

École doctorale des sciences de la vie et de la santé  
CNRS UPR 9002 : ARN

# THÈSE

présentée par Romain GASSER

soutenue le 13 juin 2016

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline : Virologie : aspects moléculaires et médicaux

**Comprendre la flexibilité génétique de la protéine  
d'enveloppe de VIH-1 à travers l'étude du réseau de  
coévolution de ses acides aminés**

**THÈSE dirigée par :**

**monsieur Matteo NEGRONI**

Docteur à l'Université de Strasbourg

**RAPPORTEURS :**

**monsieur Ali AMARA**

**madame Martine BRAIBANT**

Docteur à l'université Paris DIDEROT

Docteur à l'Université RABELAIS de Tours

---

**AUTRES MEMBRES DU JURY :**

**madame Christiane MOOG**

Docteur à l'Université de Strasbourg



*« Hope is the first step on the road to disappointment »*

*C.S. Goto*



# Sommaire

Introduction.....	11
Avant-propos.....	11
1. Classification du VIH.....	14
1.1. VIH-1 et VIH-2.....	14
1.2. Classification du VIH-1 groupe M.....	16
1.3. CRF et URF.....	18
1.4. Classification des VIH par rapport aux autres virus.....	19
1.4.1. Retroviridae.....	19
1.4.2. Orthoretrovirinae.....	20
1.4.3. Lentivirus.....	20
2. Organisation du génome des VIH.....	25
2.1. Structure des LTR.....	25
2.1.1. Région U3.....	25
2.1.2. Région R et région U5.....	28
2.2. Organisation génomique en exons et séquences épissées.....	29
2.3. Sites donneurs et accepteurs d'épissages.....	32
3. Cycle viral.....	35
3.1. Phase tardive.....	35
3.1.1. Expression du génome viral.....	35
3.1.2. Tat.....	35
3.1.3. Rev.....	37
3.1.4. Nef.....	39
3.1.5. Vpr.....	41
3.1.6. Expression de la polyprotéine Gag-Pol.....	42
3.1.7. Formation des nouveaux virions.....	45
3.1.7.1. Sélection de l'ARN génomique viral.....	45
3.1.7.2. Assemblage du virion et bourgeonnement.....	47
3.1.8. Vpu.....	49
3.1.9. Maturation du virion.....	51
3.1.10. Phase virion.....	55
3.1.11. Env.....	56
3.1.11.1 Expression, synthèse et maturation de la protéine d'enveloppe.....	56

3.1.11.2	Structure des protéines d'enveloppe .....	58
	Considération générales .....	58
	Structure de la gp120 .....	59
	Structure de la gp41 .....	61
3.1.11.3.	Recrutement de la protéine Env lors de la viriogénèse .....	61
3.1.11.4.	Système « retromer » .....	63
3.2.	Phase Précoce .....	65
3.2.1.	Entrée virale .....	65
3.2.1.1.	Liaison à CD4 .....	65
3.2.1.2.	Liaison au corécepteur .....	67
	Existence de différents corécepteurs et conséquences pour le VIH .....	67
	Liaison à CCR5 .....	69
3.2.1.3.	Fusion des membranes cellulaires et virales .....	71
	Structure native de la gp41 .....	71
	Activation de la gp41 .....	73
	Internalisation dans un endosome .....	74
3.2.2.	Libération de la capsid virale dans le cytoplasme .....	75
3.2.3.	Transcriptase inverse et transcription inverse .....	77
3.2.4.	Vif .....	81
3.2.5.	Décapsidation du virion .....	84
3.2.6.	Import nucléaire .....	85
3.2.7.	Intégrase et intégration .....	86
4.	Variabilité génétique du VIH-1 .....	92
4.1.	Origine de la variabilité .....	92
4.1.1.	Fidélité de la RT .....	92
4.1.2.	APOBEC3G .....	92
4.1.3.	Recombinaisons .....	93
4.1.4.	Autres sources de variabilité .....	94
4.2.	Conséquences de la variabilité génétique et réseaux de co-évolution .....	95
5.	Etude des réseaux de coévolution .....	96
6.	Etude de la glycoprotéine d'enveloppe de VIH-1 par ses réseaux de coévolution .....	99
	Premier article .....	105
	Introduction .....	105
	Second article .....	109

Introduction.....	109
Discussion de la thèse .....	113
Conclusion .....	121
Bibliographie .....	123



# Introduction



# Introduction

## Avant-propos

Depuis son apparition, le succès épidémiologique du virus de l'immunodéficience humaine (VIH), a été si important que ce virus est à l'origine de l'une des plus importantes pandémies touchant l'Homme qui n'ait jamais existé. Le VIH s'est répandu sur toute la surface du globe, a infecté au moins 71 millions de personnes et plus de 35 millions de personnes sont décédées des suites du syndrome d'immunodéficience d'acquise (SIDA) causé par ce virus [1]. Outre cette expansion record dans la population humaine c'est la létalité de la maladie qui rend la pandémie de SIDA particulièrement préoccupante, celle-ci avoisine en effet les 100% en absence de traitements.

Néanmoins, ce taux de létalité extrême n'explique pas la réussite épidémiologique impressionnante de ce virus. Au contraire les forts taux de létalité ont toujours plutôt été associés à des pathogènes provoquant des épidémies meurtrières, mais ayant du mal à persister de façon stable dans la population humaine, comme *Yersinia pestis*, bactérie responsable des épidémies de peste ayant frappé l'Europe médiévale, ou le virus Ebola (EBOV) responsable des épidémies de fièvres hémorragique en Afrique centrale contemporaine. Il est en effet très important pour un pathogène de modérer sa létalité car chaque hôte qui meurt est un propagateur de la maladie qui disparaît. Ainsi, ce qui explique pour une grande part l'efficacité avec laquelle le VIH s'est répandu dans la population humaine c'est avant tout la capacité de ce virus à persister de façon productive chez son hôte, des années durant, tout en ne provoquant l'apparition que de très peu de symptômes. Ainsi, même si l'hôte du VIH finit inéluctablement par mourir, le génome parasite dispose d'une quantité de temps on ne peut plus confortable pour être transmis avec succès à de nouveaux individus.

Cependant, cette information soulève une nouvelle question. Comment ce virus peut-il survivre chez son hôte pendant plusieurs années alors que *Homo sapiens* possède un système immunitaire adaptatif capable de cibler spécifiquement n'importe quelle molécule étrangère et de la neutraliser [2]? L'apparition de ce système immunitaire coïncide avec l'apparition du clade des vertébrés à mâchoires, les gnathostomes, il y a plus de 400 millions d'années, ce qui signifie que ce système a parcouru un long chemin évolutif et qu'il est hautement complexe et efficace. De nombreuses propriétés du virus doivent donc agir en synergie pour expliquer la survie du VIH face à un tel système. La première d'entre elle est la capacité du virus à contrecarrer l'action du système immunitaire de son hôte avec une grande performance. En effet, d'une part, les cellules infectées et détruites par le VIH au cours de son cycle sont des acteurs clés de la réponse immunitaire, et d'autre part l'action conjuguée de nombreuses protéines virales parvient à inhiber fortement la réponse

immunitaire de l'hôte. La seconde caractéristique du VIH qui explique son succès, et qui est beaucoup plus lié au sujet de ma thèse, est son impressionnante variabilité génétique. Celle-ci est permise à la fois par le taux de mutation élevé du génome viral et également par la faculté du virus à pouvoir effectuer des recombinaisons entre plusieurs génomes comme j'aurai l'occasion d'en parler beaucoup plus en détail dans la suite de mon introduction. Disposer d'une population très variable génétiquement assure au VIH une excellente aptitude à échapper à toutes les tentatives de cibler un de ses composants pour l'inactiver, que ce soit par le biais du système immunitaire ou même par le biais des chimiothérapies mises au point par l'Homme.

## **Première partie : classification**

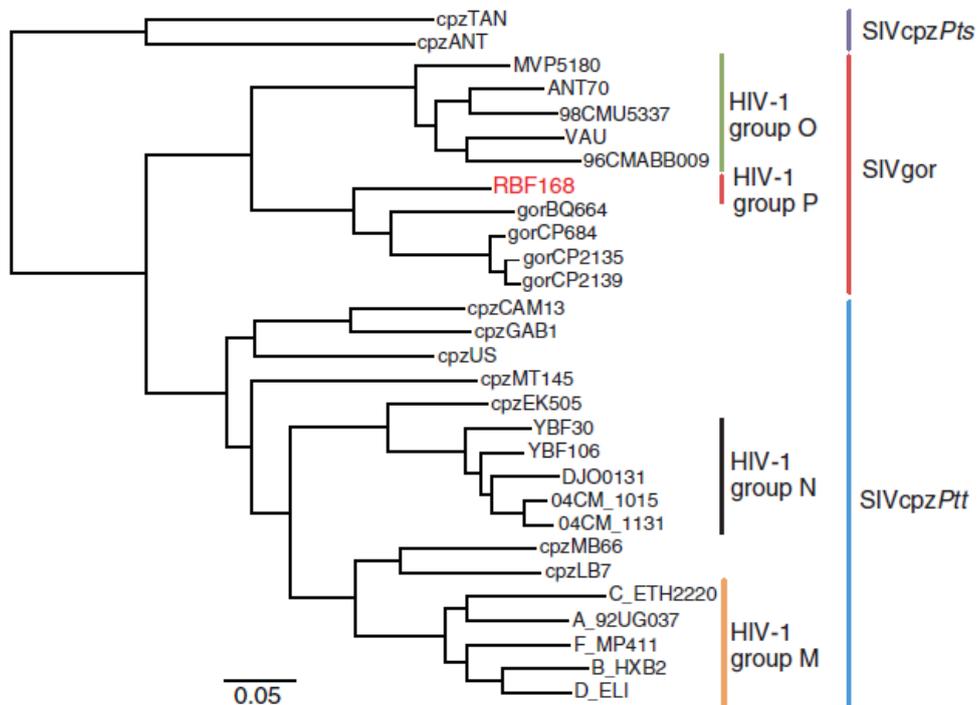
# 1. Classification du VIH

## 1.1. VIH-1 et VIH-2

De l'impressionnante variabilité du VIH résulte une classification complexe ; classification qui n'est pas simplifiée par les origines multiples de ce virus. En effet, il serait presque plus judicieux de parler des virus de l'immunodéficience humaine car ceux-ci sont issus d'espèces virales différentes et sont apparus chez *Homo sapiens* suites à de multiples évènements de passage de l'animal vers l'homme. Tous ces virus sont regroupés sous le terme de VIH par convenance en raison de la similarité de leurs cycles infectieux, des symptômes qu'ils provoquent et de leurs organisations génomiques mais, le terme « VIH » sans plus de précision ne renvoie à aucun ensemble de souches virales ayant une origine phylogénétique commune. Le terme grand public « VIH », que j'ai utilisé jusqu'à présent, englobe en réalité deux types viraux distincts : le VIH-1 et le VIH-2 ; cependant ni le terme de VIH-1 ni celui de VIH-2 ne désignent des espèces virales au sens stricte car ces deux termes regroupent encore une fois des virus ayant des origines phylogénétiques distinctes et sont tous deux, de fait, des groupes polyphylétiques.

Le VIH-1 est constitué de quatre groupes : M (pour « main ») qui est le premier groupe de VIH à avoir été découvert en 1983 par Françoise Barré-Sinoussi [3], O (pour « outlier ») identifié en 1990 [4], N (pour « non-M non-O ») découvert en 1998 [5], et P identifié en 2009 [6]. Ces quatre groupes sont, dans l'état actuel de nos connaissances, monophylétiques car ils contiennent tous une souche virale qui s'est transmise de l'animal à l'Homme ainsi que l'intégralité du lignage de cette souche (figure 1). Les deux groupes M et N dérivent du virus de l'immunodéficience simienne (VIS) infectant les chimpanzés de la sous-espèce *Pan troglodytes troglodytes* : VIS<sub>ScpzPtt</sub>. Ces deux groupes se sont formés suite à deux évènements distincts de transmission du virus VIH à l'Homme et nous savons même aujourd'hui que dans le cas du groupe M, le tout premier foyer épidémique est apparu vers 1920 en République Démocratique du Congo (appelé alors Zaïre) dans les alentours de l'agglomération de Kinshasa (appelée alors Léopoldville) [7]. Les deux groupes restants, O et P, sont issus du VIS<sub>gor</sub> infectant les gorilles de l'espèce *Gorilla gorilla* et sont eux aussi issus de deux évènements de transmission distincts depuis *Gorilla gorilla* vers *Homo sapiens* [8,9].

C'est le VIH-1 groupe M qui est seul responsable de la pandémie mondiale de SIDA, les trois autres groupes étant beaucoup moins répandus. Les cas d'infection à une souche VIH-1 appartenant au groupe N comptent pour moins de 1% des cas d'infection à VIH dans le monde et sont rarement trouvés en dehors de l'Afrique centrale (Cameroun et Gabon principalement). Les infections à VIH-1 du groupe O ou P sont quant à eux extrêmement rares avec moins de 20 cas décrits en 2011 pour les deux groupes confondus et ne sont rencontrés nulle part ailleurs dans le monde qu'au Cameroun.



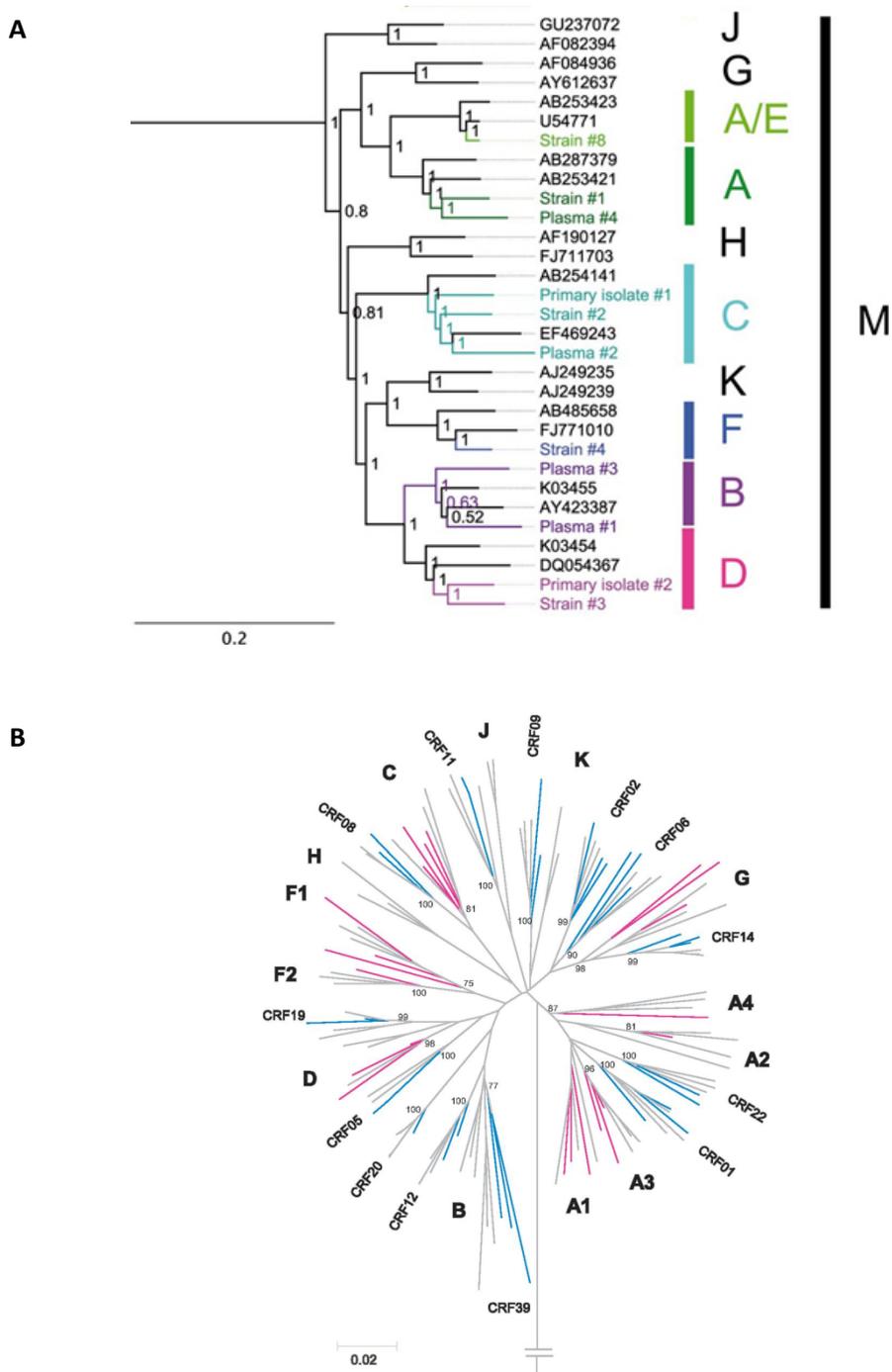
**Figure 1 : Arbre phylogénétique de VIH-1, de SIVcpzPts et de SIVgor adapté de [6]**

Cet arbre phylogénétique permet d'observer les origines multiples des souches VIH qui composent le VIH-1. L'arbre a été construit en utilisant des alignements d'acides aminés concaténés et correspondant au génome partiel de la souche SIVgorBQ664. Les souches virales SIVcpzPts infectant les chimpanzés de l'espèce *Pan troglodytes schweinfurthii* ont été choisies comme groupe externe. Les noms des souches virales étudiées sont indiqués aux extrémités des branches de l'arbre phylogénétique tandis que les groupes et espèces virales auxquels appartiennent ces souches sont indiqués par des barres verticales colorées. La souche RBF168 est indiquée en rouge car elle était la seule souche caractérisée appartenant au groupe P au moment de l'étude. L'unité de distance choisie pour les branches est de 0,05 substitutions nucléotidiques par site.

Concernant le VIH type 2, celui-ci est composé de huit groupes nommés de A jusqu'à H qui, à l'instar des quatre groupes composants le VIH-1, semblent monophylétiques selon les données disponibles actuellement. La différence majeure entre VIH-1 et VIH-2, du point de vue de leur classification, est que là où VIH-1 englobe des souches VIH ayant pour ancêtre des souches SIV assez différentes (SIVcpzPtt et SIVgor), les huit groupes de VIH-2 ont tous pour ancêtre des souches SIVsm infectant le mangabey couronné (*Cercocebus torquatus atys*)[<sup>10</sup>]. Les huit groupes du VIH-2 représentent donc huit événements indépendants d'adaptation d'une souche SIVsm à *Homo sapiens* donnant chacun naissance à un lignage de VIH-2. Aucun des huit groupes du VIH-2 n'a réussi à atteindre le stade pandémique cependant. Au contraire, seuls les sous-types A et B sont endémiques de l'Afrique de l'Ouest avec la Guinée-Bissau comme pays le plus touché par ces deux groupes [<sup>11</sup>]. En plus de la Guinée-Bissau (ancienne possession coloniale portugaise), le sous-type A est rencontré principalement en Sierra Leone et au Liberia, mais ce groupe a également été propagé, quoique de façon minoritaire, au Brésil, aux Etats-Unis et à l'Europe en étant très probablement porté par les anciennes routes commerciales du Portugal. Le sous-type B, quant à lui, est totalement circonscrit à l'Afrique de l'Ouest et couvre une région allant de la Guinée-Bissau jusqu'en Côte d'Ivoire. Tous les autres sous-types du VIH-2 n'ont, jusqu'à présent, jamais été décrits que chez un ou deux patients en Afrique de l'Ouest sans aucune trace de propagation dans la population.

## **1.2. Classification du VIH-1 groupe M**

Le groupe M est constitué de neuf sous-types : A, B, C (environ 50% des souches de VIH répertoriées dans le monde appartiennent à ce sous-type), D, F, G, H, J et K [<sup>12</sup>]. De plus, le séquençage intensif de nombreuses souches virales a conduit à augmenter notre connaissance de la phylogénie de VIH et à subdiviser le sous-type A en sous-sous-types A1, A2, A3, et A4 ainsi qu'à subdiviser le sous-type F en sous-sous-type F1 et F2. Tous ces sous-types sont des taxons monophylétiques résultant à la fois de la sélection d'une souche, par effet fondateur ou par effet goulot d'étranglement, et de l'expansion du lignage de cette souche dont on peut retracer l'histoire naturelle [<sup>13</sup>]. Par exemple les souches VIH du sous-type B sont toutes issues d'une souche ancestrale sélectionnée par effet fondateur suite à son introduction en Haïti depuis l'Afrique de l'Ouest, et dont le lignage s'est ensuite propagé aux Amériques puis à l'Europe de l'Ouest [<sup>14</sup>].



**Figure 2 : Arbres phylogénétiques du VIH-1 groupe M adaptés de [13] et [16]**

(A) Arbre phylogénétique du groupe M du VIH-1 excluant les CRF. L'arbre a été construit en utilisant des alignements de génomes complets d'une collection de souches de VIH-1 groupe M comprenant des souches de références pour chacun des sous-types de ce groupe. Les noms des souches sont indiqués aux extrémités des branches et les groupes ou sous-types viraux ont été représentés par des barres verticales. Les chiffres présents aux nœuds de l'arbre sont les scores réalisés par les nœuds en probabilité Bayésienne. L'unité de distance choisie pour les branches est de 0,2 substitutions nucléotidiques par site.

(B) Arbre phylogénétique de groupe M du VIH-1 incluant les CRF. On peut voir que les CRF sont systématiquement exclues des sous-types viraux, compliquant la classification. Cet arbre a été construit sur la base des séquences des gènes *pol*. Les souches de références VIH-1 sont en gris, les CRF en bleue et les souches non références en rose. Les scores Bootstrap de certains nœuds sont indiqués à proximité de ceux-ci. L'unité de distance choisie pour les branches est de 0,02 substitutions nucléotidiques par site.

### 1.3. CRF et URF

Il est important de préciser ici que de nombreuses autres souches VIH faisant partie du groupe M ne sont classées dans aucun de ces sous-types. En effet, le VIH est connu comme ayant la capacité de générer des virus recombinants par un mécanisme que je détaillerai plus loin dans mon introduction. Lorsqu'un virus recombinant est généré à partir de deux souches n'appartenant pas au même sous-type, le virus ainsi produit se distingue largement des deux souches parentales par sa séquence nucléotidique hybride. Lors de la construction d'arbres phylogénétiques *in silico* ces souches recombinantes sont éjectées en dehors des neufs sous-types décrits à cause de cette divergence de séquence et sont donc classées à part en tant que formes recombinantes (figure 2). Suivant le succès épidémiologique de ces recombinants, ceux-ci ne sont pas désignés de la même façon [15]. Si la souche recombinante n'est retrouvée que chez un patient unique ou des patients qui sont épidémiologiquement liés, c'est-à-dire qui ont été contaminés par la même source, on parle de forme recombinante unique (URF). Si la souche recombinante est retrouvée chez au moins trois patients qui ne sont pas épidémiologiquement liés, dans ce cas cela signifie que le virus recombinant a réussi à se propager dans la population et on parle alors de forme recombinante circulante (CRF). Ces formes recombinantes sont loin d'être négligeables du point de vue épidémiologique car environ 20% des infections à VIH-1 dans le monde sont causées par des CRF [16]. Ainsi, aujourd'hui en 2016, se sont 70 CRF et des dizaines d'URF qui ont été découvertes et décrites dans la littérature scientifique [17].

L'existence de ces CRF complique considérablement le travail de classification des souches du VIH-1 groupe M. C'est notamment à cause des CRF que les sous-types E et I sont absents du groupe M. En effet, lors du séquençage de la première CRF caractérisée, la CRF01\_AE, il est apparu que celle-ci était constituée de séquences caractéristiques du sous-type A et de séquences d'origines inconnues qui ont été attribuées à un nouveau sous-type E qui n'était pas encore décrit. Depuis ce jour aucune souche virale pure appartenant à ce sous-type E hypothétique n'a été découverte, d'où l'absence de ce sous-type dans la classification. Bien que les lettres AE soient toujours présentes dans le nom de cette CRF, les séquences anciennement désignées comme d'origines E sont maintenant désignées comme étant d'origines inconnues [17]. Pour le sous-type I, c'est en réalité la CRF04\_cpx qui lors de sa découverte a été classée dans le sous-type I, créé pour l'occasion. Après des études plus approfondies, il est apparu que cette souche était en réalité une CRF contenant des séquences d'origines multiples (sous-types A, G, H, K et des séquences de provenances inconnues). A la suite de ces études la souche a été reclassée comme CRF et le sous-type I a été retiré de la classification [17].

## 1.4. Classification des VIH par rapport aux autres virus

Même si classer les souches VIH les unes par rapport aux autres est un exercice complexe, la classification des souches VIH par rapport aux autres virus existants est en revanche beaucoup plus aisée. Malgré leur grande diversité et leurs origines multiples, les types et groupes de VIH ont tous en commun une organisation génomique et un cycle viral très similaires, de plus ils induisent des syndromes identiques. Ce sont ces caractéristiques communes qui font que la communauté scientifique a classé l'intégralité des souches VIH à l'intérieur de la même famille et du même genre dans le système de classification du comité international sur la taxonomie des virus (ICTV). En effet toutes les souches VIH sont placées dans la famille des Retroviridae, la sous-famille des Orthoretrovirinae et le genre Lentivirus [18].

### 1.4.1. Retroviridae

La famille des Retroviridae est un groupe de virus enveloppés possédant tous une enzyme particulière : la transcriptase inverse (reverse transcriptase ou RT) qui est une ADN polymérase ARN et ADN dépendante. La RT est présente chez tous les virus de cette famille à cause de leur cycle infectieux assez particulier. L'étape clé du cycle de tous les Retroviridae est l'intégration de leur génome sous forme d'ADN dans le génome de leur hôte, qui est quant à lui toujours un Eutélostome (vertébré à mâchoire et à squelette osseux) du moins dans l'état actuel de nos connaissances. Cette fusion intime du virus avec son hôte permet au génome viral d'être exprimé par la machinerie de transcription/traduction eucaryote classique et surtout de devenir un élément indissociable de la cellule qu'il parasite. Les Eutélostomes ne possèdent en effet aucune machinerie enzymatique capable de séparer le génome viral du leur une fois l'intégration effectuée [19]. Cependant, cela signifie également que la seule solution dont dispose les Retroviridae pour amplifier leurs génomes tout en le dissociant de celui de leurs hôtes est de récupérer les ARN messagers (ARNm) viraux produit par l'ARN polymérase II. C'est ensuite la RT qui permet de générer de nouveaux ADN viraux à partir de ces ARNm. Le cycle viral des Retroviridae est donc indissociable de la présence de la RT. Ce sont la combinaison de cette enzyme et l'intégration du génome ADN viral dans celui de sa cellule hôte qui, ensemble, constituent les deux éléments les plus caractéristiques des Retroviridae. Cependant il faut noter que d'autres caractéristiques sont également partagées par les membres de cette famille comme l'organisation globale du génome par exemple.

### **1.4.2. Orthoretrovirinae**

Les deux sous-familles composant les Retroviridae sont discriminées selon la nature de l'acide nucléique qui est encapsidé dans les particules virales. En effet, tous les Retroviridae utilisent la RT pour produire l'ADN indispensable à leur cycle à partir d'ARNm viraux, mais cette conversion peut avoir lieu avant le bourgeonnement du nouveau virion ou après le bourgeonnement de celui-ci. Si la rétro-transcription a lieu majoritairement avant le bourgeonnement, de l'ADN est alors encapsidé et le virus est classé dans la sous-famille des Spumaretrovirinae [20]. Dans le cas contraire la rétro-transcription a lieu après le bourgeonnement et sachant qu'un virus ne possède pas de métabolisme propre, cela signifie qu'il doit patienter jusqu'à pénétrer à l'intérieur d'une nouvelle cellule pour parasiter son pool de nucléotide et produire l'ADN viral. Dans ce cas-ci c'est majoritairement de l'ARN qui est présent dans la capsid et le virus se classe parmi les Orthoretrovirinae, ce qui est le cas des VIH. Une autre différence majeure sépare les deux sous-familles : les Spumaretrovirinae acquièrent leur enveloppe en bourgeonnant à la membrane du réticulum endoplasmique tandis que les Orthoretrovirinae bourgeonnent à la membrane plasmique des cellules.

### **1.4.3. Lentivirus**

Les six genres classés à l'intérieur des Orthoretrovirinae sont différenciés principalement sur la base de leur organisation génomique, qui varie d'un genre à l'autre même si l'organisation générale reste globalement très similaire. Cette introduction ne s'attardera pas sur les différences subtiles qui existent entre les organisations génomiques des genres viraux appartenant aux Orthoretrovirinae pour se concentrer sur d'autres aspects.

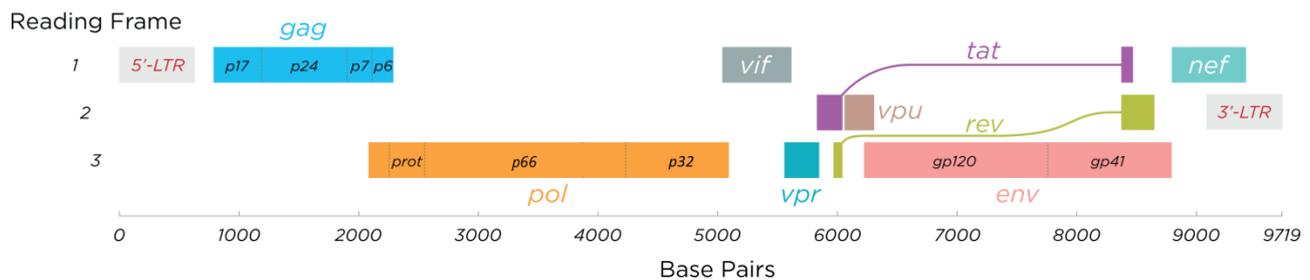
En effet, en plus d'une organisation génomique différente de celle des autres genres, les Lentivirus se distinguent également par les syndromes qu'ils induisent chez leurs hôtes. Le mot « Lentivirus », traduit du latin vers le français, peut grossièrement se comprendre comme « poison lent ». Une des propriétés des Lentivirus, et la raison qui a fait classer les VIH dans ce genre, est donc le temps extrêmement long pendant lequel ces virus arrivent à maintenir une infection asymptomatique chronique de leur hôte [21]. Pour prendre l'exemple du VIH, l'infection se découpe en trois phases que sont la primo-infection, la phase asymptomatique et la phase SIDA. La primo-infection est caractérisée par la rencontre entre l'hôte et le virus, c'est l'étape pendant laquelle l'infection débute et où on assiste à une production massive de particules virales avec une réponse immunitaire réduite. La longue période asymptomatique caractéristique des Lentivirus est la phase pendant laquelle un équilibre dynamique s'est mis en place entre le virus et le système immunitaire ;

même si des virions sont produits en continu et que de nouvelles cellules sont constamment infectées, le système immunitaire parvient tout de même à juguler l'infection sans pouvoir la stopper. Enfin la phase SIDA correspond à l'effondrement final du système immunitaire, victime de la destruction continue des cellules qui le composent, et également de son activation permanente qui le fragilise et l'épuise. En effet, lors de l'infection à VIH, la réplication virale se fait de manière ininterrompue pendant des années et induit une pathologie appelée inflammation chronique, car le système immunitaire adaptatif n'est pas conçu pour pouvoir être activé en continu sur de longues périodes [22].

Un autre signe distinctif des Lentivirus est leur capacité à pouvoir infecter des cellules quiescentes [23]. La stratégie répliquative des Retroviridae passe en effet par l'intégration irréversible de leur matériel génétique dans le génome de leur hôte qui est un Eutélostome et donc un organisme eucaryote. Il faut donc que ces virus soient capables de traverser la membrane nucléaire pour atteindre le génome hôte. Ceci s'accomplit de manière passive pour la majorité des Retroviridae qui profitent simplement de la fragmentation de cette membrane lors des événements de la mitose. Cette particularité implique que la majorité des Retroviridae ne peuvent pas infecter des cellules qui ne se répliquent pas. Les Lentivirus disposent en revanche d'un mécanisme d'import nucléaire actif qui leur permet de s'affranchir de cette contrainte et d'infecter des types cellulaires qui ne se répliquent que très peu.



## **Deuxième partie : organisation génétique**



**Figure 3 : Organisation génomique du provirus VIH-1 adapté de [24]**

Les gènes composants le génome proviral sont représentés par des rectangles colorés, *gag* en bleu, *pol* en orange, *vif* en gris, *vpr* en vert d'eau, *tat* en violet, *rev* en vert, *vpu* en marron, *env* en rose, *nef* en vert clair. Les deux LTR sont également représentés par des rectangles gris clair aux deux extrémités du génome. Une ligne joignant deux rectangles signifie que le gène est codé par deux exons et doit être reconstitué par épissage d'ARNm pour pouvoir être traduit. Dans le cas de *gag*, *pol* et *env* les protéines finales obtenues par maturation du produit de la traduction du gène sont représentées par des pointillés et leurs noms est écrit en noir et en italique.

Les différents éléments du génome sont représentés en ordonné sur trois lignes, chaque ligne correspondant à un cadre de lecture. Le codon ATG du premier gène *gag* sert de référence et définit le cadre de lecture 1. L'axe des abscisses représente la longueur du génome VIH-1 en paire de base et les rectangles représentant les éléments du génome sont positionnés sur cet axe en fonction de l'espace qu'ils occupent sur le génome.

## 2. Organisation du génome des VIH

Après avoir traité la phylogénie et le classement du VIH-1, il est nécessaire de revenir sur quelques éléments essentiels de la structure de son génome avant d'aborder son cycle. Etant donné que le génome de tous les virus de cette famille alterne entre une forme ADN et une forme ARN qui ne sont pas sensiblement identiques, il est important de commencer par préciser que par défaut il sera question dans cette section de l'organisation génomique de la forme ADN<sup>24</sup>.

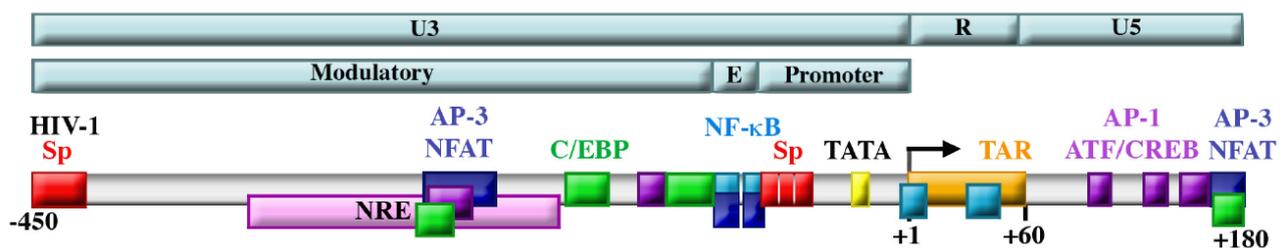
Etant donné que le génome ADN viral fusionne complètement avec celui de son hôte, une fois qu'il fait partie intégrante du génome hôte il est désigné par le terme spécifique de provirus. Cette fusion implique que les génomes viraux et humains sont compatibles entre eux et donc que le génome viral sous forme ADN est obligatoirement linéaire et double brin. De plus, le provirus se soumet à tous les changements que lui impose son hôte, en l'occurrence le génome viral est lié à des histones et compacté sous forme de chromatine et il est répliqué par la machinerie enzymatique de l'hôte lors de chaque cycle cellulaire.

L'ensemble des Retroviridae possèdent une organisation génomique assez similaire, il est donc attendu que les Lentivirus, au nombre desquels figure le VIH-1, présentent cette organisation caractéristique. Or, deux éléments du génome rétroviral retrouvés chez tous les virus de cette famille sont les longues séquences terminales répétées (long terminal repeat ou LTR) présentes aux deux extrémités du provirus (figure 3). La raison principale qui explique que ces éléments génétiques se situent aux extrémités du génome viral est le fait que les LTR contiennent tous les éléments nécessaires à l'expression des gènes du provirus. Comme leur nom le suggère la structure des deux LTR est identique et ils sont chacun constitués, d'amont en aval, d'une région U3, d'une région R et d'une région U5 [25]. Encore une fois il faut noter que cette organisation est celle du provirus, l'ARN viral en revanche n'est constitué dans sa partie amont que de la région R suivie de la région U5 tandis que la partie aval n'est constituée que de la région U3 suivie de la région R. Pour détailler la composition des LTR et du génome, c'est le cas de VIH-1 qui sera étudié ici étant donné qu'il a constitué le sujet de cette thèse.

### 2.1. Structure des LTR

#### 2.1.1. Région U3

La région U3 s'étend sur près de 450 nucléotides et contient à la fois un promoteur ainsi que des éléments modulateurs et enhancers. Le promoteur est long d'un peu moins de 100 nucléotides et est situé juste en amont de la région R. Il est constitué par une TATA box, contenant un site de fixation



**Figure 4 : Organisation d'un LTR de provirus VIH-1 adapté de [28]**

(haut) Les régions et éléments qui constituent un LTR sont représentés par des rectangles bleu clairs. Les trois régions sont U3, R et U5 et les éléments sont l'élément modulateur (Modulatory), enhancer (E) et le promoteur (Promoter)

(bas) Les sites de fixations à quelques facteurs de transcription ou protéines régulatrices sont représentés par des rectangles de différentes couleurs. Les sites de fixation à SP-1 sont représentés en rouge, la TATA box en jaune, les sites de fixation à NF-κB en bleu clair, à NFAT en bleu foncé, à C/EBP en vert. La position de l'élément TAR est indiqué par un rectangle de couleur orange. Les chiffres indiquent la numération des acides nucléiques en prenant le +1 de transcription comme référence. La flèche indique le sens de transcription du génome.

au facteur de transcription TFIIID, et par trois sites de fixation au facteur de transcription SP-1 (specificity protein 1) exprimé en permanence dans les cellules cibles de VIH [26] (figure 3). Ces éléments assurent au provirus une transcription à un niveau faible mais constitutif. L'élément « enhancer » de U3 est situé juste en aval du promoteur et se compose uniquement d'une quarantaine de nucléotides. Cet élément est constitué par des sites de fixation au facteur de transcription NF-κB [27]. Il faut préciser que ce facteur de transcription est exprimé dans les cellules du système immunitaire en réponse à une infection et qu'il favorise l'expression de gènes impliqués dans la prolifération et la différenciation de ces cellules. Il faut également souligner que la capacité du virus à lier NF-κB pour augmenter son niveau de transcription est une propriété remarquable. En effet, au tout début de l'infection à VIH, lorsque le système immunitaire s'active pour tenter d'éliminer le virus cette activation va paradoxalement augmenter de façon drastique la production de particules virales.

Enfin le dernier élément composant la région U3 est son élément modulateur. Cet élément a la capacité d'augmenter ou de diminuer le niveau de transcription du provirus [28]. Le fait de diminuer le niveau de transcription peut sembler curieux au premier abord car il est normalement toujours préférable pour un virus de produire un maximum de nouveaux virions. Cette régulation négative s'explique par le comportement tout à fait singulier du VIH vis-à-vis de son hôte. Encore une fois, contrairement à la majorité des autres virus, le VIH s'intègre de façon durable dans son hôte et s'y réplique des années durant. Comme le VIH dispose de cette capacité de pouvoir s'intégrer de façon irréversible dans les cellules qu'il parasite, faire produire de grosses quantités de virions par les cellules infectées n'est nécessaire qu'en tout début d'infection pour pouvoir infecter un grand nombre de cellules hôtes. Une fois qu'une population suffisante de cellules a ainsi été infectée de façon permanente et que la réponse immunitaire a commencé, la production de virions est régulée négativement grâce à l'élément modulateur de U3 qui sert ici véritablement de senseur. Cette régulation permet au virus de se cacher du système immunitaire et le résultat de cette modulation est la création de réservoirs viraux constitués par des cellules infectées qui ne produisent que peu ou pas de virions.

