HAL
open science

Immunité innée et contre-mesures virales : études structurales et biophysiques du complexe formé entre la cytidine désaminase humaine APOBEC3G et le facteur d'infectivité virale du VIH-1

Joelle Mezher

To cite this version:

Joelle Mezher. Immunité innée et contre-mesures virales : études structurales et biophysiques du complexe formé entre la cytidine désaminase humaine APOBEC3G et le facteur d'infectivité virale du VIH-1. Immunité innée. Université de Strasbourg, 2015. Français. NNT : 2015STRAJ053 . tel-01374420

HAL Id: tel-01374420

https://theses.hal.science/tel-01374420v1

Submitted on 30 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE Sciences de la vie et santé

Architecture et Réactivité de l'ARN

THÈSE présentée par :

Joelle Mezher

soutenue le : **29 Septembre 2015**

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

Discipline/ Spécialité : Sciences du vivant

Mention : Aspects moléculaires et cellulaires de la biologie

IMMUNITE INNEE ET CONTRE-MESURES VIRALES : ETUDES STRUCTURALES ET BIOPHYSIQUES DU COMPLEXE FORME ENTRE LA CYTIDINE DESAMINASE HUMAINE APOBEC3G ET LE FACTEUR D'INFECTIVITE VIRALE DU VIH-1

THÈSE dirigée par :

Dr. ENNIFAR Eric	Docteur, université de Strasbourg

RAPPORTEURS :

Dr. FRIBOURG Sébastien	Docteur, INSERM U869, IECB Bordeaux
Dr. TISNE Carine	Docteur, UMR8015 CNRS, université Paris-Descartes

---

AUTRES MEMBRES DU JURY :

Dr. RUFF Marc	Docteur, IGBMC, université de Strasbourg

" On vous a dit aussi que la vie est obscurité, et dans votre lassitude vous répétez ce que disent les las.

Et je vous dis que la vie est en effet obscure sauf là où il y a élan,

Et tout élan est aveugle sauf là où il y a la connaissance.

Et toute connaissance est vaine sauf là où il y a le travail,

Et tout travail est futile sauf là où il y a l'amour ;

Et quand vous travaillez avec amour vous attachez votre être à votre être, et vous aux autres, et vous à Dieu "

Le prophète, Khalil Gibran

À la mémoire de mon Grand-Père,
À mes parents,
À mes frères,
À Josette et Bébé Charbel
À tous les miens,

"Sois reconnaissant envers tous,
Tous t'enseignent "
Bouddha

Merci !

En premier lieu, je souhaite remercier les membres de mon jury qui m'ont fait l'honneur d'accepter de juger ce travail : Sébastien Fribourg, Karine Tisné, Marc Ruff et Christophe Romier.

En deuxième lieu j'aimerai remercier Sidaction et le fond de dotation de Pierre Berger pour le financement de ces quatre années de thèse.

Un grand merci à Eric Westhoff pour m'avoir accueilli au sein de l'UPR9002 et merci à tous les membres de cette unité qui ont rendu les longues journées et nuits passées au laboratoire plaisantes par leurs sourires et aides continuelles. Je n'oublierai jamais nos bbq , nos soirées et les pauses thés et cafés.

Je tiens à exprimer tous mes sincères remerciements à Philippe Dumas pour m'avoir fait une place dans son équipe et pour son soutien et ses conseils. Merci de m'avoir passé cette passion pour la recherche et la biologie structurale.

Je remercie toute l'équipe de biophysique et biologie structurale de m'avoir accueilli et supporté: Philippe Walter, Dominique Burnouf, Guillaume Bec, Philippe Wolff, Marie, cyrielle, Benoit, Sondes. Votre humour, disponibilité, conseil et savoir scientifique ont rendu cette expérience très enrichissante. Ces années passées ensemble et nos moments partagés resteront à toujours gravés dans ma mémoire et mon cœur. Merci à toi Séverine pour avoir marqué mes premier pas dans la recherche et de m'avoir appris à cloner, exprimer et purifier puis cristalliser comme on respire☺.

Je tiens également à remercier Rudy et Emma, les deux personnes qui ont participé à ce travail et que j'ai eu le bonheur de les encadrer pendant leur stages. Un merci particulier à toi Emma, ta motivation et ton enthousiasme pour la recherche m'ont ressourcé au moment où je commençais à perdre cette passion. Je te souhaite une thèse très réussie sur le ribosome, malgré que j'aie bien aimé si tu prenais le relais sur le projet APOBEC/Vif que tu as commencé durant ton stage, mais je suis sure que tu vas mener ton futur projet vers beaucoup de succès et de publications.

Comment ne pas te remercie toi Eric ? Je n'oublierai jamais la confiance que tu m'as accordée depuis le premier jour de mon stage. Merci pour ton soutien et encouragements quotidiens, ton questionnement constructif, la cohérence de tes suggestions et ton savoir scientifique très riche. Outre tout ça, merci pour ta bonne humeur et ta disponibilité en tout moment pour m'aider et répondre à mes questions. Merci aussi de m'avoir transmis l'intérêt pour la cristallographie. Cela a réellement été un grand plaisir de travailler avec toi !

Merci à nos collaborateurs qui m'ont permis de réaliser ce travail, Christophe Romier et Marc Ruff. Merci pour votre accueil dans vos laboratoires et au sein de vos équipes et pour votre confiance dans ce projet qui présentait malgré tout beaucoup de difficultés. Merci à vous pour toutes ces co-expressions !!
Un grand merci aussi à Yves Nominé, qui m'a accompagné dès le début de ma thèse. Nos longues discussions et tes conseils m'ont permis d'élargir mes champs de recherche et de surmonter les problèmes de purification.

Un grand merci à une personne qui a joué un grand rôle dans cette thèse, Robert Drillien. Je tiens dans mon cœur un merci très spécial pour toi et des souvenirs inoubliables de ces mois passés à tes cotés à l'IGBMC. J'ai eu l'honneur de travailler avec toi, d'apprendre de ton savoir scientifique remarquable. Tu m'as enrichi tant scientifiquement que moralement et tu m'as passé ta passion à la virologie. Mon expérience que j'ai acquise à l'IGBMC a été sans doute l'une des plus belles et plus utiles de ma vie. Merci pour ton aide et surtout pour l'inquiétude et l'enthousiasme que tu avais pour ce projet.

Merci à tous les membres de la plateforme de protéomique de l'IBMC : Philippe wolff, Philippe Hamann, Lauriane Kuhn, Joanna chicher. Je vous ai beaucoup embêté durant ces années afin d'identifier ce complexe et vous étiez toujours présents pour m'aider et répondre à mes questions. A part notre relation professionnelle, j'ai gagné des amis aussi et ça me fait un grand plaisir.

Merci à tous ceux qui ont fait de cette période de ma vie une belle période par leur sourire et leur amitié. Franzi , je ne sais pas ce que j'aurai fais sans toi au labo, merci pour ton amitié et pour nos pauses cafés , on s'est partagé nos joies et nos soucis, on a rigolé et pleuré , regardé beaucoup de films à l'UGC…merci parce que tu tiens à tes amitiés.
Un grand merci à toi Marianne, ma meilleur amie, merci parce que tu m'as supporté tout au long de ces années et pour ton côté maternel. Tu étais à mes côtés et toujours à l'écoute et à l'aide à n'importe quel moment. Nos pauses déjeuners quotidiens pendant les cinq années me tiennent beaucoup à cœur.

Merci également à tous mes amis, Rachelle, Omar, Charbel, Ola, Joelle, Zeinab, Mira, Patrick, Ranine , Bechara, Farouk, Hala…Sans vous mes amis , la vie n'aurait pas le même goût.

Finalement, le plus grand merci est à ma famille. Papa et maman, sans vous je ne serai pas quoi faire, votre présence dans ma vie est ma raison de vivre et d'aller encore plus loin dans mes ambitions. Merci pour vos sacrifices, votre amour éternel et inconditionnel, votre confiance et vos encouragements. Merci également à mes trois frères Joseph, Johnny et Julien, vous êtes pour moi toute la sécurité et le bonheur du monde qui s'est complété par toi Josette, la sœur que je n'ai pas eu et par mon neveu Charbel, la joie de notre famille.

Merci Dieu pour chaque pas, chaque décision et chaque personne que j'ai remerciée ci-dessus. Aucun de tout ça n'aurait été possible si ce n'était pas par votre bénédiction.

SOMMAIRE

SOMMAIRE

ABREVIATIONS 1

INTRODUCTION 1

I. Les Virus 1

A. Définitions 1

B. Les rétrovirus 2

1. Généralités 2

2. Classification 3

3. Les lentivirus 4

II. Le Virus de l'Immunodéficience Humaine de type 1 5

A. Infection et VIH 5

1. Découverte 5

2. Origine et variabilité du VIH-1 5

3. Mode de transmission 7

4. Organisation génétique du VIH-1 7

a. Régions non codantes 7

b. Structure du virion 9

5. Evolution de la maladie 10

6. La pandémie de SIDA 11

B. Le cycle réplicatif du VIH-1 12

1. La phase pré-intégrative 12

a. Entrée virale : Reconnaissance et entrée 12

b. La rétrotranscription et la décapsidation 14

c. L'intégration 16

2. La phase post-intégrative 18

a. L'expression génique 18

b. La traduction 20

c. Assemblage 21

d. Bourgeonnement et maturation (Fig. 17) 22

C. Les traitements antirétroviraux 23

D. Stratégie vaccinale 24

III. Immunité innée et facteurs de restriction virale 25

A. Les facteurs de restrictions virales 26

1. TRIM5α 26

a. Structure et fonction biologique 26

b. Action antirvirale de TRIM5α 27
2. Téthérine 28
a. Action antivirale de la téthérine 29
b. Vpu contrecarre téthérine 30
3. SAMHD1 31
a. Généralités 31
b. Action antivirale de SAMHD1 32
c. Vpx contrecarre SAMHD1 32
4. APOBEC3G et contre mesure virale 33
a. APOBEC3: Structure et fonction 33
b. Rôle d'APOBEC3G/F dans la restriction virale 36
c. Rôle de la famille des APOBEC3 dans d'autres rétrovirus 38
d. La protéine virale Vif contrecarre APOBEC 39
e. Rôle d'APOBEC3G/Vif dans le traitement antirétroviral 44

IV. Système d'expression par le virus de la vaccine *47*

A. Définition du virus de la vaccine 47
B. La réplication du virus de la vaccine 47
C. Application dans l'expression des protéines 49

OBJECTIFS DU TRAVAIL DE THESE 51

MATERIELS ET METHODES 53

I. Biologie moléculaire-Biochimie *53*

A. Techniques de biologie moléculaire 53
1. Cellules et virus 53
a. Système bactérien 53
b. Système mammifère 54
2. Amplification de l'ADN 56
a. PCR 56
b. Clonage par restriction/ligation 57
c. Transformation bactérienne 57
3. Générations des vecteurs d'expression bactériens 58
a. Clonage en vecteurs d'expression bactériens 58
4. Optimisation des codons 61
5. Harmonisation des codons (Annexe) 62
6. Génération de vecteurs d'expression en cellules de mammifère 63
a. Les clonages Biobrick 63
b. Clonage LR Gateway® 67

B. BHK21/ système d'expression par le virus de la vaccine 71
1. Principe et stratégie 71
a. Recombinaison homologue : 71
b. Isolement et purification des virus recombinants 73
2. Protocole 75
a. Transfection de cellules BHK21 infectées par le virus 75
b. Sélection des virus recombinants 76
c. Amplification des virus recombinants 76
C. Techniques de Biochimie 76
1. Tests d'expression 76
a. Principe de l'expression induite par l'IPTG 77
2. Expression à grande échelle 78
a. Expression en bactéries 78
b. Expression en BHK21 78
3. Les différentes stratégies de formation de complexe 79
a. La formation de complexe par co-expression 79
b. La formation de complexe par co-broyage 80
c. La formation du complexe *in vitro* 80
4. Purification des protéines et du complexe 80
a. La lyse des cellules 81
b. Chromatographie par affinité 81
c. Chromatographie d'exclusion de taille 85
5. Purification des acides nucléiques 85

II. Caractérisation physico-chimique et fonctionnelle des protéines 86

A. Analyse SDS-PAGE 86
B. La diffusion dynamique de lumière (DLS) 87
C. Dichroïsme circulaire 87
D. Western Blot 88
E. Spectrométrie de masse 89
1. Analyse des bandes de gel en MALDI-TOF 89
2. Analyse en solution en LC-MS/MS 90
a. En conditions dénaturantes 90
b. En conditions natives 90
F. Méthylation des Lysines 90
G. Microcalorimétrie isotherme (ITC) 91
H. Test d'activité d'APOBEC 93
I. Tests de cristallisation 94

SOMMAIRE

III. Criblage haut-débit « ESPRIT » *95*

RESULTATS PARTIE 1 : Identification des déterminants structuraux et de séquence impliqués dans l'interaction entre le facteur d'infectivité virale VIF et les ARN TAR et DIS du VIH-1 ..... **98**

Publication 1: Sequence and Structure Requirements for Specific Recognition of HIV- TAR and DIS RNA by the HIV-1 Vif Protein **100**

RESULTATS PARTIE 2 : Reconstitution du complexe Vif/APOBEC3G en cellules bactériennes E. coli **101**

I. Introduction et objectifs *101*

II. Stratégie *102*

III. Résultats *103*

A. APOBEC3G/F **103**

1. Caractérisation de APOBEC3G-CD2 103
   a. Expression et purification de hA3G-CD2 103
   b. Caractérisation biophysique de hA3G-CD2 105
   c. Test d'activité de hA3G-CD2 106
   d. Caractérisation des interactions Vif/CD2-A3G 107
2. Caractérisation de APOBEC3F-CD2 108
   a. Expression et purification de hA3F-CD2chimère 108
   b. Test d'activité de hA3F-CD2c 109
3. Recherche de constructions solubles d'APOBEC3 par criblage haut débit 110
4. Expression d'APOBEC dans les cellules ArcticExpress 112

B. Vif au sein du complexe de la E3 ubiquitine ligase **114**

1. Reconstitution du complexe *in vitro* 114
   a. Expression et purification de CBFβ 114
   b. Expression et purification du complexe EloB-EloC 119
   c. Expression et purification du complexe Cul5-Rbx2 121
   d. Expression et purification du complexe Vif-CBFβ 123
2. Reconstitution du complexe *in vivo* 127
   a. Optimisation des conditions 127
   b. Co-expression sur deux plasmides différents 128
   c. Co-expression sur un seul plasmide 132
   d. Co-expression Vif-CBFβ-EloB-EloC-Cul5-Rbx2 133
   e. Co-expression Vif-CBFβ-EloB-EloC-APOBEC3G/F 134
3. Tests de cristallisation 138

SOMMAIRE

IV. Discussion, conclusions et perspectives *140*

RESULTATS PARTIE 3 : Reconstitution du complexe Vif/APOBEC3G en cellules de mammifères **143**

I. Introduction et objectif *143*

II. Stratégie *144*

III. Résultats *145*

A. Clonage des gènes d'intérêt dans les vecteurs d'expression **145**

1. Complexe CBFβ – Vif – EloB – EloC – A3G 145
2. La protéine APOBEC3G 147

B. Sélection des virus recombinants **147**

C. Expression des protéines **149**

1. Marqueur d'expression YFP 149
2. Tests d'expression et purification 150

D. Expression et purification à grande échelle **153**

1. Tampons 153
2. Purification du complexe CBFβ – Vif – EloB – EloC – A3G 154
   a. Colonne d'affinité au Nickel 154
   b. Colonne d'affinité Streptactin® 156
3. Purification d'A3G seule 156
   a. Purification de Flag/10His A3G 156
   b. Purification de Strep-A3G 158

E. Validation par spectrométrie de masse **158**

1. Validation du système de clivage à la TEV 159
2. Identification d'A3G et du complexe 160

F. Activité d'APOBEC3G **163**

IV. Conclusions, discussion et perspectives *165*

CONCLUSION GÉNÉRALE **167**

RÉFÉRENCES BIBILOGRAPHIQUES **169**

ANNEXES **186**

INDEX DES TABLEAUX ET FIGURES

Tableaux

Tableau 1: Rôles des protéines virales. 9
Tableau 2: Liste des antirétroviraux approuvés par la FDA..... 24
Tableau 3: Les différents facteurs de restriction contre le VIH-1. 33
Tableau 4: Action des protéines de la famille APOBEC contre les rétroéléments. 38
Tableau 5: Les motifs d'interaction de la protéine Vif dans les différents APOBEC3. 43
Tableau 6: Composition du milieu LB. 54
Tableau 7: Les différentes séquences intégrées dans le virus de la vaccine par recombinaison homologue. 56
Tableau 8: Composition du mélange réactionnel pour la réaction de PCR..... 57
Tableau 9: Différentes caractéristiques des vecteurs utilisés pour l'expression et la co-expression des protéines Vif/CBFβ/EloB/EloC/APOBEC. 59
Tableau 10: Les gènes utilisés pour les tests d'expression/co-expression, originaux, optimisés ou harmonisés. 63
Tableau 11: Les sites de restrictions utilisés pour les clonages BioBrick en 5' et 3' de l'ADN. 64
Tableau 12: Les amorces en Forward et Reverse Utilisées pour le clonage dans pUC19 65
Tableau 13: les différents plasmides pUC_BB utilisés dans ce projet et les gènes insérés respectivement ainsi que leurs étiquettes de purification.. 66
Tableau 14: Composition du mélange réactionnel de la réaction LR. 69
Incuber la réaction pour une heure à 25°C puis rajouter 1 µl de la protéinase K pour arrêter la réaction. Incuber 10 minutes à 37°C puis transformer dans des cellules compétentes. 69
Tableau 15: Les différentes étiquettes utilisées pour la purification des protéines et leurs caractéristiques. 82
Tableau 16 : Les différentes compositions de tampons utilisés durant les étapes de purification du domaine C-terminal d'A3G. 83
Tableau 17 : Tableau regroupant les différentes constructions et combinaisons testées en co-expression entre Vif/CBFβ et Vif/EloB/EloC. 126
Tableau 18 : Tableau regroupant les différentes conditions testés afin d'exprimer et de purifier le complexe Vif/CBFβ/EloB/EloC/Cul5/Rbx2. 128
Tableau 19: Tableau résumant les différentes conditions de cristallisation testées pour les quatre complexes purifiés. 138
Tableau 20: Composition des tampons utilisés pour la purification des protéines produites en cellules BHK21. 154

SOMMAIRE

Figures :

Figure 1: La classification de Baltimore..... 2
Figure 2: Organisation des génomes rétroviraux..... 3
Figure 3: Phylogénie des rétrovirus..... 4
Figure 4: Arbre phylogénétique des rétrovirus humains (VIH-1 et VIH-2) et de primates (SIV) 6
Figure 5: La distribution régionale des sous-types génétiques et formes recombinantes du VIH-1 en 2008 présentée par UNAIDS..... 6
Figure 6: Schéma de la région 5'UTR du VIH-1. 8
Figure 7: Structure schématique d'une particule virale du VIH-1. 9
Figure 8: Evolution du SIDA chez les individus ne recevant pas de traitements antiviraux..... 10
Figure 9: Etude effectuée en 2013 montrant le nombre d'adultes et d'enfants vivant avec le VIH et leur répartition dans le monde 11
Figure 10: Les différentes étapes du cycle réplicatif du VIH-1..... 12
Figure 11: Mécanisme d'entrée du virus dans la cellule. 14
Figure 12: Rétrotranscription de l'ARN génomique du VIH-1. 15
Figure 13: Intégration de l'ADN viral dans l'ADN cellulaire..... 17
Figure 14: Représentation du LTR 5' sur l'ADN proviral intégré dans le génome de l'hôte. 18
Figure 15: Modèle de l'action trans-activatrice de Tat..... 19
Figure 16: Représentation d'un monomère de Rev et des différents domaines fonctionnels 20
Figure 17: Bourgeonnement de la particule virale de la membrane plasmique 22
Figure 18: Représentation schématique de la maturation $Pr55^{Gag}$ 23
Figure 19: Représentation schématique des domaines de la protéine TRIM5α. 27
Figure 20: Modèle actuel du rôle de TRIM5α sur la décapsidation précoce..... 28
Figure 21: Représentation de la dimérisation de la téthérine. 29
Figure 22: Restriction du VIH-1 par la téthérine 30
Figure 23: Structure primaire de SAMHD1 31
Figure 24: Famille des APOBEC3 cytidines désaminases et leur organisation.. 34
Figure 25: Structures cristallographiques de la protéine complète APOBEC3C, des domaines CD2 d'APOBEC3G et APOBEC3F. 35
Figure 26: Structure en solution déterminée par RMN d'un variant du domaine N-terminal d'A3G responsable de l'interaction avec Vif. 35
Figure 27: Activité désaminase-dépendante d'A3G/F.. 37
Figure 29: La structure cristallographique de Vif 41
Figure 30: Polyubiquitination et dégradation d'APOBEC3 par la protéine Vif 43

Figure 31: Modèle proposé suggérant le mode d'action d'ASK1 et son effet sur la restauration de l'activité antivirale d'A3G..... 46
Figure 33: Le cycle infectieux du virus de la vaccine..... 48
Figure 34: Génome de la souche du virus de la vaccine modifié Ankara 50
Figure 35 : Représentation schématique du virus de la vaccine..... 55
Figure 37: Carte des deux principaux vecteurs d'expression bactériens utilisés..... 60
Figure 38: Stratégie utilisée pour la co-expression du complexe chez E. coli. 61
Figure 39: Clonage Biobrick. 64
Figure 40: Le plasmide pEntr_BB-TEV-YFP et ses caractéristiques. 66
Figure 50: Le plasmide pVote0GW 68
Figure 51: Principe de la technologue Gateway®..... 69
Figure 52: Schéma simplifié des étapes de clonages effectuées pour l'expression de la protéine A3G et de la polyprotéine Vif-CBFβ-EloB-EloC-A3G dans le système mammifère. 70
Figure 53: Principe de l'infection des cellules BHK21 par le MVA..... 71
Figure 54: Recombinaison homologue entre le génome viral de la vaccine et le plasmide d'expression pVote.. 72
Figure 55: Voie de synthèse des purines. 74
Figure 56: Représentation du système de la coumerrmycine/GyrB-PKR..... 75
Figure 57: Chromatographie d'affinité Strep-tag®-Strep-tactin® 84
Figure 58: Courbes de spectres standards de dichroisme circulaire pour des peptides se retrouvant dans une structure secondaire définie..... 88
Figure 59: Réaction de méthylation de lysines..... 91
Figure 60: Montage d'une expérience de microcalorimétrie 92
Figure 62: Robot de cristallisation type Mosquito®. 95
Figure 63: Schéma de la stratégie ESPRIT. 96
Figure 64: Schéma récapitulatif de la stratégie utilisée au laboratoire pour aller du gène à la structure.. 97
Figure 65: Purification d'A3G-CD2 sur colonne d'affinité.. 104
Figure 66: Purification de A3G-CD2 par tamis moléculaire..... 104
Figure 67: Analyse de CD2-A3G par DLS.. 105
Figure 68: Chromatogrammes montrant l'activité de hA3G CD2 106
Figure 69: Caractérisation de l'interaction CD2 d'A3G avec Vif..... 107
Figure 70: Schéma de la délétion N-terminal et les substitutions des acides aminés effectués au niveau d'A3F Wildtype afin de produire la chimère soluble A3F185-373-11x. 108
Figure 71: Purification de A3F-CD2chimère par tamis moléculaire..... 109
Figure 72: Schéma de l'expérience ESPRIT que nous avons réalisé, indiquant le nombre de clones obtenus pour chaque construction. 110

SOMMAIRE

Figure 73: Exemple de matrice de colonies obtenues pour le criblage des constructions d'APOBEC3F et APOBEC3G.. 111

Figure 74: Exemple de résultats d'expression de constructions issues du criblage ESPRIT 112

Figure 75: Gel SDS-PAGE 12% montrant les résultats de l'expression de hA3F CD1, hA3F CD2 et hA3G CD1 dans les cellules ArcticExpress® 113

Figure 76: Gel SDS-PAGE 12% montrant les résultats de la purification en gel filtration. 113

Figure 77: Purification de 6xHis-CBFβ par tamis moléculaire. 116

Figure 78: Purification de 6xHis-CBFβ142 par tamis moléculaire..... 118

Figure 79: Purification de EloB(6xHis)-EloC par tamis moléculaire 120

Figure 80: Purification de Cul5-Rbx2 par tamis moléculaire 122

Figure 81: Spectre d'analyse LC-MS/MS du complexe Cul5-Rbx2.. 123

Figure 82: Résultats des expériences de co-expression entre Vif exprimée en fusion avec une étiquette de 6 Histidines en Nt et CBFβ sans tag.. 124

Figure 83: Résultats des expériences de co-expression entre CBFβ exprimée en fusion avec une étiquette de 6 Histidines en Nt et VifpNL4-3 et HXB2 sans tag..... 125

Figure 84: Construction basée sur deux vecteurs pnEAtH et pMCN utilisée pour la co-expression du complexe $Vif_{HXB2}/_{(6xHis)}CBF\beta_{1-142}$/EloB/EloC. 129

Figure 85: Purification du complexe VifHXB2/(6xHis)CBFβ/EloB/EloC.. 129

Figure 86: Analyse par spectrométrie de masse du complexe $Vif_{HXB2}/_{(6xHis)}CBF\beta_{1-142}$/EloB/EloC... 131

Figure 87: Construction basée sur un unique vecteur (obtenu après plusieurs concaténations) exprimant le complexe $Vif_{NL4.3}/_{(6xHis)}CBF\beta_{1-187}$/EloB/EloC. 132

Figure 88: Purification du complexe $Vif_{pNL4-3}/_{(6xHis)}CBF\beta_{1-187}$/EloB/EloC. 132

Figure 89: Purification du complexe Vif-CBFβ-EloB-EloC-Cul5-Rbx2. 133

Figure 90: Constructions utilisées pour la co-expression entre le complexe $Vif_{NL4.3}$-$_{(6xHis)}CBF\beta_{1-187}$-EloB-EloC et APOBEC3G/F. 134

Figure 91: Test de co-expression entre le complexe pneaTHVifpNL4-3-CBFβ-EloB-EloC et pnCSA3Gopt. 135

Figure 92: Western Blot montrant la présence d'A3G 136

Figure 93: Test de co-expression entre le complexe pneaTHVifpNL4-3-CBFβ-EloB-EloC et pnCSA3Fopt.. 136

Figure 94: Purification du complexe Vif_{HXB2}-CBFβ142-EloB-EloC-A3FCD2c. 137

Figure 95: Purification du complexe Vif_{HXB2}-CBFβ142-EloB-EloC par tamis moléculaire.. 139

Figure 96: Spectre d'analyse LC-MS/MS du complexe Vif_{HXB2}-CBFβ142-EloB-EloC en conditions dénaturantes..... 139

Figure 97: L'insert TEV – FH CBFβ – Vif – EloB – EloC –Strep A3G – YFP est représenté dans le vecteur pVote0GW pour définir le vecteur d'expression final..... 146

Figure 98: L'insert TEV-APOBEC3G-YFP a été représenté dans le vecteur pVote0GW avec une étiquette Flag(10XHis) en N-terminal d'APOBEC3G (a) et une étiquette Strep en (b). 147
Figure 99: Caractéristiques des virus sauvages et recombinants. 148
Figure 100: La visualisation de la fluorescence jaune de la protéine YFP. 149
Figure 101: Test d'expression et purification d'A3G. 150
Figure 102: Test d'expression et purification du complexe FHCBFβ-Vif-EloB-EloC-RRPA3G. 151
Figure 103: Test d'expression des protéines recombinantes Vif et A3G. 152
Figure 104: Purification du complexe-A3G sur colonne de Nickel 155
Figure 105: Purification du complexe– A3G sur colonne de streptavidine.. 156
Figure 106: Purification de Flag/10HisA3G sur colonne de Nickel 157
Figure 107: Purification de StrepA3G sur colonne de streptavidine. 158
Figure 108: La séquence de la polyprotéine TEV-Flag/10HisA3G-YFP. 160
Figure 109: Validation du système de clivage à la TEV en spectrométrie de masse. 160
Figure 110: Identification d'A3G par spectrométrie de masse. 161
Figure 111: Identification de Vif, CBFβ et EloC en spectrométrie de masse. 162
Figure 112: Test d'activité de l'activité cytidine désaminase d'APOBEC3G. 164

ABREVIATIONS

3TC	2',3'-didéoxy-3'-ThiaCytidine
A	Adénine
ABC	ABaCavir
ADN	Acide DésoxyriboNucléique
AGS	Aicardi-Goutières Syndrome
APOBEC3G	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G)
ARN	Acide RiboNucléique
ARNg	ARN génomique
ARNm	ARN messager
ARNv	ARN viral
ARV	AntiRétroViral
ASK1	Apoptosis Signal-regulating Kinase 1
attL	Attachement Left
attR	Attachement Right
AZT	AZidoThymidine
BAP	Biotin Peptide Accepteur
BST2	Bone marrow STromal antigen2
C	Cytosine
CA	CApside
CBFβ	Core Binding Factor β
CCR5	C-C chemokine Receptor type 5
CD	Dichroïsme Circulaire
CD1	Catalytic Domain 1
CD2	Catamytic Domain 2
CD4	Cluster of Differenciation 4
CDA	Cytidine Deaminase Active site
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cpn60	Chaperonin 60
CRF	HIV Circulating Recombinant Forms
CTD	Carboxy-Terminal Domain
CUL4	Cullin 4
CXCR4	C-X-C chemokine Receptor type 4

ABREVIATIONS

CypA	Cyclophiline A
DCAF1	Drosophila Melanogaster Chromatin Assembly Factor 1
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DDB1	DNA damage-binding protein 1
DDI	2′,3′-DiDeoxyInosine
dGTP	désoxyGuanosine TriPhosphate
DIS	Dimerisation Initiation Site
DLS	Dimer Linkage Structure
DLS	Diffusion Light Scattering
DSIF	DRB Sensitivity Inducing Factor
DTT	DiThioThréitol
E.coli	Escherichia coli
eIF2α	Eukaryotic Initiation Factor 2α
EloB	Elongin B
EloC	Elongin C
EMC	EncephaloMyoCarditis virus leader
Env	Enveloppe
ESCRT	Endosomal Sorting complex Recquired for Transport
ESPRIT	Expression of Soluble Proteins by Random Incremental Truncation
FDA	Food and Drug Administration
FH10P	Flag/10Histidines/Protease cleavage site
FL	Full Length
G	Guanine
Gag	Group Antigen Specific
GFP	Green fluorescent Protein
GMEM	Glasgow’s Minimum Essential Medium
GMP	Guanosine MonoPhosphate
Gp	Glycoprotein
GPI	Glycosyl-Phosphatidyl Inositol
Gpt	Guanine phosphoribosyl transférase
GST	Gluthation S-Transférase
HAARTS	Highly Active AntiRetroviral TherapieS
HA-L	HemAgglutinin gene Left
HA-R	HemAgglutinin gene Right
HEPES	Acide 4-(2-hydroxyEthyl)-1-Pipérazine Ethane Sulfonique
HMGI(Y)	High-Mobility Group Protein HMG-I/HMG-Y

ABREVIATIONS

HR	Heptad repeats
HSPG	Heparan Sulfate ProteoGlycans
IFN	Interferon
IN	Integrase
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISG	interferon-stimulated genes
ITC	Isothermal Titration Calorimetry
LC-MS/MS	Liquid Chromatography coupled to tandem Mass Spectrometry
LEDGF	Lens Epithelium-Derived Growth factor
LFA-1	Lymphocyte Function-associated Antigen 1
LTR	Long Terminal Repeats
MA	Matrice
MES	2-(N-Morpholino)EthaneSulfonic acid
MVA	Modified Vaccinia Ankara
MVB	Multivesicular endosomes
NC	Nucleocapsid
NcSRS2	Neospora caninum surface protein2
Nef	Negative Regulatory Factor
NELF	Negative Elongation Factor
NF-κB	Nuclear Factor-*kappa B*
NF-IB	HIV-1 induced Nuclear Factor IB
NLS	Nuclear Localization Signal
OMPF	Outer Membrane Protein F
ORF	Open Reading Frame
PBS	Primer Binding Site
PCAF	P300/CREB binding protein-associated factor
PCR	Polymerase Chain Reaction
PFU	Plaque-Forming Unit
PIC	Pre-Integration Complex
PKR	protéine kinase ARN dépendante d
Pol	Polymerase
PR	Protéase
p-TEFb	Positive Transcription Elongation Factor
pVote	Vaccinia virus /lac operon/T7 RNA polymerase
RBE	Rev Binding Element
RBS	Ribosome Binding Site
RMN	Résonance Magnétique Nucléaire

ABREVIATIONS

RNP	RiboNucléoProtéin
RRP	StRep-StRep-Protease cleavge site
RT	Rétrotranscriptase
RTC	Reverse Transcription Complex
SAM	Sterile alpha Motif
SAMHD1	SAM and HD domain containing protein
SIDA	Syndrome d'ImmunoDéficience Acquise
Sm23	Schistosoma mansoni
SOCS	Suppressor Of Cytokine Signaling
SP1	Specificity Protein 1
SPRY	SPla and the RYanodine Receptor
TAR	TransActivation Responsive region
Tat	Transcriptional transactivator
TEV	Tobacco Etch Virus
TIP47	Tail-Interacting protein of 47 KDa
TM	Trans-Membranaire
TRIM	TRipartite Interaction Motif
U	Uracile
UTR	UnTranslated Region
VAIE	Virus de l'Anémie Infectieuse Equine
VIB	Virus de l'Immunodéficience Bovine
VIF	Virion Infectivity Factor
VIH	Virus de l'Immunodéficience humaine
VIS	Virus de l'Immunodéficience simienne
VMW	Visna/Maedi Virus
Vpr	Viral protein r
Vpu	Viral protein u
WR	Western Reserve
XMP	Xanthosine Monophosphate
YFP	Yellow Fluorescent Protein
βTrcP	β-transducing repeat-containing protein)

INTRODUCTION

I. Les Virus

A. Définitions

André Lwolff a défini en 1953 le concept de virion qui est la forme extracellulaire des virus. C'est une particule virale mature et infectieuse libre dans le milieu extérieur. Le virus existe aussi sous forme intra-cellulaire qui peut être intégré sous forme dormante ou se réplique en utilisant la machinerie cellulaire de l'hôte. Les virus se caractérisent par 3 caractères essentiels : 1- Ils possèdent un seul type d'acide nucléique ADN ou ARN portant l'intégralité de l'information génétique. 2- ils se reproduisent à partir de leur matériel génétique et par réplication. 3- Ils sont doués d'un parasitisme intracellulaire obligatoire, et ne peuvent se reproduire qu'au sein d'une cellule hôte vivante, la multiplication virale pourra aboutir à une lyse cellulaire ou à une persistance virale, cette interaction entre le virus et la cellule engendre des lésions non létales.

Il existe 2 classifications des virus connus : La classification Baltimore proposée par David Baltimore et qui est basé sur le type d'acide nucléique des virus (ADN ou ARN) et son mode d'expression ainsi que le procédé de réplication de l'ADN (Fig. 1). La deuxième classification est du la Comité International sur la Taxonomie des Virus où les virus sont rangés par ordre, famille, sous-famille, genre et espèce. Ces classifications ne sont pas phylogénétiques car les virus ne partagent pas d'origine commune.

Selon la classification Baltimore il existe 7 groupes de virus :

– Groupe 1 : Virus à ADN à double brin. Ce type de virus se réplique dans le noyau de la cellule l'hôte (à l'exception des poxvirus) en utilisant son ADN polymérase pour la réplication du génome virale. Un exemple de ce groupe est les *Herpesviridae.*
– Groupe 2 : Virus à ADN à simple brin. Ce groupe de virus se réplique aussi dans le noyau et forme un ADN double brin intermédiaire au cours de la réplication, par exemple les *Circoviridae.*
– Groupe 3 : Virus à ARN double brin. Ces virus se répliquent dans le cytoplasme, ils n'ont pas besoin de la polymérase de la cellule hôte autant que les virus à ADN. Chez ces virus, la réplication est individuelle et monocistronique ce qui signifie que chacun des gènes code pour une seule protéine. Un exemple de ce groupe est les *Reoviridae.*

– Groupe 4 : Virus à ARN simple brin à polarité positive. C'est un virus qui se réplique dans le cytoplasme contrairement aux virus à ADN dont la réplication a lieu dans le noyau. Tous les gènes de ces virus sont définis comme antisens ce qui permet aux polymérases de l'hôte de les décrypter directement pour synthétiser des protéines. Un exemple de ce groupe est les *Astroviridae.*

– Groupe 5 : virus à ARN simple brin à polarité négative. Les gènes de ce groupe de virus ne peuvent pas être décryptés par les polymérases de l'hôte. Ils doivent êtres transcrits par les polymérases virales dans une forme lisible à polarité positive. La réplication a aussi lieu dans le cytoplasme. Un exemple de ce groupe est les *Filoviridae.*

– Groupe 6 : Rétrovirus à ARN simple brin. Ce groupe de virus utilise la transcriptase inverse pour convertir l'ARN à polarité positive en ADN pour assurer la réplication dans la cellule hôte. Un exemple de ce groupe est les *Retroviridae.*

– Groupe 7 : Pararétrovirus à ADN double brin. L'information génétique chez ce petit groupe de virus est codée sous forme d'ADN double brin. La réplication se base sur l'ARNm viral et l'ARN subgénomique. Un exemple de ce groupe est les *Hepadnaviridae.*

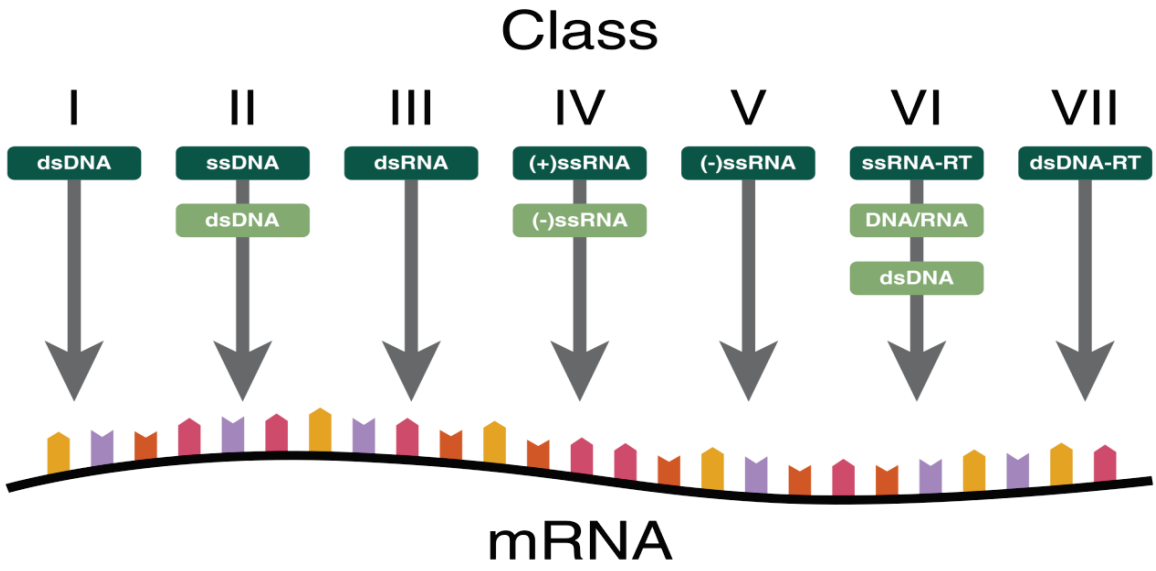


Figure 1: La classification de Baltimore.

B. Les rétrovirus

1. Généralités

Les rétrovirus constituent une famille de virus infectant les vertébrés et invertébrés et provoquant des pathologies très variées. Ce sont des virus enveloppés faisant de 80 à 120 nm de

diamètre. L'enveloppe virale provient d'une partie de la membrane plasmique durant l'étape de bourgeonnement du cycle réplicatif. Les rétrovirus possèdent un génome à ARN linéaire, monocaténaire et présent en 2 copies et dont la taille varie de 7 à 12 Kb. Chez cette famille de virus, l'ARN génomique est convertit en ADN double brin durant sa réplication par la RT virale. Après la transcription inverse, le génome du virus sera transporté dans le noyau et intégré dans l'ADN de la cellule hôte par l'intégrase.

Les rétrovirus sont divisés en 2 catégories: simple et complexe selon l'organisation de leurs génomes. Tous les rétrovirus comportent 3 domaines principaux codant pour les protéines virales: gag, pol et env codant respectivement pour les protéines de structure (MA, CA et NC), enzymatiques (PR, RT et IN) et les protéines d'enveloppe (gp41 et gp120) (Fig. 2).

Généralement, les rétrovirus simple possèdent seulement ces informations élémentaires alors que les rétrovirus complexes codent en plus pour des protéines régulatrices et auxiliaires.

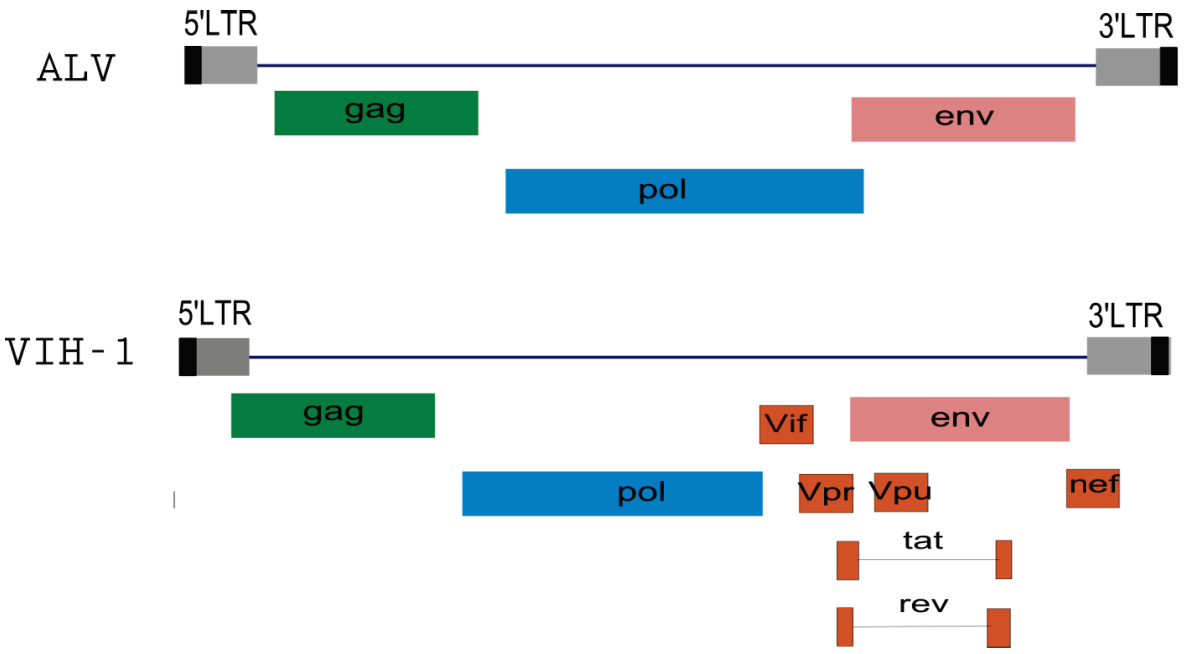


Figure 2 : Organisation des génomes rétroviraux. Les génomes du virus de l'ALV (rétrovirus simple) et du VIH-1 (rétrovirus complexe) sont représentés.

2. Classification

On classe les rétrovirus en 2 catégories : les rétrovirus exogènes et endogènes. Les virus exogènes ne sont pas présents naturellement dans l'organisme et ont besoin de l'infecter pour effectuer leur cycle de réplication. Les rétrovirus endogènes possèdent leur matériel génétique au sein du même génome de l'organisme sans être exprimé.

Seul les rétrovirus exogènes sont formellement classifiés par le comité international sur la taxonomie des virus qui les a regroupés en 7 genres. Ces genres se divisent en deux catégories (Fig. 3).

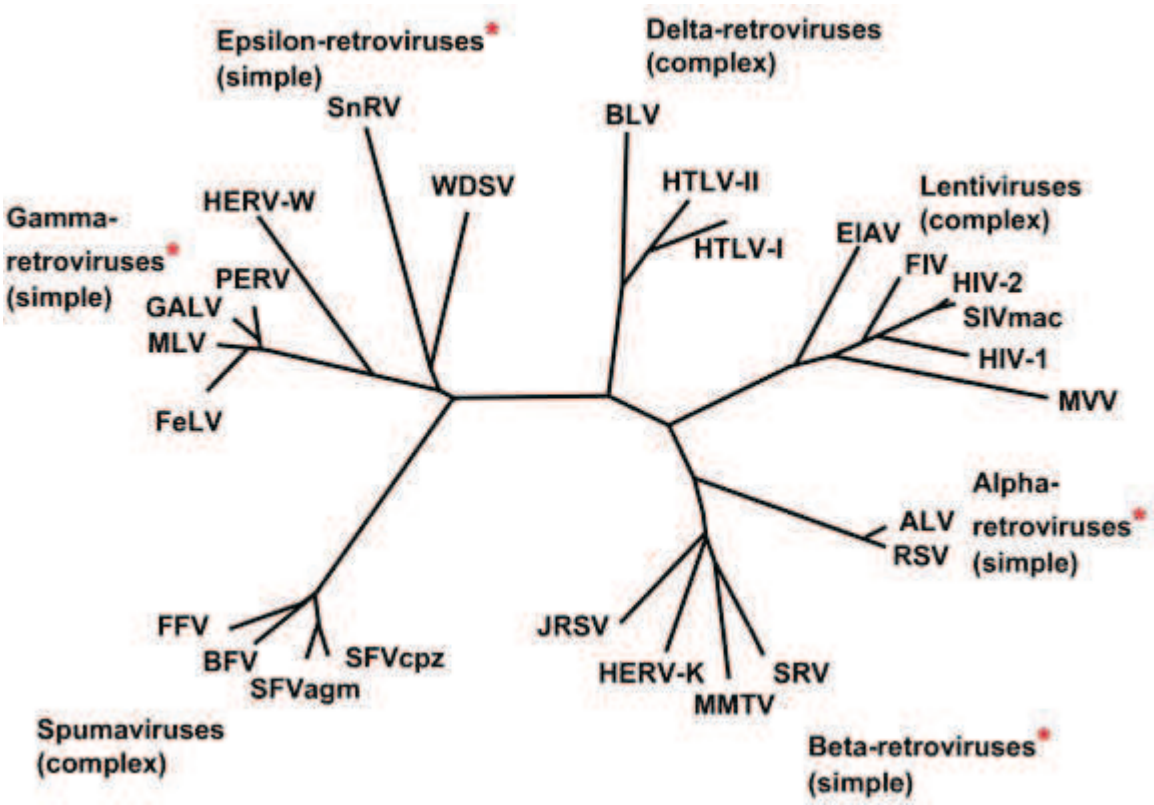


Figure 3 : Phylogénie des rétrovirus : Les astérisques indiquent des genres incluant des virus endogènes [1].

3. Les lentivirus

Les lentivirus (du latin *lenti* lent) sont des virus de la famille des rétrovirus. Leur nom s'explique par leur longue période d'incubation qui est nécessaire avant la manifestation de leur pouvoir pathogène sur les cellules. Ils infectent principalement les cellules immunitaires.

Les lentivirus sont des virus sphériques encapsulés mesurant entre 80 et 100 nm de diamètre. Ils sont composés de protéines virales structurales et d'enveloppe (gp41 et gp120) et enzymatiques (RT, IN et PR). Ces protéines représentent 60 % de la masse totale. Les lipides constituent 35 % et se situent au niveau de l'enveloppe. 3% de la masse totale du virion est composée de glucides et 2% sont occupés par le génome.

Cinq sérogroupes ont été identifiés parmi les lentivirus, infectant chacun une famille particulière des mammifères. Parmi ces groupes, les VIB, VAIE, VIF et VMV infectent respectivement les bovins,

équidés, félins et caprins. Le 5ème groupe est le VIH ou VIS qui infecte les primates. Les lentivirus rendent l'hôte immunodéprimé en provoquant la mort des cellules immunitaires infectées ce qui développe la maladie du SIDA. Seul le VAIE fait exception. Malgré une réplication rapide du rétrovirus, les animaux contrôlent l'infection en demeurant en stade chronique et qui n'aboutit pas au SIDA.

II. Le Virus de l'Immunodéficience Humaine de type 1

A. Infection et VIH

1. Découverte

C'est en Juillet 1981 que le centre pour le contrôle et la prévention des maladies (CDC) d'Atlanta relève une maladie se manifestant par une recrudescence de pneumonies et de sarcomes de Kaposi et affectant le système immunitaire. En 1982, cette maladie prend le terme AIDS pour Acquired ImmunoDeficiency Syndrome ou SIDA en français. L'agent responsable de cette maladie est isolé en 1983 dans l'équipe du professeur Luc Montagnier à l'institut Pasteur à Paris par Francoise Barré-Sinoussi qui mit en évidence une activité de rétro transcription dans un échantillon de ganglion lymphatique d'un patient atteint par la maladie [2]. En mai 1986, une commission de nomenclature virologique attribue à ce virus l'appellation de Virus de l'Immunodéficience humaine (VIH). Après la découverte d'un second virus VIH-2 découvert chez des patients originaires d'Afrique de l'Ouest [3], la dénomination VIH-1 est adoptée pour ce virus.

2. Origine et variabilité du VIH-1

L'origine du VIH-1 est probablement liée à une infection zoonotique d'un virus simien à l'homme. Les arbres phylogénétiques ont permis d'identifier le VIH-1 comme étant le lentivirus le plus apparenté au SIVcpz du chimpanzé [4]. SIV est un lentivirus qui affecte 36 espèces différentes de primates en Afrique subsaharienne [5]. Le VIH-2 est plus semblable au SIVmm (infectant les sooty mangabeys) et au SIVmac (infectant les macaques) (Fig. 4). La transmission de ce virus chez l'homme est due à l'exposition du sang contaminé par des blessures ou des morsures lors du dépeçage de ces animaux.

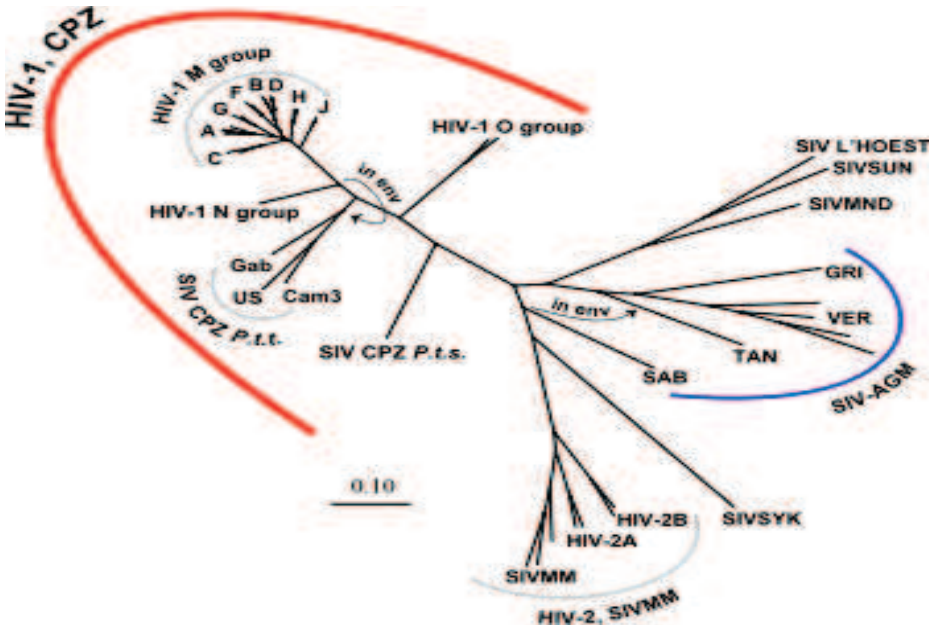


Figure 4 : Arbre phylogénétique des rétrovirus humains (VIH-1 et VIH-2) et de primates (SIV) [6].

Selon les études des relations évolutives avec le SIVcpz, le VIH-1 se divise en 3 groupes M (Main), N (Non O non M) et O (Outlier) et proviennent de trois évènements de transmission indépendants. Le groupe M est le plus répandu au monde et comprend neuf sous types (A, B, C, D, F, G, H, J, K) et plusieurs formes recombinantes (CRF) (Fig. 5). Les groupes O et N sont rares et se retrouvent presque exclusivement en Afrique centrale. L'épidémie de VIH-2 n'a pas atteint la même ampleur que celle du VIH-1 et n'est retrouvée qu'en Afrique de l'Ouest. Au sein du VIH-2, on retrouve sept sous types de A à G.

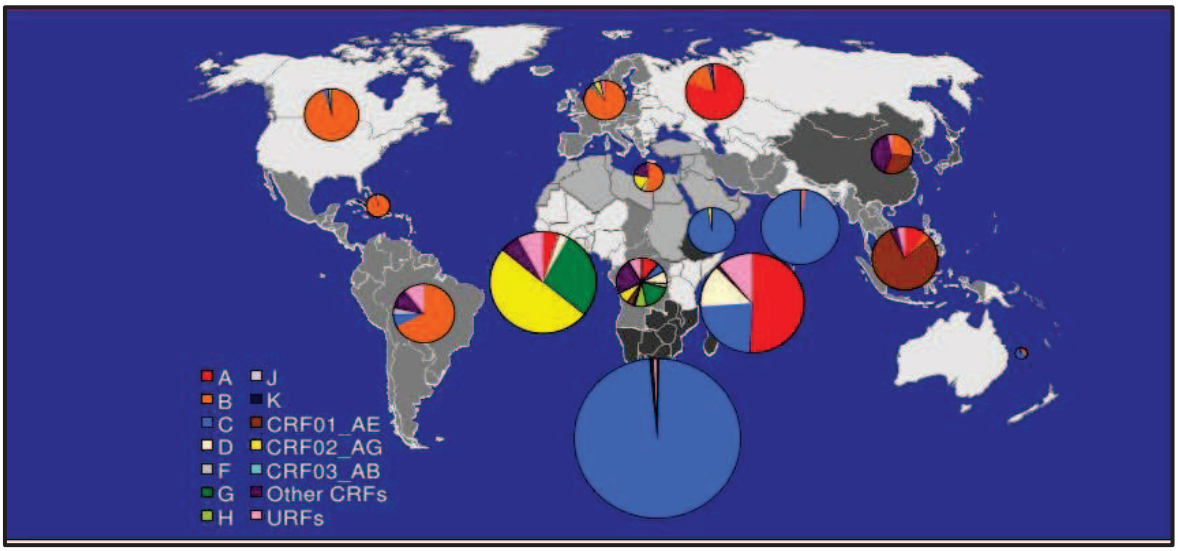


Figure 5 : La distribution régionale des sous-types génétiques et formes recombinantes du VIH-1 en 2008 présentée par UNAIDS durant la 5ème conférence francophone VIH/SIDA en 2010 à Casablanca.

3. Mode de transmission

Les modes de contamination se font :

– Majoritairement par voie sexuelle : C'est le mode de transmission de l'infection par le VIH le plus répandu au monde. Il représente 85 à 90 % des causes d'infection par le VIH.
– Par voie sanguine. L'injection de sang en provenance d'un sujet contaminé est susceptible de transmettre le VIH. Il peut s'agir d'échange de seringues entre usagers de drogue (toxicomanie), d'un acte de soin ou d'Accident d'Exposition au Sang (AES) des professionnels de santé. Le risque de transmission du VIH par voie intraveineuse est estimé à 0,67 %.
– Par voie materno-fœtale : Le sang maternel communique avec celui du fœtus à travers le placenta. La transmission du virus dans ce cas se produit lors de la grossesse (durant le dernier semestre) ou lors de l'accouchement (65% des cas de transmission). La contamination peut se produire aussi lors de l'allaitement maternel.

4. Organisation génétique du VIH-1

a. Régions non codantes

La rétrotranscription de l'ARN viral conduit à la synthèse d'un ADN proviral bicaténaire. A chacune de leurs extrémités on trouve cette répétition de séquences identiques sur plusieurs centaines de bases, d'où leur nom de LTR, Long Terminal Repeats. Suite au processus de rétrotranscription, ces séquences répétées sont également présentes dans l'ADN proviral. Ces séquences sont non-codantes, mais contiennent des régions nécessaires à l'intégration de l'ADN proviral dans le génome de la cellule hôte, à la synthèse de l'ARN viral et sa régulation (Fig. 6) :

– Une coiffe en 5'
– La région R (Redondant) de 98 nucléotides présente en deux exemplaires aux extrémités du génome. Elle a un rôle important lors de la rétrotranscription. Elle comporte une région TAR à son extrémité 5' qui intervient dans la régulation de l'expression visant la protéine tat. Enfin, la région R contient une séquence signal pour la polyadénylation qui n'est fonctionnelle qu'à l'extrémité 3' de l'ARN viral.
– La région U5 (Unique en 5'), elle est fortement conservée et très structurée. Elle contient des séquences qui interviennent dans le processus d'hybridation entre l'ARN viral et l'ARNt^{Lys}3. C'est la première partie du génome à être rétrotranscrite.
– La région PBS (Primer Binding Site) qui sert d'amorce à la transcription inverse.

– La région DLS (Dimer Linkage Structure) localisée entre le PBS et le gène gag. Elle contient en particulier la séquence DIS (Dimerisation Initiation Site) et elle est responsable de la dimérisation du génome viral.

– La région U3 (Unique en 3') est dupliquée à l'autre extrémité de l'ADN proviral après la rétrotranscription. Elle contient les signaux nécessaires à l'intégration du provirus ainsi qu'à la régulation transcriptionnelle par les protéines cellulaires.

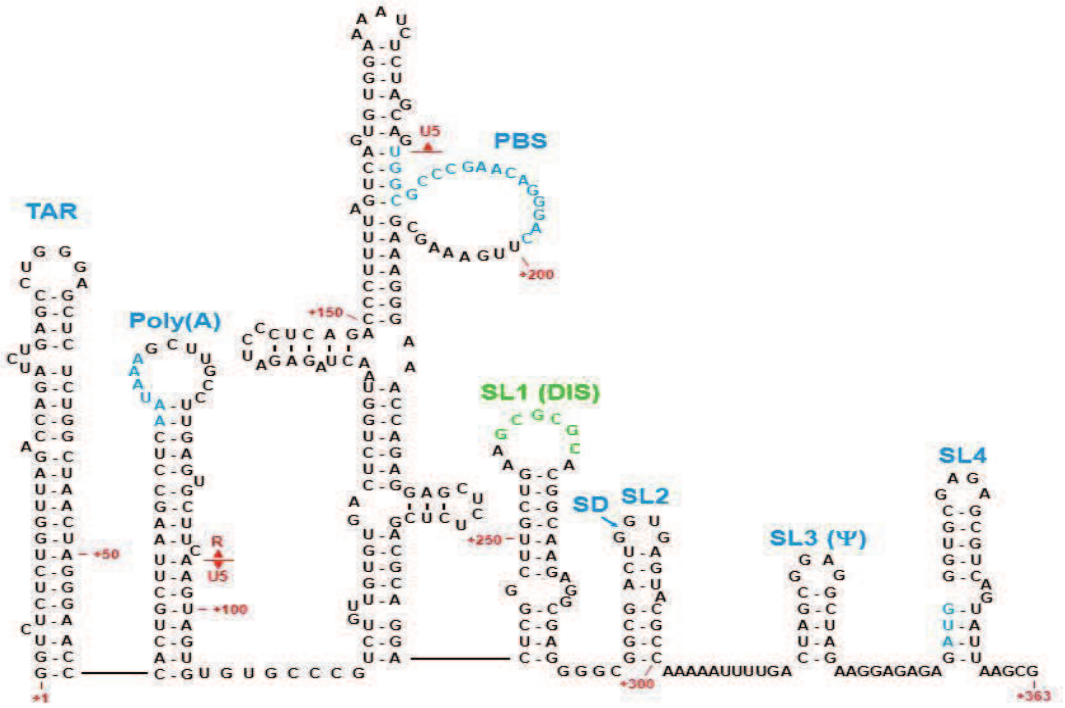


Figure 6 : Schéma de la région 5'UTR du VIH-1 [7].

Les régions codantes contiennent principalement trois gènes de protéine de structure, enzymatiques et de l'enveloppe (gag, env et pol) et six autres gènes qui codent pour des protéines régulatrices. Les rôles et fonctions de ces différentes protéines sont présentés dans le tableau ci-dessous.

Gènes	Protéines et fonctions
Gag	Protéine de structure
Pol (polymérase)	Transcriptase inverse, protéase, intégrase
Env (enveloppe)	Gp120: Attachement/gp41: fusion
Tat (transactivator)	Régulateur positif de la transcription
Rev (Regulator of Viral Expression)	Export des transcrits non épissés hors du noyau
Vif (Viral infectivity)	Empêche la désamination de l'ADN viral en liant la protéine APOBEC3G
Vpr (Viral Protein R)	Transport de l'ADN au noyau, augmente la production virale et contrôle le cycle cellulaire
Vpu (Viral Protein U)	Réduit l'expression du CD4
Nef (Negative Regulation Factor)	Augmente l'infectivité et réduit l'expression du CD4

Tableau 1 : Rôles des protéines virales.

b. Structure du virion

La particule virale du VIH, sous sa forme mature est presque sphérique avec un diamètre de 120 nm [8]. Elle est formée d'une capside conique et entourée d'une enveloppe formée par une bicouche lipidique issue de la membrane plasmique de la cellule hôte acquise lors du bourgeonnement (Fig. 7). Les glycoprotéines d'enveloppe gp120 (surface) et gp41 (transmembranaire), dérivant d'un précurseur polyprotéique gp160 issu du gène env, sont insérées dans l'enveloppe sous forme de trimères [9] mais le nombre de protéines associées est aléatoire [8]. Une matrice est associée à l'enveloppe, elle est composée de la protéine MA ou p17 produite par le gène gag. A l'intérieur de la capside, on retrouve deux copies d'ARN génomiques et les protéines nécessaires à la réplication virale : l'IN, la RT et la PR ainsi que les protéines accessoires Tat, Nef, Vif, Vpr, Vpu [10] ainsi que des protéines cellulaires incorporées de manière spécifique et aléatoire lors du bourgeonnement.

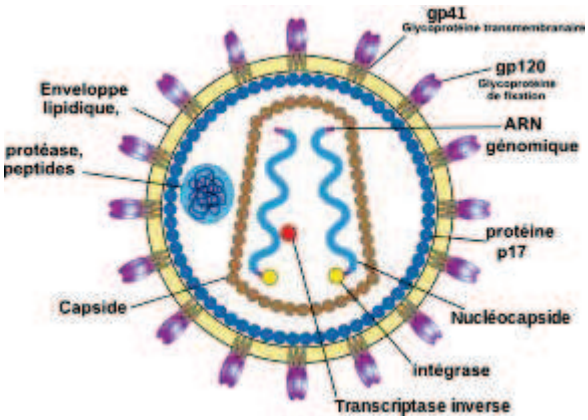


Figure 7 : Structure schématique d'une particule virale du VIH-1 (Wikipédia).

5. Evolution de la maladie

En l'absence de traitement antirétroviral, la quasi-totalité des patients infectés par le VIH évoluent vers le SIDA, phase ultime de la maladie. Le virus attaque les lymphocytes T, cellules du système immunitaire et porteurs du récepteur $CD4^+$. Cette molécule s'exprime également à la surface d'autres cellules comme les monocytes et macrophages, les cellules dendritiques et les cellules des systèmes lymphatique, hématopoïétique et nerveux [11, 12]. Lors de l'infection par le VIH-1, la dissémination se fait tout d'abord par l'intermédiaire des cellules dendritiques au niveau des muqueuses. Ensuite, ces cellules entrent en contact avec les cellules du système immunitaire [13]. Les virions pourront alors coloniser la circulation sanguine.

L'évolution de la maladie se divise en 3 phases (Fig. 8) :

– La primo-infection ou phase aigüe : Elle correspond à une période de réplication virale intense et à une dissémination du virus dans diverses cellules cibles. Cette phase dure quelques semaines. Le nombre de lymphocytes T chute d'environ 1000 cellules/mm^3. La lyse des lymphocytes provient des effets cytopathogènes du virus et l'induction de l'apoptose par la cellule infectée. A cette phase de l'infection, le système immunitaire se déclenche en diminuant le taux de virus plasmatique.
– La phase symptomatique : Cette phase peut durer plusieurs années. La charge virale est constante, il y a un équilibre entre la production de virus et son élimination par le système immunitaire.
– La phase SIDA : elle intervient suite à une diminution très importante du taux de $CD4^+$ conduisant à une immunodéficience et à un risque d'infections opportunistes. Par la suite, la production virale est stimulée et la personne infectée ne survivra pas en absence de traitement.

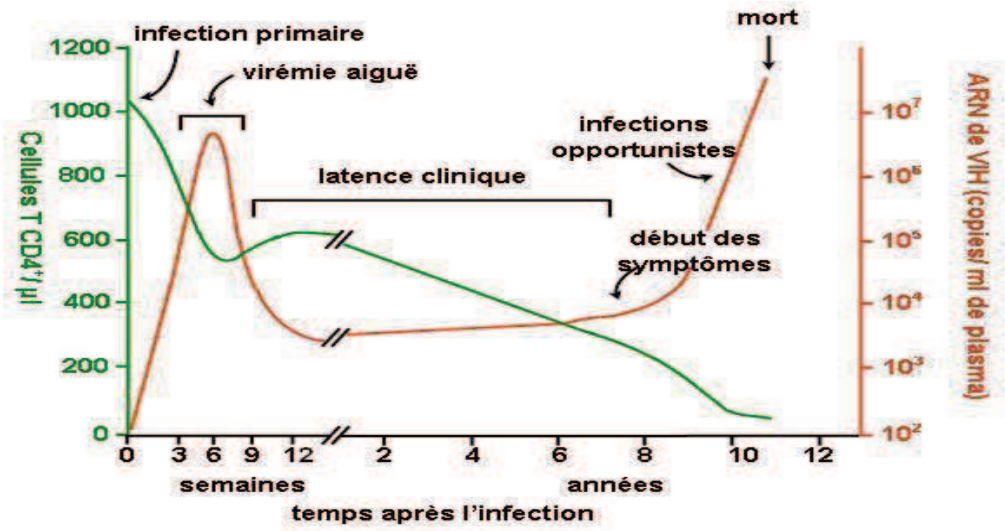


Figure 8 : Evolution du SIDA chez les individus ne recevant pas de traitements antiviraux [14].

6. La pandémie de SIDA

La pandémie de SIDA continue sa progression dans le monde avec de très importantes variations géographiques et une grande hétérogénéité des épidémies qui la compose. Depuis le début de l'épidémie environ 78 millions de personnes ont été infectées par le VIH et 39 millions de personnes sont décédées de maladies liées au sida.

Selon les données mondiales les plus récentes publiées par l'ONUSIDA, 35 millions de personnes vivent encore avec le VIH en 2013. La majorité des infections à VIH se localisent en Afrique subsaharienne (24,7 millions d'infections en 2013) (Fig. 9). Les femmes représentent 58% du nombre total des personnes vivant avec le VIH.

2,1 millions de personnes ont été nouvellement infectées par le VIH en 2013 dont 240 000 enfants contre 3,4 millions en 2001 dont 580 000 enfants d'où une chute des infections à VIH de 38% d'infections totales et 58% d'infections parmi les enfants. La prévention commencerait donc à porter ses fruits et les traitements antirétroviraux gagnent en efficacité. En plus, l'accessibilité à ces traitements a augmenté de 37 % en une année.

Figure 9 : Etude effectuée en 2013 montrant le nombre d'adultes et d'enfants vivant avec le VIH et leur répartition dans le monde. Source ONUSIDA

B. Le cycle réplicatif du VIH-1

Le cycle réplicatif du VIH-1 se divise en deux phases, une phase précoce et une phase tardive. La phase précoce est la phase non intégrative qui débute lors de l'entrée du virus dans la cellule jusqu'à son intégration dans le génome de l'hôte. La phase tardive est la phase post-intégrative qui comprend la transcription du provirus et l'expression des gènes viraux jusqu'au bourgeonnement et maturation de la particule virale (Fig. 10).

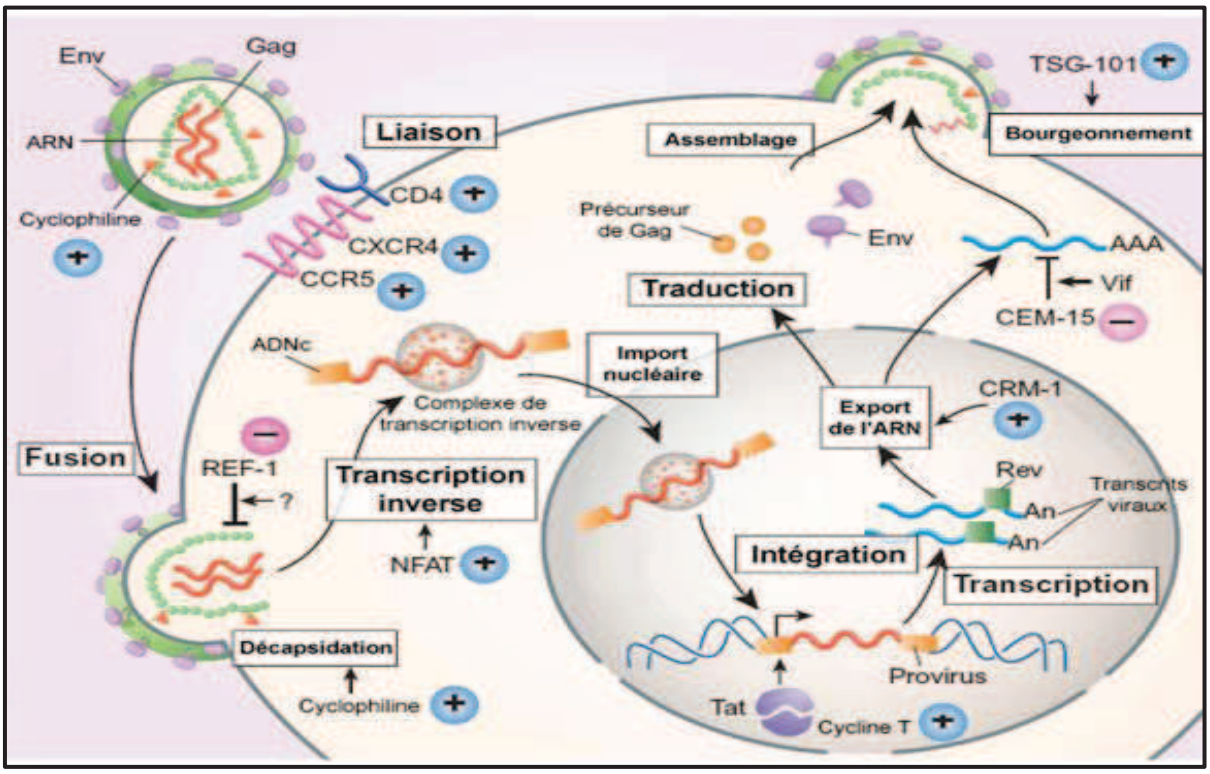


Figure 10 : Les différentes étapes du cycle réplicatif du VIH-1 et un résumé des facteurs cellulaires en jeu. Les facteurs cellulaires (+) et (-) sont indiqués à côté de l'étape de la réplication virale dont ils affectent [15].

1. La phase pré-intégrative

a. Entrée virale : Reconnaissance et entrée

L'étape initiale du cycle de réplication du VIH-1 est l'attachement d'un virion à la surface d'une cellule cible du système immunitaire majoritairement les lymphocytes T4. Le VIH-1 reconnait en premier le récepteur CD4 (récepteur principal de l'entrée) et un corécepteur CCR5/CXCR4. Diverses molécules de surface cellulaires peuvent participer à l'attachement comme par exemple les HSPG (héparanes sulfates protéoglycanes) [16] , le LFA-1 [17] et la nucléoline [18]. En effet, l'affinité de l'enveloppe virale pour CD4 est faible [19], pour cela les facteurs d'attachement serviront

à concentrer le virus à la surface de la cellule avant la liaison de CD4 au VIH-1. La glycoprotéine Env est formée d'un domaine de surface gp120 et d'un domaine transmembranaire gp41. Env s'associe en trimère lors de sa synthèse et devient mature suite au clivage de ses deux sous-unités. La fixation du virus à la cellule hôte repose sur l'interaction entre la protéine virale de surface gp120 et les récepteurs CD4. Cette liaison n'est pas suffisante pour l'entrée du virus, ; la présence d'un des deux récepteurs aux chimiokines CCR5 ou CXCR4 est indispensable [20, 21]. L'interaction entre gp120 et le corécepteur entraine la dissociation du complexe gp120-gp41 et provoque l'exposition de l'ectodomaine de la gp41 au corécepteur puis la transition de la gp41 sous sa forme fusiogénique qui va s'insérer dans la membrane plasmique de la cellule cible et entrainer la fusion des membranes cellulaires et virales [9, 22]. Le cœur viral peut ainsi entrer dans la cellule hôte (Fig. 11*)*. Dans sa forme fusiogénique, la gp41 expose les domaines HR (Heptad repeats) N- et C-terminaux, ceux-ci peuvent être liés par des peptides inhibiteurs, ce qui fait de la gp41 une cible thérapeutique [23]. Notons que la formation de pores de fusion nécessite le recrutement d'une quantité suffisante de CD4 et de co-récepteurs [9].

Un autre moyen d'entrée du virion dans la cellule est l'endocytose. La protéine Dynamine est impliquée dans cette voie d'entrée en provoquant la formation de vésicules d'endocytose en remodelant l'actine et favorisant la courbure de la bicouche lipidique. Dans ce cas la protéine gp120 du virus interagit avec une molécule d'adhésion spécifique des cellules dendritiques DC-SIGN. Ce contact permet une internalisation du virus dans des vésicules non-lysosomales. Il a aussi été observé qu'un virus produit sur une lignée cellulaire traitée à l'IFN-I3 est capable d'infecter avec une meilleur fixation une lignée CD4 négative [24].

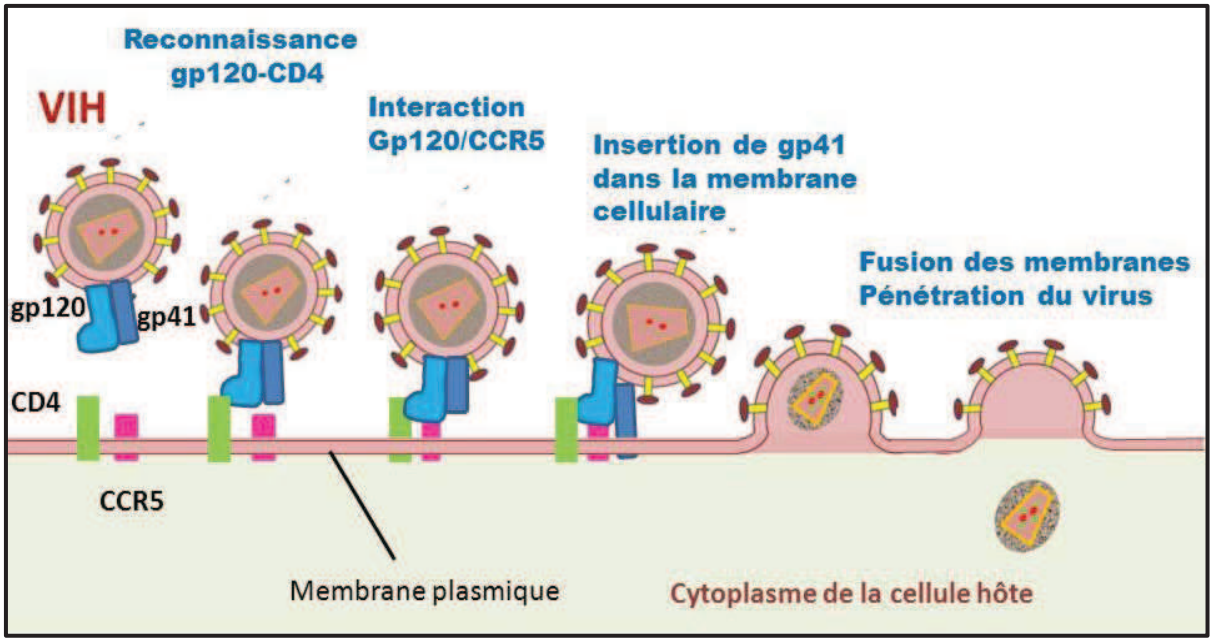


Figure 11: Mécanisme d'entrée du virus dans la cellule. La liaison spécifique entre le récepteur CD4 et la glycoprotéine d'enveloppe gp120 déclenche une deuxième interaction entre gp120 et le co-récepteur (CCR5 ou CXCR4). Ceci déclenche un changement conformationnel exposant gp41 qui va être ancrée dans la membrane de la cellule hôte. Le changement conformationnel de gp41 permettra à la suite la fusion des deux membranes. (adapté de [25]).

b. La rétrotranscription et la décapsidation

Suite à l'étape de fusion et à l'entrée du virus, le core viral est libéré dans le cytoplasme de la cellule hôte. Ces étapes sont suivies par la décapsidation du virus. Les dernières études montrent que la décapsidation est un événement qui survient après la rétrotranscription au niveau des pores nucléaires [26] .

Une fois la capside libérée dans le cytosol, l'ARN viral est rétrotranscrit pour former l'ADN double brin qui forme le génome du VIH. L'acteur majeur de la rétrotranscription est la rétro-transcriptase, une enzyme hétérodimérique de deux sous-unités p66 et p51 multifonctionnelle ayant des activités ADN polymérase ARN dépendante, ADN polymérase ADN dépendante et RNaseH.
La RT utilise comme amorce un ARNt^{lys}3 d'origine cellulaire hybridé à la région PBS (Primer Binding Site) du génome afin d'initier la synthèse d'un ADN simple brin de polarité négative à partir de l'ARN génomique viral. Ensuite, elle dégrade le brin d'ARN génomique par son activité RNaseH. Finalement, elle synthétise un brin d'ADN à polarité positive afin de former l'ADN double brin qui pourra être intégré au génome cellulaire. La RT génère à chaque extrémité du génome de longues répétitions terminales identiques (LTR) (Fig. 12). Ces séquences constituent le promoteur du VIH et sont donc essentielles à l'expression des gènes viraux lors de l'étape de transcription [27]. Une étude

récente a montré que les facteurs de nucléation IB (NF-IB) sont exprimés lors de l'infection par le VIH-1 et provoque un effet négatif sur la réplication du VIH-1 suite à leur interaction avec ces séquences LTR [28].

Vu que la RT ne possède pas d'activité exonucléase 3'-5', elle est à l'origine de nombreuses mutations dans le génome viral ce qui contribue grandement à l'apparition de quasi-espèces ainsi que de virus résistants aux antirétroviraux (ARV) [29].

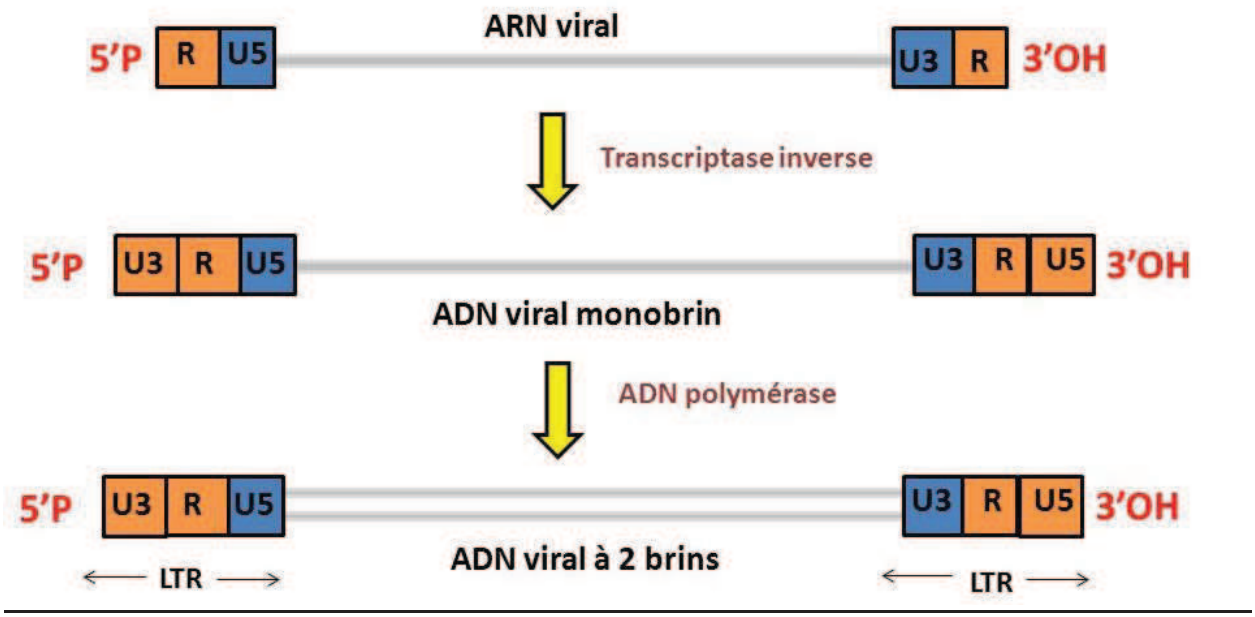


Figure 12 : Rétrotranscription de l'ARN génomique du VIH-1.

La synthèse du brin d'ADN (+) s'achève au niveau d'une séquence riche en AT et déplace le brin d'ADN déjà synthétisé sur une centaine de nucléotides. Ce recouvrement entraine la formation d'une structure à trois brins appelée « DNA flap ». La formation du « DNA flap » signalerait la fin de la rétro-transcription et déclencherait la décapsidation (23,24). La décapsidation entraine la maturation du complexe de transcription inverse (RTC) en complexe de pré-intégration (PIC) transportable au travers des pores nucléaires. Cette étape a lieu à proximité des pores nucléaires [30].

Les complexes de pré-intégration des lentivirus possèdent la propriété de pouvoir être transportés au noyau par la machinerie d'import nucléaire de la cellule. Cette caractéristique leur permet d'accéder au noyau de la cellule en division mais aussi lorsque la cellule ne prolifère pas ou son cycle cellulaire est arrêté en phase G1. De cette façon, le VIH-1 peut infecter des lymphocytes T latents, des macrophages et les cellules dendritiques [18].

c. L'intégration

Le PIC protège l'ADN viral de la dégradation et facilite son intégration dans le génome de l'hôte. La capacité du VIH-1 de se répliquer dans des cellules qui ne sont pas en division, repose sur le transport actif de l'ADN viral (ADNv) dans le noyau de la cellule infectée. Ceci dépend de la capacité du PIC formé de l'ADNv, MA, IN, TI, Vpr et des protéines de l'hôte liant l'ADN comme les HMGI(Y) [31] et LEDGF/p75 [32], à interagir avec la machinerie cellulaire. La plupart des protéines du PIC possèdent un signal de localisation nucléaire (NLS) qui leur permet d'être importées dans le noyau.

Une fois dans le noyau, l'intégrase (IN) permet l'intégration de l'ADN proviral dans le chromosome. L'IN est une protéine de 32 KDa générée suite au clivage de la polyprotéine Gag-pol. La réaction d'intégration se produit en 3 étapes (Fig. 13) :

– « 3' processing » : l'intégrase clive un nucléotide GT en 3'LTR des 2 brins de l'ADN proviral, libérant des extrémités 3'OH rentrantes.
– le transfert du brin : étape de transestérification permettant un clivage de l'ADN cellulaire et ligation de l'ADN viral aux extrémités 5' phosphate générées dans l'ADN chromatinien.
– la dernière étape de l'intégration est l'élimination des nucléotides CA et la réparation de l'ADN par la machinerie cellulaire.

Il a été démontré que le virus s'intègre plus fréquemment en des sites spécifiques, souvent à l'intérieur de gènes qui eux-mêmes sont régulés lors d'une infection par le VIH-1.

Tous les PIC n'aboutissent pas à une infection productive. l'ADN viral non intégré représente 99% de l'ADN viral total [33]. Dans ce cas, il y a 3 scénarios possibles : la liaison des deux extrémités de l'ADN pour former un cercle contenant les deux LTR, la recombinaison homologue du génome viral formant un cercle avec un seul LTR et finalement, l'ADN viral demeure linéaire.

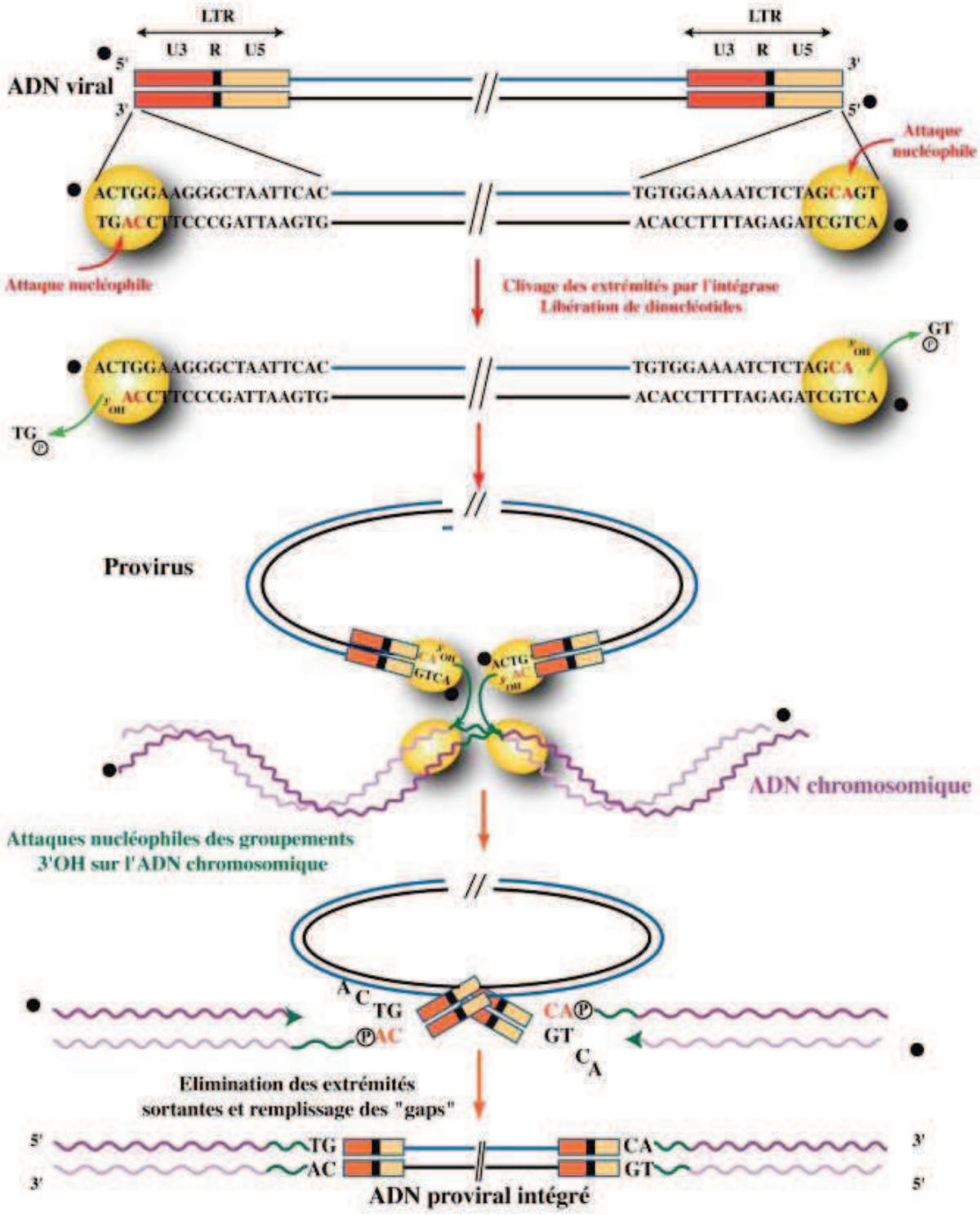


Figure 13 : Intégration de l'ADN viral dans l'ADN cellulaire (Thèse Gaelle Mercenne).

2. La phase post-intégrative

a. L'expression génique

Après l'intégration, le provirus sert de matrice pour la synthèse des ARN viraux qui vont permettre l'expression des protéines structurales, enzymatiques, régulatrices et accessoires. La transcription est divisée en 2 étapes et reposent sur les interactions entre l'ADN viral, les facteurs de transcription cellulaires et la protéine transactivatrice Tat. La région LTR en 5' permet l'initiation de la transcription et elle est composée de trois domaines fonctionnels : une région promotrice dans U3 (Unique, 3'end) qui contient une boite TATA, trois séquences consensus pour la liaison du facteur cellulaire SP1 et deux régions pour NF-κb. Ces sites de liaison permettent également le recrutement de l'ADN polymérase lors de la transcription. La deuxième région est R (Repeated) qui contient la région transactivatrice TAR impliquée dans l'activation de la transcription suite à son liaison avec la protéine Tat. Sans les facteurs de transcription cellulaires et la protéine Tat, le virus est incapable de produire des ARNm et entre dans un état de latence. La dernière région du LTR est U5 (Unique 5'end) (Fig. 14).

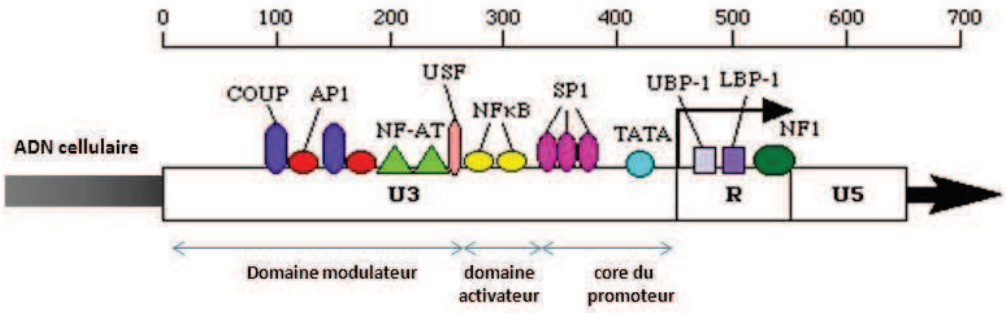


Figure 14 : Représentation du LTR 5' sur l'ADN proviral intégré dans le génome de l'hôte.

i. *La phase précoce*

Pendant la première phase (phase précoce), le provirus est transcrit par l'ARN polymérase II et produit des ARN courts non polyadénylés dû à la faible efficacité d'élongation de l'ARN polII mais persistant du fait de la présence de TAR formant une structure en tige boucle stable [34]. L'activation de la cellule permet la relocalisation de NF-κβ qui se fixe sur le domaine activateur du LTR 5'. Une histone est alors activée et le domaine carboxy-terminal (CTD) de l'ARN polymérase II est phosphorylé (29). Ceci entraine un remodelage de la chromatine conduisant à une augmentation du taux d'initiation et d'élongation de la transcription, aboutissant à une production d'ARN viraux multi-épissés qui codent pour les protéines Tat, Rev, Nef. A ce stade, les protéines Tat et Rev sont réimportées dans le noyau afin d'activer la transcription et permettre le transport des ARN génomiques (ARNg) et mono- épissés, respectivement (Fig. 15).

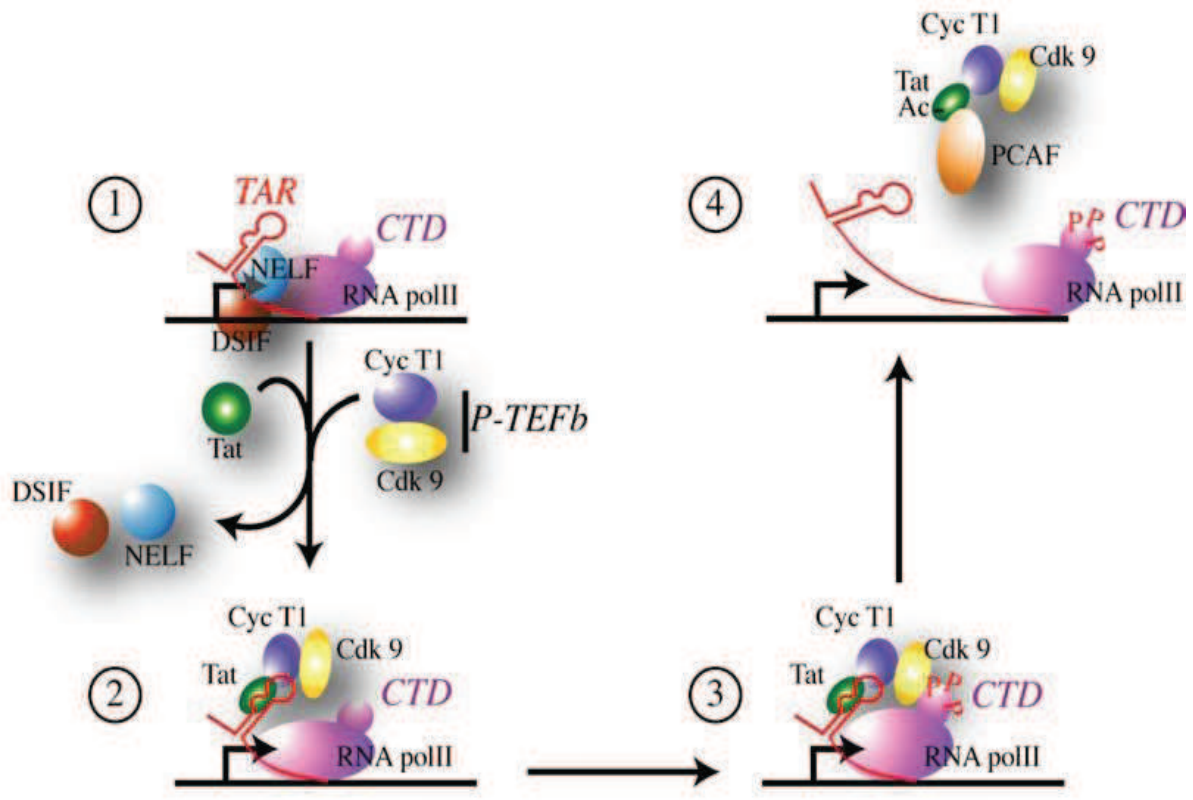


Figure 15 : Modèle de l'action trans-activatrice de Tat. 1-Hypophosphorylation du domaine C-terminal de l'ARN polymérase associée à des facteurs négatifs d'élongation de la transcription (NELF et DSIF) entrainant la synthèse de faible taux d'ARN messagers viraux ; 2- Tat se fixe sur la tige-boucle TAR en 5' de l'ARNv et recrute le complexe P-TEFb qui contient la cycline T1 et la kinase Cdk9 ; 3- Hyperphosphorylation du CTD de l'ARN polymérase II par Cdk9 ; 4- Acétylation de Tat et interaction entre Tat et PCAF (p300/CREB binding protein-associated factor) entrainent la dissociation du complexe Tat-TAR-CyclineT1-Cdk9 et par la suite une activation de la transcription et la production de protéines virales.

ii. *La phase tardive*

Durant cette étape, les ARNm multi-épissés (2 Kb) seront exportés du noyau vers le cytoplasme selon une voie endogène nucléaire tandis que les messagers viraux mono-épissés et non-épissés seront exportés par un mécanisme impliquant la protéine Rev [35, 36]. Lorsque la concentration de Rev est suffisante, elle se fixe au site RBE (Rev Binding Element) de la région RRE de ces ARNm, ce qui entraine une oligomérisation recouvrant les régions RRE. Le complexe Rev-RRE-ARN peut ainsi interagir avec la machinerie d'export nucléaire cellulaire. Les ARNm peuvent ainsi être traduits ou encapsidés , libérant Rev qui retourne dans le noyau grâce à son NLS (Nuclear Localization Signal). Les structures cristallographiques de la protéine Rev seule et en complexe avec le domaine RRE de l'ARNm ont été résolus récemment [37, 38] (Fig. 16).

