

2 m 11. 3449. 9

Université de Montréal

**Conception d'une banque de nonapeptides exprimés constitutivement en
cellules de mammifères pour l'étude *in vivo* d'interactions protéine-protéine**

par

Philippe Courtemanche-Asselin

Département de microbiologie et immunologie
Faculté des études supérieures/médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade M.Sc.
en microbiologie et immunologie

Décembre 2005

© Philippe Courtemanche-Asselin, 2005

Université de Montréal



W

4

U58

2006

V146

Direction des bibliothèques

AVIS

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

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

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

NOTICE

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

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

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

Faculté des études supérieures

Ce mémoire intitulé :

**Conception d'une banque de nonapeptides exprimés constitutivement en
cellules de mammifères pour l'étude *in vivo* d'interactions protéine-protéine**

présenté par

Philippe Courtemanche-Asselin

a été évalué par un jury composé des personnes suivantes :

Dr Georges Szatmari
président-rapporteur

Dr Éric A. Cohen
directeur de recherche

Dr Guy Lemay
membre du jury

Résumé

Depuis des millions d'années, les organismes infectieux et leurs hôtes cohabitent. Les virus ont réduit à un strict minimum leur génome et utilisent les constituants cellulaires pour effectuer leur cycle de réplication. Ces infections répétées ont permis aux cellules de s'adapter et d'adopter certains mécanismes de défense sous forme de facteurs de restriction affectant diverses étapes du cycle viral. Dans plusieurs cas, le virus a trouvé une parade à ces restrictions, dans d'autres cas, le mécanisme de défense a été efficace. À une époque où l'on voit l'apparition de virus résistant aux traitements disponibles, il devient important de mieux comprendre et d'exploiter les interactions protéiques qui s'établissent entre le virus et les cellules infectées. Ces études permettraient la découverte de cibles ou de molécules thérapeutiques inédites. Aussi, de nouveaux modèles animaux ou cellulaires qui jusqu'alors ne pouvaient pas être infectés pourraient être développés. De telles études passent par l'élaboration d'outils perfectionnés permettant la recherche fine d'interactions protéine-protéine. Certaines conditions expérimentales sont requises (soit l'analyse dans le cadre d'une infection virale, *in vivo*, en cellules de mammifères) et ne peuvent être obtenues avec les méthodes classiques (double-hybride, biopuces, banques d'ADNc, etc.). Dans le dessein de répondre aux exigences de cet outil innovateur, nous avons conceptualisé un vecteur rétroviral encodant une banque de nonapeptides. Ce vecteur doit permettre la création de cellules de mammifères exprimant constitutivement la banque de peptides. Ces nonapeptides, de forme et de séquence aléatoires, servent d'agents perturbateurs. Lors de l'infection virale, les peptides altéreront certaines interactions entre les protéines virales et les protéines des cellules hôtes. Cette perturbation affectera le phénotype de la cellule ou le cours de l'infection. Les peptides candidats peuvent servir à caractériser les protéines impliquées dans la voie métabolique étudiée et être à la base de la conception de molécules inhibitrices. Ce mémoire met en lumière les difficultés inhérentes à la création d'une banque de nonapeptides. Plusieurs méthodes sont décrites qui visent toutes à atteindre la qualité essentielle d'une banque de molécules : une grande diversité favorisant la probabilité d'une interaction avec la voie métabolique ciblée.

MOTS CLÉS : banque de peptides, vecteurs rétroviraux, sélection génétique, interactions protéine-protéine

Abstract

For millions of years, pathogens and their hosts have interacted. Viruses have maintained the strict minimum of their genome and have used the cell machinery and its numerous proteins to complete their replication cycle. Repeated infections have allowed cells to adapt and to develop defense mechanisms expressed as restriction factors affecting various steps of the viral cycle. In many cases, viruses have found ways to counter-attack these restrictions, but in other cases, the defense mechanisms are effective. We are now at a point where new strains of resistant viruses have emerged and it has become important to better understand and exploit the intimate virus-host cell protein interactions. Such studies may allow for the discovery of new targets or new kinds of therapeutic molecules. It may also allow for the development of cell or animal models previously immune to infection. Such studies are possible if new proteomic tools are created, which would allow precise and effective detection of protein-protein interactions. Some experimental conditions are required (analysis in the context of an infection, *in vivo* in mammalian cells) and can not be obtained with the actual proteomic tools available (two-hybrid, micro-array, cDNA libraries, etc.). To meet these criteria, we have designed a retroviral vector expressing a nonapeptide library. This vector should allow the establishment of a stable cell line expressing the peptide library. Following an infection of these cells with HIV-1, these peptides of random shape and sequence may interact in critical virus - host cell protein-protein interactions and alter the phenotype. These peptides may be used to characterize the proteins involved in the metabolic venue studied and serve as a model for the inhibitors. We also emphasized the difficulties one would encounter when creating a nonapeptide library. Several methods were tested in order to get the principal quality of a good random molecular library: a huge diversity.

KEYWORDS : peptides library, retroviral vector, genetic selection, protein-protein interactions

Table des matières

INTRODUCTION	1
1. L'ère protéomique	1
1.1 Contexte	1
1.2 Interactions protéiques entre le virus et la cellule hôte	3
2. Outils d'analyse des interactions protéine-protéine	5
2.1 Techniques d'analyse d'interactions protéine-protéine <i>in vitro</i>	7
2.1.1 Les banques de phages présentoirs	7
2.1.2 Les biopuces	9
2.1.3 Le système de purification sur colonne d'affinité et le spectre de masse	11
2.2 Techniques d'analyse des interactions protéine-protéine <i>in vivo</i>	12
2.2.1 Le système de double-hybride	12
2.2.2 Analyse d'interactions par la technologie de transfert d'énergie entre fluorochromes (<i>FRET</i>)	18
2.2.3 Expression d'une banque de peptides en cellules de mammifères par le biais d'un vecteur rétroviral	19
3. Objectifs de notre projet de recherche	26
MATÉRIEL ET MÉTHODES	32
1. Biologie moléculaire	32
1.1 Construction des vecteurs rétroviraux de base	32
1.1.1 Clonage des inserts Puro ^R /lien souple, Puro ^R , Zeo ^R /lien souple et Zeo ^R , dans le vecteur pBABE HAGIZ	33
1.1.2 Remplacement du promoteur SV40 par le promoteur de la phosphoglycérate kinase	35
1.1.3 Introduction de sites de restriction pour le clonage et la sélection des clones positifs	35
1.1.4 Construction d'un vecteur rétroviral contrôle pBABE P _{PGK} eYFP	38
1.2 Construction de la banque de nonapeptides	38
1.2.1 Méthode insert-vecteur	38
1.2.2 Méthode insert-vecteur par amplification d'un insert de taille moyenne	40
1.2.3 Méthode ARN-ADNc-ADN	42
1.2.4 Clonage par amplification ou extension de tout le vecteur rétroviral	47
1.3 Ligatures	49
1.4 Transformations	50
1.4.1 Bactéries chimiocompétentes <i>MAX Efficiency Stb12 Cells</i>	50

1.4.2	Bactéries électrocompétentes <i>ElectroMAX Stbl4 Cells</i>	50
1.4.3	Incubation et amplification	51
1.5	Calculs	51
1.6	Analyse des clones	51
1.7	Réserve de clones en milieu glycérol	52
2.	Production et infection virales	52
2.1	Transfection des vecteurs dans des cellules HEK 293T	52
2.2	Introduction de la banque de peptides par infection de cellules de mammifères HEK 293	54
3.	Techniques biochimiques	55
4.	Milieus de culture	56
5.	Cellules utilisées	56
5.1	Cellules de mammifères	56
5.2	Bactéries	57
	SECTION RÉSULTATS	58
1.	Conception et génération de la banque de peptides dans le vecteur rétroviral murin pBABE.	58
1.1	Configuration optimale de la protéine de fusion	58
1.2	Clonage du vecteur rétroviral pBABE P _{PGK} HA/Puro ^R /lien souple	61
2.	Synthèse par ligature insert-vecteur	62
2.1	Stratégie du clonage insert-vecteur	62
2.2	Difficultés inhérentes à la méthode insert-vecteur	63
2.3	Ligature	65
2.4	Efficacité de la transformation	66
2.5	Analyse sur gel	67
2.6	Séquençages	69
2.7	Résumé	70
3.	Insertion de la banque par amplification	72
3.1	Stratégie du clonage	72
3.2	Ligature	74
3.3	Efficacité des transformations	75
3.4	Analyse des clones	76
3.5	Séquençages	77

3.6	Résumé	78
4.	Synthèse ARN-ADNc-ADN	79
4.1	Stratégie du clonage	79
4.2	Clonage	82
4.3	Efficacité des transformations	83
4.4	Analyse des clones	83
4.5	Séquençages	84
4.6	Résumé	84
5.	Synthèse par amplification différentielle	85
5.1	Principe de la méthode	85
5.2	Clonage	86
5.3	Ligatures	87
5.4	Efficacité des transformations	88
5.5	Analyse des clones par digestion enzymatique <i>BamH I</i>	88
5.6	Analyse des séquences	90
5.7	Résumé	94
6.	Considérations techniques communes à toutes les méthodes utilisées	95
6.1	Analyse des polynucléotides encodant la banque de peptides	95
7.	Essais de ligature avec des polynucléotides biotinylés	99
7.1	Synthèse séquentielle et biotinylation des polynucléotides	99
7.2	Clonage et purification de l'ADN avec les billes de streptavidine	101
8.	Production virale et infection	102
8.1	Production de virus pseudotypés VSV-G et infection de cellules de mammifères	102
	DISCUSSION	104
1.	Conception et génération du vecteur rétroviral	105
2.	Caractéristiques des peptides	108
3.	Purification des amorces	109
4.	Technique de construction à privilégier	113
5.	Ligature et efficacité des transformations	116

6. Considérations mathématiques	117
7. Infection des cellules de mammifères	119
CONCLUSION	121
RÉFÉRENCES	125
ANNEXES	xix
- Cartes des vecteurs utilisés	

Liste des figures

Figure 1. Représentation du pouvoir antiviral de TRIM5 α , Fv1 et des dérivés TRIM sur divers rétrovirus N-MLV, B-MLV, SIV _{mac} et HIV-1.	4
Figure 2. Comparaison du ratio « efficacité versus précision » de quelques outils d'étude d'interactions protéine-protéine.	6
Figure 3. Protéines plasmatiques potentiellement observables à des concentrations de l'ordre du millimolaire (10^{-3}) au yoctomolaire (10^{-24}).	6
Figure 4. Schématisation du criblage d'une banque de peptides sur phages présentoirs	7
Figure 5. Schématisation de biopuces fonctionnelles impliquant l'analyse d'interactions protéine-protéine.	10
Figure 6. Transcription normale. L'activation de la transcription requiert l'interaction du domaine de liaison à l'ADN (BD) et le domaine d'activation (AD).	12
Figure 7. Transcription active dans l'essai double-hybride. La technique du double-hybride mesure les interactions protéine-protéine par le biais de la transcription d'un gène rapporteur.	13
Figure 8. Double-hybride en cellules de mammifères.	15
Figure 9. Principe de l'essai <i>Split-ubiquitine</i> .	16
Figure 10. Schématisation du criblage d'interactions protéine-protéine par sélection génétique à l'aide de la librairie de peptides.	20
Figure 11. Présentation d'une banque de peptides sous forme de boucle, par introduction d'une séquence SKIVILF.	21
Figure 12. Criblage de banque de peptides en cellules de mammifères sous sélection génétique permettant la caractérisation de peptides dont l'interaction avec une protéine cible en perturbera le phénotype cellulaire.	22
Figure 13. Clonage de la banque de peptides dans un vecteur rétroviral.	24

Figure 14. Conception d'une banque de peptides par intégration site-spécifique d'un polynucléotide sur un vecteur rétroviral sous forme d'ADN simple brin.	25
Figure 15. Mécanisme MMR de réparation bactérien en présence d'une bulle d'ADN non complémentaire.	26
Figure 16. L'introduction d'une banque de peptides dans une cellule mènera à une perturbation de son phénotype.	27
Figure 17. Étapes du criblage par interférence d'une banque de peptides dans les mécanismes d'interactions protéine-protéine, <i>in vivo</i> , en cellules de mammifères.	29
Figure 18. Schématisation du clonage des inserts Puro ^R /lien souple et Zeo ^R /lien souple dans le vecteur pBABE HAGIZ.	33
Figure 19. Schématisation du clonage de pBABE P _{PGK} HA/Puro ^R /lien souple/kill sites.	36
Figure 20. Représentation des protéines de fusion générées. Ces protéines de fusion sont clonées dans le vecteur rétroviral pBABE.	59
Figures 21 et 22. (21) Évaluation de la courbe de survie de la puromycine sur les cellules HEK 293T. (22) Évaluation de la courbe de survie de la puromycine sur les cellules OMK.	61
Figure 23. Vecteur pBABE HAGIZ	61
Figure 24. Vecteur pBABE P _{PGK} HA/Puro ^R /lien souple/kill sites.	62
Figure 25. Schématisation du clonage selon la méthode insert-vecteur.	63
Figure 26. Migration sur gel d'agarose des polynucléotides codant pour la banque de nonapeptides, complétés en ADN double brin par extension avec la polymérase d'ADN.	64
Figure 27. Digestion enzymatique du vecteur pBABE P _{PGK} HA/Puro ^R /lien souple.	65
Figure 28. Ligatures obtenues avec la stratégie insert-vecteur.	66

Figure 29. Électrophorèse de l'ADN des clones de banque de peptides générés par la méthode insert-vecteur.	67
Figure 30. Analyse par électrophorèse sur gel d'agarose des clones issus de la ligature aux sites <i>Dra</i> III et <i>Cla</i> I.	67
Figure 31. Analyse par électrophorèse sur gel d'agarose des clones issus de la ligature aux sites <i>Dra</i> III et <i>Cla</i> I.	68
Figure 32. Analyse des clones non coupés issus de la ligature aux sites <i>Mfe</i> I et <i>Cla</i> I.	68
Figure 33. Efficacité de digestion du site <i>Cla</i> I en fonction de la séquence en amont du site.	71
Figure 34. Schématisation du clonage par amplification de la banque de peptides.	72
Figure 35. Analyse sur gel de l'insert de la banque de nonapeptides du site <i>BsrG</i> I au site <i>Cla</i> I.	73
Figure 36. Analyse sur gel du vecteur pBABE P _{PGK} HA/Puro ^R /lien souple coupé <i>BsrG</i> I ou <i>Cla</i> I avant et après purification sur gel.	73
Figure 37. Analyse par électrophorèse des éléments requis pour la ligature insert-vecteur aux sites <i>Dra</i> I et <i>Cla</i> I.	74
Figure 38. Ligature de l'insert encodant la banque de nonapeptides selon que le vecteur ou l'insert est déphosphorylé.	74
Figure 39. Ligature de l'insert de nonapeptides entre les sites <i>Dra</i> I et <i>Cla</i> I.	75
Figure 40. Analyse par digestion enzymatique <i>BamH</i> I de clones issus de la ligature insert-vecteur aux sites <i>BsrG</i> I et <i>Cla</i> I.	76
Figure 41. Analyse par digestion <i>BamH</i> I de clones issus de la ligature insert-vecteur aux sites <i>Dra</i> I et <i>Cla</i> I.	77
Figure 42. Analyse par digestion enzymatique <i>BamH</i> I de clones issus de la ligature insert-vecteur aux sites <i>EcoR</i> I et <i>Cla</i> I.	77

- Figure 43. Schématisation de l'introduction d'un site promoteur T7 dans le vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites. 80
- Figure 44. Schématisation de l'ARNm du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites/T7. 81
- Figure 45. Rétrotranscription du vecteur avec l'amorce de la banque de peptides. 81
- Figure 46. Synthèse du second brin du vecteur ADNc encodant la banque de peptides. 81
- Figure 47. Ligature du vecteur rétroviral encodant la banque de peptides. 81
- Figure 48. Analyse sur gel d'agarose/formaldéhyde de la production d'ARN de vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple. 82
- Figure 49. Analyse sur gel de la rétrotranscription en ADNc du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/nonapeptide. 82
- Figure 50. Analyse sur gel de l'extension du second brin d'ADN du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/nonapeptide. 83
- Figure 51. Analyse par électrophorèse sur gel d'agarose des clones obtenus lors de la transformation des ADN produits par la technique ARN-ADNc-ADN. 84
- Figure 52. Analyse de l'ADN provenant de l'amplification différentielle du vecteur rétroviral 86
- Figure 53. Évaluation de l'efficacité de ligature du vecteur rétroviral amplifié. 87
- Figure 54. Amplification du vecteur à l'aide d'amorces purifiées par PAGE et ligature des ADN. 87
- Figures 55 à 61. Analyse par digestion des clones produits avec les polynucléotides non purifiés. 89
- Figure 62 et Figure 63. Analyse par digestion enzymatique *Bam*H I des clones obtenus des transformations des ADN faits avec des amorces purifiées par PAGE. 89
- Figure 64. Séquençage de l'ADN représentatif de tous les clones de la banque de peptides. 93

Figure 65. Démonstration que l'amplification du vecteur faite avec des oligonucléotides sans séquence dégénérée permet une ligature plus efficace du vecteur rétroviral amplifié.	95
Figure 66. Spectre de masse du polynucléotide antisens encodant la banque de nonapeptides.	96
Figure 67. Analyse sur gel de polyacrylamide des polynucléotides encodant la banque de peptides.	97
Figure 68. Spectre de masse de l'oligonucléotide sens complémentaire au vecteur rétroviral sous forme d'ADN simple brin.	97
Figure 69. Analyse de la séquence de l'amplification effectuée avec le polynucléotide synthétisé par la compagnie Alpha DNA inc. et purifié par PAGE.	98
Figure 70. Analyse de la séquence de l'amplification effectuée avec le polynucléotide synthétisé par la compagnie Sigma Genosys.	98
Figure 71. Spectre de masse du polynucléotide combiné encodant la banque de nonapeptides.	100
Figure 72. Analyse de la séquence de l'ADN amplifié avec le polynucléotide combiné encodant la banque de nonapeptides.	100
Figure 73. Purification des ADN sur billes de streptavidine.	101
Figure 74. Production virale par triple transfection.	102
Figure 75. Configuration d'un vecteur rétroviral comportant un IRES.	106
Figure 76. Représentation du vecteur rétroviral généré : pBabe P _{PGK} HA/Puro ^R /lien souple/nonapeptide.	107
Figure 77. Schématisation d'une immunoprécipitation.	108

Liste des tableaux

Tableau I. Évaluation qualitative de la survie cellulaire de clones de cellules HEK 293T en présence des antibiotiques puromycine ou zéocine.	60
Tableau II. Efficacité des transformations des ADN provenant des ligatures aux sites <i>EcoR I/Cla I</i> , <i>BsrG I/Cla I</i> ou <i>Dra I/Cla I</i> .	76
Tableau III. Production de particules virales pseudotypées VSV-G pour l'établissement de lignées stables en cellules de mammifères.	103
Tableau IV. Pourcentage d'oligonucléotides complets en fonction de la longueur de ceux-ci et du pourcentage d'efficacité de la réaction de synthèse.	110

Liste des abréviations

AD :	Domaine d'activation de la transcription, <i>Activation Domain</i>
ADN :	Acide désoxyribonucléique
ADNc :	ADN complémentaire
Amp :	Ampicilline
APOBEC3G :	<i>Apolipoprotein B mRNA Editing enzyme, Catalytic Polypeptide-like 3G</i>
ARN :	Acide ribonucléique
ARNm :	ARN messenger
BD :	Domaine de liaison à l'ADN, <i>Binding Domain</i>
BSA :	Albumine de sérum bovin, <i>Bovine Serum Albumin</i>
CAT :	Chloramphénicol acétyltransférase
CC ₅₀ :	Concentration cytotoxique 50 % (courbe de survie)
DEPC :	Diéthylpyrocarbonate
DMEM :	milieu de culture « <i>Dulbecco's Modified Eagle Medium</i> »
DMSO :	Diméthylsulfoxyde
D.O. :	Densité optique
dNTP :	Désoxyribonucléotides triphosphates
FACS :	Tri cellulaire activé par la fluorescence, <i>Fluorescence-Activated Cell Sorting</i>
FRET :	Transfert d'énergie résonnant de type Förster entre des fluorochromes, <i>Fluorescence Resonance Energy Transfer</i>
FV1 :	<i>Friend-virus-susceptibility-1</i>
Gag :	Protéine du VIH-1, <i>Group Antigens (Ag)</i>
GAL4 :	Galactose4
GFP :	Protéine de fluorescence verte, <i>Green Fluorescent Protein</i>
GST :	Glutathione-S-transférase
HA :	Épitope hémagglutinine A
IRES :	Site d'entrée interne du ribosome, <i>Internal Ribosome Entry Site</i>
kb :	Kilobase
kV :	Kilovolt
LB :	Milieu Luria-Bertani
LMP :	Agarose de point de fusion bas, <i>Low Melting Point</i>
LTR :	Séquences terminales répétées, <i>Long Terminal Repeat</i>
LV-1 :	Facteur de résistance aux lentivirus, <i>Lentivirus Resistance Factor-1</i>
MALDI-TOF :	<i>Matrix-assisted laser desorption/ionization time-of-flight</i>
MDR1 :	Gène de multirésistance aux drogues, <i>Multi-Drug Resistance gene-1</i>

MI :	Multiplicité d'infection, <i>Multiplicity of Infection (MOI)</i>
min :	Minute
mM :	Millimolaire
MoMuLV :	Virus de la leucémie murine de Moloney, <i>Moloney Murine Leukemia Virus</i>
NEAA :	Acides aminés non essentiels, <i>Non Essential Amino Acids</i>
NF- κ B :	Facteur de transcription <i>Kappa B</i> , <i>Nuclear Factor Kappa B</i>
O/N :	Toute la nuit, <i>Overnight</i>
PAGE :	Électrophorèse sur gel de polyacrylamide, <i>Polyacrylamide Gel Electrophoresis</i>
pb :	Paire de bases
pBABE HAZ :	Plasmide pBABE HA/Zeo ^R
pBABE HAGIZ :	Plasmide pBABE HA/GST/Zeo ^R
PBS :	Solution saline tamponnée par le phosphate, <i>Phosphate Buffered Saline</i>
PCR :	Réaction de polymérisation en chaîne, <i>Polymerase Chain Reaction</i>
PGK :	Phosphoglycérate kinase
Puro ^R :	Gène de résistance à la puromycine (<i>pac</i> ou <i>puromycine-N-acétyltransférase</i>)
hRIP :	Protéine interagissant avec Rev chez l'humain, <i>Human Rev-Interacting Protein</i>
sec :	Seconde
SIV :	Virus simien de l'immunodéficience acquise, <i>Simian Immunodeficiency Virus</i>
TAP :	Purification par affinité, <i>Tandem Affinity Purification</i>
TAP1 et TAP2 :	<i>Transporter Associated with Antigen Processing</i>
TRIM5 α :	Protéine de type TRIM, <i>TRIMpartite Motif</i>
Ref-1 :	Facteur de résistance au rétrovirus, <i>Retrovirus Resistance Factors</i>
Rev :	Protéine régulatrice de l'expression des protéines virales, <i>Regulator of expression of virion proteins</i>
SV40 :	<i>Simian Virus 40</i>
uM :	Micromolaire
Vif :	Facteur de potentiel infectieux du virion, <i>Virion Infectivity Factor</i>
Vpr :	Protéine virale accessoire R, <i>Viral Protein R</i>
VIH-1 :	Virus de l'immunodéficience humaine de type 1
VSV-G :	Enveloppe du virus de la stomatite vésiculaire, glycoprotéine G
X-GAL :	5-bromo-4-chloro-3-indolyl- β -D-galactoside
YFP :	Protéine de fluorescence jaune, <i>Yellow Fluorescent Protein</i>
ZAP :	Protéine antivirale de type doigts de zinc, <i>Zinc Finger Antiviral Protein</i>
Zeo ^R :	Gène de résistance à la zéocine (<i>Sh ble</i>)

**« Plus que bu, oui ! ... [...] quand on est parti, on s'est aperçu que la terre tournait.
[...] Comme quoi, il faut être pochard pour constater les lois de la nature ! »**

Feydeau, *Occupe-toi d'Amélie*.

**« [...] Et puis y a les savants, tiens, il paraît
que c'est les savants qui serchent le pluss.
Là je comprends pas.
S'ils serchent, c'est qu'ils ont pas trouvé,
Et s'ils ont pas trouvé, ils savent pas encore,
Et s'ils savent pas, ils sont pas encore savants!**

**Moi, je sens que le pourrais être un vrai sercheur.
Sercher c'est facile. C'est trouver qui est dur! [...] »**

Marc Favreau (Sol), *Presque tout Sol*.

Remerciements

Tout d'abord, je tiens à remercier Éric A. Cohen, mon directeur de recherche, d'avoir eu cette vision d'un projet qui avait beaucoup de potentiel et qui m'a permis d'y participer. Je remercie particulièrement ma femme Fannie Godbout, pour les petites et grandes attentions à me reconforter lorsque l'insomnie du chercheur m'a pris, pour son amour inconditionnel, pour sa patience dans l'attente du retour de l'étudiant fourbu. Merci à mon père, Normand, pour sa vision de la science comme une naissance perpétuelle, avec ses émerveillements et ses yeux grands ouverts devant l'inconnu. Merci à ma mère, Huguette, pour sa pédagogie qui me garde la colonne droite et les idées dans le même sens. Merci aussi à mes beaux-parents, pour leur fierté et leurs encouragements. Merci à tous mes collègues qui usent de l'humour comme d'un bouclier. La recherche, malgré ses ratés et ses déceptions quotidiennes, reste une affaire d'espoir, l'espoir d'un monde logique, du moins un peu moins incompréhensible.

Introduction

1. L'ÈRE PROTÉOMIQUE

1.1 Contexte

Les interactions complexes entre un agent pathogène et l'hôte sont la base sous-jacente de toute maladie infectieuse. La compréhension des détails moléculaires de ces interactions permet la caractérisation des gènes conférant la virulence du microbe et des stratégies de défense de l'hôte. Cela permet aussi d'identifier à quels signaux ces mécanismes s'activent et par quelles voies ils sont régulés. Pour progresser dans la lutte contre les infections, la recherche se doit de mieux connaître les micro-organismes qui en sont responsables et de définir avec précision, sur le plan moléculaire, leurs interactions avec l'hôte, l'homme en l'occurrence.

Les processus physiologiques et physiopathologiques, de la cellule jusqu'à l'organisme complet, reposent sur l'activité des protéines et les interactions protéine-protéine y jouent un rôle critique. En effet, l'ère génomique nous a révélé que l'humain posséderait à peine plus de gènes (20 000 à 25 000) que la drosophile ou le nématode *Caenorhabditis elegans*¹. En revanche, le nombre de protéines chez l'humain est estimé entre 300 000 et 1 000 000². La très grande complexité des voies métaboliques est donc intimement liée aux différences d'expression, de modification ou d'assemblage protéiques et aux interactions que les protéines entretiennent entre elles et avec d'autres composants cellulaires. L'étude de ces protéines a donné naissance à une nouvelle discipline communément appelée « génomique fonctionnelle », dont le but est d'attribuer une fonction aux milliers de protéines identifiées par les divers projets de séquençage. La protéomique, quant à elle, prend le relai de la génomique et analyse en détail les interactions et les structures essentielles à la fonction protéique. Contrairement au génome qui est relativement stable, la dynamique du protéome change de minute en minute en réponse aux signaux intra et extracellulaires. La caractérisation de l'activité et de la fonction de chaque protéine apparaît alors comme un labeur considérable. L'addition

de groupements chimiques et les changements de conformation suivant la traduction induisent ou modifient l'activité protéique. Ces modifications posttraductionnelles ne peuvent être reproduites exactement *in vitro*. La notion de fonction est elle-même complexe puisqu'elle dépend en partie de la position de la protéine dans les voies métaboliques auxquelles elle participe. La localisation cellulaire est également importante, car elle est déterminante pour la fonction. Les études visant à caractériser une voie métabolique particulière doivent tenir compte de ces variations constantes et considérer les résultats en rapport au contexte, dans une cellule donnée et dans une situation donnée.

Une puissante technique pour déduire la fonction d'une protéine est d'en identifier les partenaires protéiques, puisque les protéines interagissant les unes avec les autres ou faisant partie d'un même complexe sont généralement impliquées dans le même processus cellulaire³. La compréhension des interactions protéiques a beaucoup progressé grâce aux outils de la biologie moléculaire qui ont été développés ces dernières années. Diverses approches ont été utilisées. À petite échelle, le double-hybride – décrit plus bas – est utilisé pour la détection d'interactions binaires *in vivo*⁴. Une autre technique récente consiste en la purification de complexes protéiques en utilisant des marqueurs d'affinité, ou *tag* (TAP). Cette sélection est alors suivie d'une identification des protéines par spectre de masse. Cette dernière technique permet d'identifier pratiquement tout le spectre des complexes formés avec une protéine particulière³. Des techniques utilisant la fluorescence pour la détection d'interactions ont aussi été utilisées, ainsi que les biopuces (*microarray*)⁵ et les phages présentoirs⁶. Cependant, certains aspects de l'approche protéomique restent limités; la détection et les investigations fonctionnelles directes des interactions protéine-protéine *in vivo* et *in vitro* reposent en général sur des manipulations lourdes accompagnées de nombreuses limitations (protéines présentes en faible abondance, nombreux artefacts, propriétés physico-chimiques particulières, etc.). L'émergence de nouveaux virus ou de virus résistant aux traitements a mis l'accent sur l'importance de caractériser les interactions protéiques entre l'agent pathogène et la cellule hôte. Une course à l'innovation a permis à d'innombrables nouvelles techniques d'analyse des interactions protéiques de voir le jour, chacune visant à contourner les limites des méthodes traditionnelles, inadéquates dans le cadre d'une infection virale.

1.2 Interactions protéiques entre le virus et la cellule hôte

Plusieurs groupes tentent d'élucider les mécanismes protéiques régissant le cycle de réplication viral. Ces interactions protéiques essentielles régulent l'entrée et la sortie du virus, sa réplication ou sa latence et modulent son potentiel infectieux. Les millions d'années passées sous l'assaut répété d'épidémies virales ont permis aux cellules des vertébrés d'acquérir plusieurs mécanismes de défense très efficaces dont plusieurs des composantes nous sont encore inconnues. Certains de ces mécanismes n'impliquent pas le système immunitaire et sont plutôt régis par des protéines de la cellule hôte ayant des effets restrictifs sur des protéines ou des processus propres aux virus.

Un exemple d'une telle interaction apparaît entre la protéine virale Vif du VIH-1 et les cytidines désaminases cellulaires APOBEC3A-G⁷. Dans la cellule infectée, Vif agit à divers niveaux sur la protéine APOBEC; elle interfère dans la traduction des ARNm et se lie aussi directement à la protéine afin de la déstabiliser et d'en accélérer la dégradation par le protéasome 26S. Cette interaction diminue à presque zéro la concentration intracellulaire d'APOBEC et empêche son incorporation dans les virions nouvellement assemblés. D'autres mécanismes sont aussi impliqués, mais sont encore inconnus ou mal définis⁸. En absence de Vif, APOBEC est incorporée dans les virions du VIH-1 et empêche l'infection productive de cellules nouvellement infectées par l'introduction massive de mutations dans le brin négatif de l'ADN viral durant la transcription inverse⁹. La démonstration que le pouvoir infectieux d'un virus peut être altéré par des peptides prévenant des interactions protéine-protéine telles que Vif et APOBEC pourrait mener à une nouvelle classe de molécules anti-VIH-1. Des interactions protéine-protéine critiques pour l'infection virale sont retrouvées dans presque toutes les étapes du cycle d'infection et sont autant de cibles potentielles pour l'élaboration de petites molécules capables d'inhiber le pouvoir pathogène des micro-organismes. Une thérapie agissant par le biais de l'inhibition de diverses étapes de l'infection (entrée du virus dans la cellule, réplication ou latence, invasion tissulaire, processus inflammatoire destructeur, etc.) serait plus efficace et permettrait de retarder l'apparition de virus résistants.

Outre la conception d'inhibiteurs viraux, la compréhension de certains phénomènes comme la spécificité d'hôte permettrait d'étendre la connaissance du cycle viral et ouvrirait la voie à la création de nouveaux modèles animaux. Le virus de l'immunodéficience

humaine (VIH-1), par exemple, montre un tropisme d'hôte très restreint; il peut se répliquer efficacement dans les cellules T primaires ou immortalisées et dans les macrophages humains, mais pas dans les cellules de singes. Ce phénomène, depuis longtemps observé, fut récemment l'objet d'une percée majeure dans le domaine de l'étude des rétrovirus. Il fut démontré que l'existence d'une protéine intracellulaire de la famille TRIM, TRIM5 α (autrefois identifiée Ref-1 chez l'humain et LV-1 chez le singe) inhibe de manière dominante une étape précoce du cycle d'infection de divers rétrovirus en empêchant la décapsidation, bloquant ainsi l'initiation de la transcription inverse (Figure 1)^{10,11}. Ce mécanisme de défense intracellulaire propre à chaque espèce confère, en partie, la spécificité d'hôte de ces virus. Les différences en acides aminés du domaine SPRY de la molécule TRIM5 α humaine et simienne expliqueraient la spécificité d'espèce^{12,13}. Brièvement, la protéine TRIM5 α humaine aurait une plus forte affinité pour la capsid du SIV que pour celle du VIH-1, et inversement, ce qui expliquerait pourquoi le SIV ne peut se répliquer dans les cellules humaines et pourquoi le VIH-1 ne peut se répliquer dans les cellules de singes¹⁴. Le mécanisme exact de la restriction du virus par TRIM5 α est cependant toujours inconnu.

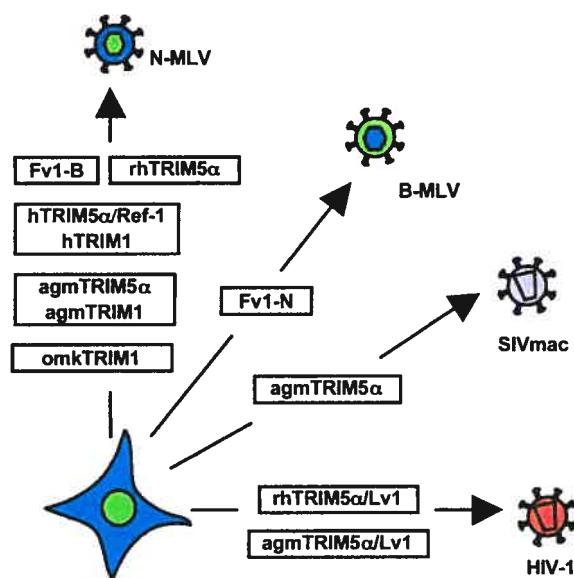


Figure 1. Représentation du pouvoir antiviral de TRIM5 α , Fv1 et des dérivés TRIM sur divers rétrovirus N-MLV, B-MLV, SIV_{mac} et HIV-1. (Schéma tiré de réf. 13)

De nouvelles défenses cellulaires semblables à TRIM5 α , comme ZAP ou le facteur de restriction cellulaire dont l'action est inhibée par la protéine virale Vpu^{15,16}, sont des cibles

potentielles pour une action thérapeutique. La longue coévolution des virus et des cellules a permis à ces dernières de déterminer les points faibles des virus. Les informations importantes qu'il serait possible d'obtenir de l'étude de cette sélection naturelle demandent que l'on s'attarde aux interactions entre l'organisme infectieux et la cellule hôte. D'autres protéines cellulaires, sans être des facteurs de restriction, ont un impact positif ou négatif sur la réplication virale. Par exemple, la protéine cellulaire hRIP agit comme cofacteur essentiel à la protéine virale Rev¹⁷. Ces protéines sont impliquées à divers stades du cycle de réplication des virus¹⁸. Une partie de ces interactions reliées au VIH-1 sont répertoriées en ligne dans le *HIV Protein Interaction Database*¹⁹. Ces exemples montrent bien l'interdépendance des protéines de l'hôte et des protéines virales. Dans ce contexte, la sélection de molécules pouvant interagir avec des facteurs de restriction cellulaire comme TRIM5 α permettrait l'exploitation de nouveaux modèles animaux ou cellulaires jusqu'à maintenant non susceptibles à l'infection. Dans le cas des cellules de singes, il est important de noter que la restriction suivant l'entrée du virus dans la cellule constitue le blocage le plus important de la multiplication du VIH dans ce type cellulaire. Une fois la restriction du récepteur passée, le virus peut se multiplier normalement si la restriction de TRIM5 α est surmontée. Ce phénomène est propre aux cellules de certaines espèces de singes seulement⁹. En effet, l'infection d'organismes plus éloignés phylogénétiquement est plus compliquée, puisque le virus doit composer avec divers autres éléments empêchant l'infection, comme des différences de récepteurs cellulaires, d'autres types de facteurs de restriction, etc.

2. OUTILS D'ANALYSE DES INTERACTIONS PROTÉINE-PROTÉINE

Afin de disséquer efficacement les interactions entre les protéines virales et les protéines cellulaires, de puissants outils protéomiques doivent être utilisés. Une vaste panoplie de ces outils existe et permet, en général, la découverte de partenaires protéiques et leur caractérisation. Le choix de la technique appropriée à une étude dépend de diverses caractéristiques. Un compromis doit être fait entre la facilité d'utilisation d'un système *in vitro* et la validité des données obtenues par ces méthodes peu physiologiques. De plus, les diverses techniques offrent le choix entre une grande couverture du spectre de toutes les interactions possibles et l'exactitude des interactions observées (Figure 2, page 6)²⁰.

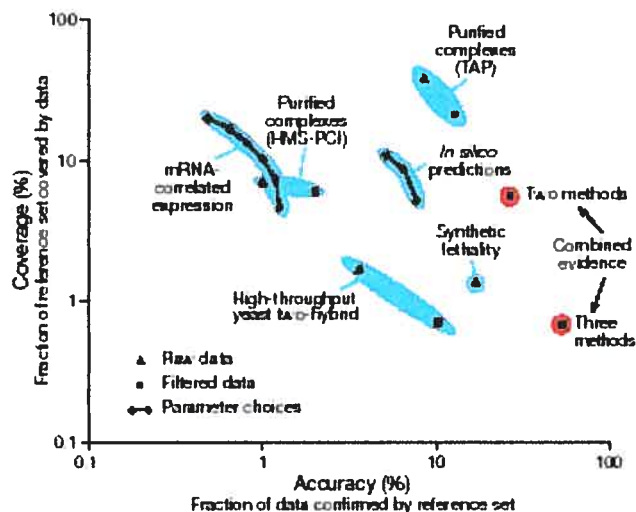


Figure 2. Comparaison du ratio « efficacité versus précision » de quelques outils d'étude d'interactions protéine-protéine. (Graphique tiré de réf. 20)

Pour des raisons historiques et pratiques, la plupart des protéines qui ont été décrites jusqu'à maintenant comptent parmi les plus abondantes de l'organisme. Or, les protéines aux propriétés les plus intéressantes pourraient justement n'être exprimées qu'en très faible quantité ou sporadiquement, à la suite d'une induction provoquée par une infection virale, par exemple. Une estimation du nombre de protéines plasmatiques, en rapport à leur abondance relative, montre clairement que la majorité des protéines existantes ne sont présentes qu'en des concentrations bien en deçà des limites de détection de plusieurs techniques d'analyse classiques, comme le gel 2D-PAGE (Figure 3)².

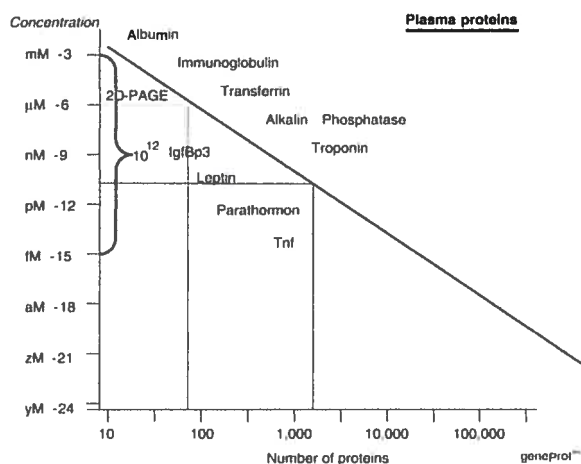


Figure 3. Protéines plasmatiques potentiellement observables à des concentrations de l'ordre du millimolaire (10^{-3}) au yoctomolaire (10^{-24}). (Illustration tirée de réf. 2)

2.1 Techniques d'analyse d'interactions protéine-protéine *in vitro*

Ce sujet sera abordé brièvement, car le choix d'outils basés sur un criblage d'interactions protéine-protéine dans un système *in vitro* est habituellement justifié pour des analyses à haut débit, comme l'étude du protéome humain ou le criblage de molécules pharmaceutiques²¹. De plus, les diverses techniques *in vitro* sont plus ou moins basées sur le même principe et comportent des avantages et inconvénients semblables.

2.1.1 Les banques de phages présentoirs

Les banques de peptides exprimés sur phages présentoirs décrivent une méthode de sélection *in vitro* au cours de laquelle des peptides ou des protéines sont fusionnés génétiquement à une protéine de coque d'un bactériophage, ce qui a pour effet d'exposer la protéine fusionnée à la surface du phage^{22, 23}.