Cette modulation est assurée par des sites de fixation à de nombreux facteurs de transcription cellulaires dont la présence est indicative de l'état physiologique dans lequel se trouve la cellule. Ces sites permettent ainsi au provirus d'avoir des informations sur son environnement et de réguler son activité en conséquence. La modulation de l'expression du provirus est néanmoins réalisée de façon très fine et est, par conséquent, un processus très complexe qui ne sera pas décrit en détail. Cependant parmi les mécanismes majeurs de modulation il est possible de citer les protéines NFAT (nuclear factor of activated T-cells) qui stimulent la transcription du provirus. Ces protéines sont connues pour être produites lors de l'activation des lymphocytes T, acteurs majeurs de la réponse

immunitaires. La capacité de VIH-1 à lier les protéines NFAT et d'être stimulé par elles, constitue un moyen supplémentaire d'activer la transcription de son génome lorsque sa cellule hôte est elle-même activée et dans l'état physiologique le plus propice à la réplication virale [29,30]. Des autres exemples de facteurs de transcription sont les protéines appartenant à la famille des glissières à leucine (leucine zipper) et plus particulièrement à la famille des C/EBP (CCATT-enhancer-binding proteins). Ces protéines s'activent en réponse à l'interféron  $\beta$  (IFN- $\beta$ ) produit au cours de la réponse immunitaire et peuvent soit agir en tant qu'activateur ou en tant que répresseur, en fonction de la quantité d'IFN-  $\beta$ , permettant ainsi de moduler l'expression du provirus en fonction de la force de la réponse du système immunitaire [31].

### **2.1.2. Région R et région U5**

La région R des LTR est située juste en aval de U3. Cet élément, composé de 60 nucléotides, est délimité dans sa partie amont par le nucléotide appelé +1 qui délimite le début de la transcription, et dans sa partie aval par le signal de polyadénylation (figure 4). C'est cette organisation qui explique que l'ARN génomique viral commence et se termine par la région R, ce qui n'est pas le cas de l'ADN proviral. Le fait que le signal de polyadénylation présent dans le LTR en 5' soit largement ignoré lors de la synthèse de l'ARNm s'explique par un double mécanisme. Le signal de polyadénylation présent dans le LTR amont est inhibé par un complexe ribonucléoprotéique : U1 snRNP [32]. Ce complexe ribonucléoprotéique cellulaire est impliqué dans l'épissage des ARNm or, il se trouve que le provirus, et par extension les ARNm de VIH, sont composés d'exons et de séquences épissées. La fin du premier exon sur un ARNm viral donné contient donc logiquement le premier site donneur d'épissage et lie le complexe U1 snRNP. Or, U1 snRNP a la capacité d'inhiber les sites de polyadénylation de VIH, et sachant qu'un site donneur d'épissage se trouve à proximité du LTR amont mais qu'aucun site donneur ne se trouve à proximité du LTR aval, cela implique que seul le site du LTR amont est inhibé par ce mécanisme.

Le second mécanisme utilisé par VIH consiste à activer le site de polyadénylation du LTR aval. Une séquence particulière se trouvant dans la partie aval des ARNm viraux en cours de synthèse, et baptisée « upstream enhancer element » (USE), permet de recruter le terminateur de transcription CPSF (cleavage and polyadenylation specificity factor) à cette extrémité. Le recrutement de CPSF à proximité du LTR aval permet une meilleure association de ce facteur à l'ARN polymérase II et permet d'améliorer l'efficacité du site de polyadénylation aval [33].

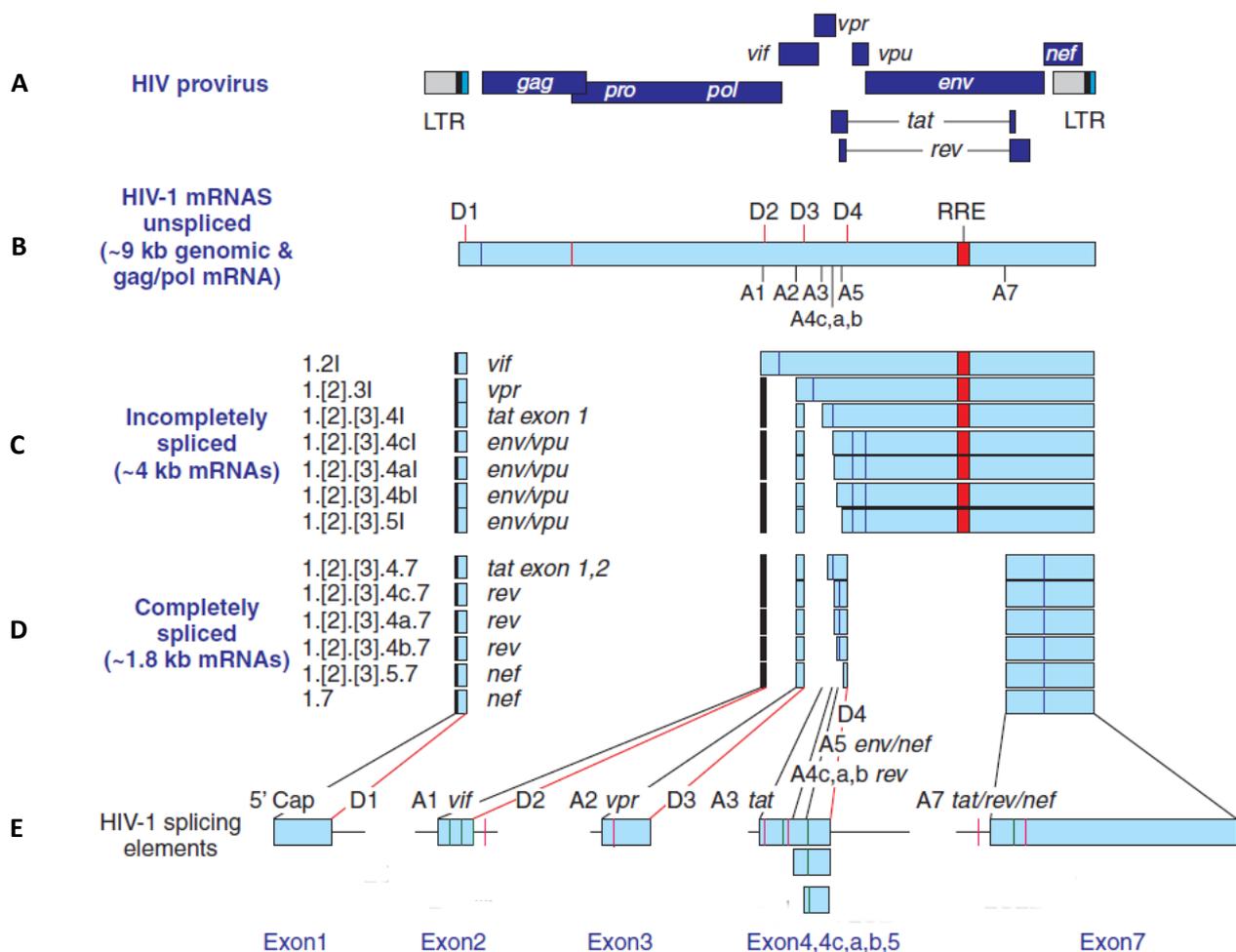
En plus du nucléotide +1 et du site de polyadénylation l'élément R est aussi constitué d'un site de haute importance pour le provirus : le trans-acting responsive element (TAR). Cet élément permet la

fixation de la protéine virale Tat (trans-activator of transcription) qui permet d'augmenter de façon drastique le niveau de transcription du provirus. Il sera à nouveau question de ce mécanisme dans la section consacrée à la protéine Tat.

Enfin le dernier élément constituant les LTR est la région U5, située directement en aval de l'élément R. Le rôle principal de ce site est, à l'instar de U3, d'augmenter le taux d'expression du provirus VIH en recrutant des facteurs de transcription tels que les protéines interferon regulatory factor (IRF) de type 1 et de type 2 entre autres [34] (figure 4). IRF-1 par exemple est une protéine exprimée de façon constitutive dans les cellules immunitaires infectées par VIH-1 mais l'expression de ce facteur de transcription augmente drastiquement lorsque ces cellules s'activent en réponse à la présence d'un pathogène. Ceci constitue une preuve de plus de l'existence du cercle vicieux qui se met en place lors de l'infection à VIH-1, et dans lequel l'activation du système immunitaire en réponse à la présence de ce virus va de pair avec une augmentation de la production de virions par les cellules infectées. Pourtant, le rôle de cette région ne se limite pas à promouvoir l'expression du provirus mais elle est également impliquée dans la rétro-transcription de l'ARN génomique viral en ADN, notamment grâce à un site : le primer activation site (PAS) dont le rôle sera décrit plus en détail dans la section consacrée à la rétro-transcription.

## **2.2. Organisation génomique en exons et séquences épissées**

Une des caractéristiques du génome VIH est qu'il contient de multiples sites donneurs et accepteurs d'épissages [35]. Ce qui rend cette caractéristique des plus intéressantes est que les gènes viraux sont codés à la fois dans les parties épissées et non épissées des ARN viraux (figure 5). Cette organisation singulière implique que le VIH doit obligatoirement posséder un mécanisme lui permettant d'exporter du noyau ses ARNm non épissés, car eux aussi codent pour des protéines virales. Ce mécanisme repose sur la présence d'une séquence particulière appelée RRE (Rev Responsive Element) présente dans une des séquences épissées du génome. Cette séquence RRE est reconnue par la protéine virale Rev lorsqu'elle est présente sur les ARNm partiellement épissés ce qui permet leur export du noyau avant que l'épissage ne se fasse entièrement, mécanisme qui sera plus détaillé dans la section consacrée à la protéine Rev. La question qui viens cependant à l'esprit est : pourquoi le VIH-1 a-t-il recourt à un tel mécanisme alors que respecter l'organisation génétique de son hôte serait a priori plus avantageux ? En effet, chez les eucaryotes les séquences épissées ne sont en général pas des séquences codantes, sauf en cas d'épissage alternatif ce qui n'est pas le cas de VIH-1 [36], alors si VIH-1 ne renfermait pas de séquences épissées codantes cela lui aurait permis l'économie d'un tel mécanisme d'export d'ARNm non épissés.



**Figure 5 : Localisation des sites donneurs et accepteurs d'épissages ainsi que de l'élément RRE dans le génome VIH-1 et principaux ARNm produits par VIH-1 adapté de [45]**

(A) Rappel de l'organisation génétique du provirus. Les régions transcrites du génome proviral sont en bleu foncé, pour les LTR la région U5 est en gris, R est en noir et U3 est en bleu clair, les noms des gènes sont également indiqués en italique.

(B) ARNm complètement non épissé de VIH-1, les positions des sites donneurs (D1 à D4) et accepteurs d'épissages (A1 à A7) retrouvés chez toutes les souches VIH-1 sont positionnés sur l'ARN. L'élément RRE est indiqué en rouge.

(C et D) Principaux ARNm partiellement et entièrement épissés de VIH-1. Les numéros indiquent les exons présents dans l'ARN et obtenus par épissage, les numéros entre crochets signifient que l'exon n'est pas retrouvé systématiquement chez tous les ARNm et la lettre finale I indique que l'ARN est incomplètement épissé. 1.2I signifie par exemple que l'ARNm contient les exons 1 et 2 qui ont été produits par épissage et que cet ARNm est incomplètement épissé. Les gènes qui sont exprimés à partir des ARNm sont indiqués en italiques.

(E) Représentation des exons codés par le génome VIH-1 avec les sites donneurs et accepteurs d'épissages qui les encadrent et le nom du ou des gènes dont ils contiennent la séquence.

Il y a deux raisons majeures, étroitement liées, qui expliquent ce mécanisme singulier. Premièrement, une des principales raisons est le souci de place. En dehors du fait que la taille de l'acide nucléique pouvant être contenu dans une capsid virale est limitée, le génome VIH n'est exprimé que lorsqu'il est sous forme ADN mais il est nécessaire que le génome du VIH passe sous une forme ARN à un stade de son cycle.

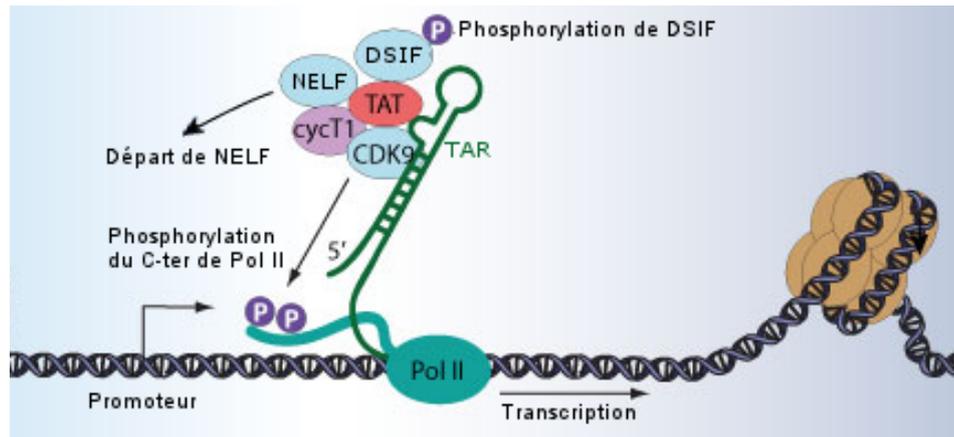
Or, il est connu que les enzymes virales permettant de répliquer les génomes ARN (les ARN polymérases ARN dépendante virales) ou de convertir ces génomes sous forme ADN (les RT) sont très peu fidèles [37]. Pour le virus cela implique de devoir limiter la taille de son génome car plus la molécule d'ARN sera grande, plus forte sera la probabilité pour ces enzymes d'insérer des mutations délétères [38]. Cette contrainte est exacerbée pour les virus infectant les eucaryotes car ces organismes sont totalement étrangers au système opéron, chaque gène est ainsi supposé avoir son propre promoteur, donc autant de place à sacrifier. Le virus aura donc tout intérêt à trouver un moyen de faire exprimer tous les gènes qui lui sont nécessaires et limiter la taille de son génome en même temps. Pour le VIH, la solution à ces contraintes est justement l'encodage de ses gènes dans les séquences normalement épissées de son génome. Ceci lui permet de produire une première série de gènes à partir des exons de son génome, puis, grâce au mécanisme viral d'export des ARNm non épissés de produire une seconde série de gènes à partir des ARNm partiellement épissés ou complètement non-épissés. L'intégralité de ce mécanisme se produit alors avec un seul promoteur et permet un gain de place.

La seconde raison à ce mécanisme particulier d'export, qui rejoint en un certain point la première, vient du besoin de coordonner dans le temps l'expression des gènes viraux sans un mécanisme complexe de promoteurs et de senseurs. Or, si le virus veut échapper efficacement à la réponse immunitaire qui va tenter de détruire la cellule hôte du virus, il est vital pour lui que les protéines régulant son cycle et celles qui inhibent la réponse immunitaire en « masquant » la cellule infectée soient exprimées les premières. Les protéines structurales, qui sont naturellement plus immunogènes, doivent en revanche être impérativement exprimées les dernières, une fois que la réponse immunitaire est entravée par le fait que la cellule hôte n'est pas reconnue efficacement comme étant infectée par les cellules immunitaires. Là encore le système d'export des ARNm non épissés répond à cette contrainte. Les protéines devant être produites en priorité sont codées dans les exons du génome tandis que les autres protéines sont codées dans les parties épissées du génome et ne sont donc produites que lorsque la machinerie virale d'export des ARNm se met en place.

### 2.3. Sites donneurs et accepteurs d'épissages

Au total se sont quatre sites donneurs d'épissages (D1 à D4) et huit sites accepteurs d'épissages (A1, A2, A3, A4c, A4a, A4b, A5 et A7) qui sont présents le long du génome viral chez la majorité des souches VIH (figure 5). Il faut encore ajouter le site donneur D5 et le site accepteur A6 qui sont spécifiques à certaines souches ainsi que certains sites cryptiques pour obtenir l'intégralité des sites d'épissages décrits chez le VIH [39,40]. Grâce à cette impressionnante quantité de sites d'épissages se sont au moins une quarantaine d'ARNm partiellement ou complètement épissés qui sont produits à partir du provirus VIH. De plus de l'ARNm complet non épissé, en plus de pouvoir être traduit pour générer des protéines virales, est également celui qui peut être encapsidé dans de nouveaux virions pour servir de génome viral et de matrice à la RT.<sup>41</sup>

## **Troisième partie : cycle viral**



**Figure 6 : Schéma du mécanisme d'action de la protéine Tat adapté de [41]**

La fixation de la protéine Tat (rouge) sur la séquence TAR d'un ARNm viral en cours de synthèse permet le recrutement du complexe P-TEFb constitué par les protéines cycT1 (violet) et CDK9 (bleu). Le recrutement de ce complexe entraîne le départ du complexe NELF, qui régule négativement la transcription, et la phosphorylation du complexe DSIF ce qui augmente l'efficacité de transcription. P-TEFb phosphoryle également l'extrémité C-ter de l'ARN polymérase II (en vert clair) ce qui augmente la processivité de cette enzyme et induit une meilleure transcription du génome viral.

## 3. Cycle viral

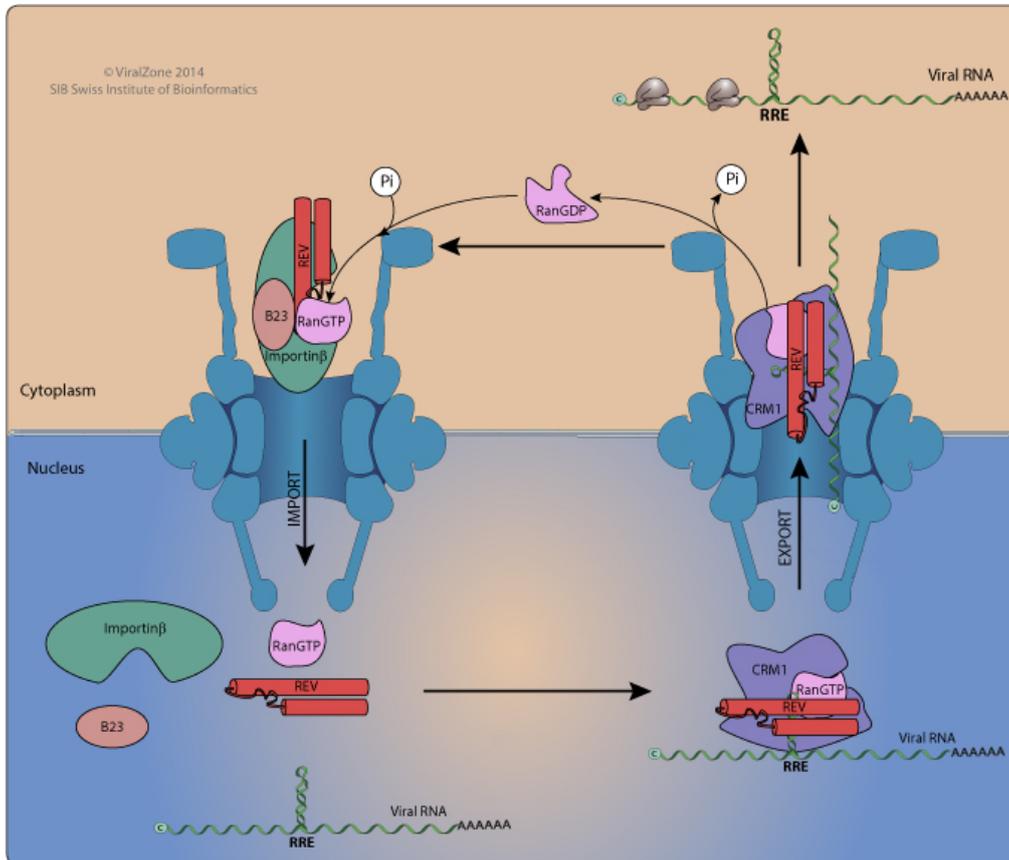
### 3.1. Phase tardive

#### 3.1.1. Expression du génome viral

Une fois qu'un provirus s'est intégré dans le génome d'une cellule hôte, l'information génétique qu'il contient est exprimée grâce au promoteur contenu dans ses LTR. Ainsi, des ARNm sont produits grâce à l'ARN polymérase II cellulaire qui reconnaît et se fixe au promoteur viral. De plus, la présence des éléments enhancers liant des facteurs de transcription tels que NF- $\kappa$ B vont assurer une excellente initiation de la traduction du provirus. Pourtant, au début de l'infection le niveau de transcription du provirus reste à un niveau très faible. En effet, même si le promoteur du VIH parvient très efficacement à initier la transcription du provirus, un problème se pose quant à la taille de celui-ci : environ 10 000 paires de bases (base pair ou bp) qui sont à comparer avec la taille moyenne d'un ORF eucaryote qui ne dépasse pas 1 500 bp [42]. Cette taille, largement en dehors des standards des eucaryotes, nécessite non seulement une bonne initiation de la transcription mais également une ARN polymérase très processive qui parvienne à réaliser des transcrits complets. Or cette processivité est absente au début de l'infection d'une cellule et ne sera obtenue que plus tard par un mécanisme d'origine viral. Néanmoins, même si globalement le niveau de transcription efficace est au départ très faible, ce niveau permet la production des premiers transcrits complets qui seront donc complètement épissés, car le processus d'export viral des ARNm non épissés n'est, lui non plus, toujours pas en place. Plusieurs types de transcrits différents sont produits par épissage alternatif à partir des différents sites donneurs et accepteurs disponibles et cette première population de transcrits est exportée du noyau pour être traduite par les ribosomes cellulaires. Ces tout premiers ARNm vont ainsi permettre la synthèse de trois protéines virales qui vont maintenant être traitée plus en détail : Tat, Rev et Nef.

#### 3.1.2. Tat

Tat est une petite protéine constituée d'environ 80 à 100 acides aminés pour une moyenne de 10 KDa selon les souches de VIH-1. Après sa synthèse dans le cytoplasme, elle est redirigée vers le noyau grâce à un signal de localisation nucléaire (nuclear localization signal ou NLS) pour y jouer un rôle des plus cruciaux. L'action de cette protéine permet d'augmenter considérablement le niveau de transcription du provirus en lui donnant ce qui lui faisait défaut tant que Tat n'était pas présente : une processivité accrue des ARNpol II [43]. Pour ce faire, cette protéine se fixe à la séquence TAR, que j'avais rapidement évoqué précédemment, présente dans l'élément R des LTR viraux (figure 6).



**Figure 7 : Schéma du mécanisme d'action de la protéine Rev adapté de [41]**

(gauche) La protéine Rev cytoplasmique est reconnue et se fixe à une Importine  $\beta$  en complexe avec la protéine B23. Le complexe va acheminer Rev jusqu'à un pore nucléaire et va induire son import nucléaire ce qui nécessite la fixation et la phosphorylation d'une protéine RanGDP en RanGTP. Dans le noyau le complexe se dissocie et libère Rev.

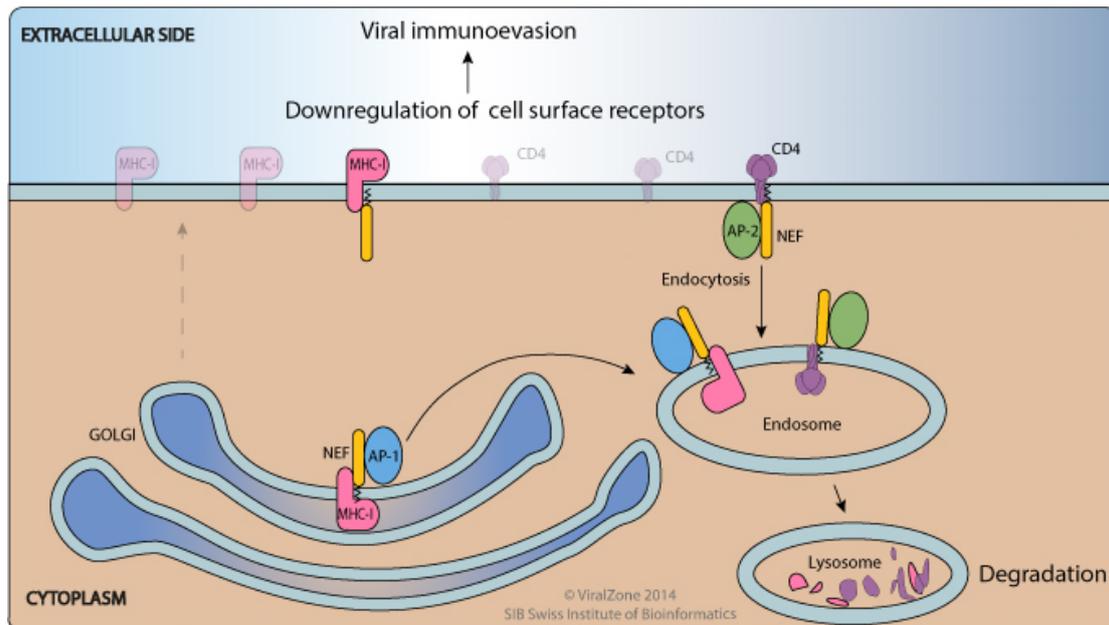
(droite) Rev reconnaît l'élément RRE présent uniquement sur les ARNm viraux incomplètement épissés et se fixe sur lui puis multimérise (non représenté sur le schéma) et recrute la protéine CRM-1 et une Ran GTP. Ce complexe transite ensuite vers un pore nucléaire et est exporté du noyau ce qui nécessite la déphosphorylation de Ran GTP en Ran GDP. Le complexe se dissocie enfin dans le cytoplasme ce qui libère l'ARNm partiellement épissé et le rend disponible pour être traduit par les ribosomes.

Cependant la fixation de Tat ne se fait pas sur le provirus mais sur la séquence TAR des transcrits en cours de synthèse. TAR présente, en effet, la particularité d'adopter une conformation en tige-boucle caractéristique lorsque qu'elle est sous forme ARN et c'est la présence de cette structure en tige-boucle qui est reconnue par la protéine Tat et qui provoque sa fixation [44]. Tat va ainsi permettre la fixation d'un facteur d'élongation : le complexe P-TEFb composé de la cycline T1 (CycT1) et de la kinase cycline-dépendante 9 (CDK9). Sans Tat, deux complexes régulant négativement l'ARNpol II sont liés à celle-ci : ce sont les complexes NELF (negative elongation factor) et DSIF (DRB sensitivity inducing factor) qui, lorsqu'ils sont présent simultanément sur l'ARNpol II, préviennent l'élongation efficace du transcrit. Le recrutement de P-TEFb par Tat permet à ce complexe cellulaire de phosphoryler NELF et de provoquer ainsi son départ de l'ARNpol II. Le complexe DSIF se retrouve ainsi seul et de facteur négatif il devient un facteur impactant positivement l'élongation. De plus, le complexe P-TEFb va également phosphoryler DSIF ainsi que le domaine carboxy-terminal (C-ter) de l'ARNpol II ce qui permet d'augmenter la processivité de la polymérase et de garantir que les transcrits viraux soient complets. Ces trois mécanismes conjugués vont ainsi accroître considérablement l'efficacité de l'élongation des transcrits viraux et, en addition de l'initiation de la transcription du provirus déjà très efficace, vont permettre une transcription abondante de la totalité du provirus [45].

### 3.1.3. Rev

Une autre protéine produite précocement est la protéine Rev, qui est tout comme Tat une protéine de petite taille avec environ 110 acides aminés pour une masse moyenne de 13 KDa. Cette protéine est l'une des clés de voûte permettant de réguler le cycle viral car c'est elle qui va permettre l'export des ARNm viraux incomplètement épissés [46]. Pour accomplir cette tâche il est nécessaire que la protéine renferme à la fois un NLS et un NES (nuclear export signal) afin de pouvoir faire la navette entre le noyau et le cytoplasme. Le phénomène qui explique l'absence d'interférence entre ces deux signaux contradictoire est intimement lié à l'activité de cette protéine.

Lorsque la protéine n'est pas en contact avec un ARNm viral seul le NLS est actif, ce qui permet l'import de cette protéine dans le noyau après reconnaissance directe par un complexe constitué par une Importine  $\beta$  et par la protéine B23 qui dirigent la protéine Rev vers les nucléoporines [47] (figure 7). Après être arrivée dans le noyau, Rev va alors reconnaître et se fixer à une séquence spécifique codée par le génome viral : la séquence RRE (Rev responsive element). Cet élément se trouve dans un intron du génome VIH, et plus précisément dans la séquence du gène *env*, condition *sine qua non* au bon fonctionnement de l'export des ARN partiellement épissés.



**Figure 8 : Schéma du mécanisme d'action de la protéine Nef adapté de [41]**

(gauche) La protéine Nef reconnaît et se fixe sur la protéine MHC-I, alors que celle-ci se trouve dans l'appareil de Golgi, et recrute la protéine adaptatrice AP-1. Cette fixation permet le recrutement ultérieur de protéines de clathrines (non représentées sur le schéma) ce qui permet de diriger MHC-I vers un endosome. La maturation naturelle de l'endosome en lysosome assure la dégradation de MHC-I

(droite) Nef reconnaît et se fixe sur la protéine CD4 située à la membrane cellulaire et recrute la protéine adaptatrice AP-2. Cette fixation permet le recrutement ultérieur de protéines de clathrines (non représentées sur le schéma) ce qui permet de diriger le CD4 vers un endosome. La maturation naturelle de l'endosome en lysosome assure la dégradation du CD4.

Exactement comme pour la séquence TAR et la protéine Tat, la séquence RRE n'est pas reconnue lorsqu'elle est sous forme ADN mais uniquement sous forme ARN. En effet la séquence ARN de RRE produit des structures secondaires spécifiques qui sont reconnues par Rev et qui provoquent la fixation de cette protéine. Rev se fixe ainsi sur la séquence RRE, qui n'est présente que sur les ARN viraux n'ayant pas subi un épissage complet. La fixation de Rev va provoquer la multimérisation de cette protéine [48] ce qui induit le masquage des NLS et l'activation des NES présent sur les protéines Rev. Les NES vont alors promouvoir l'export du complexe ARNm-Rev en étant reconnu par le récepteur nucléaire CRM-1 qui, à son tour, va acheminer le complexe vers les pores nucléaires [49].

#### **3.1.4. Nef**

La dernière protéine qui peut être produite à partir des ARNm complètement épissés est la protéine Nef : une protéine myristoylée d'une masse moléculaire comprise entre 27 et 35 KDa selon les souches de VIH-1. Les fonctions subtiles, mais vitales, de cette protéine n'ont pas été immédiatement bien comprises par la communauté scientifique et lui ont valu son nom, Nef signifiant « negative factor ». Les premiers travaux concernant Nef la décrivent effectivement comme un répresseur transcriptionnel des LTR du VIH [50,51]. De plus, le fait que le gène codant cette protéine puisse être inactivé sans que cela n'empêche la production de particules virales *in vitro* à renforcé ce rôle putatif de protéine régulant négativement le cycle viral. Même si assez rapidement de nombreux travaux contestant l'existence d'une telle activité répressive ont vu le jour [52,53], il faudra néanmoins attendre quelques années supplémentaires pour que finalement les principaux rôles de Nef soient identifiés. Le premier de ces rôle est de réguler négativement la production de la protéine membranaire cellulaire CD4 (cluster of differentiation 4) [54] (figure 8). Comme il a été précisé, Nef est une protéine myristoylée, ce qui lui permet d'être ancrée sur la face interne de la membrane plasmique cellulaire. C'est donc à la membrane que se fait la rencontre entre les protéines Nef et CD4 et c'est là que s'effectue la régulation. La protéine Nef va, en effet, lier le CD4 et provoquer son endocytose, grâce au recrutement d'une protéine adaptatrice AP-2 et de protéines de clathrines, puis sa dégradation dans les lysosomes [55]. Cette répression de CD4 est nécessaire à cause de deux propriétés de cette protéine : elle est à la fois le récepteur membranaire permettant aux particules virales de pénétrer à l'intérieur de leurs cellules cibles et également une protéine membranaire essentielle à l'activation du système immunitaire. CD4 est en effet reconnue par la protéine d'enveloppe du VIH-1, Env qui sera largement traitée dans les sections ultérieures, et cette reconnaissance permet au virion de pénétrer à l'intérieur de la cellule portant CD4 pour l'infecter. La première conséquence de la disparition de CD4 à la membrane plasmique d'une cellule infectée

consiste à défavoriser la surinfection de la cellule hôte par d'autres particules virales de VIH-1. Ceci permet de favoriser le fait que toute l'énergie de la cellule hôte sera consacrée à la production de la seule descendance du provirus déjà présent dans son génome et évitera au maximum les compétitions entre provirus. Le second intérêt qui découle de la réduction du taux de CD4 à la surface d'une cellule est l'entrave du bon déroulement de la mise en place de la réponse immunitaire au début de l'infection d'un nouvel hôte. En effet, CD4 est une protéine permettant aux différentes cellules composant le système immunitaire de se coordonner et d'activer uniquement la population de cellules la plus à même de combattre un pathogène donné [56]. Le fait de réduire l'expression de CD4 à la surface des cellules immunitaires va donc permettre de ralentir considérablement la capacité de l'hôte à aboutir à une réponse immunitaire efficace face à l'infection par VIH-1. Une preuve de cette torpeur initiale du système immunitaire est la lenteur d'apparition des anticorps dans le sérum d'un individu nouvellement infecté. Alors qu'en temps normal la séroconversion d'un individu après infection à un virus tel qu'Influenza, responsable de la grippe saisonnière, prend entre 3 et 4 jours [57], la séroconversion d'un individu après une infection à VIH met en moyenne 22 jours à s'effectuer.

Le second rôle majeur de la protéine Nef est de réguler négativement l'expression d'une seconde protéine cellulaire : MHC-I (major histocompatibility complex class I) (figure 8). La régulation s'effectue également à un niveau post-traductionnel et selon un mécanisme assez similaire que pour l'action sur CD4. Nef va lier MHC-I, non pas au niveau de la membrane plasmique mais directement dans l'appareil de Golgi [58], et provoquer l'adressage de MHC-1 aux endosomes par le recrutement d'une protéine adaptatrice AP-1 et de protéines de clathrines. Ceci permet ultimement de dégrader MHC-I dans des lysosomes comme c'est le cas pour CD4 [59].

Pourquoi est-il nécessaire pour le VIH d'induire la dégradation de cette protéine cellulaire ? La réponse est une nouvelle fois liée à la forte capacité de ce virus à échapper au système immunitaire. La protéine MHC-I est une protéine produite à la surface de toutes les cellules nucléées chez *Homo sapiens*. Cette production ubiquitaire fait écho à la fonction capitale de cette protéine qui est de renseigner le système immunitaire sur l'état de la cellule produisant la protéine. MHC-I est capable de lier n'importe quel fragment protéique issu du recyclage naturel des peptides par le protéasome. Une fois qu'un fragment est attaché à MHC-I, cette protéine est adressée à la membrane cellulaire, là où les cellules du système immunitaire, et principalement les lymphocytes cytotoxiques, vont pouvoir reconnaître et se fixer au MHC-I. Si un peptide cellulaire est présenté par le MHC-I au lymphocyte rien ne se passe, mais si un peptide d'origine étrangère, comme un peptide de VIH-1 est présenté par le MHC-I, alors le lymphocyte cytotoxique s'active et détruit la cellule qui vient d'être reconnue comme étant infectée. Il y a donc un énorme intérêt à empêcher les protéines MHC-I de subsister à la surface des cellules, et ce afin de bloquer la capacité du système immunitaire à

reconnaître les cellules produisant des virions et ainsi prévenir leur destruction [60]. Mais, dans le même temps, une absence complète de MHC-I à la membrane plasmique d'une cellule est considéré comme anormal et active la réponse immunitaire des cellules lymphocytaires dites « natural killer » (NK) qui induisent la destruction de cette cellule [61]. C'est très probablement pour cette raison que le mécanisme inhibiteur de Nef se produit par destruction des protéines *via* les endosomes. Ce mécanisme est moins efficace qu'une dégradation par le protéasome ou une inhibition traductionnelle, par exemples, mais permet justement de conserver à la membrane une quantité faible mais suffisante de protéines MHC-I afin d'éviter la destruction de la cellule infectée par les NK.

Ainsi, contrairement à ce que la communauté scientifique pensait lors des premières études de Nef, l'action de cette protéine largement dédiée à la lutte contre le système immunitaire adaptatif de l'hôte est absolument cruciale pour le virus. C'est ce rôle vital qui a paradoxalement fait que la protéine était, à l'origine, considérée comme purement accessoire. En effet, dans les systèmes *in vitro* où le système immunitaire adaptatif est absent, cette protéine devient *de facto* accessoire. Cependant, dans un hôte vivant Nef devient fondamentale, la preuve étant que si cette protéine est inactive dans toute la population virale infectant un individu, alors l'infection à VIH reste invariablement au stade asymptomatique chronique et n'évolue jamais vers le SIDA pour cet individu [62].

Pour revenir au cycle viral, c'est donc au cours de cette première phase, dites précoce, que l'action immunomodulatrice de la protéine Nef permet la création d'un terrain favorable à la réplication virale, et que la présence de la protéine Rev autorise la synthèse de nouvelles protéines virales en exportant des ARNm viraux partiellement épissés. Et parmi les nouvelles protéines qui seront synthétisées par ce mécanisme se trouve la protéine Vpr.

### 3.1.5. Vpr

Vpr est une petite protéine d'une centaine d'acides aminés et de 14 KDa environ. L'une des actions caractéristiques de Vpr est de provoquer l'arrêt du cycle cellulaire en phase G2 dans les cellules en division puis la mort de la cellule infectée [63]. Des études ont également démontrées que le fait de bloquer la cellule en division dans sa phase G2 provoque l'accroissement du niveaux de transcription du provirus par un mécanisme non-élucidé [64], ce qui a été une découverte intéressante car il est également prouvé que cette transcription accrue causée par le blocage en G2 ne semble pas avoir pour effet une réplication accrue du virus [63]. Cette contradiction entre le fait que la Vpr précipite la mort de la cellule infectée sans pour autant accélérer la réplication virale à très vite fait supposer à la communauté scientifique que l'arrêt du cycle cellulaire en phase G2 n'était

peut-être pas la véritable fonction de Vpr, ou en tout cas pas la totalité de sa fonction. Des études approfondies menées à la suite de cette constatation ont fini par aboutir à deux informations capitales : premièrement, pour pouvoir stopper le cycle cellulaire en phase G2 il est nécessaire que Vpr interagisse avec un complexe cullin-RING ubiquitin ligase, en l'occurrence le complexe cullin4A-DDB1, et deuxièmement lorsque, par mutagenèse, on empêche Vpr d'interagir avec ce complexe, on réduit également la capacité du VIH à infecter certains types cellulaires tels que les macrophages, qui sont pourtant parmi les premières cellules infectées par VIH-1 après qu'un hôte ait été infecté par le virus [65]. L'arrêt en G2 et la levée de la restriction réduisant l'infection à VIH-1 dans les macrophages se fait donc via le même mécanisme : l'ubiquitylation de facteurs cellulaires et leur dégradation par le protéasome. En revanche, tant que ces facteurs cellulaires ne sont pas identifiés, il est impossible de dire quel lien de cause à effet existe entre les deux actions. Il est tout à fait possible que Vpr dégrade ainsi deux facteurs cellulaires différents, un facteur de restriction présent chez les macrophages et un deuxième facteur impliqué dans le cycle cellulaire. Cependant, étant donné que l'avantage que le virus aurait à bloquer la cellule en G2 reste obscur, une hypothèse encore non vérifiée stipule que le facteur de restriction qui est dégradé par Vpr est peut-être également un élément impliqué, directement ou non, dans le cycle cellulaire [65]. Ainsi, selon cette hypothèse, l'arrêt en G2 serait un dommage collatéral dû à la dégradation de ce facteur de restriction mais ne serait pas le but initial « recherché » par le VIH-1.

