda et al. 2002). Interestingly, initiation of the synthesis of the F protein is reminiscent of the situation encountered with the synthesis of the L\* protein in Theiler's murine encephalomyelitis virus. This picornavirus possesses a genome organization related to HCV and the L\* protein is produced by an IRES mediated initiation in the +1 reading frame of the polyprotein, 13 nt downstream of the polyprotein AUG start site (van Eyll et Michiels 2002). The L\* protein can be expressed from an AUG or an ACG codon and is essential for viral persistence (van Eyll et Michiels 2000). The similarity in the mechanisms accounting for the initiation of the synthesis of the L\* and the F protein leads us to hypothesize that the F protein could be involved in HCV persistence.



## **ACKNOWLEDGMENTS**

This study was supported by a grant from the Canadian Institutes for Health Research. Martin Baril is a recipient of a scholarship from the Fonds de Recherche en Santé du Québec (FRSQ). We are grateful to Hugo Soudeyns for the generous gift of plasmid pCore1a-33 and to Jerry Pelletier for the gift of the edeine. We thank Hany Ghonaim for his participation in this project as a summer student. We also thank Guy Boileau, Dominic Dulude, Gerardo Ferbeyre and Nikolaus Heveker for stimulating discussions and critical reading of this manuscript.

**REFERENCES**

- Bain, C., Parroche, P., Lavergne, J.P., Duverger, B., Vieux, C., Dubois, V., Komurian-Pradel, F., Trepo, C., Gebuhrer, L., Paranhos-Baccala, G., Penin, F. and Inchauspe, G. 2004. Memory T-cell-mediated immune responses specific to an alternative core protein in hepatitis C virus infection. *J. Virol.* **78**: 10460-10469.
- Barry, J.K. and Miller, W.A. 2002. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc. Natl. Acad. Sci. USA* **99**: 11133-11138.
- Blight, K.J., Kolykhalov, A.A. and Rice, C.M. 2000. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**: 1972-1974.
- Boulant, S., Becchi, M., Penin, F. and Lavergne, J.P. 2003. Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b. *J. Biol. Chem.* **278**: 45785-45792.
- Brierley, I. and Pannel, S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. *The Ribosome*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press. pp 233-248.
- Carrasco, L., Battaner, E. and Vazquez, D. 1974. The elongation steps in protein synthesis by eukaryotic ribosomes: effects of antibiotics. *Methods Enzymol.* **30**: 282-289.
- Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G. and Houghton, M. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**: 2451-2455.
- Dever, T.E. 2002. Gene-specific regulation by general translation factors. *Cell* **108**: 545-556.
- Dinman, J.D. and Wickner, R.B. 1992. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.* **66**: 3669-3676.

- 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**: 5094-5102.
- 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**: 739-740, 742-733.
- Farabaugh, P.J. 1997a. Programmed +1 frameshifting. In: Farabaugh PJ, ed. *Programmed alternative reading of the genetic code*. Austin, TX: R.G. Landes. pp 41-67.
- Farabaugh, P.J. 1997b. Programmed readthrough of translational termination codons. In: Farabaugh PJ, ed. *Programmed alternative reading of the genetic code*. Austin, TX: R.G. Landes. pp 149-182.
- Farabaugh, P.J. 2000. Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.* **64**: 131-170.
- Guo, J.T., Bichko, V.V. and Seeger, C. 2001. Effect of alpha interferon on the hepatitis C virus replicon. *J. Virol.* **75**: 8516-8523.
- Hellen, C.U. and Pestova, T.V. 1999. Translation of hepatitis C virus RNA. *J. Viral Hepat.* **6**: 79-87.
- Herr, A.J., Atkins, J.F. and Gesteland, R.F. 2000. Coupling of open reading frames by translational bypassing. *Annu. Rev. Biochem.* **69**: 343-372.
- Hershey, J.W.B., Mathews, M. and Sonenberg, N. 1996. Translational control. *Cold Spring Harbor monograph series*, 30. Plainview, N.Y.: Cold Spring Harbor Laboratory Press. pp xi, 794.
- 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**: 51-59.
- Honda, M., Beard, M.R., Ping, L.H. and Lemon, S.M. 1999. A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.* **73**: 1165-1174.

- Honda, M., Brown, E.A. and Lemon, S.M. 1996. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2**: 955-968.
- Ikeda, M., Yi, M., Li, K. and Lemon, S.M. 2002. Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. *J. Virol.* **76**: 2997-3006.
- 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.
- Kato, N. 2000. Genome of human hepatitis C virus (HCV): gene organization, sequence diversity, and variation. *Microb. Comp. Genomics* **5**: 129-151.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283-292.
- Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**: 187-208.
- Kozak, M. and Shatkin, A.J. 1978. Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. *J. Biol. Chem.* **253**: 6568-6577.
- Lefkowitz, J.H., Schiff, E.R., Davis, G.L., Perrillo, R.P., Lindsay, K., Bodenheimer, H.C., Jr., Balart, L.A., Ortego, T.J., Payne, J. and Dienstag, J.L. 1993. Pathological diagnosis of chronic hepatitis C: a multicenter comparative study with chronic hepatitis B. The Hepatitis Interventional Therapy Group. *Gastroenterology* **104**: 595-603.
- Lo, S.Y., Masiarz, F., Hwang, S.B., Lai, M.M. and Ou, J.H. 1995. Differential subcellular localization of hepatitis C virus core gene products. *Virology* **213**: 455-461.
- Lo, S.Y., Selby, M., Tong, M. and Ou, J.H. 1994. Comparative studies of the core gene products of two different hepatitis C virus isolates: two alternative forms determined by a single amino acid substitution. *Virology* **199**: 124-131.

- Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. and Bartenschlager, R. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**: 110-113.
- Martinez-Salas, E. 1999. Internal ribosome entry site biology and its use in expression vectors. *Curr. Opin. Biotechnol.* **10**: 458-464.
- Odon, O.W., Kramer, G., Henderson, A.B., Pinphanichakarn, P. and Hardesty, B. 1978. GTP hydrolysis during methionyl-tRNA<sup>f</sup> binding to 40 S ribosomal subunits and the site of edeine inhibition. *J. Biol. Chem.* **253**: 1807-1816.
- Peabody, D.S. 1989. Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**: 5031-5035.
- Pestova, T.V. and Hellen, C.U. 2001. Functions of eukaryotic factors in initiation of translation. *Cold Spring Harb. Symp. Quant. Biol.* **66**: 389-396.
- Poulin, F. and Sonenberg, N. 2003. Mechanism of translation initiation in eukaryotes. In: Lapointe J, Brakier-Gingras L, eds. *Translation mechanisms*. Georgetown, TX: Landes Bioscience. pp 280-297.
- Ray, R.B., Lagging, L.M., Meyer, K. and Ray, R. 1996. Hepatitis C virus core protein cooperates with ras and transforms primary rat embryo fibroblasts to tumorigenic phenotype. *J. Virol.* **70**: 4438-4443.
- Reed, K.E. and Rice, C.M. 2000. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. *Curr. Top. Microbiol. Immunol.* **242**: 55-84.
- Reynolds, J.E., Kaminski, A., Carroll, A.R., Clarke, B.E., Rowlands, D.J. and Jackson, R.J. 1996. Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA* **2**: 867-878.
- Reynolds, J.E., Kaminski, A., Kettinen, H.J., Grace, K., Clarke, B.E., Carroll, A.R., Rowlands, D.J. and Jackson, R.J. 1995. Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J.* **14**: 6010-6020.

- Rijnbrand, R.C., Abbink, T.E., Haasnoot, P.C., Spaan, W.J. and Bredenbeek, P.J. 1996. The influence of AUG codons in the hepatitis C virus 5' nontranslated region on translation and mapping of the translation initiation window. *Virology* **226**: 47-56.
- Rijnbrand, R.C. and Lemon, S.M. 2000. Internal ribosome entry site-mediated translation in hepatitis C virus replication. *Curr. Top. Microbiol. Immunol.* **242**: 85-116.
- Rosenberg, S. 2001. Recent advances in the molecular biology of hepatitis C virus. *J. Mol. Biol.* **313**: 451-464.
- Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q.L., Houghton, M. and Kuo, G. 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* **87**: 6547-6549.
- Sarnow, P. 2003. Viral internal ribosome entry site elements: novel ribosome-RNA complexes and roles in viral pathogenesis. *J. Virol.* **77**: 2801-2806.
- Sasaki, J. and Nakashima, N. 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl. Acad. Sci. USA* **97**: 1512-1515.
- Scheuer, P.J., Ashrafzadeh, P., Sherlock, S., Brown, D. and Dusheiko, G.M. 1992. The pathology of hepatitis C. *Hepatology* **15**: 567-571.
- Schwab, S.R., Shugart, J.A., Horng, T., Malarkannan, S. and Shastri, N. 2004. Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* **2**: 1774-1784.
- Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A. and Frank, J. 2001. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291**: 1959-1962.
- Szer, W. and Kurylo-Borowska, Z. 1970. Effect of edeine on aminoacyl-tRNA binding to ribosomes and its relationship to ribosomal binding sites. *Biochim. Biophys. Acta* **224**: 477-486.
- Tate, W.P., Poole, E.S., Horsfield, J.A., Mannering, S.A., Brown, C.M., Moffat, J.G., Dalphin, M.E., McCaughan, K.K., Major, L.L. and Wilson, D.N. 1995.

- Translational termination efficiency in both bacteria and mammals is regulated by the base following the stop codon. *Biochem. Cell Biol.* **73**: 1095-1103.
- Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C. and Vagner, S. 2003. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell.* **95**: 169-178.
- van Eyll, O. and Michiels, T. 2000. Influence of the Theiler's virus L\* protein on macrophage infection, viral persistence, and neurovirulence. *J. Virol.* **74**: 9071-9077.
- van Eyll, O. and Michiels, T. 2002. Non-AUG-initiated internal translation of the L\* protein of Theiler's virus and importance of this protein for viral persistence. *J. Virol.* **76**: 10665-10673.
- Vassilaki, N. and Mavromara, P. 2003. Two alternative translation mechanisms are responsible for the expression of the HCV ARFP/F/core+1 coding open reading frame. *J. Biol. Chem.* **278**: 40503-40513.
- Walewski, J.L., Keller, T.R., Stump, D.D. and Branch, A.D. 2001. Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**: 710-721.
- Wang, C., Sarnow, P. and Siddiqui, A. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J. Virol.* **67**: 3338-3344.
- Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F. 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.* **52**: 687-693.
- Wilson, J.E., Pestova, T.V., Hellen, C.U. and Sarnow, P. 2000. Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**: 511-520.
- Xu, Z., Choi, J., Yen, T.S., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M.J. and Ou, J. 2001. Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *EMBO J.* **20**: 3840-3848.

- Yeh, C.T., Lo, S.Y., Dai, D.I., Tang, J.H., Chu, C.M. and Liaw, Y.F. 2000. Amino acid substitutions in codons 9-11 of hepatitis C virus core protein lead to the synthesis of a short core protein product. *J. Gastroenterol. Hepatol.* **15**: 182-191.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406-3415.



Figure 4-1 : Sequence and secondary structure of the HCV IRES (based on reference Honda et al. 1999) and beginning of the HCV coding sequence (nt 1-510 of genotype 1a).

The initiation codon of the HCV polyprotein (codon 1, nt 342-344) is highlighted. The grey arrows point to the 5' end (nt 1) of the IRES and to the 5' end (nt 290) of the HCV insertion in the constructs without IRES (see Fig. 2). The black arrows point to the 3' end of segments of different length of the HCV coding sequence that were used in this study. The number preceding the arrows corresponds to the last nt of these segments. The sites where a +1 ribosomal frameshift was previously proposed to occur (see the text) are in bold and marked by arrowheads. The brackets indicate codons of the HCV polyprotein, with their corresponding number, which are referred to in the text. The codon overlapping codon 26 in the +1 reading frame, to which we also refer in the text, is in bold. It was verified, using the *mfold* program (Zuker 2003), that the predicted secondary structure of the HCV sequence was not altered when it was fused to the luciferase coding sequence.



Figure 4-2 : Description of the luciferase vectors used for the study of the expression of the HCV F protein *in vitro* and in cultured cells.

Nt are numbered according to Figure 1.

A) The construct presented in this figure is pHCV-447-LUC where a portion of the HCV coding sequence extending to nt 447 is inserted upstream the coding sequence of the firefly luciferase reporter gene. Inserts of different length are: pHCV-387-LUC, pHCV-426-LUC, pHCV-447-LUC and pHCV-510-LUC, where the number indicates the last nt of the insertion (see Fig. 1). The AUG initiation codon (nt 342-344) of the HCV polyprotein is underlined. For the (0) constructs, the luciferase sequence is in-frame with this AUG initiation codon. For these (0) constructs, the expression of luciferase mimicks the synthesis of the polyprotein. For the (+1) constructs, an adenine was added immediately after the *Bam*HI restriction site (at position 455), so that only ribosomes translating the HCV sequence in the +1 reading frame relative to the HCV initiator AUG synthesize the luciferase. With these (+1) constructs, the expression of luciferase mimicks the synthesis of the F protein. All the constructs were cloned by inserting a PCR product containing the investigated HCV sequence between the *Kpn*I-*Bam*HI sites of the pcDNA3.1-LUC vector. This HCV sequence was obtained from a plasmid that contains the 5'UTR plus the beginning of the coding sequence of HCV genotype 1a (see Materials and Methods). In all the pHCV-LUC (0) and (+1) constructs, the HCV coding sequence is preceded either by the complete HCV 5'UTR (IRES constructs) or by a small segment (nt 290-341) of the HCV 5'UTR region (cap constructs).

The boxed sequence, which corresponds to a HCV consensus sequence, was mutated in pHCV-387-LUC, pHCV-426-LUC, pHCV-447-LUC and pHCV-510-LUC, creating derivatives with a stretch of 10A (10A constructs).

B) Sequences of the HCV coding region of the derivatives of pHCV-447-LUC constructed in this study. Nucleotides that are mutated are underlined, nt that are inserted and deleted are, respectively, in bold and indicated by a dashed line.



Figure 4-3 : Synthesis of the F protein with constructs containing insertions of different length of the HCV coding sequence.

The synthesis of the F protein was measured with constructs containing HCV segments of different length described in Figure 2. In the pHCV-387-LUC, pHCV-426-LUC, pHCV-447-LUC and pHCV-510-LUC cap and IRES constructs, the consensus HCV sequence (AAAGAAAAC, nt 364-373) occupies the site where a +1 frameshift was proposed to occur. In the 10A corresponding constructs, the consensus HCV sequence is replaced with a 10A stretch.

- A) Synthesis of the F protein in cultured cells. Synthesis of the F protein was measured after co-transfection of 293FT cells with 3  $\mu$ g of a pHCV-LUC (0) or (+1) construct and 1  $\mu$ g of pcDNA3.1/Hygro(+)/lacZ, which is used to normalize for variations in transfection efficiency.
- B) Synthesis of the F protein *in vitro*. *In vitro* translation experiments were carried out in 25  $\mu$ l of RRL with 0.2  $\mu$ g of mRNAs transcribed from the *Fba*I-digested pHCV-LUC constructs. Results are reported as the amount of the F protein synthesized (assessed by the activity of luciferase in the +1 reading frame) relative to the amount of the polyprotein synthesized (assessed by the activity of luciferase in the 0 reading frame) and were calculated as described in the text. Each value represents the mean  $\pm$  standard error of four to six independent experiments.

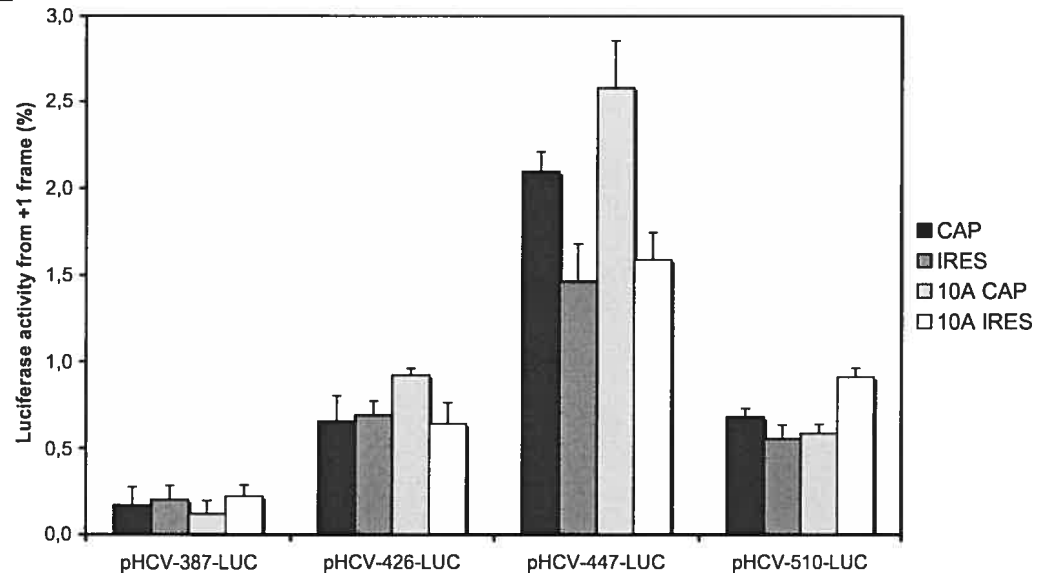
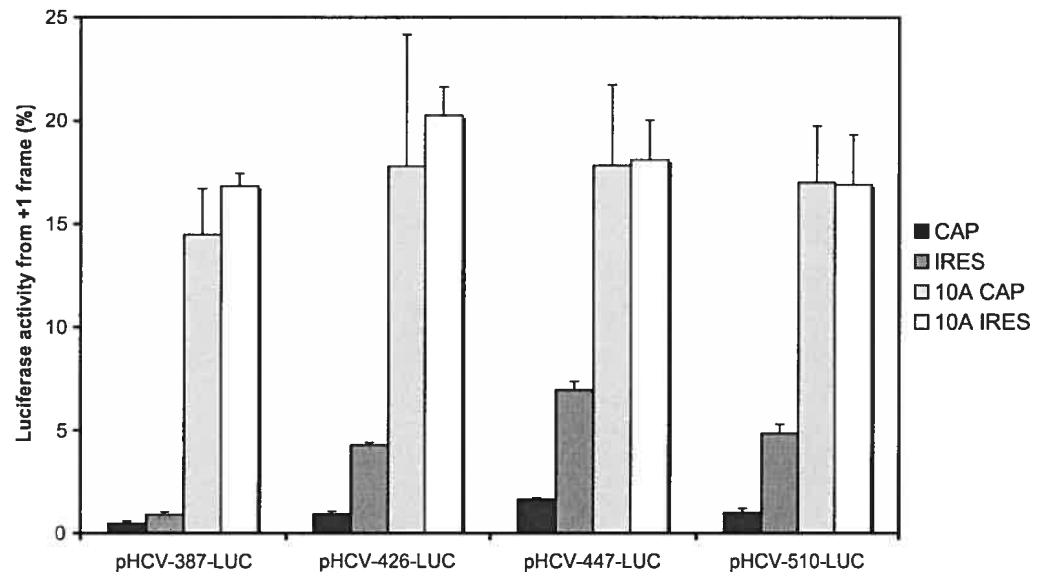
**A****B**

Figure 4-4 : Synthesis of the F protein when a stop codon is introduced in the reading frame of the polyprotein.