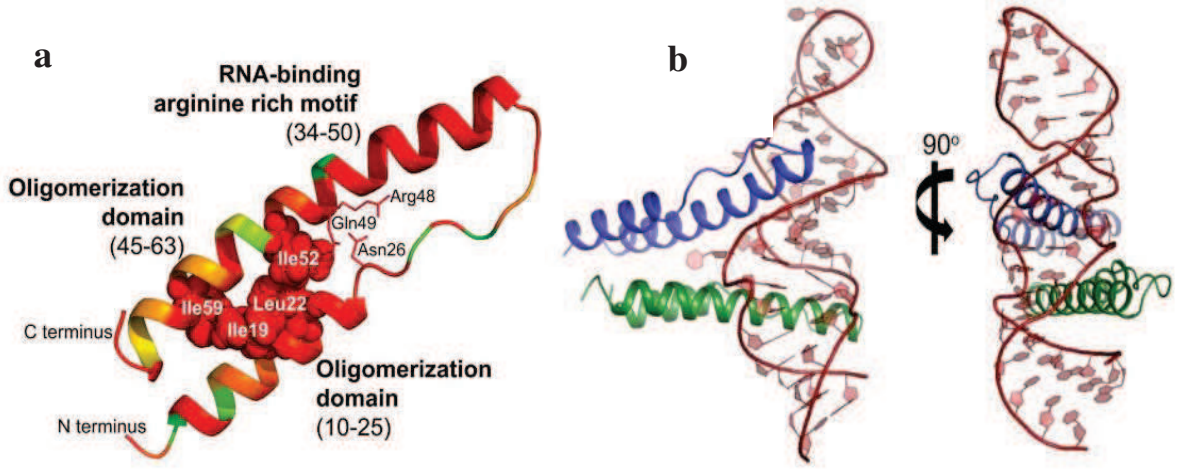


Figure 16 : (a) Représentation d'un monomère de Rev et des différents domaines fonctionnels [37]. (b) Structure du dimère de Rev en complexe avec l'ARN. L'ARN se situe entre la première molécule de Rev (bleu) et la deuxième (vert) [38].

b. La traduction

Une fois dans le cytoplasme, les ARNm viraux sont traduits par la machinerie cellulaire. Les protéines virales Tat, Rev, Nef, Vif et Vpr sont produites selon le mode de traduction classique. La synthèse des précurseurs protéiques Gag et Gag-pol se fait à partir de l'ARN non-épissé du VIH-1 alors que la synthèse des protéines d'enveloppe et Vpu se fait à partir de l'ARN mono-épissé mais dans les deux cas la traduction a lieu à partir du même transcrit. Cependant, la régulation de la traduction dans les deux cas est différente :

– La synthèse des protéines Env et Vpu est obtenue à partir d'un ARN bi-cistronique contenant deux ORF qui se suivent et se chevauchent. La synthèse de Env est permise grâce à un évènement de « Leaky scanning » permettant au ribosome de passer au travers du codon initiateur de Vpu [39, 40]. Vpu entraine la dégradation des CD4 piégés avec gp160 au niveau du réticulum endoplasmique ce qui permet le relargage des gp160 dans la membrane plasmique [41].

– La traduction de la polyprotéine Gag-pol implique un décalage du cadre de lecture en -1. La présence d'une séquence de glissement (UUUUUUA) associée à une tige boucle permet un décalage d'un nucléotide vers le 5' du cadre de lecture [42, 43]. Gag peut donc être traduite seul ou en fusion avec Pol sous forme de Gag-pol. Cet événement est vingt fois plus en faveur pour Gag seul, ce qui est critique pour la propagation du virus [44].

c. Assemblage

La formation d'une particule virale infectieuse requiert tout d'abord, le transport de Gag au niveau de la membrane plasmique et son association avec différents partenaires viraux et cellulaires. Suite à la synthèse des protéines virales, l'assemblage débute. Il a été démontré que l'assemblage de la particule virale se fait au niveau de la membrane plasmique et les endosomes tardifs (MVB) chez les macrophages [45, 46]. Par contre, il a été démontré que ce phénomène existait également chez les lymphocytes T CD4+ [47]. Plusieurs études ont démontré que le VIH-1 bourgeonne sélectivement aux radeaux lipidiques. En effet, la composition lipidique de l'enveloppe virale est enrichie de cholestérol et de sphingolipides [48]. De plus, il a été démontré que le cholestérol est important pour l'entrée virale et le bourgeonnement. L'acteur principal de l'assemblage est le précurseur de Gag, $Pr55^{Gag}$, capable de s'auto-assembler et de former des particules virales non infectieuse en l'absence d'autres protéines virales [49]. $Pr55^{gag}$ code les protéines MA, CA, NC, p6 et les peptides intermédiaires p1 et p2. Gag est composée de 3 domaines impliqués dans l'assemblage :

- Le domaine M (Membrane binding domain) situé à l'extrémité N-terminale est myristylé au niveau de la futur protéine de matrice MA ce qui permet l'ancrage du précurseur dans la membrane plasmique [50, 51]. Aussi, MA facilite l'incorporation de la protéine Env durant l'assemblage en interagissant avec son domaine TM [52, 53].
- Le domaine d'interaction I composé par les futures protéines CA et NC permet la multimérisation de Gag, important pour l'assemblage [54].
- Le domaine L (Late domain) situé en C-terminal et contient le peptide p6 requis pour l'encapsidation de la protéine Vpr [55, 56]. Une fois le virion mature, Vpr se lie à la NC [57].

Lors de l'assemblage du virus, deux copies homologues de d'ARN génomique doivent être encapsidées dans la particule virale sous forme dimérique. L'interaction spécifique entre le domaine NC de Gag et la région leader ou domaine psi localisé en 5' de l'ARN viral déclenche l'encapsidation de l'ARN [58, 59]. Cette interaction permet aussi la multimérisation de Gag induisant par la suite un changement de conformation de MA qui va exposer son groupement N-terminal myristylé [60]. Ce groupement ainsi que la présence d'une région basique permettent l'ancrage de Gag dans la bicouche lipidique de la membrane plasmique cellulaire en interagissant avec le phosphatidylinositol (4,5) bisphosphate [60].

Après glycosylation dans le réticulum endoplasmique, la polyprotéine Env ou gp160 est clivée en gp120 et gp41 dans l'appareil de Golgi, qui restent associés de manière non covalente à la membrane. Gp120 est bien ancrée à la membrane extracellulaire alors que gp41 reste transmembranaire avec une partie cytoplasmique. La protéine cellulaire TIP47 (Tail-Interacting protein of 47 KDa) sert d'adaptateur entre les protéines de l'enveloppe et Gag. TIP47 interagit avec le domaine CT de la MA [61] et la partie cytoplasmique de gp41 au niveau de la membrane plasmique afin de favoriser l'incorporation des protéines d'enveloppe dans les virions.

d. Bourgeonnement et maturation (Fig. 17)

Le virus détourne la voie cellulaire pour l'expulsion des exosomes, le système VPS (Vacuolar Protein Sorting) afin de libérer les virions. Dans cette étape, Gag a pour fonction de recruter les protéines cellulaires nécessaires à ce processus aux endosomes tardifs pour les rediriger ensuite vers la membrane plasmique. Ce système est composé des complexes ESCRT-1, 2 et 3 (Endosomal Sorting complex Recquired for Transport). Le bourgeonnement est initié par l'interaction entre le motif PTAP du domaine p6 de Gag suite à son interaction avec Tsg101 (Tumor Susceptibility gene 101) , une protéine du complexe ESCRT-1 [62]. La liaison avec les autres complexes ESCRT serait assurée par le contact entre Tsg101 et la protéine cellulaire Alix [63, 64] (Fig. 17).

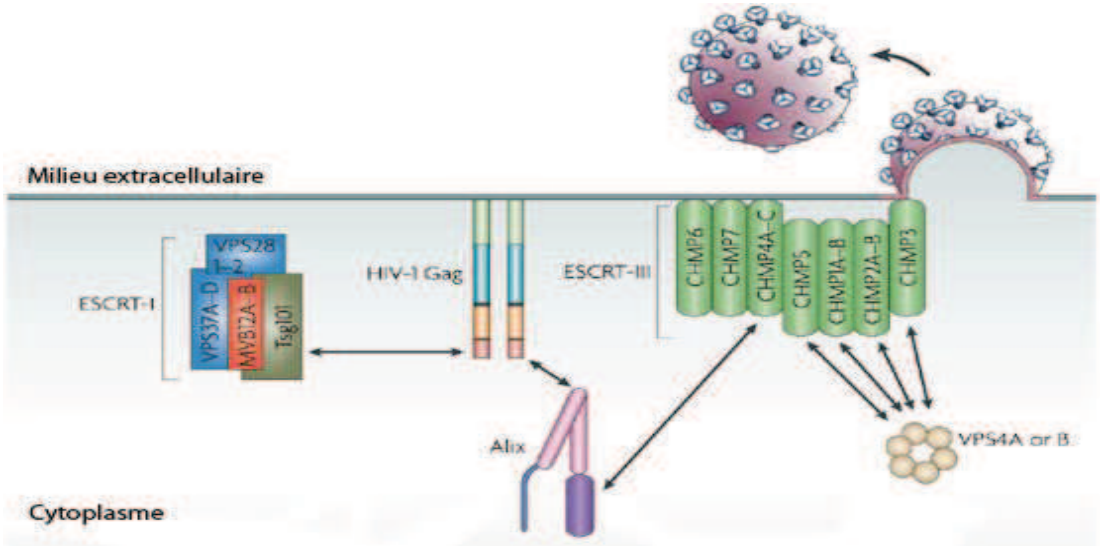


Figure 17 : Bourgeonnement de la particule virale de la membrane plasmique [65].

Lors de l'assemblage, les interactions entre précurseurs Gag-pol entrainent le rapprochement des domaines correspondant à la protéase qui vont dimériser pour former une protéase active (PR). PR va cliver tout d'abord entre les domaines p2 et NC du précurseur Gag-pol, puis dans un deuxième temps entre MA et CA pour terminer avec le clivage entre CA-p2 et NC-p1 libérant la NC et entrainant la condensation de NC et du génome d'ARN en une forme conique entourée par CA. Ces réarrangements sont nécessaires pour l'infectivité des virus.

Des images de microscopie électronique montrent deux types de particules virales différentes (Fig. 18). La particule virale immature présente, sous la bicouche lipidique une surface sphérique dense aux électrons.

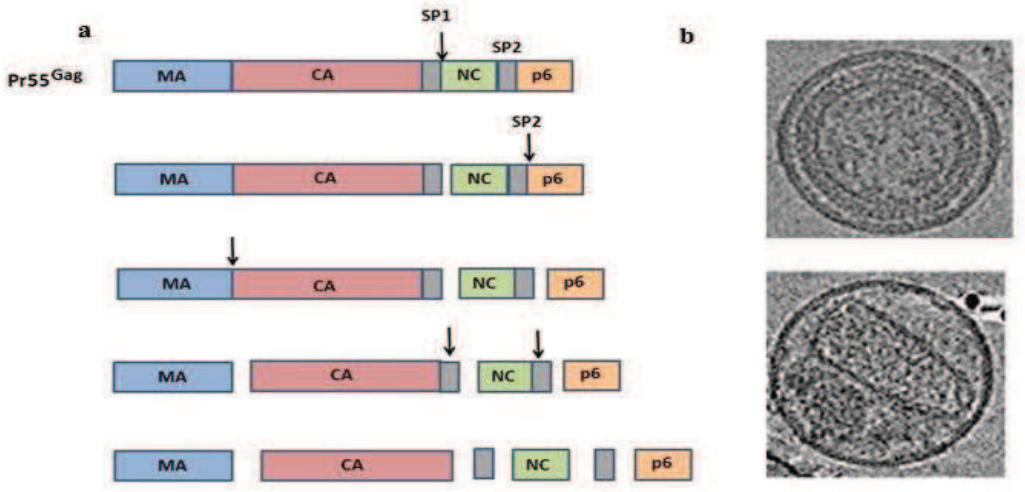


Figure 18 : (a) Représentation schématique de la maturation $Pr55^{Gag}$ suivant un ordre régulé. (b) Coupe transversale d'une particule virale immature (haut) et mature (bas) [66] .

C. Les traitements antirétroviraux

Actuellement, aucun traitement ne permet de guérir du SIDA. En 1987, l'AZT ou zidovudine a été le premier antirétroviral a été approuvé par la FDA (Food and Drug Administration). Son action porte sur l'inhibition de la RT virale. La plupart des antirétroviraux approuvés par FDA cible cette protéine.

Plus tard, d'autres antirétroviraux ont été mis au point visant l'entrée et la fusion, l'intégration et la maturation de la particule virale (Tableau 2). Le saquinavir est par exemple un inhibiteur de la protéase. Le problème majeur de ces antirétroviraux est l'apparition constante et relativement rapide de mutations de résistance du virus. Afin de limiter cet effet, ces médicaments sont administrés en combinaison (inhibiteur de RT et protéase) dans le but de maximiser la suppression de la réplication virale. Ce traitement est nommé HAARTS pour Highly active AntiRetroviral TherapieS et est le traitement le plus utilisé aujourd'hui. Il permet de ralentir la prolifération du VIH chez les personnes infectées. De plus, malgré le fait que les cellules infectées ne soient pas éradiquées par la prise d'antirétroviraux, le système immunitaire réagira par une augmentation du nombre de cellules LT4 ce qui permettra un contrôle de la réplication du VIH-1, une reconstitution partielle de l'immunité ce qui diminuera l'incidence des maladies opportunistes et du décès du patient. De plus en plus, les nouvelles études visent les facteurs de restriction cellulaires comme la protéine APOBEC3G

.

Inhibiteurs nucléosidiques de la transcriptase inverse (NRTI)	Inhibiteurs non nucléosidiques de la transcriptase inverse (NNRTI)
Retrovir(1987)	Nevirapine (1996)
Videx(1991)	Delaviridine(1997)
Hivid(1992)	Efavirenz(1998)
Tenofovir(2001)	Etravirine (2008)
Emtriva(2003)	Rilpivirine (2011)

Protéase	Intégrase	Entrée	Fusion
Saquinavir(1995)	Raltegravir(2009)	Maraviroc (2010)	Enfuvirtide(2003)
Indinavir (1996)	Elvitegravir(2013)		
Nelfinavir (1997)	Dolutégravir(2014)		
Darunavir (2006)			

Tableau 2 : Liste des antirétroviraux approuvés par la FDA, l'année de mise sur le marché et leur catégorie respective.

D. Stratégie vaccinale

Comme pour toutes les maladies infectieuses, la meilleure solution pour contrecarrer la propagation du SIDA serait le développement d'un vaccin. Jusqu'à aujourd'hui, plus de 250 essais cliniques ont été menés, la plupart en phase 1 et 2. Malgré toutes ces recherches et études cliniques encourageantes, aucun vaccin anti-VIH n'a encore été mis au point. La cause de cet échec est due aux mutations rapides des protéines virales de surface reconnues par les anticorps.

A l'heure actuelle, les approches privilégiées ont toujours visé la stimulation de la réponse immunitaire humorale [67]. Les vaccins induisant chez l'hôte des anticorps contre les protéines virales dont la gp120 comme le vaccin AidVax (Vax Gen) n'ont pas montré une protection efficace puisque des domaines structuraux essentiels aux protéines virales échappent à ces anticorps [68, 69]. Récemment, un essai clinique phase 1 d'un anticorps monoclonal 3BNC117 contre le site de CD4 chez des individus infectés et non infectés par le VIH-1 a montré des résultats très prometteurs [70].

D'autres études récentes ont recombiné un vaccin à ADN avec un virus de la vaccine recombinant (rTV) et un vaccin exprimant gp140. Ce vaccin a été testé sur des animaux et a montré qu'en recombinant ces 3 cibles, la réponse immunitaire humorale est renforcée [71].

La troisième stratégie appliquée dans ce domaine est l'induction d'immunité locale au niveau des muqueuses. Premièrement, cette réponse est très rapide, se déclenchant après à peine quelques minutes ou heures et deuxièmement cette stratégie permet de bloquer l'entrée du virus dans les cellules avant la primo-infection. Pour cela, un candidat-vaccin a été développé et testé sur des femelles macaque rhésus où il montre déjà de bons résultats [72].

Aujourd'hui, les chercheurs s'intéressent de plus en plus à l'immunité innée, sachant qu'elle représente la défense primaire contre l'infection [73]. Comprendre comment l'immunité innée régule les réponses immunitaires adaptatives serait d'une grande importance pour l'optimisation des candidats-vaccins [74]. Récemment, une jeune fille née séropositive est en rémission depuis 12 ans, alors qu'elle ne prend plus aucun traitement depuis ses 6 ans. L'explication à ce cas unique au monde est qu'elle a un système immunitaire qui a trouvé les moyens de contrôler son virus et d'éviter la multiplication. Cette rémission ne doit toutefois pas être assimilée à une guérison mais c'est un cas exceptionnel.

III. Immunité innée et facteurs de restriction virale

La réponse immunitaire innée est la première ligne de défense de l'organisme. Ainsi, le système adaptatif a le temps de développer une réponse effectrice. Cette réponse immunitaire innée est non spécifique du virus, immédiate et locale. Elle est médiée par des cellules immunitaires telles que les monocytes, macrophages, cellules dendritiques, et les cellules « Natural Killer » ainsi que des molécules comme l'interféron et les cytokines. Les stratégies et les techniques développées par les virus pour contenir la défense innée sont très variées.

En 2012, une analyse d'expression génique *in vitro* a montré qu'il existe 20,687 gènes codant des protéines dans le génome humain dont plusieurs centaines étant des ISG (interferon-stimulated genes) qui sont surexprimés durant des infections virales [75, 76]. Malgré que la fonction antivirale de certains ISG soit bien décrite, le rôle de la plupart d'eux durant les infections virales n'est pas encore bien connu. Parmi ces protéines, on retrouve des IFN (Induced by Interferon 1) tel que TRIM5α, téthérine , APOBEC3G, qui sont des joueurs clés dans la réponse immunitaire innée et représentent la première lignée de défense contre les lentivirus.

A. Les facteurs de restrictions virales

1. TRIM5α

a. Structure et fonction biologique

TRIM5α est une protéine codée par le gène TRIM5 et faisant partie de la famille des TRIM (TRipartite Interaction Motif). Elle a été identifiée dans des cellules de mammifères comme un facteur de restriction antirétroviral. Elle est caractérisée par la présence d'un motif en trois parties appelé RBCC, il contient un domaine RING (R), un domaine B-Box2 (B) et un domaine Coil-Coiled (CC) (Fig. 19a). Le domaine RING est un domaine de liaison au zinc impliqué dans l'interaction avec d'autres protéines plus particulièrement les protéines du complexe E3 ubiquitine ligase. Le domaine B-box est aussi un domaine de liaison au zinc médiant les interactions protéines–protéines. Le domaine Coiled-coil est responsable de la multimérisation de ces protéines [77] (Fig. 19b)

La majorité des gènes TRIM codent un domaine additionnel en C-terminal, le domaine SPRY, responsable de la spécificité de TRIM et de son implication dans la défense immunitaire innée contre le VIH-1 [78] Le gène TRIM5 code pour plusieurs isoformes grâce à un mécanisme d'épissage alternatif permettant le raccourcissement du domaine SPRY. TRIM5α est l'isoforme possédant le domaine SPRY le plus long capable d'interagir avec la CA du VIH-1 (Fig. 19c).

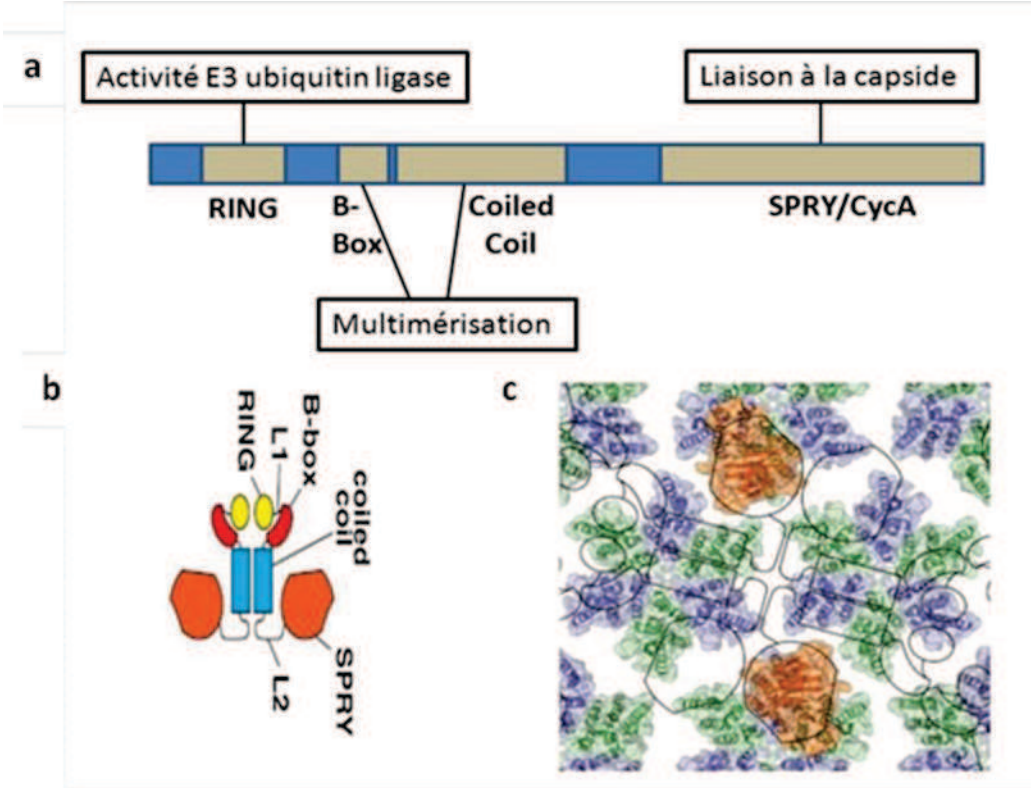


Figure 19 : (a) Représentation schématique des domaines de la protéine TRIM5α et leurs fonctions respectives connues. (b) TRIM5alpha forme un dimère, stabilisé par le domaine « Coiled-Coil ». (c) La structure du domaine SPRY de rhTRIM5α (orange) avec les interactions variables orientés vers la surface de la capside.

b. Action antirvirale de TRIM5α

On ne sait pas encore parfaitement comment TRIM5α contrecarre la réplication du VIH-1 mais de nombreuses études ont pu identifier deux voies primordiales entrainant cet effet. Lorsque la capside du VIH-1 pénètre dans le cytoplasme des cellules simiennes, TRIM5α reconnait la protéine de capside et bloque la réplication du virus dans la cellule hôte. Ceci entrainerait une décapsidation précoce de la capside [77, 79]. La découverte de 2 protéines de fusion entre TRIM5α et CypA chez les singes appuient l'idée que CypA intervient dans le mécanisme de restriction [80]. Mais, cette protéine semblerait seulement inhiber la réplication virale dans les cellules simiennes, alors que dans les cellules humaines , sa présence permet au virus d'effectuer un cycle viral normal [81].

Le domaine SPRY ou CypA de TRIM5α et TRIMCyp se lient de façon directe avec la capside un peu de temps suite à l'entrée du virus dans la cellule. Suite à cette interaction, le mécanisme de décapsidation de la capside survient. Ceci requiert l'activité de la E3 ubiquitine ligase et la machinerie de dégradation par le protéasome (Fig. 20).

TRIM5α est capable aussi, par son domaine RING, de recruter le complexe E3 ubiquitine ligase ce qui va activer la voie de signalisation de la réponse immunitaire innée par AP-1 et NF-kβ, entrainant ainsi une réponse inflammatoire à l'infection par le VIH-1.

Récemment, des études cliniques sur des macaques ont suggéré la thérapie génique comme approche contre la réplication virale. Cette cohorte a consisté à introduire une version modifiée du gène TRIM5α dans les cellules des macaques capable d'inhiber la réplication virale chez les macaques infectés par SIV [82].

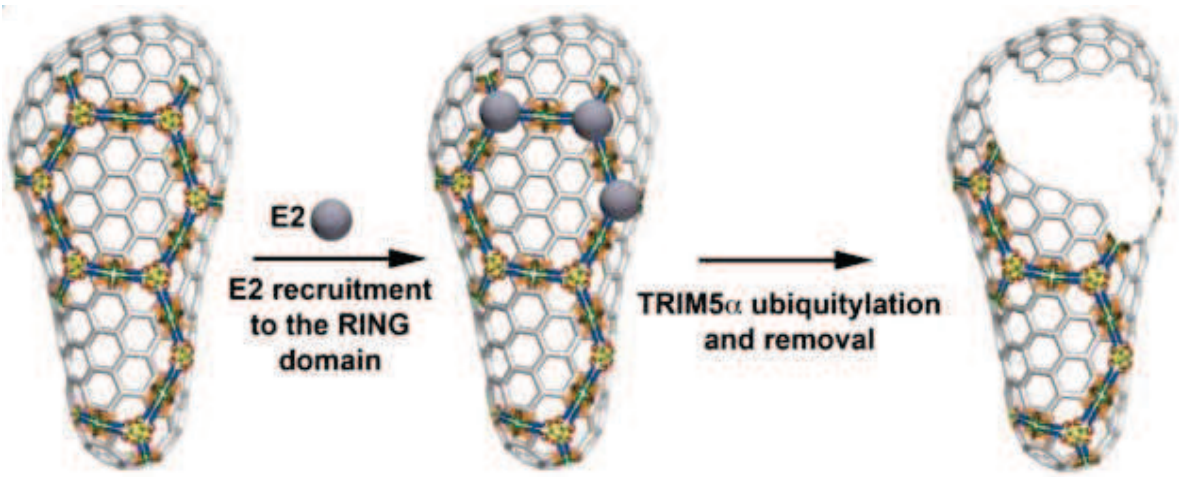


Figure 20 : Modèle actuel du rôle de TRIM5α sur la décapsidation précoce. D'abord, TRIM5α bloque la capside selon le modèle proposé par Ganser-Pornillos [83]. Le fait que le domaine RING est exposé à la surface de la capside, les enzymes E2 seront recrutées ce qui va polyubiquitiner TRIM5α. Le détachement de TRIM5α polyubiquitinilée de la surface de la capside pour sa dégradation entrainerait des dommages au niveau de la capside et une décapsidation précoce immature [84].

2. Téthérine

La téthérine a été identifiée d'abord comme BST2, (Bone marrow STromal antigen2) protéine responsable de la restriction du VIH-1 en absence de la protéine Vpu, en interférant avec le détachement et le bourgeonnement des particules virales matures des membranes des cellules infectées par le VIH-1 [85, 86]. En effet, Des mutations au niveau de Vpu entrainent une inhibition importante de la charge virale [87]. Récemment, il est devenu plus clair que la téthérine induit aussi l'expression des gènes proinflammatoire par la voie NF-κb-dépendante dans les cellules infectées par le virus [88]. La téthérine est une glycoprotéine transmembranaire de type II formée d'un domaine N-terminal court suivi par un site transmembranaire et un ectodomaine coiled-coil flexible ainsi qu'un site de clivage pour l'addition d'une ancre GPI (Glycosyl-PhosphatidylInositol) en C-terminal (Fig. 21). La téthérine

partage cette topologie structurale particulière et rare entre les protéines présentes chez mammifères avec une forme spéciale de la protéine prion qui est la ponticuline [89], Sm23 [90] et NcSRS2 [91].

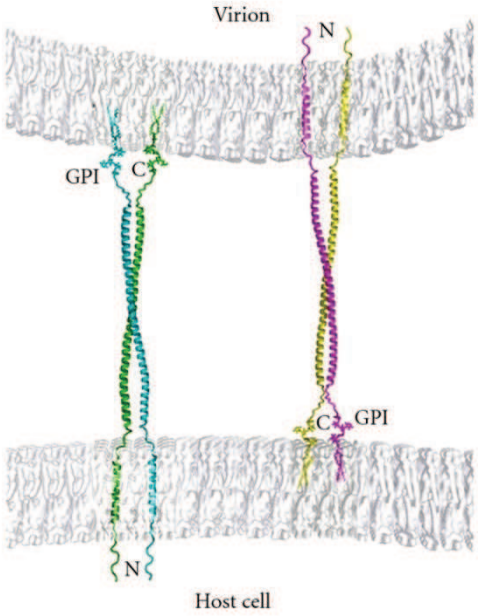


Figure 21: Représentation de la dimérisation de la téthérine. La téthérine forme un pont entre les 2 membranes. Elle contient un domaine transmembranaire en N-terminal et elle est ancrée à la membrane par un Linker GPI en C-terminal. La téthérine dimérise dû à la structure parallèle des domaines coiled-coil entre les extrémités de la protéine. Cet ancrage aux deux membranes entraine le blocage des particules virales et empêchent leur libération [92].

a. Action antivirale de la téthérine

La téthérine est exprimée dans différentes lignées cellulaires du fait qu'elle cible la bicouche lipidique virale. Son expression est induite ou régulée par les interférons de type I et II [85].

En absence de Vpu, la téthérine entraine la rétention des virions complet et matures à la surface des cellules infectées par le VIH-1 [86, 93]. Les virions s'accumulent à la surface cellulaire, dont une fraction subit l'endocytose via un mécanisme Clathrine-dépendant et par la suite seront dégradés [85]. Des études suggèrent qu'une extrémité de la téthérine est ancrée dans la membrane plasmique lipidique de la cellule et l'autre extrémité s'incorpore dans l'enveloppe de la particule virale, ce qui entraine l'homodimérisation de la surface cellulaire de la téthérine avec la téthérine associée au virion ; cette liaison est médiée par des ponts disulfide ou des régions coiled-coil dans le domaine extracellulaire [94]. De cette façon-là, les virions restent bloqués à la surface de la cellule infectée et liés l'un à l'autre par la téthérine (Fig. 22).

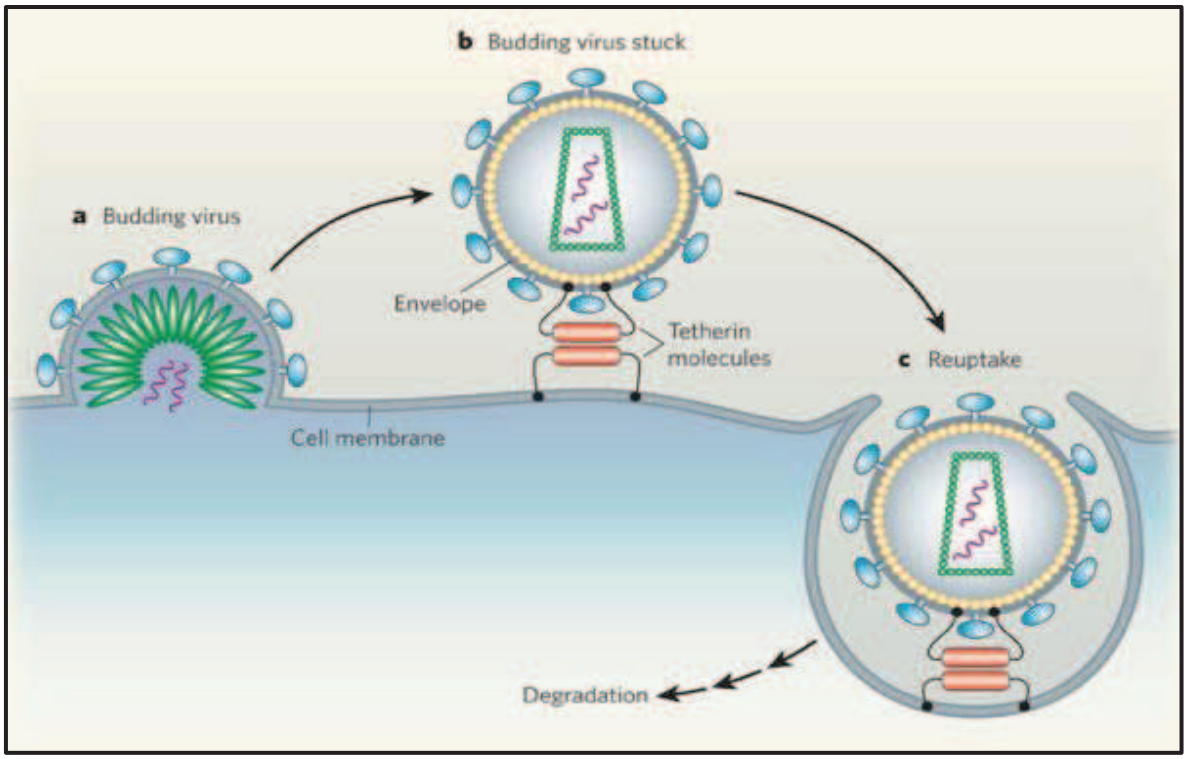


Figure 22: Restriction du VIH-1 par la téthérine. (a) Les virions VIH-1 s'assemblent à la surface cellulaire et bourgeonnent de la membrane cellulaire. (b) La téthérine est localisée à l'extérieur de la surface cellulaire, mais ancrée dans la membrane par ses deux extrémités. La téthérine s'attache au virion bourgeonnant en s'ancrant à son enveloppe virale empêchant par la suite sa libération de la surface de la cellule. (c) La tétherine se connecte à la machinerie cellulaire spécifique à l'endocytose par l'un de ces extrémités ce qui entraine l'endocytose des virions matures dans la cellule infectée pour finalement être dégradés par le système digestif cellulaire [95].

b. Vpu contrecarre téthérine

Vpu (Viral Protein U) est une phosphoprotéine membranaire de classe I de 16 KDa exprimée lors de la phase tardive du cycle rétroviral. Elle est absente chez le VIH-2 et les virus simiens où elle est remplacée par la protéine Nef concernant son activité anti-téthérine [96]. Vpu est composée d'un domaine trans-membranaire TM en hélice α et d'un domaine cytoplasmique (CT) dans lequel les 2 hélices α sont reliées par une boucle flexible.

Ces 2 protéines interagissent par leurs domaines transmembranaires. Vpu colocalise avec la téthérine et réduit son expression sur la surface cellulaire en induisant sa dégradation [93]. Un rôle très connu de Vpu est la dégradation du récepteur CD4 du VIH-1 dans le réticulum endoplasmique. Après formation du complexe Vpu/CD4, Vpu recrute la protéine βTrcP (β-transducing repeat-containing protein) , une sous-unité d'un complexe de la E3 ubiquitine ligase (SCF), ce qui va entrainer l'ubiquitination de CD4 et sa dégradation dans le cytoplasme par la voie du protéasome [97, 98]. Vpu est aussi capable de détourner la téthérine de sa voie d'acheminement normal du RE vers le cytoplasme en passant par

l'appareil de Golgi vers le compartiment intracellulaire par un mécanisme de séquestration [99]. Cette séquestration intracellulaire est suivie d'une dégradation lysosomale partielle de la téthérine et Vpu.

3. SAMHD1

a. Généralités

SAMHD1 (Sam and HD domain containing protein) est un facteur de restriction découvert récemment [100]. Cette protéine est très exprimée et fonctionnelle dans les cellules myéloïdes et dendritiques. Cette protéine est impliquée dans le syndrome Aicardi-Goutières (AGS) qui est une encéphalopathie caractérisée par une production d'IFN et des conséquences telles qu'une infection virale congénitale. Ceci peut être dû à un rétrovirus endogène n'ayant pas pu être inhibé par des mutants de SAMHD1 dans les cellules myéloïdes [101]. SAMHD1 est la seule protéine humaine qui possède un motif α (SAM) et un domaine HD en tandem. SAM est un domaine d'interaction protéine-protéine mais pouvant aussi lier les ARN. Le domaine HD est caractérisé par un motif à double cations coordonnant des résidus His et Asp. Ces domaines sont conservés durant l'évolution et jouent des rôles cruciaux dans le métabolisme des acides nucléiques (Fig. 23). SAMHD1 intervient dans l'étape précoce du cycle viral. Il a été décrit qu'une déplétion de SAMHD1 entraine une accumulation de l'ADN viral entier ce qui suppose que SAMH1 pourrait avoir un rôle durant l'étape de la transcription inverse [100].

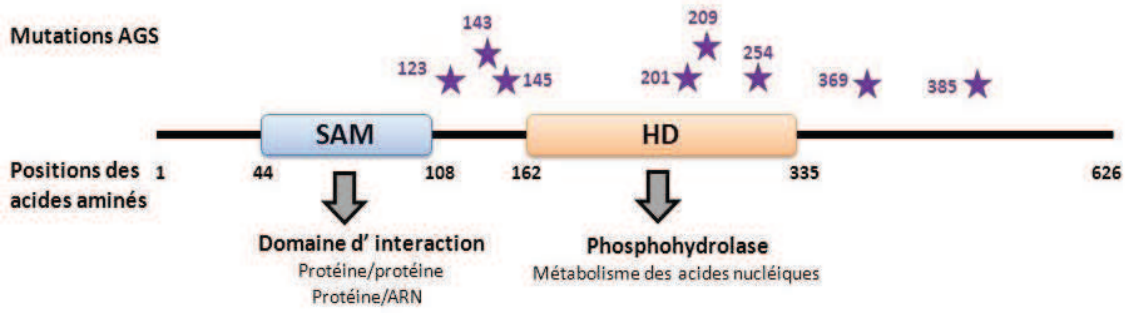


Figure 23 : Structure primaire de SAMHD1. SAMHD1 est une protéine de 626 acides aminés composée de deux domaines SAM et HD en tandem. Le domaine HD possède une activité phosphohydrolase. Les étoiles en violet représentent les acides aminés mutés dans le cas de l'AGS. (Adapté de [102])

b. L'action antivirale de SAMHD1

Contrairement aux autres facteurs de restriction, VIH-1 ne possède pas de moyen pour contrecarrer SAMHD1. Ce facteur de restriction a été identifié suite à des études montrant que certaines souches de lentivirus comme le VIH-2 ou le SIV codant la protéine auxiliaires Vpx , (absente dans le VIH-1) sont capables d'enlever le blocage de la réplication virale dans les cellules myéloïdes résistants à l'infection virale [103, 104]. Ainsi, Vpx a été utilisée pour identifier SAMHD1 par immuno-précipitation comme étant le facteur de restriction virale des cellules myéloïdes et dendritiques [105]. L'activité antivirale de SAMHD1 est liée à son activité phosphohydrolase, en hydrolysant les dNTP en deoxynucléosides et triphosphate inorganique par la voie dGTP-dépendante. Ce mécanisme va réduire la concentration cellulaire des dNTP sous le seuil nécessaire pour la transcription inverse créant un environnement cellulaire non favorable à la synthèse de l'ADN viral [100, 105, 106]. En effet, SAMHD1 a été montrée active dans les cellules dendritiques et les macrophages qui sont des cellules différenciées non divisibles ayant une concentration cellulaire de dNTP faible par rapport aux cellules prolifératives où SAMHD1 n'est pas efficace [107].

c. Vpx contrecarre SAMHD1

Vpx est une protéine nucléaire comme SAMHD1. Vpx du VIH-2 ou SIV est capable de contrecarrer SAMHD1 en induisant sa dégradation par le protéasome. Vpx stimule la formation du complexe de l'E3 ubiquitine ligase composé de DCAF1, DDB1, et CUL4. Ce complexe induira la polyubiquitination de SAMHD1 et sa dégradation par le protéasome [100, 105].

Dans le VIH-1, Vpr est exprimée et très proche du niveau structural et fonctionnel de Vpx du VIH-2. Une modélisation de ces 2 protéines montre une structure 3D très similaire ainsi qu'une homologie de séquences très élevée. Vpr interagit aussi avec les sous unités du complexe de l'E3 ubiquitine ligase mais n'induit pas la dégradation de SAMHD1. Par contre, cette interaction induit d'autres mécanismes comme le blocage du cycle cellulaire à la transition G2/M, ce qui n'est pas le cas de Vpx [108]. Vpr est aussi impliquée dans l'import nucléaire du complexe de préintégration viral et possède un impact sur l'expression des gènes viraux et cellulaires. Malgré le fait que Vpr ne dégrade pas SAMHD1, une étude récente a montré que la protéine ancestrale Vpr était capable de neutraliser SAMHD1 avant qu'elle n'évolue en Vpx dans VIH-2. Cette capacité à induire la dégradation de SAMHD1 a donc été perdue chez la protéine Vpr actuelle du VIH-1 [109].

Ces dernières années, les scientifiques ont approfondi leurs recherches sur les facteurs de restriction virales contre le VIH-1. Plusieurs facteurs de restriction autres que ceux mentionnées ci-dessus ont été découverts, bloquant différentes étapes de la réplication virale. Des études plus poussées

sont importantes pour comprendre mieux le mécanisme d'action de ces facteurs et leur rôle *in vivo* (Tableau 3).

Facteurs de restriction humains contre le VIH-1	Etape d'intervention durant le cycle viral	Virus contrecarre le facteur de restriction	Références
IFITM	Fusion	Inconnu	[110]
CNP	Assemblage	Inconnu	[111]
Viperin	bourgeonnement	Inconnu	[112]
MOV10	Transcription Inverse	Inconnu	[113, 114]
MxB	Intégration	Inconnu	[115]
RNaseL	ARN viral	Inconnu	[116]
ZAP	Synthèse dARNm viraux	Inconnu	[117]
SLFN11	Synthèse des protéines virales	Inconnu	[118]

Tableau 3 : Les différents facteurs de restriction contre le VIH-1 avec l'étape de la réplication virale à laquelle ils interviennent, la présence d'un facteur qui contrecarre leur effet et les références respectives.

4. APOBEC3G et contre mesure virale

a. APOBEC3: Structures et fonctions

APOBEC3G (Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G) appartient à la famille des APOBEC/AID composée de 11 membres dont le plus étudié est APOBEC1, responsable de l'édition de l'ARNm de l'apolipoprotéine B (ApoB), essentielle au métabolisme des lipides. Au cours de l'évolution, la famille des APOBEC a subit une expansion, la plus récente chez les primates a formé la famille des APOBEC3 en réponse à des anciens rétrovirus pathogènes [119]. Chez l'homme, sept représentants de cette sous-famille, APOBEC3, ont été identifiés: A3A, A3B, A3C, A3DE, A3F, A3G, A3H. Ces protéines représentent des facteurs de restriction contre la réplication des rétroéléments [120]. La diversité des A3 constituerait un moyen de réponse efficace à l'évolution rapide de ces rétroéléments. La famille des A3 est capable d'inhiber la réplication des rétrovirus comme le VIH-1 [121–124] et le virus de l'hépatite B [125].

Tous les membres de la famille APOBEC3 présentent un ou deux domaines CDA (Cytidine DeAminase). Seul le domaine C-terminal est catalytiquement actif, alors que le domaine N-terminal est impliqué dans différentes fonctions tel que la liaison à l'ARN, la multimérisation ou l'encapsidation (Fig. 24). A3B est une exception du fait qu'elle pourrait avoir deux domaines catalytiques actifs [126, 127]. Ces deux domaines ont une similarité de séquences très importante et portent un motif CDA conservé composé des résidus His-X-Glu-X_{23-28}-Pro-Cys-X_{2-4}-Cys capable de coordonner le Zn^{2+} indispensable à l'activité catalytique de la protéine [128]. Ce domaine permet l'activation d'une molécule d'eau afin de produire un résidu hydroxyle qui, par attaque nucléophile du groupement amine de la désoxycytidine, pourra induire la réaction de désamination des cytidines.

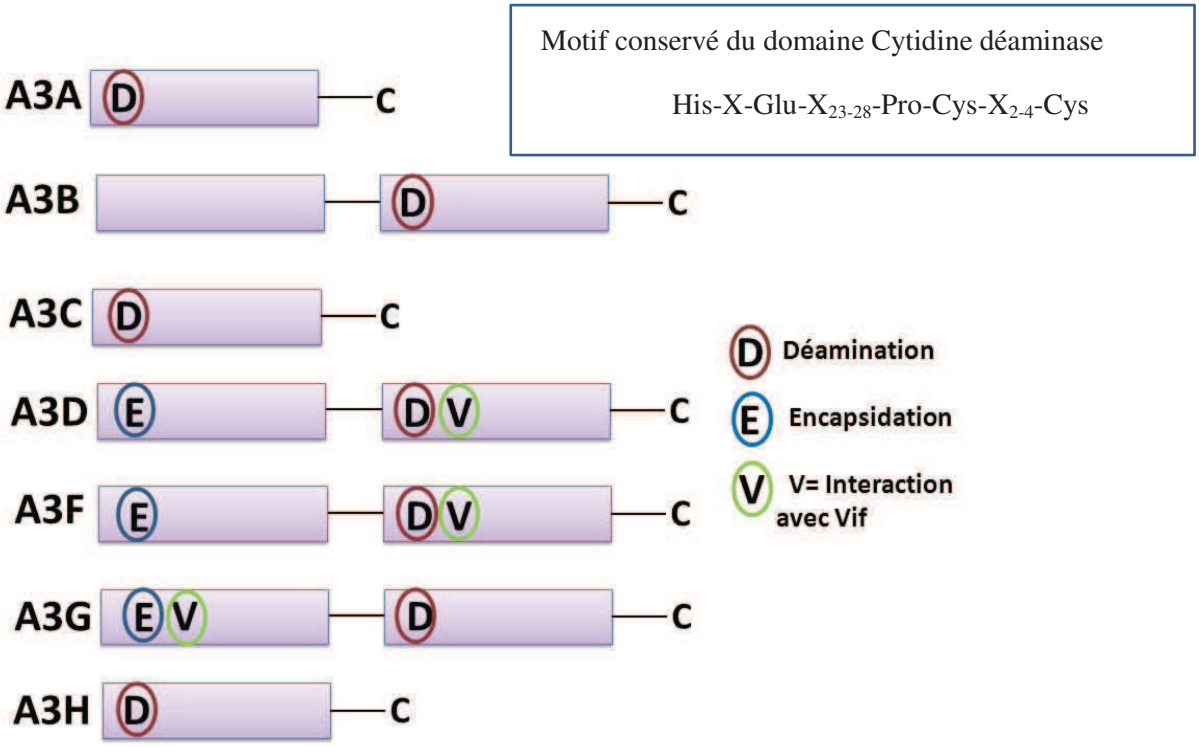


Figure 24 : Famille des APOBEC3 cytidines désaminases et leur organisation. La famille A3 est formée de un ou deux domaines cytidines désaminases. Le N-terminal (NTD) est responsable de l'encapsidation et le C-terminal (CTD) est responsable de l'activité de désamination. Les deux domaines sont capables de lier les acides nucléiques et se caractérisent par un motif à doigts de zinc. Le site d'interaction d'APOBEC3 avec Vif se trouve dans le NTD pour A3G et dans le CTD pour A3D et A3F.

APOBEC3G est une protéine de 46 KDa, composée de 384 acides aminés. A3G est principalement exprimée dans les cellules hématopoïétiques (Lymphocytes T CD4+, monocytes, macrophages, cellules dendritiques). Au niveau structural, on connait la structure tridimensionnelle d'APOBEC2 [129], la structure cristallographique d'APOBE3C [130], le CTD d'APOBEC3G obtenu par RMN d'abord puis par cristallographie à 2.3 A de résolution constituée d'un feuillet β à 5 brins entouré de 6 hélices α [131]. Récemment la structure du CTD d'APOBEC3F a aussi été résolue par

cristallographie. Le problème de solubilité de ce domaine a été résolue en remplaçant une partie du CTD d'A3F par une séquence du CTD d'A3G qui était soluble [132]. On ne connait pas encore les structures des domaines entiers des protéines APOBEC3G et APOBEC3F. Les structures mentionnées ci-dessus sont représentées dans la (Fig. 25).

Récemment, la structure en solution d'un variant du domaine CD1 d'APOBEC3G a été également déterminée par RMN [133] (Fig. 26).

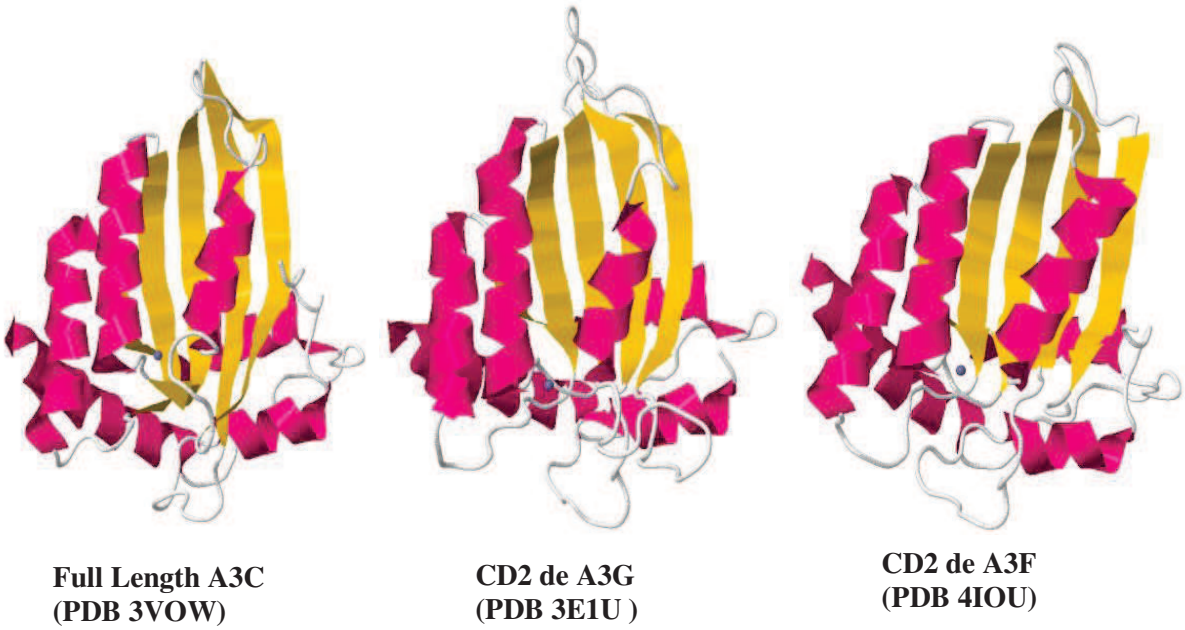


Figure 25 : Structures cristallographiques de la protéine complète APOBEC3C, des domaines CD2 d'APOBEC3G et APOBEC3F. Les hélices α sont représentées en rose et les brins β en cyan. Le zinc est représenté en sphère grise.

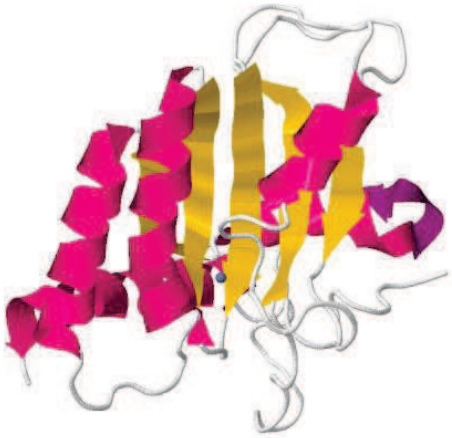


Figure 26 : Structure en solution déterminée par RMN d'un variant du domaine N-terminal d'A3G responsable de l'interaction avec Vif (2MZZ).

b. Rôle d'APOBEC3G/F dans la restriction virale

i. *Activité désaminase dépendante*

A3G est encapsidée dans les particules virales et agit comme facteur de restriction viral au niveau du cycle réplicatif suivant. Il a été montré par la suite que A3F possède lui aussi un rôle dans la restriction de la réplication virale [128]. L'activité d'A3F est indépendante d'A3G et cette protéine est exprimée dans plusieurs tissus humains. Cependant, A3G est la protéine de la famille des APOBEC3 ayant le pouvoir de restriction du VIH-1 le plus puissant [134].

En absence de Vif, A3G/A3F sont encapsidées dans les virions suite à leur interaction avec Gag et l'ARN. Parmi ces ARN on retrouve l'ARN7SL (qui fait partie du complexe de reconnaissance du peptide signal lors de la synthèse des protéines), l'ARN Y1 (petit ARN cytoplasmique) et l'ARN génomique. L'interaction avec l'ARN nécessite le domaine NC de Gag et l'importance de la liaison avec l'ARN a été montrée dans plusieurs études montrant la sensibilité des complexes Gag-A3 au traitement à la RNaseA [135]. Après que le virus soit entré dans la prochaine cellule cible, A3G/F interviennent dans la transcription inverse. Bien que A3G/F soient localisées dans le cytoplasme, une étape d'encapsidation est nécessaire pour la restriction du virus car elles ne sont pas capables d'atteindre le brin négatif de l'ADN viral d'un virus incomplet [136]. Ceci est probablement dû à la structure de la capside du VIH-1 ou au fait que les protéines A3 sont localisées dans des régions d' « ARN processing » comme les granules de stress ou P-bodies où elles ont un rôle dans la séquestration de l'ARN rétrotransposon humain pour empêcher la transposition [137]. Une fois dans la particule virale, A3G/F va induire des mutations dans l'ADN viral. Elle va transformer les cytosines en uraciles sur le premier brin d'ADN viral négatif lors de l'étape de la transcription inverse. L'hypothèse de base était que les uraciles sont alors reconnues et excisées par des enzymes cellulaires de réparation de l'ADN (UDG), générant des sites abasiques qui vont conduire à la coupure, puis à la dégradation de l'ADN viral [138], mais une étude récente a montré l'improbabilité de cette hypothèse , puisqu'une diminution du taux de l'ADNc viral est toujours observé même en absence d'UDG [139]. D'autre part, cette substitution dG→dA dans le génome viral pourrait aboutir à une accumulation d'hypermutations ce qui représente un effet négatif sur la réplication du VIH-1 en altérant les cadres ouverts de lecture et/ou introduisant des codons stop prématurés, conduisant à la synthèse de protéines aberrantes [138, 140]. Le provirus n'est alors plus capable de coder pour des virions infectieux (Fig. 27).

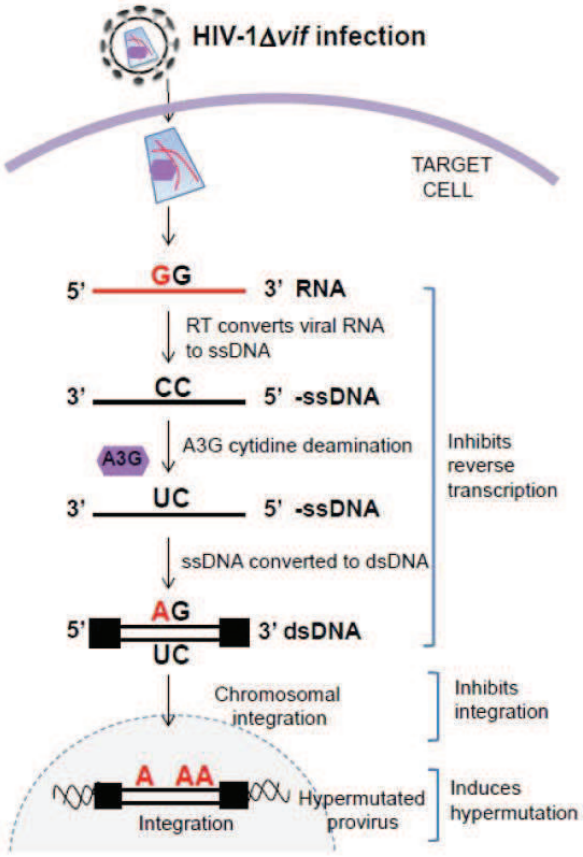


Figure 27 : Activité désaminase-dépendante d'A3G/F. En absence de Vif, A3G est encapsidée dans les virions bourgeonnants et agit comme inhibiteur efficace de la réplication du VIH-1 en désaminant les dC dans le brin ADN (-) provoquant des hypermutations de dG en dA dans le brin d'ADN proviral (+) et la destruction de l'ADN [141].

ii. *Activité désaminase indépendante*

Principalement, A3G inhibe la réplication du VIH-1 par son activité désaminase. Pourtant, plusieurs études ont montré la capacité d'A3G d'inhiber d'autres processus comme la transcription inverse [139, 142] ou l'intégration de l'AND proviral (116). En effet, l'expression du domaine N-terminal qui est catalytiquement inactif seul ou bien l'expression des mutants de A3G délétés du domaine catalytiquement actif est suffisante pour contrecarrer la réplication virale [143]. Le mécanisme initial proposé était qu'A3G interagit avec l'ARN génomique viral , empêchant la RT de s'y lier [142]. Mais ce mécanisme est improbable vu que seulement 3-11 molécules d'A3G sont encapsidées alors qu'il y a approximativement 100 molécules de RT, ce qui suggère que ces observations d'inhibition de la réplication virale par l'activité désaminase indépendante d'A3G ne sont que des artefacts. En effet, une seule molécule d'A3G inactive un provirus du VIH-1 par son activité cytosine désaminase alors que le mécanisme désaminase indépendant dépend beaucoup plus de la concentration d'A3G [144]. Plusieurs études ont montré plus tard l'importance de cette activité dans la réplication du VIH-1 [139, 145] indépendamment de l'activité désaminase dépendante.

c. Rôle de la famille des APOBEC3 dans d'autres rétrovirus

Les effets antiviraux d'APOBEC3G ne sont pas limités au VIH-1 mais ciblent également d'autres rétrovirus tel que le MLV (Murine Leukemia Virus), le MMTV (Mouse Mammary Tumor Virus), le SIV et le EIAV [122, 146, 147] (Tableau 4). De plus, l'expression d'A3G bloque la réplication du virus de l'hépatite B, un hepadnavirus dont le cycle de vie implique la rétrotranscription d'un ARN prégénomique [148]. L'encapsidation d'APOBEC3G dans de tels virus différents et distincts suggère que son encapsidation dans les virus est probablement un processus non spécifique ou bien qu'elle utilise des signaux partagés par ces virus.

	Rétrovirus				Rétrotransposons			Autres virus
APOBEC	VIH-1	SIV	MLV	EIAV	IAP	Ty1	LINE-1	HBV
APO1	Non		Non					Oui
APO2	Non	Non						
APO3A	Non		Non		Oui		Oui	
APO3B	Oui	Oui	Oui		Oui		Oui	Oui
APO3C	Faible	Oui	Non		Oui Faible	Oui	Oui Faible	Oui
APO3DE	Oui	Oui	Non					
APO3F	Oui	Oui	Non		Oui Faible	Oui	Oui/ Non	Oui
APO3G	Oui	Oui	Oui	Oui	Oui Faible	Oui	Non	Oui
APO3H	Non/ oui	Non	Non				Non	
APO4	Non							
AID	Non	Non	Non					Oui

Tableau 4 : Action des protéines de la famille APOBEC contre les rétroéléments. « Oui » indique une inhibition de la réplication du rétroélément, « Non » correspond à une non inhibition, «Oui/non » représente des données dans des références différentes, « Faible » indique une inhibition modérée, les cases vides correspondent à une information non déterminée. Pour les virus VIH-1 et SIV, l'inhibition est en plus qualifiée comme sensible à Vif (rouge) et insensible à Vif (Bleu).

d. La protéine virale Vif contrecarre APOBEC

i. *Caractéristiques générales*

Vif (Virion Infectivity Factor) est une petite protéine de 192 acides aminés (23 KDa) hautement basique (pI=10,7), exprimée en grande quantité dans la phase tardive du cycle viral à partir d'un ARNm monoépissé. Elle est conservée chez tous les lentivirus, excepté chez le virus de l'anémie infectieuse du cheval (EIAV) [149] . La protéine de Vif des différentes souches proches du VIH-1 est très conservée (89% d'identité entre les isolats HXB2 et NL4-3 du VIH-1). Même entre les souches un peu plus éloignées comme le VIH-1 et SIV, on retrouve un maximum de 30% de différence entre les séquences [150]. L'alignement de séquences de Vif entre le VIH-1, VIH-2 et SIV montre des motifs très conservés comme le N-terminal riche en tryptophane, le motif HCCH, ainsi que le motif SOCS-box.

La protéine Vif joue un rôle important dans le cycle viral. Elle est encapsidée dans les virions grâce à son interaction avec l'ARN génomique viral et le domaine NC du précurseur Gag et contrecarre l'action antivirale des protéines APOBEC3G/F d'une façon qui sera détaillée par la suite. L'interaction Vif-Gag requiert le domaine C-terminal de Vif [151] alors que la liaison à l'ARN se fait grâce à son domaine N-terminal. Vif se lie à l'ARN génomique du VIH-1 dans le cytoplasme des cellules infectées via son domaine N-terminal [152]. Au laboratoire nous avons montré que cette interaction a lieu préférentiellement sur la partie 5' de l'ARN génomique et il existe des sites primaires de forte affinité à Vif. Parmi ces sites, on retrouve DIS (Dimerisation Iniation Site) et TAR (TransActivation Responsive region), deux régions très conservées formant des tige-boucles, respectivement au milieu et à l'extrémité de la région 5' non traduite (5'UTR) de l'ARN génomique du VIH-1 [153, 154]. (Article joint).

Deux autres domaines sont aussi importants: le domaine de liaison à la Cullin5 qui possède un motif HCCH de coordination du zinc, ainsi que le domaine SOCS qui renferme à la fois un motif de liaison à l'ElonginC et un motif PPLP de dimérisation [149]. Le domaine N-terminal de Vif est responsable de l'interaction avec les protéines A3G/F (Fig. 28).

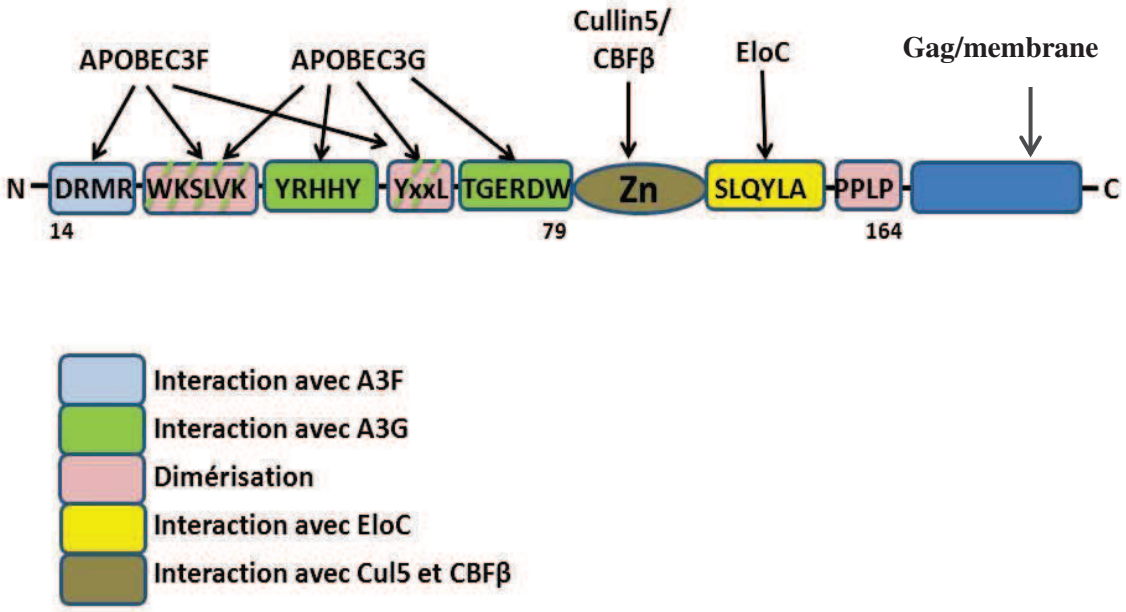


Figure 28 : Représentation schématique des différents domaines d'interaction de Vif du VIH-1 et leurs motifs d'interactions avec les différentes protéines.

ii. *La structure 3D de Vif dans le complexe E3 ubiquitine ligase*

Pendant longtemps, la structure de Vif était encore inconnue, essentiellement du fait de sa faible solubilité qui la rendait difficile à purifier en conditions natives [155]. En 2012, il a été découvert que Vif interagit avec le cofacteur CBFβ afin de la stabiliser dans les cellules [156–158]. Cette interaction avec CBFβ est médiée par les acides aminés ^{84}GxSIEW^{89} et ^{102}LADQLI^{107} de Vif [159]. Le complexe Vif/CBFβ est indispensable pour une meilleur stabilité *in vitro* de Vif et permet donc l'expression recombinante de Vif soluble en large quantité dans *E.coli* qui peut être purifiée par la suite pour des études biochimiques [158].

Afin d'inhiber l'activité d'APOBEC3G/F, Vif interagit directement avec APOBEC3 et recrute le complexe E3 ubiquitine ligase composé de CBFβ, EloB, EloC, Culline5-Rbx2. Vif interagit avec l'Elongin C (EloC) qui forme un hétérodimère avec l'Elongin B (EloB) et la cullin5 (Cul5) [160, 161]. L'interaction de Vif avec CBFβ augmente la stabilité de Vif dans les cellules et *in vitro* ainsi qu'elle promeut la recrutement d'EloB/EloC [159]. Le motif SLQ présent dans Vif est responsable de l'interaction avec EloC. Le motif ^{108}HCCH^{139} est composé de deux His-Cys conservés au sein d'une hélice α formée d'un groupe de résidus hydrophobes conservés et se lie au zinc. Il a été observé que les mutants F115A, I120S et AL123/124SS ne peuvent plus interagir avec la Cullin5 ce qui suggère l'importance de ce motif dans l'interaction avec la Cullin5 (133). Le complexe Vif/CBFβ/ELoB/EloC subit un changement conformationnel qui promeut l'interaction à la Cul5 ce qui suggère un ordre bien particulier dans l'assemblage du complexe de la E3 CRL5 ligase [162].

Récemment, la structure cristallographique de Vif au sein du complexe E3 ubiquitine ligase (CBFβ/EloB/EloC/Cul5) a été résolue [163] (Fig. 29). Elle montre que Vif possède une structure conique allongée et contient deux domaines avec un motif à doigt de zinc dans le centre des deux domaines. CBFβ se lie au domaine α-β en N-terminal et EloC et Cul5 se lient au domaine α du C-terminal de Vif. EloC ainsi que la Cul5 interagissent avec les interfaces hydrophobes de Vif dans des hélices α distinctes. La structure cristallographique montre la stabilité de l'interaction entre Vif et CBFβ puisque l'aire de la surface d'interaction totale est de 4797 A^2 et forment un feuillet β antiparallèle avec un brin β de chaque protéine. CBFβ se lie à Vif avec la même face d'interaction que lorsqu'il se lie au facteur de transcription RUNX1, son partenaire naturel suggérant une liaison mutuelle exclusive , bien que plusieurs études avaient suggéré que CBFβ lie Vif et RUNX1 sur deux surfaces distinctes [164, 165]. La région hydrophobe du domaine α-β de Vif est responsable de l'interaction avec CBFβ mais la région exposée est très positivement chargée et pourrait être responsable de l'interaction avec les protéines APOBEC3. La structure du complexe avec la protéine APOBEC n'est pas encore connue et ceci représente un objet d'étude principale et important du point de vue pharmacologique et thérapeutique.

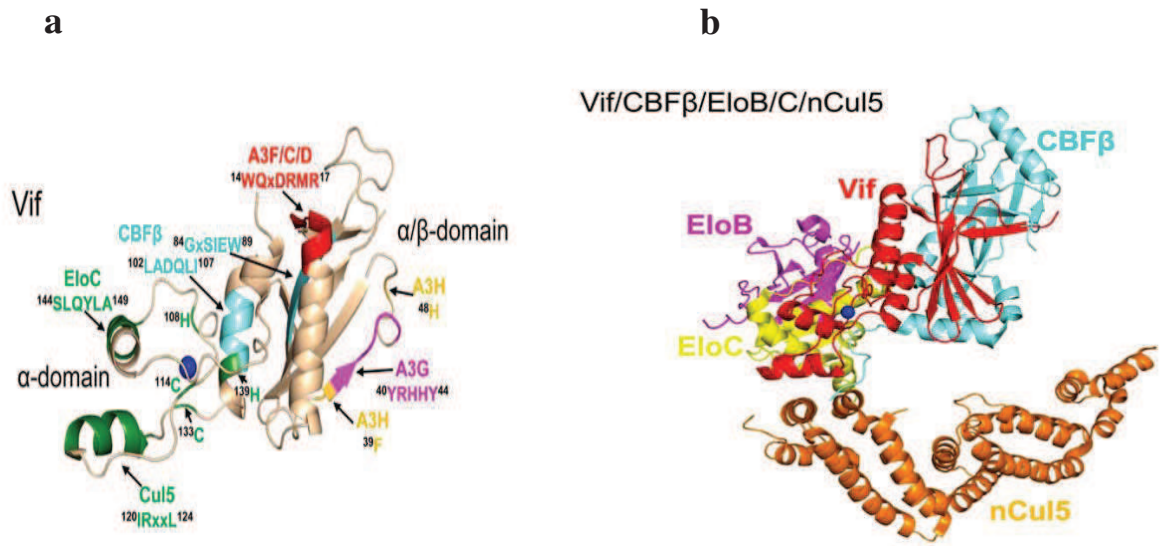


Figure 29 : (a) La structure cristallographique de Vif (PDB : 4N9F) montre qu'elle a deux domaines des deux côtés du Zinc (bleu). Le N-terminal est composé de cinq feuillets β, un brin β discontinu et trois hélices α. Ce domaine α-β comporte l'interface de liaison à CBFβ et les protéines A3. Le C-terminal de vif est structuré en deux hélices α responsables de l'interaction avec EloC(vert) et Cul5 (vert). (b) Structure de Vif en complexe avec CBFβ (cyan), EloC (jaune), le domaine N-terminal de la Cul5 (orange). EloB (magenta) dimérise avec EloC [166].

iii. *Vif et contre mesure virale*

Vif diminue de manière considérable le niveau de A3G et de A3F dans les particules virales , limitant ainsi leur action antivirale [138]. La diminution du taux d'APOBEC3G en présence de Vif a été démontrée par cytométrie de flux sur des cellules exprimant une protéine de fusion APOBEC3G-GFP (Green fluorescent Protein) et d'autres techniques biochimiques [167]. A l'heure actuelle, deux mécanismes sont proposés :

- ***Dégradation d'APOBEC3G dépendante du protéasome***

L'activité antivirale d'APOBEC3G est inhibée par Vif permettant au virus de se répliquer dans la cellule hôte. Vif cible également d'autres APOBEC malgré la différence dans leur sensibilité à Vif et il existe une spécificité significative liée à l'espèce qui permet à Vif provenant d'un certain virus d'attaquer APOBEC d'une espèce hôte particulière et non pas une autre (Fig. 30). Les différentes relations entre Vif et APOBEC sont toujours sujettes à différentes hypothèses mais il y a un accord général que l'inhibition de l'activité antivirale d'APOBEC3G par Vif est médiée par une interaction physique avec APOBEC3G qui résulte de l'exclusion de la désaminase des virions. Les motifs d'interaction de Vif dans différentes APOBEC sont résumés dans le tableau 5. Cet effet de Vif est accompagné d'une réduction du taux d'APOBEC3G dans les cellules T infectées par le VIH-1 alors que le taux de son messager n'est pas affecté. APOBEC3G est une protéine stable, sa demi-vie est estimée à plus que huit heures dans des cellules HeLa ou 293T transfectées et Vif réduit sa demi-vie entre 5min et 4 heures suivant le protocole expérimental utilisé. Ce changement au niveau de la stabilité d'APOBEC3G a été attribué à sa dégradation par la machinerie du protéasome ubiquitine-dépendante [167–170]. Ceci est confirmé par le fait que des inhibiteurs de protéasome (MG132 ou lactacystine) stabilisent le niveau intracellulaire de hA3G et augmentent son effet antiviral [170, 171]. Ce mécanisme a été largement étudié et il est relativement bien compris aujourd'hui (Fig. 30). Plusieurs domaines dans Vif sont impliqués dans le recrutement du complexe de l'E3 ubiquitine ligase et ont été détaillés en haut (Fig. 29). Des études ont suggéré que l'interaction de Vif-Cullin5/EloB/EloC/Rbx1 en complexe avec APOBEC3G accélère la polyubiquitination de la désaminase et conduit à sa destruction par le protéasome 26S [168–170, 172]. La capacité de Vif d'induire la polyubiquitination d'APOBEC3G a été prouvée par des études *in vitro* dans lesquels Vif, coexprimée avec la Cullin5, EloB, EloC et Rbx1 s'assemblent tous ensemble en un complexe E3 ubiquitine ligase fonctionnel et induit la polyubiquitination de la protéine APOBEC3G immunopurifiée *in vitro* [173]. D'un autre côté, Vif est une protéine instable (30 minutes) et est elle-même aussi dégradée par des protéasomes cellulaires mais, il a été observé que les délétions des motifs SLQ et HCCH impliqués dans les interactions avec les protéines de la E3 ubiquitine ligase induisent la dégradation d'APOBEC3G mais n'augmentent pas la stabilité de Vif ou empêchent sa

polyubiquitination (144,147). Ces résultats suggèrent que Vif n'est pas dégradée via le complexe de la Cullin5 E3 ubiquitine ligase comme APOBEC3G ce qui élimine l'hypothèse que Vif et APOBEC3G soient co-dégradées par ce système. Le mécanisme de dégradation de Vif est toujours inconnu.

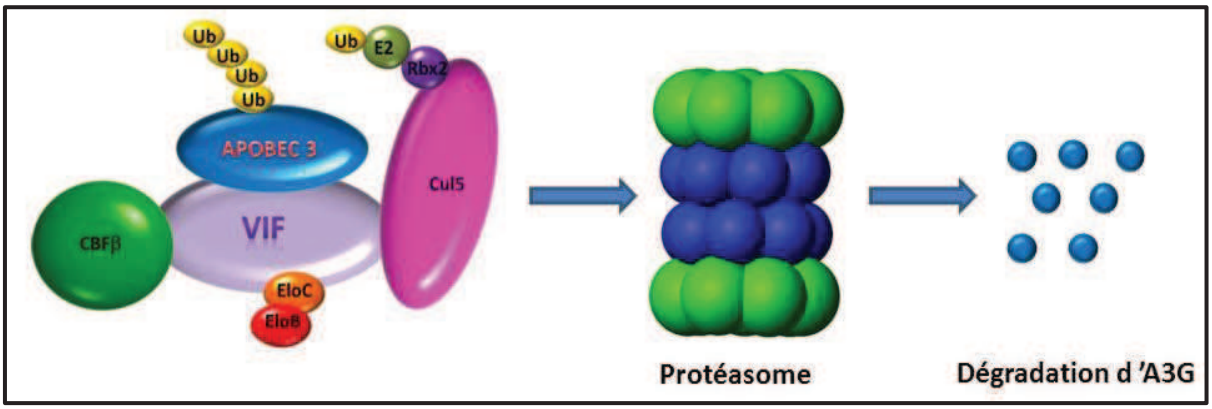


Figure 30 : Polyubiquitination et dégradation d'APOBEC3 par la protéine Vif . Représentation schématique de tous les composants de ce mécanisme de dégradation d'A3G. Il y a une interaction directe entre CUL5 et EloC ainsi qu'entre CUL5 et Rbx2. Vif s'associe au complexe CUL5-ELoB-ELoC-Ebx2-E2 par une interaction directe entre son motif HCCH et la Cul5 et entre le motif SOCS box et EloC. A3G est recrutée à ce complexe par Vif, par la suite A3G est polyubiquitinée et dégradée dans le protéasome 26S. CBFβ, un cofacteur de Vif, a été démontré important pour stabiliser les interactions Vif-Cul5 et inhiber l'activité antivirale d'A3G dans l'infection par le VIH-1

APOBEC3	Motifs d'interaction avec Vif	Références
APOBEC3C	I72, ^{75}FCxxILS^{81}, Y86, ^{106}EF^{107}, H111, P129, E141	[130]
A3D-CTD	L268, ^{271}FC^{272}, ^{275}ILS^{277}, Y282, ^{302}EF^{303}, H307, E337	[130]
A3F-CTD	L255, ^{258}FC^{259}, S264, Y269, F290,H294, E286, E289, E316, S320, E324	(105, 139,140)
A3G-NTD	^{128}DPD^{130}	[176–179]
A3H haplotype II	E121	[180, 181]

Tableau 5 : Les motifs d'interaction de la protéine Vif dans les différents APOBEC3.

- ***Inhibition d'APOBEC3G indépendante de la dégradation***

D'autres mécanismes ne mettant pas en jeu le protéasome existent pour inhiber l'activité antivirale d'APOBEC3G. L'identification d'une forme d'APOBEC3G résistante à la dégradation a pu confirmer cette hypothèse [182]. Malgré son caractère résistant à la dégradation, cette forme d'APOBEC3G a bien été encapsidée dans des virions de VIH-1 déficient en Vif et possédait des propriétés antivirales. Mais, Vif empêchait l'encapsidation de ce variant d'APOBEC3G et restaurait l'infectivité virale. Hormis la dégradation dépendante du protéasome, plusieurs mécanismes peuvent avoir lieu pour inhiber APOBEC3G. Ainsi, Vif associée au RNP viral (RiboNucléoProtéin), interagit avec A3G et empêche son encapsidation dans les virions [140]. Vif inhibe également l'activité d'A3G dans les virions et ceci est lié à la capacité de Vif à s'associer aux acides nucléiques au cours de l'étape de rétrotranscription, inhibant ainsi l'activité désaminase de hA3G en limitant son accès au brin d'ADN viral nouvellement synthétisé [151].

Enfin, Vif pourrait inhiber la synthèse d'A3G au niveau traductionnel. L'équipe de Jean Christophe Paillart et Roland Marquet dans notre laboratoire a pu démontrer que Vif se lie à l'ARNm d'A3G et inhibe sa traduction *in vitro* [183]. En effet, des expériences de rétention sur filtre et de fluorescence ont pu montrer que Vif se lie à l'ARNm d'A3G avec une forte affinité (Kd=120 ±10nM). L'affinité de Vif pour le 3'UTR est plus importante que pour le 5'UTR.

e. Rôle d'APOBEC3G/Vif dans le traitement antirétroviral

La résistance aux traitements contre le VIH est un déterminant majeur dans l'évolution virale durant la maladie. A ce jour, une centaine de sites de mutations impliqués dans la résistance aux traitements a été identifiée dans le génome du VIH [184]. Ces sites de mutations fonctionnent directement en modifiant les cibles des antirétroviraux ou indirectement en modifiant les voies ce qui contribue à cette résistance. Plusieurs de ces mutations ont été trouvés dans des cibles potentielles à l'activité d'A3G.

Une étude bioinformatique a évalué la probabilité des mutations causées par A3G au niveau des sites connus dans la résistance aux inhibiteurs, prenant en considération les mutations induites par A3G dans tout le génome viral, sur les deux brins d'ADN. Sur les 52,000 mutations de G en A, seulement 695 (1.3%) sont localisées au niveau d'un site de résistance aux inhibiteurs [185]. Dans ce contexte, les investigateurs ont conclu qu'il existe une corrélation relativement modeste entre l'activité d'A3G et la génération des mutations aux antirétroviraux si on prend en compte l'effet global d'A3G sur le génome viral [185, 186]. Néanmoins, des évidences expérimentales récentes montrent l'implication d'A3G dans la génération de ces mutations. Par exemple, la mutation très commune M1841(V) dans la RT causant la résistance aux inhibiteurs 3TC, ABC et DDI est localisée au niveau

d'une cible d'A3G (TCCAT en TCUAT) et est produite par A3G *in vitro* durant la réplication du VIH-1 dans des systèmes de culture cellulaire [187]. Ceci représente un exemple du rôle parallèle d'A3G en aidant simultanément l'hôte et le virus : de la même manière avec laquelle il agit comme facteur d'immunité innée préexistant qui fortifie la défense de l'hôte après exposition virale, A3G augmente la capacité inhérente du VIH-1 à acquérir la résistance même avant traitement aux inhibiteurs.

Plusieurs pistes ont été suggérées à ce jour et/ou poursuivies vers une exploitation d'A3G en tant que thérapie antivirale. Une de ces approches serait le développement de petites molécules qui inhibent l'interaction entre Vif et A3G et/ou inhibent l'interaction avec les facteurs cellulaires qui agissent après le recrutement de Vif. La molécule RN-18, identifiée en 2008, inhibe l'interaction A3-Vif. RN-18 réduit le taux de Vif et donc augmente le taux des différents A3 quand Vif et les protéines A3 sont présentes ensembles [188]. L'identification de la voie de dégradation d'A3G par le protéasome et la détermination de la structure tridimensionnelle du complexe de l'ubiquitine ligase ouvrent la porte à des stratégies potentielles de blocage de la réplication virale en visant cette voie. Par exemple, la petite molécule (VEC-5) a été identifiée en 2012 ciblant l'interaction de Vif avec la machinerie du protéasome, spécifiquement l'ElonginC..

Il a également été démontré récemment que la protéine ASK1 (Apoptosis Signal-regulating Kinase 1) se lie à Vif pour bloquer son interaction avec les composants du complexe de l'ubiquitine ligase, induisant une stabilisation d'A3G et une réactivation de l'activité de défense médiée par A3G. Des cellules surexprimant ASK1 montrent une restauration de l'activité antivirale d'A3G et forment très peu de particules virales infectieuses dans l'infection secondaire [189]. Pour augmenter la concentration d'ASK1, exprimée en faibles taux dans certains cellules et tissus, dont les LT, cette étude a montré que des concentrations physiologiques d'AZT stimule l'expression d'ASK1 dans les cellules T, d'où l'inhibition de capacité de Vif à neutraliser A3G (Fig. 31).

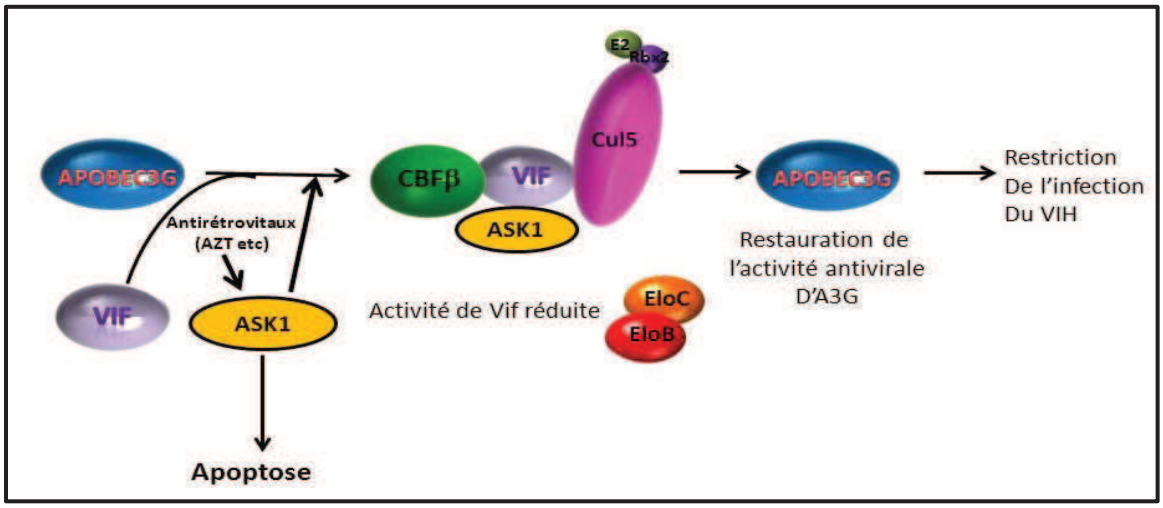


Figure 31 : Modèle proposé suggérant le mode d'action d'ASK1 et son effet sur la restauration de l'activité antivirale d'A3G. Des agents externes tel l'AZT pourrait induire l'expression d'ASK1. ASK1 interagit physiquement avec Vif et interfère avec la formation du complexe Vif-E3 ubiquitine ligase. En conséquence, l'activité antivirale d'A3G est restaurée et la réplication du virus est inhibée (adapté de [189]).

Il n'existe jusqu'à ce jour aucune molécule agissant sur ces interactions testées en phase préclinique ou clinique. Malgré le fait que le développement de petites molécules semble être le domaine le plus prometteur, les progrès effectués dans la thérapie génique anti-VIH-1 suggèrent que les stratégies basées sur A3 pourraient être inclus dans les thérapies géniques anti-VIH-1 actuelles [190, 191]. En plus, ces approches suggèrent l'augmentation de taux d'A3G par un traitement aux interférons ou par thérapie génique en livrant du A3G avec le facteur de restriction TRIM5α [192, 193]. Ces stratégies reposent sur l'utilisation des variantes d'A3 résistantes à la dégradation médiée par Vif. Par exemple, le mutant naturel A3G D128K trouvé chez les singes verts africains ne permet pas l'interaction à Vif et donc sa dégradation par le protéasome [194]. Ce variant pourrait représenter la base des nouvelles stratégies de thérapies géniques anti-VIH-1.

IV. Système d'expression par le virus de la vaccine

A. Définition du virus de la vaccine

Le virus de la vaccine appartient à la famille des *Poxviridae*, il fait partie des virus les plus grands et les plus complexes connus qui infectent des mammifères. La vaccine est transmissible à l'homme, pour qui elle est le plus souvent bénigne, sauf pour les personnes immunodéprimées. La constatation par Edward Jenner que la vaccine bénigne (ou cowpox) protégeait ceux qui étaient infectés de la variole, maladie grave, par immunité croisée, a conduit à l'utiliser dans la prévention de la variole, et donc en vaccination [195]. Le virus de la vaccine mesure entre 300 et 400 nm avec des membranes lipoprotéiques qui entourent un corps central à structure complexe contenant une molécule d'ADN bicaténaire, et linéaire de 200 Kbps (Fig. 32). Ce virus se réplique dans le cytoplasme de façon relativement indépendante des fonctions cellulaires [196]. La réplication de l'ADN et la transcription de la cellule se passant normalement à l'intérieur du noyau, ces virus ont par conséquent besoin d'un génome de grande taille, permettant de coder les enzymes et les protéines nécessaires aux étapes de réplication du virus.

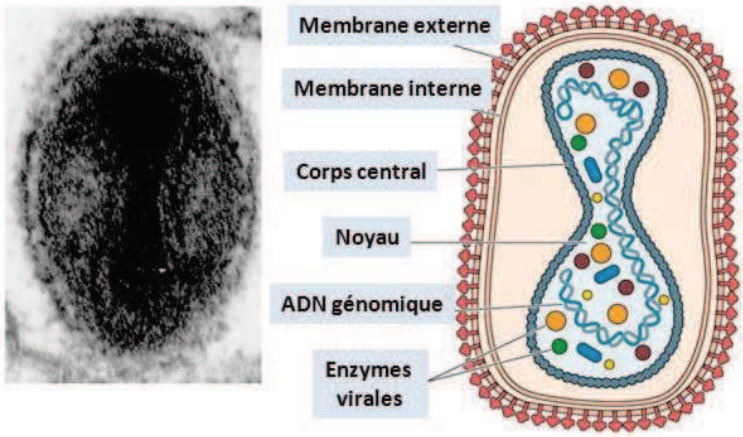


Figure 32 : Morphologie du virus de la vaccine.

B. La réplication du virus de la vaccine

La première étape après l'adsorption du virus à la membrane de la cellule hôte est la fusion de sa membrane à la membrane cellulaire ce qui aboutit à l'injection du corps central viral dans le cytoplasme (première décapsidation). Cette première étape peut aussi de dérouler après internalisation

du virus dans une vésicule d'endocytose. Puis, les transcrits précoces sont synthétisés et sortent de la particule virale (sans nécessité de synthèse de protéines). La majorité des protéines précoces ayant une fonction connue sont liées à la synthèse de l'ADN ou l'ARN. Les protéines associées aux corps centraux infectant sont éliminées (décapsidation secondaire, l'ADN parental devient sensible à la DNase). Cette décapsidation requiert la synthèse des protéines mais ce mécanisme n'est pas encore bien connu. Aussi, au premier stade de l'infection, le télomère de l'ADN viral est clivé par une endonucléase spécifique et la réplication virale démarre ainsi, dans des régions précises du cytoplasme appelés virosomes. Les virosomes sont des régions denses du cytoplasme visibles par microscopie électronique ou photonique après marquage fluorescent de l'ADN ou des protéines virales. Le modèle actuel de la réplication de l'ADN des poxvirus implique la formation de concatémères suivi par l'élongation de l'ADN par l'ADN polymérase du virus de la vaccine débutant du terminus 3'OH [197]. La répétition inversée peut alors se replier et poursuivre la synthèse de l'ADN ce qui entrainera la formation d'intermédiaire de concatémères. Ces concatémères sont par la suite transformés en des molécules d'ADN matures, incorporées dans les particules virales dans les étapes tardives de l'infection [198] (Fig. 33).

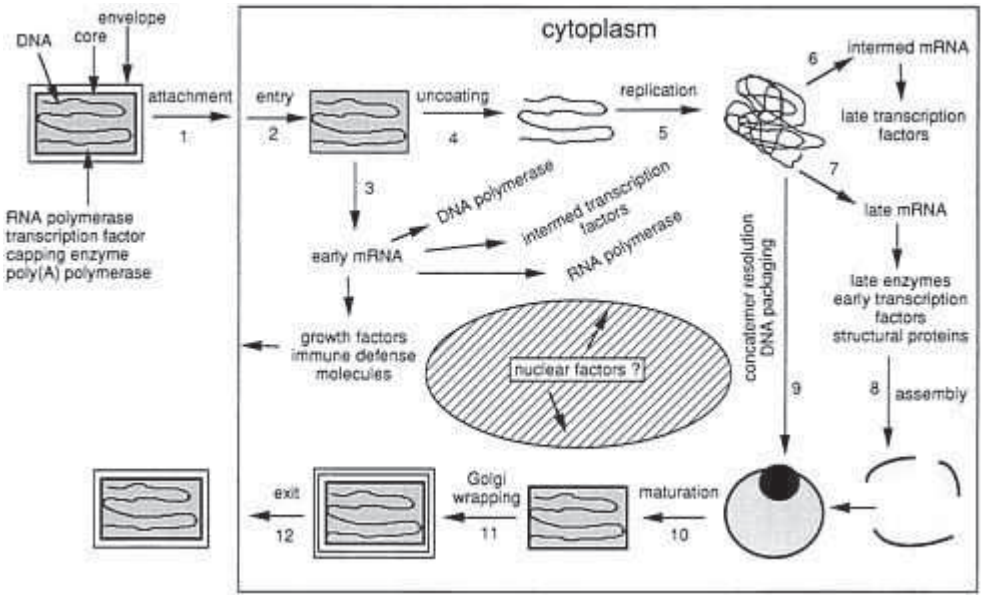


Figure 33 : Le cycle infectieux du virus de la vaccine (Moss, B. (1996) in Poxviridae: The Viruses and Their Replication)

C. Application dans l'expression des protéines

Le virus de la vaccine, utilisé comme vaccin pour protéger de la variole pendant les deux derniers siècles a permis l'éradication de cette maladie en 1980. Depuis, le virus de la vaccine a été utilisé comme vecteur efficace pour la vaccination et pour l'expression des protéines [199, 200]. Un grand nombre d'études ont montré la capacité des virus de vaccine recombinant codant des protéines étrangères à induire des niveaux significatifs d'expression à la fois *in vivo* et *in vitro*. En effet, le génome viral est capable d'intégrer de grands fragments d'ADN (au moins 25 Kbps) et possède des sites multiples d'insertion de l'ADN étranger (Fig. 34). Ces caractéristiques du virus de la vaccine permettent l'intégration simultanée de plusieurs séquences codantes différentes. En plus, le virus de la vaccine est capable d'infecter une variété d'espèces animales et de lignées cellulaires *in vitro* provenant de différentes espèces, donc des protéines peuvent être spécifiquement produites dans la lignée cellulaire hôte appropriée.

Les premières investigations faisant appel au virus de la vaccine comme système d'expression ont exploité la machinerie de transcription du virus en intégrant des gènes d'intérêts après les promoteurs viraux [199–201]. Un développement majeur dans l'expression de protéines par le virus de la vaccine a été réalisé en intégrant le gène de l'ARN polymérase du bactériophage T7dans le génome du virus de la vaccine derrière un promoteur du même virus et le gène d'intérêt en aval d'un promoteur du bactériophage T7 au sein du même génome viral [202]. Suite à l'infection virale, la T7 RNA polymérase est synthétisée et va transcrire l'ADN en ARNm à partir du promoteur T7. Cette transcription est tellement efficace qu'elle interfère avec la transcription et la réplication du virus de la vaccine. La régulation de la transcription et de la traduction est assurée par des éléments de control bactériophage, l'opéron lactose de *E.coli* et l'EMC (encephalomyocarditis virus leader) respectivement. L'expression des protéines est alors induite par l'ajout de l'IPTG, un analogue du lactose.