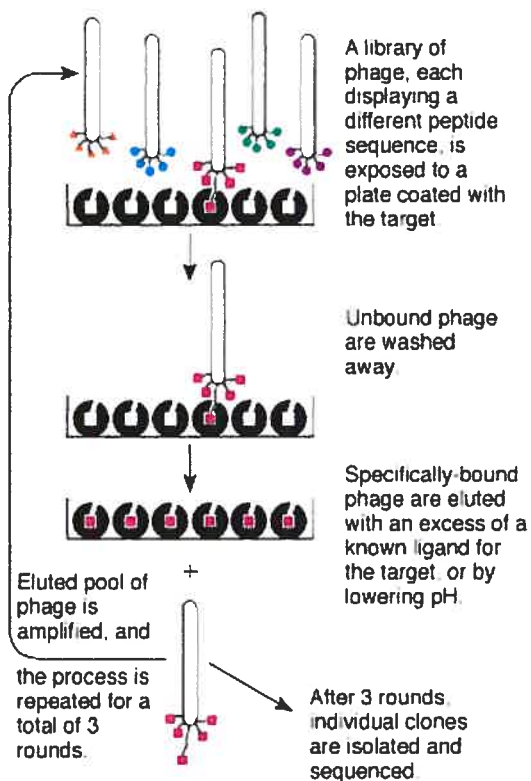


Figure 4. Schématisation du criblage d'une banque de peptides sur phages présentoirs (*phage display*). (Illustration tirée de réf. 24)

Les banques de phages présentoirs peuvent exprimer diverses variantes de peptides ou de protéines de l'ordre de 10^{12} . Cette large diversité de peptides présentés, soit 10^6 fois plus que dans le cas d'une banque d'ADNc en système de double-hybride chez la levure, permet un large éventail de détection d'interactions peptide-protéine. Plusieurs études ont tiré profit de ce système très puissant. Par exemple, dans une expérience visant à déterminer si la protéine Vpr du VIH-1 joue un rôle dans l'incorporation de protéines dans le virion, il fut démontré, par une approche de criblage de peptides sur phages présentoirs, que la majorité des peptides interagissant avec Vpr possèdent un motif commun WXXF. Il fut aussi démontré que le motif WXXF de la protéine uracile ADN-glycosylase est impliqué dans son interaction avec Vpr dans le contexte cellulaire. La fusion d'un dimère du motif WXXF avec la protéine chloramphénicol acétyle transférase (CAT) résulte en une activité CAT à l'intérieur du virion en présence de Vpr, démontrant ainsi le rôle d'incorporation de la protéine virale Vpr²⁵. La technique de criblage d'une banque de phages présentoirs implique généralement une interaction protéique basée sur l'affinité de séquence ou de conformation, à la suite de l'exposition de la banque de phages à la cible. Plusieurs rondes de sélection permettent d'enrichir la population de phages interagissant avec la cible (Figure 4, page 7). Parce que l'ADN encodant la protéine ou le peptide de fusion exprimé à la surface du phage est encapsidé dans le virion, il devient très facile de définir les séquences encodant les protéines exprimées à la surface des phages ayant interagi avec la protéine cible. Il est alors possible de procéder à des expériences complémentaires pour confirmer les interactions trouvées par la technique de phages présentoirs²⁶.

Les inconvénients de la technique de présentation de peptides ou de protéines sur phages impliquent, entre autres, la nature chimérique des protéines ou des peptides de fusion. La présence du phage fusionné à la protéine impose la contrainte d'une certaine rigidité et d'une conformation peu représentative d'une condition physiologique. Cette rigidité diminue la probabilité de certaines interactions demandant une flexibilité des protéines impliquées. L'orientation dans laquelle la protéine sera fusionnée influencera aussi sa capacité à interagir avec sa cible. De plus, la technique étant *in vitro*, les protéines impliquées n'auront pas subi les modifications posttraductionnelles propres à une synthèse *in vivo*. Ces modifications peuvent être d'une importance primordiale pour certaines interactions. Malgré la forte diversité des banques de protéines pouvant être présentées à la surface des phages, plusieurs de ces protéines de fusion seront délétères

à la réplication du phage et ne pourront être amplifiées. D'autres auront une conformation favorisant la protéolyse lors de la propagation du phage. Toute méthode impliquant des lavages destinés à éliminer les protéines n'ayant pas interagi avec la cible comporte également le risque de voir disparaître les protéines dont les interactions sont faibles. Les protéines ayant la meilleure spécificité pour la cible ne sont pas nécessairement celles qui ont les plus fortes interactions. De multiples lavages auront comme finalité l'enrichissement de ces dernières, au détriment de multiples protéines plus intéressantes du point de vue de la fonction²⁷. Sans nécessairement constituer un remplacement à d'autres méthodes d'analyse d'interactions protéine-protéine, les banques de phages présentoirs sont cependant un excellent complément. Cette méthode permet le criblage très rapide et exhaustif de centaines de millions de peptides ou de protéines pour leur affinité envers une protéine cible. La présélection, par la technique de phages présentoirs, de peptides ou de protéines montrant une affinité pour la cible peut permettre, par exemple, d'orienter le type de structures de banques qui pourront être réalisées en double-hybride ou par d'autres méthodes *in vivo*.

2.1.2 Les biopuces

De multiples dérivés de la technique des biopuces d'ADN ont été développés récemment, incluant des biopuces d'ADN double brin permettant de détecter et de quantifier les interactions entre ADN et protéines, ainsi que des biopuces de protéomes de divers organismes²⁸. Une technique fréquemment employée consiste à analyser le profil de l'expression génique globale du génome de la cellule dans sa condition normale et dans son état suivant une infection. Les gènes étant transcrits seulement lorsque leur fonction est requise, ces données peuvent être un indicateur important de la fonction d'une protéine cellulaire en réponse à l'infection²⁹. Un exemple de l'utilisation de cette technique appliquée à l'étude de l'influence de la protéine virale accessoire Nef du VIH-1, sur le profil d'expression des protéines de la cellule hôte, a été décrit par Shaheduzzaman S. et coll., en 2002³⁰. La comparaison du profil d'expression sur biopuces d'ADNc de cellules HeLa exprimant Nef par rapport aux mêmes cellules n'exprimant pas cette protéine a permis d'identifier divers gènes cellulaires dont l'expression est modifiée par la présence de Nef. Certaines des protéines identifiées sont reconnues comme importantes pour la réplication virale et la pathogenèse. Cet exemple montre que, par l'analyse indirecte de l'expression

concomitante de protéines, il est possible de définir des interactions protéiques potentielles. Ces interactions devront ensuite être confirmées par des analyses d'interactions physiques directes. D'autres biopuces, les biopuces de protéines, permettent de mesurer plus directement les interactions physiques entre protéines. La méthode généralement adoptée est basée sur le format planaire où les agents de capture (protéines ou peptides) sont fixés sur une surface (PVDF, verre, plastique, etc.). Ceux-ci peuvent alors interagir avec un partenaire (protéine, peptide ou petite molécule) et générer un signal permettant la détection de cette interaction (Figure 5)³¹. Par exemple, la biotinylation des échantillons testés et l'utilisation de streptavidine couplée à une molécule fluorescente peuvent être utilisées pour la détection des interactions entre les protéines de l'échantillon et les protéines cibles fixées à la matrice. Il faut noter que les biopuces de protéines ne ressemblent en rien aux biopuces d'ADN classiques. La taille des biopuces est beaucoup plus grosse et la densité des protéines fixées à la matrice est beaucoup moins importante, ne comportant que quelques dizaines à quelques centaines de dépôts seulement. Outre l'aspect haut débit, le format biopuces permet d'aborder *in vitro* des classes d'interactions difficiles à détecter par des approches classiques plus lourdes à mettre en œuvre, comme le double-hybride.

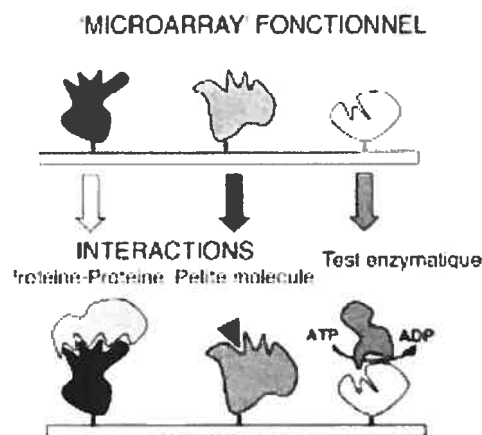


Figure 5. Schématisation de biopuces fonctionnelles impliquant l'analyse d'interactions protéine-protéine. Des protéines natives ou des peptides sont purifiés ou synthétisés, puis fixés à une surface appropriée, formant une biopuce. Ces biopuces recouvertes de protéines sont criblées pour leur affinité envers une banque de protéines ou autres molécules. Avec un système de détection approprié, ce système permet de détecter des interactions protéine-protéine. (Illustration tirée de réf. 31)

Les inconvénients de ce système sont surtout reliés au cadre artificiel propre à toutes les techniques sur support solide. En outre, comme c'est le cas pour les banques de phages présentoirs, les multiples étapes de lavage ont tendance à éliminer toutes les interactions de faible liaison. Aussi, il est difficile de détecter les interactions de protéines peu abondantes dans la cellule. De plus, il n'est pas rare d'éprouver des difficultés lors de la construction des vecteurs d'expression et de la purification de protéines. Celles-ci doivent posséder une conformation adéquate et suffisamment stable pour être représentatives de la conformation naturelle de la protéine retrouvée *in vivo*. L'obtention et la purification en quantité adéquate d'une protéine possédant une conformation active et stable sont difficiles à générer. Finalement, la fixation des protéines à la matrice représente un véritable défi, si l'on considère la variété des propriétés physico-chimiques des protéines³².

2.1.3 Le système de purification sur colonne d'affinité et le spectre de masse

La spectrométrie de masse par MALDI-TOF appliquée aux protéines est une technique permettant l'identification des interactions protéiques sur la base de leur masse³³. Le concept consiste à exprimer, dans la cellule, une protéine d'intérêt, en fusion avec un épitope. Après l'interaction de la protéine de fusion avec la protéine cible, le complexe protéique peut être purifié sur gel ou sur une colonne d'affinité (*TAP*) et ensuite être analysé par spectre de masse. L'énorme avantage de la spectrométrie de masse sur des techniques comme le double-hybride, par exemple, est qu'elle permet de caractériser les complexes protéiques d'ordre supérieur tels qu'ils existent dans la cellule. Aussi, les protéines sont générées dans un contexte *in vivo*, donc physiologique, si l'on fait abstraction de l'épitope fusionné à la protéine appât. Aussi, le fait que les interactions ont lieu dans la cellule, dans un environnement normal, représente un net avantage par rapport aux autres techniques *in vitro* et au double-hybride en levures. Cependant, contrairement aux techniques de double-hybride, cette approche ne permet pas d'appréhender les interactions protéiques individuelles qui conditionnent la formation du complexe. Cette méthode n'indique pas non plus l'affinité de ces interactions. De plus, comme pour toute technique dont l'analyse est *in vitro*, son inconvénient majeur réside dans la nécessité de purifier ces complexes par différentes méthodes physiques auxquelles les interactions les plus faibles ne résistent pas toujours. Il semble donc que la spectrométrie de masse et les techniques de double-hybride – qui sera présenté ci-

dessous – constituent des approches complémentaires dans l'exploration des interactions protéiques³.

2.2 Techniques d'analyse des interactions protéine-protéine *in vivo*

Le survol des techniques *in vitro* a montré que l'analyse exhaustive des interactions protéine-protéine ne peut se limiter à des observations faites dans un environnement artificiel. Il importe donc d'explorer également des avenues *in vivo*. Ces dernières sont plus appropriées à l'étude des interactions entre les protéines virales et les protéines des cellules hôtes. En effet, les protéines y sont exprimées dans un environnement plus physiologique et certaines techniques permettent même l'étude d'interactions dans une situation d'infection virale.

2.2.1 Le système de double-hybride

Le système de double-hybride est une technique génétique qui a été développée chez la levure *Saccharomyces cerevisiae* pour mettre en évidence des interactions protéine-protéine⁴. Ce système, très puissant, flexible et très précis est rapidement devenu l'une des méthodes de choix pour découvrir des protéines interagissant physiquement. Cette méthode repose sur la nature modulaire (qui peut être dissociée) de plusieurs types de facteurs de transcription eucaryotes contenant un domaine de liaison à l'ADN (BD) et un domaine d'activation (AD) qui recrute le complexe protéique de transcription (Figure 6).

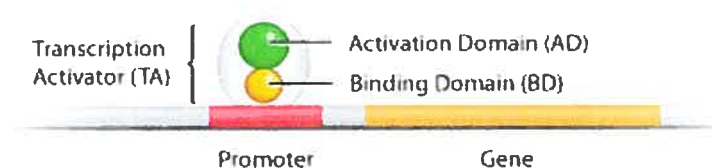


Figure 6. Transcription normale. L'activation de la transcription requiert l'interaction du domaine de liaison à l'ADN (BD) et le domaine d'activation (AD). (Illustration tirée de réf. 34)

Le principe du double-hybride consiste à exprimer dans la cellule deux protéines dont on veut tester l'interaction sous la forme de protéines hybrides fusionnées. L'une est

composée d'une protéine X et du domaine de liaison à l'ADN (X-BD) et l'autre protéine de fusion est composée d'une protéine Y et du domaine d'activation de la transcription (Y-AD). L'interaction physique entre les protéines chimériques X et Y rapproche topographiquement les domaines BD de AD et reconstitue l'activité du facteur de transcription, ce qui permet l'expression d'un gène rapporteur comportant un site de reconnaissance pour le domaine BD utilisé (Figure 7). Aussi, dans le cas où on introduit un gène rapporteur comme l'activation de *Lac Z*, l'addition de X-Gal dans le milieu de culture mène à l'hydrolyse de ce substrat chromogène par la β -galactosidase et colore les colonies positives en bleu. L'intensité de la coloration bleue des colonies, traduisant l'activité β -galactosidase, peut être corrélée à la force de l'interaction, ce qui permet de discriminer les interactions faibles, intermédiaires ou fortes³⁵.

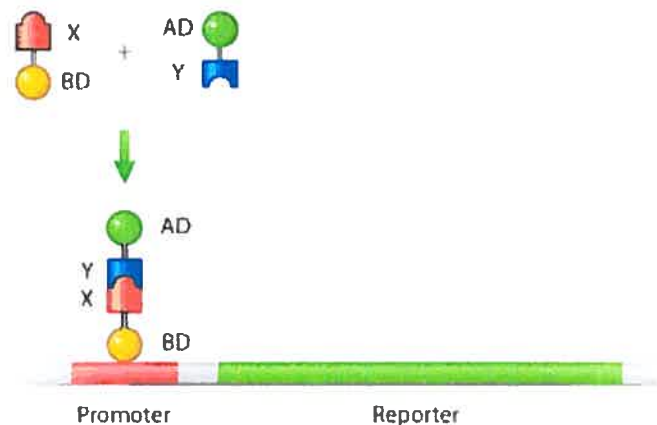


Figure 7. Transcription active dans l'essai double-hybride. La technique du double-hybride mesure les interactions protéine-protéine par le biais de la transcription d'un gène rapporteur. Si la protéine X et la protéine Y interagissent, les domaines de liaison à l'ADN et d'activation de la transcription seront réunis pour former un facteur de transcription fonctionnel permettant l'expression du gène rapporteur. (Illustration tirée de réf. 34)

Cette approche double-hybride a très rapidement connu un grand succès pour la mise en évidence d'interactions entre deux protéines connues, pour la caractérisation de changements phénotypiques suivant la mutation de l'une des protéines étudiées et pour l'établissement de cartes d'interactions d'une protéine connue. De nombreuses déclinaisons sont apparues ces dernières années, illustrant la flexibilité de la méthode. Ainsi, il est possible de calibrer la détection d'interactions d'affinité variable par la variation des degrés d'expression des protéines hybrides. Le nombre et la nature des domaines de

liaison à l'ADN ainsi que la composition des milieux de sélection peuvent être modifiés. Il est aussi d'usage de varier les facteurs de transcription ou les gènes rapporteurs et même d'appliquer le double-hybride en bactéries³⁶ ou en cellules de mammifères³⁷. Le double-hybride est applicable à l'étude de l'interaction de deux protéines d'identité connue³⁸, d'une protéine connue avec les protéines provenant d'une banque génomique ou d'ADNc³⁹ et finalement de protéines issues de deux banques génomiques ou d'ADNc⁴⁰. Plusieurs de ces banques, constituées de l'ADN de divers organismes, sont maintenant commercialisées.

Parmi les avantages du double-hybride se trouve la très grande sensibilité du système qui permet la détection d'interactions ayant une constante de dissociation de 10^{-7} M, ce qui se trouve dans le spectre des interactions les plus faibles d'une cellule. Cette technique permet même la détection de certaines interactions transitoires. Le double-hybride est donc beaucoup plus puissant que la purification sur colonne d'affinité, qui requiert la stabilité du complexe protéique pour les dilutions des lysats cellulaires et les étapes de lavage subséquentes. Cette sensibilité provient d'une panoplie de facteurs, dont la surproduction de protéines *in vivo* et le concept dirigeant toute la réaction dans un seul compartiment : le noyau. Il faut aussi considérer qu'un très grand nombre d'inserts variables peuvent être introduits en même temps dans la cellule et que ce système permet la sélection génétique, qui ne conserve que les clones positifs⁴¹.

Les principales limitations de la technique du double-hybride incluent l'inévitable occurrence de nombreux faux positifs et faux négatifs, dont la causalité est multiple⁴². Diverses analyses de double-hybride génomique montrent qu'environ 50 % des résultats obtenus sont corrects; ce qui implique logiquement que les autres 50 % sont faux. Ceci est dû à la grande sensibilité du système qui augmente le risque de détecter des artefacts. Par exemple, de nombreuses protéines fusionnées au domaine de liaison à l'ADN (BD) activent spontanément la transcription des gènes rapporteurs, indépendamment de toute interaction avec la protéine chimérique liée au domaine d'activation (AD) (ce qui représente environ 10 % des protéines-appâts provenant d'une banque d'ADNc). Il est en partie possible de pallier ce problème par l'utilisation conjointe de deux gènes rapporteurs. En effet, pour que l'interaction soit positive, les clones sélectionnés devront être capables d'activer simultanément les deux gènes rapporteurs bien qu'ils soient munis de promoteurs différents³⁸. L'identification des faux positifs et l'augmentation de l'intervalle de

confiance des résultats obtenus requièrent l'accumulation progressive de nombreuses données expérimentales et des calculs statistiques tenant compte de recouvrements de données issues d'autres méthodes. Ces expériences complémentaires peuvent être basées sur le fait que les protéines impliquées dans les mêmes fonctions sont souvent exprimées de façon concomitante. Ces protéines sont aussi parfois associées à une colocalisation dans des compartiments cellulaires^{43,44,45}. D'autres inconvénients liés à la technique du double-hybride laissent dans l'ombre un grand nombre de protéines qui ne peuvent être détectées par cette méthode. Ainsi, les protéines dont la fusion aux domaines BD ou AD ne permet pas le repliement correct ou les protéines dont les interactions dépendent de modifications posttraductionnelles ne pourront pas être détectées. En effet, ces modifications, importantes pour la fonction protéique, ne se produisent pas chez la levure. Afin de pallier ce problème, la compagnie Stratagene a mis sur le marché, en 2002, un système de double-hybride exprimé dans des cellules de mammifères³⁷. Le principe demeure le même que le double-hybride en levures, à cette exception que le test détecte l'activation de la transcription d'un gène rapporteur comme la luciférase ou la β -galactosidase située en aval des multiples sites de liaison à l'ADN GAL4. Les protéines chimériques sont constituées de GAL4 BD-protéine proie et NF- κ B AD-protéine cible. Un plasmide rapporteur est transfecté concomitamment avec les protéines de fusion pour la détection des interactions positives (Figure 8).

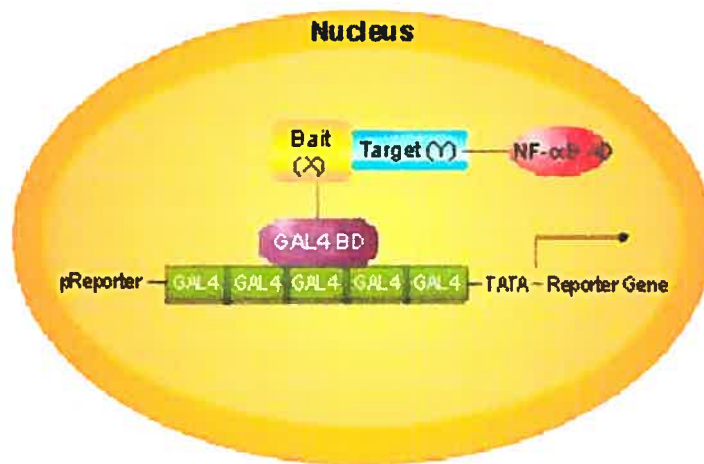


Figure 8. Double-hybride en cellules de mammifères. Les protéines-appâts sont fusionnées au domaine de liaison à l'ADN et les protéines-proies sont fusionnées au domaine d'activation de NF- κ B. Ces protéines chimériques sont transfectées simultanément avec un gène rapporteur et permettent la détection d'interactions protéine-protéine en cellules de mammifères. (Illustration tirée de réf. 37)

Parmi les interactions protéiques qui ne seront pas détectées figurent aussi les protéines dont l'expression est toxique pour la cellule et celles ne pouvant être localisées dans le noyau. Celles-ci sont principalement les protéines membranaires, représentant 40 % des protéines cellulaires totales, et les protéines de sécrétion⁴¹. Une solution permettant de réduire le nombre de ces résultats faussement négatifs consiste à utiliser des domaines protéiques, plutôt que la protéine complète, dans la construction des protéines de fusion. En effet, des résultats expérimentaux ont montré que les interactions sont favorisées lors de l'utilisation de fragments de protéines. Dans ce cas, les protéines de fusion sont plus petites et certains domaines de la protéine, auparavant inaccessibles, sont maintenant exposés⁴². Les dernières années ont aussi vu la conception d'une multitude de techniques dérivées du double-hybride et visant à contourner les divers inconvénients associés au double-hybride. Parmi celles-ci, on retrouve l'intégration de promoteurs inductibles ou encore l'utilisation de systèmes de complémentation protéique qui s'affranchissent de l'étape d'activation de la transcription d'un gène rapporteur. L'un de ces systèmes consiste en la reconstitution de chimères de la protéine ubiquitine (Figure 9)⁴⁶. D'autres méthodes, encore, se basent sur la reconstitution de deux parties de la protéine fluorescente GFP.

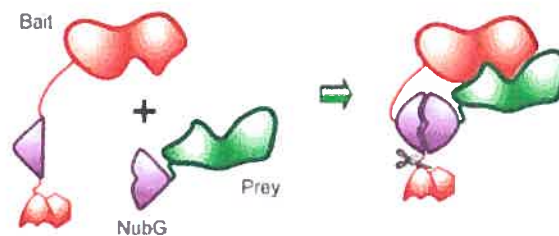


Figure 9. Principe de l'essai *Split-ubiquitine*. La protéine-appât est fusionnée au domaine Cub lié à une protéine rapporteuse. La protéine-proie est en fusion avec le domaine NubG muté. Lorsque les protéines-proies et appâts interagissent, les domaines Cub et NubG sont rapprochés topographiquement et reconstituent l'ubiquitine. Les ubiquitine protéases reconnaissent cette configuration et coupent le lien liant la protéine rapporteuse au domaine Cub. (Illustration tirée de réf. 47)

Les expériences de double-hybride ont mis en évidence un très grand nombre d'interactions entre protéines. Elles ont aussi fourni de nombreuses informations sur la fonction de ces protéines dans les mécanismes moléculaires. Au-delà de la simple observation des interactions binaires, ce type d'analyse se révèle utile dans la prédiction de la fonction d'un groupe de protéines de même famille. Cependant, il convient de ne pas oublier les nombreuses limitations de la technique du double-hybride qui laissent dans

l'ombre un grand nombre d'interactions importantes. Entre autres, il faut se rappeler que quel que soit le système dérivé du double-hybride, les protéines sont liées physiquement aux domaines BD et AD, ce qui leur confère une rigidité et une conformation qui, dans ces conditions, ne sont pas représentatives d'une situation physiologique et nuisent à l'interaction des protéines étudiées. Il faut aussi considérer que l'une des limitations les plus fondamentales des diverses techniques de double-hybride réside dans le fait que seules les interactions binaires sont analysées, ce qui ne permet pas, dans la plupart des cas, de cartographier les différentes régions d'interaction et de modéliser l'architecture d'un complexe multiprotéique.

Dans le cadre d'une étude sur les interactions protéiques retrouvées dans des cellules humaines infectées par le virus VIH-1, l'utilisation de ce type de banque peut se révéler très utile. La caractérisation d'interactions entre deux protéines sera possible en criblant les protéines provenant de banques génomiques ou d'ADNc humains et de VIH-1. Cependant, certaines études particulières ne pourront être réalisées avec ce système. Par exemple, les protéines qui ne sont exprimées qu'en réponse à une induction engendrée par l'infection virale de la cellule ou celles qui ne s'associent que sous la forme de complexes protéiques. Finalement, la rareté de certaines protéines impliquées dans les mécanismes d'infection, de réplication du virus ou de défense antivirale pourrait ne pas être compatible avec le système du double-hybride. En effet, les banques d'ADNc contiennent environ 3×10^6 ADNc indépendants. Ces ADNc proviennent d'ARNm abondants dans la cellule. Les ADNc dont la fréquence est rare sont sous-représentés dans la banque d'ADNc. Il pourrait alors être difficile d'en apprécier les interactions parmi l'ensemble des protéines exprimées. Ce système fournit des informations sur les interactions de protéines abondantes. Cependant, il pourrait ne pas être approprié à l'analyse d'événements rares ou de protéines qui ne sont pas exprimées constitutivement, donc non représentées dans les banques d'ADNc commerciales.

Pour pallier la diversité limitée rencontrée dans le système de double-hybride en levures, certains utiliseront une technique alternative, basée sur l'expression d'une banque de peptides. Ainsi, Sang-Hyun Oark et Ronald T. Raines ont développé une technique tirant avantage de la nature aléatoire d'une banque de nonapeptides. La diversité théorique d'une banque de nonapeptides est de plus de 10^{11} . Les restrictions imposées par les techniques de clonage leur ont permis d'obtenir une banque d'une diversité de 3×10^8

peptides. Une expérience préliminaire a eu pour résultat 300 peptides capables d'inhiber la formation de la structure quaternaire de la protéase du VIH-1, essentielle à son activité catalytique. Cette technique montre à quel point une forte diversité de la banque des protéines criblées est essentielle à la découverte d'événements rares. Si l'on considère qu'un peptide sur 10^6 interfère dans une voie métabolique donnée, il serait peu probable qu'une banque d'ADNc permettant l'expression d'environ 10^6 protéines parvienne à cribler efficacement un phénotype impliquant des protéines rares. Basée sur un principe légèrement différent du double-hybride d'ADNc en levures, la banque de peptides implique l'interaction d'un peptide inconnu avec des partenaires protéiques, menant à la perturbation d'une voie métabolique donnée. La sélection génétique et l'amplification des clones dans lesquels le phénotype est altéré permettent la récupération et la caractérisation des peptides impliqués. Une recherche par homologie de séquence et l'immunoprécipitation de ces peptides permettent d'identifier les protéines partenaires. Les peptides de la banque ne sont pas restreints à l'interaction avec une protéine précise, mais peuvent intervenir dans une panoplie d'interactions protéine-protéine en aval et en amont d'une voie métabolique cible. Cette particularité permet une vue d'ensemble des diverses protéines impliquées dans les voies métaboliques qui mènent à un phénotype particulier. Ainsi, une caractéristique importante des banques de peptides, outre le fait qu'elles permettent l'identification d'interactions protéine-protéine, est la découverte de peptides capables de moduler ces interactions. Ces informations peuvent enfin servir de base à l'élaboration de molécules pharmaceutiques.

2.2.2 Analyse d'interactions par la technologie de transfert d'énergie entre fluorochromes (*FRET*)

Une variante intéressante de la technique du double-hybride utilise la technologie de transfert d'énergie résonnant de type Förster (*FRET*) entre des fluorochromes fusionnés aux protéines. La technique de *FRET* est un procédé basé sur la variation de l'énergie émise. Par exemple, lorsqu'un fluorochrome comme GFP est excité, elle transmet son énergie excédentaire à une autre protéine fluorescente comme YFP, si ces deux protéines sont situées à une distance de moins de 60 Å. Les avantages du *FRET* sont nombreux et incluent le fait que les analyses peuvent être effectuées *in vivo* dans n'importe quel type cellulaire. Les interactions ont lieu dans les compartiments cellulaires où elles sont

normalement retrouvées et dans un environnement physiologique. La présence de fluorochromes associés aux protéines permet aussi l'identification des diverses protéines évoluant dans le même compartiment cellulaire et apporte une donnée supplémentaire en faveur d'interactions impliquant ces protéines. En principe, il devrait être possible de cribler une banque de protéines cellulaires fusionnées à un fluorochrome, en interaction avec une banque de protéines virales fusionnées à un second fluorochrome. En pratique, cependant, le nombre de faux négatifs ne permet pas de faire de telles analyses. Ces faux négatifs proviennent de l'orientation inadéquate des protéines fusionnées aux fluorochromes ne permettant pas les interactions protéine-protéine⁴⁸. Il arrive aussi que le bruit de fond de fluorescence émis par les molécules excitées ne permette pas la détection du *FRET*. Ce phénomène est observé lorsque les protéines fusionnées à un fluorochrome sont exprimées à un plus haut niveau que les protéines fusionnées à l'autre fluorochrome.

L'utilisation de la technologie de *FRET* pour étudier des interactions protéine-protéine *in vivo* est de plus en plus populaire. Par exemple, il a été montré par des observations *in vivo* résultant de cette méthode que la multimérisation de la protéine Gag du VIH-1 a principalement lieu à la membrane plasmique de la cellule, mais que d'autres interactions Gag-Gag ont lieu à des sites distincts. Des interactions Gag-Tsg101 redirigeant cette protéine cellulaire vers la membrane plasmique ont aussi été observées. Plus important, cette expérience démontre que le système de *FRET* est approprié à l'étude d'interactions protéine-protéine dans un contexte *in vivo* en cellules de mammifères⁴⁹.

2.2.3 Expression d'une banque de peptides en cellules de mammifères par le biais d'un vecteur rétroviral

Au-delà de l'étude des interactions directes protéine-protéine, la caractérisation des voies métaboliques menant à un phénotype donné est d'une plus grande importance encore. Une telle analyse est généralement rendue possible par la perturbation de ces voies métaboliques. L'altération des processus cellulaires peut être effectuée à l'aide de mutants ou par déplétion de protéines impliquées dans ces voies métaboliques, lorsque l'identité de ces protéines est connue. Une autre technique consiste à introduire dans la cellule une banque aléatoire de peptides qui interagissent avec les protéines impliquées dans le

processus cellulaire et en modifieront la fonction. Un système de sélection génétique permettra d'isoler et d'amplifier uniquement les cellules au phénotype altéré. La caractérisation des peptides ayant altéré la voie métabolique, une recherche par homologie de séquence et une immunoprécipitation permettent alors d'identifier les protéines dont la fonction a été altérée par l'interaction avec le peptide. La caractérisation de toutes les protéines ainsi identifiées permet de dresser une carte d'interactions protéiques et donne une idée générale de l'ensemble des protéines essentielles à un processus cellulaire particulier. Une telle méthode, combinant divers éléments des techniques *in vitro* et *in vivo* décrites plus haut, a été développée dans le laboratoire du Dr Gary Nolan, de l'Université de Stanford⁵⁰. Cette méthode utilise la très grande efficacité des rétrovirus à introduire leur matériel génétique dans la cellule hôte pour créer des lignées de cellules exprimant constitutivement une banque de peptides aléatoires. Le principe consiste à cloner une banque de peptides dans un vecteur rétroviral. Ce dernier sert à la production de particules rétrovirales qui introduiront la banque de peptides dans les cellules cibles. Le pseudotypage des vecteurs rétroviraux avec l'enveloppe VSV-G (enveloppe du virus de la stomatite vésiculaire, glycoprotéine G) permet d'infecter une grande variété de types cellulaires. De plus, cette enveloppe apporte au virus une stabilité permettant ainsi de le concentrer par ultracentrifugation. L'expression stable de la banque de peptides permet l'interaction *in vivo* avec les protéines cellulaires et en modifiera la fonction, menant à un changement du phénotype (Figure 10).

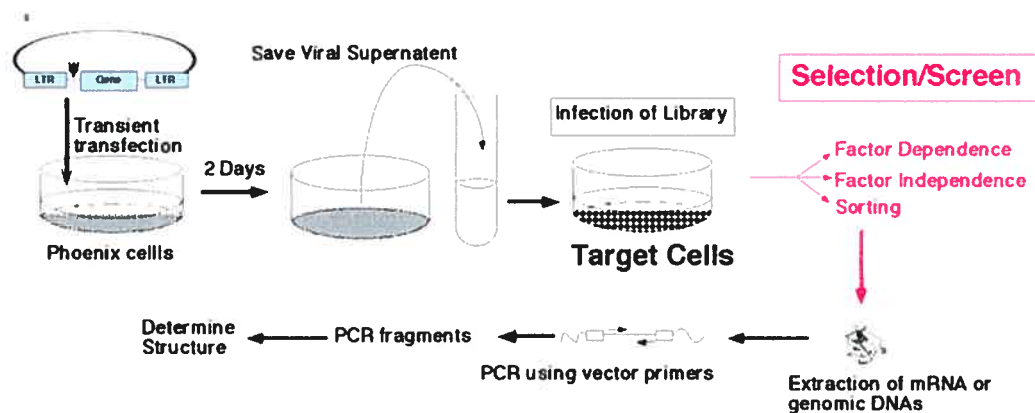


Figure 10. Schématisation du criblage d'interactions protéine-protéine par sélection génétique à l'aide de la banque de peptides. Des cellules sont infectées avec des vecteurs rétroviraux afin de générer une lignée cellulaire exprimant constitutivement les peptides. Une sélection génétique permettra la caractérisation des peptides qui auront interagi avec des protéines cellulaires et en auront altéré le phénotype. (Illustration tirée de réf. 51)

À partir de cet outil, divers autres types de banques de peptides ont été produits, certaines en fusion avec la protéine CCR5 pour une situation cellulaire à la membrane⁵², d'autres en conformation cyclique⁵³ et d'autres, encore, sous forme de peptides libres repliés en boucle afin d'éviter une dégradation rapide du peptide par des peptidases cellulaires ou le transport dans le réticulum endoplasmique par TAP1 et TAP2 (Figure 11)⁵⁴.

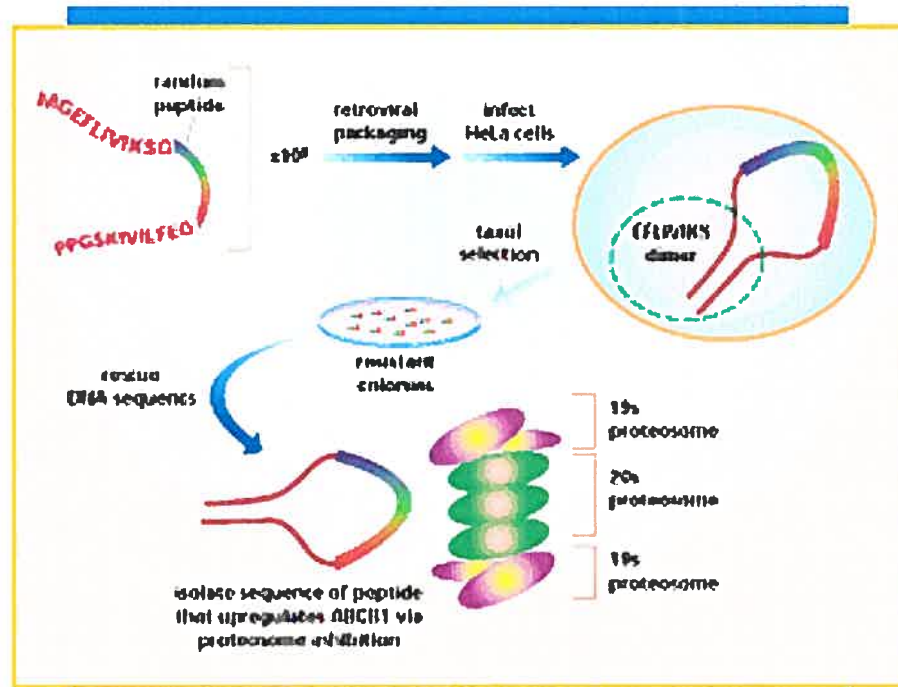


Figure 11. Présentation d'une banque de peptides sous forme de boucle, par introduction d'une séquence SKIVILF. Cette conformation évite la dégradation du peptide, mais entraîne une rigidité du peptide présenté. (Illustration tirée de réf. 54)

L'utilisation de cette banque de peptides a permis l'identification de diverses molécules d'intérêt, dont la caractérisation de peptides qui augmentent le niveau de ABCB1 (MDR1) en inhibant la dégradation d'un facteur de transcription par le protéasome impliqué dans la résistance de cellules cancéreuses à l'action du TAXOL (Figure 12)⁵⁰. Cet exemple illustre toute la puissance du criblage avec une banque de peptides. À partir d'un phénotype donné – ici la résistance d'une cellule à l'action du Taxol, un agent toxique utilisé dans le traitement du cancer – il est possible de cribler tous les mécanismes impliqués dans ce phénomène. Normalement, des cellules traitées au Taxol entrent en apoptose. Ces

mêmes cellules, infectées avec un vecteur rétroviral qui leur confère la propriété d'exprimer constitutivement une banque de peptides (un peptide par cellule), pourront réagir différemment en présence du Taxol. En effet, dans certaines cellules, le peptide exprimé aura la conformation et la séquence adéquates pour interagir avec l'une ou l'autre des protéines impliquées dans le mécanisme de mort cellulaire en présence du Taxol. Cette interaction perturbera la fonction de cette protéine et permettra la survie de la cellule. La survie de ces quelques cellules permettra que l'on puisse les isoler et les amplifier afin de caractériser le ou les peptides qui auront permis la survie cellulaire en présence de Taxol. Par une recherche d'homologie de séquence et une immunoprécipitation, il est possible d'identifier les protéines avec lesquelles les peptides ont interagi. Il est donc permis de penser que ces protéines ont un rôle important à jouer dans le phénotype initialement observé, ici la résistance cellulaire au Taxol, hypothèse qui peut être confirmée par des expériences complémentaires (interférence à ARN, double-hybride, etc.).

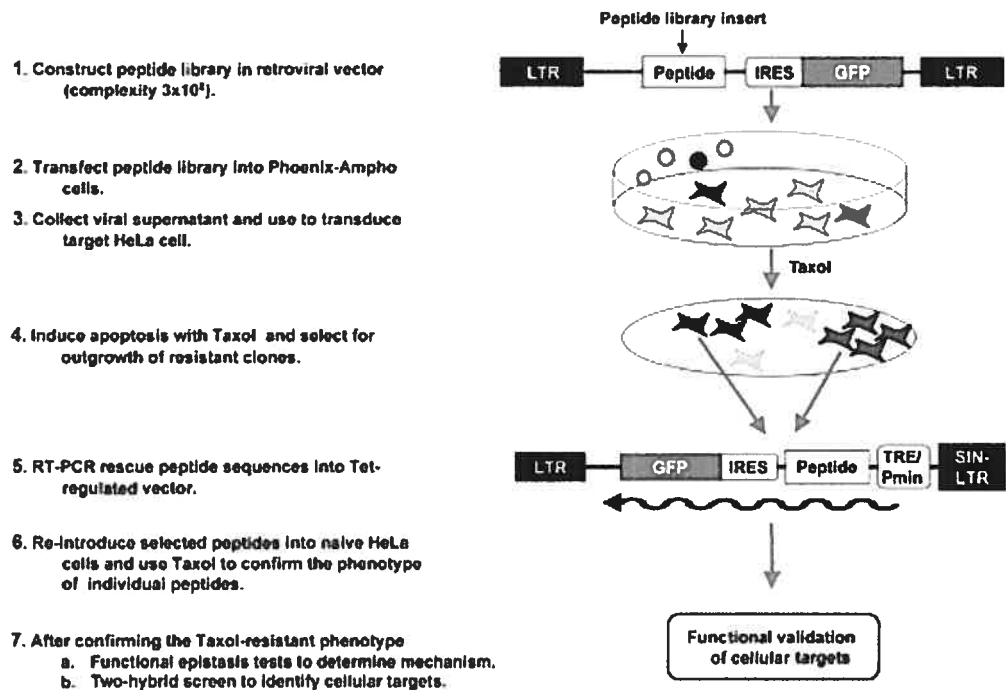


Figure 12. Criblage de banque de peptides en cellules de mammifères sous sélection génétique permettant la caractérisation de peptides dont l'interaction avec une protéine cible en perturbera le phénotype cellulaire. (Illustration tirée de réf. 50)

Le criblage par banque de peptides est une façon unique et puissante d'étudier un phénomène particulier. En effet, à partir d'un simple phénotype, il devient possible de définir les protéines qui sont responsables de ce phénotype (cibles thérapeutiques potentielles), de caractériser des peptides qui altèrent cette voie métabolique (agents thérapeutiques) et d'identifier des protéines impliquées dans diverses voies métaboliques, mais toutes reliées au phénotype étudié (compréhension des métabolismes fondamentaux). Un outil de recherche aussi puissant ne se retrouve pas dans les systèmes de criblage classiques comme le double-hybride ou les biopuces. Cette banque possède certaines caractéristiques intéressantes qui correspondent aux critères recherchés pour l'étude d'interactions protéine-protéine dans le cadre d'une infection virale. Premièrement, la diversité de séquences de la banque de peptides qu'il est possible d'obtenir ne dépend que de l'efficacité de la transformation lors de l'étape de clonage. Un peptide aléatoire constitué de neuf acides aminés encode une diversité théorique de plus de 10^{11} peptides différents. Une telle diversité ne peut être obtenue avec une banque d'ADNc, d'une diversité moyenne de 10^6 . De plus, la banque de peptides possède une distribution aléatoire, contrairement à l'ADNc qui encode les protéines les plus abondantes de la cellule au détriment de protéines rares, mais importantes. Une telle diversité pourra se traduire par la découverte d'interactions à des niveaux beaucoup plus fins, permettant la caractérisation de protéines exprimées même très faiblement. L'élément le plus important de cette technique, en rapport avec l'étude des interactions virus-cellule hôte, est la possibilité d'analyser les interactions protéine-protéine dans le contexte d'une infection virale. Si les méthodes décrites précédemment permettent d'étudier les interactions de banques de protéines virales en présence de banques de protéines cellulaires, aucune ne permet d'interférer directement dans le processus d'une infection virale réelle. Le processus d'infection déclenche des réponses cellulaires qui ne peuvent être étudiées que dans un contexte physiologique. L'établissement d'une lignée cellulaire exprimant constitutivement une banque de peptides et l'infection subséquente de ces cellules, avec des virus rapporteurs, permettent l'étude de voies métaboliques importantes à la réplication virale et à la riposte cellulaire.