A UGA stop codon was introduced in pHCV-447-LUC, in the reading frame of the polyprotein, upstream or across the previously proposed +1 ribosomal frameshift site (Xu et al. 2001), at codon 3, 9, 10, 11 or 12. This generated pHCV-UGA3-LUC, pHCV-UGA9-LUC, pHCV-UGA10-LUC, pHCV-UGA11-LUC and pHCV-UGA12-LUC. All these constructs contained or not the HCV IRES (IRES and cap construct, respectively), except pHCV-UGA3-LUC, for which the HCV IRES was not constructed since mutating nt 348-350 of the HCV RNA destabilizes a hairpin structure that is important for the IRES function (Honda et al. 1996). Assays were made in cultured cells, as described in the legend to Figure 3. The synthesis of the F protein in the mutant constructs is expressed relative to that obtained with pHCV-447-LUC (+1), which is arbitrarily set at 100%. Each value represents the mean  $\pm$  standard error of four to six independent experiments.



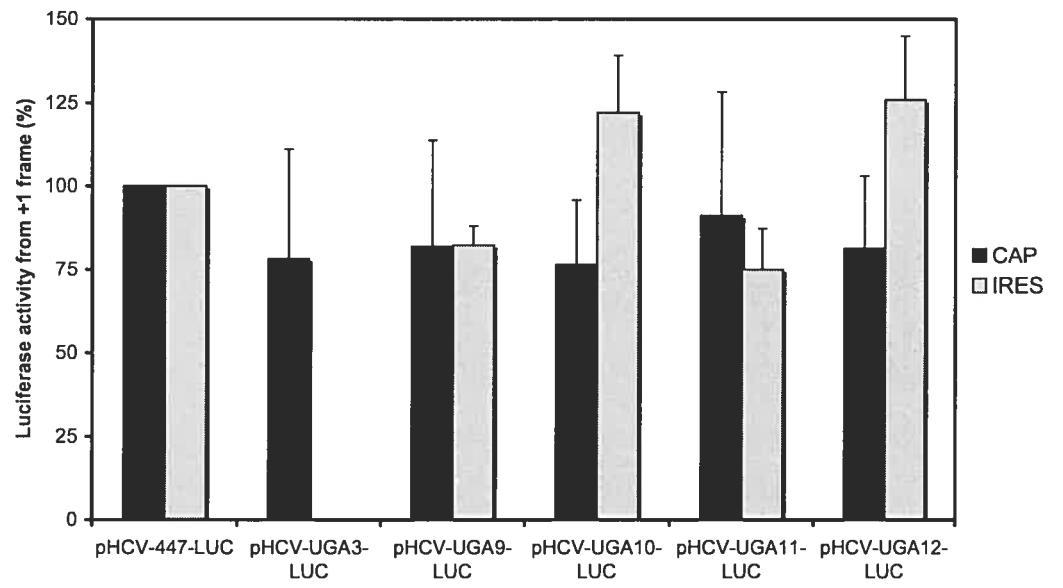


Figure 4-5 : Synthesis of the polyprotein and the F protein when the AUG start codon of the polyprotein or its context is mutated.

Derivatives of pHCV-447-LUC were constructed by mutating the AUG (nt 342-344) polyprotein start codon to UAC or its upstream context from ACC to UUU, generating pHCV-accUAC-LUC and pHCV-uuuAUG-LUC, respectively. Translation efficiencies are indicated on a logarithmic scale for the polyprotein and for the F protein. The synthesis of the polyprotein measured with the mutant (0) constructs and that of the F protein measured with the (+1) constructs are expressed relative to the synthesis of the polyprotein with pHCV-447-LUC (0), which is arbitrarily set at 100%. Assays were made in cultured cells, as described in the legend to Figure 3. Each value represents the mean  $\pm$  standard error of four to six independent experiments.

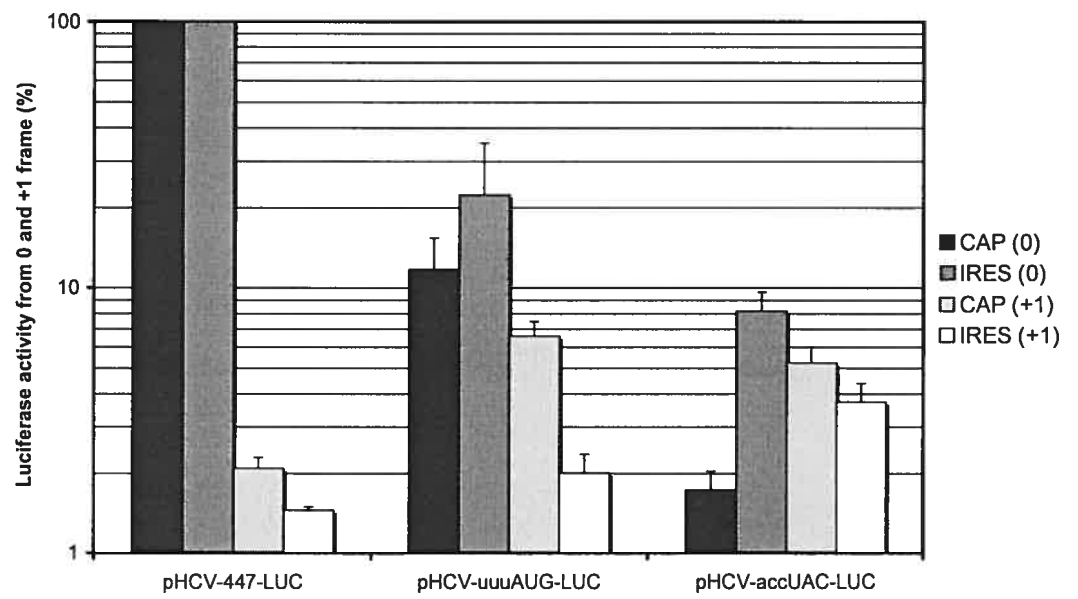


Figure 4-6 : Identification of the translation start site of the F protein.

A stop codon was introduced in pHCV-447-LUC, at codons overlapping codon 23, 24, 25 or 26 of the polyprotein, in the +1 reading frame, generating pHCV-UAA23+1-LUC, pHCV-UGA24+1-LUC, pHCV-UGA25+1-LUC and pHCV-UGA26+1-LUC. Assays were made in cultured cells, as described in the legend to Figure 3. The efficiency of the synthesis of the F protein from construct pHCV-447-LUC (+1) is arbitrarily set at 100%. Each value represents the mean  $\pm$  standard error of four to six independent experiments.

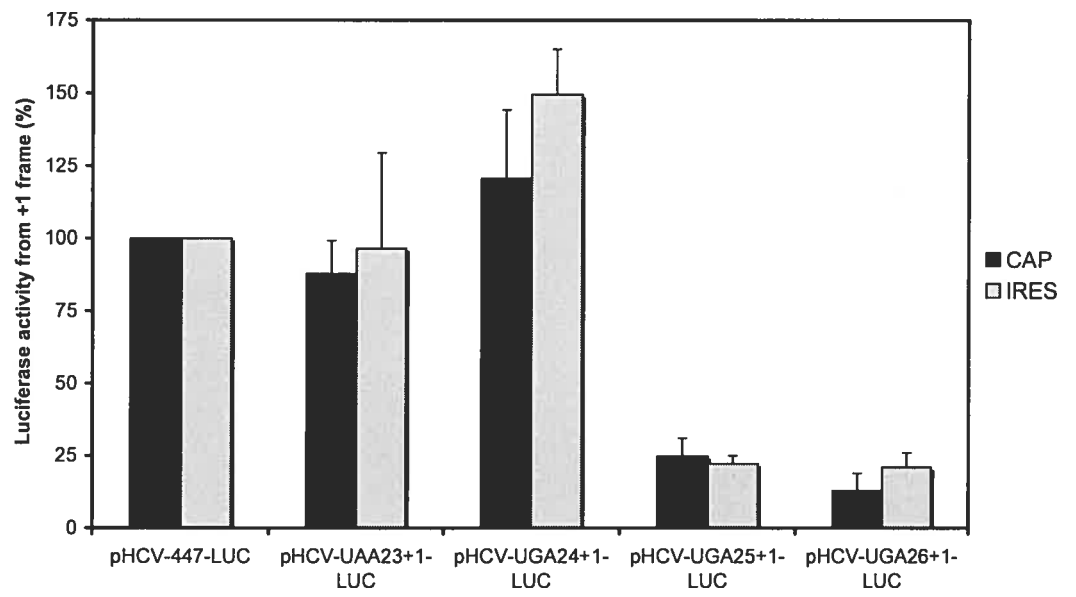


Figure 4-7 : Confirmation of the location of the translation start site of the F protein.

- A) Description of the mutants used for this confirmation. A nt was added in pHCV-447-LUC (+1) immediately after codon 25(+1) or 26(+1), generating, respectively, pHCV+U418-LUC and pHCV+U421-LUC, where the reading frame of luciferase, which monitor the synthesis of the F protein, is shifted by one nt relative to its reading frame in pHCV-447-LUC (+1). Nt 447 was deleted (dashed line) in pHCV+U421-LUC, generating pHCV+U421 $\Delta$ U447-LUC, in which the luciferase reading frame is restored.
- B) Assays were made in cultured cells, as described in the legend to Figure 3. The efficiency of synthesis of the F protein from construct pHCV-447-LUC (+1) is arbitrarily set at 100%. Each value represents the mean  $\pm$  standard error of four to six independent experiments.

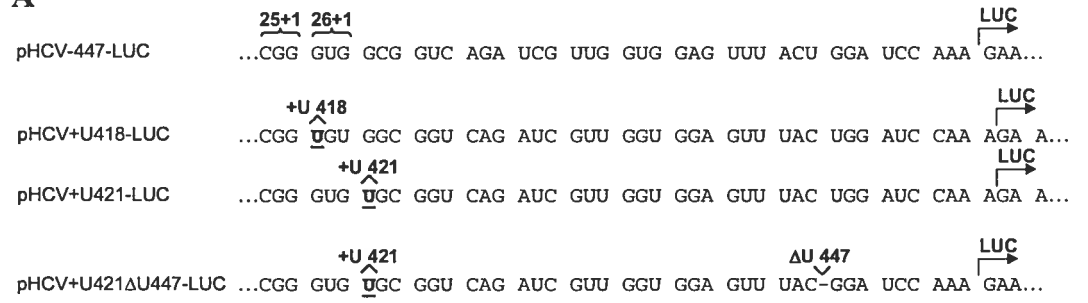
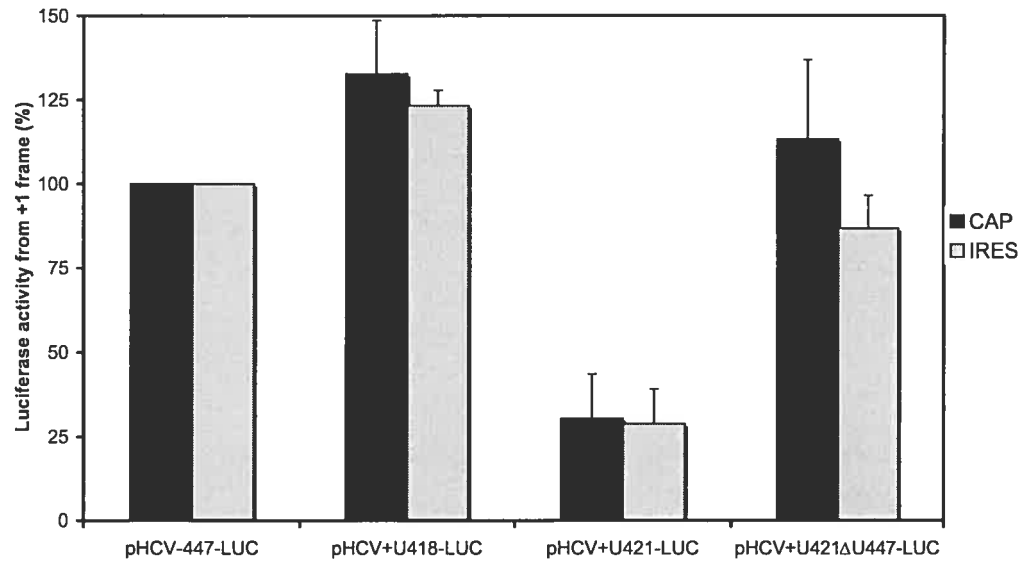
**A****B**

Figure 4-8 : Further details on the initiation of translation of the F protein.

In pHCV-GCG26+1-LUC, the initiation codon of the F protein, which is GUG in pHCV-447-LUC, was mutated to GCG. In pHCV-AUG23-LUC, an AUG codon with an optimal Kozak context for initiation was introduced in the 0 frame at codon 23 of pHCV-447-LUC. In pHCV-AUG26+1-LUC, the GUG initiation codon of the F protein was mutated to AUG. In pHCV-accGUG26+1-LUC and pHCV-accGCG26+1-LUC, the Kozak context of codon 26(+1) was mutated to ACC, the initiator codon being, respectively, GUG and GCG. The synthesis of the F protein is indicated on a logarithmic scale. Assays were made in cultured cells, as described in the legend to Figure 3. The synthesis of the F protein from pHCV-447-LUC (+1) is arbitrarily set at 100%. Each value represents the mean  $\pm$  standard error of four to six independent experiments.



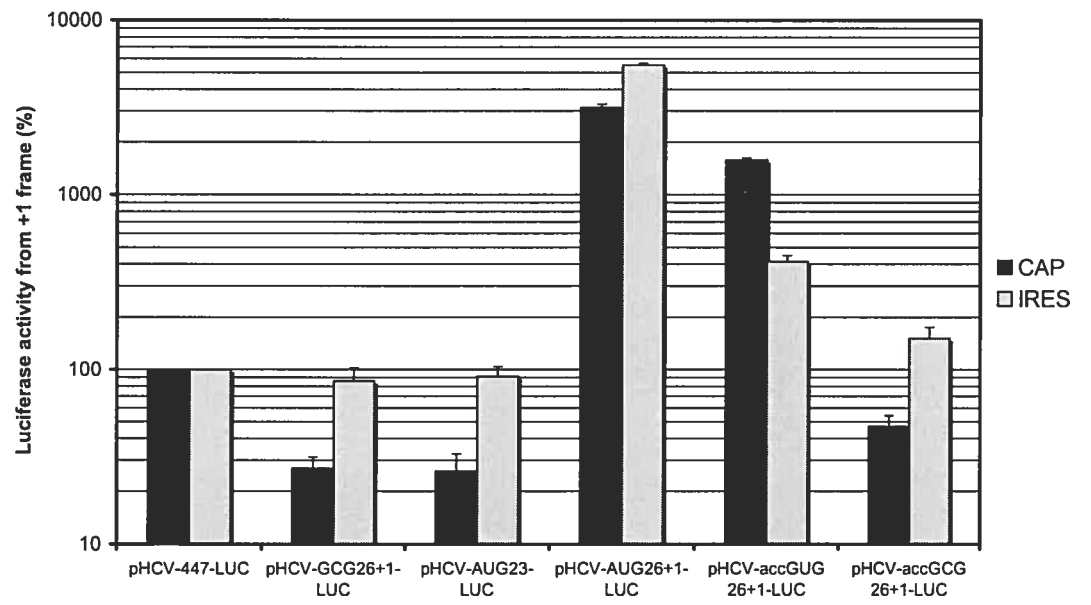
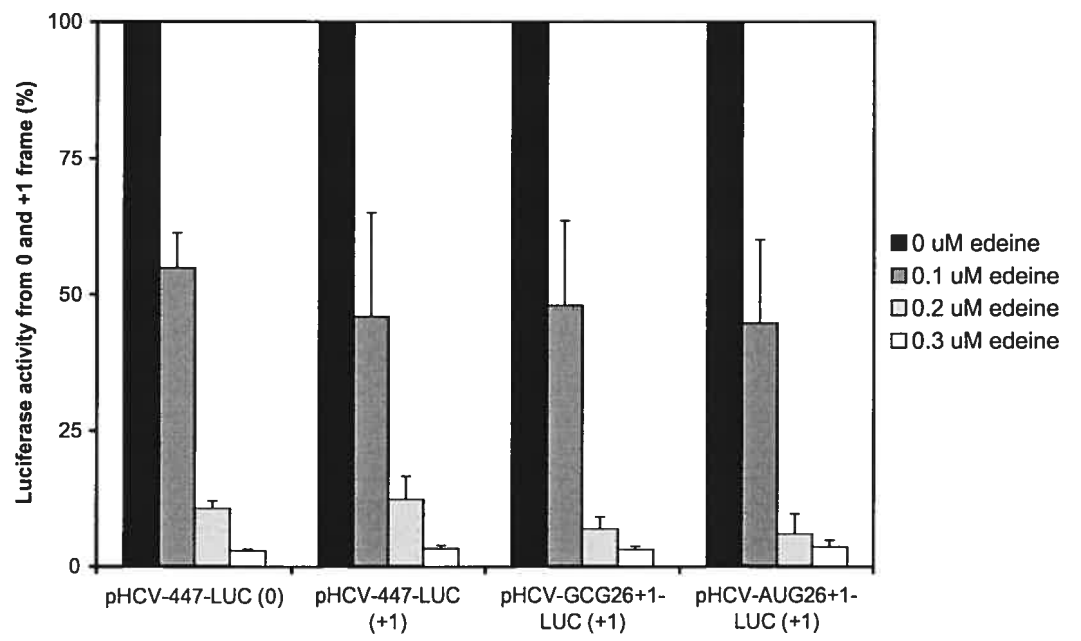


Figure 4-9 : Inhibition of the synthesis of the polyprotein and the F protein by edeine.