Pour éviter les risques et dangers liés à la manipulation d'une souche de virus de la vaccine (Western Reserve) pathogène pour les souris et potentiellement dangereuse pour les personnes immunodéprimées et donc nécessitant le travail dans un laboratoire de niveau de sécurité 2, une équipe à l'IGBMC a transféré le système d'expression par l'ARN polymérase du bactériophage T7 dans le génome de la souche d'un virus de la vaccine entièrement sure appelé MVA : Modified Vaccinia Ankara. Le virus de la vaccine utilisé dans cette stratégie est un virus Ankara modifié (MVA). De cette manière la production de protéines dans cette souche peut être réalisée dans des conditions de laboratoire standard parce que ce virus ne représente pas de risque d'infection pour l'homme. Cette souche peut désormais être utilisée sous les conditions de sécurité de premier niveau [203]. Il s'est avéré plus tard que ce nouveau système d'expression est quatre fois plus efficace que le précédent et il

a été utilisé pour exprimer une variété de protéines ayant un rôle très important dans la régulation de la transcription chez des cellules humaines (estrogen receptor-α, steroid-receptor coactivator-1, coactivator-associated arginine methyltransferase-1), et des membres de familles de protéines affectés dans le cancer comme la tubulin tyrosine ligase. Plusieurs protéines exprimées par le biais de ce nouveau vecteur ont été purifiées pour des études structurales.

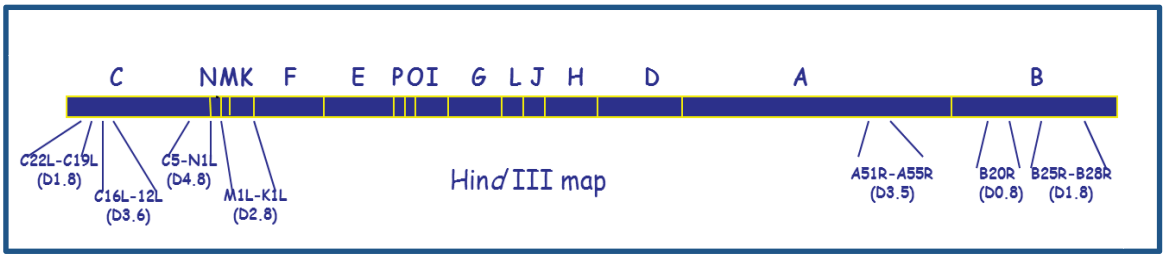


Figure 34 : Génome de la souche du virus de la vaccine modifié Ankara avec les différentes délétions au niveau des six régions non essentielles au génome viral (la délétion à l'extrémité gauche est dupliquée à l'extrémité droite).

OBJECTIFS DU TRAVAIL DE THESE

Depuis la découverte du facteur de restriction virale A3G, de nombreux laboratoires ont essayé de l'étudier plus profondément et ont tenté d'élucider la manière dont A3G inhibe l'infectivité virale ainsi que les mécanismes d'actions de Vif afin de contrecarrer son effet. La protéine Vif du VIH-1 permet la production d'infection dans des cellules non-permissives, (dont les cellules cibles naturelles du VIH-1) en contrecarrant des cytidines désaminases cellulaires A3G et A3F et par la suite empêchant leur incorporation dans les particules virales. Vif inhibe A3G/F par deux mécanismes principaux : 1) Vif entraine la dégradation d'A3G/F par le protéasome et 2) Vif induit l'inhibition de la traduction d'A3G/F. Notre équipe s'intéresse au premier mécanisme, qui représente la cause principale de la réduction du taux des APOBEC3 dans les particules virales.

Pendant de nombreuses années, les données structurales permettant de comprendre les différentes interactions entre Vif et A3G ainsi qu'entre Vif et ses partenaires du complexe de l'E3 ubiquitine ligase étaient inconnues. Malgré le nombre important d'études biochimiques et structurales réalisées sur les protéines Vif et APOBEC, la structure tridimensionnelle caractérisant l'interaction entre ces deux protéines n'est pas encore connue à ce jour et aucune structure n'était encore disponible au début de ma thèse. L'élucidation de la structure du complexe représente une étape essentielle dans la recherche d'inhibiteurs de ces interactions.

La difficulté majeure dans l'obtention de ce type de structure est avant tout la production en grande quantité et dans de bonnes conditions de solubilité, de chacune de ces protéines ainsi que le complexe de l'E3 ubiquitine ligase. Le but de mon projet de thèse était de pouvoir obtenir les protéines Vif et APOBEC3G/F individuellement ou au sein du complexe de l'E3 ubiquitine ligase, caractériser ces protéines avec des techniques physico-chimiques et de vérifier leur bon repliement et leur pureté, tester la fonctionnalité de ces protéines par des tests d'activité pour finalement cristalliser le complexe et déterminer la structure tridimensionnelle.

Cette tâche s'est avérée compliquée du fait des problèmes de solubilité des protéines d'une part et des problèmes de dissociation du complexe durant la purification d'autre part. Durant la première partie, je me suis intéressée à la protéine Vif seule, sa production et l'étude de son interaction

avec l'ARN génomique viral, plus spécifiquement les régions DIS et TAR et ceci en utilisant la technique ITC (Isothermal Titration Calorimetry) disponible au laboratoire.

Dans la deuxième partie de mon travail de thèse, le facteur CBFβ cofacteur de la protéine Vif et qui aide à sa solubilisation a été découvert. Mon projet a été orienté par la suite vers la co-expression de Vif avec CBFβ ainsi que les autres composants de l'E3 ubiquitine ligase dans le système procaryotique en exprimant les protéines dans les cellules *E. coli*. Le but de ces co-expressions était d'obtenir la protéine Vif soluble et bien repliée pour des caractérisations biophysiques et des tests de cristallisation. Ce travail a été effectué en collaboration avec le Dr. Christophe Romier à l'IGBMC qui nous a donné les plasmides et la stratégie de co-expression, tous les deux développés par lui-même.

La troisième partie de ma thèse a été réalisée en collaboration avec Dr. Marc Ruff à l'IGBMC et le travail quotidien expérimental était effectué sous l'encadrement et l'aide du Dr. Robert Drillien. Cette partie consiste en l'expression de la protéine APOBEC3G seule en ou forme de polyprotéine avec Vif et les autres composants de l'E3 ubiquitine ligase dans le système eucaryotique et ceci en utilisant comme vecteur viral, le virus de la vaccine et les cellules BHK21 de hamster pour l'expression. Cette technique nous permettra de reconstituer le complexe *in vivo*, directement dans les cellules ce qui pourrait favoriser le bon repliement et la solubilisation des protéines.

MATERIELS ET METHODES

Notre but est de caractériser structuralement et fonctionnellement les différentes protéines et complexes que nous étudions. Ceci nécessite la mise en place d'un large éventail de techniques recouvrant la biologie moléculaire, la production, et la purification de protéines recombinantes,la caractérisation biophysique, la caractérisation fonctionnelle et structurale des protéines et complexes étudiés. Ma thèse a porté sur deux systèmes d'expression, le premier procaryotique en exprimant dans les bactéries *E. coli* et le deuxième eucaryotique dans des cellules de mammifères et plus particulièrement les cellules de hamster BHK21. Nous avons essayé par ces deux biais de produire la protéine APOBEC3G ainsi que le complexe Vif-CBFβ-EloB-EloC-APOBEC3G de façon soluble. Une fois l'optimisation des techniques de production et de purification achevées, des caractérisations fonctionnelles sont alors envisagées afin de comprendre leurs rôles. Cependant, le processus qui mène à la cristallisation des protéines n'est pas linéaire et nécessite un grand nombre d'essais. Cette partie décrit les principales techniques qui ont été utilisées au cours de cette thèse.

I. Biologie moléculaire-Biochimie

A. Techniques de biologie moléculaire

1. Cellules et virus

a. Système bactérien

Durant ce travail de thèse quatre souches de bactéries *E.coli* ont été utilisées :

– La souche DH5α est utilisée pour l'amplification de tous les vecteurs excepté ceux qui contiennent le gène ccdB, et pour les étapes de clonage nécessaire au projet.

– La souche DB3.1 est utilisée pour amplifier les vecteurs contenant le gène ccdB qui bloque la croissance bactérienne d'*E.coli* en produisant une protéine interférant avec la gyrase d'*E.coli*. en effet, la souche DB3.1 contient l'allèle de la gyrA462 qui rend la souche résistante aux effets toxiques du gène ccdB.

– Les souches BL21(DE3RIL) et BL21 (DE3 Rosetta2) ont permis l'expression des différentes protéines nécessaires au projet. Les deux souches utilisées ont des caractéristiques communes décrites ci-dessous.

DE3 signifie que cette souche a intégré le lysogène du bactériophage λDE3 qui contient le gène codant l'ARN polymérase du bactériophage T7 sous contrôle du promoteur lacUV5 inductible à l'IPTG. De plus ces souches sont déficientes pour deux principales protéases responsables de la dégradation des protéines recombinantes exprimées : les protéases *lon* et *ompT*. Ces deux protéases, si elles étaient présentes pourraient dégrader les protéines recombinantes exprimées. La souche BL21 (Rosetta2) est utilisée pour augmenter l'expression des protéines recombinantes d'origine eucaryote dont les gènes correspondant contiennent des codons rarement utilisés chez *E. coli*. Cette souche contient un plasmide pRARE2 fournissant les ARNt correspondant à 6 codons rares : (AGG, AGA, AUA, CUA, CCC, GGA), en plus d'une résistance au chloramphénicol.

– La souche ArcticExpress est utilisée pour l'expression des protéines insolubles en activant leur repliement et leur solubilité par l'expression de chaperones *d'Oleispira antarctica.* Cette souche pousse à température basse.

Composition	LB liquide	LB solide
Bacto-tryptone ou Peptone	10g	10g
Extrait de levure	5g	5g
NaCl	5g	5g
Bacto-agar	/	15g
Eau Millipore	1 litre qsp	1 litre qsp

Tableau 6 : Composition du milieu LB.

b. Système mammifères

– BHK-21 [C-13] ATCC ® CCL-10™ *Mesocricetus auratus* : ce sont des cellules de reins de hamster doré, créés en 1966 par I.A Macpherson et M.G.P Stoker. Cette lignée cellulaire a été utilisée pour l'infection virale avec le virus de la vaccine pour surexprimer des protéines humaines recombinantes. Les cellules BHK21 étaient cultivés dans le milieu GMEM contenant 10% FCS, 1.5 mg/ml bacto-tryptose phosphate, 100 Unités/ml pénicilline et 100 µg/ml Streptomycine à 37°C. Pour le décollement cellulaire, les cellules sont lavées avec le PBS et trypsinisées pour 10 min.

– BHK-21 : lignée en suspension

– Virus de la vaccine modifié souche Ankara MVA-Koom (Fig. 35) *(Pradeau-Aubreton, article en préparation).* Ce virus est utilisé pour infecter les cellules BHK-21. Il est très fortement atténué par des passages successifs sur cellules d'embryons de poulet, et permet de travailler dans des conditions standard de laboratoire. Le virus MVA ne se multiplie pas dans la plupart des lignées de cellules de mammifères (sauf les cellules de hamster BHK-21 et les cellules de singe MA104). Certains éléments

de l'opéron lactose (répresseur I et opérateur Lac O) ont été introduits dans le génome de ce virus ainsi que le gène de l'ARN polymérase du phage T7 (Tableau 7).

– Glasgow's Minimum Essential Medium (GMEM) de *Life technologies (*21710-082*)*, milieu de culture pour les cellules BHK21. Ce milieu a été développé comme étant une modification du milieu « Eagle ». C'est un milieu MEM modifié en rajoutant du L-glutamine et Phenol Red, sans Tryptose phosphate broth et HEPES.

– Serum bovin fœtal de *Fisher Scientific (11550356).*

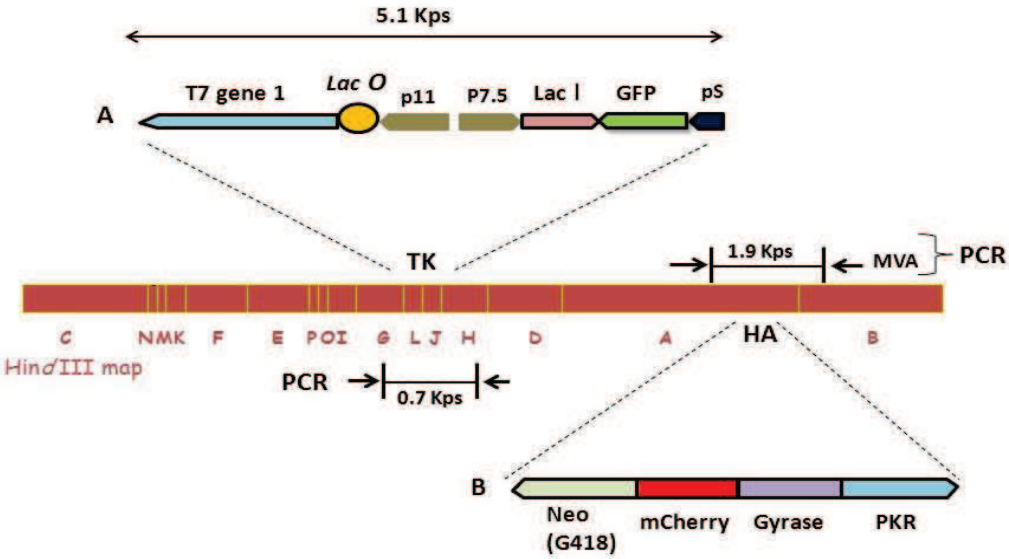


Figure 35 : Représentation schématique du virus de la vaccine (Pradeau-Aubreton, article en préparation). Le génome du virus de la vaccine est représenté en ligne horizontale épaisse avec la taille et la position relative des fragments de restriction Hind III produits pour la souche MVA. Les positions des sites d'insertion TK et HA ciblé par les plasmides A et B respectivement sont indiqués par des flèches. Le génome du MVA comporte des gènes importants pour la transcription et traduction des protéines d'intérêt ainsi que pour la sélection des virus recombinants (Tableau 7).

Gène	Protéine/Fonction
T7	T7 RNA polymérase
lacI	Répresseur lac
LacO	Opérateur : cible de lac I (régulation de la transcription
GFP	Protéine fluorescente verte
mCherry	Protéine fluorescente rouge
Neo	Résistance à la Néomycine
Gyrase	Domaine de dimérisation de la gyrase d'*E. coli*
PKR	Domaine actif de la protéine kinase ARN dépendante des cellules eucaryotes

Tableau 7 : Les différentes séquences intégrées dans le virus de la vaccine par recombinaison homologue.

2. Amplification de l'ADN

a. PCR

L'amplification du gène à partir d'un génome ou d'un vecteur de départ est réalisée par réaction de polymérisation en chaîne. Cette réaction est catalysée par une polymérase, l'enzyme Taq polymérase, qui va ajouter des dNTP par complémentarité à la matrice à partir des amorces sens et anti-sens. La composition du milieu réactionnel est présentée dans le tableau 8. Cette réaction est réalisée dans un thermocycleur. Le produit de la réaction est ensuite déposé sur gel d'agarose, les bandes correspondantes au produit de l'amplification sont découpées et l'ADN est extrait du gel à l'aide du kit Nucleospin® Extract II (Macherey Nagel).

La réaction de PCR a été réalisée pour amplifier les différents gènes et en insérant en même temps les sites de restriction nécessaires pour leur clonage ensuite dans les vecteurs d'expression et leur concaténation.

Composant	Volume	Concentration ou quantité
Matrice	10 ng	10 ng
Tampon Taq pol (10x)	10 µl	1X
Primer 5' (10 µM)	1 µl	10 pmol
Primer 3' (10 µM)	1 µl	10 pmol
dNTPs (10 mM)	4 µL	
Taq DNA polymérase (2U/µl)	0.2 µL	0.4 U
H20 MQ	100 qsp	100 qsp

Tableau 8 : Composition du mélange réactionnel pour la réaction de PCR

b. Clonage par restriction/ligation

Après la PCR et la purification des fragments, ces derniers sont digérés avec les enzymes de restriction dont les sites de reconnaissance ont été placés sur les amorces sens et anti-sens. La digestion se produit sur 1µg d'ADN dans 20 µL de volume total et en rajoutant 1 unité de chaque enzyme. La réaction de déroule à 37°C pendant 4 heures. Après une nouvelle purification sur colonne Nucleospin® Extract pour éliminer les sels et enzymes présentes, une réaction de ligation est réalisée. 50 ng de plasmide est mélangé à une quantité d'insert suivant la réaction : m= taille insert/taille plasmide x 50. A ce mélange est rajouté la T4 ADN ligase de Thermo Scientific ainsi que son tampon. Cette réaction est laissée à 18°C sur la nuit puis transformée dans les bactéries DH5α. Le lendemain , des colonies auront poussé sur les boites LB agar ; certaines seront choisies et mises en culture dans 2 mL de milieu LB contenant l'antibiotique convenable dilué au 1000ème et mises sous agitation à 37°C pendant 16h. Le lendemain, les plasmides contenus dans ces bactéries sont extraits à l'aide du kit Nucleospin Plasmid QuickPure (Macherey-Nagel). La présence de l'insert est vérifiée par digestion du vecteur avec les enzymes de restriction utilisées lors du clonage et l'insert est séquencé.

c. Transformation bactérienne

Une fraction aliquote de bactéries chimio-compétentes (50 µl) est mélangée à l'ADN plasmidique (50 ng) et le mélangée est laissé 30 min sur la glace. Un choc thermique est réalisé à 42°C pendant 1 min. Ensuite, le mélange est laissé dans la glace pendant 5 min pour après rajouter 100 µl de milieu LB. La réaction est placée à 37°C pendant une heure. Finalement, ce mélange sera étalé sur une boite LB agar supplémenté de l'antibiotique adéquat.

3. Générations des vecteurs d'expression bactériens

Généralement, les premiers tests d'expression sont réalisés dans le système bactérien pour différentes raisons. D'abord, l'expression de protéines en bactéries permet d'obtenir à très faible coût de grandes quantités de protéines utilisables pour des études structurales. De plus, l'utilisation d'hôtes bactériens permet de tester facilement et très rapidement la solubilité de constructions différentes des protéines d'intérêt, surtout quand il s'agit de complexes protéiques.

Dans cette partie, plusieurs constructions ont été testées en expression ainsi qu'en co-expression. Cette phase a donc necesité un important travail de clonage compte tenu du nombre de gènes mis en jeu et des différentes combinaisons possibles. Les gènes doivent d'abord être amplifiés par PCR avant de rajouter les sites de restriction en N et C-terminal nécessaire pour leur passage ultérieur dans les vecteurs d'expression.

a. Clonage en vecteurs d'expression bactériens

Une fois que l'insert est obtenu et vérifié par séquençage, le gène codant pour la protéine d'intérêt peut être inséré dans un vecteur d'expression par restriction/ligation. Les vecteurs d'expression utilisés dans le système d'expression en bactéries vont permettre d'introduire en amont de la protéine un site de fixation de la polymérase T7 qui va transcrire le gène, un site de fixation des ribosomes qui vont traduire l'ARN messager produit au cours de la transcription ainsi qu'une séquence codant pour une étiquette de purification et un site de coupure par une protéase. Des codons stop et un terminateur se trouvent en aval du gène pour arrêter la traduction et la transcription.

i. *Les vecteurs d'expression bactériens de la suite pET-MCN*

pET-MCN de « pET-Multi Cloning and expressioN » [204], ces vecteurs utilisés pour cette partie sont construits par le C. Romier. Cette série de vecteurs permettent de tester différentes étiquettes de purifications sur les deux extrémités du gène ainsi que des sites de clivage différents. La caractéristique particulière de cette suite de vecteurs est qu'elle permet de réaliser des co-expressions puisque chacun de ces vecteurs porte des gènes de résistance à un antibiotique et des origines de réplication différentes. Une autre particularité de ces vecteurs est la possibilité de concaténer plusieurs gènes sur un même vecteur grâce à des sites de restriction spécifiques. Pour les co-expressions, les vecteurs utilisés étaient : le vecteur pnEA-tH qui porte une étiquette de purification (6xHis) suivi d'un site de coupure par la thrombine au côté N-terminal en plus de la présence d'un gène de résistance à l'ampicilline, ainsi que le vecteur pnCS qui ne porte pas d'étiquette de purification et qui sert à l'expression des protéines natives. Il porte un gène de résistance à la spectinomycine. Ces vecteurs possèdent également un promoteur T7 inductible par l'IPTG.

Un autre vecteur de type pET-MCN qui a été utilisé pour l'expression du complexe EloB-EloC est le vecteur pnEA-vH, qui est un vecteur codant pour l'étiquette 6xHis du coté C-terminal de la protéine et ayant un site de clivage à la TEV.
Le clonage des séquences codantes d'intérêt a été réalisé en utilisant des sites de restriction NdeI (5'-CATATG-3') et BamHI (5'-GGATCC-3'). NdeI est le premier site de restriction de la cassette de clonage et la présence de ce site rajoute deux acides aminés supplémentaires en amont de la protéine.

Nom	Abréviation	Base	Promoteur	Résistance	Tag	Protéase
pETMCN-EATNH	pneaTH	pET15b	T7	Amp	His(Nt)	Thrombin
pETMCN-EAVCH	pneavH	pET15b	T7	Amp	His(Ct)	TEV
pETMCN-CS	pnCS	pCDF-DuET	T7	Spec	-	-

Tableau 9 : Différentes caractéristiques des vecteurs utilisés pour l'expression et la co-expression des protéines Vif/CBFβ/EloB/EloC/APOBEC.

- ***Stratégie de co-expression sur des vecteurs différents***

Les vecteurs d'expression bactériens qui vont permettre l'expression des protéines d'intérêt dans les bactéries ont été développés par Christophe Romier à l'IGBMC [205] . Pour cela, les séquences nécessaires à la production de la protéine dans cet hôte sont placées sur le vecteur d'expression de part et d'autre de la cassette de clonage. En amont du gène, le promoteur de la polymérase T7 est placé afin d'assurer la transcription du gène, ainsi qu'un site de fixation des ribosomes qui permet la traduction de la protéine. En aval du gène, des codons stop vont permettre l'arrêt de la traduction et une séquence terminatrice qui va servir au décrochage de la polymérase T7.

Les deux vecteurs présentés dans la Fig. 37 ont principalement été utilisés pour cette partie de la thèse (expression en cellules de bactéries). Ces vecteurs permettent la co-expression de complexes protéiques, le vecteur pnEA-X permettant la production d'une protéine portant une étiquette de purification et le vecteur pnCS permettant l'expression d'une protéine native. De plus, les sites de restriction qui entourent l'insert et qui sont présents dans ce vecteur permettent la concaténation des gènes qu'ils contiennent.

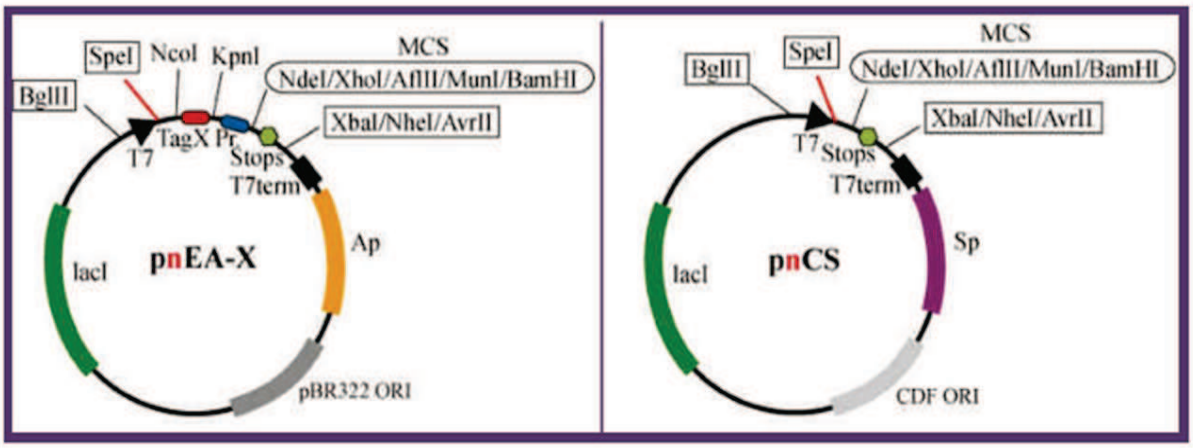


Figure 37 : Carte des deux principaux vecteurs d'expression bactériens utilisés.

- ***Concaténation des vecteurs pour l'expression de complexes bactériens***

Les vecteurs ont été conçus de manière à pouvoir aligner plusieurs séquences codantes au sein d'un même vecteur de façon directe et rapide et en utilisant les sites de restrictions. L'avantage de cette stratégie par rapport à la co-expression par plusieurs vecteurs est la simplification de l'étape de transformation bactérienne ainsi que la diminution du nombre d'antibiotiques qui limite la croissance bactérienne s'il est élevé. Ces avantages sont importants surtout quand on passe à des complexes de plus de deux protéines, comme dans notre cas. La concaténation va utiliser les sites de restriction en amont et en aval du gène (Fig. 38). Le vecteur receveur sera digéré par BglII et SpeI ce qui entrainerait l'élimination de son promoteur et la conservation de son ossature ; le vecteur donneur sera digéré par BglII et XbaI, le promoteur ainsi que l'insert seront conservés à la suite de cette digestion. L'insert est ligué dans le vecteur et les sites SpeI et XbaI sont ligués ensemble du fait de la compatibilité de ligation de ces deux sites. La concaténation de ces deux fragments va recréer un vecteur comportant un seul promoteur qui contrôle l'expression de deux gènes. Cette étape de « couper-coller » sera répétée pour mettre jusqu'à 5 à 6 différents gènes sur un même vecteur. Avant chaque gène, il y a un RBS (Ribosome Binding Site) ce qui permettra la transcription et traduction de chaque gène indécemment des autres.

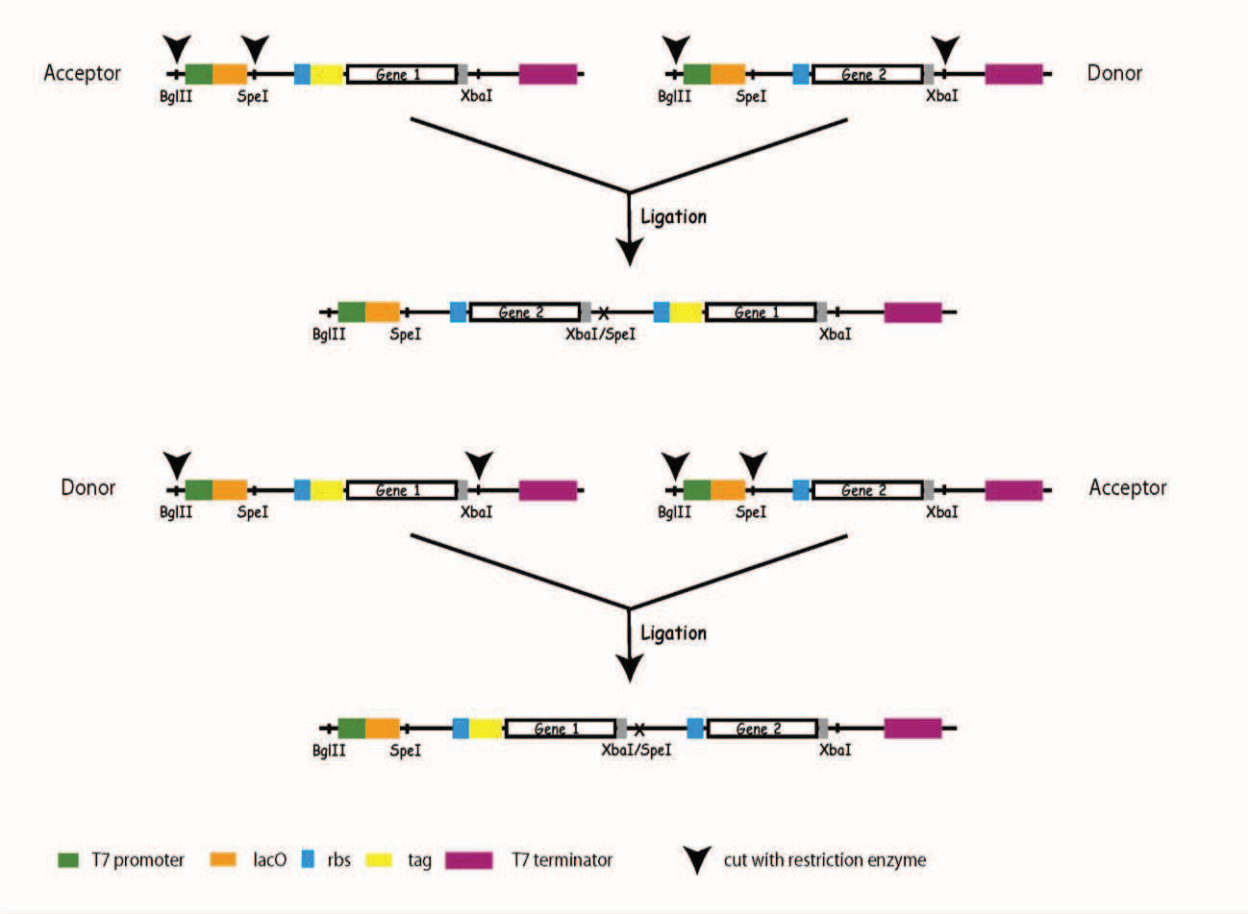


Figure 38 : Stratégie utilisée pour la co-expression du complexe chez E. coli [204].

ii. *Le vecteur pGEx*

Dans le cadre de la réalisation des tests de GST pulldown pour la protéine APOBEC3G, les protéines ont été produites avec une étiquette de purification GST à leur extrémité N-terminale et par un site de clivage à la protéase PreScission permettant de détacher la GST. APOBEC3G a été inséré entre les sites de restriction NdeI et XhoI. Ce vecteur ne peut pas être utilisé pour les co-expressions ou les concaténations. Comme les autres vecteurs utilisés pour l'expression, la transcription est réalisée par la polymérase T7 et induite par l'IPTG.

4. Optimisation des codons

La bactérie *E. coli* est le premier choix en tant qu'organisme hôte pour l'expression de protéines recombinantes parce que sa culutre est rapide, simple, à haute densité et à faible coût, sa génétique est très bien connue. De plus, de nombreux outils sont compatibles avec son utilisation, tels que des souches mutées, des partenaires de fusion et des plasmides recombinants [206, 207]. Pourtant,

tous les gènes ne peuvent être efficacement exprimé dans *E. coli*, probablement dû à la structure unique du gène d'intérêt, la faible stabilité et traduction de l'ARNm, le repliement difficile des protéines, la dégradation de la protéine cible par les protéases de *E .coli*, la différence d'usage des codons entre l'organisme du gène et *E. coli* native, ou la toxicité de la protéine cible exprimée [208]. Il a été montré que les codons rares causent des problèmes d'expression dans *E. coli* ou d'autres organismes, plutôt au niveau traductionnel que transcriptionnel [209–211]. Les problèmes principaux causés par les codons rares sont :

- Les codons rares réduisent le taux de traduction du gène d'intérêt.
- Le taux d'expression de la protéine d'intérêt est faible ou pas détectable.
- Les acides aminés sont mal incorporés dans la protéine
- Des peptides ou des protéines tronqués ou délétés d'acides aminés pourraient être synthétisés.
- Des peptides ou des protéines ayant un décalage du cadre de lecture pourraient être synthétisés.

Un codon rare est un codon synonyme qui n'est pas seulement utilisé rarement, mais est aussi codé par un ARNt non abondant (ARNt rare).
L'usage des codons a été défini comme étant le nombre de fois ou la fréquence avec laquelle un codon est traduit par unité de temps dans la cellule d'un organisme [212]. L'usage des codons n'est pas le même chez tous les organismes ; dans les organismes unicellulaires comme *E. coli*, les codons traduits par les ARNt les plus abondants sont les plus fréquemment utilisés. D'où l'idée de modifier les codons favorisant les codons synonymes ayant le gène d'ARNt avec un nombre de copie élevé pour améliorer l'expression.
Certains gènes ont été optimisés pour améliorer l'expression dans *E. coli ;* les codons rares ont été substitués par des codons utilisés plus fréquemment dans *E. coli* tout en conservant la même séquence protéique (mutations silencieuses).

5. Harmonisation des codons (Annexe)

La substitution des codons synonymes peut améliorer l'expression de la protéine. Mais le fait de changer des codons non fréquents pour des codons fréquents dans des régions où la traduction du messager doit être lente peut affecter le repliement correct de la protéine. En effet, les substitutions des codons synonymes ne sont pas toujours silencieuses et changer la fréquence d'usage des codons peut affecter la structure de la protéine ainsi que sa fonction. L'harmonisation des codons repose sur le principe d'ajuster la disponibilité de l'isoaccepteur de l'ARNt trouvé dans l'hôte naturel à celle de l'hôte d'expression ; ceci est important pour obtenir des cinétiques de traduction optimales dans l'hôte

d'expression. Le concept de l'harmonisation des codons provient d'observations observées dans *E. coli* où il a été montré que les codons rares se regroupent dans des régions de l'ARNm qui codent pour des boucles ou des régions qui séparent des éléments à structure très ordonnée. Ces segments sont formés d'une quinzaine de résidus et sont codés par des clusters de codons rares. Ralentir le temps de transit du ribosome dans ces régions permet l'acquisition d'une structure correcte pour certains domaines [213, 214]. Dans ce contexte, le gène VifHXB2 a été synthétisé après harmonisation des codons dans sa séquence (annexe) et cette séquence harmonisée a été testée dans différentes expériences de co-expression.

	Vif pNL4-3	Vif HXB2	CBFβ	EloB	EloC	Cul5	A3G/ A3F
Non optimisé	x		x				x
Optimisé		x	x	x	x	x	x
Harmonisé		x					x

Tableau 10 : Les gènes utilisés pour les tests d'expression/co-expression, originaux, optimisés ou harmonisés.

6. Génération de vecteurs d'expression en cellules de mammifère

Le but final pour pouvoir exprimer les protéines dans les cellules BHK21 est de les insérer dans le vecteur d'expression pVoteOGW en réalisant une réaction LR Gateway entre le vecteur final pEntr_BB-TEV et le pVote.

Une série d'étapes de clonage est nécessaire au préalable :

a. Les clonages Biobrick

Les plasmides pUB_BB sont des vecteurs BioBrick construits par Karine Pradeau-Aubreton à l'IGBMC [215], contenant une région de réplication bactérienne, les sites de restriction BioBrick, 2 sites consécutifs de clivage à la TEV et le gène AmpR. Les pUC_BB utilisés dans ce système portent soit une étiquette Flag-Histidine (pUC_BBFH10P), soit un double Strep appelé également « Twin Strep » (pUC_BBRRP).

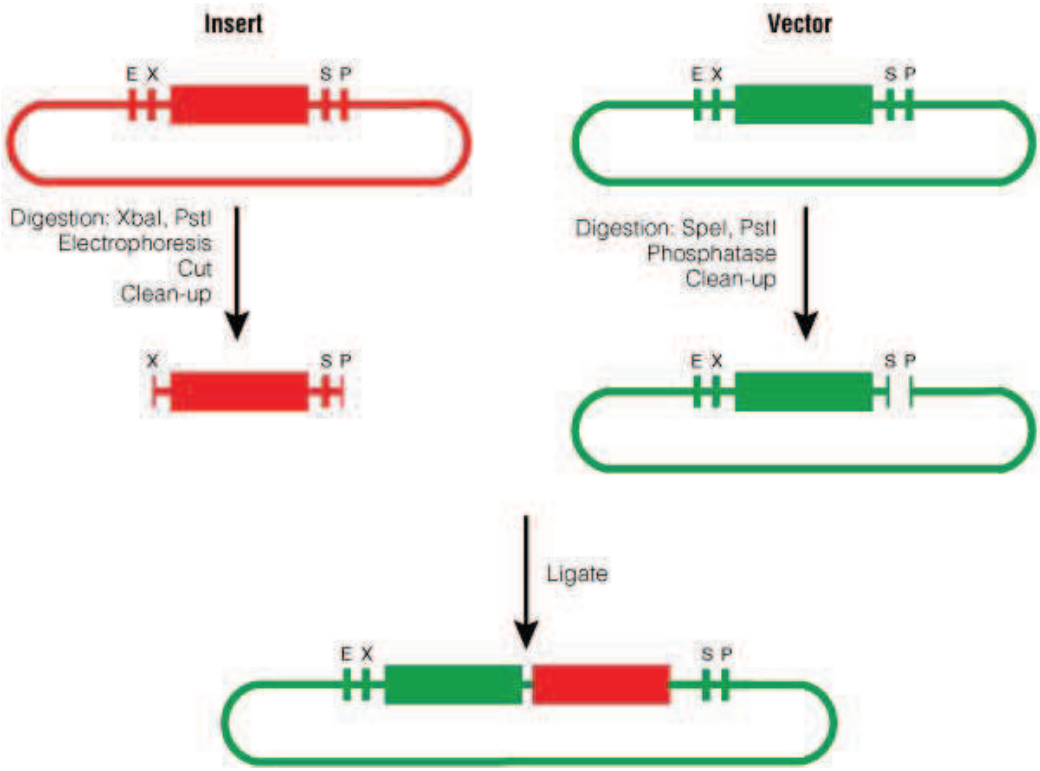


Figure 39 : Clonage Biobrick. Les plasmides possèdent les sites de restriction EcoRI, XbaI, SpeI et PstI. Les inserts sont flanqués des sites EcoRI / XbaI en 5' et des sites SpeI / PstI en 3'. Le plasmide vecteur (vert) est digéré par SpeI / PstI et l'insert (rouge) par XbaI/pstI ; l'insertion se fait en « suffixe » dans le plasmide vecteur grâce à la compatibilité des sites XbaI et SpeI qui disparaissent donc après la ligation. Les sites BioBrick uniques dans le plasmide sont utilisés pour répéter cette étape et insérer d'autres séquences d'ADN.

i. *Clonage en pUC19*

Pour effectuer les clonages en BioBrick une série de sites de restriction est indispensable des côtés 5' et 3' de l'ADN codant chaque protéine. Pour cela, chaque gène sera cloné en premier lieu dans le plasmide pUC19 tout en insérant ces sites par PCR à partir des amorces spécifiques.

5'	**EcoR1-Not1-Xba1**	5' GGAATTCGCGGCCGCTTCTAGA 3'
3'		5' AACTGCAGCGGCCGCTACTAGT 3'
Reverse Complement	**Spe1-Not1-Pst1**	5' ACTAGTAGCGGCCGCTGCAGTT 3'

Tableau 11 : Les sites de restrictions utilisés pour les clonages BioBrick en 5' et 3' de l'ADN.

Gène	Primer Forward (5'-3')	Primer Reverse (5'-3')
Vif	GGAATTCGCGGCCGCTTCTAGA ATGGAGAACCGGTGGCAG	AACTGCAGCGGCCGCTACTAGT GTGTCCATTCATTGTATGG
CBFβ	GGAATTCGCGGCCGCTTCTAGA ATGCCGCGCGTCGTGCC	AACTGCAGCGGCCGCTACTAGT ACGAAGTTTGAGGTCATCACCAC
EloB	GGAATTCGCGGCCGCTTCTAGA ATGGACGTGTTCCTCATGAT	AACTGCAGCGGCCGCTACTAGT CCAATGAACAAGCCGTGCAG
EloC	5'GGAATTCGCGGCCGCTTCTAGA ATGTATGTCAAATTGATATCAT	AACTGCAGCGGCCGCTACTAGT TGGCTGCGAACTTCTTAGATTGT
A3G	GGAATTCGCGGCCGCTTCTAGA ATGAAGCCTCACTTCAGAAA	AACTGCAGCGGCCGCTACTAGT ATTCTCCAGAATCAGGAAAA

Tableau 12 : Les amorces en Forward et Reverse Utilisées pour le clonage dans pUC19

ii. *Passage en pUC_BB*

- *Protocole*

1 µg d'ADN est digéré par 1U de chaque enzyme de restriction : XbaI / PstI pour les inserts dans pUC19 et SpeI / PstI pour les vecteurs plasmidiques pUC_BB. La réaction est menée 1h à 37 °C dans un tampon NEB3 1X, additionné de BSA 1X (NEB), dans un volume final de 20 µL. Un traitement à la phosphatase alcaline est réalisé pour les vecteurs plasmidiques pendant 10 min à 37 °C et la réaction est interrompue par chauffage à 70°C pendant 5 min. Les 20 µL de produits de ligation, additionnés de 5 µL de bleu de dépôt (glycérol 40 %, xylène cyanol 0,05 %, bleu de bromophénol 0,05 %), sont migrés sur gel d'agarose 1,2 % supplémenté de BET (Roth) à 0,5 µg/µL, dans un tampon TBE 1X (Euromedex) pendant 45 min à 100 V (système Mupid, Advance). Les fragments sont visualisés sous UV dans le Gel Doc EZ Imager (Biorad) puis purifiés à partir du gel grâce au kit Nucleospin Gel and PCR Clean-up (Macherey-Nagel). Les ADN sont élués dans un volume final de 20 µL et la concentration massique est directement obtenue après lecture de l'absorbance à 260 nm au NanoDrop.

50 ng de plasmide vecteur sont mélangés à l'insert, avec un rapport vecteur/insert entre ½ et 1/10, dans un tampon T4 ADN ligase 1X (Thermo Scientific) additionné d'ATP 10 mM dans un volume final de 20 µL. La ligation est réalisée par la T4 ADN ligase 1 h à 22 °C.

- *Vecteurs utilisés*

pUC_BB	Tag	Gène
T1	Flag10His	APOBEC3G
	TwinStrep	
T2	Flag10His	CBFβ
	TwinStrep	
T3	-	Vif
T4	-	EloB
T5	-	EloC

Tableau 13 : Les différents plasmides pUC_BB utilisés dans ce projet et les gènes insérés respectivement ainsi que leurs étiquettes de purification. Chaque gène a été inséré dans un pUB_BB selon l'ordre prévu pour la polyprotéine. La différence entre les différents pUC_BB, autres que le tag, est le site de clivage à la TEV.

iii. *Concaténation des gènes en polyprotéine.*

Le même protocole sera réalisé pour les digestions des inserts et plasmides afin de concaténer premièrement deux à deux les gènes, puis les quatre sur le même pUB_BB.

Une fois tous les clonages dans pUC_BB réalisés, chacun de ces clones ainsi que le gène codant la polyprotéine sont digérés par XbaI/PstI et inséré dans le plasmide pEntr_BB-TEV-YFP.

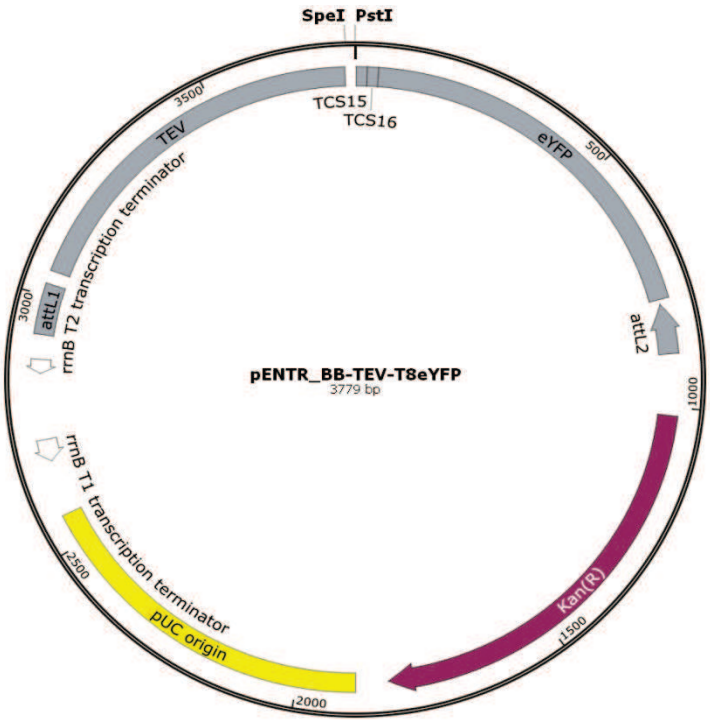


Figure 40 : Le plasmide pEntr_BB-TEV-YFP et ses caractéristiques. Les protéines seront insérées entre les sites SpeI/PstI du pEntr résistant à la Kanamycine. Ce plasmide code pour la protéine TEV en N-terminal, protéase qui va cliver entre les différents gènes de la polyprotéine au niveau du site de clivage à la TEV. En C-terminal, le gène codant pour la YFP est présent. La YFP servira comme marqueur de sélection en suivant la coloration jaune des virus recombinants après induction à l'IPTG.Ce plasmide comporte aussi les sites attL1 et attL2, spécifiques pour la réaction LR Gateway avec le plasmide d'expression pVote.

b. Clonage LR Gateway®

Après avoir inséré les gènes codant des protéines individuelles ou un gène codant la polyprotéine dans le vecteur pEntr, une réaction LR gateway entre ce vecteur et le vecteur d'expression pVote va pouvoir créer un plasmide d'expression pVote0GW prêt à être transfécter dans les cellules BHK21 préalablement infectées par le virus de la vaccine.

- ***Le vecteur d'expression pVoteOGW***

pVote correspond à (Vaccinia virus /lac operon/T7 RNA polymerase/EMC), et OGW signifie l'absence d'une étiquette de purification présente dans son génome (l'étiquette est déjà inséré dans les PUB_BB). Il est couramment utilisé dans le système d'expression en cellules de mammifères. Ce plasmide comporte dans son génome plusieurs séquences essentielles à la réplication des virus et à l'expression des protéines d'intérêt :

- Les séquences « HA-R » et « HA-L » essentielles pour la recombinaison homologue entre le génome viral et le plasmide pVote pour la formation du virus recombinant.
- Un promoteur « T7-lacO » inductible, important pour la transcription.
- Un Terminateur T7 pour arrêter la transcription.
- La séquence « EMC1 » (Encephalomyocarditis) est une séquence non codante qui va renforcer la stabilité et la traduction des transcrits T7 non coiffés produits par le virus de la vaccine.
- Les sites « attR1 » et « attR2 », pour la réaction de LR Gateway avec pEntr.
- La cassette Gateway (Un gène de résistance au Chloramphénicol « Cm(R) » et un gène létal « ccdB » entourés par les sites attR1 et attR2). La présence de ces gènes permettra la sélection des colonies ayant bien effectué la réaction de Gateway et donc qui ont perdu cette cassette de Gateway en la remplaçant par la séquence codante dans pEntr. En plus, le « ccdB » est un gène létal pour la souche DH5α.
- Un gène de résistance à l'ampicilline « Amp(R) » qui servira aussi à la sélection des colonies positives.
- Le gène « gpt » codant pour la protéine guanine phosphoribosyltransférase. Il sera utilisé pour la sélection positive des virus recombinants.

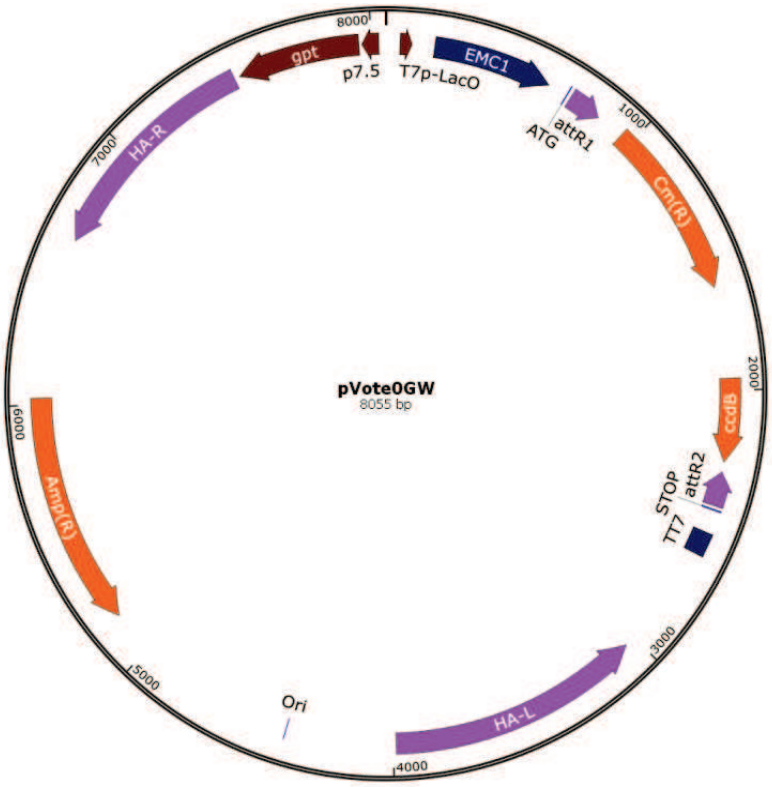


Figure 50 : Le plasmide pVote0GW et ses caractéristiques.

- ***Principe de la réaction LR Gateway***

Le clonage Gateway est un système de clonage par recombinaison homologue proposé et commercialisé par Invitrogen™. C'est un procédé basé sur la recombinaison homologue à spécificité de site du phage λ que ce dernier utilise pour s'intégrer dans le génome d'*E.coli*, et donc de passer d'un cycle lytique à un cycle lysogénique. Le principe de la méthode est présenté dans la Fig. 51 :

La technique de clonage Gateway permet le transfert de fragments d'ADN entre différentes vecteurs tout en gardant la bonne ORF. Pour développer un clone d'entrée Gateway, le système requiert l'insertion d'un fragment d'ADN dans un plasmide avec 2 séquences flanquantes appelés « attL1 » et « attL2 ».

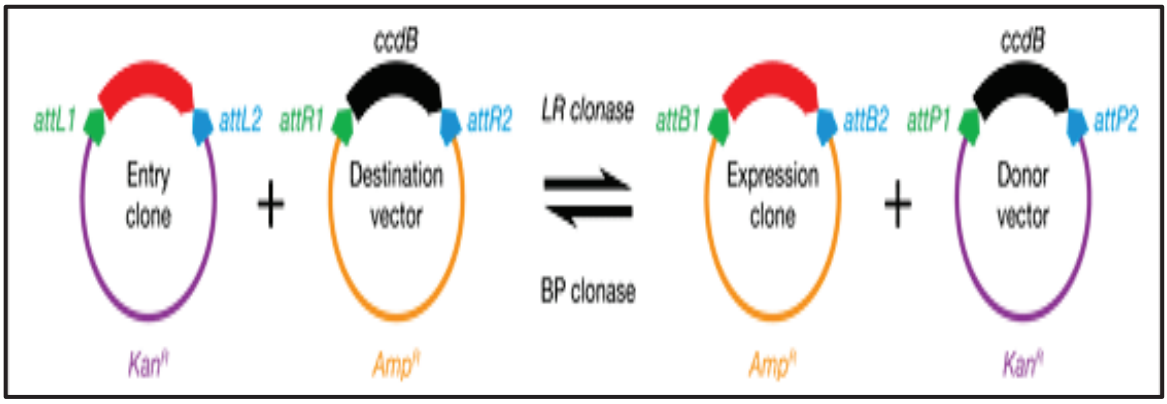


Figure 51 : Principe de la technologue Gateway®.

Le gène *ccdB* permet la sélection négative des vecteurs de destination et donneur et surtout une sélection positive des vecteurs d'entrée et d'expression ayant intégré le gène d'intérêt après une recombinaison homologue et une transformation. Les produits secondaires et les plasmides n'ayant pas subi la réaction LR, expriment la protéine CcdB qui interfère avec l'ADN gyrase d'E.coli en inhibant la croissance des bactéries [216]. L'enzyme LR clonase va catalyser l'échange des deux brins, permettant le transfert du gène dans le vecteur d'expression.

- ***Protocole de la réaction***

	Concentration	Volume
Clone d'entrée (pEntr)	50-150 ng	1-7 µl
Vecteur de destination (pVote)	150 ng/µl	1 µl
Tampon TE , pH 8.0		8 µl
LR Clonase™ II		2 µl

Tableau 14 : Composition du mélange réactionnel de la réaction LR.

Incuber la réaction pour une heure à 25°C puis rajouter 1 µl de la protéinase K pour arrêter la réaction. Incuber 10 minutes à 37°C puis transformer dans des cellules compétentes.

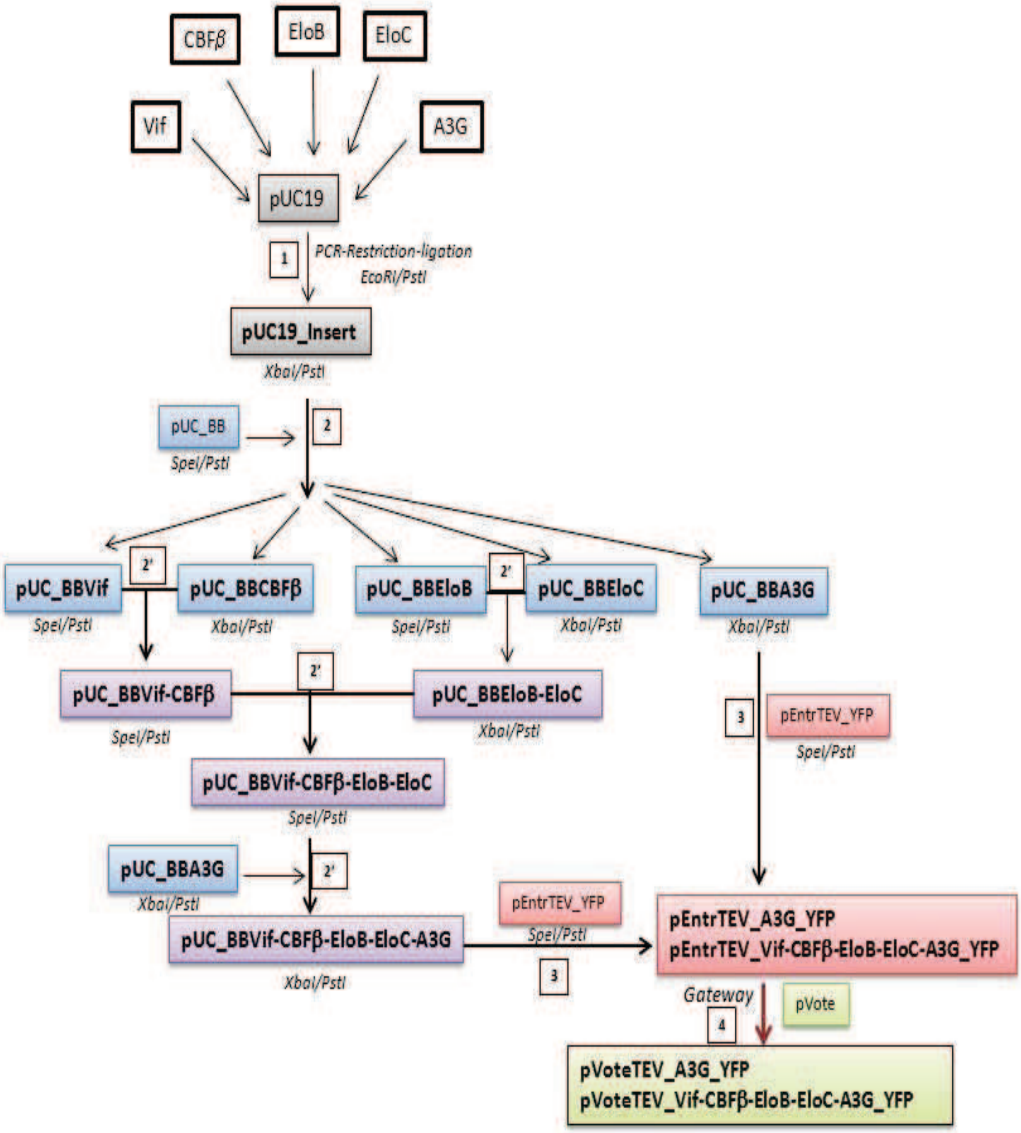


Figure 52 : Schéma simplifié des étapes de clonages effectuées pour l'expression de la protéine A3G et de la polyprotéine Vif-CBFβ-EloB-EloC-A3G dans le système mammifère.
1-Clonage des différents gènes dans le pUC19 entre EcoRI/PstI avec insertion par PCR des sites de restriction nécessaires pour la suite et pour la concaténation. 2- Passages des gènes dans les vecteurs pUC_BB possédant ou pas des étiquettes en N-terminal. 2'- Concaténation des gènes un par un, puis deux par deux, et enfin addition d'A3G à la fin de la séquence. 3- Passage des gènes dans pEntr codant pour la TEV et YFP. 4- La dernière étape des clonages consiste en un Gateway entre pEntr et pVote afin de transférer la séquence codante dans un plasmide d'expression.

B. BHK21/ système d'expression par le virus de la vaccine

1. Principe et stratégie

Le plasmide contenant la séquence d'intérêt flanquée de régions homologues à un locus du génome viral est intégré dans une cellule de mammifère préalablement infectée par un viurs de la vaccine (MVA-Koom). Par un mécanisme de recombinaison homologue, la séquence d'intérêt est transmise au génome viral. Les virus recombinants obtenus sont isolés et utilisés pour réinfecter de nouvelles cellules. L'expression de l'ARN T7 polymérase, contrôlée par l'opérateur LacO de l'opéron lactose, est induite par l'IPTG. Les gènes d'intérêts sous la dépendance d'un promoteur T7 seront transcrits et les protéines produites avec la machinerie eucaryote.

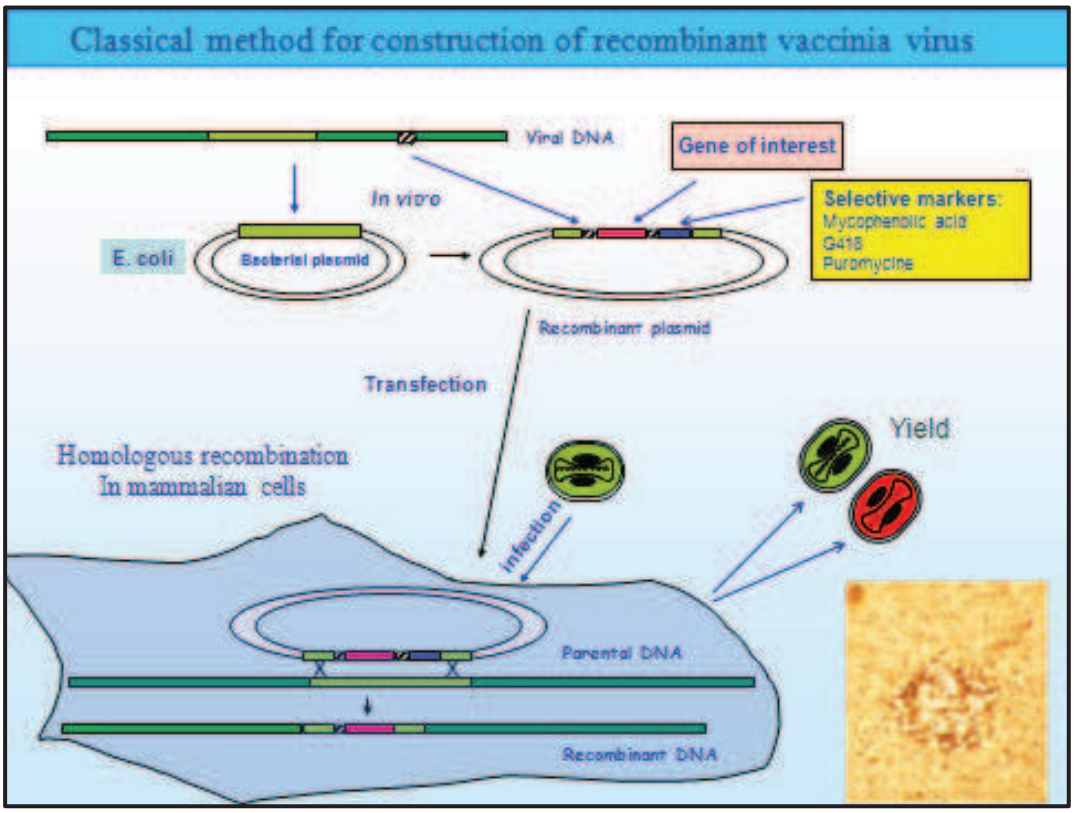


Figure 53 : Principe de l'infection des cellules BHK21 par le MVA. (Avec permission Drillien).

a. Recombinaison homologue :

Les procédures des modifications génétiques des poxvirus sont liées à plusieurs aspects de leur biologie, en particulier, la taille du génome viral et le fait que l'ADN viral ne soit pas infectieux. En règle générale, les virus de la vaccine recombinants sont générés dans les cellules par recombinaison

homologue entre l'ADN génomique viral et le plasmide transfecté [217–220]. De façon alternative, la ligation *in vitro* peut être tout de même utilisée afin de générer des génomes de vaccine recombinants [221, 222].

En général, dans la recombinaison *in vivo*, et pour faciliter la bonne insertion de l'ADN dans une position précise du génome, des séquences virales sont insérées dans l'ADN transfecté. Dans notre système, les sites de recombinaison homologue HA (gène viral non essentiel codant une hémagglutinine du virus) ont été employés pour assurer la recombinaison homologue entre le plasmide d'expression et l'ADN viral parental (Fig. 53). Le résultat de cette recombinaison est la formation d'un génome viral modifié ayant intégré les séquences situées entre HA-R et HA-L du plasmide d'expression. Le génome viral modifié permet l'expression constitutive de la GFP et l'ARN T7 polymérase dont les gènes sont sous la dépendance du promoteur du virus et l'expression du gène d'intérêt ainsi que l'YFP sous la dépendance du promoteur T7 bactériophage.

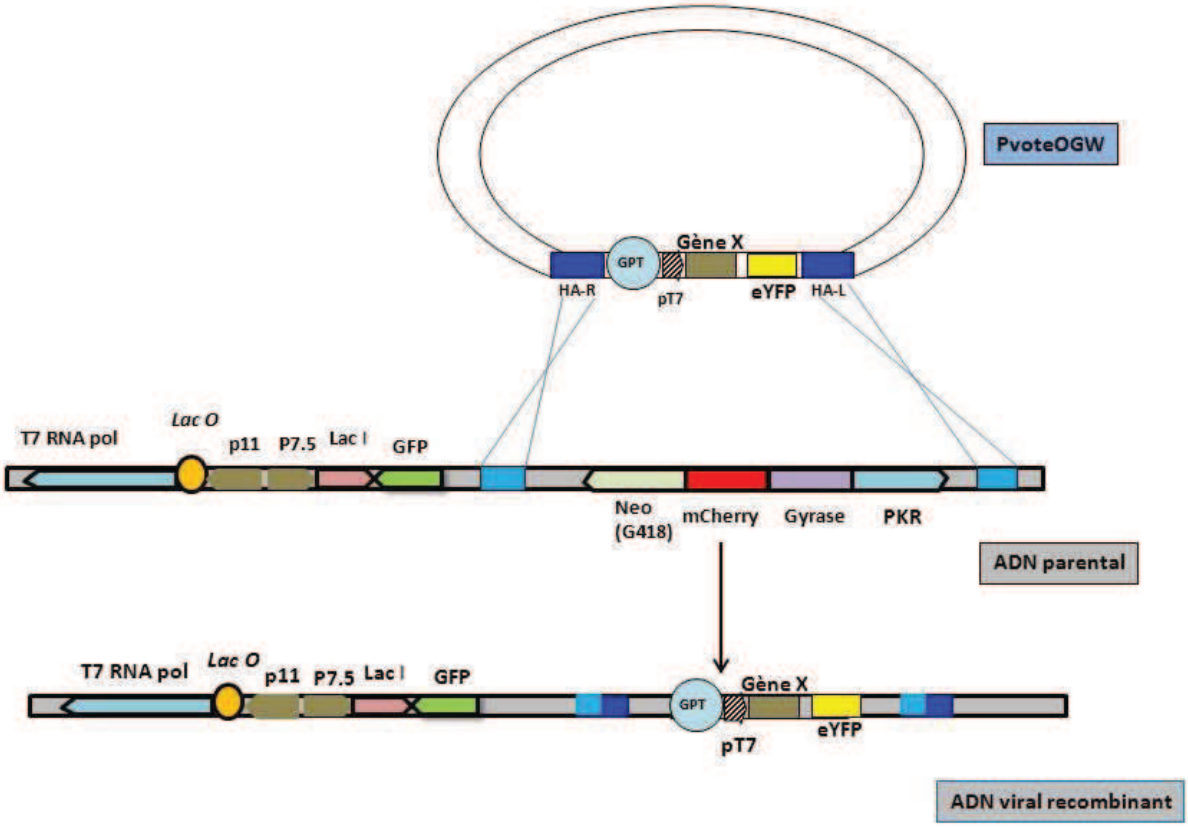


Figure 54 : Recombinaison homologue entre le génome viral de la vaccine et le plasmide d'expression pVote. HA-R : portion droite du gène de l'hémagglutinine, HA-L : portion gauche du gène de l'hémagglutinine.

b. Isolement et purification des virus recombinants

Dans une expérience typique d'infection/transfection, les virus recombinants ne représentent qu'une très faible proportion (10^{-4} à 10^{-3}) du virus produit. Pour cela, des marqueurs génétiques sont souvent rajoutés dans le plasmide transfecté pour faciliter la sélection des virus recombinants [223]. Dans ce système utilisé, il existe deux procédures de sélection différentes : la première est une sélection positive pour les virus recombinants repose sur la résistance à l'antibiotique en utilisant le gène de la xanthine-guanine phosphoribosyl transférase d'*E.coli* (gpt) [224] ; la deuxième sélection est une sélection négative contre le virus parental basée sur la résistance à la coumermycine en utilisant les gènes de fusion Gyrase-PKR [225].

Une considération importante est la différence entre les marqueurs dont le but est de détecter les virus recombinants et les marqueurs de sélection qui permettent une vraie sélection des virus recombinants. Dans les deux cas, l'isolement des virus recombinants peut être accompli par la reprise répétée de plages individuelles du virus, mais généralement, les systèmes de sélection permettent des procédures d'isolement plus rapides et plus efficaces.

i. *Marqueurs de séléction*

- ***Sélection à l'acide mycophénolique***

L'acide mycophénolique est un puissant inhibiteur sélectif, non compétitif de l'inosine monophosphate déshydrogénase ; il inhibe donc sans être incorporé à l'ADN, la synthèse de novo des nucléotides guanine. Il a été montré qu'il bloque la réplication du virus de la vaccine dans des lignées cellulaires normales. Cette observation a entrainé le développement d'une étape de sélection basée sur l'expression du gène gpt d'*E. coli* pour l'isolement du virus recombinant de vaccine. La synthèse de la xanthine-guanine phosphoribosyltransférase permet uniquement au virus recombinant de former des plaques larges dans un milieu sélectif contenant l'acide mycophénolique, la xanthine et l'hypoxanthine.

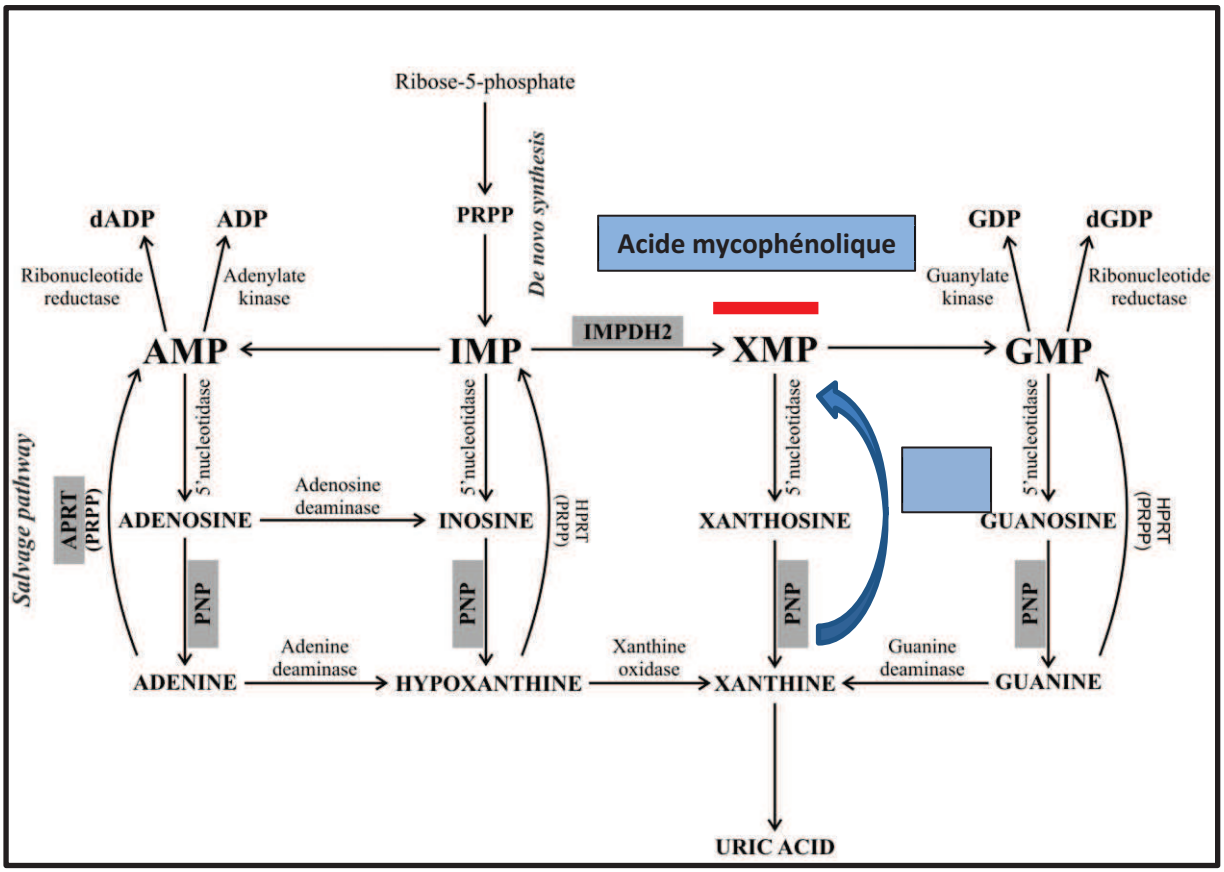


Figure 55 : Voie de synthèse des purines. La gpt est utilisée comme marqueur de sélection positive. Elle assure la conversion de la xanthine ou la guanine en XMP ou GMP. En présence de l'acide mycophénolique, la voie de synthèse des purines est inhibée mais l'enzyme gpt peut compenser cette inhibition en permettant la synthèse de XMP, et le cycle de synthèse des purines est de nouveau établi.

- ***Sélection à la coumermycine***

La protéine de fusion GyrB-PKR est constituée de la sous-unité B, domaine de dimérisation de la protéine gyrase *d'E.coli* fusionnée au domaine catalytique de la protéine humaine PKR. Le domaine de dimérisation de la gyrase est dépendant de la coumermycine. En présence de coumermycine, GyrB dimérise ce qui active PKR et par la suite entraine la phosphorylation du facteur d'initiation de la traduction eIF2α. La phosphorylation de ce facteur inhibe la synthèse protéique et par la suite inhibe la réplication virale. En présence de coumermycine, les recombinants sont isolés du fait qu'ils ont perdu le gène qui confère une sensibilité à la coumermycine par la recombinaison homologue.

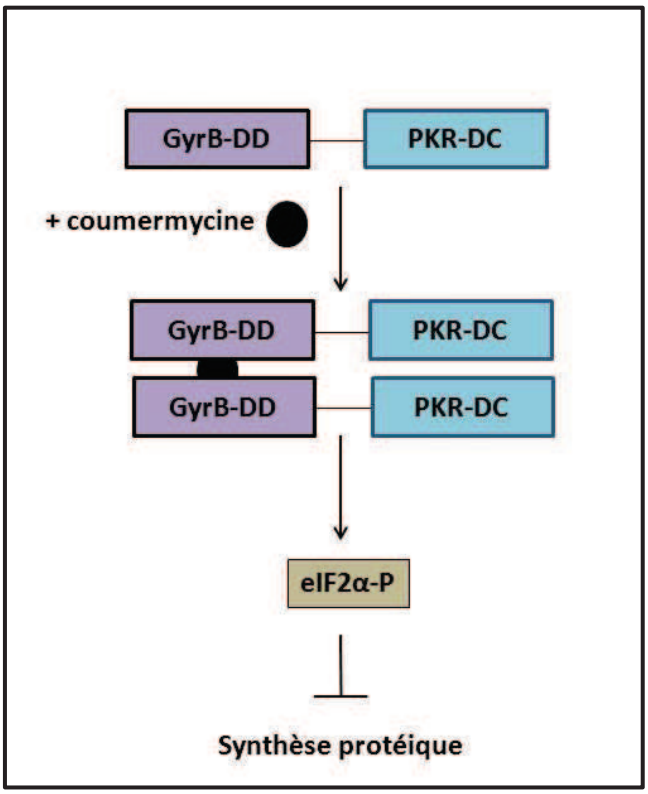


Figure 56 : Représentation du système de la coumermycine/GyrB-PKR comme marqueur de sélection négatif pour l'isolement des recombinants du MVA. GyrB-DD : Domaine de dimérisation de la sous-unité B de la gyrase. PKR-DC : Domaine catalytique de la PKR

- ***Marqueur de sélection fluorescent***

La détection des virus recombinants sera suivie d'après la fluorescence à partir des deux gènes GFP et mCherry. Un virus recombinant doit perdre sa fluorescence rouge parce que, suite à la recombinaison, le gène codant pour la mCherry n'appartient plus à l'ADN recombinant. Ce n'est pas le cas pour le gène codant la protéine fluorescente verte GFP, qui est située dans le génome viral à proximité du gène codant l'ARN polymérase du phage T7. Au fur et à mesure des étapes de sélections aux antibiotiques, un virus recombinant sera reconnu par la perte de sa fluorescence rouge tout en conservant la fluorescence verte.

2. Protocole

a. Transfection de cellules BHK21 infectées par le virus

Des plaques 6 puits de 35 mm^2 contenant entre 0,5x106 et 106 cellules BHK21 / puits sont incubées à 37 °C sous une atmosphère comportant 5 % de CO2 pendant 24 h. Les cellules sont infectées par 1 mL de virus à une concentration telle qu'on aura 0,1 PFU/cellule, puis sont incubées 1

h à 37 °C, 5 % CO2. Pour le mélange de transfection, 1µg d'ADN est ajouté à 5 µL d'un agent de transfection, la lipofectamine, dilué dans 100 µL de milieu sans sérum Opti-MEM (milieu optimisé pour les transfections). Après 15 minutes, 200 µL du mélange sont ajoutés aux cellules dans chaque puits, contenant 800 µL de nouveau milieu. Les plaques sont incubées une nuit à 37 °C, 5 % CO2.

b. Sélection des virus recombinants

Après observation des plages de lyse, 3 étapes de congélation / décongélation sont effectuées pour lyser les cellules et libérer les virus dans le milieu. Des dilutions en cascade 10-1, 10-2 et 10-3 sont réalisées à partir des suspensions pour obtenir des concentrations finales différentes de virus. 200 µL de chaque dilution sont mélangés à 1,8 mL de milieu G-MEM, 2 % SVF (Fisher Scientific). 1 mL de chaque dilution est transféré sur des cellules, préparées la veille en flasque T75, et la plaque 6 puits est placée 1h à 37 °C, 5 % CO2. Pour la sélection à l'acide mycophénolique, 1 mL de milieu G-MEM, 2 % SVF contenant de la xanthine (0,5 mg/mL), de l'hypoxanthine (1,5 mg/mL) et de l'acide mycophénolique (50 µg/mL) est ajouté au 1 mL de virus. Pour la sélection à la coumermycine, 1 mL de milieu G-MEM 2 % SVF contenant de la coumermycine (400 ng/mL) est ajouté au 1 mL de virus. Les plaques sont incubées 2 jours à 37 °C, 5 % CO2 et les cellules sont observées au microscope optique à épifluorescence EVOS® FL (Life Technologies) pour visualiser la fluorescence verte (GFP) et rouge (mCherry) des plages de lyse.

c. Amplification des virus recombinants

La dilution adéquate de lysat de cellules infectées est centrifugée à 3000 g, 5 minutes pour éliminer les débris cellulaires. Le surnageant est déposé sur des cellules préparées la veille en flasque T75, préalablement additionnées de 10 mL de milieu sans sérum. La flasque est placée environ 45 minutes à 37 °C, 5 % CO2. Puis 10 mL de milieu GMEM 4 % SVF, additionné de coumermycine à une concentration finale de 50 µg/mL, sont ajoutés. Pour une bonne amplification, la flasque est incubée à 37 °C entre 2 et 3 jours.

C. Techniques de Biochimie

1. Tests d'expression

Les tests d'expression vont permettre de tester de façon rapide et directe les différentes constructions testées et déterminer les conditions optimales d'expression et de purification des protéines. A partir de ces tests, on pourra identifier la solubilité de la protéine en fonction de la température d'induction, le tampon, la concentration en sel et plusieurs autres facteurs. Ces tests ont

été réalisés dans la bactérie et dans les cellules de mammifères BHK21. Ces tests serviront aussi à identifier les protéines interagissant ensemble dans une co-expression. Les tests d'expression consistent à produire la protéine dans un petit volume. Une mini-purification est ensuite réalisée sur une résine en batch. Les protéines sont éluées et déposées sur gel acrylamide dénaturant.

a. Principe de l'expression induite par l'IPTG

Les deux systèmes d'expression utilisés dans ce projet sont adaptés à régulation de l'expression par l'IPTG. L'IPTG est un analogue du lactose non métabolisable par *E. coli* utilisé comme inducteur de l'opéron lactose. La concentration d'IPTG reste constante ainsi le taux d'expression des gènes sous la dépendance de l'opéron lactose ne varie pas. En absence d'inducteur, le répresseur produit par le gène lacI se fixe spécifiquement sur le site de l'opérateur Lac ce qui empêche l'ARN polymérase de se fixer au promoteur Lac, réprimant la transcription/traduction de l'ensemble des gènes. L'IPTG induit un changement conformationnel du répresseur LacI qui devient incapable de se fixer sur le site opérateur Lac. L'ARN polymérase T7 peut alors se fixer sur le promoteur T7 présent sur le vecteur et le gène cible est exprimé.

i. *Test d'expression en bactéries*

Après transformation de bactéries *E. coli* BL21(DE3) RIL ou Rosetta2 par le plasmide d'intérêt, 2mL de LB supplémenté de(s) antibiotiques approprié(s) par une colonie isolée est incubé pendant 6 heures à 37°C sous agitation. Cette pré-culture est ensuite utilisée pour inoculer une culture liquide de 50 mL de LB. Lorsque les cellules atteignent une densité entre 0.6 et 0.8 à 600 nm, la température modifiée ou non selon la ou les protéines en question, puis l'expression est induite en rajoutant 0.5 mM d'IPTG et les cultures sont laissées sous agitation quelques heures selon la température. Les tests d'expression permettent également de tester plusieurs tampons différents et différentes concentrations salines pour trouver les conditions optimales. Une fois l'optimisation effectuée, on peut passer à l'expression dans des plus grands volumes.

ii. *Tests de co-expression en bactéries*

Les tests de co-expression vont permettre de vérifier la formation d'un complexe et de définir les zones d'interaction en testant différentes régions de chacune des protéines. La co-expression présente plusieurs d'avantages pour les complexes protéiques. En effet, parfois des protéines insolubles dans *E. coli* se rendent solubles en leur ajoutant leur partenaire ce qui améliore leur repliement. La co-expression consiste en l'expression de deux ou plus protéines dans une même bactérie ce qui permet leur co-solubilisation. Trouver les conditions optimales est une étape cruciale dans la co-expression puisque chaque protéine mise en jeu a des caractéristiques différentes. Comme dans l'expression de protéines seules, des tests d'expression sont nécessaires dans la co-expression

surtout pour tester les concentrations salines qui influencent beaucoup la formation de complexes. Deux concentrations salines sont testées au minimum. Un autre point important dans la co-expression est le type de l'étiquette de purification et sa position. En effet, l'étiquette peut masquer le site d'interaction de la protéine avec son partenaire. Il est donc souvent nécessaire de tester des combinaisons différentes d'étiquette et de varier leurs positions.

2. Expression à grande échelle

a. Expression en bactéries

Les étapes de transformation bactérienne sont similaires au test de co-expression. Par contre, la préculture se fait pendant la nuit dans un volume de 100 ml de culture supplémenté de l'antibiotique approprié en agitation à 37 °C. Le lendemain, la culture de 1 litre se fait dans une fiole de 5 litres pour une meilleure oxygénation des cellules dans les mêmes conditions. Le volume d'ensemencement est calculé pour obtenir une DO à 600nm finale de 0.1 dans le milieu de culture. Quand la DO atteint entre 0.4 et 0.6 la température est réglée suivant les besoins de la protéine. L'IPTG est ensuite ajouté à une concentration finale de 0.5 mM ; la phase de production commence alors.

b. Expression en BHK21

Une préculture de cellules BHK21 est infectée à une concentration initiale de virus de 0,1 PFU/cellule et la préculture est incubée 2 jours à 37 °C, 5 % CO2. Une culture de cellules BHK21 à une concentration de cellules d'environ 10^6 – $1,5x10^6$ cellules/mL est ensuite infectée par la préculture à raison de 1/10ème du volume final et incubée 1 h à 37 °C, 5 % CO2. L'expression est induite à 1 mM d'IPTG pendant 24 h à 37 °C, 5 % CO2 sous agitation. Des fractions aliquantes de cellules sont prélevées et déposées dans des plaques 6 puits pour visualiser l'expression avant et après induction. La fluorescence jaune de la protéine eYFP est un marqueur visible d'une expression effective de la polyprotéine.

Les cellules sont culottées par une centrifugation de 20 min à 3000 rpm à 4 °C et le culot est lavé par du PBS (NaCl 140 mM, KCl 2,7 mM, Na2HPO4 8 mM, KH2PO4 1,8 mM, Tween 80 0,02 %) stérile. Après une nouvelle centrifugation et élimination du surnageant, le culot est congelé dans de la carboglace et conservé à – 80 °C.

3. Les différentes stratégies de formation de complexe

La formation d'un complexe *in vivo* et/ou *in vitro* est basée sur la stabilisation des protéines par leurs partenaires. Par cette méthode, on évite les techniques de solubilisation artificielle à l'aide de détergents ou des protéines de fusion telles que la GST ou la MBP qui solubilisent parfois les protéines en formant des agrégats solubles, ce qui est nuisible à la cristallisation. Nous avons ici cherché à former le complexe Vif-CBFβ-EloB-EloC-Cul5-A3G/F de manière à s'approcher au plus près des conditions naturelles d'association de ces protéines. Plusieurs stratégies de formation de complexe ont été envisagées : par co-expression, par co-broyage, par co-purification. Compte tenu du nombre important de combinaisons possibles, il n'était pas réalisable de tester l'ensemble des constructions, ce qui m'a poussé à chercher la plus forte probabilité d'obtenir des résultats rapidement en se focalisant sur les différents tests effectués.

a. La formation de complexe par co-expression

i. *Formation d'un ARNm polycistronique*

Cette partie a été expliquée en détails dans les paragraphes précédents. Brièvement, cette stratégie a été tentée dans le cas de l'expression du complexe dans l'organisme *E. coli* et consiste en l'expression des différents partenaires du complexe à partir d'un même vecteur ou de différents vecteurs exprimant séparément les protéines du complexe dans l'expectative que le complexe se forme au sein de la bactérie. Ces vecteurs de co-expression possèdent un « RBS» (Ribosome Binding Site) avant chaque phase codante ainsi qu'un codon start et codon stop pour chaque protéine. Il ne s'agit pas ici d'une polyprotéine mais de différentes protéines traduites indépendamment l'une de l'autre.

Différentes conditions ont été testées (tampon, sels, tag, position du tag, ordre des gènes, température d'induction, souche d'*E. coli*, etc) pour arriver à obtenir le complexe entier mais toutes les tentatives se sont avérées peu prometteuses à cause de l'insolubilité d'A3F dans le système procaryotique.

ii. *Formation d'une polyprotéine*

Cette méthode a été appliquée dans le cas de l'expression du complexe en cellules de mammifères et a été détaillée précédemment. Contrairement à la méthode utilisée pour la co-expression dans *E.coli*, ici tous les gènes codants pour les différentes protéines seront concaténés l'un à la suite de l'autre mais la transcription et la traduction seront régulées de la même façon pour les différentes protéines. Le plasmide d'expression comporte un seul codon d'initiation au début de la phase codante et un seul codon stop après le dernier gène. De cette façon, une polyprotéine sera formée puis clivée *in vivo* au niveau des sites de clivage à la protéase TEV présents entres les gènes, la protéase étant exprimée en premier. Après individualisation des protéines dans la cellule, le complexe

sera formé *in vivo*, ce qui aide à la structuration et au bon repliement du complexe, donc à sa solubilisation.

b. La formation de complexe par co-broyage

Cette technique consiste à produire séparément les différents partenaires du complexe dans *E. coli*. Compte tenu des conditions optimales d'expression de chaque protéine, comme par exemple la température d'induction, temps d'induction, concentration en IPTG, il peut être avantageux d'exprimer les protéines séparément puis de les mélanger pour la lyse cellulaire.
Les cellules sont mélangées et lysées dans des proportions adéquates en fonction de la stœchiométrie du complexe et du taux d'expression lors de la production.
La limitation de cette technique est le manque de reproductibilité ; les quantités de protéines varient entre chaque préparation, ce qui nécessite une mise au point pour chaque nouvelle production et pour chaque protéine du complexe.

c. La formation du complexe *in vitro*

Plus le complexe contient de protéines différentes, plus le nombre de plasmides utilisés dans la co-expression augmente. Les transformations bactériennes seront donc plus difficiles à cause des différents antibiotiques utilisés dans la sélection des colonies recombinantes. Dans ce cas, il est intéressant de combiner l'expression et la co-expression. Par exemple, dans le cas du complexe Vif-CBFβ-EloB-EloC/Cul5/A3Fct.

4. Purification des protéines et du complexe

La purification des protéines est réalisée à partir des cultures réalisées précédemment. La purification commence par une étape de lyse cellulaire. Parmi les moyens diversifiés qui existent pour la lyse, nous avons choisi deux systèmes : la French Press pour les cellules *E. coli* et la sonication pour les cellules BHK21. Après l'étape de lyse, une purification sur une colonne d'affinité pour les étiquettes 6xHis, Strep et GST est d'abord réalisée. Après cette étape, une purification par tamis moléculaire est nécessaire pour augmenter la pureté.

a. La lyse des cellules

i. *French Press*

Les culots bactériens provenant d'un litre de culture d'*E. coli* sont suspendus avec 25 mL de tampon de lyse défini au cours des tests d'expression. La French Press est un appareil qui va casser la membrane plasmique des cellules suite à leurs passages dans une valve à haute pression (1,6 mPa). Elle utilise une pompe hydraulique extérieure qui poussera un piston dans un large cylindre contenant l'échantillon liquide, ce dernier sera soumis à une pression très élevée, les cellules passeront dans un espace plus petit qu'elles, ce qui déchire leurs membranes.

ii. *La sonication*

Les culots des cellules BHK21 seront resuspendus dans 25 mL de tampon de lyse adéquat additionné de DNase (2.5 µg/mL) et d'inhibiteurs de protéase 1X (Halt Protease inhibitor Cocktail 100 x, Thermo Scientific). La lyse s'effectue à l'aide d'un sonicateur équipé d'une sonde de diamètre varié, durant 1 min de temps effectif par des impulsions d'ultra-sons toutes les 3 sec à 40 % d'amplitude à 4°C. La sonication dégage de la chaleur, afin d'éviter la détérioration de la protéine, l'échantillon doit rester au froid. En fin de lyse, l'aspect général du lysat doit être homogène et fluide.

Après la lyse des cellules, les débris bactériens seront culotés par ultracentrifugation pendant une heure à 45000 g à 4°C. Le surnageant contenant l'extrait soluble est récupéré, et immédiatement injecté sur la colonne d'affinité.

b. Chromatographie par affinité

La chromatographie d'affinité est basée sur l'interaction réversible entre une protéine (ou un complexe) et un ligand spécifique couplé à une matrice de la résine. Ce ligand doit être spécifique du tag d'affinité fusionné à la protéine d'intérêt. La protéine sera retenue grâce à son tag puis éluée spécifiquement par un ligand libre compétitif. Cette technique permet de séparer sélectivement la protéine d'intérêt des protéines bactériennes. Les trois types de tag utilisés pour les purifications sont les étiquettes Histidines (6X) et (10X) ou Flag-His, le Twin-strep (RRP) et la Gluthation S-Transférase (GST).

TAG	Séquence	Ligand couplé à la matrice	Ligand Compétiteur
Flag-Histidine (FH10P)	DYKDDDDK HHHHHHHHHH	Cations métalliques divalents Ni^{2+} ,Co^{2+} , Zn^{2+}	Imidazole pH
Twin-Strep (RRP)	SAWSHPQFEK(GGGS)2GGSA WSHPQFEK	Strep-tactin	Desthiobiotine
Gluthation S-Transférase (GST)	Protéine GST de 211 acides aminés (26KDa)	Gluthation	Gluthation réduit

Tableau 15 : Les différentes étiquettes utilisées pour la purification des protéines et leurs caractéristiques.

i. *La Gluthation S-transférase (GST)*

La GST est une protéine de fusion de 26000 Da qui, exprimée dans *E. coli*, conserve l'activité enzymatique et la capacité de dimériser. Le glutathion immobilisé sur une matrice de sépharose est utilisé comme colonne d'affinité pour purifier le lysat bactérien. L'élution de la protéine se fait avec du glutathion réduit (50 mM), qui va prendre la place de la GST sur la résine de glutathion.

La GST se replie rapidement en une protéine stable et très soluble après sa traduction, ce qui peut permettre d'améliorer l'expression et la solubilité de certaines protéines difficiles à replier. La GST est utilisée comme étiquette de purification dans le cas des protéines insolubles.

Cette méthode de purification a été utilisée pour l'expression des protéines APOBEC3G/3F (complet ou en domaines séparés).

Etapes de purification de CD2-A3G	Tampons
Lyse	50 mM Na_2HPO_4/NaH_2PO_4 pH7 100 mM NaCl 1 mM DTT
Elution de la colonne Glutathion	50 mM TrisHCl pH7 10 mM Gluthation réduit 10 mM DTT
Tampon PreScission	20 mM TrisHCl pH7 150 mM NaCl 1 mM DTT 0.5 mM EDTA
Tampon de stockage de CD2-A3G	50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5 100 mM NaCl 1 mM DTT
Tampon de régénération des billes de Glutathion	2V guanidine hydrochloride 5V PBS 4V EtOH 70 % 5V PBS 5V EtOH 20%
Tampon de test d'activité	20 mM Tris pH 7.5 30 mM NaCl 5 mM DTT 10 µM $ZnCl_2$

Tableau 16 : Les différentes compositions de tampons utilisés durant les étapes de purification du domaine C-terminal d'A3G.

ii. *Chromatographie d'affinité par Ni^{2+}*

Ce type de chromatographie permet la purification des protéines portant un tag polyhistidine. Le cation métallique Ni^{2+} (ou autre) est immobilisé par des groupements imido-diacétique greffés via un espaceur de sept atomes sur une matrice de type sépharose. Le Nickel interagit avec les Histidines

de la protéine purifiée. Des étapes de lavages avec de faibles concentrations en imidazole permettent ensuite de purifier la protéine et d'éliminer les protéines se liant non spécifiquement à la résine. L'élution se fait à une concentration plus élevée (200-400 mM Imidazole) et suffisante pour séparer la protéine des cations divalents. Si l'on ne connait pas la concentration d'élution à l'imidazole et pour optimiser la purification, une élution par gradient d'imidazole est réalisée.

iii. *Strep-tag®-Strep-tactin®*

Ce système développé par IBA repose sur le principe de l'interaction biotine-streptavidine. Le strep-tag® est capable de se lier à la streptavidine au site d'interaction de la biotine. Le Strep-tag® est un petit peptide dont ni la longueur ni la séquence n'affecte la protéine recombinante. Il n'interfère pas non plus avec le repliement ou l'activité de la protéine. La strep-tactin® a été conçue comme alternative à la streptavidine. Elle possède 100 fois plus d'affinité pour le Strep-tag que la streptavidine. Après des étapes de lavages, l'élution de la protéine recombinante purifiée se fait par addition de la desthibiotine (2.5 mM) au tampon. La desthiobiotine est un analogue réversible à la biotine, stable et de faible coût.