La technique initialement utilisée pour la construction d'une banque de peptides consiste à compléter, à l'aide d'une enzyme Klenow, un polynucléotide encodant la banque de

peptides dégénérés. Ce nouvel ADN double brin peut alors être introduit dans un vecteur rétroviral linéarisé, par ligature des extrémités (Figure 13, page 24)⁵⁵.

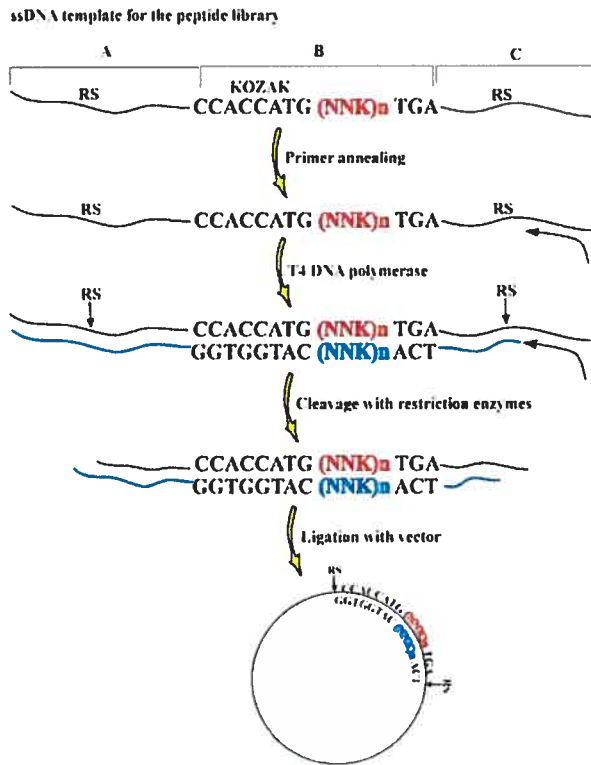


Figure 13. Clonage de la banque de peptides dans un vecteur rétroviral. (Schéma tiré de réf. 55)

Il a été admis, par l'instigateur même de cette technique, que le clonage par insert-vecteur d'un insert d'aussi petite taille, comme décrit à la Figure 13, est une étape particulièrement difficile à réaliser⁵⁶, une difficulté qui se traduit par l'exclusivité de l'utilisation de ces banques de peptides en vecteur rétroviral dans les laboratoires du Dr Nolan. Afin de contourner les problèmes liés à cette méthode, une technique alternative de clonage fut décrite par ce même groupe de recherche. Celle-ci consiste en l'intégration, à un site déterminé, d'un polynucléotide dans un vecteur rétroviral sous forme d'ADN simple brin (Figure 14)⁵⁶. Cette technique est basée sur la technique de la mutagenèse dirigée qui est couramment utilisée pour l'introduction de mutations dans un plasmide d'ADN double brin. L'efficacité de la mutagenèse dirigée (1 % à 80 %, selon la méthode utilisée)⁵⁷ est appropriée lorsque l'on vise l'obtention de quelques clones positifs qui seront amplifiés subséquemment. Cependant, cette technique se révèle inadéquate lorsque l'on vise une diversité de centaines de millions de clones différents. La méthode consistant à utiliser une matrice d'ADN simple brin, générée grâce au bactériophage M13KO7, a permis à

l'équipe de recherche du Dr Nolan d'introduire la séquence d'une banque de peptides aléatoires de quatre acides aminés avec une efficacité permettant l'obtention de 10^6 clones indépendants.

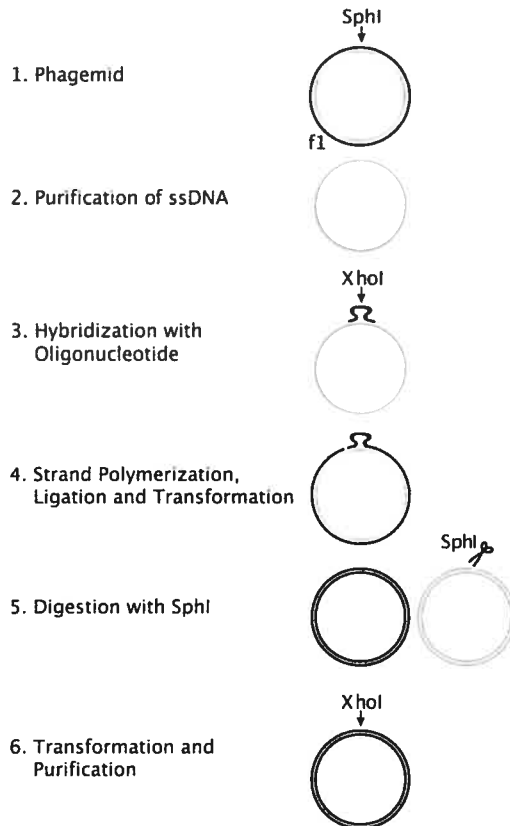


Figure 14. Conception d'une banque de peptides par intégration site-spécifique d'un polynucléotide sur un vecteur rétroviral sous forme d'ADN simple brin. (Illustration tirée de réf. 56)

Malgré cette intéressante solution de rechange au clonage classique, la diversité obtenue avec cette méthode est inférieure à celle obtenue avec la méthode précédente de ligature insert-vecteur. De plus, cette diversité n'a été obtenue que pour une banque de quatre acides aminés. Ce même article mentionne qu'une banque de dix acides aminés a aussi été réalisée, mais les données concernant la diversité obtenue avec une telle banque n'ont pas été rapportées. On remarquera, à l'étape 4 de la Figure 14, que le produit engendré par l'hybridation du polynucléotide de la banque dégénérée à la matrice d'ADN simple brin forme une bulle d'ADN ne trouvant pas son équivalent sur le brin complémentaire. Il est permis de se demander si, en présence d'une telle bulle d'ADN, les mécanismes de réparation bactériens ne procéderont pas à des réarrangements massifs,

voire même à la délétion de la banque de peptides. En effet, le mécanisme de réparation bactérien connu sous le nom de « *Mismatch DNA repair* », ou MMR, élimine le brin non méthylé de l'ADN, sur une portion entourant la bulle d'ADN, et utilise le brin méthylé pour la synthèse correcte de l'ADN (Figure 15)⁵⁸. Dans ce cas-ci, seul le brin de la matrice produit en bactéries est méthylé. La banque de peptides nouvellement introduite par amplification pourrait alors être éliminée.

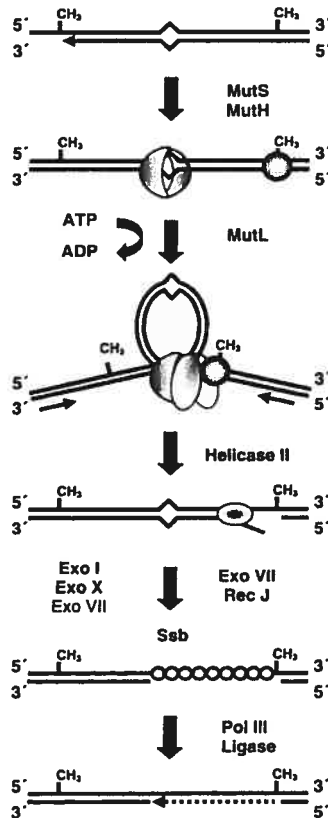


Figure 15. Mécanisme MMR de réparation bactérien en présence d'une bulle d'ADN non complémentaire. (Illustration tirée de réf. 58)

3. OBJECTIFS DE NOTRE PROJET DE RECHERCHE

L'objectif ultime de ce projet de recherche consiste à développer un outil permettant d'identifier et de caractériser les protéines cellulaires dont les interactions avec les protéines virales ont un impact sur le cycle de réplication lors d'une infection par le VIH-1

dans des cellules humaines. Cette étude est effectuée au niveau cellulaire et vise la compréhension des voies métaboliques importantes pour la réplication virale. Elle tente aussi de décrire les mécanismes de défense internes acquis par les cellules pour se prémunir des infections virales. Ces connaissances fondamentales sont indispensables en amont de certaines recherches appliquées. Dans le contexte précis de l'étude d'interactions protéiques virus-hôte, il est important d'apprécier ces interactions en se rapprochant au plus près des conditions physiologiques, c'est-à-dire dans le cadre d'une infection virale, *in vivo*, en cellules de mammifères.

La technique la plus appropriée et la plus puissante pour la caractérisation d'interactions protéine-protéine nous a semblé être un système de sélection génétique par inhibition peptidique de voies métaboliques. Nous avons vu que les peptides ont la capacité, par affinité de séquence ou de conformation, d'interagir avec des protéines cellulaires et d'en altérer la fonction, ce qui génère un phénotype typique (Figure 16).

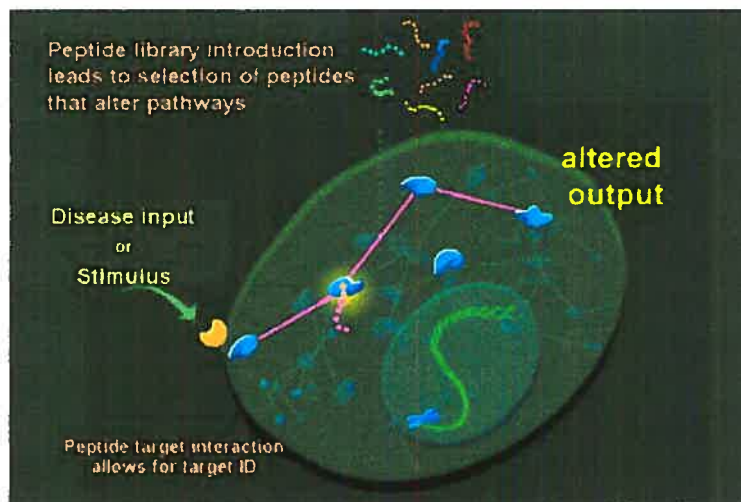


Figure 16. L'introduction d'une banque de peptides dans une cellule mènera à une perturbation de son phénotype. (Illustration tirée de réf. 59)

Un avantage important lors de l'utilisation d'une banque de peptides est que seuls ceux-ci seront en fusion avec un épitope, ce qui permettra de les récupérer et de les caractériser. Les protéines impliquées dans le phénotype étudié sont, elles, sous leur forme naturelle. Un second avantage de la banque de peptides est que sa séquence aléatoire n'est pas biaisée par des différences d'expressions ou de sélection naturelle comme le sont les

ADNc. Toutefois, ce même caractère aléatoire implique que la séquence peut contenir l'un des trois codons de terminaison (TAG, TAA et TGA). Si ceux-ci apparaissent dans le peptide, ils en termineront prématurément la séquence. Une configuration NNK permet d'éliminer l'occurrence de deux des trois codons non-sens (ocre et opale). N correspond à n'importe quel des quatre nucléotides (A, T, C ou G) et K est T, C ou G. Considérant qu'il y a vingt codons différents, des peptides constitués de neuf acides aminés aléatoires permettront, en théorie, la composition de plus de 5×10^{11} nonapeptides différents (20⁹). Cependant, quelle que soit la configuration de la séquence que l'on choisisse, il est pratiquement impossible d'obtenir une banque de peptides vraiment aléatoires. Un biais de séquence est toujours présent. En effet, certains acides aminés sont représentés par plusieurs codons alors que d'autres sont encodés par un seul. De plus, certains nucléotides ont une affinité plus forte pour la matrice de synthèse et seront représentés majoritairement lors de la génération des peptides dégénérés. Ainsi, le choix du nombre d'acides aminés composant la banque de peptides est généralement le résultat d'un compromis entre une grande diversité théorique et une longueur favorisant des interactions avec la protéine cible. La littérature rapporte l'utilisation de banques de peptides variant généralement entre 4 à 18 acides aminés. Enfin, la nature aléatoire des peptides entraîne aussi des désavantages. Certains peptides seront toxiques pour la cellule, d'autres auront une conformation entraînant leur dégradation, leur sécrétion ou leur séquestration dans un compartiment cellulaire. La forte diversité de la banque de peptides a pour fonction de compenser pour ces pertes inévitables.

Certaines conditions expérimentales sont requises pour une étude des interactions protéiques entre un virus et la cellule hôte dans le cadre d'une infection virale. Ces conditions sont la possibilité d'immunoprécipiter le peptide et les protéiques partenaires et l'expression des peptides *in vivo* en cellules de mammifères. Aussi, il est important de pouvoir sélectionner les clones cellulaires exprimant la banque de peptides par le biais d'un gène de résistance à un antibiotique. La méthode pour générer une banque de peptides se rapprochant le plus près de ces conditions consiste en l'utilisation de la technologie de transfert de gènes par des vecteurs rétroviraux encodant la banque de peptides. Ces vecteurs, lorsque complétés avec les protéines structurales Gag, Pol et Env, génèrent des particules rétrovirales contenant l'ARN du vecteur. Par infection des cellules cibles, le vecteur rétroviral encodant la banque de peptides sera rétrotranscrit et

intégré dans le génome de ces cellules. Les cellules infectées exprimeront alors la banque de peptides (Figure 17). Cette technique permet d'obtenir une banque de peptides exprimés constitutivement en cellules de mammifères, telle que décrite au point 2.2.3 de l'introduction.

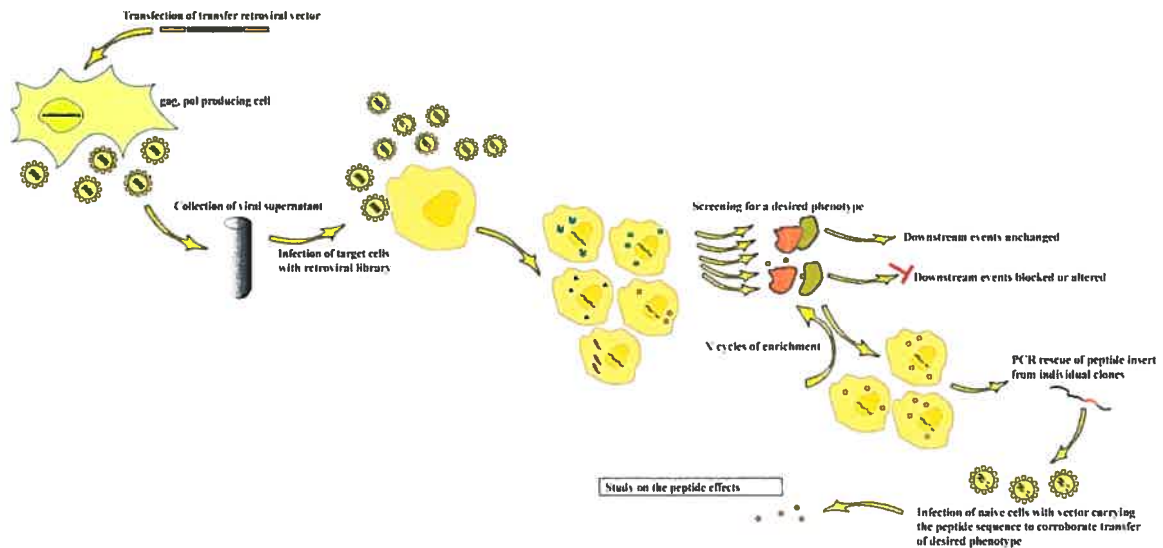


Figure 17. Étapes du criblage par interférence d'une banque de peptides dans les mécanismes d'interactions protéine-protéine, *in vivo*, en cellules de mammifères. (Illustration modifiée de réf. 55)

Dans le cadre de son projet de maîtrise, Alexandre Ostiguy, un membre de notre laboratoire, entreprit le clonage d'une banque de nonapeptides dans le vecteur rétroviral pBAGE HAGIZ, selon la méthode insert-vecteur. Les difficultés inhérentes à cette technique ainsi que la conception bicistronique du vecteur utilisé n'ont pas permis l'obtention d'une diversité clonale satisfaisante. Le nombre de clones indépendants obtenus fut d'environ 10^4 peptides⁶⁰, un nombre insuffisant pour ce type d'étude d'interactions protéine-protéine. En effet, les expériences réalisées par divers groupes utilisant des banques de peptides, que ce soit par phages présentoirs, par double-hybride en levures ou par l'utilisation de vecteurs rétroviraux, ont montré qu'environ un peptide sur un million interagira avec la protéine ciblée lors de l'étude d'un métabolisme particulier. Ceci nous conduit à considérer que la diversité finale de notre banque de peptide devra être d'au moins 10^6 peptides différents si nous voulons être en mesure d'obtenir au moins un clone positif lors de notre étude. Il serait présomptueux de considérer qu'une banque de peptides de seulement 10^6 représentants permettra de récupérer le peptide candidat.

En effet, les diverses étapes de clonage, de transfection, d'infection et d'amplification des clones ne sont pas efficaces à 100 % et impliquent même des pertes quelques fois importantes de la diversité théorique de la banque. Lors du clonage d'une banque de peptides par la méthode insert-vecteur, le groupe du Dr Nolan a obtenu une diversité clonale de 3×10^8 peptides présentés sous forme de boucle. De ces 3×10^8 peptides criblés, seulement 56 colonies ont pu être isolées, desquelles 8 clones seulement encodaient un peptide interagissant effectivement dans la voie métabolique ciblée (résistance au Taxol)⁵⁰. La diversité clonale à atteindre n'est pas une valeur absolue; elle est essentiellement déterminée par l'étape de transformation bactérienne avec le plasmide encodant la banque. Selon le type de bactéries compétentes, il est possible d'obtenir des efficacités de transformation de plus de 10^8 colonies/ug d'ADN circulaire transformé. Cette valeur est représentative de la diversité que l'on peut espérer obtenir, à la suite de quelques transformations d'un ADN de bonne qualité.

Les résultats obtenus au sein de notre laboratoire par la technique insert-vecteur nous ont obligé à repenser l'approche du clonage et la conception générale du vecteur rétroviral, de façon à générer une banque de nonapeptides fonctionnels dont la diversité se compte en centaines de millions d'exemplaires distincts. Deux options sont possibles : tenter le clonage de la banque de peptides selon la technique d'intégration d'un polynucléotide sur un vecteur rétroviral sous forme d'ADN simple brin, comme décrite au point 2.2.3 de l'introduction, ou alors créer une méthode de clonage originale. Divers éléments nous amènent à favoriser cette deuxième option. Premièrement, le seul exemple de clonage réussi d'une banque de peptides par la méthode des ADN simple brin a permis d'obtenir une diversité relativement faible de 1×10^6 clones. Cette diversité fut obtenue pour le clonage d'une banque de quatre acides aminés, les résultats de la banque de décapeptides n'ayant pas été rapportés dans la littérature. Ensuite, nous devons considérer la possibilité d'un mécanisme de réparation bactérien affectant l'intégration de longs ADN dégénérés, comme décrit plus haut.

Notre premier objectif est donc de conceptualiser une technique originale qui permettra d'obtenir une diversité de peptides adéquate pour un système de criblage fonctionnel. Les diverses méthodes utilisées afin d'atteindre cet objectif seront décrites à la section « Résultats ».

Plus spécifiquement, les étapes du projet consistent à :

1) Cloner une banque de peptides fusionnés à l'épitope HA et à un gène de résistance. Cette construction sera insérée dans le vecteur rétroviral pBABE.

2) Établir une lignée cellulaire stable exprimant la banque de peptides. Pour réaliser cette étape, des échantillons représentatifs de l'ensemble de la diversité de la banque peptidique sont utilisés afin de transfecter des cellules HEK 293T en conjugaison avec l'ADN encodant les gènes *gag-pol* du MoMuLV ainsi que les protéines d'enveloppe VSV-G. Ceci permet d'obtenir des banques de rétrovirus pseudotypés, encodant l'ensemble de la diversité de la banque peptidique. Ces virus sont utilisés pour établir, par infection, des lignées de cellules de mammifères exprimant constitutivement la banque de peptides. Le gène de résistance à l'antibiotique fusionné aux nonapeptides permet de sélectionner les clones ayant intégré les peptides au génome cellulaire.

Enfin, l'efficacité de ce nouvel outil qu'est la banque de nonapeptides devra être validée. Cette validation pourrait consister à évaluer la capacité de la banque de peptides à interférer dans un processus cellulaire connu, par exemple la phosphorylation du ganciclovir dans des cellules exprimant la thymidine kinase. Ultérieurement, ce nouvel outil de sélection génétique pourra être appliqué au contexte d'une infection de cellules par le VIH-1 afin de sélectionner, d'identifier et de caractériser les séquences peptidiques interagissant dans une voie métabolique essentielle à la réplication virale ou à la défense cellulaire.

Matériel et méthodes

1. BIOLOGIE MOLÉCULAIRE

1.1 Construction des vecteurs rétroviraux de base

Afin d'établir une lignée de cellules de mammifères exprimant constitutivement une banque de nonapeptides, nous avons construit un vecteur rétroviral basé sur le vecteur pBABE HAGIZ (carte du vecteur n° 2, en annexe). Ce vecteur fut développé par Alexandre Ostiguy, un étudiant à la maîtrise au sein de notre laboratoire. Ce vecteur est lui-même conçu à partir du vecteur rétroviral pBABE-puro modifié par le Dr S. Goff⁶¹ et initialement développé par le Dr JP. Morgenstern⁶² (carte du vecteur n° 1, en annexe). Le but de cette étape de clonage est de créer des vecteurs rétroviraux qui serviront de matrice à l'introduction d'une banque de nonapeptides, sous la forme de polynucléotides aléatoires de séquence (NNK)₉.

Parmi les éléments qui ont été ajoutés ou modifiés au vecteur rétroviral pBABE HAGIZ se trouve une protéine de fusion composée de l'épitope HA, du gène de résistance à la puromycine ou à la zéocine et d'un lien souple (Ser-Gly)₃. Les gènes de résistance à la puromycine ou du gène de résistance à la zéocine sont donc fusionnés en C terminal et N terminal. Ne sachant pas lequel de ces gènes restera fonctionnel dans cette configuration, nous avons construit les vecteurs correspondants à ces deux possibilités. Ces diverses constructions ont été testées fonctionnellement en présence de puromycine ou de zéocine afin de déterminer quelles protéines de fusion conservent leurs propriétés (section 3 de ce chapitre). De plus, deux sites de restriction uniques ont été ajoutés devant le site *Cla* I. Quant à la séquence IRES, elle est éliminée du vecteur rétroviral car le gène permettant la sélection des clones positifs est fusionné au peptide qui sera exprimé sous le contrôle d'un promoteur placé en amont de celle-ci, ici le promoteur de SV40. Ce promoteur a ensuite été remplacé par celui de la phosphoglycérate kinase (PGK), un promoteur constitutif fort. Nous appellerons ces vecteurs modifiés pBABE HA/Puro^R, pBABE HA/Puro^R/lien souple, pBABE HA/Zéo^R et pBABE HA/Zeo^R/lien souple.

1.1.1 Clonage des inserts Puro^R/lien souple, Puro^R, Zeo^R/lien souple et Zeo^R, dans le vecteur pBABE HAGIZ

L'épitope HA étant déjà présent dans le vecteur pBABE HAGIZ, nous avons exploité le site unique *EcoR* I qui termine la séquence de HA afin d'y introduire les inserts dans le même cadre de lecture que celui-ci (Figure 18).

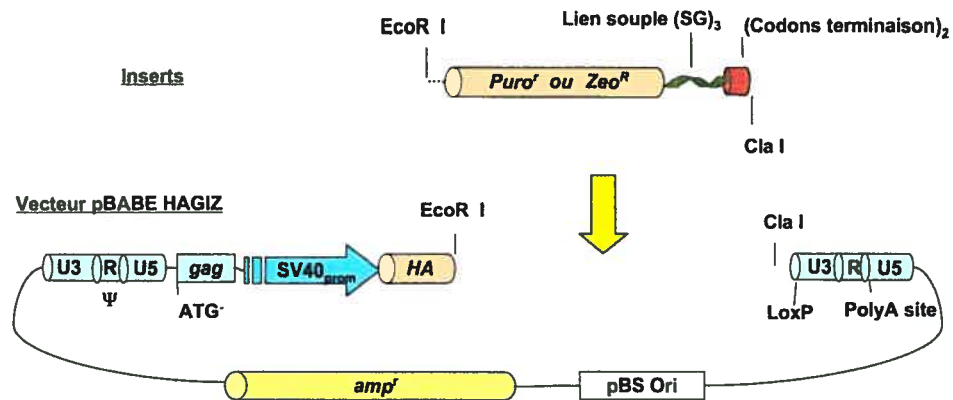


Figure 18. Schématisation du clonage des inserts Puro^R/lien souple et Zeo^R/lien souple dans le vecteur pBABE HAGIZ.

Les inserts d'ADN codant les gènes de résistance à la zéocine ou à la puromycine ont été construits par amplification. La séquence du gène Puro^R fut amplifiée à partir du plasmide pBABE-puro et la séquence du gène Zéo^R fut amplifiée à partir du plasmide pBABE HAGIZ (cartes des vecteurs n° 1 et n° 2 respectivement, en annexe). Les paires d'amorces sens (F') et antisens (R') suivantes ont été utilisées. Les sites de restriction *EcoR* I et *Cla* I, utilisés pour le clonage, sont en italique. La séquence du lien souple est soulignée et les codons de terminaison sont en caractères gras :

Insert « Zéo^R »

F' : 5'-TGATAAGAATTGCGCAAGTTGACCAGTGCCGTTCCGGTGCTC

R' : 5'-TGATAAATCGATCTATCAGTCCTGCTCCTCGGCCACGAAGTGCACG

Insert « Zéo^R/lien souple »

F' : 5'-TGATAAGAATTGCGCAAGTTGACCAGTGCCGTTCCGGTGCTC

R' : 5'-TGATAAATCGATTCATCACCCAGAGCCGGACCCAGAGTCCTGCTCCTCGGCC
ACGAAGTGCACG

Insert « Puro^R »

F' : 5'-TGATAAGAATTCACCGAGTACAAGCCCACGGTGCGCCTCG

R' : 5'-TAGTAAATCGATCTATCAGGCACCGGGCTTGCGGGTCATGCACC

Insert « Puro^R/lien souple »

F' : 5'-TGATAAGAATTCACCGAGTACAAGCCCACGGTGCGCCTCG

R' : 5'-TAGTAAATCGATTCATCACCCAGAGCCGGACCCAGAGGCACCGGGCTTGCGG
GTCATGCACC

Les amplifications ont été effectuées avec la polymérase d'ADN *Cloned Pfu* (Stratagene) dans 100 µl d'un mélange réactionnel composé du tampon *Cloned Pfu* (1X), d'un mélange des quatre dNTP (0,3 mM), de MgSO₄ (1 mM supplémentaire), des amorces R' et F' (0,3 µM chacun), de matrice pBABE-puro ou pBABE HAGIZ (200 ng), de polymérase *Cloned Pfu* (5 U) et de solution *PCR_x Enhancer* (2X) (Invitrogen). L'appareil *Thermocycler PTC-200* (Peltier Thermal cycler) et programme suivant ont été utilisés:

94°C, 45 sec	→	1 cycle
94°C, 45 sec	}	5 cycles
66°C, 45 sec		
72°C, 90 sec		
94°C, 30 sec	}	30 cycles
63°C, 45 sec		
72°C, 90 sec		
72°C, 10 min	→	1 cycle

L'ADN a été purifié sur gel LMP 2 % avec l'ensemble *QIAquick Gel Extraction Kit* (Qiagen). Les extrémités des inserts et le vecteur pBABE HAGIZ^a ont été digérés 4 heures à 37°C par les enzymes de restriction *EcoR* I et *Cla* I (NEB). Les inserts ont été purifiés sur colonne *QIAquick PCR Purification Kit* (Qiagen) et le vecteur coupé fut purifié sur gel d'agarose LMP 1,2 %. Les inserts ont été ligaturé au vecteur pBABE HAGIZ

^a L'efficacité de digestion de l'enzyme de restriction *Cla* I est sensible aux séquences entourant le site de reconnaissance de l'enzyme. Des essais de digestion ont été tentés avec les vecteurs pBABE HAZ et pBABE HAGIZ (cartes n° 3 et n° 2 respectivement, en annexe) et seul ce dernier est coupé efficacement par l'enzyme *Cla* I.

préalablement déphosphorylé selon le protocole décrit à la section 1.3. Les produits de ligature ont été transformés dans des bactéries chimiocompétentes *MAX Efficiency Stbl2* (Invitrogen) selon le protocole décrit à la section 1.4.

L'ADN de clones isolés pour chaque nouveau vecteur a été extrait sur des colonnes *QIAprep Spin Miniprep Kit* (Qiagen) et analysé par digestion enzymatique et par séquençage (Service de séquençage de l'Institut Armand-Frappier, Québec). L'amorce utilisée pour le séquençage des jonctions s'hybride au site Kozak, en amont de l'épitope HA (F' : 5'-AAGCTTGCCACCATGGCTTATCCA).

1.1.2 Remplacement du promoteur SV40 par le promoteur PGK

Le promoteur de SV40 n'est pas un promoteur très fort en cellules HEK 293 et il peut aussi être inactivé dans le temps par la méthylation. Nous avons donc décidé de le remplacer par le promoteur de la phosphoglycérate kinase (PGK), qualifié de « *house keeping gene* » ou promoteur d'expression continue.

Dans ce dessein, nous avons coupé le promoteur PGK du vecteur pPGK PURO entre les sites *Sal I* et *Hind III* (carte du vecteur n° 4 en annexe) et l'avons inséré dans les vecteurs pBABE HA/Puro^R, pBABE HA/Puro^R/lien souple, pBABE HA/Zéo^R et pBABE HA/Zeo^R/lien souple, générés précédemment. Les ligatures ont été faites selon le protocole décrit à la section 1.3. L'insertion correcte du promoteur PGK fut vérifiée par séquençage avec une amorce s'hybridant à la fin du gène *gag* et précédant le site d'insertion du promoteur (F' : 5'-GCCCTTTGTACACCCTAAGCCTCCGC). L'ADN des clones positifs a été amplifié par maxipreps⁶³.

1.1.3 Introduction de sites de restriction pour le clonage et la sélection des clones positifs

Ce clonage consiste à introduire un site de restriction unique *Mfe I*. Ce site permet de couper le vecteur parental ayant servi de matrice à l'insertion de la banque de peptides. Ceci réduit le nombre de clones qui n'auront pas intégré le nonapeptide. Pour cette raison,

l'insert fut nommé « *kill sites* ». De plus, un second site de restriction unique (*Nhe* I) a été ajouté afin de servir de site de clonage alternatif à *Cla* I.

La synthèse de l'insert contenant les sites *Mfe* I et *Nhe* I fut effectuée par l'amplification d'une partie du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple avec un oligonucléotide antisens comprenant les sites dans sa séquence (Figure 19).

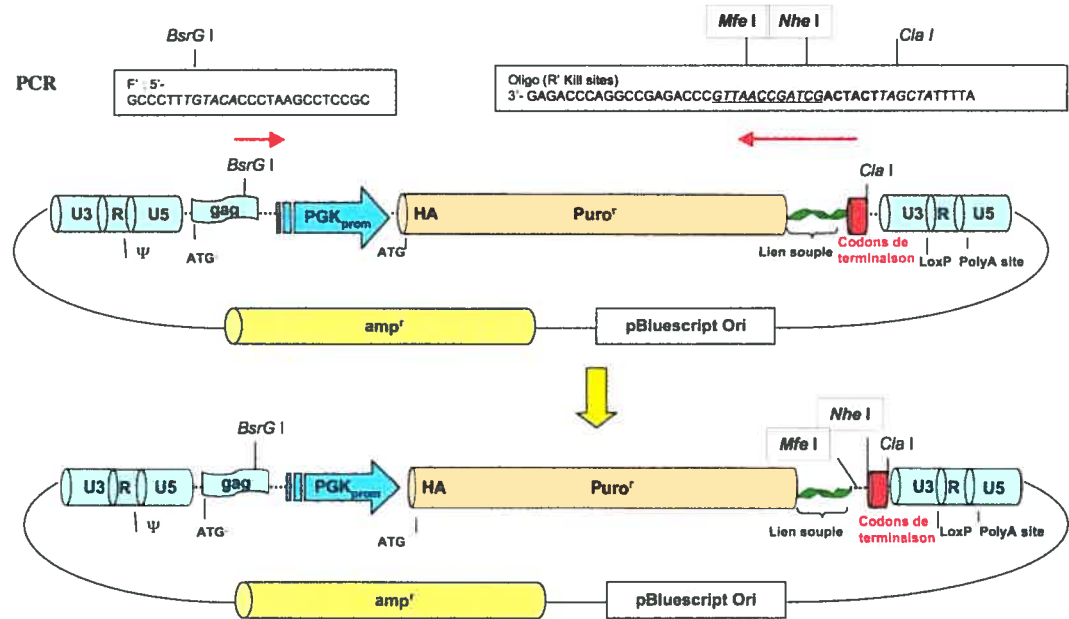


Figure 19. Schématisation du clonage de pBABE P_{PGK} HA/Puro^R/lien souple/kill sites.

La paire d'amorces sens (F') et antisens (R') suivante a été utilisée pour effectuer ce clonage. Les sites *Mfe* I et *Nhe* I sont en italique et soulignés. Les sites de restriction *BsrG* I et *Cla* I, utilisés pour le clonage, sont en italique. Les codons de terminaison sont en caractères gras :

Insert « *kill sites* »

F' : 5'-GCCCTTTGTACACCCTAAGCCTCCGC

R' : 5'-ATTTTATCGATTCATCAGCTAGCCAATTGCCAGAGCCGGACCCAGAG

Les amplifications ont été effectuées avec l'ensemble *Accuprime Supermix Pfx polymerase* (Invitrogen) dans 100 ul d'un mélange réactionnel composé de 90 ul de

polymérase *Accuprime Supermix Pfx*, des amorces R' et F' (0,2 uM) et de vecteur pBABE P_{PGK} HA/Puro^R/lien souple (100 ng). L'amplification a été effectuée selon le programme suivant :

95°C, 5 min	→	1 cycle
95°C, 15 sec	}	15 cycles
68°C à 58°C (-0,3°C/sec), 30 sec		
68°C, 1 min		
95°C, 15 sec	}	20 cycles
58°C à 68°C (+0,5°C/cycle), 30 sec		
68°C, 1 min		
68°C, 10 min	→	1 cycle

L'insert « *kill sites* » et le vecteur pBABE P_{PGK} HA/Puro^R/lien souple^b furent digérés O/N à 37°C par les enzymes *Cla* I et *BsrG* I (NEB). L'insert et le vecteur furent purifiés sur gel LMP 1,2 %. L'insert fut ensuite ligaturé au vecteur préalablement déphosphorylé avec la CIP (NEB). La ligature de l'insert « *kill sites* » dans le vecteur fut effectuée selon le protocole décrit à la section 1.3. Le produit de ligature (50 ng) fut transformé dans 100 ul de bactéries chimiocompétentes *MAX Efficiency Stbl2* selon le protocole décrit à la section 1.4.

L'analyse de l'ADN de clones isolés a été réalisée par électrophorèse sur gel d'agarose 1 % de digestion enzymatique des ADN, à 37°C pour 3 heures avec chacune des enzymes *Nhe* I (tampon 2 + *BSA* 1X) et *Mfe* I (tampon 4). L'intégration de l'insert dans le vecteur rétroviral a été confirmée par séquençage. L'amorce utilisée pour le séquençage s'hybride à la séquence qui correspond au dernier tiers du gène de résistance à la puromycine (F' : 5'-TGGTTCCTGGCCACCGTC). L'ADN d'un clone positif fut amplifié par maxiprep et extrait avec le *Qiagen Maxiprep Extraction kit* (Qiagen). Cet ADN de vecteur rétroviral a servi au clonage de la banque de nonapeptides.

^b Les tests fonctionnels en présence d'antibiotique ont montré que ce vecteur est le plus approprié pour le clonage de la banque de nonapeptides.

1.1.4 Construction d'un vecteur rétroviral contrôle pBABE P_{PGK} eYFP

Ce vecteur contrôle a été fait dans le but d'être utilisé comme témoin de l'efficacité de transfection des cellules, lors de l'étape de production de stocks viraux. Ce vecteur peut aussi servir à évaluer rapidement le titre viral, par analyse par FACS des cellules infectées. La carte du vecteur est présentée en annexe (carte n° 5)

Le clonage fut réalisé en insérant le gène eYFP du vecteur commercial pEYFP-N1 (BD Biosciences Clontech) dans le vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites, en remplacement du gène de fusion HA/Puro^R/lien souple. Les sites de restriction utilisés sont *Hind* III et *Mfe* I. La séquence *Hind* III fait partie des nombreux sites de restriction aux fins de clonage (MCS) du vecteur pEYFP-N1 et coupe devant le gène eYFP. *Mfe* I est un site qui se trouve après le gène eYFP et devant la séquence polyA de SV40 de ce vecteur. Après digestion des sites de restriction, l'insert et le vecteur rétroviral ont été purifiés sur gel d'agarose LMP 1,2 %. La ligature a été faite selon le protocole décrit au point 1.4 de cette section. Le produit de ligature fut transformé dans des bactéries électrocompétentes en suivant la technique détaillée à la section 1.5. L'intégrité de la séquence de l'insert et des sites de ligature fut confirmée par séquençage.

1.2 Construction de la banque de nonapeptides

1.2.1 Méthode insert-vecteur

Cette technique consiste à compléter le polynucléotide de la banque de nonapeptides en ADN double brin et à l'insérer dans le vecteur rétroviral coupé.

De nombreux essais de synthèse du brin complémentaire des polynucléotides furent effectués, avec diverses amorces et polymérases d'ADN (*Pfx*, KOD et *TAQ2000*). Deux inserts ont été générés par cette méthode, l'un s'incorporant entre les sites de restriction

Dra III et *Cla* I, l'autre se insérant aux sites *Mfe* I et *Cla* I^c. Les meilleurs résultats furent obtenus avec la polymérase d'ADN *KOD Hot Start* (Novagen). Brièvement, l'extension des polynucléotides fut réalisée dans un mélange réactionnel composé de tampon KOD (1X), de MgSO₄ (1 mM), d'un mélange des quatre dNTP (0,2 mM), de polymérase *KOD Hot Start* (1 U / 50 ul PCR), de *PCR_x Enhancer Solution* (2X) (Invitrogen) et des amorces R' et F' (0,5 uM). La paire de polynucléotides sens (F') et antisens (R') suivante a été utilisée pour effectuer ce clonage. La séquence du lien souple est soulignée, le site *Bam*H I qui sert à l'analyse des clones est en italique et souligné. Les sites de restriction *Dra* III et *Cla* I, utilisés pour le clonage, sont en italique. Les codons de terminaison sont en caractères gras :

Insert « nonapeptides *Dra* III-*Cla* I »^d

F' : 5'-ACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCTCTGGGTCCGGATCC
GGG

R' : 5'-AAACGCATCGATCTATCAVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNCCCGGATC
CGGACCCAGAGGCACCGGGCTTGC

L'élongation des polynucléotides a été réalisée dans l'appareil *Thermocycler PTC-200* (Peltier Thermal cycler), selon le programme suivant :

94°C, 2 min	→	1 cycle
72°C à 50°C (-0,1°C/sec)	}	30 cycles
72°C, 30 sec		
72°C, 15 min	→	1 cycle

^c Seul le clonage de l'insert « nonapeptides *Dra* III - *Cla* I » est détaillé. À titre informatif, l'insert « nonapeptides *Mfe* I - *Cla* I » a été fait avec la paire d'oligonucléotides suivante : R' : 5'-TTATTTTATCGATTCATCA et F' : 5'-GCTCTGGGCAATTGNNBNNBNNBNNBNNBNNBNNBNNBTGATGAATCGATAAAAATAA. Ces amorces, synthétisées par Alpha DNA Inc, n'ont pas été purifiées. La procédure de clonage est sensiblement la même, mais les clones ne peuvent pas être criblés par digestion enzymatique avec *Bam*H I puisque l'insert « *Mfe* I - *Cla* I » n'encode pas le lien souple.