*In vitro* translation experiments were carried out in 25  $\mu$ l of RRL, containing 0, 0.1, 0.2 or 0.3  $\mu$ M edeine, with 0.2  $\mu$ g of mRNAs transcribed from the *FbaI*-digested pHCV-LUC IRES constructs. For each construct, the protein synthesis in the absence of edeine (0  $\mu$ M) is arbitrarily set at 100%. Each value represents the mean  $\pm$  standard error of three independent experiments.



## **Chapitre V**

## **Discussion**

## 5.1 Efficacité du *frameshift* -1 chez les différents sous-types du groupe M du VIH-1

Le *frameshift* -1 du VIH-1 permet la synthèse de la protéine Gag-Pol, le précurseur des enzymes du virus. Un article publié précédemment par notre groupe avait démontré que le signal stimulateur du *frameshift* -1 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 en Amérique et en Europe, est une tige-boucle irrégulière où deux tiges sont séparées par une protubérance de trois purines (Dulude et al. 2002, voir section 1.4.4). Dans cette structure, la tige-boucle supérieure correspond à la tige-boucle classique, qui était auparavant considérée comme le signal stimulateur, et la tige inférieure résulte d'un pairage entre l'espaceur, qui sépare la séquence glissante et la tige-boucle supérieure, et une séquence complémentaire en aval de cette tige-boucle (Figure 1-19B).

La première partie de mon projet de recherche a permis de démontrer que le signal stimulateur des autres sous-types du groupe M (A, C, D, F, G, H, J et K) peut adopter la structure en tige-boucle irrégulière caractérisée pour le sous-type B, malgré les variations dans sa séquence primaire (Figure 2-3). La stabilité thermodynamique de chacun de ces signaux stimulateurs, prédite à l'aide du programme *mfold* (Zuker 2003), varie cependant de -17.2 kcal/mole pour le sous-type F à -26.2 kcal/mole pour le sous-type K. Nos résultats obtenus après transfection de vecteurs permettant de mesurer l'efficacité de *frameshift* -1 dans des cellules HEK 293T en culture ont toutefois démontré que cette efficacité ne varie pas de façon parallèle à la stabilité du signal stimulateur et qu'elle a une valeur similaire pour tous les sous-types du groupe M, la variation maximale étant d'environ 35% par rapport à la valeur moyenne de l'efficacité de *frameshift* -1 de tous les sous-types. Dans un système de traduction *in vitro*, par contre, l'efficacité de *frameshift* -1 varie en fonction de la stabilité de la tige-boucle supérieure. Ces résultats *in vitro* s'expliquent par l'étude précédente faite avec le sous-type B qui avait montré que la tige inférieure de la tige-boucle

irrégulière influence l'efficacité de *frameshift* -1 dans des cellules en culture, mais pas dans un système de traduction *in vitro* (Dulude et al. 2002). Ces résultats indiquent qu'il faut être prudent dans l'interprétation des résultats obtenus *in vitro*, puisque des différences dans la cinétique de traduction peuvent produire des résultats biaisés. Le maintien d'une efficacité de *frameshift* -1 semblable dans les différents sous-types du groupe M dans des cellules en culture résulterait donc de la présence de la tige inférieure. Nos résultats suggèrent aussi que la protubérance de trois purines peut interagir spécifiquement avec le ribosome, car la mutation des trois purines en pyrimidines entraîne une diminution importante de l'efficacité de *frameshift* -1. De plus, un variant naturel du sous-type B, où une délétion de trois nucléotides empêche la formation de cette protubérance, possède une efficacité de *frameshift* -1 réduite qui rendrait vraisemblablement ce variant non-infectieux.

L'efficacité de *frameshift* -1 contrôle le rapport Gag-Pol/Gag, où Gag est le précurseur des protéines de structure du virus (voir section 1.4.4). La conservation de la structure du signal stimulateur et de l'efficacité de *frameshift* -1 dans les différents sous-types du groupe M confirme l'importance du rapport Gag-Pol/Gag. Nous pouvons donc affirmer que la variation génétique ayant mené à la formation des différents sous-types du groupe M a subi une pression sélective afin de maintenir une efficacité de *frameshift* -1 dont la valeur ne peut varier qu'à l'intérieur d'une fenêtre étroite. La conservation de la structure du signal stimulateur pourrait également permettre la sélection d'agents qui moduleraient l'efficacité de *frameshift* -1 en liant le signal stimulateur du *frameshift* -1 et qui pourraient agir sur les différents sous-types du groupe M du VIH-1. Pour diminuer l'infectivité du VIH-1, ces agents anti-*frameshift* devraient au minimum déplacer l'efficacité de *frameshift* -1 en dehors de la fenêtre de variabilité de 35% observée pour les différents sous-types du groupe M. Plus spécifiquement, des études précédentes ont observé une diminution de l'infectivité du VIH-1 de 10 à 100 fois lorsque le rapport Gag-Pol/Gag était augmenté (Shehu-Xhilaga et al. 2001) ou diminué (Telenti et al. 2002) d'un facteur trois. Les résultats de mon premier article (Baril et al. 2003a) (voir Chapitre II) confirment

donc le *frameshift* -1 comme cible potentielle pour le développement de nouveaux agents anti-VIH.

Jusqu'à maintenant, plusieurs groupes ont tenté de développer des agents anti-*frameshift*. Certaines de ces études ont montré que des oligonucléotides, modifiés pour résister aux attaques des nucléases, qui s'hybrident en aval de la séquence glissante de la région de *frameshift* -1 pouvaient stimuler l'efficacité de *frameshift* -1 du VIH-1 (Howard M. T. et al. 2004; Olsthoorn et al. 2004). Cependant, ces études ont été réalisées en absence du signal stimulateur de *frameshift* -1 et dans des systèmes de traduction *in vitro*, mais non dans des cellules en culture. Or, une étude réalisée il y a plusieurs années avait montré que des oligonucléotides modifiés s'hybridant en aval de la tige-boucle classique du VIH-1 stimulaient l'efficacité de *frameshift* -1 *in vitro*, mais pas dans des cellules en culture (Vickers et Ecker 1992). D'autres études ont montré que certains inhibiteurs de la traduction peuvent agir comme agents anti-*frameshift*. Dans ces études effectuées avec le virus L-A de la levure, l'anisomycine et la preussine, des inhibiteurs affectant l'étape de l'accommodation de l'aminoacyl-ARNt au site A du ribosome, réduisaient l'efficacité de *frameshift* -1, alors que la sparsomycine, un inhibiteur de l'activité peptidyl transférase du ribosome, augmentait l'efficacité de *frameshift* -1 (Dinman et al. 1997; Goss Kinzy et al. 2002). Plusieurs agents permettant d'augmenter l'efficacité de *frameshift* -1 du VIH-1 et de d'autres virus ont également été identifiés après une sélection à haut débit de plus de 56 000 molécules chimiques (Hung et al. 1998). La toxicité de ces agents chimiques, qui agissent probablement sur le ribosome, empêche toutefois leur utilisation thérapeutique. Notre groupe de recherche avait aussi montré que des oligonucléotides modifiés s'hybridant en un site spécifique de l'ARNr 18S pouvaient diminuer l'efficacité de *frameshift* -1 du VIH-1 *in vitro* et dans des cellules en culture, en plus d'altérer la réplication virale (Brakier-Gingras et al. 1998). Cependant, les traitements ciblant le ribosome entraîneraient probablement des effets secondaires importants, puisque le ribosome joue un rôle primordial dans le fonctionnement cellulaire.

En ce moment, notre équipe développe un système de sélection de peptides liant spécifiquement la tige-boucle irrégulière qui stimule le *frameshift* -1 du VIH-1 (Dulude et Brakier-Gingras, en préparation). Cette sélection se fait à partir d'une librairie de peptides qui est triée par cytométrie de flux. Afin de faciliter et d'accélérer le tri de plusieurs millions de peptides, ce système utilise un vecteur d'expression procaryote plutôt qu'eucaryote pour mesurer l'efficacité de *frameshift* -1. L'utilisation de bactéries a été validée par une étude antérieure de notre équipe de recherche qui a montré que le *frameshift* -1 du VIH-1 répond de la même façon au signal stimulateur du sous-type B du groupe M chez les eucaryotes et chez les procaryotes (Léger et al. 2004). Dans ce système de sélection, la protéine fluorescente rouge (RFP) est synthétisée par traduction conventionnelle, alors que la protéine fluorescente verte (GFP) est exprimée après *frameshift* -1 dans la région de *frameshift* étudiée. Ce système permet d'identifier des peptides qui affectent significativement le *frameshift* -1 (GFP), sans affecter la traduction conventionnelle (RFP). Ces peptides pourraient diminuer l'efficacité de *frameshift* -1 en masquant des sites d'interactions entre le signal stimulateur et le ribosome, comme potentiellement la protubérance de trois purines, ou encore augmenter l'efficacité de *frameshift* -1 en stabilisant la structure du signal stimulateur. La preuve de concept de cette approche a été fournie par une étude antérieure où le signal stimulateur du VIH-1 avait été remplacé par l'élément de réponse aux ions fer (IRE), une tige-boucle présente sur l'ARNm de la ferritine, qui est lié par des protéines de contrôle de la concentration en ions fer (IRP) lors de carence en fer (Kollmus et al. 1996). Dans cette étude, l'efficacité de *frameshift* -1 augmentait d'environ trois fois lors de la liaison de l'IRP à l'IRE.

La méthode de sélection de peptides développée par notre équipe utilise la région de *frameshift* -1 du sous-type B du groupe M. L'efficacité de ces peptides contre les autres sous-types du groupe M sera ensuite évaluée. La conservation de la structure en tige-boucle irrégulière, notamment la protubérance de purines, permet de suggérer que ces peptides pourront affecter l'efficacité de *frameshift* -1 de l'ensemble des sous-types du groupe M. Si



ce n'est pas le cas, des sélections indépendantes de peptides anti-*frameshift* pour chacun des sous-types pourraient être nécessaires.

## **5.2 Le signal stimulateur du *frameshift* -1 du groupe O du VIH-1 est un pseudonoëud**

La deuxième partie de mes travaux de recherche a porté sur la caractérisation de la région de *frameshift* -1 du groupe O du VIH-1. L'examen des séquences disponibles pour le génome du groupe O a révélé que son signal stimulateur de *frameshift* -1 ne pouvait pas former la tige-boucle irrégulière retrouvée chez le groupe M, puisqu'il n'y a pas d'appariement possible entre l'espaceur, séparant la séquence glissante de la tige-boucle, et une séquence en aval de la tige-boucle. Par des études de mutagénèse et d'attaques enzymatiques contre l'ARN correspondant à la région de *frameshift* -1 du groupe O, nous avons caractérisé une structure en pseudonoëud pour le signal stimulateur du *frameshift* -1 du groupe O. Dans cette structure, la boucle coiffant une tige de huit paires de bases forme une deuxième tige de huit paires de bases avec une séquence complémentaire en aval (voir Figure 3-1a) (Baril et al. 2003b).

Comme mentionné à la section 1.4.1, cinq sous-types sont présentement proposés pour le groupe O (Yamaguchi et al. 2002). Malheureusement, l'étude du *frameshift* -1 du groupe O est limitée à deux sous-types, puisque sur les huit séquences disponibles qui couvrent la région de *frameshift* -1, une appartient au sous-type ANT70, le premier membre du groupe O isolé (De Leys et al. 1990), et sept appartiennent à un second sous-type, MVP5180, qui semble être le sous-type du groupe O le plus répandu. Nous avons observé que le signal stimulateur de tous les membres du sous-type MVP5180 peut former la structure en pseudonoëud. Par contre, des variations dans la séquence en aval de la tige-boucle dans le sous-type ANT70, par rapport au sous-type MVP5180, empêchent la formation d'un pseudonoëud. Cette incapacité à former le pseudonoëud chez le sous-type ANT70 se reflète dans l'efficacité de son *frameshift* -1 qui est deux fois plus faible que

celle du sous-type MVP5180 dans des cellules en culture. Il est toutefois possible que la seule séquence disponible pour la région de *frameshift* -1 du sous-type ANT70 ne soit pas représentative des autres membres de ce sous-type. Une autre possibilité est que le pseudonoeud soit formé avec une séquence plus éloignée en aval de la tige-boucle, puisque la distance entre la tige-boucle et la séquence en aval est variable dans les pseudonoeuds stimulant le *frameshift* -1, atteignant par exemple 4000 nucléotides chez le BYDV (Barry et Miller 2002). Le séquençage de d'autres régions de *frameshift* -1 du groupe O est nécessaire afin de déterminer si le signal stimulateur du groupe O adopte une conformation en pseudonoeud pour l'ensemble des sous-types.

En plus de la divergence au niveau structural entre le signal stimulateur des groupes M et O, il y a également une différence importante entre leur efficacité de *frameshift* -1. En effet, le pseudonoeud du groupe O permet une efficacité de *frameshift* -1 deux fois plus élevée que la tige-boucle irrégulière du groupe M. Cette différence dans l'efficacité de *frameshift* -1 entre les groupes M et O est significativement supérieure à celle observée entre les différents sous-types du groupe M (voir section 5.1) et répond vraisemblablement à un besoin spécifique au groupe O. La réplication du groupe O pourrait nécessiter la production de plus d'enzymes virales que le groupe M. Un besoin accru en protéase virale pourrait par exemple s'expliquer par une activité plus faible de cette protéase chez le groupe O comparativement au groupe M. Cette activité réduite pourrait être compensée par une synthèse et une incorporation accrues de protéase dans les virions, afin de maintenir une cinétique adéquate de maturation du virus. Aucune étude sur l'activité enzymatique de la protéase du groupe O n'a été effectuée jusqu'à maintenant, mais sa séquence renferme plusieurs différences comparativement à la protéase du groupe M et certaines de ces différences correspondent à des mutations secondaires qui favorisent l'apparition de résistance aux PI chez le groupe M. Ces mutations secondaires diminuent l'activité enzymatique de la protéase du groupe M et pourraient entraîner un effet similaire sur la protéase du groupe O (de Baar et al. 2000; Yang et al. 2000; Luk et al. 2001). La nécessité pour le groupe O de produire plus de Gag-Pol pourrait également être due à des variations

de séquence pouvant affecter l'activité des peptides codés dans la région de *frameshift* -1 du VIH-1. Ces peptides, p1 et p6 dans le cadre de lecture de Gag et TFP et p6\* dans le cadre de lecture de Pol, jouent aussi un rôle important dans la formation de particules virales infectieuses (Louis et al. 1999; Demirov et al. 2002; Hill et al. 2002). Une caractérisation détaillée de l'activité des enzymes virales du groupe O et une comparaison avec les données disponibles pour les enzymes virales du groupe M seront essentielles pour expliquer la nécessité du rapport Gag-Pol/Gag plus élevé chez le groupe O.

En plus de caractériser la structure du pseudonoeud du groupe O par mutagénèse et par attaques enzymatiques contre l'ARN, une collaboration avec le Dr S. Steinberg de notre département nous a également permis de modéliser la structure de ce pseudonoeud *in silico*. Cette modélisation a montré que la séquence AACAA dans la boucle 2 du pseudonoeud pouvait potentiellement former un triplex avec la tige 1 (voir Figure 3-6). Dans ce triplex, le résidu C formerait une protubérance qui serait disponible pour une interaction avec le ribosome. La formation d'un triplex similaire a été observée chez SRV-1 (Michiels et al. 2001) et chez BWYV (Su et al. 1999), où le triplex a été proposé comme étant la caractéristique principale qui permet la stimulation de l'efficacité du *frameshift* -1 (voir section 1.3.1.1.2). Cette stimulation proviendrait à la fois d'une pause accrue du ribosome au niveau du site de glissement, en raison de la difficulté pour l'activité hélicase du ribosome de défaire le triplex, et d'une interaction spécifique entre ce triplex et le ribosome qui affecterait son activité de décodage par un mécanisme encore inconnu.