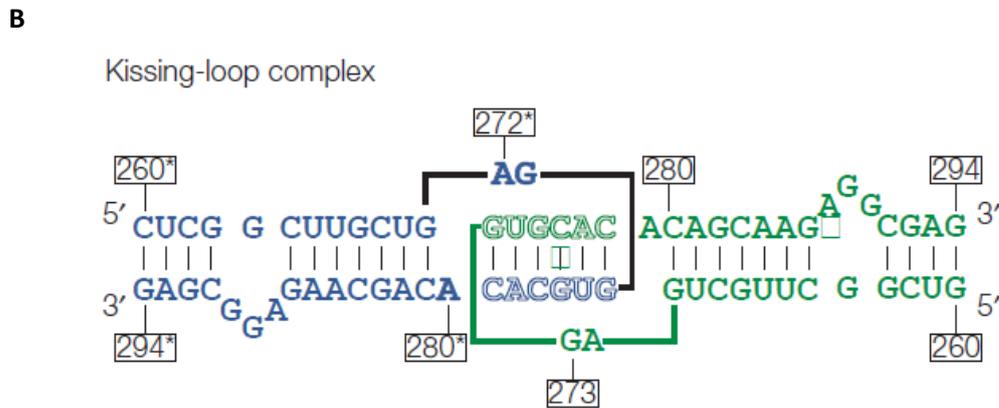
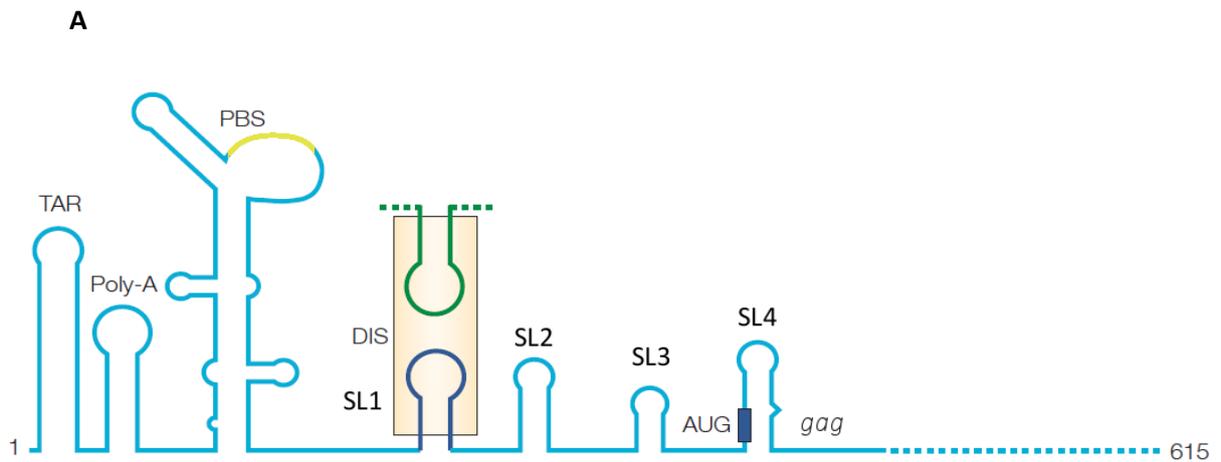
### **3.1.6. Expression de la polyprotéine Gag-Pol**

En plus d'un large éventail de protéines régulatrices à action anti-immunitaire, le provirus VIH-1 code également pour les protéines structurales et enzymatiques qui vont servir à la production de virions capables d'infecter de nouvelles cellules. Toutes ces protéines sont produites à partir du même ARNm, et cet ARNm est assez particulier car il est entièrement non-épissé. Ceci signifie que l'ARNm produisant les protéines structurales et enzymatiques est également l'ARN génomique qui sera encapsidé dans les virions au moment du bourgeonnement. Cette particularité s'explique par le fait que l'intron codant pour les protéines structurales et enzymatiques est le premier intron présent sur le génome VIH-1. Cet intron est ainsi flanqué par des sites donneur et accepteur d'épissage fort [66], ce qui implique que si la protéine Rev exporte un transcrit qui n'a pas subi cet épissage précis, alors il est très peu probable que les autres événements d'épissage aient eu lieu.

La question qui se pose, sachant que les ribosomes eucaryotes ne peuvent d'ordinaire pas aller au-delà du premier cistron d'un ARNm, est : comment le VIH-1 peut-il faire exprimer toute la batterie de protéines structurales et enzymatiques à partir de ce seul ARNm ? La réponse trouvée par le VIH-1

à ce problème n'est pas inédite dans le sens où d'autres virus, qui partagent avec VIH-1 le fait de posséder un génome à ARN non segmenté, ont également sélectionné cette stratégie : posséder un ARNm codant pour une polyprotéine. Cette polyprotéine est clivée, le plus souvent par une protéase virale comme c'est le cas pour VIH-1 [67], afin de libérer les différentes protéines qui la composent. Une autre subtilité vient se greffer à ce mécanisme : étant donné qu'il n'est pas nécessaire que les protéines enzymatiques soient produites dans des quantités aussi importantes que les protéines structurales, un codon stop sépare la partie de la polyprotéine codant pour les protéines structurales (correspondant à un gène appelé *gag* sur le provirus), qui sont en amont du codon, et les protéines enzymatiques (gène *pol* du provirus) situées en aval du codon qui, de plus, ne sont pas codées dans le même cadre de lecture que *gag*. Si rien n'est fait, la traduction de l'ARNm s'arrête donc après la synthèse de la polyprotéine codant uniquement les protéines structurales donnant ainsi naissance à une polyprotéine Gag. Un mécanisme typique de « frameshifting » d'origine viral permet de forcer le ribosome cellulaire à s'arrêter sur une structure secondaire présente sur l'ARNm viral un peu en amont du codon stop situé en fin de *gag* et de changer de cadre de lecture [68]. Ce changement de cadre de lecture permet au ribosome de franchir le codon stop, qui est désormais dans un cadre de lecture différent, et permet la synthèse d'une polyprotéine complète Gag-Pol contenant à la fois les protéines structurales et les protéines enzymatiques, au nombre desquelles figure la protéase nécessaire à la maturation des polyprotéines. Étant donné que l'évènement de « frameshifting » ne se produit qu'environ 5 à 10 fois sur 100, la production de la polyprotéine Gag dépasse nettement la production de Gag-Pol, et ce mécanisme permet ainsi de réguler la quantité de protéines enzymatiques virales produites [69].

Cependant, même si la protéase est produite en même temps que la polyprotéine, il est plus avantageux pour le VIH-1 que la maturation du précurseur polyprotéique ne se fasse qu'après le bourgeonnement des particules virales. En effet, conserver la polyprotéine entière jusqu'au bourgeonnement permet de faciliter considérablement l'assemblage des virions car toutes les protéines structurales et enzymatiques qui doivent être présentes dans la nouvelle particule virale sont regroupées dans le précurseur. Ainsi, plutôt que de devoir coordonner l'assemblage de multiples protéines structurales et enzymatiques, sans oublier la protéine d'enveloppe Env ainsi que le génome viral sous forme ARN, il est seulement nécessaire de coordonner l'assemblage de la polyprotéine, sous ces deux formes Gag et Gag-Pol, avec la protéine Env et le génome ARN.



**Figure 9 : Schémas de la structure secondaire de la région 5' de l'ARN génomique du VIH-1 et d'un « kissing loop complex » adaptés de [73]**

(A) Les principaux domaines structuraux de la région 5' d'un ARN génomique sont représentés en bleu clair, depuis la position 1 jusqu'à la position 615, notamment la tige boucle composant l'élément TAR, la tige boucle renfermant le site de polyadénylation (Poly-A), le primer binding site (PBS) représenté en jaune, ainsi que les quatre tiges boucles SL1 (bleu foncé), SL2, SL3 et SL4. La tige boucle SL1 d'un second ARN génomique est représentée en vert pour montrer l'interaction entre deux ARN génomiques. Les « dimerization initiation sites » (DIS) des deux ARN génomiques correspondent à la partie encadrée en fond jaune. La position du codon initiateur du gène *gag* est représentée par un rectangle bleu foncé.

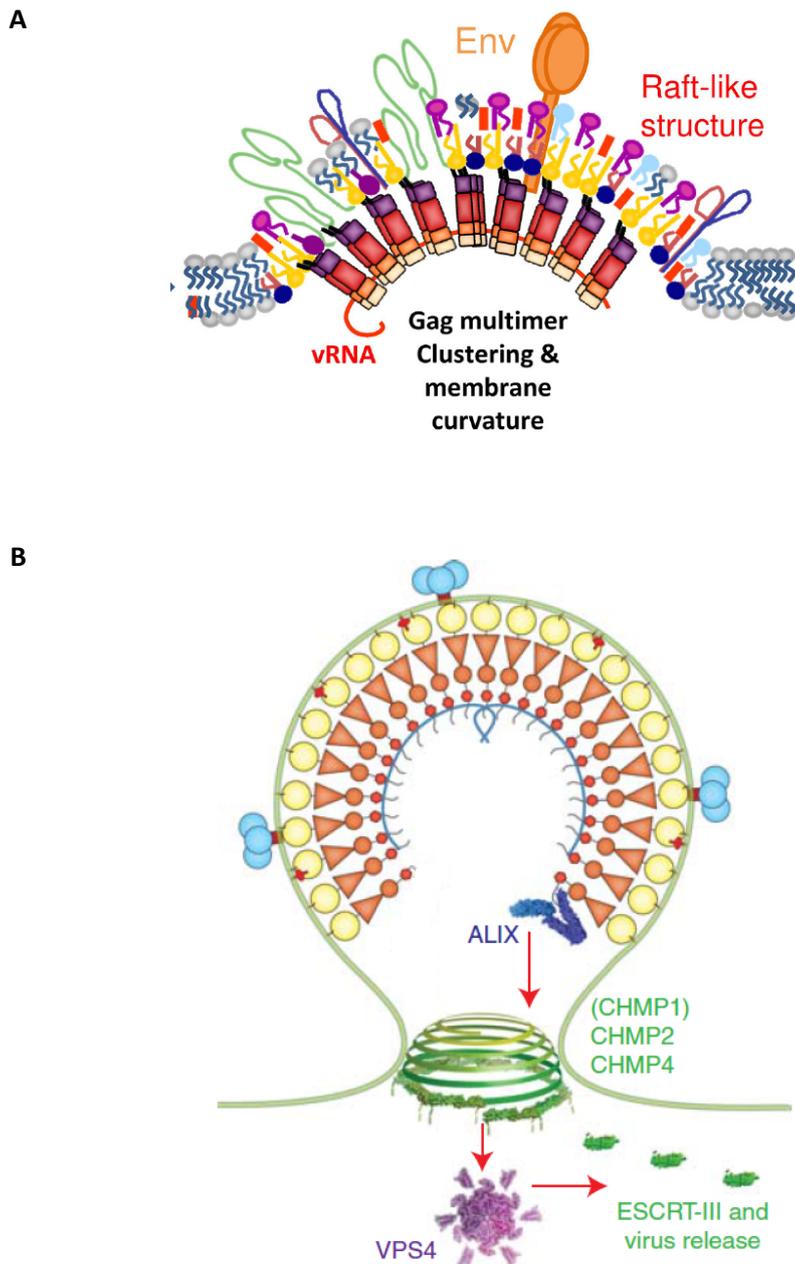
(B) Un « kissing loop complex » est représenté. Le DIS du premier ARN génomique est représenté en bleu et la numération de ses nucléotides est indiquée avec une étoile tandis que celui du second ARN génomique est en vert et la numération est indiquée sans étoiles. Les nucléotides indiqués avec un remplissage blanc correspondent à la région double brin constituée des séquences des deux ARN génomiques viraux.

Cet assemblage est d'autant facilité que les précurseurs Gag et Gag-Pol subissent une myristoylation dans leur partie N-ter lors de leur synthèse dans le réticulum endoplasmique [70], ils sont donc naturellement ancrés dans la membrane et n'ont plus qu'à être dirigés là où se fait le bourgeonnement du VIH-1. De plus la polyprotéine Gag contient à la fois la séquence peptidique permettant de recruter la protéine Env dans sa partie N-ter et une structure permettant de lier l'ARN génomique viral dans sa partie C-ter, ce qui permet de concentrer efficacement tous les éléments nécessaires à la formation d'un virion infectieux au même endroit. Cependant, seul le précurseur est nécessaire à la production de virions, c'est-à-dire que même en l'absence de protéines Env et d'ARN viral, la polyprotéine seule est suffisante pour qu'il y ait bourgeonnement de particules virales. De tels virions seraient néanmoins non-infectieux car privés à la fois de protéines Env et de génome.

### **3.1.7. Formation des nouveaux virions**

#### **3.1.7.1. Sélection de l'ARN génomique viral**

Les premières étapes du bourgeonnement consistent en la liaison du futur ARN génomique viral par plusieurs polyprotéines Gag et/ou Gag-Pol. L'ARN viral est sélectionné parmi l'énorme population d'ARN existant à l'intérieur de la cellule hôte grâce à des structures secondaires présentes en 5', notamment des structures en tige-boucle, ainsi que des séquences nucléotidiques particulières en amont et en aval de ces tige-boucles (figure 9). Ces structures en tige-boucle, nommées de SL1 à SL4 se forment sur l'ARN viral à partir des séquences nucléotidiques comprises entre la fin de l'élément U5 du LTR amont et le début du premier gène, c'est à dire *gag*. Afin de sélectionner spécifiquement les ARN viraux, le précurseur Gag se lie grâce à sa partie C-ter à ces tige-boucles et plus spécifiquement à SL1 [71]. Cependant, ces tige-boucles sont présentes sur tous les ARN viraux, car formées à partir du premier exon en aval du +1 de transcription, donc un mécanisme supplémentaire doit exister pour permettre spécifiquement la reconnaissance des ARN viraux complètement non-épissés, qui sont les seuls à coder l'intégralité de l'information génétique nécessaire à la nouvelle génération de virus. Ce mécanisme repose sur les deux séquences nucléotidiques en amont et en aval des tige-boucles : la séquence amont, correspondant à l'élément U5, est présente dans tous les transcrits viraux et inhibe fortement la liaison au précurseur Gag, tandis que la séquence aval, composée du début du gène *gag* qui est codé uniquement par le premier intron, permet de lever cette inhibition [71]. Étant donné que la liaison de l'ARN à la polyprotéine Gag est dépendante de la présence du premier intron, qui est pourtant le plus favorablement épissé, non seulement seul un ARNm viral non-épissé peut se lier <sup>72</sup>



**Figure 10 : Schémas du mécanisme de bourgeonnement viral adaptés de [72] et [79]**

(A) Schéma du mécanisme de formation d'un bourgeon viral. Après avoir reconnu et s'être fixé sur le dimère d'ARN génomique viral, les polyprotéines Gag (rectangles violet, rouge, orange et beige) multimérisent à l'endroit d'un « lipid raft » caractérisé par une composition lipidique différente du reste de la membrane plasmique et symbolisé dans le schéma par la présence de lipides colorés en jaune et violet et de cholestérol (rectangle rouge). Cette multimérisation provoque la courbure de la membrane cellulaire et la formation d'un bourgeon viral.

(B) Schéma du mécanisme de séparation des membranes cellulaires et virales. La polyprotéine Gag (cercle jaune, triangle et cercle orange et cercle rouge) recrute la protéine ALIX qui à son tour recrute les protéines CHMP2 et CHMP4 de la famille de protéine ESCRT-III. Ces protéines ferment le bourgeon viral en multimérisant et en formant une structure en forme de glissière spiralée qui glisse le long des parois membranaires pour les joindre. La libération finale du nouveau virion est permise par la protéine VPS4 qui déstabilise les membranes cellulaires et virales et les sépare définitivement.

au précurseur Gag, mais cet ARNm sera de façon presque certaine un ARNm entièrement non-épissé et renfermant toute l'information génétique nécessaire aux futurs virions.

Phénomène intéressant, ce n'est pas une molécule d'ARN viral qui est ainsi reconnu par la polyprotéine Gag mais un dimère d'ARN viral. Les virions matures contiennent ainsi en réalité deux ARN génomiques viraux, cependant on ne parle pas ici de diploïdie mais de pseudo-diploïdie car un seul des allèles sera conservé après la rétro-transcription. Cette pseudo-diploïdie n'est pas unique au VIH-1 mais est au contraire une caractéristique conservée chez tous les Retroviridae.

Pour terminer cette section, il faut ajouter que le mécanisme moléculaire régissant la dimérisation des ARN viraux repose sur la tige-boucle SL1. Cette tige-boucle contient en effet une séquence appelée « dimerization initiation site » (DIS) qui permet à deux ARN viraux d'interagir et de se lier ensemble par complémentarité de séquence en formant une structure particulière : le « kissing-loop complex » [73]. Plus précisément ce sont les deux séquences simples brins des deux domaines SL1 qui interagissent et s'associent pour former ensemble une courte région double brin, dont chacun des brins provient donc d'un des deux ARN génomiques viraux. Cette structure n'est que transitoire cependant, car la formation de ce « kissing-loop complex » induit la formation d'un « extended duplex » dans lequel les deux boucles SL1 des deux ARN génomiques viraux s'associent pour former un double brin d'ARN [74].

### **3.1.7.2. Assemblage du virion et bourgeonnement**

Le processus d'assemblage des nouveaux virions est un phénomène assez complexe qui ne sera ici traité que très superficiellement. Les premières étapes de l'assemblage d'un nouveau virion constituent en la liaison du précurseur Gag sur le dimère d'ARN viral. Cette étape est assurée par le domaine C-ter des polyprotéines Gag, domaine qui constituera la future protéine de nucléocapside après maturation du précurseur, en reconnaissant et se fixant aux tige-boucles SL1 (voir section « sélection de l'ARN génomique »). Cette première étape de fixation entraîne la multimérisation des précurseurs Gag et Gag-Pol grâce à un domaine commun à ces deux polyprotéines situé en amont du C-ter de Gag et qui constituera la protéine de capsid après maturation des précurseur [75].

La liaison des premiers précurseurs Gag à l'ARN va ainsi déclencher la multimérisation graduelle des précurseurs Gag et Gag-Pol, qui forment alors des structures hexamériques caractéristiques s'accumulant à la membrane plasmique [76] (figure 10). Ces hexamères interagissent entre eux et s'assemblent à leur tour pour former une sphère protéique creuse entourant le génome. Il est important de préciser que cette étape de multimérisation ne prend pas place à n'importe quel endroit de la membrane cellulaire mais a lieu dans des « lipid rafts » [77]. Ces « lipid rafts » sont

caractérisés par une composition lipidique différente de celle du reste de la membrane cellulaire et caractérisés par une forte concentration en lipides saturés et en cholestérol. Cette composition singulière donne des propriétés particulières aux « lipid rafts » qui sont plus résistants aux détergents et aux basses températures que le reste de la membrane cellulaire. La ou les raisons qui forcent le virus à bourgeonner à ces endroits précis de la membrane cellulaire ne sont pour l'instant pas entièrement comprises. Toutefois, les données les plus récentes indiquent que la composition lipidique particulière des « lipid rafts » est essentielle pour garantir une fluidité optimale de la membrane virale, fluidité qui favorisera ensuite l'étape de fusion des membranes des virions avec les membranes de leurs nouvelles cellules hôtes au moment de l'entrée virale [78].

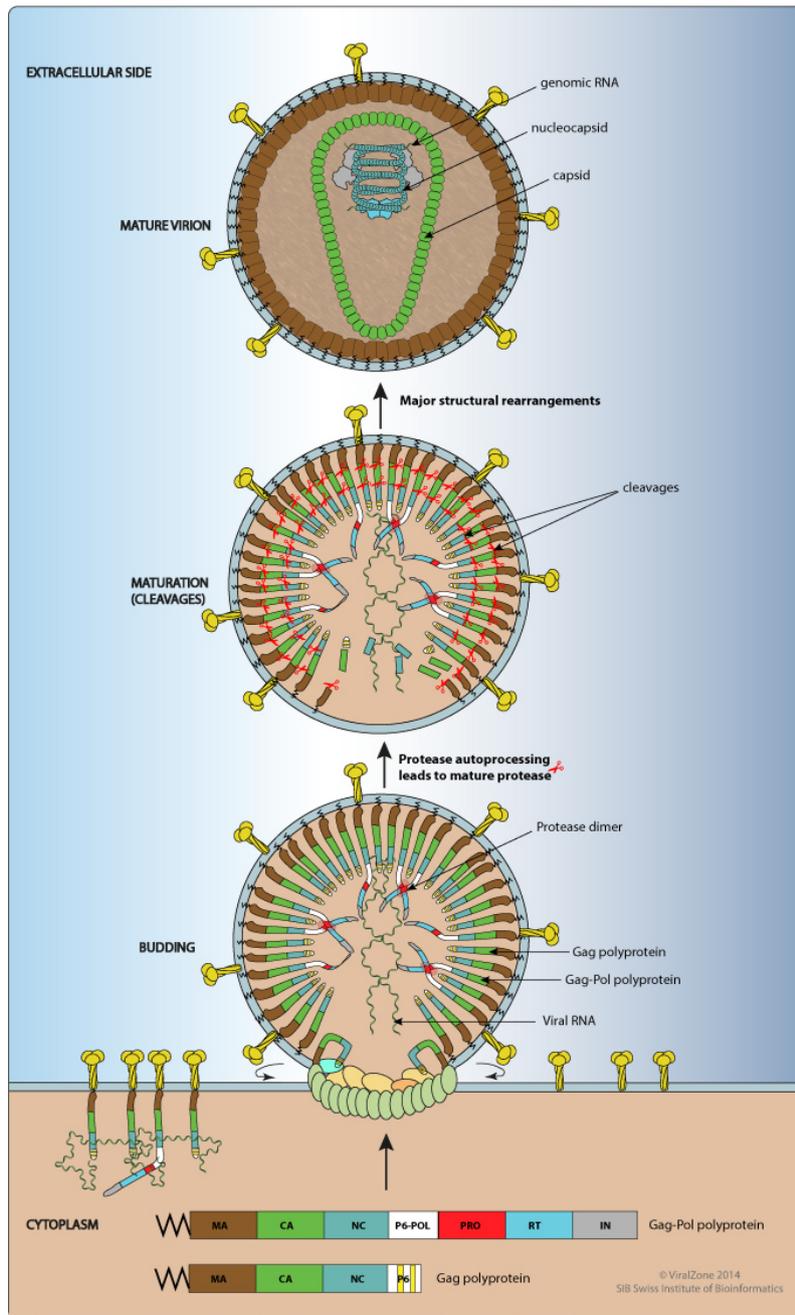
Pour revenir au bourgeonnement, l'ancrage de Gag dans la membrane est capital pour celui-ci car la courbure générée par l'assemblage des hexamères et la formation de la coque protéique sphérique se répercute sur la membrane cellulaire qui se courbe elle-même. La formation de cette excroissance membranaire par la multimérisation des précurseurs Gag et Gag-Pol est le mécanisme permettant de créer le bourgeon viral. Pourtant, même si l'assemblage de Gag permet de créer une excroissance membranaire, il n'en demeure pas moins que séparer la membrane cellulaire avec la membrane qui deviendra la membrane virale est un mécanisme thermodynamiquement défavorable et coûteux en énergie. Aussi il est nécessaire pour la particule virale de s'aider de facteurs cellulaires pour désolidariser le bourgeon de la membrane plasmique, et en l'occurrence c'est la machinerie protéique ESCRT qui est recrutée au niveau du bourgeon et qui va accomplir cette tâche en consommant de l'ATP [79] (figure 10). Cette machinerie est d'ordinaire utilisée par la cellule lors de la formation des compartiments endosomaux à partir de la membrane plasmique ou lors de la cytokinèse pour séparer les membranes des deux futures cellules filles issues d'une mitose ou d'une méiose.

Le mécanisme complet permettant la libération de la particule virale par ce cette machinerie protéique n'est pas encore totalement compris et est complexe mais, brièvement, c'est la protéine ALIX, un facteur de la machinerie ESCRT qui est recrutée par un signal peptidique présent en C-ter de Gag. Cette protéine ALIX recrute à son tour des protéines appartenant à la famille des protéines ESCRT-III, notamment CHMP2 et CHMP4, ainsi que la protéine VPS4, à l'endroit où se forme le bourgeon. Ces protéines ESCRT-III vont s'assembler pour former une structure en spirale à l'endroit où la membrane plasmique forme un goulot et ensuite, par le glissement du filament protéique en forme de spirale le long du goulot, le complexe va graduellement rapprocher les bords du cylindre membranaire pour le fermer. A ce stade néanmoins, seule une structure membranaire en hémifusion se forme et il faut encore l'action de la protéine VPS4 pour déstabiliser les membranes hémifusionnées et libérer le virion nouvellement formé.

### 3.1.8. Vpu

Le bon déroulement du bourgeonnement des virions repose en grande partie sur l'action de la protéine virale Vpu. C'est une petite protéine régulatrice composée de 80 acides aminés environ pour une masse avoisinant les 16 KDa [80]. L'une des actions capitales de cette protéine est d'inhiber la production de la protéine cellulaire tetherin [81]. Cette protéine cellulaire est produite en réponse à la stimulation par l'IFN- $\alpha$  et présente la particularité de posséder deux domaines transmembranaires séparés par une large région hydrophile. Cette structure remarquable permet à la tetherin de contrer l'étape de libération des particules virales. En effet, lorsque le bourgeonnement des virus enveloppés se fait à la membrane, la capsidie protéique de ceux-ci se place sous la membrane plasmique et s'extrait de la cellule infectée en s'entourant d'une portion de la membrane plasmique qui devient par le fait la membrane virale. Grâce aux deux domaines transmembranaires séparés de la tetherin, lors du bourgeonnement d'un virion l'un des domaines transmembranaires peut se trouver dans la membrane virale nouvellement formée alors que le second reste dans la membrane cellulaire. La tetherin agit alors comme une ancre en retenant le virion à proximité de la cellule infectée et en l'empêchant d'aller infecter une nouvelle cellule. Ce mécanisme de défense cible une étape générale commune à tous les virus enveloppés bourgeonnant à la membrane plasmique et fait de la tetherin une protéine du système immunitaire innée particulièrement redoutable. Il est donc vital pour le VIH-1 de contrer l'action de cette protéine et c'est la protéine virale Vpu qui est chargé d'entraver ce mécanisme de défense cellulaire. Pour ce faire Vpu induit la dégradation de la tetherin, par un moyen relativement similaire à l'inhibition du CD4 par la protéine Nef. Vpu se lie en effet à la tetherin au niveau de la membrane plasmique et provoque l'ubiquitinylation de la protéine cellulaire puis, une fois ubiquitinylée, la tetherin est endocytée et dégradée dans des lysosomes [82]. Comme pour Nef et CD4, la dégradation de la tetherin se produit uniquement après que celle-ci n'ait atteint la membrane et n'empêche donc pas la production continue de cette protéine par la cellule et son adressage à la membrane. Aussi, Vpu n'assure pas l'absence complète de la tetherin à la membrane mais seulement une baisse significative du taux de cette protéine ; baisse néanmoins suffisante pour assurer la production massive de virions libres non-ancrés à la cellule.

L'autre action capitale de Vpu est toujours liée au bourgeonnement du virion, mais de façon moins directe, car il s'agit d'inhiber la production de la protéine CD4. Cette protéine sert de récepteur au VIH-1 et permet son entrée dans sa cellule hôte par un mécanisme qui sera détaillé ultérieurement. Il faut rappeler ici que l'action de la protéine Nef permettait déjà de dégrader les protéines CD4, et ce pour plusieurs raisons, en les redirigeant de la membrane plasmique vers les lysosomes (voir section « Nef »).



**Figure 11 : Schéma du mécanisme global de la maturation virale adapté de [41]**

(*Cytoplasm*) Les protéines matures composants les polyprotéines Gag et Gag-Pol sont représentés sur le schéma. Le groupement myristoyl est représenté par une ligne brisée noire, les protéines de : matrice (MA) en marron, capsid (CA) en vert, nucléocapsid (NC) en bleu-vert, P6 en blanc, protéase (PRO) en rouge, transcriptase inverse R(T) en bleu et intégrase (IN) en gris. Après reconnaissance de l'ARN génomique viral, Gag et Gag-Pol s'assemblent pour former le core viral ce qui provoque le bourgeonnement.

(*Budding*) Lors du bourgeonnement des dimères de PRO se forment dans le virion grâce à leur rapprochement physique.

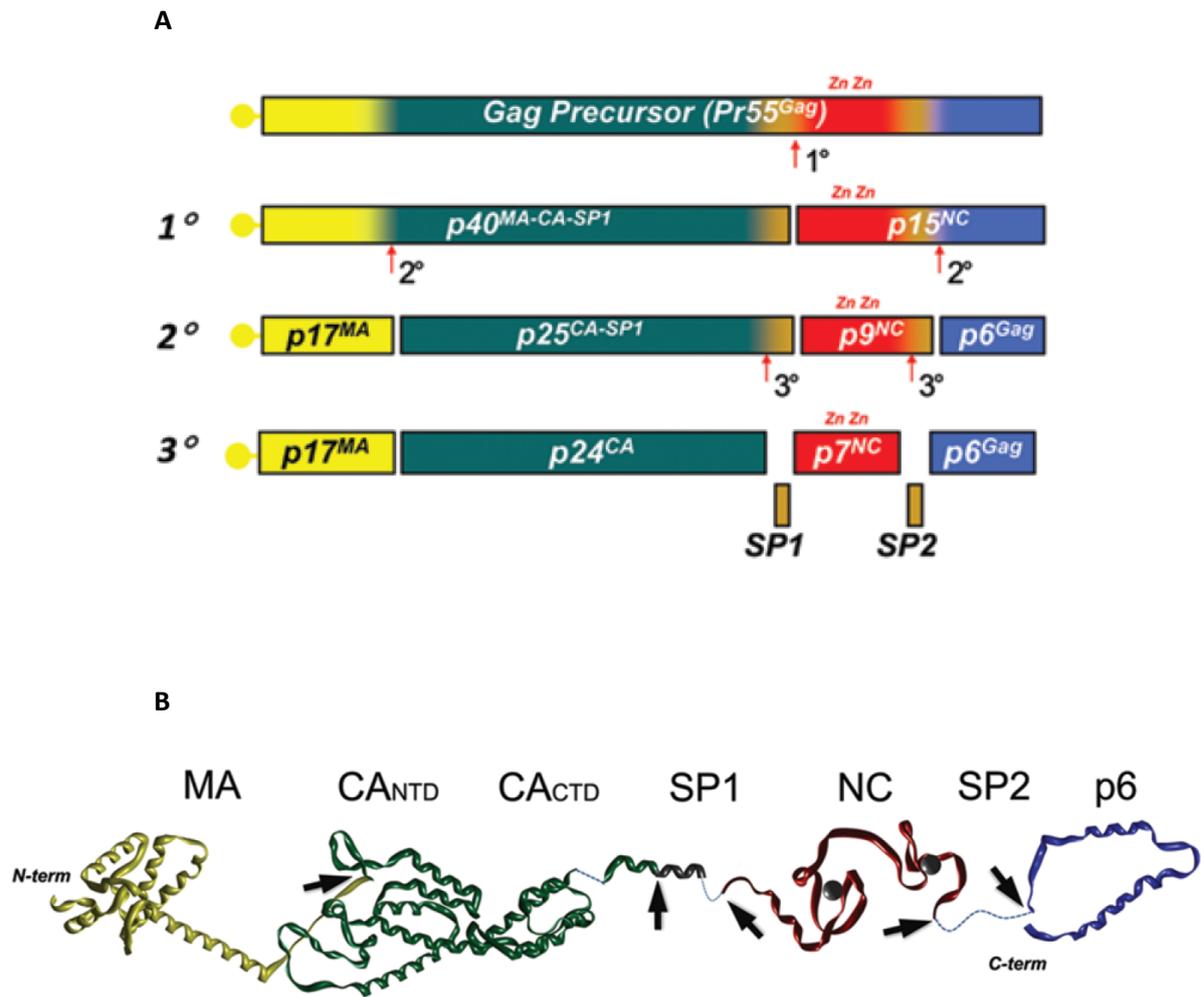
(*Maturation*) Les dimères actifs de PRO catalysent le clivage des polyprotéines Gag et Gag-Pol pour former les différentes protéines matures. Les clivages sont symbolisés par des ciseaux rouges.

(*Mature virion*) Les protéines matures s'assemblent : les protéines CA s'assemblent pour former la capsid en forme de cône typique des VIH et les protéines NC s'assemblent pour former la nucléocapsid qui condense le génome virale dans la capsid. MA reste associée à la membrane virale à cause de la présence du groupement myristoyl dans son extrémité N-ter.

La question qui se pose alors est pourquoi le VIH aurait-il besoin d'une seconde protéine permettant d'inhiber la production de CD4 ? La réponse vient du mécanisme d'inhibition par Nef qui se produit obligatoirement à la membrane. En effet, il est nécessaire que Nef exerce son action de façon post-traductionnelle, non pas à cause de CD4 mais à cause de l'inhibition de la protéine MHC-I, qui ne doit être que partielle. Ceci implique que la présentation de CD4 à la membrane cellulaire, tout comme MHC-I, n'est inhibée que partiellement et que celle-ci se fait uniquement après que CD4 ait été acheminée jusqu'à la membrane. Nef n'empêche donc aucunement la synthèse de CD4 par le ribosome ni le transit de cette protéine à travers le réticulum endoplasmique et le Golgi. C'est ce transit qui devient gênant pour le bon déroulement du cycle viral car il implique que CD4 tout comme la protéine virale Env qui reconnaît et se fixe à CD4, vont transiter par les mêmes compartiments. Si rien n'est fait pour dégrader rapidement CD4 une fois que Env est produite, la protéine cellulaire va s'associer avec Env à l'intérieur même de la cellule et va, de fait, l'inactiver en la séquestrant. C'est donc une fois que la production d'Env est permise par Rev qu'il devient nécessaire d'inhiber la production de la protéine CD4 de façon plus complète et plus en amont dans sa production que ne le fait Nef, et c'est à cette étape qu'intervient Vpu [83,84]. En se liant à CD4, la protéine Vpu entraîne la dissociation des complexes Env-CD4, qui se forment dès la synthèse de ces deux protéines à l'intérieur du réticulum endoplasmique, puis Vpu provoque la dégradation de CD4 [85]. Cette dégradation est effectuée grâce à un mécanisme assez complexe faisant intervenir à la fois le protéasome et au moins une voie de dégradation des protéines du réticulum endoplasmique (ERAD) [86,87]. Etant donné que ce mécanisme ne concerne pas cette thèse, celui-ci ne sera pas traité plus en détail.

### **3.1.9. Maturation du virion**

Après que la protéine Vpu ait rendu possible la libération de la particule virale et l'incorporation de protéines d'enveloppes sur sa membrane, le virion est complet et il n'est plus nécessaire de conserver les précurseur Gag et Gag-Pol sous leur forme polyprotéine. Au contraire, il devient alors vital de libérer les différentes protéines pour qu'elles soient en mesure d'accomplir leurs tâches respectives une fois que le virion aura rencontré une nouvelle cellule hôte potentielle (figure 11). Le mécanisme qui permet de favoriser le clivage protéolytique des différents composants des polyprotéines seulement après que le nouveau virion se soit formé est d'une simplicité redoutable : pour que la protéase virale soit active il faut qu'elle soit sous forme de dimère [67,88].



**Figure 12 : Schéma de la cinétique de maturation de Gag et structure 3D de Gag d'après [91]**

(A) Le précurseur Gag est constitué des protéines de matrice (MA en jaune), de capsid (CA en vert), de nucléocapside (NC en rouge), p6 (bleu) et des protéines spacer (SP1 et SP2 en marron). Gag est maturé en trois étapes principales. Un site de clivage de haute affinité (1°) est reconnu par la protéase et permet de produire les précurseur p40 et p15. Deux sites de clivage secondaires (2°) sont ensuite reconnus et permettent de générer les protéines matures MA et p6 et les précurseurs p25 et p9. Enfin des sites de clivage tertiaires de basse affinité (3°) sont reconnus et clivés pour former les protéines matures CA, SP1, NC et SP2. La position approximative des deux domaines en doigt de zinc de NC sont indiqués par les lettres Zn en rouge.

(B) Reconstitution informatique de la structure 3D du précurseur Gag. Les futures protéines matures sont colorées avec le même code couleur que précédemment et avec la même nomenclature à l'exception de la protéine de capsid dont sont renseignés les domaines N-ter (CA<sub>NTD</sub>) et C-ter (CA<sub>CTD</sub>). Les sites de clivages par la protéase virale sont indiqués par des flèches.

Or, c'est la reconnaissance de l'ARN génomique par les polyprotéines qui initie l'assemblage des précurseurs ensemble et cet assemblage initie à son tour le bourgeonnement. L'endroit où les polyprotéines Gag-Pol sont concentrées et où la protéase virale (toujours incluse dans la polyprotéine à cette étape) a la plus de chance de dimériser est donc l'intérieur d'un virion qui vient de bourgeonner [89]. En revanche il n'est pas impossible que de petites quantités de protéases parviennent à former des dimères à l'intérieur de la cellule hôte, comme en témoigne le fait qu'il est possible de mettre en évidence des protéines structurales et enzymatiques matures dans le cytoplasme d'une cellule hôte. Mais, même si cette maturation précoce existe, elle reste minoritaire et dans des proportions suffisamment faibles pour ne pas entraver le bon déroulement du cycle viral.

Une fois que deux protéases virales ont formé un dimère, la première action réalisée est le clivage auto-catalytique des protéases qui permet de libérer le dimère ainsi que les deux fragments Gag et Pol sans la protéase. La maturation de Gag par la protéase a été très bien étudiée et est connue dans sa quasi-totalité. Sans trop rentrer dans les détails du processus, la protéase virale est une protéase à aspartate qui ne reconnaît pas de séquences peptidiques particulières mais d'avantage des structures tridimensionnelles présentes sur les précurseurs Gag et Gag-Pol [90]. Le clivage s'effectue par la reconnaissance successive de trois catégories de sites de clivages qui sont reconnus avec une efficacité décroissante par la protéase : les sites de hautes affinités, les sites à affinité médiane et les sites de faible affinité [91]. Au total se sont ainsi cinq sites de clivages qui sont présents le long du précurseur Gag et qui délimitent six peptides (figure 12). Le premier de ces peptides, en commençant par l'extrémité N-ter de Gag-Pol, est la protéine de matrice, aussi appelée p17 en raison de sa masse de 17 KDa en moyenne, ou MA. Dans la particule virale mature cette protéine reste ancrée dans la membrane virale grâce à la myristoylation en N-ter qu'elle a héritée du précurseur Gag. Il est important de préciser que la p17 contient un NES qui trouvera son utilité plus tard dans le cycle viral.

La protéine venant en seconde position est la protéine de capsid p24 ou CA. C'est elle qui est responsable de la formation des structures en hexamères lors du bourgeonnement de la particule virale grâce à des domaines d'association qui lui permettent de multimériser. Le clivage de cette protéine lors de la maturation par la protéase lui permet de passer de cette structure purement hexamérique à une structure conique typique de la capsid virale de VIH-1. Ce cône arrondi est généré par la création des douze pentamères caractéristiques des capsides virales, qui se forment à la base (7 pentamères) et au sommet (5 pentamères) de celui-ci, et qui vont ensemble lui donner sa courbure [92].

Juste après la p24 vient la nucléocapsid également appelée la protéine NCp7 ou NC. Le rôle de cette protéine devient évident dès que l'on observe sa structure qui contient deux domaines en doigt de zinc. Ce sont en effet ces domaines qui permettent de lier l'ARN et qui rendent possible la reconnaissance de l'ARN génomique viral par le précurseur Gag lors des toutes premières étapes de

l'assemblage viral. Après clivage de Gag par la protéase, la nucléocapside compacte l'ARN viral pour former un cœur dense aux électrons à l'intérieur de la capsid, qu'il est possible d'observer en microscopie électronique à transmission [93].

Enfin la protéine située directement à l'extrémité C-ter du précurseur Gag est la protéine p6. Cette protéine permet, entre autres, de recruter la protéine virale Vpr pour l'encapsider au moment du bourgeonnement grâce à une séquence peptidique particulière [94]. En effet, l'un des rôles connus de cette protéine est la levée d'un mécanisme d'inhibition présent chez les macrophages (voir section « Vpr »). Il est donc nécessaire que Vpr soit encapsidée pour être délivrée dès le début de l'infection de ces cellules et s'assurer que les facteurs cellulaires empêchant l'infection soient inactivés, même si ce n'est pas la seule raison de son encapsidation.