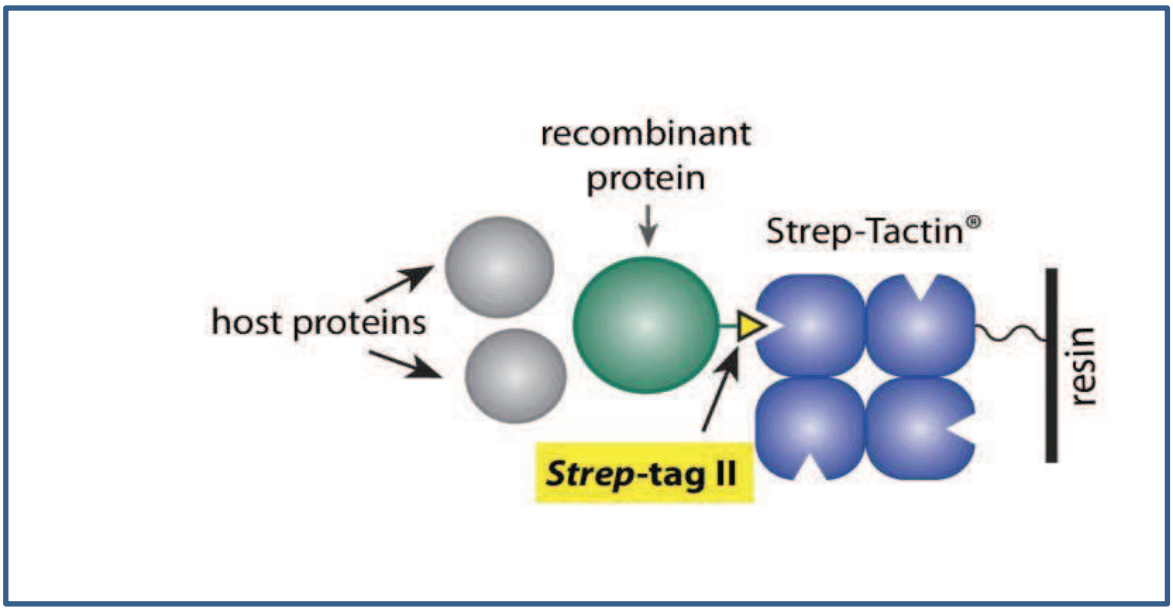


Figure 57 : Chromatographie d'affinité Strep-tag®-Strep-tactin®

c. Chromatographie d'exclusion de taille

Cette technique provient en deuxième étape après la purification sur une colonne d'affinité. Elle permet de séparer les protéines en fonction de leur taille et leur forme. Elle a également été utilisée au cours de ma thèse pour le dessalage d'échantillons protéiques et l'élimination des agrégats. Au laboratoire, on utilise les colonnes Superdex 75 et 200 de GE Healthcare selon la taille des protéines ou complexes d'intérêt. Les protéines sont éluées dans l'ordre décroissant de leur poids moléculaire dans la gamme de résolution de la colonne. Les protéines très encombrées ou très grosses passeront à l'extérieur du tamis de la résine constituée d'une matrice de dextrane et d'agarose formant des granules poreux et donc sortiront les premières de la colonne. Les molécules plus petites passeront à travers les pores et seront retardées.

5. Purification des acides nucléiques

La colonne échangeuse d'ions DNA-PAC100 (Dionex, Thermo Fisher Science) a été utilisée dans le cadre de la purification des acides nucléiques : (1) les ARN étudiés lors des expériences d'ITC effectués entre les différentes régions de l'ARN viral et la protéine Vif, (2) les ADN purifiés afin de mettre en évidence l'activité des protéines APOBEC3G produits. Les acides nucléiques ont été purifiés dans des conditions dénaturantes. Deux tampons sont utilisés pour appliquer un gradient de sels : un tampon A contenant 20 mM MES pH 6.5 et 4M urée et un tampon B contenant 400 mM $NaClO_4$, 20 mM MES pH 6.5 et 4 mM urée. La purification est effectuée avec un système HPLC, à un débit de 0.5 mL/min et les acides nucléiques sont élués à un gradient de 40% à 50 % de tampon B pendant 30 min. Pour les ARN, les fractions purifiées sont par la suite lavées et reconcentrées sur un centricon 10 K (Millipore) pour enlever toute trace de sel, puis placées dans un tampon identique à celui de la protéine Vif pour les expériences d'ITC (détails dans la partie 1 des Résultats).

Pour les ADN, les expériences étaient purement analytique pour vérifier que l'ADN a été bien clivé : test d'activité détaillé dans (paragraphe 5.f ci-dessous).

II. Caractérisation physico-chimique et fonctionnelle des protéines

Différents paramètres physico-chimiques vont permettre de caractériser l'échantillon avant de réaliser les tests de cristallisation. Parmi ces tests, la technique la plus souvent utilisée durant ma thèse pour caractériser l'homogénéité et la monodispersité de l'échantillon est la DLS. L'identification des protéines et la détermination de la stœchiométrie d'un complexe est effectué par la spectrométrie de masse. Le repliement des protéines est étudié par dichroïsme circulaire et les activités des protéines ont été vérifiées par ITC ou par des tests d'activités spécifiques aux protéines. Une fois les protéines bien caractérisées et fonctionnelles, des tests de cristallisation sont menés afin d'obtenir des cristaux en vue de déterminer la structure tridimensionnelle.

A. Analyse SDS-PAGE

L'analyse s'effectue par électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE). Pour chaque construction ou combinaison de vecteurs testés, plusieurs dépôts sont effectués :

- Le témoin avant induction (AI) qui permet de vérifier qu'il n'y a pas de fuite du vecteur et de comparer contre le témoin après induction.
- Le témoin après induction (Post-induction PI) en fin de culture permet d'évaluer l'expression des protéines.
- L'extrait brut (EB) permet d'évaluer la solubilité en comparant contre le surnageant (partie soluble).
- Le culot (C) permet d'identifier la partie non soluble qui forme des corps d'inclusions dans le culot.
- Le surnageant (S) est l'extrait soluble et permet d'évaluer la solubilité des protéines par comparaison contre l'extrait brut.
- Le Flow-Through (FT) représente ce qui passe à travers les colonnes d'affinité utilisés au cours de ce projet et permet d'évaluer le repliement des protéines ainsi que la formation du complexe.
- Les lavages (Wash W) représente les protéines se décrochant des colonnes dans des conditions de sels plus élevées et permet d'évaluer la pureté de l'échantillon.
- L'Elution (E) représente les élutions des colonnes d'affinité.

B. La diffusion dynamique de lumière (DLS)

Les coefficients de diffusion sont mesurés avec l'appareil Zetasizer Nano S (Malvern). Après ultracentrifugation, la protéine est disposée dans une cuvette trUView de Biorad puis irradiée par un faisceau laser à une longueur d'onde de 780 nm.

Lors d'une expérience de DLS (Dynamic Light Scattering), on observe la fluctuation de l'intensité de la lumière monochromatique diffusée par les particules en solution sujettes au mouvement brownien au cours du temps. Cette fluctuation de l'intensité lumineuse peut être ensuite transformée en une fonction d'autocorrélation qui permet de déterminer le coefficient de diffusion (D). En faisant l'hypothèse que les particules en solution ont une forme sphérique, ce dernier est relié au rayon hydrodynamique (Rh) des particules diffusantes à travers l'équation de Stockes-Einstein :

$$D = \frac{KT}{6\pi\mu Rh}$$

Pour une viscosité et une température données, les particules de petite taille se déplaçant plus rapidement que celles de grande taille. Enfin, une distribution de tailles pour les particules en solution est calculée. Une solution propice à la cristallisation et aux études biophysiques doit être monodisperse, sans présenter d'agrégation. La mesure du rayon hydrodynamique permet une estimation de la masse moléculaire du multimère. Toutes les données sont corrigées en tenant compte de la viscosité du solvant et de l'indice de réfraction.

C. Dichroïsme circulaire

Une analyse de la structure secondaire de la protéine Vif permet d'évaluer si le protocole de renaturation conduit bien à une protéine correctement structurée. Le CD est une spectroscopie d'absorption dans l'UV proche et lointain qui donne des informations sur la structure secondaire dans les macromolécules biologiques en solution. Le CD se base sur le principe qu'une molécule chirale absorbe différemment une onde polarisée circulaire gauche et une onde polarisée circulaire droite. Cette différence d'absorption sera ensuite mesurée et le spectre CD obtenu varie suivant le contenu de la protéine en hélices α, feuillet β, coudes et en régions non structurées. Ces spectres sont obtenus en utilisant des longueurs d'onde variant généralement entre 260 et 190 nm. Une analyse par déconvolution des spectres de CD permet de quantifier la répartition en éléments de structure

secondaire de la protéine. Le spectromètre JASCO J815 CD disponible à l'Ecole Supérieure de Biotechnologies de Strasbourg (ESBS) a permis une étude de la structure secondaire de Vif.

.

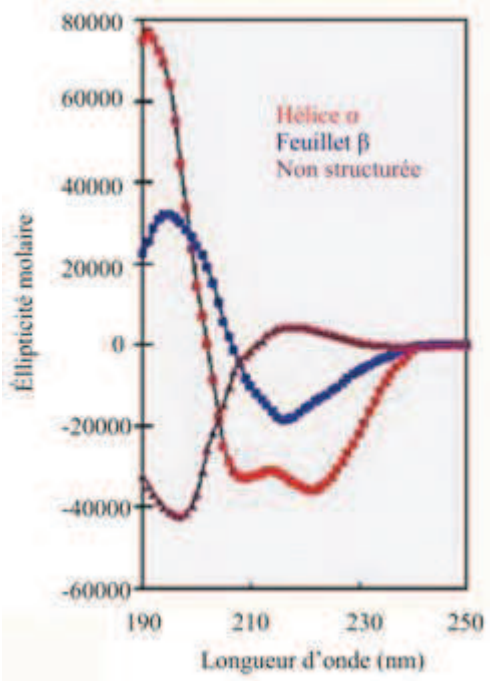


Figure 58 : Courbes de spectres standards de dichroisme circulaire pour des peptides se retrouvant dans une structure secondaire définie

D. Western Blot

Cette technique a été uniquement utilisée dans le cadre de l'identification des protéines après leur production dans le système eucaryotes à partir des cellules BHK21.

Les protéines, migrées sur gel dénaturant SDS-PAGE, sont transférées en 10 min sur une membrane commerciale de nitrocellulose ou PVDF (Biorad) à l'aide de l'appareil Trans-Blot Turbo Transfer System (Biorad). La membrane est ensuite saturée par une solution de lait 5 % dilué dans du TBST (Tris HCl pH 7,5 50 mM, NaCl 150 mM, Tween 0,02 %) pendant 1 h à 4 °C sous agitation. La membrane est incubée avec un anticorps primaire de souris dilué dans du lait 5 % - TBST (anti-Vif 1/10000ème ; anti-Flag 1/1000ème) pendant une nuit à 4 °C sous agitation. La membrane est ensuite lavée deux fois 5 min avec du TBST puis mise en présence, pendant 2 h à 4 °C sous agitation, d'un anticorps secondaire anti-souris couplé à la péroxydase HRP, dilué au 1/10000ème dans du TBST. La membrane est finalement lavée deux fois 5 min dans du TBST puis de même dans du TBS. Les

protéines sont révélées par chemiluminescence après incubation de la membrane dans la solution d'ECL (Biorad) et exposition sur film photographique.

E. Spectrométrie de masse

Les protéines produites individuellement ou en complexe ont été identifiées tout au long de ce projet par la spectrométrie de masse à la plateforme de l'IBMC.

Trois méthodes ont été principalement utilisées : (i) l'identification des protéines à partir de bandes de gel SDS-PAGE en MALDI-TOF, (ii) l'identification des protéines en solution en LC-MS/MS, (iii) l'identification des protéines à partir de la masse entière.

1. Analyse des bandes de gel en MALDI-TOF

J'ai réalisé le dessalage des échantillons, isolés sur gel SDS-PAGE, par des bains successifs de bicarbonate d'ammonium 25 mM et d'acétonitrile 50 %. Des étapes de lavages de 15 minutes à température ambiante et sous agitation ont été répétées deux fois et le liquide est enlevé complètement avant l'addition du solvant. Après les lavages, les bandes sont séchées 20 minutes avant réduction (10 mM DTT/25 mM NH_4HCO_3 à 56°C pour 45 minutes) et alkylation (25 mM iodoacétamide/25 mM NH_4CO_3, 45 minutes) à température ambiante. Après encore trois autres étapes de lavage, le gel est séché sous vide pendant 20 minutes et puis réhydraté par une solution de trypsine (Promega, V5111) à une concentration finale de 10 ng/µL et incubé sur la nuit à température ambiante. Les peptides trypsiques sont extraits du gel par des agitations de 30 minutes avec 30 µL de 35%H2O/60%CAN/5%HCOOH puis par une sonication de 15 minute à 180 W.

Pour une analyse MALDI, 1µL du surnageant de digestion et 4µL de 0.1% TFA (Trifluoroacetic acid) est déposé sur une cible Anchorchip (Format spots 96 de Bruker) sur laquelle a déjà été apposée de la matrice HCCA (acide α-cyano-4-hydroxycinnamique) et d'un calibrant standard (Bradykinin (1-7) m/z = 757,39916 ◊ACTH-clip (7-38) m/z = 3617,92890). Les peptides co-cristallisent avec la matrice resolubilisée par le 0,1% TFA. Grâce au cœur hydrophobe des emplacements de la cible Anchorchip, l'échantillon est plus concentré au centre de chaque dépôt lors de l'analyse. Ce type de plaque est propice à l'identification des protéines. L' -HCCA, matrice utilisée en routine, est préconisée pour les analyses de peptides sur une assez large gamme de masse. Elle permet d'effectuer des mesures de

masse sans discriminer les plus petits peptides des plus gros. La cible est ensuite insérée ensuite dans le spectromètre de masse de type MALDI-TOF/TOF (UltraflexII, Bruker Daltonics).

2. Analyse en solution en LC-MS/MS

a. En conditions dénaturantes

Pour les échantillons liquides, les analyses ont été réalisées en LC-ESI-TOF par l'équipe de la plateforme. Pour la digestion trypsique, l'échantillon est précipité avec 0.1 M acétate d'ammonium dans 100 % méthanol et les protéines suspendues dans 50 mM de bicarbonate d'ammonium. Après une étape de réduction-alkylation, les protéines seront digérées sur la nuit avec 1/25 (W/W) de trypsine.

Après la digestion, les peptides sont resuspendus dans 15 µL de solvant composé d'eau et de 0.1 % FA. Le mélange de peptides est analysé en utilisant un système NanoLC-2DPlus (ABSciex, Concord, Ontario, Canada) couplé à un spectromètre de masse TripleTOF de ABSciex

L'interrogation de la banque de données Mascott a permis l'identification des protéines par comparaison avec les mesures des masses des fragments trypsiques obtenues.

b. En conditions natives

En conditions non dénaturantes, les mesures de la masse du complexe protéique non digéré à la trypsine sont réalisées dans 50 mM d'acétate d'ammonium avec 10% d'alcool isopropylique pour conserver la conformation native. Le complexe est injecté à 2 µM.

F. Méthylation des Lysines

Les modifications chimiques peuvent faciliter la cristallisation des protéines ou de domaines solubles ou désordonnées. La méthylation des lysines affecte les propriétés physico-chimiques de la protéine et diminue sa solubilité [226]. Seules les lysines exposées au solvant vont être modifiées. Cette approche a été testé afin d'essayer de cristalliser le complexe Vif/CBFβ/EloB/EloC en utilisant un kit de méthylation (JBS Methylation Kits).

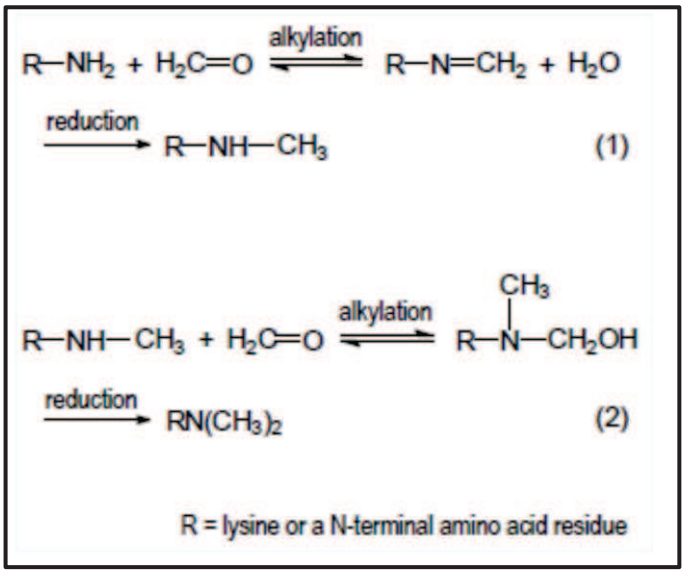


Figure 59 : Réaction de méthylation de lysines. (1) La première étape de la réaction consiste en une formation de base Schiff entre le groupement amine de la lysine et le formaldéhyde, réduit après en une amine secondaire. (2) la lysine monométhylée réagit rapidement avec le formaldéhyde du fait que le son pKa est légèrement inférieur à celui de la lysine même (JBS Methylation Kit).

Le protocole de la méthylation des lysines consiste à dialyser les protéines contre un tampon 50 mM HEPES pH 7,5, 250 mM NaCl. Les tampons contenant des groupements amines ne peuvent pas être utilisés. On ajoute à la protéine concentrée 20 à 100 µL de réactif A (1 M Diméthyl borane complex) et 40 à 200 µL de réactif B (1 M formaldéhyde). La réaction se déroule pendant 24 heures. Après, l'échantillon est centrifugé pour éliminer les agrégats causés par la réaction de méthylation, puis chargée sur une colonne de tamis moléculaire. Les fractions qui correspondent à la bonne taille du complexe seront collectées et concentrées. La méthylation de toutes les lysines est vérifiée par spectrométrie de masse en native.

G. Microcalorimétrie isotherme (ITC)

La microcalorimétrie ITC (Isothermal Titration Calorimetry) est une technique qui permet d'obtenir en une seule mesure l'ensemble des paramètres thermodynamiques d'une interaction entre deux molécules : stœchiométrie (n), constante d'affinité (Ka), variation d'enthalpie (ΔH) et variation d'entropie (ΔS). Contrairement à d'autres approches comme la fluorescence ou la résonance plasmonique de surface, elle ne nécessite ni marquage des protéines (ou acides nucléiques), ni fixation sur une surface solide.

L'appareil d'lITC est composé d'une enceinte adiabatique contenant deux cellules : une cellule contenant l'échantillon et une cellule de référence qui servira comme standard interne. Le ligand est placé dans la seringue. Une différence de température constante ΔT est maintenue entre la cellule de l'échantillon et la cellule de référence (calorimètre à compensation).L'appareil utilisé dans ce projet est l'ITC-200 de Microcal disponible au laboratoire.

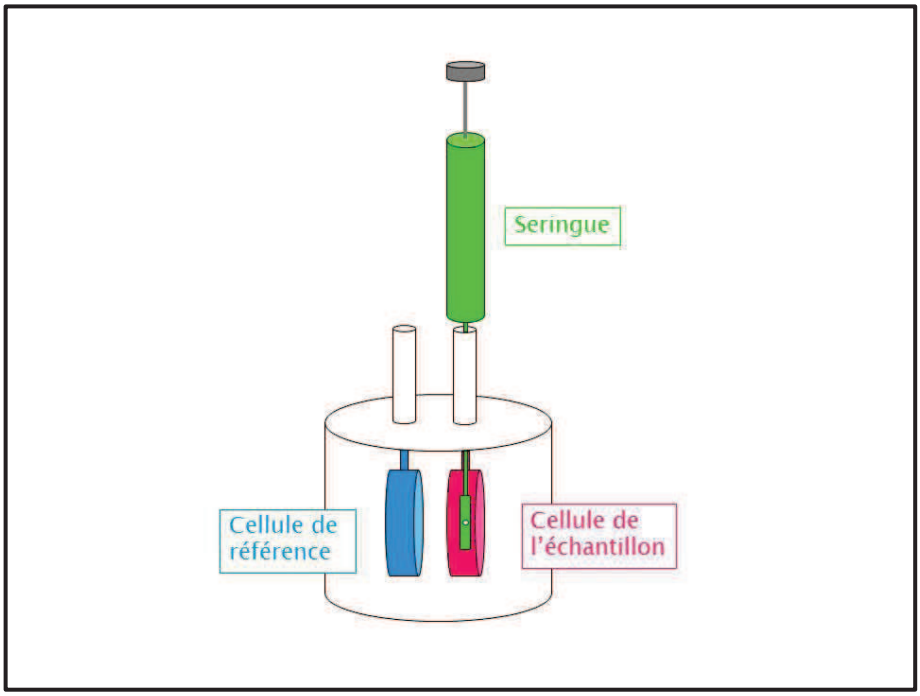


Figure 60 : Montage d'une expérience de microcalorimétrie

Le but de cette expérience est de quantifier le dégagement ou l'absorption de chaleur causé par l'injection du ligand dans l'échantillon en fonction du temps. Lorsque l'échantillon est saturé par le ligand, un plateau apparaitra du fait d'une simple dilution du ligand..Ces valeurs sont ensuite ramenées au ratio molaire de ligand injecté par rapport à la quantité de protéines dans la cellule, ce qui permet d'obtenir une courbe isotherme. Suite à chaque injection de ligand dans la cellule, l'appareil va mesurer la chaleur à absorber ou fournir afin que la différence ΔT entre cellule de mesure et de référence reste constante.

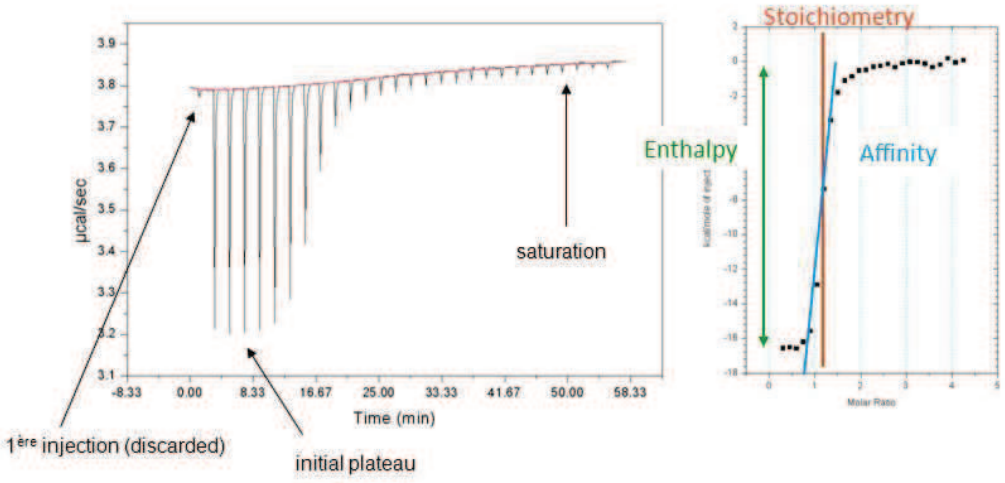


Figure 61 : Résultats typiques d'une expérience de microcalorimétrie

L'ITC a été utilisée pour évaluer les paramètres thermodynamiques d'interaction entre la protéine Vif et les différents domaines DIS et TAR de l'ARN génomique viral (Article joint, « Résultats Partie 1»)

H. Test d'activité d'APOBEC

Ce test d'activité a été réalisé sur les protéines APOBEC produits dans les deux systèmes, bactérien et en cellules de mammifères, afin de vérifier l'activité de désamination d'APOBEC. Ce test consiste à mettre la protéine A3G produite (10 µM) en présence d'un ADN substrat simple brin (entre 10 et 40 µM), de 39 nucléotides possédant le motif CCC de désamination d'A3G, dans un tampon contenant du zinc (20 mM Tris HCl pH 7,5, 30 mM NaCl, 10 µM MgCl2, 5 mM DTT) et le tout est incubé 2 h à 37 °C. Une extraction phénol-chloroforme est effectuée suivi d'une précipitation à l'éthanol pour ne récupérer que l'ADN, qui sera repris dans 20,5 µL de tampon zinc. 10 U d'uracile ADN glycosylase (NEB) et 2,5 µL de tampon uracile ADN glycosylase ou UDG (NEB) sont ajoutés, et le milieu réactionnel de 25 µL est incubé 2 h à 37 °C. Finalement, une hydrolyse alcaline est réalisée par un chauffage 20 min à 95 °C en présence de 0,2 M de NaOH, afin de cliver le brin d'ADN au niveau du site abasique crée par l'action de l'UDG. En parallèle, des contrôles sont réalisés : l'ADN substrat non soumis à la réaction de désamination, utilisé comme contrôle négatif, et un ADN produit possédant déjà le motif final CCU, utilisé comme contrôle positif, sont incubés dans les mêmes conditions avec de l'UDG et soumis à l'hydrolyse alcaline.

Les échantillons sont injectés sur une colonne échangeuse d'anions Nucleopac PA-100 (Dionex) chauffée à 75 °C et l'ADN retenu sur la colonne sera élué par un gradient de sel de 0 % à 80 % de tampon B, effectué par un mélange des tampons A (20 mM MES pH 6,2 et 4 M urée) et B (20 mM MES pH 6,2, 4 M urée et 400 mM NaClO4).

ADN substrat : 5'-AAAGAAAAGAGAAA**CCCA**AAGAGGAAAGGTGAGGAGGA-3'
ADN produit : 5'-AAAGAGAAAGAGAAA**CCUA**AAGAGGAA AGGTGAGGAGGA-3'

I. Tests de cristallisation

La résolution de la structure d'une macromolécule permet de mieux comprendre son mode d'action au niveau moléculaire. Pour déterminer la structure tridimensionnelle des molécules microbiologiques, notre laboratoire utilise la cristallographie comme technique structurale.

La cristallographie requiert l'obtention de protéines en grandes quantités et de bonnes qualités (pures, monodisperses, etc) pour l'obtention de cristaux.

La cristallisation de la protéine a pour but de passer d'un état liquide à un état solide ordonné selon un diagramme de phase qui comprend autant de dimensions que de paramètres testés. Cet état est obtenu suite à l'addition d'agents de cristallisation qui vont diminuer la solubilité de la protéine. D'autres facteurs jouent aussi un rôle important dans la cristallisation des protéines, tels que la concentration de la protéine, son pH, la température, la pureté et l'homogénéité de l'échantillon ainsi que la présence d'additifs tel que des petites molécules molécules ou des ligands de la protéine.

Les conditions initiales de cristallisation sont déterminées par un criblage de conditions en utilisant des kits commerciaux (Hampton Research) qui rassemblent des compostions variées en agents cristallisants, en sels, pH et tampons. Le criblage est réalisé par diffusion de vapeur en gouttes assises d'un volume de 100 à 300 nL dans des plaques de 96 puits. On utilise pour cela le robot de cristallisation disponible au laboratoire (Mosquito®, TTP Lab Tech Inc.). Les plaques seront aussi incubées à des températures différentes (4°C, 25°C, 37°C).

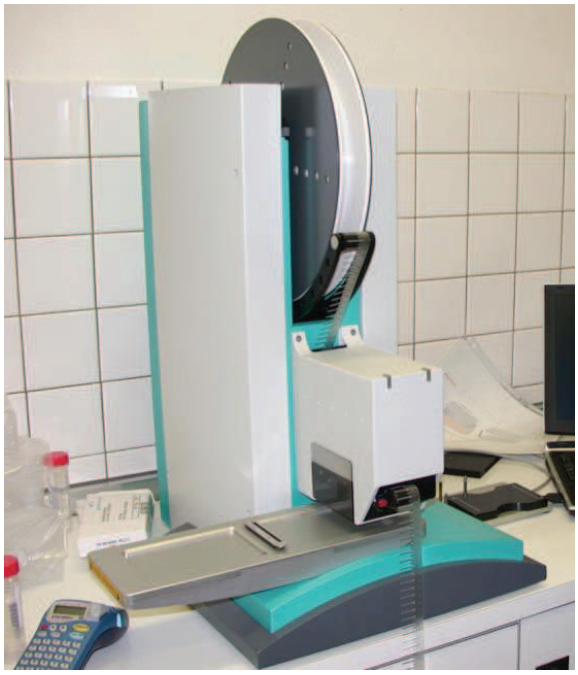


Figure 62 : Robot de cristallisation type Mosquito®. Matériel utilisé pour le criblage des conditions de cristallisation.

III. Criblage haut-débit « ESPRIT »

De nombreuses approches biophysiques sont limitées par le besoin de grandes quantités de protéines. Pour améliorer le rendement d'expression et la solubilité d'une protéine, il est fréquent de rechercher des constructions plus courtes qui correspondent à des assemblages de domaines ou des domaines isolés. La technologie ESPRIT (Expression of Soluble Proteins by Random Incremental Truncation), développée à l'EMBL de Grenoble par Darren Hart, permet la recherche rapide et hautement automatisée de constructions solubles par criblage sur des cibles particulièrement difficiles. La stratégie ESPRIT se compose de trois étapes (Fig. 63) : (1) construction d'une bibliothèque aléatoire de plasmides par troncation (en 5' et/ou en 3') avec une exonucléase ; cette étape va générer des milliers de colonies après transformation dans E.coli ; (2) expression des protéines sur une matrice de colonies à haute densité (jusqu'à 56 000 colonies, généralement 28 000 en duplicata), chaque colonie exprimant une seule construction ; cette étape permet un premier crible des constructions solubles ; (3) au cours de la dernière étape, les protéines apparemment solubles à l'étape précédente sont exprimées à petite échelle et purifiées. Les clones correspondant aux constructions sélectionnées sont séquencés.

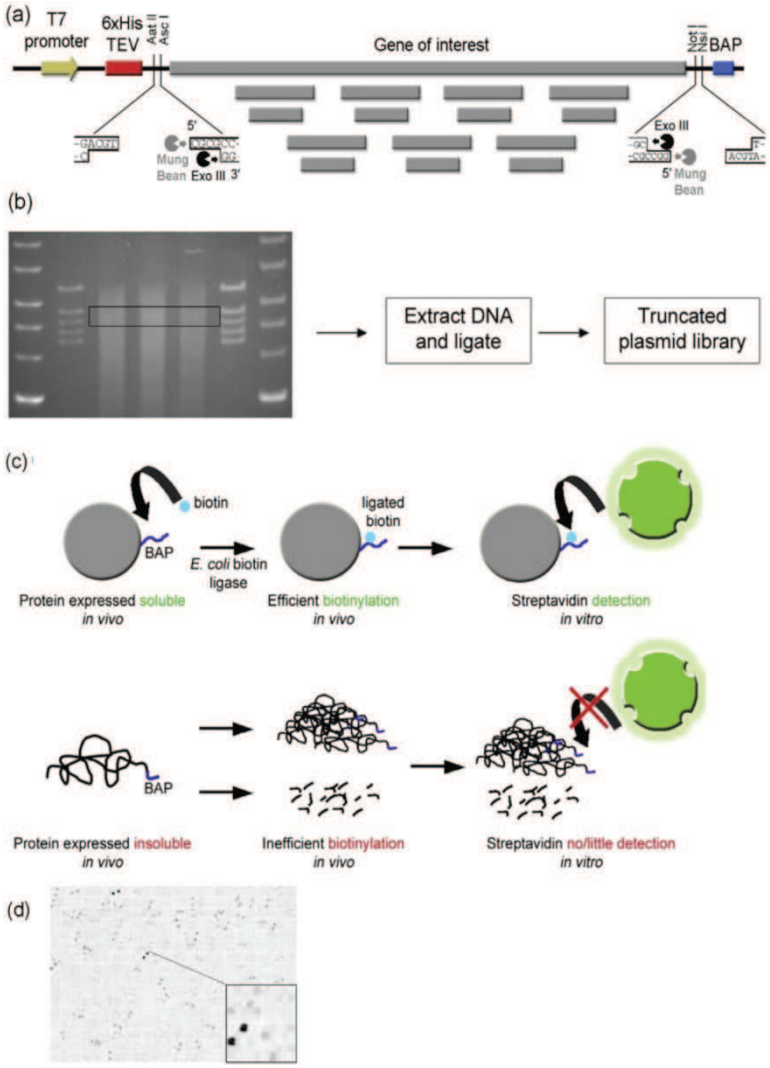


Figure 63: Schéma de la stratégie ESPRIT. (a) L'insert est bordé d'une étiquette histidine et d'un site protéase TEV d'une part, et d'un site peptide accepteur de biotine (BAP) d'autre part. Une troncation 5'→3' et/ou 3'→5' du vecteur est réalisée en utilisant l'exonucléase III du haricot Mungo. Deux paires de sites de restriction (AatII et AscI, NotI et NsiI) permettent des réactions unidirectionnelles séquentielles, générant la librairie de constructions. (b) Les vecteurs qui contiennent la librairie d'inserts de la taille désirée sont excisés d'un gel d'agarose (afin d'éviter la recherche de constructions de taille aberrantes). (c) Dans la matrice de colonies, les protéines solubles sont mieux biotinylées que les protéines insolubles ou agrégées. Une streptavidine fluorescente interagit avec le peptide signal BAP et permet une détection aisée des protéines biotinylées. (d) Exemple d'expression de protéine à l'aide d'une librairie de 26,880 constructions aléatoires. Adapté de [227, 228]

L'efficacité de la stratégie ESPRIT a déjà été prouvée sur la polymérase du virus de la grippe [228, 229] et pourrait faciliter l'identification de domaines solubles d'APOBEC3.

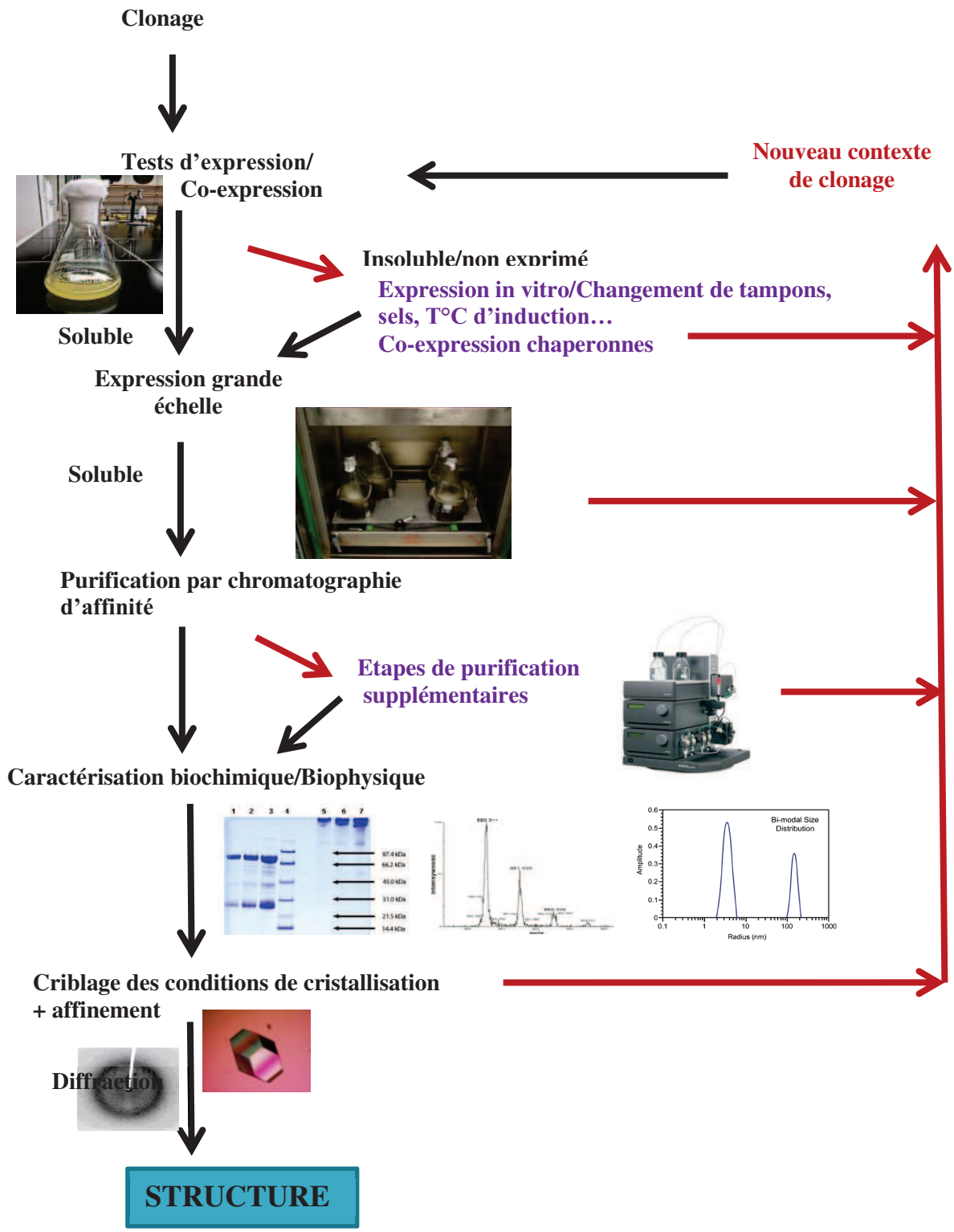


Figure 64 : Schéma récapitulatif de la stratégie utilisée au laboratoire pour aller du gène à la structure.

RESULTATS PARTIE 1 : IDENTIFICATION DES DETERMINANTS STRUCTURAUX ET DE SEQUENCE IMPLIQUES DANS L'INTERACTION ENTRE LE FACTEUR D'INFECTIVITE VIRALE VIF ET LES ARN TAR ET DIS DU VIH-1

Grâce aux grandes quantités de protéine Vif de bonne qualité que nous avons pu produire par dialyses successives dans un tampon dénaturant, nous avons caractérisé les interactions Vif/ARN viral par microcalorimétrie ITC. L'équipe de R. Marquet dans notre UPR avait montré précédemment que Vif interagit spécifiquement avec plusieurs structures localisées dans la région 5' non-traduite du génome viral, en particulier avec les séquences TAR et DIS [151, 153, 230]. Afin de disséquer les interactions entre Vif et ces deux séquences, nous avons utilisé des fragments d'ARN viral qui correspondent à ces deux tiges-boucles et entamé une étude par ITC.

Tout d'abord, il était primordial de vérifier la monodispersité et le bon repliement de la protéine Vif produite. Après une identification de Vif par spectrométrie de masse, des analyses en DLS (Dynamic Light Scattering) ont montré que Vif est bien monodisperse (indice de polydispersité de 16%) avec un rayon hydrodynamique de 12 nm en faveur d'une oligomérisation de la protéine. Afin d'analyser la structure secondaire de Vif, j'ai réalisé des mesures en dichroïsme circulaire dans l'UV lointain. Les résultats obtenus entre 186 et 260 nm montrent que la protéine Vif produite est partiellement structurée , composée de 20 % d'hélices α, 26 % de brins β, 23% de boucles et 30 % de structures désordonnées, ce qui est en accord d'autres données de la littérature [230, 231]. Ces observations nous permettent de mener les expériences prévues en ITC entre Vif et l'ARN génomique viral étant donné que la protéine est en bonne qualité.

Nous avons pu observer une interaction spécifique de Vif avec les séquences TAR et DIS, mais aucune interaction avec un ARN correspondant à la boucle de la Sarcine-Ricine de l'ARN 23 S qui a servi de témoin négatif. Des expériences ITC entre Vif et TAR ont montré une interaction exothermique de forte affinité (Kd=159 nM à 30°C dans nos conditions expérimentales) et spécifiques

(stœchiométrie 1 :1). En utilisant différents ARN mutants de la séquence TAR, nous avons pu montrer que la boucle latérale UCU de TAR est essentielle à l'interaction avec Vif car sa suppression abolit l'interaction spécifique de Vif avec TAR. La mutation de la boucle apicale de TAR n'a qu'un effet modéré sur l'interaction Vif/TAR. Cette boucle UCU étant le site d'interaction bien connu d'une autre protéine régulatrice de HIV-1, Tat, notre résultat suggère que des inhibiteurs connus de Tat/TAR qui ciblent la boucle UCU pourraient inhiber également l'interaction Vif/TAR et réduire l'encapsidation de Vif [232–234].

Des expériences ITC entre Vif et DIS ont montré une interaction exothermique (Kd=61 nM à 30°C). Cette interaction est spécifique mais de plus faible affinité en comparant avec l'interaction de Vif avec TAR. En procédant de la même manière avec différents mutants de la séquence DIS, nous avons pu montrer que la boucle latérale de DIS n'est pas reconnue par Vif, tandis que la boucle apicale joue un rôle important, probablement *via* des interactions hydrophobes car l'interaction est peu affectée par une augmentation de la concentration en sels. La tige adjacente de la boucle apicale du DIS semble également reconnue par Vif. Cette région du DIS ayant également été identifiée comme site d'interaction d'APOBEC3G, notre étude suggère que l'interaction de Vif avec le DIS pourrait participer à l'exclusion d'APOBEC3 des particules virales en l'empêchant d'interagir avec l'ARN viral afin d'être encapsidé. Un manuscrit décrivant ces résultats sur les interactions Vif/TAR et Vif/DIS a été publié dans la revue *RNA Biology* [154].

Publication 1: Sequence and Structure Requirements for Specific Recognition of HIV- TAR and DIS RNA by the HIV-1 Vif Protein

Séverine Freisz, Joelle Mezher, Lamine Hafirassou, Philippe Wolff, Yves Nominé, Christophe Romier, Philippe Dumas & Eric Ennifar

RESEARCH PAPER

RNA Biology 9:7, 1-12; July 2012; © 2012 Landes Bioscience

Sequence and structure requirements for specific recognition of HIV-1 TAR and DIS RNA by the HIV-1 Vif protein

Séverine Freisz,[1] Joelle Mezher,[1] Lamine Hafirassou,[1] Philippe Wolff,[1] Yves Nominé,[2] Christophe Romier,[3] Philippe Dumas[1] and Eric Ennifar[1,*]

[1]Architecture et Réactivité des ARN; CNRS/Université de Strasbourg; Institut de Biologie Moléculaire et Cellulaire; Strasbourg, France; [2]Equipe Oncoprotéines; Institut de Recherche de l'Ecole de Biotechnologie de Strasbourg; Université de Strasbourg; Illkirch, France; [3]Département de Biologie et Génomique Structurales; IGBMC, CNRS/INSERM Université de Strasbourg; Illkirch, France

Keywords: HIV, biophysics, RNA-protein interactions, microcalorimetry

The HIV-1 Vif protein plays an essential role in the regulation of the infectivity of HIV-1 virion and in vivo pathogenesis. Vif neutralizes the human DNA-editing enzyme APOBEC3 protein, an antiretroviral cellular factor from the innate immune system, allowing the virus to escape the host defense system. It was shown that Vif is packaged into viral particles through specific interactions with the viral genomic RNA. Conserved and structured sequences from the 5′-noncoding region, such as the Tat-responsive element (TAR) or the genomic RNA dimerization initiation site (DIS), are primary binding sites for Vif. In the present study we used isothermal titration calorimetry to investigate sequence and structure determinants important for Vif binding to short viral RNA corresponding to TAR and DIS stem-loops. We showed that Vif specifically binds TAR and DIS in the low nanomolar range. In addition, Vif primarily binds the TAR UCU bulge, but not the apical loop. Determinants for Vif binding to the DIS loop-loop complex are likely more complex and involve the self-complementary loop together with the upper part of the stem. These results suggest that Tat-TAR inhibitors or DIS small molecule binders might be also effective to disturb Vif-TAR and Vif-DIS binding in order to reduce Vif packaging into virions.

This manuscript has been published online, prior to printing. Once the issue is complete and page numbers have been assigned, the citation will change accordingly.

© Do not distribute.

Introduction

The HIV-1 Virion Infectivity Factor (Vif) is a small basic protein of 192 amino acids (23 KDa, pI = 10.7). It has been known for a long time to be essential for viral replication in some, but not all, cell lines.[1-5] In absence of Vif, human antiviral proteins from the APOBEC3 cytidine deaminase family are packaged into virions and exhibit anti-HIV-1 activity mostly by two mechanisms: (1) deamination of cytidines in the minus strand of the proviral DNA,[6-8] and (2) inhibition of the reverse transcription through a deaminase-independent mechanism.[9-11] APOBEC3G, formerly known as CEM15, was first identified as a potent endogenous inhibitor for replication of several retroviruses, including HIV-1.[8,12,13] Later on, other members of the APOBEC3 family have been identified as HIV-1 restriction factors: APOBEC3F,[12,14,15] APOBEC3B,[12,16] APOBEC3DE[17] and APOBEC3H.[18] The antiretroviral activity of APOBEC3DE[19] and APOBEC3H[17,20] is however still controversial and, with the exception of APOBEC3F, these proteins are poorly expressed in human lymphoid tissues and only lead to a moderate effect on HIV-1 infection.

To bypass the host defense system and counteract APOBEC3 action, Vif dramatically reduces APOBEC3G incorporation into the virion using several pathways (for a review on Vif-APOBEC3 interactions see ref. 21–23). First, Vif interacts with APOBEC3F/G and hijacks the human E3 ubiquitin ligase complex consisting of Cullin5, ElonginB, ElonginC and a zinc RING finger protein (Rbx), leading to a polyubiquitination of APOBEC3F/G and to its degradation by the 26 S proteasome.[24-28] It was recently shown that, in addition to the E3 ubiquitin ligase complex, the transcription cofactor CBF-β is also recruited by Vif and is required for the APOBEC3G degradation.[29-31] Interestingly, APOBEC3B[16] and APOBEC3H[18] have been reported to escape this Vif-mediated degradation. Besides APOBEC3 degradation, Vif negatively regulates APOBEC3G transcription through binding to 5'-UTR of APOBEC3G mRNA.[28,32,33] Finally, Vif also reduces APOBEC3G packaging into the virion by competing for the same Gag and viral RNA binding sites, thus leading to a physical exclusion of APOBEC3G from the budding virion.[27,32,34,35] It was indeed shown by several studies that Vif is an RNA-binding protein that specifically recognizes the viral RNA in vitro and in infected cells.[36-39]

Several conserved domains have been identified in Vif:[40] (1) a tryptophan-rich N-terminal domain responsible for RNA and APOBEC3 binding, (2) a HCCH motif that coordinates a Zn^{2+} ion and which is responsible for the binding with Cullin5, (3) an ^{144}SLQ(Y/F)LA^{149} motif, named "SOCS-box", responsible for

*Correspondence to: Eric Ennifar; Email: e.ennifar@ibmc.u-strasbg.fr
Submitted: 03/20/12; Revised: 04/18/12; Accepted: 04/23/12
http://dx.doi.org/10.4161/rna.20483

the interaction with Elongin B/C, (4) a proline-rich region containing a [161]PPLP[164] motif involved into Vif multimerization and (5) a C-terminal domain interacting with Gag and the cellular membrane (**Fig. 1A**).

Using RNase-T1 footprinting, an enzyme that cleaves after a guanine residue accessible to solvent, preferential Vif RNA binding sites have been mapped in the 5'-region of the genomic RNA.[37,41] The strongest affinity was observed for the apical part of the TAR stem-loop, the poly(A) stem and a short region in *gag*, whereas the primer-binding site (PBS), the SL1 and SL2 stem-loops, as well as a purine-rich single-stranded region adjacent to the Gag start codon have been identified as lower-affinity Vif-binding sites.[37] Binding to the upper part of the TAR stem-loop was subsequently confirmed on short RNA fragments using steady-state fluorescence spectroscopy.[41,42]

In the present work, we used Isothermal Titration Calorimetry (ITC) to study the binding thermodynamics of Vif binding to the HIV-1 genomic RNA TAR and Dimerization Initiation Site (DIS) stem-loops (**Fig. 1B**). Using various mutants of TAR and DIS mutants (**Fig. 1C**), we could define the sequence and structure requirements for the Vif specific binding to these regions of the viral RNA. Our results show that Vif specifically and tightly binds both TAR and DIS sequences in the low nanomolar range. Vif primarily binds the UCU lateral bulge of the TAR, and secondarily the apical loop. Regarding binding to the DIS loop-loop complex, we found that mutations of the DIS loop preventing DIS dimerization also affect Vif binding, whereas the purine-rich internal loop of the DIS stem is clearly dispensable. Our data also indicate that the upper part of the DIS stem might be important for protein binding. Taken together, these results open the perspective of developing drugs that would prevent Vif binding to the TAR and DIS stem-loops and, consequently, reduce Vif packaging. They also suggest that Vif might compete with APOBEC3G for binding the DIS stem-loop, leading to an exclusion of the restriction factor from the viral particle.

©2012 Landes Bioscience. Do not distribute.

Results

Biophysical characterization of the full-length HIV-1 Vif protein. It has been shown that Vif monomers self-associate into multimers both in vitro and in infected cells.[43,44] The [161]PPLP[164] motif is important for this Vif oligomerization, for viral infectivity,[42,45-47] and was also reported to interact with Elongin B.[48] Mutation of the PPLP motif into AALA reduces, but does not abolish, oligomerization.[42] This strong tendency of Vif to self-associate and aggregate[33,42,49,50] has also thwarted attempts to obtain large quantities of high-quality and homogenous protein. Here, we successfully expressed and purified milligrams of recombinant full-length wild-type Vif, typically about 13 mg of purified protein per liter of cell culture, thus allowing detailed ITC analysis of Vif/RNA interaction. Because the recombinant protein was always found insoluble when expressed in *E. coli*,[33,41,42,49-51] Vif was solubilized using guanidine hydrochloride and subsequently refolded by dialysis. The proper folding of the protein was assessed using far-UV circular dichroism (CD). Secondary structure calculations from the CD spectra yielded 20% α-helix, 26% β-strands, 23% turns and 30% of disordered structure (which was expected in the Gag/Membrane and Elongin B/C binding domains in absence of their partners), with good agreement between calculated and experimental data (**Fig. 2A**). This result shows that the produced protein is correctly folded after denaturation/renaturation and observed secondary structure is approximately in agreement with recently published Vif CD data.[42,49,52] However, a significant increase in α-helix and decrease in β-strands contents is noticeable compared with previous reports (**Table 1**). This difference might be responsible for the increased solubility of our Vif sample, in line with a previous study showing that formation of β-sheets might be responsible for Vif aggregates.[53]

We next analyzed the homogeneity of Vif using Dynamic Light Scattering (DLS) to ensure that our sample was monodisperse and free of aggregates, two necessary conditions required for reliable and good quality ITC analysis. Based on DLS measurements performed at various protein concentrations (4 to 200 μM), we found that the protein sample was highly monodisperse with a polydispersity index of 16.3% and a hydrodynamic radius of 7 nm (**Fig. 2B**). This is in good agreement with the recent report of Bernacchi et al.[42] but quite different from Gallerano et al. who obtained a hydrodynamic radius of 24 nm with a polydispersity index of 51%. Our DLS measurements performed at 10%, 5% and 2.5% glycerol concentrations did not show any significant variations down to 2.5% of glycerol, whereas massive aggregation was observed in absence of glycerol (not shown). To investigate the oligomerization state of our Vif construct, the average molecular weight of the sample was evaluated using Static Light Scattering (SLS). In a buffer containing 5% glycerol, we observed an average molecular weight of 270 kDa, which corresponds to multimers containing 10 to 11 Vif proteins and is in agreement with a recent similar study based on both size exclusion chromatography and DLS.[42] We concluded from these experiments that, even in absence of the co-factor CBF-β, the produced wild-type Vif protein can be obtained under a soluble, homogenous and stable form at concentrations up to 400 μM, and is therefore suitable for ITC measurements.

Vif binding to HIV-1 TAR RNA stem-loop. In a further step, isothermal titration calorimetry (ITC) was used to characterize Vif binding to the TAR stem-loop. ITC is a true in-solution technique and is considered the "gold standard" assay for binding since it directly provides, in one single experiment, the complete binding profile between two molecules, without any labeling requirement (reviewed in ref. 54–56). Typically, the ligand solution is repetitively injected in small aliquots from an automatic syringe into a thermally isolated stirred cell containing the protein. In an ITC experiment, one physically measures the heat generated or absorbed during a binding reaction. A direct measurement of the binding thermodynamics allows determination of the binding affinity (K_a), enthalpy and entropy changes (ΔH and ΔS) and stoichiometry (N) between two molecules.

Several Vif aliquots (1.5 to 2.0 μl at a ~200 μM concentration) were injected into 15 μM TAR RNA. This protocol was used since the reverse titration (injections of concentrated TAR aliquots into diluted Vif protein) systematically led to aggregation

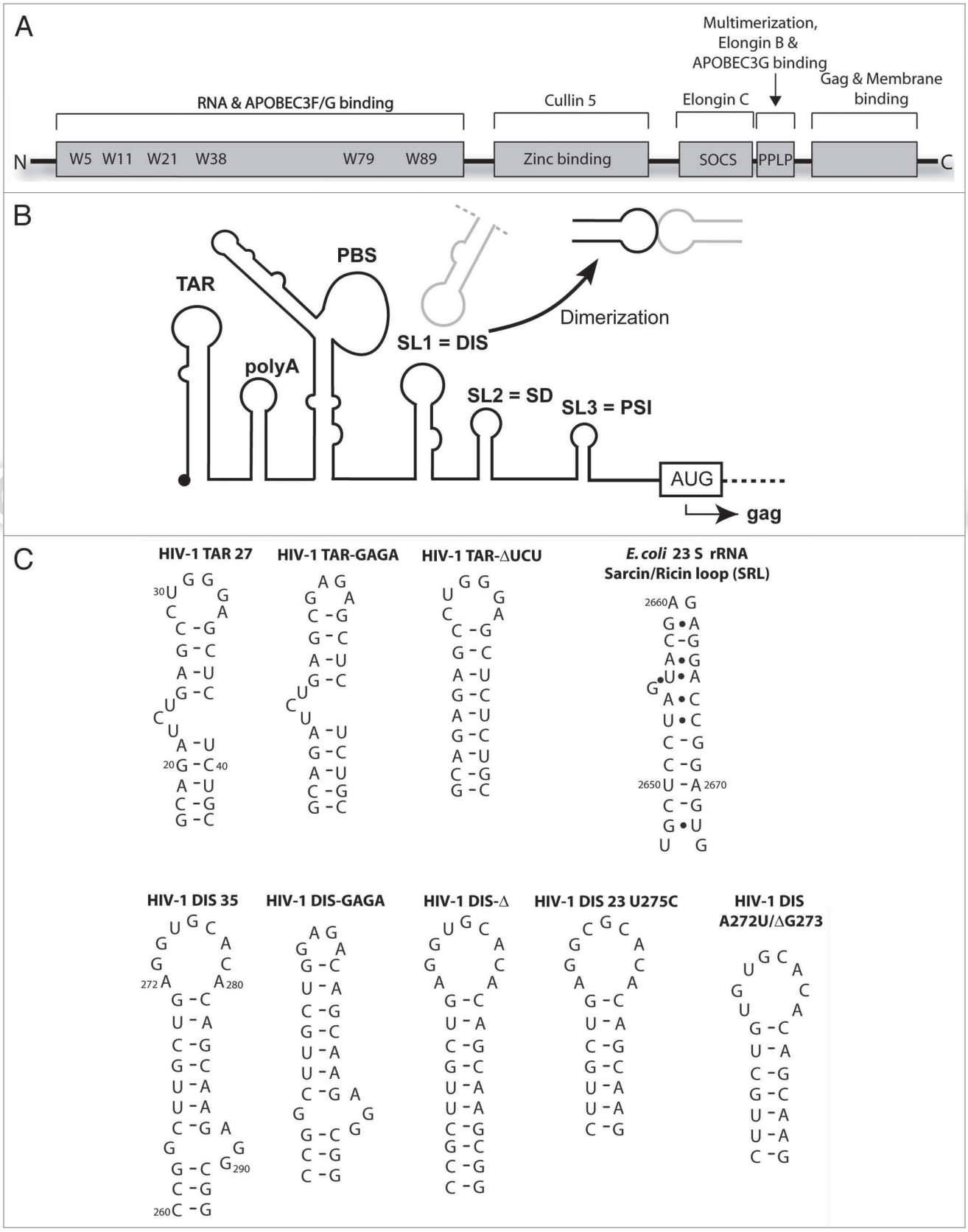


Figure 1. (A) Schematic representation of identified Vif functional regions. (B) Secondary structure of the 5′-untranslated region of the HIV-1 RNA genome. (C) RNA sequences used in this study.

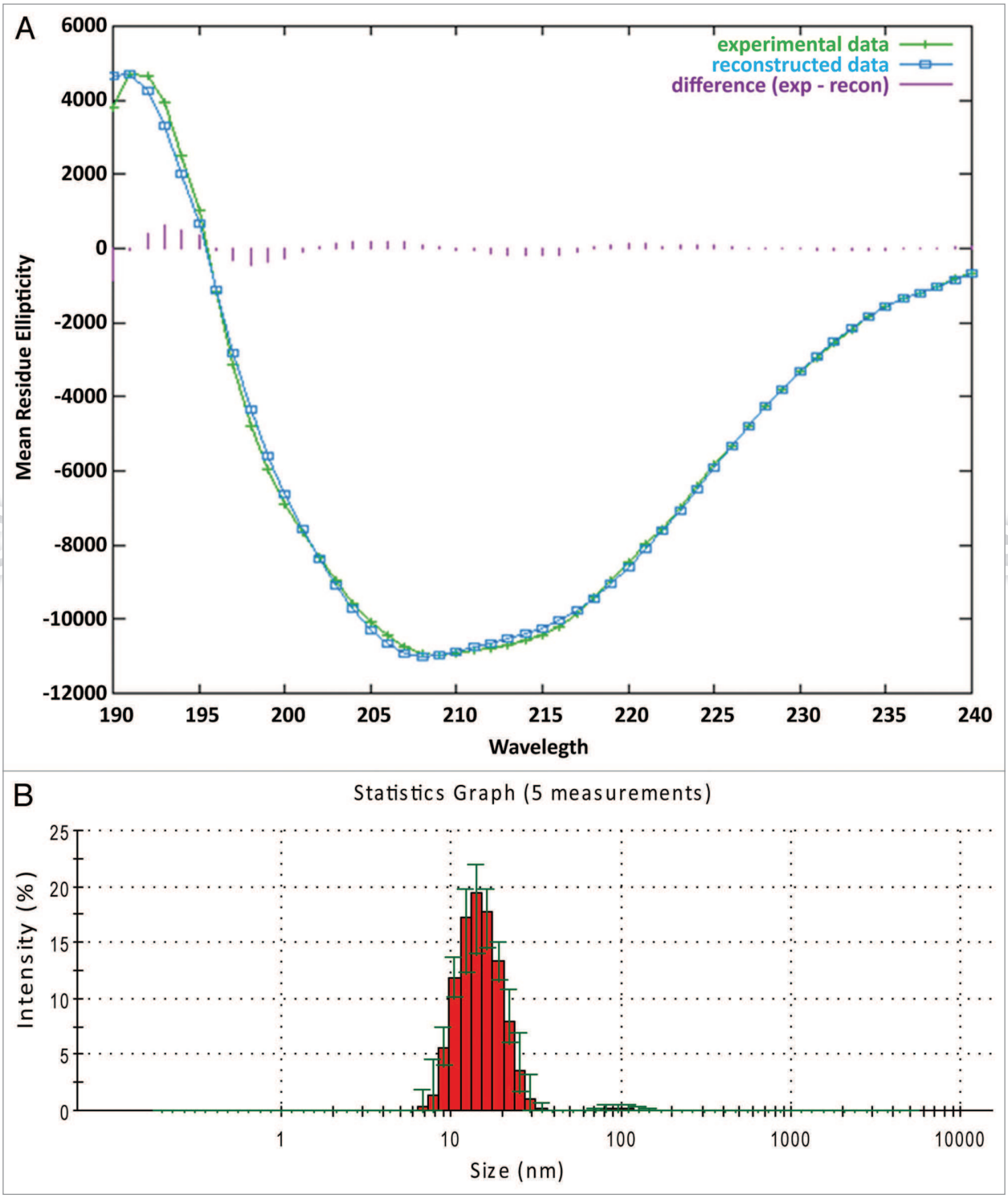


Figure 2. Biophysical characterization of purified Vif protein. (A) Far-UV spectra of the Vif protein and reconstructed data. (B) Dynamic light scattering profiles of the HIV-1 Vif showing a mean diameter of 14 nm.

of the protein/RNA complex in the ITC cell as judged by an abnormally high exothermic signal and baseline drift. In addition, most of our data were acquired at 30°C since baselines were more stable at this temperature. Some data were nevertheless obtained at 25°C, providing access to a temperature dependence of the binding enthalpy (see **Sup. data**). In our experimental conditions and in presence of 200 mM KCl, we observed a specific Vif binding to the TAR27 RNA (**Figs. 1C and 3**). The

observed dissociation constant (K_d ~ 160 nM) is associated with a rather large negative enthalpy variation (ΔH = -10.6 kcal/mol, corresponding to an exothermic process) and a slightly negative entropy variation (**Fig. 3 and Table 2**). The specificity of the protein/RNA binding is assessed by the 1:1 observed stoichiometry. This implies either that each Vif protein within a multimer is able to bind independently one TAR RNA, or that Vif multimers dissociate upon TAR binding. In a previous report it was found that, upon Vif binding to short RNA constructs, the mean hydrodynamic radius of resulting Vif/RNA complexes significantly increased compared with the free protein.[42] This observation therefore suggests that RNA binding does not promote Vif multimer dissociation. An increase of monovalent salt concentration to 300 mM, in order to reduce potential unspecific electrostatic Vif/RNA interactions, only led to a moderate decrease of the dissociation constant (1.8-fold). This suggests a significant contribution of hydrophobic interactions in the Vif/TAR interaction, likely through the tryptophan-rich RNA binding region (Trp5, Trp11, Trp21, Trp38, Trp79 and Trp89).

To further confirm the specificity of the Vif/TAR interaction, we used the *E. coli* 23 S RNA sarcin-ricin loop (SRL), a highly structured RNA stem-loop similar in size, but unrelated to TAR (**Fig. 1C**), as a control for unspecific RNA binding. Mixing of Vif into SRL RNA only led to a weak endothermic signal (ΔH = ~5.0 kcal/mol), that could not be fitted using a single site binding model (**Fig. 4** and left). Furthermore, increasing the salt concentration to 300 mM led to a complete loss of interaction (**Fig. 4** and right).

In order to define structure and sequence determinants for Vif binding to TAR RNA, we next used two TAR mutants (**Fig. 1C**): TAR-GAGA, where the wild-type five nucleotide apical loop was replaced by a GAGA tetraloop (a member of thermostable GNRA tetraloops), and TAR-ΔUCU, where the apical three nucleotide lateral bulge was deleted (**Fig. 1C**). ITC data obtained on the TAR-GAGA (in 200 mM KCl) mutant did not reveal any significant change in Vif binding. The observed stoichiometry (n = 0.7) was still close to 1.0 and both the dissociation constant (1.2-fold decrease) and the enthalpy variation (ΔH = -8.9 kcal/mol) were only moderately affected (**Fig. 3 and Table 2**). Notably, the small decrease in the enthalpic component is compensated by a slightly positive entropic term. Likewise, the protein binding to this mutant TAR sequence was only poorly affected in a buffer containing 300 mM KCl: the affinity decreased ~2-fold compared with binding to the TAR27 in the same conditions, together with a slight decrease of ΔH (**Fig. 3 and Table 2**).

On the contrary to the mutation of the apical loop, deletion of the UCU lateral bulge severely affected Vif binding as revealed by ITC data obtained with the TAR-ΔUCU sequence. In conditions with 200 mM KCl, the stoichiometry significantly increases to n = 2.0 and the K_d decreased to about 1 μM, which represents a 6.7-fold loss compared with wild-type sequence. In a buffer with 300 mM KCl, no Vif binding could be detected anymore (**Fig. 3 and Table 2**). We conclude from this set of experiments that Vif likely binds primarily the UCU lateral bulge of TAR through specific hydrophobic interactions. This interaction might be reinforced by subsequent (mostly electrostatic) interactions with the TAR apical loop.

Table 1. Comparison of Vif secondary structure composition derived from CD data

Reference	α-helix	β-strands	turns	disordered
33	6.5%	41.5%	12%	40%
49	11%	33%	28%	28%
52	6%	32%	62%	
This work	20%	26%	23%	30%

Our data and those from references 33 and 49 were obtained using the same CDSSTR algorithm and reference set database, thus ruling out any influence of the data processing.

Vif binding to HIV-1 DIS RNA kissing-loop. We next investigated the Vif binding to the DIS stem-loop. Because of the self-complementarity in the loop, the wild-type 35-mer DIS monomer forms DIS loop-loop dimers (also called kissing-loop) in native conditions (reviewed in ref. 57 and 58). Alternatively, an extended duplex dimeric form can also be obtained in vitro depending on RNA folding conditions. In this study, care was taken to use folding conditions leading essentially to DIS loop-loop complexes (see materiel and methods). To define RNA determinants for the protein binding, two mutants were designed (**Fig. 1C**): (1) the DIS-Δ mutant, where the internal purine-rich loop was deleted. It was shown that this internal loop is important for viral infectivity[59] and might be involved into binding of the NCp7 nucleocapsid protein;[60] (2) the DIS-GAGA mutant containing a GAGA tetraloop instead of the nine nucleotide apical loop. As a consequence for this mutation, the DIS-GAGA mutant was not able to form dimers in vitro in our conditions as evaluated by native polyacrylamide gels (not shown).

ITC data performed in presence of 200 mM KCl on the wild-type DIS 35-mer showed that Vif specifically and tightly (K_d = 63 nM) binds the DIS loop-loop complex with a 1:1 stoichiometry (one Vif binding per DIS stem-loop monomer, **Fig. 5 and Table 2**). As seen previously for TAR binding, the Vif/DIS interaction is exothermic and mostly driven by enthalpic contributions (ΔH = -9.1 kcal/mol, -TΔS = -0.9 kcal/mol). An increase of the monovalent salt concentration to 300 mM KCl led to a 2.8-fold loss in affinity (K_d = 177 nM) but still preserved the specificity of the interaction (n = 0.8). Deletion of the purine-rich internal loop of the DIS (sequence DIS-Δ, **Fig. 1C**) does not significantly affects Vif binding, neither in 200 mM KCl, nor in 300 mM KCl. The affinity of Vif for this DIS mutant only drops about 1.2-fold compared with the wild-type DIS and the stoichiometry remains close to 1:1 (**Fig. 5 and Table 2**). As seen previously for the TAR-GAGA mutant, a loss in the negative binding enthalpy is compensated by a positive entropic contribution for this DIS-Δ sequence. Experiments performed on the DIS-GAGA sequence revealed more significant differences in Vif binding. In conditions based on 200 mM KCl, a 2-fold loss in affinity was observed. In higher salt concentration, it was not possible to properly fit ITC corrected injection heats using a single site binding model used for all other ITC data (**Fig. 5**), suggesting a more complex protein binding.

©2012 Landes Bioscience. Do not distribute.

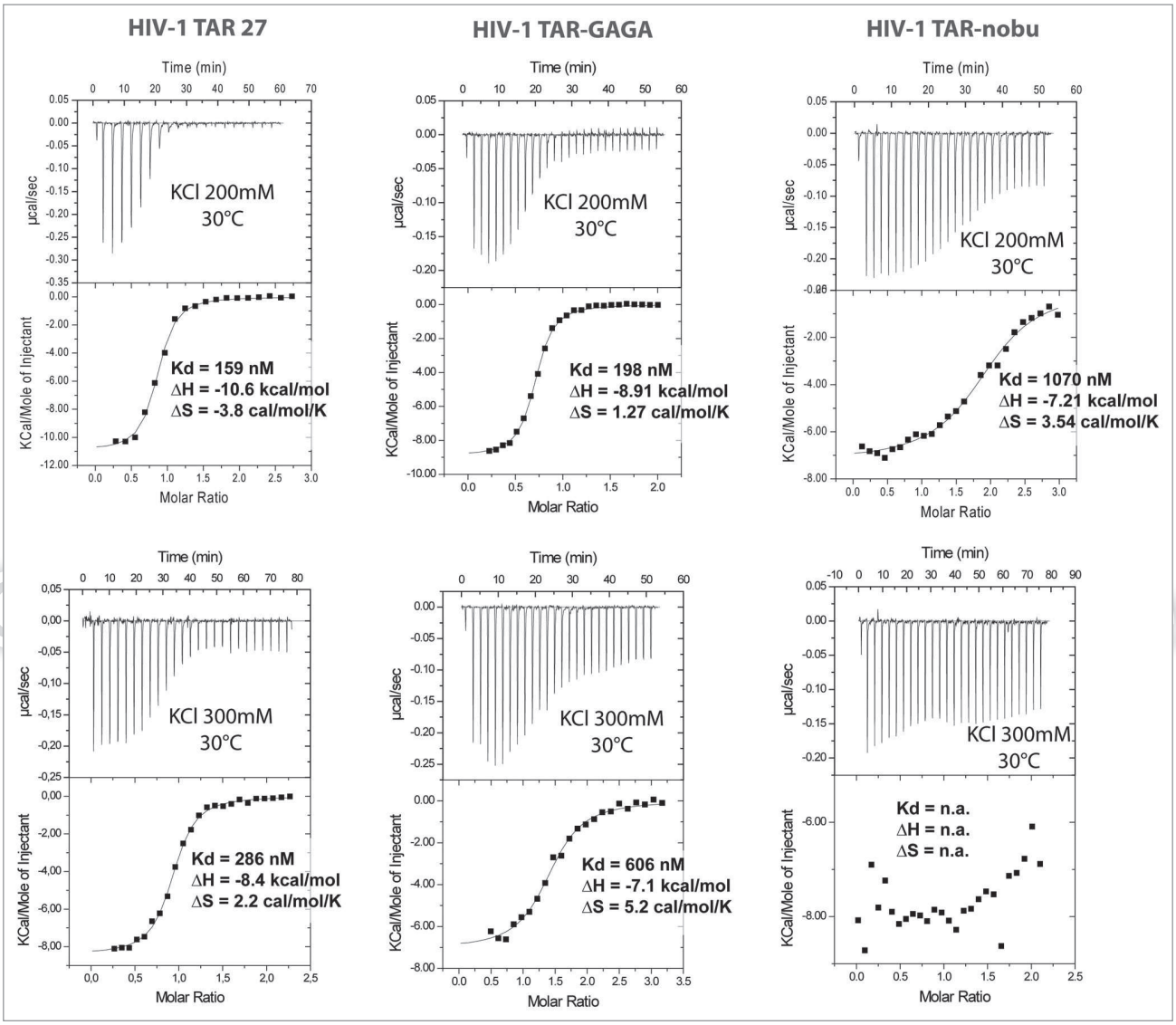


Figure 3. ITC profiles for the titration of HIV-1 Vif on the HIV-1 TAR27 stem-loop and mutant TAR sequences at 30°C and in presence of 200 or 300 mM of KCl. Each heat burst curve is the result of a 1.5 or 2.0 µl injection of 200 µM Vif in a 15 µM RNA solution. The corrected injection heats were derived by integration of the ITC profiles and were fitted with a one site binding model. "n.a." denotes "not applicable" since no binding was observed.

Table 2. Summary of binding parameters derived from ITC measurements for Vif binding to HIV-1 TAR and DIS wild-type and mutant sequences

200 mM KCl *300 mM KCl*	**DIS-WT**		**DIS-Δ**		**DIS-GAGA**		**TAR-WT**		**TAR-GAGA**		**TAR-ΔUCU**	
N	**1.1**	*0.8*	**1.0**	*0.7*	**1.1**	*n.d.*	**0.8**	*0.9*	**0.7**	*1.4*	**2.0**	*n.d.*
ΔH (kcal/mol)	**-9.1 ± 0.1**	*7.1 ± 0.1*	**-8.0 ± 0.1**	*-5.9 ± 0.1*	**-7.1 ± 0.1**	*n.d.*	**-10.6 ± 0.2**	*-8.4 ± 0.1*	**-8.9 ± 0.1**	*-7.1 ± 0.2*	**-7.2 ± 0.2**	*n.d.*
ΔS (cal/mol/deg)	**2.9**	*7.4*	**6.3**	*11.0*	**8.3**	*n.d.*	**-3.8**	*2.2*	**1.3**	*5.2*	**3.5**	*n.d.*
K_d (nM)	**63 ± 3**	*177 ± 21*	**75 ± 9**	*229 ± 16*	**121 ± 10**	*n.d.*	**159 ± 20**	*286 ± 22*	**198 ± 13**	*606 ± 81*	**1070 ± 122**	*n.d.*

N.d. denotes "not determinable" since accurate values could not be obtained.

In order to understand the role of the DIS loop dimerization in Vif binding, we tested the protein interaction on a DIS loop mutant with a U275C mutation that prevents self-dimerization (sequence DIS 23 U275C, **Fig. 1C**). In addition, this mutant is deprived of the lower part of the DIS stem, including the purine-rich internal loop that is dispensable for Vif binding as shown previously. In conditions with 200 mM KCl, the obtained ITC binding profile is quite similar to the one observed with the DIS-GAGA mutant: a slight increase in the stoichiometry (n = 1.2) together with a ~2-fold drop of the affinity constant (**Fig. S1** and top). In presence of 300 mM KCl, the affinity decreased to 1.3 μM, highlighting the role of electrostatic interactions in the Vif/DIS U275C interaction. We also noticed that ITC data collected with this DIS mutant were rather noisy (**Fig. S1** and top), which might suggest some aggregation of protein/RNA complexes. Finally, to evaluate a possible binding of Vif to the DIS stem, we used a DIS A272U/ΔG273 mutant (**Fig. 1C**). Because of the perfect complementarity of the eight-nucleotide loop, this sequence forms a 22 base-pairs duplex in vitro instead of a loop-loop complex (**Fig. S1**). In a buffer containing 200 mM KCl, Vif binds rather tightly to this sequence since we observed only a ~2-fold loss in affinity compared with wild-type DIS, however with a substantial increase in the stoichiometry (n = 1.6). As observed previously for the DIS U275C mutant, the Vif affinity for the DIS A272U/ΔG273 mutant was severely affected in presence of 300 mM KCl (K_d = 0.7 μM).

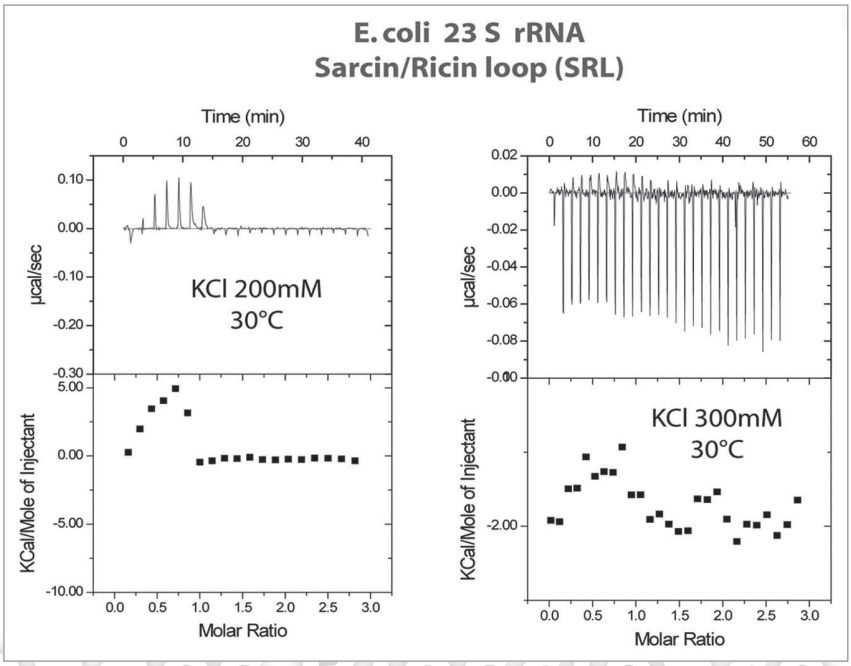


Figure 4. ITC profiles for the titration of HIV-1 Vif on the *E. coli* 23 S rRNA Sarcin/Ricin Loop at 30°C and in presence of 200 or 300 mM of KCl. Each heat burst curve is the result of a 1.5 or 2.0 μl injection of 200 μM Vif in a 15 μM RNA solution. Data could not be fitted using binding models available in the Origin ITC software and are likely relevant of unspecific Vif/RNA binding.

©2012 Landes Bioscience. Do not distribute.

It thus appears that Vif behaves quite similarly with DIS A272U/ΔG273 and DIS U275C, i.e., in DIS sequences where the 7-base-pair upper part of the stem is preserved. A detailed look in the binding thermodynamics of the Vif-DIS A272U/ΔG273 and Vif-DIS U275C interactions in 200 mM KCl (**Fig. S1**) reveals an important loss in the binding enthalpy (ΔH = -5.0 and -4.3 kcal/mol, respectively) compensated by an important gain in binding entropy (ΔS = 15.4 and 17.1 cal/mol/K, respectively) compared with binding to the wild-type DIS kissing-loop complex. This suggests that Vif binds the wild-type DIS and mutant sequences using slightly different binding mechanisms, likely mixed hydrophobic and electrostatic interactions on the wild-type DIS, but mostly electrostatic interactions in DIS loop mutants since binding is more sensitive to monovalent salt concentration.

We therefore conclude from this series of Vif/DIS experiments that Vif interacts both with the DIS loop-loop helix and the 7-base-pair stem adjacent to the loop, but not with the purine-rich internal loop. An increase in salt concentration only moderately affects Vif binding to the DIS wild-type and DIS-Δ sequences, but more severely the binding to DIS-GAGA, DIS UC275 and DIS A272U/ΔG273 mutants, i.e., all sequences affecting the loop-loop complex integrity but preserving the 7-base-pair stem (**Fig. 5 and Table 2**). As a consequence, Vif likely binds the DIS kissing-loop mostly through hydrophobic interactions and the 7-base-pair stem through electrostatic interactions. Because A272 and G273 are observed in an extrahelical conformation in crystal structures of the DIS kissing-loop complex,[61,62] we can hypothesize that they are involved into stacking interactions with some Vif tryptophan residues. Furthermore, because of structural similarities between the DIS kissing-loop complex and DIS extended duplex forms,[63,64] which is putatively present in mature virions and obtained after isomerization by the NCp7 protein,[65,66] we can anticipate that Vif should also specifically interact with this form of the DIS dimer.

Discussion

Here we show using Isothermal Titration Calorimetry that Vif specifically binds HIV-1 TAR stem-loop and DIS kissing-loop complex. Using a set of TAR and DIS mutants, we elucidated RNA determinants of Vif interaction to both sequences.

We found that Vif primarily recognizes the TAR UCU lateral bulge, which appears to be essential for the protein binding. The TAR apical loop might play a secondary role in the Vif binding since it contributes positively, but moderately, to the interaction. Interestingly, it is well-established that the TAR UCU lateral bulge is recognized by an arginine-rich RNA-binding motif of the HIV-1 Tat protein,[67,68] another small regulatory basic protein. This Tat/TAR interaction is essential for Tat-dependent transcriptional activation.[69] Several low molecular-weight compounds have been reported to block the Tat/TAR interaction in

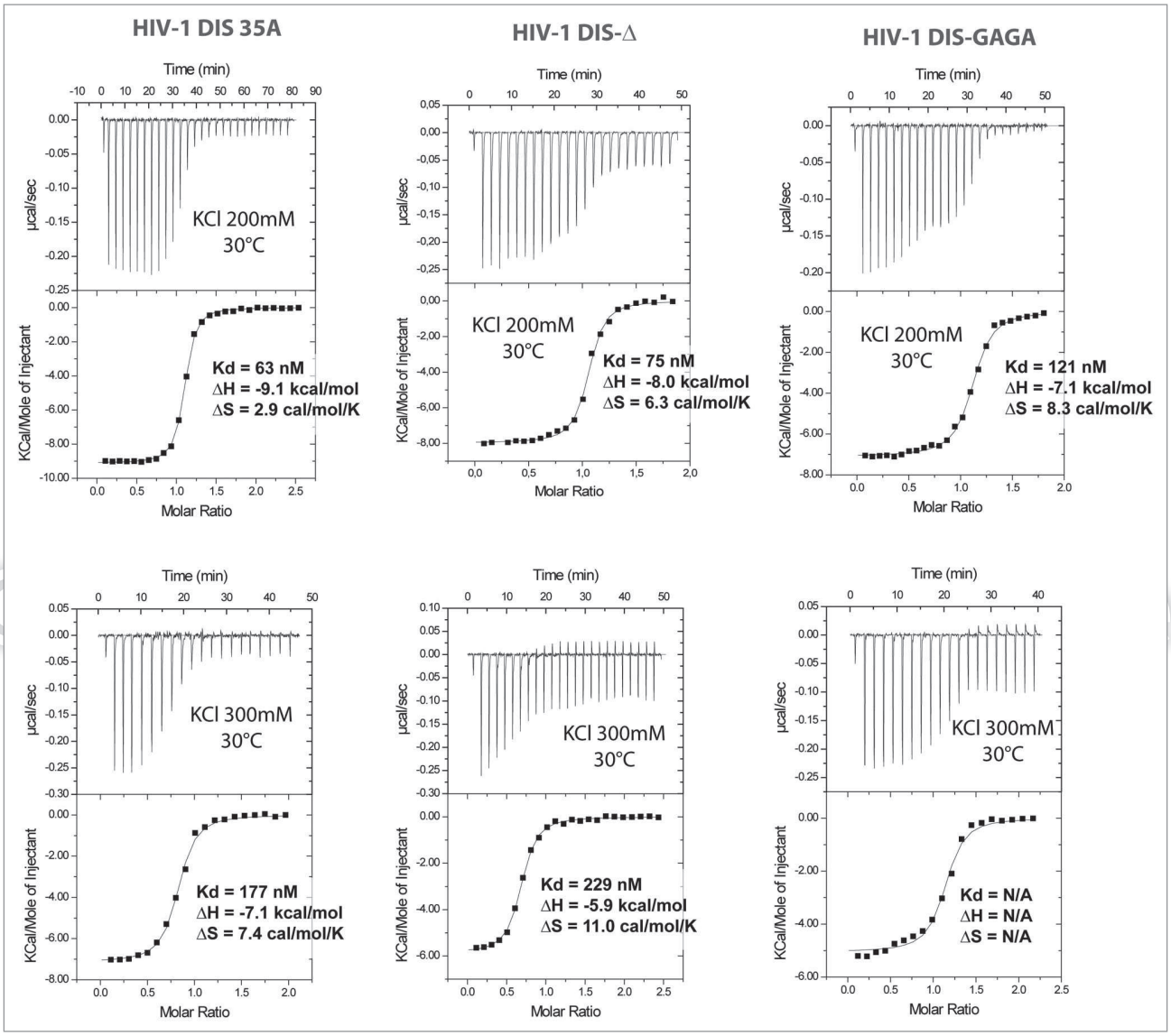


Figure 5. ITC profiles for the titration of HIV-1 Vif on the HIV-1 DIS loop-loop complex and mutant DIS sequences at 30°C and in presence of 200 or 300 mM of KCl. Each heat burst curve is the result of a 1.5 or 2.0 µl injection of 200 µM Vif in a 15 µM RNA solution. The corrected injection heats were derived by integration of the ITC profiles and were fitted with a one site binding model. N/A. denotes "not applicable" since accurate values could not be obtained using a single site binding model.

vitro and in infected cells with inhibition constants in the low nanomolar-range by competing with Tat for binding to the TAR UCU bulge.[70-73] We can therefore postulate that such compounds should also interfere with the binding of Vif to TAR (**Fig. 6A**). This hypothesis however remains to be tested.

Regarding the DIS RNA, our results show that Vif interacts both with the loop-loop helix and the 7-base-pair stem adjacent to the apical loop. Interestingly, it was shown that this stem is important for APOBEC3G binding to the viral RNA.[74] An attractive hypothesis is therefore that Vif could compete with APOBEC3G for binding the DIS, thus physically excluding the antiviral factor from packaging (**Fig. 6B**). Finally, we have previously shown that some 4,5-disubstituted aminoglycoside antibiotics specifically and tightly bind DIS dimers through interactions with the loop-loop helix and stem.[57,75-78] We therefore plan in a next future to investigate a putative interference of such aminoglycosides on the Vif/DIS interaction.

Materials and Methods

Expression and purification of recombinant Vif. A polymerase chain reaction (PCR) on the pcDNA-HVif vector encoding the full-length HIV-1 NL4-3 Vif protein[79] was used to amplify the Vif gene and to insert *Nde*I and BamHI restriction sites at the extremities. The digested PCR product was then inserted between *Nde*I and BamHI sites into the bacterial expression

plasmid pnEAtH[80] encoding a 6-His tag followed by a thrombin cleavage site, leading to a Vif construct containing a histidine tag and a protease cleavage site fused with the Vif N terminus (MGS SHH HHH HSS GTG SGL VPR GSH-Vif). The sequence of the Vif protein used is shown on **Supplemental Figures**. These plasmids pnEAtH-Vif were used to chemically transform *E. coli* Rosetta 2 strains. Production of Vif proteins was induced in LB liquid medium by addition of 0.5 mM IPTG to log phase bacterial culture (OD 600 nm = 0.5–0.6). After 4 h at 37°C bacteria were harvested by centrifugation at 4,000 g during 15 min and and lysed by an overnight stirring at room temperature in the denaturing buffer (6 M guanidine Hydrochloride, 100 mM sodium phosphate, 50 mM Tris, pH 8.0). Cellular debris were separated by centrifugation at 100,000 g during 1 h at 4°C and the cleared lysate was loaded onto a Ni-NTA agarose column. The column was washed with the lysis buffer and elution was performed by applying 25 ml of the same solution at decreasing pH values (pH 6.5, pH 6.0, pH 5.8, pH 5.5 and pH 5.0). The Vif containing fractions were identified on a 12% SDS-PAGE, pooled and the volume was completed to 50 ml (~0.3 mg ml^{-1} of Vif) using elution buffer. The protein was renaturated by successive dialysis for 1 h at 4°C against 500 ml of a buffer containing 50 mM MOPS pH 6.5, 150 mM NaCl, 10% glycerol, 1 mM dithiothreitol and decreasing guanidium chloride concentrations (3.0 M, 1.5 M, 0.75 M, 0.42 M, 0.21 M and finally 0 M). The protein was then concentrated to ~5 mg ml^{-1} on an Amicon Ultra-10 and stored at -80°C. Stock solutions were centrifuged at 100,000 g for 1 h immediately prior to use. The molecular mass of the purified protein was checked by mass spectrometry (Calculated mass: 25,035 Da, Observed mass: 25,036 Da, see **Fig. S3A**) and the identity of the protein was assessed by peptide mass fingerprinting (PMF, see **Fig. S3B**). The purified protein was > 95% pure based on SDS-PAGE analysis (**Fig. S4**).

Prior measurements, Vif was dialysed in either the ITC buffer (50 mM MOPS pH 6.5, 200–300 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 1 mM dithiothreitol) or the CD buffer [50 mM MOPS pH 6.5, 150 mM NaCl, 2.5% glycerol, 1 mM tris(2-carboxyethyl)phosphine]. The protein was then concentrated to 100 μM for CD and 150–280 μM for ITC measurements. All buffers were argon-degased.

Mass spectrometry. Protein samples were dialyzed against 200 mM ammonium acetate and subsequently diluted in 1:10 water acetonitrile (v/v) mixture acidified with 1% formic acid to achieve a concentration of 0.3 M. Mass spectrometry was performed on an ESI-TOF mass spectrometer Q-TOF micro (Waters, MA USA). Calibration was achieved in the positive ion mode, using denatured horse heart myoglobin (Sigma), using denatured horse heart myoglobin (Sigma). For PMF analysis, Vif (10 μl, 5 mg ml^{-1}) was digested with trypsin (1 μl, 10 ng μl^{-1}) in native conditions in 50 mM ammonium carbonate for 30 min at 37°C. Data were collected on an Autoflex III MALDI-TOF mass spectrometer (Bruker).

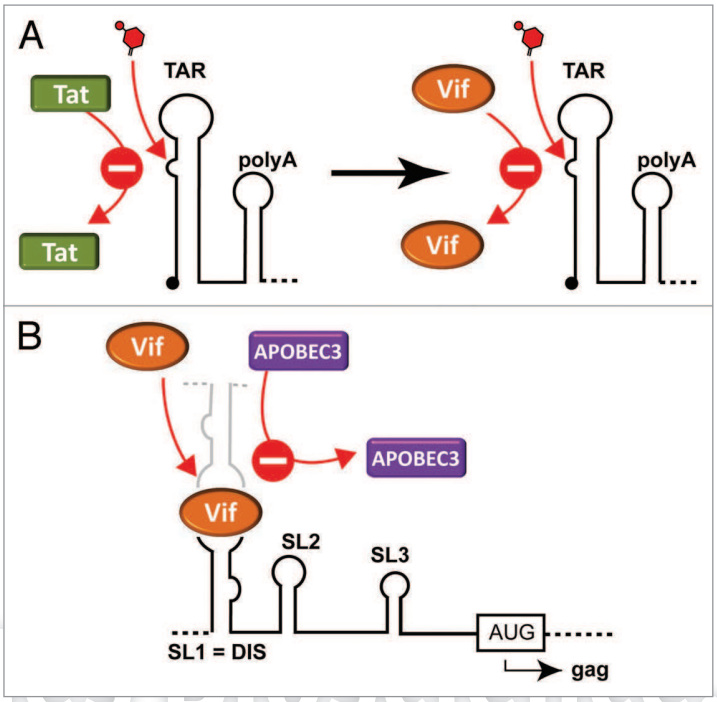


Figure 6. Consequences of our results for Vif-RNA interactions. (A) The HIV-1 Tat protein is known to bind the TAR lateral bulge. Small molecules are known to prevent this binding by inhibiting the Tat binding. Since Vif also recognizes the TAR lateral bulge, we anticipate that such Tat/TAR inhibitors could also prevent Vif binding to TAR. (B) Scheme showing that Vif could reduce APOBEC3 binding to the viral genome by competing for the DIS sequence.

©2012 Landes Bioscience. Do not distribute.

Light scattering experiments. Dynamic Light Scattering (DLS) was performed using a Zetasizer Nano S (Malvern). Static Light Scattering (SLS) was performed on a Dynapro Nanostar (Wyatt Technology Europe). For DLS experiments, protein sample was adjusted to 200 μM in the buffer 50 mM MOPS, pH 6.5, 150 mM NaCl, 0–10% glycerol, 1 mM dithiothreitol. For SLS experiments, a 0.1 mg ml^{-1} protein concentration was used in the same buffer containging 5% glycerol. Sample temperature was maintained at 30°C. Fluctuations of the scattered intensity due to Brownian motions were recorded at microsecond time intervals. An autocorrelation function was derived, thus leading to the determination of diffusion coefficients. Assimilating proteins in solution to spheres, diffusion coefficients were related to the hydrodynamic radius of the particles, R_h via the Stokes-Einstein equation:

$$D = k_B T / 6\pi\eta R_h$$

where k_B is the Boltzmann constant, T the temperature, η the solvent viscosity and D the translational diffusion coefficient. All experimental data were corrected for solvent viscosity using an AMVn automated micro viscosimeter (Anton Paar) and for refractive index using an Abbé refractometer. Data were analyzed using DLS analysis software (Malvern).

Circular dichroism. Vif protein sample was prepared in 50 mM MOPS, 150 mM NaCl, 2.5% glycerol, 1 mM

tris(2-carboxyethyl)phosphine, at a concentration of 100 μM and diluted again in the same buffer prior to the experiment at a final concentration of 25 μM. Far-UV CD measurements were performed with a J-850 Jasco spectropolarimeter equipped with a Peltier temperature controller and calibrated with ammonium d-10 camphorsulfonate. Spectra (186–260 nm) were measured at 30°C, with a constant bandwidth of 1 nm and a 3–5 sec integration time using a quartz cell of path length 0.1 mm. Spectra were averaged over ten scans and corrected for buffer contributions. Quantitative estimations of the secondary structure contents were performed using Dichroweb[81] and the reference CDSSTR database 4 containing the CD spectra of 43 soluble proteins ranging from 190 to 240 nm..

RNA sample preparation. Chemically-synthetized wild-type HIV-1 subtype B 27-mer TAR RNA, HIV-1 subtype A 35-mer DIS RNA, 27-mer Sarcin-Ricin Loop (SRL) from the *E. coli* 23 S rRNA, as well as TAR and DIS mutant sequences were purchased from Integrated DNA technologies or from Dharmacon. RNA were purified as described in reference 82. In short, RNA was loaded on a Nucleopac PA-100 column (Dionex) heated at 70°C equilibrated in 4 M urea, 20 mM Mes [2-N(-Morpholine) ethane sulfonic acid], pH 6.2 and eluted using a $NaClO_4$ gradient.

DIS kissing-loop complexes were obtained as follows: RNA was diluted to 2 μM concentration in water, heated at 90°C for 5 min and then cooled on ice for 5 min. As shown previously, this protocol essentially leads to the formation of DIS loop-loop dimers and prevents the formation of extended duplex dimers.[83,84] A similar protocol was performed for folding of TAR and SRL hairpins. DIS-A272U/ΔG273 was diluted to 60 μM in water, heated at 90°C for 5 min and then cooled at room temperature. This protocol was used in order to favor the formation of a duplex instead of hairpin monomers. Folding buffer (50 mM MOPS pH 6.5, 200 mM KCl, 2 mM $MgCl_2$) was then added and RNA samples were dialysed against ITC buffer (50 mM MOPS pH 6.5, 200–300 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 1 mM dithiothreitol) and adjusted to a final concentration of 15 μM.

Isothermal titration calorimetry (ITC). ITC measurements were performed on a MicroCal ITC_{200} (GE Healthcare) in ITC buffer containing 200 or 300 mM KCl. In a typical experiment, 27 injections of 1.5 μl aliquots of Vif protein at 150–280 μM were injected (at 0.5 μl s^{-1}) into 203 μl of RNA at 15 μM in the sample cell. The delay between injections was 120 sec. All ITC curves were analyzed using the software Origin (OriginLab). All data could be fitted with a single-site model. Standard free energies of binding (ΔG) were obtained using the equation

$$\Delta G = -RT \ln Ka$$

Entropic contributions were obtained from the relationship

$$\Delta G = \Delta H - T \Delta S$$

In our experimental conditions, the product Ka x [RNA] x N, where N is the number of binding sites, was in the 0.1–1,000 range, allowing an accurate and simultaneous determination of binding parameters by ITC.[56] Because of a too fast Vif-RNA binding, it was not possible to use our recently-developed *kinITC* approach to derive kinetic parameters of binding in addition to thermodynamic data.[85]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

©2012 Landes Bioscience. Do not distribute.

Acknowledgements

We thank Dr. Bernard Lorber for his assistance in D.L.S. and S.L.S. experiments, Dr. Guillaume Bec and Guillaume Hoffmann for their technical assistance, Dr. Carine Tisné and Dr. Serena Bernacchi for fruitful discussions and comments. The following reagent was obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: pcDNA-HVif from Dr. Stephan Bour and Dr. Klaus Strebel.

Funding

This work was supported by the "Agence Nationale de Recherche sur le SIDA" (ANRS to E.E.). S.F. is an ANRS fellow. J.M. is a SIDACTION/Fondation Pierre Bergé fellow.