^d L'amorce F' a été synthétisée par Alpha DNA Inc. et l'amorce R' fut synthétisée par BioCorp Inc. Ces oligonucléotides n'ont pas été purifiés.

À la suite de la synthèse des polynucléotides en ADN double brin, l'insert « nonapeptides » est purifié par précipitation phénol/chloroforme et précipité O/N à -20°C à l'éthanol 100 % supplémenté de glycogène, de MgCl₂ (0,01 M) et de sodium acétate (0,3 M). L'insert et le vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites sont ensuite coupés aux extrémités 5' et 3' pendant 3 heures à 37°C par les enzymes *Dra* III et *Cla* I dans le tampon H (Roche). Après une seconde purification, l'insert « nonapeptides » fut dosé sur gel d'agarose NuSieve GTG 4 %. L'insert fut ligaturé au vecteur rétroviral préalablement déphosphorylé et purifié sur gel d'agarose LMP 1,2 %. La ligature des inserts « nonapeptides » au vecteur rétroviral a été effectuée selon le protocole décrit à la section 1.3. La transformation de la banque de nonapeptides a été faite dans les bactéries chimiocompétentes *MAX Efficiency Stbl2 Cells* selon le protocole détaillé à la section 1.4. L'analyse des clones a été réalisée selon les protocoles décrits aux sections 1.5 et 1.6.

1.2.2 Méthode insert-vecteur par amplification d'un insert de taille moyenne

Cette technique consiste à amplifier une partie du vecteur rétroviral par amplification, à l'aide de polynucléotides encodant la banque de nonapeptides. Cette étape permet de générer un insert de taille moyenne (1500 pb) qui peut être facilement purifié sur gel d'agarose et inséré dans le vecteur rétroviral. Des essais de clonage ont été tentés avec trois différents inserts. L'un des inserts correspond à une séquence allant du site *EcoR* I à *Cla* I. Un second insert va du site *BsrG* I à *Cla* I. Finalement, un troisième insert va du site *Dra* I à *Cla* I. Seuls les clonages faits avec l'insert *BsrG* I à *Cla* I donnent des résultats satisfaisants pour le nombre de clones obtenus^e.

Les paires d'amorces suivantes furent utilisées pour ces expériences. Les sites de restriction sont indiqués en italique et les codons de terminaison sont en caractères gras.

^e Puisque seul le clonage de l'insert *BsrG* I à *Cla* I a donné des résultats intéressants, seules les conditions de clonage de cet insert sont détaillées ici. Le clonage des autres inserts est cependant très semblable. À titre informatif, les amorces utilisées pour les autres inserts sont :

1) Insert *EcoR* I / *Cla* I F' : 5'-TGATAAGAATT**CACCGAGTACAAGCCCACGGTGC**CGCTCG

2) Insert *Dra* I / *Cla* I F' : 5'-AGAACTTTAA**AGTGCTCATC**ATTGG

Les amorces antisens R' sont les mêmes que pour l'insert *BsrG* I / *Cla* I (Site *Cla* I. Oligonucléotides non purifiés)

La séquence du lien souple est soulignée et le site *BamH I* utilisé pour l'analyse des clones positifs est en italique souligné :

Insert « *BsrG I/Cla I* » (Site de ligature *Cla I*. Polynucléotides non purifiés)

F' : 5'-GCCCTTTGTACACCCTAAGCCTCCGC

R' : 5'-AAACGCATCGATCTATCAVNNVNNVNNVNNVNNVNNVNNVNNVNNCCCGGATC
CGGACCCAGAGGCACCGGGCTTGC

Insert « *BsrG I/Nhe I* » (Site de ligature *Nhe I*. Polynucléotides R' purifiés par PAGE)

F' : 5'-GCCCTTTGTACACCCTAAGCCTCCGC

R' : 5'-ATCTTCCTTGCTAGCTTAVNNVNNVNNVNNVNNVNNVNNVNNVNNCCCGGATC
CGGACCCAGAGGC

La synthèse de l'insert « *BsrG I/Cla I* » fait avec l'amorce R' non purifiée a été effectuée avec la polymérase d'ADN *Platinum Pfx* (Invitrogen). Le mélange réactionnel est composé d'un tampon d'amplification *Pfx* (2X), de solution *PCR_x Enhancer* (2X), d'un mélange des quatre dNTP (0,3 mM), de $MgSO_4$ (1 mM), de vecteur pBABE P_{PGK} HA/Puro^R/lien souple (100 ng/50 ul de PCR), de polymérase *Platinum Pfx* (2 U par 50 ul de PCR) et de chacune des amorces F' et R' (0,3 mM).

L'amplification a été réalisée selon le programme d'amplification suivant :

94°C, 4 min	→	1 cycle
94°C, 15 sec	}	30 cycles
60°C à 53°C (-0,2°C/sec), 30 sec		
68°C, 90 sec		
68°C, 10 min	→	1 cycle

La synthèse de l'insert « *BsrG I / Nhe I* » fait avec l'amorce R' purifiée par PAGE a été effectuée avec la polymérase d'ADN *KOD Hot Start polymerase* (Novagen). Le mélange réactionnel est composé de tampon d'amplification pour polymérase d'ADN *KOD* (1X), de solution *PCR_x Enhancer* (1X) (Invitrogen), d'un mélange des quatre dNTP (0,24 mM), de $MgSO_4$ (1mM), de vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites (20 ng/50 ul de

PCR), de polymérase *KOD Hot Start* (2 U par 50 ul de PCR) et de chacune des amorces F' et R' (0,2 mM).

L'amplification a été réalisée selon le programme suivant :

94°C, 2 min	→	1 cycle
94°C, 15 sec	}	10 cycles
60°C à 50°C (-0,1°C/cycle), 30 sec		
68°C, 1 min		
94°C, 15 sec	}	30 cycles
68°C à 51°C (-0,1°C/sec)		
51°C, 20 sec		
68°C, 1 min	}	1 cycle
68°C, 10 min		

À la suite de l'amplification, les extrémités de l'insert et du vecteur rétroviral furent digérées pendant 6 heures à 37°C à l'aide des enzymes de restriction *BsrG I* et *Cla I* ou *Nhe I*. Les inserts et le vecteur digérés furent purifiés sur gel d'agarose LMP 1,2 %. Ces inserts ont finalement été ligaturé au vecteur rétroviral préalablement déphosphorylé. La ligation a été effectuée selon le protocole décrit à la section 1.3. Les transformations de la banque de nonapeptides ont été faites dans les bactéries chimiocompétentes *MAX Efficiency Stbl2 Cells* et dans les bactéries électrocompétentes *ElectroMAX Stbl4 Cells* (Invitrogen) selon le protocole détaillé à la section 1.4. L'analyse des clones a été réalisée selon les protocoles décrits aux sections 1.5 et 1.6.

1.2.3 Méthode ARN-ADNc-ADN

Cette méthode consiste à produire, en conditions *in vitro*, des ARN du vecteur rétroviral avec de la polymérase d'ARN T7. Cet ARN simple brin servira alors de matrice pour l'hybridation de l'amorce contenant la banque de peptides. Une étape de transcription inverse permet de former des ADNc puis des ADN double brin par digestion du brin ARN avec l'ARNase H et synthèse du second brin d'ADN avec une polymérase d'ADN.

Création du vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites/T7

La première étape consistait à introduire un site promoteur T7 dans le vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites. Le clonage consiste à préparer un insert d'ADN composé du promoteur T7. Dans ce dessein, des amorces complémentaires codant pour le promoteur T7 avec des extensions codant le site de restriction *Cla* I ont été hybridées. Une étape d'extension des amorces avec la polymérase d'ADN *Accuprime Supermix Pfx* permet de compléter les brins d'ADN. La paire d'amorces suivante fut utilisée. La séquence du promoteur T7 est soulignée et les sites *Cla* I sont en italique :

Insert « T7 »

F' : 5'-TCCGACATCGATGTTAACTGAGCGCGCGTAATACGACTCAC

R' : 5'-TGTTTCGATCGATAATTCGCCCTATAGTGAGTCGTATTACGCGCGC

En bref, 0,5 nmol des amorces F' et R' fut ajoutée à 90 ul de polymérase *Accuprime Supermix Pfx*. Le programme d'hybridation suivant fut utilisé :

95°C, 5 min	} 1 cycle
68 à 57°C (-0,1°C/sec)	
68°C, 60 min	

À la suite de l'extension des amorces, l'insert « T7 » fut digéré aux deux extrémités pendant 16 heures à 37°C par l'enzyme *Cla* I (NEB) dans le tampon d'hybridation, additionné d'un volume équivalent de tampon 4 (1X) de BSA (1X). L'insert étant très petit, il ne peut être purifié sur colonne. Il fut donc déposé sur gel d'agarose *GTG NuSieve* 4,5 %. À la suite de l'électrophorèse, la bande d'ADN correspondant à l'insert fut coupée dans le gel. L'agarose fut digéré avec 15 U de β -Agarase (Cambrex Bio Science Rockland Inc.) pendant 1 heure à 42°C dans le tampon β -Agarase (1X). L'insert « T7 » fut ligaturé au vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites préalablement linéarisé au site *Cla* I et déphosphorylé. La ligature a été faite selon le protocole décrit à la section 1.3. Le produit de ligature fut introduit dans des bactéries chimiocompétentes *MAX Efficiency Stbl2 Cells* selon le protocole de transformation décrit à la section 1.4. L'analyse de clones ayant intégré le promoteur T7 dans le vecteur rétroviral a été effectuée par criblage par amplification directement sur les colonies de clones. L'amorce F' utilisée pour ce criblage

s'hybride au vecteur rétroviral dans la partie terminale de *gag* alors que l'amorce R' s'hybride au promoteur T7 nouvellement introduit. La paire d'amorces sens et antisens suivante fut utilisée :

F' : 5'-GCCCTTTGTACACCCTAAGCCTCCGC

R' : 5'-TGTTTCGATCGATAATTCGCCCTATAGTGAGTCGTATTACGCGCGC

Pour le criblage, chaque colonie est diluée dans 10 ul d'eau. Le criblage par amplification se fait dans 20 ul d'un mélange réactionnel composé de 10 ul de clones dilués, du tampon pour *Cloned Pfu* (1X), d'un mélange des quatre dNTP (0,2 mM), de chacune des amorces F' et R' (0,5 uM) et de polymérase *Cloned Pfu* (1 U) (*Stratagene*). Le programme suivant fut utilisé :

94°C, 45 sec	→	1 cycle
94°C, 45 sec	}	25 cycles
72°C à 55°C (-0,2°C/sec)		
72°C, 2 min		
72°C, 10 min	→	1 cycle

Les ADN ont été analysés sur gel d'agarose et les clones positifs ont été amplifiés par minipreps. Afin de confirmer la présence du promoteur T7 dans le vecteur rétroviral, 1 ug d'ADN de clones positifs fut coupé avec l'enzyme de restriction *Hpa* I (ce site unique est contenu dans l'insert « T7 ») et *Bsr*G I (un site dans *gag*). L'insertion correcte du site T7 fut confirmée par séquençage en utilisant l'amorce s'hybridant au dernier tiers du gène de résistance à la puromycine (F' : 5'- TGGTTCCTGGCCACCGTC). L'ADN d'un clone positif fut amplifié par maxiprep.

Produire l'ARNm du vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites/T7

La seconde étape de cette méthode consiste à produire l'ARNm du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites/T7.

1.2.4 Clonage par amplification ou extension de tout le vecteur rétroviral

Cette technique a été effectuée de plusieurs façons, toutes visant à obtenir un bon rendement de synthèse d'ADN et une incorporation correcte de la banque de nonapeptides⁹. Le principe consiste à utiliser un polynucléotide antisens encodant la banque de peptides et terminé par un site de restriction unique. Cette amorce sert à amplifier tout le vecteur rétroviral. Utilisée seule ou en excès, elle génère l'ADNc du vecteur avec la séquence du nonapeptide à son extrémité. L'ajout d'un oligonucléotide sens, complémentaire à la séquence du site de restriction unique de la première amorce, permet de synthétiser le second brin du vecteur rétroviral. La digestion enzymatique du site de restriction unique et la ligature subséquente génèrent le vecteur rétroviral complet, sous sa forme circulaire.

La première étape de cette technique consiste en la création de l'ADNc du vecteur rétroviral avec un polynucléotide contenant la banque de nonapeptides et une séquence complémentaire au vecteur. Cette élongation nécessite l'utilisation de polymérase d'ADN robustes. Nous avons essayé une dizaine de polymérase d'ADN commerciales et les plus efficaces pour une élongation de cette longueur ont été la *KOD Hot Start DNA polymerase* (Novagen) et la *Phusion High-Fidelity DNA Polymerase* (Finnzymes). Pour cette amplification en présence d'un excès d'amorces de la librairie de peptides, nous avons utilisé un mélange réactionnel composé du tampon *Phusion GC* (1X)^h, de DMSO (3 %), d'un mélange des quatre dNTP 0,2 mM, de polynucléotide R' (0,5 uM), de l'amorce F' (0,05 uM), de vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites (50 ng/50 ul PCR) et de polymérase *Phusion High-Fidelity* (2 U/50 ul de PCR).

L'amorce sens (F' : 5'-ATCTGCAATTGGCTAGCTGATGAATCG) s'hybride à l'extrémité 3' du nouveau vecteur rétroviral synthétisé par l'extension complète de l'amorce antisens. Ce polynucléotide fut synthétisé par Alpha DNA inc. et fut purifié par PAGE.

⁹ Seule la technique de l'amplification avec excès de l'amorce antisens est détaillée. L'extension pendant de multiples cycles avec seulement l'amorce antisens ne permet pas d'obtenir suffisamment d'ADN pour le clonage d'une banque de vaste diversité peptidique.

^h Le tampon *HF* ne donne aucun résultat dans ces conditions expérimentales.

Toutes les amorces encodant la banque de peptides en notre possession ont été essayées avec cette technique. Néanmoins, les polynucléotides qui ont donné les meilleurs résultats lors de la ligature sont l'amorce purifiée par PAGE (Alpha DNA inc.) et le polynucléotide synthétisé par étapes successives (Sigma Genosys). Ce dernier comporte 81 nucléotides d'orientation antisens. On y retrouve, à l'extrémité 3', une séquence de 21 nucléotides complémentaires à la fin du gène Puro^R en fusion avec le lien souple. La banque est encodée par une séquence (NNK)₉ et par un codon non-sens TAA. Un site de restriction *Pac* I (TTAATTA) suit le codon de terminaison et peut servir au criblage des clones. Les sites *Nhe* I (GCTAGC) et *Mfe* I (CAATTG) sont présents à l'extrémité 5' et permettent de circulariser le vecteur. Finalement, une molécule de biotine est placée en 5' pour la purification sur billes magnétiques recouvertes de streptavidine.

La synthèse du polynucléotide a été effectuée par une synthèse séquentielle. En effet, la longue séquence dégénérée rend la synthèse de ce type d'amorces inefficace et a pour résultat une majorité de produits de synthèse tronqués. Afin d'obtenir un maximum de polynucléotides complets, la génération de l'amorce fut effectuée en deux étapes. La première partie correspond à la partie 5' du polynucléotide dont la séquence d'orientation est 5'-Biotine-TTGGACGCTAGCCAATTGTTAATTAATTA. Cette amorce, fixée à la matrice de synthèse, fut transférée dans une autre colonne et un autre appareil. La suite de l'addition des nucléotides a permis de compléter le polynucléotide dont la séquence complète est la suivante :

5'-Biotine-TTGGACGCTAGCCAATTGTTAATTAATTAVNNVNNVNNVNNVNNVNNVNNV
 NNVNNCCCGGATCCGGACCCAGAGGCACCG.

Les extensions ont été effectuées selon le programme d'extension suivant :

98°C, 30 sec	→	1 cycle
98°C, 8 sec	}	25 cycles
68°C, 30 sec		
72°C, 2 min 30 sec		
72°C, 10 min	→	1 cycle

À la suite de cette amplification préférentielle d'un brin du vecteur rétroviral, un surplus d'amorce sens (1,5 μ M de F'), 0,5 U de polymérase *Phusion High* et 0,2 mM de chacun des quatre dNTP sont ajoutés au mélange réactionnel. Un seul cycle d'extension est fait afin de compléter le second brin d'ADN. Le programme suivant fut utilisé :

98°C, 30 sec

72°C, 10 min

72°C à 25°C (-0,1°C/sec), 10 min

72°C, 10 min

Après cette extension, l'ADN est analysé sur gel d'agarose 1 %. La purification du produit d'amplification a été faite avec des billes magnétiques recouvertes de streptavidine (*Dynabeads KibobaseBinder*, Dynal Biotech). Brièvement, les billes sont conditionnées dans le tampon de liaison. L'ADN biotinylé est mis en contact avec les billes pendant 3 heures à 25°C. Les ADN non biotinylés, dont le vecteur parental, sont éliminés par deux lavages dans le tampon de lavage et un lavage dans de l'eau. Les billes sont retenues par un aimant collé contre la paroi du tube. L'ADN attaché aux billes est ensuite libéré de sa biotine par digestion du site *Mfe* I pendant 6 heures à 37°C. Le vecteur rétroviral digéré à ses extrémités, maintenant dans le surnageant, est récupéré. Cet ADN est ligaturé en suivant le protocole décrit à la section 1.3. L'ADN est ensuite transformé dans les bactéries chimiocompétentes *MAX Efficiency Stbl2 Cells* et dans les bactéries électrocompétentes *ElectroMAX Stbl4 Cells* selon le protocole de la section 1.4. L'analyse des clones a été réalisée selon les protocoles décrits aux sections 1.5 et 1.6.

1.3 Ligatures

Avant la ligature, tous les ADN ont été purifiés sur une résine liquide *StrataClean* (Stratagene). Cette méthode est beaucoup plus douce pour l'ADN qu'une purification phénol/chloroforme. En bref, l'ADN est mélangé à la résine. Les protéines s'y collent par affinité, laissant l'ADN dans le surnageant. Les sels sont éliminés par une précipitation de l'ADN à l'éthanol 100 %. Sauf avis contraire, les ligatures ont été effectuées dans 20 μ l d'un mélange réactionnel composé de 1 μ l d'ADN Ligase T4 (400 U/ μ l, NEB), de tampon de ligature (1X), d'ATP (1 mM supplémentaire) et de l'ADN à ligaturer. Le protocole de

ligature fut le suivant : 2 heures à température pièce (~25°C), suivies par une incubation O/N à 16°C. Les ligatures sont gardées à 4°C jusqu'à leur utilisation. L'efficacité de la ligature est vérifiée sur gel d'agarose 1,2 % et le produit de ligature est purifié sur résine *StrataClean* et précipité à l'éthanol 100 %. La concentration en ADN du produit purifié est déterminée par densité optique à 260nm. L'ADN purifié est alors transformé dans les bactéries compétentes.

1.4 Transformations

1.4.1 Bactéries chimiocompétentes *MAX Efficiency Stbl2 Cells*

Le produit de ligature purifié est mélangé à 100 ul de bactéries chimiocompétentes *MAX Efficiency Stbl2* conditionnées au laboratoire. Le protocole de transformation est le suivant : 30 minutes sur glace, 90 secondes à 42°C, puis 2 minutes sur glace. À la suite de la transformation, 900 ul de milieu SOC sont ajoutés aux cellules transformées. Ces clones sont amplifiés suivant le protocole décrit au point 1.4.3.

1.4.2 Bactéries électrocompétentes *ElectroMAX Stbl4 Cells*ⁱ

La transformation des produits de ligature purifiés s'est effectuée dans 20 ul de cellules électrocompétentes avec 50 ng d'ADN (10 pg pour le contrôle pUC19). L'électroporation a été faite dans un appareil *Bio-Rad* avec des cuvettes de 0,1 cm d'espacement. Les paramètres suivants ont été utilisés : 1,2 kV/cm, 200 Ω , 25 μ F. Ces conditions nous ont permis d'obtenir une constance de temps (τ) entre 3,6 et 4,2 msec. Après les transformations, le volume de chaque transformation est immédiatement complété à 1 ml avec du milieu SOC.

ⁱ D'autres types de bactéries électro-compétentes ont été utilisées ; *GeneHogs E. coli* (Invitrogen), *ElectroTen-Blue* (Stratagene) et *DH5 α -E* (Invitrogen). Les paramètres indiqués ne sont bons que pour les bactéries *Stbl4*. Les bactéries *ElectroMAX Stbl4* réduisent le nombre de recombinaisons entre les séquences répétées (LTR du vecteur rétroviral). Ces bactéries ont donc été privilégiées pour l'amplification de la banque de peptides.

1.4.3 Incubation et amplification

Les clones dilués à un volume de 1 ml dans le milieu SOC sont incubés 90 minutes à 30°C avec agitation à 225 rpm. Afin de déterminer l'efficacité des transformations, une série de dilutions (10^{-3} à 10^{-6}) sont étalées sur pétris LB ampicilline et mises à incuber 24 à 30 heures à 30°C. Le reste des clones est ajouté à 100 ml de milieu LB ampicilline et incubé dans les mêmes conditions, mais avec agitation à 225 rpm. Une fraction de ces clones amplifiés est congelée dans le glycérol et servira à la mise en réserve de la banque de peptides. L'ADN du restant des clones amplifiés est extrait avec l'ensemble QIAGEN *Plasmid Midi Kit* (Qiagen). Cet ADN servira aux transfections de la banque de nonapeptides.

1.5 Calculs

Les calculs permettant une évaluation de la diversité de la banque sont les suivants :

(nombre de colonies/volume étalé sur pétri) x dilution x volume total de la transformation = nombre de clones/transformation

Nombre de clones/transformation x nombre de transformation = diversité clonale finale de la banque de nonapeptides

1.6 Analyse des clones

Pour chaque transformation, l'ADN de 20 clones isolés fut extrait par minipreps, selon la technique de lyse alcaline⁶³. Ces ADN ont été analysés par digestion enzymatique avec l'enzyme *BamH I*. Le vecteur rétroviral ne possède qu'un seul site *BamH I* dans le gène *gag*. Les vecteurs rétroviraux possédant la banque de peptides possèdent deux sites *BamH I*, le second ayant été introduit dans la séquence du lien souple. Les clones trouvés positifs par digestion enzymatique (bandes à 1400 pb et 4000 pb) ont été séquencés (Service de séquençage du Centre d'Innovation Génome Québec et Université McGill, Québec). L'amorce utilisée pour le séquençage s'hybride au dernier tiers du gène de

résistance à la puromycine et permet de vérifier la présence du peptide ainsi que l'intégrité des sites de ligature (F' : 5'-TGGTTCCTGGCCACCGTC). Les séquences *ABI* sont analysées à l'aide du logiciel *4Peaks* (par A. Griekspoor et Tom Groothuis, mekentosj.com).

Afin de vérifier que la diversité des clones n'est pas altérée par une incubation longue en milieu liquide et par la congélation en milieu glycérol, un échantillon de clones congelés a été étalé sur pétri LB ampicilline et incubé à nouveau 30 heures à 30°C. L'ADN de clones isolés fut extrait par minipreps et analysé par digestion enzymatique *BamH I* et par séquençage.

1.7 Réserve de clones en milieu glycérol

Un volume de 1/100 des clones amplifiés en milieu LB ampicilline est ajouté à un volume équivalent de milieu SOC supplémenté de glycérol 40 %. Cette réserve de clones en milieu glycérol est immédiatement congelée à -80°C. La diversité de la banque de peptides encodée par cette réserve de clones en milieu glycérol a été analysée par digestion enzymatique *BamH I* et par séquençage.

2. PRODUCTION ET INFECTION VIRALES

2.1 Transfection des vecteurs dans des cellules HEK 293T

Cette étape comporte certains risques inhérents au caractère infectieux des virus générés. Toutes les manipulations qui sont décrites ont été effectuées selon les paramètres de bioprotection de niveau 2.

Une étape d'optimisation des conditions de transfection et de production virales a été effectuée avec les vecteurs rétroviraux pBABE P_{PGK} Puro^R/lien souple/kill sites et pBABE P_{PGK} eYFP (carte du vecteur n° 5, en annexe) et pBABE GFP. Le protocole optimisé décrit ci-dessous fut suivi afin de générer des virus par transfection de cellules HEK 293T avec

les vecteurs rétroviraux ainsi que les plasmides encodant les gènes *gag-pol* et l'enveloppe VSV-G :

- 24 heures avant la transfection, $0,3 \times 10^6$ cellules HEK 293T¹ sont placées dans chacun des puits d'une plaque 6 puits, dans un volume de 2 ml de milieu DMEM complet sans antibiotique. Lorsque les cellules atteignent environ 80 % de confluence, les ADN à transférer et l'agent de transfection *Lipofectamine 2000* (Invitrogen) sont préparés selon le protocole suivant : 2 ug de vecteur rétroviral murin pBABE, 2 ug de plasmide pCIG3B (*gag-pol* de MLV) et 0,4 ug de plasmide pSVCMV-IN^k (VSV-G) sont ajoutés à du milieu sans sérum *Opti-MEM* (Gibco), dans un volume final total de 250 ul. Les contrôles ont été faits avec le vecteur rétroviral pBABE P_{PGK} eYFP et pSVCMV-IN VSVG. Le plasmide encodant les gènes *gag-pol* a été omis et remplacé par une quantité équivalente du plasmide pBluescript SK. Ces contrôles ont permis d'estimer, par FACS, l'efficacité de la triple transfection du vecteur rétroviral.

Dans un tube séparé, 6 ul de *Lipofectamine 2000* sont ajoutés à 244 ul de milieu *Opti-MEM* sans sérum. (2,5 fois plus de *Lipofectamine 2000*, en ul, que la quantité en ug d'ADN à transférer.)

L'ADN dilué et la *Lipofectamine 2000* diluée sont incubés pendant 5 minutes à la température ambiante. Après l'incubation, l'ADN et la *Lipofectamine 2000* sont combinés et incubés à nouveau pendant 20 minutes à température ambiante. Le mélange *Lipofectamine 2000*/ADN est ajouté aux cellules, sans enlever le milieu de culture déjà présent. La transfection est poursuivie pendant 16 heures à 37°C dans une atmosphère contrôlée de CO₂ à 5 %. Le lendemain de la transfection, le milieu de culture est changé pour du milieu DMEM complet. 36 à 48 heures après la transfection, selon la confluence, les cellules sont réensemencées dans des pétris de 10 cm dans un volume total de 10 ml de milieu DMEM complet. Pour la production virale, les cellules ont été mises à 32°C sous atmosphère contrôlée à 5 % de CO₂.

¹ Les cellules GP2-293 (BD Bioscience Clontech) (don du Dr A. Veillette) ont aussi été utilisées pour la génération de virus. Ces cellules fonctionnent très bien aussi. Ces cellules expriment de façon stable les gènes *gag-pol* de MLV. La sélection de ces cellules est faite par ajout de méthotrexate (100nM) au milieu de culture DMEM complet.

^k Le VSV-G est toxique pour la cellule et doit donc être ajouté en quantité moindre que les autres plasmides.

À intervalles de 12 heures et jusqu'à 72 heures après la transfection, le milieu de culture contenant les virions est prélevé et remplacé par du milieu frais. Le volume prélevé est filtré sur un filtre de $0,45\mu$ et congelé à -80°C sans agent cryoprotecteur.

Les cellules ayant été transfectées par les vecteurs contrôles ont été analysées par FACS. Les cellules ont été préalablement lavées 2 fois au PBS 1X et centrifugées à $900 \times g$ pour 10 minutes. Les cellules sont ensuite fixées 16 heures dans une solution de paraformaldéhyde 1 %. Les données ont été générées sur les appareils *FACScan* et *FACScalibur*. Les analyses ont été réalisées grâce au logiciel *Cellquest Pro*.

2.2 Introduction de la banque de peptides par infection de cellules de mammifères HEK 293

Pour éviter que plusieurs peptides soient exprimés par la même cellule, l'infection avec les virions encodant la banque de peptides doit être faite à basse multiplicité d'infection (MI). Une MI allouant l'infection de 20 % des cellules permettra l'obtention d'une majorité de cellules infectées n'exprimant qu'un peptide. Il est donc inutile, dans ces conditions, de concentrer le virus. La détermination du titre viral a été faite dans les cellules HEK 293. Ces cellules sont les mêmes qui seront utilisées pour la validation de la méthode, à une étape ultérieure du projet.

Le titre viral a été déterminé par la dilution sérielle 10^{-1} à 10^{-6} du concentré viral, dans un volume final de 200 μl . Les dilutions sont faites dans du milieu DMEM complet. Dans un tube de polypropylène de 15 ml, 2×10^5 cellules HEK 293 dans un volume de 800 μl sont mises en contact avec les 200 μl des différentes dilutions de virus. À ce mélange, nous avons ajouté du *polybrene* (8 $\mu\text{g/ml}$). Les virus et les cellules sont incubés 4 heures à 37°C dans un environnement contrôlé à 5 % de CO_2 , en agitant doucement le mélange toutes les 30 minutes. À la suite de l'incubation, le milieu de culture des cellules est changé pour 3 ml de milieu DMEM complet et les cellules sont étalées dans les puits d'une plaque de 6 puits (1 puits par dilution). Un puits de cellules non infectées est gardé comme contrôle de l'efficacité de sélection.

L'efficacité de l'infection peut être déterminée de deux façons. La méthode rapide est l'évaluation, par FACS, du nombre de cellules exprimant la YFP ou la GFP avec les vecteurs rétroviraux contrôles. 48 heures après l'infection, les cellules sont récoltées et fixées avec de la paraformaldéhyde 1 %, pour 16 heures. L'analyse par FACS révèle quelle dilution a permis d'infecter 20 % des cellules. La méthode la plus longue est la sélection sous antibiotique. La puromycine fusionnée à la banque de peptides permet la survie des cellules infectées. Pour cette dernière technique, 36 heures après l'infection, le milieu de culture est remplacé par du milieu DMEM complet supplémenté de l'agent de sélection, ici la puromycine (5 ug/ml). La sélection est exercée pendant 4 jours. Le puits contrôle permet de vérifier l'efficacité de la sélection. Le puits de la plus petite dilution contenant des colonies résistantes correspond au titre du virus concentré. Le nombre de colonies résistantes dans chacun des puits permet de déterminer le titre viral avec la formule suivante : (nombre de colonies/facteur de dilution) x 5 = nombre de particules infectieuses/ml

3. TECHNIQUES BIOCHIMIQUES

3.1 Test de cytotoxicité CC_{50}

Une première expérience a consisté à déterminer la concentration la plus efficace pour chaque antibiotique. Des cellules HEK 293T ont été placées dans chacun des puits d'une plaque à 6 puits. Après 48 heures, la puromycine fut ajoutée à des concentrations de 0 à 4 ug/ml. Les courbes de survie (CC_{50}) pour la puromycine furent déterminées après 24, 48, 72 et 96 heures d'exposition à l'antibiotique. La même expérience fut répétée dans les cellules de singes OMK, car ces cellules devaient servir à une étape ultérieure du projet. Dans ces cellules, pour le même effet cytotoxique, les concentrations en puromycine ont dû être augmentées à 1, 5, 25, 50 et 200 ug/ml. La courbe de survie de la zéocine dans les cellules HEK 293T avait déjà été déterminée par A. Ostiguy⁶⁰. La concentration efficace de la zéocine est de 50 ug/ml durant une exposition de 9 jours.

Une seconde expérience a permis de déterminer quelles protéines de fusion ont gardé les propriétés de résistance à l'antibiotique. Les vecteurs rétroviraux encodant les protéines

chimériques HA/Puro^R/lien souple, HA/Puro^R, HA/Zeo^R/lien souple et HA/Zeo^R ont été testés. Chacun de ces vecteurs fut transfecté dans des cellules HEK 293T avec l'agent de transfection *Lipofectamine 2000*. 48 heures après la transfection, nous avons ajouté les antibiotiques au milieu de culture. La puromycine fut ajoutée à une concentration de 5 ug/ml et la zéocine fut ajoutée à une concentration de 50 ug/ml. L'évaluation qualitative de la survie cellulaire a été faite lorsque les cellules témoins transfectées avec le vecteur pBABE GFP étaient toutes mortes. Pour la puromycine, cette observation a été faite à 72 heures après l'exposition à l'antibiotique. Pour la zéocine, la survie cellulaire a été évaluée à 9 jours après l'exposition à l'antibiotique.

4. MILIEUX DE CULTURE

DMEM complet : DMEM, FBS inactivé à la chaleur (10 %), NEAA (1X), L-glutamine (1X ou 2 mM), sans antibiotique

LB ampicilline : 10 g de NaCl, 10 g de tryptone, 5 g d'extrait de levures. Le volume est ajusté à 1 litre avec de l'eau. Autoclaver. Lorsque le mélange est refroidi, ajouter stérilement 100 ug/ml d'ampicilline filtrée. Pour les pétris de LB ampicilline, ajouter 20 g d'agar avant d'autoclaver.

5. CELLULES UTILISÉES

5.1 Cellules de mammifères

HEK 293 : les cellules HEK 293 ont été générées par la transformation de cellules de reins d'embryons humains avec l'ADN d'adénovirus 5⁶⁴.

Numéro ATCC : CRL-1573

HEK 293T : Les HEK 293T sont des cellules dérivées de la lignée HEK 293. Le gène thermosensible de l'antigène T de SV40 fut introduit dans ces cellules. Elles sont fortement transfectables. Numéro ATCC : CRL-11268

OMK(637-69) : Cellules de reins de singes (Kidney, owl monkey, *Aotus trivirgatus*).
Numéro ATCC : CRL-1556

5.2 Bactéries

Max Efficiency Stbl2 Cells (Invitrogen) : Bactéries chimiocompétentes dérivées de JM109. Ces cellules permettent la stabilisation d'inserts instables comme les vecteurs rétroviraux et les séquences répétées. Ces bactéries doivent être amplifiées à 30°C.

Génotype : F⁻, *mcrA*, $\Delta(mcrBC-hsdRMS-mrr)$, *recA1*, *endA1*, *gyrA96*, *thi-1*, *supE44*, *relA1*, λ^- , $\Delta(lac-proAB)$

ElectroMAX Stbl4 Cells (Invitrogen) : Bactéries électrocompétentes dérivées de JM109.

Génotype : *mcrA*, $\Delta(mcrBC-hsdRMS-mrr)$, *recA1*, *endA1*, *gyrA96*, *gal*, *thi-1*, *supE44*, λ^- , *relA1*, $\Delta(lac-proAB)$, F'[*proAB+ lacI^q*, Z Δ M15, Tn10(Tet^R)]

Section résultats

1. CONCEPTION ET GÉNÉRATION DE LA BANQUE DE PEPTIDES DANS LE VECTEUR RÉTROVIRAL MURIN pBABE.

Quel que soit le système utilisé (double-hybride en levures, phages présentoirs, etc.), la réussite d'un criblage d'interactions peptide-protéine repose sur l'expression d'une banque de peptides d'une grande diversité. Un système efficace doit aussi permettre d'isoler et d'identifier les éléments protéiques impliqués dans les interactions observées. À ces caractéristiques particulières s'ajoute, dans notre modèle d'étude, la nécessité que la banque soit exprimée *in vivo*, de façon stable et constitutive, dans des cellules de mammifères. Les sections qui suivent rapportent, selon les méthodes utilisées, les résultats obtenus lors de la création du vecteur rétroviral encodant la banque de peptides.

1.1 Configuration optimale de la protéine de fusion

Le gène de résistance à un antibiotique permet de sélectionner les clones cellulaires ayant intégré et exprimant l'ADN introduit, dont la banque de peptides. Dans le but de nous assurer que chaque cellule qui exprime le facteur de sélection exprime aussi un peptide de la banque, nous avons choisi de fusionner la banque de peptides au gène de résistance à l'antibiotique. La fusion de protéines en C et N terminal du facteur de sélection peut en altérer la fonction. Nous avons donc choisi de tester diverses configurations de protéines chimériques en milieu sélectif. Dans ce dessein, nous avons cloné deux types de vecteurs rétroviraux. Le premier encode le gène de résistance à la puromycine (*puromycine-N-acétyl-transférase, PAC ou Puro^R*) et le second encode le gène de résistance à la zéocine (*Sh ble ou Zeo^R*). À ces protéines de fusion, nous avons aussi ajouté un épitope HA qui sert à l'immunoprécipitation des complexes peptide-protéine. Nous avons évalué l'efficacité, lors de la sélection sous antibiotiques, des diverses configurations de protéines de fusion illustrés à la page suivante :

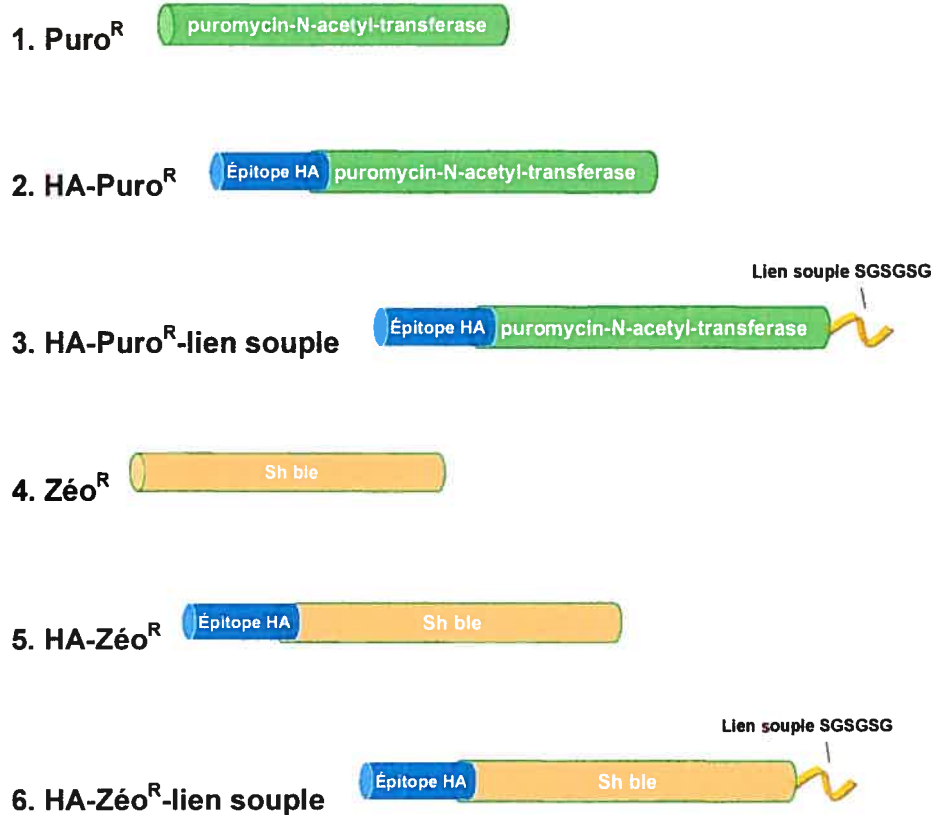


Figure 20. Représentation des protéines de fusion générées. Ces protéines de fusion sont clonées dans le vecteur rétroviral pBABE. Leur expression est sous le contrôle du promoteur de la phosphoglycérate kinase (PGK), sauf les vecteurs contrôles pBABE-puro (Puro^R seul) et pBUD CE4 (Zéo^R seul) qui sont sous le contrôle du promoteur SV40.

Nous avons transfecté des cellules HEK 293T avec les diverses constructions de vecteurs. Quarante-huit heures après la transfection, les antibiotiques puromycine ou zéocine ont été ajoutés au milieu de culture. Lorsque toutes les cellules contrôles transfectées avec le vecteur pBABE GFP sont mortes, la survie des cellules exprimant les protéines de fusion a été évaluée. Il a fallu 72 heures à la puromycine pour tuer toutes les cellules témoins. La zéocine requiert plus de temps et a nécessité une incubation de 9 jours en présence de l'antibiotique. Seule la protéine de fusion codant la résistance à la puromycine conserve pleinement sa fonction en présence de l'antibiotique, lorsque celle-ci est fusionnée en C et N terminal. L'efficacité de la protéine de résistance à la zéocine est influencée par la présence de protéines qui lui sont fusionnées en C terminal. Les observations qualitatives de la survie cellulaire des cellules transfectées avec les diverses constructions sont rapportées au tableau I à la page suivante.

**Transfections
(n=2)**

pBABE-puro (Contrôle +)	+	Puromycine (5 ug/ml) 72h post-transfection
pBABE P_{PGK} HA/Puro^R	+++	
pBABE P_{PGK} HA/Puro^R/lien souple	+++	
pBABE GFP (Contrôle -)	-	

pBUD CE4 (Contrôle +)	+	Zéocine (50 ug/ml) 9 jours post-transfection
pBABE P_{PGK} HA/Zeo^R	+++	
pBABE P_{PGK} HA/Zeo^R/lien souple	+	
pBABE GFP (Contrôle -)	-	

pBABE GFP	88 % GFP+ (FACS)	48h post-transfection
------------------	-----------------------------	-----------------------

Tableau I. Évaluation qualitative de la survie cellulaire de clones de cellules HEK 293T en présence des antibiotiques puromycine ou zéocine. Les protéines de résistance qui sont exprimées dans ces cellules ont été fusionnées en N terminal ou fusionnées en N et C terminal. Le vecteur pBABE GFP fut utilisé pour évaluer l'efficacité des transfections. Les cellules transfectées avec ce dernier ont été utilisées comme témoins négatifs.