En ce qui concerne la cinétique du clivage protéolytique, celle-ci est très bien caractérisée [95]. Le premier clivage a lieu entre la capsid et la nucléocapsid grâce à la présence d'un site de haute affinité reconnu par la protéase. Ensuite, se sont deux sites de clivage secondaires, d'affinité médiane, qui sont reconnus par la protéase et qui entraînent deux coupures presque simultanées, la première entre la matrice et la capsid et la seconde entre la nucléocapsid et la protéine p6. Enfin, la reconnaissance tardive de deux sites de clivages tertiaires, de faible affinité, permet de cliver les extrémités C-ter des protéines de capsides et de nucléocapsides afin de parachever la maturation. Les peptides ainsi libérés par cet ultime clivage sont les peptides SP1 et SP2 qui n'ont chacun aucune activité identifiée à l'heure actuelle mais permettent, par leur clivage, des réarrangements structuraux nécessaires à la formation du virion mature. Ainsi c'est le clivage de SP1 qui permet à la capsid virale de former les pentamères qui vont lui donner sa forme conique. C'est également le clivage de SP2 qui modifie l'interaction entre la nucléocapsid et l'ARN viral et permet la stabilisation d'une conformation particulière du dimère d'ARN dans le cœur du virion grâce à une activité chaperon de NCp7 [79].

La cinétique du clivage du précurseur Pol est, par contre, plus nébuleuse. Après l'auto-clivage de la protéase il ne reste que deux protéines formant le précurseur Pol, mais la question est de savoir si la maturation des protéines enzymatiques se fait avant, pendant ou après la maturation des protéines structurales, et malheureusement aucune certitude n'existe pour le moment à ce sujet. Ce qui est sûr c'est qu'à un moment donné les deux protéines formant le précurseur Pol sont séparées pour former la transcriptase inverse (reverse transcriptase ou RT), aussi appelée p66, et l'intégrase ou p32. Une fois que la RT est clivée elle dimérise car à l'instar de la protéase elle n'est active que sous cette forme. En revanche le dimère de RT ne constitue pas encore la forme mature de ce complexe enzymatique, la protéase catalyse une nouvelle coupure sur une et une seule des sous-unités du dimère pour donner naissance à une forme tronquée de la protéine appelée alors la p51. L'autre sous-unité est protégée du clivage par la protéase par un phénomène d'encombrement

stérique qui masque le site de clivage [95], ainsi un dimère de RT est en réalité un hétérodimère formé d'une sous-unité p66 et d'une sous-unité p51. La sous-unité p66 porte grâce à ce mécanisme une activité supplémentaire dont il sera question dans la section consacrée à la transcription inverse.

Concernant l'intégrase, celle-ci reste liée à la transcriptase inverse même après clivage, grâce à des interactions non covalentes entre son domaine C-ter et la RT [96]. Cependant cette liaison n'existe qu'à l'intérieur du virion et aussi lors de la rétro-transcription, après quoi l'intégrase se dissocie de la RT pour se fixer sur l'ADN double brin viral et former un tétramère qui constitue la forme active de l'intégrase [97] (voir section « intégrase et intégration »).

### **3.1.10. Phase virion**

Après que le bourgeonnement de la particule virale ait eu lieu, le virion est enfin constitué de toutes les protéines nécessaires à l'infection d'une nouvelle cellule et, suite à la maturation du virion par la protéase virale, toutes ces protéines deviennent pleinement fonctionnelles. Arrivé à ce stade le virion entre alors dans une phase de son cycle répliatif que l'on pourrait presque appeler quiescente. C'est dans cette phase que la véritable nature du virus se révèle : celle d'une entité chimique possédant les mêmes constituants que le vivant mais qui ne peut pas être appelée vivante au sens strict du terme. Par comparaison, une spore bactérienne ou un cyste eucaryote sont des organismes vivants qui se trouvent dans un état de cryptobiose à première vue similaire à celui de la particule virale. Pourtant, les spores et les cystes sont toujours des organismes autonomes capables de sortir de cet état de latence par leurs propres moyens pour atteindre à nouveau un état physiologique permettant la réplication. Le virion libre en revanche ne possède aucun métabolisme propre et devient totalement dépendant de la rencontre avec une cellule vivante pour poursuivre son cycle. Contrairement aux autres parasites connus qui possèdent une vie propre, même si ils ont besoin d'un hôte pour compléter leur cycle de vie, les virus se fondent totalement dans leurs hôtes et cessent purement d'exister dès qu'ils sont séparés définitivement de leurs hôtes. Le VIH étant un virus il n'échappe pas à cet état de fait.

Il est nécessaire de rappeler ce constat fondamental pour comprendre les contraintes de la prochaine étape du cycle répliatif du VIH : la rencontre d'un virion avec une cellule vivante. En effet, arrivé à ce stade le virion se trouve à proximité immédiate de tous les composants biologiques qu'il pourra utiliser pour s'animer et passer d'un assemblage de protéines devenues inertes à une entité dynamique capable de se répliquer. Pour parvenir à ce but il faut néanmoins que le virion traverse la membrane plasmique de la cellule.

Or, une membrane cellulaire, de par la disposition des lipides qui la constitue et qui sont ordonnés en « tête-bêche », est un cristal liquide de type lyotropique lamellaire, doté de propriétés tout à fait uniques [98]. Ce cristal liquide est hautement stable, doté de bonnes propriétés de résistance mécanique et d'une grande imperméabilité tout en étant remarquablement fluide et déformable. La membrane plasmique d'une cellule est donc en réalité une barrière redoutable et un obstacle certain dans le cycle répliatif des virus en général. La contrainte imposée par la membrane est d'autant plus forte que le virus doit la franchir dans l'état où il est le plus désavantagé. Sans un métabolisme pour lui fournir de l'énergie, les options de la particule virale sont très limitées et le sont d'autant plus que le virion ne peut pas non plus espérer s'aider des protéines de l'hôte. En effet, toutes les protéines qui pourraient éventuellement faciliter le passage du virion ne sont pas présentes dans le milieu extracellulaire mais sont hors de portée, de l'autre côté des membranes cellulaires.

### **3.1.11. Env**

Le challenge qui pèse sur le virion au moment de l'entrée virale est conséquent, pourtant de façon assez surprenante une unique protéine virale va répondre à cette contrainte et va lever le blocage physique qui empêche l'entrée du virus dans sa cellule cible. Cette protéine est la protéine d'enveloppe Env. Pour pouvoir remplir sa fonction il est nécessaire que cette protéine soit exposée à la surface de la particule virale, là où elle est prête à interagir avec les composants de la membrane de la cellule cible du virus. Les propriétés uniques et le rôle crucial de cette protéine en font un objet de curiosité fort, d'autant plus que son exposition à la surface de la particule virale en fait une cible de choix pour n'importe quelle molécule inhibitrice et donc une cible idéale pour un traitement. Ce sont ces raisons qui ont incité notre laboratoire à prendre cette protéine comme sujet d'étude et qui ont fait d'elle le centre de ma thèse.

#### **3.1.11.1 Expression, synthèse et maturation de la protéine d'enveloppe**

Avant de rentrer plus en avant dans les fonctions de la protéine Env, il est nécessaire de revenir sur sa production par la cellule infectée. En effet, même si le chemin qui permet à cette protéine d'atteindre la membrane virale lors de la production des virions peut être deviné au travers de la mosaïque des précédents paragraphes, il est maintenant nécessaire de retracer ce chemin. Tout d'abord, l'ARNm qui permet la production de Env est le même qui permet la production de Vpu, protéine régulatrice impliquée dans le bon déroulement de la synthèse de Env (voir section « Vpu »).

Aussi, le fait que ces deux protéines soient codées par le même ARNm bicistronique a probablement été favorisé par l'évolution. Le mécanisme permettant de synthétiser les deux protéines n'est pas le même que pour le duo Gag/Gag-Pol : il n'est pas ici question de frameshift. En revanche, ici le codon initiateur AUG de Vpu est dans un contexte Kozak faible et est souvent ignoré par les ribosomes lors du scan du 5' de l'ARNm au moment de l'initiation de la traduction. Par contre, le codon initiateur de l'enveloppe se trouve dans un contexte Kozak fort et est très bien reconnu par les ribosomes, ce qui se traduit par un phénomène de « leaky scanning » permettant aux ribosomes de passer outre la séquence de Vpu et de continuer à scanner l'ARNm jusqu'à rencontrer le début de la séquence codant la protéine Env [99].

Une fois la synthèse de la protéine commencée, le chemin qu'elle emprunte devient logique si on garde à l'esprit que Env est ancrée dans la membrane de la particule virale, ce qui implique qu'elle est une protéine possédant au moins un domaine transmembranaire et donc que sa synthèse passe par le réticulum endoplasmique de la cellule [100]. La synthèse de Env dans la lumière du réticulum endoplasmique est permis grâce à un peptide signal, présent en N-ter de la protéine Env, qui permet aux ribosomes cytoplasmiques de synthétiser le peptide directement dans la lumière du réticulum en se fixant à un pore présent à la surface de celui-ci. Puis, un second signal présent juste après le domaine transmembranaire de la protéine permet au ribosome de stopper brièvement la synthèse le temps de se décrocher du pore du réticulum endoplasmique puis de poursuivre la synthèse dans le cytoplasme donnant ainsi naissance à une protéine transmembranaire. Le peptide signal est ensuite clivé à l'intérieur même du réticulum et ne sera donc pas présent sur la protéine mature.

Pour récapituler, à la fin de la traduction de l'ARNm, la protéine est constituée par un domaine N-ter se trouvant dans la lumière du réticulum séparé d'une courte extrémité C-ter se trouvant dans le cytoplasme par un domaine transmembranaire ancré dans la membrane du réticulum. Cette forme ne constitue pas la protéine finale mais n'est qu'un précurseur d'environ 160 KDa qui sera clivé ultérieurement par des protéases. De plus, il est important de préciser que Env n'est active que sous la forme d'un trimère et la formation de ces trimères débute immédiatement après synthèse par le ribosome par l'assemblage dans le réticulum des précurseurs encore non clivés. C'est encore dans le réticulum que débute une série d'importantes modifications post-traductionnelles de ces trimères d'enveloppes qui vont se poursuivre tout le long du transit de cette protéine vers la membrane cellulaire. Ces modifications consistent en de nombreuses N-glycosylation ainsi que quelques O-glycosylation, qui ensemble, vont compter pour près de la moitié de la masse de la protéine finale [101]. Ces modifications post-traductionnelles sont absolument nécessaire au bon fonctionnement de la protéine que ce soit de façon directe, car ces modifications contribuent à son activité, mais également de façon indirecte comme j'aurai l'occasion d'y revenir. Une fois que la protéine quitte le réticulum endoplasmique elle est ensuite dirigée vers l'appareil de Golgi où les glycosylations se

poursuivent et se complexifient et où se fait simultanément la maturation du précurseur par une protéase cellulaire : la furine. Il est nécessaire que ce soit une protéase d'origine cellulaire qui effectue cette tâche car les glycoprotéines d'enveloppe sont exposées à la surface des virions après le bourgeonnement de ceux-ci, et sont donc hors d'atteinte des protéases virales confinées dans les particules virale. Cette maturation par la furine va permettre de dissocier les deux protéines qui composent le précurseur et qui sont respectivement la gp120 et la gp41. Dans cette nomenclature, « gp » signifie « glycoprotéine » et le chiffre indique la masse moyenne de la protéine (glycosylations comprises), bien que comme toujours cette masse peut assez largement varier selon les souches de VIH-1 que l'on considère.

### **3.1.11.2 Structure des protéines d'enveloppe**

#### Considération générales

Une fois que les trimères quittent l'appareil de Golgi ils sont matures et fonctionnels et sont constitués de six sous-unités : trois gp120 et trois gp41. La séparation entre gp120 et gp41 traduit une séparation dans l'espace des deux fonctions que doit posséder la protéine d'enveloppe. La première activité est la capacité de reconnaître spécifiquement une cellule capable de soutenir la réplication virale et de s'y fixer. Cette fonction est remplie par la gp120 qui, afin de réaliser cette reconnaissance, est située entièrement à l'extérieur du virion sans même être ancré dans la membrane virale. Cette disposition un peu particulière s'explique par le fait que la gp120 est formée par la partie N-ter du précurseur gp160 et n'hérite pas de l'unique domaine transmembranaire du précurseur. Une autre « fonction » que doit remplir la gp120 est dictée par sa position. En effet, la position très exposée de la gp120 va de pair avec une contrainte supplémentaire : il est nécessaire que cette protéine soit capable d'échapper à l'action des anticorps neutralisants générés par le système immunitaire. Il est donc attendu que le gène codant la gp120 présente une diversité de séquences soutenue entre les souches VIH existantes. Produire de la diversité en réponse à l'action du système immunitaire est en effet un mécanisme utilisé largement par les pathogènes afin de contrer l'action des anticorps produits par le système immunitaire adaptatif [102,103].

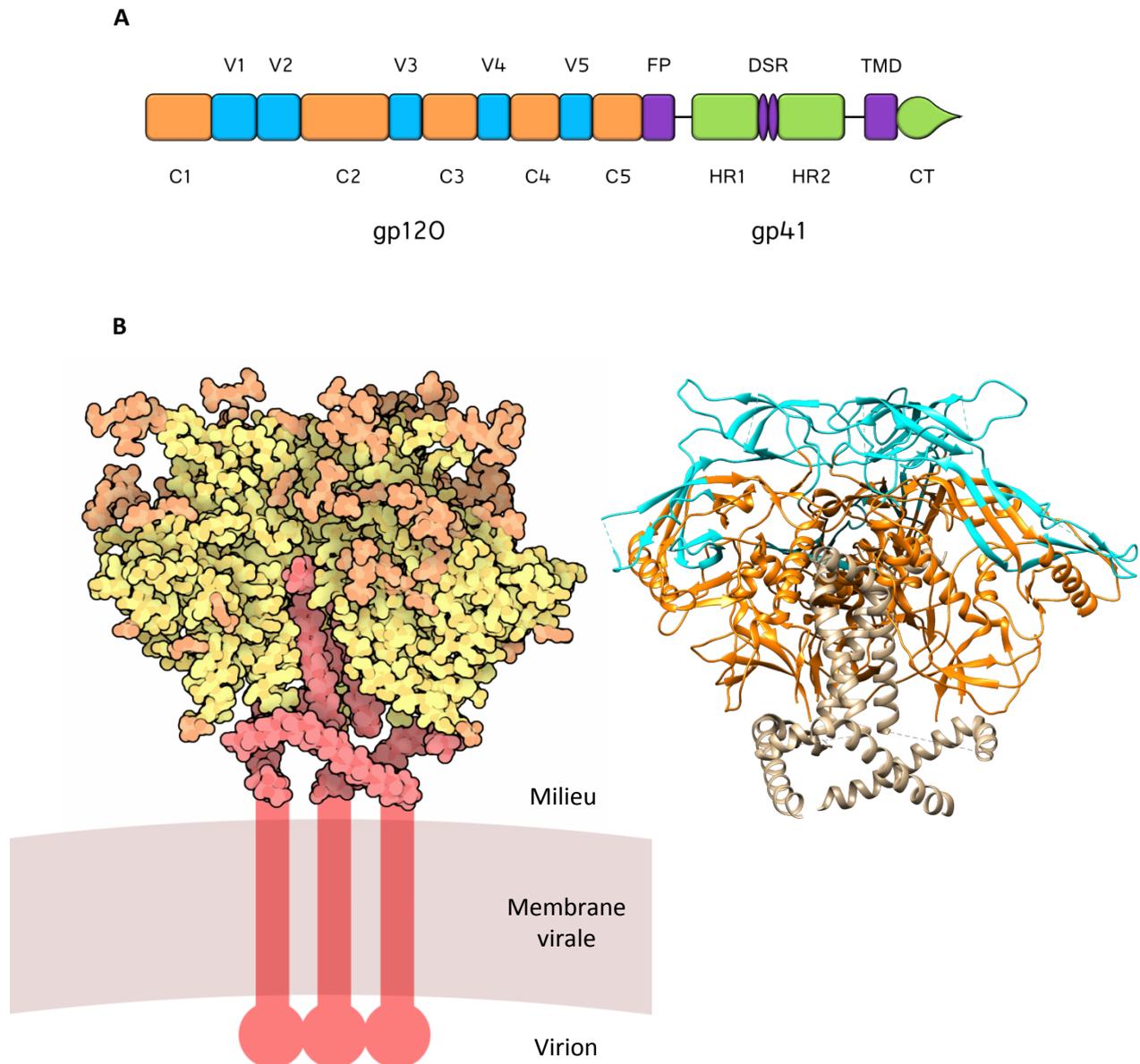
La seconde activité que doit porter un trimère est d'être capable de faire fusionner les membranes cellulaires et virales et cette activité est portée par la gp41. Pour réaliser cette action et agir mécaniquement sur les membranes cellulaires et virales, il est nécessaire que la gp41 dispose d'un ancrage physique dans chacune d'elle, il est donc naturel que ce soit elle qui ait hérité du domaine transmembranaire de la gp160. En revanche, étant donné que la gp120 ne dispose d'aucun domaine transmembranaire, il est donc nécessaire que la gp120 et la gp41 disposent d'une interface

permettant des interactions et empêchant le détachement de la gp120 [104]. Cette disposition implique que la gp41 soit en partie cachée par la présence de la gp120 et explique que la gp41 soit beaucoup moins sujette à l'action neutralisante des anticorps produits par l'hôte, même si elle n'en est pas complètement exempte. Ceci explique aussi que le gène de la gp41 présente moins de variabilité génétique que celui de la gp120.

### Structure de la gp120

Ensemble, les contraintes imposées par les besoins de pouvoir reconnaître et de se fixer à une cellule hôte potentielle et d'échapper aux anticorps produits par le système immunitaire dictent la structure de la gp120 aussi bien au niveau de son gène que de la protéine en elle-même. Le gène de la gp120 est en effet constitué d'une succession de régions dont la séquence est assez bien conservée entre les souches de VIH-1 existantes et de régions beaucoup plus variables. Pour cette raison ces régions sont nommées « constantes » (ou C en abrégé) et « variables » (ou V) et se suivent dans cet ordre C1, V1, V2, C2, V3, C3, V4, C4, V5 et C5 [105] (figure 13, A). Cette organisation se reflète sur la structure de la protéine car ce sont les régions constantes qui constituent le cœur de la protéine tandis que les régions variables émergent de ce cœur pour l'envelopper [106] (figure 13, B). On peut voir ici le résultat de l'action du système immunitaire qui impose une plus grande variabilité aux régions les plus exposées de la protéine. Cependant, c'est aussi l'action du système immunitaire qui a très certainement favorisé le fait que les activités de reconnaissance d'une cellule cible et de fixation à cette cellule de la gp120 ne se fassent que rarement par le biais de régions particulière mais soient d'avantage réalisées par des motifs conformationnels. De plus, ces motifs ne sont pas présents en permanence sur la gp120 mais au contraire se forment au fur et à mesure de l'activation de la protéine après la rencontre avec une cellule hôte potentielle.

Cette particularité complique d'avantage l'action des anticorps neutralisants. Même si des anticorps sont capables de se fixer sur un site crucial présent sur la gp120 afin d'inactiver cette protéine, il est très probable que ce site d'importance ne se formera que pendant le processus d'activation de cette protéine et donc pendant une fenêtre temporelle réduite. En contrepartie, la gp120 se doit d'être une molécule très flexible pour pouvoir garantir les changements de conformations nécessaires à la création de ces sites.<sup>107</sup>



**Figure 13 : Schéma d'un précurseur de glycoprotéine d'enveloppe et structure 3D d'un trimère de glycoprotéine d'enveloppes matures adaptés de [107] et [222]**

(A) Le précurseur gp160 est constitué de deux glycoprotéines : gp120 (orange et bleu) et gp41 (violet et vert). La gp120 est constituée d'une succession de régions constantes (notées C, en orange) et de régions variables (notées V, en bleu). La gp41 est constituée d'un peptide de fusion (FP), d'une première région constituée de répétitions d'heptades (HR1), d'une région contenant un pont disulfure (DSR), d'une seconde région constituée de répétitions d'heptades (HR2), d'une région transmembranaire (TMD) et d'une queue « cytoplasmique » (CT).

(B) Structure 3D d'un trimère de glycoprotéine d'enveloppe. (*gauche*) Structure 3D d'un trimère d'enveloppe où les atomes sont représentés. La gp120 est coloré en jaune et la gp41 en rouge. Les glycosylations de la gp120 ont été représentées en orange. Les formes rouges en têtes d'épingles représentent les régions de la gp41 qui n'ont jamais pu être cristallisées et qui correspondent grossièrement aux régions TMD et CT. (*droite*) Structure 3D d'un trimère d'enveloppe où les structures secondaires de la protéine sont représentées. La gp120 est en orange et bleu et la gp41 en beige. Les régions constantes de la gp120 sont en orange et les régions variables en bleu.

Une contradiction apparaît alors ici car visiblement la gp120 doit maintenir en même temps un taux de mutations élevé qui peut rapidement se révéler très délétère pour son activité et la capacité de réaliser des réarrangements complexes de sa structure, et ce sur demande, pour remplir son rôle vital pour le virus. Il est important ici de garder à l'esprit qu'une telle dichotomie existe, car comprendre comment celle-ci peut exister constitue le sujet de mon doctorat et ce sujet sera traité de façon beaucoup plus détaillée ultérieurement.

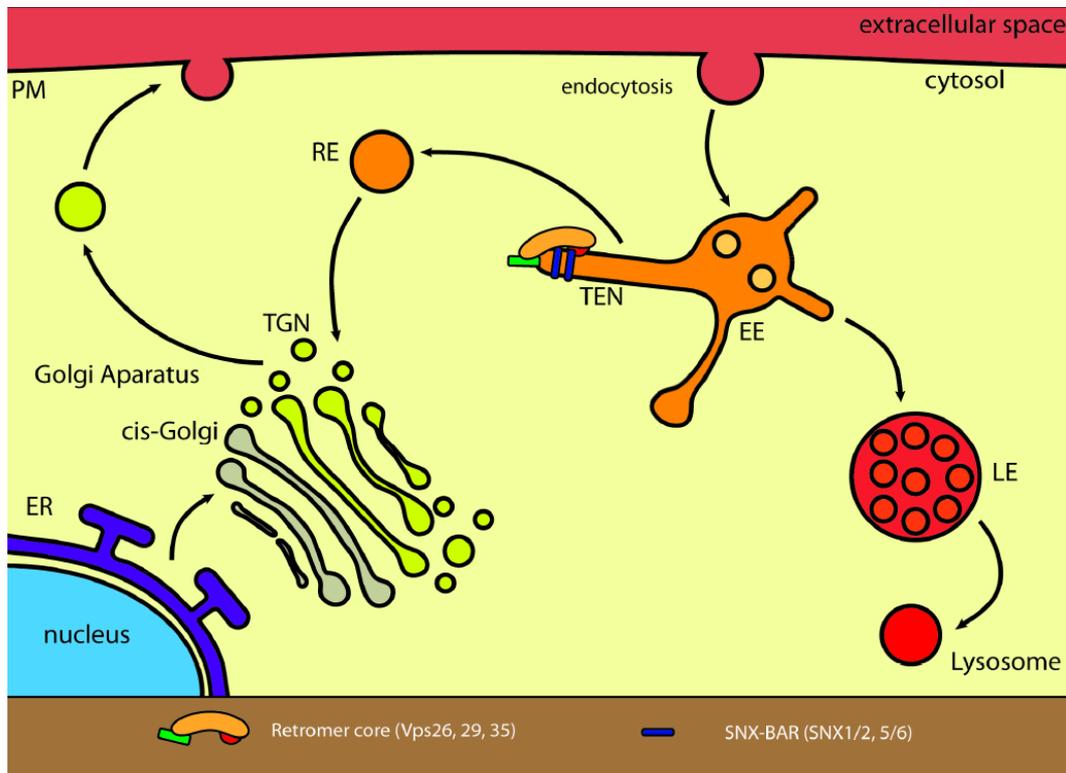
### Structure de la gp41

Comme précisé dans les sections précédentes, la gp41 est beaucoup moins impactée par l'action des anticorps neutralisants que la gp120. Ainsi, contrairement à la gp120, l'organisation de la gp41 est plus « simple » dans le sens où les régions qui la constituent sont d'avantage associées à une activité particulière et il n'est pas question ici de motifs conformationnels. Ces régions sont, de façon purement descriptive et sans détailler leurs fonctions pour l'instant, en N-ter de la gp41 le peptide de fusion (FP), ensuite vient une première région composée de répétitions d'heptades (HR1 pour heptad repeat one), séparée par la région DSR (disulfide bonded region) d'une deuxième région constituée de répétitions d'heptades (HR2) (figure 13, A). Après cette région HR2 se trouve la région MPER (membrane proximal external region) qui comme son nom l'indique est la dernière portion de la gp41 située à l'extérieur du virion. Enfin ce sont, d'une part le domaine transmembranaire (TMD), et d'autre part une queue cytoplasmique (CT) relativement longue pour un Retroviridae [108], qui viennent constituer l'extrémité C-ter de la protéine gp41.

#### **3.1.11.3. Recrutement de la protéine Env lors de la viriogenèse**

Une fois que le clivage protéolytique et les glycosylations sont parachevées dans l'appareil de Golgi, les trimères matures, constitués de trois gp41 sur lesquels sont maintenus trois gp120, sont dirigés vers la membrane plasmique. C'est à la membrane que se fait le recrutement des trimères par le précurseur Gag et c'est plus précisément la longue queue cytoplasmique de la gp41 qui interagit avec un domaine particulier présent dans la partie du précurseur Gag (domaine qui sera ensuite hérité par la protéine de matrice). Ce recrutement se fait au même endroit que se produit le bourgeonnement des virions, c'est-à-dire au niveau des « lipid rafts » [75].

Il est cependant très intéressant de noter que si la protéine Env n'est pas recrutée rapidement par un précurseur Gag après qu'elle soit parvenue à la membrane plasmique, elle est alors internalisée



**Figure 14 : Schéma du mécanisme de recyclage des glycoprotéines d'enveloppe de VIH-1 par le système « retromer » adapté de [111]**

Les protéines membranaires synthétisées dans le réticulum endoplasmique (ER) sont transférées dans l'appareil de Golgi pour y être maturée (TGN : réseau trans-golgien). Après maturation les protéines sont dirigées vers la membrane plasmique. A la membrane, les glycoprotéines virales sont endocytées si elles ne sont pas recrutées par un précurseur Gag. Dans le compartiment endosomal précoce (EE), les glycoprotéines sont reconnues par le système « retromer » constitué des protéines Vps 26, 29 et 35 qui en sont le core et de deux nexines parmi quatre possibles : SNX 1 ou 2 en complexe avec SNX 5 ou 6. Les nexines permettent de réarranger la membrane de l'endosome précoce pour créer une vésicule de recyclage endosomal (RE) qui est redirigé vers l'appareil de Golgi. Arrivé dans l'appareil de Golgi les protéines membranaires sont à nouveau dirigées vers la membrane plasmique et sont effectivement recyclées. Les protéines qui n'ont pas été reconnues par le système « retromer » restent dans l'endosome précoce. Celui-ci est mûri pour devenir un endosome tardif (LE) puis un lysosome dans lequel les protéines membranaires non recyclées sont dégradées.

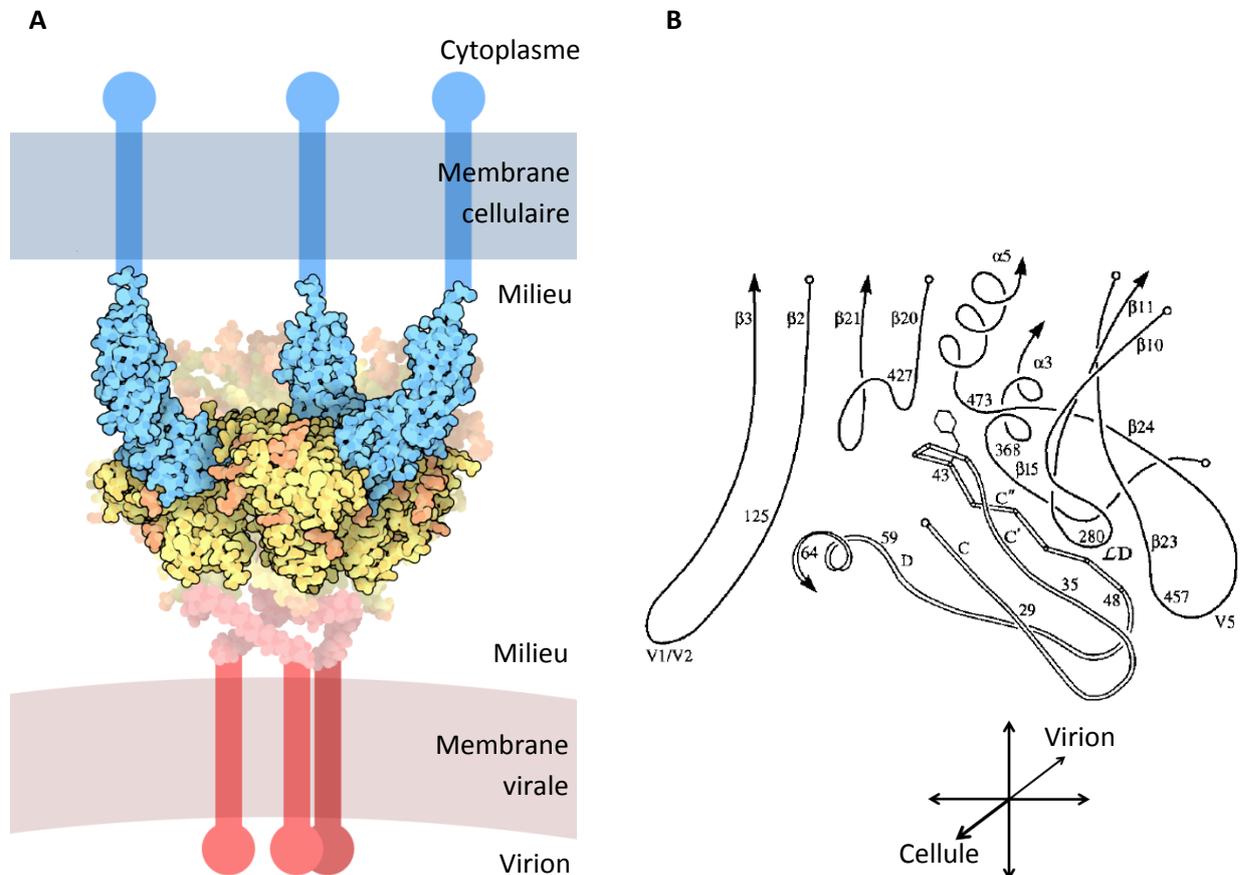
dans des endosomes [109]. Ce mécanisme peut sembler étrange au premier abord car sa conséquence immédiate est que la protéine d'enveloppe n'est présente qu'en faible quantité à la membrane cellulaire et, par voie de fait, est également incorporée en faible quantité à la surface des virions. Pourtant ce mécanisme s'explique très bien car la présence de cette protéine à la surface des cellules infectées et des particules virales font d'elles des cibles faciles pour le système immunitaire et notamment pour les anticorps neutralisants. Pour éviter ce problème, la solution sélectionnée par l'évolution consiste à réduire la visibilité de ces protéines en baissant leur présentation à la membrane cellulaire mais également à la membrane virale.

#### **3.1.11.4. Système « retromer »**

Pour éviter que les glycoprotéines d'enveloppes, assez immunogène, n'induisent une trop forte réaction immunitaire contre les virions, la présence de ces glycoprotéines aux membranes cellulaires comme virales est limitée. Néanmoins, dans le cas présent la mécanistique employée n'est pas la même que lorsque le virus réduisait la présence à la membrane cellulaire des molécules CD4 et CMH-I. En effet, il n'est pas question ici de dégrader une protéine gênante mais bien de camoufler une protéine essentielle. C'est pourquoi, après avoir été internalisées dans les endosomes, les protéines d'enveloppes ne sont pas dégradées par formation d'un lysosome mais sont captées et redirigées vers l'appareil de Golgi [109]. Ce mécanisme permet ainsi de limiter le temps d'exposition des protéines d'enveloppes à la membrane plasmique tout en évitant leur dégradation car, une fois dans le Golgi, ces protéines vont à nouveau être adressées à la membrane (figure 14).

Les trimères sont internalisés dans des endosomes grâce à un signal présent dans la partie cytoplasmique de la gp41. Ce signal est une séquence caractéristique de type YXX $\phi$  basée sur la présence d'une tyrosine séparée d'un acide aminé hydrophobe (une leucine dans la majorité des cas pour VIH-1) par deux acides aminés quelconques. Cette séquence est reconnue par la protéine cellulaire adaptatrice AP-2 qui permet le recrutement de nombreuses protéines de clathrines [110]. En polymérisant, les protéines de clathrines sont connues pour remodeler la membrane plasmique et permettre la création d'un endosome qui contiendra majoritairement ici des glycoprotéines d'enveloppes virales. Un second mécanisme moléculaire entre ensuite en jeu pour capter les protéines d'enveloppes présentes dans les endosomes et les rediriger vers l'appareil de Golgi.

C'est encore une fois la partie cytoplasmique de la gp41 qui va remplir ce rôle grâce à deux séquences appelées IS1 et IS2, composées respectivement d'une dizaine et d'une vingtaine d'acides aminés, qui vont recruter une autre série de facteurs cellulaires : les composants du complexe retromer [109].



**Figure 15 : Structure en 3D d'un trimère de Env en liaison avec trois molécule de CD4 et schéma de la zone de contact entre la gp120 et CD4 adaptés de [104] et [105]**

(A) La structure en 3D représente un trimère dont les gp120 sont colorés en jaune et les gp41 en rouge. Les glycosylations de la gp120 sont représentées en orange. Les trois molécules de CD4 sont colorées en bleu. Les formes en têtes d'épingles représentent les parties des molécules qui n'ont pas pu être cristallisées.

(B) Schéma de la zone de contact entre gp120 et CD4. La perspective adoptée correspond à un trimère de Env vu depuis la membrane cellulaire. Le domaine représenté en trait plein correspond aux séquences de la gp120 et celui en trait ouvert au CD4. Les cercles ouverts aux extrémités des séquences indiquent l'extrémité N-ter et les flèches l'extrémité C-ter. Les numéros indiquent les acides aminés principaux impliqués dans les interactions gp120-CD4. La seule chaîne latérale représentée est celle de la phénylalanine 43 du CD4 qui est l'acteur majeur de la liaison gp120-CD4. Les interactions principales impliquent les acides aminés 43 et 59 du CD4 ainsi que les résidus 368, 370 (non représenté pour des raisons de lisibilité), 427 et 473 de la gp120.

Ce complexe cellulaire est un hétéropentamère qui peut-être subdivisé en un trimère de protéines, chargées de sélectionner les protéines qui vont être redirigée vers l'appareil de Golgi, ainsi que deux autres protéines capable de remodeler la membrane de l'endosome pour séparer physiquement les protéines sélectionnées qui vont être envoyée vers l'appareil de Golgi du reste de l'endosome qui deviendra un lysosome. Le trimère est constitué des protéines Vps-26, Vps-29 et Vps-35 [111] et parmi ces protéines se sont les protéines Vps-26 et Vps-35 qui vont interagir avec les deux domaines IS1 et IS2 de la queue cytoplasmique de la gp41. Les deux autres protéines constituant un complexe retromer actif, et responsable du réarrangement membranaire, sont des nexines et sont soit SNX1 soit SNX2 en association avec SNX5 ou SNX6.

## **3.2. Phase Précoce**

### **3.2.1. Entrée virale**

#### **3.2.1.1. Liaison à CD4**

Grâce à l'utilisation combinée des signaux d'endocytose et de recrutement du complexe « retromer », les protéines d'enveloppes sont incorporées en quantités réduites, mais suffisantes, à la surface des virions. On estime en effet que ce sont entre 5 et 15 trimères d'enveloppes qui sont incorporés à la surface d'une particule virale [112-114]. Ces protéines vont alors rester en attente à la membrane virale et ne s'activeront que lorsque le hasard amènera un virion à proximité d'une cellule capable de supporter la réplication du virus. C'est la gp120 qui est capable de discriminer une cellule capable de soutenir la réplication du virus de toutes les autres en reconnaissant un marqueur membranaire présent exclusivement à la surface des cellules pouvant servir d'hôte : la protéine CD4 (figure 15). Cet état de fait avait déjà été introduit précédemment mais il était nécessaire de le rappeler par souci de clarté.

D'un point de vue moléculaire, le CD4 présente de nombreuses ressemblances avec les anticorps aussi bien du point de vue de l'organisation génétique que de la structure protéique. Le gène est en effet composé dans sa partie amont de deux régions présentant de nombreuses similitudes avec les régions V et J des gènes des immunoglobulines [115]. Ces similarités se répercutent sur la structure de la protéine et la font classer dans la superfamille des immunoglobulines. La protéine CD4 est en effet constituée d'une large région extracellulaire en N-ter, elle-même composée de quatre domaines immunoglobulines stabilisées par des ponts disulfures qui rappellent l'organisation des chaînes lourdes des anticorps [116]. Après cette large région se trouvent un domaine transmembranaire permettant de fixer la protéine sur la membrane plasmique puis une queue cytoplasmique qui vient constituer la partie C-ter de la protéine.

La région de la protéine CD4 reconnue par la gp120 se trouve à son extrémité N-ter et correspond au premier domaine immunoglobuline, celui qui est le plus éloigné de la membrane cellulaire. Se sont notamment 22 acides aminés, compris entre l'acide aminé numéro 25 et le numéro 64, qui établissent un contact avec 26 résidus présents sur la gp120 pour former de nombreuses liaisons hydrogènes et interactions van der Waals [106]. Le domaine de la gp120 qui entre en contact avec le CD4 est plus délicat à décrire. A cause de la nature conformationnelle des sites de la gp120, les 26 acides aminés établissant des interactions avec le CD4 sont proches les uns des autres dans l'espace mais ils sont très épars lorsque l'on se réfère à la structure primaire de la protéine, ils sont répartis entre les régions V2, C2, C3, C4 et C5 avec les régions C4 et C5 comme constituants majoritaires de ce domaine.

Il est nécessaire pour le virus d'engager autant de contacts avec le récepteur cellulaire car la petite masse des virions représente un avantage lors de la dissémination (les particules virales sont sensibles aux moindres remous du milieu extracellulaire et sont portées par lui), mais cette petite masse se transforme en désavantage dès que les virions doivent se maintenir au même endroit pour réaliser une action. Les nombreuses interactions entre le CD4 et la gp120 servent donc à fournir un ancrage permettant au virion de rester en contact avec sa cellule cible le temps que la fusion des membranes se produise ; phénomène qui prend une quantité non négligeable de temps car la durée nécessaire pour que la moitié d'une population virale donnée parvienne à réaliser l'étape de fusion membranaire ( $t_{1/2}$  de fusion) avoisine les 50 minutes même dans les cinétiques les plus rapides [117,118].

Parmi les nombreuses et complexes interactions qui se font entre la gp120 et le CD4, il existe un ensemble d'interactions qui sont absolument cruciales non seulement pour l'ancrage mais aussi pour le bon déroulement de l'entrée virale. Il s'agit de contacts multiples existant entre la phenylalanine 43 et l'arginine 59 du CD4 d'une part et l'asparagine 368, l'acide aspartique 370, le tryptophane 427 et la glycine 473 de la gp120 d'autre part (figure 15, B). Les quatre acides aminés appartenant à la gp120 forment le fond d'une cavité et lorsque le hasard amène un virion au contact d'une cellule, la phenylalanine 43 du CD4 peut se glisser dans cette cavité et engager des contacts avec les quatre acides aminés composants le fond de celle-ci. Cette interaction centrale est renforcée par la chaîne latérale de l'arginine 59, située non loin, et qui à son tour initie des contacts avec les quatre acides aminés de la gp120 [106]. Cette série d'interactions permet de favoriser des changements de conformation de la gp120 et surtout stabiliser la gp120 dans sa nouvelle conformation [119]. L'une des différences majeures entre le trimère d'enveloppe non activé et la conformation liée au CD4 est la création d'une structure secondaire en feuillet  $\beta$  antiparallèle appelée le « bridging sheet » qui se forme à la verticale de la gp41 à partir de séquences des régions constantes C1, C2 et C4. L'autre

évènement majeur est le fait que la boucle variable V3 quitte sa position initiale dans laquelle elle est plaquée contre les parties constantes de la gp120 pour être projetée vers la membrane cellulaire.