Supplemental Materials

Supplemental material can be found at:
www.landesbioscience.com/journals/rnabiology/article/20483

References

1. Fisher AG, Ensoli B, Ivanoff L, Chamberlain M, Petteway S, Ratner L, et al. The sor gene of HIV-1 is required for efficient virus transmission in vitro. Science 1987; 237:888-93; PMID:3497453; http://dx.doi.org/10.1126/science.3497453.
2. Gabuzda DH, Lawrence K, Langhoff E, Terwilliger E, Dorfman T, Haseltine WA, et al. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J Virol 1992; 66:6489-95; PMID:1357189.
3. Gallo R, Wong-Staal F, Montagnier L, Haseltine WA, Yoshida M. HIV/HTLV gene nomenclature. Nature 1988; 333:504; PMID:2836736; http://dx.doi.org/10.1038/333504a0.
4. Sodroski J, Goh WC, Rosen C, Tartar A, Portetelle D, Burny A, et al. Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions. Science 1986; 231:1549-53; PMID:3006244; http://dx.doi.org/10.1126/science.3006244.
5. Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA. The HIV 'A' (sor) gene product is essential for virus infectivity. Nature 1987; 328:728-30; PMID:2441266; http://dx.doi.org/10.1038/328728a0.
6. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. Cell 2003; 113:803-9; PMID:12809610; http://dx.doi.org/10.1016/S0092-8674(03)00423-9.
7. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 2003; 424:94-8; PMID:12808465; http://dx.doi.org/10.1038/nature01707.
8. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 2003; 424:99-103; PMID:12808466; http://dx.doi.org/10.1038/nature01709.
9. Bishop KN, Holmes RK, Malim MH. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol 2006; 80:8450-8; PMID:16912295; http://dx.doi.org/10.1128/JVI.00839-06.
10. Bishop KN, Verma M, Kim EY, Wolinsky SM, Malim MH. APOBEC3G inhibits elongation of HIV-1 reverse transcripts. PLoS Pathog 2008; 4:1000231; PMID:19057663; http://dx.doi.org/10.1371/journal.ppat.1000231.
11. Holmes RK, Malim MH, Bishop KN. APOBEC-mediated viral restriction: not simply editing? Trends Biochem Sci 2007; 32:118-28; PMID:17303427; http://dx.doi.org/10.1016/j.tibs.2007.01.004.
12. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol 2004; 14:1392-6; PMID:15296758; http://dx.doi.org/10.1016/j.cub.2004.06.057.

13. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002; 418:646-50; PMID:12167863; http://dx.doi.org/10.1038/nature00939.
14. Wiegand HL, Doehle BP, Bogerd HP, Cullen BR. A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO J 2004; 23:2451-8; PMID:15152192; http://dx.doi.org/10.1038/sj.emboj.7600246.
15. Zheng YH, Irwin D, Kurosu T, Tokunaga K, Sata T, Peterlin BM. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. J Virol 2004; 78:6073-6; PMID:15141007; http://dx.doi.org/10.1128/JVI.78.11.6073-6.2004.
16. Doehle BP, Schäfer A, Cullen BR. Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. Virology 2005; 339:281-8; PMID:15993456; http://dx.doi.org/10.1016/j.virol.2005.06.005.
17. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. J Virol 2006; 80:10522-33; PMID:16920826; http://dx.doi.org/10.1128/JVI.01123-06.
18. Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng YH. Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem 2008; 283:11606-14; PMID:18299330; http://dx.doi.org/10.1074/jbc.M707586200.
19. Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 2006; 281:16837-41; PMID:16648136; http://dx.doi.org/10.1074/jbc.M602367200.
20. OhAinle M, Kerns JA, Malik HS, Emerman M. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol 2006; 80:3853-62; PMID:16571802; http://dx.doi.org/10.1128/JVI.80.8.3853-62.2006.
21. Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol 2008; 26:317-53; PMID:18304004; http://dx.doi.org/10.1146/annurev.immunol.26.021607.090350.
22. Henriet S, Mercenne G, Bernacchi S, Paillart JC, Marquet R. Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors. Microbiol Mol Biol Rev 2009; 73:211-32; PMID:19487726; http://dx.doi.org/10.1128/MMBR.00040-08.
23. Wissing S, Galloway NL, Greene WC. HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors. Mol Aspects Med 2010; 31:383-97; PMID:20538015; http://dx.doi.org/10.1016/j.mam.2010.06.001.
24. Conticello SG, Harris RS, Neuberger MS. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. Curr Biol 2003; 13:2009-13; PMID:14614829; http://dx.doi.org/10.1016/j.cub.2003.10.034.
25. Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 2003; 9:1398-403; PMID:14528301; http://dx.doi.org/10.1038/nm946.
26. Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D. Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem 2004; 279:7792-8; PMID:14672928; http://dx.doi.org/10.1074/jbc.M313093200.
27. Sheehy AM, Gaddis NC, Malim MH. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med 2003; 9:1404-7; PMID:14528300; http://dx.doi.org/10.1038/nm945.
28. Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell 2003; 12:591-601; PMID:14527406; http://dx.doi.org/10.1016/S1097-2765(03)00353-8.
29. Hultquist JF, Binka M, LaRue RS, Simon V, Harris RS. Vif proteins of human and simian immunodeficiency viruses require cellular CBFβ to degrade APOBEC3 restriction factors. J Virol 2012; 86:2874-7; PMID:22205746; http://dx.doi.org/10.1128/JVI.06950-11.
30. Jäger S, Kim DY, Hultquist JF, Shindo K, LaRue RS, Kwon E, et al. Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. Nature 2012; 481:371-5; PMID:22190037.
31. Zhang W, Du J, Evans SL, Yu Y, Yu XF. T-cell differentiation factor CBF-β regulates HIV-1 Vif-mediated evasion of host restriction. Nature 2012; 481:376-9; PMID:22190036.
32. Kao S, Khan MA, Miyagi E, Plishka R, Buckler-White A, Strebel K. The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. J Virol 2003; 77:11398-407; PMID:14557625; http://dx.doi.org/10.1128/JVI.77.21.11398-407.2003.
33. Mercenne G, Bernacchi S, Richer D, Bec G, Henriet S, Paillart JC, et al. HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. Nucleic Acids Res 2010; 38:633-46; PMID:19910370; http://dx.doi.org/10.1093/nar/gkp1009.
34. Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, et al. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 2003; 114:21-31; PMID:12859895; http://dx.doi.org/10.1016/S0092-8674(03)00515-4.
35. Opi S, Kao S, Goila-Gaur R, Khan MA, Miyagi E, Takeuchi H, et al. Human immunodeficiency virus type 1 Vif inhibits packaging and antiviral activity of a degradation-resistant APOBEC3G variant. J Virol 2007; 81:8236-46; PMID:17522211; http://dx.doi.org/10.1128/JVI.02694-06.
36. Dettenhofer M, Cen S, Carlson BA, Kleiman L, Yu XF. Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. J Virol 2000; 74:8938-45; PMID:10982337; http://dx.doi.org/10.1128/JVI.74.19.8938-45.2000.
37. Henriet S, Richer D, Bernacchi S, Decroly E, Vigne R, Ehresmann B, et al. Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. J Mol Biol 2005; 354:55-72; PMID:16236319; http://dx.doi.org/10.1016/j.jmb.2005.09.025.
38. Khan MA, Akari H, Kao S, Aberham C, Davis D, Buckler-White A, et al. Intravirion processing of the human immunodeficiency virus type 1 Vif protein by the viral protease may be correlated with Vif function. J Virol 2002; 76:9112-23; PMID:12186895; http://dx.doi.org/10.1128/JVI.76.18.9112-23.2002.
39. Zhang H, Pomerantz RJ, Dornadula G, Sun Y. Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. J Virol 2000; 74:8252-61; PMID:10954522; http://dx.doi.org/10.1128/JVI.74.18.8252-61.2000.
40. Barraud P, Paillart JC, Marquet R, Tisné C. Advances in the structural understanding of Vif proteins. Curr HIV Res 2008; 6:91-9; PMID:18336256; http://dx.doi.org/10.2174/157016208783885056.
41. Bernacchi S, Henriet S, Dumas P, Paillart JC, Marquet R. RNA and DNA binding properties of HIV-1 Vif protein: a fluorescence study. J Biol Chem 2007; 282:26361-8; PMID:17609216; http://dx.doi.org/10.1074/jbc.M703122200.
42. Bernacchi S, Mercenne G, Tournaire C, Marquet R, Paillart JC. Importance of the proline-rich multimerization domain on the oligomerization and nucleic acid binding properties of HIV-1 Vif. Nucleic Acids Res 2011; 39:2404-15; PMID:21076154; http://dx.doi.org/10.1093/nar/gkq979.
43. Auclair JR, Green KM, Shandilya S, Evans JE, Somasundaran M, Schiffer CA. Mass spectrometry analysis of HIV-1 Vif reveals an increase in ordered structure upon oligomerization in regions necessary for viral infectivity. Proteins 2007; 69:270-84; PMID:17598142; http://dx.doi.org/10.1002/prot.21471.
44. Yang S, Sun Y, Zhang H. The multimerization of human immunodeficiency virus type I Vif protein: a requirement for Vif function in the viral life cycle. J Biol Chem 2001; 276:4889-93; PMID:11071884; http://dx.doi.org/10.1074/jbc.M004895200.
45. Donahue JP, Vetter ML, Mukhtar NA, D'Aquila RT. The HIV-1 Vif PPLP motif is necessary for human APOBEC3G binding and degradation. Virology 2008; 377:49-53; PMID:18499212; http://dx.doi.org/10.1016/j.virol.2008.04.017.
46. Miller JH, Presnyak V, Smith HC. The dimerization domain of HIV-1 viral infectivity factor Vif is required to block virion incorporation of APOBEC3G. Retrovirology 2007; 4:81; PMID:18036235; http://dx.doi.org/10.1186/1742-4690-4-81.
47. Yang B, Gao L, Li L, Lu Z, Fan X, Patel CA, et al. Potent suppression of viral infectivity by the peptides that inhibit multimerization of human immunodeficiency virus type 1 (HIV-1) Vif proteins. J Biol Chem 2003; 278:6596-602; PMID:12480936; http://dx.doi.org/10.1074/jbc.M210164200.
48. Bergeron JR, Huthoff H, Veselkov DA, Beavil RL, Simpson PJ, Matthews SJ, et al. The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS Pathog 2010; 6:1000925; PMID:20532212; http://dx.doi.org/10.1371/journal.ppat.1000925.
49. Gallerano D, Devanaboyina SC, Swoboda I, Linhart B, Mittermann I, Keller W, et al. Biophysical characterization of recombinant HIV-1 subtype C virus infectivity factor. Amino Acids 2011; 40:981-9; PMID:20809132; http://dx.doi.org/10.1007/s00726-010-0725-x.
50. Yang X, Goncalves J, Gabuzda D. Phosphorylation of Vif and its role in HIV-1 replication. J Biol Chem 1996; 271:10121-9; PMID:8626571; http://dx.doi.org/10.1074/jbc.271.17.10121.
51. Marcsisin SR, Engen JR. Molecular insight into the conformational dynamics of the Elongin BC complex and its interaction with HIV-1 Vif. J Mol Biol 2010; 402:892-904; PMID:20728451; http://dx.doi.org/10.1016/j.jmb.2010.08.026.
52. Marcsisin SR, Narute PS, Emert-Sedlak LA, Kloczewiak M, Smithgall TE, Engen JR. On the solution conformation and dynamics of the HIV-1 viral infectivity factor. J Mol Biol 2011; 410:1008-22; PMID:21763503; http://dx.doi.org/10.1016/j.jmb.2011.04.053.
53. Paul I, Cui J, Maynard EL. Zinc binding to the HCCH motif of HIV-1 virion infectivity factor induces a conformational change that mediates protein-protein interactions. Proc Natl Acad Sci USA 2006; 103:18475-80; PMID:17132731; http://dx.doi.org/10.1073/pnas.0604150103.
54. Feig AL. Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry. Methods Enzymol 2009; 468:409-22; PMID:20946780; http://dx.doi.org/10.1016/S0076-6879(09)68019-8.
55. Salim NN, Feig AL. Isothermal titration calorimetry of RNA. Methods 2009; 47:198-205; PMID:18835447; http://dx.doi.org/10.1016/j.ymeth.2008.09.003.
56. Velázquez Campoy A, Freire E. ITC in the post-genomic era...? Priceless. Biophys Chem 2005; 115:115-24; PMID:15752592; http://dx.doi.org/10.1016/j.bpc.2004.12.015.

©2012 Landes Bioscience. Do not distribute.

57. Ennifar E, Paillart JC, Bernacchi S, Walter P, Pale P, Decout JL, et al. A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. Biochimie 2007; 89:1195-203; PMID:17434658; http://dx.doi.org/10.1016/j.biochi.2007.03.003.
58. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. Nat Rev Microbiol 2004; 2:461-72; PMID:15152202; http://dx.doi.org/10.1038/nrmicro903.
59. Shen N, Jetté L, Wainberg MA, Laughrea M. Role of stem B, loop B and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization. J Virol 2001; 75:10543-9; PMID:11581429; http://dx.doi.org/10.1128/JVI.75.21.10543-9.2001.
60. Hagan NA, Fabris D. Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. J Mol Biol 2007; 365:396-410; PMID:17070549; http://dx.doi.org/10.1016/j.jmb.2006.09.081.
61. Ennifar E, Dumas P. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. J Mol Biol 2006; 356:771-82; PMID:16403527; http://dx.doi.org/10.1016/j.jmb.2005.12.022.
62. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. Nat Struct Biol 2001; 8:1064-8; PMID:11702070; http://dx.doi.org/10.1038/nsb727.
63. Ennifar E, Walter P, Dumas P. Cation-dependent cleavage of the duplex form of the subtype-B HIV-1 RNA dimerization initiation site. Nucleic Acids Res 2010; 38:5807-16; PMID:20460458; http://dx.doi.org/10.1093/nar/gkq344.
64. Ennifar E, Yusupov M, Walter P, Marquet R, Ehresmann B, Ehresmann C, et al. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. Structure 1999; 7:1439-49; PMID:10574792; http://dx.doi.org/10.1016/S0969-2126(00)80033-7.
65. Muriaux D, De Rocquigny H, Roques BP, Paoletti J. NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. J Biol Chem 1996; 271:33686-92; PMID:8969239; http://dx.doi.org/10.1074/jbc.271.52.33686.
66. Takahashi KI, Baba S, Chattopadhyay P, Koyanagi Y, Yamamoto N, Takaku H, et al. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. RNA 2000; 6:96-102; PMID:10668802; http://dx.doi.org/10.1017/S1355838200991635.
67. Roy S, Delling U, Chen CH, Rosen CA, Sonenberg N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev 1990; 4:1365-73; PMID:2227414; http://dx.doi.org/10.1101/gad.4.8.1365.
68. Weeks KM, Ampe C, Schultz SC, Steitz TA, Crothers DM. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. Science 1990; 249:1281-5; PMID:2205002; http://dx.doi.org/10.1126/science.2205002.
69. Karn J. Tackling Tat. J Mol Biol 1999; 293:235-54; PMID:10550206; http://dx.doi.org/10.1006/jmbi.1999.3060.
70. Davis B, Afshar M, Varani G, Murchie AI, Karn J, Lentzen G, et al. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". J Mol Biol 2004; 336:343-56; PMID:14757049; http://dx.doi.org/10.1016/j.jmb.2003.12.046.
71. Hamy F, Felder ER, Heizmann G, Lazdins J, Aboulela F, Varani G, et al. An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. Proc Natl Acad Sci USA 1997; 94:3548-53; PMID:9108013; http://dx.doi.org/10.1073/pnas.94.8.3548.
72. Lalonde MS, Lobritz MA, Ratcliff A, Chamanian M, Athanassiou Z, Tyagi M, et al. Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-transactivating response element (TAR) RNA. PLoS Pathog 2011; 7:1002038; PMID:21625572; http://dx.doi.org/10.1371/journal.ppat.1002038.
73. Mei HY, Mack DP, Galan AA, Halim NS, Heldsinger A, Loo JA, et al. Discovery of selective, small-molecule inhibitors of RNA complexes-I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. Bioorg Med Chem 1997; 5:1173-84; PMID:9222511; http://dx.doi.org/10.1016/S0968-0896(97)00064-3.
74. Khan MA, Kao S, Miyagi E, Takeuchi H, Goila-Gaur R, Opi S, et al. Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. J Virol 2005; 79:5870-4; PMID:15827203; http://dx.doi.org/10.1128/JVI.79.9.5870-4.2005.
75. Bernacchi S, Freisz S, Maechling C, Spiess B, Marquet R, Dumas P, et al. Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion. Nucleic Acids Res 2007; 35:7128-39; PMID:17942426; http://dx.doi.org/10.1093/nar/gkm856.
76. Ennifar E, Paillart JC, Bodlenner A, Walter P, Weibel JM, Aubertin AM, et al. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. Nucleic Acids Res 2006; 34:2328-39; PMID:16679451; http://dx.doi.org/10.1093/nar/gkl317.
77. Ennifar E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C, Dumas P, et al. HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. J Biol Chem 2003; 278:2723-30; PMID:12435744; http://dx.doi.org/10.1074/jbc.M205726200.
78. Freisz S, Lang K, Micura R, Dumas P, Ennifar E. Binding of aminoglycoside antibiotics to the duplex form of the HIV-1 genomic RNA dimerization initiation site. Angew Chem Int Ed Engl 2008; 47:4110-3; PMID:18435520; http://dx.doi.org/10.1002/anie.200800726.
79. Nguyen KL, llano M, Akari H, Miyagi E, Poeschla EM, Strebel K, et al. Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression. Virology 2004; 319:163-75; PMID:15015498; http://dx.doi.org/10.1016/j.virol.2003.11.021.
80. Romier C, Ben Jelloul M, Albeck S, Buchwald G, Busso D, Celie PH, et al. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Acta Crystallogr D Biol Crystallogr 2006; 62:1232-42; PMID:17001100; http://dx.doi.org/10.1107/S0907444906031003.
81. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 2008; 89:392-400; PMID:17896349; http://dx.doi.org/10.1002/bip.20853.
82. Ennifar E, Walter P, Dumas P. A crystallographic study of the binding of 13 metal ions to two related RNA duplexes. Nucleic Acids Res 2003; 31:2671-82; PMID:12736317; http://dx.doi.org/10.1093/nar/gkg350.
83. Bernacchi S, Ennifar E, Tóth K, Walter P, Langowski J, Dumas P. Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. J Biol Chem 2005; 280:40112-21; PMID:16169845; http://dx.doi.org/10.1074/jbc.M503230200.
84. Ennifar E, Bernacchi S, Wolff P, Dumas P. Influence of C-5 halogenation of uridines on hairpin versus duplex RNA folding. RNA 2007; 13:1445-52; PMID:17630326; http://dx.doi.org/10.1261/rna.408507.
85. Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, et al. kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. J Am Chem Soc 2012; 134:559-65; PMID:22126339; http://dx.doi.org/10.1021/ja209057d.

©2012 Landes Bioscience. Do not distribute.

RESULTATS PARTIE 2 : RECONSTITUTION DU COMPLEXE VIF/APOBEC3G EN CELLULES BACTERIENNES E. COLI

I. Introduction et objectifs

Vif diminue le niveau d'APOBEC3G/F dans les particules virales, limitant ainsi leur action antivirale comme il a été décrit dans l'introduction. Le mode d'action de Vif a été étudié et il est aujourd'hui bien connu. Vif entraine la diminution du taux d'A3G cellulaire sous différentes manières mais dans ma thèse je m'intéresserai plutôt à la dégradation d'APOBEC par la machinerie du protéasome 26S suite à sa liaison avec Vif et le recrutement du complexe de l'E3 ubiquitine ligase. Cette voie d'inhibition d'A3G présente une piste thérapeutique très importante contre le VIH-1, mais pour cela une élucidation des détails des interactions entre Vif et APOBEC3G/F serait indispensable. En effet, et pour accéder à cet objectif, une détermination de la structure cristallographique du complexe de la E3 ubiquitine ligase formé de CBFβ-Vif-EloB-EloC-Cul5-A3G/F serait d'une grande priorité.

Au cours des années précédentes, de nombreux sites de liaison d'A3G et A3F ont été mis en évidence sur Vif. On sait aujourd'hui qu'A3G/F se lie préférentiellement au domaine N-terminal de Vif, domaine de liaison à l'ARN. Parmi ces sites de reconnaissance on cite le motif ^{40}YRHHY^{44} de Vif qui est important pour la liaison d'A3Get le motif ^{14}DRMR^{17} pour la liaison d'A3F. Les nombreuses études réalisés sur le complexe Vif/A3 ont montré que les interactions entre ces deux protéines sont plutôt de nature électrostatique et que le N-terminal d'A3G/F est principalement impliqué dans l'interaction avec Vif. Les hypothèses et les études peuvent élucider beaucoup de ces mécanismes de reconnaissance mais néanmoins seule une structure du complexe formé entre Vif et A3G ou A3F permettrait de comprendre cette interaction.

II. Stratégie

La découverte du role majeur joué par le facteur de transcription CBFβ en 2012 représente une étape clé dans les recherches menées sur la protéine Vif. Le complexe Vif/CBFβ permet la stabilisation de la protéine Vif au sein du complexe de l'E3 ubiquitine ligase et permet son expression de façon soluble dans la souche bactérienne *E. coli* et en quantité adéquate avec les études biochimiques [158, 165, 235].

Depuis cette découverte, les efforts se sont multipliés afin de résoudre la structure du complexe Vif/CBFβ/EloB/EloC/Cul5/A3G/F. Réussir à obtenir la protéine Vif soluble est en effet essentiel afin de conduire des analyses structurales telles que la cristallographie, RMN, cryo-microscopie électronique, etc. Produire la protéine en la dénaturant puis la renaturant comme décrit dans la partie 1 des résultats rendait extrêmement diffficle la tache de cristallisation, vu que Vif oligomérise puis s'agrège, ce qui représente une énorme limitation pour la cristallogenèse.

A cette étape de ma thèse, la structure cristallographique du complexe Vif-CBF-EloB-EloC-Cul5 n'était pas encore connue et mes efforts étaient donc concentrés sur l'obtention de ce complexe en vue d'en résoudre la structure. La cristallisation de ce complexe avec A3G/F constituait le but ultime de ma thèse. Plusieurs étapes étaient nécessaires pour arriver à cet objectif, dont l'optimisation des conditions d'expression, de la purification et d'association de ce complexe dans *E. coli*:

1- Clonage des gènes dans des vecteurs d'expression et de co-expression
2- Tests d'expression et de co-expression
3- Optimisation des conditions d'expression et de purification
4- Former le complexe *in vivo* ou *in vitro*
5- Evaluer la qualité des protéines par des études structurales
6- Tests d'activité d'A3G
7- Tests de cristallisation

III. Résultats

A. APOBEC3G/F

Le domaine N terminal d'A3G/F est responsable de l'interaction avec la protéine Vif alors que le domaine C-terminal est responsable de l'activité catalytique. Vu le manque de données structurales concernant les protéines A3G/F, une surexpression des protéines entières ainsi que les domaines N et C terminaux séparés était nécessaire pour des études de cristallisation et des études structurales. Une fois les protéines exprimées, une évaluation de la pureté et du repliement correct des protéines sont réalisés, ainsi qu'un test d'activité montrant que les protéines sont bien fonctionnelles. D'autre part, il était aussi important d'étudier les aspects thermodynamiques de l'interaction entre APOBEC et la protéine Vif.

La deuxième partie de ma thèse consistait en la réalisation de tests de co-expression entre Vif et APOBEC afin d'obtenir ce complexe pure et le cristalliser.

1. Caractérisation de APOBEC3G-CD2

a. Expression et purification de hA3G-CD2

Le domaine C-terminal d'A3G responsable de l'activité de déamination de l'ADN viral a été cloné dans le vecteur pGex l'exprimant en fusion avec la GST en N-terminal et un site de clivage à la protéase Prescission (GE Healthcare). Des cellules Rosetta2 ont été transformées par ce vecteur. Une fois les cellules lysées et le surnageant isolé, nous l'ajoutons à la solution de billes d'agarose préalablement resuspendues avec du tampon APOBEC (Tableau 16). Les billes d'agarose sont couplées avec le glutathion. Après centrifugation, le culot va contenir les billes liées à la GST qui est liée à la protéine recherchée. Les autres protéines se trouveront dans le surnageant. Nous avons pu obtenir 6 mg de protéine pour 2 l de culture. Les analyses en spectrométrie de masse réalisées (effectuées par Philippe Wolff au sein de l'équipe) ont confirmé l'identité de la protéine. La GST a pu être clivée de la protéine par l'addition de la Prescission 16h à 4°C (Fig. 65 piste +Pre). Pour améliorer la pureté de la protéine, une purification sur tamis moléculaire avec une colonne Superdex 75 a été réalisée et les fractions les plus pures, contenant le moins de contaminant selon le gel SDS, sont regroupées (Fig. 66). La protéine CD2-hA3G a été produite en grande quantité dans l'objectif de tester son activité et de réaliser une étude cinétique par ITC.

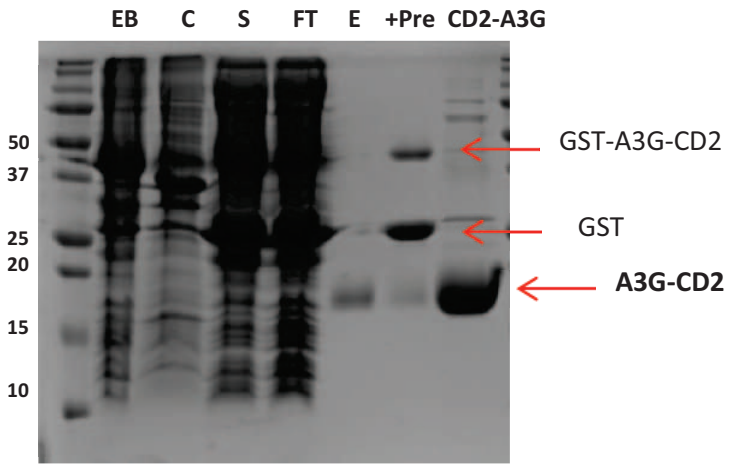


Figure 65 : Purification d'A3G-CD2 sur colonne d'affinité. Le gel SDS-PAGE montre que la protéine A3G est bien exprimée en fusion avec la GST. Après addition de la protéase PreScission, la GST a été clivée et on récupère A3G-CD2 avec encore un peu de contaminant.

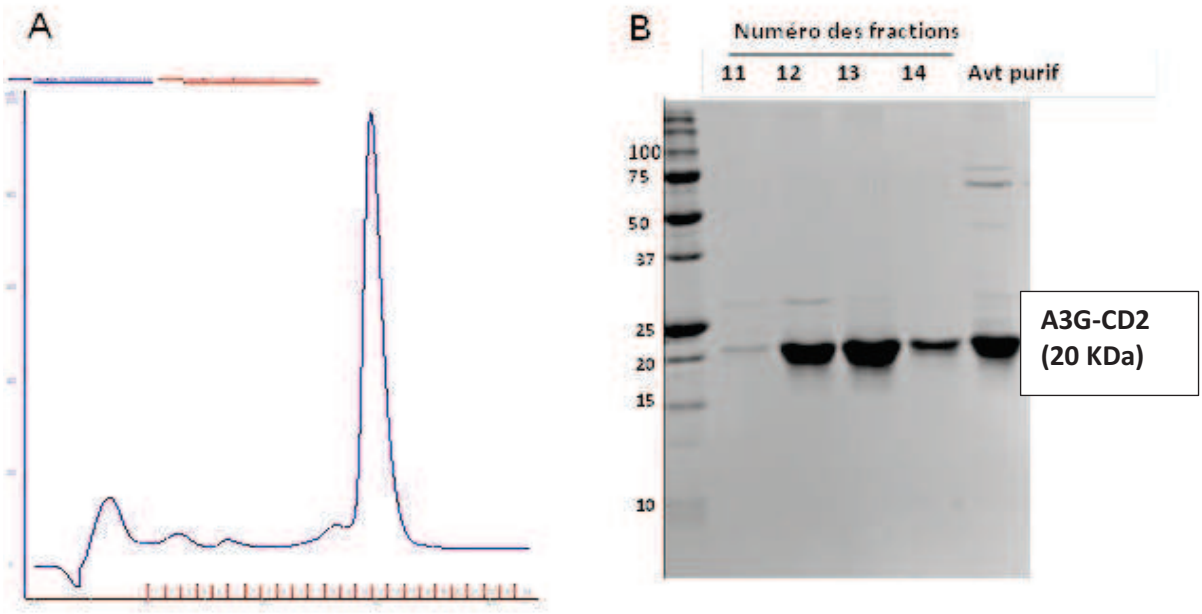


Figure 66 : Purification de A3G-CD2 par tamis moléculaire.
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant hA3G-CD2. Le pic d'absorbance apparait des fractions 11 à 14 (courbe bleue). Le gel montre que les fractions 12 à 14 sont relativement pures en comparaison à la protéine avant purification (dernier puits).

b. Caractérisation biophysique de hA3G-CD2

Afin d'étudier les aspects thermodynamiques de l'interaction entre A3G-CD2 et la protéine Vif purifiée dans des conditions dénaturantes, il était indispensable d'évaluer la qualité de l'échantillon pour les mesures d'ITC. Ces études ont été réalisées sur la protéine A3G encore fusionnée avec la GST. Des mesures en DLS (Fig. 67) ont montré que la protéine purifiée CD2-A3G fusionnée avec la GST est monodisperse et sans agrégats avec un diamètre hydrodynamique de 11 nm. Etant donné que les structures de ce domaine d'A3G et de la protéine GST sont connues, il est possible d'estimer le diamètre total de la protéine de fusion. La GST formant un dimère en solution (la dimérisation s'effectuant par les deux extrémités C-terminales), nous pouvons estimer que le dimère GST-CD2-A3G aura un diamètre de 11.2 nm (Fig. 67), ce qui est parfaitement cohérent avec le diamètre hydrodynamique observé en DLS.

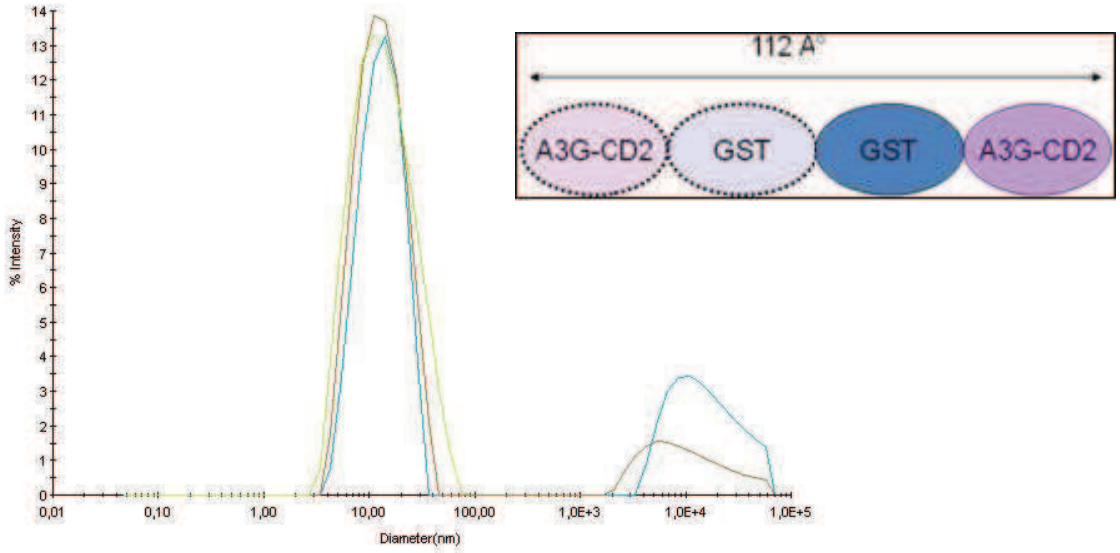
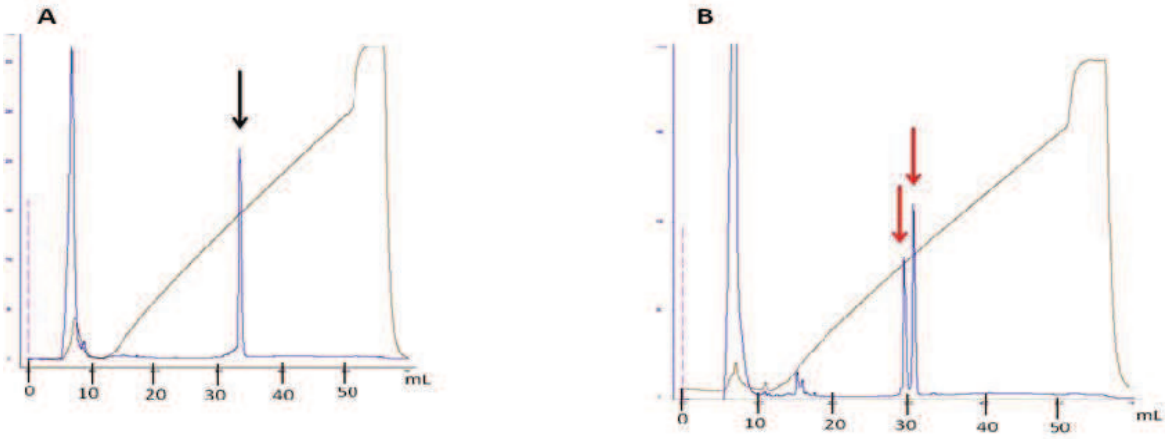


Figure 67 : Analyse de CD2-A3G par DLS. Les résultats de 3 mesures à 1 mg/ml de GST-CD2-A3G sont présentés (par répartition en intensité ; les agrégats visibles sur deux courbes ne sont donc pas significatif ni en masse, ni en nombre). Insert : Modèle de dimérisation de GST-CD2-A3G.

c. Test d'activité de hA3G-CD2

Ce test consiste à vérifier que le domaine hA3G-CD2 possède bien une activité cytidine désaminase (Matériels et Méthodes §II-H). Pour cela la protéine va être mise en présence d'un ADN substrat qui contient le motif CCC, motif cible pour la désamination. Ainsi si hA3G-CD2 est active elle sera capable de convertir le motif CCC en motif CCU. L'apparition de 2 pics sur le chromatogramme signifie que l'UDG a bien reconnue et excisé une Uracyle dans le motif CCU, générant une coupure du brin après traitement alcalin. D'après ce test, la protéine produite s'est avérée active dans du tampon contenant du zinc puisque elle était capable de désaminer l'ADN (Fig. 68C). Les contrôles positifs et négatifs ont été aussi réalisés (Fig. 68A-B). Par contre, l'activité d'A3G était très minimale en absence de zinc. A3G est une protéine à doigts de zinc, il est possible qu'une partie des protéines aient perdues le zinc dans un tampon ne contenant pas ce cation, et soient donc inactives. La présence de zinc dans le tampon permet le repliement correct d'A3G et restaure donc son activité catalytique.



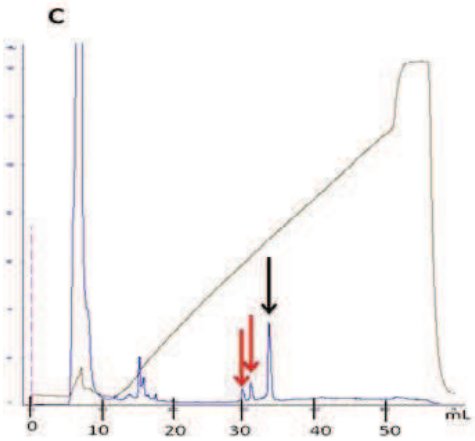


Figure 68 : Chromatogrammes montrant l'activité de hA3G CD2.
A : chromatogramme obtenu pour le contrôle négatif. B : Chromatogramme obtenu pour le contrôle positif. C : chromatogramme obtenu pour le test effectué sur hA3G CD2.
La flèche noire montre le pic correspondant à l'ADN intact ; les flèches rouges montrent les deux fragments d'ADN obtenus après que l'UDG et le traitement alcalin aient clivé l'ADN contenant le motif CCU.

d. Caractérisation des interactions Vif/CD2-A3G

D'après la littérature, le domaine CD1 de A3G est nécessaire à l'interaction avec Vif ex vivo, mais aucune étude *in vitro* n'a porté sur de possibles interactions Vif/CD2-A3G. Disposant de grandes quantités de chaque protéine, nous avons utilisé la microcalorimétrie ITC afin de tester cette interaction. L'expérience a été réalisée à 15°C en ajoutant Vif (200 µM) dans CD2-A3G (20µM). Malgré un signal très faible et bruité, la courbe de titration montre clairement une interaction faiblement exothermique (ΔH = -6.2 kcal/mol) entre les deux protéines, avec un Kd de 164 ± 58 nM dans nos conditions expérimentales (Fig. 69). Pour vérifier que l'interaction observée ne provient pas d'un artefact d'interaction entre la protéine de fusion GST et Vif, nous avons réalisé la même expérience dans les mêmes conditions expérimentales mais cette fois en ajoutant Vif (200µM) dans la GST seule (20 µM). Il n'y a pas d'interaction entre ces deux protéines. Ce qui nous permet de conclure que le domaine CD2 d'A3G responsable de l'activité catalytique pourrait interagir également avec la protéine Vif. Ceci reste un résultat préliminaire qui mérite d'être confirmé par une étude approfondie.

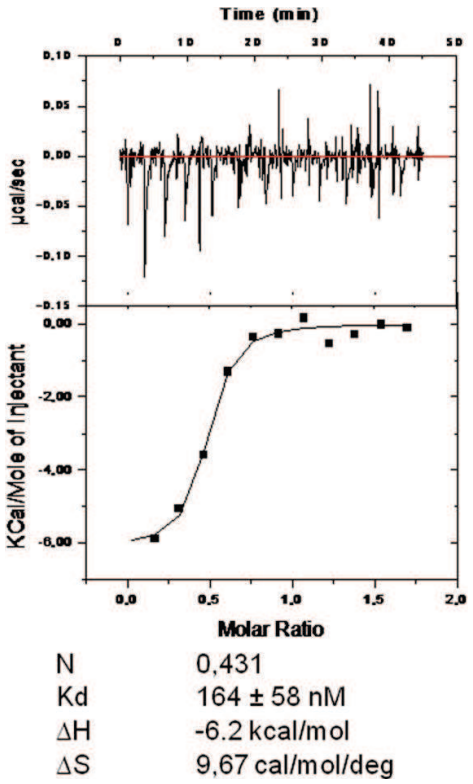


Figure 69 : Caractérisation de l'interaction CD2 d'A3G avec Vif. Malgré un signal bruité dû à une variation d'enthalpie très faible, les résultats obtenus en ITC à 15°C suggèrent que le domaine CD2 de A3G interagit avec Vif avec une affinité sub-micromolaire

2. Caractérisation de APOBEC3F-CD2 [132, 175]

Le domaine C-terminal d'APOBEC3F est insoluble dans *E. coli* et ne peut donc être produit pour des études structurales. A3F-CTD partage des motifs structuraux avec des portions d'A3G-CTD, A3C et A2. Les résidus identifiés importants pour la dégradation d'A3F dépendante de Vif se situent dans une région chargée négativement à la surface d'A3F-CTD. Après alignement avec A3G-CD2, une construction soluble du domaine C-terminal d'A3F a été conçu en effectuant des mutations au niveau de la séquence en comparant avec A3G-CTD soluble sans affecter l'activité de la protéine. Plusieurs étapes de mutagenèse ont donné lieu à 11 substitutions d'acides aminés dans $A3F_{185-373.}$ Ce variant d'A3F était soluble et sa structure a été déterminée par cristallographie [132, 175].

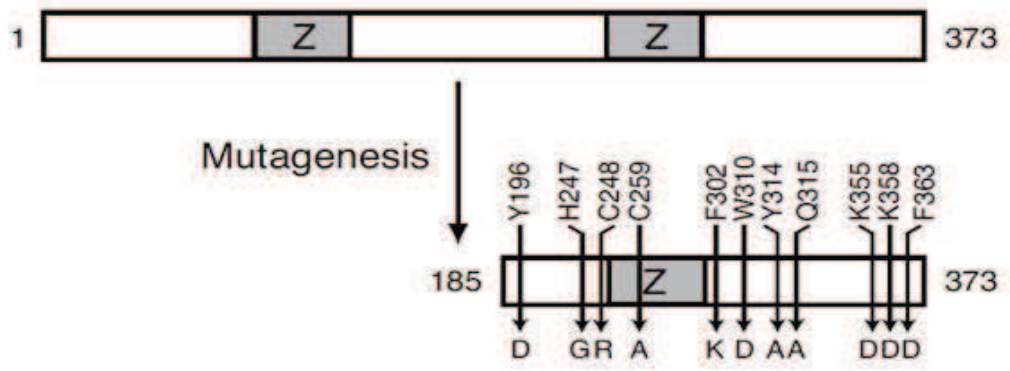


Figure 70 : Schéma de la délétion N-terminal et les substitutions des acides aminés effectués au niveau d'A3F Wildtype afin de produire la chimère soluble A3F185-373-11x.

Avoir une protéine A3F-CDT soluble était très intéressant pour mon projet afin de réaliser la co-expression *in vivo* et *in vitro* entre le complexe Vif-CBFβ-EloB-EloC soluble et ce domaine d'A3F pour cristalliser et résoudre la structure du complexe. De plus, la surexpression de A3F-CD2c pourrait servir pour des expériences d'ITC avec le complexe de Vif afin d'identifier les paramètres thermodynamiques et cinétiques de cette interaction.

a. Expression et purification de hA3F-CD2chimère

Après le clonage de ce mutant dans le plasmide pneaTH ayant un His-tag en N-terminal de la séquence codante et un site de clivage à la thrombine, l'expression de la protéine a été induite par l'IPTG à 25°C sur la nuit. La protéine a été purifiée en premier lieu sur une colonne Nickel puis sur tamis moléculaire (Superdex75). Le tampon utilisé contenait 50 mM Tris pH8, 300 mM NaCl. J'ai réussi à obtenir 20 mg de protéine pour 4 litres de culture.

Pour les expériences d'ITC, A3F-CD2c a été dialysée contre le même tampon du complexe de Vif : 50 mM HEPES pH 7.5, 500 mM NaCl, 10% Glycérol, 1 mM DTT.

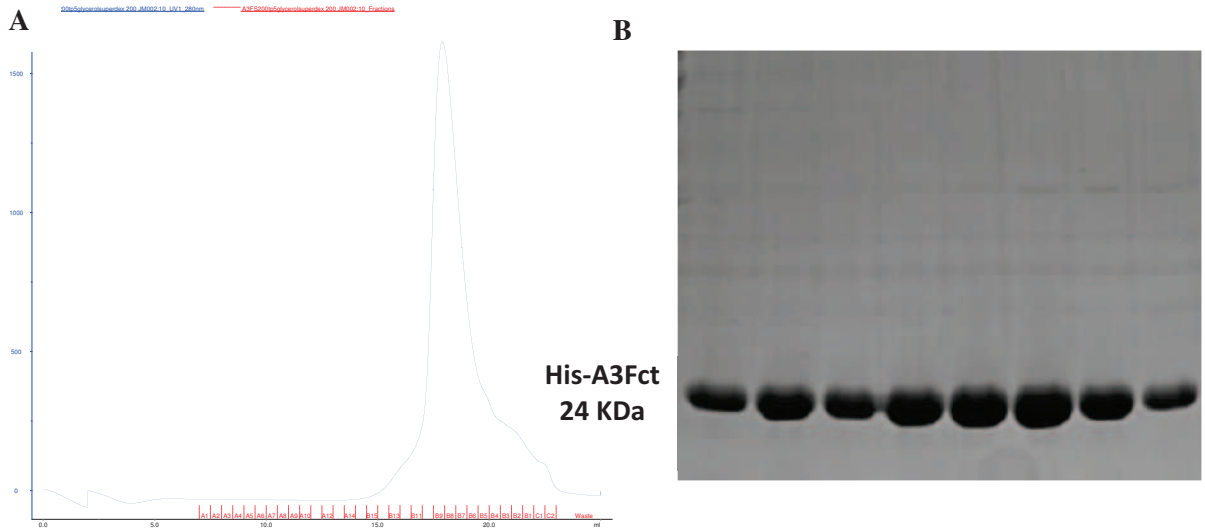


Figure 71 : Purification de A3F-CD2chimère par tamis moléculaire
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant hA3F-CD2c. Les fractions correspondant au pic d'absorbance (courbe bleue) ont été collectées. Le gel montre que les fractions sont pures.

b. Test d'activité de hA3F-CD2c

Le même test d'activité que vu précédemment a été effectué sur la protéine A3F-CD2c purifiée afin de vérifier que le domaine C-terminal chimère d'A3F possède bien une activité désaminase.

Nos résultats ont montré que la protéine n'était pas active. Ceci a probablement affecté les analyses ITC effectués entre le complexe Vif-CBFβ-EloB-EloC (20 µM) dans la cuve et la protéine A3F-CD2c (200 µM) dans la seringue. Nous n'avons pas réussi à mettre en évidence une interaction.

L'absence d'interaction entre le complexe de Vif et la chimère d'A3F-CD2 pourrait s'expliquer par le fait que la protéine produite n'est pas active et pourrait donc être repliée de manière incorrecte, ce qui empêcherait son interaction avec Vif.

Les résultats de co-expression avec Vif-CBFβ-EloB-EloC seront montrés dans les paragraphes suivants.

3. Recherche de constructions solubles d'APOBEC3 par criblage haut débit

Notre recherche par la technique ESPRIT (Matériels et Méthodes III) a été focalisée sur l'obtention de constructions correspondant au domaine CD1 d'hA3G et d'hA3F, ou au domaine CD2 d'hA3F. Des délétions unidirectionnelles en C-terminal (l'extrémité N-terminale étant donc fixe) ont été effectuées sur les deux protéines complètes ainsi que sur une version tronquée des 12 premiers acides aminés en N-terminal, ceci afin de donner plus de variabilité au système. De plus, des délétions unidirectionnelles en N-terminal ont également été effectuées sur hA3F (Fig. 72). Un total de 9216 clones a ainsi été obtenu et testé sur une matrice de colonies (Fig. 73), laissant apparaitre un certain nombre de clones positifs, c'est-à-dire dont les deux extrémités semblent correctement repliées. Finalement, les 96 constructions donnant le meilleur résultat sont isolées et exprimées dans 4 ml de milieu de culture et purifiées par affinité sur une colonne NiNTA.

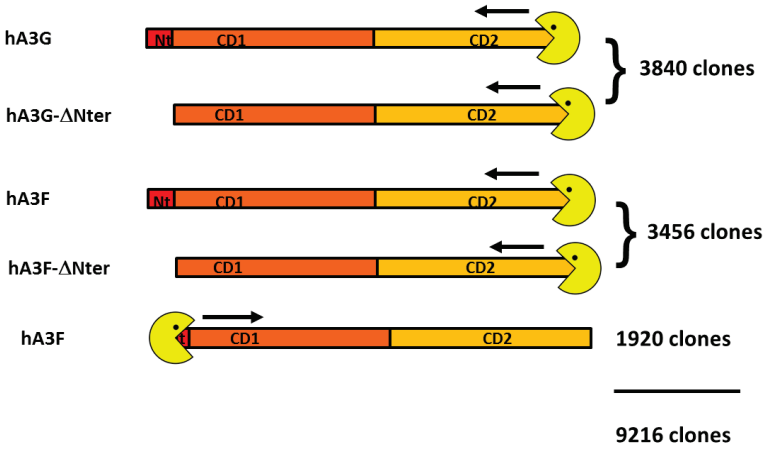


Figure 72 : Schéma de l'expérience ESPRIT que nous avons réalisé, indiquant le nombre de clones obtenus pour chaque construction.

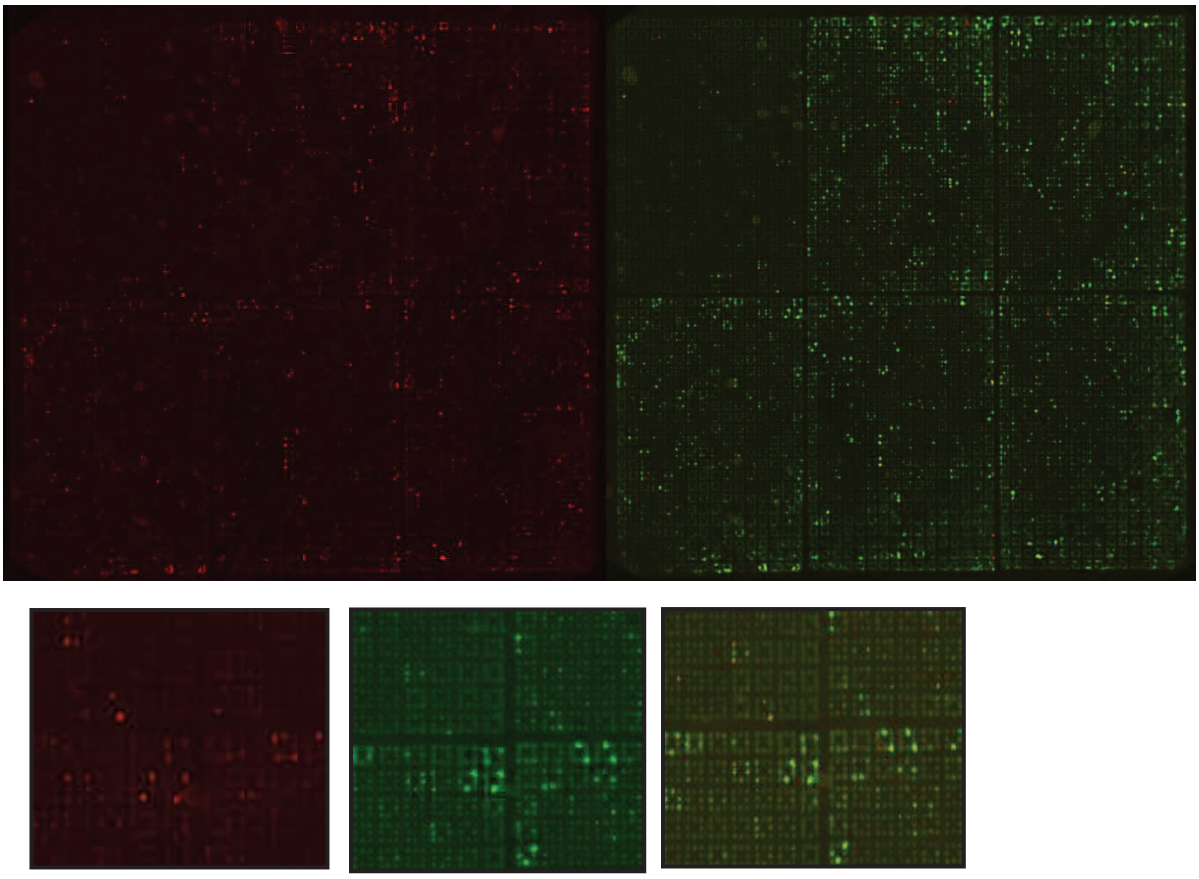


Figure 73 : Exemple de matrice de colonies obtenues pour le criblage des constructions d'APOBEC3F et APOBEC3G. La Figure du haut montre les images de la matrice de colonies scannée pour rechercher un repliement correct de la partie C-terminale (signal vert si présentation correcte du peptide accepteur de biotine – BAP scan) et N-terminale (signal rouge si présentation correcte de l'étiquette histidine – His scan). En bas est représentée une vue détaillée de la matrice avec à droite le signal composite généré par la superposition des signaux rouge et vert. Une construction correctement repliée doit être positive dans les deux canaux et apparaitre donc en jaune sur ce signal composite.

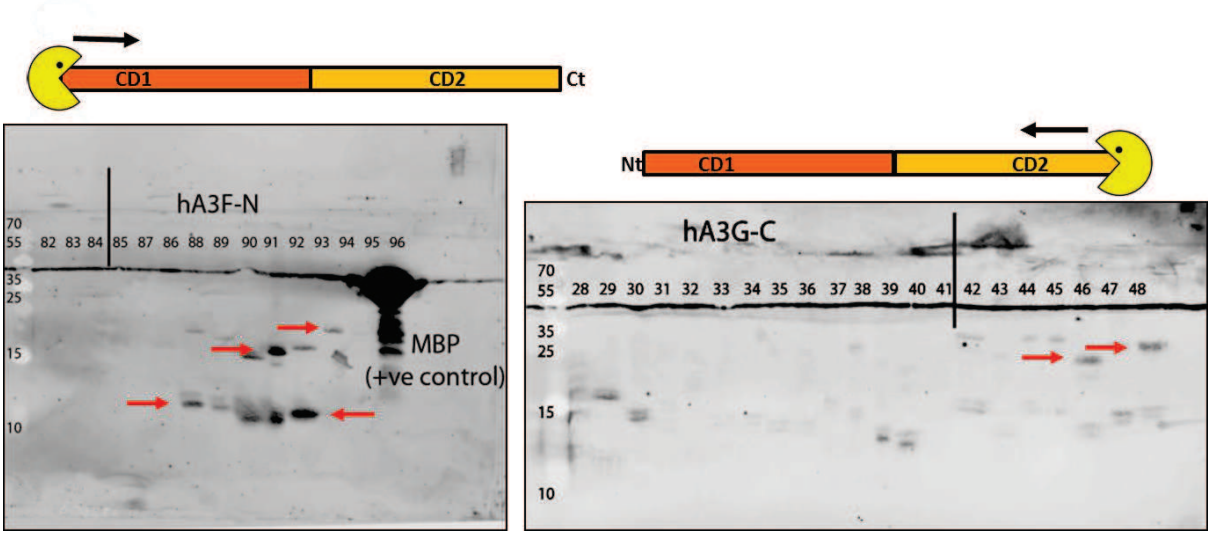


Figure 74 : Exemple de résultats d'expression de constructions issues du criblage ESPRIT pour la recherche d'un domaine soluble du CD2 d'hA3F (à gauche) et du CD1 d'hA3F (à droite) analysées sur un gel SDS. Les flèches rouges indiquent des constructions solubles.

Les constructions sont ensuite analysées sur un gel SDS PAGE (Fig. 74). Les résultats obtenus mettent en évidence des constructions solubles correspondant à des domaines de 12 à 20 KDa du CD2 d'hA3F, ce qui représente des domaines inférieures à ce nous attendions. Concernant le CD1 d'hA3G, les résultats ont montré des domaines solubles de 22 à 25 KDa mais à une très faible expression. Ces données ne nous ont pas permis d'aller plus loin et nous avons donc cherché d'autres solutions pour les expressions d'APOBEC de façon soluble dans *E. coli*.

4. Expression d'APOBEC dans les cellules ArcticExpress

Ce travail a été effectué avec *Cyrielle Silva da Veiga*, ingénieure d'étude au sein de notre équipe. Cyrielle avait cloné Le domaine CD1 et CD2 de A3F et A3G dans le vecteur Pgex6P1 qui porte une GST en N-terminal. De même, des domaines N-terminaux tronqués de 27 Acides aminés d'A3G et A3F. L'expression et la solubilité de ces 5 protéines ont été testées dans des cellules BL21(DE3)RIL et des cellules ArcticExpress (Agilent technonolgies). Cette souche surexprime des protéines chaperonne de l'organisme psychrophile *Oleispira antartica* à basse température (12°C) ce qui permet d'améliorer le repliement de certaines protéines. Les protéines ont été purifiées sur des billes de gluthation et la GST ensuite clivée par addition de la protéase PreScission. Une deuxième étape de gel filtration a permis de mieux purifier la protéine mais n'a pas servi pour la séparation de la chaperonne liée trop fortement aux protéines d'intérêts. Les tests d'expression et de solubilité dans les cellules BL21(DE3)RIL n'ont pas permis d'obtenir des fragments solubles alors que dans les cellules

ArcticExpress on a pu obtenir une petite fraction soluble pour les fragments A3F CD1 , A3F CD2 et hA3G CD1 (Fig. 75) et leur présence dans le surnageant a été confirmée en spectrométrie de masse.

Dans les dépôts effectués pour le test d'expression et de solubilité de hA3F-CD1, une protéine de taille comprise entre 50 et 75 KDa est exprimée après induction. Elle est également présente dans le culot et pourrait correspondre à notre protéine qui couplée à la GST a une taille de 53 KDa ; toutefois l'intensité de la bande étant très faible de nouveaux dépôts ont été effectués avec des quantités plus importantes de culot et surnageant (résultats non montrés) ce qui a permis de visualiser une protéine dont la taille correspond à hA3F-CD1 dans le surnageant.

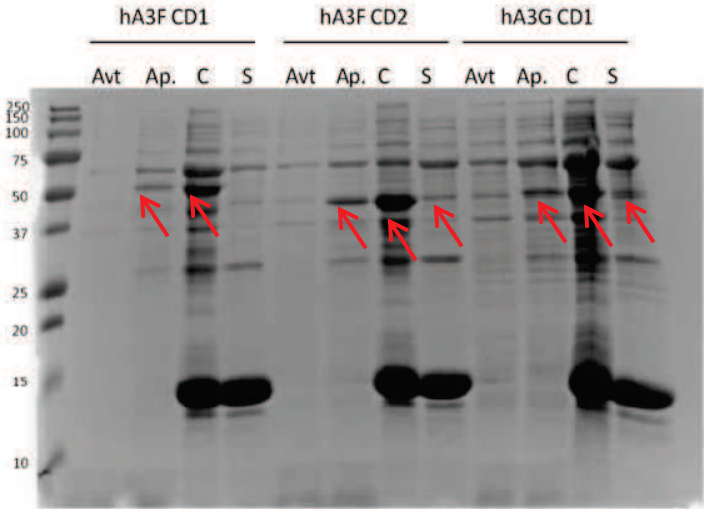


Figure 75 : Gel SDS-PAGE 12% montrant les résultats de l'expression de hA3F CD1, hA3F CD2 et hA3G CD1 dans les cellules ArcticExpress®
Les protéines hA3F CD1, hA3F CD2 et hA3G CD1 en fusion avec la GST sont exprimées dans ce type de cellules. De plus, une partie semble être soluble car une fraction est retrouvée dans le surnageant.

Le problème majeur rencontré est l'association de la chaperonne Cpn60 à nos protéines (Fig.76). Plusieurs tentatives de dissociation ont été tentées en vue de se débarrasser des chaperonnes (ajout d'ATP, de Mg, présence de 2M urée, autres colonnes de purification, etc) mais sans résultats. Nous avons donc abandonné cette voie pour l'expression de protéine soluble.

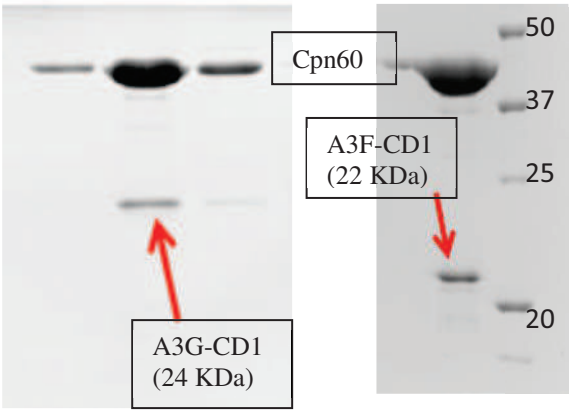


Figure 76 : Gel SDS-PAGE 12% montrant les résultats de la purification en gel filtration. Les domaines N-terminaux d'A3G et A3F ont pu être obtenus solubles mais contaminés par la chaperonne Cpn60.

B. Vif au sein du complexe de la E3 ubiquitine ligase

Suite à la découverte relativement récente de l'importance joué par la protéine CBF-β pour la stabilisation de Vif et sa fonction de dégradation d'APOBEC3G, il était très intéressant d'appliquer la stratégie de co-expression pour produire le complexe entier Vif/CBFβ/EloB/EloC/Cul5 sans et/ou avec APOBEC, ceci afin de le cristalliser. Ce travail a été réalisé parallèlement aux tentatives d'expression des protéines APOBEC seules. J'ai investi beaucoup de temps et de travail sur l'expression du complexe en essayant plusieurs constructions et combinaisons pour la co-expression. La reconstitution du complexe a été effectuée *in vivo* et *in vitro* en utilisant deux isolats de Vif pNL4-3 et HXB2 qui diffèrent de 21 acides aminés *(Annexe)*. J'ai également utilisé deux constructions de CBFβ : la protéine complète de 187 acides aminés, ainsi qu'une version tronquée en N-terminal de 142 acides aminés qui est suffisante pour l'interaction avec Vif [158, 165]. Outre CBF-β, j'ai également co-exprimé et purifié le complexe EloB/EloC, lui aussi important pour la stabilisation de Vif. La protéine Culline5 avec sa protéine chaperonne Rbx2 ont été également exprimées et purifiés pour être ensuite rajoutés au complexe entier.

L'ensemble de ces constructions permet déjà d'accéder à une grande variabilité de complexes possibles en combinant les différentes constructions. Ainsi j'ai réussi des quantités appréciables (2 mg par litre de culture) des deux types de complexes Vif/CBFβ/EloB/EloC sur lesquels j'ai déjà réalisé de nombreux essais de cristallisation.

1. Reconstitution du complexe *in vitro*

En premier lieu, je montrerai les résultats obtenus en essayant de former le complexe E3 ubiquitine ligase *in vitro* en exprimant et purifiant indépendamment les différentes protéines et les différents complexes (CBFβ, EloB-EloC, Cul5-Rbx2, Vif-CBFβ) puis en reconstituant le complexe après une colonne de gel filtration . Les détails et résultats de chacune des expressions est présenté ci-dessous.

a. Expression et purification de CBFβ

i. *CBFβ Full Length*

Après une PCR pour insérer les sites NdeI/BamHI compatibles avec les plasmides de co-expression, CBFβ a été inséré dans le plasmide pnEAtH exprimant la protéine en fusion avec un his-tag en N-terminal de la protéine. Un site de clivage à la thrombine est inséré entre le tag et le gène codant pour permettre le clivage du His-tag.

Les cellules BL21DE3RIL d'E. *coli* sont transformées par le plasmide pmcnEATNH exprimant les protéines CBFβ. L'induction a été réalisée 4 heures à 37°C avec 1 mM IPTG. Les cellules ont été lysées avec 20 mM TrisHCl pH8, 150 mM NaCl, la protéine purifiée sur une colonne d'affinité puis par tamis moléculaire. On a récupéré 5 mg de protéine pure pour 4 litres de culture (Fig. 77).

Le problème rencontré est la dégradation de CBFβFL comme cela est visible sur le gel SDS-PAGE (Fig. 77B). L'expérience a été répétée plusieurs fois sans pouvoir limiter ce problème. Afin d'identifier les régions dégradées, nous avons effectué des analyses en spectrométrie de masse sur la bande la plus intense. La recherche a pu identifier les peptides correspondant au His-tag en N-terminal alors que la région C-terminal de CBFβ a montré la présence de dégradations. Ce résultat n'était pas très étonnant, des études ayant montré que la région C-terminale de CBFβ est flexible et sujette à la dégradation durant les purifications [158].

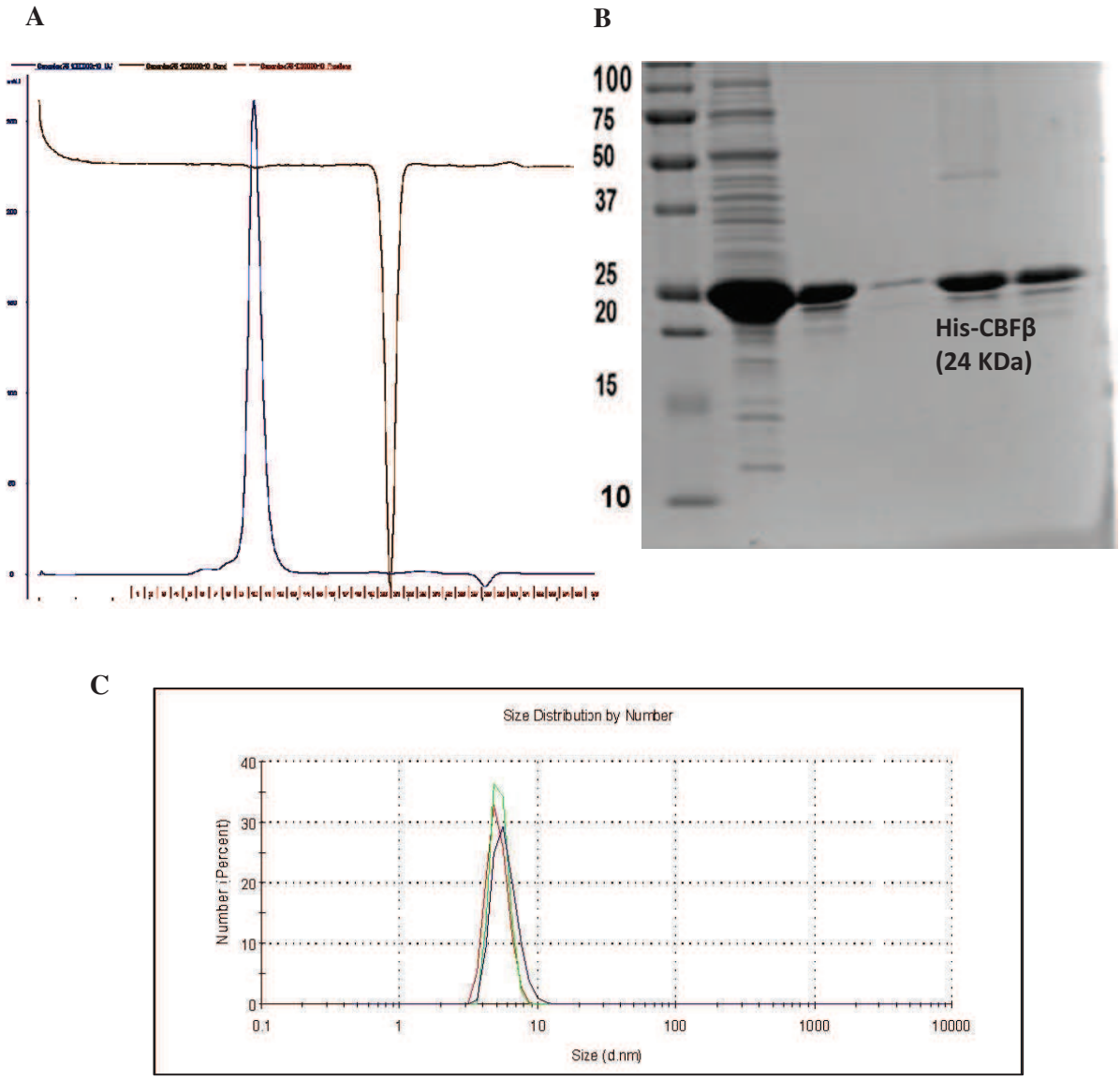


Figure 77 : Purification de 6xHis-CBFβ par tamis moléculaire.
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant CBFβ.
Le pic d'absorbance apparait des fractions 10 à 12 (courbe bleue). Le gel montre que les fractions 10 à 12 sont relativement pures en comparaison à la protéine avant purification (premier puits) mais on observe une dégradation au niveau de CBFβ. C : Analyse de CBFβ par diffusion dynamique de la lumière (DLS). Les 3 courbes correspondent à trois mesures indépendantes. Chaque courbe représente la distribution des tailles de la molécule en suspension (CBFβ) en fonction de l'intensité de lumière diffusée. La protéine est monodisperse sans agrégats.

ii. ***CBFβ142***

Après une mutagenèse dirigée, CBFβ a été délétée de sa partie C-terminale pour éviter les dégradations dû à la flexibilité de cette région. CBFβ142 est la forme tronquée de CBFβ en C-terminal.

Les cellules BL21DE3RIL d'*E. coli* sont transformées par le plasmide pmcnEATNH exprimant les protéines CBFβ142 en fusion avec un his-tag en N-terminal de CBFβ. Les conditions et tampons d'expression et de purification sont les mêmes qu'avec CBFβFL.

5 mg de protéines pures et monodisperses ont été obtenus après la purification (Fig. 78) ; on remarque que cette forme de la protéine n'est pas dégradrée, ce qui nous a poussé à utiliser cette construction pendant les coexpressions, d'autant plus que la partie N-terminale de CBFβ est responsable de l'interaction avec Vif ; la délétion de la partie C-terminale n'affectera donc pas la formation du complexe Vif-CBFβ.

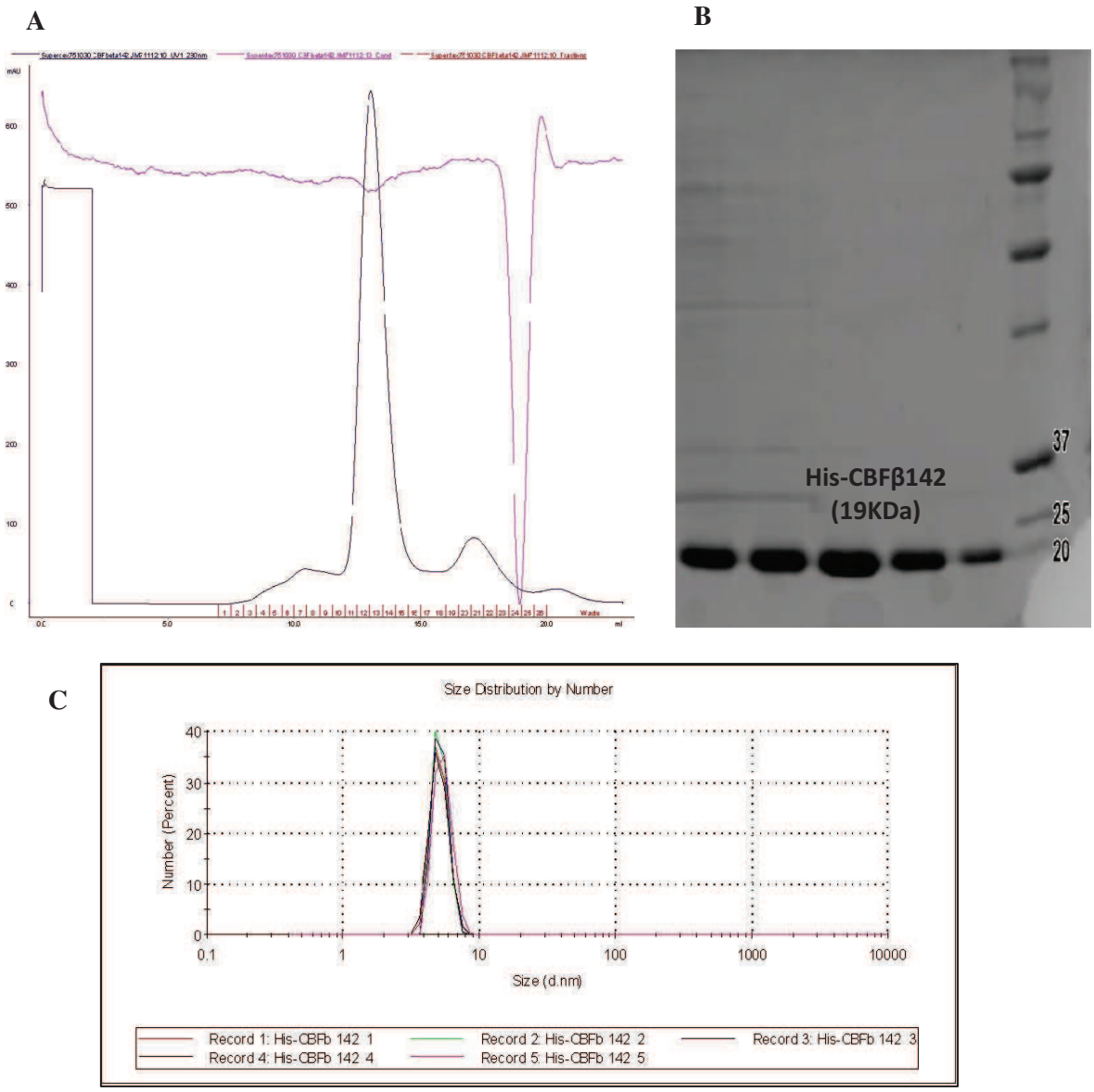


Figure 78 : Purification de 6xHis-CBFβ142 par tamis moléculaire
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant 6xHis-CBFβ142.
Le pic d'absorbance apparait dans les fractions 12 à 16 (courbe bleue). Le gel montre que les fractions 12 à 16 sont pures (dernier puits) et nous n'observons plus la dégradation au niveau du C-terminal de CBFβ. C : Analyse de CBFβ par diffusion dynamique de la lumière (DLS). Les 4 courbes correspondent à quatre mesures indépendantes. Chaque courbe représente la distribution des tailles de la molécule en suspension (CBFβ142) en fonction de l'intensité de lumière diffusée. La protéine est monodisperse sans agrégats.

b. Expression et purification du complexe EloB-EloC

Les gènes codant pour les protéines ElongineB et ElongineC ont été concaténés suivant la méthode de concaténation mise au point par Christophe Romier, avec qui nous collaborons sur ce projet *(Méthodes et méthodes § I-A-3),* et insérés dans un plasmide possédant un his-tag en C-terminal et un site de clivage à la protéase TEV.

Pour l'expression, la même souche d' *E. coli* est transformée par un ce plasmide. Pour la purification, le même protocole a été appliqué pour les 4 protéines. La culture a eu lieu sur la nuit à 18°C après induction avec 1 mM IPTG lorsque la densité optique à 600 nm atteint 0.6. Les cellules sont lysées par Frenchpress dans le tampon 50 mM Tris-HCl pH 7.5, 150 mM NaCl et 1 mM DTT. Les protéines sont purifiées sur une colonne d'affinité puis éluées sur gradient d'imidazole avant une étape de tamis moléculaire qui purifiera les protéines par exclusion de taille. La pureté des protéines est observée sur un gel SDS-PAGE et la monodispersité en DLS (Fig. 79).

Le clivage du his-tag a été réalisé par addition de la protéase TEV ON à 4°C sous agitation et le clivage de la TEV a été vérifié par spectrométrie de masse. Après clivage, 3 mg de protéine pure ont été obtenues pour 4 litres de culture.

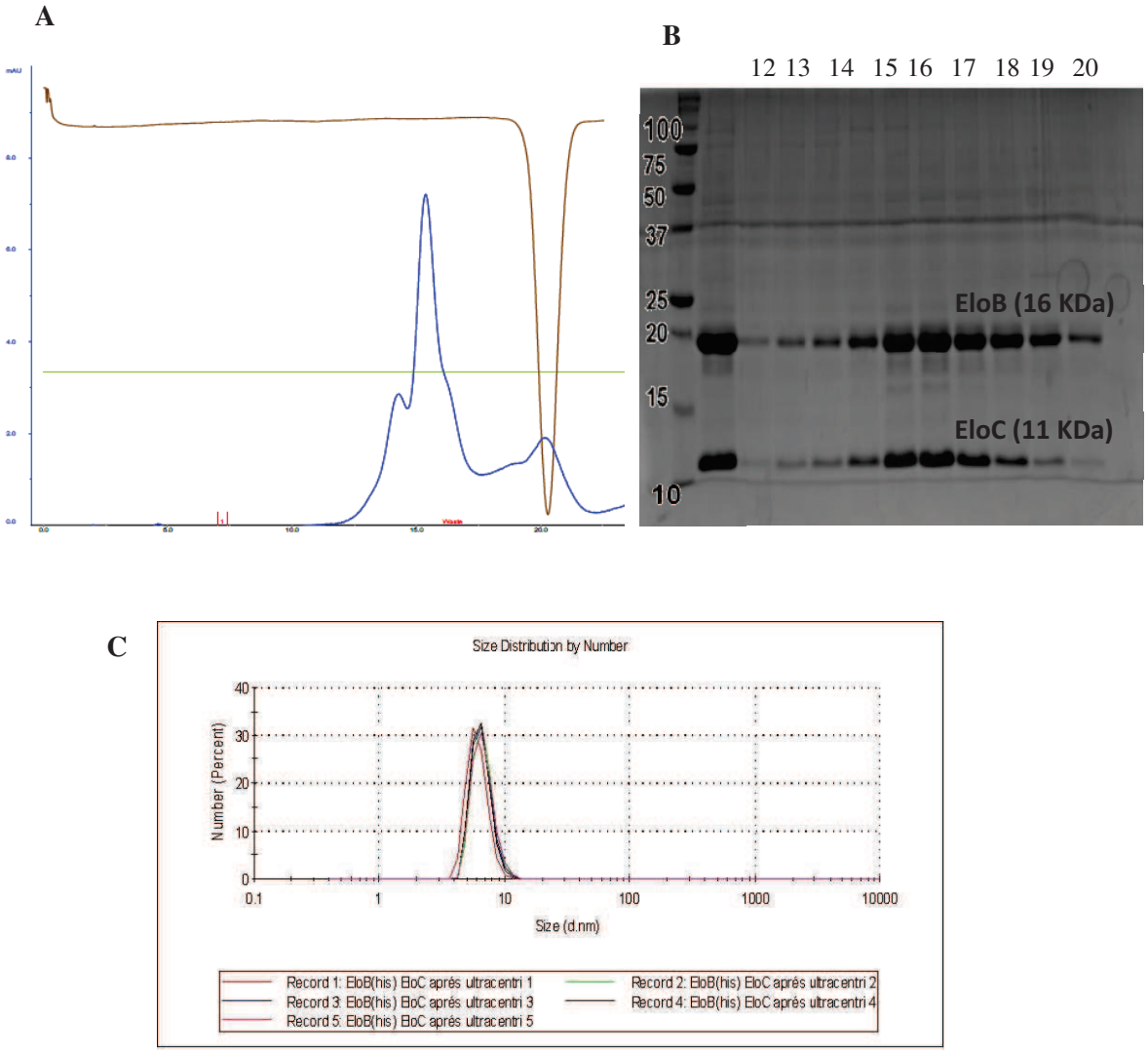


Figure 79 : Purification de EloB(6xHis)-EloC par tamis moléculaire
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant le complexe EloB-EloC.
Le pic d'absorbance apparait dans les fractions 12 à 22 (courbe bleue). Le gel montre que les fractions 12 à 14 sont relativement pures en comparaison à la protéine avant purification (premier puits). C : Analyse d'EloB-EloC par diffusion dynamique de la lumière (DLS). Les 5 courbes correspondent à cinq mesures indépendantes. Chaque courbe représente la distribution des tailles de la molécule en suspension (EloB-EloC) en fonction de l'intensité de lumière diffusée. La protéine est monodisperse sans agrégats.

c. Expression et purification du complexe Cul5-Rbx2

Pour l'expression du complexe Cul5-Rbx2, les gènes codant pour la Cul5 et Rbx2 ont été clonés respectivement dans les deux plasmides pnEAtH et pnCS (sans tag) compatibles avec la co-expression. La sélection de la transformation se fait grâce à la double sélection par les deux antibiotiques ampicilline et spectinomycine.

Le complexe Cul5-Rbx2 est co-exprimée à 20°C avec 0.5 mM IPTG, resuspendu avec un tampon contenant 50 mM HEPES pH 7.5, 500 mM NaCl, 10% Glycérol, purifié une première fois lieu sur une colonne d'affinité ; le his-tag est clivé par addition de la thrombine puis le complexe est purifié une seconde fois sur un tamis moléculaire. Le chromatogramme, qui présente les deux pics attendus pour le complexe et la thrombine, nous indique les fractions à vérifier sur gel SDS PAGE à 15% (Fig. 80A). On dépose des aliquotes des fractions 14 à 23 correspondant au pic du complexe. La révélation nous permet de voir deux bandes visibles correspondant aux tailles de la Culline5 et de RBX2 à environ 95KDa et 10KDa respectivement (Fig. 80B). Pour compléter cette vérification, une analyse par spectrométrie de masse LC-MS/MS a été réalisée (Fig. 81). Elle a confirmé que la taille de l'hétérodimère présent dans la solution correspondait bien à celle attendue pour le complexe Culline5-Rbx-2 sans His tag. La solution est enfin concentrée à 10mg/mL et conservée dans le tampon de stockage avant l'expérience de DLS. Nous avons pu obtenir 2 mg de complexe pure et monodisperse avec un rayon hydrodynamique du complexe d'environ 10nm (Fig. 80C). Cette valeur est probablement surévaluée du fait de la forme allongée de la Culline.

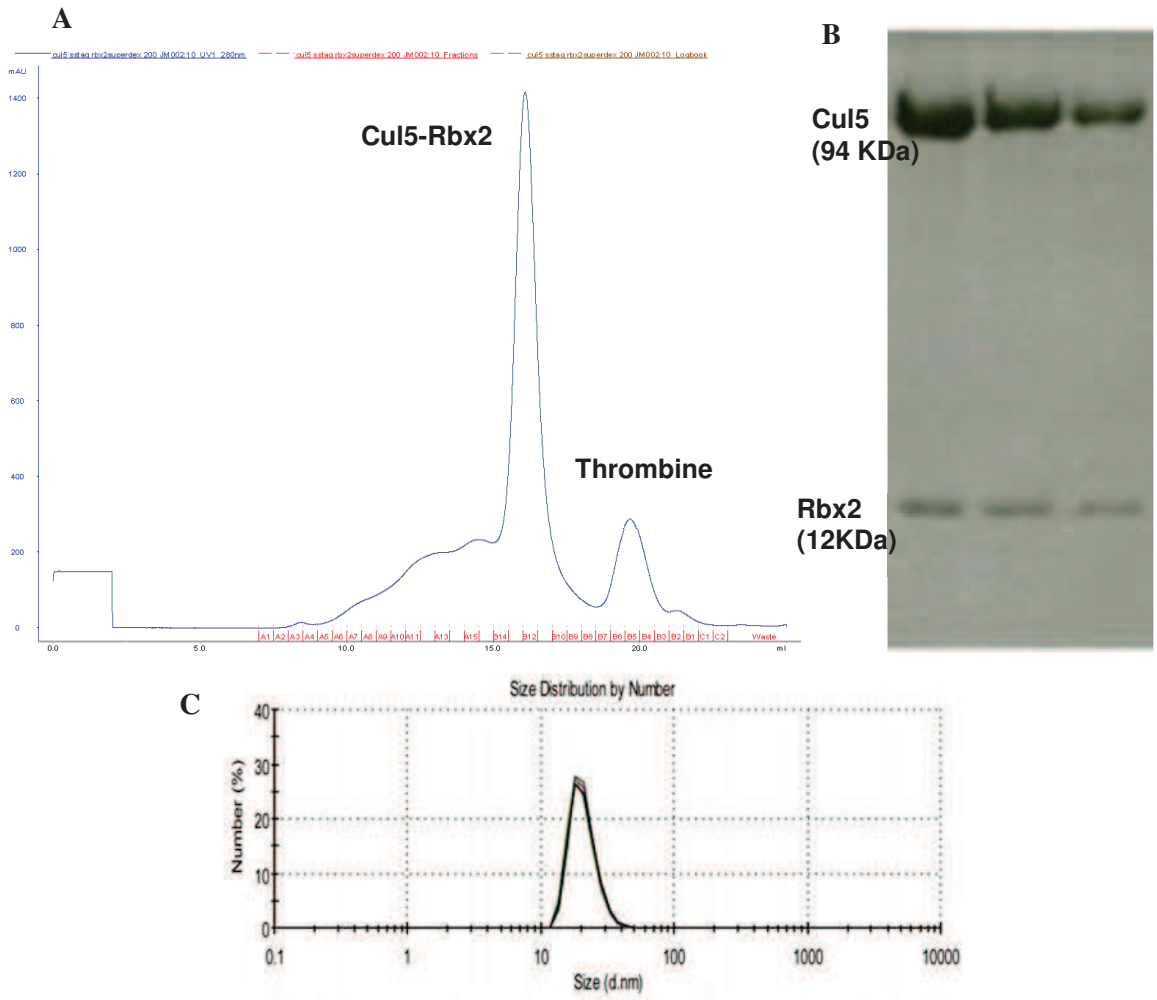


Figure 80 : Purification de Cul5-Rbx2 par tamis moléculaire
A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution B : Gel SDS-PAGE montrant les différentes fractions contenant le complexe Cul5-Rbx2. C : Analyse de Cul5-Rbx2 par diffusion dynamique de la lumière (DLS). Les 4 courbes correspondent à quatre mesures indépendantes. Chaque courbe représente la distribution des tailles de la molécule en suspension (Cul5-Rbx2) en fonction de l'intensité de lumière diffusée. La protéine est monodisperse sans agrégats.

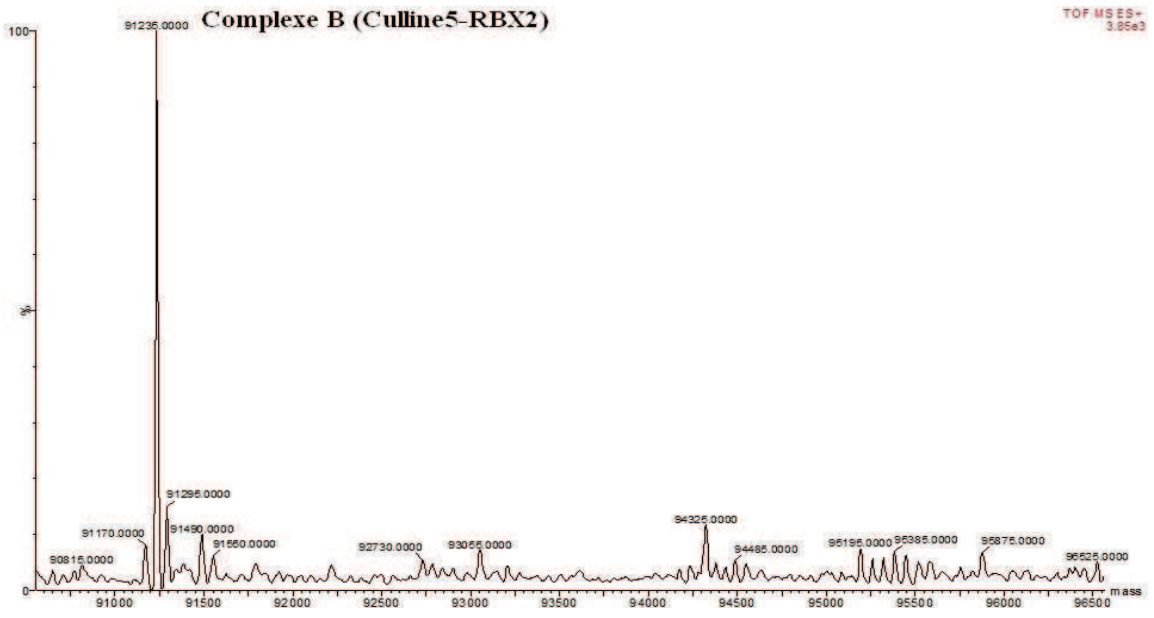


Figure 81 : Spectre d'analyse LC-MS/MS du complexe Cul5-Rbx2. On retrouve la taille exacte attendue pour le complexe après clivage du his-tag par la thrombine (91 KDa).

d. Expression et purification du complexe Vif-CBFβ

Pour compléter le complexe, il reste à exprimer et purifier la protéine Vif. Vif seule n'est pas soluble, d'où l'importance et la nécessité d'y ajouter le cofacteur CBFβ capable de la stabiliser et la solubiliser. Les premiers essais de co-expression du complexe Vif-CBFβ n'ont pas réussi malgré le nombre importants de combinaisons de constructions essayés ainsi que des conditions d'expression et purification.

Deux exemples de co-expression présentés ci-dessous sont :

i. *pMCNCBF/pnEAtHVif*

Dans ce cas, j'ai utilisé le gène de CBFβFL en fusion avec un his-tag en N-terminal en co-expression avec la protéine Vif non taggée. Les gels SDS-PAGE (Fig.82) montrent que les deux protéines ont été bien exprimées seuls ou en co-expression (vérifié par spectrométrie de masse) mais ne sont pas solubles. Le complexe se trouve en majorité dans le culot dans les différentes conditions.

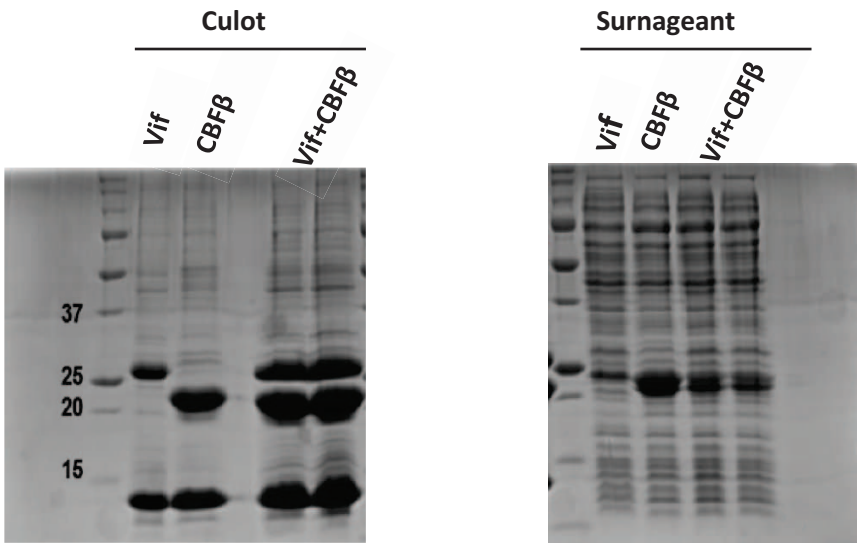
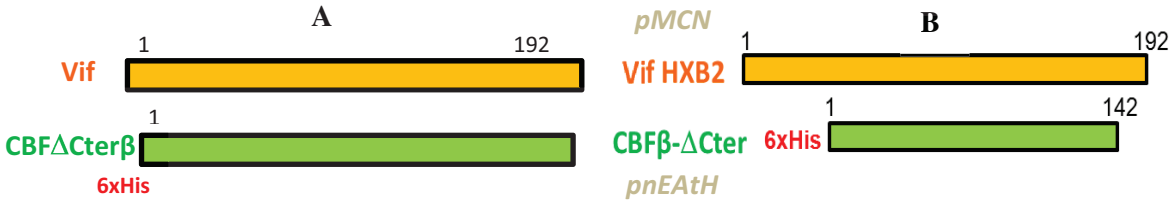


Figure 82: Résultats des expériences de co-expression entre Vif exprimée en fusion avec une étiquette de 6 Histidines en Nt et CBFβ sans tag. Les fractions Culot et Surnageant de chaque expression : Vif seule, CBFseule, Vif/CBF en coexpression ont été déposées sur gel SDS PAGE 12% et les protéines ont été colorées au bleu de Coomassie .

ii. *pMCNVif/pnEAtHCBF*

Dans tous les essais d'expression de Vif taggée seule ou en complexe, la protéine n'est pas soluble ce qui nous a poussé à déplacer le tag de Vif vers CBFβ ou les autres protéines pour la suite des tests de co-expressions. D'autre part, considérant les problèmes de dégradation de CBFFL et sa co-migration avec Vif (vu leur tailles quasiment identiques), nous avons utilisé pour la suite la construction de CBFβ tronquée en C-terminal.



Deux isolats de Vif ont été testés, Vif pNL4-3 et Vif HXB2 qui diffèrent de 21 acides aminés (Annexe 1). Dans les deux cas nous avons travaillé avec des codons optimisées ou harmonisés pour *E. Coli* en vue d'améliorer l'expression des protéines.

Les tests de co-expression effectués entre les différents Vif et CBFβ n'ont pas abouti à des résultats très prometteurs. CBFβ est surexprimée et soluble mais Vif n'est pas surexprimée dans nos conditions ; le complexe ne peut donc pas se former (Fig. 83).

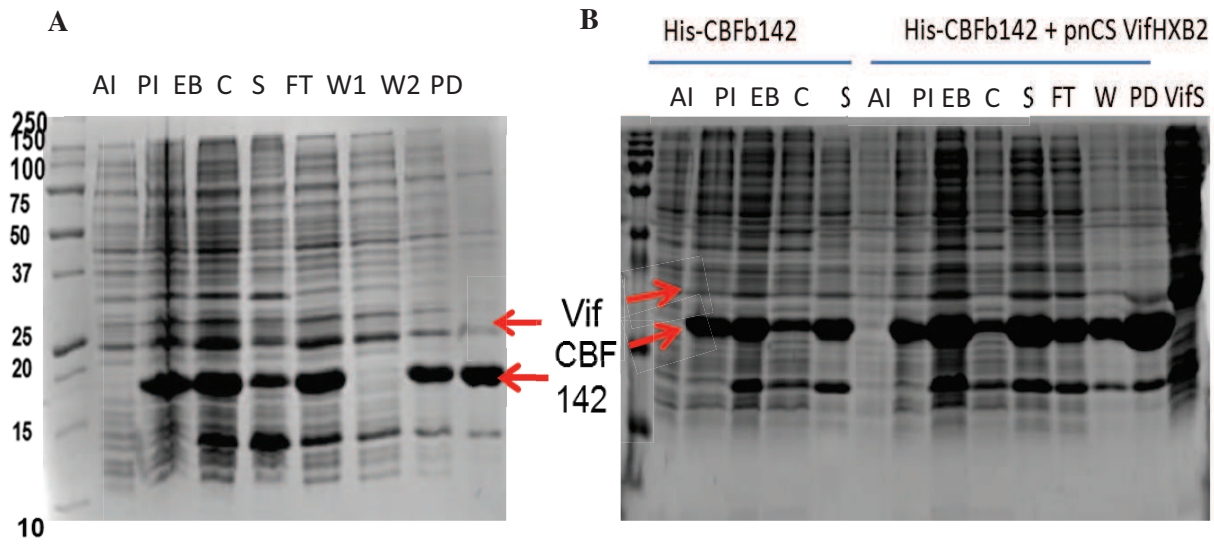


Figure 83 : Résultats des expériences de co-expression entre CBFβ exprimée en fusion avec une étiquette de 6 Histidines en Nt et VifpNL4-3 et HXB2 sans tag. Les fractions collectées tout au long de l'expression et de la purification ont été déposées sur gel SDS PAGE 12% et les protéines ont été colorées au bleu de Coomassie . A : gel SDS PAGE correspondant à la co-expression VifpNL4-3/CBFβ142 ,B : gel SDS PAGE correspondant à l'expression de CBFβ142 seul, puis la co-expression VifHXB2/CBF142.
AI : avant induction, PI : Post-induction, EB : Extrait brut, C : Culot, S : surnageant, FT : Flow through, W : Wash, PD : Pull down

iii. *Essais de coexpression*

Il était important de produire le complexe Vif/CBFβ avant de compléter le complexe avec les autres protéines. Pour cela, j'ai essayé une combinaison de surexpression entre différentes constructions de Vif et de CBFβ, ainsi qu'avec différentes conditions que je ne détaillerais pas dans cette partie (tampon, température d'induction, sels, etc).

	pnCS Vif pNL4-3	pneaTH Vif pNL4-3	pneaVHX Vif pNL4-3	pnCS Vif HXB2opt	pnCS Vif HXB2 Harm	pneaTH VifHXB2 Harm	pneaVHX VifHXB2 harm
pneaTH CBFβ	Non exprimée		Peu exprimée		dissociée		
pneaVH CBFβ	dissociée						
pneaVHX CBFβ	dissociée						
pneaVx-V_H CBFβ	dissociée						
pnCS CBFβ		Insoluble					
pneaTHC BFβ IF1 N.O.				Peu /Non exprimée			
PnCSCBFβ IF1 N.O						Insoluble	
pneaTH CBFβ142	Peu exprimée/ dissociée	Insoluble			dissociée		
pnCS CBFβ142			Non exprimée	Peu/Nonexprimée		Insoluble	Peu exprimée (17°C) Insoluble (37°C)
pneaVHX CBFβ142	Peu exprimé/ Insoluble				dissociée		
pnCSVif+ pneaTH CBFβ	dissociée						
pneaVHEloB/EloC	Insoluble			Peu/Nonexprimée			

Tableau 17 : Tableau regroupant les différentes constructions et combinaisons testés en co-expression entre Vif/CBFβ et Vif/EloB/EloC ; pneaVH : plasmide comportant un his-tag en C-terminal avec un site de clivage à la TEV. pneaVHX : plasmide comportant un His-tag et un tag thioredoxine en N-terminal. pneaVx-VH : Plasmide comportant un his-tag en N-terminal et un tag thioredoxine en C-terminal.