Pour le groupe O du VIH-1, nous avons proposé que la formation du triplex est favorisée par le repliement d'une partie de la boucle 2 pour former une troisième tige-boucle de trois paires de bases (Figure 5-1A). Cette troisième tige-boucle permet de réduire la longueur de la boucle 2 du pseudonoeud de 20 à huit nucléotides, ce qui assurerait un positionnement adéquat des nucléotides de la boucle 2 pour former le triplex. Notre article était le premier à suggérer la formation d'une troisième tige-boucle dans les pseudonoeuds qui stimulent le *frameshift* -1. Récemment, des études sur le pseudonoeud qui stimule

l'efficacité de *frameshift* -1 chez le virus du SRAS ont caractérisé la formation d'une troisième tige-boucle à partir des nucléotides de la boucle 2 du pseudonoeud (Ramos et al. 2004; Baranov et al. 2005). Dans le cas du SRAS, cette troisième tige-boucle est formée de 10 paires de bases interrompues par deux protubérances contenant chacune une adénine, laissant six des 32 nucléotides de la boucle 2 disponibles pour la formation potentielle d'un triplex (Figure 5-1B). Des études de mutagenèse ont toutefois démontré que la délétion de cette troisième tige-boucle n'affecte pas l'efficacité de *frameshift* -1. Par contre, certaines mutations, notamment le remplacement des adénines des deux protubérances par des pyrimidines, abolissaient totalement le *frameshift* -1 (Ramos et al. 2004). Ces résultats suggèrent que la troisième tige-boucle n'est pas requise pour une stimulation maximale du *frameshift* -1, mais que des mutations qui affectent sa structure pourraient interférer avec le repliement adéquat du pseudonoeud. Ces observations appuient notre hypothèse qui suggère que le rôle d'une troisième tige-boucle serait de minimiser la longueur de la boucle 2 afin de permettre la formation plus efficace d'un triplex avec la tige 1 du pseudonoeud. Il serait intéressant d'effectuer des études de mutagenèse semblables sur la troisième tige-boucle du pseudonoeud stimulant le *frameshift* -1 du groupe O.

Les résultats de ce deuxième article (voir Chapitre III) mettent en évidence la divergence de structure observée entre les signaux stimulateurs de *frameshift* -1 des groupes M et O du VIH-1. Cette divergence au niveau structural appuie les études phylogénétiques qui suggèrent que l'apparition des groupes M et O proviennent de transmissions indépendantes du chimpanzé à l'homme (Gao et al. 1999). L'utilisation d'un pseudonoeud comme signal stimulateur du *frameshift* -1 chez le groupe O suggère également que les agents anti-*frameshift* développés pour la tige-boucle irrégulière du groupe M ne seront pas efficaces contre le pseudonoeud du groupe O. Notre système de sélection procaryote détaillé à la section 5.1, devra donc être adapté afin de permettre la sélection de peptides spécifiques au pseudonoeud du groupe O. Les signaux stimulateurs de *frameshift* -1 chez les bactéries étant généralement des tiges-boucles et non des pseudonoeuds, il faudra également vérifier que le *frameshift* -1 du groupe O du VIH-1

répond de la même façon au pseudonoeud chez les procaryotes et chez les eucaryotes. Le développement de nouveaux agents anti-VIH est particulièrement important pour le groupe O en raison de sa résistance intrinsèque aux NNRTI (Descamps et al. 1997) et de variations dans sa séquence, correspondant à des mutations secondaires dans la protéase du groupe M, qui peuvent mener à une résistance rapide aux PI (de Baar et al. 2000).

Nous n'avons pas effectué d'étude sur le signal stimulateur du troisième groupe du VIH-1, le groupe N, qui est peu répandu et pour lequel seulement deux séquences couvrant la région de *frameshift* -1 sont disponibles. Nous avons toutefois observé que son signal stimulateur peut adopter la structure en tige-boucle irrégulière retrouvée chez le groupe M. Cette observation n'est pas surprenante, puisque le groupe N provient d'un événement de recombinaison entre les gènes *gag* et *pol* du groupe M du VIH-1 et le gène *env* du SIV<sub>cpz</sub> (Simon et al. 1998; Corbet et al. 2000).

Figure 5-1 : Structure secondaire de signaux stimulateurs de *frameshift* -1 formant un pseudonoeud et une troisième tige-boucle.

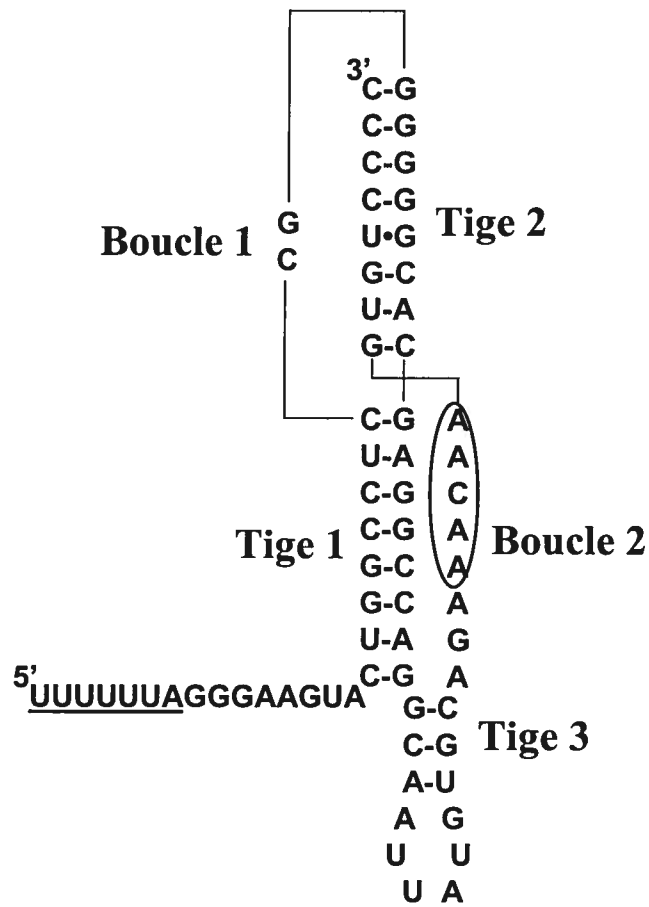
A) Structure secondaire du pseudonoeud stimulant le *frameshift* -1 chez le groupe O du VIH-1 (adapté de Baril et al. 2003b).

La séquence glissante est soulignée et la séquence AACAA qui peut participer à la formation d'un triplex avec la tige 1 est encerclée. Les tiges 1, 2 et 3 et les boucles 1 et 2 sont identifiées.

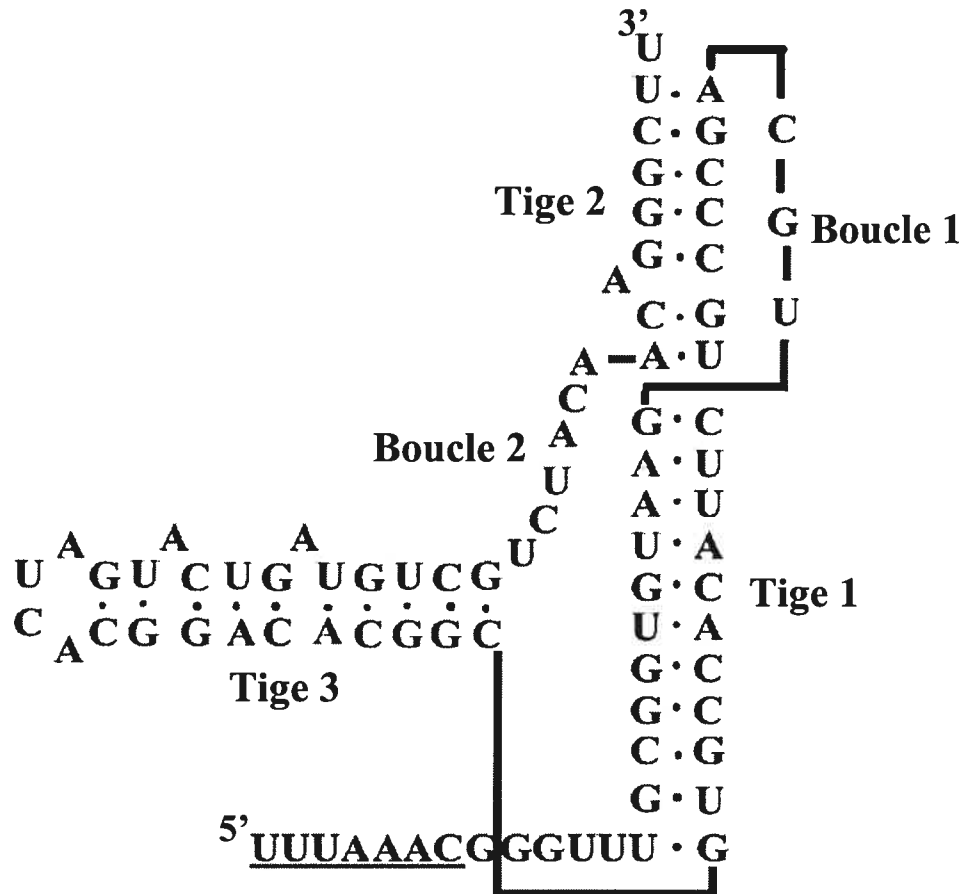
B) Structure secondaire du pseudonoeud stimulant le *frameshift* -1 chez le SRAS (basé sur Ramos et al. 2004; Baranov et al. 2005).

La séquence glissante est soulignée. Les tiges 1, 2 et 3 et les boucles 1 et 2 sont identifiées.

**A**



**B**



### 5.3 La protéine F du VHC est synthétisée après initiation de la traduction à un codon non-AUG dans un cadre de lecture +1 par rapport à la polyprotéine virale

Dans la troisième partie de mes travaux de recherche, j'ai étudié le mécanisme menant à la synthèse d'une nouvelle protéine du VHC, la protéine F, dont l'existence a récemment été découverte. La première étude qui a caractérisé l'expression de la protéine F avait montré, par séquençage N-terminal de la protéine F produite *in vitro*, qu'elle était synthétisée par un *frameshift* +1 au niveau d'une séquence de 10 adénines consécutives (10A) correspondant aux codons 9 à 11 de la polyprotéine du VHC (Xu et al. 2001). Toutefois, il est connu que les séquences répétitives, comme ces 10A, induisent des *frameshifts* -1 et +1 non-spécifiques dans des systèmes de traduction *in vitro*, mais non dans des cellules en culture (Weiss et al. 1987; Farabaugh 1997b). La première partie de notre étude a comparé l'expression de la protéine F dans des cellules de mammifères en culture (cellules rénales HEK 293FT dans l'article du Chapitre IV, mais nos résultats ont été reproduits dans des cellules hépatiques Huh7) et dans un système de traduction *in vitro* en utilisant une séquence du VHC contenant soit la suite de 10A utilisée dans les études précédentes, soit la séquence consensus (AAAGAAAAC) correspondant à cette région. Nos résultats ont démontré que la protéine F est synthétisée par un *frameshift* +1 uniquement lorsque la traduction de la séquence du VHC contenant la suite de 10A est effectuée *in vitro* (Baril et Brakier-Gingras 2005). Cette observation montre clairement que les résultats obtenus dans un système de traduction *in vitro* doivent être interprétés avec prudence. Nous avons également montré que la synthèse de la protéine F n'est pas affectée lorsque la synthèse de la protéine *core* du VHC est diminuée par l'introduction de codons d'arrêt au début de sa séquence codante ou par la mutation de son codon d'initiation. Cette indépendance de la synthèse de la protéine F vis-à-vis de la protéine *core* confirme que la protéine F n'est pas produite après un *frameshift* +1 dans des cellules en culture. Ces résultats nous ont également permis de suggérer que la protéine F était en fait produite



après initiation de la traduction directement dans le cadre de lecture +1 par rapport à la polyprotéine du VHC.

Des études de mutagenèse, qui introduisaient des codons d'arrêt à différents sites dans le cadre de lecture +1 de la protéine *core* du VHC, ont localisé le site d'initiation de la protéine F au codon chevauchant le codon 26 de la protéine *core* (codon 26+1). Dans nos constructions, ce codon 26+1 est GUG, comme dans 18% des séquences virales disponibles dans la banque de séquences du VHC (<http://s2as02.genes.nig.ac.jp/>), alors que ce codon est GCG dans 81% des séquences virales. Nous avons donc muté le codon GUG 26+1 de nos constructions en GCG et observé que la synthèse de la protéine F n'est pas diminuée lorsque l'initiation de la traduction est médiée par l'IRES du VHC, une condition qui correspond à la situation réelle lors de l'infection. L'initiation de la traduction au codon GCG, qui diffère du codon initiateur traditionnel AUG par deux nucléotides, soulevait la possibilité qu'un mécanisme d'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-indépendant au site A du ribosome (voir section 1.3.2.2) puisse être utilisé pour synthétiser la protéine F.

Pour vérifier cette possibilité, nous avons muté le codon 26+1 en codon AUG et observé que l'efficacité de synthèse de la protéine F augmente d'environ 50 fois, à un niveau similaire à l'efficacité de synthèse de la protéine *core*. Étant donné que l'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-indépendante au site A du ribosome ne devrait pas être affectée par la nature du codon d'initiation, ces résultats suggèrent fortement que l'initiation de la traduction se produit avec un Met-ARNt<sub>i</sub><sup>Met</sup> décodant un codon non-AUG au site P du ribosome, mais moins efficacement qu'un codon AUG. Pour confirmer ce résultat, nous avons également démontré que la synthèse de la protéine F augmente de deux à quatre fois lorsque la séquence flanquant le codon GUG ou GCG 26+1 est mutée pour créer un contexte Kozak optimal. La stimulation de l'efficacité d'initiation de la traduction par la séquence Kozak étant normalement associée à une initiation avec un Met-ARNt<sub>i</sub><sup>Met</sup> au site P du ribosome, ces résultats appuient un mécanisme d'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-dépendant pour la synthèse de la protéine F. Finalement, nous avons effectué des études de traduction *in vitro* en

présence d'édéine, un inhibiteur spécifique à la liaison du Met-ARNt<sub>i</sub><sup>Met</sup> au site P du ribosome (Szer et Kurylo-Borowska 1970; Odon et al. 1978). Une étude précédente sur le CrPV avait en effet démontré que l'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-indépendante au site A du ribosome n'était pas affectée par la présence d'édéine (Wilson et al. 2000). Nos résultats ont montré que l'édéine inhibe l'initiation de la traduction de la protéine F et de la protéine *core* de façon identique, confirmant qu'un mécanisme Met-ARNt<sub>i</sub><sup>Met</sup>-dépendant est responsable de l'initiation de la traduction de la protéine F, que le codon soit GUG ou GCG. Des cas d'initiation de la traduction à un codon GUG, qui diffère d'AUG par un nucléotide, ont déjà été observés, mais l'initiation de la traduction de la protéine F au codon GCG, qui diffère d'AUG par deux nucléotides, identifie pour la première fois une initiation de la traduction où un Met-ARNt<sub>i</sub><sup>Met</sup> décode un codon qui diffère d'AUG par plus d'un nucléotide.

Par contre, quoique le codon GCG diffère d'AUG par deux nucléotides, il peut quand même former deux paires de bases, dont une *wobble*, avec l'anticodon 3'UAC<sup>5'</sup> du Met-ARNt<sub>i</sub><sup>Met</sup>. La formation de deux paires de bases pourrait donc être requise pour permettre l'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-dépendante. Pour tester cette hypothèse, nous avons mesuré l'efficacité de synthèse de la protéine F lorsque son codon initiateur 26+1 est remplacé par différents codons dont le nombre de mésappariements avec l'anticodon du Met-ARNt<sub>i</sub><sup>Met</sup> varie. Des six codons sens testés, le codon GUG forme trois paires de base dont une *wobble*, les codons GCG, GAG et GUC forment deux paires de bases dont une *wobble*, le codon GCC forme une paire de bases *wobble* et le codon UAC ne forme aucune paire de bases. Nous avons aussi testé un codon d'arrêt (non-sens) UGA qui ne forme aucune paire de bases avec l'anticodon du Met-ARNt<sub>i</sub><sup>Met</sup>. Nos résultats montrent que chacun des codons sens permet l'initiation de la traduction de la protéine F avec une efficacité similaire au codon GUG, qui est présent dans notre construction sauvage, peu importe le nombre de mésappariements entre le codon non-AUG et l'anticodon du Met-ARNt<sub>i</sub><sup>Met</sup> (Figure 5-2, résultats non-publiés). Le seul codon qui ne permet pas une initiation efficace de la traduction de la protéine F est le codon d'arrêt UGA. Il semble donc y avoir

une discrimination entre les codons sens et non-sens comme site d'initiation. L'absence d'initiation de la traduction à un codon d'arrêt pourrait résulter d'un blocage du site P du ribosome par le facteur de terminaison de la traduction eRF1, qui empêcherait la liaison du Met-ARNt<sup>Met</sup>. Il est également important de rappeler qu'un codon AUG à la position 26+1 permet de synthétiser 50 fois plus de protéine F que le codon GUG qui forme également trois paires de bases, dont une *wobble*, avec l'anticodon du Met-ARNt<sup>Met</sup>. Il y a donc discrimination efficace entre un codon AUG et un codon non-AUG au niveau de l'initiation de la traduction de la protéine F, mais il n'y a pas de discrimination entre les différents codons sens non-AUG. Il est intéressant de signaler que deux publications récentes sur le *bypassing* du gène 60 du bactériophage T4 montrent que la complémentarité du lien codon:anticodon n'est pas un facteur déterminant pour la liaison d'un ARNt au site P (Herr et al. 2004; Bucklin et al. 2005). Dans le mécanisme du *bypassing*, le peptidyl-ARNt au site P du ribosome en élongation se dissocie de l'ARNm à un codon de décollage, le ribosome parcourt l'ARNm en direction 5' et reprend la traduction à un site d'atterrissage, qui est normalement un codon synonyme au site de décollage (décrit en détail à la section 1.3). Les publications récentes ont montré que lors du *bypassing*, le ribosome portant un peptidyl-ARNt au site P peut reprendre la traduction à des codons qui ne possèdent aucune complémentarité avec le peptidyl-ARNt. Ces études démontrent donc que la complémentarité du lien codon:anticodon n'est pas un facteur déterminant pour la liaison d'un ARNt au site P, contrairement à la liaison au site A du ribosome. L'emplacement de la reprise de la traduction serait plutôt dicté par la présence de séquences primaires et secondaires interagissant potentiellement avec le ribosome. Un mécanisme semblable pourrait être utilisé par le VHC pour contrôler l'emplacement du site d'initiation de la traduction de la protéine F. Ce contrôle serait principalement exercé par l'IRES du VHC qui permet à la sous-unité 40S du ribosome de se lier directement au codon initiateur de la protéine F, puisque l'initiation de la traduction de la protéine F à partir du codon GCG 26+1 est efficace en présence de l'IRES. Par contre, l'initiation de la traduction au codon GCG n'est pas efficace en absence de l'IRES, lorsqu'un mécanisme traditionnel de balayage

est utilisé (voir Figure 4-8 de l'article). Cependant, la synthèse de la protéine F à partir du codon GUG 26+1 est aussi efficace en présence qu'en absence de l'IRES, ce qui suggère que d'autres séquences de l'ARN du VHC joueraient également un rôle dans la détermination de l'emplacement du site d'initiation de la traduction de la protéine F. Des études préliminaires de mutagenèse visant à identifier des séquences proximales au codon 26+1 pouvant interagir avec l'ARNr 18S ou des protéines du ribosome ne nous ont toutefois pas permis d'identifier de motif nécessaire à l'initiation de la traduction de la protéine F.