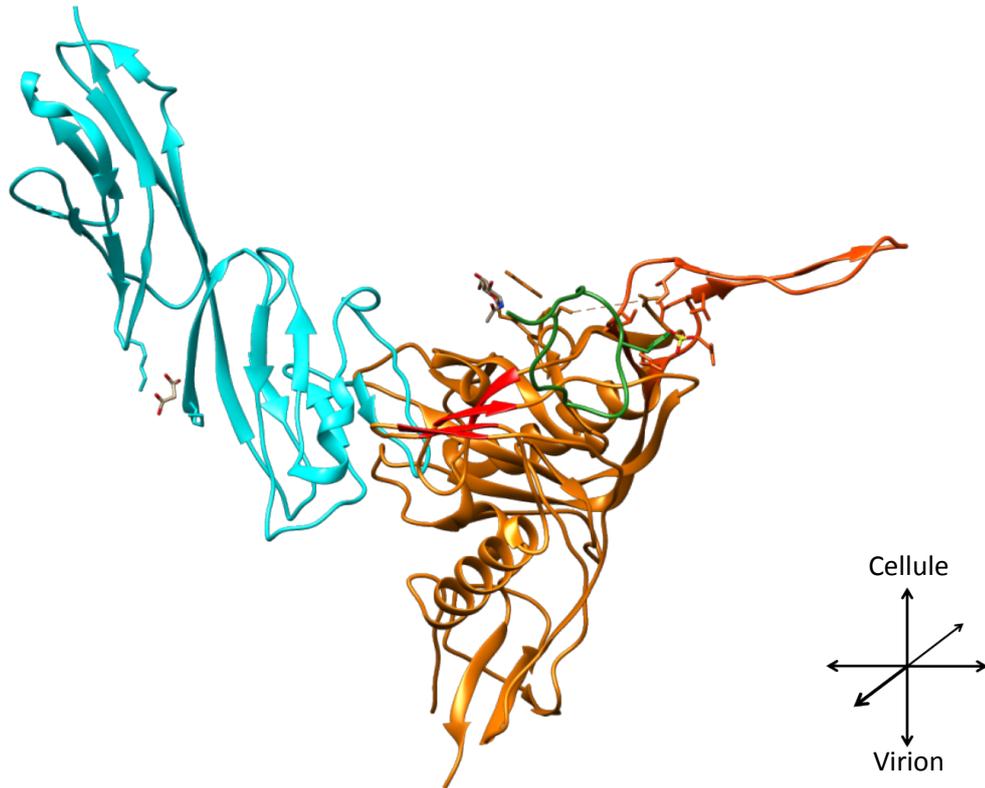
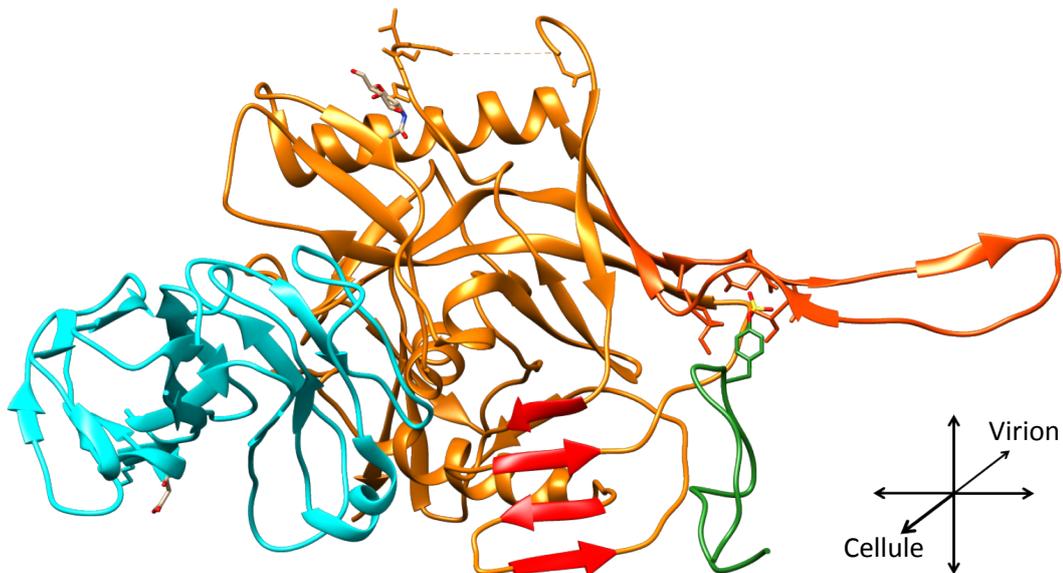
### 3.2.1.2. Liaison au corécepteur

Existence de différents corécepteurs et conséquences pour le VIH

Le « bridging sheet » et la région V3 sont les acteurs principaux de l'étape suivante du processus d'entrée viral qui est la reconnaissance du corécepteur. Ce corécepteur peut varier selon les souches de VIH-1 mais deux corécepteurs membranaires majeurs existent et sont CCR5 et CXCR4, qui définissent ainsi deux grandes familles de souches virales : les souches dites R5 tropiques qui lient le CCR5 et les souches X4 tropiques qui reconnaissent le CXCR4 [120-123].

Ces deux grandes familles de souches virales se distinguent donc naturellement par les types cellulaires qu'elles sont capables d'infecter mais également par une autre propriété : la capacité à induire la formation de syncytia ou non [124]. En l'occurrence les souches R5 tropiques infectent majoritairement les monocytes et les macrophages, de plus ces souches n'induisent pas la formation de syncytia quand leurs glycoprotéines d'enveloppes sont exprimées seules à la surface de cellules en culture. Les souches X4 tropiques en revanche infectent principalement les lymphocytes T CD4<sup>+</sup>. Ces cellules sont une population de cellules lymphocytaires possédant à la fois le récepteur CD4 et le corécepteur CXCR4. Les glycoprotéines d'enveloppes de ces souches virales X4 tropiques sont connues pour induire la formation de syncytia lorsqu'elles sont exprimées à la surface de cellules en culture.

Cependant, ces deux types de virus ne représentent pas deux groupes phylogénétiques distincts. Que ce soit lors de la primo-infection ou lors de la phase asymptomatique, les souches VIH-1 qui infectent un patient sont essentiellement R5 tropiques, alors qu'en revanche lors de la phase SIDA, ou tout du moins lorsque le système immunitaire est suffisamment affaibli, des souches X4 tropiques ont tendances à spontanément apparaître [125,126]. Beaucoup plus intéressant, si ces souches X4 tropiques sont transmises à un nouvel individu lors d'une infection, elles disparaissent assez rapidement chez ce nouvel hôte et seules les souches R5 tropiques persistent et donnent naissance aux lignées virales à l'origine de l'infection chronique [127]. Le fait que les souches R5 tropiques puissent spontanément muter pour s'adapter à l'utilisation du CXCR4 s'explique par les nombreux points communs qui existent entre ces deux corécepteurs et qui facilitent le changement de tropisme [128]. En dehors de leur rôle dans l'infection à VIH-1, CCR5 et CXCR4 sont tous les deux des récepteurs couplés à des protéines G et appartiennent tous les deux à la famille des récepteurs de chemokines.

**A****B**

**Figure 16 : Structures en 3D d'une gp120 en liaison avec une molécule de CD4 et l'extrémité N-ter d'une molécule de CCR5 adaptés de [131]**

(A) La vue a été prise « de profil » avec la membrane cellulaire en haut et la membrane virale. Le CD4 est coloré en bleu, la gp120 en orange et l'extrémité N-ter de CCR5 en vert. La région V3 de la gp120 est coloré en orange-rouge et le « bridging sheet » en rouge.

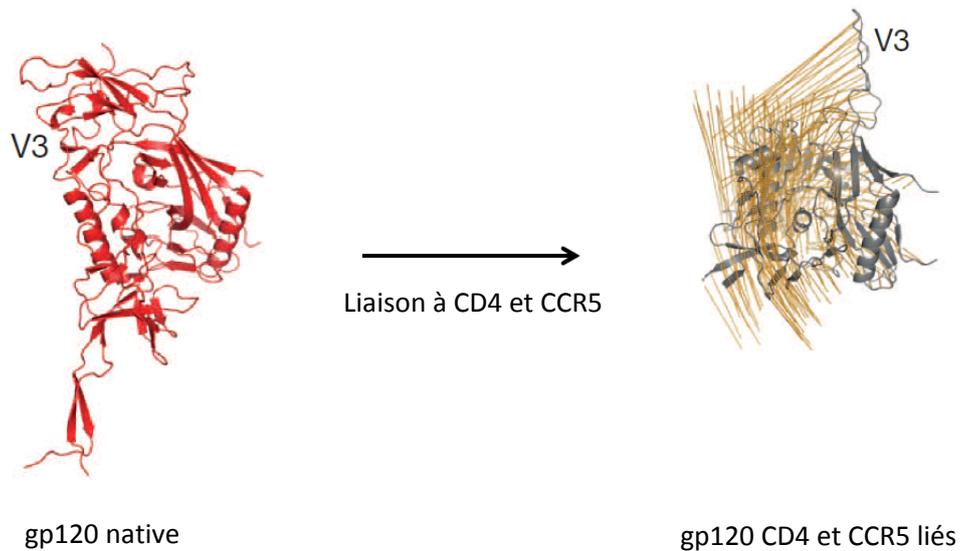
(B) La vue a été prise depuis le point de vue de la membrane cellulaire en regardant vers la membrane virale. Le code couleur est le même que précédemment. La seule chaîne latérale représentée de CCR5 (vert) correspond à la sulfotyrosine qui initie le contact avec V3. Les résidus de V3 qui interagissent avec la sulfotyrosine ont leurs chaînes latérales visibles (orange-rouge).

Ces deux protéines partagent une organisation très similaire et comptent toute deux sept domaines transmembranaires, une longue extrémité N-ter extracellulaire permettant la reconnaissance du ligand naturel, une chemokine en l'occurrence, et une longue extrémité C-ter intracellulaire permettant la liaison aux protéines G et l'activité du récepteur [129].

Néanmoins, ce sont les souches R5 tropiques qui jouent le rôle le plus important dans la transmission de la maladie, donc par défaut lorsqu'il sera question de glycoprotéines d'enveloppes dans la suite de cette thèse, il sera question de glycoprotéines issues de souches R5 tropiques. Le fait de passer sous silence les souches X4 tropiques n'est pas un problème majeur car à part les différences qui ont été décrites plus en amont, le processus d'activation lors de l'entrée virale, les mécanismes de neutralisation par les anticorps et le comportement général des deux types de glycoprotéines d'enveloppes sont presque en tout point identiques.

### Liaison à CCR5

La liaison à CCR5 s'effectue principalement par l'interaction de CCR5 avec la boucle V3 de la gp120 (figure 16, A). Ce sont les acides aminés 298, 302 et 303, situés non loin de la base de V3, qui vont initier des interactions avec des sulfo-tyrosines présente dans la partie N-ter de CCR5 [130] (figure 16, B). La sulfatation d'une tyrosine est une modification post-traductionnelle assez rare, aussi les tyrosines qui subissent ces modification ne sont pas toujours les mêmes selon les molécules de CCR5 : la partie N-ter de CCR5 comporte quatre tyrosine en positions 3, 10, 14 et 15 et n'importe laquelle de ces tyrosine est susceptible de subir cette modification post-traductionnelle [131]. En revanche il a été décrit qu'une sulfatation des tyrosines 10 et 14 ensemble est suffisante pour permettre l'activité de la gp120. L'extrémité de la boucle V3 va elle aussi participer à la reconnaissance de CCR5 en interagissant avec un domaine appelé ECL2 qui est constitué par les quelques acides aminés servant de jonction entre les domaines transmembranaires 4 et 5 de CCR5 [132]. D'autres interactions, moins bien décrites, existent aussi entre l'extrémité N-ter de CCR5 et des acides aminés composant le « bridging sheet » de la gp120 et viennent compléter l'ensemble des interactions nécessaires à la poursuite du processus de fusion membranaires. Ces contacts entre la boucle variable V3, le « bridging sheet » et CCR5 permettent d'ultimes réarrangements structuraux de la gp120 qui s'accompagnent d'un rapprochement des membranes cellulaires et virales [133].



**Figure 17 : Changements conformationnels réalisés par une sous-unité gp120 au cours de l'entrée virale adapté de [135]**

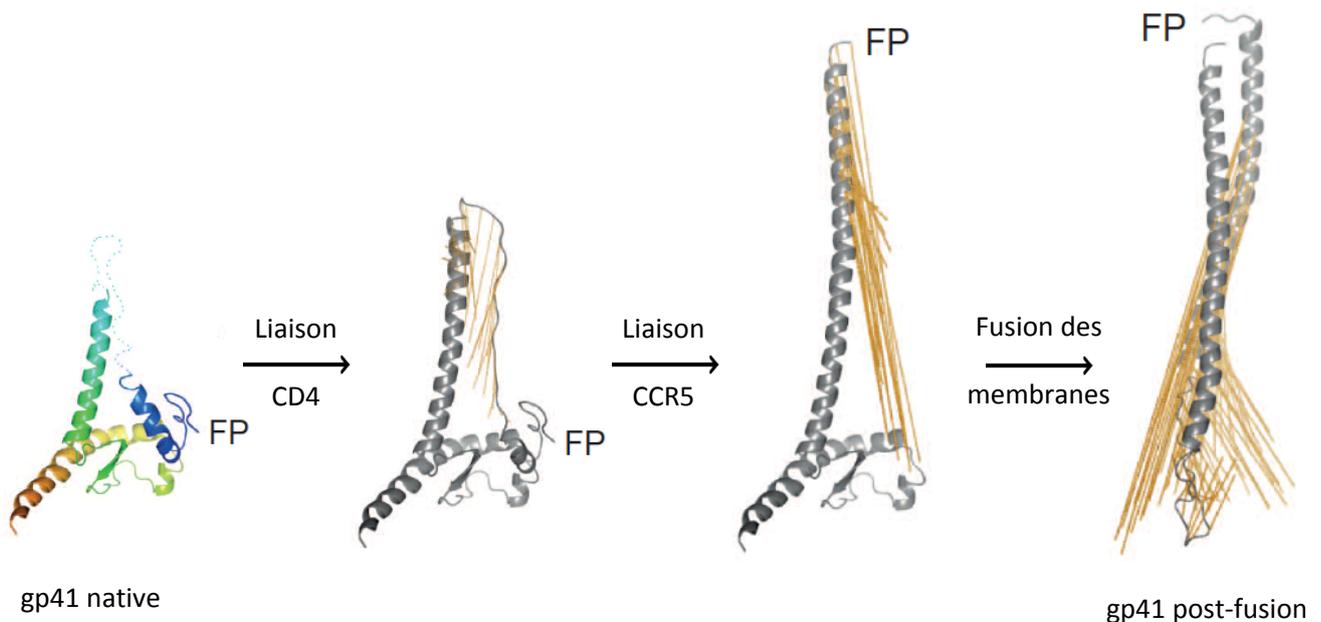
La structure en rouge est celle d'une sous-unité gp120 dans sa conformation native tandis que la structure en gris correspond à une gp120 dans sa conformation CD4 et CCR5 lié. Les lignes orangées représentent les trajectoires des carbones  $\alpha$  des acides aminés entre la structure de gauche et celle de droite. La position de la boucle V3 est signalée par son nom. Les changements de conformation ont été modélisés par ordinateur à partir de structures cristallographiques publiées et ne prennent en compte que les changements entre l'état natif et l'état final, CD4 et CCR5 lié, de la gp120.

Ces ultimes changements de conformations de la gp120 (figure 17) permettent d'une part de dévoiler la gp41, qui était jusque-là enfouie sous la gp120, et d'autre part permettent à la gp41 de s'activer. Aussi, le rapprochement des membranes cellulaires et virales, permis par les changements de conformations de la gp120, offre à la gp41 un contexte favorable à son activité. Malheureusement les changements structuraux qui permettent à la gp120 de dévoiler la gp41 sont assez mal connus. La seule certitude est que l'entrée en action de la gp41 s'accompagne très souvent d'un détachement complet de la gp120 qui est perdue dans le milieu extracellulaire, un phénomène appelé shedding [134]. Il est donc logique de penser que ces réarrangements mal connus, mais nécessaires, de la gp120 induisent une diminution des interactions avec la gp41. D'ailleurs, il n'est pas impossible que le shedding de la gp120 fasse non seulement partie du processus normal d'activation du trimère d'enveloppe mais qu'il soit nécessaire à cette activation bien qu'aucune preuve n'existe formellement pour soutenir cette hypothèse.

### **3.2.1.3. Fusion des membranes cellulaires et virales**

#### Structure native de la gp41

Lorsqu'elle est dans son état natif, la partie extracellulaire de la gp41 est composée essentiellement par quatre hélices alpha,  $\alpha 6$ ,  $\alpha 7$ ,  $\alpha 8$  et  $\alpha 9$  reliées entre elles par des séquences peu structurées qui encerclent les extrémités N-ter (inclue dans C1) et C-ter (inclue dans C5) de la gp120 à la manière d'un anneau [135]. On peut retrouver dans cette disposition particulière le moyen qui permet à la gp120 d'être maintenue à la surface de la particule virale et qui a été évoqué précédemment. Pour revenir au mécanisme de fusion des membranes, la clé qui permet de comprendre le fonctionnement de cette protéine réside dans cette conformation native. En effet, à ce stade le problème majeur qui est l'absence apparente d'une source d'énergie permettant de subvenir au coût énergétique du remaniement des membranes n'a toujours pas trouvé de réponse. Or il se trouve que la conformation native de la gp41 n'est que métastable et elle n'est maintenue dans cet état que par les interactions qu'elle entretient avec la gp120 [136]. Lors de l'affaiblissement de ces interactions au cours du processus de reconnaissance du récepteur et du corécepteur, la gp41 est libre de quitter cet état thermodynamiquement défavorable et d'effectuer les changements de conformations qui lui donneront un état stable. C'est donc l'énergie potentielle contenue dans la conformation native métastable des protéines gp41 qui est utilisée par le virus comme source d'énergie lors de la fusion des membranes.



**Figure 18 : Changements conformationnels réalisés par une sous-unité gp41 au cours de l'entrée virale adapté de [135]**

La structure multicolore est celle d'une sous-unité gp41 en conformation native, seules les régions HR 1 et HR2 sont représentées. Les structures en gris sont celles d'une sous-unité gp41 au cours de ses différents changements de conformations. Les lignes orangées représentent les trajectoires des carbones  $\alpha$  des acides aminés entre les structures. La position du peptide de fusion (FP) est signalée par son nom. Les changements de conformation ont été modélisés par ordinateur à partir de structures cristallographiques publiées.

Sur la structure de la gp41 native, l'hélice  $\alpha$  en bleu foncé est  $\alpha_6$  ;  $\alpha_7$  est en bleu clair-vert ;  $\alpha_8$  est en vert (en partie caché derrière  $\alpha_6$ ) ;  $\alpha_9$  est en jaune-rouge. La liaison au CD4 permet à l'hélice  $\alpha_7$  de s'allonger à partir de régions peu structurées. La liaison à CCR5 permet la formation complète de la région HR1 à partir des hélices  $\alpha_6$  et  $\alpha_7$  et permet d'ancrer le peptide de fusion dans la membrane cellulaire. Enfin la formation de HR1 entraîne la formation de HR2 à partir des hélices  $\alpha_8$  et  $\alpha_9$  ce qui permet de déclencher la fusion des membranes cellulaires et virales.

## Activation de la gp41

Selon les plus récentes études [135] les tout premiers changements de conformations de la gp41 s'effectuent immédiatement après la liaison de CCR5 par la gp120 et permettent à la région HR1 de former une longue hélice  $\alpha$  par le biais du dépliement et du repliement de deux hélices préexistantes :  $\alpha_6$  et  $\alpha_7$  (figure 18). Ce changement de conformation, dont la cinétique évoque l'allongement d'une tige en tire-bouchon, permet au peptide de fusion, FP, de s'ancrer dans la membrane cellulaire.

Le changement conformationnel suivant de la gp41 est complètement inféodée au relâchement des interactions qu'elle entretient avec la gp120, voir au shedding de cette dernière. La diminution de la force de ces interactions permet aux hélices  $\alpha_8$  et  $\alpha_9$  d'être libérées et de former l'hélice  $\alpha$  HR2. Les trois domaines HR2 nouvellement formés des trois gp41 d'un trimère d'enveloppe vont ensuite s'associer avec les trois hélices HR1 pour former une structure en super-hélice, ou structure « coiled-coil » (figure 18). Dans la forme stable finale de cette super-hélice les parties N-ter des domaines HR1 et les parties C-ter des HR2 sont à proximité immédiates les unes des autres. Ceci signifie que dans la conformation « coiled coil » de la gp41, les peptides de fusion ancrés dans la membrane cellulaire se trouvent à proximité immédiate des domaines transmembranaires ancrés dans la membrane virale. C'est donc la formation de la super-hélice à partir des domaines HR1 et HR2 qui donne l'impulsion nécessaire à la fusion des membranes cellulaires et virales en cherchant à rapprocher dans l'espace les régions FP et les régions transmembranaires des gp41 d'un trimère.

De façon tout à fait remarquable sur la dizaine de trimères d'enveloppes incorporée à la surface de la particule virale lors du bourgeonnement, seule l'activation d'une partie d'entre eux est nécessaire à une fusion membranaire efficace. Bien que le nombre de trimère dont l'activation est nécessaire varie selon les souches de VIH-1 étudiées, ce nombre varie généralement entre 4 et 7 pour les souches les plus répandues et peut descendre jusqu'à 2 trimères seulement pour certaines souches réalisant une fusion membranaire particulièrement rapide [113,114].

Pourtant, l'étape de formation de la super-hélice, même si elle marque bel et bien la fin de l'implication du trimère d'enveloppe, ne constitue pas tout à fait la fin du processus d'entrée virale. Il est nécessaire que la cellule cible, elle-même, intervienne à un moment donné et parachève le processus de fusion membranaire initié par le trimère d'enveloppe. En effet, la fusion de deux vésicules membraneuses passe par des étapes intermédiaires dont la première est la formation d'une structure appelée hémifusion. Dans cette structure les deux couches lipidiques externes des vésicules se mélangent sans que les couches lipidiques internes des deux vésicules ne le fassent [137]. Ce n'est qu'ensuite qu'un pore se forme qui permet de faire communiquer le contenu des deux vésicules, avant l'étape finale qui consiste en l'effondrement des deux vésicules qui fusionnent pour

n'en donner qu'une [138]. Chacune de ces étapes est en équilibre dynamique avec la suivante ou la précédente, aussi le rôle principal du trimère d'enveloppe, en plus de catalyser le remaniement des membranes, est de déplacer ces équilibres pour que le processus n'évolue que dans le sens de la fusion complète des membranes. En revanche il a été démontré que l'action du trimère d'enveloppe seul ne permet pas de franchir l'étape de l'hémifusion [139,140]. Sans action extérieure le système ne progresse pas jusqu'à la création d'un pore ni à l'effondrement de la membrane virale dans la membrane cellulaire, ceci interdit à la capsid virale de passer dans le cytoplasme. Ce constat peut paraître étrange car la conclusion est qu'un virion seul est incapable de parvenir à infecter une nouvelle cellule hôte sans intervention extérieure. Cependant, compte tenu des conditions extrêmement précaires dans lesquelles se trouve le virion, réussir à provoquer la formation d'une structure en hémifusion permettant un contact étroit avec la cellule cible en ne puisant uniquement que dans de faibles réserves d'énergie potentielle témoigne déjà de la sophistication de la glycoprotéine d'enveloppe. De plus parvenir à l'hémifusion des membranes est suffisant pour le virus, car une fois parvenu à cette étape c'est la cellule cible qui va elle-même se charger de compléter la fusion des membranes et va, ironiquement, ainsi participer à sa propre infection comme décrit dans la section suivante.

### Internalisation dans un endosome

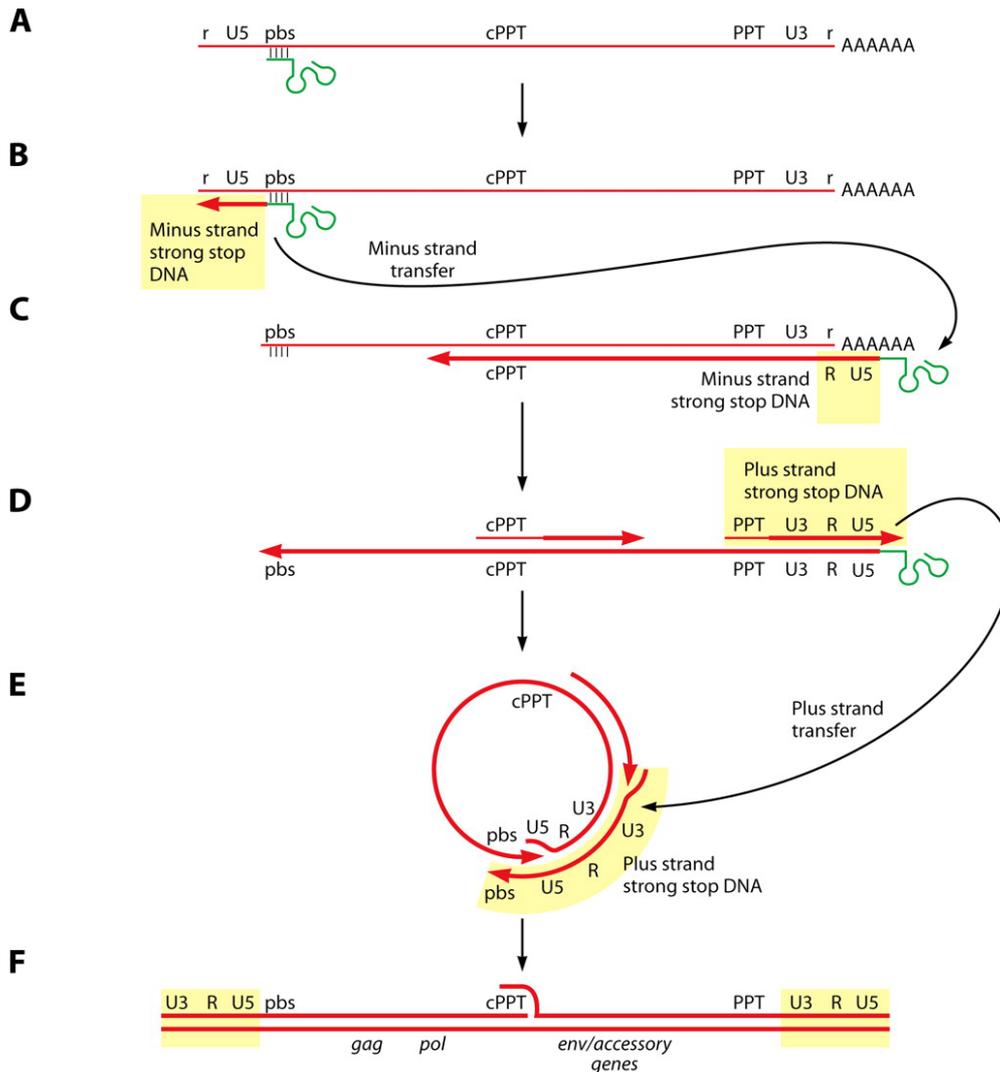
Avant d'en venir aux ultimes étapes du processus de fusion des membranes cellulaires et virales il est d'abord nécessaire de réaliser un récapitulatif et surtout de quitter l'échelle moléculaire pour utiliser l'échelle cellulaire. La rencontre entre le virus et sa cellule cible s'effectue à la membrane plasmique de celle-ci. La reconnaissance du récepteur CD4 et du corécepteur par la gp120 se font aussi au niveau de la membrane plasmique selon tous les modèles actuels. En revanche il a longtemps été soutenu que la fusion des membranes se faisait également à la membrane plasmique, en vertu de la capacité du VIH à promouvoir la fusion des membranes sans avoir besoin d'un milieu acide [141]. Jusqu'à assez récemment, les modèles de fusion membranaires établissaient une distinction nette entre les virus nécessitant un pH acide pour réaliser leur entrée, et qui avaient donc besoin d'être internalisé dans un endosome afin de profiter de l'acidification progressive de celui-ci, et les virus dit pH indépendants dont on pensait qu'ils réalisaient l'étape d'entrée directement à la membrane plasmique [142,143]. Les travaux les plus récents contestent cet état de fait et ce plus principalement pour le VIH-1. Nous savons aujourd'hui que le VIH a besoin d'être internalisé dans un endosome pour que la fusion des membranes parvienne à son terme [139]. En revanche ce virus est bel et bien pH indépendant, et même sensible au pH faible, ce qui représente une contrainte car le

virion doit parvenir à pénétrer le cytoplasme de la cellule cible avant que l'endosome ne s'acidifie, ce qui lui serait fatal.

L'action du trimère d'enveloppe seul n'est pas suffisante pour faire progresser la fusion des membranes au-delà de l'étape d'hémifusion. Cependant, la fixation de trimères d'enveloppe à CCR5 induit l'internalisation du virion dans un endosome, ce qui permet la résolution de la structure en hémifusion et la fusion complète des membranes [139]. Le mécanisme moléculaire qui se tient derrière cet état de fait reste obscur, plusieurs théories circulent sans qu'aucune d'elle n'ait été clairement prouvée ou réfutée aujourd'hui en 2016. Parmi ces théories deux grandes idées s'affrontent, la première veut que l'internalisation dans un endosome fournisse seulement les conditions environnementales nécessaires à une reprise plus ou moins automatique de la fusion. En effet, la membrane plasmique est un milieu assez complexe et surtout « hostile » à l'entrée virale : de très nombreuses protéines sont présentes et encombrant la surface, la présence de rafts lipidiques modifie localement les propriétés physico-chimique de la membrane plasmique et enfin, juste derrière cette membrane se trouve un réseau dense de filament d'actine qui est un frein supplémentaire à l'entrée virale [140]. Ces contraintes sont en revanche absentes ou largement atténuées dans l'environnement endosomal. La seconde proposition pour expliquer le besoin d'un endosome stipule que c'est un facteur cellulaire, encore à identifier, qui agit sur la structure en hémifusion et induit la reprise de la fusion des membranes par son action. Dans ce modèle ce facteur cellulaire inconnu serait vraisemblablement un membre de la machinerie protéique cellulaire pilotant la formation et/ou le devenir des endosomes.

### **3.2.2. Libération de la capsid virale dans le cytoplasme**

La conséquence directe de la réussite de l'entrée virale est la libération de la capsid virale ainsi que des protéines de matrice dans le cytoplasme d'une nouvelle cellule hôte. Une fois que la capsid se trouve dans le cytoplasme, la présence de diverses protéines et éléments cellulaires vont induire le déclenchement de trois évènements intimement liés : le démarrage des réactions de rétro-transcription du génome viral, la désintégration de la capsid virale, phénomène appelé « uncoating » ou « décapsidation », et enfin le transport du génome viral vers le noyau cellulaire. Il a été démontré que ce sont les réactions de rétro-transcription ayant lieu à l'intérieur de la capsid virale qui initient sa désintégration [144], et il a aussi été établi que l'interaction entre la capsid virale et les microtubules qui servent à transporter le génome vers le noyau est essentielle au bon déroulement de la décapsidation [145]. C'est pourquoi, malgré le découpage artificiel qui sera fait entre ces étapes dans la suite de ce manuscrit, celles-ci ne doivent pas être vues comme trois phases



**Figure 19 : Schéma du mécanisme de transcription inverse chez VIH-1 adapté de [155]**

(A) L'ARN génomique viral (trait rouge fin) est associé à un ARNt (vert) au niveau du site PBS.

(B) La RT utilise l'ARNt comme amorce pour synthétiser l'ADN complémentaire (-) à la région 5' de l'ARN génomique (trait rouge épais). Une fois arrivé à l'extrémité 5', la RT stoppe faute de matrice : c'est le « minus strand strong stop ». La partie 5' de l'ARN viral est dégradé par l'activité RNase H de la RT.

(C) Un changement d'ARN matrice permet à la région R « ADN » de s'associer à la région R « ARN » présente en 3' de l'ARN, ceci permet de poursuivre la synthèse du brin d'ADN. L'AR viral est dégradé par l'activité RNase H de la RT à l'exception de deux régions, le PPT central (cPPT) et le PPT 3' (PPT).

(D) Le brin ADN (+) est synthétisé en utilisant les PPT comme amorces. L'ARNt est dégradé par l'activité RNase H de la RT. La synthèse stoppe à l'extrémité faute de matrice, c'est le « plus strand strong stop ».

(E) Un nouveau changement de matrice s'opère par association des deux régions PBS complémentaires et par circularisation de l'ADN. Chaque brin d'ADN sert de matrice pour compléter la synthèse de son brin complémentaire en utilisant les deux séquences PBS comme amorce.

(F) En fin de synthèse, l'ADN viral double brin comporte deux LTR entier à chacune de ses extrémités et une courte région triple brin à l'endroit du PPTc.

séparées dans l'espace et le temps mais bien comme un continuum d'événements qui s'entrelacent.

### **3.2.3. Transcriptase inverse et transcription inverse**

La rétro-transcription, phase du cycle viral caractéristique aux rétrovirus consiste à convertir le génome viral encapsidé sous forme d'ARN pour en faire un ADN double brin, qui va s'intégrer dans le génome de la cellule hôte (figure 19). Les réactions de rétro-transcription démarrent dès que le dimère de transcriptase inverse entre en contact avec le pool de nucléotides présent dans le cytoplasme de la cellule. Des événements précoces de décapsidation permettent en effet aux nucléotides d'accéder au core viral. La transcriptase inverse est une ADN polymérase capable d'utiliser à la fois des molécules d'ARN et d'ADN comme matrice. Comme toutes les ADN polymérases connues, cette enzyme virale ne peut pas démarrer la synthèse du brin complémentaire d'une matrice donnée sans la présence d'une région partiellement double brin. Il est en effet, indispensable que cette enzyme dispose d'une extrémité libre 3'OH sans quoi la première substitution nucléophile  $S_N2$ , qui permet d'incorporer le tout premier nucléotide, ne peut se faire [146].

La solution sélectionnée au cours de l'évolution pour résoudre ce problème consiste à utiliser un ARN de transfert (ARNt) issu de la cellule hôte comme amorce pour les premières réactions de polymérisation. Il est important de noter que l'ARNt n'est pas sélectionné une fois que le virus arrive dans le cytoplasme de sa nouvelle cellule hôte mais est encapsidé avec l'ARN viral au moment du bourgeonnement du virion. Ce sont deux sites présents sur l'ARN viral, le primer binding site (PBS) situé en aval de la région U5, et le primer association site (PAS), situé dans la région U5 qui ensemble, lient physiquement l'ARN de transfert au génome viral en interagissant avec lui par complémentarité de séquence. Cette complémentarité de séquence implique que l'ARNt n'est pas choisi au hasard parmi la population d'ARNt présente dans le cytoplasme mais est sélectionné spécifiquement par l'ARN génomique. Dans le cas de tous les virus de l'immunodéficience humaine, y compris VIH-1, cet ARNt est toujours l'ARNt<sup>lys</sup><sub>3</sub> [147,148]. Cependant il est important de préciser que même si l'association entre l'ARN génomique et l'ARNt est thermodynamiquement favorable, elle nécessite l'action d'un chaperon, la NCp7, qui va stabiliser l'ARN viral dans la conformation requise pour lier l'ARNt simultanément au niveau des sites PBS et PAS [149].

Les réactions de transcription inverse démarrent grâce à l'extrémité 3'OH de l'ARNt qui s'hybride au PBS de l'ARN génomique viral et qui fournit l'amorce nécessaire à la toute première réaction de polymérisation [150]. Par convention le brin d'ADN qui est produit par ces premières réactions de transcription inverse est nommé ADN (-) car il code les séquences complémentaires à l'ARN viral.

L'ADN (-) est donc, à cette étape, en réalité une molécule hybride ARN/ADN car l'ARN de transfert ayant servi d'amorce constitue l'extrémité 5' de l'ADN (-). Cet état de fait soulève, déjà à ce stade, un problème dont le virus devra s'affranchir : il est nécessaire de parvenir à se débarrasser des molécules d'ARN devenues inutiles et même gênantes une fois qu'elles ont rempli leur rôle de matrice pour la transcription inverse.

L'élément qui répond à cette nécessité est l'une des deux sous-unités de la transcriptase inverse et plus précisément la p66. J'avais évoqué plus en amont que le clivage asymétrique du dimère de transcriptase inverse générerait une sous-unité plus longue portant une activité supplémentaire par rapport à la sous-unité p55, plus courte. Cette activité supplémentaire est une activité RNase H, qui par définition permet de dégrader une molécule ARN si elle forme un duplex ARN/ADN [151].

Pour revenir aux premières étapes de transcription inverse, étant donné que le PBS se situe en aval de la région U5 de l'ARN viral, le brin d'ADN (-) en cours de synthèse s'allonge en « remontant » la région U5 puis la région R de l'ARN génomique viral. Arrivé à la région R qui marque l'extrémité 5' de l'ARN génomique viral, la transcription inverse s'arrête faute d'avoir une matrice pour poursuivre, ce qui constitue une pause dans la transcription inverse appelée « minus strand strong stop ». Pendant ce temps de pause, c'est l'activité RNase H du dimère de RT qui permet de dégrader la séquence ARN associée à l'ADN nouvellement synthétisé. Le résultat de cette dégradation est la formation d'une molécule partiellement double brin assez complexe à décrire. Cette molécule est composée dans sa partie amont d'une molécule hybride ADN/ARN constituée par un ADN simple brin lié fusionné dans sa partie 3' à l'ARNt ayant servi de matrice initiale par une liaison phosphodiester. L'ARNt forme un duplex avec la séquence PBS présente sur l'ARN génomique viral, et cette région constitue le seul endroit double brin. Enfin l'ARN génomique viral n'est plus constitué dans sa partie 5' que par la séquence PBS, les régions R et U5 ayant été dégradées, mais toutes les séquences en aval de ce site sont toujours intactes.

Pourtant, cet arrêt de la synthèse n'est que temporaire et il est levé par un mécanisme de transfert de brin. En effet, sur l'ARN génomique viral complet la région R se trouve à la fois au début et à la fin de l'ARN génomique viral. Ainsi l'ADN (-), qui a hérité de la séquence R située en 5' de l'ARN génomique lors de la transcription inverse, a la possibilité de changer de matrice et de s'associer avec la région R située à la fin de l'ARN génomique [152]. Ce phénomène est favorisé par le fait que les régions R et U5 situées en 5' de l'ARN viral aient été dégradées au préalable, en effet comme présenté plus haut ces séquences ARN étant associées avec l'ADN (-) nouvellement synthétisé sont reconnues et dégradées par l'activité RNase H du dimère de RT lors du « minus strand strong stop ». J'insiste également sur le fait que la séquence PBS, en revanche, est associée avec ARNt<sup>Lys</sup><sub>3</sub> et forme donc un ARN double brin qui est épargné par la digestion. Ainsi, après digestion, la courte région double brin formée par la paire ARNt-PBS n'étant pas très stable et pouvant se dénaturer assez

facilement permet de libérer l'ADN (-). La région R présente sur l'ADN (-) peut alors former un duplex avec la région R située à la fin de l'ARN viral et le duplex ARN/ADN ainsi formé fournit une nouvelle matrice à la transcriptase inverse qui peut reprendre sa tâche.