Le tableau 17 regroupe les différentes combinaisons testées et le résultat obtenu pour chacun d'elle. A noter que trois séquences différentes de Vif et de CBFβ ont été utilisées pour les clonages dans les plasmides fournis par C. Romier pour la coexpression. Néanmoins, je n'ai jamais réussi à obtenir le complexe Vif/CBFβ malgré toutes les combinaisons et conditions testées. Les problèmes rencontrés dans ces tests de co-expression sont les suivants :

- Lorsque Vif n'est pas taggée , elle n'est pas surexprimée et donc sa présence est négligeable devant CBFβ.
- Lorsque Vif est taggée, elle est surexprimée mais n'est pas soluble. On retrouve cette protéine majoritairement dans le culot des cellules bactériennes centrifugées.
- Le manque de stœchiométrie entre l'expression de Vif et CBFβ laisse supposer que l'optimisation des codons de CBFβ entraine l'utilisation de la machinerie de traduction de la bactérie au profit de l'exression CBFβ, d'où le manque d'expression de Vif. Pour éviter cela, nous avons essayé une version de CBFβ non optimisée et une version de Vif optimisée , en espérant, mais en vain , rétablir l'équilibre.
- En absence d'EloB-EloC, le complexe Vif-CBFβ ne se forme pas ou bien se dissocie rapidement. Il est connu que EloB-EloC stabilisent la protéine Vif en présence de CBFβ et que cette dernière ne suffit pas pour la solubilisation complète de Vif.

2. Reconstitution du complexe *in vivo*

Je n'ai pas réussi à reconstituer le complexe complet Vif/CBFβ/EloB/EloC/Cul5/Rbx2/A3F/G *in vitro* à cause des deux problèmes majeurs que j'ai rencontré durant les purifications :

- L'insolubilité de la protéine Vif
- L'insolubilité des protéines APOBEC

Pour ces raisons, nous avons décidé de changer de stratégie et d'essayer de produire le complexe et de le reconstituer *in vivo* par co-expression entre les différentes protéines. Cette partie du projet a pris beaucoup de temps d'optimisation afin d'arriver à trouver les bonnes conditions d'expression et de purification du complexe.

a. Optimisation des conditions

Afin d'obtenir le complexe Vif/CBFβ/EloB/EloC/Cul5/Rbx2, j'ai dû essayer plusieurs constructions, combinaisons de co-expressions, tampons, température d'induction, souches de *E. coli* etc.

De même, en ce qui concerne les tests d'expression pour la formation de ce complexe, j'ai réalisé plusieurs constructions et testé plusieurs combinaisons afin de solubiliser et de former le complexe Vif/CBFβ/EloB/EloC/Cul5/Rbx2 *in vivo*. Le tableau 18 résume les différentes combinaisons testées. Avant l'identification des conditions optimales pour l'expression du complexe, toutes ces combinaisons donnaient des résultats décevant (protéines non exprimées, insolubles, complexe dissocié durant la purification, mauvaise stœchiométrie, etc).

Température d'induction	**Concentration tampon**	**Tampon**	**pH**	**Sels**
17, 18, 20, 25, 37°C	20, 50 , 100 mM	Tris, HEPES Na_2HPO4/NaH_2PO4	7, 7,5, 8	100, 150, 200, 300 mM NaCl
Temps d'induction	**Concentration IPTG**	**Lyse des cellules**	**Compléments tampons**	**Purification**
4h, 12h (selon la T°C)	0.2, 0.5, 1 mM	French Press,Sonication, Congélation+Glycogène	Glycérol,Imidazole, zinc, Mg^{2+},DTT, βmercaptoéthanol	En batch, Akta/Co-broyage

Tableau 18 : Tableau regroupant les différentes conditions testés afin d'exprimer et de purifier le complexe Vif/CBFβ/EloB/EloC/Cul5/Rbx2.

b. Co-expression sur deux plasmides différents

Les conditions optimales ont été mises en point et seront détaillées dans la suite :

i. *Expression et purification*

Les cellules BL21 Rosetta 2 d'*E. coli* sont transformées par un plasmide (pmcnEATNH AmpR) exprimant la protéine Vif et la protéine CBFβ142 tronquée en son extrémité Cterminale et en fusion avec un His-tag en Nterminal ainsi que par un plasmide (pmcn-CS SpectR) codant pour l'Elongine B et l'Elongine C (Fig. 84). L'induction se fait avec 0.5 M IPTG lorsque la DO de la culture bactérienne à 600 nm atteint 0.6. La culture reste sous agitation sur la nuit à 15°C. Les cellules sont lysées par FrenchPress dans le tampon : 50 mM HEPES pH 7.5, 500 mM NaCl, 10 % Glycérol et 1 mM DTT en présence d'inhibiteur de protéase. Ensuite, l'extrait brut est ultracentrifugé 1 heure à 30000 g. Le surnageant contenant le complexe est appliqué sur une colonne d'affinité et le complexe sera élué par un gradient d'imidazole. Ensuite, le complexe est dialysé contre un tampon contenant 300 mM NaCl et il est ensuite purifié sur tamis moléculaire (Superdex75) (Fig. 85A). Les fractions

sont collectées et déposés sur gel SDS 12% (Fig. 85B). 1.7 mg de protéines pures et monodisperses (Fig. 85C) ont été obtenus à partir de 2 litres de culture.

Figure 84 : construction basée sur deux vecteurs pnEAtH et pMCN utilisée pour la co-expression du complexe $Vif_{HXB2}/_{(6xHis)}CBF\beta_{1-142}$/EloB/EloC.

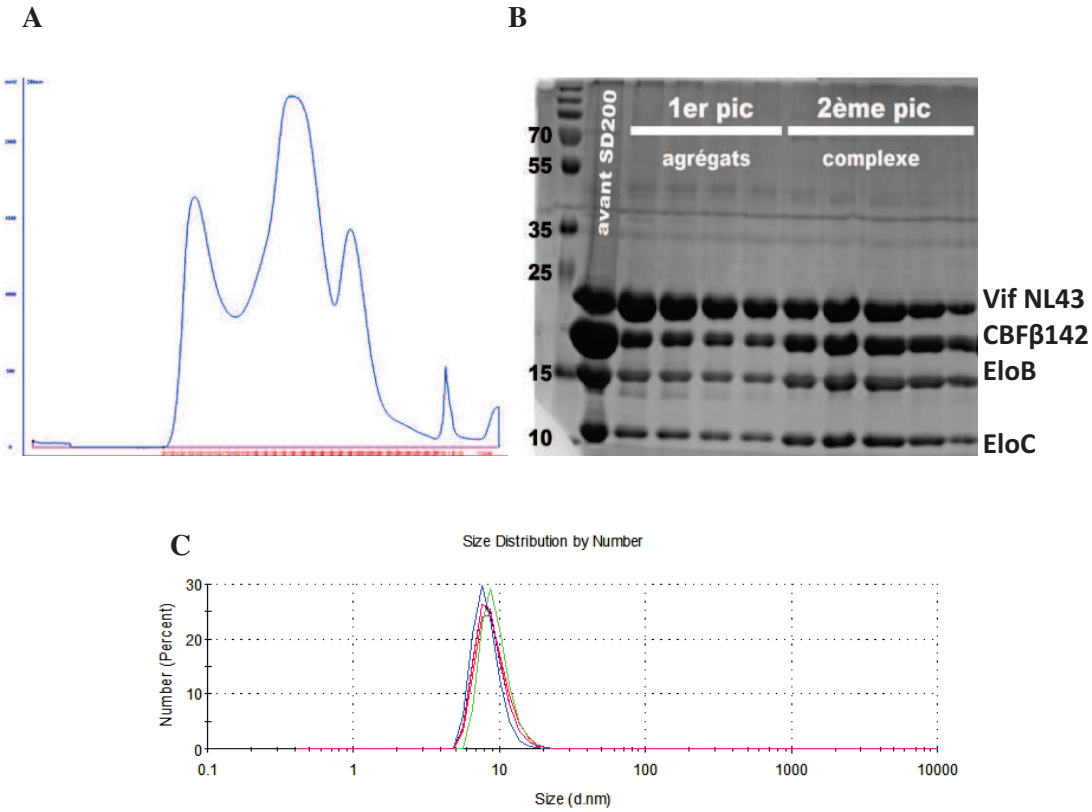


Figure 85: Purification du complexe VifHXB2/(6xHis)CBFβ/EloB/EloC. A : le complexe élué de la colonne de nickel est appliqué sur un tamis moléculaire Superdex200. Le gel SDS en haut à droite montre que le premier pic contient les 4 protéines mais sous forme d'agrégats. Le 3ème pic contient un excès de (6xHis)β pas montré sur le gel). Le second pic contient le complexe quaternaire recherché. La caractérisation par DLS de ce complexe (en bas) montre qu'il est parfaitement monodisperse avec le rayon hydrodynamique attendu, donc apte à des expériences de cristallisation.

ii. *Analyse du complexe par spectrométrie de masse*

L'analyse par spectrométrie de masse en conditions dénaturantes a montré que l'échantillon purifié contient bien toutes les protéines du complexe et avec la taille attendue de chacune (Fig. 86A). D'autre part, le complexe a été aussi analysé dans des conditions natives pour vérifier que toutes les protéines sont bien associées pour former le complexe entier (Fig. 86B). Les analyses ont pu détecter la présence d'un complexe majeur dans l'échantillon à une taille de 53,140 Daltons, cette masse ne correspond pas à la taille attendue du complexe aux environs de 66,600 Da. On suppose que le complexe détecté par la spectrométrie de masse est celui de Vif-CBFβ142-EloC (53,100 Da) et que l'ElongineB s'est dissociée du complexe.

Deux hypothèses sont possibles ; ou bien l'EloB a été dissociée du complexe durant les étapes de purifications, ou bien que les conditions de déssalage du complexe pour les analyses de spectrométrie de masse ne sont pas favorables à l'association du complexe, ce qui a entrainé la dissociation de l'EloB.

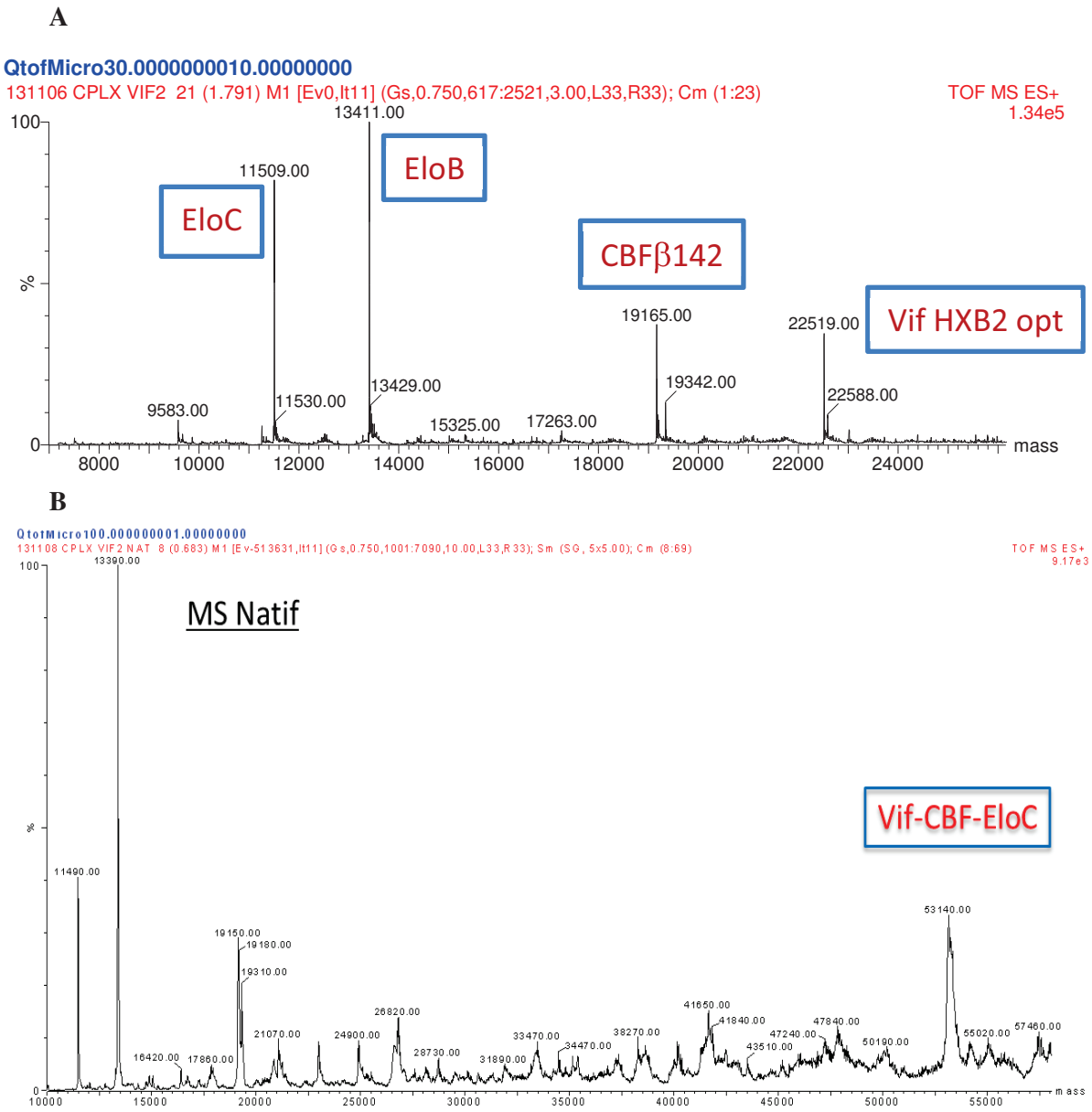


Figure 86 : Analyse par spectrométrie de masse du complexe $Vif_{HXB2}/_{(6xHis)}CBF\beta_{1-142}$/EloB/EloC. A : Spectre d'analyse LC-MS/MS du complexe Vif/CBFβ142/EloB/EloC en conditions dénaturantes. On retrouve la taille exacte attendue pour chacune des protéines. B : Spectre d'analyse LC-MS/MS du complexe Vif/CBFβ142/EloB/EloC en conditions natives pour une détermination de la masse entière. La masse detectée de 53,140 Da correspond au complexe Vif-CBFβ142-EloC sans l'EloB.

c. Co-expression sur un seul plasmide

Pour exprimer le complexe formé entre Vif pNL4-3, CBFβFL, EloB, et EloC, j'ai concaténé les différents gènes sur un seul vecteur pneaTH ayant le his-tag sur l'extrémité N-terminale de CBFβ (Fig. 87).

Figure 87 : construction basée sur un unique vecteur (obtenu après plusieurs concaténations) exprimant le complexe $Vif_{NL4.3}/_{(6xHis)}CBF\beta_{1-187}$/EloB/EloC.

Pour l'expression et la purification du complexe, nous avons appliqué les mêmes conditions que celles décrites dans le paragraphe précédent en ce qui concerne la souche bactérienne, la température d'induction, la concentration en sels.

J'ai réussi à produire 3 mg de protéines avec une pureté encore insuffisante mais monodisperse pour les tests de cristallisation. On observe une dégradation de CBFβ du fait de l'utilisation dea séquence FL sensible à la protéolyse.

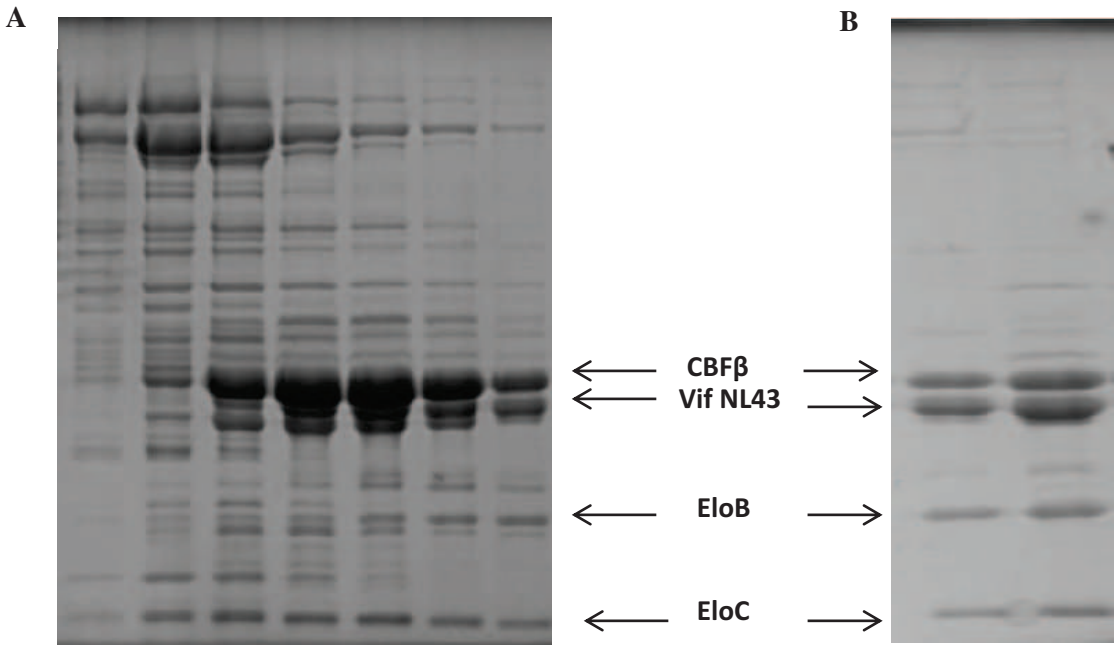


Figure 88 : Purification du complexe $Vif_{pNL4-3}/_{(6xHis)}CBF\beta_{1-187}$/EloB/EloC.

d. Co-expression Vif-CBFβ-EloB-EloC-Cul5-Rbx2

Après avoir trouvé les conditions optimales pour la co-expression du complexe Vif/$_{(6xHis)}$CBFβ/EloB/EloC avec les deux isolats différents de Vif (HXB2 et pNL4-3) ainsi qu'avec les deux constructions de CBFβ FL et tronquée en C-terminal, il fallait compléter le complexe E3 ubiquitine ligase en rajoutant la Culline5 avec sa protéine chaperonne Rbx2. Les tentatives de coexpressions entre ces deux complexes n'ont pas réussi pour plusieurs raisons :

1- Les températures d'expressions après induction à l'IPTG des deux complexes sont différentes : Vif-CBFβ-EloB-EloC (15°C) et Cul5-Rbx2 (20°C).
2- Le nombre de gènes en question est limite pour la co-expression.

Pour ces raisons, les deux complexes ont été exprimés par co-expression séparément , purifiés chacun isolément, puis mélangés *in vitro* pour 2 heures à 4°C et le complexe complet est passé sur une colonne gel filtration pour récupérer le complexe avec une stœchiométrie équivalente pour toutes les protéines (Fig. 89).

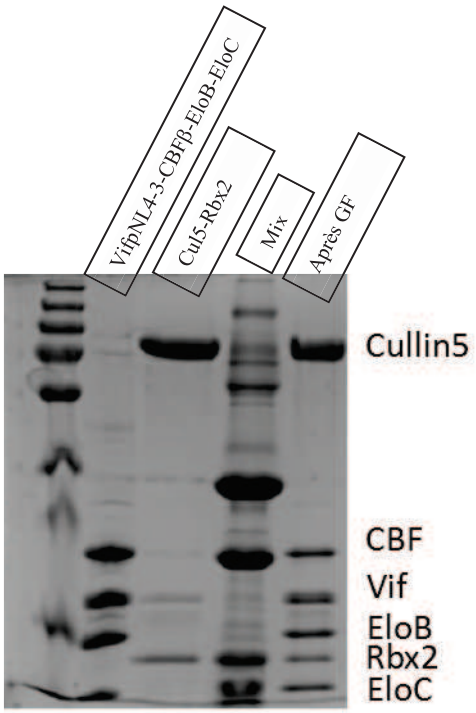


Figure 89 : Purification du complexe Vif-CBFβ-EloB-EloC-Cul5-Rbx2. Gel SDS-PAGE 15% montrant les purifications respectives des deux complexes seuls puis du mélange ;la dernière piste montre la purification du complexe complet après tamis moléculaire.

e. Co-expression Vif-CBFβ-EloB-EloC-APOBEC3G/F

Les protéines APOBEC3G/F FL seules sont insolubles dans *E. coli.* Des tests de co-expression entre le complexe E3 ubiquitine ligase et A3G/F ont donc été réalisés afin de solubiliser APOBEC, pour, par la suite, tenter de cristalliser le complexe entier. Pour ces tests, nous avons utilisé des séquences d'APOBEC dont les gènes ont été optimisés pour *E. coli*, afin d'améliorer l'expression dans les bactéries. Les tests de co-expressions ont été réalisés dans des conditions différentes : trois températures (37°C, 25°C, 15°C) et deux concentrations salines (150 mM et 500 mM NaCl).

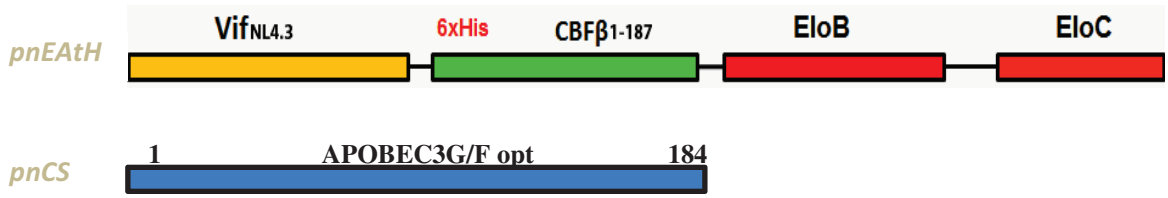


Figure 90 : Constructions utilisées pour la co-expression entre le complexe $Vif_{NL4.3}/_{(6xHis)}CBF\beta_{1-187}$/EloB/EloC et APOBEC3G/F.

Les cellules BL21 de *E. coli* ont été co-transformées par les deux plasmides (Fig. 90). 50 mL de milieu de LB liquide a été inoculé par colonie puis l'expression des protéines a été induite à l'IPTG à 18°C pour 15 heures. Le tampon phosphate pH 7.5 a été utilisé durant les étapes de purification additionné de 150 ou 500 mM NaCl. La purification a été réalisée sur des billes de Nickel en ciblant le his-tag de CBFβ et les différents aliquotes ont été déposés sur gel SDS-PAGE pour identifier le complexe (Fig. 91).

La bande correspondant à la taille d'APOBEC se retrouve à chaque fois dans le culot (Fig. 91 piste C). Cette bande visualisée sur le gel a été analysée par spectrométrie de masse pour vérifier l'identité de la protéine et son abondance dans l'échantillon. Les résultats ont révélé la présence minoritaire d'A3G dans cette bande avec seulement 3 spectres retrouvés alors qu'on retrouve la protéine OMPF avec 664 spectres détectés. Dans cette bande on retrouve aussi 23 protéines contaminantes d'*E. coli* ce qui signifie que A3G/A3F ne sont pas surexprimées.

Pour plus de vérification, nous avons réalisé un Western blot avec des anticorps anti-A3G afin d'identifier A3G dans le gel. La révélation a montré qu'A3G est exprimée (résultat qui n'est pas quantitatif) mais se retrouve dans le culot (Fig. 92). Par contre, on retrouve bien le complexe Vif-CBFβ-EloB-EloC dans les purifications.

On peut conclure que dans nos différents tests de co-expression, A3G/F n'est pas ou est peu surexprimée et la petite quantité exprimée n'est pas soluble. Pour ces raisons, le complexe Vif/A3G n'as pas pu été obtenu.

i. *A3Gopt*

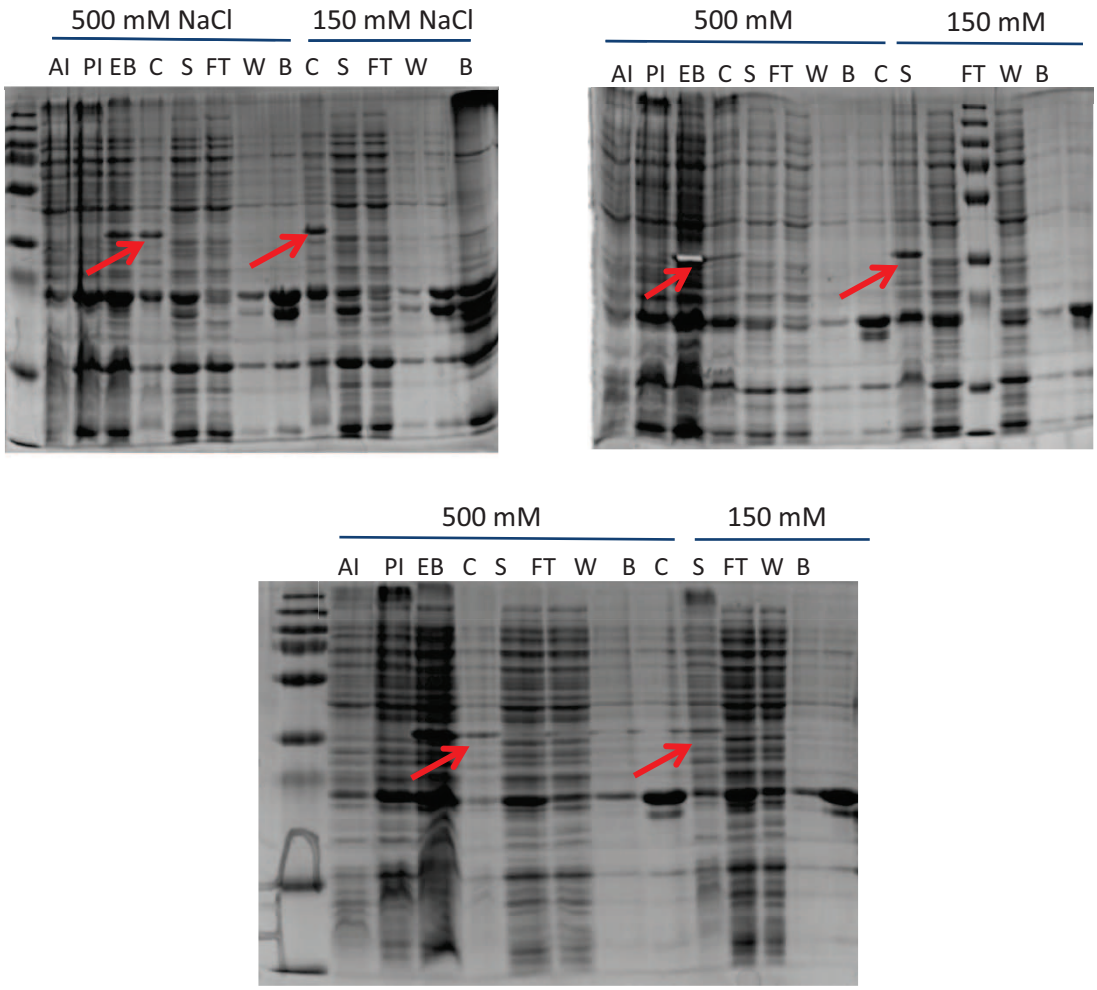


Figure 91 : Test de co-expression entre le complexe pneaTHVifpNL4-3-CBFβ-EloB-EloC et pnCSA3Gopt. Gel SDS-PAGE 15% montrant des aliquotes prélevés à différentes étapes de l'expression et la purification du complexe à trois températures différentes et deux concentrations salines. Dans tous les tests effectués, A3G est retrouvée dans le culot.

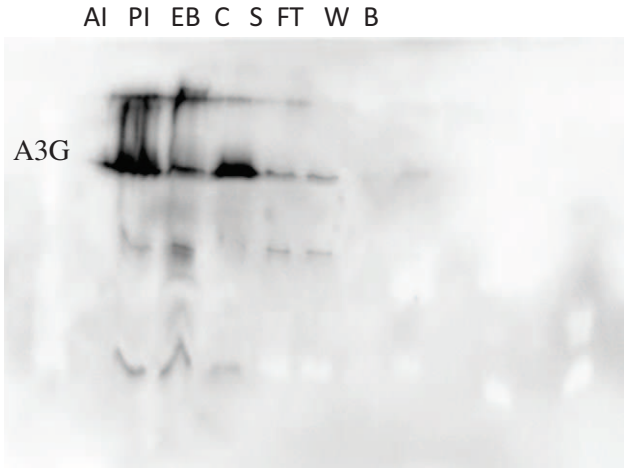


Figure 92 : Western Blot montrant la présence d'A3G dans les différentes pistes par action de l'anticorps anti-A3G. A3G se retrouve dans le culot.

ii. *A3Fopt*

Les mêmes tests de co-expression ont été réalisés entre pneaTHVifpNL4-3-CBFβ-EloB-EloC et pnCSA3Fopt (Fig. 90). Le complexe a été purifié sur des billes de Nickel et déposé sur gel SDS-PAGE. Les résultats montrent qu'A3F se retrouve dans le culot (Fig. 93). Le complexe E3 ubiquitine ligase n'as pas pu solubiliser A3F qui n'est pas non plus surexprimée.

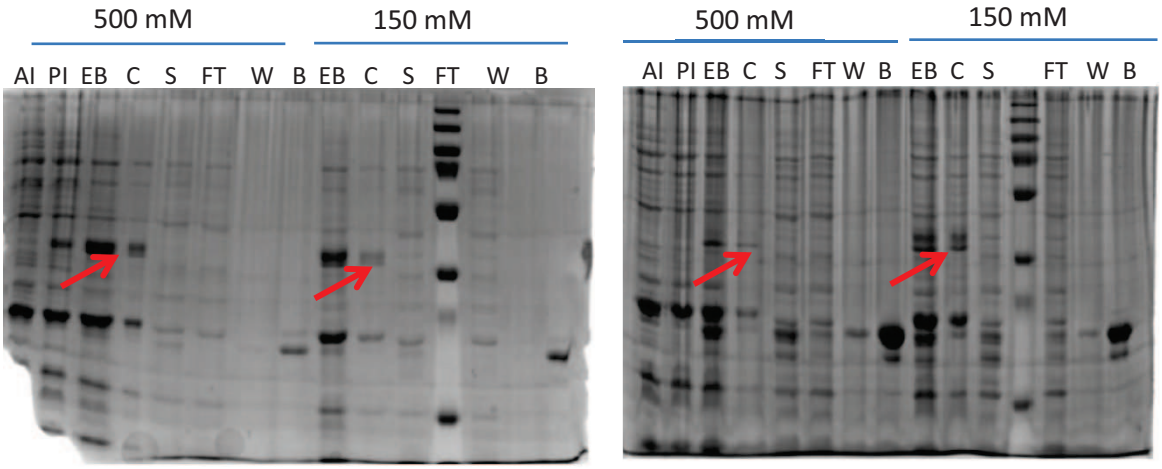


Figure 93 : Test de co-expression entre le complexe pneaTHVifpNL4-3-CBFβ-EloB-EloC et pnCSA3Fopt. Gel SDS-PAGE 15% montrant des aliquotes prélevés à différentes étapes de l'expression et la purification du complexe à deux températures différentes et deux concentrations salines. Dans tous les tests effectués, A3F se retrouve toujours dans le culot.

iii. *A3F-CD2c*

Etant donné que le domaine C-terminal d'A3F interagit avec Vif contrairement au C-terminal d'A3G, et que le domaine C-terminal chimère d'A3F est soluble, nous avons essayé de former le complexe Vif-CBF-EloB-EloC-A3Fct par co-broyage (Matériels et Méthodes §1-C-3b).

Après avoir exprimé le complexe et A3Fct séparément, on mélange les culots cellulaires et on purifie le mélange d'abord sur une colonne Nickel en visant le his-tag de CBFβ puis sur tamis moléculaire. Le complexe correspond aux fractions A14 jusqu'à B13 sur le gel SDS-PAGE (Fig. 94). Cependant, même après le tamis moléculaire, le complexe reste contaminé comme c'est visible sur le gel. Il est important d'améliorer la qualité de ce complexe comportant Vif et A3F-CD2 pour pouvoir effectuer des tests de cristallisation et d'autres études structurales.

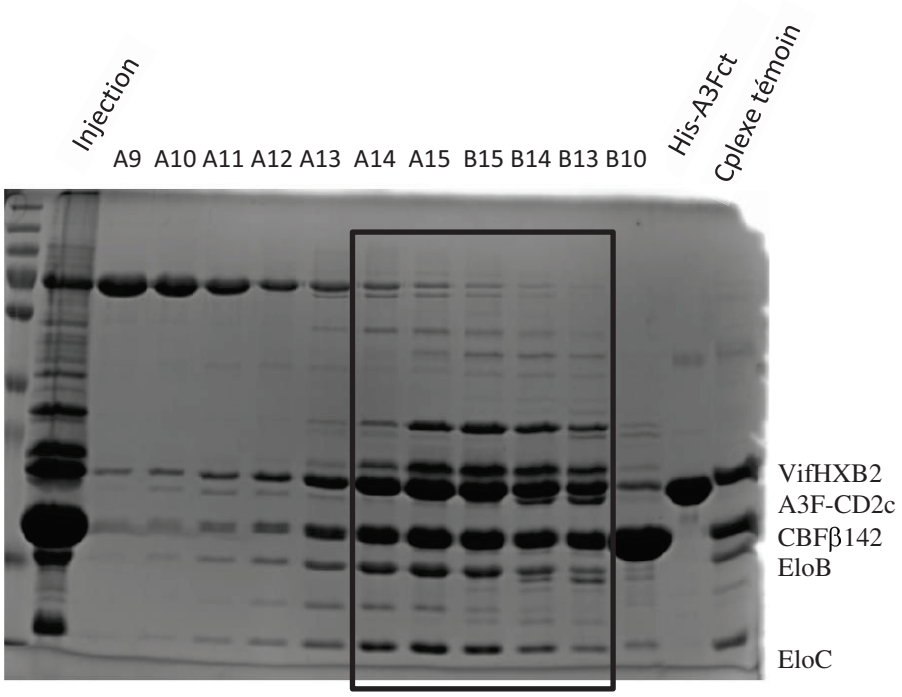


Figure 94 : Purification du complexe Vif_{HXB2}-CBFβ142-EloB-EloC-A3FCD2c. On retrouve les différentes protéines du complexe entre les fractions A14 et B13 de la chromatographie d'exclusion de taille.

3. Tests de cristallisation

Les quatre complexes : Vif_{HXB2}-CBFβ142-EloB-EloC (1), $Vifp_{NL4-3}$-CBFβFL-EloB-EloC (2), $Vifp_{NL4-3}$-CBFβFL-EloB-EloC-Cul5-Rbx2 (3) et Vif_{HXB2}-CBFβ142-EloB-EloC-A3FCD2c (4) ont été exprimés, purifiés en deux étapes successives (affinité et tamis moélculaire) afin d'obtenir un échantillon homogène puis concentrés pour des tests de cristallisation. Un grand nombre de conditions de cristallisation, de températures de cristallisation et de concentration protéique ont ainsi été testées mais aucun cristal n'a pu être obtenu.

Tampon	Concentrations protéiques	Conditions de cristallisation	Températures	Volume goutte
HEPES, Tris, Phosphate	6, 10, 12 mg/mL	Screen , Crystal, PEG ION Screen	4°C, 20°C,	100, 200 , 300 nL

Tableau 19: Tableau résumant les différentes conditions de cristallisation testées pour les quatres complexes purifiés.

Méthylation des lysines

Les protéines formant ce complexe possèdent toutes des lysines. VifHXB2 (14 K), CBFβ142 (6K), EloB (7K), EloC (5K). La flexibilité des chaines latérales des résidus protéiques ainsi que leur charge peut expliquer l'impossibilité de cristalliser ces complexes. Afin de tenter de remédier à ce problème, le complexe Vif_{HXB2}-CBFβ142-EloB-EloC a subit une méthylation sur ces lysines selon un protocole décrit dans la partie « Matériels et Méthodes ». L'introduction de ce groupement méthyl peut permettre la formation de contacts hydrophobes. Une fois les méthylations réalisées, une chromatographie d'exclusion de taille permet la purification du complexe puis une DLS permet de vérifier la monodispersité de l'échantillon pour les tests de cristallisation (Fig. 95). Les méthylations ont été vérifiées par spectrométrie de masse pour s'assurer que toutes les lysines des quatre protéines ont été méthylées (Fig. 96). En effet, une méthylation rajoute 28 Dalton à la masse de la protéine. Cette augmentation a été observée par LC-MS/MS en analysant le complexe dans des conditions dénaturantes. Le complexe a été concentré à 6 mg/mL et utilisé pour des tests de cristallisation (Screen et Crystal screen) à 4 et 20°C mais aucun cristal n'a été observé.

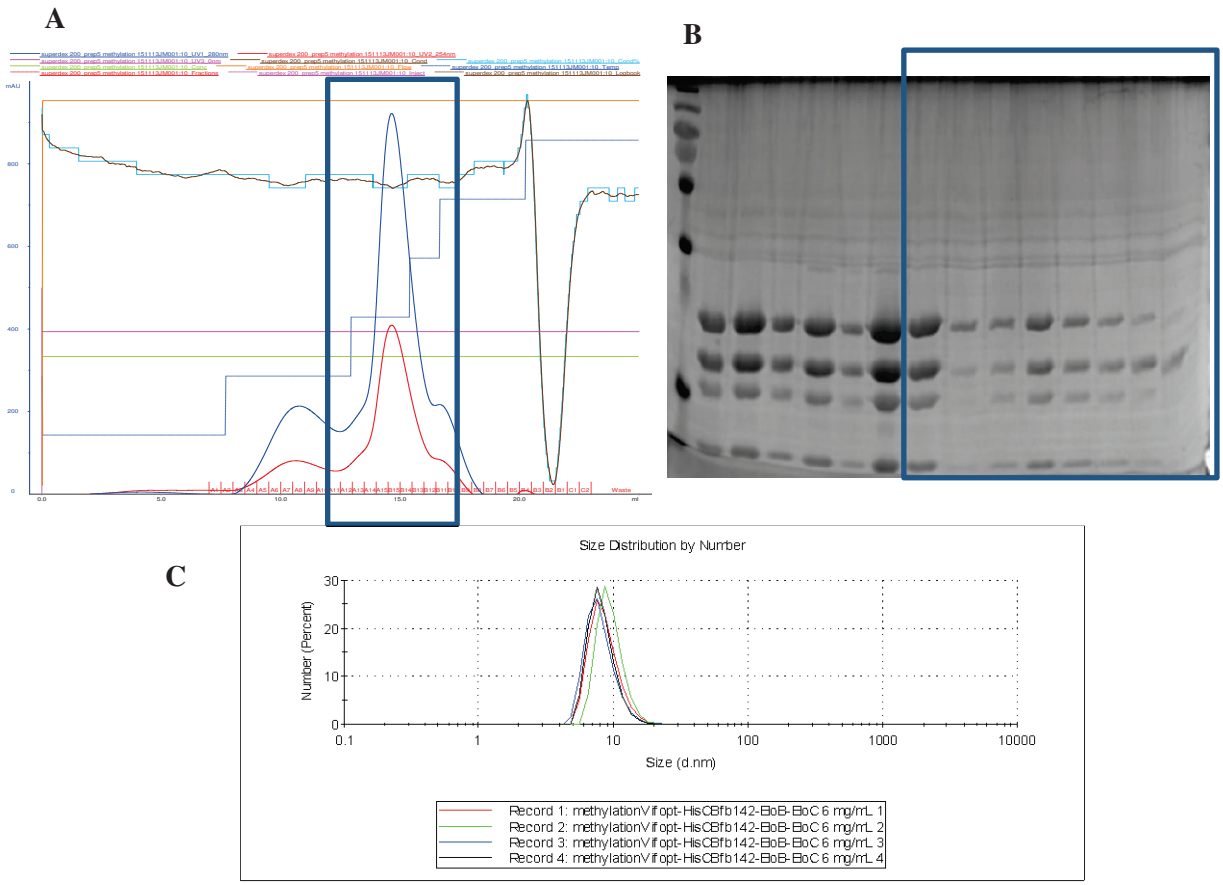


Figure 95 : Purification du complexe Vif_{HXB2}-CBFβ142-EloB-EloC par tamis moléculaire.A : chromatogramme représentant l'évolution de la DO en fonction du volume d'élution. Le premier pic représente les agrégats du au protocole expérimental de méthylation. Le deuxième pic correspond au complexe méthylé B : Gel SDS-PAGE montrant les différentes fractions correspondantes aux deux pics. On collecte seulement les fractions du deuxième pic. C : Analyse du complexe méthylé par diffusion dynamique de la lumière (DLS). Les 4 courbes correspondent à quatre mesures indépendantes. La protéine est monodisperse sans agrégats.

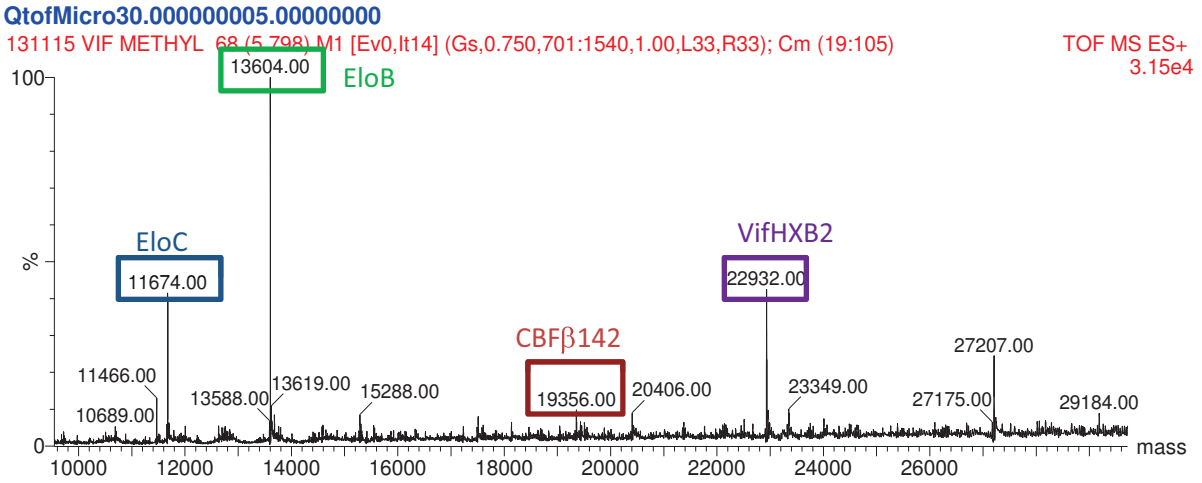


Figure 96 : Spectre d'analyse LC-MS/MS du complexe Vif_{HXB2}-CBFβ142-EloB-EloC en conditions dénaturantes. On retrouve la taille exacte attendue pour chacune des protéines ayant toutes ses lysines méthylées.

IV. Discussion, conclusions et perspectives

Dans cette deuxième partie de ma thèse, le travail a porté sur la production d'APOBEC seul ainsi qu'au sein du complexe E3 ubiquitine ligase formé de Vif-CBFβ-EloB-EloC-Cul5-Rbx2 et ceci dans le système procaryotique (plus précisément la bactérie *E. coli).*

Auparavant, le seul moyen de produire la protéine Vif était dans des conditions dénaturantes (6M Chlorure de guanidinium). Vif était ensuite renaturée par dialyses successives contre un tampon dépourvu d'agent dénaturant. Cette protéine était par la suite, utilisée pour des expériences biochimiques et structurales. L'inconvénient de cette méthode d'expression de Vif est que la protéine multimérise en solution par son domaine PPLP [230] et son domaine de liaison au zinc [236] ce qui rend sa cristallisation très difficile, voire impossible.

La découverte du facteur CBFβ en 2012 a orienté le travail sur Vif seule vers une autre direction. CBFβ est capable de stabiliser Vif et de la solubiliser. Dorénavant, on ne parle plus de Vif seule mais du complexe Vif-CBFβ-EloB-EloC, qu'on a essayé de surproduire au laboratoire et d'y rajouter la protéine APOBEC3G/F. Dans cet objectif, nous avons appliqué la technique de co-expression développée par Dr. Christophe Romier à l'IGBMC pour co-exprimer l'ensemble des protéines formant le complexe.

Concernant la partie APOBEC, j'ai réussi à exprimer et produire en bonne quantité le domaine C-terminal des protéines A3G et A3F et caractérisé leurs activités désaminases. J'ai pu aussi démonter une interaction potentielle entre A3G C-terminal et Vif, mais cette hypothèse reste encore à vérifier. J'ai aussi réussi à produire le domaine N-terminal d'A3G/F responsable de l'interaction avec Vif dans des cellules *d'E. coli* (Arctic Express). Le problème rencontré a été la présence inévitable des protéines chaperonnes Cpn60 que nous n'avons pas pu séparer d'APOBEC malgré les nombreuses tentatives. Les domaines complets des protéines APOBEC sont demeurés insolubles dans *E. coli,* même après l'optimisation ou l'harmonisation des codons.

Cependant, il est important de signaler que la structure du C-terminal d'A3F a pu être résolue par cristallographie après avoir produit une protéine chimère soluble avec A3F. La détermination des domaines à remplacer a été identifié après alignement des séquences et des analyses bio-informatiques [132, 175]. La même stratégie a été appliquée récemment pour la résolution de la structure du N-terminal d'A3G ; un variant soluble du NTD d'A3G, identique à 80% à la protéine Wild-type a été construit pour cet objectif et cela en appliquant le principe de conservation de séquences et la mutagenèse tout en conservant l'activité de la protéine [133]. Il serait très intéressant d'essayer d'appliquer la même stratégie sur le domaine N-terminal d'A3G et les domaines complets d'A3G et

A3F tout en conservant la structure et la fonctionnalité de l'enzyme, ces protéines présentant une très forte homologie de séquences.

Concernant la protéine Vif, tous les tests de co-expressions effectués entre Vif et CBFβ n'ont pas permis d'obtenir un complexe stable et soluble dans les conditions testées. Ceci n'était pas étonnant étant donné qu'il a par la suite été montré que Vif a besoin aussi des Elongines B et C pour une meilleure stabilité. Co-exprimer le complexe Vif-CBFβ-EloB-EloC en bonne quantité et qualité a nécessité de nombreuses optimisations. Les meilleures conditions qui ont permis finalement d'obtenir ce complexe en grande quantité pour les tests de cristallisation sont une induction à faible température (15°C) et un tampon contenant une concentration élevée en sel (500 mM NaCl). Ceci n'est pas surprenant au vu de la structure cristallographique qui a été obtenue [163]. En effet, celle-ci montre l'importance des interactions hydrophobes entre les protéines ; une forte concentration saline permet donc de stabiliser le complexe.

Malgré les nombreux tests de cristallisation effectués et les différentes conditions et concentrations protéiques testées, aucun cristal n'a été formé pour les deux complexes Vif-CBFβ-EloB-EloC et Vif-CBFβ-EloB-EloC-Cul5-Rbx2. Deux isolats de Vif HXB2 et pNL4-3 ont été testés ainsi que deux versions de CBFβ , complet et tronqué en C-terminal vu la dégradation causée par l'extrémité flexible de ce domaine.

La cristallisation du complexe Vif-CBFβ-EloB-EloC-Cul5-Rbx2 par l'équipe de Huang en 2014 a été réalisée après avoir obtenu le complexe à partir de conditions très similaires à celles que nous avons optimisées, mais après avoir réalisé une protéolyse ménagée qui a entrainé la perte du domaine C-terminal de la Culline5, qui était donc très labile et empêchait nos essais de cristallisation.

Dans la deuxième partie, je me suis intéressée à l'obtention du complexe Vif-CBFβ-EloB-EloC-APOBEC. Dans un premier temps, je n'ai pas inclus la Culline5 dans les co-expressions, sa présence n'est pas primordiale à l'interaction entre Vif et APOBEC. Dans nos résultats obtenus, A3G ainsi que A3F complet n'était pas solubles et demeuraient dans le culot dans plusieurs conditions testées. Vif n'était pas capable de solubiliser APOBEC ; une hypothèse probable expliquant ce résultat est que les résidus conservés de Vif important pour l'interaction avec A3G/F ne sont pas exposés et donc sont regroupés dans la structure de Vif ou bien à l'interface avec CBFβ suggérant une superposition dans les domaines de Vif responsables de l'interaction avec CBFβ et A3. Il a été montré que l'interaction d'A3F avec le complexe Vif-CBFβ-EloB-EloC déplace CBFβ appuyant l'idée que la liaison de CBF et A3F est mutuellement exclusive [162]. Des études contredisant cette théorie ont montré que l'interaction de CBFβ à Vif n'empêche pas l'interaction avec A3G, suggérant que ces deux interactions ne sont pas mutuellement exclusives [165]. Ces hypothèses restent à confirmer par la

détermination de la structure tridimensionnelle du complexe Vif/E3 ubiquitine ligase en interaction avec A3G/F. Une autre hypothèse potentielle serait que les protéines APOBEC ne sont pas surexprimées dans *E. coli* et donc la co-expression n'est pas stœchiométrique entre toutes les protéines du complexe ce qui empêche la solubilisation d'APOBEC. Il serait pour cela intéressant d'essayer d'exprimer APOBEC3G/F seule ou au sein de ce complexe dans un autre organisme plus proche de l'organisme humain, hôte principal des protéines APOBEC3G. Dans notre cas, nous avons décidé de partir sur une collaboration avec l'équipe de Marc Ruff à l'IGBMC et produire ces protéines dans un système mammifère utilisant des cellules de hamster.

RESULTATS PARTIE 3 : RECONSTITUTION DU COMPLEXE VIF/APOBEC3G EN CELLULES DE MAMMIFERES

I. Introduction et objectif

L'insolubilité d'APOBEC3G/F dans le système procaryote notamment Escherichia Coli, et ce malgré toutes les tentatives d'optimisation des conditions d'expression, nous a poussé a essayé d'autres systèmes d'expression pour pouvoir obtenir le complexe Vif/CBFβ/EloB/EloC/A3G/F soluble. Le système d'expression utilisant les cellules de mammifères pour la production de protéines recombinantes peut permettre un repliement correct des protéines, l'introduction des modifications post-traductionnelles essentielles ou un bon assemblage du complexe.

La production de protéines recombinantes en cellules mammifères est requise pour de nombreuses applications scientifiques et commerciales et ce besoin a abouti au développement de plusieurs systèmes d'expression. Un outil très connu pour la production des protéines en cellules eucaryotes est basé sur une souche recombinante du virus de la vaccine (Western Reserve, WR) codant pour l'ARN polymérase du bactériophage T7 [237]. La production de protéines recombinantes en cellules de mammifères n'est pas une stratégie simple à mettre en œuvre compte tenu de son fort besoin en main d'œuvre, le temps necessaire pour sa mise en place que son coût élevé. Cependant, le virus de la vaccine atténué (MVA) possède la capacité de transférer un ADN plasmidique exogène dans les cellules mammifères durant le processus d'infection ce qui rend la production des protéines eucaryotiques simple, rapide et peu coûteuse, rendant cette approche intéressante.

Ce travail a été effectué en collaboration avec l'équipe dirigée par Marc Ruff à l'institut de génétique et biologie moléculaire et cellulaire (IGBMC) à Strasbourg. Cette équipe utilise le virus de la vaccine pour infecter les cellules mammifères BHK21 suivant une stratégie détaillée dans la partie « Matériels et Méthodes » [237, 238]. L'équipe de Marc Ruff a réussi à caractériser l'intégrase du VIH-1 seule ainsi qu'en complexe avec la protéine humaine LEDGF à partir de ce système [239]. L'objectif de cette stratégie est la production des protéines APOBEC3G/F seules et/ou en complexe avec Vif/CBF/EloB/EloC en grande quantité pour des études de cristallographie ou RMN ainsi que la cryo-microscopie électronique.

II. Stratégie

Dans le but d'exprimer ces protéines en cellules de mammifères infectées par le virus de la vaccine, le travail a été séparé en quatre grands axes de travail :

A- Clonages des gènes d'intérêt dans les vecteurs d'expression

B- Purification et amplification des virus recombinant

C- Expression et purification des protéines d'intérêt

D- Caractérisation des protéines d'intérêt et test d'activité

La première phase de travail consistait à produire des vecteurs d'expression adaptés à la stratégie de la vaccine et donc à insérer chacune des protéines APOBEC3G et APOBEC3F dans un vecteur de destination pVoteOGW. Ce vecteur possède les éléments essentiels pour l'expression avec le virus de la vaccine. En plus des protéines A3G/F, nous avons construit la polyprotéine CFBb-Vif-EloB-EloC-A3G grâce à la technique BioBrick, pour transférer ce plasmide dans le vecteur d'expression pVoteOGW exprimant en N-terminal la protéase TEV et en C-terminal le marqueur de sélection fluorescent YFP, ainsi qu'un promoteur T7 inductible à l'IPTG. Ces étapes de clonages ont nécessité plusieurs sous clonages pour chaque gène avant de pouvoir le transférer dans le pDest final par un dernier clonage Gateway. Cette phase de travail a été effectuée en cinq mois.

Une fois les vecteurs d'expression obtenus on passe à la deuxième étape. Il s'agit maintenant de transfecter les cellules BHK21 par ces vecteurs et les infecter par le virus de la vaccine. Ensuite des étapes de purification sont importantes pour purifier des virus recombinants ayant intégré notre insert par recombinaison homologue. Finalement, les virus isolés sont amplifiés par infection de nouvelles cellules pour avoir un stock de virus suffisant pour l'expression. Ce travail a été réalisé en deux mois.

L'étape suivante est d'infecter un grand nombre de cellules BHK21 adhérentes ou en suspension par le virus purifié et amplifié portant nos gènes codant pour les protéines A3G/F ou la polyprotéine CBFβ/Vif/EloB/ELoC/A3G puis induire par l'IPTG. Les protéines seront exprimées puis purifiées sur des colonnes d'affinité (Nickel et Streptactine) puis identifiées par Western Blot et spectrométrie de masse ; un test d'activité sera réalisé pour évaluer leur activité *in vitro*. Les dernières phases de production des protéines et leur caractérisation a nécessité quatre mois de travail.

III. Résultats

A. Clonage des gènes d'intérêt dans les vecteurs d'expression

1. Complexe CBFβ – Vif – EloB – EloC – A3G

Les séquences codantes des gènes ont été insérées les unes derrière les autres sous la dépendance d'un unique codon d'initiation pour former un ARNm polycistronique. Ainsi les protéines sont exprimées sous la forme d'une polyprotéine. Les gènes ont été insérés dans un même plasmide, selon l'ordre CBFβ – Vif – EloB – EloC – A3G, grâce à la technique du clonage BioBrick (Fig. 39). Par la suite, le système de recombinaison *in vitro*, Gateway, fondé sur l'utilisation de protéines du bactériophage lambda et d'*E. coli* a été employé pour recombiner les plasmides entre eux en évitant l'utilisation des enzymes de restriction. L'insert est donc introduit dans un vecteur d'entrée Gateway entre un gène codant pour la protéase TEV et un gène codant pour la protéine eYFP. L'insert, contenant nos gènes d'intérêt flanqués des gènes TEV et eYFP, a finalement été transféré dans un vecteur de destination, pVote0GW (Fig. 97) Ce vecteur possède les éléments essentiels pour l'expression avec le virus de la vaccine, et notamment des régions homologues à un locus du génome viral ce qui permettra l'intégration de notre insert. L'expression de l'ARN polymérase T7 par le

génome viral, inductible à l'IPTG, permet la transcription des gènes qui sont sous la dépendance du promoteur T7.

Dans cette construction, deux étiquettes de purification (un « Flag10xHis-tag » et un « Twin strep-tag») ont été insérées en amont de CBFβ et d'A3G respectivement pour permettre une meilleur purification du complexe. Un site de clivage à la protéase TEV est inséré entre les différents gènes pour permettre l'individualisation des protéines par la protéase TEV exprimée en amont et permettre ainsi une reconstitution *in vivo* du complexe.

Les mêmes clonages ont été effectués pour la protéine APOBEC3F seule (His-tag/Strep-tag) ou en polyprotéine CBFβ – Vif – EloB – EloC – A3F, mais je n'ai pas eu le temps de produire la protéine en cellules de mammifères, donc tous les résultats que je vais montrer dans cette partie concernent la protéine APOBEC3G.

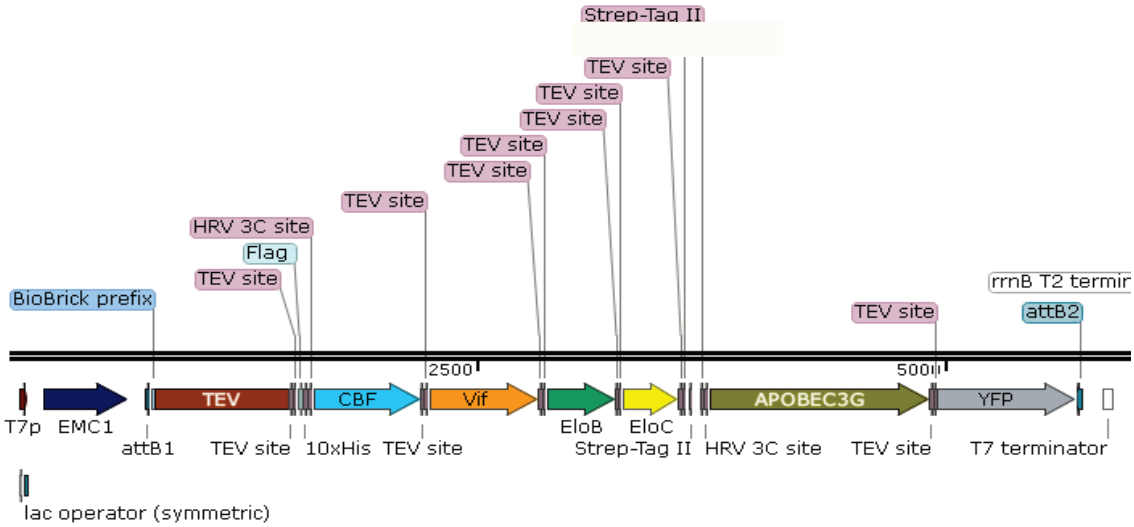


Figure 97 : L'insert TEV – FH CBFβ – Vif – EloB – EloC –Strep A3G – YFP est représenté dans le vecteur pVote0GW pour définir le vecteur d'expression final. Le promoteur T7 (T7p) est en amont d'une séquence leader EMC1 utile pour la transcription du transgène. Les sites de clivage à la TEV et le site de clivage HRV 3C sont représentés en rose (TEV site) ; la séquence BioBrick est représentée en bleu ; les sites *attB1* et *attB2* résultant de la réaction LR sont également indiqués. L'expression du transgène est inductible à l'IPTG grâce à la présence de l'opérateur LacO.

2. La protéine APOBEC3G

Le même clonage a été effectué pour APOBEC3G seule. Nous avons donc deux constructions d'APOBEC3G : Flag (10xHis)-A3G et Strep-tag-A3G.

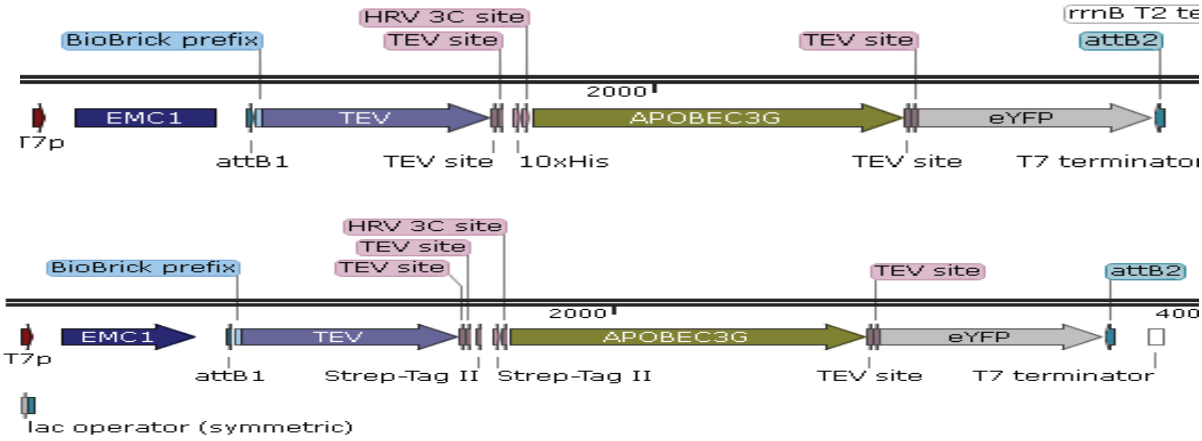


Figure 98 : L'insert TEV-APOBEC3G-YFP a été représenté dans le vecteur pVote0GW avec une étiquette Flag(10XHis) en N-terminal d'APOBEC3G (a) et une étiquette Strep en (b).

B. Sélection des virus recombinants

La souche de virus utilisée, *MVA-Koom*, élaborée par l'équipe de Marc Ruff, a été optimisée pour les systèmes d'expression et de sélection utilisés, notamment par l'insertion dans le génome viral des fluorophores GFP et mCherry. Des cellules BHK21 infectées par le virus sauvage ont été transfectées par les plasmides d'expression contenant nos gènes d'intérêt et ces derniers ont été transférés au génome viral par recombinaison homologue (Fig. 54). Dans ces cellules on retrouve deux types de virus : des virus sauvages et des virus, dits recombinants, qui auront intégré notre insert. A raison d'environ un virus recombinant pour mille sauvages, il a été indispensable d'enrichir les suspensions en virus recombinants grâce à des marqueurs présents dans la séquence transférée ou dans le génome viral. La sélection de ces virus s'est faite en deux temps : une sélection positive à l'acide mycophénolique grâce au gène marqueur GPT [240] et une sélection négative à la coumermycine grâce au gène de fusion Gyr-PKR [225]. Ces méthodes de sélection ont été décrits dans la partie « Matériels et Méthodes ». Cette sélection engendre la disparition de la fluorescence rouge de la mCherry, située au niveau du site de recombinaison dans le génome viral. Cette fluorescence rouge, spécifique des virus sauvages, permet de les discriminer des recombinants qui ne fluorescent qu'en vert grâce à l'expression constitutive de la GFP.

Pour chaque étape de sélection, la fluorescence rouge et verte des plages de lyse a été observée au microscope à epifluorescence (Fig. 99) pour des suspensions de virus dilués. Les deux étapes de sélection ont été réalisées successivement jusqu'à ce que les suspensions de virus ne présentent plus, ou une quantité très négligeable, de plage de lyse rouge. En général, une sélection à l'acide mycophénolique en présence de xanthine et d'hypoxanthine et deux étapes de sélection à la coumermycine sont suffisantes pour obtenir une suspension de virus recombinants assez pure. Par la suite, les virus isolés ont été amplifiés, par infection de nouvelles cellules, de manière à avoir un stock de virus suffisant pour l'expression. Pendant cette amplification de l'acide mycophénolique peut être ajouté au milieu pour éliminer les quelques virus sauvages résiduels.

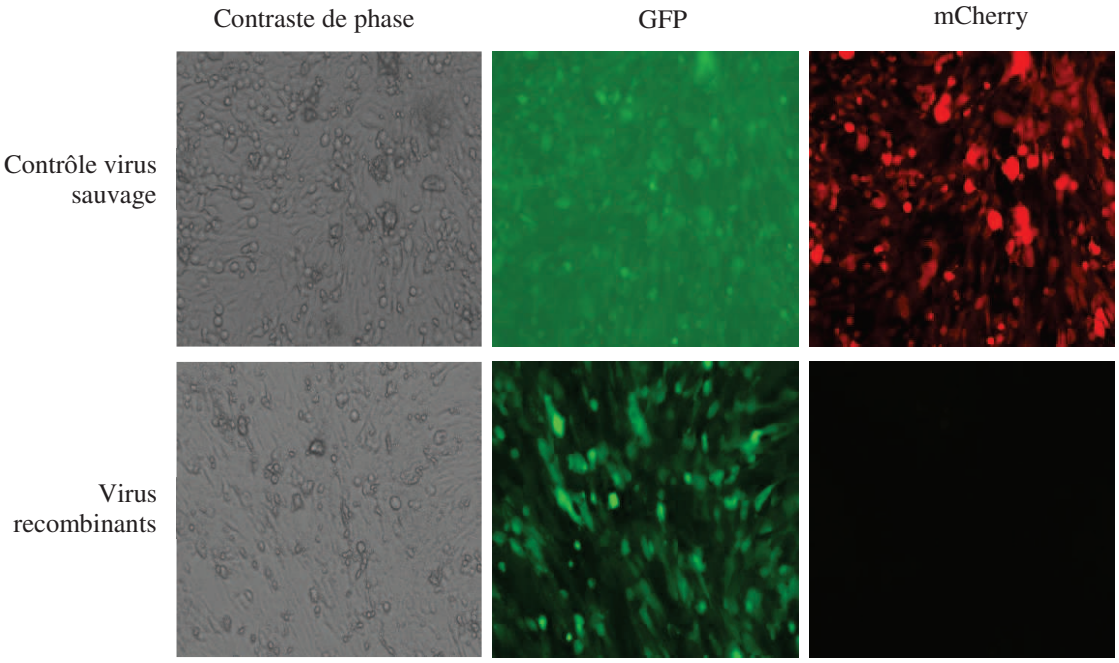


Figure 99 : Caractéristiques des virus sauvages et recombinants.
Des cellules BHK21 adhérentes, infectées par un virus sauvage MVA-Koom en contrôle (ligne du dessus) et par un virus recombinant (ligne du dessous) sont observées au microscope à épifluorescence (objectif x20). Les cellules infectées sont visualisées en contraste de phase (1ère colonne) ; la fluorescence verte, rouge, jaune (2ème, 3ème et 4ème colonne) est visualisée à l'aide des lasers. Les virus sauvages qui n'ont pas subi l'évènement de recombinaison, conservent la fluorescence rouge (ligne du dessus mCherry) alors que les virus recombinants ont perdu la fluorescence rouge (ligne du dessous mCherry). Les deux types de virus conservent la fluorescence de la GFP (2ème colonne).

C. Expression des protéines

Après isolement des virus recombinants sur les milieux sélectifs plusieurs fois pour avoir le moins de virus sauvage possible et le plus de virus recombinants et après amplification de ces virus, il était indispensable de réaliser des tests d'expression sur un petit volume de culture de cellule BHK21 adhérentes afin d'évaluer la solubilité des protéines dans les différentes constructions avant de commencer les expressions à grande échelle.

1. Marqueur d'expression YFP

Les virus recombinants ont été utilisés pour infecter de nouvelles cellules BHK21, et l'ARN polymérase T7, exprimée par le génome viral, a permis la production d'un ARNm polycistronique puis la synthèse des protéines recombinantes (Fig. 53).

Les cellules infectées présentant une fluorescence jaune (Fig. 100), on peut supposer que les protéines codées en amont de la eYFP ont dû être exprimées. Les cellules infectées fluorescent bien en vert grâce à la GFP exprimée par le virus. Dans le cas de la polyprotéine, la fluorescence jaune représente aussi un indicateur de la bonne expression de toutes les protéines.

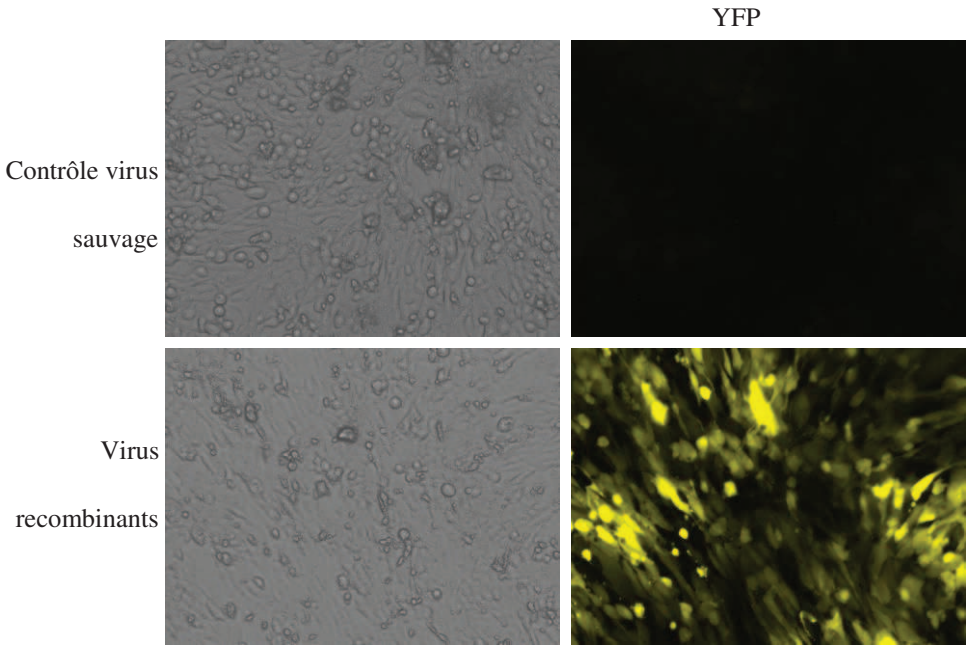


Figure 100 : La visualisation de la fluorescence jaune de la protéine YFP. L'expression de la YFP, présente en N-terminal de la polyprotéine et induite par l'ajout d'IPTG, est visible uniquement pour les virus recombinant, synonyme d'une expression effective de nos protéines.

2. Tests d'expression et purification

Après avoir induit l'expression à l'IPTG, lysé les cellules et vérifié leur expression à partir du marqueur de fluorescence YFP, les protéines ont été purifiées en batch sur des billes de résine d'affinité (Nickel et Streptactin) compte tenu de la petite quantité de protéines produites. Les différents échantillons seront déposés sur un gel SDS PAGE 10 ou 15 % et un test de western blot préliminaire permettra de détecter la présence des protéines. Dans les tests d'expression montrés ici, A3G a été purifiée à 150 mM NaCl sur des billes de Nickel ou Streptactine alors que le complexe FHCBFβ-Vif-EloB-EloC-RRPA3G a été purifié à 500 mM NaCl sur des billes de Streptactine en ciblant le Twin-Strep (RRP) dans le N-terminal de la protéine A3G. Dans ces expériences, on a utilisé comme contrôle des cellules BHK21 non infectées.

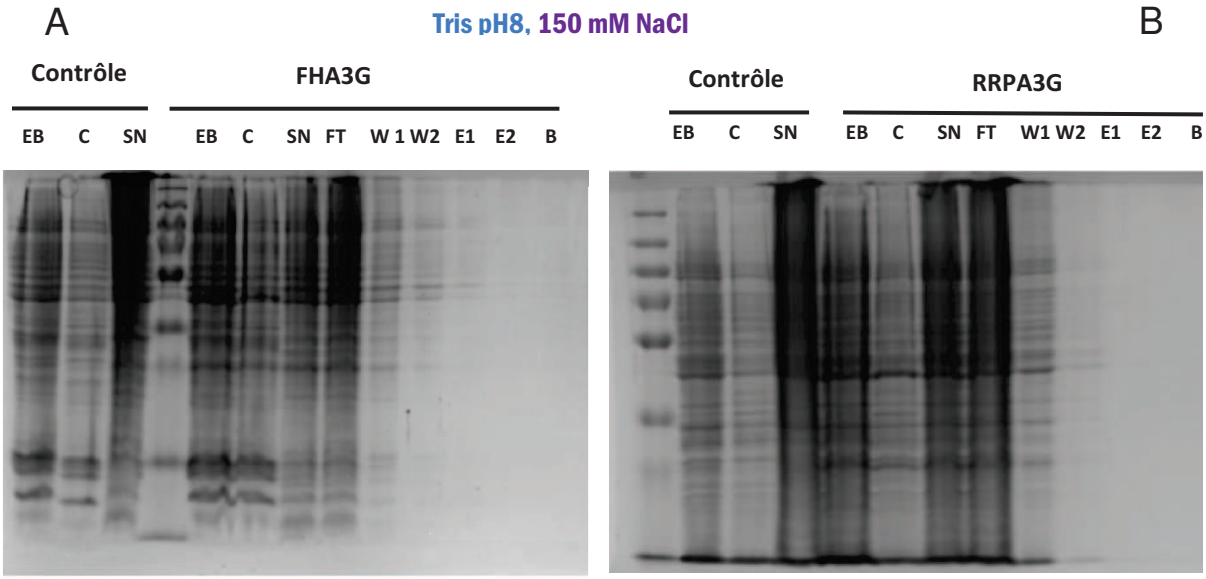


Figure 101 : Test d'expression et purification d'A3G.
Les cellules BHK21, infectées par les virus recombinants exprimant (après induction IPTG) les protéines recombinantes Flag/10His A3G (A) et RRPA3G (B), ont été récoltées et lysées. Les extraits bruts purifiés sur des billes de Nickel/Streptactin et les protéines ont été éluées par 400 mM d'imidazole/2.5 mM Desthiobiotine. Les différents échantillons ont ensuite été migrés sur gel de polyacrylamide dénaturant 15 % puis les protéines ont été révélées au bleu de Coomassie.
EB : extrait brut ; C : culot ; SN : surnageant ; FT : flow through ; W1 : lavage; W2 : lavage excessif ; E1 : élution ; E2: élution E concentrée ;B : billes de résine.

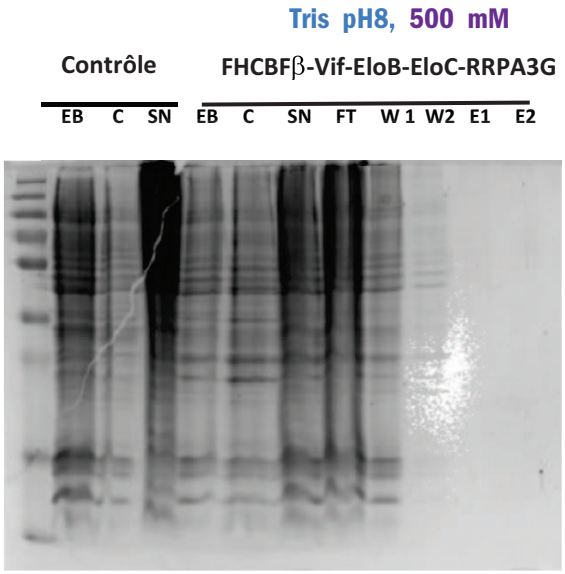


Figure 102 : Test d'expression et purification du complexe FHCBFβ-Vif-EloB-EloC-RRPA3G. Le complexe a été exprimé dans les cellules BHK21 et purifié sur des billes de Streptactine. Les échantillons ont été déposés sur un gel polyacrylamide 15% selon le même ordre qu'A3G seule.

Les échantillons déposés sur les gels SDS PAGE correspondent à des échantillons collectés tout au long de la purification : de l'extrait brut, au culot et surnageant ainsi qu'aux lavages et élutions des billes de nickel ou Streptactine. Le but est d'évaluer d'abord l'expression des protéines en les identifiant dans l'extrait brut, puis d'évaluer leur solubilité en comparant les profils « C » et « SN ». Finalement, la fraction soluble se fixera par affinité sur la résine ; en déposant les éluats sur gels on déterminera la solubilité (donc le probable bon repliement de la protéine) ainsi que sa pureté. Dans notre cas nous cherchons surtout à identifier dans les éluats les différentes protéines du complexe.

Les résultats des tests d'expression réalisés sur les différentes constructions étaient similaires et ressemblent aux résultats obtenus avec le contrôle. On ne remarque pas de surexpression dans aucune des pistes dans ces conditions. Toutes les bandes observées dans les pistes de purification des protéines d'intérêts se retrouvent aussi dans le contrôle et donc correspondent à des protéines endogènes des cellules BHK21.

Après discussion avec Robert Drillien, notre collaborateur dans ce projet, il n'est pas très surprenant de ne pas voir des bandes bien précises sur les gels dans les pistes « EB », « C » et « S » ; nos protéines pourraient être noyées parmi les autres protéines de BHK21. Pour vérifier la présence des protéines même en faible quantité et afin d'éliminer la probabilité de problèmes provenant des étapes en amont (Clonage, recombinaison, amplification des virus, …), nous avons réalisé un Western Blot. Celui-ci (Fig. 103) a été réalisé à partir de tests d'expression pour détecter la protéine Vif dans le complexe CBFβ-Vif-EloB-EloC-A3G grâce à un anticorps anti-Vif disponible au laboratoire. Nous avons également détecté la protéine A3G dans la construction Flag/10His A3G avec un anticorps anti-

Flag également disponible au laboratoire. Vif a effectivement été identifiée dans le complexe à la taille attendue, environ 23 KDa (Fig. 103A), tout comme A3G, à la taille d'environ 50 KDa pour la construction A3G seule (Fig. 103B). Nous avons aussi pu identifier la protéine Vif dans l'élution pour le complexe (Fig. 103A piste E). En effet, les échantillons utilisés pour ce Western blot sont issus d'une purification du complexe qui ciblait l'étiquette d'A3G ; donc si Vif a été récupérée dans l'élution cela indique qu'A3G est bien soluble et présente. A noter cependant que pour l'expression d'A3G seule, aucune bande n'a été détectée dans l'élution (Fig. 103B piste E), certainement du fait d'une quantité de cellules de départ trop faible. Vif semble bien interagir avec A3G dans notre complexe. Ceci doit cependant être validé par l'identification de chaque protéine du complexe dans l'élution après purification à partir d'une quantité plus importante de protéines. Ce Western blot a été réalisé sans contrôle de charge interne, qui aurait permis de s'assurer de la quantité équivalente de protéines déposées dans chaque puits, puisque nous nous sommes intéressés à l'aspect qualitatif et non quantitatif de l'expérience.

Figure 103 : Test d'expression des protéines recombinantes Vif et A3G.
Un Western blot pour détecter la protéine Vif dans le complexe (A) et la protéine A3G pour la construction A3G seule (B) a été réalisé. Les échantillons issus de la purification effectuée à partir d'un test d'expression ont été migrés sur gel de polyacrylamide dénaturant 15 % puis transférés sur une membrane de nitrocellulose. La protéine Vif (~23 KDa) a été détectée à l'aide d'un anticorps anti-Vif et A3G (~50 KDa) à l'aide d'un anticorps anti-Flag. EB : extrait brut ; C : culot ; SN : surnageant ; FT : flow through ; E : elution

D. Expression et purification à grande échelle

Après l'expérience de Western blot, nous avons pu vérifier que le système de clivage à la protéase TEV pour un complexe de si grande taille fonctionne bien. Nous avons de plus démontré que les problèmes de surexpression ne proviennent pas des clonages ou des virus recombinants puisque les protéines ont été identifiées grâce à leurs anticorps spécifiques. L'étape suivante était de produire les protéines sur une large échelle pour augmenter les rendements et optimiser les tampons de purification.

1. Tampons

Pour les purifications à grande échelle, sur des volumes de 4, 6 et 10 litres de culture de cellules BHK21 en suspension, j'ai essayé plusieurs tampons pour trouver les plus convenables pour les purifications d'APOBEC3G seule et du complexe CBFβ-Vif-EloB-EloC-A3G. Je me suis inspirée des conditions établies pour le complexe CBFβ – Vif – EloB - EloC exprimée avec *E. coli* (1ère ligne du tableau 20).

Quelques modifications aux tampons utilisés pour la purification dans *E. coli* ont été réalisées dans ce système, dont l'addition d'un détergent qui permet de mieux solubiliser A3G. Le CHAPS est un détergent zwitterionique non dénaturant mais il ne pourrait pas être rajouté au tampon de lyse du complexe car cela risquerait de dissocier ce dernier. Par contre, 7 mM de CHAPS a été rajouté à tous les tampons de purification d'A3G seule.

Les tampons utilisés par la suite et durant la lyse ainsi que la purification d'A3G et du complexe sont représentés dans le tableau 20.

	Complexe A3G	A3G
Test d'expression et purification Nickel	50 mM Tris HCl pH8 0,2 mM β-mercaptoéthanol 10 % glycérol 0,1 mM EDTA	
	500 mM NaCl	150 mM NaCl 7mM CHAPS
Test d'expression sur billes de streptavidine	50 mM Tris HCl pH8 0,2 mM β-mercaptoéthanol	
	500 mM NaCl	150 mM NaCl
Purification sur colonne de streptavidine	50 mM Tris HCl pH8 0,2 mM β-mercaptoéthanol 150 mM NaCl 10 µM $ZnCl_2$ 10 µM $MgCl_2$	
		9 mM CHAPS

Tableau 20: Composition des tampons utilisés pour la purification des protéines produites en cellules BHK21.

2. Purification du complexe CBFβ – Vif – EloB – EloC – A3G

a. Colonne d'affinité au Nickel

La purification a été réalisée pour le complexe Flag/10HisCBFβ – Vif – EloB – EloC – StrepA3G en utilisant l'étiquette histidine en N-ter de CBFβ. Les 4 litres de culture cellulaire ont été récupérés et les cellules lysées puis centrifugées avant injection sur la colonne de chromatographie. En comparant les profils de migration (Fig. 104B) de l'extrait brut, du culot et du surnageant (Fig. 104B pistes EB, C et SN), on observe les mêmes profils. Les protéines d'intérêt sont noyées dans les protéines totales. La purification a été réalisée sur une colonne Nickel avec des lavages à 40 mM Imidazole ; Les protéines sont éluées à 400 mM imidazole. Après purification, des bandes pouvant

correspondre aux tailles des protéines du complexe ont pu être détectées (Fig. 104B B1 à B5). Les résultats de spectrométrie de masse ont bien permis d'identifier sans équivoque la protéine CBFβ pour B1, Vif pour B2 et EloC pour B4. Malheureusement, A3G et EloB n'ont pas pu être identifiées. De plus, il faut également remarquer que la majorité de nos protéines d'intérêt sont déjà décrochées de la colonne à faible concentration d'imidazole dans le lavage et que l'élution est très contaminée. Enfin, même si l'on a récupéré une partie du complexe, nous n'avons pas pu identifier A3G.

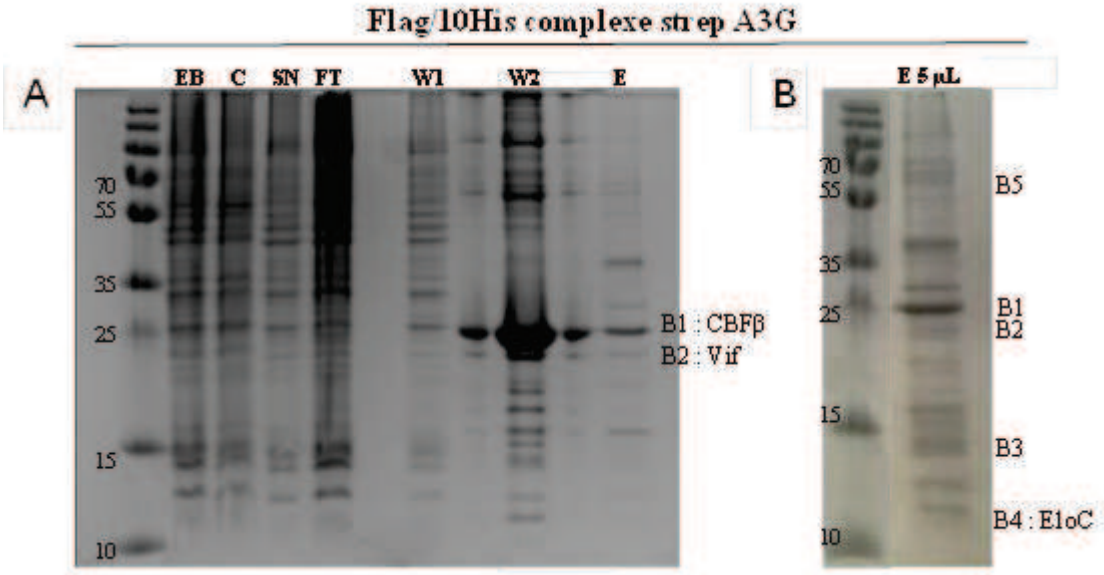


Figure 104 : Purification du complexe-A3G sur colonne de Nickel . Les cellules BHK21, infectées par les virus recombinants exprimant (après induction IPTG) les protéines recombinantes formant le complexe Flag/10His CBFβ – Vif – EloB – EloC – strep A3G, ont été récoltées et lysées. (A) Les extraits bruts ont été injectés sur une colonne de Nickel de 1 mL et les protéines ont été éluées par 400 mM d'imidazole. Les différents échantillons ont ensuite été migrés sur gel de polyacrylamide dénaturant 15 % puis les protéines ont été révélées au bleu de Coomassie. (B) l'éluat a été concentré pour une meilleur identification des protéines. Les bandes B1 à B5 pouvant correspondre à nos protéines d'intérêts ont ensuite été analysées par spectrométrie de masse. AI : avant induction ; PI : après induction ; EB : extrait brut ; C : culot ; SN : surnageant ; FT : flow through ; W et W1 : lavage à 40 mM imidazole ; W2 : lavage à 133 mM imidazole ; E : élution à 400 mM imidazole ; E 5 µL : élution E concentrée.

b. Colonne d'affinité Streptactin®

Par la suite, dans le but de savoir si A3G est bien éluée avec tout le complexe et pour essayer de la récupérer en plus grande quantité, nous avons changé la stratégie de purification en ciblant A3G, qui porte l'étiquette Twin-Strep. De plus, la composition des tampons a été adaptée (4ème ligne du Tableau 20), notamment en ajoutant du zinc et en diminuant la concentration en sels. L'élution concentrée (Fig. 105B piste Complexe A3G), déposée sur gel n'est pas pure. Cependant, une analyse par spectrométrie de masse de la bande intense entre 35 et 55 KDa a permis d'identifier A3G (Fig. 110A). Mais elle n'est pas la seule protéine retrouvée dans cette bande qui contient également de l'actine (Fig. 110B). L'intensité de la bande ne traduit donc pas la quantité d'A3G réelle récupérée. Contrairement aux résultats obtenus après la première purification, nous avons réussi à identifier sur gel la protéine A3G. Des analyses de spectrométrie de masse ont également permis d'identifier CBFβ et Vif (Fig. 111A et B). EloB et EloC n'ont pas pu être identifiées dans ce cas. Néanmoins, ce résultat montre que nous avons réussi à exprimer A3G de manière soluble avec le complexe.

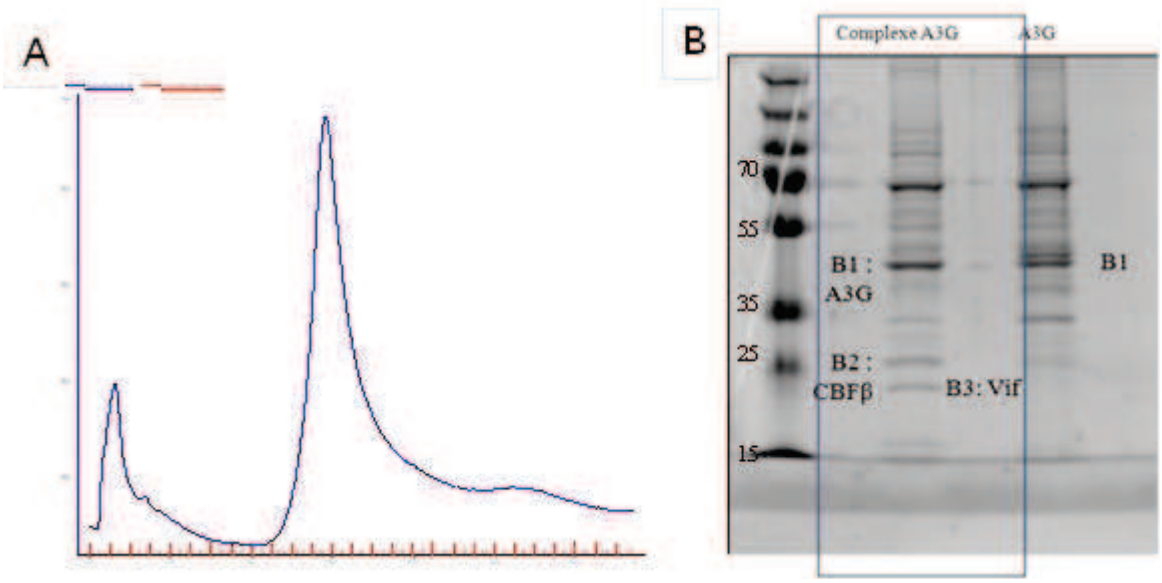


Figure 105 : Purification du complexe– A3G sur colonne de streptavidine.Les cellules BHK21, infectées par les virus recombinants exprimant (après induction IPTG) les protéines recombinantes formant le complexe Flag/10His CBFβ – Vif – EloB – EloC – strep A3G, ont été récoltées et lysées. Les extraits bruts ont été injectés sur une colonne de streptavidine StrepTactin® de 1 mL et les protéines ont été éluées par 2,5 mM de desthiobiotine. Les profils d'élutions sont présentés pour A3G avec le complexe (A) en donnant l'absorbance à 280 nm en fonction du temps. Les élutions concentrées ont ensuite été migrées sur gel de polyacrylamide dénaturant 15 % puis les protéines ont été révélées au bleu de Coomassie (B). Les bandes (B1 à B3) pouvant correspondre à nos protéines d'intérêts ont ensuite été analysées par spectrométrie de masse.

3. Purification d'A3G seule

a. Purification de Flag/10His A3G

La purification d'A3G a d'abord été réalisée pour la construction Flag/10His A3G à partir d'une culture cellulaire de 1L. Le chromatogramme nous a montré que la quantité de protéines au départ n'est pas encore suffisante car aucun pic d'élution n'a été détecté. De la même manière que pour A3G avec le complexe, les profils de migration (Fig. 106) des échantillons extrait brut, culot et surnageant (Fig. pistes EB, C et SN) présentent les mêmes bandes de protéines. Aucune bande ne se détache des autres y compris pour les pistes des extraits avant et après induction (Fig. 106 pistes AI et PI). Après induction, aucune surexpression des protéines n'est donc visible. Dans l'élution (Fig.106 piste E) on peut voir parmi les nombreuses bandes contaminantes une bande plus intense vers 15 KDa (Fig. 106 B1). Malheureusement, il ne s'agit pas d'A3G dégradée comme l'on aurait pu l'espérer mais de BSA présente dans le milieu de culture.

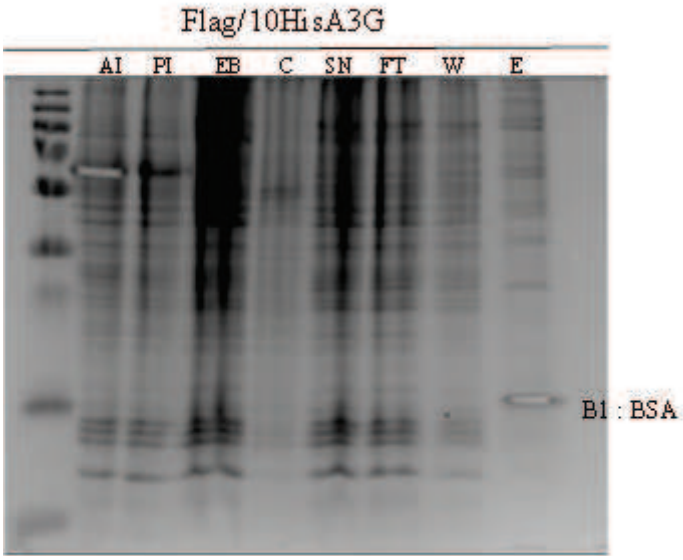


Figure 106 : Purification de Flag/10HisA3G sur colonne de Nickel .Les cellules BHK21, infectées par les virus recombinants exprimant (après induction IPTG) la protéine recombinante Flag/10His A3G, ont été récoltées et lysées. (A) Les extraits bruts ont été injectés sur une colonne de Nickel de 1 mL et les protéines ont été éluées par 400 mM d'imidazole. Les différents échantillons ont ensuite été migrés sur gel de polyacrylamide dénaturant 15 % puis les protéines ont été révélées au bleu de Coomassie. La bande B1 pouvant correspondre à notre protéine d'intérêt a ensuite été analysée par spectrométrie de masse mais ce n'était que la contamination par la BSA . AI : avant induction ; PI : après induciotn ; EB : extrait brut ; C : culot ; SN : surnageant ; FT : flow through ; W et W1 : lavage à 40 mM imidazole ; W2 : lavage à 133 mM imidazole ; E : élution à 400 mM imidazole ; E 5 µL : élution E concentrée.