Plus de cellules ont survécu lorsqu'elles exprimaient les protéines chimériques que lorsqu'elles exprimaient les protéines natives des contrôles positifs. La différence majeure réside dans le fait que le promoteur des contrôles est SV40 alors que le promoteur des protéines fusionnées est PGK. Nous avons aussi évalué la survie cellulaire lorsque les cellules sont transfectées avec le plasmide SR α -puro, un autre plasmide exprimant la puromycine sous le contrôle de SV40, et nous avons obtenu les mêmes résultats (données non montrées). Il semble que dans nos conditions de sélection, le promoteur PGK permette une expression plus forte de la protéine de résistance à l'antibiotique, ce qui se traduit par une meilleure survie cellulaire. Nous avons montré que le gène de résistance à la puromycine peut être fusionné en N et C terminal. L'efficacité rapide à basse concentration de la puromycine en fait un agent de sélection intéressant (Figures 21 et 22, page 61).

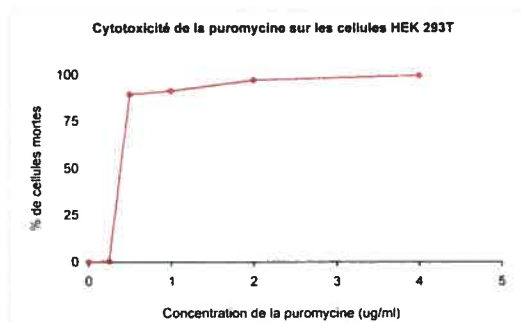


Figure 21

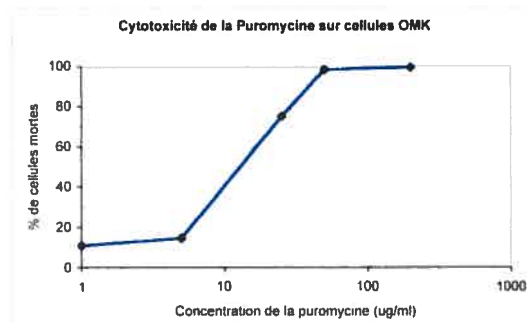


Figure 22

Figures 21 et 22. (21) Évaluation de la courbe de survie de la puromycine sur les cellules HEK 293T. (22) Évaluation de la courbe de survie de la puromycine sur les cellules OMK. Exposition de 96 h (HEK 293T) et de 72 h (OMK) en présence de l'antibiotique.

1.2 Clonage du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple

L'introduction de la banque de nonapeptides dans le vecteur rétroviral pBABE a nécessité l'adaptation de celui-ci. Premièrement, nous avons retiré tous les gènes situés entre les sites *EcoR* I et *Cla* I du vecteur rétroviral pBABE HAGIZ (Figure 23). Ensuite, nous y avons introduit, dans le même cadre de lecture que le gène HA, le gène encodant la résistance à la puromycine provenant du vecteur pPGK PURO.

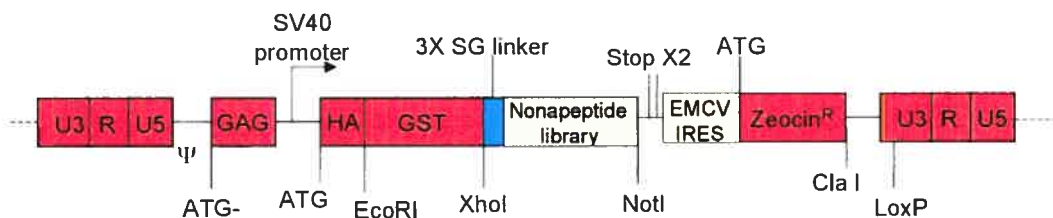


Figure 23. Vecteur pBABE HAGIZ (Illustration tirée de réf. 60)

À partir de ce nouveau vecteur, nous avons retiré le promoteur SV40, qui fut remplacé par le promoteur PGK. À l'aide d'une amorce codant une série de trois sérines-glycines (SGSGSG), nous avons introduit le lien souple à la suite du gène de résistance à la puromycine. Plus tard dans le projet, deux sites de restriction uniques, *Mfe* I et *Nhe* I, ont aussi été ajoutés à la fin du lien souple. Ces sites ont été utilisés comme sites de clonage

alternatifs. Le vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites, tel qu'illustré à la Figure 24, a servi de base pour l'introduction de la banque de peptides selon diverses techniques qui seront détaillées dans les prochaines sections.

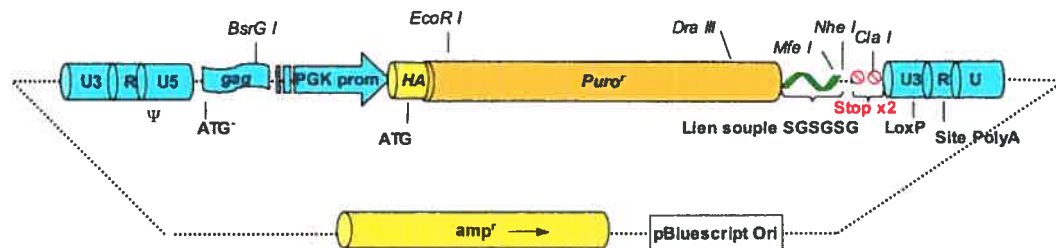


Figure 24. Vecteur pBABE P_{PGK} HA/Puro^R/lien souple/kill sites. Ce vecteur a servi de base à la construction des divers vecteurs rétroviraux qui ont été utilisés afin d'y introduire la banque de nonapeptides.

2. SYNTHÈSE PAR LIGATURE INSERT-VECTEUR

2.1 Stratégie du clonage insert-vecteur

La stratégie utilisée par plusieurs pour introduire une banque de peptides dans un plasmide consiste à produire la banque sous forme de polynucléotides simple brin bordés par des sites de restriction uniques⁵⁵. Le polynucléotide est ensuite transformé en petit ADN double brin par complémentarité avec l'enzyme Klenow ou une polymérase d'ADN. La procédure consiste alors à couper les extrémités de la banque avec les enzymes de restriction, à purifier le fragment d'ADN et à insérer celui-ci dans un vecteur rétroviral ouvert (Figure 25, page 63). Cette méthode, bien qu'elle semble simple, comporte des difficultés majeures qui influenceront la diversité finale de la banque de peptides. Malgré que cette stratégie de clonage ait déjà été essayée sans succès dans notre laboratoire, nous avons tout de même repris cette expérience en parallèle avec d'autres techniques. En effet, elle reste l'une des seules qui ont permis à d'autres groupes d'obtenir une banque de peptides exprimés dans un vecteur rétroviral.

Cependant, cette stratégie oblige l'emploi de très longs polynucléotides qui couvrent une partie du gène Puro^R et le lien souple. L'amorce antisens doit comporter une séquence recoupant le lien souple, en plus de la banque de peptides, des codons de terminaison et de sites de restriction flanqués de six nucléotides (Figure 25, page 63). La fabrication de polynucléotides de cette dimension (~70 à 80 nucléotides) est très inefficace et il en résulte une vaste majorité de produits tronqués. La question de la synthèse de ces polynucléotides encodant une séquence dégénérée de 27 nucléotides sera abordée plus en détail à la fin de ce chapitre (section 6).

Une autre difficulté propre à cette technique, souvent rapportée dans la littérature, est qu'il est impossible de distinguer l'insert coupé de l'insert non coupé ou partiellement coupé. En effet, le fragment d'ADN codant un nonapeptide flanqué des deux sites de restriction et des codons de terminaison est d'environ 60 paires de bases, ce qui ne peut être résolu que sur gel de polyacrylamide. Étant donné la nature dégénérée des bases, l'ADN aura des propriétés de migration légèrement variables et ne se présentera pas sous la forme d'une bande de grosseur distincte, mais plutôt comme une bande floue. Nous pouvons aussi observer la présence de la bande formée par les polynucléotides qui seront restés simple brin (Figure 26).

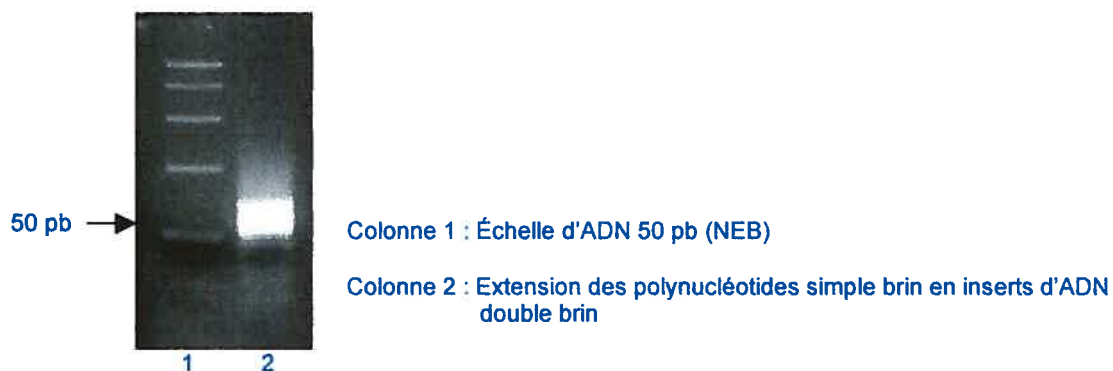


Figure 26. Migration sur gel d'agarose GTG Nusieve 4,5 % des polynucléotides codant pour la banque de nonapeptides, complétés en ADN double brin par extension avec la polymérase d'ADN.

Cette incapacité de bien distinguer les inserts coupés de ceux non coupés rend compliquée, voire impossible, la purification sur gel. En plus de la faible résolution sur gel, la limite inférieure de liaison de fragments d'ADN à la matrice des colonnes de purification est de 70 paires de bases. Conséquemment, nous avons purifié l'ADN par purification

phénol/chloroforme et précipitation à l'éthanol à -80°C en présence d'additifs comme le sodium acétate, le MgCl_2 et le glycogène. Cependant, la précipitation efficace de ces petits inserts entraîne aussi la précipitation des extrémités coupées, donc ne règle en rien le problème de la purification des inserts d'ADN.

À la fin du lien souple de l'amorce qui sert à cloner l'insert, nous avons modifié deux nucléotides afin de créer un site de restriction *BamH* I, sans perturber l'ordre ou la nature des codons. Une autre séquence *BamH* I se trouve dans le vecteur rétroviral pBABE (à la fin du gène *gag*). Nous avons criblé les clones de la banque en utilisant cette enzyme. La digestion d'un clone qui possède le peptide correspond à deux bandes, l'une à 1400 pb et l'autre à 4000 pb. Les difficultés de digestion du site *Dra* III utilisé pour le clonage nous ont amené à tenter de cloner l'insert avec le site alternatif *Mfe* I placé après le lien souple. Le site *Mfe* I fut utilisé comme solution de rechange, car *Dra* III a une séquence de reconnaissance peu commune (CACNNGTG) et il semble digérer de façon peu spécifique (Figure 27). Cependant, l'utilisation du site *Mfe* I introduit inévitablement deux codons supplémentaires entre le lien souple et la banque de peptides. De plus, l'utilisation du site *Mfe* I comme site de clonage ne permet pas le criblage des clones par digestion enzymatique du site *BamH* I.



Figure 27. Digestion enzymatique du vecteur pBABE P_{PGK} HA/Puro^R/lien souple. Le vecteur fut coupé avec les enzymes *Dra* III et *Cla* I.

2.3 Ligature

En utilisant cette méthode d'insert-vecteur, que ce soit en incluant le lien souple pour utiliser les sites de clonage *Dra* III et *Cla* I ou sans lien souple pour utiliser les séquences

Mfe I et *Cla* I, la ligature ne semble pas se produire puisque peu de vecteurs circularisés apparaissent sur le gel d'électrophorèse des produits de ligature (Figure 28).



Figure 28. Ligatures obtenues avec la stratégie insert-vecteur. Des inserts constitués de la banque de nonapeptides (ligature aux sites *Mfe* I et *Cla* I) ou constitués du lien souple et de la banque de nonapeptides (ligature aux sites *Dra* III et *Cla* I) sont intégrés au vecteur rétroviral pBABE. Cette expérience est représentative de la meilleure ligature obtenue avec la stratégie de clonage insert-vecteur.

2.4 Efficacité de la transformation

Étant donné que la ligature était peu efficace, il était important de vérifier la proportion de clones ayant intégré la banque et l'intégrité de l'ADN. De plus, l'évaluation de l'efficacité de transformation des produits de ligature sert à déterminer le nombre total de clones que nous pouvons obtenir. Ce nombre, multiplié par le pourcentage de clones ayant intégré la banque, correspond à la diversité réelle de la banque de peptides. Dans ce but, 2 μ l de ligature ont été utilisés pour transformer des bactéries chimiocompétentes *Max Efficiency Stbl2 Cells*. Le nombre total de clones obtenus fut largement en dessous du nombre attendu. Alors que le contrôle positif pUC19 a été transformé avec une efficacité de 10^8 colonies/ μ g, l'ADN ligaturé nous a permis d'obtenir seulement 10^3 colonies par transformation. Ce nombre est insuffisant pour créer une banque de forte diversité.

Nous avons ensuite analysé par électrophorèse l'ADN combiné de clones amplifiés en milieu LB ampicilline. Le gel montre que l'ADN des clones fait avec les sites *Dra* III et *Cla* I ne ressemble en rien à l'ADN rétroviral circulaire ni à celui des clones faits avec les enzymes *Mfe* I et *Cla* I (Figure 29).

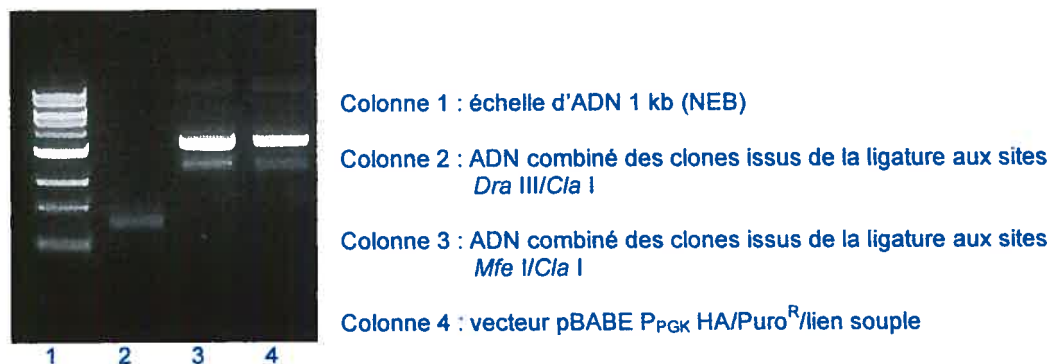


Figure 29. Électrophorèse de l'ADN des clones de banque de peptides générés par la méthode insert-vecteur. Les ADN des clones provenant des ligatures aux sites *Dra* III et *Cla* I ou *Mfe* I et *Cla* I ont été combinés et analysés par électrophorèse sur gel d'agarose 1 %. L'ADN des clones provenant de l'insertion aux sites *Dra* III et *Cla* I montre un taux très élevé de recombinaisons.

2.5 Analyse sur gel

L'ADN digéré *Bam*H I des quelques clones ayant poussé sur pétri ampicilline fut analysé sur gel d'agarose et par séquençage. La digestion *Bam*H I des clones positifs donne deux bandes : l'une à 4000 pb et l'autre à 1400 pb. La digestion *Bam*H I des vecteurs parentaux n'ayant pas intégré la banque produit une seule bande à 5400 pb.

La majorité des clones issus de la ligature aux sites *Dra* III et *Cla* I semblent avoir été tronqués (Figures 30 et 31, pages 67 et 67). Pourtant, ils ont poussé sur milieu ampicilline. Nous pouvons présager que le vecteur rétroviral a été entièrement ou partiellement éliminé du plasmide.

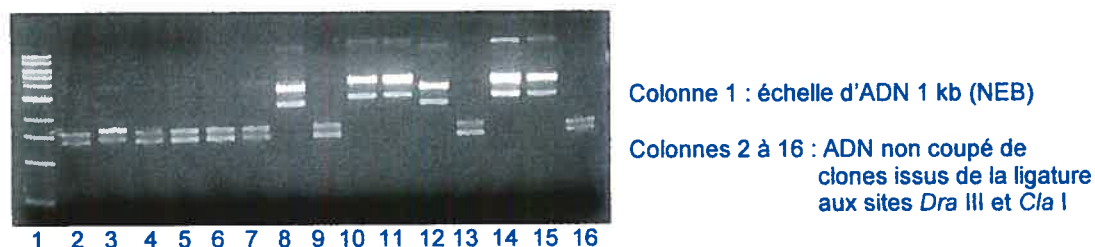


Figure 30. Analyse par électrophorèse sur gel d'agarose des clones issus de la ligature aux sites *Dra* III et *Cla* I. Ici, l'ADN n'a pas été digéré avec l'enzyme *Bam*H I. Cette figure illustre le nombre élevé de recombinaisons issues de la transformation d'ADN partiellement ligaturés.

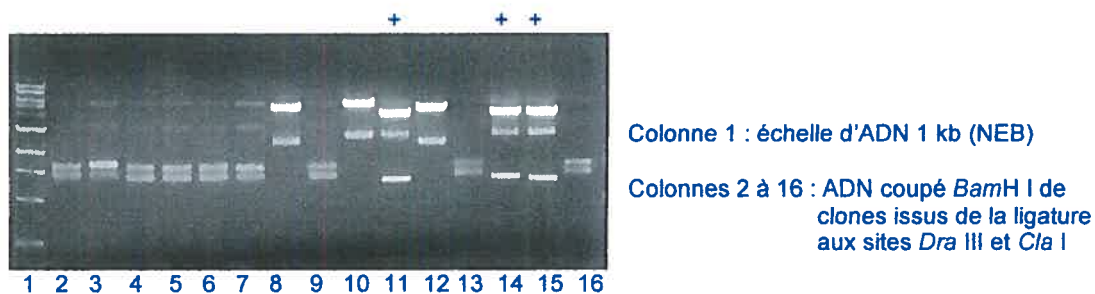


Figure 31. Analyse par électrophorèse sur gel d'agarose des clones issus de la ligature aux sites *Dra* III et *Cla* I. Ici, l'ADN a été digéré par l'enzyme *Bam*H I. Trois clones sur quinze possèdent le site *Bam*H I dans le lien souple, donc présumément la banque de peptides.

Le criblage de nombreux clones avec l'enzyme *Bam*H I nous a permis d'établir qu'environ 16 % des clones ont intégré l'insert de la banque de peptides. L'ADN de nombreux clones semble s'être recombinaisonné, conséquence possible de la transformation d'ADN dont la ligature est incomplète. Nous avons aussi analysé par électrophorèse l'ADN des clones issus de la ligature *Mfe* I/*Cla* I (Figure 32). Cependant, ces clones ne peuvent être analysés par digestion. En effet, le site de restriction *Bam*H I se trouve dans le lien souple et celui-ci ne fait pas partie de l'insert lorsque le site *Mfe* I est utilisé. Dans ce cas, les clones doivent être analysés par séquençage.



Figure 32. Analyse des clones non coupés issus de la ligature aux sites *Mfe* I et *Cla* I. Les clones positifs ne peuvent être criblés à l'aide de l'enzyme *Bam*H I, car l'insert ne contient pas le lien souple.

Contrairement aux clones obtenus avec l'insert *Dra* III/*Cla* I, cette méthode engendre peu de recombinaisons. La migration de l'ADN de ces clones est identique à celle du vecteur rétroviral pBABE contrôle (voir Figure 29, colonne 4, page 67). Le nombre de clones possédant réellement l'insert pourra être déterminé par séquençage.

suivi par les codons d'arrêt ou par un peptide de trois acides aminés seulement. Une séquence inconnue de 88 nucléotides s'est même introduite entre la fin de l'insert et le vecteur de l'un des clones.

Ligature *Mfe* I/*Cla* I

Séquence attendue

[...]TCTGGGTCCGGCTCTGGGCAATTGNNKNNKNNKNNKNNKNNKNNKNNKNNKNNK**TGATGA**ATCGATAA[...]

Clone 120804(1.2) : Ce clone ne possède pas de peptide.

[...]JGCCTCTGGGTCCGGCTCTGGGCAATT**TGATGA**ATCGATAAAAT[...]

Clone 120804(2.1) : Ce clone possède un peptide de 3 acides aminés seulement. De plus, une séquence étrangère de 88 nucléotides s'est insérée dans le site *Cla* I.

[...]JGCCTCTGGGTCCGGCTCTGGGCAATTGGAGCAGCCG**TGATGA**ATCGTTATTTTATCGCGATTCATCAC
AAGAGGAAGGACTAGAAGACCAGCCACAATTCGATTCATCAAGGGACGCAGCCCGCGCGGATACCGCCC
AA[...]

2.7 Résumé

Il s'est avéré que le clonage insert-vecteur introduit des difficultés non négligeables. Étant donné le nombre important de séquençages qu'un tel projet requiert, il était important que la technique utilisée permette un criblage des clones à l'aide d'une enzyme de restriction. Cet aspect nous a forcé à abandonner le clonage aux sites *Mfe* I et *Cla* I, et cela, quoique les clones issus de cette méthode soient semblables, sur gel d'électrophorèse, au vecteur non digéré, donc présentant moins de recombinaisons. Les deux séquences qui ont été analysées ont cependant révélé des peptides tronqués. À ces problèmes s'ajoute la difficulté d'obtenir une ligature efficace. Ceci se traduit par un très faible nombre de clones et plusieurs recombinaisons. L'efficacité de la ligature est un critère primordial pour l'obtention d'une diversité importante de la banque. Un point commun aux deux essais de clonage insert-vecteur est le site de ligature *Cla* I. Ayant des doutes sur l'efficacité de digestion de l'enzyme *Cla* I, nous avons procédé à un test de clivage du site en fonction du vecteur utilisé. Il nous est apparu que l'enzyme *Cla* I coupe à des efficacités variables dépendamment de la séquence en amont et en aval du site de restriction (Figure 33, page 71).

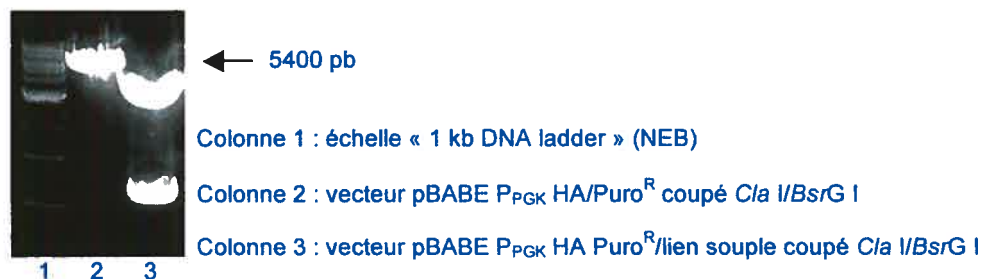


Figure 33. Efficacité de digestion du site *Cla* I en fonction de la séquence en amont du site. Deux vecteurs, pBABE P_{PGK} HA/Puro^R et pBABE P_{PGK} HA/Puro^R/lien souple, sont coupés avec les enzymes *BsrG* I (site unique dans la séquence *gag*) et *Cla* I (site unique en aval de Puro^R ou du lien souple, selon le vecteur).

Il semble évident que le site *Cla* I est inadéquat au clonage de la banque de peptides. L'absence de la séquence du lien souple devant le site *Cla* I semble complètement éliminer la capacité de l'enzyme à couper son site de restriction. Nous devons donc nous demander si cette enzyme sera fonctionnelle lorsque le site de restriction sera précédé de la séquence aléatoire des peptides. Afin de contourner le problème posé par le site *Cla* I, nous avons pensé introduire l'insert entre les sites *Dra* III et *Mfe* I. Cependant, nous avons vu que la digestion par l'enzyme de restriction *Dra* III est peu spécifique pour ce vecteur et son utilisation est à proscrire. Il serait aussi possible de cloner la banque entre les sites *Mfe* I et *Nhe* I, mais la digestion de ces deux sites consécutifs risque d'être peu efficace. En effet, *Nhe* I est très peu actif lorsque le site de restriction n'est pas bordé par plusieurs nucléotides. À ces considérations s'ajoute la difficulté de faire la synthèse de longs polynucléotides. À 70 paires de bases, une majorité des amorces seront tronquées, un phénomène relié à l'efficacité de la synthèse. Seule une purification sur gel de polyacrylamide permet de sélectionner les ADN de pleine longueur. Toutefois, la nature aléatoire de la séquence peptidique ne permet pas une résolution parfaite de la bande d'ADN et complique la purification. La problématique de la synthèse des amorces sera détaillée à la fin de ce chapitre, à la section 6. Bien que l'utilisation de cette méthode insert-vecteur comporte des difficultés majeures, elle reste l'une des seules techniques ayant permis à d'autres groupes de créer des banques de peptides. Plusieurs des problèmes rencontrés lors de nos expériences sont relatifs au concept d'une banque de peptides en fusion avec d'autres protéines et pourraient ne pas être un obstacle en d'autres circonstances.

3. INSERTION DE LA BANQUE PAR AMPLIFICATION

3.1 Stratégie du clonage

L'insertion de la banque par complémentarité des polynucléotides comporte des difficultés lors de la purification des petits inserts. Nous avons donc pensé amplifier un insert d'ADN plus long, allant du site *Cla* I¹ au site *BsrG* I situé dans le gène *gag* du vecteur rétroviral. Dans ce dessein, une grande partie du vecteur rétroviral est amplifiée en utilisant un polynucléotide antisens encodant la banque de peptides et un oligonucléotide sens correspondant à une région éloignée dans le vecteur rétroviral. D'autres expériences ont été tentées en utilisant divers sites alternatifs au site *BsrG* I (*EcoR* I situé après l'épitope HA, *Sca* I et *Dra* I situés dans le gène de résistance à l'ampicilline et *Nde* I situé après le promoteur PGK). Quels que soient les sites utilisés, le principe demeure le même. Une étape de ligature insert-vecteur classique permet de compléter le vecteur rétroviral et d'introduire ainsi la banque de nonapeptides (Figure 34).

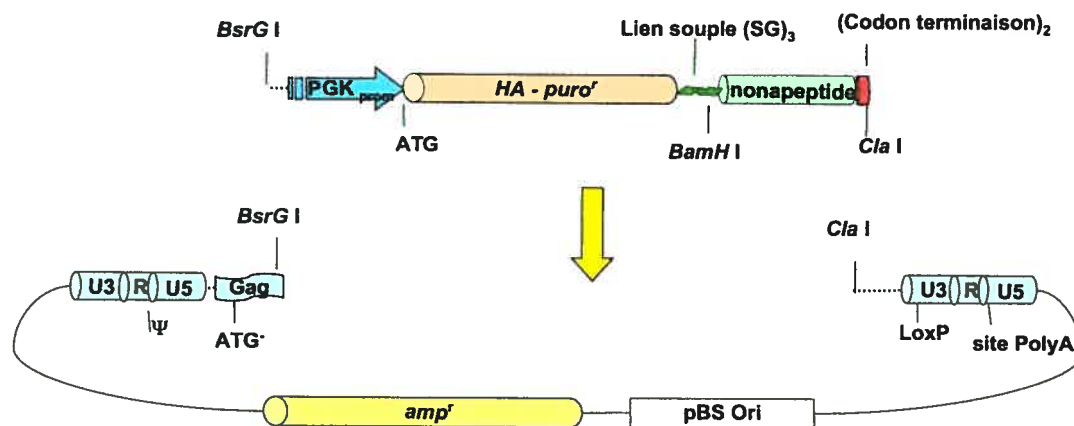


Figure 34. Schématisation du clonage par amplification de la banque de peptides. Cette méthode insert-vecteur génère un insert de grosseur moyenne incluant la banque de nonapeptides. Ici, l'insert est inséré aux sites *BsrG* I et *Cla* I du vecteur rétroviral linéarisé.

Le choix des sites d'insertion a été dicté par certaines considérations pratiques. Nous possédions déjà les oligonucléotides pour le site *EcoR* I qui fut utilisé précédemment pour

¹ Il est à noter que plusieurs des expériences ont été faites de façon concomitante. Nous n'avons pas, à ce moment, les données montrant les difficultés inhérentes à l'utilisation du site *Cla* I.

la construction des vecteurs contrôles pBABE P_{PGK} HA/Puro^R/lien souple. Dans le vecteur rétroviral, ce site est relativement proche de *Cla* I et ne permet pas une purification sur gel très efficace du vecteur coupé. Ceci se traduit par un nombre important de vecteurs parentaux lors de l'analyse des clones. Nous avons donc opté pour le site *BsrG* I, situé dans le gène *gag* du vecteur rétroviral. Les problèmes d'autoligature (Figure 38) de ce site nous ont fait ensuite opter pour le site *Dra* I. Cette séquence est située après le gène de résistance à l'ampicilline. Contrairement au vecteur coupé *Dra* III / *Cla* I utilisé précédemment, la digestion enzymatique à des sites distants permet une purification efficace du vecteur linéarisé et une quantité appréciable d'ADN amplifié (Figures 35 et 36, page 73 et Figure 37, page 74). Ces conditions optimales devaient permettre une ligature plus efficace.

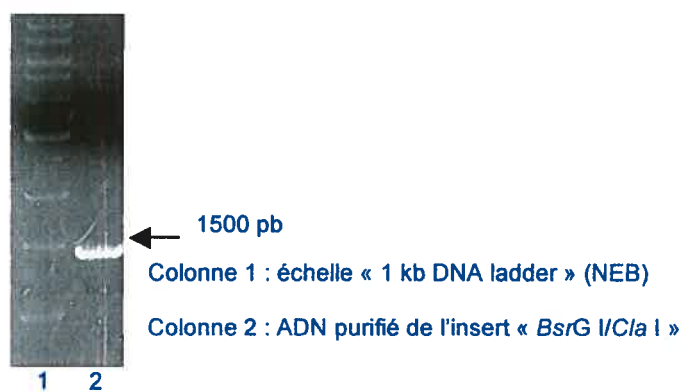


Figure 35. Analyse sur gel de l'insert de la banque de nonapeptides du site *BsrG* I au site *Cla* I.

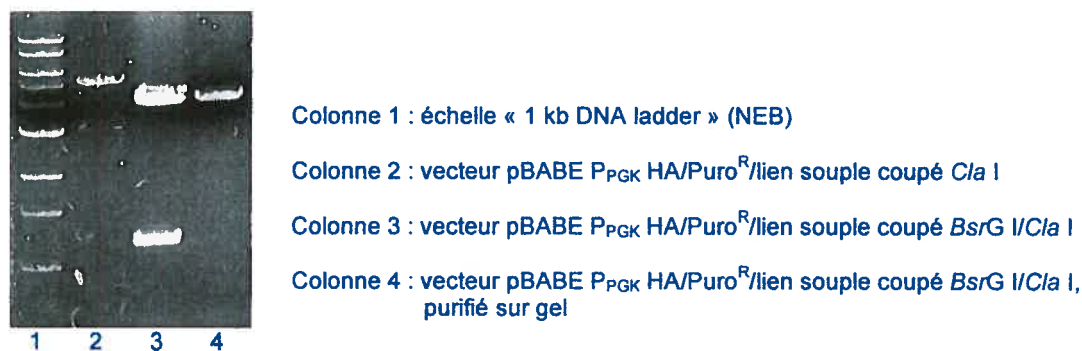


Figure 36. Analyse sur gel du vecteur pBABE P_{PGK} HA/Puro^R/lien souple coupé *BsrG* I ou *Cla* I avant et après purification sur gel.

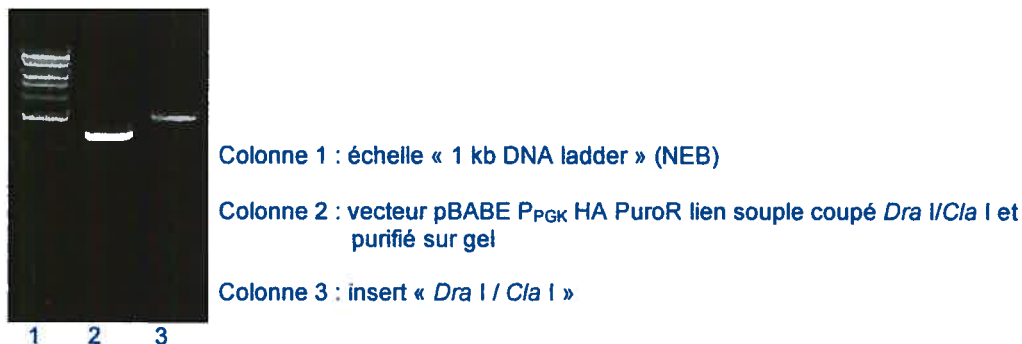


Figure 37. Analyse par électrophorèse des éléments requis pour la ligature insert-vecteur aux sites *Dra* I et *Cla* I. Nous retrouvons le vecteur pBABE P_{PGK} HA/Puro^R/lien souple coupé *Dra* I et *Cla* I après purification sur gel, ainsi que l'insert de la banque de nonapeptides entre les sites *Dra* I et *Cla* I.

3.2 Ligature

La ligature des inserts au vecteur rétroviral linéarisé s'est révélée décevante. Divers essais de déphosphorylation du vecteur ou de l'insert montrent que ceux-ci se lient à eux-mêmes s'ils ne sont pas déphosphorylés, malgré le fait que deux enzymes différentes aient été utilisées à chaque extrémité (*Bsr*G I : T|GTACA et *Cla* I : AT|CGAT) (Figure 38). Nous voyons que si nous déphosphorylons le vecteur, l'insert se lie à lui-même. Si nous déphosphorylons l'insert, c'est le vecteur qui fait de l'autoligature. Dans ces conditions, il est impossible d'obtenir un bon rendement de ligature insert-vecteur.

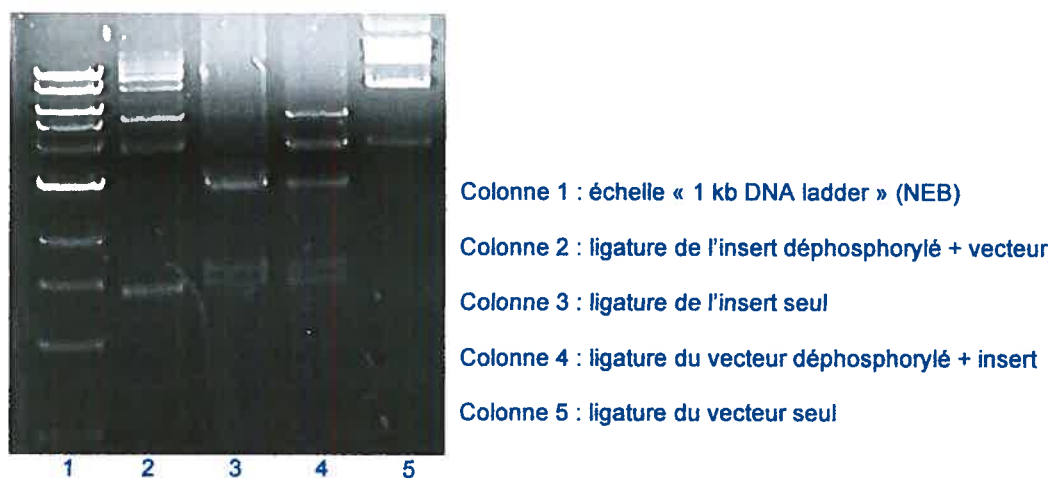


Figure 38. Ligature de l'insert encodant la banque de nonapeptides selon que le vecteur ou l'insert est déphosphorylé. Cette figure représente la ligature entre les sites *Bsr*G I et *Cla* I.

La ligature aux sites *Dra* I et *Cla* I s'est révélée très inefficace. Cependant, les vecteurs linéarisés avec ces mêmes enzymes ont été efficacement ligaturés (Figure 39). Les conditions de ligature ne semblent donc pas être en cause. La ligature des inserts et du vecteur révèle une bande à 5400 pb, ce qui démontre qu'une partie des ADN se sont liés à l'une des extrémités. Cependant, aucune ligature du vecteur ne semble se produire. L'absence ou la digestion incomplète de l'un des deux sites de restriction semble l'explication probable de ce phénomène. L'expérience d'efficacité de digestion du site *Cla* I illustrée à la figure 33 montre que les difficultés de digestion de cette enzyme peuvent être en partie responsables de cette ligature partielle. De plus, les analyses des polynucléotides utilisés pointent vers une absence du site *Cla* I. Cet aspect du problème est analysé en détail à la section 6.

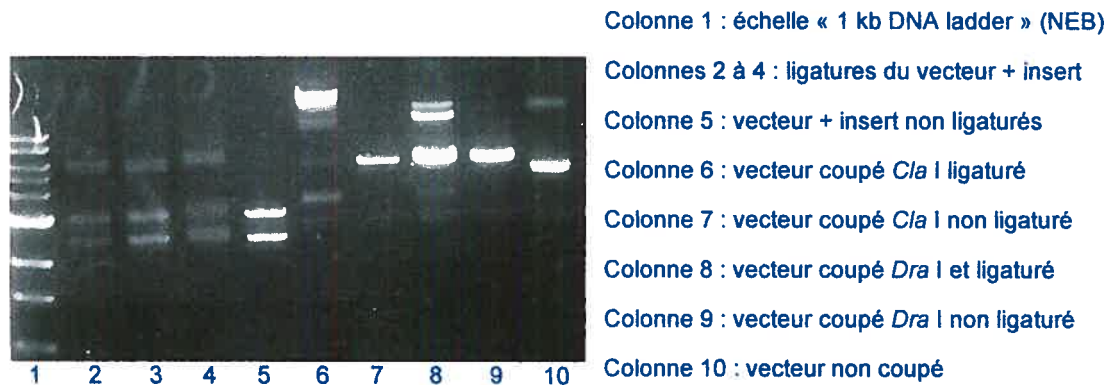


Figure 39. Ligature de l'insert de nonapeptides entre les sites *Dra* I et *Cla* I. Les vecteurs rétroviraux coupés seulement *Cla* I ou *Dra* I, puis ligaturés ont servi de contrôle de l'efficacité de la ligature.

3.3 Efficacité des transformations

Les efficacités des transformations montrent que ces ADN ne peuvent pas être utilisés pour générer une banque de peptides de grande diversité (Tableau II, à la page 76). La transformation de produit d'ADN partiellement ligaturé explique les piètres résultats que nous avons obtenus.

Transformations	Bactéries chimiocompétentes <i>MAX Efficiency Stbl2</i>	Bactéries électrocompétentes <i>ElectroMAX Stbl4</i>
Ligature insert-vecteur aux sites <i>EcoR I/Cla I</i>		2×10^4 colonies/transformation
Ligature insert-vecteur aux sites <i>BsrG I/Cla I</i>	212 colonies/transformation	$9,4 \times 10^4$ colonies/transformation
Ligature insert-vecteur aux sites <i>Dra I/Cla I</i>	270 colonies/transformation	200 colonies/transformation
pUC19	$1,2 \times 10^7$ colonies/ug	$2,3 \times 10^9$ colonies/ug

Tableau II. Efficacité des transformations des ADN provenant des ligatures aux sites *EcoR I/Cla I*, *BsrG I/Cla I* ou *Dra I/Cla I*. Le plasmide pUC19 fut utilisé comme contrôle positif.

3.4 Analyse des clones

La digestion *BamH I* des clones positifs produit deux bandes : l'une à 4000 pb et l'autre à 1400 pb. La digestion enzymatique des clones obtenus révèle que seulement 30 % des clones provenant de la ligature *BsrG I/Cla I* semblent avoir intégré l'insert de la banque de peptides. Ce nombre est encore moins élevé (7 %) pour les clones provenant de la ligature *Dra I/Cla I*. La ligature *EcoR I/Cla I*, quant à elle, n'a donné que des vecteurs parentaux. Le site de restriction *EcoR I* étant proche de *Cla I*, la purification sur gel du vecteur coupé est moins efficace que pour les autres sites.

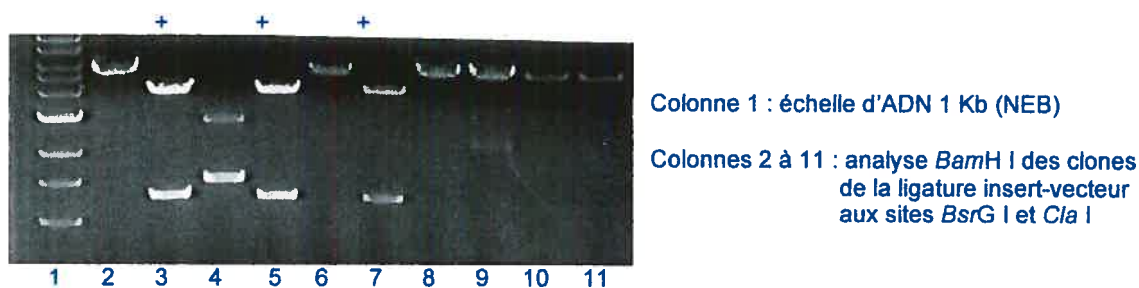


Figure 40. Analyse par digestion enzymatique *BamH I* de clones issus de la ligature insert-vecteur aux sites *BsrG I* et *Cla I*. Ici, trois clones sur dix semblent posséder le lien souple et, vraisemblablement, la banque de peptides.



Figure 41. Analyse par digestion *Bam*H I de clones issus de la ligation insert-vecteur aux sites *Dra* I et *Cla* I. Un seul clone sur quinze semble posséder la banque de peptides.