De manière très intéressante, ce mécanisme peut aboutir à un changement de matrice complet. Or, ce sont deux ARN génomiques viraux qui sont contenus dans la capsid, il est donc possible que lors du saut réplcatif la RT « choisisse » la région R du même ARN, qui a donc déjà servi à la synthèse de l'ADN (-), ou au contraire se fixe sur la région R du second ARN. Cette capacité de pouvoir changer de matrice, en dehors du fait que cela permette de lever la pause du « minus strand strong stop », présente l'avantage d'offrir également une solution à l'instabilité des ARN génomiques viraux. Comme toutes les molécules d'ARN, l'ARN génomique viral est assez instable et peut être facilement dégradé, le fait pour la RT de pouvoir changer d'ARN matrice lors de la transcription inverse permet d'offrir une solution à une telle dégradation. Ainsi si un ARN présente une cassure en un point donné, il est possible pour la RT de poursuivre néanmoins la transcription inverse en prenant pour matrice l'autre ARN qui, de façon très probable, ne présentera pas de cassure identique et permettra la reprise de la synthèse de l'ADN viral [153].

Quoi qu'il en soit, ce transfert permet de redémarrer la transcription inverse pour aboutir à la synthèse de l'ADN (-) complet, qui est par nature l'anti sens de l'ARN génomique viral. C'est également à cette étape qu'un LTR complet est reconstitué dans la partie 5' de l'ADN(-). En effet l'ADN est constitué dans sa partie 5' des régions R et U5 et le transfert de brin permet à la région R de cet ADN de s'apparier avec la région R présente en 3' de l'ARNm, cette même région R qui est située juste en aval de la région U3. Ainsi la reprise des réactions de transcription inverse permet d'ajouter la région U3 en 3' des régions R et U5 présentes sur l'ADN (-) et reforme un LTR complet dans la partie 5' de l'ADN (-). Ensuite, les réactions de transcription inverse vont se poursuivre en utilisant tout l'ARN génomique comme matrice et vont se terminer par la transcription inverse du site PBS présent en 5' de l'ARN génomique ayant échappé à la dégradation par l'activité RNase H avant le transfert de brin.

Une nouvelle fois, le dimère ARN/ADN engendré par la synthèse de l'ADN (-) complet est reconnu par le site RNase H de la p66, ce qui entraîne la dégradation quasi-complète de l'ARN génomique viral. A cette étape seul subsiste l'ADN (-) qui comporte dans sa partie 5' un LTR complet et se termine en 3' par un site PBS. Ici, le problème rencontré au tout début de la rétro-transcription se repose : pour constituer le brin complémentaire de l'ADN (-), appelé par convention ADN (+) car étant l'équivalent en ADN de l'ARN sens viral, il est nécessaire que la RT dispose d'une amorce. La solution à ce problème est apportée par des séquences riches en purines, ou poly purine tract (PPT), présentes sur l'ARN génomique et qui possèdent la particularité d'être résistantes à la dégradation effectuée par le domaine RNase H de la RT [154]. Ainsi la persistance de ces fragments d'ARN après la

fin de la synthèse de l'ADN (-) permet de redémarrer rapidement la synthèse en jouant le rôle de primer. Ces PPT sont au nombre de deux : le PPT central, qui comme son nom l'indique est situé à peu près au milieu de l'ARN viral, et le PPT 3', situé juste en amont de la région U3 sur l'ARN viral et proche de l'extrémité 5' de l'ADN (-).

En revanche, même si l'ADN (+) est partiellement synthétisé grâce à ces amorces, la synthèse se heurte à un obstacle similaire à celui rencontré lors de la synthèse de l'ADN (-) à savoir aucun PPT n'est complémentaire à l'extrémité 3' de l'ADN (-). Ceci implique qu'avec ces seules amorces, l'ADN (+) ne peut être synthétisé que partiellement, conduisant à une nouvelle pause dans la synthèse de l'ADN viral appelée « plus strand strong stop ». En l'occurrence à ce stade, l'ADN (+) n'est constitué que des séquences entre le PPT central et le PPT 3' ainsi que des régions U3, R et U5 se trouvant en aval du PPT 3'. Précision vitale à soulever ici : l'ADN(-) était toujours en fusion avec l'ARNt dans sa partie 5' lors de la synthèse du brin d'ADN (+). Ceci implique que la transcriptase inverse va synthétiser un site PBS en 3' de l'ADN (+) en se servant de l'ARNt comme matrice, puis va dégrader cette ARNt grâce à son activité RNaseH, libérant ainsi l'ADN (-) de cette séquence ARN et laissant le site PBS sous forme simple brin à cet endroit. Il faut encore ici noter qu'une région où deux molécules d'ADN (+) se chevauchent, appelée « DNA flap », est engendrée par le fait que la RT synthétisant l'ADN (+) à partir du PPT central fini par heurter le PPT 3' et le déplace pour continuer sa synthèse, formant ainsi une courte région triple brin.

Un mécanisme de transfert de brin assez similaire à celui utilisé précédemment pour constituer un ADN (-) complet est mis en œuvre pour terminer la synthèse de l'ADN (+). Une séquence PBS est présente à chacune des extrémités de l'ADN viral encore partiellement double brin à ce stade : le premier se trouve à l'extrémité 3' de l'ADN (-) tandis que le second se trouve à l'extrémité 3' de l'ADN (+) accolé à la région U5 du LTR complet présent ici. Ce phénomène rend possible une circularisation de l'ADN par appariement des séquences PBS ; circularisation qui permet alors à la RT de poursuivre la synthèse en se servant simultanément des séquences situées dans la partie 3' de l'ADN (-) comme matrice pour constituer un ADN (+) complet et des séquences situées en 5' de l'ARN (+), qui codent un LTR complet, comme matrice pour synthétiser un LTR complet en 3' de l'ADN (-).

Une fois arrivé à cette étape on peut comprendre les raisons qui ont favorisé la mise en place d'un système de rétro-transcription aussi complexe. Le fait même que le VIH soit obligé d'encapsider un ARNm pour servir de génome viral le met face à un paradoxe : dans la forme provirus les séquences promotrices contenues dans U3 sont situées avant le +1 de transcription, qui lui est situé dans la région R ; ces promoteurs ne peuvent donc pas être présents en 5' de la séquence des ARNm formés à partir du provirus. Pourtant c'est un ARNm qui va constituer le génome des virions et qui va servir à reconstituer un nouveau provirus, or il faudra que ces séquences promotrices essentielles se trouvent en amont des gènes à exprimer dans la nouvelle forme provirale. Or les LTR du provirus

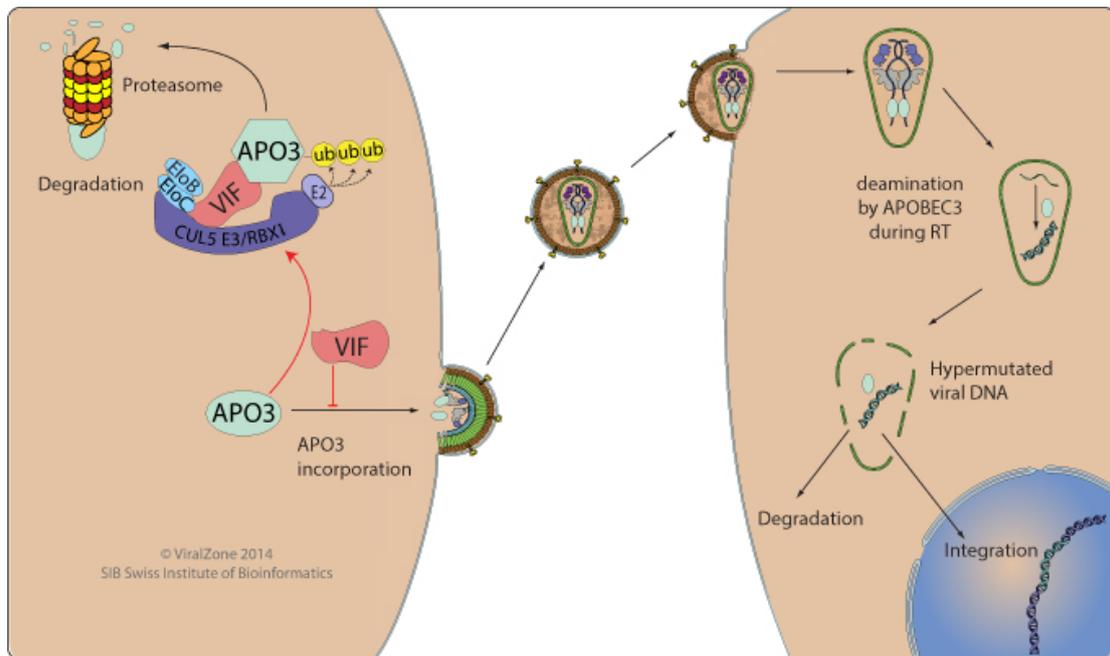
sont constitués, dans cet ordre, des régions U3, R et U5 et c'est la région R qui contient à la fois le +1 de transcription et le site de terminaison de la transcription. Le génome ARN est donc constitué dans sa partie 5' de la région R puis U5, et dans sa région 3' il est constitué de la région U3 puis R comme il a été décrit précédemment. La rétro-transcription permet donc de reformer deux LTR complet en début et fin de l'ADN double brin : un LTR complet est reconstitué dans la partie aval du génome (en 5' de l'ADN (-) plus précisément) postérieurement au « minus strand strong stop » et au transfert de brin, tandis que la circularisation qui met fin au « plus strand strong stop » permet de « déplacer » le LTR complet présent en 5' de l'ADN (+) pour la positionner au début du génome viral [155].

Une fois les réactions de rétro-transcriptions terminées, c'est un ADN linéaire dont un des brins est le complémentaire de l'ARN génomique viral qui est obtenu (nommé pour cette raison ADNc) et celui-ci possède, comme le provirus dont il est issu, deux LTR : un à chaque extrémité de sa séquence. Une caractéristique majeure différencie cependant cet ADNc d'un provirus, outre le fait qu'il ne soit pas encore intégré au génome hôte, c'est la présence du DNA flap qui n'a pas été corrigé par la RT et qui persiste chez l'ADNc. Le brin (+) de l'ADNc est donc en réalité constitué de deux molécules d'ADN qui se chevauchent sur une courte région et cette particularité jouera un rôle plus tard au moment de l'import nucléaire du génome viral.

Pour conclure cette section, j'ajouterais que le processus de rétro-transcription n'est pas effectué exclusivement par des protéines d'origines virales mais que des cofacteurs cellulaires sont également requis [156,157]. Même si ces facteurs ont été volontairement omis dans cette thèse pour des raisons de simplicités, il me faut cependant au moins les évoquer ici pour garder une vision d'ensemble aussi proche de la réalité que possible du mécanisme de rétro-transcription.

#### **3.2.4. Vif**

Une protéine qui n'a pas été détaillée plus tôt dans cette introduction est la protéine Vif, dont l'acronyme signifie « Virion Infectivity Factor ». Il a, en effet, été décrit relativement tôt dans l'histoire de l'étude de VIH que la présence de cette protéine était essentielle au bon déroulement de l'infection des cellules cibles de VIH-1 [158], ce qui lui a valu son nom. Certaines expériences ont suggéré que cette protéine d'environ 190 acides aminés était incorporée sélectivement dans les virions, notamment grâce à des interactions encore mal comprises avec le précurseur Gag [159].



**Figure 20 : Schéma du mécanisme d'action de la protéine Vif adapté de [41]**

(gauche) La protéine Vif peut reconnaître et se fixer sur la protéine cellulaire APOBEC3G (APO3). Ce faisant elle recrute le complexe Culline 5 – Ring Box 1 (CUL5 E3/RBX1) ainsi que d'autres facteurs cellulaires tels que les élongines B (EloB) et C (EloC). Ce complexe va recruter l'ubiquitine ligase E2 qui va ubiquitinyler APOBEC3G et ainsi conduire à la dégradation de APOBEC3G par le protéasome. Cette dégradation empêche l'incorporation de APOBEC3G dans les virions en cours de bourgeonnement.

(droite) En l'absence de Vif, APO3 est incorporée dans les virions. Lors de la transcription inverse (RT) du génome de VIH-1, APO3 va massivement désaminer les cytosines et les convertir en uraciles, ce qui conduit à une hypermutation létale du génome viral. Le génome muté est alors soit dégradé, si les mutations empêchent la transcription inverse de se poursuivre normalement, soit intégré dans le génome hôte, auquel cas les nombreuses mutations conduisent à la production de protéines virales essentiellement tronquée ou non fonctionnelles. Dans les deux cas le cycle viral devient très souvent abortif suite à l'action de APOBEC3G.

La possibilité que Vif soit incorporée sélectivement dans le virion lors du bourgeonnement a été également discutée par de nombreuses études [160,161] ce qui laisse la question en suspens. Concernant l'action de cette protéine, bien que des rôles aussi divers que d'aider à la maturation des protéines structurales après bourgeonnement ou de servir de chaperon pour le dimère d'ARN génomique aient été attribués à Vif, le fait que cette molécule soit entièrement dispensable dans certains types cellulaires donnés [162] soulève la question de l'importance de ces rôles, mais sans remettre complètement en cause leur existence. Question d'autant plus légitime qu'il y a toujours débat sur le fait que Vif soit incorporé de façon sélective dans la particule virale. C'est pourquoi ces questions délicates sont laissées à d'autres, plus experts en la matière, et il ne sera question ici uniquement que du rôle le mieux décrit de Vif : celui de contrecarrer l'action d'un facteur de restriction de l'hôte visant à empêcher VIH-1 de se multiplier dans ses cellules cibles. Ce facteur de restriction est APOBEC3G (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G), une protéine membre de la famille des cytidine déaminases [163].

Avant de m'avancer plus avant dans la description de cette protéine, il me faut préciser que d'autres membres de la même famille ont également été décrits comme pouvant contrecarrer l'infection à VIH-1 en absence de Vif, telle que APOBEC3F [164]. Néanmoins, pour ne pas me perdre sur le sujet il ne sera question ici que de APOBEC3G. Comme le nom de la famille à laquelle appartient cette protéine le laisse penser, APOBEC3G est une enzyme capable de désaminer les cytosines pour les convertir en uraciles. En revanche, contrairement à ce que le nom de la protéine pourrait suggérer, APOBEC3G ne cible pas les ARNm mais les ADN simple brins [165].

En l'absence de Vif, lors de l'infection d'une cellule par VIH-1, APOBEC3G est encapsidé dans les virions lors du bourgeonnement et c'est lors de la rétro-transcription que ce facteur de restriction joue un rôle, et plus précisément après la synthèse du brin d'ADN (-) du VIH et la dégradation de l'ARN génomique (figure 20). APOBEC3G se fixe sur l'ADN (-), presque entièrement simple brin, et hydrolyse les fonctions amines présentes en position 4 des cytosines de cet ADN pour générer des uraciles. Ensuite, lors de la synthèse du brin ADN (+), la présence de désoxyuridines va provoquer l'incorporation de désoxyadénosines comme base complémentaire. Le résultat net, si on compare les séquences génétiques du provirus et des ARNm qui ont servi à sa synthèse, est l'incorporation de très nombreuses mutations C-G vers T-A. Il faut en effet noter que les désoxyuridines produites par APOBEC3G ne persisteront pas dans le génome proviral car les uraciles sont reconnus comme étant des bases anormales de l'ADN par la machinerie enzymatique de l'hôte, ainsi les désoxyuridines sont systématiquement remplacées par des désoxythymidines.

L'hypermutation du génome viral induite par l'action de APOBEC3G est presque toujours fatale au virus, en partie à cause des diverses mutations faux-sens qui sont engendrées et qui perturbent le fonctionnement des protéines codées par les gènes touchés. Mais, et surtout, ce qui rend ce

phénomène d'hypermutation particulièrement létal est le fait que les mutations induites par APOBEC3G, C-G vers T-A, ont une chance extrêmement forte de générer des codons non-sens [166]. En effet, pour prendre un exemple, à chaque anti-codon tryptophane ACC de l'ADN (-) n'importe quel remplacement de désoxycytidine par une désoxythymidine génère un de ces trois anti-codons stop : ATT, ATC ou ACT sur l'ADN (-). Le résultat est la création d'un des trois codons stop TAA, TAG ou TGA sur le brin ADN (+) complémentaire. C'est pourquoi, lors de l'hypermutation d'un génome de VIH-1 par APOBEC3G, il n'est pas rare que tous les cadres ouverts de lectures du génome se terminent prématurément à cause d'une mutation non-sens.

Afin d'éviter une telle catastrophe, il est absolument nécessaire pour le virus de prévenir l'encapsidation de APOBEC3G dans les virions. Pour comprendre comment Vif agit il va donc falloir procéder à un retour en arrière dans le cycle viral et revenir à l'instant où toutes les protéines virales sont produites grâce à Rev à partir de la transcription d'un provirus, mais avant que le bourgeonnement des nouveaux virions ne se produise. En effet, Vif exerçant son action avant le bourgeonnement des virions et l'action de APOBEC3G étant indissociable de la rétro-transcription, il est extrêmement difficile de présenter l'ensemble du mécanisme en suivant le déroulement temporel du cycle viral, à moins d'effectuer une coupure dans la linéarité.

C'est donc avant que des virions ne soient générés par une cellule infectée que la protéine Vif entre en action et prévient l'encapsidation d'APOBEC3G [167]. Pour ce faire, la méthode utilisée passe une fois encore par l'ubiquitinylation et la dégradation du facteur de restriction par le protéasome. En l'occurrence, Vif est capable d'interagir directement avec la protéine APOBEC3G [168], puis de recruter le complexe Culline-5-RBX-1 (ring-box-1) qui va à son tour recruter des enzymes de conjugaison de l'ubiquitine (E2) et ainsi conduire à l'ubiquitinylation de APOBEC3G [169]. Comme toutes les protéines ubiquitinylées, APOBEC3G est ensuite dégradée très rapidement par le protéasome cellulaire.

La dégradation d'APOBEC3G par Vif garanti ainsi la production de virions majoritairement exempts de ce facteur de restriction cellulaire et évite le problème d'hypermutation du génome viral.

### **3.2.5. Décapsidation du virion**

En marge des événements de rétro-transcription se déroule la décapsidation de la particule virale. Comme précisé plus en amont, la rétro-transcription nécessite des facteurs cellulaires qui doivent par conséquent être capable de franchir la capsid. Une désintégration au moins partielle de la capsid est donc nécessaire pour le bon déroulement de la transcription inverse ce qui suppose une certaine concomitance des deux phénomènes. Ces deux événements ne sont cependant *a priori* pas

complètement interdépendants. Il a, en effet, été démontré que même si le déroulement de la rétro-transcription est nécessaire au déroulement de la décapsidation, en revanche l'initiation de la décapsidation est indépendante des premiers événements de rétro-transcription [170]. Le modèle généralement admis aujourd'hui propose une décapsidation se déroulant en au moins deux étapes : une étape de décapsidation précoce et partielle qui commence avant la rétro-transcription et s'arrête, ou au moins ralenti considérablement, puis reprend dans un second temps pour aller de pair avec la rétro-transcription [171].

La fin de l'étape de décapsidation est en revanche moins bien connue. Conceptuellement, il est en effet vital que des protéines de capsidite restent associées au génome viral tant qu'il est situé dans le cytoplasme et cela pour deux raisons. La première est que la présence d'un ADN double brin dans le cytoplasme d'une cellule humaine est reconnue comme étant signe d'une infection virale et active une voie de production des IFN- $\beta$  ; la présence de protéines associées au génome viral permet ainsi de le protéger de la reconnaissance cellulaire. Ce fait est confirmé par l'expérience car, dans le cas où le virus possède des mutations qui entraînent une décapsidation trop rapide, il a été démontré que de l'interféron était produit par la cellule hôte et que le cycle viral devenait abortif [170]. La seconde raison vient indirectement du fait que le VIH-1 soit capable d'infecter des cellules quiescentes chez qui la membrane nucléaire est intacte, or les expériences ont montré que la p24 est très impliquée dans le mécanisme actif d'import du génome viral dans le noyau de la cellule hôte [172], il est donc nécessaire que la p24 reste associée au génome ADN pour garantir le succès de cette étape. Cependant, même si la conservation de protéines de capsidite associées au génome viral ADN est *a priori* nécessaire à l'infection, jusqu'à très récemment aucune expérience n'avait permis de mettre en évidence que de la p24 restait associée à l'ADN viral après rétro-transcription. Ce problème apparent a été résolu il y a quelques années par une étude démontrant sans ambiguïté que des protéines de capsidites restent associées à l'ADN viral après que la transcription inverse soit terminée et même après l'import nucléaire [173].

### **3.2.6. Import nucléaire**

Les prémices de l'import nucléaire du génome viral consistent tout d'abord à acheminer celui-ci à travers le cytoplasme jusqu'à un pore nucléaire. Il a été élucidé que ce transport se fait par le biais des microtubules et plus précisément par l'interaction de la capsidite virale avec la dynéine [174]. Ce transport est nécessaire à la décapsidation et rejoint le fait que les événements de rétro-transcription, décapsidation et transport cytoplasmique se déroulent de manière concomitante et sont interconnectés.

Arrivé à proximité des pores nucléaires et une fois que la rétro-transcription est terminée, le génome viral est constitué par un ADN double brin linéaire, mais cet ADN n'est pas nu dans le cytoplasme : il reste associé avec des protéines de capsidie comme discuté plus haut mais aussi avec d'autres protéines qui forment ensemble le complexe de pré-intégration (PIC). Les constituants les mieux décrits de ce complexe en sont les protéines virales, aussi il sera exclusivement question d'elles dans la suite de cette section. Mais encore une fois par souci de rigueur il faut garder à l'esprit que le PIC renferme également certains facteurs cellulaires comme la protéine BAF (barrier to autointegration factor) qui garantit l'intégration de l'ADN viral dans le génome hôte en l'empêchant de s'intégrer à lui-même [175].

Le PIC est composé d'un ensemble de protéines virales de matrice, de capsidie et par des protéines Vpr, auxquelles viennent s'ajouter l'ADN viral et l'intégrase. De façon surprenante ce n'est pas une protéine particulière qui va être chargée de l'import nucléaire mais ce sont toutes ces protéines, et même l'ADN viral, qui vont chacun participer à cette étape [176]. Il a en effet été décrit que toutes les protéines virales composant le PIC possède au moins un NLS jouant un rôle dans l'import nucléaire du PIC. On voit ici que c'est l'implication de Vpr dans l'import nucléaire qui explique qu'elle soit encapsidée dans la particule virale, en plus du fait que cette protéine semble jouer un rôle de contre-mesure virale lors de l'infection des macrophages comme cela avait été précisé dans la section consacrée à cette protéine. L'ADN viral est lui aussi impliqué dans son propre import, non pas par le biais d'un NLS bien entendu, mais grâce à un motif particulier : le DNA flap [177], qui permet de faciliter l'import par un mécanisme encore mal compris.

Les différents NLS présents sur les protéines virales sont reconnus par des importines  $\alpha$  qui s'associent au PIC, ces interactions provoquent l'hétéro-dimérisation des importines  $\alpha$  avec des importines  $\beta$  et la translocation du complexe vers les pores nucléaires [178]. Une fois parvenu au pore nucléaire, des interactions supplémentaires s'établissent entre certaines protéines virales et les protéines cellulaires constituant le pore. Pour exemple, les mieux décrites de ces interactions sont celles qui s'établissent entre la protéine de capsidie et les nucléoporines NUP 158 et NUP 358 et qui permettent la translocation du PIC vers le noyau [179].

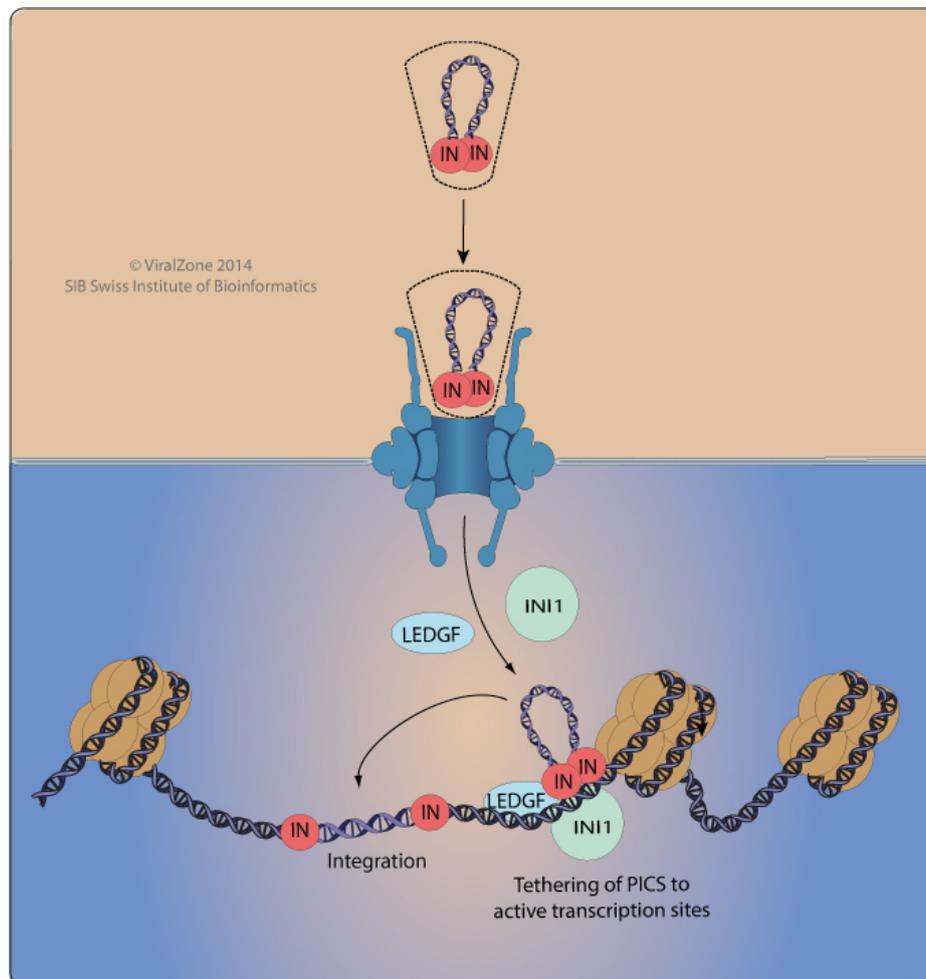
### **3.2.7. Intégrase et intégration**

Arrivé dans le noyau cellulaire, la dernière étape qui permet de compléter le cycle viral consiste à intégrer le génome viral, constitué désormais par de l'ADN double brin linéaire, dans celui de la cellule hôte. Cette action est catalysée par l'intégrase virale.

Jusqu'à assez récemment, un consensus existait sur le fait que l'intégration du génome VIH se faisait de manière aléatoire dans le génome hôte. Ce modèle était principalement soutenu par le fait qu'aucun site préférentiel n'existe pour l'intégration de VIH comme cela est le cas par exemple pour les adénovirus qui possède un site précis sur le chromosome 19 [180]. Toutefois, ce modèle a évolué suite à un nombre croissant d'études réalisées sur le sujet et nous savons aujourd'hui que VIH-1 intègre son génome, certes de manière assez aléatoire, mais préférentiellement dans des sites activement transcrits du génome hôte, c'est-à-dire dans l'euchromatine. Un tel ciblage est tout à l'avantage du virus car pouvoir transcrire le très long ADN viral constitue déjà une problématique en soi pour le virus comme nous l'avons vu, aussi une intégration se faisant dans des régions trop condensées du génome cellulaire pour être facilement transcrite serait très désavantageux pour le virus. Le facteur permettant de diriger l'intégration vers un site adéquat n'est pas d'origine virale mais cellulaire, il s'agit de la protéine LEDGF/p75 [181] (figure 21). Pour être complet, il faut préciser que cette protéine est recrutée dans le cytoplasme de la cellule par le PIC et est également impliquée dans l'import nucléaire.

Un dernier élément très intéressant est le fait que certaines nucléoporines sont également impliquées dans le processus d'intégration en plus de jouer un rôle dans le passage vers le noyau [182]. Ceci suggère que l'intégration se produit préférentiellement dans les régions périphériques du noyau, à proximité des pores nucléaires, et pas seulement dans des régions spécifiques du génome hôte.

Concernant la réaction en elle-même, celle-ci est catalysée par l'intégrase virale présente sous forme de tétramère. En effet, des dimères d'intégrases sont présents à chaque extrémité du génome viral et la jonction de ces extrémités ensemble permet la formation d'un tétramère, bien qu'on puisse aussi considérer que c'est la formation du tétramère d'intégrase qui permet la jonction des extrémités. La toute première action de l'intégrase consiste à cliver deux nucléotides aux extrémités de l'ADN viral pour convertir les deux extrémités franches en deux extrémités 5' phosphate sortantes. Les deux extrémités 3' OH sont terminées par deux nucléotides, CA3'OH, très conservés chez les VIH [97]. La réaction chimique catalysée par le tétramère est une transestérification obtenue par l'attaque du groupe hydroxyle présent en 3' de l'ADN viral sur la liaison phosphodiester qui joint normalement deux nucléotides du génome hôte. Cette attaque chimique est réalisée en parallèle sur les deux brins du génome hôte, mais étant donné qu'elle ne se fait que par les 3'OH, à chaque extrémité du génome VIH seul le brin possédant le groupement hydroxyle se lie de manière covalente avec le génome eucaryote tandis que les deux nucléotides de l'extrémité 5' phosphate restent non appariés et non liés au génome hôte [183]. C'est la machinerie enzymatique de l'hôte qui va reconnaître le mésappariement ainsi que la non-jonction des nucléotides et qui va « réparer » ces sites pour finaliser l'intégration du génome viral.



**Figure 21 : Schéma du mécanisme d'adressage de l'ADN viral aux régions activement transcrites du génome adapté de [41]**

(haut) Dans le cytoplasme la transcription inverse aboutit à la formation d'un ADN double brin linéaire. Un dimère d'intégrase (IN) est fixé à chaque extrémité de l'ADN viral et la formation d'un tétramère d'IN permet de circulariser l'ADN sans joindre les extrémités. Cet ADN est en complexe avec plusieurs protéines virales non représentées telles que la protéine de capsid ou Vpr, ce qui forme le complexe de préintégration (PIC). Ce complexe dispose de NLS, aussi il est reconnu par la machinerie d'import cellulaire et est dirigé vers les pores nucléaires.

(bas) Après passage du PIC dans le noyau la protéine LEDGF permet de diriger l'ADN viral en complexe avec le tétramère d'intégrase vers les zones activement transcrites du génome hôte. Le tétramère d'intégrase catalyse alors l'intégration du génome viral dans le génome hôte.

Une fois intégré dans le génome hôte, le provirus va pouvoir être transcrit et traduit pour former toutes les protéines virales que nous venons de voir. Ces protéines vont permettre de donner naissance à une nouvelle génération de virions ce qui complète le cycle viral du VIH-1.



## **Quatrième partie : variabilité génétique et réseaux de coévolution**

## 4. Variabilité génétique du VIH-1

### 4.1. Origine de la variabilité

#### 4.1.1. Fidélité de la RT

Les exposés des sections précédentes de mon introduction permettent d'aborder des sujets plus étroitement liés à cette thèse en commençant par la variabilité génétique du VIH-1. Il était en effet nécessaire de traiter ces deux grandes parties que sont la classification des VIH et le cycle des VIH-1 pour aborder cette variabilité. Variabilité qui transparait au travers de la classification complexe de la vaste quantité de souches existantes de VIH, et dont on peut percevoir la cause en observant de plus près le cycle de réplication de ce virus. En effet, de nombreux éléments du cycle de réplication favorisent grandement l'apparition de variation génétique, au premier desquels la faible fidélité de la transcriptase inverse, qui incorpore en moyenne une à trois mutations tous les dix mille nucléotides lors des réactions de transcription inverses [184,185]. Ce taux de mutation peut être comparé avec celui des ADN polymérases humaines telles que les polymérases  $\delta$  et  $\sigma$  qui incorporent en moyenne une mutation tous les un million de nucléotides [186]. Bien que moins fidèle que d'autres ADN polymérases, la fidélité de la transcriptase est tout de même satisfaisante pour le virus car le génome viral n'est constitué que par environ 10kbp, ce qui signifie tout de même qu'il y a au minimum une mutation qui est incorporée par la RT lors de chaque évènement de transcription inverse. En revanche cette fidélité doit également être considérée en parallèle avec la quantité de virions produite chez l'hôte. En effet, on estime que ce sont environ 10 milliard de nouveaux virions qui sont produits chaque jour dans un hôte infecté par VIH-1 [187]. Ainsi, ce sont potentiellement 10 milliards d'évènements de transcription inverse qui ont lieu chaque jour chez un hôte infecté. C'est pourquoi le taux de fidélité de la RT, en prenant 1 mutation tous les 10kbp, lorsqu'il est combiné au fait que ce sont théoriquement 10 milliards d'évènements de transcription inverse qui se déroulent chaque jour dans l'hôte, ont tous deux pour résultat une diversification génétique rapide du VIH-1 *via* l'intégration potentielle de 10 millions de mutations dans la population virale globale toutes les 24 heures.

#### 4.1.2. APOBEC3G

Des facteurs cellulaires peuvent également contribuer à augmenter le taux de mutation de VIH, l'exemple le mieux décrit est celui de la protéine APOBEC3G. Même si la dégradation de cette protéine est induite par la protéine virale Vif, il a été décrit que des quantités non létales d'APOBEC3G peuvent néanmoins être encapsidées et, incidemment, contribuer à augmenter la

variabilité génétique de VIH-1 [188]. Dans le cas d'une encapsidation d'APOBEC3G en faible quantité, en effet, la faible quantité de mutations induites par cette protéine n'est pas suffisante pour provoquer systématiquement l'apparition de mutations délétères [189]. Dans le cas où les mutations apparues n'entravent pas le déroulement du cycle viral, ces mêmes mutations viennent ainsi *de facto* augmenter la diversité génétique du VIH-1. Ce mécanisme contribue largement à la variabilité génétique de VIH-1 comme en témoigne le fait que la variabilité observée chez ce virus provient en bonne partie de transitions G vers A sur le brin ADN sens, qui est la mutation signature d'APOBEC3G [188].

#### **4.1.3. Recombinaisons**

A ces sources de variabilité génétique viennent s'ajouter les phénomènes de recombinaison pouvant se produire au cours de la rétro-transcription [190]. Comme énoncé lors de la description de la rétro-transcription, la RT est capable de changer d'ARN matrice pendant la transcription inverse car ce sont deux ARN génomiques qui sont encapsidés dans la particule virale, un exemple classique de ce phénomène est le changement de matrice pouvant survenir au cours du « minus strand strong stop ». Cependant, ces changements de matrice ne sont pas uniquement susceptibles d'avoir lieu lors du « minus strand strong stop », mais peuvent survenir pendant toute la synthèse de l'ADN (-).

Pourtant, ces changements de matrice ne surviennent pas de façon complètement aléatoire mais ont lieu préférentiellement en certains points du génome viral qui sont qualifiés de « hot spots » [191]. Ce sont principalement des ralentissements ou des pauses effectuées par la RT lors la synthèse du brin d'ADN (-) qui favorisent le phénomène de changement d'ARN matrice en cours de synthèse et qui induisent l'apparition d'un « hot spot » [153,192]. Ces pauses ou ces ralentissements sont majoritairement induits par la présence de structures secondaires sur l'ARN génomique viral, par exemple il a déjà été décrit que la présence d'une structure en tige-boucle sur l'ARN favorise le changement d'ARN matrice [193]. De tels « hot spots » ont ainsi été décrits : entre les gènes *gag* et *pol*, à l'intérieur de la séquence du gène de la RT, et à chaque extrémité du gène *env*, pour n'en citer que quelques-uns [194,195].

Concernant les conséquences que peuvent avoir un tel changement d'ARN matrice survenant pendant la synthèse du brin d'ADN (-), si les deux ARNm contenus dans le virion proviennent de la transcription du même provirus, alors le changement de molécule matrice est sans conséquence car c'est la même information génétique qui est contenue dans ces deux ARNm. Toutefois, le changement de matrice porte à conséquence lorsque les deux ARNm encapsidés dans le virion ont chacun pour origine un provirus différent. La présence de deux ARNm d'origines différentes dans un

même virion est en effet possible, si la cellule qui a produit ce virion était co-infectée et contenait dans son génome deux provirus différents. Dans ce cas l'encapsidation dans la même particule virale de deux ARNm viraux d'origines différentes est possible, à la condition bien sûr que les protéines Gag produites soient capables de reconnaître et de lier les deux types d'ARN génomiques lors de l'assemblage des nouveaux virions. Dans ce cas, tout changement d'ARN matrice lors de la transcription inverse conduit à une recombinaison des séquences issues des deux provirus parentaux.

Le résultat d'un tel changement d'ARN matrice entre deux ARN ne portant pas la même information génétique est la formation d'un ADN viral contenant des séquences génétiques provenant des deux ARNm qui ont été encapsidés dans la particule virale, eux-mêmes issus de deux provirus différents. Même si cette recombinaison ne génère pas de mutations à proprement parler, elle permet de réassortir des séquences génétiques d'origines différentes ensemble, ce qui permet de générer de nouvelles souches de VIH-1. Pour faire écho à la section de l'introduction consacrée aux CRF et aux URF, c'est par ce mécanisme que sont générées ces types de souches virales recombinantes. Le fait que près de 20% des infections à VIH-1 dans le monde sont causées par des CRF témoigne de l'importance du mécanisme de recombinaison chez VIH-1 qui non seulement contribue à créer de nouvelles souches de VIH-1 mais également grandement au succès épidémiologique de ce virus.

#### **4.1.4. Autres sources de variabilité**

Pour être rigoureux, il faut encore préciser que VIH-1 est sensible à tous les mutagènes ou événements extérieurs qui peuvent modifier l'information génétique d'un ARN ou d'un ADN en général, le génome VIH-1 passant par ces deux formes. Ainsi, par exemple toute exposition du génome viral ADN aux ultraviolets est potentiellement mutagène, les UV étant connus pour induire la formation de dimère de pyrimidine [196] ; l'intégrité génétique du génome proviral est également inféodée au taux de fidélité des ADN polymérases humaines lors de la duplication de l'ADN nucléaire à chaque phase S [186] ; et enfin les acides nucléiques viraux sont sujets aux mutations spontanées car les protons constituant les bases azotées sont des objets quantiques et peuvent se déplacer par effet tunnel en changeant, de par le fait, la nature chimique de la base [197].

Bien que VIH-1 soit sensible à toutes ces sources de variabilité génétiques, comme n'importe quel autre organisme, celles-ci ne contribuent que de façon minoritaire à la diversité génétique globale de VIH-1. Les trois sources principales de variabilité génétique sont la faible fidélité de la transcriptase inverse, la présence d'APOBEC3G en quantité non létale dans les virions et les recombinaisons

possibles entre ARN génomiques d'origines phylogénétique différentes. Ainsi, la fidélité de la transcriptase inverse et APOBEC3G sont majoritairement responsables du fort taux de mutation associé à VIH-1 et induisent l'apparition rapide de nouvelles souches virales tandis que les phénomènes de recombinaisons participent à la création de nouvelles souches, et notamment de CRF, sans générer de nouvelles mutations mais en combinant différents jeux de mutations préexistants ensembles.