Afin de savoir si A3G était tout de même présente dans l'élution, la totalité de l'élution liquide a été analysée en spectrométrie de masse. Etonnamment, les résultats obtenus indiquent qu'A3G est bien retrouvée dans l'élution en tant que protéine majoritaire. En conclusion, A3G est bien soluble et présente dans l'élution mais l'absence de bandes visibles sur gel montre que la quantité purifiée n'est pas très importante.

b. Purification de Strep-A3G

Une seconde purification a ensuite été menée pour la construction Strep A3G à partir de 4 litres de culture cellulaire. L'élution, pour laquelle un pic a été observé sur le chromatogramme (Fig. 107A), a été concentrée directement après sortie de colonne et déposée sur gel. On peut constater (Fig. 107B piste A3G) qu'elle est encore très contaminée. Néanmoins, on remarque la même bande intense (Fig. 107B B1) observée dans l'élution d'A3G avec le complexe et pour laquelle on a bien identifié A3G. Ainsi, nous avons réussi à solubiliser A3G et en récupérer une quantité suffisante pour être détectée.

Ces deux purifications ont été menées en présence d'un détergent (CHAPS) (Tableau 20) pour essayer d'améliorer la solubilité de la protéine A3G seule.

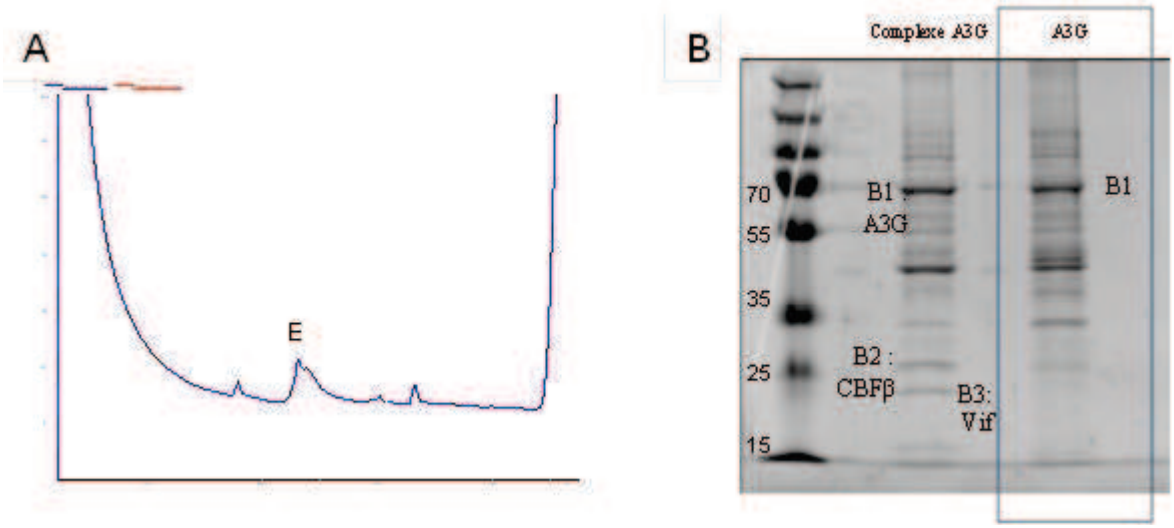


Figure 107 : Purification de StrepA3G sur colonne de streptavidine.
Les cellules BHK21, infectées par les virus recombinants exprimant (après induction IPTG) la protéine recombinante StrepA3G, ont été récoltées et lysées. Les extraits bruts ont été injectés sur une colonne de streptavidine de 1 mL et les protéines ont été éluées par 2,5 mM de desthiobiotine. Le profil d'élutions est présenté en A (pic E) en donnant l'absorbance à 280 nm en fonction du temps. L'élution concentrée de l'éluat d'A3G est représenté en B dans la piste A3G encadrée en bleu et comparée à l'éluat du complexe. La bande B1 pouvant correspondre à A3G a été ensuite analysée par spectrométrie de masse.

E. Validation par spectrométrie de masse

Afin de vérifier la présence des protéines dans les purifications réalisées au cours de cette partie, nous avons eu recours à la technique de spectrométrie de masse détaillée dans la partie « Matériels et Méthodes ». APOBEC3G a été identifiée avec un très bon recouvrement, qu'elle soit exprimée seule ou au sein du complexe.

1. Validation du système de clivage à la TEV

En premier lieu, il était primordial de vérifier que le système de clivage à la protéase TEV est bien fonctionnel sachant ce système n'a été testé à ce jour que pour un complexe de deux protéines par l'équipe de Marc Ruff sur le complexe intégrase/LEDGF. Dans notre cas le complexe étant formé de 5 protéines (CBFβ, Vif, EloB, EloC et A3G), il n'était pas évident que les protéines seraient exprimées puis clivées correctement.

Un résultat primaire de la validation de ce système était le Western blot dans lequel nous avons pu identifier Vif et A3G aux tailles attendues (Fig. 103). Cela signifie que la protéase TEV a bien clivé entre les différentes protéines avant qu'elles ne se rassemblent *in vitro* pour former le complexe. Le système de clivage a été vérifié ensuite en spectrométrie de masse.

L'étape suivante était de vérifier ce clivage par spectrométrie de masse. Pour cela, nous avons analysé en solution l'éluat de la purification de Flag/10HisA3G sur la colonne de Nickel après une digestion à la trypsine. Les analyses effectuées ont montré la présence des trois protéines TEV, A3G et YFP (Fig. 109). Le premier résultat intéressant est l'identification d'A3G dans les analyses de l'échantillon en solution malgré l'absence d'une bande de surexpression.

Sachant que la spécificité de coupure à la trypsine se situe au niveau des résidus arginine et lysine, tous les peptides d'une protéine digérée à la trypsine doivent théoriquement comporter R ou K à l'extrémité C-terminale. Cependant , nous pouvons constater dans notre analyse que nous retrouvons des peptides trypsiques conventionnels (Fig. 109) mais aussi deux peptides non-trypsiques en positions 263-268 et 697-703 (Fig. 109 rectangles). Ce peptide non-trypsique identifié dans ces deux positions est le site de clivage à la protéase TEV (ENLYFQ) , présent entre la protéase TEV et A3G et entre A3G et YFP (Fig. 108 marron) . Nous pouvons donc en déduire que la protéase TEV a bien clivé entre les protéines et que ce système de polyprotéine TEV-A3G-YFP était bien fonctionnelle.

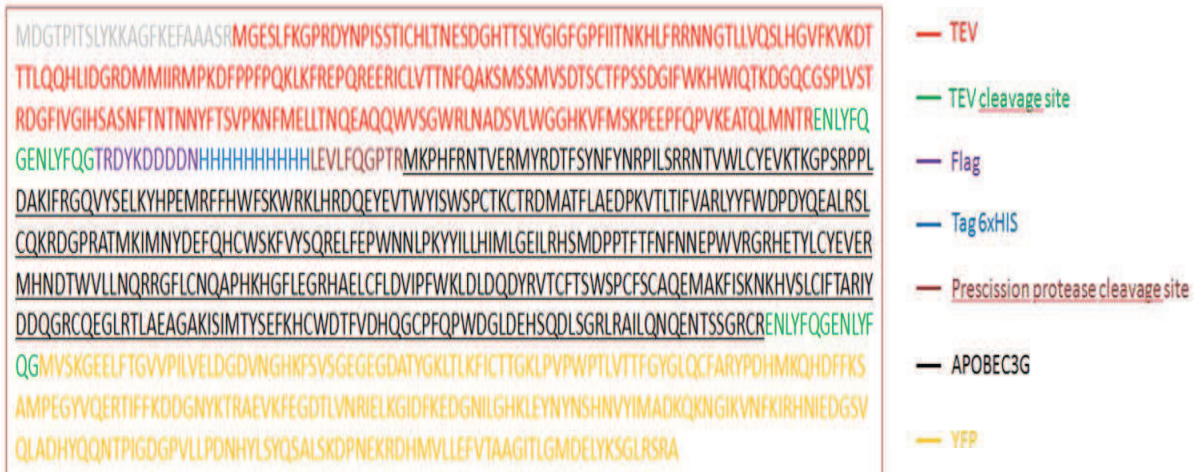


Figure 108 : La séquence de la polyprotéine TEV-Flag/10HisA3G-YFP.
Cette séquence a été utilisée en spectrométrie de masse pour l'identification des peptides trypsiques correspondant dans les banques de données appropriée.

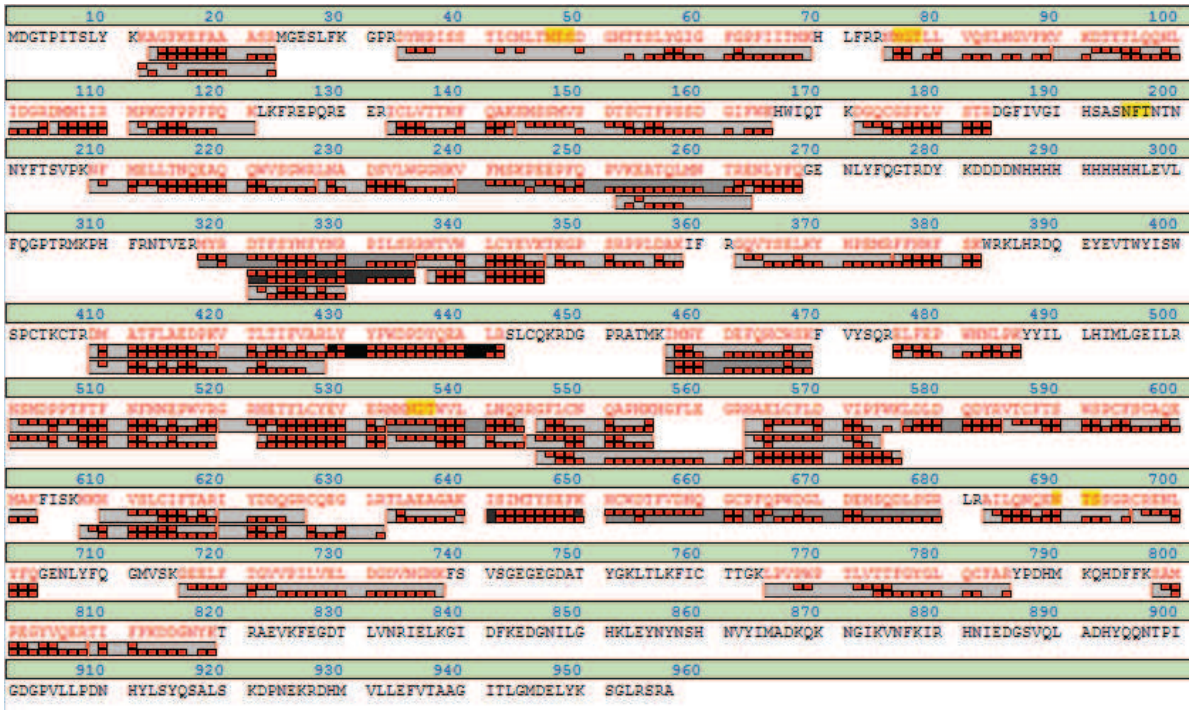
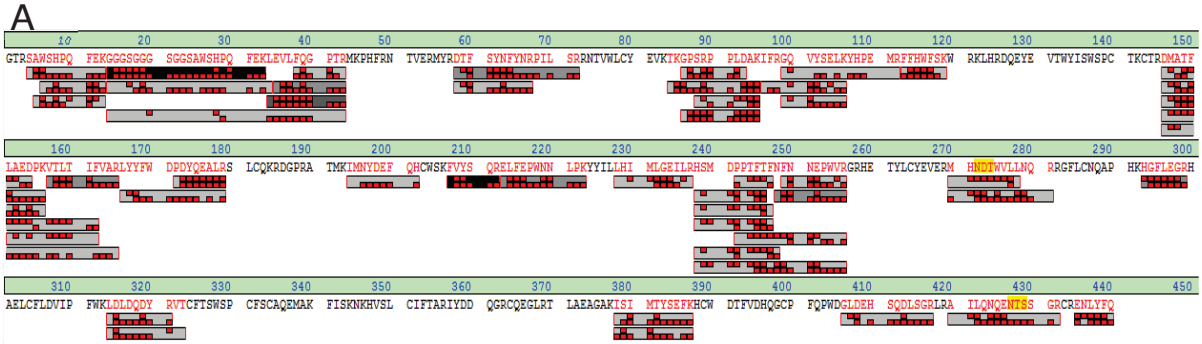


Figure 109 : Validation du système de clivage à la TEV en spectrométrie de masse.
La piste a été analysée en spectrométrie de masse et des peptides correspondant à la TEV, A3G et YFP ont été retrouvés avec un bon recouvrement de séquence. L'identification du peptide ENLYQ (site de clivage à la TEV) dans les analyses indique que la protéase TEV a bien clivée entre les différentes protéines.

2. Identification d'A3G et du complexe

Après avoir vérifié que le système de clivage à la TEV fonctionne bien, la deuxième étape consistait à vérifier la présence des différentes protéines formant le complexe ainsi que leur abondance dans l'échantillon. Cette identification a été réalisée en appliquant les deux techniques MALDI TOF et/ou LC-MS/MS. Une étude préliminaire de modifications post-traductionnelles sur APOBEC a été menée. Nous avons pu identifier une monoubiquitinylation pour A3G exprimée seule (en position K609 de la polyprotéine) (résultat pas montré). Ces analyses ne sont encore que préliminaires et il est probable que d'autres modifications soient présentes sur APOBEC aidant ainsi à sa solubilisation (phosphorylation, glycosylation, etc).

A



B

Accession Number	Protein Description	MW [kDa]	pI	Score	# Peps	# Spectra	#Spectra/MW	SC [%]
APOBEC3G_TagStrepNter	>APOBEC3G_TagStrepNter	52,4	8.5	2699.8	51	243	4,64	56.8
ACTG_HUMAN	Actin, cytoplasmic 2 OS=Homo sapiens GN=ACTG1 PE=1 SV=1	41,8	5.2	1856.5	25	180	4,31	55.5
ACTH_HUMAN	Actin, gamma-enteric smooth muscle OS=Homo sapiens GN=ACTG2 PE=1 SV=1	41,8	5.2	1134.4	17	116	2,78	30.3
ABC3G_HUMAN	DNA dC->dU-editing enzyme APOBEC-3G OS=Homo sapiens GN=APOBEC3G PE=1 SV=1	46,4	9.2	1151.3	19	93	2,00	38.5
VIME_HUMAN	Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4	53,6	4.9	961.0	14	26	0,49	29.0
PGK1_HUMAN	Phosphoglycerate kinase 1 OS=Homo sapiens GN=PGK1 PE=1 SV=3	44,6	9.2	476.1	5	7	0,16	15.8
ILF2_HUMAN	Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=2	43	5.0	261.3	3	5	0,12	8.7
CISY_HUMAN	Citrate synthase, mitochondrial OS=Homo sapiens GN=CS PE=1 SV=2	51,7	9.1	52.3	1	2	0,04	2.4
IDHC_HUMAN	Isocitrate dehydrogenase [NADP] cytoplasmic OS=Homo sapiens GN=IDH1 PE=1 SV=2	46,6	6.6	128.6	1	2	0,04	2.7

Figure 110 : Identification d'A3G par spectrométrie de masse.
La bande B1 identifié sur le gel présenté Figure 13C a été analysé en spectrométrie de masse LC-ESI-TOF MS/MS. Les résultats ont été comparés à la séquence de notre protéine A3G recombinante qui possède l'étiquette Twin-Strep en N-terminal. Le recouvrement de séquence (56,8%) est traduit par le nombre de peptides identifiés (boîtes grises) et les spectres obtenus sont représentés par les carrés rouges. Une recherche dans la banque de donnée Mascott (B) a permis d'identifier Strep A3G comme majoritaire dans la bande (1ère ligne du tableau B) mais a révélé de l'actine comme contaminant (3ème et 4ème ligne du tableau B) dans des proportions presque équivalentes.

Figure 111: Identification de Vif, CBFβ et EloC en spectrométrie de masse.
Les bandes B2, B3 et B4 identifié sur les gels présentés dans les Figures ont été analysés en spectrométrie de masse LC-ESI-TOF MS/MS. Les résultats ont été comparés aux sequences de nos protéines recombinantes Vif, CBF et EloC. On retrouve bien les protéines recherchées avec plus ou moins un bon recouvrement traduits par le nombre de peptides identifiés.). Nous n'avons pas réussi à identifier EloB dans les bandes choisies pour les analyses en spectrométrie de masse

F. Activité d'APOBEC3G

A3G a bien été exprimée mais n'a pas pu être purifiée correctement. Afin de savoir si elle était bien repliée et fonctionnelle, nous avons testé son activité cytidine désaminase (Matériels et Méthodes). Un contrôle positif permet de voir sur le profil d'élution que les fragments de 17 et 22 nucléotides générés comme produits des trois réactions du test (cytidine désaminase, uracyle ADN glycosylase et clivage de la chaine phosphodiester en milieu basique sous l'effet de la chaleur) sont élués respectivement vers 31 mL et 32 mL (Fig. 112A). L'ADN substrat de 39 nucléotides est élué vers 34 mL (Fig. 112B). L'ADN incubé en présence d'A3G seule (Fig. 112C) présente les 2 pics d'élution vers 31 et 32 mL. On peut donc en déduire qu'A3G est bien active, la présence des deux fragments clivés attendus étant observée. Un pic d'élution vers 34 mL correspondant à l'ADN non clivé résiduel est également observé.

Dans les banques de données, il n'existe aucune désaminase d'ADN simple brin identifiée pour le hamster doré duquel proviennent les cellules utilisées. Seule APOBEC1 est identifiée mais elle est responsable de la désamination du codon CAA en UAA dans l'ARNm de l'apolipoprotéine B. En effet, un alignement de séquences ne m'a pas permis d'identifier dans cette espèce de protéines homologues d'A3G ou de protéines douées d'activité de désamination des dC en dU. Autrement dit, il n'existe aucune preuve de la présence d'une désaminase chez les hamsters dorés capable de désaminer de l'ADN simple brin avec le motif de reconnaissance spécifique CCCA. Mais pour confirmer l'absence d'activité désaminase endogène et que les résultats observés sont bien dus à l'A3G recombinante, nous avons réalisé un test dans les mêmes conditions avec un extrait brut de cellules BHK21 non infectées. L'absence de fragments clivés a permis de conclure que l'activité cytidine désaminase observée est bien due à A3G recombinante produite dans nos conditions. Ces résultats préliminaires restent purement qualitatifs car A3G, n'étant pas pure, il est difficile de connaître la quantité d'enzyme mise en jeu dans nos expériences.

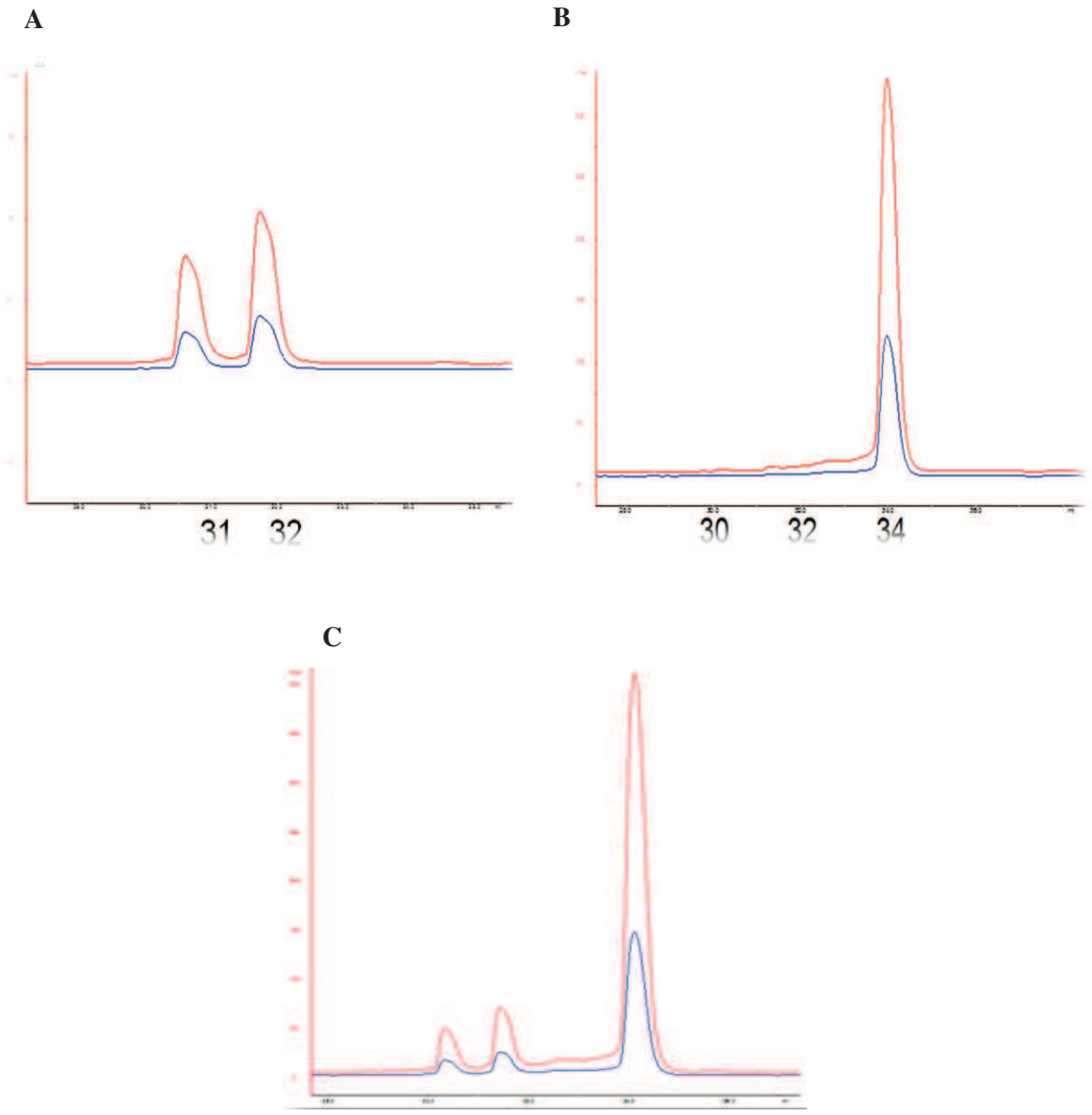


Figure 112 : Test d'activité de l'activité cytidine désaminase d'APOBEC3G. (C) Un test d'activité d'activité cytidine désaminase UDG-dépendant a été réalisé avec l'élution d'A3G purifiée sur colonne de streptavidine. (A) Un contrôle positif avec un brin d'ADN possédant le motif CCU, (B) négatif avec un brin d'ADN possédant le motif CCC. Les flèches bleues claires et foncées indiquent respectivement les pics d'élution des fragments de 17 et 22 nucléotides issus de l'hydrolyse alcaline au niveau du site abasique créent par l'uracile ADN glycosylase (UDG).

IV. Conclusions, discussion et perspectives

Les résultats obtenus dans cette troisième partie de ma thèse m'ont permis d'obtenir la protéine A3G soluble et fonctionnelle en cellules mammifères, contrairement aux travaux menés dans *E. coli*. Or, il est notoire que les modifications post-traductionnelles pourraient jouer un rôle important dans la solubilité des protéines. Pour vérifier cette hypothèse, des analyses préliminaires ont été effectuées à la plateforme de protéomique par spectrométrie de masse. Nous avons pu identifier un peptide monoubiquitinylé pour A3G exprimée seule (en position K609 de la polyprotéine) mais des recherches plus poussées et ciblant d'autres modifications (phosphorylation, glycosylation, …) devront être envisagées. Cependant, ces études nécessiteront une amélioration de la pureté des protéines. D'autre part, l'expression d'une polyprotéine avec le système de la vaccine a été testée pour la première fois sur un complexe de grande taille pendant ma thèse. Les résultats obtenus ont permis de valider cette stratégie et devraient faire l'objet d'une publication.

Des études structurales sur le complexe constituaient le but final de ce projet. Les faibles rendements obtenus ont cependant représenté un obstacle pour atteindre cet objectif. Ces faibles rendements peuvent être liés à une expression hétérogène des protéines par les virus recombinants. En effet, tous les virus montrant une fluorescence verte ne montrent pas une fluorescence jaune (indicateur de l'expression de la polyprotéine). On peut suspecter que la présence de sept protéines dépendant d'un unique RBS limite les rendements de synthèse des protéines par le ribosome, le temps nécessaire à la traduction étant d'autant plus important que la polyprotéine est longue. De plus, des études ont montré une diminution du taux de virus dans les cellules après induction (3 à 6 fois moins) ce qui pourrait également expliquer les faibles rendements d'expression [203]. Pour pallier ces problèmes, il faudrait envisager des plus gros volumes de culture. L'optimisation de cette stratégie est encore à l'étude par l'équipe de Marc Ruff.

La température peut également avoir un impact sur les rendements d'expression. En effet, les études précédentes menées avec *E. coli (Résultats partie2)* ont montré que les températures optimales d'expression se situaient entre 15 et 25 °C pour le complexe CBFβ – Vif – EloB - EloC, et à 18 °C pour le domaine C-terminal d'A3G. Or, dans nos conditions de culture, une sensibilité des cellules BHK21 à la diminution de la température nous contraint à réaliser l'induction à 37°C. Cette grande différence de température pourrait avoir un effet sur l'expression des protéines. De plus, A3G étant exprimée avec le complexe Vif - E3 ubiquitine ligase, son éventuelle dégradation par le protéasome

est aussi à envisager. En effet, même s'il manque des éléments importants pour ce mécanisme, tels que la Cul5 et les facteurs de polyubiquitination, leurs homologues pourraient être présents dans la cellule de hamster. Rajouter des inhibiteurs de protéasome au milieu de culture durant l'infection aurait pu permettre de tester cette hypothèse. Malheureusement, des études ont montré que l'utilisation d'inhibiteurs de la voie de dégradation ubiquitine-dépendant par le protéasome 26S, comme le MG132, inhibe la réplication du virus de la vaccine en bloquant l'expression des gènes intermédiaires et tardifs [241].

Des pertes durant les étapes de purification et notamment lors de la sonication et l'ultracentrifugation sont également probables. En effet, une identification de la protéine A3G aussi bien dans le culot que dans le surnageant des cellules dans l'expérience de Western Blot (Fig. 103) suggère qu'une partie d'A3G n'est pas soluble. La quantité de protéines d'intérêt injectée sur les colonnes de purification serait alors réduite. Ceci peut favoriser la liaison de protéines contaminantes aux sites non saturés de la résine et expliquer la contamination importante observée dans les profils d'élution. Une façon de surmonter ce problème serait d'ajouter des détergents dans les tampons de purification ce qui pourrait améliorer d'avantage la solubilité d'A3G. Mais ceci ne peut être applicable dans le cas du complexe qui serait dissocié en présence de détergents. Cette hypothèse a été testée en rajoutant du CHAPS pour A3G seule mais les résultats obtenus ne semblent pas indiquer une meilleure solubilité d'A3G en présence de détergent par rapport à A3G avec le complexe. Il est cependant possible que la concentration en CHAPS utilisée dans ce cas n'était pas adaptée, ou que la solubilité a effectivement été augmentée pour A3G seule mais que le complexe compense l'efficacité du détergent en favorisant également la solubilité d'A3G. Il sera indispensable d'optimiser les conditions de purification même si les rendements restent limitants pour y parvenir. Il pourra également être intéressant de réaliser deux purifications successives avec le complexe en utilisant les deux étiquettes Flag/10His et Twin-Strep.

Une activité spécifique d'A3G, capable de catalyser la réaction irréversible de désamination des dC en dU, a été mise en évidence par le test d'activité préliminaire. Ceci montre que la protéine produite est correctement repliée. Toutefois, une évaluation quantitative est difficile en raison de la forte contamination avec d'autres protéines. Ainsi, il est difficile de savoir si la présence d'ADN non clivé est synonyme d'une activité partielle (une partie seulement de la protéine est bien repliée) ou d'une très faible quantité d'enzyme au départ. Il sera nécessaire de réaliser des tests enzymatiques plus poussés afin de déterminer la fraction de protéines actives à partir des paramètres connus de l'activité enzymatique d'A3G.

CONCLUSION GENERALE

Vif joue un rôle très important dans l'infectivité du VIH-1 et dans la dégradation du facteur de défense innée APOBEC3G/F. La structure cristallographique de Vif en complexe avec les composants de la machinerie d'ubiquitinylation ouvrira un nouvel épisode dans l'investigation des antirétroviraux.

Concernant les retombées de ce travail, les publications très prolifiques ces derniers temps sur les protéines Vif et APOBEC3 montrent à quel point des informations biophysiques, et structurales en particulier, sont très attendues. Les exemples de structures cristallographiques ayant contribuées au développement rationnel d'antiviraux sont désormais nombreux : les inhibiteurs de protéase de HIV-1 saquinavir (Invirase ®), amprenavir (Agesenase ®) et nelfinavir (Viracept ®), ou la découverte du zanamivir (Relenza ®) et de l'oseltamivir (Tamiflu ®) contre le virus de la grippe ont tous été développés à partir de structures cristallographiques. Nous espérons que l'obtention d'une structure tridimensionnelle de bonne qualité du complexe Vif/APOBEC, permettra de mieux comprendre le mécanisme de reconnaissance d'APOBEC par Vif et pourra alors servir de base pour l'élaboration de molécules capables de contrer la dégradation d'APOBEC par le virus.

Au cours de ma thèse, j'ai réussi à produire le complexe Vif-CBFβ-EloB-EloC-Cul5-Rbx2 ainsi que le complexe Vif-CBFβ-EloB-EloC-A3FCD2Chimère en combinant la co-expression *in vivo* et *in vitro* dans *E. coli*. Malheureusement, nos tests de cristallisation sont restés négatifs. Concernant la protéine APOBEC complète, nos essais de production d'APOBEC seul ou au sein du complexe de l'ubiquitine ligase dans *E. coli* n'ont pas donné des résultats probants.

Des études biophysiques et structurales compatibles avec des petites quantités de protéines peuvent tout de même être envisagées. Si l'on obtient des quantités de protéines pures avec nos rendements actuels, nous pouvons espérer faire des études en cryomicroscopie électronique. En effet, les dernières innovations liées aux détecteurs et microscopes ont permis d'améliorer considérablement la résolution et les limites inférieures de résolution. Ainsi, de hautes résolutions ont pu être atteintes récemment (2,2 Å) [242]. Notre complexe ayant une taille d'environ 150 KDa et A3G seule d'environ 50 KDa, de nouvelles possibilités d'études structurales s'offrent alors à nous. Des analyses de résonance plasmonique de surface ou de rétention sur filtre pourront également être considérées de manière complémentaire pour étudier la cinétique d'association du complexe.

A3G, posant des problèmes de solubilité, sa surproduction reste difficile. D'autres études ont été menées pour surexprimer la protéine A3G entière en utilisant le système d'expression Baculovirus mais les rendements étaient limités aux microgrammes. Récemment, Kouno et al. (2015) ont réussi à produire dans *E. coli* un variant du domaine N-terminal d'A3G pur en grande quantité. Ceci a permis de résoudre la structure par RMN, grâce à des mutations dans le domaine N-terminal en s'inspirant du domaine C-terminal qui est soluble dans *E. coli*. Il serait alors intéressant de savoir si ces mêmes mutations favoriseraient également la solubilité de la protéine entière

De nombreuses études récentes ont apporté des informations précieuses sur les mécanismes et les structures de Vif ainsi qu'A3G, qui viennent enrichir les cibles thérapeutiques potentielles. La résolution de la structure de Vif a offert de nouvelles perspectives dans le développement de molécules inhibitrices de l'interaction avec le complexe E3 ubiquitine ligase. Une meilleure connaissance des détails des interactions entre A3G et Vif dans le complexe permettrait d'affiner les connaissances du mécanisme de polyubiquitination et de développer des molécules plus à même de bloquer l'action de Vif, ceci toujours dans le but de rétablir l'action antivirale d'A3G.

REFERENCES BIBILOGRAPHIQUES

1. Weiss RA (2006) The discovery of endogenous retroviruses. Retrovirology 3:67

2. Barré-Sinoussi F, Chermann JC, Rey F, et al (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868–871

3. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C (1986) Isolation of a new human retrovirus from West African patients with AIDS. Science 233:343–346

4. Sharp PM, Shaw GM, Hahn BH (2005) Simian immunodeficiency virus infection of chimpanzees. J Virol 79:3891–3902

5. Bibollet-Ruche F, Bailes E, Gao F, et al (2004) New simian immunodeficiency virus infecting De Brazza's monkeys (Cercopithecus neglectus): evidence for a cercopithecus monkey virus clade. J Virol 78:7748–7762

6. Weiss RA (2006) The discovery of endogenous retroviruses. Retrovirology 3:67

7. Russell RS, Liang C, Wainberg MA (2004) Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? Retrovirology 1:23

8. Kuznetsov YG, Victoria JG, Robinson WE, McPherson A (2003) Atomic force microscopy investigation of human immunodeficiency virus (HIV) and HIV-infected lymphocytes. J Virol 77:11896–11909

9. Clapham PR, McKnight A (2002) Cell surface receptors, virus entry and tropism of primate lentiviruses. J Gen Virol 83:1809–1829

10. Cantin R, Méthot S, Tremblay MJ (2005) Plunder and stowaways: incorporation of cellular proteins by enveloped viruses. J Virol 79:6577–6587

11. Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T, Gluckman JC, Montagnier L (1984) T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312:767–768

12. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312:763–767

13. Wu L, KewalRamani VN (2006) Dendritic-cell interactions with HIV: infection and viral dissemination. Nat Rev Immunol 6:859–868

14. (2015) Syndrome d'immunodéficience acquise. Wikipédia

15. Stevenson M (2003) HIV-1 pathogenesis. Nat Med 9:853–860

16. Mondor I, Ugolini S, Sattentau QJ (1998) Human immunodeficiency virus type 1 attachment to HeLa CD4 cells is CD4 independent and gp120 dependent and requires cell surface heparans. J Virol 72:3623–3634

17. Fortin JF, Cantin R, Tremblay MJ (1998) T cells expressing activated LFA-1 are more susceptible to infection with human immunodeficiency virus type 1 particles bearing host-encoded ICAM-1. J Virol 72:2105–2112

18. Nisole S, Saïb A (2004) Early steps of retrovirus replicative cycle. Retrovirology 1:9

19. Moore JP, McKeating JA, Huang YX, Ashkenazi A, Ho DD (1992) Virions of primary human immunodeficiency virus type 1 isolates resistant to soluble CD4 (sCD4) neutralization differ in sCD4 binding and glycoprotein gp120 retention from sCD4-sensitive isolates. J Virol 66:235–243

20. Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM, Berger EA (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 272:1955–1958

21. Berger EA, Doms RW, Fenyö EM, Korber BT, Littman DR, Moore JP, Sattentau QJ, Schuitemaker H, Sodroski J, Weiss RA (1998) A new classification for HIV-1. Nature 391:240

22. Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67:1–25

23. Wang Q, Bi W, Zhu X, Li H, Qi Q, Yu F, Lu L, Jiang S (2015) Non-neutralizing antibodies induced by HIV-1 gp41 NHR domain gain neutralizing activity in the presence of HIV fusion inhibitor enfuvirtide: a potential therapeutic vaccine strategy. J Virol. doi: 10.1128/JVI.00791-15

24. Castilletti C, Capobianchi MR, Fais S, Abbate I, Ficociello B, Ameglio F, Cordiali Fei P, Santini SM, Dianzani F (1995) HIV type 1 grown on interferon gamma-treated U937 cells shows selective increase in virion-associated intercellular adhesion molecule 1 and HLA-DR and enhanced infectivity for CD4-negative cells. AIDS Res Hum Retroviruses 11:547–553

25. Melikyan GB (2008) Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology 5:111

26. McDonald D, Vodicka MA, Lucero G, Svitkina TM, Borisy GG, Emerman M, Hope TJ (2002) Visualization of the intracellular behavior of HIV in living cells. J Cell Biol 159:441–452

27. Jones KA, Peterlin BM (1994) Control of RNA initiation and elongation at the HIV-1 promoter. Annu Rev Biochem 63:717–743

28. Vemula SV, Veerasamy R, Ragupathy V, Biswas S, Devadas K, Hewlett I (2015) HIV-1 induced nuclear factor I-B (NF-IB) expression negatively regulates HIV-1 replication through interaction with the long terminal repeat region. Viruses 7:543–558

29. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH (2010) Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol 84:9864–9878

30. Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, Prévost M-C, Allen TD, Charneau P (2007) HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. EMBO J 26:3025–3037

31. Farnet CM, Bushman FD (1997) HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. Cell 88:483–492

32. Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM (2006) An essential role for LEDGF/p75 in HIV integration. Science 314:461–464

33. Chun TW, Carruth L, Finzi D, et al (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature 387:183–188

34. Kao SY, Calman AF, Luciw PA, Peterlin BM (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature 330:489–493

35. Fernandes J, Jayaraman B, Frankel A (2012) The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. RNA Biol 9:6–11

36. Purcell DF, Martin MA (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67:6365–6378

37. Daugherty MD, Liu B, Frankel AD (2010) Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nature Structural & Molecular Biology 17:1337–1342

38. Jayaraman B, Crosby DC, Homer C, Ribeiro I, Mavor D, Frankel AD RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev–Rev response element complex. eLife. doi: 10.7554/eLife.04120

39. Schwartz S, Felber BK, Fenyö EM, Pavlakis GN (1990) Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. J Virol 64:5448–5456

40. Schwartz S, Felber BK, Pavlakis GN (1992) Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. Mol Cell Biol 12:207–219

41. Bour S, Strebel K (2003) The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. Microbes Infect 5:1029–1039

42. Bidou L, Stahl G, Grima B, Liu H, Cassan M, Rousset JP (1997) *In vivo* HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. RNA 3:1153–1158

43. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature 331:280–283

44. Karacostas V, Wolffe EJ, Nagashima K, Gonda MA, Moss B (1993) Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. Virology 193:661–671

45. Pelchen-Matthews A, Kramer B, Marsh M (2003) Infectious HIV-1 assembles in late endosomes in primary macrophages. J Cell Biol 162:443–455

46. Raposo G, Moore M, Innes D, Leijendekker R, Leigh-Brown A, Benaroch P, Geuze H (2002) Human macrophages accumulate HIV-1 particles in MHC II compartments. Traffic 3:718–729

47. Nydegger S, Foti M, Derdowski A, Spearman P, Thali M (2003) HIV-1 egress is gated through late endosomal membranes. Traffic 4:902–910

48. Brown D (2002) Structure and function of membrane rafts. Int J Med Microbiol 291:433–437

49. Gheysen D, Jacobs E, de Foresta F, Thiriart C, Francotte M, Thines D, De Wilde M (1989) Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell 59:103–112

50. Göttlinger HG, Sodroski JG, Haseltine WA (1989) Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 86:5781–5785

51. Bryant M, Ratner L (1990) Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc Natl Acad Sci USA 87:523–527

52. Freed EO, Martin MA (1996) Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol 70:341–351

53. Freed EO, Martin MA (1995) The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. J Biol Chem 270:23883–23886

54. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI, Hill CP (1997) Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science 278:849–853

55. Kondo E, Mammano F, Cohen EA, Göttlinger HG (1995) The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. J Virol 69:2759–2764

56. Paxton W, Connor RI, Landau NR (1993) Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. J Virol 67:7229–7237

57. Selig L, Pages JC, Tanchou V, Prévéral S, Berlioz-Torrent C, Liu LX, Erdtmann L, Darlix J, Benarous R, Benichou S (1999) Interaction with the p6 domain of the gag precursor mediates incorporation into virions of Vpr and Vpx proteins from primate lentiviruses. J Virol 73:592–600

58. Dannull J, Surovoy A, Jung G, Moelling K (1994) Specific binding of HIV-1 nucleocapsid protein to PSI RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues. EMBO J 13:1525–1533

59. Lever A, Gottlinger H, Haseltine W, Sodroski J (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J Virol 63:4085–4087

60. Ono A, Ablan SD, Lockett SJ, Nagashima K, Freed EO (2004) Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. Proc Natl Acad Sci USA 101:14889–14894

61. Lopez-Vergès S, Camus G, Blot G, Beauvoir R, Benarous R, Berlioz-Torrent C (2006) Tail-interacting protein TIP47 is a connector between Gag and Env and is required for Env incorporation into HIV-1 virions. Proc Natl Acad Sci USA 103:14947–14952

62. Pornillos O, Higginson DS, Stray KM, Fisher RD, Garrus JE, Payne M, He G-P, Wang HE, Morham SG, Sundquist WI (2003) HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. J Cell Biol 162:425–434

63. Strack B, Calistri A, Craig S, Popova E, Göttlinger HG (2003) AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell 114:689–699

64. von Schwedler UK, Stuchell M, Müller B, et al (2003) The protein network of HIV budding. Cell 114:701–713

65. Fujii K, Hurley JH, Freed EO (2007) Beyond Tsg101: the role of Alix in "ESCRTing" HIV-1. Nat Rev Microbiol 5:912–916

66. Ott DE, Coren LV, Sowder II RC, Adams J, Schubert U (2003) Retroviruses Have Differing Requirements for Proteasome Function in the Budding Process. J Virol 77:3384–3393

67. Miedema F (2008) A brief history of HIV vaccine research: stepping back to the drawing board? AIDS 22:1699–1703

68. Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF, rgp120 HIV Vaccine Study Group (2005) Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J Infect Dis 191:654–665

69. Pitisuttithum P, Gilbert P, Gurwith M, Heyward W, Martin M, van Griensven F, Hu D, Tappero JW, Choopanya K, Bangkok Vaccine Evaluation Group (2006) Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. J Infect Dis 194:1661–1671

70. Caskey M, Klein F, Lorenzi JCC, et al (2015) Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. Nature. doi: 10.1038/nature14411

71. Liu C, Wang S, Ren L, Hao Y, Zhang Q, Liu Y (2014) [A novel immunization strategy to induce strong humoral responses against HIV-1 using combined DNA, recombinant vaccinia virus and protein vaccines]. Bing Du Xue Bao 30:645–651

72. Bomsel M, Tudor D, Drillet A-S, et al (2011) Immunization with HIV-1 gp41 subunit virosomes induces mucosal antibodies protecting nonhuman primates against vaginal SHIV challenges. Immunity 34:269–280

73. Pulendran B, Ahmed R (2006) Translating innate immunity into immunological memory: implications for vaccine development. Cell 124:849–863

74. Shattock RJ, Haynes BF, Pulendran B, Flores J, Esparza J, Working Group convened by the Global HIV Vaccine Enterprise (2008) Improving defences at the portal of HIV entry: mucosal and innate immunity. PLoS Med 5:e81

75. Pennisi E (2012) Genomics. ENCODE project writes eulogy for junk DNA. Science 337:1159, 1161

76. Fernández-Suárez XM, Rigden DJ, Galperin MY (2014) The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. Nucleic Acids Res 42:D1–6

77. Grütter MG, Luban J (2012) TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. Curr Opin Virol 2:142–150

78. Rhodes DA, de Bono B, Trowsdale J (2005) Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? Immunology 116:411–417

79. Towers GJ (2007) The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology 4:40

80. Johnson WE, Sawyer SL (2009) Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics 61:163–176

81. Towers GJ, Hatziioannou T, Cowan S, Goff SP, Luban J, Bieniasz PD (2003) Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat Med 9:1138–1143

82. Wu F, Ourmanov I, Riddick N, Matsuda K, Whitted S, Plishka RJ, Buckler-White A, Starost MF, Hirsch VM (2015) TRIM5α restriction affects clinical outcome and disease progression in simian immunodeficiency virus-infected rhesus macaques. J Virol 89:2233–2240

83. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M (2011) Hexagonal assembly of a restricting TRIM5alpha protein. Proc Natl Acad Sci USA 108:534–539

84. Diaz-Griffero F (2011) Caging the beast: TRIM5α binding to the HIV-1 core. Viruses 3:423–428

85. Neil SJD, Sandrin V, Sundquist WI, Bieniasz PD (2007) An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. Cell Host Microbe 2:193–203

86. Neil SJD, Zang T, Bieniasz PD (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451:425–430

87. Strebel K, Klimkait T, Martin MA (1988) A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science 241:1221–1223

88. Galão RP, Le Tortorec A, Pickering S, Kueck T, Neil SJD (2012) Innate sensing of HIV-1 assembly by Tetherin induces NFκB-dependent proinflammatory responses. Cell Host Microbe 12:633–644

89. Hitt AL, Hartwig JH, Luna EJ (1994) Ponticulin is the major high affinity link between the plasma membrane and the cortical actin network in Dictyostelium. J Cell Biol 126:1433–1444

90. Köster B, Strand M (1994) Schistosoma mansoni: Sm23 is a transmembrane protein that also contains a glycosylphosphatidylinositol anchor. Arch Biochem Biophys 310:108–117

91. Nishikawa Y, Tragoolpua K, Makala L, Xuan X, Nagasawa H (2002) Neospora caninum NcSRS2 is a transmembrane protein that contains a glycosylphosphatidylinositol anchor in insect cells. Vet Parasitol 109:191–201

92. Yang H, Wang J, Jia X, McNatt MW, Zang T, Pan B, Meng W, Wang H-W, Bieniasz PD, Xiong Y (2010) Structural insight into the mechanisms of enveloped virus tethering by tetherin. Proc Natl Acad Sci USA 107:18428–18432

93. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, Stephens EB, Guatelli J (2008) The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell Host Microbe 3:245–252

94. Fitzpatrick K, Skasko M, Deerinck TJ, Crum J, Ellisman MH, Guatelli J (2010) Direct restriction of virus release and incorporation of the interferon-induced protein BST-2 into HIV-1 particles. PLoS Pathog 6:e1000701

95. Göttlinger HG (2008) HIV/AIDS: virus kept on a leash. Nature 451:406–408

96. Jia B, Serra-Moreno R, Neidermyer W, Rahmberg A, Mackey J, Fofana IB, Johnson WE, Westmoreland S, Evans DT (2009) Species-Specific Activity of SIV Nef and HIV-1 Vpu in Overcoming Restriction by Tetherin/BST2. PLoS Pathog. doi: 10.1371/journal.ppat.1000429

97. Bour S, Geleziunas R, Wainberg MA (1995) The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. Microbiol Rev 59:63–93

98. Margottin F, Bour SP, Durand H, Selig L, Benichou S, Richard V, Thomas D, Strebel K, Benarous R (1998) A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. Mol Cell 1:565–574

99. Dubé M, Roy BB, Guiot-Guillain P, Binette J, Mercier J, Chiasson A, Cohen EA (2010) Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment. PLoS Pathog 6:e1000856

100. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, Yatim A, Emiliani S, Schwartz O, Benkirane M (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature 474:654–657

101. Rice GI, Bond J, Asipu A, et al (2009) Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. Nat Genet 41:829–832

102. Laguette N, Benkirane M (2012) How SAMHD1 changes our view of viral restriction. Trends Immunol 33:26–33

103. Sharova N, Wu Y, Zhu X, Stranska R, Kaushik R, Sharkey M, Stevenson M (2008) Primate lentiviral Vpx commandeers DDB1 to counteract a macrophage restriction. PLoS Pathog 4:e1000057

104. Goujon C, Arfi V, Pertel T, Luban J, Lienard J, Rigal D, Darlix J-L, Cimarelli A (2008) Characterization of simian immunodeficiency virus SIVSM/human immunodeficiency virus type 2 Vpx function in human myeloid cells. J Virol 82:12335–12345

105. Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, Florens L, Washburn MP, Skowronski J (2011) Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. Nature 474:658–661

106. St Gelais C, de Silva S, Amie SM, Coleman CM, Hoy H, Hollenbaugh JA, Kim B, Wu L (2012) SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. Retrovirology 9:105

107. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, Rice CM (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. Nature 472:481–485

108. Planelles V, Benichou S (2009) Vpr and its interactions with cellular proteins. Curr Top Microbiol Immunol 339:177–200

109. Lim ES, Fregoso OI, McCoy CO, Matsen FA, Malik HS, Emerman M (2012) The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. Cell Host Microbe 11:194–204

110. Lu J, Pan Q, Rong L, Liu S-L, Liang C (2011) The IFITM Proteins Inhibit HIV-1 Infection. J Virol 85:2126–2137

111. Wilson SJ, Schoggins JW, Zang T, Kutluay SB, Jouvenet N, Alim MA, Bitzegeio J, Rice CM, Bieniasz PD (2012) Inhibition of HIV-1 particle assembly by 2',3'-cyclic-nucleotide 3'-phosphodiesterase. Cell Host Microbe 12:585–597

112. Duschene KS, Broderick JB (2012) Viperin: a radical response to viral infection. Biomol Concepts 3:255–266

113. Wang X, Han Y, Dang Y, Fu W, Zhou T, Ptak RG, Zheng Y-H (2010) Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. J Biol Chem 285:14346–14355

114. Abudu A, Wang X, Dang Y, Zhou T, Xiang S-H, Zheng Y-H (2012) Identification of molecular determinants from Moloney leukemia virus 10 homolog (MOV10) protein for virion packaging and anti-HIV-1 activity. J Biol Chem 287:1220–1228

115. Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, Cen S, Guo F, Liang C (2013) The interferon-inducible MxB protein inhibits HIV-1 infection. Cell Host Microbe 14:398–410

116. Maitra RK, Silverman RH (1998) Regulation of Human Immunodeficiency Virus Replication by 2′,5′-Oligoadenylate-Dependent RNase L. J Virol 72:1146–1152

117. Zhu Y, Chen G, Lv F, Wang X, Ji X, Xu Y, Sun J, Wu L, Zheng Y-T, Gao G (2011) Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. Proc Natl Acad Sci USA 108:15834–15839

118. Li M, Kao E, Gao X, et al (2012) Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. Nature 491:125–128

119. LaRue RS, Jónsson SR, Silverstein KAT, Lajoie M, Bertrand D, El-Mabrouk N, Hötzel I, Andrésdóttir V, Smith TPL, Harris RS (2008) The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC Mol Biol 9:104

120. Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature 433:430–433

121. Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418:646–650

122. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH (2003) DNA deamination mediates innate immunity to retroviral infection. Cell 113:803–809

123. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 424:94–98

124. Richardson SR, Narvaiza I, Planegger RA, Weitzman MD, Moran JV (2014) APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife 3:e02008

125. Lucifora J, Xia Y, Reisinger F, et al (2014) Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA. Science 343:1221–1228

126. Bogerd HP, Wiegand HL, Doehle BP, Cullen BR (2007) The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. Virology 364:486–493

127. Haché G, Liddament MT, Harris RS (2005) The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. J Biol Chem 280:10920–10924

128. Harris RS, Liddament MT (2004) Retroviral restriction by APOBEC proteins. Nat Rev Immunol 4:868–877

129. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature 445:447–451

130. Kitamura S, Ode H, Nakashima M, et al (2012) The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. NatStructMolBiol 19:1005–1010

131. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature 456:121–124

132. Bohn M-F, Shandilya SMD, Albin JS, et al (2013) Crystal Structure of the DNA Cytosine Deaminase APOBEC3F: The Catalytically Active and HIV-1 Vif-Binding Domain. Structure 21:1042–1050

133. Kouno T, Luengas EM, Shigematsu M, Shandilya SMD, Zhang J, Chen L, Hara M, Schiffer CA, Harris RS, Matsuo H (2015) Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. Nat Struct Mol Biol 22:485–491

134. Chaipan C, Smith JL, Hu W-S, Pathak VK (2013) APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. J Virol 87:444–453

135. Apolonia L, Schulz R, Curk T, Rocha P, Swanson CM, Schaller T, Ule J, Malim MH (2015) Promiscuous RNA Binding Ensures Effective Encapsidation of APOBEC3 Proteins by HIV-1. PLoS Pathog 11:e1004609

136. Harris RS, Hultquist JF, Evans DT (2012) The Restriction Factors of Human Immunodeficiency Virus. Journal of Biological Chemistry 287:40875–40883

137. Chiu Y-L, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC (2006) High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci USA 103:15588–15593

138. Henriet S, Mercenne G, Bernacchi S, Paillart J-C, Marquet R (2009) Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors. Microbiol Mol Biol Rev 73:211–232

139. Gillick K, Pollpeter D, Phalora P, Kim E-Y, Wolinsky SM, Malim MH (2013) Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination. J Virol 87:1508–1517

140. Chiu Y-L, Greene WC (2008) The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements. Annual Review of Immunology 26:317–353

141. Desimmie BA, Delviks-Frankenberrry KA, Burdick RC, Qi D, Izumi T, Pathak VK (2014) Multiple APOBEC3 Restriction Factors for HIV-1 and One Vif to Rule Them All. Journal of Molecular Biology 426:1220–1245

142. Iwatani Y, Chan DSB, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG (2007) Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res 35:7096–7108

143. Newman ENC, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM (2005) Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol 15:166–170

144. Browne EP, Allers C, Landau NR (2009) Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. Virology 387:313–321

145. MacMillan AL, Kohli RM, Ross SR (2013) APOBEC3 Inhibition of Mouse Mammary Tumor Virus Infection: the Role of Cytidine Deamination versus Inhibition of Reverse Transcription. J Virol 87:4808–4817

146. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424:99–103

147. Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, Münk C, Nymark-McMahon H, Landau NR (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 114:21–31

148. Turelli P, Mangeat B, Jost S, Vianin S, Trono D (2004) Inhibition of hepatitis B virus replication by APOBEC3G. Science 303:1829

149. Barraud P, Paillart J-C, Marquet R, Tisné C (2008) Advances in the structural understanding of Vif proteins. Curr HIV Res 6:91–99

150. Oberste MS, Gonda MA (1992) Conservation of amino-acid sequence motifs in lentivirus Vif proteins. Virus Genes 6:95–102

151. Henriet S, Richer D, Bernacchi S, Decroly E, Vigne R, Ehresmann B, Ehresmann C, Paillart J-C, Marquet R (2005) Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. J Mol Biol 354:55–72

152. Dettenhofer M, Cen S, Carlson BA, Kleiman L, Yu X-F (2000) Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. Journal of virology 74:8938–8945

153. Bernacchi S, Henriet S, Dumas P, Paillart J-C, Marquet R (2007) RNA and DNA binding properties of HIV-1 Vif protein: a fluorescence study. J Biol Chem 282:26361–26368

154. Freisz S, Mezher J, Hafirassou L, Wolff P, Nominé Y, Romier C, Dumas P, Ennifar E (2012) Sequence and structure requirements for specific recognition of HIV-1 TAR and DIS RNA by the HIV-1 Vif protein. RNA Biology. doi: 10.4161/rna.20483

155. Yang X, Goncalves J, Gabuzda D (1996) Phosphorylation of Vif and its role in HIV-1 replication. Journal of Biological Chemistry 271:10121–10129

156. Zhang W, Du J, Evans SL, Yu Y, Yu X-F (2011) T-cell differentiation factor CBF-β regulates HIV-1 Vif-mediated evasion of host restriction. Nature. doi: 10.1038/nature10718

157. Jäger S, Kim DY, Hultquist JF, et al (2011) Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. Nature. doi: 10.1038/nature10693

158. Zhou X, Evans SL, Han X, Liu Y, Yu X-F (2012) Characterization of the Interaction of Full-Length HIV-1 Vif Protein with its Key Regulator CBFβ and CRL5 E3 Ubiquitin Ligase Components. PLoS ONE 7:e33495

159. Wang H, Lv G, Zhou X, Li Z, Liu X, Yu X-F, Zhang W (2014) Requirement of HIV-1 Vif C-terminus for Vif-CBF-β interaction and assembly of CUL5-containing E3 ligase. BMC Microbiol 14:290

160. Stanley BJ, Ehrlich ES, Short L, Yu Y, Xiao Z, Yu X-F, Xiong Y (2008) Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. J Virol 82:8656–8663

161. Bergeron JRC, Huthoff H, Veselkov DA, Beavil RL, Simpson PJ, Matthews SJ, Malim MH, Sanderson MR (2010) The SOCS-box of HIV-1 Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS Pathog 6:e1000925

162. Fribourgh JL, Nguyen HC, Wolfe LS, DeWitt DC, Zhang W, Yu X-F, Rhoades E, Xiong Y (2014) Core Binding Factor Beta Plays a Critical Role by Facilitating the Assembly of the Vif-Cullin 5 E3 Ubiquitin Ligase. Journal of Virology 88:3309–3319

163. Guo Y, Dong L, Qiu X, Wang Y, Zhang B, Liu H, Yu Y, Zang Y, Yang M, Huang Z (2014) Structural basis for hijacking CBF-β and CUL5 E3 ligase complex by HIV-1 Vif. Nature 505:229–233

164. Hultquist JF, Binka M, LaRue RS, Simon V, Harris RS (2012) Vif Proteins of Human and Simian Immunodeficiency Viruses Require Cellular CBF To Degrade APOBEC3 Restriction Factors. Journal of Virology 86:2874–2877

165. Zhang W, Du J, Evans SL, Yu Y, Yu X-F (2011) T-cell differentiation factor CBF-β regulates HIV-1 Vif-mediated evasion of host restriction. Nature. doi: 10.1038/nature10718

166. Feng Y, Baig TT, Love RP, Chelico L (2014) Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. Front Microbiol. doi: 10.3389/fmicb.2014.00450

167. Kao S, Khan MA, Miyagi E, Plishka R, Buckler-White A, Strebel K (2003) The Human Immunodeficiency Virus Type 1 Vif Protein Reduces Intracellular Expression and Inhibits Packaging of APOBEC3G (CEM15), a Cellular Inhibitor of Virus Infectivity. J Virol 77:11398–11407

168. Mehle A, Strack B, Ancuta P, Zhang C, McPike M, Gabuzda D (2004) Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. J Biol Chem 279:7792–7798

169. Stopak K, de Noronha C, Yonemoto W, Greene WC (2003) HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell 12:591–601

170. Sheehy AM, Gaddis NC, Malim MH (2003) The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med 9:1404–1407

171. Shao Q, Wang Y, Hildreth JEK, Liu B (2010) Polyubiquitination of APOBEC3G Is Essential for Its Degradation by HIV-1 Vif. J Virol 84:4840–4844

172. Marin M, Rose KM, Kozak SL, Kabat D (2003) HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 9:1398–1403

173. Kobayashi M, Takaori-Kondo A, Miyauchi Y, Iwai K, Uchiyama T (2005) Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. J Biol Chem 280:18573–18578

174. Albin JS, Harris RS (2010) Interactions of host APOBEC3 restriction factors with HIV-1 *in vivo*: implications for therapeutics. Expert Reviews in Molecular Medicine. doi: 10.1017/S1462399409001343

175. Siu KK, Sultana A, Azimi FC, Lee JE (2013) Structural determinants of HIV-1 Vif susceptibility and DNA binding in APOBEC3F. Nature Communications. doi: 10.1038/ncomms3593

176. Huthoff H, Malim MH (2007) Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. J Virol 81:3807–3815

177. Lavens D, Peelman F, Van der Heyden J, et al (2010) Definition of the interacting interfaces of Apobec3G and HIV-1 Vif using MAPPIT mutagenesis analysis. Nucleic Acids Res 38:1902–1912

178. Mangeat B, Turelli P, Liao S, Trono D (2004) A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem 279:14481–14483

179. Schröfelbauer B, Chen D, Landau NR (2004) A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci USA 101:3927–3932

180. Letko M, Silvestri G, Hahn BH, Bibollet-Ruche F, Gokcumen O, Simon V, Ooms M (2013) Vif proteins from diverse primate lentiviral lineages use the same binding site in APOBEC3G. J Virol 87:11861–11871

181. Zhen A, Wang T, Zhao K, Xiong Y, Yu X-F (2010) A single amino acid difference in human APOBEC3H variants determines HIV-1 Vif sensitivity. J Virol 84:1902–1911

182. Opi S, Kao S, Goila-Gaur R, Khan MA, Miyagi E, Takeuchi H, Strebel K (2007) Human immunodeficiency virus type 1 Vif inhibits packaging and antiviral activity of a degradation-resistant APOBEC3G variant. J Virol 81:8236–8246

183. Mercenne G, Bernacchi S, Richer D, Bec G, Henriet S, Paillart J-C, Marquet R (2010) HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. Nucleic Acids Res 38:633–646

184. Johnson VA, Calvez V, Günthard HF, Paredes R, Pillay D, Shafer R, Wensing AM, Richman DD (2011) 2011 update of the drug resistance mutations in HIV-1. Top Antivir Med 19:156–164

185. Jern P, Russell RA, Pathak VK, Coffin JM (2009) Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog 5:e1000367

186. Suspène R, Rusniok C, Vartanian J-P, Wain-Hobson S (2006) Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic Acids Res 34:4677–4684

187. Mulder LCF, Harari A, Simon V (2008) Cytidine deamination induced HIV-1 drug resistance. Proc Natl Acad Sci USA 105:5501–5506

188. Nathans R, Cao H, Sharova N, Ali A, Sharkey M, Stranska R, Stevenson M, Rana TM (2008) Small-molecule inhibition of HIV-1 Vif. Nature Biotechnology 26:1187–1192

189. Miyakawa K, Matsunaga S, Kanou K, et al (2015) ASK1 restores the antiviral activity of APOBEC3G by disrupting HIV-1 Vif-mediated counteraction. Nat Commun. doi: 10.1038/ncomms7945

190. Glazkova DV, Bogoslovskaia EV, Markelov ML, Shipulin GA, Pokrovskiĭ VV (2012) [Treatment of HIV-infection by means of gene therapy]. Vestn Akad Med Nauk SSSR 16–23

191. Hoxie JA, June CH (2012) Novel cell and gene therapies for HIV. Cold Spring Harb Perspect Med. doi: 10.1101/cshperspect.a007179

192. Soros VB, Greene WC (2006) APOBEC3G and HIV-1: Strike and counterstrike. Curr Infect Dis Rep 8:317–323

193. Sakuma R, Noser JA, Ohmine S, Ikeda Y (2007) Inhibition of HIV-1 replication by simian restriction factors, TRIM5alpha and APOBEC3G. Gene Ther 14:185–189

194. Xu H, Svarovskaia ES, Barr R, Zhang Y, Khan MA, Strebel K, Pathak VK (2004) A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci USA 101:5652–5657

195. Chastel C (1998) L'éradication de la variole : menaces persistantes et développements nouveaux. Virologie 2:9–16

196. Tolonen N, Doglio L, Schleich S, Krijnse Locker J (2001) Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. Mol Biol Cell 12:2031–2046

197. Traktman P (1990) The enzymology of poxvirus DNA replication. Curr Top Microbiol Immunol 163:93–123

198. DeLange AM, McFadden G (1990) The role of telomeres in poxvirus DNA replication. Curr Top Microbiol Immunol 163:71–92

199. Panicali D, Paoletti E (1992) Construction of poxviruses as cloning vectors: insertion of the thymidine kinase gene from herpes simplex virus into the DNA of infectious vaccinia virus. 1982. Biotechnology 24:503–507

200. Mackett M, Smith GL, Moss B (1982) Vaccinia virus: a selectable eukaryotic cloning and expression vector. Proc Natl Acad Sci USA 79:7415–7419

201. Moss B, Earl PL (2002) Overview of the vaccinia virus expression system. Curr Protoc Mol Biol Chapter 16:Unit16.15

202. Elroy-Stein O, Moss B (2001) Gene expression using the vaccinia virus/T7 RNA polymerase hybrid system. Curr Protoc Mol Biol Chapter 16:Unit16.19

203. Hebben M, Brants J, Birck C, et al (2007) High level protein expression in mammalian cells using a safe viral vector: Modified vaccinia virus Ankara. Protein Expression and Purification 56:269–278

204. Diebold M-L, Fribourg S, Koch M, Metzger T, Romier C (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. Journal of Structural Biology 175:178–188

205. Romier C, Ben Jelloul M, Albeck S, et al (2006) Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Acta Crystallographica Section D Biological Crystallography 62:1232–1242

206. Gold L (1990) Expression of heterologous proteins in Escherichia coli. Meth Enzymol 185:11–14

207. Sørensen HP, Mortensen KK (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb Cell Fact 4:1

208. Jonasson P, Liljeqvist S, Nygren P-A, Ståhl S (2002) Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. Biotechnol Appl Biochem 35:91–105

209. Pedersen S (1984) Escherichia coli ribosomes translate *in vivo* with variable rate. EMBO J 3:2895–2898

210. Chen D, Duggan C, Ganley JP, Kooragayala LM, Reden TB, Texada DE, Langford MP (2004) Expression of enterovirus 70 capsid protein VP1 in Escherichia coli. Protein Expr Purif 37:426–433

211. Choi AH-C, Basu M, McNeal MM, Bean JA, Clements JD, Ward RL (2004) Intranasal administration of an Escherichia coli-expressed codon-optimized rotavirus VP6 protein induces protection in mice. Protein Expr Purif 38:205–216

212. Zhang MY, Schillberg S, Prins M, Fischer R (1999) Optimizing expression of a rare codon-rich viral protein in Escherichia coli using the IMPACT system. Anal Biochem 271:202–204

213. Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. Biotechnol J 6:650–659

214. Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS ONE 3:e2189

215. Ho-Shing O, Lau KH, Vernon W, Eckdahl TT, Campbell AM (2012) Assembly of standardized DNA parts using BioBrick ends in E. coli. Methods Mol Biol 852:61–76

216. Information NC for B, Pike USNL of M 8600 R, MD B, Usa 20894 Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. - PubMed - NCBI. http://www.ncbi.nlm.nih.gov/pubmed/?term=bernard+couturier+1992. Accessed 6 Jul 2015

217. Moss B, Flexner C (1987) Vaccinia Virus Expression Vectors. Annual Review of Immunology 5:305–324

218. Moss B, Flexner C (1989) Vaccinia Virus Expression Vectorsa. Annals of the New York Academy of Sciences 569:86–103

219. Moss B (1992) Vaccinia virus vectors. Biotechnology 20:345–362

220. Moss B (1996) Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety. Proc Natl Acad Sci U S A 93:11341–11348

221. Merchlinsky M, Moss B (1992) Introduction of foreign DNA into the vaccinia virus genome by *in vitro* ligation: recombination-independent selectable cloning vectors. Virology 190:522–526

222. Scheiflinger F, Dorner F, Falkner FG (1992) Construction of chimeric vaccinia viruses by molecular cloning and packaging. Proc Natl Acad Sci USA 89:9977–9981

223. Isaacs SN (ed) (2004) Construction and Isolation of Recombinant Vaccinia Virus Using Genetic Markers - Springer.

224. Falkner FG, Moss B (1988) Escherichia coli gpt gene provides dominant selection for vaccinia virus open reading frame expression vectors. J Virol 62:1849–1854

225. White SD, Conwell K, Langland JO, Jacobs BL (2011) Use of a negative selectable marker for rapid selection of recombinant vaccinia virus. BioTechniques 50:303–309

226. Rauert W, Eddine AN, Kaufmann SHE, Weiss MS, Janowski R (2007) Reductive methylation to improve crystallization of the putative oxidoreductase Rv0765c from Mycobacterium tuberculosis. Acta Crystallogr Sect F Struct Biol Cryst Commun 63:507–511

227. Yumerefendi H, Tarendeau F, Mas PJ, Hart DJ (2010) ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. J Struct Biol 172:66–74

228. Tarendeau F, Boudet J, Guilligay D, et al (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat Struct Mol Biol 14:229–233

229. Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RWH (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 458:914–918

230. Bernacchi S, Mercenne G, Tournaire C, Marquet R, Paillart J-C (2011) Importance of the proline-rich multimerization domain on the oligomerization and nucleic acid binding properties of HIV-1 Vif. Nucleic Acids Res 39:2404–2415

231. Gallerano D, Devanaboyina SC, Swoboda I, Linhart B, Mittermann I, Keller W, Valenta R (2011) Biophysical characterization of recombinant HIV-1 subtype C virus infectivity factor. Amino Acids 40:981–989

232. Weeks et al. Fragments of the HIV-1 Tat Protein Specifically Bind TAR RNA.

233. Hamy F, Felder ER, Heizmann G, Lazdins J, Aboul-ela F, Varani G, Karn J, Klimkait T (1997) An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. Proc Natl Acad Sci USA 94:3548–3553

234. Mei HY, Mack DP, Galan AA, et al (1997) Discovery of selective, small-molecule inhibitors of RNA complexes--I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. Bioorg Med Chem 5:1173–1184

235. Salter JD, Lippa GM, Belashov IA, Wedekind JE (2012) Core-Binding Factor β Increases the Affinity between Human Cullin 5 and HIV-1 Vif within an E3 Ligase Complex. Biochemistry 51:8702–8704

236. Giri K, Scott RA, Maynard EL (2009) Molecular Structure and Biochemical Properties of the HCCH-Zn2+ Site in HIV-1 Vif. Biochemistry 48:7969–7978

237. Jester BC, Drillien R, Ruff M, Florentz C (2011) Using Vaccinia's innate ability to introduce DNA into mammalian cells for production of recombinant proteins. J Biotechnol 156:211–213

238. Pradeau-Aubreton K, Ruff M, Garnier J-M, Schultz P, Drillien R (2010) Vectors for recombinational cloning and gene expression in mammalian cells using modified vaccinia virus Ankara. Anal Biochem 404:103–105

239. Maillot B, Lévy N, Eiler S, et al (2013) Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex. PLoS ONE 8:e60734

240. Mulligan RC, Berg P (1981) Selection for animal cells that express the Escherichia coli gene coding for xanthine-guanine phosphoribosyltransferase. Proc Natl Acad Sci U S A 78:2072–2076

241. Satheshkumar PS, Anton LC, Sanz P, Moss B (2009) Inhibition of the ubiquitin-proteasome system prevents vaccinia virus DNA replication and expression of intermediate and late genes. J Virol 83:2469–2479

242. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JLS, Subramaniam S (2015) 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science 348:1147–1151

ANNEXES

1- Isolats de Vif

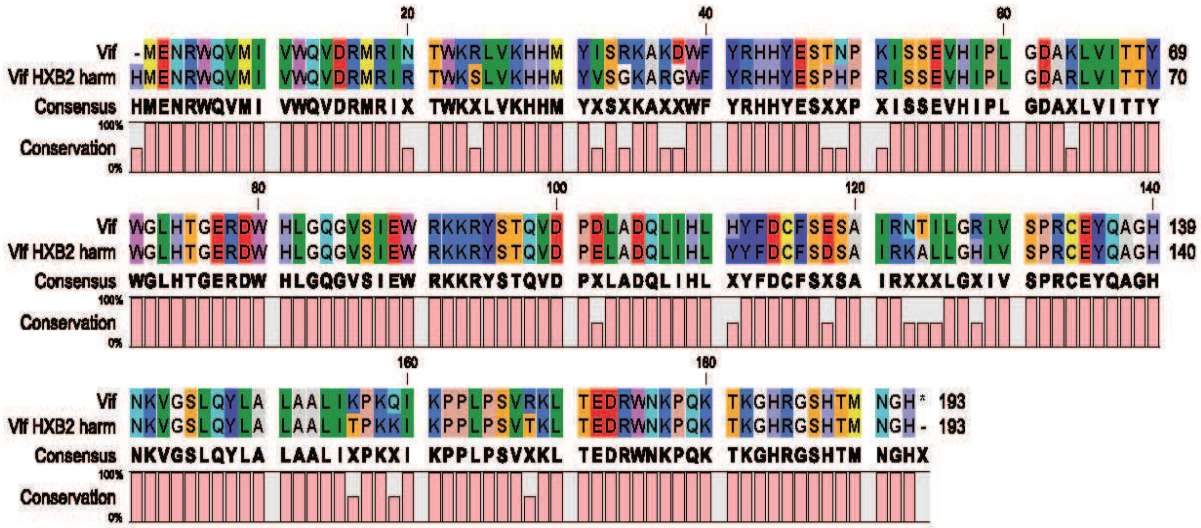
a. Séquence de la protéine Vif de l'isolat NL4-3

MENRWQVMIVWQVDRMRINTWKRLVKHHMYISRKAKDWFYRHHYESTNPKISSEVHIPLG
DAKLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPDLADQLIHLHYFDCFSESAIRN
TILGRIVSPRCEYQAGHNKVGSLQYLALAALIKPKQIKPPLPSVRKLTEDRWNKPQKTKGHRG
SHTMNGH

b. Séquence de la protéine Vif de l'isolat HXB2

HMENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHYESPHPRISSEVHIPL
GDARLVITTYWGLHTGERDWHLGQGVSIEWRKKRYSTQVDPELADQLIHLYYFDCFSDSAIR
KALLGHIVSPRCEYQAGHNKVGSLQYLALAALITPKKIKPPLPSVTKLTEDRWNKPQKTKGH
RGSHTMNGH

c. Alignement des deux séquences



Annexe 1 : Séquences des protéines de l'isolat NL4-3 (1) et HXB2 (2) et alignement des deux séquences (3). Il existe une 89,1% d'identité et 94,3% de similarité entre les deux séquences.

2- Harmonisation des codons : Exemple Protéine VifHXB2

a. VifHXB2 séquence originale et fréquence d'usage des codons chez *E .coli*

a

Name : >Vif_HBX2_database

No. of Codons : 193 Rare codons : 28 Ratio : 14.508%

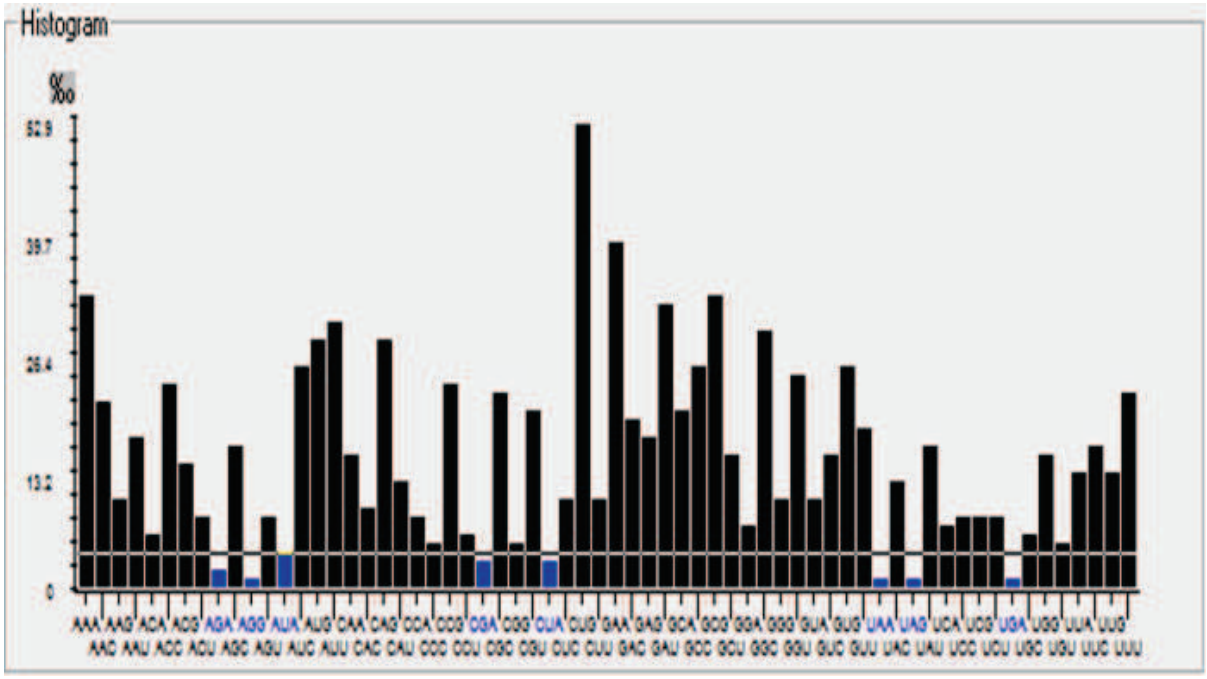
Table 3' (64x64) codons Sum up of Quantification

Redirect by : Escherichia coli

0 0.000%

0 0.000%

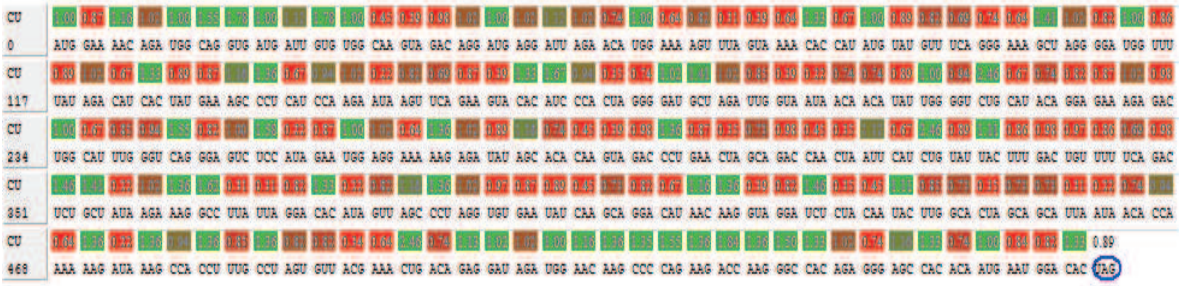
b



Annexe 2 : (a) Séquence de VifHXB2 non harmonisée avec les codons rares entourés en bleu, les codons ayant une fréquence d'usage (appelée valeur RSCU :*Relative synonymous codon usage*) elevée sont indiqués en vert, et ceux ayant une RSCU faible en rouge. (b) Fréquence d'usage des codons présents dans la séquence de VifHXB2 dans *E . coli*. Les codons rares sont définis comme étant ceux qui ont une fréquence d'usage en dessous de la limite (trait vertical)

b. VifHXB2 : Séquence originale et fréquence d'usage des codons chez *H. sapiens :*

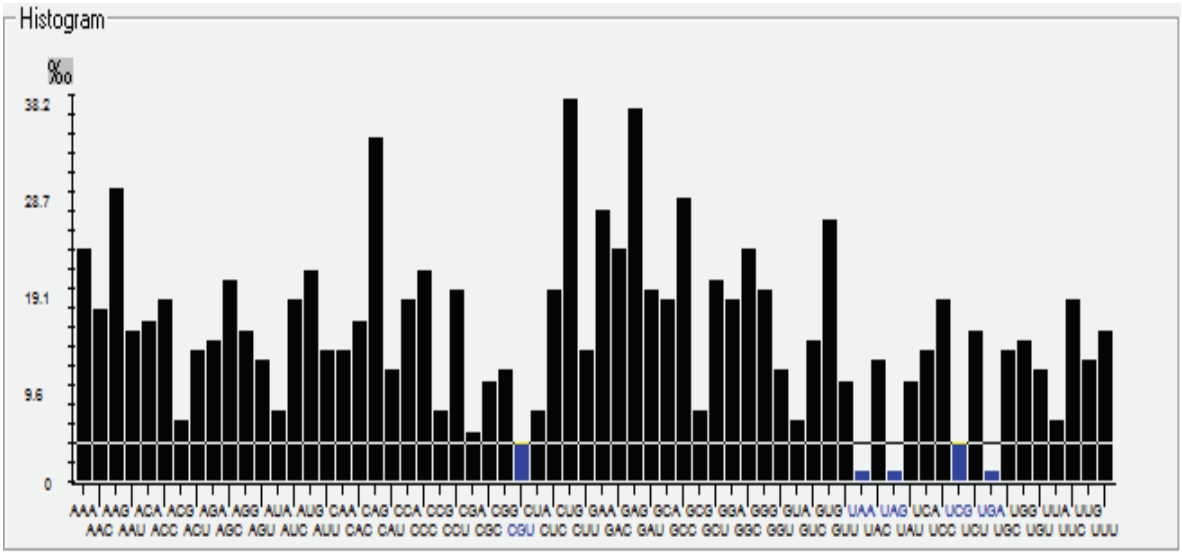
Name : >Vif_HBX2_database
No. of Codons : 193 Rare codons : 1 Ratio : 0.518%
Table 3' (64x64) codons Sum up of Quantification
Redirect by : Homo_sapiens
0 0.000%
0 0.000%



Annexe 3 : Séquence de VifHXB2 non harmonisée et la fréquence d'utilisation des codons dans l'organisme hôte (original) H. sapiens. Il n'existe pas de codon rare dans ce cas.

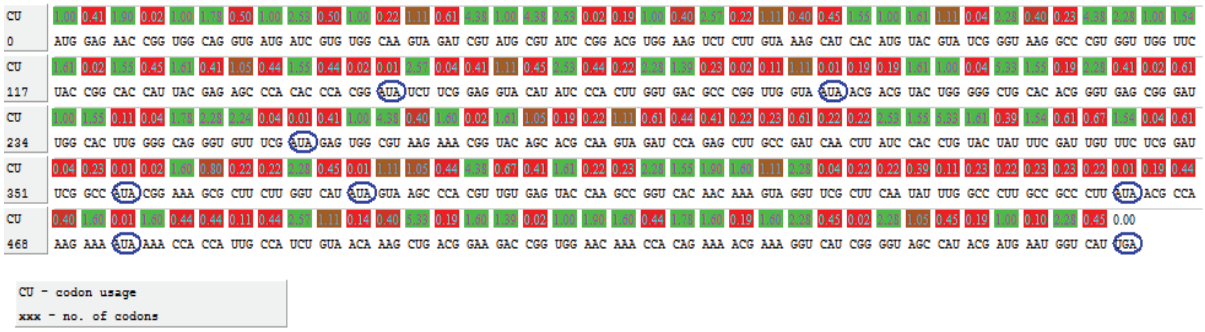
c. VifHXB2 harmonisée pour *E. coli*.

a



Name : >Vif_HBX2_database
No. of Codons : 193 Rare codons : 8 Ratio : 4.145%
Table 3' (64x64) codons Sum up of Quantification
Redirect by : Escherichia coli
0 0.000%
0 0.000%

b



Annexe 4 : Harmonisation des codons de VifHXB2 pour l'expression dans *E. coli*. Tous les codons n'ont pas été optimisés, il reste encore cinq codons rares conservés dans la séquence modifiée de leur importance pour le bon repliement de la protéine.

Chapter 9

Isothermal Titration Calorimetry: Assisted Crystallization of RNA–Ligand Complexes

[AU1] **Cyrielle da Veiga, Joelle Mezher, Philippe Dumas, and Eric Ennifar**

Abstract

The success rate of nucleic acids/ligands co-crystallization can be significantly improved by performing preliminary biophysical analyses. Among suitable biophysical approaches, isothermal titration calorimetry (ITC) is certainly a method of choice. ITC can be used in a wide range of experimental conditions to monitor in real time the formation of the RNA– or DNA–ligand complex, with the advantage of providing in addition the complete binding profile of the interaction. Following the ITC experiment, the complex is ready to be concentrated for crystallization trials. This chapter describes a detailed experimental protocol for using ITC as a tool for monitoring RNA/small molecule binding, followed by co-crystallization.

Key words Microcalorimetry, Crystallization, RNA–ligand complex, ITC, Thermodynamics

1 Introduction

Co-crystallization of nucleic acid/ligand complexes is frequently problematic, even on established biological systems. Much time, efforts, and material can be spared by using preliminary biophysical methods to assess a correct formation of the complex in conditions that are optimal for co-crystallization. Among these approaches, ITC is well-suited for analysis of nucleic acid interactions [1–4] and numerous examples include binding of small ligands [5–10], proteins [11–17] or nucleic acid–nucleic acid interactions [18–21]. A major advantage of ITC over other equivalent biophysical approaches (such as microscale thermophoresis, surface plasmon resonance, nuclear magnetic resonance, or steady-state fluorescence) is that ITC directly measures the heat that is released or absorbed upon a binding interaction [22–25]. Consequently, no labeling of the sample is needed and this sample can be directly used for crystallization experiments. In addition, ITC allows the simultaneous determination of all binding parameters in one single experiment: the affinity constant (K_a), the binding enthalpy (ΔH), the entropy change (ΔS), the Gibbs free energy change (ΔG), and

Eric Ennifar (ed.), *Nucleic Acid Crystallography: Methods and Protocols*, Methods in Molecular Biology, vol. 1320,
DOI 10.1007/978-1-4939-2763-0_9, © Springer Science+Business Media New York 2015

Uncorrected Proof

the stoichiometry (n). Unlike other techniques, ITC has a very wide range of applications: it is not subjected to any upper or lower molecular weight limits and can be used virtually in any buffers, provided the RNA/DNA buffer is rigorously identical to the ligand buffer. A real limitation of ITC is the requirement of large amount of material; however, this is obviously not an issue if the sample is dedicated for crystallization experiments since crystallization and ITC requirements are quite comparable. For most situations, the complex formed following the ITC titration can thus be used for crystallization by concentrating the sample about ten times. Finally, when complex RNA molecules are used, ITC also provides the opportunity to assess proper folding of the nucleic acids by monitoring the stoichiometry of the interaction: it is indeed commonly observed during large-scale preparation of RNA that a significant portion is misfolded and does not allow ligand binding (the same holds true for proteins). A suboptimal stoichiometry is a strong indicator of such a frequent problem.

In this chapter, we provide a detailed protocol for implementing ITC as a guide to improve co-crystallization of nucleic acid–ligand complexes. An example is presented with the titration of the aminoglycoside antibiotic lividomycin into the homodimeric HIV-1 genomic RNA dimerization initiation site (DIS) loop–loop complex (Figs. 1 and 2) [26–29], followed by co-crystallization experiments. This RNA/ligand interaction also provides a good example of salt-dependent binding, as frequently observed with nucleic acids. Indeed, as many nucleic acid binders, aminoglycosides are positively charged and unspecific interactions due to electrostatic interactions can be observed at low salt concentration, whereas excessively high salt concentration can even prevent the specific ligand binding. It might therefore be advisable to perform preliminary ITC experiments in various salt conditions in order to exclude (or at least reduce) unspecific binding while maintaining specific interactions. For a study of the binding of the HIV DIS RNA to lividomycin, ITC experiments performed in a low-salt buffer (25 mM of potassium chloride) clearly showed two binding events [28] and cannot be fitted with a single site model (Fig. 3). However, in higher-salt conditions (200 mM potassium chloride), only the specific binding was observed (Fig. 3). Such experimental conditions leading to a homogenous RNA/ligand complex should therefore be preferred in the search for crystallization conditions. In addition, such an experiment provides the opportunity to stop the ITC experiment close to the expected stoichiometry (close to 1.0 in this case, depicted by the black arrow in Fig. 3) in order to avoid the use of an excess of ligand. Crystallization experiments can then be performed in the most favorable conditions. In the following example, a more general situation is described with a significant nonspecific binding of a positively charged small molecule ligand and without any prior knowledge of the preferable

Uncorrected Proof

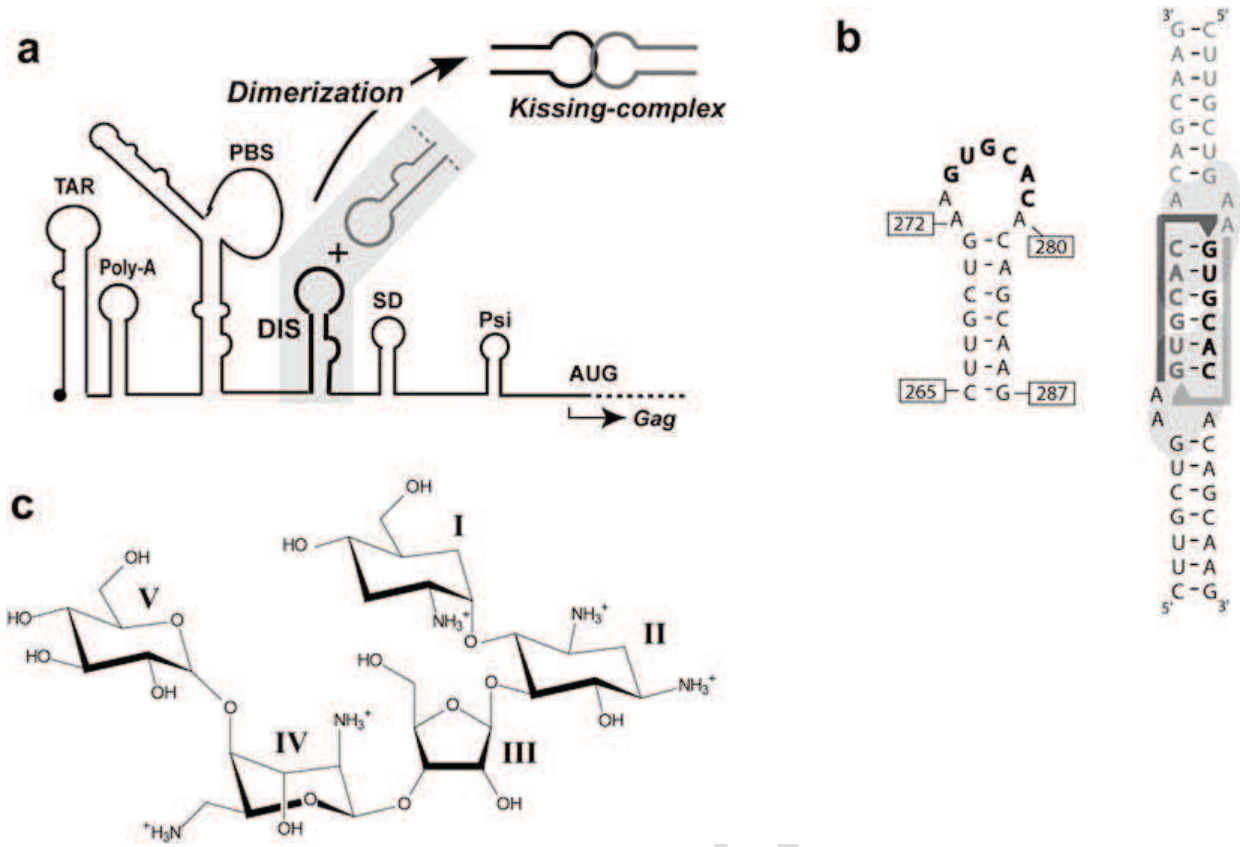


Fig. 1 HIV-1 DIS RNA and aminoglycosides. (**a**) Localization of the DIS in the 5′-untranslated region of the HIV-1 viral genomic RNA and dimerization mechanism through formation of a loop–loop "kissing" complex. (**b**) Sequence and secondary structure of the HIV-1 subtype F stem-loop used in this study. The self-complementary sequence is *bolded*. The loop–loop complex is depicted on the *right*, with the "A-site motif" circled in *light gray*. (**c**) Chemical structure of the aminoglycoside lividomycin

salt concentration to use. A brominated RNA (on the third uridine) is used in order to perform a MAD or SAD phasing following co-crystallization.

2 Materials

2.1 Instrumentation

1. iTC_{200} (Microcal-Malvern, Malvern, UK) or equivalent Isothermal Titration Calorimeter.
2. Crystallization robot Mosquito Crystal (TPP LabTech, Melbourn, UK) or equivalent.
3. NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) or equivalent.
4. Refrigerated Centrifuge.

2.2 Instrument Accessories

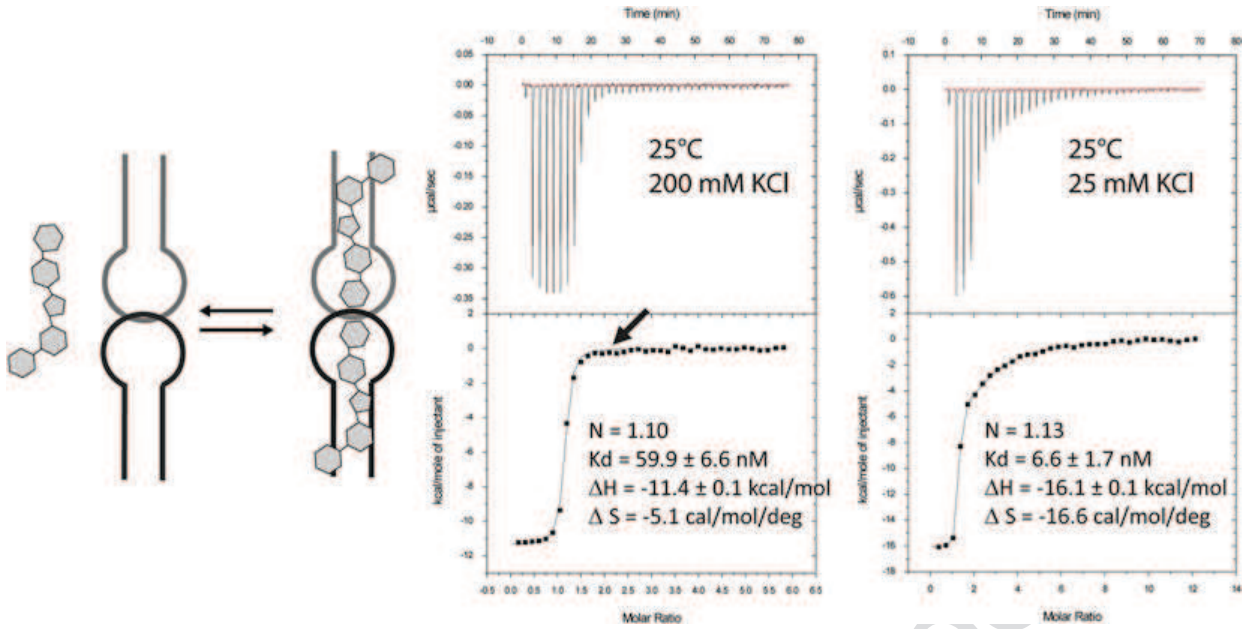
1. Hamilton loading syringes for iTC_{200} (Hamilton Company, Reno, NV, USA).
2. BD Luer-Lok 1 ml syringe.

Uncorrected

Fig. 2 Crystal structure of the free HIV-1 DIS RNA loop–loop complex (*left*) and bound to two lividomycin aminoglycosides (*right*). The two RNA stem loops are shown in *black* and *gray*, respectively

2.3 RNA Preparation

1. Chemically synthesized (purity >98 %) HIV-1 DIS 23-nucleotides RNA (Br-U3 modified): 5′-CU[BrU] GCUGAAGUGCACACAGCAAG-3′.
2. Dry bath capable of maintaining 90 °C.
3. KCl, powder.
4. 1 M $MgCl_2$. Store at 4 °C.



this figure will be printed in b/w

Fig. 3 Examples of titration of DIS RNA with lividomycin at 25 °C. On the *left*, the buffer contains 200 mM KCl and the experiment was performed with 12 μM RNA (in strands) and 350 μM aminoglycoside. On the *right*, the buffer contains 25 mM KCl and the experiment was performed with 6 μM RNA and 400 μM aminoglycoside. The *black arrow* depicts the saturation point where the ITC experiment might be stopped in order to perform crystallization trials in optimized RNA/ligand ratios

5. 1 M Sodium cacodylate, pH 7.0. Store at 4 °C.
6. 10× ITC Buffer: 1,500 mM KCl, 50 mM $MgCl_2$, 200 mM Sodium cacodylate, pH 7.0. Store at room temperature. Prepare 10 ml before use.
7. Amicon Ultra-4 centrifugal filter unit 10,000 MWCO (Millipore, Billerica, MA, USA).