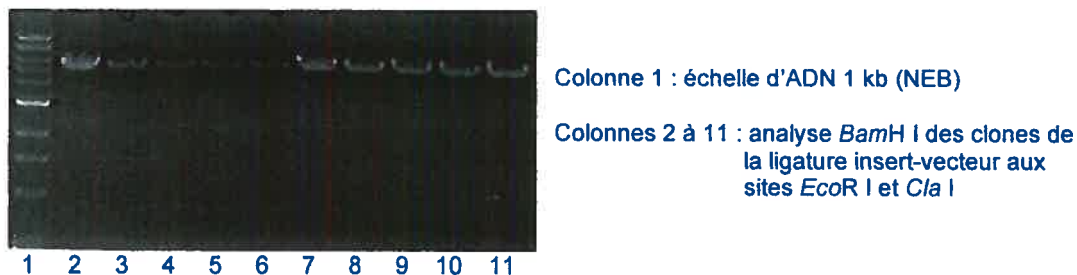


Figure 42. Analyse par digestion enzymatique *Bam*H I de clones issus de la ligation insert-vecteur aux sites *Eco*R I et *Cla* I. Aucun clone ne possède la banque de peptides.

3.5 Séquençages

Séquence attendue

[...]CICTGGGTCCGGATCCGGGNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNK**TGATGA**ATCGATAAAAT[...]

Ligation *Bsr*G I/*Cla* I

Clone 220404(1)

[...]CICTGGGTCCGGATCCGGGTTTCCTATGGTCTGTCCCGTCAGGCTG**TGATAG**ATCGATAAAAT[...]

Clone 220404(2)

[...]CTCTGGGTCCGGATCCGGGACGGACAAGTGAATTA**CTCGACTCGT****TGATAG**ATCGATAAAAT[...]

Clone 220404(4)

[...]CTCTGGGTCCGGATCCGGGCCTCTTACTTTCTTCATGAATCACCTG**TGATAG**ATCGATAAAAT[...]

Ligature *Dra* I/*Cla* I

Clone 181104(5)

[...]CTCTGGGTCCGGATCCGGGACGTTTCCTCATACGCCGTATGGGTGGT**GATGA**ATCGATAAAAAT[...]

Clone 181104(8) : Le site de restriction *Cla* I est aussi altéré.

[...]CTCTGGGTCCGGATCCGGGTATTCCGGAGTGTTGTGGGGCGCCTGTG**TATGA**ATCG**CG**ATAAAAAT[...]

Clone 181104(10) : Ce clone ne possède pas de peptide ni de codon d'arrêt. De plus, le site *Cla* I est incomplet.

[...]CTCTGGGTCCGGATCCGGGCGATAAAAAT[...]

Le séquençage de clones obtenus par la transformation des produits de ligature montre que la majorité des peptides obtenus sont de la bonne longueur (neuf acides aminés). Cependant, l'un des clones ne possède pas de peptide, malgré la présence du site *Bam*H I dans le lien souple. Ceci signifie que la séquence encodant le peptide n'était pas présente dans le polynucléotide utilisé pour l'amplification. Une autre possibilité est que le peptide a été clivé à la suite de la transformation. Ceci pourrait être une conséquence possible de la transformation d'ADN partiellement ligaturés.

3.6 Résumé

Les problèmes inhérents à l'autoligature de l'insert et du vecteur rendent la ligature peu efficace. Cela se traduit par l'obtention de peu de clones, un nombre trop faible pour que nous considérions cette méthode pour la construction de la banque de peptides. Malgré les problèmes de ligature et de transformation, la majorité des clones séquencés possèdent un nonapeptide. De plus, la séquence de ces peptides semble aléatoire. Encore une fois, un point commun aux divers essais de clonage insert-vecteur est le site de ligature *Cla* I. Comme il a été mentionné précédemment, les expériences ont été faites de façon concomitante. Les résultats démontrant l'inefficacité de digestion du site *Cla* I ne nous étaient donc pas encore connus. Au problème de clivage du site *Cla* I s'ajoutent les difficultés de synthèse des polynucléotides. Ce dernier aspect, commun à toutes les méthodes, sera discuté à la fin de ce chapitre à la section 6.

Une solution au problème de l'autoligature consiste à n'utiliser qu'un seul site de restriction. Ceci implique que l'on doit amplifier tout le vecteur rétroviral (~5,4 kb) en utilisant des amorces possédant le même site de restriction à leurs extrémités. L'amplification comporte cependant une donnée incertaine. Puisque la séquence de la banque est de nature aléatoire, cette séquence ne trouvera pas son équivalent sur le brin d'ADN complémentaire. Ceci induit la formation d'une bulle d'ADN de 27 nucléotides dont les bases ne sont pas appariées. Il est possible que les systèmes de réparation de l'ADN des bactéries compétentes procèdent au réarrangement de cette bulle d'ADN lors de la transformation. Le système de réparation le plus probable pour cette opération est le mécanisme MMR, décrit au point 2.2.3 de l'introduction (Figure 15, page 26). Ce mésappariement des bases n'a pas lieu lorsque les polynucléotides sont simplement complétés en ADN double brin. En nous basant sur ce principe de l'élongation d'un ADN simple brin, nous avons pensé à une technique basée sur la formation d'un vecteur rétroviral sous forme d'ADNc. Cette technique est décrite à la section suivante.

4. SYNTHÈSE ARN-ADNc-ADN

4.1 Stratégie du clonage

La procédure consistant à introduire un insert d'ADN dans un vecteur comporte de nombreuses difficultés. Dans le but de pallier les problèmes de ligature, de purification et de multiples insertions engendrés par cette méthode, nous avons opté pour l'utilisation d'un seul site de restriction permettant de circulariser le plasmide. L'amplification de tout le vecteur rétroviral permet d'introduire la banque de nonapeptides incluse dans l'une des amorces. Cependant, afin d'éviter la formation de séquences d'ADN mésappariées telles que décrites précédemment, nous avons pensé générer un vecteur rétroviral sous forme d'ADNc. Celui-ci servira alors de matrice à l'amorce contenant la banque pour l'hybridation et l'extension du brin complémentaire.

Deux méthodes se sont imposées pour la synthèse de ce vecteur rétroviral sous forme d'ADN simple brin. La première consiste à produire d'abord des ARN de vecteur rétroviral. Une étape de rétrotranscription avec l'amorce encodant la banque de peptides permet de

généraliser des ADNc du vecteur. S'ensuit la destruction de l'ARN du complexe ADN-ARN avec l'ARNase H et la synthèse du brin complémentaire par l'extension d'une amorce. L'autre technique consiste à produire des ADN simple brin du vecteur linéarisé avec amplification préférentielle d'un seul brin. Ceci peut être réalisé en utilisant un ratio inégal de l'amorce sens et de l'amorce antisens. Cette dernière technique sera abordée en détail à la section 5. Encore une fois, à la suite de l'obtention des ADNc, il suffit de compléter le brin simple par l'extension d'un l'oligonucléotide complémentaire par le biais d'une polymérase d'ADN. Par la suite, le vecteur rétroviral contenant la banque est circularisé par ligature du site de restriction unique présent dans les amorces sens et antisens utilisées.

La première étape de la technique ARN-ADNc-ADN consiste à générer des ARN de vecteur rétroviral grâce à un site promoteur T7 (Figure 43). En présence d'une polymérase d'ARN T7, le vecteur rétroviral est transcrit en ARN simple brin (Figure 44, page 81). Cet ARN sert alors de matrice afin d'y hybrider l'amorce encodant la banque de peptides. En présence de transcriptase inverse, la banque est introduite dans la séquence de l'ADN rétrotranscrit (Figure 45, page 81). La matrice d'ARN est alors éliminée grâce à l'ARNase H, qui élimine spécifiquement l'ARN d'un hétéroduplexe ARN-ADN. Une simple étape d'élongation à partir d'une amorce complémentaire permet de compléter le second brin d'ADN (Figure 46, page 81). La digestion enzymatique des extrémités de l'ADN néo-formé suivi de la ligature des extrémités digérées permet de circulariser le vecteur rétroviral (Figure 47, page 81).

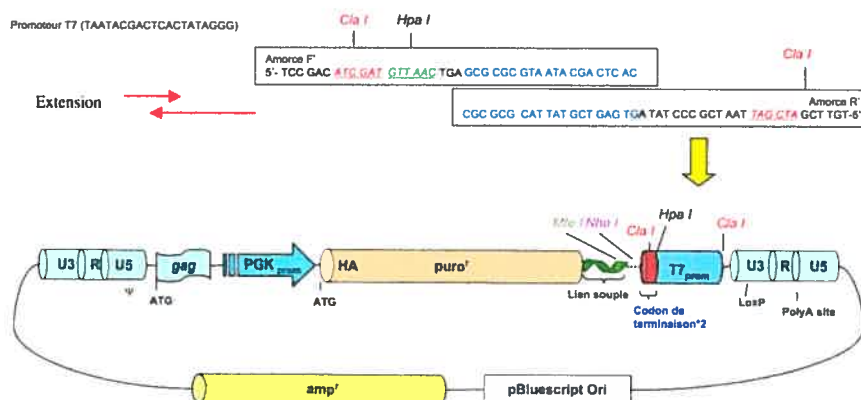


Figure 43. Schématisation de l'introduction d'un site promoteur T7 dans le vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites. Le vecteur pBABE P_{PGK} HA/Puro^R/lien souple/T7 sert à la formation de vecteurs ARN simple brin.

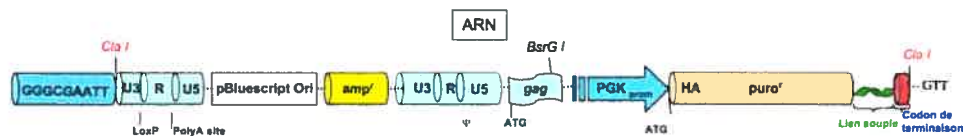


Figure 44. Schématisation de l'ARNm du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/kill sites/T7.

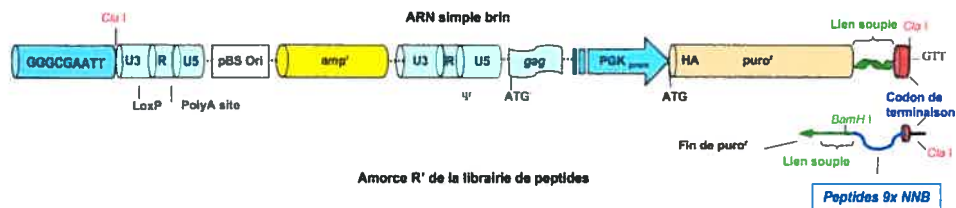


Figure 45. Rétrotranscription du vecteur avec l'amorce de la banque de peptides.

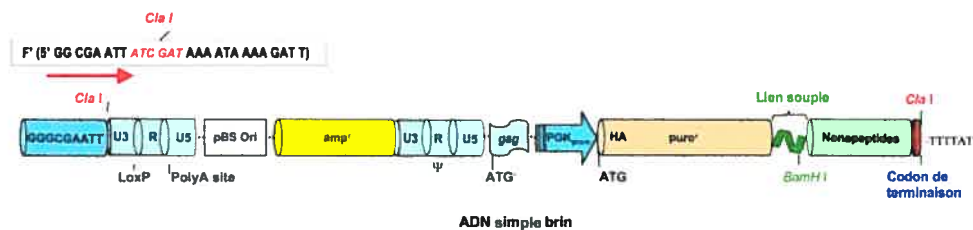


Figure 46. Synthèse du second brin du vecteur ADNC encodant la banque de peptides.

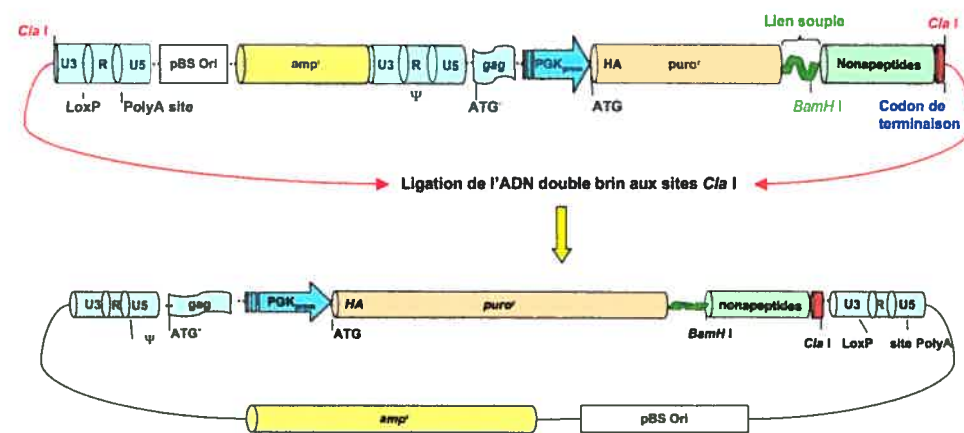


Figure 47. Ligation du vecteur rétroviral encodant la banque de peptides.

4.2 Clonage

Cette technique de génération d'un ARN simple brin et sa rétrotranscription en ADNc, puis en ADN double brin est limitée par la polymérase d'ADN. La formation des ARN ne présente aucune difficulté et est très efficace (Figure 48).

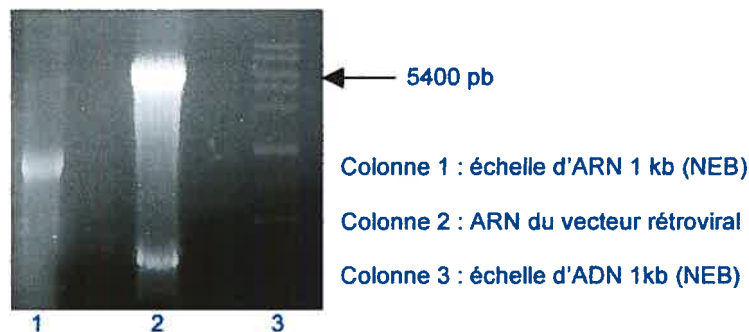


Figure 48. Analyse sur gel d'agarose/formaldéhyde de la production d'ARN de vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple.

La principale limitation se trouve dans la complémentation du brin complémentaire, car, à haute concentration d'ADN, l'activité de la polymérase se trouve inhibée. Il faut mentionner que le protocole de la rétrotranscription requiert d'utiliser une fraction de l'ADNc produit pour la synthèse du brin complémentaire. Cette limitation expérimentale amène une forte réduction de la diversité, critère réduisant l'intérêt pour cette méthode.



Figure 49. Analyse sur gel de la rétrotranscription en ADNc du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/nonapeptide.

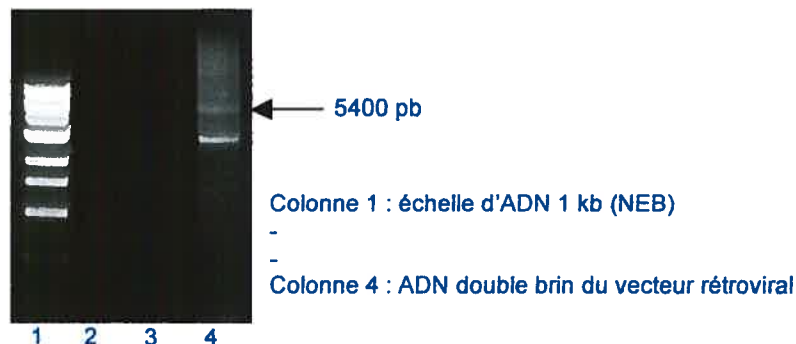


Figure 50. Analyse sur gel de l'extension du second brin d'ADN du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/nonapeptide.

La quantité d'ADN double brin obtenue lors de cette dernière étape est insuffisante pour effectuer de nombreuses transformations bactériennes. Le peu d'ADN récupéré de cette technique a été ligaturé. La ligature s'est révélée semblable à celle obtenue avec les autres techniques, c'est-à-dire peu efficace (données non montrées).

4.3 Efficacité des transformations

La transformation de bactéries chimiocompétentes *MAX Efficiency Stbl2* avec le peu d'ADN obtenu de ce clonage produit seulement 100 colonies par transformation. Les efficacités des transformations obtenues sont bien en deçà du nombre de clones requis pour la création d'une banque de nonapeptides de grande diversité. La difficulté à obtenir des ADN circularisés se traduit par une efficacité de transformation réduite et de nombreuses recombinaisons.

4.4 Analyse des clones

La digestion enzymatique des clones obtenus révèle que moins de 7 % des clones semblent posséder la banque de peptides (figure 51, page 84). Le nombre de recombinaisons est particulièrement élevé (14/15). Ce résultat montre encore une fois que la ligature efficace des ADN est primordiale pour l'obtention de vecteurs rétroviraux intègres.

effectuée en une étape n'arrive pas à rétrotranscrire et amplifier l'ADN. Un protocole à deux étapes (rétrotranscription et amplification séparée) donne de meilleurs résultats. Cependant, la quantité d'ADN obtenu est très faible. Seules les enzymes robustes comme *px*, *Phusion* et la polymérase d'ADN *KOD* peuvent amplifier ce vecteur avec une relative efficacité lorsqu'elles sont supplémentées d'additifs en forte concentration. Les conditions d'amplification optimisées permettent d'obtenir une bande visible sur gel. Néanmoins, la simple élongation d'un brin, quelle que soit la quantité de matrice ajoutée au mélange réactionnel, ne permet pas d'obtenir suffisamment d'ADN pour la création d'une banque de peptides.

5. SYNTHÈSE PAR AMPLIFICATION DIFFÉRENTIELLE

5.1 Principe de la méthode

Comme il a été dit précédemment, la technique ARN-ADNc-ADN ne permet pas d'obtenir suffisamment d'ADN pour suffire à toutes les étapes de la création d'une banque de peptides. Dans le but d'obtenir des ADNc par une méthode alternative, nous avons convenu d'utiliser une technique qui consiste à produire des ADN simple brin par amplification. Ceci est rendu possible en utilisant des proportions différentes d'amorces sens et antisens. L'utilisation de 0,5 uM de polynucléotides de banque de peptides avec 0,05 uM d'oligonucléotides correspondant à l'autre extrémité du vecteur rétroviral permet de générer une majorité d'ADN simple brin contenant la banque de peptides. Un cycle d'élongation en présence d'un excès d'oligonucléotides complémentaires génère les ADN double brin. La digestion enzymatique d'un site de restriction commun aux deux extrémités permet de circulariser le vecteur rétroviral.

L'une des difficultés rencontrées dans cette méthode est de générer suffisamment d'ADN amplifié. En effet, la synthèse d'un vecteur de 5,4 kb nécessite des conditions d'amplification optimisées. Le peptide se trouvant dans le polynucléotide, chaque réaction avortée correspond alors à une baisse équivalente de la diversité de la banque de nonapeptides.

5.2 Clonage

Les premiers essais d'amplification se sont basés sur les protocoles de mutagenèse dirigée utilisant une matrice méthylée afin d'y hybrider les polynucléotides contenant la banque de peptides. Le brin méthylé ne contenant pas la banque est détruit lors de la transformation dans les bactéries compétentes, éliminant la contamination par le vecteur parental. Lors de cette expérience, les polymérases d'ADN *pfx* et *TAQ High fidelity* ont été essayées. Quels que soient les conditions d'amplification, les enzymes ou les additifs utilisés, aucune amplification satisfaisante n'a été obtenue en employant l'ADN méthylé comme matrice. La même amplification avec la matrice non méthylée a permis l'obtention de meilleurs résultats. Sur cette base, nous avons utilisé le vecteur non méthylé comme matrice pour l'hybridation des polynucléotides.

Diverses expériences ont été tentées afin d'obtenir une quantité de vecteurs suffisante pour les étapes subséquentes de transformation et d'analyse des clones. Nous avons tenté d'amplifier un seul brin d'ADN pendant une centaine de cycles afin de former un ADNc comprenant la banque de peptides, pour ensuite compléter le brin formé à l'aide d'un oligonucléotide qui lui est complémentaire. Une autre technique utilisée consistait à utiliser un ratio 10 : 1 d'amorces, ce qui permet à la fois une amplification du vecteur de départ et la formation d'ADN simple brin. Cette technique s'est révélée la plus efficace pour le rendement final (Figure 52).



Figure 52. Analyse de l'ADN provenant de la réaction d'amplification différentielle du vecteur rétroviral. Électrophorèse sur gel d'agarose 1,2 %.

5.3 Ligatures

L'analyse de la ligature subséquente à l'amplification du vecteur montre que très peu de vecteurs se sont circularisés (Figure 53).

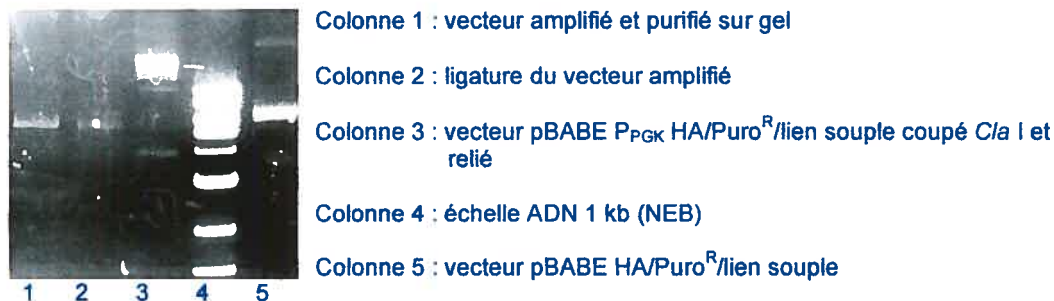


Figure 53. Évaluation de l'efficacité de ligature du vecteur rétroviral amplifié.

Les éléments communs à toutes les méthodes sont les polynucléotides comprenant la banque de nonapeptides et le site *Cla* I utilisé pour la circulariser le vecteur. Il semble que ces deux facteurs sont responsables du défaut de ligature et ont d'importantes conséquences sur le nombre de clones obtenus et sur la qualité des peptides.

Les conclusions auxquelles nous sommes arrivé à la lumière des expériences précédentes pointent vers l'intégrité des polynucléotides que nous avons utilisés. Afin d'obtenir la garantie que ceux-ci possèdent la banque de nonapeptides et les sites de restrictions requis, nous avons fait purifier les amorces sur gel de polyacrylamide. Aussi, nous avons utilisé le site *Nhe* I pour circulariser le vecteur comme solution de remplacement au site *Cla* I.

L'amplification du vecteur rétroviral avec les polynucléotides purifiés par PAGE a permis d'obtenir une bonne efficacité de ligature (Figure 54).

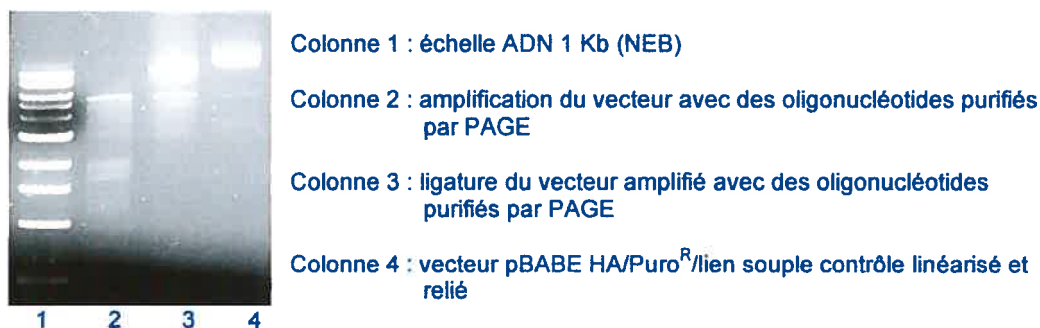


Figure 54. Amplification du vecteur à l'aide d'amorces purifiées par PAGE et ligature des ADN.

5.4 Efficacité des transformations

Une première série de transformations ont été effectuées avec les ADN partiellement ligaturés provenant des expériences basées sur l'utilisation de polynucléotides non purifiés. Les efficacités de transformations obtenues dans des bactéries chimiocompétentes *MAX Efficiency Stbl2* sont de $0,6 \times 10^3$ à $2,3 \times 10^3$ colonies/transformation, selon les expériences. Dans les bactéries électrocompétentes *ElectroMAX Stbl4*, les transformations sont de l'ordre de $1,2 \times 10^4$ à $1,3 \times 10^6$ colonies/transformation. Ces résultats montrent que les ADN partiellement ligaturés ne peuvent pas être utilisés pour obtenir une banque de peptides de grande diversité.

Une seconde série de transformations ont été faites avec les ADN ligaturés issus des amplifications effectuées avec les polynucléotides purifiés par PAGE. L'obtention d'une bonne efficacité de ligature a permis d'obtenir un plus grand nombre de clones dans les bactéries électrocompétentes (de $1,6 \times 10^6$ à $2,8 \times 10^6$ colonies/transformation). Ce nombre de clones obtenus nous permet de croire qu'une banque de forte diversité peut être générée par cette technique par le biais de quelques dizaines de transformations.

5.5 Analyse des clones par digestion enzymatique *BamH I*

L'analyse par digestion *BamH I* des clones produits avec les polynucléotides non purifiés révèle que 15 clones sur 95 analysés (16 %) semblent posséder la banque de peptides. Le nombre de recombinaisons est, encore une fois, très élevé (Figures 55 à 61).

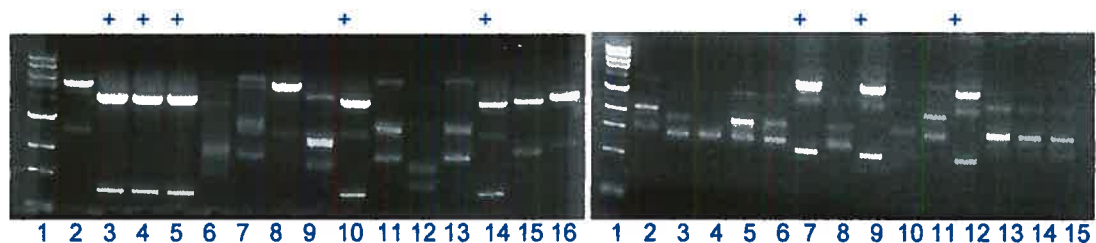


Figure 55

Figure 56

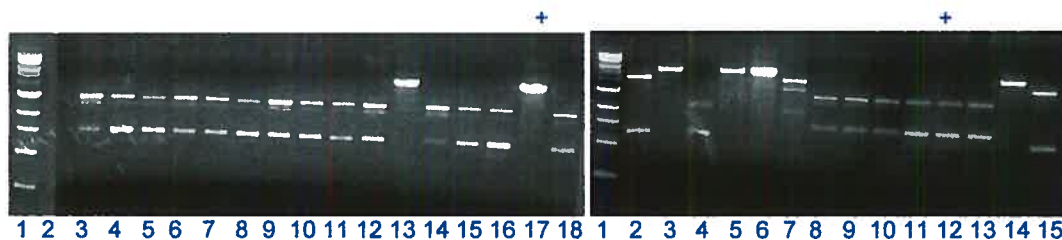


Figure 57

Figure 58

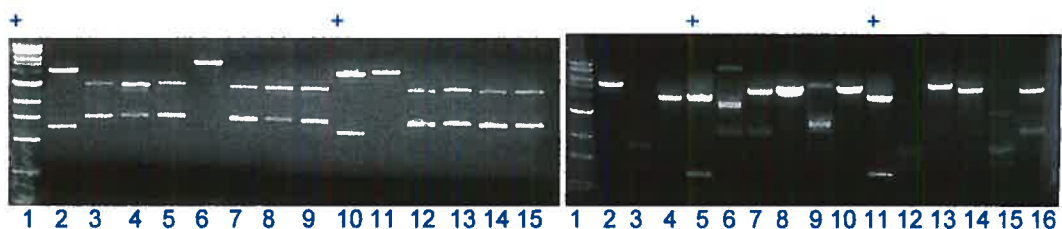


Figure 59

Figure 60



Colonne 1 : échelle d'ADN 1 kb (NEB)

Colonnes suivantes : ADN des transformants
coupé *Bam*H I

Figure 61. Analyse par digestion *Bam*H I des clones produits avec les polynucléotides non purifiés. Au total, 15 clones sur 95 (16 %) possèdent, présumément, le nonapeptide.

L'analyse par digestion des sites *Bam*H I de clones issus des amplifications effectuées avec les amorces purifiées par PAGE révèle que la majorité des 200 clones analysés possèdent la banque de peptides (66 % à 84 %, selon les expériences). Il faut aussi noter que le nombre de recombinaisons est très faible. Ces résultats contrastent fortement avec ceux qui sont obtenus précédemment avec les polynucléotides non purifiés.

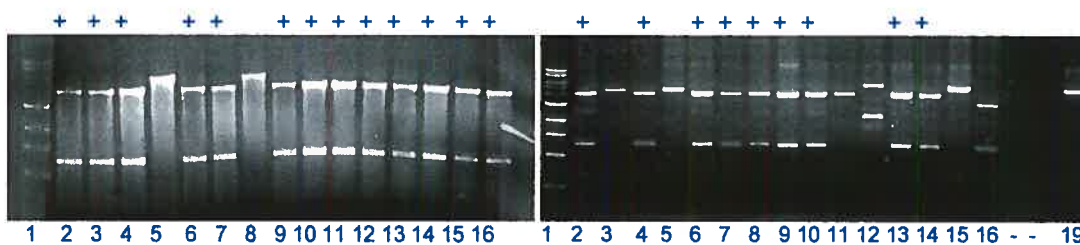


Figure 62

Figure 63

Figure 62 et Figure 63. Analyse par digestion enzymatique *Bam*H I des clones générés avec les amorces purifiées par PAGE. Plus de 200 clones ont été analysés (les gels d'analyse ne sont pas tous montrés). La colonne 1 de chaque gel représente une échelle d'ADN de 1 kb (NEB). La colonne 19 de la figure 63 représente le vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple linéarisé.

Clone 181104(8E) : L'ADN de ce clone semble avoir subi des réarrangements. On y retrouve deux liens souples : celui du vecteur et celui de l'amorce, avec la séquence du nonapeptide.

[...]CTCTGGGTCCGGCTCTGGGCAATCGATAAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGGAA
TGAAAGACCCACCTGTAGTTTGGCAAGCTAGATAACTTCGTATAATGTATGCTATACGAAGTTATCTAG
CTTAAGTGCACCTGGTGCATGACCCGCAAGCCGGTGCCCTCTGGGTCCGGATCCGGGCCTGCGGAGAG
GTGGTGTGGGATGAGTGATGAATCGCGCCGACATCGAT[...]

Clone 181104(10E) : Le site de restriction *Cla* I est altéré.

[...]CTCTGGGTCCGGATCCGGGTTGAAGAATTTTCGGGGAGGCATATGTGATGAATCGCGAT[...]

Clone 181104(12E) : L'ADN de ce clone semble avoir subi des réarrangements. On y retrouve deux liens souples : celui du vecteur et celui de l'amorce, avec la séquence du nonapeptide. Les codons de terminaison sont absents et le site de restriction *Cla* I est altéré.

[...]CTCTGGGTCCGGCTCTGGGCAATTCGATAAAAATAAAAGATTTTATTTAGTCTCCACCGAGAATGAAAAG
CCGCAAGCCCGGTGCCCTCTGGGTCCGGATCCGGGTCGAGGGTTTAGGACTGGCTTCAGCGAT[...]

Clone 181104(2A)

[...]CTCTGGGTCCGGATCCGGGCGGCTGTGTAATTTGCTTGTGCATGATTGATGAATCGAT[...]

Clone 181104(3AN) : Ce clone possède un peptide incomplet. De plus, les codons non-sens sont absents. Le site de restriction *Cla* I est aussi altéré.

[...]CTCTGGGTCCGGATCCGGGTGTCCGAT[...]

Clone 181104(3AO)

[...]CTCTGGGTCCGGATCCGGGACGTTTCCTCATACGCCGTATGGGTGGTGAATCGAT[...]

Clone 181104(4A)

[...]CTCTGGGTCCGGATCCGGGACGCGGTGTGTGAAGGCGCGGGTGGGTGATGAATCGAT[...]

Clone 181104(7A) : Le site de restriction *Cla* I est altéré.

[...]CTCTGGGTCCGGATCCGGGTATTCGGAGTGTGTGGGGCGCCTGTGTGATGAATCGCGAT[...]

Clone 181104(9A)

[...]CTCTGGGTCCGGATCCGGGACGAAGTTGAGGGGTTCAAGGTTTATTGATGAATCGAT[...]

Cette fois encore, nous pouvons observer une ligature anormale des extrémités du vecteur. Le site *Nhe* I de l'amorce encodant le nonapeptide n'a pas été digéré et s'est ligaturé de force. Le site *Nhe* I de l'autre amorce (oligonucléotide sens) a, quant à lui, été digéré efficacement. Dans le but d'évaluer la qualité de la banque finale obtenue, nous avons aussi fait séquencer 23 clones issus de la réserve en milieu glycérol. À notre grande surprise, les séquences des 23 clones analysés sont toutes identiques, à l'exception d'une séquence. Face à ce constat, nous avons fait séquencer tout l'ensemble (*pool*) des ADN qui constituaient la banque de peptides, soit l'ADN de plus de $6,6 \times 10^7$ clones (Figure 64).

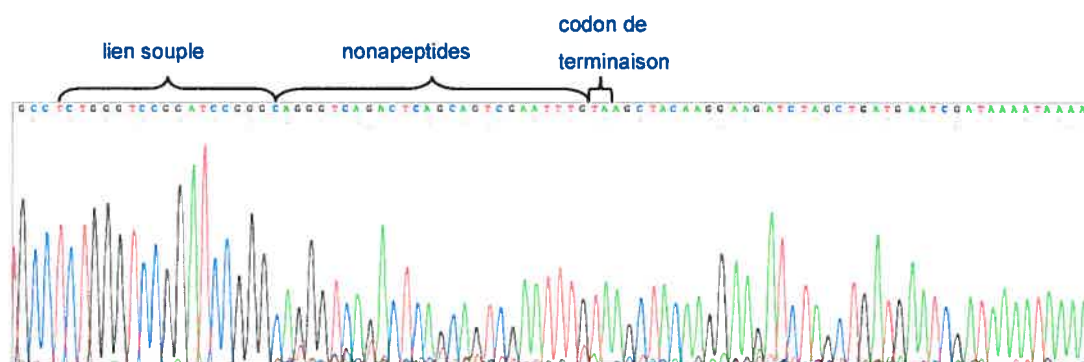


Figure 64. Séquençage de l'ADN représentatif de tous les clones de la banque de peptides ($>6,6 \times 10^7$). La même séquence se retrouve dans les analyses des clones du stock glycérol (donnée non montrée).

Au site correspondant aux peptides dégénérés, nous nous attendions à observer une quantité équivalente des quatre types de bases, donc quatre pics superposés. En effet, la séquence du peptide est encodée par la séquence $(NNK)_9$. Le séquençage de la banque a plutôt révélé une faible diversité au site correspondant au peptide. Une séquence particulière a été amplifiée préférentiellement, malgré une diversité théorique de la banque de peptides de plus de 10^{11} . De plus, le site *Nhe* I est clairement altéré dans l'ensemble des vecteurs rétroviraux amplifiés. Il n'est pas surprenant, dans ces conditions, que la ligature des vecteurs soit aussi difficile. Afin de comprendre pourquoi la banque générée à partir des amorces purifiées par PAGE montre une si faible diversité, nous avons analysé en détail la qualité de ces polynucléotides. Cette analyse fait l'objet de la section 6.

5.7 Résumé

L'introduction de la banque de peptides par amplification de tout le vecteur rétroviral a été tentée par diverses méthodes. Pour éviter la formation d'une séquence d'ADN mésapparié, nous avons développé une stratégie consistant à générer un vecteur rétroviral sous forme simple brin et encodant la banque de peptides. Un cycle final en présence d'un oligonucléotide complémentaire ramène le vecteur rétroviral sous sa forme double brin. Malgré les divers essais proposés, l'efficacité de transformation des vecteurs s'est toujours butée à une faible proportion d'ADN ligaturés. La purification sur gel de polyacrylamide des amorces utilisées nous a permis d'améliorer l'amplification des ADN et l'efficacité de la ligature. Cependant, l'analyse subséquente des clones obtenus a montré que la diversité de la banque de peptides était presque nulle.

Ces résultats étonnants sont probablement le fait de la méthode utilisée. En effet, l'amplification de tout le vecteur avec une amorce sans peptide dégénéré a permis d'obtenir une quantité appréciable d'ADN amplifié (Figure 65). Ce résultat est très différent de l'amplification avec le polynucléotide du peptide dégénéré qui génère peu d'ADN. Ceci démontre que le polynucléotide contenant une séquence dégénérée réduit l'efficacité d'amplification. Diverses raisons peuvent expliquer cette différence. Entre autres, la longueur de la séquence peptidique qui ne s'hybride pas à la matrice forme des structures secondaires qui nuisent à l'amplification. Parmi toutes les amorces de la banque de peptides dégénérés, celles qui encodent un peptide tronqué ou qui auront une conformation secondaire faible seront favorisées lors de l'amplification du vecteur. Ceci a pour conséquence de réduire la diversité finale de la banque et d'introduire un biais important. De plus, la purification PAGE des polynucléotides élimine une bonne partie des amorces initialement synthétisées. Cette sélection réduit, à la base, la diversité des peptides.

Nous avons aussi tenté d'élucider ce qui, dans toutes ces expériences, pouvait empêcher la ligature du vecteur rétroviral. Dans ce dessein, de nombreuses optimisations des conditions de ligature ont été essayées (données non montrées). Aucune condition de ligature n'a permis d'augmenter significativement l'efficacité de la ligature de notre vecteur. Divers sites de restriction ont aussi été utilisés pour ligaturer le vecteur. Aucun des sites alternatifs n'a amélioré les résultats de la ligature. Nous nous sommes alors interrogé sur

l'intégrité des polynucléotides utilisés. Dans ce dessein, nous avons procédé à l'amplification du vecteur rétroviral. Cette fois, cependant, nous avons utilisé un oligonucléotide sans la séquence dégénérée du nonapeptide. Cette expérience a permis d'élucider une partie du problème. En effet, la ligature du produit d'amplification obtenu à l'aide de l'oligonucléotide sans peptide dégénéré est relativement efficace en comparaison à celle obtenue à la suite de l'amplification de la banque de peptides (Figure 65). Ceci nous porte à croire que les polynucléotides non purifiés qui ont servi à introduire la banque sont en majorité tronqués et ne possèdent pas de site de restriction permettant la ligature du vecteur amplifié. Sur la base de ces observations, nous avons conclu que nous devons analyser les polynucléotides qui ont été utilisés pour l'introduction de la banque de nonapeptides. Cette analyse est rapportée en détail à la section 6 qui suit.

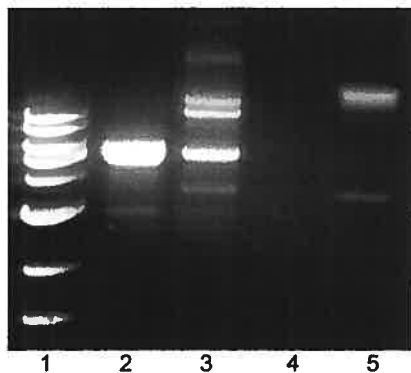


Figure 65. Démonstration que l'amplification du vecteur faite avec des oligonucléotides sans séquence dégénérée permet une ligature plus efficace du vecteur rétroviral amplifié. Une telle efficacité de ligature ne se rencontre pas lorsque les polynucléotides comprenant la banque de peptides sont utilisés.

6. CONSIDÉRATIONS TECHNIQUES COMMUNES À TOUTES LES MÉTHODES UTILISÉES

6.1 Analyse des polynucléotides encodant la banque de peptides

Toutes les tentatives de construction de la banque de nonapeptides ont été effectuées en utilisant des polynucléotides dont la synthèse suit les procédures standard. L'analyse par spectre de masse (MS) montre que cette synthèse produit une majorité d'amorces tronquées (Figure 66). Le spectre de masse du polynucléotide antisens encodant la banque de nonapeptides montre que la majorité de la synthèse s'est arrêtée entre les

nucléotides 28 à 35 (masse équivalant à 9 000–11 000 daltons). La synthèse des oligonucléotides se faisant dans l'orientation 3' vers 5', la partie conservée correspond à la séquence s'hybridant au vecteur rétroviral. La partie aléatoire de l'oligonucléotide est, en tout ou en partie, tronquée. La partie en 5' correspondant aux codons d'arrêt et au site de restriction est aussi absente. Les résultats de cette analyse expliquent les difficultés de ligature et les séquences de peptides tronqués obtenus dans les diverses stratégies de clonage.