#### **4.2. Conséquences de la variabilité génétique et réseaux de co-évolution**

Le fait pour VIH-1 de disposer d'une grande diversité génétique et de pouvoir l'entretenir au travers de sources importantes de variabilité génétique est un avantage décisif pour le VIH-1 dans sa lutte contre le système immunitaire de son hôte. En effet, même si le système immunitaire adaptatif humoral est doué d'une grande capacité neutralisante, la vitesse à laquelle ce système peut s'adapter pour neutraliser une souche virale nouvellement apparue est définie par la vitesse d'apparition des mutations somatiques des cellules B [198]. Or, la vitesse à laquelle le virus accumule des mutations est supérieure à la capacité du système immunitaire à produire des mutations somatiques. De plus, conceptuellement, le virus dispose de « l'avantage de l'offensive », étant donné que le système immunitaire ne peut pas anticiper qu'elle va être la nature des souches virales qui vont apparaître au fur et à mesure du temps, il est contraint de subir l'apparition des nouvelles souches VIH-1 et de s'adapter à elles uniquement après que celles-ci se soient disséminées dans l'organisme. Cependant, ces deux phénomènes ne garantissent pas à eux seuls au VIH-1 de pouvoir échapper efficacement à la réponse humorale car le système immunitaire est connu pour pouvoir produire des anticorps neutralisants à large spectre, qui sont capables de neutraliser simultanément plusieurs souches de VIH-1 [199–201]. De plus les anticorps, contrairement au VIH-1, ne sont théoriquement pas limités dans leur capacité à muter, car là où les souches mutantes de VIH-1 subissent une sélection négative pour leur fonctionnalité car elles doivent en permanence être capable d'accomplir le cycle viral, les anticorps ne remplissent aucun autre rôle que celui de neutraliser des antigènes ; il n'y a donc aucune limite théorique aux mutations qu'ils peuvent incorporer. Pour le VIH-1, c'est uniquement par l'action combinée de la grande variabilité génétique du virus et de la batterie de protéines virales chargées de contrer, ou tout du moins de ralentir, l'action du système immunitaire que VIH-1 peut empêcher son élimination par la réponse immunitaire. Produire et maintenir une diversité génétique accrue est donc une condition essentielle à la survie à long terme du VIH-1 chez son hôte et est une des raisons du succès épidémiologique important de ce virus.

Posséder un taux de mutations élevé n'est pas sans désavantage pour VIH-1, ou pour tout organisme de façon générale. En effet, il n'est pas suffisant pour un virus de pouvoir muter rapidement, il est également nécessaire que ce fort taux de mutations ne compromette pas l'intégrité génétique des gènes qui composent le génome viral. Il est vital que les protéines virales soient capables de remplir leurs rôles, surtout au regard des rôles cruciaux que chacune d'entre elles remplit et qui les rendent toutes indispensables pour le succès répliatif du virus chez son hôte. Pourtant il est assez rare que les mutations qui apparaissent dans un gène donné aient des effets positifs sur l'activité de la protéine codée par lui, au contraire la grande majorité des mutations apparaissant dans un gène impactent très souvent négativement la fonctionnalité de la protéine correspondante [202].

Si VIH-1 possède un taux de mutation très élevé lui permettant d'échapper au système immunitaire de son hôte, comment peut-il dans le même temps empêcher ces mutations de détruire les protéines qui le composent ? La réponse à cette question vient en partie du phénomène de co-évolution qui permet de concilier ces deux besoins. Dans le cas d'une protéine virale dont un des acides aminés vient de varier suite à une mutation, si cette mutation particulière rend la protéine virale capable d'échapper aux anticorps produits par le système immunitaire, elle sera positivement sélectionnée. Mais un problème se pose si cette mutation impacte également négativement la fonctionnalité de cette protéine. Le fait que des mutations ayant un effet négatif sur la fonction d'une protéine d'un pathogène donné, mais qui sont tout de même maintenues dans la population car permettant d'échapper aux anticorps neutralisants, est un phénomène non seulement possible, mais survenant assez couramment et largement documenté [203-208]. Etant donné que toute réversion de la mutation est hautement improbable, car un tel événement augmenterait la sensibilité du pathogène aux anticorps, la solution la plus probable est donc que d'autres mutations soient incorporées pour restaurer la fonctionnalité de la protéine tout en conservant la capacité d'échapper aux anticorps. Ces événements définissent des ensembles d'acides aminés qui coévoient ensemble, générant ainsi des réseaux de coévolution qui ont pour but principal de maintenir un équilibre entre le besoin d'échapper à la réponse immunitaire de l'hôte et le besoin de conserver la fonctionnalité des protéines virales. C'est ce phénomène complexe de co-évolution entre les différents acides aminés qui composent une protéine donnée, qui a été au centre de mon travail de thèse.

## **5. Etude des réseaux de coévolution**

Etudier les réseaux de coévolution qui se sont mis en place dans une protéine de pathogène donnée permet donc, en théorie, d'étudier comment cette protéine accumule les mutations, dans le

but à la fois d'échapper au système immunitaire et de conserver la capacité à remplir la tâche qui lui incombe.

Mettre en évidence de tels liens coévolutifs peut se faire de différents moyens. Le plus utilisé à l'heure actuelle est de réaliser une approche *in silico* visant à analyser un grand nombre de séquences génétiques et/ou protéiques disponibles. Le but de ces approches consiste, en utilisant différents algorithmes et différentes méthodes, telles que le « direct-coupling analysis » [209] ou la méthode « MetaPSICOV » [210], à trouver des acides aminés ou des codons qui covarient. Par exemple, dans un gène donné si deux codons coévoluent ensemble, il est attendu que si un de ces codons présente une mutation, alors le codon coévoluant avec lui contiendra très certainement lui aussi au moins une mutation compensatoire. Lors de l'étude des séquences disponibles de ces gènes, ces deux mutations seront alors toujours trouvées ensemble dans une partie de la population et seront toutes les deux absentes de l'autre partie, permettant de les mettre en évidence. Cette approche présente les avantages d'être facile à mettre en place, d'être peu coûteuse et de donner des résultats rapides. Cependant un des premiers désavantages associé à cette technique et qu'elle nécessite de disposer d'un très grand nombre de séquences disponibles pour pouvoir donner des résultats statistiquement significatifs. Si la population étudiée est trop petite il en résulte un nombre élevé de faux négatifs et de faux positifs [211]. Ceci pose problème même dans le cas de VIH-1, car malgré le fait que de nombreux génomes VIH-1 aient été séquencés, certains sous-types viraux, tel que le sous-type G, sont peu répandus et ne comportent que peu de génomes séquencés ce qui complique leur étude par ces méthodes. De plus le fait que plusieurs méthodes d'analyse *in silico* existent et qu'elles ne soient pas toutes équivalentes implique que les résultats obtenus en utilisant l'une d'elles ne seront pas forcément retrouvés en utilisant une autre méthode, ce qui ne signifie en aucun cas que ce type d'étude ne donne pas de résultats exploitables mais ce qui signifie que les analyses des résultats obtenus sont complexes et demandent d'utiliser plusieurs méthodes ou algorithmes en parallèle afin d'éliminer au maximum les faux positifs. De plus, ces études *in silico* demandent systématiquement une conformation expérimentale car par nature ce ne sont que des prédictions faites sur la base d'analyses de base de données.

C'est à cause de ces désavantages que nous avons cherché à utiliser une méthode alternative permettant d'étudier les réseaux de coévolution et c'est pourquoi le principe des études *in silico* ne sera pas détaillé d'avantage dans le cadre de cette thèse.

La méthodologie choisie ici pour étudier les réseaux de coévolution repose sur une observation : les événements de recombinaisons survenant entre différentes souches virales ont parfois pour résultat la génération d'une souche virale déficiente alors que les deux souches virales parentales sont infectieuses [195]. La seule explication à un tel phénomène est que des liens coévolutifs unissant différents acides aminés présents de part et d'autre du point de recombinaison ont été brisés lors de

cet événement qui met en contact des parties du génome n'ayant pas la même origine phylogénétique et qui au fil de leur histoire évolutive ont divergées, finissant par devenir incompatibles. En effet, l'un des rôles de la coévolution est de maintenir la fonctionnalité des protéines tant au niveau intraprotéique, car les différentes régions constituant la protéine considérée doivent être compatibles entre elles, qu'à un niveau interprotéique car les différentes protéines constituant un organisme doivent pouvoir interagir entre elles. Il est donc attendu que briser des liens coévolutifs impacte négativement la fonctionnalité des protéines étudiées.

Pour prendre un exemple concret, imaginons une protéine donnée existant dans deux versions phylogénétiquement différentes dans deux organismes différents et dont la fonctionnalité est garantie par l'interaction entre deux acides aminés nommés A et B. Dans le cas de la première version de la protéine l'interaction entre ces deux acides aminés est garantie par une liaison hydrophobe car l'acide aminé A est une valine tandis que le B est une isoleucine. Pour la version B en revanche l'interaction entre les acides aminés A et B prend la forme d'une liaison ionique car A est une lysine et B est un acide aspartique. Dans ce cas une protéine issue d'une recombinaison qui réassortirait l'acide aminé A de la première protéine avec l'acide aminé B de la seconde protéine ne serait pas fonctionnelle car la valine de la protéine A et l'acide aspartique de la protéine B ne peuvent interagir. Le résultat net pour la protéine recombinante contenant ces deux acides aminés est que celle-ci est non fonctionnelle car la rupture du lien coévolutif entre ces résidus empêche leur interaction.

L'exemple prit ici pour expliciter la relation entre réseau de coévolution et fonctionnalité ne doit cependant pas être généralisé à tous les réseaux de coévolution. Un lien coévolutif existant entre deux acides aminés ne signifie pas nécessairement que ces acides aminés interagissent physiquement [212]. De façon similaire un lien de coévolution liant de façon stricte deux acides aminés est possible, mais l'existence de réseaux de coévolution impliquant de multiples résidus a également déjà été décrit [213].

Concrètement, l'approche que nous avons donc utilisée pour notre étude des réseaux de coévolution consiste à générer par ingénierie génétique des protéines chimériques, telles qu'elles auraient pu être produites dans la nature par le mécanisme de recombinaison, et qui combinent des régions ayant des origines phylogénétiques différentes. Ainsi, si des protéines chimériques sont non fonctionnelles, cela implique que un ou plusieurs réseaux existants entre les régions qui la composent et qui ne partagent plus une origine phylogénétique commune ont été brisés.

## 6. Etude de la glycoprotéine d'enveloppe de VIH-1 par ses réseaux de coévolution

Nous avons ainsi entrepris de mener une étude des réseaux de coévolution présents chez VIH-1 et nous avons choisi les glycoprotéines d'enveloppe de VIH-1 comme modèle. La première raison expliquant ce choix est que ces protéines sont directement exposées à la surface de la particule virale et donc, sans protection face aux anticorps neutralisants se trouvant dans le milieu extracellulaire. De plus, ces glycoprotéines sont responsables, presque à elles seules, de la réalisation de l'étape cruciale d'entrée virale. Les contraintes qui s'exercent sur ces protéines pour qu'elles soient simultanément capable de se soustraire à la neutralisation par les anticorps et qu'elles puissent remplir leur tâche avec succès sont donc attendues être particulièrement fortes. Il est donc logique que les événements de coévolution permettant de concilier ces deux contraintes soient d'autant plus fréquents et faciles à mettre en évidence. Cette assertion est soutenue par le fait que de nombreux articles parus dans la littérature scientifique démontrent que les mutations permettant à la protéine d'enveloppe de VIH-1 d'échapper aux anticorps neutralisants induisent dans un premier temps une perte de fitness de la souche virale qui porte ces mutations [204,206–208]. Ceci soutient le fait que le mécanisme de coévolution participe à l'accumulation des mutations dans le gène de la glycoprotéine Env de VIH-1. En effet, les avantages apportés par ces mutations qui impactent pourtant le fitness sont indéniables et provoquent le maintien de ces mutations dans la population virale. Les effets négatifs sont donc corrigés ultérieurement par le réseau de coévolution qui permet d'insérer des mutations compensatoires ailleurs dans le génome viral.

Le fait que l'enveloppe du VIH-1 soit un bon modèle pour notre étude car les réseaux de coévolution qui la compose sont supposés être faciles à mettre en évidence n'a pas été la seule raison pour laquelle nous avons porté notre choix sur cette protéine. Comprendre par quel mécanisme et par la mise en jeu de quels réseaux de coévolution se fait l'évolution moléculaire de la glycoprotéine d'enveloppe de VIH-1 est une problématique peu étudiée et très intéressante. En effet, cette glycoprotéine cristallise autour d'elle la majorité des tentatives de mise d'un vaccin protecteur contre VIH-1 en l'utilisant comme immunogène [214]. Aussi comprendre un mécanisme aussi fondamental que celui qui façonne la manière dont cette protéine évolue dans le temps en réponse à l'action du système immunitaire pourrait être d'une grande utilité pour améliorer la façon dont nous concevons les vaccins.

De plus, bien que l'enveloppe du VIH-1 ait été très étudiée, des interrogations et des incompréhensions persistent et pourraient être levées par une approche différente de celles employées précédemment, en l'occurrence une approche basée sur l'étude du réseau de coévolution. L'exemple le plus probant de ces interrogations vient de l'organisation dans l'espace du

trimère d'enveloppe, qui a pourtant été très étudié par différentes techniques, notamment par la cristallisation et la diffraction aux rayons X ainsi que la cryo-microscopie électronique. Ces techniques et les efforts combinés de plusieurs équipes de chercheurs ont ainsi permis de générer des données spectaculaires et d'accroître considérablement nos connaissances de l'organisation du trimère. Pourtant même si ces données sont les meilleures que nous ayons actuellement à notre disposition, elles possèdent leurs limites. Les trimères d'enveloppes constituent notamment des assemblages de protéines très flexibles qui transitent en permanence entre plusieurs états conformationnels [119,215]. L'enveloppe se prête donc assez mal à de telles études car pour obtenir des structures détaillées demande de figer ces protéines.

Or, le fait de devoir figer ces protéines demande bien souvent l'utilisation de molécule servant d'échafaudage, tels des anticorps, qui présentent le défaut de ne pas être des ligands « naturels » et de pouvoir éventuellement induire des conformations qui ne sont pas des conformations physiologiques de la glycoprotéine. De plus, l'utilisation de ces échafaudages protéiques n'est pas toujours suffisante et de nombreuses études ont eu recours à l'utilisation de glycoprotéines Env modifiées par ingénierie génétique pour augmenter leur stabilité.

Ainsi, pour les premières études structurales par diffraction aux rayons X qui ont été publiées dans les années 1990, ce sont uniquement des gp120 dont on avait retiré toutes les régions variables qui étaient utilisées [106]. Il était alors techniquement impossible d'étudier le trimère natif à cause des liaisons trop faibles qui unissent les gp41 aux gp120, de plus retirer les régions faiblement structurées, telles que les régions variables, permet de baisser grandement la difficulté qu'il y a à obtenir des cristaux de protéines. Cependant des modifications aussi lourdes de la protéine d'intérêt soulevaient naturellement des questions sur la pertinence des résultats obtenus, questions qui étaient d'ailleurs soulevées par les auteurs des travaux eux-mêmes.

Depuis cette époque, une avancée majeure dans le domaine a été réalisée avec l'obtention des enveloppes dites SOSIP 140. Les trimères composés de ces glycoprotéines d'enveloppes sont stabilisés par un pont disulfure entre chaque hétérodimère de gp120/gp41 [216], ainsi que par la mutation I559P qui stabilise les interactions entre les trois sous-unités gp41 [217]. Ces trimères ont également été rendu solubles par optimisation d'un site de clivage de la furine [218,219] qui permet de cliver les gp41 entre leur région HR2 et leur région MPER, ce qui permet de dissocier les gp41 de la membrane au prix de la perte de la région transmembranaire et de la queue cytoplasmique. Cet ensemble de modifications définissent les enveloppes virales dites SOSIP 140 qui présentent les énormes avantages d'être des trimères stables de gp120/gp41 sous forme soluble et de présenter une organisation tridimensionnelle globale assez similaire à celle des protéines sauvages [220]. En revanche même si le trimère obtenu est relativement stable, il nécessite toujours l'utilisation de ligands pour pouvoir être cristallisé.

Même si la mise au point des enveloppes SOSIP 140 a permis des avancées majeures dans l'obtention de structures de trimères, il reste que les SOSIP 140 sont des enveloppes mutantes et qu'il est toujours très difficile de s'affranchir de la présence des ligands. Des études ont notamment avancé le fait que l'introduction de la mutation I559P rendait certaines enveloppes sauvages incapable de réaliser l'entrée virale [221], questionnant le degré de validité des résultats obtenus. Les propos qui sont tenus ici doivent bien être nuancés : ce discours n'a en aucun cas pour but de minorer les résultats des études cristallographiques obtenues avec les enveloppes SOSIP 140 et encore moins à essayer de démontrer que ces études apportent peu au domaine. Au contraire, ces travaux ont permis aussi bien des avancées de nos connaissances que des avancées techniques. De plus les résultats obtenus par les différents travaux se recoupent assez bien les uns avec les autres ce qui témoigne de leur validité générale. En effet, les différentes études qui ont été réalisées avec diverses techniques, que ce soit la diffraction aux rayons X [135,222] ou la cryo-microscopie électronique [223], et sur des sous-types viraux variés [224], présentent des résultats remarquablement similaires lorsque que l'on regarde l'organisation globale des régions constituant les trimères d'enveloppes. Cependant, des divergences apparaissent lorsque l'on observe les structures à de plus petites échelles et notamment lorsqu'on cherche à étudier la disposition d'acides aminés d'intérêts, ce qui démontre que malgré tout, ces études ne sont pas exemptes de défauts et de biais.

Ces divergences peuvent avoir plusieurs origines : d'une part des artefacts provenant des méthodes utilisées pour stabiliser les trimères d'enveloppes, et d'autre part ces divergences peuvent avoir une origine naturelle et provenir de la flexibilité de la protéine d'enveloppe connue pour pouvoir adopter plusieurs conformations différentes. Connaître la cause de ces divergences serait un atout certain car cela permettrait de connaître les limites de fiabilité d'un modèle structural donné.

Pour ce faire, utiliser une approche passant par la caractérisation des réseaux de coévolution qui constituent l'enveloppe serait une solution intéressante. Les événements de coévolution apparaissent en réponse aux contraintes conjuguées du besoin de fonctionnalité et du besoin d'échapper à la réponse immunitaire, leur étude pourrait donc permettre de caractériser des interactions vitales entre acides aminés et de guider la lecture des structures existantes de la protéine d'enveloppe.

L'avantage majeur d'une telle étude est de pouvoir rendre compte des liens qui unissent des acides aminés même si ceux-ci interagissent de façon transitoire au cours d'un changement de conformation, le seul prérequis est que cette interaction soit suffisamment importante pour nécessiter une coévolution des partenaires. Une telle approche n'est donc pas limitée par notre capacité à figer l'enveloppe dans un état donné. De plus, contrairement aux approches structurales

pures, il est éventuellement ici possible de caractériser des interactions indirectes entre acides aminés, qui fonctionneraient de concert sans pour autant présenter de proximité spatiale.

Pour conclure, et en s'éloignant de la finalité structurale qui est une thématique intéressante mais qui ne constitue pas notre objectif principal, il faut insister sur le fait que l'un des points les plus importants est qu'une telle étude fournirait des informations sur la façon dont l'enveloppe de VIH-1 évolue et concilie les besoins de variabilité et de stabilité.

# Premier Article

**Genetic diversity of the highly variable V1 region interferes  
with Human Immunodeficiency Virus type 1 envelope  
functionality**

*Meriem HAMOUDI, Etienne SIMON-LORIERE, Romain GASSER and Matteo NEGRONI*



## Premier article

### Introduction

Le tout premier projet qui constitue cette thèse a consisté à étudier les réseaux de coévolution qui existent entre les régions V1-V2 de la gp120 et le reste de la glycoprotéine d'enveloppe. Antérieurement à notre étude, le rôle le mieux décrit de ces régions variables était de servir de protection contre la réponse immunitaire humorale grâce à leurs capacités à varier rapidement pour échapper à tout anticorps neutralisant nouvellement produit. Ainsi les régions V1-V2 étaient surtout perçues comme étant des régions faiblement impliquées dans l'activité de la glycoprotéine d'enveloppe et presque entièrement libres de varier afin de promouvoir l'échappement aux anticorps [225].

Nous avons donc décidé de nous intéresser à ces régions dans le but de savoir si celles-ci possédaient, malgré tout, des liens coévolutifs avec le reste de la glycoprotéine qui pourraient limiter leurs capacités à accumuler des mutations. Etant donné que ces régions sont activement ciblées par le système immunitaire et sont impliquées dans la mise au point d'un vaccin efficace contre VIH-1 [226], pouvoir ainsi mettre en évidence les limites avec lesquelles ces domaines peuvent varier présente un intérêt scientifique fort.

Afin d'étudier les liens de coévolution entre V1-V2 et le reste de la glycoprotéine d'enveloppe nous avons eu recours à la méthode citée dans l'introduction : nous avons construit des gènes *env* chimériques dans lesquels les régions V1 ou V2 ou V1-V2 ensembles et le reste de *env* étaient de deux origines phylogénétiques différentes et nous avons quantifiés le niveau de fonctionnalité de ces chimères en testant leur capacité à réaliser l'étape d'entrée virale.

Les résultats ainsi obtenus nous ont permis de mettre en évidence non seulement le fait que les régions V1 et V2 coévoluent en réalité avec le reste de la glycoprotéine d'enveloppe mais aussi que les liens coévolutifs entre V1 et le reste de la glycoprotéine sont surprenamment forts car peu de chimères ont été trouvées être fonctionnelles lorsque V1 n'était pas de la même origine phylogénétique que le reste de la protéine.

Après avoir obtenu ce résultat, nous avons tenté de caractériser la nature du ou des partenaires coévolutifs de ces régions variables. Nous avons réussi à démontrer que ce partenaire n'était pas la région variable V3, contrairement à ce que laissaient supposer les structures tridimensionnelles disponibles à ce moment, et qui plaçaient les régions V1-V2 et V3 à proximité les unes des autres. Ceci a pu être démontré par le fait que les chimères dans lesquelles les régions V1-V2 et V3 partagent une origine phylogénétique différente de celle du reste de la glycoprotéine d'enveloppe ne sont pas

plus fonctionnelles que les chimères où seules V1-V2 ont une origine phylogénétique différente des autres régions.

Enfin, le dernier apport majeur de cette étude a consisté à démontrer que le défaut fonctionnel lié à la perturbation du réseau de coévolution de V1-V2 était lié à un problème de shedding prématuré de la gp120. Des données disponibles alors laissaient suspecter que les trois régions V1-V2 d'un trimère étaient impliquées dans la stabilité du trimère d'enveloppe mais sans en apporter la preuve [227,228]. Les résultats confirment l'hypothèse et tendent à démontrer que si V1-V2 coévoluent avec le reste de la glycoprotéine c'est en grande partie afin de garantir la stabilité des interactions entre les sous-unités gp120 et gp41 ; et ce malgré le fait que ces régions soient spatialement très éloignées des zones d'interactions entre ces deux sous-unités.

# Second Article

**Buffering deleterious polymorphisms in highly constrained parts of HIV-1 envelope by flexible regions: the case of V3 and C2**

*Romain GASSER, Meriem HAMOUDI, Martina PELLICCIOTTA, Zhicheng ZHOU, Philippe COLIN, Bernard LAGANE and Matteo NEGRONI*



## Second article

### Introduction

Après nous être intéressés aux réseaux de coévolution reliant les régions V1-V2 avec les autres régions de la glycoprotéine d'enveloppe, nous avons conduit une étude similaire sur la région C2. L'idée de réaliser ce second projet a trouvé son impulsion dans un résultat obtenu précédemment par le laboratoire et décrivant le fait qu'un hot spot de recombinaison existe dans la région C2 de la gp120, mais que celui-ci ne génère quasi exclusivement que des protéines peu ou pas fonctionnelles [195].

Nous avons ainsi supposé que la raison se tenant derrière cette observation est qu'un réseau de coévolution absolument vital pour le maintien de la fonctionnalité de la glycoprotéine d'enveloppe relie des acides aminés présents dans C2 à d'autres acides aminés présents ailleurs dans Env. Pour tester cette hypothèse, la même stratégie expérimentale a été utilisée, à savoir la création de protéines d'enveloppes chimériques dans lesquelles la région C2 n'as pas la même origine phylogénétique que le reste de la glycoprotéine d'enveloppe. Ces premières constructions ont permis de confirmer les résultats précédemment obtenus car dans toutes nos constructions, dès que la région C2 ne possède pas la même origine phylogénétique que le reste de Env, la glycoprotéine chimérique est très peu ou pas fonctionnelle.

Par la génération extensive de glycoprotéines chimériques dans lesquelles des sous-régions, puis des acides aminés ponctuels de C2 ont été remplacés nous avons réussi à mettre en évidence le fait que l'acide aminé clé du réseau de coévolution de la région C2 est l'acide aminé 202. En effet, si cet acide aminé 202 partage la même origine phylogénétique que toutes les autres régions de Env, exceptée C2, cela est suffisant pour garantir la fonctionnalité de la glycoprotéine d'enveloppe.

La suite logique de ce résultat a été de chercher à caractériser le ou les partenaires coévolutifs de cet acide aminé 202. Ainsi nous avons pu démontrer que le partenaire coévolutif du résidu 202 se trouve dans la région variable V3. L'étude a pu être poussée plus avant et a également permis de définir les acides aminés 305 et 328 de la région V3 comme étant les partenaires de l'acide aminé 202. Cependant, et comme il sera discuté dans l'article, bien que le fait qu'il est certain que les résidus 305 et 328 soient des partenaires coévolutifs de l'acide aminé 202 soit une certitude, il est en revanche moins certain que ces deux résidus le soient pour toutes les souches de VIH-1. Ainsi le fait que V3 et 202 coévoluent ensemble mais que les acides aminés de V3 qui coévoluent avec le résidu 202 puissent varier suivant les souches de VIH-1 considérées est un phénomène intéressant. Ceci démontre la plasticité génétique de V3 et cela tend à donner à cette région le rôle de « tampon »

visant à accommoder les mutations survenant dans des régions disposant de moins de plasticité génétique, en l'occurrence dans C2.

La seconde partie du projet a consisté à caractériser le défaut fonctionnel associé à la perturbation du réseau de coévolution de l'acide aminé 202. Etant donné que la littérature scientifique indique que l'acide aminé fait partie du « bridging sheet », qui est une structure jouant un rôle dans la liaison au corécepteur, nous avons rapidement suspecté que le défaut pouvait être l'impossibilité, pour les glycoprotéines chimériques non fonctionnelles, de lier le corécepteur après liaison au CD4. Toutefois, les différents tests d'incubation, de pharmacocinétiques, et de liaison de virions entiers à des cellules cibles ont démontrés que le défaut n'était pas dû à l'absence de liaison au corécepteur mais était vraisemblablement dû à l'impossibilité pour les glycoprotéines chimériques non fonctionnelles de réaliser les changements de conformations survenant normalement après la liaison au corécepteur. Ainsi, ce blocage empêche l'activation de la gp41 et, incidemment, la fusion des membranes cellulaires et virales.

C'est la première fois qu'est décrite une mutation ne touchant pas un site de liaison à un récepteur ou corécepteur et empêchant la gp120 de réaliser un changement de conformation. Les acides aminés membres du réseau de coévolution que nous avons identifiés, bien qu'étant proches des résidus chargés de la reconnaissance et de la liaison au corécepteur, ne sont pas décrits comme établissant des interactions directes avec le corécepteur [131]. Ceci sous-entend que lors de la reconnaissance de CCR5 par V3, ces deux partenaires peuvent s'associer selon différentes conformations dont certaines seulement sont fonctionnelles et compatibles avec la poursuite des changements de conformations nécessaires à la fusion des membranes cellulaires et virales. Ceci implique également que le mécanisme régissant les changements de conformation induits par la liaison à CCR5 est assez complexe et ne peut être déduit des structures tridimensionnelles du trimères d'enveloppes présentes actuellement dans la littérature.

# Discussion



## Discussion de la thèse

Les deux études qui constituent cette thèse s'inscrivent dans la même optique, celle de caractériser les réseaux de coévolution présents dans la glycoprotéine d'enveloppe du VIH-1. L'intérêt d'une telle approche s'explique par le fait que les réseaux de coévolution qui apparaissent dans la glycoprotéine d'enveloppe ont pour but de concilier les besoins opposés du maintien de la fonctionnalité de la protéine et de la variation antigénique. Etudier les réseaux de coévolution apporte donc théoriquement le double intérêt de permettre de comprendre comment la glycoprotéine d'enveloppe accumule des mutations et comment elle parvient à échapper aux anticorps neutralisants produits par le système immunitaire. Les deux travaux diffèrent cependant quant à leurs objectifs et les résultats qu'ils exposent, car là où le premier article se concentre sur les régions V1-V2 et conclue que le rôle de ces régions n'est pas seulement de fournir une protection à la glycoprotéine d'enveloppe contre les anticorps neutralisants, le second met à jour un réseau de coévolution essentiel incluant des acides aminés appartenant aux régions C2 et V3 et attribue à V3 le rôle de région tampon capable d'accommoder les mutations apparaissant dans la région C2.

Le premier article traduit notre volonté d'explorer la possibilité que les régions variables, et en l'occurrence les régions V1-V2, puissent jouer un rôle crucial dans la fonctionnalité de la glycoprotéine d'enveloppe du VIH-1. Il était déjà décrit dans la littérature que ces régions étaient impliquées dans les mécanismes d'échappement de cette protéine au système immunitaire et donc que ces régions jouaient un rôle important pour la glycoprotéine d'enveloppe [225]. Il avait également déjà été montré que la délétion complète de ces régions impactait négativement la fonctionnalité de la glycoprotéine d'enveloppe sans pour autant l'abolir [229]. Néanmoins, cette étude a réussi à démontrer que les régions V1-V2 sont également nécessaires au bon fonctionnement de la glycoprotéine d'enveloppe et qu'elles doivent par conséquent coévoluer avec les autres régions qui constituent cette protéine.

La seconde étude, centrée sur la région C2, n'avait, quant à elle, pas pour but véritable de comprendre comment cette région était impliquée dans la fonctionnalité de la glycoprotéine d'enveloppe. En effet, cette région est très conservée et, de fait, il avait déjà été suspecté puis démontré par d'autres travaux que cette région joue un rôle non négligeable pour la glycoprotéine d'enveloppe afin d'expliquer cette conservation. Il a ainsi été décrit dans la littérature que cette région est impliquée dans la formation du « bridging sheet » [106], et qu'elle joue un rôle dans la liaison au corécepteur [131]. Le but de notre étude était donc plutôt de caractériser les partenaires coévoluant étroitement avec C2, et ce dans l'optique de comprendre par quel mécanisme cette région peut maintenir son activité lorsqu'elle subit des mutations.

Les résultats présentés dans la première étude démontrent que les régions V1-V2 coévoluent avec les autres régions de la glycoprotéine d'enveloppe. En effet, V1-V2 perturbent la fonctionnalité de Env lorsqu'elles ne partagent pas la même origine phylogénétique que le reste de la protéine. Cependant, le ou les partenaires coévolutifs de V1-V2 n'ont pas pu être mis en évidence lors de ces expériences. Les travaux publiés antérieurement indiquaient que V1-V2 était à proximité immédiate de V3 dans la conformation native du trimère, ce qui laissait supposer que ces régions pouvaient coévoluer ensemble. Les résultats ont infirmé cette supposition car les chimères dans lesquelles V1-V2 et V3 partagent la même origine phylogénétique ne présentent pas de regain de fonctionnalité significatifs.

A ce jour le ou les partenaires coévolutifs de V1-V2 sont toujours inconnus et aucune information présente dans la littérature scientifique ne permet de diriger les recherches en permettant de supposer quelle région peut ou ne peut pas être le partenaire. Cette question demeure aujourd'hui en suspens.

Concernant le deuxième article, comprendre comment la région C2 pouvait maintenir son activité tout en étant sujet à des mutations a trouvé un élément de réponse. La glycoprotéine chimérique dans laquelle la région C2 et la région V3 partagent une origine phylogénétique commune présente un regain de fonctionnalité par rapport à la chimère dans laquelle la région C2 est d'une origine phylogénétique différente du reste de l'enveloppe (figure 4 de l'article 2). Les résultats ont ensuite pu être poussés plus avant et ont permis l'identification précise des acides aminés 202, de la région C2, et 305 et 328 de la région V3 comme faisant partie du même réseau de coévolution.

Cependant, ces résultats n'excluent pas la possibilité que ce réseau de coévolution ne soit pas encore complètement caractérisé. Pour des raisons techniques, la recherche des partenaires coévolutifs de l'acide aminé 202 a dû se faire en utilisant une glycoprotéine Env de sous-type B ayant une région V3 issue d'un sous-type viral C. Pour confirmer nos résultats nous nous sommes donc affranchis dans un second temps de cette région V3 de sous-type C pour reprendre comme modèle une enveloppe sauvage de sous-type B. Comme indiqué dans l'article, l'insertion des acides aminés 202, 305 et 328 d'un sous-type viral G dans une glycoprotéine sauvage de sous-type B n'a pas eu pour résultat un regain de fonctionnalité par rapport à la glycoprotéine non fonctionnelle où seul l'acide aminé 202 est de sous-type viral G. Il est donc très probable que d'autres acides aminés fassent partie de ce réseau de coévolution et restent encore à caractériser.

Comme évoqué dans la discussion de l'article 2, une autre interprétation possible de ce résultat négatif est que le réseau de coévolution identifié soit celui correspondant aux régions variables de sous-type C. Dans cette hypothèse les régions V3 de sous-type viral B et de sous-type viral C ont

chacune un réseau de coévolution différent dans lequel les acides aminés partenaires de 202 ne sont pas les mêmes. Mais même dans cette hypothèse, le fait que des acides aminés de V3 et l'acide aminé 202 coévoluent ensemble de façon soutenue constitue un résultat remarquable et inédit. De plus, toujours dans l'hypothèse où chaque sous-type, voire chaque souche virale, dispose d'un réseau de coévolution particulier entre V3 et l'acide aminé 202, ceci constituerait une preuve de la plasticité de V3 qui serait capable d'accommoder les mutations apparaissant dans C2 par la mise en place de différents réseaux de coévolution. Le fait que les régions variables puissent servir de « tampon » permettant d'accroître la flexibilité génétique des domaines constants a déjà été décrit dans d'autres travaux et sur d'autres protéines que celles de VIH-1 [230-233], mais notre étude permet d'apporter une observation supplémentaire qui tend à généraliser cette mécanistique.

En ce qui concerne la région C2, il est également très probable que d'autres réseaux de coévolution incluant des acides aminés de C2 restent également à identifier. Cette supposition est soutenue par les constructions dans lesquelles les deux sous-régions terminales de C2, à savoir R2 et R3, partagent la même origine que la glycoprotéine d'enveloppe mais que les autres parties de C2 sont d'une origine phylogénétique différente (figure 2a et 2b de l'article 2). Ces chimères demeurent peu ou pas fonctionnelles en comparaison de la glycoprotéine parentale sauvage, ce qui signifie que les réseaux de coévolution putatifs qui composent ces sous-régions n'ont pas un effet suffisamment fort pour provoquer un regain significatif de fonctionnalité, mais cela ne démontre en aucune façon que ces réseaux n'existent pas. Néanmoins, le point négatif soulevé par ces résultats peut être minoré car le résultat principal reste que l'acide aminé 202 peut à lui seul garantir la fonctionnalité de la glycoprotéine d'enveloppe à condition qu'il soit de la même origine phylogénétique que le reste de la glycoprotéine, peu importe l'origine de C2. De plus, des données non publiées obtenues au laboratoire soutiennent que la position 202 revêt une grande importance dans le réseau de coévolution de C2. Nous avons construit une série de chimères, incomplète à l'heure actuelle, qui consiste à insérer un résidu 202 dans un contexte phylogénétique qui n'est pas le sien. Tous les résultats obtenus jusqu'à maintenant montrent que lorsque l'acide aminé présent précisément en position 202 n'est pas de la même origine phylogénétique que le reste de la protéine il en résulte une baisse significative du niveau de fonctionnalité de la glycoprotéine d'enveloppe, ce qui renforce l'importance de cette position dans le réseau de coévolution de la région C2 et tend à généraliser cette importance au moins aux souches du groupe M de VIH-1.

Après identification du réseau de coévolution impliquant les acides aminés 202, 305 et 328, la caractérisation fonctionnelle du défaut induit par la perturbation de ce réseau a été entreprise. Le but de cette caractérisation était de démontrer que le défaut était bien lié à la reconnaissance du corécepteur. En effet, étant donné que l'acide aminé 202 est un constituant du « bridging sheet »,

élément conformationnel impliqué dans la liaison du corécepteur, et que la région V3 est l'acteur majeur participant à la liaison à ce même corécepteur, il était attendu que le défaut soit en rapport avec cette étape de l'entrée virale. Le résultat qui a été donné par les différents tests fonctionnels a été assez surprenant et inédit. Les quatre tests principaux effectués, et dont les résultats sont détaillés dans les figures 6 et 7 de l'article 2, permettent de circonscrire la nature du défaut aux changements de conformations de la gp120 ayant lieu après la liaison du corécepteur. Bien que cette déduction soit faite sur la base de preuve indirecte c'est une conclusion solide.

Il a été démontré que les glycoprotéine d'enveloppes dont le réseau de coévolution est perturbé sont néanmoins incorporées en quantité suffisantes à la surface des particules virales (figure 6a et 6b de l'article 2), et qu'elles sont capables de lier CD4 comme CCR5, que ce soit en tests pharmacocinétiques effectués sur des gp120 solubles (figure 6c et 6d de l'article 2) ou en culture cellulaire en analysant la fixation de virions entiers sur des cellules cibles (figure 7a et 7b de l'article 2). La seule hypothèse restante autre qu'un blocage des changements de conformation effectués après fixation à CCR5 est un défaut associé à la gp41 qui serait incapable de réaliser l'étape de fusion membranaire. Cette hypothèse semble pouvoir être rapidement écartée car en aucune façon la gp41 n'a été manipulée dans nos expériences, elle est donc toujours restée sauvage et *a priori* fonctionnelle. Une confirmation directe de ce blocage dans les changements de conformations de la gp120 aurait pu être obtenue en utilisant des techniques telles que le « deuterium exchange » qui a déjà permis de mettre en évidence des changements de conformations effectués par la gp120 [119,215,234–236], mais ceci n'a pas été entrepris pour des raisons de temps.

Ce résultat est inédit car c'est la première fois qu'est décrite une mutation capable de bloquer la glycoprotéine d'enveloppe dans un de ses changements de conformations et sans être directement impliquée dans la reconnaissance du récepteur ou du corécepteur. L'intérêt de cette découverte est autre que purement académique car des applications pratiques peuvent éventuellement en découler. La production de trimères d'enveloppes rendus non fonctionnels à cause de la perturbation du réseau de coévolution associé à l'acide aminé 202 peut ainsi être envisagé. Le but serait dans ce cas de produire des glycoprotéines d'enveloppes qui se comportent en tout point comme des protéines sauvages en ce qui concerne la liaison au récepteur et au corécepteur, mais restent bloquées dans la conformation « corécepteur lié ». De telles protéines pourraient permettre d'exposer de façon prolongée des épitopes qui sont normalement présents de façon transitoire à la surface des protéines sauvages, et ainsi faciliter la production des anticorps qui ciblent ces épitopes.

L'intérêt thérapeutique d'une telle approche se trouve dans la possibilité d'utiliser ces glycoprotéines non fonctionnelles comme immunogènes vaccinaux. Etant donné que les épitopes exposés par les protéines immunogènes non fonctionnelles ne sont en revanche exposés que de façon transitoire par les protéines présentes dans la nature, il n'est pas certain que ces immunogènes

puissent aider l'hôte à produire des anticorps neutralisants efficaces. En effet, les épitopes ciblés par ces nouveaux anticorps ne seront exposés que de façon éphémères par les protéines virales sauvages. Cependant, même si l'utilité de favoriser la production de tels anticorps dans le cadre d'une thérapie vaccinale est sujette à question, considérer une telle approche n'est pas sans intérêt. Les précédents essais vaccinaux n'ont eu qu'un succès très limité dans la génération d'anticorps protecteurs chez les personnes vaccinées [237-239], et ce principalement parce que les anticorps générés peinent à neutraliser efficacement la grande diversité de souches de VIH-1 existantes. Aussi, tenter de favoriser la production de classes d'anticorps qui ne sont pas normalement produits lors de l'infection peut se révéler être un complément intéressant visant à diversifier la population d'anticorps qu'est capable de produire un individu. Dans cette optique les glycoprotéines immunogènes construites sur la base de la perturbation du réseau de coévolution que nous avons décrit ne remplaceraient pas les immunogènes actuels mais viendraient les compléter.