En collaboration avec le Dr Y. Durocher de l'Institut de recherche en biotechnologie, nous avons entamé des expériences visant à purifier la protéine F suite à son expression dans un système de culture de cellules de mammifères en suspension, afin de séquencer l'extrémité N-terminale de la protéine F. La protéine F sera isolée par double purification utilisant les étiquettes polyhistidine et Strep-tag II (Durocher et al. 2002). Deux groupes de recherche ont déjà séquencé la protéine F, mais l'utilisation de systèmes d'expression inadéquats faussait leurs résultats (voir section 1.5.3). En effet, une étude utilisant un système d'expression *in vitro*, dans un RRL, a montré que la protéine F est synthétisée par un *frameshift* +1 au niveau d'une suite de 10A (Xu et al. 2001). Toutefois, cette suite de 10A induit des *frameshifts* -1 et +1 non-spécifiques dans des systèmes de traduction *in vitro* et nous avons démontré que le *frameshift* +1 ne se produit pas dans des cellules en culture. Une autre étude exprimant la protéine F chez *Escherichia coli* a montré qu'elle était synthétisée après un *frameshift* +1 au codon 42 (AGG) de la polyprotéine (Boulant et al. 2003). Cependant, ce *frameshift* +1 se produit chez la bactérie, mais non chez l'homme, puisque le codon AGG est un codon rare uniquement chez la bactérie. Le séquençage de la protéine F produite dans des cellules humaines en culture nous permettra de confirmer, par une méthode indépendante, la localisation du site d'initiation de la traduction de la protéine F identifié par mutagenèse.

Finalement, plusieurs études ont montré que l'initiation de la traduction de la polyprotéine est régulée négativement par l'expression de la protéine *core* qui lie spécifiquement l'IRES du VHC (Zhang et al. 2002; Li D. et al. 2004; Boni et al. 2005). Nous avons voulu vérifier si l'inhibition de l'initiation de la traduction de la polyprotéine par la protéine *core* bloque également la synthèse de la protéine F, puisque ces deux protéines sont en compétition pour les mêmes sous-unités 40S liées à l'IRES du VHC. Une expérience pilote où nous avons surexprimé la protéine *core* dans des cellules en culture montre que la surexpression de la protéine *core* diminue la synthèse de la polyprotéine jusqu'à 30%, alors que la synthèse de la protéine F reste inchangée (Figure 5-3). Ce résultat suggère que la protéine *core* lierait l'IRES du VHC dans la région de la tige-boucle contenant l'AUG initiateur de la polyprotéine (domaine IV de l'IRES, voir Figure 1-4B). Cette interaction masquerait l'AUG initiateur de la polyprotéine sans affecter la liaison de la sous-unité 40S à l'IRES, qui pourrait initier la traduction de la protéine F. Cette expérience pilote suggère donc que la synthèse de la protéine F est régulée indépendamment de la polyprotéine, ce qui appuie l'hypothèse voulant que la protéine F joue un rôle important dans le cycle de réplication du VHC.

Les résultats de ce troisième article (voir Chapitre IV) ont permis d'identifier le site d'initiation de la traduction de la protéine F (Baril et Brakier-Gingras 2005). La production de la protéine F avec la séquence N-terminale exacte contribuera à déterminer sa localisation cellulaire, puisque la protéine F utilisée dans les études de localisation précédentes était fusionnée à la portion N-terminale de la protéine *core*, ce qui pouvait biaiser les résultats (Roussel et al. 2003; Xu et al. 2003). Il sera également très important de déterminer la fonction de la protéine F, puisque plusieurs indices laissent croire qu'elle pourrait jouer un rôle important dans l'infection virale. Premièrement, la protéine F est la seule protéine du VHC qui puisse être produite indépendamment des autres protéines virales synthétisées sous forme de polyprotéine. Plus précisément, la protéine F pourrait jouer un rôle dans l'établissement de la persistance virale puisque son cadre de lecture +1 est conservé à travers les différents génotypes du VHC et chez le GB virus B qui font partie

des rares virus à ARN(+) pouvant établir une infection persistante (Martin et al. 2003). Un rôle potentiel de la protéine F dans la persistance du VHC est également suggéré par son mécanisme de synthèse qui est similaire à celui de la protéine L\*. En effet, la protéine L\* est essentielle à la persistance du TMEV et est synthétisée après initiation de la traduction à un codon ACG dans le cadre de lecture +1 par rapport à la polyprotéine virale (van Eyll et Michiels 2000, 2002). De plus, la protéine F pourrait être impliquée dans le développement de lésions au foie, puisque son expression semble augmentée dans les patients ayant une cirrhose ou un cancer du foie (Branch et al. 2005). Pour toutes ces raisons, la synthèse de la protéine F, une fois son rôle défini, pourrait représenter une cible efficace pour le développement de traitements thérapeutiques. Par exemple, une version eucaryote de l'approche expérimentale développée pour la sélection de peptides liant spécifiquement le signal stimulateur de *frameshift* -1 du VIH-1 (voir section 5.1) pourrait être développée afin de sélectionner des peptides liant spécifiquement la tige-boucle conservée qui renferme le codon 26+1. Ces peptides pourraient inhiber spécifiquement la synthèse de la protéine F en masquant son site d'initiation de la traduction.

Figure 5-2 : Efficacité d'initiation de la traduction de la protéine F du VHC lorsque le codon 26+1 est remplacé par différents codons.

Rappelons que la synthèse de la protéine F est évaluée en mesurant la synthèse d'une fusion entre le début de la protéine F et la luciférase (voir détails dans l'article du Chapitre IV). La synthèse de la protéine F dans les constructions mutantes est exprimée en pourcentage de la synthèse de la protéine F dans la construction sauvage (GUG), dont la valeur est normalisée à 100%. L'initiation de la traduction est dépendante de l'IRES dans cette expérience. Les paires de bases Watson-Crick entre le codon 26+1 et l'anticodon  $3'UAC^{5'}$  du Met-ARNt<sub>i</sub><sup>Met</sup> (en bleu) sont identifiées par un tiret (-) et les paires de bases *wobble* sont identifiées par un point (•). Les résultats correspondent à la moyenne de trois expériences indépendantes.

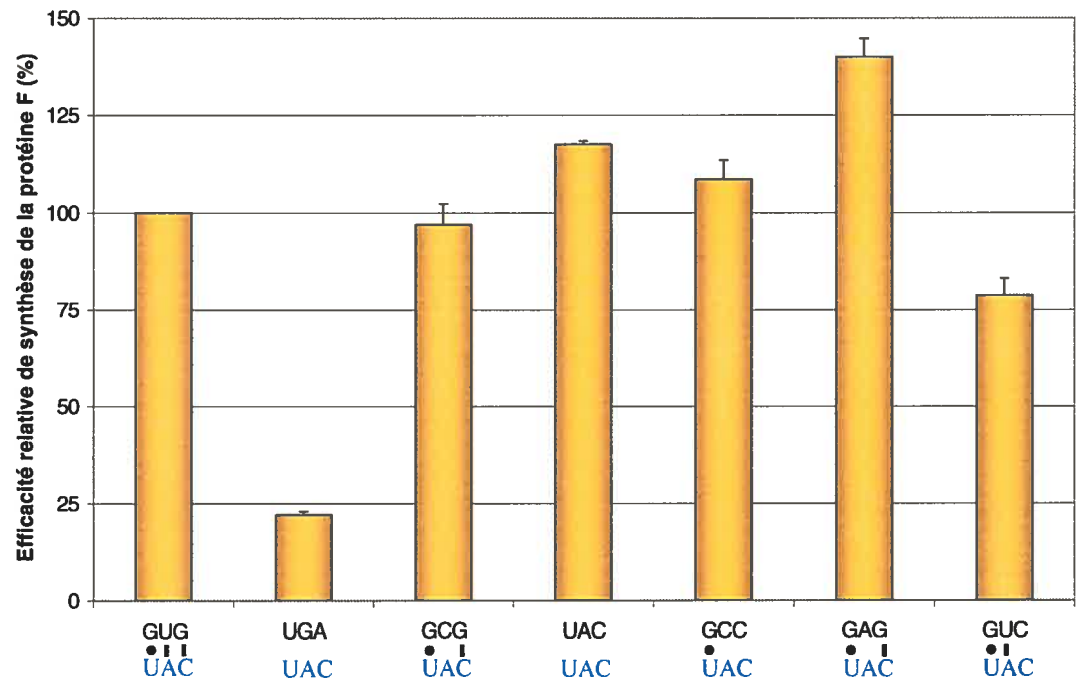
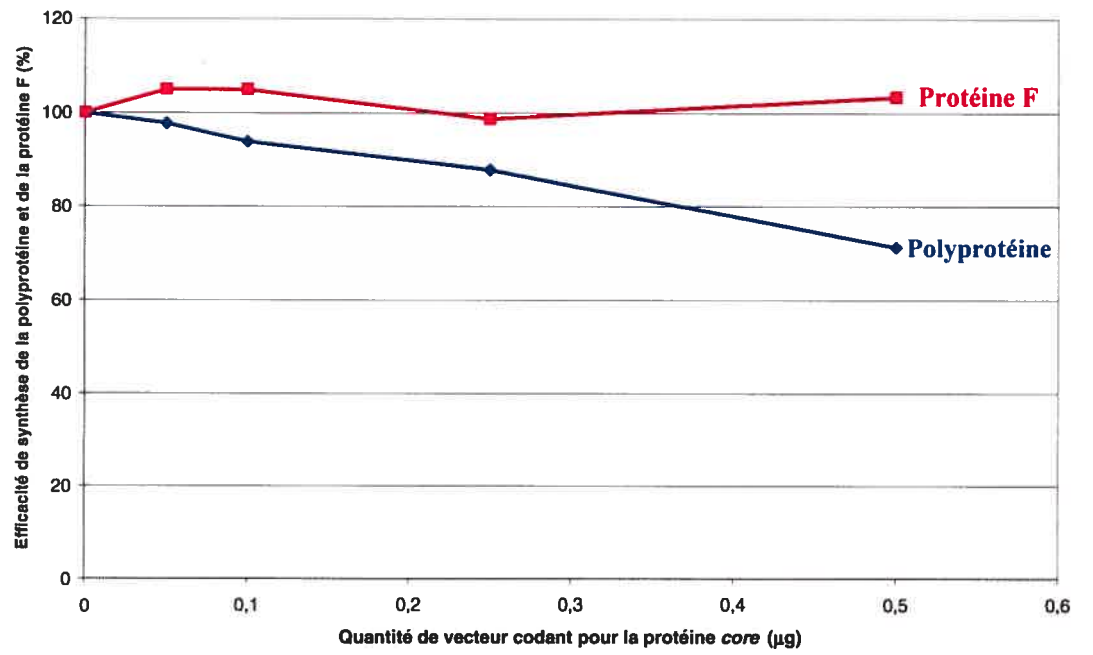




Figure 5-3 : Effet de la surexpression de la protéine *core* du VHC sur la synthèse de la polyprotéine et de la protéine F.

Des quantités croissantes du vecteur pcDNA3.1-CORE codant pour la protéine *core* (0 à 0,5 µg) ont été cotransfectées dans des cellules HEK 293FT avec le vecteur servant à mesurer la synthèse de la polyprotéine ou de la protéine F du VHC. L'efficacité de synthèse de la polyprotéine et de la protéine F est mesurée en utilisant des fusions entre le début de la séquence de ces protéines et la luciférase (voir détails dans l'article du Chapitre IV). Leur efficacité de synthèse en présence de pcDNA3.1-CORE est exprimée en pourcentage de leur efficacité de synthèse en absence de pcDNA3.1-CORE. L'initiation de la traduction est dépendante de l'IRES dans cette expérience.



## 5.4 Conclusions

Les travaux de recherche effectués au cours de mon doctorat ont porté sur des mécanismes non-conventionnels de traduction chez le VIH-1 et le VHC, deux virus responsables de pandémies mondiales qui entraînent la mort de plusieurs millions de personnes annuellement. Mes travaux ont permis de caractériser les régions de *frameshift* – 1 des différents sous-types du groupe M et du groupe O du VIH-1 et ont fourni des informations importantes sur le mécanisme menant à la synthèse de la protéine F du VHC.

Dans le cas du VIH-1, mes études peuvent être utilisées pour le développement de nouveaux agents anti-VIH ciblant spécifiquement le signal stimulateur du *frameshift* – 1 des différents sous-types du groupe M ou du groupe O. Le développement d'un agent anti-*frameshift* permettrait le traitement des patients séropositifs infectés par des souches de VIH-1 résistantes aux traitements présentement disponibles. En attendant un vaccin efficace, le développement de nouveaux agents anti-VIH reste une priorité internationale.

Quant à mes études sur le VHC, la caractérisation de l'initiation de la traduction à un codon non-AUG menant à la synthèse de la protéine F a identifié pour la première fois un mécanisme d'initiation Met-ARNt<sub>i</sub><sup>Met</sup>-dépendant se produisant à un codon qui diffère d'AUG par plus d'un nucléotide. En plus de cet aspect fondamental, mes travaux de recherche permettent d'identifier la séquence N-terminale exacte de la protéine F, fournissant une information importante aux différentes équipes de recherche qui tentent présentement de déterminer le rôle de la protéine F. Par exemple, la protéine F pourrait jouer un rôle important dans la progression de la maladie, puisqu'elle semble être surexprimée chez les patients qui développent une cirrhose ou un cancer du foie. Il serait donc intéressant de développer un agent thérapeutique qui lierait spécifiquement la tige-boucle conservée renfermant le site d'initiation de la protéine F. La liaison de cet agent pourrait bloquer spécifiquement la synthèse de la protéine F.

## Bibliographie

- Agnello, V., Abel, G., Elfahal, M., Knight, G.B. et Zhang, Q.X. 1999. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc. Natl. Acad. Sci. USA* **96**: 12766-12771.
- Alam, S.L., Wills, N.M., Ingram, J.A., Atkins, J.F. et Gesteland, R.F. 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.
- Ali, N. et Siddiqui, A. 1995. Interaction of polypyrimidine tract-binding protein with the 5' noncoding region of the hepatitis C virus RNA genome and its functional requirement in internal initiation of translation. *J. Virol.* **69**: 6367-6375.
- Ali, N. et Siddiqui, A. 1997. The La antigen binds 5' noncoding region of the hepatitis C virus RNA in the context of the initiator AUG codon and stimulates internal ribosome entry site-mediated translation. *Proc. Natl. Acad. Sci. USA* **94**: 2249-2254.
- Andries, K., Azijn, H., Thielemans, T., Ludovici, D., Kukla, M., Heeres, J., Janssen, P., De Corte, B., Vingerhoets, J., Pauwels, R. et de Bethune, M.P. 2004. TMC125, a novel next-generation nonnucleoside reverse transcriptase inhibitor active against nonnucleoside reverse transcriptase inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **48**: 4680-4686.
- 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.
- Bain, C., Parroche, P., Lavergne, J.P., Duverger, B., Vieux, C., Dubois, V., Komurian-Pradel, F., Trepo, C., Gebuhrer, L., Paranhos-Baccala, G., Penin, F. et Inchauspe, G. 2004. Memory T-cell-mediated immune responses specific to an alternative core protein in hepatitis C virus infection. *J. Virol.* **78**: 10460-10469.

- Baldwin, C.E., Sanders, R.W., Deng, Y., Jurriaans, S., Lange, J.M., Lu, M. et Berkhout, B. 2004. Emergence of a drug-dependent human immunodeficiency virus type 1 variant during therapy with the T20 fusion inhibitor. *J. Virol.* **78**: 12428-12437.
- Baranov, P.V., Gesteland, R.F. et Atkins, J.F. 2004. P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* **10**: 221-230.
- Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. et Howard, M.T. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**: 498-510.
- Baril, M. et Brakier-Gingras, L. 2005. Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein. *Nucleic Acids Res.* **33**: 1474-1486.
- 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, S.V. 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.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dautet, 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.
- Barry, J.K. et Miller, W.A. 2002. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc. Natl. Acad. Sci. USA* **99**: 11133-11138.
- Bartenschlager, R., Kaul, A. et Sparacio, S. 2003. Replication of the hepatitis C virus in cell culture. *Antiviral Res* **60**: 91-102.
- Bartenschlager, R. et Lohmann, V. 2000. Replication of hepatitis C virus. *J. Gen. Virol.* **81**: 1631-1648.