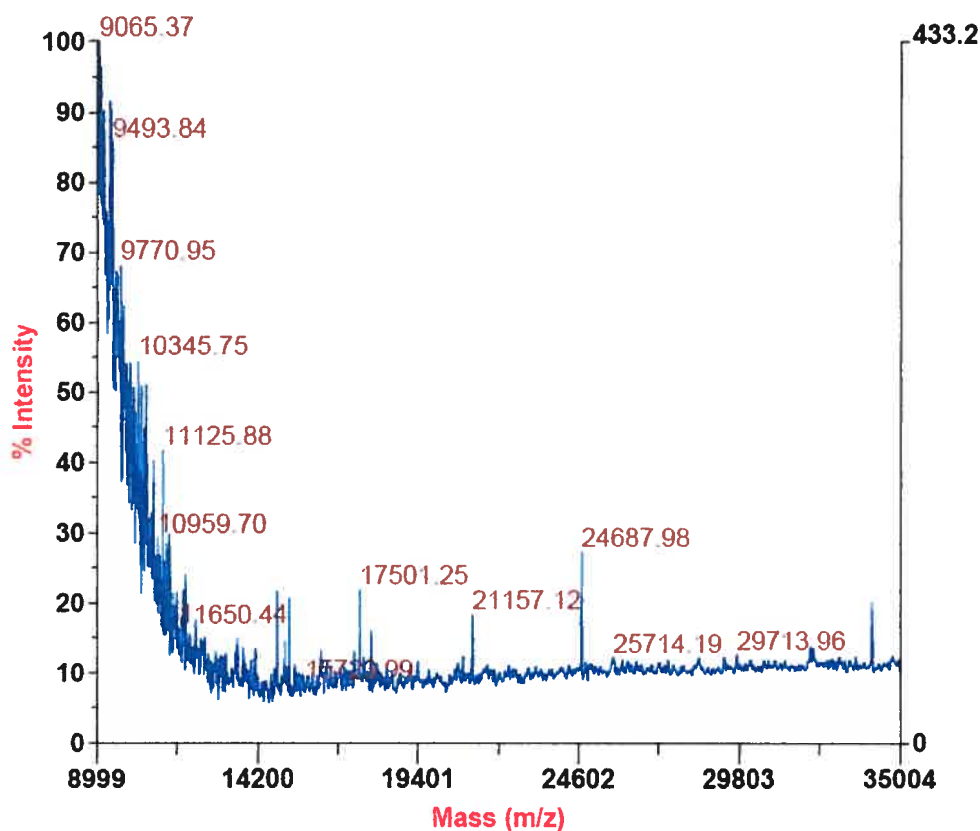


Figure 66. Spectre MS du polynucléotide antisens encodant la banque de nonapeptides. Ce spectre de masse comprend un pic majoritaire à 9 000-11 000 daltons, correspondant à un polynucléotide de 30 nucléotides. L'amorce contenant la séquence de peptides aléatoires devait comporter 82 nucléotides (24 000 daltons).

Nous pouvons également constater sur le gel de polyacrylamide de ce même oligonucléotide que, comme pour le spectre de masse, à 80 bases nous ne pouvons pas discriminer la bande d'ADN du bruit de fond (Figure 67, page 97).

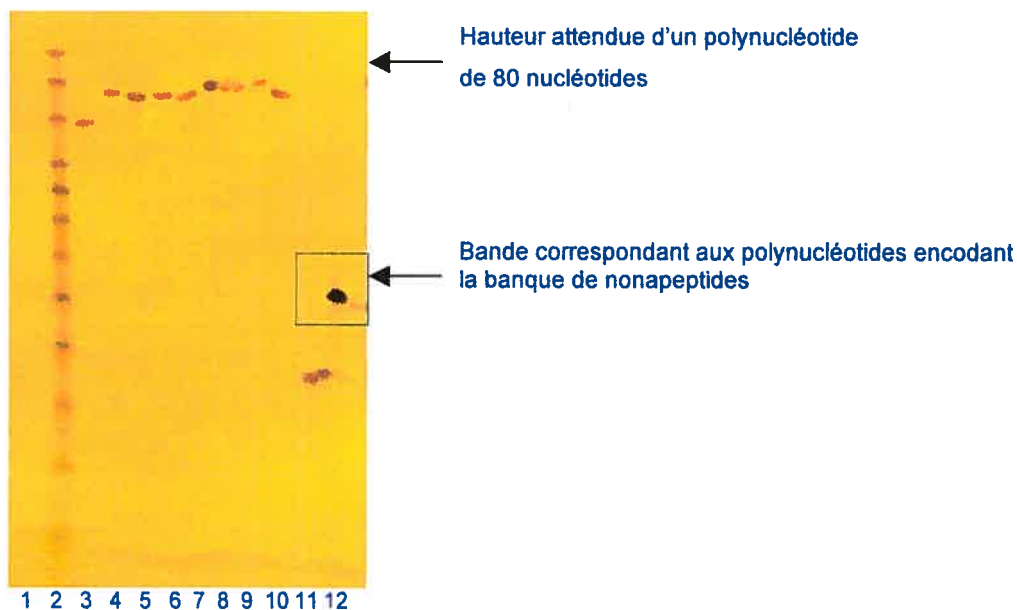


Figure 67. Analyse sur gel de polyacrylamide des polynucléotides de 82 nucléotides encodant la banque de peptides. Seule la colonne 12 contient le polynucléotide de la banque.

L'oligonucléotide complémentaire destiné à s'hybrider à l'ADN simple brin du vecteur rétroviral fut aussi analysé par spectre de masse (Figure 68). Le pic correspondant aux nucléotides de pleine longueur (27 nucléotides) est présent à 9 000 daltons. L'intégrité de cet oligonucléotide n'est donc pas mise en doute pour expliquer les difficultés de ligature.

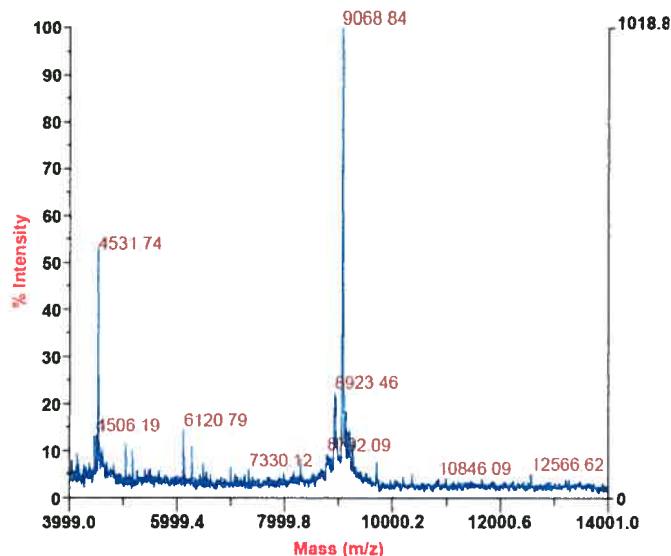


Figure 68. Spectre de masse de l'oligonucléotide sens complémentaire. Ce spectre de masse comprend un pic majoritaire à 9 000 daltons qui correspond à l'oligonucléotide de pleine longueur (27 nucléotides). (Analyse effectuée par Sigma Genosys inc.)

La synthèse plus ou moins efficace des polynucléotides peut varier selon la compagnie choisie pour la conception de ceux-ci. Nous avons fait séquencer l'ADN provenant d'amplifications faites avec des amorces de banque de peptides de deux compagnies distinctes : Alpha DNA inc. et Sigma Genosys (Figures 69 et 70). Dans les deux cas, la séquence montre que le gène Puro^R et le lien souple sont bien présents. Cependant, la séquence est fortement altérée dès le début de la banque dégénérée. Les codons de terminaison et les sites de restriction sont aussi absents ou décalés d'une base. Ces résultats corroborent les données obtenues par spectre de masse et par migration sur gel de polyacrylamide.

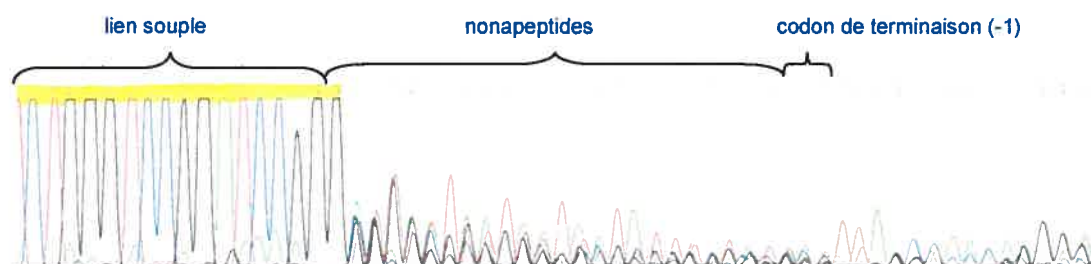


Figure 69. Analyse de la séquence de l'amplification effectuée avec le polynucléotide synthétisé par la compagnie Alpha DNA inc. et purifié par PAGE.

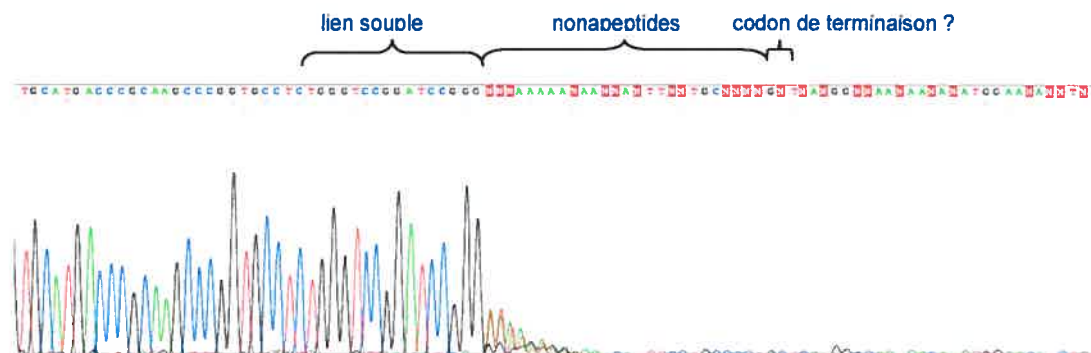


Figure 70. Analyse de la séquence de l'amplification effectuée avec le polynucléotide synthétisé par la compagnie Sigma Genosys.

7. ESSAIS DE LIGATURE AVEC DES POLYNUCLÉOTIDES BIOTINYLÉS

7.1 Synthèse séquentielle et biotinylation des polynucléotides

Dans le but de pallier le problème de l'arrêt prématuré de la synthèse des amorces, un essai original fut tenté par la compagnie Sigma Genosys. Cette nouvelle synthèse fut effectuée en deux parties. La première consiste en la génération d'un polynucléotide correspondant à la fin du gène Puro^R, au lien souple et à la banque de peptides. La séquence de cette partie est la suivante : AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNCC CGGATCCGGACCCAGAGGCACCG

La résine sur laquelle est fixée cette partie d'oligonucléotide est alors transférée pour la synthèse du reste de la séquence. Les éléments qui sont ajoutés correspondent à trois codons non-sens dans les trois cadres de lecture possibles et aux sites de restriction *Mfe* I et *Nhe* I. Le polynucléotide est complété par une molécule de biotine. La biotine est un ajout au concept original de l'amorce et pourra servir à purifier sur billes recouvertes de streptavidine les ADN qui auront intégré la banque de peptides. La séquence de la partie ajoutée est celle-ci : 5'-Biotine-TTGGACGCTAGCCAATTGTTAATTAATTA

Le polynucléotide complet (82 nucléotides) correspond, en lecture antisens 5'-3', à l'amorce suivante : Biotine-TTGGACGCTAGCCAATTGTTAATTAATTA AVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNCCCGGATCCGGACCCAGAGGCACCG

Le spectre de masse de ce polynucléotide combiné (Figure 71, page 100) montre clairement une bande enrichie à 82 nucléotides, ce qui correspond à une masse de 25 000 daltons. Cependant, il est clairement visible que la majorité des amorces sont encore tronquées, et cela, malgré la technique de synthèse séquentielle. Ce polynucléotide n'a pas été purifié, ce qui aurait pu enrichir davantage la proportion d'amorces de pleine longueur.

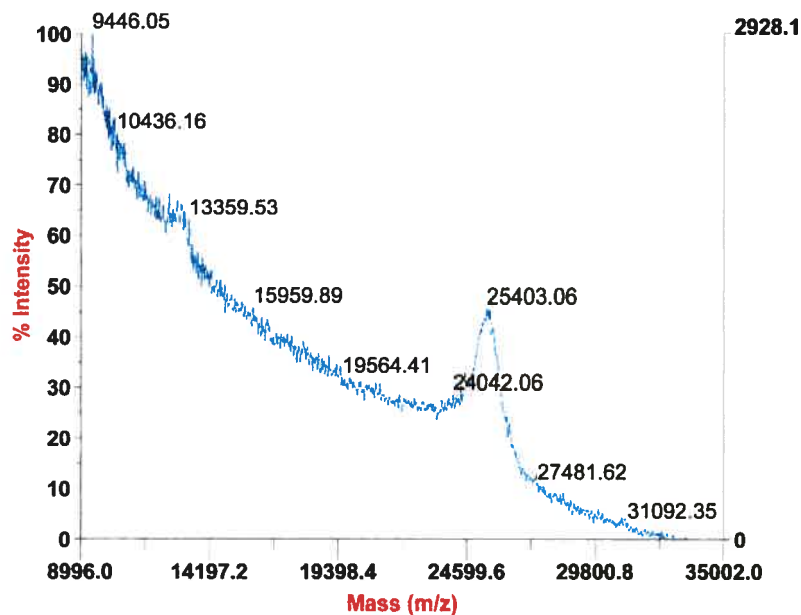


Figure 71. Spectre de masse du polynucléotide combiné encodant la banque de nonapeptides. Ce spectre de masse comprend un pic majoritaire allant de 9 000 à 20 000 Daltons, correspondant à un polynucléotide de moins de 70 nucléotides. L'amorce contenant la séquence de peptides aléatoires de 82 nucléotides apparaît à 25 000 Daltons. (Analyse effectuée par Sigma Genosys).

Le séquençage de l'ADN amplifié avec ce polynucléotide (Figure 72) montre un enrichissement important de la proportion d'inserts contenant la banque de peptides dégénérés. De plus, une fraction des ADN amplifiés semblent posséder les sites de restriction requis pour le criblage des clones positifs ainsi que les sites nécessaires à la ligature du vecteur. Cependant, le spectre de masse et le séquençage du polynucléotide montrent clairement que même avec cette nouvelle technique de synthèse, une fraction majoritaire de polynucléotides sont tronqués et ne pourront être ligaturés.

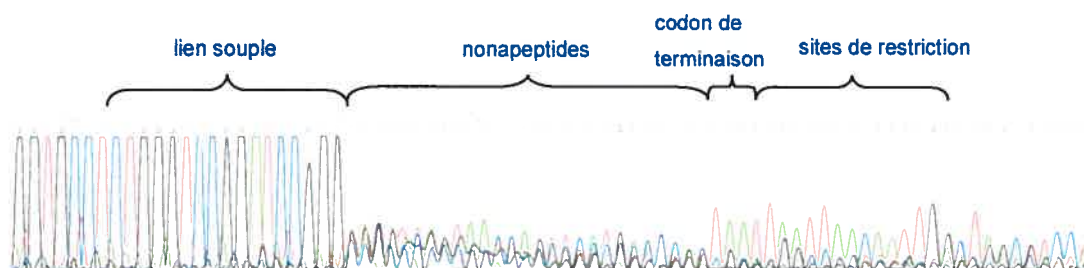


Figure 72. Analyse de la séquence de l'ADN amplifié avec le polynucléotide combiné encodant la banque de nonapeptides.

7.2 Clonage et purification de l'ADN avec les billes de streptavidine

La biotine est ajoutée à l'extrémité 5' du polynucléotide encodant la banque de peptides à la toute fin de la synthèse de l'amorce. Cette particularité permet d'isoler, grâce à l'affinité de la biotine pour la streptavidine, uniquement les vecteurs rétroviraux contenant la biotine, donc présumément la banque de peptides complète. Dans ce dessein, nous avons utilisé des billes magnétiques recouvertes de streptavidine. La digestion par l'enzyme *Mfe* I permet de libérer le vecteur rétroviral de la bille magnétique et ainsi de récupérer le vecteur rétroviral. Cependant, nous avons constaté que cette technique, si elle nous permet de purifier les vecteurs complets, élimine du même coup plus de 90 % du produit d'amplification. Cette sélection draconienne des vecteurs complets réduit d'autant la diversité de la banque de nonapeptides (Figure 73). Cette forte proportion de vecteurs ne se liant pas aux billes de streptavidine concorde avec le spectre de masse de ce polynucléotide combiné présenté plus haut à la figure 71.

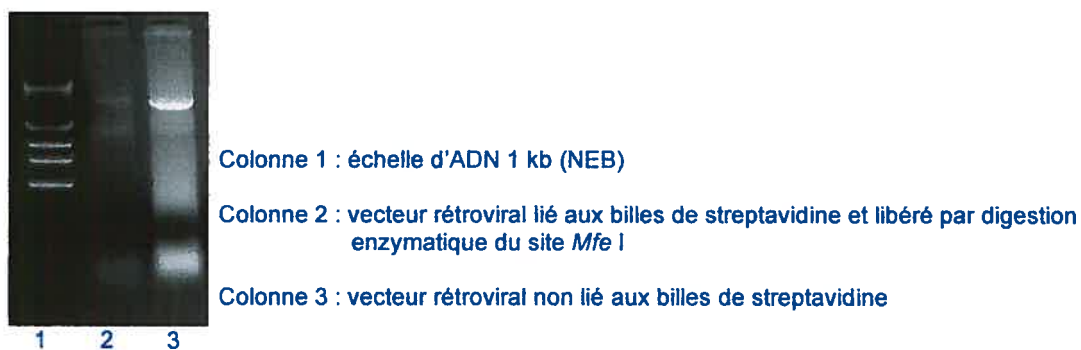


Figure 73. Purification des ADN sur billes de streptavidine.

Les résultats combinés du spectre de masse et de la purification sur billes de streptavidine montrent que la majorité des polynucléotides sont incomplets. Malgré la synthèse séquentielle, la longue séquence de nucléotides dégénérés ne permet pas l'obtention d'un produit final qui assurera la conception d'une banque de nonapeptides de forte diversité, dans la forme actuelle du vecteur rétroviral pBABE P_{PGK} HA/Puro^R/lien souple/nonapeptide.

L'étape subséquente à un clonage réussi serait la production de virions rétroviraux qui serviront à la transduction efficace des cellules cibles. Ces virions, contenant le vecteur rétroviral encapsidé, peuvent être conservés congelés, ce qui constitue une réserve utile représentant l'ensemble de la diversité de la banque de peptides. La production de ces

virions doit être suffisamment efficace pour ne pas perdre davantage de diversité de la banque de peptides et doit permettre le transfert effectif de l'ADN. Nous avons évalué quelle technique de production virale permettrait d'obtenir un rendement suffisant en terme de titre viral, pour la poursuite du projet, advenant que le clonage réussisse. Les résultats des expériences de production virale sont présentés à la section suivante.

8. PRODUCTION VIRALE ET INFECTION

8.1 Production de virus pseudotypés VSV-G et infection de cellules de mammifères

La production de particules virales a été faite dans les cellules HEK 293T. Les virions sont issus de la triple transfection du vecteur rétroviral additionné des plasmides encodant les protéines structurales et l'enveloppe VSV-G. Les productions virales ont été réalisées avec le vecteur rétroviral pBABE P_{PGK} eYFP (Figure 74).

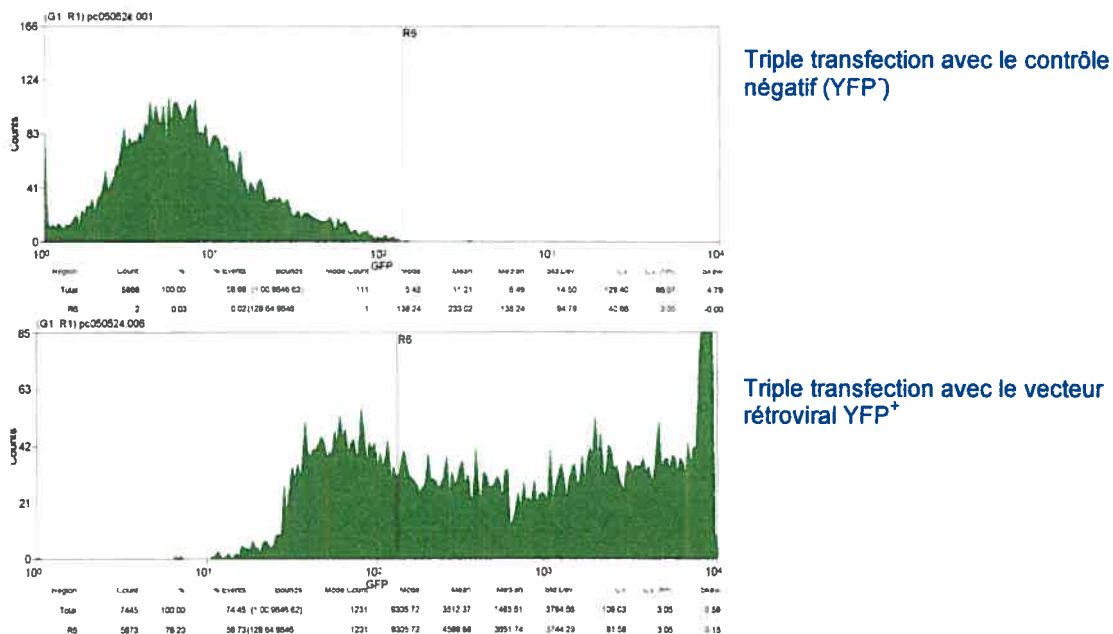


Figure 74. Production virale par triple transfection. Les plasmides pBABE P_{PGK} eYFP, pCIG3B (*gag/pol*) et pSVCMV-VSVG ont été transfectés dans les cellules HEK 293T. L'efficacité des transfections est évaluée par FACS.

La méthode présentée pour la production virale permet d'obtenir des titres viraux de l'ordre de 3×10^5 particules infectieuses par ml (Tableau III, page 103). Cette quantité de particules virales est suffisante pour infecter un million de cellules HEK 293 avec une MI résultant en 20 % de cellules infectées. Dans le but d'obtenir l'intégration d'un seul vecteur rétroviral par cellule, il est nécessaire d'infecter les cellules cibles à une MI aussi basse.

	Dilution	Nombre de clones obtenus à cette dilution	Titre de la production virale
pBABE-puro (Témoin positif)	10^{-4}	2	1×10^5 particules infectieuses/ml
pBABE P _{PGK} HA/Pu ro ^R /lien souple	10^{-4}	6	3×10^5 particules infectieuses/ml

Tableau III. Production de particules virales pseudotypées VSV-G pour l'établissement de lignées stables en cellules de mammifères. Ici, 200 μ l de dilution de surnageant de cellules productrices ont été utilisés pour l'infection de cellules HEK 293. Comme il a été mentionné au point 2.2 de la section Matériel et méthodes, un puits de cellules non infectées est gardé comme contrôle de l'efficacité de sélection sous antibiotique. Seuls les puits de la plus petite dilution contenant des colonies résistantes sont présentés ici. Ce titre viral est une valeur approximative.

Discussion

Les dernières années ont vu la réapparition d'agents infectieux que l'on croyait disparus, de nouveaux types de virus et de virus résistant à la panoplie de molécules antivirales. Les connaissances de base acquises sur ces organismes pathogènes ne suffisent pas à expliquer la dynamique complexe du cycle infectieux et de la spécificité d'hôte. Les premières études réalisées ont porté sur la caractérisation macromoléculaire des virus et sur le cycle viral dans son ensemble. Maintenant, la recherche est rendue à un point où les interactions fines qui se sont établies entre les organismes infectieux et leurs hôtes doivent être analysées. Nous avons vu dans l'introduction quelques-une des découvertes récentes du point de vue des interactions protéine-protéine entre le virus et la cellule infectée (TRIM5 α , APOBEC, etc.) Ces quelques exemples ne représentent qu'une infime partie de ces relations, mais ils montrent à quel point l'infection et la réplication virales sont dépendantes des protéines cellulaires.

Plusieurs groupes exploitent les outils traditionnels d'études protéiques afin de caractériser les innombrables interactions existant entre un agent infectieux et son hôte. Dans l'optique d'apporter notre contribution à cette ère de découvertes protéomiques, nous avons tenté, depuis quelques années, de mettre au point un système original pour l'étude de ces interactions protéine-protéine. Le choix de développer un nouvel outil plutôt que d'utiliser les méthodes classiques se base sur le fait que celles-ci présentent certaines lacunes lorsqu'il s'agit d'étudier les interactions protéiques dans le cadre d'une infection virale. Les expériences qui ont été faites dans le but d'identifier et de caractériser les facteurs de restriction cellulaire, dont TRIM5 α , ont démontré que plusieurs des protéines cellulaires interagissant avec les protéines virales sont exprimées à de faibles taux. De plus, il a été établi qu'elles interagissent avec la forme mature des protéines et que ces interactions ont lieu dans certains compartiments cellulaires déterminés. Nous croyons que la création d'un outil permettant l'étude *in vivo* des interactions protéine-protéine dans un système où leurs propriétés physico-chimiques, leur situation cellulaire et leur niveau d'expression correspondent à une réalité physiologique, permettrait de contourner les limitations retrouvées dans les modèles d'études protéomiques classiques. Dans cet objectif, nous avons conceptualisé un système de sélection génétique qui permet l'établissement de cellules de mammifères exprimant une banque de nonapeptides. Lors

d'une infection virale, ces peptides aléatoires pourront, par leur conformation ou leur séquence, interagir avec les protéines cellulaires ou virales et altérer le cours de l'infection. La caractérisation de ces peptides et des protéines avec lesquelles ils ont interagi permet de cartographier les protéines des voies métaboliques impliquées dans l'infection virale. Cette section résume le concept général du vecteur rétroviral que nous avons conçu et rapporte les étapes qui ont mené à sa synthèse. Nous y discuterons aussi des considérations pratiques et les difficultés qui ont jalonné la conception de ce vecteur.

1. CONCEPTION ET GÉNÉRATION DU VECTEUR RÉTROVIRAL

Nous avons choisi d'utiliser une variante du vecteur rétroviral pBABE comme construction de départ pour produire un vecteur rétroviral répondant à nos exigences. Le vecteur pBABE provient de la séquence génomique du virus de la leucémie murine de Moloney (MoMuLV), sans les gènes structuraux. Ce vecteur rétroviral est constitué de divers éléments. Premièrement, nous y retrouvons les deux longues séquences terminales répétées (LTR) et une partie du gène gag (la conservation de cette séquence favorise l'encapsidation du vecteur rétroviral). Nous y retrouvons ensuite un promoteur SV40 qui permet l'expression du gène qui est placé en aval de celui-ci. Dans notre construction, le gène qui est exprimé est constitué d'une longue protéine de fusion constituée d'un épitope HA, du gène de résistance à la puromycine, d'un lien souple et de la banque de nonapeptides. Pour notre outil de sélection génétique, nous voulions que la banque de peptides soit sous le contrôle d'un promoteur fort et stable. Il a donc été décidé que le promoteur SV40 du vecteur rétroviral soit remplacé par le promoteur stable de la phosphoglycérate kinase (PGK). Ce promoteur permet une expression ininterrompue de la protéine qu'il contrôle. Comme l'ont montré les résultats des cellules transfectées mises sous sélection antibiotique, le promoteur PGK a permis d'obtenir une expression très forte du gène de résistance à l'antibiotique. Le reste du vecteur rétroviral est composé d'une séquence polyA ainsi que d'une séquence LoxP. Cette dernière caractéristique permet de répliquer la région LoxP lors de la transcription inverse et permet, en présence de la Cre recombinase, d'extraire le peptide intégré au génome cellulaire. Le reste du plasmide est composé d'une origine de répllication bactérienne et du gène de résistance à l'ampicilline. Ce dernier permet de sélectionner les clones bactériens issus de la transformation.

Les éléments du vecteur qui sont particuliers à un système d'étude d'interactions protéine-protéine sont la banque de peptides fusionnés à un épitope et le gène de résistance à l'antibiotique. Ce dernier est essentiel dans la perspective de pouvoir isoler les clones cellulaires ayant intégré et exprimant la banque de peptides. Ce gène de résistance doit absolument être exprimé concomitamment avec les peptides. À cette fin, la plupart des laboratoires se tournent vers l'utilisation d'un site d'entrée interne du ribosome (IRES)^{65,66} (Figure 75). Cependant, nous avons rejeté cette possibilité en fonction des résultats obtenus lors d'études précédentes qui ont démontré que l'utilisation de ces IRES a pour résultats, dans des proportions variables, une expression relativement indépendante des deux protéines⁶⁰. Une telle conformation, si elle permet de sélectionner les cellules exprimant l'agent de sélection, ne garantit pas que le peptide y soit exprimé. Inversement, les cellules qui expriment les peptides, mais pas l'agent de sélection seront éliminées lors de la sélection. L'expression réduite ou la non-expression du gène de sélection en aval de l'IRES risquerait d'entraîner la perte de la diversité de la banque, élément critique de notre système de criblage.

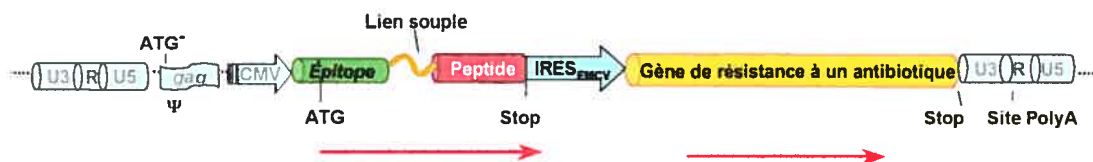


Figure 75. Configuration d'un vecteur rétroviral comportant un IRES. Celui-ci permet l'expression de la protéine de résistance à un antibiotique. La traduction de la protéine en aval de l'IRES est concomitante de l'expression du peptide. Dans cette configuration, les peptides exprimés sont fusionnés en C terminal à une protéine inerte telle que GST qui sert à la précipitation des protéines partenaires.

Comme mentionné précédemment, le modèle de conception que nous avons retenu repose quant à lui sur un gène encodant une longue protéine de fusion comprenant les éléments mentionnés plus haut (c'est-à-dire l'épitope HA, un gène de résistance à un antibiotique, un lien souple constitué d'une série de trois sérines-glycines et la banque de nonapeptides). La décision de générer une telle protéine chimérique repose sur l'hypothèse que dans cette configuration, l'expression de la protéine de résistance à l'antibiotique est garante de l'expression du peptide. Pour ce projet, nous avons testé la fonctionnalité de la protéine de résistance à la puromycine (Puro^R ou *Pac*) et de la protéine

de résistance à la zéocine (Zéo^R ou *Sh Ble*). Nos résultats indiquent que seule la protéine de fusion conférant la résistance à la puromycine conserve pleinement sa fonction en présence de l'antibiotique. La fonction de la protéine de résistance à la zéocine est influencée par la présence de protéines qui lui sont fusionnées en C terminal. De plus, la sélection à la puromycine a l'avantage d'agir en peu de temps, ce qui facilite la sélection des clones. Nous avons noté que la sélection à la zéocine est particulièrement difficile. Alors que les cellules traitées à la puromycine meurent et disparaissent, les cellules traitées avec la zéocine s'arrondissent et flottent dans le milieu. Dans ces conditions, il est plus compliqué de compter ou d'isoler les clones survivants avec toutes ces cellules en suspension.

Le désavantage de la fusion d'une protéine à la banque de peptides est qu'elle entraîne la présence d'une molécule relativement grosse à proximité du nonapeptide. En effet, la puromycine-N-acétyl-transférase comporte 200 acides aminés. Cette conformation pourrait générer un encombrement stérique et empêcher les peptides d'interagir avec leurs partenaires protéiques. Le lien souple constitué d'une série alternée de trois glycines et de trois sérines, introduit entre la protéine de résistance à l'antibiotique et le peptide, devrait conférer une certaine flexibilité et une indépendance au peptide. Le vecteur rétroviral que nous avons conçu est illustré à la figure 76.

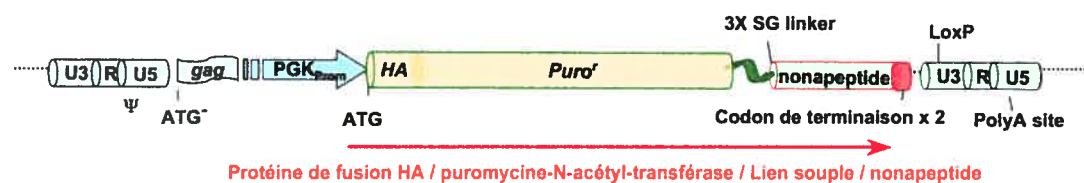


Figure 76. Représentation du vecteur rétroviral généré : pBabe P_{PGK} HA/Puro^R/lien souple/nonapeptide. À ce vecteur rétroviral s'ajoutent l'origine de répllication bactérienne et le gène de résistance à l'ampicilline du plasmide pBluescript (les divers constituants représentés ne sont pas à l'échelle).

Un élément important du vecteur rétroviral que nous avons généré est l'épitope qui est fusionné à la banque de peptides. Cet élément a pour fonction de permettre l'immunoprécipitation des protéines ayant interagi avec les peptides. Cet épitope devait correspondre à une molécule inerte et qui ne se retrouve pas dans des protéines cellulaires. Aussi, nous avons utilisé l'épitope HA, un peptide de neuf acides aminés

(YPYDVPDYA) de l'hémagglutinine du virus de l'influenza. Cet épitope est un peptide ayant une distribution ubiquitaire dans la cellule. Son faible poids moléculaire (1,1 kDa) réduit les chances que l'épitope nuise aux interactions peptide-protéine. L'utilisation d'un anticorps commercial anti-HA permet d'immunoprécipiter le complexe formé de l'épitope HA, du peptide fusionné et des protéines interagissant avec ce peptide (Figure 77). Les nonapeptides sont placés en C terminal de la protéine de fusion. Cette configuration élimine la possibilité d'une terminaison prématurée de la protéine de fusion. En effet, la banque étant composée de codons dégénérés et aléatoires, certains seront des codons de terminaison. Si les protéines HA et Puro^R étaient placées en C terminal de la banque, plusieurs des protéines exprimées s'arrêteraient au peptide et ne contiendraient pas les protéines situées en aval. Cette situation a été rapportée dans la littérature. Lorsque la banque de peptides fut placée en N terminal d'une protéine de fusion comprenant le peptide et la protéine CCR5, 34 % des protéines de fusion étaient tronquées et ne contenaient pas la molécule CCR5⁵².

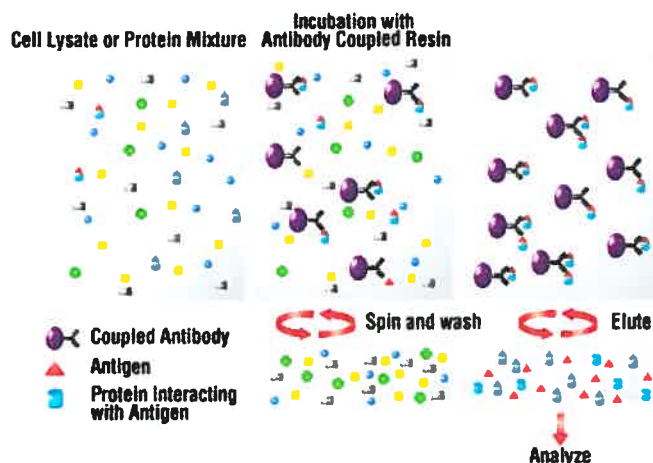


Figure 77. Schématisation d'une immunoprécipitation. Les HA-peptides de la banque de peptides, ici représentés par les antigènes, interagissent avec des partenaires protéiques cellulaires. L'incubation de lysats cellulaires avec des anticorps anti-HA permet d'immunoprécipiter et subséquemment de caractériser le peptide et les partenaires protéiques. (Illustration tirée de réf. 67)

2. CARACTÉRISTIQUES DES PEPTIDES

La qualité d'une banque de peptides n'est pas caractérisée par sa diversité, mais par sa capacité d'altérer un phénotype étudié. Une grande diversité, cependant, augmente les

chances qu'une banque de peptides bien conceptualisée interagisse avec un plus grand nombre de protéines impliquées dans les voies métaboliques menant à ce phénotype. De plus, une grande diversité permettra de compenser les pertes causées par les peptides qui seront toxiques pour la cellule ou qui posséderont une séquence qui provoquera leur excrétion. Autant l'aspect fonctionnel que la diversité de la banque sont dépendants de la conception des peptides. La longueur du peptide, le choix des codons encodant la banque aléatoire et la configuration statique ou libre du peptide influenceront sa capacité d'interagir avec les protéines. Nous avons choisi de créer une banque de peptides libres^m d'une longueur de neuf acides aminés encodés par une séquence (NNK)₉. Comme il a été dit dans l'introduction, cette configuration NNK permet d'éliminer deux des trois codons de terminaison. Cependant, la synthèse du peptide par l'utilisation d'un mélange équimolaire des quatre nucléotides induit un biais dans la séquence. Ce biais est causé par les diverses affinités des codons pour la matrice de synthèse et par le fait que les acides aminés sont encodés par un à six codons différents. Une étude a fait l'analyse exhaustive de ce biais et propose une solution à ce problème par la combinaison de codons semi-aléatoires dont la fréquence est calculée sur la base de calculs statistiques⁶⁸. Un laboratoire qui voudrait, comme nous, créer une banque de peptides aura tout avantage à tenir compte de ces informations dans la conceptualisation des polynucléotides encodant les peptides. Un autre élément qui influencera la diversité de la banque est la longueur du peptide. Alors qu'une banque d'ADNc peut atteindre 10⁶ molécules différentes, une banque de nonapeptides a une diversité potentielle de plus de 10¹¹. Plus longs seront les peptides, plus grande sera la diversité théorique de la banque. Cependant, les restrictions liées aux techniques de clonage limitent la diversité réelle de la banque qui atteint habituellement une diversité de 10⁶ à 10⁸. Cette limitation est principalement due à l'étape de transformation.

3. PURIFICATION DES AMORCES

La question de la synthèse et de la purification des polynucléotides qui encodent la banque de peptides est le point central de cette discussion. Cet aspect a un impact majeur

^m La banque de peptides est fusionnée à son extrémité N-terminale, mais elle n'est pas contrainte à une forme statique comme le sont les banques présentées sous forme de boucles.

sur la réussite du clonage et a eu des conséquences importantes dans ce projet de recherche. Il est généralement reconnu que la synthèse de longs polynucléotides s'accompagne d'une quantité croissante de produits de synthèse arrêtés prématurément. Ceci est dû au fait que l'addition des nucléotides n'est pas efficace à 100 %. Le tableau IV montre le pourcentage d'oligonucléotides avortés pour différentes longueurs de synthèse.

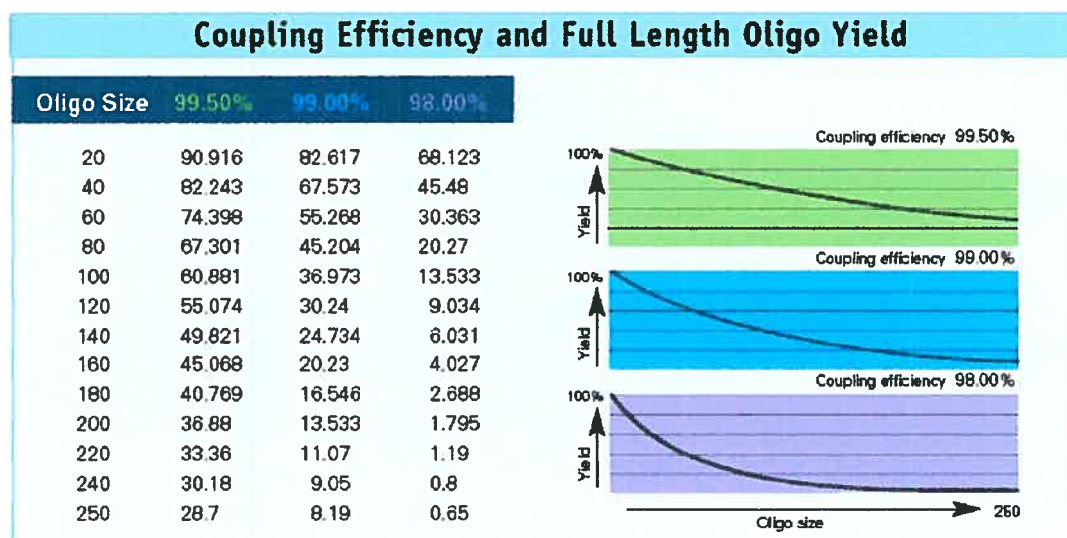


Tableau IV. Pourcentage d'oligonucléotides complets en fonction de la longueur de ceux-ci et du pourcentage d'efficacité de la réaction de synthèse. (Illustration tirée de la réf. Erreur ! Signet non défini.)

L'oligonucléotide que nous avons utilisé devait comporter 82 bases. Nous voyons qu'à cette longueur, avec une efficacité de synthèse de 98 %, seulement 20 % des oligonucléotides seront complets. Il semblerait que les difficultés de synthèse sont plus présentes lorsque la séquence est dégénérée. En effet, le spectre de masse du polynucléotide encodant notre banque de nonapeptides montre une proportion d'oligonucléotides complets bien en deçà de 20 % du total (Figure 66, page 96). Une solution recommandée par l'ensemble des compagnies qui font la synthèse des oligonucléotides est de purifier ces amorces sur gel de polyacrylamide. En effet, nous avons vu que les meilleurs résultats de ligature ont été obtenus avec la purification PAGE des polynucléotides servant à introduire la banque de nonapeptides. Après discussion avec les techniciens de Sigma Genosys, nous en avons conclu qu'une telle purification risque cependant d'entraîner un biais important de la diversité de la banque de peptides. En effet, la nature dégénérée de la séquence encodant le peptide fait que le

polynucléotide apparaît sur le gel sous forme d'une bande floue. Si nous ne purifions que la seule partie correspondant à un ADN standard de même longueur, ceci ne sélectionnera qu'une infime partie de la banque. Pour la plupart des expériences qui ont été décrites dans la section « Résultats », nous avons fait le choix d'utiliser seulement les amorces non purifiées. Nous espérons ainsi conserver le maximum de diversité. Les amorces tronquées ne possédant pas de site de restriction, les vecteurs ne devaient pas se ligaturer. Nous pouvions alors profiter du fait que la transformation des ADN linéaires est peu efficace par rapport à des ADN ligaturés pour amplifier préférentiellement les vecteurs contenant l'ADN encodant des peptides complets. Nous ne nous attendions pas à ce qu'aussi peu de polynucléotides soient de pleine longueur. Si peu, en fait, que les bactéries compétentes ont été saturées de vecteurs linéaires. Ceci a permis d'obtenir très peu de clones. Un nombre insuffisant pour une banque de peptides de bonne qualité.