2.4 Aminoglycosides Preparation

1. Lividomycin A Sulfate, powder. Store at 4 °C.
2. 1× ITC buffer: 150 mM KCl, 5 mM $MgCl_2$, 20 mM Sodium cacodylate, pH 7.0.

2.5 ITC Experiment

1. 1× ITC buffer: 150 mM KCl, 5 mM $MgCl_2$, 20 mM Sodium cacodylate, pH 7.0. Prepare 10 ml to wash the sample cell.
2. ITC EXPERT, the software controlling the iTC_{200} calorimeter.
3. Microcal Origin software for ITC data analysis (or equivalent software, such as NITPIC [30] or AFFINImeter: http://www.affinimeter.com/, *see* **Note 1**).

2.6 Crystallization of RNA–Aminoglycosides Complexes

1. Natrix HT crystallization screening kit (Hampton research, Aliso Viejo, CA, USA). Store at 4 °C.
2. 96-well CrystalEX™ microplate, conical flat bottom (Corning, Corning, NY, USA) or equivalent.

Uncorrected Proof

3. Crystal Clear Sealing Films (Hampton Research) or equivalent.
4. Amicon Ultra 0.5 ml centrifugal filter unit 10,000 MWCO (Millipore, Billerica, MA, USA).

3 Methods

3.1 RNA and Aminoglycosides Preparation for ITC Experiment

1. In order to prepare 300 μl of RNA at 120 μM for ITC experiment, RNA is refolded in water. To prevent the formation of duplex dimers and favor the hairpin fold, RNA concentration at this step must not exceed 3 μM. RNA is aliquoted into 12 Eppendorf tubes, each tube containing 1 ml of RNA at 3 μM in water.
2. Heat the RNA sample for 5 min at 90 °C and snap cool in water at 0 °C.
3. Add 111 μl of 10× ITC buffer in each Eppendorf tube. The RNA will then be in 1× ITC buffer.
4. Concentrate the RNA solution to about 300 μl with an Amicon Ultra-4 concentrator: centrifuge 10 min at 7,500 rpm at 10–12 °C. Higher temperature should be avoided since it might promote the hairpin to duplex conversion during the RNA concentration [31]. [AU2]
5. Read the absorbance at 260 nM using a NanoDrop spectrophotometer. An RNA concentration of 120 μM corresponds to ca. 20 OD.
6. Prepare 1 ml of lividomycin in 2.4 mM of 1× ITC buffer. [AU3]

3.2 Designing the ITC Experiment

3.2.1 Experimental Considerations

In an ITC experiment, the calorimeter directly measures the heat evolved in liquid samples as a result of mixing precise amount of reactants. Heat is produced (exothermic interaction) or absorbed (endothermic interaction) as a result of the interaction between the ligand and the sample molecule (*see* **Note 2**).

ITC monitors these heat changes by measuring the differential power applied to the cell heaters required to maintain a small temperature difference between the reference and sample cells as the binding partners are mixed. This differential power compensation per ligand injection is the raw signal recorded by the experiment. Each ligand injection results in a heat pulse that is integrated with time and normalized for concentration. This will produce a titration curve of kcal/mol vs. molar ratio (ligand/sample). The resulting isotherm is fitted to a binding model to obtain the affinity constant ($K_a = 1/K_d$), stoichiometry (N), and enthalpy of interaction (ΔH). The Gibbs free energy change (ΔG) and the entropy change (ΔS) are obtained via the thermodynamic relationships:

Uncorrected Proof

$$\Delta G = -\mathrm{R\,T}\ln K_a = \mathrm{R\,T}\ln K_d$$

$$\Delta G = \Delta H - T\Delta S$$

The appropriate concentration of the component placed in the sample cell (M_T) depends on the binding affinity, the stoichiometry, and heat of binding ΔH [32]. In order to derive accurate binding constant and ΔH, the Wiseman coefficient $c = N \cdot K_a \cdot M_T$ should be larger than 1 but smaller than 1,000 (preferably between 10 and 500) [33, 24].

In practice, required concentrations of sample in the cell for a "traditional" ITC experiment are ranging from 5 to 100 μM. The molar concentration of the ligand in the injection syringe is typically 10–20 times higher (and even higher in the case of low c value experiments [34]). This will ensure that the cell material will become saturated by the end of the titration experiment. As a consequence, sample requirement in order to perform an ITC experiment in ideal conditions ($c \sim 100$) might be significant for biological system with a low affinity. This should however not be an obstacle for ITC experiments performed prior crystallization.

The key to successful ITC experiments, one never insists too much on that, is to minimize the heat signal arising from mismatches between the solutions that are mixed (*see* **Note 3**).

3.2.2 Instrument Parameters

On the iTC_{200} microcalorimeter, the titration syringe holds up to 40 μl. Typical injections must not exceed 2 μl and must not be lower than 0.2 μl. In an ITC experiment performed prior crystallization, sample concentrations are large and rather small injections (~1 μl) should be sufficient to ensure a satisfactory signal/noise ratio and a correct sampling of the nucleic acid/ligand saturation. It is recommended to perform a first small injection (0.3–0.5 μl) that may be disregarded in data analysis (*see* **Note 4**). The number of injections per experiment depends on the volume of the injections and should be set so that the sample in the cell will be saturated at the end of the titration. Usually, the complete volume of the syringe is injected into the cell. However, it might be appropriate to stop the titration when the nucleic acid is saturated with the ligand in order to perform crystallization trials with the minimum excess of ligand (Fig. 3). The injection time should be set to twice the injection volume in μl, thus corresponding to a 0.5 μl/s injection speed (i.e. 2 s for 1 μl injections). The spacing between the injections is generally set to 120 s by default, which is often adequate to ensure a correct return to the baseline after each injection. However, the equilibration time is strongly dependent on the rate constants k_{on} and k_{off} for each system [35] and it is not uncommon to have to seriously increase this spacing value to 240 s (even much higher for very slow systems [15]). The filter value corresponds to

Uncorrected Proof

the integration time that will be used to produce a single data point. Usually set to 5 by default, the filter period might be set to 3 for a better sampling of each peak in view of performing a kinetic analysis of the thermodynamic data [35].

Regarding the calorimeter setup, the feedback mode should be set to "high gain," which corresponds to the fastest response time. The "Reference Power" value corresponds in practice to the baseline (in μCal/s). In the high gain mode, the instrument's reference power ranges from 0 to 12.25 μCal/s. The anticipated size and direction of the titration peaks will determine the baseline setting. For exothermic interactions, a large reference power setting is required (~5 μcal/s) and for endothermic interactions, a small reference power setting is required (0.5 μCal/s). It is essential to set the reference power so that the signal generated by each peak remains positive at any time. The stirring speed is set to 1,000 rpm by default (*see* **Note 5**).

Regarding the cell temperature, the instrument's theoretical operating range is 2–80 °C but most of the experiments are performed between 4 and 37 °C. The temperature of the experiment is often an overlooked parameter that might be very useful to exploit in order to improve the results through an optimization of the Wiseman c value by playing on two parameters: ΔH and K_d. Indeed, most biomolecular interactions are characterized by temperature dependence of the ΔH due to the heat capacity change of the event:

$$\Delta C_p = \left(\partial \Delta H / \partial T\right)$$

This temperature dependence is generally negative and associated with a local folding following binding [36]. It is a valuable tool for optimizing the study of any binding event: for an exothermic reaction, ΔH will be more negative at 35 °C than 25 °C and, consequently, the signal will be increased at 35 °C. Regarding the effect on the affinity, the K_d of an exothermic reaction should decrease at lower temperature (in agreement with van't Hoff's relationship), which might improve data in systems where the c value is too low.

3.3 Performing the ITC Experiment

1. Fill the reference cell with ultra pure water using a 0.5 ml glass Hamilton syringe.
2. In the instrument controls tab, set the temperature at 12 °C in "Thermostat Control," and click "Set Jacket Temp." This temperature is selected in order to improve the affinity of the aminoglycoside for the DIS RNA and to reduce any potential hairpin to duplex conversion of the RNA in presence of magnesium cations and rather high RNA concentration [31].
3. Rinse the sample cell, two or three times, with 1× ITC buffer.
4. Gently load the sample cell with 300 μl of DIS F BrU3 RNA at 0.120 mM. The small excess should be removed. The final

Uncorrected Proof

sample volume in the cell is ca. 280 μl. This step is delicate since no bubble should be present in the sample cell.

5. To load the titration syringe, place a microcentrifuge tube with at least 70 μl of lividomycin sample at 2.4 mM. Place the syringe titration above the microcentrifuge tube so that the tip of the syringe is immerged in ligand solution. Manual loading of the syringe is by far much safer than the semi-automatic procedure, especially if the sample is viscous (i.e. with a buffer containing some glycerol). Connect the BD syringe and tubing to the titration syringe and load the sample manually by aspiration into the titration syringe (before aspiration, make sure that the plunger is in the open position). When the sample gets to the plunger, click "Close Fill Port" in the instrument controls tab; this will move the plunger tip down so that it blocks the fill port. Perform the "Purge/ReFill" procedure to remove small air bubbles in the syringe. Unplug the BD syringe from the titration syringe.
6. Place the titration syringe into the sample cell.
7. Set the reference power to 5 μCal/s.
8. Set the initial delay to 100 s and the stirring speed to 1,000 rpm.
9. Set the number of injections to 40, the sample cell concentration to 0.12 mM and the concentration in the syringe to 2.4 mM.
10. Set the titration parameters (the number and volume of the injections as well as the spacing between two injections) in the "Injection Parameters" window into the "Advanced Experimental Design" tab. Set the first injection volume to 0.3 μl and the next injections volume to 1 μl. Set the spacing between each injection to 180 s.
11. Press the "Start" button to begin the experiment. The experiment will start when both cells are thermostated at 12 °C.
12. At the end of the titration, recover the RNA/ligand sample, wash the sample cell, reload the sample cell with 1× ITC buffer, reload the syringe with lividomycin solution and perform a blank experiment using the same parameters. A signal (here endothermic) will be produced due to the dilution of the ligand into the buffer. This signal should be taken into account for subsequent data analysis.

3.4 ITC Data Analysis

ITC data analysis using the Microcal Origin7 software will be described here.

Start the Origin software. The program opens and displays the rawITC plot window.

1. Click on "Read Data" to open the titration experiment (file with .itc extension). Origin will integrate the injection peaks

Uncorrected Proof

and a new "Delta H" window will appear (Fig. 4, step 1). It is possible at this stage to modify the concentrations of your samples by using the "Concentration button."

2. The software will automatically generate a baseline important for peak integration. If required, it can be manually adjusted by using the "adjust integration" button in the "rawITC" window in the "window" menu.
3. Go to the "rawITC" window.
4. Click again "Read data."
5. Select the blank file generated by injection of lividomycin in the buffer.
6. Remove aberrant data points and/or adjust the baseline if necessary.
7. Go to the "Math" menu and select "linear regression." A red line will be drawn (Fig. 4, step 2).
8. Double click on the "1" button located on the top left corner of the window (red arrow in Fig. 4, step 1). A new window labeled "layer 1" appears, with two lists: "available data" on the left and "Layer content" on the right. The "Layer content" list contains two items: one "xxx_ndh" corresponding to integrated heats of the blank experiment and a "linearfit1_xxxndh" corresponding to the linear regression.
9. Select data corresponding to integrated heats ("yyy_ndh") generated by the injection of ligand onto the RNA and click on the right arrow placed between both lists. The data set will then appear in the "Layer content" list. Click "OK" on the "Layer 1" window and on the "Rescale" button on the button bar of the Origin window (red arrow in Fig. 4, step 2). Integrated heats of both the ligand/RNA and ligand/buffer experiments are displayed (Fig. 4, step 3).
10. Go to the "Math" menu and select "Simple math." A new "Math on/Between data Set" window appears.
11. In the "Available data" list, select the "yyy_NDH" item and click on the right arrow close to "Y1."
12. In "operator," write the minus "-" operation.
13. In the "Available data" list, select the "linearfit1_xxxndh" item and click on the right arrow close to "Y2" (Fig. 4, step 4). Press "OK" in the "Math on/Between data Set" window. This will subtract the blank experiment to the RNA/ligand titration.
14. Double click on the "1" layer button located on the top left corner of the window. In the "Layer content" list, select "xxx_ndh" and "linearfit1_xxxndh" items and click on the left window (Fig. 4, step 5). This will remove the linear regression and

Uncorrected Proof

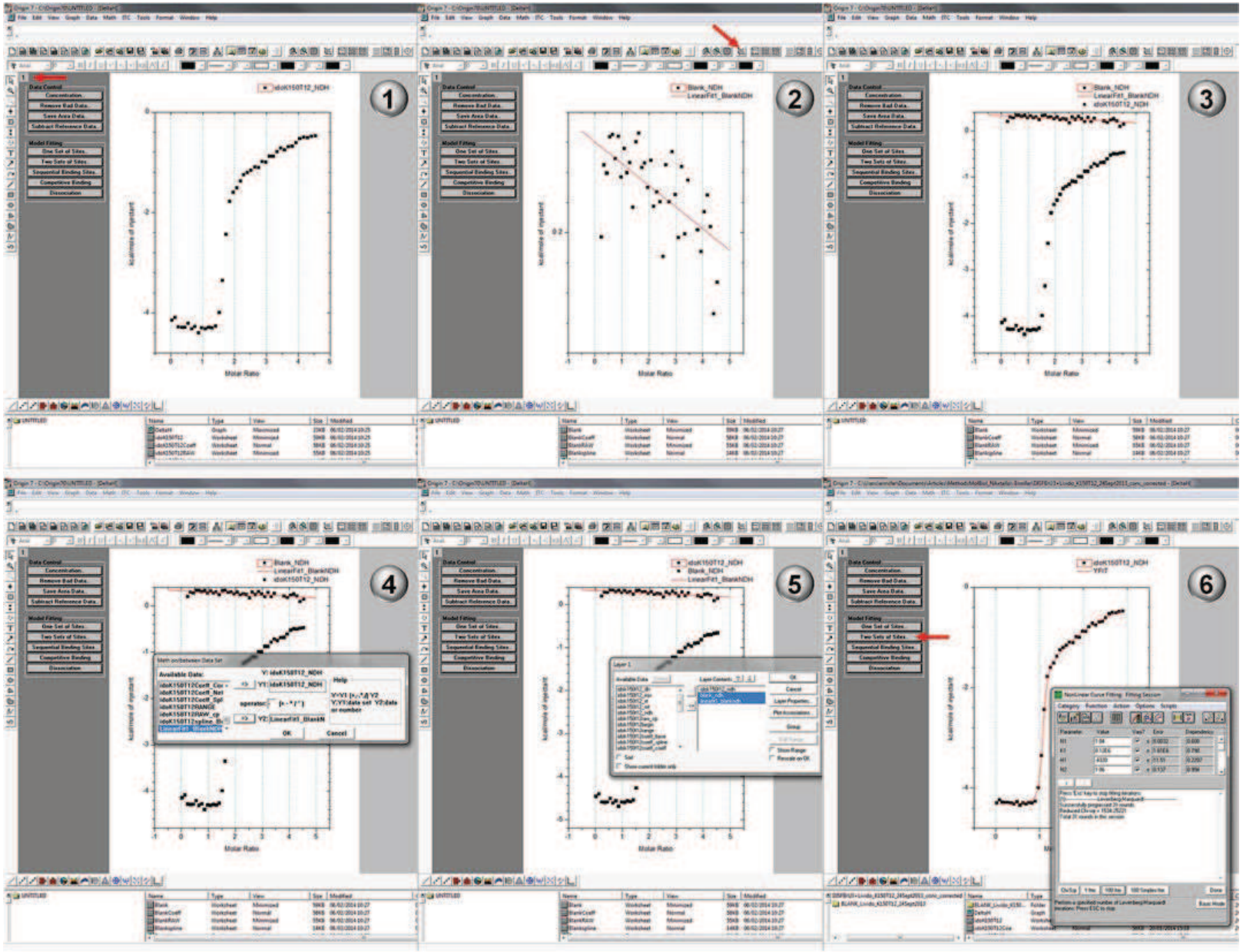


Fig. 4 Screenshots of six consecutive steps during ITC data processing with Microcal Origin

integrated heats of the blank experiment from graph in the main window. Click on the "Rescale" button. Now it is time to fit the data using an available model.

15. Remove the first injection by clicking on the "Remove Bad Data" button and a double click on the data point to be removed.
16. On the left part of the Origin main window, select the fitting model that should be used. Here the "Two Sets of Sites" model should be used since both specific and nonspecific interactions are observed (red arrow on Fig. 4, step 6).
17. Click on the "100 Iter" button to start the fitting. Repeat this until the reduced chi square remains constant (Fig. 4, step 6). The error should be smaller than the fitted value. The correlation between each fitted parameter is indicated in "Dependency": lower is better.
18. Once the fitting is finished, the stoichiometry (N), the association equilibrium constant (K), enthalpy change (ΔH), and entropy change (ΔS) values will be displayed next to the plot. Since the "two sets of sites" model was used, thermodynamic data are displayed for the two different binding sites. The specific binding site should have a stoichiometry close to 1.0 and the best affinity.
19. Save the project in the "File/Save Project" menu.
20. A publication-quality figure can be obtained in the "ITC/Final Figure" menu. Be careful: the baseline cannot be changed anymore after this step! It is possible to display data from both the RNA/ligand titration and the blank experiment on this figure by selecting the desired items on the "1" and/or "2" layers button located on the top left corner of the window (Fig. 5).

3.5 Considerations About Obtained Thermodynamic Parameters

In the example presented here, parameters for two binding sites are found (Fig. 5). The binding site with the better affinity (here binding site n°1) corresponds to the specific binding of the aminoglycoside to the RNA dimer as observed in crystal structures (Fig. 2). As expected, a stoichiometry close to 1.0 is observed since each aminoglycoside binds independently to the RNA dimer. The Wiseman c value for this binding site is $N_1 \cdot (1/K_{d1}) \cdot [\text{RNA}] = 1 \times 8.1 \times 10^6 \times 120 \times 10^{-6} = 972$, which is in the high range of acceptable c values to determine accurately both the affinity and the binding enthalpy. The second binding site has a very weak affinity (~140 μM). It can be assigned to a nonspecific component of binding (here also with a stoichiometry close to 1.0). Since crystallization experiments will be carried out at 37 °C, affinity of this nonspecific binding site will drop close to 400 μM in crystallization conditions (this estimation can be obtained using the known heat capacity change for the DIS RNA/lividomycin interaction [28] and the van't Hoff's relationship). It should therefore not interfere with co-crystallization of the specific RNA/ligand complex.

Uncorrected Proof

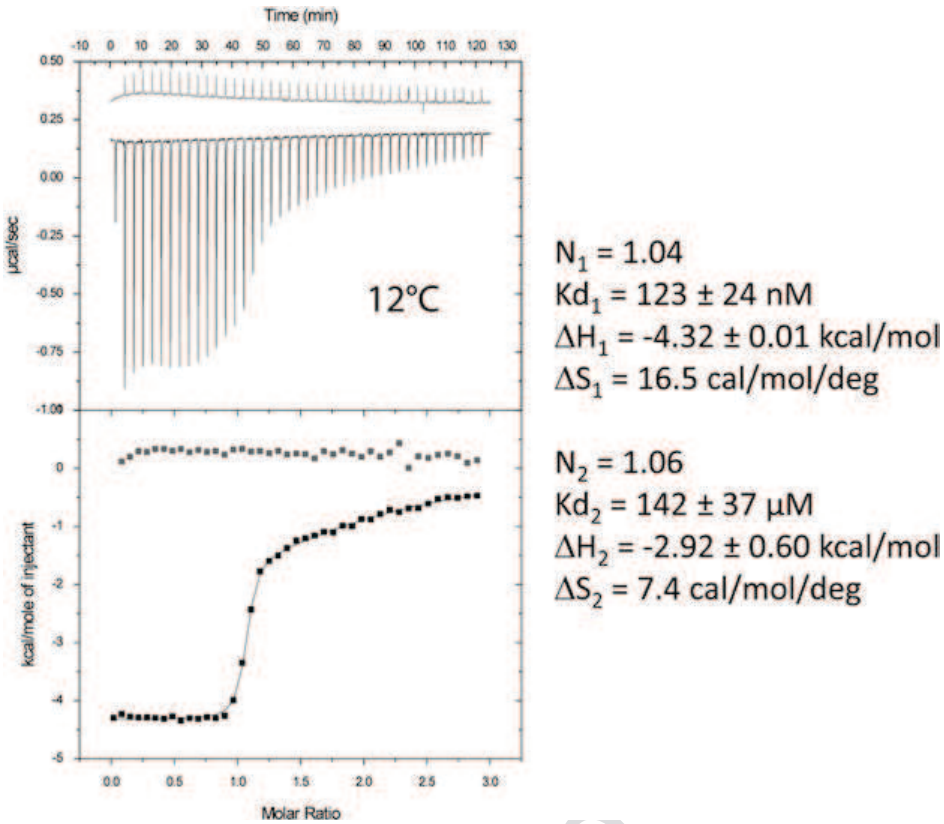


Fig. 5 Final figure obtained after the titration of DIS RNA with lividomycin at 12 °C. The blank experiment corresponding to injection of the aminoglycoside into ITC buffer is shown in *grey* and yield an endothermic dilution signal. Values obtained for the specific and the nonspecific binding sites after fitting the data using the "two set of sites" model are shown

Uncorrected

3.6 RNA–Aminoglycosides Crystallization

The RNA–lividomycin complex obtained after the ITC experiment is concentrated to a final concentration of 0.33–0.40 mM with an Amicon Ultra 0.5 ml concentrator (10,000 MWCO) at 4 °C.

3.6.1 RNA/Lividomycin Co-crystallization

1. Prior the crystallization trials, the Natrix HT kit should be placed at room temperature for at least 1 h. Dispense 0.16 ml of Natrix HT crystallization kit solutions in each well of the 96-well CrystalEX™ microplate.
2. Using the Mosquito crystallization robot, mix 0.2 µl of the RNA–lividomycin sample with 0.2 µl of the crystallization buffer in the top left (elevated) drop wells of the CrystalEX™ microplate. Alternatively, this step can be performed manually on a larger scale using 1.0 + 1.0 µl drops.
3. Seal the crystallization plate with the Crystal clear sealing film.
4. Store the crystallization plates at 37 °C overnight.
5. Check for crystals ~16 h later. After 24 h at 37 °C, place the crystallization plates containing crystals at 20 °C. Large monocrystals suitable for diffraction should grow from Natrix HT conditions E4, E8, G3, G4, G6, G11, G12 and H9, which corresponds to a remarkable score of 16.6 % successful crystallization rate (Fig. 6). Many other conditions lead to microcrystals.

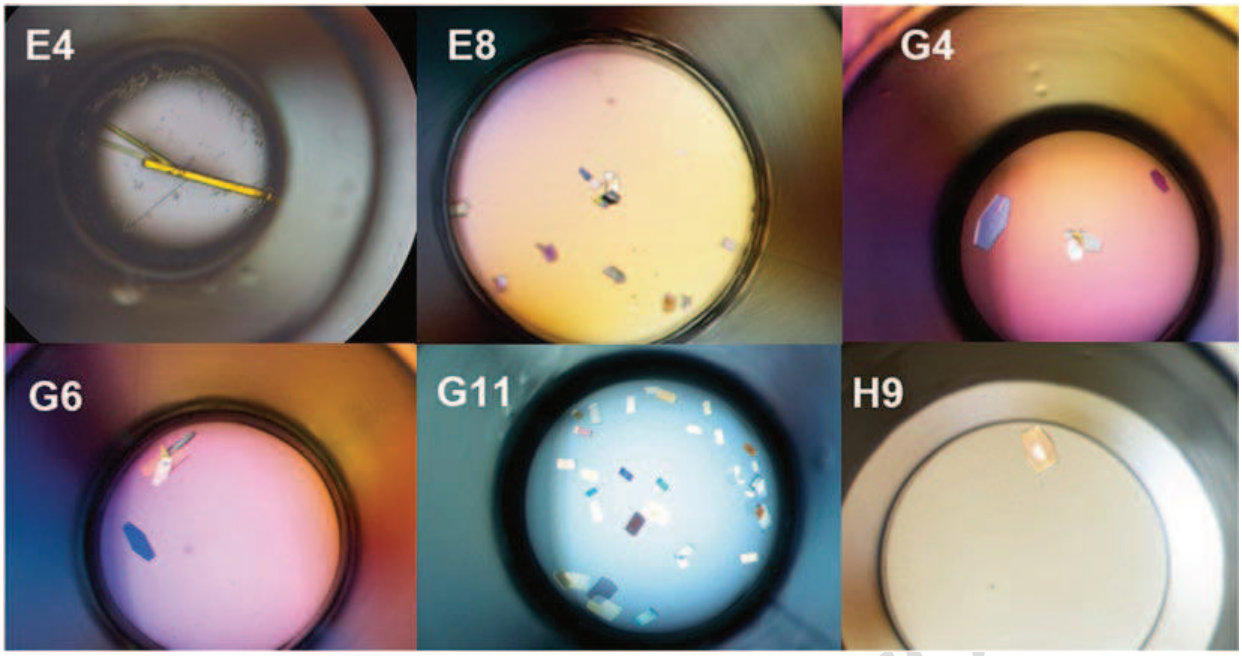


Fig. 6 Some of the crystals obtained using a Natrix HT kit with a DIS RNA loop–loop complex/lividomycin complex. The RNA/ligand complex was recovered from the ITC experiment shown in Fig. 5. The crystallization condition leading is indicated in the *top-left part* of each picture [AU4]

Uncorrected Proof

3.6.2 Crystals Harvest and Freezing

1. Harvest crystals that are deemed suitable for diffraction. Crystals grown in 2-methyl-2,4-pentanediol conditions do not require further cryoprotection. Otherwise glycerol can be used as a cryoprotectant for other precipitants. [AU5]
2. Flash freeze crystals in liquid ethane or propane (not directly in liquid nitrogen!).
3. Store the frozen crystals in liquid nitrogen.

3.6.3 X-Ray Diffraction and Data Collection

Diffraction of several crystals directly obtained from this screen were diffracted on the bending magnet X06DA beam line at the SLS synchrotron. Complete and good data sets were collected at the bromine K-edge up to 1.7 Å resolution and the structure could be solved ab initio using the SAD approach. [AU6]

4 Notes

1. Note that AFFINImeter offers many automatic procedures and is the only one allowing to obtain potential kinetic information by using *kinITC* [35].
2. It is extremely important of not being misled by situations where no heat signal can be detected. Indeed, two situations are possible: either there is no interaction at all (and a co-crystallization will be useless), or there is interaction, but the ΔH is null in the conditions of the experiment. The problem is thus to distinguish the two situations. Since the most important parameter influencing the ΔH of a reaction is the temperature

(which is characterized by the term $\Delta C_p = \partial \Delta H / \partial T$), one should immediately repeat the same experiment at, at least, another temperature or, preferably, at two temperatures, one above and the other one below the initial temperature T_0. In situations where $|\Delta C_p|$ is of order 0.5 kcal/mol/K, which is quite a common order of magnitude, an overall change by +10 K (from $T_0 - 5$ K to $T_0 + 5$ K) would induce an overall shift in ΔH by $+(\text{sign } \Delta C_p) \times 5$ kcal/mol, which would be readily detectable. Furthermore, the sign of ΔH on either side of T_0 should be opposite. Very importantly, following the van't Hoff equation $[\partial \ln K_d / \partial T = -\Delta H / (RT^2)]$, a null ΔH at some temperature implies that $\ln K_d$, and thus K_d too, are extremum at that temperature. If ΔC_p is positive (i.e. if ΔH increases when the temperature increases), a null ΔH is the mark of a minimum affinity (K_d is maximum) at that temperature. Conversely, if ΔC_p is negative (i.e. if ΔH decreases when the temperature increases), a null ΔH is the mark of a maximum affinity (K_d is minimum) at that temperature.

3. The mismatch may be produced by pH difference between the ligand solution and the sample solution, as well as any difference in salt concentration, buffer concentration or additives. To minimize these differences, the interactants should be prepared in identical solutions. In addition, because of the high sensitivity of the microcalorimeter to any temperature change, it is highly recommended to place the instrument far from any air flow, especially air conditioning, and in a room exposed to minimal temperature changes (avoid heaters, large windows or other instruments generating lots of heat such as freezers). To obtain steady baselines, the instrument should be cleaned regularly and the syringe should be in a very good state (even very slightly bent syringe will generate a significant noise). Obviously, the sample cell should be kept RNase free for experiments involving RNA.
4. Usually, this first injection is not correct due to artifacts introduced by the mechanical backlash of the screw mechanism in the injector syringe and by a frequent looseness in the Teflon tip in the syringe.
5. Reducing the stirring speed generally slightly improve the baseline stability but might introduce artifacts due to insufficient mixing.

Acknowledgments

This work was supported by grants from the Agence Nationale pour la Recherche (grant ANR-12-BS07-0007-03 "ClickEnARN") and the Agence Nationale de Recherches sur le SIDA (ANRS). The authors would like to thank Vincent Olieric (Paul Scherrer Institute/Swiss Light Source, Villigen, Switzerland), Natalia Markova and Peter Gimeson (Microcal-Malvern, Uppsala, Sweden).

Uncorrected Proof

References

1. Feig AL (2007) Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. Biopolymers 87:293–301
2. Feig AL (2009) Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry. Methods Enzymol 468:409–422
3. Gilbert SD, Batey RT (2009) Monitoring RNA-ligand interactions using isothermal titration calorimetry. Methods Mol Biol 540: 97–114
4. Salim NN, Feig AL (2009) Isothermal titration calorimetry of RNA. Methods 47:198–205
5. Pilch DS, Kaul M, Barbieri CM, Kerrigan JE (2003) Thermodynamics of aminoglycoside-rRNA recognition. Biopolymers 70:58–79
6. Kulshina N, Edwards TE, Ferre-D'Amare AR (2010) Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. RNA 16:186–196
7. Pikovskaya O, Polonskaia A, Patel DJ, Serganov A (2011) Structural principles of nucleoside selectivity in a 2′-deoxyguanosine riboswitch. Nat Chem Biol 7:748–755
8. Sokoloski JE, Dombrowski SE, Bevilacqua PC (2012) Thermodynamics of ligand binding to a heterogeneous RNA population in the malachite green aptamer. Biochemistry 51: 565–572
9. Ennifar E, Aslam MW, Strasser P, Hoffmann G, Dumas P, van Delft FL (2013) Structure-guided discovery of a novel aminoglycoside conjugate targeting HIV-1 RNA viral genome. ACS Chem Biol 8:2509–2517
10. Trausch JJ, Batey RT (2013) A disconnect between high-affinity binding and efficient regulation by antifolates and purines in the tetrahydrofolate riboswitch. Chem Biol 1(2): 205–216
11. Datta K, LiCata VJ (2003) Thermodynamics of the binding of Thermus aquaticus DNA polymerase to primed-template DNA. Nucleic Acids Res 31:5590–5597
12. Recht MI, Williamson JR (2004) RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex. J Mol Biol 344: 395–407
13. Bauer WJ, Heath J, Jenkins JL, Kielkopf CL (2012) Three RNA recognition motifs participate in RNA recognition and structural organization by the pro-apoptotic factor TIA-1. J Mol Biol 415:727–740
14. Freisz S, Mezher J, Hafirassou L, Wolff P, Nomine Y, Romier C, Dumas P, Ennifar E (2012) Sequence and structure requirements for specific recognition of HIV-1 TAR and DIS RNA by the HIV-1 Vif protein. RNA Biol 9:966–977
15. Bec G, Meyer B, Gerard MA, Steger J, Fauster K, Wolff P, Burnouf D, Micura R, Dumas P, Ennifar E (2013) Thermodynamics of HIV-1 reverse transcriptase in action elucidates the mechanism of action of non-nucleoside inhibitors. J Am Chem Soc 135:9743–9752
16. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, Baralle FE, Damberger FF, Allain FH (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. Nat Struct Mol Biol 20:1443–1449
17. Neuenfeldt A, Lorber B, Ennifar E, Gaudry A, Sauter C, Sissler M, Florentz C (2013) Thermodynamic properties distinguish human mitochondrial aspartyl-tRNA synthetase from bacterial homolog with same 3D architecture. Nucleic Acids Res 41:2698–2708
18. Vander Meulen KA, Davis JH, Foster TR, Record MT Jr, Butcher SE (2008) Thermodynamics and folding pathway of tetraloop receptor-mediated RNA helical packing. J Mol Biol 384:702–717
19. Reymond C, Bisaillon M, Perreault JP (2009) Monitoring of an RNA multistep folding pathway by isothermal titration calorimetry. Biophys J 96:132–140
20. Salim N, Lamichhane R, Zhao R, Banerjee T, Philip J, Rueda D, Feig AL (2012) Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex. Biophys J 102:1097–1107
21. Zhang J, Ferre-D'Amare AR (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. Nature 500:363–366
22. Ladbury JE, Chowdhry BZ (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. Chem Biol 3:791–801
23. Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11:560–566

Uncorrected Proof

24. Velazquez Campoy A, Freire E (2005) ITC in the post-genomic era…? Priceless. Biophys Chem 115:115–124
25. Privalov PL, Dragan AI (2007) Microcalorimetry of biological macromolecules. Biophys Chem 126:16–24
26. Ennifar E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C, Dumas P, Walter P (2003) HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics. J Biol Chem 278:2723–2730
27. Ennifar E, Paillart JC, Bodlenner A, Walter P, Weibel JM, Aubertin AM, Pale P, Dumas P, Marquet R (2006) Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. Nucleic Acids Res 34:2328–2339
28. Bernacchi S, Freisz S, Maechling C, Spiess B, Marquet R, Dumas P, Ennifar E (2007) Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion. Nucleic Acids Res 35:7128–7139
29. Ennifar E, Paillart JC, Bernacchi S, Walter P, Pale P, Decout JL, Marquet R, Dumas P (2007) A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site. Biochimie 89:1195–1203
30. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal Chem 84: 5066–5073
31. Bernacchi S, Ennifar E, Toth K, Walter P, Langowski J, Dumas P (2005) Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. J Biol Chem 280:40112–40121
32. Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179:131–137
33. Tellinghuisen J (2005) Optimizing experimental parameters in isothermal titration calorimetry. J Phys Chem B 109:20027–20035
34. Tellinghuisen J (2008) Isothermal titration calorimetry at very low c. Anal Biochem 373: 395–397
35. Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, Disdier F, Baltzinger M, Dumas P (2012) kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. J Am Chem Soc 134:559–565
36. Spolar RS, Record MT Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science 263:777–784

Uncorrected Proof

NOTE D'APPLICATION

Microcalorimetry as a tool for structural biology

MICROCALORIMÉTRIE

STRUCTURE MOLÉCULAIRE

ANALYSE SANS MARQUAGE

BINDING AFFINITY

Eric Ennifar, Cyrielle Da Veiga, Dominique Burnouf, Joelle Mezher, Barbara Puffer-Enders and Philippe Dumas. Corresponding author: e.ennifar@unistra.fr
Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire, CNRS/Université de Strasbourg.

Introduction

Crystallization of complexes (protein/protein, protein/nucleic acid, protein/ligand, nucleic acids/nucleic acid, nucleic acid/ligand), even on well-characterized biological systems, are frequently tedious and either time or sample consuming. The success rate of complex crystallization can be significantly improved if a proper preliminary characterization of the complex using biophysical methods is performed. It well established that Dynamic Light Scattering (DLS) is key to evaluate sample crystallizability (1). Likewise, DSC thermofluor-based optimization strategy have been developed to facilitate protein/ligand crystallization (2). Isothermal Titration Calorimetry (ITC) is the "gold standard" technique for investigating molecular interactions and we show here that it can be a valuable technique to improve crystallization of complexes. ITC is a true in-solution technique that directly provides, in one single experiment, the complete binding profile between two molecules: binding affinity (Ka), enthalpy and entropy changes (ΔH and ΔS) and stoichiometry (N) between two molecules are obtained very accurately (3-5). A major advantage of ITC over other similar biophysical approaches is that it is not restricted by macromolecule upper or lower size limit, there is no buffer restriction, and, most importantly for structural studies, it does not required any labeling. In addition, we have recently shown that modern ITC apparatus and new processing methods also allow obtaining a complete kinetic description on more diverse systems than usually thought, ranging from simple ligand binding to complex RNA folding (6).

©2015 Malvern Instruments Limited

The main limitation of ITC is the relatively large amount of sample required for an experiment. This should however not be a bottleneck for structural biologists involved into NMR or X-ray crystallography studies since sample requirement are quite similar to ITC analysis. Because the sample is not damaged during the ITC experiment, it can then be recovered and concentrated for subsequent crystallogenesis experiment. Finally, ITC can also be used to assess the proper folding of a nucleic acid or a protein used in a complex. We will present here some examples of such ITC-guided crystallization. All experiments were carried out on Microcal™ iTC200 (Malvern Instruments Ltd).

Example I: Crystallization of the HIV-1 RNA Dimerization Initiation Site bound to an aminoglycoside antibiotic

The HIV-1 genomic RNA Dimerization Initiation Site (DIS) is a highly conserved sequence with a stem-loop structure. The loop contains a self-complementary sequence that promotes dimerization of the viral genome by forming a loop-loop (kissing-loop) complex between two DIS hairpins. Crystal structures of the DIS loop-loop complex (7) revealed unexpected similarities with the bacterial 16 S ribosomal aminoacyl-tRNA site (A site), which is the target of aminoglycoside antibiotics (8). Consequently, it was shown that HIV-1 DIS RNA also binds aminoglycosides and can thus be an attractive target for the rationale design of new drugs targeting the HIV-1 genomic RNA (9). Here we will show how ITC can be used to guide co-crystallization of a complex made with the aminoglycoside Lividomycin and the HIV-1 DIS RNA kissing-loop complex.

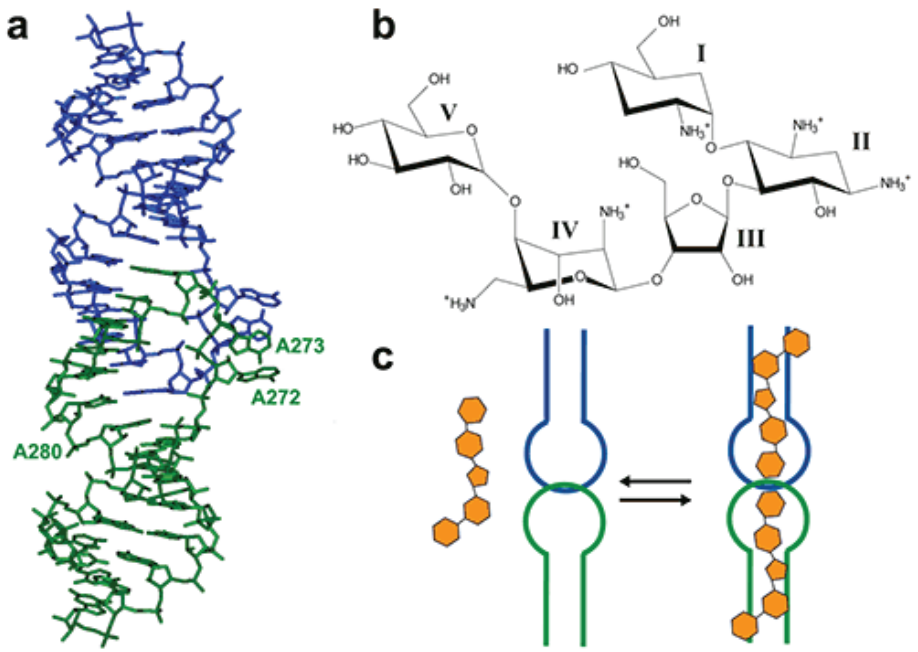


Figure 1: (a) Crystal structure of the HIV-1 genomic RNA Dimerization Initiation Site (DIS) kissing-loop complex (PDB ID 1ZC1, (10)). The two RNA stem-loops are shown in green and blue. Highly conserved adenines 5' (A272 and A273) and 3' (A280) of the 9-nucleotide loop are shown. (b) Chemical structure of the aminoglycoside antibiotic Lividomycin. (c) Schematic drawing of Lividomycin binding to the DIS kissing-loop complex with a 2:2 stoichiometry.

The DIS/lividomycin interaction provides an example of salt-dependent binding, as frequently observed with nucleic acids. As many nucleic acids ligands, aminoglycosides are positively charged and unspecific interactions due to electrostatic interactions are predominant at low salt concentration. However, excessive increase in salt concentration might prevent ligand binding due to electrostatic screening. One might therefore perform preliminary ITC experiments in various salt conditions in order to decrease unspecific binding while maintaining specific interactions. Experiments performed in a low-salt buffer (25 mM KCl, 2 mM MgCl2, 25 mM Na cacodylate pH 7.0) show two binding events and cannot be fitted with a single site model (Fig. 2A). Fitting with a two independent binding sites model shows a first very tight binding with a 1:1 stoichiometry, as expected (1 ligand for each stem-loop, 2 ligands for a kissing-loop complex), and a second binding site with an affinity in the low micromolar range that corresponds to unspecific interactions. This non-specific binding might therefore preclude formation of an homogenous RNA/ligand complex in crystallization conditions containing >100 mM concentration of the complex. However, in higher-salt conditions (200 mM KCl, 2 mM MgCl2, 25 mM Na cacodylate pH 7.0) only the specific and tight binding is observed (Fig. 2B). Such experimental conditions leading to a homogenous RNA/ligand complex should therefore be preferred in the search for crystallization conditions.

Misfolding is a frequent problem in structural studies of RNA. Here, the expected stoichiometry is obtained and these ITC experiments therefore also assess the proper folding of the RNA and, additionally, that the correct antibiotic concentration was used.

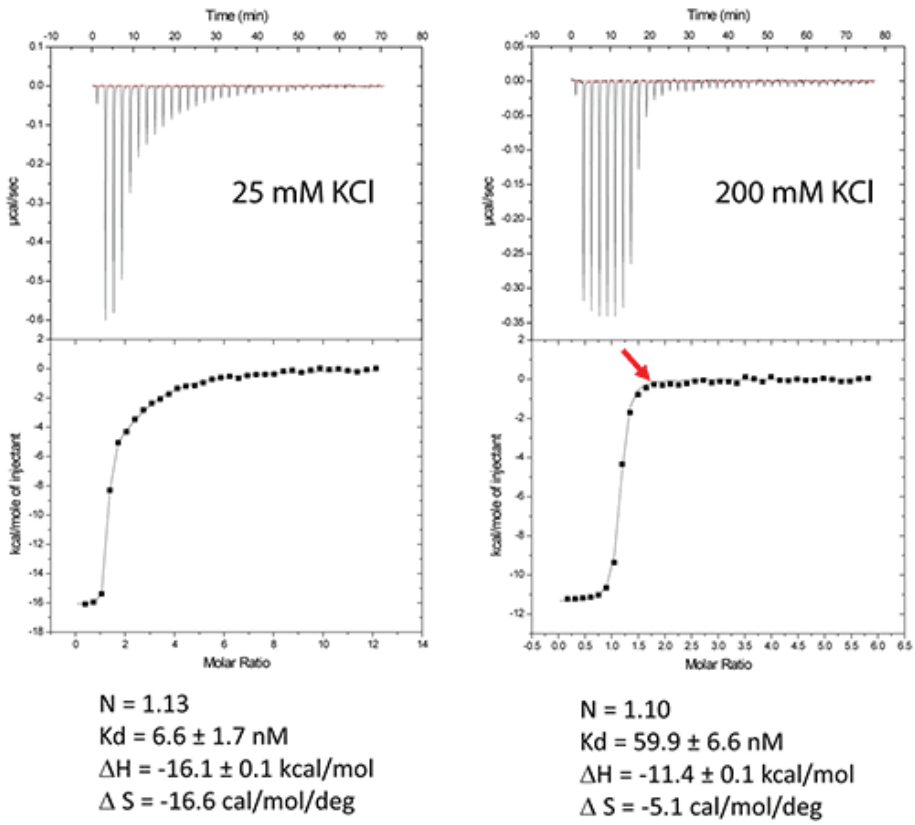


Figure 2: Titration of HIV-1 DIS kissing-loop complex (12 or 6 µM RNA in strands) with Lividomycin (350 or 400 µM) at 25°C in a buffer made with 25 mM KCl (left) or 200 mM KCl (right), respectively. Data were fitted with a two binding site model (left) or a single set of sites model using Origin analysis module for MicroCal (OriginLab Corp.). Thermodynamic parameters for the specific aminoglycoside binding site are shown.

Another interesting perspective of ITC is the real-time monitoring of the stoichiometry in order to stop the experiment when the desired RNA/ligand ratio is reached (supposedly ≥ 1.0). For example in the DIS RNA/Lividomycin titration at 200 mM KCl, the experiment could be stopped at the 12th injection when RNA is saturated with the ligand (red arrow). This approach should be considered in situation where the stoichiometry can be easily estimated, i.e. when the slope of the resulting sigmoid is rather steep (Wieseman coefficient c ≥ 100) and when the off-rate constant (koff) is supposedly slow. In such a case, this approach prevents the presence of an excess of ligand in the sample and crystallization is performed with an optimal RNA/ligand complex ratio.

We use in the following a standard example with non-optimized salt concentration for the DIS kissing-loop RNA/Lividomycin interaction (ITC buffer made with 150

mM KCl, 5 mM MgCl2, 20 mM Sodium Cacodylate, pH 7.0.). The RNA sequence used is a chemically-synthesized 23-nucleotide long and to the HIV-1 subtype A DIS containing a bromouridine on the third position(7) in order to facilitate crystal structure determination using the multiwavelength anomalous dispersion (MAD) approach. The RNA sample at 120 µM concentration (in RNA strands) is prepared in the ITC buffer and placed in the ITC cell. The syringe is filled with Lividomycin 2.4 mM in the ITC buffer. ITC analysis is performed at 12°C in order to improve the complex formation (since the interaction is exothermic) and preserve the RNA from possible degradation.

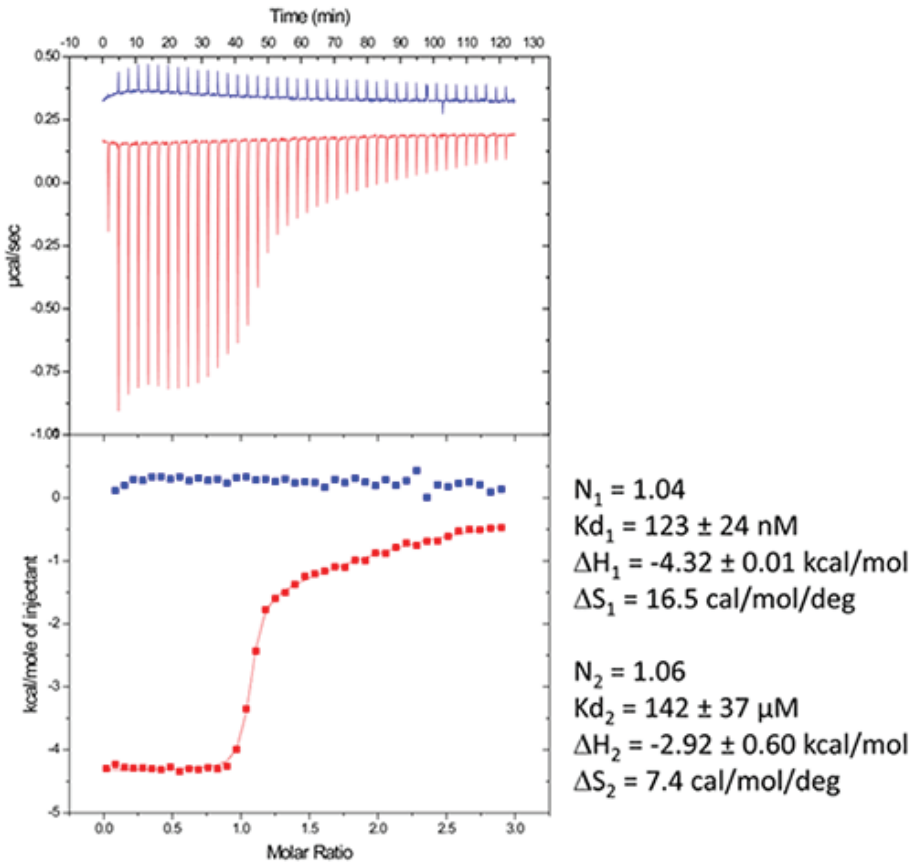


Figure 3: Raw data and binding isotherm for the interaction of [BrU-3] HIV-1 DIS RNA kissing-loop with Lividomycin (red). Experiments carried out in MicroCal iTC200. Injection of Lividomycin into ITC buffer is shown in blue. Thermodynamic parameters for specific (site 1) and unspecific (site 2) interactions are shown on the right.

Following the ITC experiment, the complex formed in the ITC cell is then recovered and concentrated about three times (at 4°C to preserve the RNA/ligand interaction) to about 400 µM using microcentrators. This solution is then mixed (1/1 volume) in crystallization trays using commercial crystallization sparse matrix screens (containing 96 different crystallization conditions) and a crystallization robot. The crystallization reservoir is filled with the sparse matrix solution and trays are placed for incubation at 37°C. Large monocrystals suitable for diffraction (up to 1.6 Å resolution using a synchrotron X-ray source) and structure

determination (Fig. 4) should grow after 16-24h incubation using Hampton Research® Natrix HT sparse matrix conditions (Fig. 4).

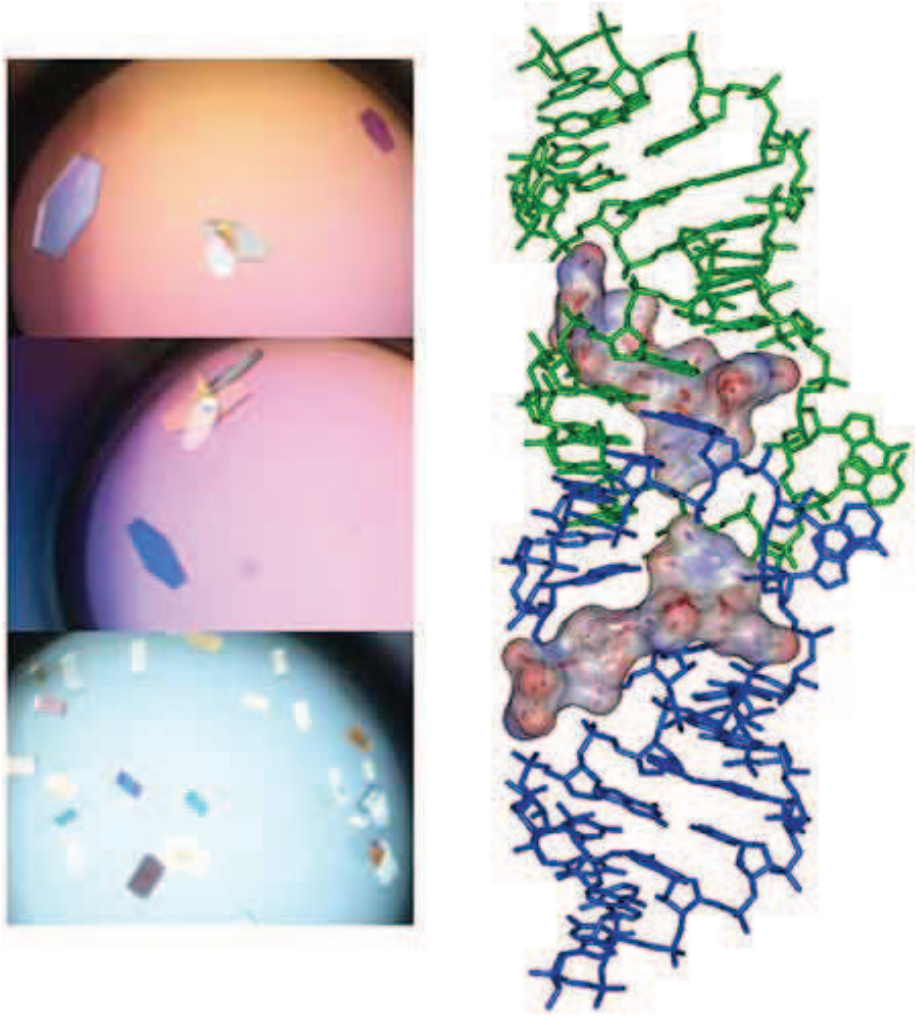


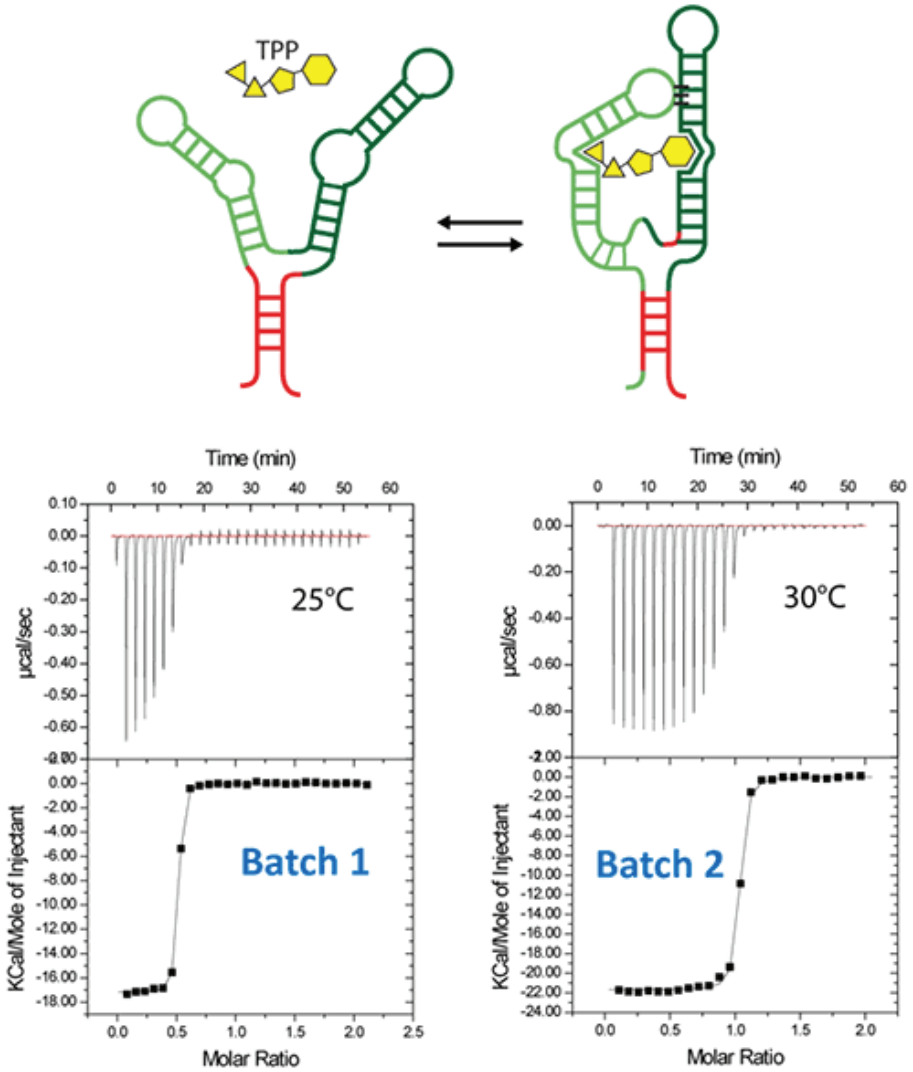
Figure 4: Crystals and X-ray crystal structure of BrU3-DIS kissing-loop complex RNA bound to Lividomycin antibiotic (11).

Example II: Crystallization of the E.coli thiamine pyrophosphate (TPP) riboswitch bound to its ligand.

Riboswitches are untranslated regions of messenger RNA, which bind specific metabolites. Upon binding, a conformational switch is observed, thus regulating the expression of proteins involved in the biosynthesis of riboswitch substrates (12). The E.coli ThiC TPP riboswitch responds to the coenzyme thiamine pyrophosphate (TPP), an active form of vitamin B1 (13) (Fig. 5, top). We use in this example a 85-nucleotides synthetic RNA fragment corresponding to the Ecoli TPP aptamer domain(6). A major problem of such a rather large RNA with complex secondary and tertiary structures is to recover properly folded molecules after overproduction, purification, and in vitro renaturation process.

In a first step, the following refolding protocol was tried (RNA Batch 1): after purification, the RNA was heat-deturated, flash-cooled on ice and placed in the ITC buffer made with Na cacodylate pH 6.5 50mM, potassium acetate 100 mM, magnesium acetate 5 mM. Following RNA refolding, an ITC experiment is performed at 25°C: RNA at 30 µM is placed in the ITC cell and the syringe is loaded with TPP 500 µM. This ITC experiment shows that only about half of the RNA is able to bind the TPP ligand, which is a strong indicator of serious misfolding problem in the RNA (Fig. 5, middle, see “Batch 1”). The RNA/TPP complex formed in the ITC cell is then recovered, concentrated and used for crystallization trials as described above for the HIV-1 DIS kissing-loop RNA. However, only clear drops are observed from these crystallization assays (Fig. 5, bottom left), in agreement with the hypothesis that a major fraction of the RNA was misfolded.

In a second step, the RNA folding protocol was changed: after heat denaturation at 90°C for 3 min in H_2O, the riboswitch was cooled to room temperature for 40 min. 10X ITC buffer was added to a final 1X concentration and the mix was further incubated for 40 min at room temperature. An ITC experiment performed on this second RNA batch (30 µM RNA, 500 µM TPP) show that this RNA refolding protocol lead to a fully active riboswitch since 100% of the RNA is able to bind the TPP ligand (Fig. 5, middle, see “Batch 2”). After concentration and setting up of crystallization drops, large monocrystals diffracting up to 5 Å resolution are obtained (Fig. 5, bottom right). This clearly ascertains the interest of using ITC as a quality control of the complex formation prior setting up crystallization trials.


TPP
Time (min)
µcal/sec
25°C
KCal/Mole of Injectant
Batch 1
Molar Ratio
Time (min)
µcal/sec
30°C
KCal/Mole of Injectant
Batch 2
Molar Ratio

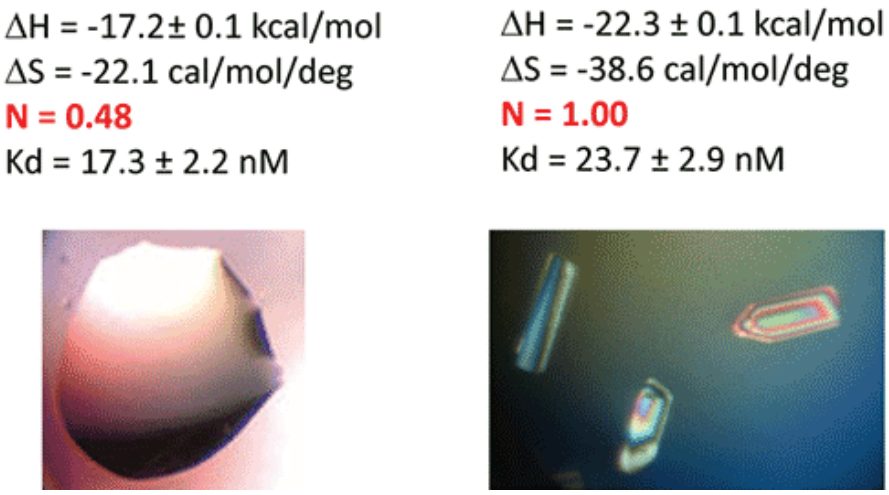


Figure 5: ITC-guided crystallization of the TPP riboswitch bound to its ligand. The secondary structure of the TPP riboswitch is shown on the top (each domain has a different color), the TPP ligand is shown in yellow. ITC experiments performed on a Microcal iTC200 microcalorimeter on RNA Batch 1 and Batch 2 are shown in the middle. The stoichiometry is highlighted in red. The result of crystallization trials is shown on the bottom.

Example III: Crystallization of the Pseudomonas aeruginosa sliding clamp bound to short peptide ligands.

In bacteria, the DNA polymerase processivity factor, or β sliding clamp, is a homodimer that confers high processivity to DNA polymerases and other factors involved in DNA metabolism, through their interaction via a small conserved peptide into a hydrophobic pocket of the clamp (14). Short synthetic peptides have been developed in order to interfere with this interaction to prevent binding of the sliding clamp to the DNA polymerase, thus acting as antibiotics (15). The purpose of this study was to cocrystallize P.aeruginosa clamp bound to a synthetic peptide to get structural insights into this interaction and to aid the structure-based rationale design of new antibiotic peptides with a better affinity. Prior crystallization, an ITC experiment was performed by placing the β sliding clamp (35 µM in Hepes pH 7.4 10 mM, NaCl 150 mM, EDTA 3 mM) in the ITC cell and the peptide (300 µM) in the syringe (Fig. 6a). This complex was further used for crystallization trials after concentration to 47 mg/ml using a microcentrator. Crystallization was set by mixing one volume of protein/peptide complex with one volume of a reservoir solution made with 0.1 M sodium Mes pH 6.0 100 mM, CaCl2 100 mM and 20-30% (w/v) polyethylene glycol (PEG) 400. After several days of incubation at 20°C, large monocrystals diffracting up to 2.2 Å using a synchrotron radiation were obtain (Fig. 6b) and the X-ray structure was solved (Fig. 7).

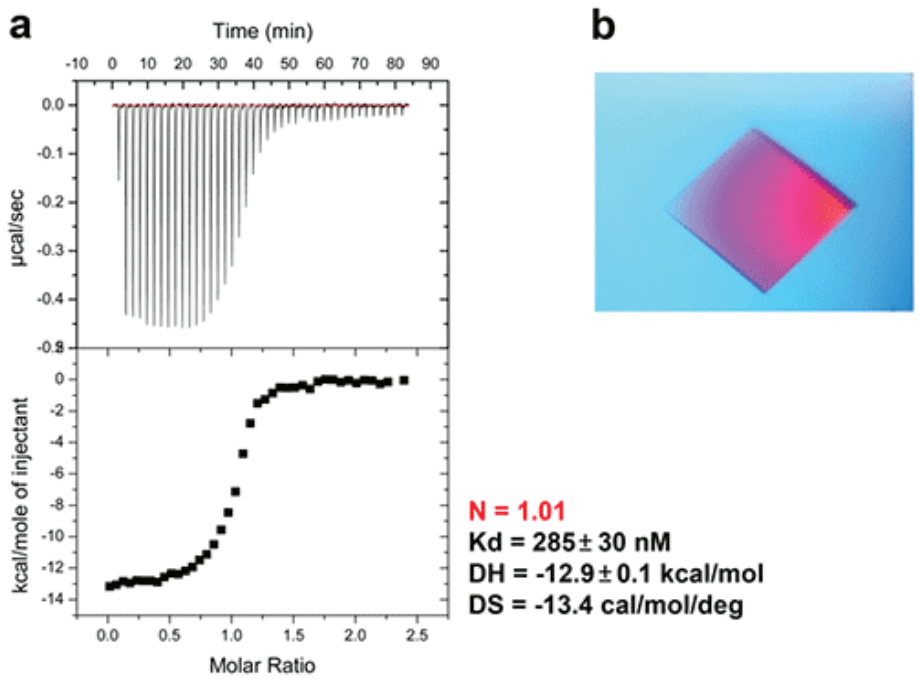


Figure 6: ITC-guided crystallization of the P.aeruginosa β sliding clamp bound to a short peptide. (a) ITC experiment performed at 30°C showing the peptide binding to the homodimeric sliding clamp. (b) Crystal of the peptide/sliding clamp complex obtained from the solution extracted from the ITC cell of MicroCal iTC200.

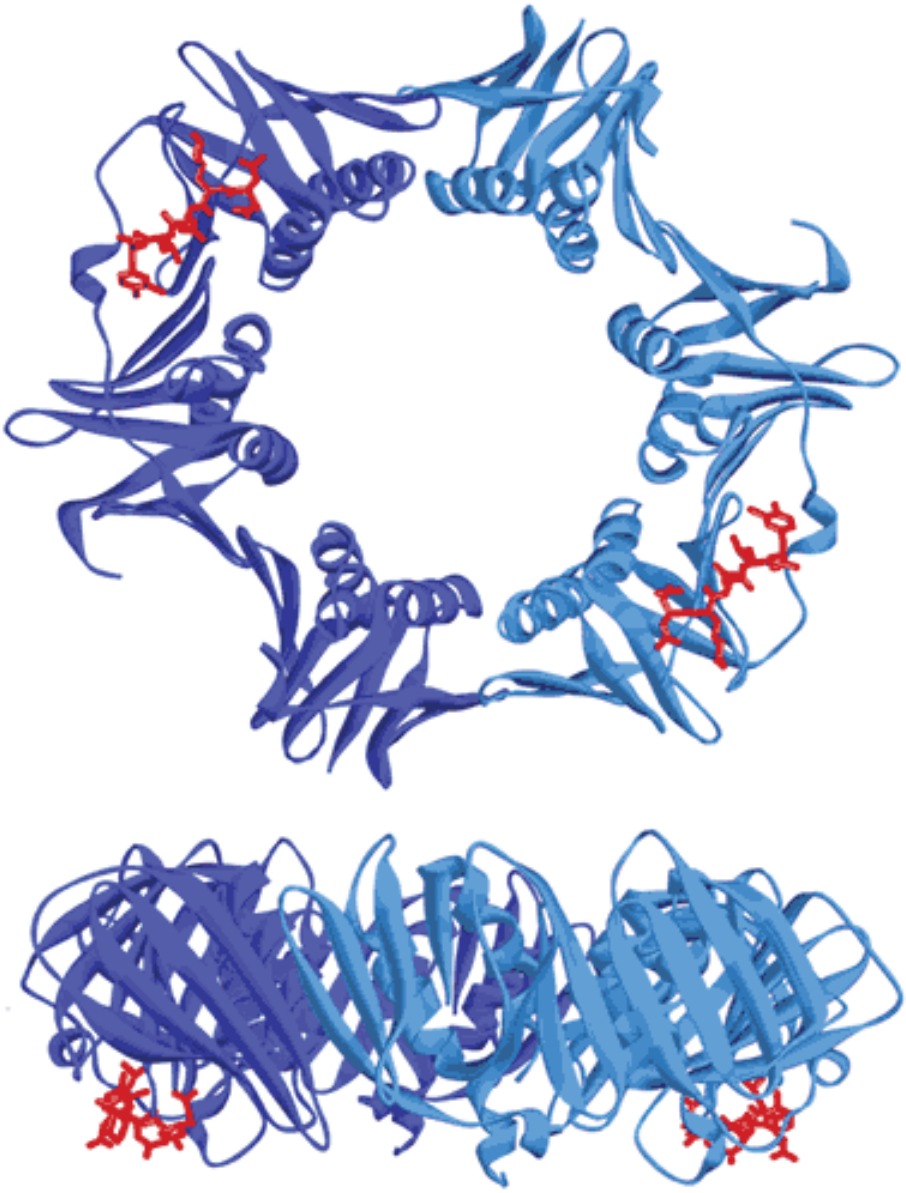


Figure 7: Two views showing the crystal structure of the replicative sliding clamp bound to a peptide with a 2:2 stoichiometry (PDB ID 4TSZ) (16). Protein monomers forming the homodimeric clamp are shown in light and dark blue. Synthetic peptide ligands are shown in red.

Conclusions

The ITC technique can be implemented in the crystallization workflow of biological macromolecules complexes and can be used as a guide to improve the success rate of crystallization:

- Sample requirement for structural studies are well-suited for ITC analysis
- Assessment of the real active protein and/or ligand concentration (provided by the observed stoichiometry)

- Real-time determination of an optimized protein/ligand ratio to be used for crystallization of the complex by monitoring the stoichiometry
- Help in optimization of the complex formation (sample homogeneity, affinity, specificity)
- Complete thermodynamic profile of the interaction is provided “for free”. One just needs to concentrate the sample after ITC analysis.

References

1. Ferre-D'Amare, A. R., and Burley, S. K. (1994) Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies, Structure 2, 357-359.

2. Dupeux, F., Rower, M., Seroul, G., Blot, D., and Marquez, J. A. (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples, Acta Crystallogr D Biol Crystallogr 67, 915-919.

3. Ladbury, J. E., and Chowdhry, B. Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions, Chem Biol 3, 791-801.

4. Leavitt, S., and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry, Curr Opin Struct Biol 11, 560-566.

5. Privalov, P. L., and Dragan, A. I. (2007) Microcalorimetry of biological macromolecules, Biophys Chem 126, 16-24.

6. Burnouf, D., Ennifar, E., Guedich, S., Puffer, B., Hoffmann, G., Bec, G., Disdier, F., Baltzinger, M., and Dumas, P. (2012) kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry, J Am Chem Soc 134, 559-565.

7. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site, Nat Struct Biol 8, 1064-1068.

8. Ennifar, E., Paillart, J. C., Marquet, R., Ehresmann, B., Ehresmann, C., Dumas, P., and Walter, P. (2003) HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics, J Biol Chem 278, 2723-2730.

9. Bernacchi, S., Freisz, S., Maechling, C., Spiess, B., Marquet, R., Dumas, P., and Ennifar, E. (2007) Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion, Nucleic Acids Res.

10. Ennifar, E., and Dumas, P. (2006) Polymorphism of Bulged-out Residues in HIV-1 RNA DIS Kissing Complex and Structure Comparison with Solution Studies, J Mol Biol 356, 771-782.

11. Ennifar, E., Paillart, J. C., Bodlenner, A., Walter, P., Weibel, J. M., Aubertin, A. M., Pale, P., Dumas, P., and Marquet, R. (2006) Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell, Nucleic Acids Res 34, 2328-2339.

12. Winkler, W. C., and Breaker, R. R. (2005) Regulation of bacterial gene expression by riboswitches, Annu Rev Microbiol 59, 487-517.

13. Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression, Nature 419, 952-956.

14. Burnouf, D. Y., Olieric, V., Wagner, J., Fujii, S., Reinbolt, J., Fuchs, R. P., and Dumas, P. (2004) Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases, J Mol Biol 335, 1187-1197.

15. Wolff, P., Olieric, V., Briand, J. P., Chaloin, O., Dejaegere, A., Dumas, P., Ennifar, E., Guichard, G., Wagner, J., and Burnouf, D. Y. (2011) Structure-based design of short peptide ligands binding onto the E. coli processivity ring, J Med Chem 54, 4627-4637.

16. Wolff, P., Amal, I., Olieric, V., Chaloin, O., Gygli, G., Ennifar, E., Lorber, B., Guichard, G., Wagner, J., Dejaegere, A., and Burnouf, D. Y. (2014) Differential modes of peptide binding onto replicative sliding clamps from various bacterial origins, J Med Chem 57, 7565-7576

Thermodynamics of the HIV-1 Vif protein binding to HIV-1 viral genomic RNA fragments

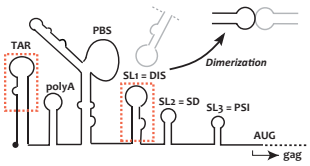
Joelle MEZHER[1], Séverine FREISZ[1], Lamine HAFIRASSOU[1], Philippe WOLFF[1], Yves NOMINE[2], Guillaume BEC[1], Christophe ROMIER[3], Philippe DUMAS[1] & Eric ENNIFAR[1]

1. Architecture et Réactivité de l'ARN, CNRS / Université de Strasbourg (France) 2. Oncoprotéines, Ecole de Biotechnologie de Strasbourg, Illkirch (France)
3. Département de Biologie et Génomique Structurales, IGBMC, CNRS / INSERM / Université de Strasbourg, Illkirch (France)

INTRODUCTION

The HIV-1 Vif protein plays an essential role in the regulation of the infectivity of HIV-1 virion and *in vivo* pathogenesis. Vif neutralizes the human DNA-editing enzyme APOBEC3 protein, an antiretroviral cellular factor from the innate immune system, allowing the virus to escape the host defense system [1]. Vif is packaged into viral particles through interactions the the viral genomic RNA [2,3]. It was shown that conserved and very structured regions such as the Tat-responsive element (TAR) or the genomic RNA dimerization initiation site (DIS) in the 5'-untranslated region of the genomic RNA are primary binding sites for Vif [4,5].

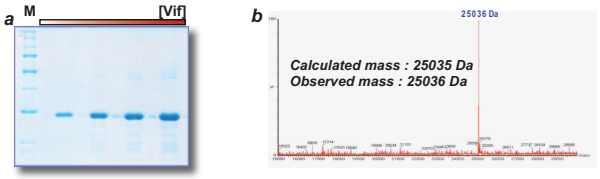
Here we purified milligram amounts of Vif protein. The quality of the protein was assessed by biophysical approaches. We then used isothermal titration calorimetry (ITC) to investigate HIV-1 Vif binding to short viral RNA fragments corresponding to TAR and DIS stem-loops to determine RNA determinants important for Vif binding.



Secondary structure of the 5'-untranslated region of the HIV-1 RNA genome. The 35-nucleotide DIS fragment and the 27-nucleotide TAR fragment used in this study are boxed.

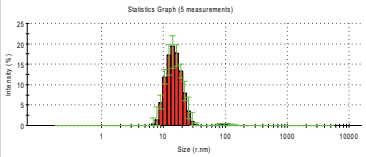
Expression, Purification and biophysical characterization of wild type HIV-1 Vif

Expression plasmids pETMCN-EATNH-Vif [6] were used to transform E. coli Rosetta 2 cells. Expression was induced by addition of 0.5 mM IPTG at 37°C, and the cells zere lysed in denaturating lysis buffer containing 6 M guanidine hydrochloride[7].After ultracentrifugation the supernatant was loaded on a Ni-NTA column; elution was performed by pH decreasing (from 8 to 5). Proteins where then renaturated by dialysis, decreasing guanidine hydrochloride concentration.

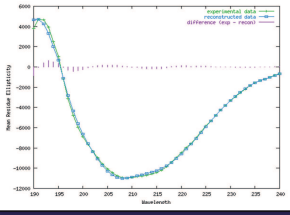


a) Coomassie-stained SDS Page gel showing fraction containing increasing concentration of Vif.
b) Mass spectrometry analysis of the purified Vif protein.

The described expression and purification procedure yielded ***17 mg of highly purified Vif*** **per liter of *E. coli* culture**.



Dynamic Light Scattering measurements were performed, showing that the protein is monodisperse (polydispersity index = 16%) and without agregates, with a mean hydrodynamic radius of 12 nm. This corresponds to polymers containing between 6 and 8 Vif proteins.



Far-UV CD spectra of the Vif protein was acquired between 186 and 260 nm. The secondary structure content was calculated with the program CDSSTR [8] using the reference database 4. The analysis yielded **20 % α-helix, 26 % β-strands, 23 % turns and 30 % disordered structure**, which is excellent agreement with recently published data [9,10], thus showing that the protein is properly refolded after denaturation.

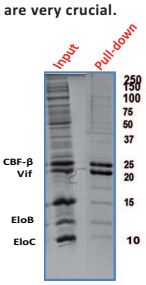
Perspectives

Recently, it has been shown that Vif recruits the transcription cofactor CBF-β to the ubiquitin ligase complex containing Cul5, EloC, EloB and a ring-box protein to target APOBEC3G for degradation. CBF-β which normally functions in concert with RUNX DNA binding proteins, allows the reconstitution of the recombinant six –protein assembly that elicits specific polyubiquitination activity with APOBEC3G. As a consequence, CBF-β is required for Vif-mediated degradation of APOBEC3G and for processing HIV-1 activity [11,12,13].

Methods of disrupting the CBF-β-Vif interaction might enable HIV-1 restriction and provide a supplement to current antiviral therapies that primarily target viral proteins. For this purpose, structural studies of the complex Vif-CBF-β are very crucial.

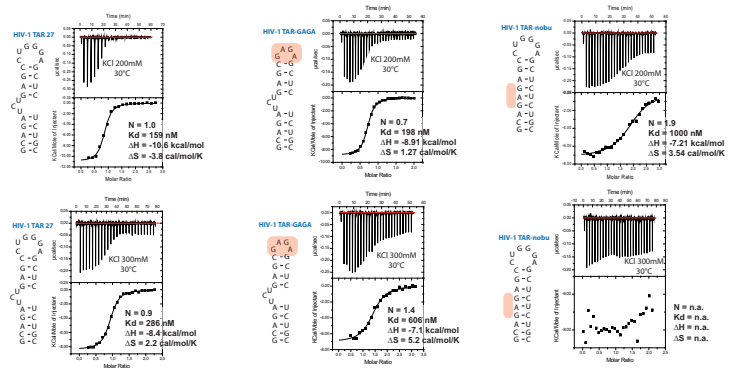
Because structural and functional analysis of full length Vif is limited by difficulty in obtaining suitable quantities of soluble Vif protein, we tried the co-expression with binding partners Vif-CBF-β-EloB-EloC[14].

Preliminary results showed that CBF-β can improve the solubility of Vif and we obtain a complex not prone to aggregate. The purification of this complex will facilitate future structural and biochemical studies of Vif functions.



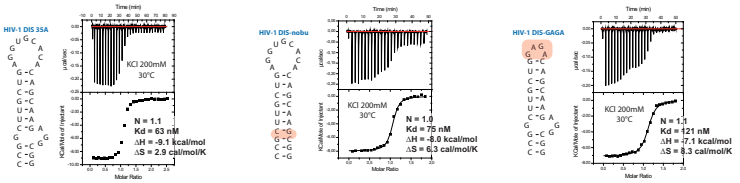
Thermodynamic characterization of the Vif protein binding to HIV-1 genomique RNA fragments.

Isothermal Titration Calorimetry (ITC) is the gold standard for measuring biomolecular interactions. ITC simultaneously determines all binding parameters (n, K, ΔH and ΔS) **in one single experiment** - *information that cannot be obtained from any other method*. We have use ITC to investigate Vif binding to the TAR and DIS stem-loops of the HIV-1 viral genomic RNA. DIS and TAR mutants with substitution of the apical loop by a GAGA tetraloop or with deletion of the internal bulge were used to define RNA determinants required for Vif binding.



ITC data obtained for Vif binding to the wild-type and mutant TAR sequences. Experiments were conducted at 200 mM KCl (top) but higher salt concentrations (300 mM KCl, bottom) were also used to get rid of unspecific electrostatic interactions.

We show that Vif interacts specifically with TAR and DIS RNA. An exothermic process is associated with the specific Vif binding, whereas a weak endothermic reaction is obtained using the *E coli* 30S rRNA sarcin/ricin loop used here as a negative control. Alteration of the TAR apical loop only moderately affects Vif binding, whereas **deletion of the lateral UCU bulge revealed that it is essential for Vif binding**. About Vif/DIS interactions, deletion of the DIS internal bulge does not affects Vif binding, whereas mutation of the apical loop (required for dimerization through formation of a kissing-loop complex) into a GAGA tetraloop significantly reduces Vif binding.



200 mM KCl 300 mM KCl	DIS-WT		DIS-nobu		DIS-GAGA		TAR-WT		TAR-GAGA		TAR-nobu	
N	1.1	0.8	1.0	0.7	1.1	1.3	1.0	0.9	1.7	1.4	2.0	n.d.
ΔH (kcal/mol)	-9.1	-7.1	-8.0	-5.9	-7.1	-5.3	-10.6	-8.4	-6.1	-7.1	-7.2	n.d.
ΔS (cal/mol/deg)	2.9	7.4	6.3	11.0	8.3	12.6	-3.8	2.2	9.5	5.2	3.5	n.d.
Kd (nM)	63	177	75	229	121	267	159	286	330	606	1000	n.d.

Table showing ITC data obtained for Vif binding to DIS and TAR sequences used in this study. Data obtained at 200 and 300 mM KCl are shown in green and blue, respectively.

REFERENCES

[1] Sheehy & al, Nature (2002)
[2] Dettenhofer & al, J Virology (2000)
[3] Zhang & al, J Virology (2000)
[4] Henriet & al, J Mol Biol (2005)
[5] Bernacchi & al, J Biol Chem (2007)
[6] Bieniossek & al, Nat Methods (2009)
[7] Yang & al, J Biol Chem. (1996)
[8] Whitmore & al, Biopolymers (2008)
[9] Bernacchi & al, Nucl Acids Res (2011)
[10] Gallerano & al, Amino Acids (2010)
[11] Zhang & al, Nature (2012)
[12] Jager & al, Nature (2011)
[13] Hultquist & al, JVI (2011)
[14] Zhou & al, PLoSOne (2012)

Fonds de dotation PIERRE BERGÉ
Sidaction

Structural analysis of viral infectivity factor of HIV-1 and its interaction with A3G:
Reconstitution of the E3 ubiquitin ligase complex using the coexpression strategy

Joelle MEZHER[1], Cyrielle DAVEIGA[1], Séverine FREISZ[1], Christophe ROMIER[2], Philippe DUMAS[1] & Eric ENNIFAR[1]

1. Architecture et Réactivité de l'ARN, CNRS UPR 9002/UDS, Strasbourg (France)

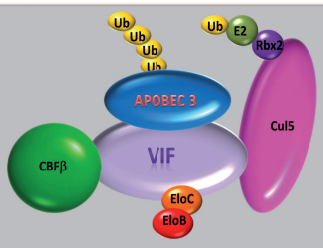
2. Département de Biologie et Génomique Structurales, IGBMC, UDS, CNRS, INSERM, Illkirch (France)

INTRODUCTION

The HIV-1 protein Vif (virion infectivity factor) plays an essential role in the assembly of virions and in the infectivity of the virus. Vif counteracts the action of the cellular protein APOBEC3G involved in viral defense [1]. Detailed molecular mechanisms that govern Vif/hAPOBEC-3 interactions remain however still unknown. In addition, the structural knowledge about hAPOBEC3 is limited to the C-terminal part of hA3G [2,3] and hA3F [4].

Interaction with Vif results in recruitment of APOBEC3G to an E3 ubiquitin ligase complex containing Cullin5, ElonginB , ElonginC and CBFβ, promoting APOBEC3G polyubiquitination and degradation and thereby damping APOBEC3G mediated cellular defences. Previous studies showed that the CBFβ is recquired for assembling a well ordered CRL5-Vif complex by inhibiting Vif oligomerization [5,6,7].

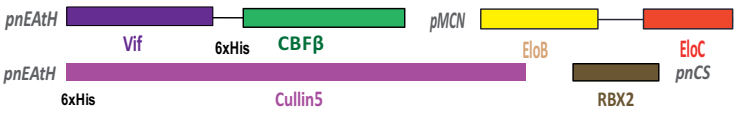
Recently, the structure of Vif-CBFβ-ELOB-ELOC-CUL5-RBX2 has been established by crystallography [8].



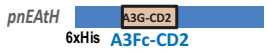
Constructions used in this study

The constructions used in this study are pET derived vectors used for coexpression (Collaboration with Christophe Romier). His-tagged CBFβ142 and full length Vif HXB2 are sub-cloned into vector pETMCN-EATNH(pnEATH) with a Lac promotor and resistance to Ampicillin while ELOB and ELOC were sub -cloned into a second vector pETMCN (pMCN) resistant to Spectinomycin.

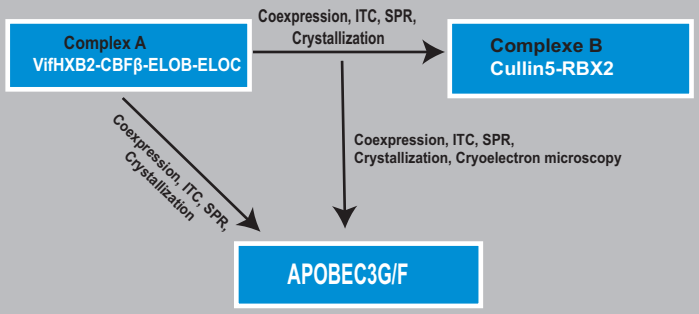
The complex A is formed by the coexpression of pnEAtH(VifHXB2-CBFβB142) and pMCN(ELOB-ELOC) and the complex B is formed by the coexpression of pnEAtH(CUL5) and pMCN(RBX2).



For solubility purposes, an A3G solubilization linker (residues 197-221) was inserted in the A3Fct gene (residues 218-373) [9] . A3Fc-CD2 gene was cloned into the pnEAtH vector containing an N-terminal His-tag.

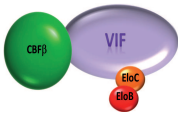


Objectives : Reconstitution of the Vif-CBFβ-ELOB-ELOC-CUL5-Rbx2 complex in interaction with APOBEC3G/F in order to define the structural determinants of HIV-1 Vif/APOBEC3 interactions



Expression and purification of complex A

The Vif, CBFβ and ELOB-ELOC constructs were co-expressed in Rosetta-2 (DE3) E.coli cells. Expression of the recombinant proteins was induced by addition of 0,5 mM IPTG at 15°C. After overnight induction, the cells were lysed in Phosphate buffer. After ultracentrifugation the supernatant was loaded on a Ni-NTA column; elution was performed by an Imidazole gradient from 40 mM To 600 mM Imidazole. The fractions were pooled , concentrated and the complex was subjected to size exclusion chromatography (Superdex200). The complex was concentrated to 10 mg/mL for crystallization.



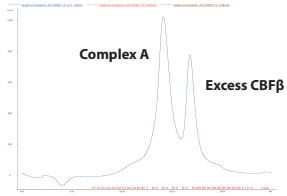


Fig 1 Size Exclusion chromatogram showing the purified fractions of the complex Vif-CBFβ142-ELOB-ELOC (67 Kda).

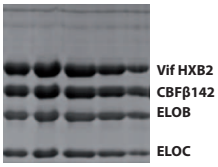
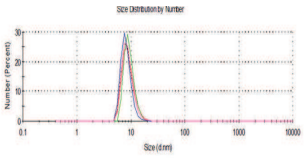


Fig 2 : Formation of the Vif-CBFβ-ELOB-ELOC complex in gel filtration.SDS-PAGE and Coomasie blue staining of Vif-CBFβ-ELOB-ELOC complex of the peak fractions from gel filtration.



Dynamic Light Scattering measurements were performed, showing that the complex A is monodisperse (polydispersity index=10%) and without aggregates, with a mean radius of 4 nm.

The complex is suitable for crystallization.

Expression and purification of complex B

The two plasmids pneaTHCullin5 and PNCSRBX2 were co-transformed into BL21 Rosetta2 E.Coli cells and expression was induced at 20°C with 0.5 mM IPTG. Clarified cell lysates were purified by IMAC and Thrombin protease was added to the resin after complete cleavage of His tag , the heterodimer of CUL5 and RBX2 was eluted from Nickel Resin and subjected to additional purification by a Superdex200 column. The Cleavage of the His tag was verified by Mass Spectrometry.

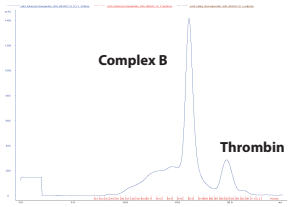


Fig 3: Size exclusion chromatogram showing the purification of the Complex CUL5-RBX2 (104 Kda).

Fig 4: SDS PAGE and Coomasie staining of CUL5-RBX2 complex of the peak fractions from gel filtration showing a pure complex

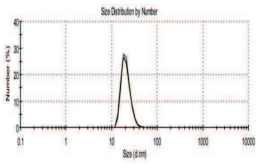


Fig 5: Dynamic Light Scattering measurements showed that the complex B is monodisperse with a mean radius of 10 nm. This value is overestimated due to the elongated form of the CUL5.

Expression and purification of APOBEC3F

A3Fc-CD2 was expressed in Rosetta-2 (DE3) E.coli cells. Cell cultures were induced by addition of 0,5 mM IPTG at 25°C for 18 hours. After cell lysis , the protein was purified on an Ni-NTA resin and then concentrated and purified furthermore on a Superdex200 column equilibrated in Buffer containing 10 mM Tris-HCl pH 7,5, 0,3 M NaCl, 1 mM DTT [9].

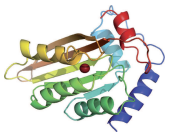


Fig 6: Ribbon diagram of A3Fc-CD2. The chain is coloured in a rainbow gradient from red (N terminus) to blue (C terminus). The catalytic zinc atom is labelled and shown as a grey sphere.

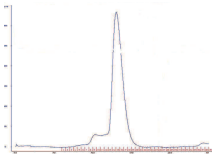


Fig 7: Size exclusion chromatogram showing the purification of the protein A3Fc-CD2

Fig 8: SDS PAGE and Coomasie staining of the fractions of the peak from Gel filtration (Fig 7). It shows that the protein purified is pure (>95%)

Perspectives

The goal of our study will be the determination of the X-ray structure of the Vif/hA3 complexes. We obtained a very good expression of the complexes A and B as well as the C-terminal domain of hA3F in good quality and quantity.

The next step will be to study the interaction between the complex Vif and Cullin5 by ITC or SPR, as well as the interaction with APOBEC3G/F and to determine thermodynamic and kinetic parameters of this interaction.

Furthermore, the aim of this project is to solve structure of the complex Vif with APOBEC3 by crystallization or cryoelectron microscopy.

REFERENCES

[1] Holmes & al Trends Biochem Sci (2007)
[2] Chen & al Nature (2008)
[3] Holden & al Nature (2008)
[4] Bohn & al Structure (2013)
[5] Jäger & al Nature (2011)
[6] Kim & al Acta Mol cell (2013)
[7] Zhang & al Nature (2012)
[8] Guo & al Nature (2013)
[9] Siu & al Nat Comm (2013)

UNIVERSITÉ DE STRASBOURG

Joelle MEZHER

IMMUNITE INNEE ET CONTRE-MESURES VIRALES : ETUDES STRUCTURALES ET BIOPHYSIQUES DU COMPLEXE FORME ENTRE LA CYTIDINE DESAMINASE HUMAINE APOBEC3G ET LE FACTEUR D'INFECTIVITE VIRALE DU VIH-1

Résumé

La protéine Vif (Viral infectivity factor) joue un rôle essentiel dans la réplication virale et dans le pouvoir infectieux du VIH-1. Elle est aussi qualifiée de « protéine de contre-défense » car elle neutralise un facteur cellulaire à action antivirale, APOBEC3 en recrutant le complexe de la E3 ubiquitine ligase formé de CBFβ/EloB/EloC/Cul5/Rbx2, qui va entrainer sa polyubiquitinilation et sa dégradation. APOBEC3 comprend une famille de protéines présentes dans les mammifères et capable d'induire des mutations dans l'ADN viral.

La structure tridimensionnelle des molécules et complexes macromoléculaires étant intimement liée à la fonction, il est primordial de connaître la place d'A3G/F dans le complexe Vif-CBFβ-EloB-EloC ainsi que les détails des interactions entre Vif et A3G/F, ceci dans une perspective thérapeutique.

La co-expression des différentes protéines de ce complexe a été réalisée dans deux systèmes d'expression : bactérien (*E. coli*) et eucaryote (BHK21) dû au manque de solubilité d'APOBEC et de la difficulté de sa cristallisation. Une fois les protéines exprimées et le complexe Vif-CBFβ-EloB-EloC-A3G/F formé dans les deux systèmes, des tests de cristallisation et des études structurales seront réalisés pour caractériser la structure et la fonction des protéines.

Dans *E. coli*, j'ai réussi à obtenir le complexe Vif/CBFβ/EloB/EloC/Cul5 *in vivo* et *in vitro* de manière soluble, pure, monodisperse ainsi qu'en quantité adéquate avec les études biophysiques envisagées. Par contre, je n'ai pas réussi à surexprimer A3G. En passant dans les cellules de hamster BHK21 et en appliquant la technique du virus de la vaccine, nous avons réussi à obtenir le complexe Vif-CBFβ-EloB-EloC-A3G/F mais en quantité insuffisante pour la cristallisation.

Les perspectives de ce projet seront d'améliorer : (1) les rendements du complexe pour des tests de cristallisation et (2) la pureté du complexe pour réaliser la cryo-microscopie électronique.

Mots-clés : VIH-1, Immunité innée, facteurs de restriction virale, Vif, APOBEC, structure, cristallographie.

Résumé en anglais

APOBEC3 are human antiviral cytidine deaminase and RNA/DNA-editing enzymes that belong to the innate immune defense system, targeting retroviruses or retrotransposons. Among all APOBEC3 proteins, hA3B, hA3C, hA3DE, hA3F and hA3G are able to interfere negatively with HIV-1 infectivity: they induce a deamination of dC to dU in the minus strand DNA, resulting in G to A hypermutation in the plus strand DNA. This hypermutation results either into a degradation of U-rich DNA strands by the uracyl-DNA glycosylase or into the production of aberrant viral protein. In addition, a deaminase-independent mechanism is able to inhibit the HIV-1 reverse transcription through a yet unknown mechanism.

To evade the host defense system, HIV expresses the virion infectivity factor, Vif, which causes the degradation of APOBEC3G by the proteasome by recruiting the E3 ubiquitin ligase complex composed of the CBFβ, EloB, EloC, Cullin5, Rbx2 proteins. The goal of this study will be the determination of the X-ray structure of the Vif/APOBEC-3 complex. Understanding this molecular recognition might be useful to direct structure-based design of anti-HIV drugs that act by inhibiting the action of Vif and lead to new anti-HIV drugs.

To overcome the solubility problems of Vif and APOBEC3G/F, a co-expression strategy had been applied with the different E3 ubiquitin ligase proteins both in prokaryotic *(E. coli)* and eukaryotic (BHK21) systems. Once the complex obtained, we will perform several structural and biophysical studies as well as crystallization trials.

In *E. coli,* I managed to obtain the Vif/CBFβ/EloB/EloC/Cul5 complex *in vivo* and *in vitro,* soluble, monodisperse and in good quantities for structural studies. The protein A3G was not obtained in *E. coli* even by co-expression with the complex.

On the other side, I succeeded in obtaining the polyprotein Vif-CBFβ-EloB-EloC-A3G/F complex in the hamster cells (BHK21) by applying the vaccinia virus strategy. Optimizing the yield and the purity is necessary for crystallization and more structural studies.

Keywords: HIV-1, innate immunity, viral restriction factors, Vif, APOBEC, structure, crystallization.