- Bartenschlager, R., Lohmann, V., Wilkinson, T. et Koch, J.O. 1995. Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J. Virol.* **69**: 7519-7528.
- Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A. et Cosset, F.L. 2003. Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J. Biol. Chem.* **278**: 41624-41630.
- Basu, A., Steele, R., Ray, R. et Ray, R.B. 2004. Functional properties of a 16 kDa protein translated from an alternative open reading frame of the core-encoding genomic region of hepatitis C virus. *J. Gen. Virol.* **85**: 2299-2306.
- Behne, D. et Kyriakopoulos, A. 2001. Mammalian selenium-containing proteins. *Annu Rev Nutr* **21**: 453-473.
- Behrens, S.E., Tomei, L. et De Francesco, R. 1996. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J.* **15**: 12-22.
- Beier, H. et Grimm, M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**: 4767-4782.
- Belcourt, M.F. et Farabaugh, P.J. 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* **62**: 339-352.
- Bertrand, C., Prere, M.F., Gesteland, R.F., Atkins, J.F. 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.
- Bidou, L., Stahl, G., Grima, B., Liu, H., Cassan, M. et 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.
- Blight, K.J., Kolykhalov, A.A. et Rice, C.M. 2000. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**: 1972-1974.
- Boden, D., Pusch, O., Lee, F., Tucker, L. et Ramratnam, B. 2003. Human immunodeficiency virus type 1 escape from RNA interference. *J. Virol.* **77**: 11531-11535.

- Boni, S., Lavergne, J.P., Boulant, S. et Cahour, A. 2005. Hepatitis C virus core protein acts as a trans-modulating factor on internal translation initiation of the viral RNA. *J. Biol. Chem.*
- Boulant, S., Becchi, M., Penin, F. et Lavergne, J.P. 2003. Unusual multiple recoding events leading to alternative forms of hepatitis C virus core protein from genotype 1b. *J. Biol. Chem.* **278**: 45785-45792.
- Brakier-Gingras, L., Melançon, P., Côté, M. et Payant, C. 1998. Use of antisense DNA oligomers for inhibition of HIV by decreasing ribosomal frameshifting. US patent no. 5707866.
- Branch, A.D., Stump, D.D., Gutierrez, J.A., Eng, F. et Walewski, J.L. 2005. The Hepatitis C Virus Alternate Reading Frame (ARF) and Its Family of Novel Products: The Alternate Reading Frame Protein/F-Protein, the Double-Frameshift Protein, and Others. *Semin Liver Dis* **25**: 105-117.
- Braut, V. et Miller, W.A. 1992. Translational frameshifting mediated by a viral sequence in plant cells. *Proc Natl Acad Sci U S A* **89**: 2262-2266.
- Brierley, I. 1995. Ribosomal frameshifting viral RNAs. *J. Gen. Virol.* **76**: 1885-1892.
- Brierley, I., Digard, P. et Inglis, S.C. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**: 537-547.
- Brierley, I., Jenner, A.J. et Inglis, S.C. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227**: 463-479.
- Brierley, I. et Pennel, S. 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic -1 ribosomal frameshifting. *The Ribosome*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press. pp 233-248.
- Brierley, I., Rolley, N.J., Jenner, A.J. et Inglis, S.C. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **220**: 889-902.
- Brunelle, M.N., 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.
- Bucklin, D.J., Wills, N.M., Gesteland, R.F. et Atkins, J.F. 2005. P-site pairing subtleties revealed by the effects of different tRNAs on programmed translational bypassing where anticodon re-pairing to mRNA is separated from dissociation. *J. Mol. Biol.* **345**: 39-49.
- Buratti, E., Tisminetzky, S., Zotti, M. et Baralle, F.E. 1998. Functional analysis of the interaction between HCV 5'UTR and putative subunits of eukaryotic translation initiation factor eIF3. *Nucleic Acids Res* **26**: 3179-3187.
- Carrasco, L., Battaner, E. et Vazquez, D. 1974. The elongation steps in protein synthesis by eukaryotic ribosomes: effects of antibiotics. *Methods Enzymol.* **30**: 282-289.
- Cassan, M., Delaunay, N., Vaquero, C. et 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.
- Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L.A., Offenhauser, N., Marchisio, P.C. et Biffo, S. 2003. Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**: 579-584.
- Cevallos, R.C. et Sarnow, P. 2005. Factor-independent assembly of elongation-competent ribosomes by an internal ribosome entry site located in an RNA virus that infects penaeid shrimp. *J. Virol.* **79**: 677-683.
- Chamorro, M., Parkin, N. et 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. USA* **89**: 713-717.
- Chandler, M. et Fayet, O. 1993. Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* **7**: 497-503.
- Chang, S.Y., Sutthent, R., Auewarakul, P., Apichartpiyakul, C., Essex, M. et Lee, T.H. 1999. Differential stability of the mRNA secondary structures in the frameshift site of various HIV type 1 viruses. *AIDS Res. Hum. Retroviruses* **15**: 1591-1596.
- Chen, C.Y. et Sarnow, P. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**: 415-417.



- Chen, X., Kang, H., Shen, L.X., Chamorro, M., Varmus, H.E. 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.
- Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. et Houghton, M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**: 359-362.
- Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G. et Houghton, M. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**: 2451-2455.
- Ciccaglione, A.R., Costantino, A., Marcantonio, C., Equestre, M., Geraci, A. et Rapicetta, M. 2001. Mutagenesis of hepatitis C virus E1 protein affects its membrane-permeabilizing activity. *J. Gen. Virol.* **82**: 2243-2250.
- Coffin, J.M., Hughes, S.H. et Varmus, H. 1997. *Retroviruses*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press.
- Condra, J.H. 1998. Resistance to HIV protease inhibitors. *Haemophilia* **4**: 610-615.
- Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M. et et al. 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374**: 569-571.
- Condrón, B.G., Atkins, J.F. et Gesteland, R.F. 1991. Frameshifting in gene 10 of bacteriophage T7. *J. Bacteriol.* **173**: 6998-7003.
- Corbet, S., Muller-Trutwin, M.C., Versmisse, P., Delarue, S., Ayoub, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F. et Maucclere, P. 2000. env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* **74**: 529-534.
- Craig, W.J. et Caskey, C.T. 1986. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* **322**: 273-275.

- Craigien, W.J., Cook, R.G., Tate, W.P. et Caskey, C.T. 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc Natl Acad Sci US A* **82**: 3616-3620.
- Cullen, B.R. 1998. HIV-1 auxiliary proteins: making connections in a dying cell. *Cell* **93**: 685-692.
- Curran, J.F. et Yarus, M. 1988. Use of tRNA suppressors to probe regulation of Escherichia coli release factor 2. *J. Mol. Biol.* **203**: 75-83.
- de Baar, M.P., Janssens, W., de Ronde, A., Fransen, K., Colebunders, R., Kestens, L., van der Groen, G. et Goudsmit, J. 2000. Natural residues versus antiretroviral drug-selected mutations in HIV type 1 group O reverse transcriptase and protease related to virological drug failure in vivo. *AIDS Res. Hum. Retroviruses* **16**: 1385-1394.
- De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wauters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B. et et al. 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J. Virol.* **64**: 1207-1216.
- Demirov, D.G., Orenstein, J.M. et Freed, E.O. 2002. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *J. Virol.* **76**: 105-117.
- Descamps, D., Collin, G., Letourneur, F., Apetrei, C., Damond, F., Loussert-Ajaka, I., Simon, F., Saragosti, S. et Brun-Vezinet, F. 1997. Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. *J. Virol.* **71**: 8893-8898.
- Dever, T.E. 2002. Gene-specific regulation by general translation factors. *Cell* **108**: 545-556.
- Dinman, J.D., Icho, T. et Wickner, R.B. 1991. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci US A* **88**: 174-178.
- Dinman, J.D., Richter, S., Plant, E.P., Taylor, R.C., Hammell, A.B. et 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.

- Dinman, J.D., Ruiz-Echevarria, M.J., Czaplinski, K. et Peltz, S.W. 1997. Peptidyl-transferase inhibitors have antiviral properties by altering programmed -1 ribosomal frameshifting efficiencies: development of model systems. *Proc. Natl. Acad. Sci. USA* **94**: 6606-6611.
- Dinman, J.D., Ruiz-Echevarria, M.J. et Peltz, S.W. 1998. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.* **16**: 190-196.
- Dinman, J.D. et Wickner, R.B. 1992. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.* **66**: 3669-3676.
- Dixit, N.M., Layden-Almer, J.E., Layden, T.J. et Perelson, A.S. 2004. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. *Nature* **432**: 922-924.
- Drabkin, H.J., Estrella, M. et Rajbhandary, U.L. 1998. Initiator-elongator discrimination in vertebrate tRNAs for protein synthesis. *Mol. Cell. Biol.* **18**: 1459-1466.
- Du, Z., Giedroc, D.P. et Hoffman, D.W. 1996. Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. *Biochemistry* **35**: 4187-4198.
- Du, Z., Holland, J.A., Hansen, M.R., Giedroc, D.P. et Hoffman, D.W. 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.
- Durocher, Y., Perret, S. et Kamen, A. 2002. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res.* **30**: E9.
- Egli, M., Minasov, G., Su, L. et Rich, A. 2002. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc. Natl. Acad. Sci. USA* **99**: 4302-4307.

- Eustice, D.C., Feldman, P.A., Colberg-Poley, A.M., Buckery, R.M. et Neubauer, R.H. 1991. A sensitive method for the detection of beta-galactosidase in transfected mammalian cells. *Biotechniques* **11**: 739-740, 742-733.
- Falk, H., Mador, N., Udi, R., Panet, A. et Honigman, A. 1993. Two cis-acting signals control ribosomal frameshift between human T-cell leukemia virus type II gag and pro genes. *J. Virol.* **67**: 6273-6277.
- Farabaugh, P.J. 1997a. Programmed -1 frameshifting in eukaryotes. In: Farabaugh PJ, ed. *Programmed alternative reading of the genetic code*. Austin, TX: R.G. Landes. pp 69-102.
- Farabaugh, P.J. 1997b. Programmed +1 frameshifting. In: Farabaugh PJ, ed. *Programmed alternative reading of the genetic code*. Austin, TX: R.G. Landes. pp 41-67.
- Farabaugh, P.J. 1997c. Programmed readthrough of translational termination codons. In: Farabaugh PJ, ed. *Programmed alternative reading of the genetic code*. Austin, TX: R.G. Landes. pp 149-182.
- Farabaugh, P.J. 2000. Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.* **64**: 131-170.
- Farabaugh, P.J. et Bjork, G.R. 1999. How translational accuracy influences reading frame maintenance. *Embo J* **18**: 1427-1434.
- Farabaugh, P.J., 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.
- Frankel, A.D. et Young, J.A. 1998. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* **67**: 1-25.
- Frank-Kamenetskii, M.D. et Mirkin, S.M. 1995. Triplex DNA structures. *Annu. Rev. Biochem.* **64**: 65-95.
- Futterer, J., Kiss-Laszlo, Z. et Hohn, T. 1993. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**: 789-802.
- Gale, M., Jr., Blakely, C.M., Kwieciszewski, B., Tan, S.L., Dossett, M., Tang, N.M., Korth, M.J., Polyak, S.J., Gretch, D.R. et Katze, M.G. 1998. Control of PKR protein kinase

- by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol. Cell. Biol.* **18**: 5208-5218.
- Gallant, J., Bonthuis, P. et Lindsley, D. 2003. Evidence that the bypassing ribosome travels through the coding gap. *Proc. Natl. Acad. Sci. USA* **100**: 13430-13435.
- Gallant, J.A. et Lindsley, D. 1998. Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream. *Proc. Natl. Acad. Sci. USA* **95**: 13771-13776.
- Gallego, O., de Mendoza, C., Perez-Elias, M.J., Guardiola, J.M., Pedreira, J., Dalmau, D., Gonzalez, J., Moreno, A., Arribas, J.R., Rubio, A., Garcia-Arata, I., Leal, M., Domingo, P. et Soriano, V. 2001. Drug resistance in patients experiencing early virological failure under a triple combination including indinavir. *AIDS* **15**: 1701-1706.
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. et Hahn, B.H. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* **397**: 436-441.
- Gesteland, R.F. et Atkins, J.F. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741-768.
- Giedroc, D.P., Theimer, C.A. et Nixon, P.L. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.* **298**: 167-185.
- Gil, J. et Esteban, M. 2000. Induction of apoptosis by the dsRNA-dependent protein kinase (PKR): mechanism of action. *Apoptosis* **5**: 107-114.
- Giorgi, C. et Kolakofsky, D. 1984. Effect of poliovirus superinfection on Sendai virus protein synthesis. *J. Virol.* **52**: 298-299.
- Gitlin, L., Karelsky, S. et Andino, R. 2002. Short interfering RNA confers intracellular antiviral immunity in human cells. *Nature* **418**: 430-434.
- Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H.E., Bienz, K. et Moradpour, D. 2003. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J. Virol.* **77**: 5487-5492.

- Goss Kinzy, T., Harger, J.W., Carr-Schmid, A., Kwon, J., Shastry, M., Justice, M. et Dinman, J.D. 2002. New targets for antivirals: the ribosomal A-site and the factors that interact with it. *Virology* **300**: 60-70.
- Grakoui, A., McCourt, D.W., Wychowski, C., Feinstone, S.M. et Rice, C.M. 1993. A second hepatitis C virus-encoded proteinase. *Proc. Natl. Acad. Sci. USA* **90**: 10583-10587.
- Gray, N.K. et Wickens, M. 1998. Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.* **14**: 399-458.
- Guo, J.T., Bichko, V.V. et Seeger, C. 2001. Effect of alpha interferon on the hepatitis C virus replicon. *J. Virol.* **75**: 8516-8523.
- Gupta, K.C. et Patwardhan, S. 1988. ACG, the initiator codon for a Sendai virus protein. *J. Biol. Chem.* **263**: 8553-8556.
- Hahn, B.H., Shaw, G.M., De Cock, K.M. et Sharp, P.M. 2000. AIDS as a zoonosis: scientific and public health implications. *Science* **287**: 607-614.
- Han, H. et 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.
- Hansen, T.M., Baranov, P.V., Ivanov, I.P., Gesteland, R.F. et Atkins, J.F. 2003. Maintenance of the correct open reading frame by the ribosome. *EMBO Rep* **4**: 499-504.
- Haseltine, W.A. 2001. Beyond chicken soup. *Sci Am* **285**: 56-63.
- Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C. et Miller, M.D. 2000. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**: 646-650.
- Hellen, C.U. et Pestova, T.V. 1999. Translation of hepatitis C virus RNA. *J. Viral Hepat.* **6**: 79-87.
- Hellen, C.U. et Sarnow, P. 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15**: 1593-1612.
- Hellen, C.U., Witherell, G.W., Schmid, M., Shin, S.H., Pestova, T.V., Gil, A. et Wimmer, E. 1993. A cytoplasmic 57-kDa protein that is required for translation of

- picornavirus RNA by internal ribosomal entry is identical to the nuclear pyrimidine tract-binding protein. *Proc. Natl. Acad. Sci. USA* **90**: 7642-7646.
- Herr, A.J., Atkins, J.F. et Gesteland, R.F. 2000. Coupling of open reading frames by translational bypassing. *Annu. Rev. Biochem.* **69**: 343-372.
- Herr, A.J., Wills, N.M., Nelson, C.C., Gesteland, R.F. et Atkins, J.F. 2001. Drop-off during ribosome hopping. *J. Mol. Biol.* **311**: 445-452.
- Herr, A.J., Wills, N.M., Nelson, C.C., Gesteland, R.F. et Atkins, J.F. 2004. Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *J. Biol. Chem.* **279**: 11081-11087.
- Hershey, J.W.B., Mathews, M. et Sonenberg, N. 1996. Translational control. *Cold Spring Harbor monograph series*, 30. Plainview, N.Y.: Cold Spring Harbor Laboratory Press. pp xi, 794.
- Hill, M.K., Shehu-Xhilaga, M., Crowe, S.M. 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.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. et Pease, L.R. 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.
- Honda, M., Beard, M.R., Ping, L.H. et Lemon, S.M. 1999. A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.* **73**: 1165-1174.
- Honda, M., Brown, E.A. et Lemon, S.M. 1996. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2**: 955-968.

- Honda, M., Kaneko, S., Matsushita, E., Kobayashi, K., Abell, G.A. et Lemon, S.M. 2000. Cell cycle regulation of hepatitis C virus internal ribosomal entry site-directed translation. *Gastroenterology* **118**: 152-162.
- Howard, C.R. 2002. Hepatitis C virus: clades and properties. *J. Gastroenterol. Hepatol.* **17 Suppl**: S468-470.
- Howard, M.T., Gesteland, R.F. et Atkins, J.F. 2004. Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA* **10**: 1653-1661.
- Huang, W.M., Ao, S.Z., 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.
- Hung, M., Patel, P., Davis, S. et Green, S.R. 1998. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J. Virol.* **72**: 4819-4824.
- Ikeda, M., Yi, M., Li, K. et Lemon, S.M. 2002. Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. *J. Virol.* **76**: 2997-3006.
- Ito, T., Tahara, S.M. et Lai, M.M. 1998. The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *J. Virol.* **72**: 8789-8796.
- Ivanov, I.P., Gesteland, R.F. et Atkins, J.F. 1998a. A second mammalian antizyme: conservation of programmed ribosomal frameshifting. *Genomics* **52**: 119-129.
- Ivanov, I.P., Gesteland, R.F., Matsufuji, S. et Atkins, J.F. 1998b. 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, I.P., Gurvich, O.L., Gesteland, R.F. et Atkins, J.F. 2003. Recoding: site- or mRNA-specific alteration of genetic readout utilized for gene expression. In: Lapointe J, Brakier-Gingras L, eds. *Translation mechanisms*. Georgetown, TX: Landes Bioscience. pp 354-369.