Le deuxième objectif principal de l'étude reportée dans le second article était, au travers de la caractérisation des réseaux de coévolution, de pouvoir mettre à jour des ensembles d'acides aminés interagissant et permettant de guider la lecture des structures tridimensionnelles présentes dans la littérature. Pour resituer la problématique, le défaut principal des structures existantes est le fait que pouvoir les obtenir nécessite presque dans tous les cas de modifier la glycoprotéine d'enveloppe par mutagenèse ou de lui adjoindre des molécules stabilisatrices, comme des anticorps, qui risquent de modifier sa structure pour lui donner une conformation non naturelle. Trouver un ensemble d'acides aminés qui coévoilent et dont nous sommes sûrs qu'ils interagissent physiquement les uns avec les autres permettrait de comparer et d'avoir une lecture critique des structures existantes sur la base de la présence ou de l'absence de ce regroupement spatial hypothétique d'acides aminés.

Réaliser une telle approche n'est pas possible en utilisant le réseau de coévolution de l'acide aminé 202. Comme précisé dans mon introduction, toutes les structures existantes s'accordent sur le fait que le « bridging sheet » comprenant l'acide aminé 202 et la boucle V3 qui contient les acides aminés 305 et 328 sont éloignés dans l'espace au moment de la liaison au corécepteur. Ceci implique que le phénomène de coévolution qui existe entre ces trois acides aminés n'a pas pour base une interaction physique devant être maintenue entre ces résidus mais se fait par l'intermédiaire d'interactions indirectes ou par des phénomènes de compensations. Dans ces conditions il devient très difficile d'utiliser les liens qui unissent ces trois acides aminés pour tenter de discriminer les structures tridimensionnelles entre elles.

A *contrario*, cette étude trouve quand même un intérêt structural. En effet, les résultats démontrent que les trois acides aminés que nous avons caractérisés doivent coévoluer pour maintenir le bon déroulement de changements de conformations cruciaux, pourtant ils n'ont pas été

décrits dans la littérature comme étant directement impliqués dans la liaison au corécepteur. Or, le fait que des acides aminés qui ne participent pas à la liaison d'un ligand donné soient nécessaires à la réalisation d'un changement conformationnel est une observation très intéressante. Le consensus existant à l'heure actuel soutient d'avantage le fait que ce sont les acides aminés qui engagent une interaction avec le ligand qui provoquent, par cette interaction, le changement de conformation [240]. Pour soutenir notre observation, les expériences de liaisons à CCR5 qui ont été réalisées montrent bien que la liaison au corécepteur est correctement effectuée même par la protéine non fonctionnelle B 202G chez qui le réseau de coévolution a été rompu (figure 7d de l'article). Il est donc effectivement peu probable que les trois acides aminés caractérisés par cette étude interagissent directement avec CCR5. Ce qui est démontré ici est donc que les changements de conformations réalisés par les protéines peuvent impliquer des acides aminés qui n'entrent pas en contact direct avec des ligands mais qui sont pourtant cruciaux. Caractériser de tels acides aminés n'aurait pas été possible sur la base des structures tridimensionnelles présentes dans la littérature et seul l'étude de la coévolution pouvait permettre d'arriver à une telle conclusion.

Ces constatations viennent globalement renforcer notre idée que l'étude d'une protéine au travers de la caractérisation de ses réseaux de coévolution permet de fournir des données hautement intéressantes qui vont au moins compléter les structures publiées et, éventuellement, vont permettre une lecture critique de ces mêmes structures.

# Conclusion



## Conclusion

Les deux études menées pendant cette thèse permettent de commencer à dessiner les réseaux de coévolution qui composent la glycoprotéine d'enveloppe du VIH-1.

Au travers des observations que nous avons obtenues nous pouvons remarquer que de nouveaux rôles importants s'affirment pour les régions variables qui composent la gp120. Ces régions avaient déjà été décrites dans la littérature comme possédant des rôles cruciaux, tels que favoriser l'échappement à la réponse immunitaire humorale pour les régions V1-V2 ou être l'élément central de la liaison au corécepteur dans le cas de la région V3.

Cependant, nos travaux ont permis de démontrer de façon certaine qu'il y a une limite à laquelle les régions V1-V2 peuvent varier car ces deux régions variables sont non seulement impliquées dans les mécanismes d'échappement au système immunitaire mais également dans la stabilité du trimère d'enveloppe. Elles doivent en conséquence coévoluer avec un ou des autres partenaires non identifiés pour concilier ce dernier rôle avec les variations constantes de leur séquence protéique.

Il a également été possible de démontrer qu'un ensemble d'acides aminés appartenant aux régions C2 et V3 coévoluent de concert afin de garantir le bon déroulement des changements de conformations qui surviennent après la liaison de la gp120 à son corécepteur. Cet ensemble d'acides aminés comprend le résidu 202 de la région C2 avec pour partenaires les acides aminés 305 et 328 de la région V3, bien qu'il semble que la nature des partenaires de la région V3 puisse varier selon la souche de VIH-1 considérée et l'histoire évolutive de celle-ci. Ce résultat sous-entend également que l'un des mécanismes par lesquels la région constante C2 incorpore des mutations, alors qu'elle-même tolère très mal l'apparition de variation dans sa séquence, consiste à coévoluer avec des régions variables, en l'occurrence V3.

Cet ensemble de résultats contribue à faire évoluer notre regard sur les régions variables de la glycoprotéine d'enveloppe de VIH-1 et permet de moins les percevoir comme étant les acteurs exclusifs permettant l'échappement à la réponse immunitaire. Ces régions peuvent en effet aussi bien participer à la stabilisation du trimère d'enveloppe dans sa conformation quaternaire, comme c'est le cas pour V1-V2, ou permettre d'augmenter la robustesse génétique de la protéine en offrant des partenaires coévolutifs aux régions constantes comme nous l'avons décrit pour V3.

Pour asseoir d'avantage et parfaire les conclusions exposées ici, d'autres travaux sont à envisager au premier desquels la caractérisation du partenaire coévolutif de V1-V2, qui nous échappe toujours. Caractériser complètement le réseau de coévolution de V1-V2 permettrait très certainement de définir un domaine fonctionnel contribuant à la stabilisation du trimère d'enveloppe et permettrait de comprendre par quel mécanisme cette stabilisation peut avoir lieu. Ensuite, réussir à caractériser

le réseau de coévolution de l'acide aminé 202 dans le contexte de régions V3 provenant d'autres sous-types viraux permettrait d'acquérir une connaissance approfondie de mécanisme de coévolution et des prérequis nécessaires au bon déroulement des changements de conformations ayants lieux après liaison au corécepteur. De plus, il serait intéressant d'étudier plus en avant la possibilité que toutes les régions variables de la glycoprotéine d'enveloppe puissent servir à augmenter la robustesse génétique de cette protéine en coévoluant avec les régions centrales et constantes de Env.

Enfin même si les travaux menés dans le cadre de cette thèse offrent beaucoup de perspectives dans le domaine de la recherche fondamentale, il faut souligner que des applications potentielles existent dans le domaine appliqué. Il est en effet envisageable d'utiliser le réseau de coévolution de l'acide aminé 202 dans le cadre de la recherche vaccinale pour produire une nouvelle gamme d'immunogènes permettant d'exposer de façon prolongée les épitopes formés suite à la reconnaissance du corécepteur.

## Bibliographie

1. AIDS by the numbers 2015. Available at: [http://www.unaids.org/en/resources/documents/2015/AIDS\\_by\\_the\\_numbers\\_2015](http://www.unaids.org/en/resources/documents/2015/AIDS_by_the_numbers_2015). (Accessed: 7th April 2016)
2. Bleyzac, P., Exbrayat, J.-M. & Fellah, J. S. [On the origin of the adaptative immune system (AIS): the hypothesis]. *Médecine Sci. MS* **21**, 210–215 (2005).
3. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). 1983. *Rev. Investig. Clínica Organo Hosp. Enfermedades Nutr.* **56**, 126–129 (2004).
4. De Leys, R. *et al.* Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J. Virol.* **64**, 1207–1216 (1990).
5. Simon, F. *et al.* Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**, 1032–1037 (1998).
6. Plantier, J.-C. *et al.* A new human immunodeficiency virus derived from gorillas. *Nat. Med.* **15**, 871–872 (2009).
7. Faria, N. R. *et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014).
8. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011).
9. D'arc, M. *et al.* Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1343–1352 (2015).
10. Lemey, P. *et al.* Tracing the origin and history of the HIV-2 epidemic. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6588–6592 (2003).
11. de Silva, T. I., Cotten, M. & Rowland-Jones, S. L. HIV-2: the forgotten AIDS virus. *Trends Microbiol.* **16**, 588–595 (2008).
12. Buonaguro, L., Tornesello, M. L. & Buonaguro, F. M. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.* **81**, 10209–10219 (2007).
13. Gall, A. *et al.* Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* **50**, 3838–3844 (2012).
14. Junqueira, D. M. *et al.* Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One* **6**, e27489 (2011).
15. Hemelaar, J. Implications of HIV diversity for the HIV-1 pandemic. *J. Infect.* **66**, 391–400 (2013).

16. Pyne, M. T., Hackett, J., Holzmayer, V. & Hillyard, D. R. Large-scale analysis of the prevalence and geographic distribution of HIV-1 non-B variants in the United States. *J. Clin. Microbiol.* **51**, 2662–2669 (2013).
17. HIV Databases. Available at: <http://www.hiv.lanl.gov/content/index>. (Accessed: 7th April 2016)
18. ICTV Virus Taxonomy 2014. Available at: <http://www.ictvonline.org/virustaxonomy.asp>. (Accessed: 7th April 2016)
19. Coffin, J. M. Retroviral DNA integration. *Dev. Biol. Stand.* **76**, 141–151 (1992).
20. Rethwilm, A. & Bodem, J. Evolution of foamy viruses: the most ancient of all retroviruses. *Viruses* **5**, 2349–2374 (2013).
21. Haase, A. T. Pathogenesis of lentivirus infections. *Nature* **322**, 130–136 (1986).
22. Appay, V. & Sauce, D. Immune activation and inflammation in HIV-1 infection: causes and consequences. *J. Pathol.* **214**, 231–241 (2008).
23. Weinberg, J. B., Matthews, T. J., Cullen, B. R. & Malim, M. H. Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. *J. Exp. Med.* **174**, 1477–1482 (1991).
24. Splettstoesser, T. SciStyle. Available at: <http://www.scistyle.com/>. (Accessed: 29th March 2016)
25. Joshi, S. & Joshi, R. L. Molecular biology of human immunodeficiency virus type-1. *Transfus. Sci.* **17**, 351–378 (1996).
26. Jones, K. A., Kadonaga, J. T., Luciw, P. A. & Tjian, R. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**, 755–759 (1986).
27. Kawakami, K., Scheidereit, C. & Roeder, R. G. Identification and purification of a human immunoglobulin-enhancer-binding protein (NF-kappa B) that activates transcription from a human immunodeficiency virus type 1 promoter in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4700–4704 (1988).
28. Kilareski, E. M., Shah, S., Nonnemacher, M. R. & Wigdahl, B. Regulation of HIV-1 transcription in cells of the monocyte-macrophage lineage. *Retrovirology* **6**, 118 (2009).
29. Siekevitz, M. *et al.* Activation of the HIV-1 LTR by T cell mitogens and the trans-activator protein of HTLV-I. *Science* **238**, 1575–1578 (1987).
30. Bosque, A. & Planelles, V. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood* **113**, 58–65 (2009).
31. Pereira, L. A., Bentley, K., Peeters, A., Churchill, M. J. & Deacon, N. J. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res.* **28**, 663–668 (2000).
32. Ashe, M. P., Furger, A. & Proudfoot, N. J. Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA N. Y. N* **6**, 170–177 (2000).

33. Gilmartin, G. M., Fleming, E. S., Oetjen, J. & Graveley, B. R. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev.* **9**, 72–83 (1995).
34. Van Lint, C. *et al.* Transcription factor binding sites downstream of the human immunodeficiency virus type 1 transcription start site are important for virus infectivity. *J. Virol.* **71**, 6113–6127 (1997).
35. Cullen, B. R. Regulation of HIV-1 gene expression. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **5**, 2361–2368 (1991).
36. Black, D. L. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
37. Holmes, E. C. What does virus evolution tell us about virus origins? *J. Virol.* **85**, 5247–5251 (2011).
38. Lauber, C. *et al.* The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog.* **9**, e1003500 (2013).
39. Kammler, S. *et al.* The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology* **3**, 89 (2006).
40. Schwartz, S., Felber, B. K., Benko, D. M., Fenyő, E. M. & Pavlakis, G. N. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J. Virol.* **64**, 2519–2529 (1990).
41. ViralZone. Available at: <http://viralzone.expasy.org/>. (Accessed: 30th March 2016)
42. Gerhard, D. S. *et al.* The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.* **14**, 2121–2127 (2004).
43. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**, 489–493 (1987).
44. Verhoef, K., Tijms, M. & Berkhout, B. Optimal Tat-mediated activation of the HIV-1 LTR promoter requires a full-length TAR RNA hairpin. *Nucleic Acids Res.* **25**, 496–502 (1997).
45. Karn, J. & Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med.* **2**, a006916 (2012).
46. Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254–257 (1989).
47. Truant, R. & Cullen, B. R. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol. Cell. Biol.* **19**, 1210–1217 (1999).
48. Rausch, J. W. & Le Grice, S. F. J. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses* **7**, 3053–3075 (2015).

49. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
50. Ahmad, N. & Venkatesan, S. Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. *Science* **241**, 1481–1485 (1988).
51. Niederman, T. M., Thielan, B. J. & Ratner, L. Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1128–1132 (1989).
52. Kim, S., Ikeuchi, K., Byrn, R., Groopman, J. & Baltimore, D. Lack of a negative influence on viral growth by the nef gene of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9544–9548 (1989).
53. Hammes, S. R., Dixon, E. P., Malim, M. H., Cullen, B. R. & Greene, W. C. Nef protein of human immunodeficiency virus type 1: evidence against its role as a transcriptional inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9549–9553 (1989).
54. Garcia, J. V. & Miller, A. D. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* **350**, 508–511 (1991).
55. Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E. & Trono, D. Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **76**, 853–864 (1994).
56. Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
57. Audibert, F. *et al.* La grippe : stratégies de vaccination. (1994). Available at: <http://lara.inist.fr/handle/2332/1381>. (Accessed: 7th April 2016)
58. Hurley, J. H. & Bonifacino, J. S. Nef-arious goings-on at the Golgi. *Nat. Struct. Mol. Biol.* **19**, 661–662 (2012).
59. Schwartz, O., Maréchal, V., Le Gall, S., Lemonnier, F. & Heard, J. M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat. Med.* **2**, 338–342 (1996).
60. Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D. & Baltimore, D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391**, 397–401 (1998).
61. Farrell, H. E. *et al.* Inhibition of natural killer cells by a cytomegalovirus MHC class I homologue in vivo. *Nature* **386**, 510–514 (1997).
62. Kirchhoff, F., Greenough, T. C., Brettler, D. B., Sullivan, J. L. & Desrosiers, R. C. Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *N. Engl. J. Med.* **332**, 228–232 (1995).
63. He, J. *et al.* Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. *J. Virol.* **69**, 6705–6711 (1995).
64. Goh, W. C. *et al.* HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. *Nat. Med.* **4**, 65–71 (1998).

65. Malim, M. H. & Emerman, M. HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388–398 (2008).
66. Mandal, D., Feng, Z. & Stoltzfus, C. M. Excessive RNA splicing and inhibition of HIV-1 replication induced by modified U1 small nuclear RNAs. *J. Virol.* **84**, 12790–12800 (2010).
67. Seelmeier, S., Schmidt, H., Turk, V. & von der Helm, K. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6612–6616 (1988).
68. Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280–283 (1988).
69. Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. 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 (2001).
70. Jacobs, E., Gheysen, D., Thines, D., Francotte, M. & de Wilde, M. The HIV-1 Gag precursor Pr55gag synthesized in yeast is myristoylated and targeted to the plasma membrane. *Gene* **79**, 71–81 (1989).
71. Abd El-Wahab, E. W. *et al.* Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* **5**, 4304 (2014).
72. Mariani, C., Desdouits, M., Favard, C., Benaroch, P. & Muriaux, D. M. Role of Gag and lipids during HIV-1 assembly in CD4(+) T cells and macrophages. *Front. Microbiol.* **5**, 312 (2014).
73. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–472 (2004).
74. Mundigala, H., Michaux, J. B., Feig, A. L., Ennifar, E. & Rueda, D. HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. *Nucleic Acids Res.* **42**, 7281–7289 (2014).
75. Bukrinskaya, A. HIV-1 matrix protein: a mysterious regulator of the viral life cycle. *Virus Res.* **124**, 1–11 (2007).
76. Ganser-Pornillos, B. K., Yeager, M. & Sundquist, W. I. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* **18**, 203–217 (2008).
77. Ono, A. & Freed, E. O. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13925–13930 (2001).
78. Hawkes, D. *et al.* Properties of HIV-1 associated cholesterol in addition to raft formation are important for virus infection. *Virus Res.* **210**, 18–21 (2015).
79. Sundquist, W. I. & Kräusslich, H.-G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924 (2012).
80. Strebel, K., Klimkait, T. & Martin, M. A. A novel gene of HIV-1, vpu, and its 16-kilodalton product. *Science* **241**, 1221–1223 (1988).

81. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425–430 (2008).
82. Mitchell, R. S. *et al.* Vpu antagonizes BST-2-mediated restriction of HIV-1 release via beta-TrCP and endo-lysosomal trafficking. *PLoS Pathog.* **5**, e1000450 (2009).
83. Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J. Virol.* **66**, 226–234 (1992).
84. Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J. Virol.* **66**, 7193–7200 (1992).
85. Schubert, U. *et al.* Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. *FEBS Lett.* **398**, 12–18 (1996).
86. Magadán, J. G. *et al.* Multilayered mechanism of CD4 downregulation by HIV-1 Vpu involving distinct ER retention and ERAD targeting steps. *PLoS Pathog.* **6**, e1000869 (2010).
87. Roy, N., Pacini, G., Berlioz-Torrent, C. & Janvier, K. Mechanisms underlying HIV-1 Vpu-mediated viral egress. *Front. Microbiol.* **5**, 177 (2014).
88. Hansen, J., Billich, S., Schulze, T., Sukrow, S. & Moelling, K. Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. *EMBO J.* **7**, 1785–1791 (1988).
89. Fun, A., Wensing, A. M. J., Verheyen, J. & Nijhuis, M. Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology* **9**, 63 (2012).
90. You, L., Garwicz, D. & Rögnvaldsson, T. Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease. *J. Virol.* **79**, 12477–12486 (2005).
91. Mirambeau, G., Lyonnais, S. & Gorelick, R. J. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol.* **7**, 724–734 (2010).
92. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144 (2013).
93. Darlix, J.-L. *et al.* Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **410**, 565–581 (2011).
94. Solbak, S. M. Ø. *et al.* HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. *Biochim. Biophys. Acta* **1828**, 816–823 (2013).
95. Hill, M., Tachedjian, G. & Mak, J. The packaging and maturation of the HIV-1 Pol proteins. *Curr. HIV Res.* **3**, 73–85 (2005).
96. Wu, X. *et al.* Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex. *J. Virol.* **73**, 2126–2135 (1999).

97. Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* **276**, 23213–23216 (2001).
98. Ruckenstein, E. & Manciu, M. On the Stability of Lyotropic Lamellar Liquid Crystals and the Thicknesses of Their Lamellae. *Langmuir* **17**, 5464–5475 (2001).
99. Schwartz, S., Felber, B. K., Fenyö, E. M. & Pavlakis, G. N. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64**, 5448–5456 (1990).
100. Checkley, M. A., Luttge, B. G. & Freed, E. O. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* **410**, 582–608 (2011).
101. Raska, M. *et al.* Glycosylation patterns of HIV-1 gp120 depend on the type of expressing cells and affect antibody recognition. *J. Biol. Chem.* **285**, 20860–20869 (2010).
102. Voeten, J. T. *et al.* Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J. Virol.* **74**, 6800–6807 (2000).
103. Vanhamme, L., Pays, E., McCulloch, R. & Barry, J. D. An update on antigenic variation in African trypanosomes. *Trends Parasitol.* **17**, 338–343 (2001).
104. Pancera, M. *et al.* Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1166–1171 (2010).
105. Modrow, S. *et al.* Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. *J. Virol.* **61**, 570–578 (1987).
106. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
107. RCSB Protein Data Bank - RCSB PDB. Available at: <http://www.rcsb.org/pdb/home/home.do>. (Accessed: 7th April 2016)
108. Murakami, T. & Freed, E. O. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 343–348 (2000).
109. Groppelli, E., Len, A. C., Granger, L. A. & Jolly, C. Retromer regulates HIV-1 envelope glycoprotein trafficking and incorporation into virions. *PLoS Pathog.* **10**, e1004518 (2014).
110. Bu, Z. *et al.* Enhancement of immunogenicity of an HIV Env DNA vaccine by mutation of the Tyr-based endocytosis motif in the cytoplasmic domain. *Virology* **328**, 62–73 (2004).
111. Klinger, S. C., Siupka, P. & Nielsen, M. S. Retromer-Mediated Trafficking of Transmembrane Receptors and Transporters. *Membranes* **5**, 288–306 (2015).
112. Zhu, P. *et al.* Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15812–15817 (2003).

113. Brandenberg, O. F., Magnus, C., Rusert, P., Regoes, R. R. & Trkola, A. Different infectivity of HIV-1 strains is linked to number of envelope trimers required for entry. *PLoS Pathog.* **11**, e1004595 (2015).
114. Brandenberg, O. F., Magnus, C., Regoes, R. R. & Trkola, A. The HIV-1 Entry Process: A Stoichiometric View. *Trends Microbiol.* **23**, 763–774 (2015).
115. Landau, N. R., Warton, M. & Littman, D. R. The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4. *Nature* **334**, 159–162 (1988).
116. Li, Y., Yin, Y. & Mariuzza, R. A. Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. *Front. Immunol.* **4**, 206 (2013).
117. Reeves, J. D. *et al.* Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16249–16254 (2002).
118. Platt, E. J., Durnin, J. P. & Kabat, D. Kinetic factors control efficiencies of cell entry, efficacies of entry inhibitors, and mechanisms of adaptation of human immunodeficiency virus. *J. Virol.* **79**, 4347–4356 (2005).
119. Munro, J. B. & Mothes, W. Structure and Dynamics of the Native HIV-1 Env Trimer. *J. Virol.* **89**, 5752–5755 (2015).
120. Deng, H. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661–666 (1996).
121. Dragic, T. *et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* **381**, 667–673 (1996).
122. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
123. Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S. & Landau, N. R. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *J. Exp. Med.* **185**, 621–628 (1997).
124. Tersmette, M. *et al.* Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J. Virol.* **62**, 2026–2032 (1988).
125. Schuitemaker, H. *et al.* Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. *J. Virol.* **66**, 1354–1360 (1992).
126. Esté, J. A. *et al.* Shift of clinical human immunodeficiency virus type 1 isolates from X4 to R5 and prevention of emergence of the syncytium-inducing phenotype by blockade of CXCR4. *J. Virol.* **73**, 5577–5585 (1999).
127. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7557 (2008).

128. Berger, E. A., Murphy, P. M. & Farber, J. M. CHEMOKINE RECEPTORS AS HIV-1 CORECEPTORS: Roles in Viral Entry, Tropism, and Disease. *Annu. Rev. Immunol.* **17**, 657–700 (1999).
129. Gorry, P. R. & Ancuta, P. Coreceptors and HIV-1 Pathogenesis. *Curr. HIV/AIDS Rep.* **8**, 45–53 (2010).
130. Farzan, M. *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* **96**, 667–676 (1999).
131. Huang, C.-C. *et al.* Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317**, 1930–1934 (2007).
132. Lee, B. *et al.* Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J. Biol. Chem.* **274**, 9617–9626 (1999).
133. Gallo, S. A. *et al.* The HIV Env-mediated fusion reaction. *Biochim. Biophys. Acta* **1614**, 36–50 (2003).
134. Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250**, 1139–1142 (1990).
135. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
136. Kesavardhana, S. & Varadarajan, R. Stabilizing the native trimer of HIV-1 Env by destabilizing the heterodimeric interface of the gp41 postfusion six-helix bundle. *J. Virol.* **88**, 9590–9604 (2014).
137. Kozlovsky, Y. & Kozlov, M. M. Stalk model of membrane fusion: solution of energy crisis. *Biophys. J.* **82**, 882–895 (2002).
138. Harrison, S. C. Viral membrane fusion. *Virology* **479-480**, 498–507 (2015).
139. Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V. & Melikyan, G. B. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* **137**, 433–444 (2009).
140. Melikyan, G. B. HIV entry: a game of hide-and-fuse? *Curr. Opin. Virol.* **4**, 1–7 (2014).
141. Stein, B. S. *et al.* pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* **49**, 659–668 (1987).
142. White, J. M. Membrane fusion. *Science* **258**, 917–924 (1992).
143. Hernandez, L. D., Hoffman, L. R., Wolfsberg, T. G. & White, J. M. Virus-cell and cell-cell fusion. *Annu. Rev. Cell Dev. Biol.* **12**, 627–661 (1996).
144. Hulme, A. E., Perez, O. & Hope, T. J. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9975–9980 (2011).
145. Lukic, Z., Dharan, A., Fricke, T., Diaz-Griffero, F. & Campbell, E. M. HIV-1 uncoating is facilitated by dynein and kinesin 1. *J. Virol.* **88**, 13613–13625 (2014).

146. Swart, M., Solà, M. & Bickelhaupt, F. M. Energy landscapes of nucleophilic substitution reactions: a comparison of density functional theory and coupled cluster methods. *J. Comput. Chem.* **28**, 1551–1560 (2007).
147. Guo, F., Cen, S., Niu, M., Javanbakht, H. & Kleiman, L. Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1. *J. Virol.* **77**, 9817–9822 (2003).
148. Kleiman, L. tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life* **53**, 107–114 (2002).
149. Sleiman, D., Barraud, P., Brachet, F. & Tisne, C. The Interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. *PLoS One* **8**, e64700 (2013).
150. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. & Marquet, R. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.* **247**, 236–250 (1995).
151. Hansen, J., Schulze, T., Mellert, W. & Moelling, K. Identification and characterization of HIV-specific RNase H by monoclonal antibody. *EMBO J.* **7**, 239–243 (1988).
152. Chen, Y., Balakrishnan, M., Roques, B. P., Fay, P. J. & Bambara, R. A. Mechanism of minus strand strong stop transfer in HIV-1 reverse transcription. *J. Biol. Chem.* **278**, 8006–8017 (2003).
153. Galetto, R. & Negroni, M. Mechanistic features of recombination in HIV. *AIDS Rev.* **7**, 92–102 (2005).
154. Yi-Brunozzi, H. Y. *et al.* High-resolution NMR analysis of the conformations of native and base analog substituted retroviral and LTR-retrotransposon PPT primers. *Chem. Biol.* **15**, 254–262 (2008).
155. Onafuwa-Nuga, A. & Telesnitsky, A. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev. MMBR* **73**, 451–480, Table of Contents (2009).
156. Warren, K., Warrilow, D., Meredith, L. & Harrich, D. Reverse Transcriptase and Cellular Factors: Regulators of HIV-1 Reverse Transcription. *Viruses* **1**, 873–894 (2009).
157. Warren, K. *et al.* Eukaryotic elongation factor 1 complex subunits are critical HIV-1 reverse transcription cofactors. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9587–9592 (2012).
158. Gabuzda, D. H. *et al.* Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. *J. Virol.* **66**, 6489–6495 (1992).
159. Bardy, M. *et al.* Interaction of human immunodeficiency virus type 1 Vif with Gag and Gag-Pol precursors: co-encapsidation and interference with viral protease-mediated Gag processing. *J. Gen. Virol.* **82**, 2719–2733 (2001).
160. Dettenhofer, M. & Yu, X. F. Highly purified human immunodeficiency virus type 1 reveals a virtual absence of Vif in virions. *J. Virol.* **73**, 1460–1467 (1999).

161. Sova, P., Volsky, D. J., Wang, L. & Chao, W. Vif is largely absent from human immunodeficiency virus type 1 mature virions and associates mainly with viral particles containing unprocessed gag. *J. Virol.* **75**, 5504–5517 (2001).
162. Madani, N. & Kabat, D. An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J. Virol.* **72**, 10251–10255 (1998).
163. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
164. Holmes, R. K., Koning, F. A., Bishop, K. N. & Malim, M. H. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. *J. Biol. Chem.* **282**, 2587–2595 (2007).
165. Yu, Q. *et al.* Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat. Struct. Mol. Biol.* **11**, 435–442 (2004).
166. Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **4**, 868–877 (2004).
167. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **9**, 1404–1407 (2003).
168. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat. Med.* **9**, 1398–1403 (2003).
169. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056–1060 (2003).
170. Ambrose, Z. & Aiken, C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology* **454-455**, 371–379 (2014).
171. Xu, H. *et al.* Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. *Retrovirology* **10**, 70 (2013).
172. Lee, K. *et al.* Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7**, 221–233 (2010).
173. Peng, K. *et al.* Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife* **3**, e04114 (2014).
174. Pawlica, P. & Berthou, L. Cytoplasmic dynein promotes HIV-1 uncoating. *Viruses* **6**, 4195–4211 (2014).
175. Chen, H. & Engelman, A. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15270–15274 (1998).
176. Levin, A., Loyter, A. & Bukrinsky, M. Strategies to inhibit viral protein nuclear import: HIV-1 as a target. *Biochim. Biophys. Acta* **1813**, 1646–1653 (2011).
177. Zennou, V. *et al.* HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173–185 (2000).

178. Suzuki, Y. & Craigie, R. The road to chromatin - nuclear entry of retroviruses. *Nat. Rev. Microbiol.* **5**, 187–196 (2007).
179. Matreyek, K. A. & Engelman, A. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses* **5**, 2483–2511 (2013).
180. Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11288–11294 (1996).
181. Maertens, G. *et al.* LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* **278**, 33528–33539 (2003).
182. Di Nunzio, F. New insights in the role of nucleoporins: a bridge leading to concerted steps from HIV-1 nuclear entry until integration. *Virus Res.* **178**, 187–196 (2013).
183. Grandgenett, D. P., Pandey, K. K., Bera, S. & Aihara, H. Multifunctional facets of retrovirus integrase. *World J. Biol. Chem.* **6**, 83–94 (2015).
184. Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168–1171 (1988).
185. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988–25997 (1992).
186. Loeb, L. A. & Monnat, R. J. DNA polymerases and human disease. *Nat. Rev. Genet.* **9**, 594–604 (2008).
187. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996).
188. Sadler, H. A., Stenglein, M. D., Harris, R. S. & Mansky, L. M. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* **84**, 7396–7404 (2010).
189. Geller, R. *et al.* The external domains of the HIV-1 envelope are a mutational cold spot. *Nat. Commun.* **6**, 8571 (2015).
190. Smyth, R. P., Davenport, M. P. & Mak, J. The origin of genetic diversity in HIV-1. *Virus Res.* **169**, 415–429 (2012).
191. Archer, J. *et al.* Identifying the important HIV-1 recombination breakpoints. *PLoS Comput. Biol.* **4**, e1000178 (2008).
192. Simon-Loriere, E., Martin, D. P., Weeks, K. M. & Negroni, M. RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J. Virol.* **84**, 12675–12682 (2010).
193. Galetto, R. *et al.* The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J. Biol. Chem.* **279**, 36625–36632 (2004).
194. Simon-Loriere, E., Rossolillo, P. & Negroni, M. RNA structures, genomic organization and selection of recombinant HIV. *RNA Biol.* **8**, 280–286 (2011).

195. Simon-Loriere, E. *et al.* Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus. *PLoS Pathog* **5**, e1000418 (2009).
196. Goodsell, D. S. The molecular perspective: ultraviolet light and pyrimidine dimers. *The Oncologist* **6**, 298–299 (2001).
197. Löwdin, P.-O. Proton Tunneling in DNA and its Biological Implications. *Rev. Mod. Phys.* **35**, 724–732 (1963).
198. Burton, D. R., Stanfield, R. L. & Wilson, I. A. Antibody vs. HIV in a clash of evolutionary titans. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14943–14948 (2005).
199. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336–343 (2011).
200. Pancera, M. *et al.* Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. *Nat. Struct. Mol. Biol.* **20**, 804–813 (2013).
201. Scharf, L. *et al.* Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep.* **7**, 785–795 (2014).
202. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
203. Crawford, H. *et al.* Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B\*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J. Virol.* **81**, 8346–8351 (2007).
204. Bunnik, E. M., Pisas, L., van Nuenen, A. C. & Schuitemaker, H. Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J. Virol.* **82**, 7932–7941 (2008).
205. Brookes, D. W., Miah, S., Lackenby, A., Hartgroves, L. & Barclay, W. S. Pandemic H1N1 2009 influenza virus with the H275Y oseltamivir resistance neuraminidase mutation shows a small compromise in enzyme activity and viral fitness. *J. Antimicrob. Chemother.* **66**, 466–470 (2011).
206. Sather, D. N. *et al.* Broadly neutralizing antibodies developed by an HIV-positive elite neutralizer exact a replication fitness cost on the contemporaneous virus. *J. Virol.* **86**, 12676–12685 (2012).
207. Lynch, R. M. *et al.* HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* **89**, 4201–4213 (2015).
208. Li, L. *et al.* A broad range of mutations in HIV-1 neutralizing human monoclonal antibodies specific for V2, V3, and the CD4 binding site. *Mol. Immunol.* **66**, 364–374 (2015).
209. Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E1293–1301 (2011).
210. Jones, D. T., Singh, T., Kosciółek, T. & Tetchner, S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinforma. Oxf. Engl.* **31**, 999–1006 (2015).

211. Codoñer, F. M., O’Dea, S. & Fares, M. A. Reducing the false positive rate in the non-parametric analysis of molecular coevolution. *BMC Evol. Biol.* **8**, 106 (2008).
212. Fares, M. A. & Travers, S. A. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* **173**, 9–23 (2006).
213. Travers, S. A. A., Tully, D. C., McCormack, G. P. & Fares, M. A. A study of the coevolutionary patterns operating within the env gene of the HIV-1 group M subtypes. *Mol. Biol. Evol.* **24**, 2787–2801 (2007).
214. Mann, J. K. & Ndung’u, T. HIV-1 vaccine immunogen design strategies. *Viol. J.* **12**, 3 (2015).
215. Guttman, M. *et al.* CD4-induced activation in a soluble HIV-1 Env trimer. *Struct. Lond. Engl. 1993* **22**, 974–984 (2014).
216. Binley, J. M. *et al.* A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* **74**, 627–643 (2000).
217. Sanders, R. W. *et al.* Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76**, 8875–8889 (2002).
218. Binley, J. M. *et al.* Enhancing the proteolytic maturation of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* **76**, 2606–2616 (2002).
219. Klasse, P. J. *et al.* Influences on trimerization and aggregation of soluble, cleaved HIV-1 SOSIP envelope glycoprotein. *J. Virol.* **87**, 9873–9885 (2013).
220. Khayat, R. *et al.* Structural characterization of cleaved, soluble HIV-1 envelope glycoprotein trimers. *J. Virol.* **87**, 9865–9872 (2013).
221. Alsaifi, N., Debbiche, O., Sodroski, J. & Finzi, A. Effects of the I559P gp41 change on the conformation and function of the human immunodeficiency virus (HIV-1) membrane envelope glycoprotein trimer. *PLoS One* **10**, e0122111 (2015).
222. Julien, J.-P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, (2013).
223. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 Env trimer. *Science* **342**, 1484–1490 (2013).
224. Julien, J.-P. *et al.* Design and structure of two HIV-1 clade C SOSIP.664 trimers that increase the arsenal of native-like Env immunogens. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11947–11952 (2015).
225. Tobin, G. J. *et al.* Deceptive imprinting and immune refocusing in vaccine design. *Vaccine* **26**, 6189–6199 (2008).
226. Sadanand, S., Suscovich, T. J. & Alter, G. Broadly Neutralizing Antibodies Against HIV: New Insights to Inform Vaccine Design. *Annu. Rev. Med.* (2015). doi:10.1146/annurev-med-091014-090749

227. Mao, Y. *et al.* Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. *Nat. Struct. Mol. Biol.* **19**, 893–899 (2012).
228. Mao, Y. *et al.* Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12438–12443 (2013).
229. Cao, J. *et al.* Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* **71**, 9808–9812 (1997).
230. Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog.* **10**, e1004529 (2014).
231. Aït-Bara, S., Carpousis, A. J. & Quentin, Y. RNase E in the  $\gamma$ -Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains. *Mol. Genet. Genomics MGG* **290**, 847–862 (2015).
232. Lauber, C., Kazem, S., Kravchenko, A. A., Feltkamp, M. C. W. & Gorbalenya, A. E. Interspecific adaptation by binary choice at de novo polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region. *Nucleic Acids Res.* **43**, 4800–4813 (2015).
233. Yuan, J. & Xue, B. Role of structural flexibility in the evolution of emerin. *J. Theor. Biol.* **385**, 102–111 (2015).
234. Guttman, M., Weis, D. D., Engen, J. R. & Lee, K. K. Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. *J. Am. Soc. Mass Spectrom.* **24**, 1906–1912 (2013).
235. Guttman, M. *et al.* Antibody potency relates to the ability to recognize the closed, pre-fusion form of HIV Env. *Nat. Commun.* **6**, 6144 (2015).
236. Do Kwon, Y. *et al.* Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol.* **22**, 522–531 (2015).
237. Karasavvas, N. *et al.* The Thai Phase III HIV Type 1 Vaccine trial (RV144) regimen induces antibodies that target conserved regions within the V2 loop of gp120. *AIDS Res. Hum. Retroviruses* **28**, 1444–1457 (2012).
238. Esparza, J. A brief history of the global effort to develop a preventive HIV vaccine. *Vaccine* **31**, 3502–3518 (2013).
239. Wang, H.-B., Mo, Q.-H. & Yang, Z. HIV vaccine research: the challenge and the way forward. *J. Immunol. Res.* **2015**, 503978 (2015).
240. Goh, C.-S., Milburn, D. & Gerstein, M. Conformational changes associated with protein-protein interactions. *Curr. Opin. Struct. Biol.* **14**, 104–109 (2004).





## Comprendre la flexibilité génétique de la protéine d'enveloppe de VIH-1 à travers l'étude du réseau de coévolution de ses acides aminés

### Résumé

Une des caractéristiques du Virus de l'Immunodéficience Humaine de type 1 (VIH-1) est sa diversification génétique extensive, qui lui permet d'échapper au système immunitaire. Néanmoins, il est nécessaire que le taux de mutation requis pour à cette évolution rapide ne compromette pas la fonctionnalité de ses protéines.

Les travaux présentés ici ont eu pour objectif l'étude des réseaux de coévolution qui composent les glycoprotéines d'enveloppe (Env) afin de comprendre les règles qui sous-tendent leur évolution. Il a été mis en évidence que les régions variables de ces protéines, grâce à leur flexibilité structurale, peuvent aussi servir à faciliter l'incorporation de mutations touchant les régions plus constantes. De plus, un réseau de coévolution impliqué dans les changements de conformations nécessaires à l'activité de Env a été identifié, soutenant le fait que ces régions variables ont un rôle central dans ces changements.

Ces études démontrent le rôle crucial joué par les régions variables en dévoilant un nouvel aspect de leur contribution à l'évolution du VIH-1.

**Mots-clés** : VIH-1, glycoprotéine d'enveloppe, réseau de coévolution, fonctionnalité

### Abstract

The Human Immunodeficiency Virus type 1 (HIV-1) is characterized by an extensive genetic diversification of its strains that allows the virus to escape the immune system. However, the mutation rate needed for this rapid evolution must not compromise the functionality of the viral proteins.

The aim of the work presented here has been to study the coevolution networks that constitute the envelope glycoproteins (Env) in order to understand the rules driving their evolution. The results have highlighted that variable regions, thanks to their structural freedom, can facilitate the incorporation of mutations in more constant regions. Moreover, a coevolution network involved in the conformational changes required for the activity of Env has been identified, underlining the central role played by variable regions in these processes.

Besides underscoring the crucial role played by variable regions in the functionality of Env, these studies unveil a new aspect of their contribution to HIV-1 evolution.

**Keywords**: HIV-1, envelope glycoprotein, coevolution network, functionality