- Jacks, T., Madhani, H.D., Masiarz, F.R. et Varmus, H.E. 1988a. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**: 447-458.
- Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. et Varmus, H.E. 1988b. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**: 280-283.
- Jacks, T. et Varmus, H.E. 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**: 1237-1242.
- Jacque, J.M., Triques, K. et Stevenson, M. 2002. Modulation of HIV-1 replication by RNA interference. *Nature* **418**: 435-438.
- Jan, E. et Sarnow, P. 2002. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324**: 889-902.
- Ji, H., Fraser, C.S., Yu, Y., Leary, J. et Doudna, J.A. 2004. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc. Natl. Acad. Sci. USA* **101**: 16990-16995.
- Jin, D.Y., Wang, H.L., Zhou, Y., Chun, A.C., Kibler, K.V., Hou, Y.D., Kung, H. et Jeang, K.T. 2000. Hepatitis C virus core protein-induced loss of LZIP function correlates with cellular transformation. *EMBO J.* **19**: 729-740.
- Jordan, M., Schallhorn, A. et Wurm, F.M. 1996. Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res.* **24**: 596-601.
- 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. 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, L.D. et Lorsch, J.R. 2004. The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **73**: 657-704.
- Karacostas, V., Wolffe, E.J., Nagashima, K., Gonda, M.A. 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.
- Kato, N. 2000. Genome of human hepatitis C virus (HCV): gene organization, sequence diversity, and variation. *Microb. Comp. Genomics* **5**: 129-151.
- Kawakami, K., Pande, S., Faiola, B., Moore, D.P., Boeke, J.D., Farabaugh, P.J., Strathern, J.N., Nakamura, Y. et Garfinkel, D.J. 1993. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in *Saccharomyces cerevisiae*. *Genetics* **135**: 309-320.
- Kieft, J.S., Grech, A., Adams, P. et Doudna, J.A. 2001. Mechanisms of internal ribosome entry in translation initiation. *Cold Spring Harb Symp Quant Biol* **66**: 277-283.
- Kilby, J.M., Hopkins, S., Venetta, T.M., DiMassimo, B., Cloud, G.A., Lee, J.Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Matthews, T., Johnson, M.R., Nowak, M.A., Shaw, G.M. et Saag, M.S. 1998. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **4**: 1302-1307.
- Kim, D.W., Gwack, Y., Han, J.H. et Choe, J. 1995. C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem. Biophys. Res. Commun.* **215**: 160-166.
- Kim, Y.G., Maas, S. et 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.
- Knapp, G. 1989. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* **180**: 192-212.
- Kollmus, H., Hentze, M.W. et Hauser, H. 1996. Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA* **2**: 316-323.
- Kolupaeva, V.G., Pestova, T.V. et Hellen, C.U. 2000. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J Virol* **74**: 6242-6250.
- Komurian-Pradel, F., Rajoharison, A., Berland, J.L., Khouri, V., Perret, M., Van Roosmalen, M., Pol, S., Negro, F. et Paranhos-Baccala, G. 2004. Antigenic

- relevance of F protein in chronic hepatitis C virus infection. *Hepatology* **40**: 900-909.
- Kontos, H., Naphthine, 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.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S. et Bhattacharya, T. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**: 1789-1796.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283-292.
- Kozak, M. 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell. Biol.* **9**: 5073-5080.
- Kozak, M. 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **87**: 8301-8305.
- Kozak, M. 1991. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr.* **1**: 111-115.
- Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**: 187-208.
- Kozak, M. et Shatkin, A.J. 1978. Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. *J. Biol. Chem.* **253**: 6568-6577.
- Kurland, C.G. 1992. Translational accuracy and the fitness of bacteria. *Annu Rev Genet* **26**: 29-50.
- Kurotani, A., Kiyotani, K., Kato, A., Shioda, T., Sakai, Y., Mizumoto, K., Yoshida, T. et Nagai, Y. 1998. Sendai virus C proteins are categorically nonessential gene products but silencing their expression severely impairs viral replication and pathogenesis. *Genes Cells* **3**: 111-124.
- Lamarre, D., Anderson, P.C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bos, M., Cameron, D.R., Cartier, M., Cordingley, M.G., Faucher, A.M., Goudreau, N., Kawai, S.H., Kukolj, G., Lagace, L., LaPlante, S.R., Narjes, H., Poupart, M.A., Rancourt, J., Sentjens, R.E., St George, R., Simoneau, B., Steinmann, G., Thibeault,

- D., Tsantrizos, Y.S., Weldon, S.M., Yong, C.L. et Llinas-Brunet, M. 2003. An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**: 186-189.
- Larder, B.A., Darby, G. et Richman, D.D. 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* **243**: 1731-1734.
- Large, M.K., Kittlesen, D.J. et Hahn, Y.S. 1999. Suppression of host immune response by the core protein of hepatitis C virus: possible implications for hepatitis C virus persistence. *J Immunol* **162**: 931-938.
- Latorre, P., Kolakofsky, D. et Curran, J. 1998. Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol. Cell. Biol.* **18**: 5021-5031.
- Lau, J.Y., Tam, R.C., Liang, T.J. et Hong, Z. 2002. Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. *Hepatology* **35**: 1002-1009.
- Le Quesne, J.P., Stoneley, M., Fraser, G.A. et Willis, A.E. 2001. Derivation of a structural model for the c-myc IRES. *J. Mol. Biol.* **310**: 111-126.
- Le, S.Y., Shapiro, B.A., Chen, J.H., Nussinov, R. et Maizel, J.V. 1991. RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet. Anal. Tech. Appl.* **8**: 191-205.
- Lefkowitz, J.H., Schiff, E.R., Davis, G.L., Perrillo, R.P., Lindsay, K., Bodenheimer, H.C., Jr., Balart, L.A., Ortego, T.J., Payne, J. et Dienstag, J.L. 1993. Pathological diagnosis of chronic hepatitis C: a multicenter comparative study with chronic hepatitis B. The Hepatitis Interventional Therapy Group. *Gastroenterology* **104**: 595-603.
- 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.
- Levin, M.E., Hendrix, R.W. et Casjens, S.R. 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol* **234**: 124-139.

- Li, D., Lott, W.B., Martyn, J., Haqshenas, G. et Gowans, E.J. 2004. Differential effects on the hepatitis C virus (HCV) internal ribosome entry site by vitamin B12 and the HCV core protein. *J. Virol.* **78**: 12075-12081.
- Li, Z., Stahl, G. et Farabaugh, P.J. 2001. Programmed +1 frameshifting stimulated by complementarity between a downstream mRNA sequence and an error-correcting region of rRNA. *RNA* **7**: 275-284.
- Lim, A.C. et Barton, J.K. 1998. Rh(phen)<sub>2</sub>phi<sub>3</sub><sup>+</sup> as a shape-selective probe of triple helices. *Biochemistry* **37**: 9138-9146.
- Lindenbach, B.D., Evans, M.J., Syder, A.J., Wolk, B., Tellinghuisen, T.L., Liu, C.C., Maruyama, T., Hynes, R.O., Burton, D.R., McKeating, J.A. et Rice, C.M. 2005. Complete replication of hepatitis C virus in cell culture. *Science* **309**: 623-626.
- 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.
- Lo, S.Y., Masiarz, F., Hwang, S.B., Lai, M.M. et Ou, J.H. 1995. Differential subcellular localization of hepatitis C virus core gene products. *Virology* **213**: 455-461.
- Lo, S.Y., Selby, M., Tong, M. et Ou, J.H. 1994. Comparative studies of the core gene products of two different hepatitis C virus isolates: two alternative forms determined by a single amino acid substitution. *Virology* **199**: 124-131.
- Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. et Bartenschlager, R. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**: 110-113.
- Lopinski, J.D., Dinman, J.D. et Bruenn, J.A. 2000. Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Mol. Cell. Biol.* **20**: 1095-1103.
- Los Alamos National Laboratory. 2002. A compilation and analysis of nucleic acid and amino acid sequences.: Los Alamos National Laboratory.
- Lott, W.B., Takyar, S.S., Tuppen, J., Crawford, D.H., Harrison, M., Sloots, T.P. et Gowans, E.J. 2001. Vitamin B12 and hepatitis C: molecular biology and human pathology. *Proc. Natl. Acad. Sci. USA* **98**: 4916-4921.

- Louis, J.M., Clore, G.M. et Gronenborn, A.M. 1999. Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nat. Struct. Biol.* **6**: 868-875.
- Luk, K.C., Kaptue, L., Zekeng, L., Soriano, V., Gurtler, L., Devare, S.G., Schochetman, G. et Hackett, J., Jr. 2001. Naturally occurring sequence polymorphisms within HIV type 1 group O protease. *AIDS Res. Hum. Retroviruses* **17**: 1555-1561.
- Luo, G., Hamatake, R.K., Mathis, D.M., Racela, J., Rigat, K.L., Lemm, J. et Colonna, R.J. 2000. De novo initiation of RNA synthesis by the RNA-dependent RNA polymerase (NS5B) of hepatitis C virus. *J. Virol.* **74**: 851-863.
- Macejak, D.G., Jensen, K.L., Jamison, S.F., Domenico, K., Roberts, E.C., Chaudhary, N., von Carlowitz, I., Bellon, L., Tong, M.J., Conrad, A., Pavco, P.A. et Blatt, L.M. 2000. Inhibition of hepatitis C virus (HCV)-RNA-dependent translation and replication of a chimeric HCV poliovirus using synthetic stabilized ribozymes. *Hepatology* **31**: 769-776.
- 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.
- Marczinke, B., Bloys, A.J., Brown, T.D., Willcocks, M.M., Carter, M.J. et Brierley, I. 1994. The human astrovirus RNA-dependent RNA polymerase coding region is expressed by ribosomal frameshifting. *J. Virol.* **68**: 5588-5595.
- Marczinke, B., Fisher, R., Vidakovic, M., Bloys, A.J. et Brierley, I. 1998. Secondary structure and mutational analysis of the ribosomal frameshift signal of rous sarcoma virus. *J. Mol. Biol.* **284**: 205-225.
- Marquez, V., Wilson, D.N., Tate, W.P., Triana-Alonso, F. et Nierhaus, K.H. 2004. Maintaining the ribosomal reading frame: the influence of the E site during translational regulation of release factor 2. *Cell* **118**: 45-55.
- Martin, A., Bodola, F., Sangar, D.V., Goettge, K., Popov, V., Rijnbrand, R., Lanford, R.E. et Lemon, S.M. 2003. Chronic hepatitis associated with GB virus B persistence in a tamarin after intrahepatic inoculation of synthetic viral RNA. *Proc. Natl. Acad. Sci. USA* **100**: 9962-9967.

- Martinez-Salas, E. 1999. Internal ribosome entry site biology and its use in expression vectors. *Curr. Opin. Biotechnol.* **10**: 458-464.
- Martinez-Salas, E., Ramos, R., Lafuente, E. et Lopez de Quinto, S. 2001. Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J. Gen. Virol.* **82**: 973-984.
- Mathews, D.H., Sabina, J., Zuker, M. et Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**: 911-940.
- Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J.F., Gesteland, R.F. et Hayashi, S. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**: 51-60.
- McCutchan, F.E. 2000. Understanding the genetic diversity of HIV-1. *AIDS* **14 Suppl 3**: S31-44.
- McQuade, T.J., Tomasselli, A.G., Liu, L., Karacostas, V., Moss, B., Sawyer, T.K., Heinrikson, R.L. et Tarpley, W.G. 1990. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* **247**: 454-456.
- Meerovitch, K., Svitkin, Y.V., Lee, H.S., Lejbkowitz, F., Kenan, D.J., Chan, E.K., Agol, V.I., Keene, J.D. et Sonenberg, N. 1993. La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *J. Virol.* **67**: 3798-3807.
- Merluzzi, V.J., Hargrave, K.D., Labadia, M., Grozinger, K., Skoog, M., Wu, J.C., Shih, C.K., Eckner, K., Hattox, S., Adams, J. et et al. 1990. Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor. *Science* **250**: 1411-1413.
- Merrick, W.C. 2004. Cap-dependent and cap-independent translation in eukaryotic systems. *Gene* **332**: 1-11.
- Merrick, W.C. et Nyborg, J. 2000. The protein biosynthesis elongation cycle. In: Sonenberg N, Hershey JWB, Mathews M, eds. *Translational control of gene expression*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 89-125.

- Michiels, P.J., Versleijen, A.A., Verlaan, P.W., Pleij, C.W., Hilbers, C.W. et Heus, H.A. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**: 1109-1123.
- Mitsuya, H. et Broder, S. 1987. Strategies for antiviral therapy in AIDS. *Nature* **325**: 773-778.
- Moore, J.P., Parren, P.W. et Burton, D.R. 2001. Genetic subtypes, humoral immunity, and human immunodeficiency virus type 1 vaccine development. *J. Virol.* **75**: 5721-5729.
- Moriya, K., Fujie, H., Shintani, Y., Yotsuyanagi, H., Tsutsumi, T., Ishibashi, K., Matsuura, Y., Kimura, S., Miyamura, T. et Koike, K. 1998. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat. Med.* **4**: 1065-1067.
- Morris, D.K. et Lundblad, V. 1997. Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr Biol* **7**: 969-976.
- Namy, O., Rousset, J.P., Naphine, S. et Brierley, I. 2004. Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell* **13**: 157-168.
- Naphine, 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.
- Nixon, P.L., Rangan, A., Kim, Y.G., Rich, A., Hoffman, D.W., Hennig, M. et Giedroc, D.P. 2002. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* **322**: 621-633.
- Noller, H.F., Yusupov, M.M., Yusupova, G.Z., Baucom, A. et Cate, J.H. 2002. Translocation of tRNA during protein synthesis. *FEBS Lett* **514**: 11-16.
- Novina, C.D., Murray, M.F., Dykxhoorn, D.M., Beresford, P.J., Riess, J., Lee, S.K., Collman, R.G., Lieberman, J., Shankar, P. et Sharp, P.A. 2002. siRNA-directed inhibition of HIV-1 infection. *Nat. Med.* **8**: 681-686.
- Nowakowski, J. et Tinoco Jr, I. 1999. RNA structure in solution. In: Press OU, ed. *Oxford Handbook of Nucleic Acid Structure*. New-York: Oxford University Press. pp 567-602.



- O'Brien, W.A. 2000. Resistance against reverse transcriptase inhibitors. *Clin. Infect. Dis.* **30 Suppl 2**: S185-192.
- Odon, O.W., Kramer, G., Henderson, A.B., Pinphanichakarn, P. et Hardesty, B. 1978. GTP hydrolysis during methionyl-tRNA<sup>f</sup> binding to 40 S ribosomal subunits and the site of edeine inhibition. *J. Biol. Chem.* **253**: 1807-1816.
- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P. et Ramakrishnan, V. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**: 897-902.
- Ogle, J.M. et Ramakrishnan, V. 2004. Structural Insights into Translational Fidelity. *Annu. Rev. Biochem.*
- Ohlmann, T., Rau, M., Pain, V.M. et Morley, S.J. 1996. The C-terminal domain of eukaryotic protein synthesis initiation factor (eIF) 4G is sufficient to support cap-independent translation in the absence of eIF4E. *EMBO J.* **15**: 1371-1382.
- Olsthoorn, R.C., Laurs, M., Sohet, F., Hilbers, C.W., Heus, H.A. et Pleij, C.W. 2004. Novel application of sRNA: stimulation of ribosomal frameshifting. *RNA* **10**: 1702-1703.
- Op De Beeck, A., Montserret, R., Duvet, S., Cocquerel, L., Cacan, R., Barberot, B., Le Maire, M., Penin, F. et Dubuisson, J. 2000. The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J. Biol. Chem.* **275**: 31428-31437.
- Pallaoro, M., Lahm, A., Biasiol, G., Brunetti, M., Nardella, C., Orsatti, L., Bonelli, F., Orru, S., Narjes, F. et Steinkuhler, C. 2001. Characterization of the hepatitis C virus NS2/3 processing reaction by using a purified precursor protein. *J. Virol.* **75**: 9939-9946.
- Park, J. et 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.
- Parkin, N.T., Chamorro, M. et 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.

- Pavlovic, D., Neville, D.C., Argaud, O., Blumberg, B., Dwek, R.A., Fischer, W.B. et Zitzmann, N. 2003. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proc. Natl. Acad. Sci. USA* **100**: 6104-6108.
- Peabody, D.S. 1989. Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**: 5031-5035.
- Pearlman, D.A. 1995. AMBER 4.1. San-Francisco: University of California.
- Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J.P., Badombena, W., Luo, N., Vanden Haesevelde, M. et Delaporte, E. 1997. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **11**: 493-498.
- Peeters, M., Liegeois, F., Torimiro, N., Bourgeois, A., Mpoudi, E., Vergne, L., Saman, E., Delaporte, E. et Saragosti, S. 1999. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J. Virol.* **73**: 7368-7375.
- Peeters, M. et Sharp, P.M. 2000. Genetic diversity of HIV-1: the moving target. *AIDS* **14 Suppl 3**: S129-140.
- Pelletier, J. et Sonenberg, N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**: 320-325.
- Pestova, T.V., Borukhov, S.I. et Hellen, C.U. 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**: 854-859.
- Pestova, T.V. et Hellen, C.U. 2001. Functions of eukaryotic factors in initiation of translation. *Cold Spring Harb. Symp. Quant. Biol.* **66**: 389-396.
- Pestova, T.V. et Kolupaeva, V.G. 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **16**: 2906-2922.
- Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E. et Hellen, C.U. 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**: 332-335.

- Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G. et Abrignani, S. 1998. Binding of hepatitis C virus to CD81. *Science* **282**: 938-941.
- Plant, E.P. et Dinman, J.D. 2005. Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res.* **33**: 1825-1833.
- Plant, E.P., Jacobs, K.L., Harger, J.W., Meskauskas, A., Jacobs, J.L., Baxter, J.L., Petrov, A.N. et Dinman, J.D. 2003. The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA* **9**: 168-174.
- Poulin, F. et Sonenberg, N. 2003. Mechanism of translation initiation in eukaryotes. In: Brakier-Gingras L, ed. *Translation mechanisms*. Georgetown, TX: Landes Bioscience. pp 280-297.
- Prats, A.C., Vagner, S., Prats, H. et Amalric, F. 1992. cis-acting elements involved in the alternative translation initiation process of human basic fibroblast growth factor mRNA. *Mol. Cell. Biol.* **12**: 4796-4805.
- Preiss, T. et Hentze, M.W. 2003. Starting the protein synthesis machine: eukaryotic translation initiation. *Bioessays* **25**: 1201-1211.
- Ramos, F.D., Carrasco, M., Doyle, T. et Brierley, I. 2004. Programmed -1 ribosomal frameshifting in the SARS coronavirus. *Biochem Soc Trans* **32**: 1081-1083.
- Randall, G., Grakoui, A. et Rice, C.M. 2003. Clearance of replicating hepatitis C virus replicon RNAs in cell culture by small interfering RNAs. *Proc. Natl. Acad. Sci. USA* **100**: 235-240.
- Ray, R.B., Lagging, L.M., Meyer, K. et Ray, R. 1996. Hepatitis C virus core protein cooperates with ras and transforms primary rat embryo fibroblasts to tumorigenic phenotype. *J. Virol.* **70**: 4438-4443.
- Reed, K.E. et Rice, C.M. 2000. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. *Curr. Top. Microbiol. Immunol.* **242**: 55-84.
- Reil, H. et Hauser, H. 1990. Test system for determination of HIV-1 frameshifting efficiency in animal cells. *Biochim. Biophys. Acta* **1050**: 288-292.

- Reil, H., Kollmus, H., Weidle, U.H. et Hauser, H. 1993. A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J. Virol.* **67**: 5579-5584.
- Reynolds, J.E., Kaminski, A., Carroll, A.R., Clarke, B.E., Rowlands, D.J. et Jackson, R.J. 1996. Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA* **2**: 867-878.
- Reynolds, J.E., Kaminski, A., Kettinen, H.J., Grace, K., Clarke, B.E., Carroll, A.R., Rowlands, D.J. et Jackson, R.J. 1995. Unique features of internal initiation of hepatitis C virus RNA translation. *EMBO J.* **14**: 6010-6020.
- Richman, D., Shih, C.K., Lowy, I., Rose, J., Prodanovich, P., Goff, S. et Griffin, J. 1991. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proc. Natl. Acad. Sci. USA* **88**: 11241-11245.
- Rijnbrand, R.C., Abbink, T.E., Haasnoot, P.C., Spaan, W.J. et Bredenbeek, P.J. 1996. The influence of AUG codons in the hepatitis C virus 5' nontranslated region on translation and mapping of the translation initiation window. *Virology* **226**: 47-56.
- Rijnbrand, R.C. et Lemon, S.M. 2000. Internal ribosome entry site-mediated translation in hepatitis C virus replication. *Curr. Top. Microbiol. Immunol.* **242**: 85-116.
- Roberts, J.D., Bebenek, K. et Kunkel, T.A. 1988. The accuracy of reverse transcriptase from HIV-1. *Science* **242**: 1171-1173.
- Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S. et Korber, B. 2000. HIV-1 nomenclature proposal. *Science* **288**: 55-56.
- Rogers, G.W., Jr., Edelman, G.M. et Mauro, V.P. 2004. Differential utilization of upstream AUGs in the beta-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc. Natl. Acad. Sci. USA* **101**: 2794-2799.
- Rogers, G.W., Jr., Richter, N.J., Lima, W.F. et Merrick, W.C. 2001. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **276**: 30914-30922.

- Rom, E. et Kahana, C. 1994. Polyamines regulate the expression of ornithine decarboxylase antizyme in vitro by inducing ribosomal frame-shifting. *Proc Natl Acad Sci U S A* **91**: 3959-3963.
- Rose, K.M., Marin, M., Kozak, S.L. et Kabat, D. 2004. The viral infectivity factor (Vif) of HIV-1 unveiled. *Trends Mol. Med.* **10**: 291-297.
- Rosenberg, S. 2001. Recent advances in the molecular biology of hepatitis C virus. *J. Mol. Biol.* **313**: 451-464.
- Roussel, J., Pillez, A., Montpellier, C., Duverlie, G., Cahour, A., Dubuisson, J. et Wychowski, C. 2003. Characterization of the expression of the hepatitis C virus F protein. *J. Gen. Virol.* **84**: 1751-1759.
- Ryabova, L.A. et Hohn, T. 2000. Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. *Genes Dev.* **14**: 817-829.
- Ryabova, L.A., Pooggin, M.M. et Hohn, T. 2002. Viral strategies of translation initiation: ribosomal shunt and reinitiation. *Prog. Nucleic Acid Res. Mol. Biol.* **72**: 1-39.
- Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q.L., Houghton, M. et Kuo, G. 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* **87**: 6547-6549.
- Salemi, M., Strimmer, K., Hall, W.W., Duffy, M., Delaporte, E., Mboup, S., Peeters, M. et Vandamme, A.M. 2001. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *Faseb J* **15**: 276-278.
- Sarnow, P. 2003. Viral internal ribosome entry site elements: novel ribosome-RNA complexes and roles in viral pathogenesis. *J. Virol.* **77**: 2801-2806.
- Sasaki, J. et Nakashima, N. 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc. Natl. Acad. Sci. USA* **97**: 1512-1515.
- Sato, K., Okamoto, H., Aihara, S., Hoshi, Y., Tanaka, T. et Mishiro, S. 1993. Demonstration of sugar moiety on the surface of hepatitis C virions recovered from the circulation of infected humans. *Virology* **196**: 354-357.

- Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. et Vitelli, A. 2002. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* **21**: 5017-5025.
- Scheuer, P.J., Ashrafzadeh, P., Sherlock, S., Brown, D. et Dusheiko, G.M. 1992. The pathology of hepatitis C. *Hepatology* **15**: 567-571.
- Schwab, S.R., Shugart, J.A., Horng, T., Malarkannan, S. et Shastri, N. 2004. Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* **2**: 1774-1784.
- Sen, N., Cao, F. et Tavis, J.E. 2004. Translation of duck hepatitis B virus reverse transcriptase by ribosomal shunting. *J. Virol.* **78**: 11751-11757.
- Sharp, P.M. 2002. Origins of human virus diversity. *Cell* **108**: 305-312.
- Sharp, P.M., Bailes, E., Gao, F., Beer, B.E., Hirsch, V.M. et Hahn, B.H. 2000. Origins and evolution of AIDS viruses: estimating the time-scale. *Biochem Soc Trans* **28**: 275-282.
- Shehu-Xhilaga, M., Crowe, S.M. 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.
- Shen, L.X. et 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.
- Shibuya, N., Nishiyama, T., Kanamori, Y., Saito, H. et Nakashima, N. 2003. Conditional rather than absolute requirements of the capsid coding sequence for initiation of methionine-independent translation in *Plautia stali* intestine virus. *J. Virol.* **77**: 12002-12010.
- Shigemoto, K., Brennan, J., Walls, E., Watson, C.J., Stott, D., Rigby, P.W. et 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.
- Shimoike, T., Mimori, S., Tani, H., Matsuura, Y. et Miyamura, T. 1999. Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J. Virol.* **73**: 9718-9725.

- Simon, F., Mauciere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F. et Brun-Vezinet, F. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**: 1032-1037.
- Simons, J.N., Pilot-Matias, T.J., Leary, T.P., Dawson, G.J., Desai, S.M., Schlauder, G.G., Muerhoff, A.S., Erker, J.C., Buijk, S.L., Chalmers, M.L. et et al. 1995. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc. Natl. Acad. Sci. USA* **92**: 3401-3405.
- Sizova, D.V., Kolupaeva, V.G., Pestova, T.V., Shatsky, I.N. et Hellen, C.U. 1998. Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J Virol* **72**: 4775-4782.
- Somogyi, P., Jenner, A.J., Brierley, I. et Inglis, S.C. 1993. Ribosomal pausing during translation of an RNA pseudoknot. *Mol. Cell. Biol.* **13**: 6931-6940.
- Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G. et Frank, J. 2001a. Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions. *Cell* **107**: 373-386.
- Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A. et Frank, J. 2001b. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291**: 1959-1962.
- Strizki, J.M., Xu, S., Wagner, N.E., Wojcik, L., Liu, J., Hou, Y., Endres, M., Palani, A., Shapiro, S., Clader, J.W., Greenlee, W.J., Tagat, J.R., McCombie, S., Cox, K., Fawzi, A.B., Chou, C.C., Pugliese-Sivo, C., Davies, L., Moreno, M.E., Ho, D.D., Trkola, A., Stoddart, C.A., Moore, J.P., Reyes, G.R. et Baroudy, B.M. 2001. SCH-C (SCH 351125), an orally bioavailable, small molecule antagonist of the chemokine receptor CCR5, is a potent inhibitor of HIV-1 infection in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* **98**: 12718-12723.
- Su, L., Chen, L., Egli, M., Berger, J.M. et Rich, A. 1999. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol.* **6**: 285-292.

- Svitkin, Y.V., Meerovitch, K., Lee, H.S., Dholakia, J.N., Kenan, D.J., Agol, V.I. et Sonenberg, N. 1994. Internal translation initiation on poliovirus RNA: further characterization of La function in poliovirus translation in vitro. *J. Virol.* **68**: 1544-1550.
- Szer, W. et Kurylo-Borowska, Z. 1970. Effect of edeine on aminoacyl-tRNA binding to ribosomes and its relationship to ribosomal binding sites. *Biochim. Biophys. Acta* **224**: 477-486.
- Takehisa, J., Zekeng, L., Ido, E., Yamaguchi-Kabata, Y., Mboudjeka, I., Harada, Y., Miura, T., Kaptu, L. et Hayami, M. 1999. Human immunodeficiency virus type 1 intergroup (M/O) recombination in cameroon. *J. Virol.* **73**: 6810-6820.
- Tan, S.L., Pause, A., Shi, Y. et Sonenberg, N. 2002. Hepatitis C therapeutics: current status and emerging strategies. *Nat Rev Drug Discov* **1**: 867-881.
- Tate, W.P., Poole, E.S., Horsfield, J.A., Mannering, S.A., Brown, C.M., Moffat, J.G., Dalphin, M.E., McCaughan, K.K., Major, L.L. et Wilson, D.N. 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.
- Tellinghuisen, T.L., Marcotrigiano, J., Gorbalenya, A.E. et Rice, C.M. 2004. The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem.* **279**: 48576-48587.
- 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, E.B., Verlaan, P.W. et Pleij, C.W. 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.



- Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C. et Vagner, S. 2003. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell.* **95**: 169-178.
- Tsuchihashi, Z. et 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.
- Tu, C., Tzeng, T.H. et Bruenn, J.A. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc. Natl. Acad. Sci. USA* **89**: 8636-8640.
- Valasek, L., Nielsen, K.H., Zhang, F., Fekete, C.A. et Hinnebusch, A.G. 2004. Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol. Cell. Biol.* **24**: 9437-9455.
- van Eyll, O. et Michiels, T. 2000. Influence of the Theiler's virus L\* protein on macrophage infection, viral persistence, and neurovirulence. *J. Virol.* **74**: 9071-9077.
- van Eyll, O. et Michiels, T. 2002. Non-AUG-initiated internal translation of the L\* protein of Theiler's virus and importance of this protein for viral persistence. *J. Virol.* **76**: 10665-10673.
- Varaklioti, A., Vassilaki, N., Georgopoulou, U. et Mavromara, P. 2002. Alternate translation occurs within the core coding region of the hepatitis C viral genome. *J. Biol. Chem.* **277**: 17713-17721.
- Vassilaki, N. et Mavromara, P. 2003. Two alternative translation mechanisms are responsible for the expression of the HCV ARFP/F/core+1 coding open reading frame. *J. Biol. Chem.* **278**: 40503-40513.
- Velazquez-Campoy, A., Todd, M.J., Vega, S. et Freire, E. 2001. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc. Natl. Acad. Sci. USA* **98**: 6062-6067.
- Vergne, L., Peeters, M., Mpoudi-Ngole, E., Bourgeois, A., Liegeois, F., Toure-Kane, C., Mboup, S., Mulanga-Kabeya, C., Saman, E., Jourdan, J., Reynes, J. et Delaporte, E. 2000. Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor

- drug resistance mutations in treatment-naive patients. *J Clin Microbiol* **38**: 3919-3925.
- Vickers, T.A. et Ecker, D.J. 1992. Enhancement of ribosomal frameshifting by oligonucleotides targeted to the HIV gag-pol region. *Nucleic Acids Res.* **20**: 3945-3953.
- Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Krausslich, H.G., Mizokami, M., Bartenschlager, R. et Liang, T.J. 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* **11**: 791-796.
- Walczak, R., Westhof, E., Carbon, P. et Krol, A. 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *Rna* **2**: 367-379.
- Walewski, J.L., Keller, T.R., Stump, D.D. et Branch, A.D. 2001. Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**: 710-721.
- Wang, C., Sarnow, P. et Siddiqui, A. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J. Virol.* **67**: 3338-3344.
- Wang, Y., Wills, N.M., Du, Z., Rangan, A., Atkins, J.F., Gesteland, R.F. et Hoffman, D.W. 2002. Comparative studies of frameshifting and nonframeshifting RNA pseudoknots: a mutational and NMR investigation of pseudoknots derived from the bacteriophage T2 gene 32 mRNA and the retroviral gag-pro frameshift site. *RNA* **8**: 981-996.
- Watashi, K. et Shimotohno, K. 2003. The roles of hepatitis C virus proteins in modulation of cellular functions: a novel action mechanism of the HCV core protein on gene regulation by nuclear hormone receptors. *Cancer Sci* **94**: 937-943.
- Weiss, R.B., Dunn, D.M., Atkins, J.F. et Gesteland, R.F. 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.* **52**: 687-693.
- Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. et Gesteland, R.F. 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, R.B., Huang, W.M. et Dunn, D.M. 1990. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* **62**: 117-126.
- Westerhout, E.M., Ooms, M., Vink, M., Das, A.T. et Berkhout, B. 2005. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res.* **33**: 796-804.
- Willems, M., Metselaar, H.J., Tilanus, H.W., Schalm, S.W. et de Man, R.A. 2002. Liver transplantation and hepatitis C. *Transpl Int* **15**: 61-72.
- Williams, B.R. 1999. PKR; a sentinel kinase for cellular stress. *Oncogene* **18**: 6112-6120.
- Wills, N.M., Gesteland, R.F. et Atkins, J.F. 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, J.E., Pestova, T.V., Hellen, C.U. et Sarnow, P. 2000. Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**: 511-520.
- Witherell, G.W. 2001. ISIS-14803 (Isis Pharmaceuticals). *Curr Opin Investig Drugs* **2**: 1523-1529.
- Xu, Z., Choi, J., Lu, W. et Ou, J.H. 2003. Hepatitis C virus F protein is a short-lived protein associated with the endoplasmic reticulum. *J. Virol.* **77**: 1578-1583.
- Xu, Z., Choi, J., Yen, T.S., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M.J. et Ou, J. 2001. Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *EMBO J.* **20**: 3840-3848.
- Yamaguchi, J., Vallari, A.S., Swanson, P., Bodelle, P., Kaptue, L., Ngansop, C., Zekeng, L., Gurtler, L.G., Devare, S.G. et Brennan, C.A. 2002. Evaluation of HIV type 1 group O isolates: identification of five phylogenetic clusters. *AIDS Res. Hum. Retroviruses* **18**: 269-282.
- Yanagi, M., St Claire, M., Emerson, S.U., Purcell, R.H. et Bukh, J. 1999. In vivo analysis of the 3' untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. *Proc. Natl. Acad. Sci. USA* **96**: 2291-2295.

- Yang, C., Gao, F., Funjungo, P.N., Zekeng, L., van der Groen, G., Pieniazek, D., Schable, C. et Lal, R.B. 2000. Phylogenetic analysis of protease and transmembrane region of HIV type 1 group O. *AIDS Res. Hum. Retroviruses* **16**: 1075-1081.
- Yeh, C.T., Lo, S.Y., Dai, D.I., Tang, J.H., Chu, C.M. et Liaw, Y.F. 2000. Amino acid substitutions in codons 9-11 of hepatitis C virus core protein lead to the synthesis of a short core protein product. *J. Gastroenterol. Hepatol.* **15**: 182-191.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. et Noller, H.F. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**: 883-896.
- Zein, N.N. 2000. Clinical significance of hepatitis C virus genotypes. *Clin. Microbiol. Rev.* **13**: 223-235.
- Zhang, J., Yamada, O., Yoshida, H., Iwai, T. et Araki, H. 2002. Autogenous translational inhibition of core protein: implication for switch from translation to RNA replication in hepatitis C virus. *Virology* **293**: 141-150.
- Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D.R., Wieland, S.F., Uprichard, S.L., Wakita, T. et Chisari, F.V. 2005. Robust hepatitis C virus infection in vitro. *Proc. Natl. Acad. Sci. USA* **102**: 9294-9299.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406-3415.