Devant ce constat, nous avons fait purifier les polynucléotides par PAGE, quitte à perdre une partie de la diversité de la banque, une diminution de diversité qu'il est impossible de chiffrer. Les résultats obtenus avec ces amorces purifiées semblaient encourageants. Nous avons ainsi généré une banque de plus de 6×10^7 clones. Le séquençage des clones obtenus sur pétris directement après la transformation montre une certaine diversité des ADN encodant les peptides, malgré un changement dans le cadre de lecture du codon de terminaison et un site de restriction modifié. La contre-vérification des clones constituant l'ensemble de la banque congelée dans le glycérol a cependant révélé une diversité presque nulle. Comment expliquer cette absence de diversité de peptides pourtant encodés par des polynucléotides aléatoires ? La réponse à cette question est composée d'un ensemble de facteurs. Premièrement, la méthode utilisée ici était l'amplification de tout le vecteur rétroviral à l'aide des amorces dégénérées. À titre d'essais, nous avons procédé à deux amplifications du vecteur. L'une a été faite avec l'amorce encodant le peptide dégénéré et l'autre avec une amorce sans le peptide dégénéré. Cette dernière amplification a généré près de cent fois plus de vecteurs que celle faite avec l'amorce dégénérée (données non montrées). Il est évident que la présence d'une très longue séquence de plus de 40 nucléotides qui ne s'hybrident pas à la matrice influence l'efficacité de l'amplification. Une amorce de plus de 80 nucléotides adoptera nécessairement des structures secondaires qui peuvent aller jusqu'à empêcher l'hybridation à la matrice. Il faut donc considérer que seuls les amorces qui possèdent une séquence aléatoire conférant une structure secondaire faible ou dont la séquence

encodant le peptide est absente seront amplifiés préférentiellement. Un second facteur qui peut expliquer la perte de diversité est la purification par PAGE des amorces utilisées dans cette expérience. Les analyses par spectre de masse et sur gel PAGE des polynucléotides non purifiés (Figures 66 et 67, pages 96 et 97) montrent clairement qu'une infime partie des amorces sont complètes. En fait, elles sont tellement peu nombreuses que nous ne pouvons pas les distinguer du bruit de fond. La purification de cette fraction de polynucléotides rend peu probable l'obtention d'une forte diversité clonale.

Comment alors obtenir des polynuclotides complets tout en conservant la diversité de la banque de peptides ? Nous avons vu que de meilleurs résultats de synthèse furent obtenus lorsque le polynucléotide fut synthétisé séquentiellement. Une première partie de l'oligonucléotide est synthétisé sur une colonne, puis la matrice de synthèse est transférée dans une autre colonne pour la poursuite de la génération des amorces. Cette méthode de synthèse a permis l'obtention d'un pic visible composé de polynucléotides complets. Cependant, même de cette façon, la proportion de produits tronqués est franchement plus importante. Ces difficultés de synthèse sont reconnues par les techniciens qui ont eu pour mandat de synthétiser nos polynucléotides. Cependant, nous nous sommes interrogé à savoir comment les banques de peptides de la littérature ont été faites. Divers essais de construction de banque de peptides ont été décrits dans la littérature, ceux-ci ayant généralement entre quatre et dix-huit acides aminés, selon l'expérience réalisée. Par exemple, la méthode de construction de banque de peptides par l'utilisation de phages présente la création de banques constituées de quatre et dix acides aminés. Cependant, seule la diversité obtenue avec le quadrapeptide est présentée, les résultats du décapeptide sont passés sous silence⁵⁶. Faut-il y voir une difficulté inhérente à la longueur du peptide ? Quoi qu'il en soit, des polynucléotides de plus de 18 nucléotides ont été déjà générées et nous devons considérer qu'il est possible d'obtenir des amorces complètes encodant des banques de peptides aléatoires.

Quelles sont alors les options qui permettraient la construction de la banque de peptides ? L'analyse MS et celle par gel de polyacrylamide montrent clairement que la synthèse de la séquence dégénérée n'est pas fonctionnelle. À chaque nucléotide dégénéré, une partie importante des amorces ne sont pas allongées. La solution la plus facile consisterait à concevoir un peptide plus petit. En effet, la limite imposée par les méthodes de transformation et de transfection est de moins de 1×10^8 clones. L'obtention de plus de

clones exigerait toutefois une quantité de transformations et de transfections représentant une somme de travail considérable et exigerait des équipements spécialisés. Nous pouvons donc considérer qu'il est superflu d'avoir un peptide codant plus de 10^8 possibilités. Un hexapeptide permettra d'obtenir 6×10^7 clones différents. Cette différence de trois codons représente neuf nucléotides dégénérés de moins; une différence majeure à l'échelle de la synthèse d'un polynucléotide. L'utilisation d'une banque de peptides de six acides aminés plutôt que de neuf permet d'obtenir la même diversité clonale finale. Le choix d'utiliser des peptides plus longs se justifie principalement par la capacité d'un long peptide à adopter des structures tridimensionnelles favorisant certaines interactions avec les protéines. Une seconde voie de réflexion nous amène à considérer la conception de polynucléotides encodant les peptides basée sur l'utilisation de séquences semi-aléatoires telles que décrites dans la référence mentionnée plus haut, au point 2 de la présente section. Une séquence semi-aléatoire induit peut-être moins de difficultés de synthèse que lorsqu'un mélange des quatre nucléotides est utilisé. À cette conception différente du polynucléotide s'ajoute la purification des amorces. En effet, nous avons vu que la production de polynucléotides biotinylés selon la méthode séquentielle permet d'obtenir une quantité substantielle d'ADN de pleine longueur. La purification large de la bande correspondant au polynucléotide complet permettrait d'inverser le ratio des amorces complètes versus les amorces tronquées. Cette méthode ne garantit pas que la purification sur gel n'induisse pas un biais de séquence, mais nous croyons que la combinaison d'un polynucléotide plus petit, d'une meilleure synthèse et d'une purification sur gel qui n'est pas trop stricte permettra de conserver suffisamment de diversité à la banque de peptides.

4. TECHNIQUE DE CONSTRUCTION À PRIVILÉGIER

Outre les difficultés de synthèse des polynucléotides, nous nous sommes interrogé sur la méthode à privilégier pour la genèse d'une banque de peptides. Dans le cadre de ce projet, nous avons expérimenté de nombreuses techniques de clonage, passant de la méthode classique de la ligature insert-vecteur à la création de vecteurs rétroviraux sous forme d'ADNc. Ces diverses méthodes exploratoires nous ont permis de cerner les forces et les faiblesses de chacune et d'en dégager certains concepts intéressants.

Premièrement, nous avons vu que la méthode insert-vecteur possède certains atouts importants, le plus important étant la longueur des polynucléotides qui sont requis. Cette technique ne nécessite pas, dans les amorces, la présence de séquences complémentaires au vecteur puisque l'insert est complété en ADN double brin par un oligonucléotide s'hybridant au site de restriction. Cette caractéristique permet de générer des polynucléotides beaucoup plus petits, donc plus faciles à synthétiser. Ensuite, cette méthode peut être considérée comme la méthode de référence puisqu'elle a permis à d'autres groupes de générer diverses banques de peptides. Les défauts de cette technique résident essentiellement dans la difficulté de purifier les petits inserts et dans la conception du vecteur rétroviral qui doit être adapté pour pouvoir insérer ces inserts sans perturber la conformation de la protéine de fusion. En ce qui nous concerne, nous n'avons pas poussé plus avant l'exploration de cette technique et cette décision a été motivée par différentes raisons. D'abord, la conception de la protéine de fusion rendait difficile l'utilisation de cette méthode. Ensuite, nos résultats ont montré que les sites de restrictions disponibles pour la ligature d'un insert dans notre vecteur rétroviral sont soit trop peu (*Dra* III) ou trop spécifiques (*Cla* I). Finalement, une telle technique fut employée précédemment au sein de notre laboratoire et n'avait pas permis d'obtenir une diversité clonale suffisante pour une banque de peptides.

Une seconde méthode proposée consiste à amplifier des inserts plus longs ou même tout le vecteur rétroviral. Nous avons vu que cette technique induit la possibilité de formation d'une bulle d'ADN. En effet, la séquence de l'ADN encodant la banque de peptides étant de nature aléatoire, l'amplification fait s'hybrider des séquences qui ne sont pas complémentaires. Ceci a pour conséquence la formation d'une séquence d'ADN non apparié correspondant au site du peptide. Cet aspect particulier causé par la nature dégénérée de l'ADN encodant le peptide ainsi que les mécanismes de réparation bactériens impliqués ont été décrits au point 2.2.3 de l'introduction (page 19). Rappelons simplement que cette bulle devrait être réparée lorsque le plasmide est transformé dans les bactéries. Étant donné la longue séquence d'ADN mésapparié (27 nucléotides), il est probable que ce soient les systèmes de réparation bactériens connus sous le nom de MMR qui élimineront la bulle d'ADN. Normalement, ce mécanisme de réparation utilise le brin méthylé comme matrice lors de la synthèse du brin complémentaire. Dans ce cas-ci, le brin méthylé est le brin du vecteur parental, donc ne possédant pas la séquence encodant le peptide. Sans avoir la preuve que de tels mécanismes sont impliqués, nous

considérons la possibilité trop importante pour utiliser ces méthodes dans le but de construire la banque de peptides. Ce phénomène ne se produit pas dans la technique précédente qui consistait à introduire l'insert de polynucléotides double brin, car une seule étape d'extension est employée pour compléter le second brin de l'amorce.

Finalement, une dernière méthode consiste en la production d'un vecteur rétroviral sous forme d'ADNc. Cette technique possède des avantages importants. Premièrement, un seul site de restriction est impliqué. Ceci élimine la possibilité de la ligature de multiples inserts comme cela fut observé dans nos expériences impliquant une ligature de type insert-vecteur. De plus, cette technique évite la formation de bulles d'ADN telles que décrites précédemment. Ensuite, la purification du vecteur est fort simple puisqu'elle implique un ADN de bonne dimension (5,4 kb). Cependant, malgré l'aspect attrayant de cette méthode, certaines données obtenues dans nos expériences nous font douter de la validité de cette technique. Nous avons vu que l'extension ou l'amplification différentielle de tout le vecteur est une opération qui comporte de multiples étapes d'optimisation dans le but d'obtenir suffisamment d'ADN amplifié pour les étapes subséquentes. De plus, nous avons vu au point 3 de la présente section que l'amorce encodant le peptide dégénéré s'hybride moins bien que l'amorce complémentaire ou qu'une amorce tronquée. Cette caractéristique risque d'introduire un biais de séquence important dans la banque.

La méthode la plus sûre est donc la méthode de l'insert-vecteur. Comme nous l'avons mentionné précédemment, le problème principal rencontré avec cette méthode est la purification des petits inserts. Une solution originale à ce problème consiste en l'ajout d'une biotine à l'extrémité 5' du polynucléotide encodant les peptides. Après la synthèse du second brin du polynucléotide, les amorces sont ajoutées à des billes magnétiques recouvertes de streptavidine. L'extrémité 3' des polynucléotides peut alors être coupée à l'aide d'enzymes de restriction. L'extrémité coupée est éliminée facilement par un lavage, les polynucléotides étant retenus aux billes magnétiques, elles-mêmes retenues par un aimant. La digestion du site en 5' peut alors être faite et les polynucléotides récupérés dans le surnageant. L'extrémité coupée en 5' reste collée sur les billes magnétiques et ne pourra contaminer les inserts. De cette façon, aucun gel d'agarose ou de polyacrylamide n'est requis pour la purification de l'insert. Il suffit alors d'intégrer l'insert au vecteur linéarisé. Aucun changement de tampon n'est même requis puisque la ligature peut être effectuée dans le tampon utilisé pour couper le second site de restriction. En plus de

résoudre tous les problèmes liés à la purification des petits inserts, cette méthode comporte de nombreux autres avantages. En premier lieu, les amorces tronquées ne possèdent pas la biotine puisqu'elle est ajoutée à la toute fin de la synthèse des oligonucléotides. Lors de la fixation des inserts aux billes magnétiques, les amorces incomplètes et les amorces complémentaires libres seront éliminées. Ceci permet une purification supplémentaire des polynucléotides. Ensuite, la purification sur billes magnétiques se fait dans des conditions très douces comparées à une purification phénol/chloroforme. Finalement, diverses options sont possibles. Par exemple, la ligature du vecteur linéarisé à l'extrémité 3' du polynucléotide peut être effectuée alors que l'extrémité 5' est encore fixée aux billes magnétiques. Un lavage permet alors d'éliminer les vecteurs qui ne se seront pas ligaturés ou ceux qui se seront autoligaturés. De plus, ceci empêche toute insertion multiple du polynucléotide. Il suffit alors de couper l'extrémité 5' et de lier le vecteur se trouvant maintenant dans le surnageant. Cette méthode innovatrice, combinée aux modifications qui doivent être apportées aux polynucléotides comme nous l'avons vu aux points 2 et 3 de la présente section, pourrait être la clé de la réalisation de la banque de peptides et représente une nouvelle approche résolvant les problèmes rencontrés dans les techniques traditionnelles de clonage.

5. LIGATURE ET EFFICACITÉ DES TRANSFORMATIONS

Les bactéries électrocompétentes les plus aptes à recevoir un vecteur rétroviral ont dans les meilleures conditions une efficacité de transformation théorique de 10^{10} colonies/ug d'ADN transformé. Ce résultat peut être obtenu lorsque le vecteur contrôle pUC19 sous forme superenroulée est utilisé. Les résultats obtenus avec un vecteur coupé puis ligaturé sont plutôt de l'ordre de 10^8 colonies/ug. La transformation de plasmides linéaires est encore moins efficace et permet d'obtenir seulement quelques milliers de clones par transformation. L'ensemble des expériences que nous avons effectuées a révélé que la ligature simple des vecteurs encodant la banque de peptides est très inefficace et génère une majorité de vecteurs sous forme linéaire. Nous avons aussi vu que la piètre qualité des polynucléotides est responsable de ces difficultés de ligature. En effet, nous avons démontré par l'analyse des ADN encodants la banque de peptides que ceux-ci ne possèdent pas les sites de restriction requis pour la ligature. En plus de produire peu de clones, ces produits partiellement ligaturés feront l'objet, dans la bactérie, de

recombinaisons et de délétions. On serait porté à croire qu'une optimisation des conditions de ligature pourrait aider à résoudre le problème, mais il n'en est rien. Nous avons testé divers types de ligases et d'enzymes de restriction, avec ou sans phosphorylation des extrémités, avec des extrémités cohésives ou directes, avec ou sans additifs et à diverses températures (4°, 16° et 22°C). (Ces données ne sont pas rapportées dans cet ouvrage.) Une telle optimisation a généré la ligature forcée d'extrémités non compatibles, ayant pour résultat la ligature de vecteur sans peptide. Nous avons aussi constaté l'introduction de séquences étrangères au vecteur et la ligature de banques à des sites alternatifs, même sans séquences complémentaires. Bref, la ligature dans des conditions normales est à privilégier, car cela permet d'emblée d'éliminer les vecteurs partiellement digérés et les inserts incomplets et évite la formation de vecteurs chimériques impossibles à éliminer lors des étapes subséquentes. Une faible efficacité de transformation sera alors le signe d'un problème associé à l'insert introduit et permettra les rectifications nécessaires.

6. CONSIDÉRATIONS MATHÉMATIQUES

L'obtention d'une ligature efficace à plus de 90 % est l'étape clé d'une transformation réussie. En effet, nous visons une diversité de peptides maximale. Puisque la transformation de plasmides correctement ligaturés permet d'obtenir environ 10^8 clones différents par μg de vecteur, nous sommes en droit de nous attendre à une telle diversité clonale. Les bactéries électrocompétentes étant saturées à environ 20 ng d'ADN/transformation, il faudra effectuer 50 transformations pour atteindre notre objectif de 10^8 clones. Ceci est donc un objectif réaliste. Si une forte proportion de l'ADN transformé n'est pas ligaturé, l'efficacité de transformation sera réduite et rendra les procédures suivantes difficilement réalisables, voire impossibles. En effet, en plus du nombre de clones totaux à obtenir, la diversité finale doit tenir compte du pourcentage de clones qui possèdent une banque fonctionnelle (bon cadre de lecture et présence de codons de terminaison à la fin du nonapeptide). Le résultat du séquençage de quelques-uns des ADN encodant les peptides provenant des méthodes utilisées pour ce projet montre de nombreux peptides tronqués, absents ou sans codon d'arrêt. Si pour une expérience particulière nous obtenons seulement 10 % de clones comportant une séquence correcte, pour le même nombre de transformations nous perdrons un log de diversité.

L'une des raisons pour lesquelles nous visons une diversité clonale de 10^8 peptides est que les étapes subséquentes de transfection, de production virale, d'infection et d'isolement de clones sous sélection génétique impliquent des pertes inévitables. En effet, aucune de ces techniques ne permet d'obtenir un rendement de 100 %. Nous avons vu que la littérature rapporte un ratio moyen d'un peptide interagissant dans une voie métabolique donnée pour un million de peptides testés. La diversité finale de la banque, au moment de l'étude des interactions peptide-protéine, devra donc être d'au moins 10^6 . Il serait hasardeux d'espérer récupérer LE clone positif après une ronde de sélection et nous espérons obtenir une diversité plus importante afin de multiplier d'autant le nombre de clones positifs. L'élimination des faux positifs et des artéfacts d'interaction peut être obtenue par de nombreux cycles de sélection, ce qui nécessite d'obtenir au moins quelques clones positifs pour éviter le risque de les perdre lors de la sélection génétique. À première vue, on pourrait se demander si l'obtention d'une diversité de 10^8 clones est nécessaire. En effet, d'autres groupes ont caractérisé des interactions protéine-protéine avec des banques d'ADNc de diversités bien moindres. Il faut prendre en considération qu'une banque d'ADNc génère un produit protéique ayant été raffiné par des millions d'années d'évolution. La séquence et la forme de ces protéines sont optimales pour sa fonction. Lorsque mises dans un contexte favorable, les protéines issues d'une banque d'ADNc ont toutes les chances d'interagir avec leur cible. Une banque de peptides, en revanche, sera de forme et de séquence aléatoires. La vaste majorité de ceux-ci seront inertes, d'autres seront toxiques pour la cellule ou encore iront dans des compartiments cellulaires hors de toute possibilité d'interaction avec la cible. Une diversité maximale est alors primordiale pour favoriser l'hypothèse qu'un ou quelques peptides interviennent au sein d'un processus métabolique particulier pour en modifier la fonction.

Une piste de réflexion intéressante consiste à se demander si l'étape de transformation est bien nécessaire. En effet, partant d'une diversité théorique de 10^{11} molécules, les limites obligatoires de cette première étape font perdre entre trois et quatre logs à la diversité de la banque. De plus, elle ne sert qu'à amplifier des milliers de fois les mêmes séquences encodant les peptides, ce qui n'augmente pas la diversité de la banque. Cependant, il faut se demander s'il ne serait pas avantageux de transférer directement le produit de ligature dans les cellules afin de créer une banque de la banque sous la forme de stocks viraux plutôt que sous la forme ADN. Nous avons calculé qu'une quantité de 3,5

ug de vecteur rétroviral couvre l'entière diversité théorique de la banque de nonapeptides, soit 10^{11} peptides. L'obtention de cette quantité de vecteurs ligaturés est très facile. Si la confection de vecteurs rétroviraux ligaturés exige un travail de longue haleine, il serait impensable de reprendre ces étapes de clonage chaque fois qu'une banque de peptides doit être créée. Cependant, si l'obtention de polynucléotides courts, biotinylés, purifiés et de bonne qualité permet un clonage insert-vecteur aussi facile que celui décrit précédemment, alors il faudrait envisager la possibilité d'éviter l'étape de la transformation. L'omission de cette étape ne semble pas avoir été tentée par d'autres groupes et elle permettrait, si cela est possible, l'obtention d'une diversité finale inégalée. La vérification que les peptides sont dans une configuration correcte et qu'ils sont de séquence aléatoire peut être faite par la transformation d'une partie de l'ADN ligaturé ou par la dilution limite de cellules infectées suivie de l'amplification et du séquençage de la séquence peptidique introduite.

7. INFECTION DES CELLULES DE MAMMIFÈRES

Les cellules qui serviront à établir une lignée stable exprimant la banque de peptides doivent être infectées à basse multiplicité d'infection (10 à 20 % de cellules infectées) afin d'éviter que plus d'un vecteur par cellule ne soit intégré. À une MI résultant en 20 % de cellules infectées, il faudra infecter environ 100 pétris de 5×10^6 cellules/pétri afin de couvrir l'intégralité d'une banque de 10^8 peptides. L'utilisation de chambres de culture à paliers multiples permet d'infecter un nombre suffisant de cellules pour couvrir toute la diversité de la banque. Les résultats que nous avons obtenu lors de l'infection de cellules de HEK 293 avec des stocks viraux encodant les vecteurs rétroviraux contrôles pBABE P_{PGK} eYFP nous permettent de croire que les conditions de production virale sont adéquates. Il est possible d'utiliser directement le surnageant contenant les virions pour infecter les cellules de mammifères, sans devoir passer par des étapes de purification. La réussite de cette étape est cependant conditionnelle à l'obtention d'un stock viral encodant exclusivement la banque de peptides. En effet, la présence de vecteurs parentaux ou de banques déficientes ne pourra être éliminée. Le vecteur parental et les banques déficientes comprennent tous les deux le gène de résistance à la puromycine. L'infection des cellules de mammifères avec le stock viral et la sélection subséquente à la puromycine conservera autant les clones cellulaires exprimant la banque de peptides que

les clones sans peptide. La présence de nombreux clones cellulaires n'exprimant pas la banque de peptides rendrait irréalisable les études d'interactions protéine-protéine. Cela met l'accent sur la nécessité d'obtenir un produit de ligature qui soit exempt de vecteurs chimériques ou parentaux. Dans le but d'avoir la possibilité d'éliminer tous les vecteurs parentaux contaminateurs, il serait sage de considérer l'inclusion dans le vecteur parental d'un gène toxique pour la cellule. Lors du clonage par la méthode insert-vecteur, la ligature de l'insert de la banque de peptides devra remplacer ce gène toxique. Cette méthode simple a deux avantages. Elle permet de sélectionner seulement les clones cellulaires qui possèdent la banque de peptides et elle introduit une séquence dans le vecteur rétroviral qui permettra d'évaluer, lors du clonage, si celui-ci a bien été digéré par les enzymes de restriction.

Conclusion

Dans le cadre de ce projet de recherche, nous visions à développer un outil de recherche protéomique qui permettrait l'étude d'interactions protéine-protéine importantes lors d'une infection virale. Certaines caractéristiques du système que nous avons conçu vont au-delà des possibilités rencontrées dans les méthodes classiques. Malheureusement, des limites techniques ne nous ont pas permis d'atteindre les objectifs que nous nous étions fixés. Néanmoins, les essais et erreurs qui ont jalonné les expérimentations que nous avons entreprises ont permis de dégager certains concepts qui permettront, nous l'espérons, la conception d'une banque de peptides basée sur une réflexion mûrie.

Cette banque de peptides, quel que soit le concept retenu, devra être validée. En effet, l'utilité d'une banque de peptides ne se mesure pas sous l'angle de diversité clonale, mais en fonction de la capacité d'interférer dans une voie métabolique. Cette fonctionnalité peut être démontrée par le criblage d'un phénotype connu. Premièrement, il faut préciser que les applications possibles d'une banque de peptides exprimés en cellules de mammifères sont nombreuses, mais sont quand même limitées par certaines contraintes. La caractérisation des peptides ayant mené à l'altération d'une voie métabolique implique nécessairement que toutes les cellules seront éliminées lors de la sélection génétique sauf celles où l'interaction a eu lieu. Dans le but de valider la méthode, la voie métabolique connue que nous avons choisie est la phosphorylation du ganciclovir par la thymidine kinase (TK). Nous savons que la phosphorylation du ganciclovir rend cette molécule toxique pour la cellule. Une banque de peptides qui interfèrerait dans cette étape de phosphorylation et qui permettrait la survie de la cellule pourrait être considérée comme fonctionnelle. En effet, elle aura permis d'interférer avec des protéines impliquées dans une voie métabolique ciblée, ce qui est la propriété essentielle d'un système d'étude d'interactions protéine-protéine. Les étapes de cette méthode de validation consistent à générer des cellules exprimant la banque de peptides. Ces cellules seront ensuite infectées avec un virus rapporteur encodant le gène de la thymidine kinase (TK). Les cellules qui exprimeront la thymidine kinase, en présence de ganciclovir, mourront. Les cellules qui expriment un peptide qui interagit dans le processus de phosphorylation du ganciclovir ou dans le processus de mort cellulaire survivront. Les cellules qui n'auront pas été infectées par le virus ou dans lesquelles le peptide aura altéré l'une des étapes

précoces menant à l'expression des protéines encodées dans le vecteur rétroviral (ici la TK) survivront aussi. Une série de criblages permettent d'éliminer facilement les cellules qui auront survécu simplement parce qu'elles n'ont pas été infectées. La caractérisation subséquente des peptides et des protéines impliquées dans les interactions peptide-protéine des cellules survivantes permettra alors de situer dans quelle voie métabolique se trouvent les interactions qui auront permis la survie de la cellule. Il est alors possible de discriminer les interactions que nous croyons reliées à la phosphorylation du ganciclovir, des interactions qui sont reliées aux étapes précoces de l'infection. Si l'on trouve par la recherche de séquences homologues ou par l'immunoprécipitation, des peptides reliés à la thymidine kinase, nous aurons validé la méthode. Ceci aura prouvé que cet outil de recherche permet d'interférer avec une protéine cible, *in vivo*, et que les peptides et protéines impliquées dans ces interactions peuvent être caractérisés.

Les peptides qui ont été trouvés lors de la validation de méthode peuvent aussi servir à l'étude des étapes précoces de l'infection virale. Nous avons vu que les peptides affectant ces étapes permettent aussi la survie des cellules infectées avec les virus rapporteurs TK+. Dans le but de confirmer que ces peptides sont impliqués dans les étapes précoces de l'infection, il est possible de remplacer le virus rapporteur TK+ par un virus encodant un fluorochrome. Si dans les cellules exprimant les peptides sélectionnés aucun fluorochrome n'est exprimé, cela confirmera que les peptides identifiés sont effectivement impliqués dans les étapes précoces du cycle viral.

Nous avons décrit une méthode qui permet de valider l'outil qu'est la banque de peptides et qui, dans un deuxième temps, permet aussi l'étude des étapes précoces du cycle viral dans le cadre d'une infection de cellules de mammifères. Une autre technique peut être appliquée pour l'identification de facteurs de restriction affectant les étapes précoces de l'infection. Par exemple, les cellules exprimant la banque de peptides pourraient être soumises à un antibiotique. Seules les cellules qui seront infectées par un virus encodant une protéine de résistance à l'antibiotique pourront survivre. Les cellules dont les facteurs de restriction auront empêché les étapes précoces de l'infection mourront, à défaut d'exprimer la protéine de résistance à l'antibiotique. Les cellules exprimant un peptide qui interfère avec le facteur de restriction survivront. Encore une fois, les cellules dont les peptides interfèrent avec l'action de l'antibiotique survivront aussi. C'est le défaut d'une méthode qui se base sur l'observation indirecte des interactions protéine-protéine. Tous

les peptides étrangers à la voie métabolique étudiée, mais qui permettront à la cellule de survivre lors de la sélection génétique seront des peptides qu'il faudra discriminer par des expériences complémentaires. C'est le cumul des données de plusieurs expériences différentes qui permettra de caractériser exactement les interactions protéine-protéine.

Les études des cycles tardifs du cycle viral sont plus compliquées. En effet, le principe de la sélection génétique nous oblige à trouver le moyen de rendre une étape du cycle tardif toxique pour la cellule. Une application possible pourrait être, par exemple, l'identification du facteur de restriction affectant affectant l'assemblage du VIH-1 et qui est contrecarré par la protéine *Vpu*¹⁵. Pour cette étude, les virus rapporteurs utilisés pour l'infection devront être dépossédés de la protéine *Vpu*. De plus, une protéine virale devra être fusionnée à un agent toxique pour la cellule tout en conservant sa fonction. L'accumulation du virus dans la cellule, causée par le défaut d'assemblage provoqué par le facteur de restriction, causera la mort de la cellule. Les peptides qui interagiront avec le facteur de restriction et qui permettront l'assemblage normal du virus pourront survivre. La caractérisation subséquente de ces peptides et des protéines partenaires permettra l'identification de ce facteur de restriction encore inconnu.

Comme nous venons de le voir, l'utilisation de cet outil d'étude protéomique pave la voie à une panoplie de découvertes potentielles. Cet outil a l'avantage énorme qu'il agit sur des protéines exprimées dans leur environnement cellulaire naturel. L'étude faite *in vivo* dans une situation d'infection virale permet d'interagir dans des voies métaboliques activées souvent exclusivement à la suite d'une infection. Il s'agit d'un outil qui permet d'interférer indirectement, par la voie de peptides, dans des interactions protéine-protéine rares ou sporadiques. La sélection génétique des clones candidats permet de caractériser les peptides impliqués et les protéines avec lesquelles ils ont interagi. La découverte de peptides qui ont la faculté d'altérer une voie importante dans le cycle viral constitue une base importante pour la conception de molécules antivirales. Ces peptides peuvent être appliqués comme tels ou servir à la conception de molécules chimiques peptidomimétiques. Avec l'émergence continue de souches virales résistantes qui constituent une menace constante dans le contrôle efficace de l'épidémie de SIDA⁷⁰, la découverte et la conception de nouvelles classes de drogues antirétrovirales sont expressément requises. En effet, les agents anti-VIH-1 présentement disponibles agissent exclusivement en inhibant l'activité de protéines virales comme la transcriptase inverse, la

protéase et la glycoprotéine d'enveloppe gp-41. La compréhension des points faibles du virus, souvent exploités par la cellule elle-même par le biais de millions d'années de coévolution, pourrait promouvoir la création d'inhibiteurs viraux inédits⁷¹ et l'exploitation de nouveaux modèles animaux⁷².

Références

1. U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE, OFFICE OF BIOLOGICAL AND ENVIRONMENTAL RESEARCH, HUMAN GENOME PROGRAM (2004) [En ligne] http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml, site consulté en nov. 2005.
2. KENYON GL., DEMARINI DM., FUCHS E., GALAS DJ., KIRSCH JF., LEYH TS., MOOS WH., PETSKO GA., RINGE D., RUBIN GM., SHEAHAN LC. (2002) Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report National Academy of Sciences, Washington, D.C., USA. *Mol. Cell. Proteomics*, **1** : 763-780.
3. GAVIN AC., BOSCHE M., KRAUSE R., GRANDI P., MARZIOCH M., BAUER A., SCHULTZ J., RICK JM., MICHON AM., CRUCIAT CM., REMOR M., HOFERT C., SCHELDER M., BRAJENOVIC M., RUFFNER H., MERINO A., KLEIN K., HUDAK M., DICKSON D., RUDI T., GNAU V., BAUCH A., BASTUCK S., HUHSE B., LEUTWEIN C., HEURTIER MA., COPLEY RR., EDELMANN A., QUERFURTH E., RYBIN V., DREWES G., RAID A., BOUWMEESTER T., BORK P., SERAPHIN B., KUSTER B., NEUBAUER G., SUPERTI-FURGA G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**(6868) : 141-7.
4. FIELDS S., SONG O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340** : 245-246.
5. ESPEJO A., CÔTÉ J., BEDNAREK A., RICHARD S., BEDFORD MT. (2002) A protein-domain microarray identifies novel protein-protein interactions. *Biochem J.*, **367**(Pt 3) : 697-702.
6. SACHDEV SS., FAIRBROTHER WJ., DESHAYES K. (2003) Exploring Protein-Protein Interactions with Phage Display. *ChemBioChem*, **4**(1) : 14-25.

-
7. 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**(3) : 591-601.
 8. GONCALVES J., SANTA-MARTA M. (2004) HIV-1 Vif and APOBEC3G: multiple roads to one goal. *Retrovirology*, **1**(1) : 28.
 9. MARIANI R., CHEN D., SCHROFELBAUER B., NAVARRO F., KONIG R., BOLLMAN B., MUNK C., NYMARK-McMAHON H., LANDAU NR. (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell*, **114** : 21-31.
 10. HATZIOANNOU T., PEREZ-CABALLERO D., YANG A., COWAN S., BIENIASZ PD. (2004) Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. *Proc. Natl. Acad. Sci. USA*, **101**(29) : 10774-9.
 11. KECKESOVA Z., YLINEN LM., TOWERS GJ. (2004) The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. *Proc. Natl. Acad. Sci. USA*, **101**(29) : 10780-5.
 12. OWENS CM., SONG B., PERRON MJ., YANG PC., STREMLAU M., SODROSKI J. (2004) Binding and susceptibility to postentry restriction factors in monkey cells are specified by distinct regions of the human immunodeficiency virus type 1 capsid. *J. Virol.*, **78**(10) : 5423-37.
 13. LEE K, KEWALRAMANI VN. (2004) In defense of the cell: TRIM5alpha interception of mammalian retroviruses. *Proc. Natl. Acad. Sci. USA*, **101**(29) : 10496-7.
 14. STREMLAU M, OWENS CM, PERRON MJ, KIESSLING M, AUTISSIER P, SODROSKI J. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature*, **427**(6977) : 848-53.

-
15. VARTHAKAVI V., SMITH RM., BOUR SP., STREBEL K., SPEARMAN P. (2003) Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. *Proc. Natl. Acad. Sci. USA*, **100**(25) : 15154-9.

 16. BINETTE J., COHEN EA. (2004) Recent advances in the understanding of HIV-1 Vpu accessory protein functions. *Curr. Drug Targets Immune Endocr. Metabol. Disord.*, **4** : 297-307.

 17. YU Z., SANCHEZ-VELAR N., CATRINA IE., KITTLER EL., UDOFIA EB., ZAPP ML. (2005) The cellular HIV-1 Rev cofactor hRIP is required for viral replication. *Proc. Natl. Acad. Sci. USA*, **102**(11) : 4027-32.

 18. TOWERS G., BOCK M., MARTIN S., TAKEUCHI Y., STOYE JP., DANOS O. (2000) A conserved mechanism of retrovirus restriction in mammals. *Proc. Natl. Acad. Sci. USA*, **97**(22) : 12295-9.

 19. DIVISION OF ACQUIRED IMMUNODEFICIENCY SYNDROME (DAIDS) OF THE NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES (NIAID), SOUTHERN RESEARCH INSTITUTE (SRI), NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI), (2005) *NIAID HIV-1 Human Protein Interaction Database* [En ligne] <http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/index.html>, Site consulté en nov. 2005.

 20. VON MERING C., KRAUSE R., SNEL B., CORNELL M., OLIVER SG., FIELDS S., BORK P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417**(6887) : 399-403.

 21. ZHU H., BILGIN M., BANGHAM R., HALL D., CASAMAYOR A., BERTONE P., LAN N., JANSEN R., BIDLINGMAIER S., HOUFEK T., MITCHELL T., MILLER P., DEAN RA., GERSTEIN M., SNYDER M. (2001) Global analysis of protein activities using proteome chips. *Science*, **293**(5537) : 2101-5.

-
22. ZANGHI CN., LANKES HA., BRADEL-TRETHERWAY B., WEGMAN J., DEWHURST S. (2005) A simple method for displaying recalcitrant proteins on the surface of bacteriophage lambda. *Nucleic Acids Res.*, **33**(18) : e160.
23. FERRER M., SULLIVAN BJ., GODBOUT KL., BURKE E., STUMP HS., GODOY J., GOLDEN A., PROFY AT., VAN SCHRAVENDIJK MR. (1999) Structural and functional characterization of an epitope in the conserved C-terminal region of HIV-1 gp120. *J. Pept. Res.*, **54**(1) : 32-42.
24. NEW ENGLAND BIOLABS INC. (2004) *PH.D. Phage Display Library Kits*, [En ligne] <http://www.neb.com/nebecomm/TechBulletinFiles/techbulletinE8100.pdf>, site consulté en nov. 2005.
25. BOUHAMDAN M., XUE Y., BAUDAT Y., HU B., SIRE J., POMERANTZ R.J., DUAN L.-X. (1998). Diversity of HIV-1 Vpr Interactions Involves Usage of the WXXF Motif of Host Cell Proteins. *J. Biol. Chem.*, **273** : 8009-8016.
26. HOOGENBOOM HR., DE BRUINE AP., HUFTON SE., HOET RM., ARENDS JW., ROOVERS RC. (1998) Antibody phage display technology and its applications. *Immunotechnology*, **4**(1) : 1-20.
27. WILLATS WG. (2002) Phage display: practicalities and prospects. *Plant Mol. Biol.*, **50**(6) : 837-54.
28. POETZ O., SCHWENK JM., KRAMER S., STOLL D., TEMPLIN MF., JOOS TO. (2005) Protein microarrays : catching the proteome. *Mech. Ageing Dev.*, **126**(1) : 161-70.
29. CUMMINGS CA., RELMAN DA. (2000) Using DNA microarrays to study host-microbe interactions. *Emerg. Infect. Dis.*, **6**(5) : 513-25. Review.
30. SHAHEDUZZAMAN S., KRISHNAN V., PETROVIC A., BITTNER M., MELTZER P., TRENT J., VENKATESAN S., ZEICHNER S. (2002) Effects of HIV-1 Nef on cellular gene expression profiles. *J. Biomed. Sci.*, **9**(1) : 82-96.

-
31. RENUCCI, A. (2005) 'Microarray', (puces), appliqués à l'analyse des protéines. *Développements technologiques et applications, un aperçu aux États-Unis*. Rapports d'ambassade **smm05_035**, 23 pages.
32. OFIR K., BERDICHEVSKY Y., BENHAR I., AZRIEL-ROSENFELD R., LAMED R., BARAK Y., BAYER EA., MORAG E. (2005) Versatile protein microarray based on carbohydrate-binding modules. *Proteomics*, **5(7)** : 1806-14.
33. AEBERSOLD R., MANN M. (2003) Mass spectrometry-based proteomics. *Nature*, **422(6928)** : 198-207. Review.
34. SOBHANIFAR S. (2005) The Yeast Two-Hybrid Assay An Exercise in Experimental Eloquence. *BioTeach Online Journal*, accessible en ligne à l'adresse <http://www.bioteach.ubc.ca/MolecularBiology/AYeastTwoHybridAssay/>, site consulté en nov. 2005.
35. ESTOJAK J., BRENT R., GOLEMIS EA. (1995) Correlation of two-hybrid affinity data with *in vitro* measurements, *Mol. Cell. Biol.*, **15** : 5820-5829.
36. CLARKE P., Ó CUÍV P., O'CONNELL M. (2005) Novel mobilizable prokaryotic two-hybrid system vectors for high-throughput protein interaction mapping in *Escherichia coli* by bacterial conjugation, *Nucleic Acids Res.*, **33** : e18.
37. HOSFIELD T., CHANG C. (2002) New Mammalian Two-Hybrid System Detects Protein-Protein Interactions. *Stratagene technical bulletin*, accessible en ligne < <http://www.biocompare.com/techart.asp?id=27> >.
38. BRENT R., FINLEY RLJ. (1997) Understanding gene and allele function with two-hybrid methods., *Annu. Rev. Genet.*, **31** : 663-704.
39. FROMONT-RACINE M., RAIN JC., LEGRAIN P. (1997) Towards a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.*, **16** : 277-81.

-
40. ROECKLEIN JA., SENGUPTA D. (1996) A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet.*, **12** : 72-7.
41. PHIZICKY E., BASTIAENS P. I.H., ZHU H., SNYDER M., FIELDS S., (2003) Protein analysis on a proteomic scale, *Nature*, **422** : 208-215.
42. COLAS P. (2000) Exploration des interactions protéiques à l'échelle génomique. *médecine/sciences 2000*, **16** : 50-6.
43. MROWKA R., PATZAK A., HERZEL H. (2001) Is there a bias in proteome research? *Genome Res.*, **11** : 1971-73.
44. DEANE C.M., SALWINSKI L., XENARIOS I., EISENBERG D. (2002) Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol. Cell. Proteomics*, **1**, 349-356.
45. KEMMEREN P., VAN BERKUM NL., VILO J., BIJMA T., DONDEERS R., BRAZMA A., HOLSTEGE FC. (2002) Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell.*, **9**(5) : 1133-43.
46. JOHNSON N., VARSHAVSKY A. (1994). Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. USA*, **91** : 10340-10344.
47. DUALSYSTEMS BIOTECH AG (2005) *The split-ubiquitin system* [En ligne] <http://www.dualsystems.com/technologies/split.asp>, site consulté en nov. 2005.
48. LI H., YING L., REN X., BALASUBRAMANIAN S., KLENERMAN D. (2004) Fluorescence studies of single biomolecules. *Biochem Soc. Trans.*, **32**(Pt 5) : 753-6. Review.
49. DERDOWSKI A., DING L., SPEARMAN P. (2004) A novel fluorescence resonance energy transfer assay demonstrates that the human immunodeficiency virus type 1 Pr55Gag I domain mediates Gag-Gag interactions. *J. Virol.*, **78**(3) : 1230-42.

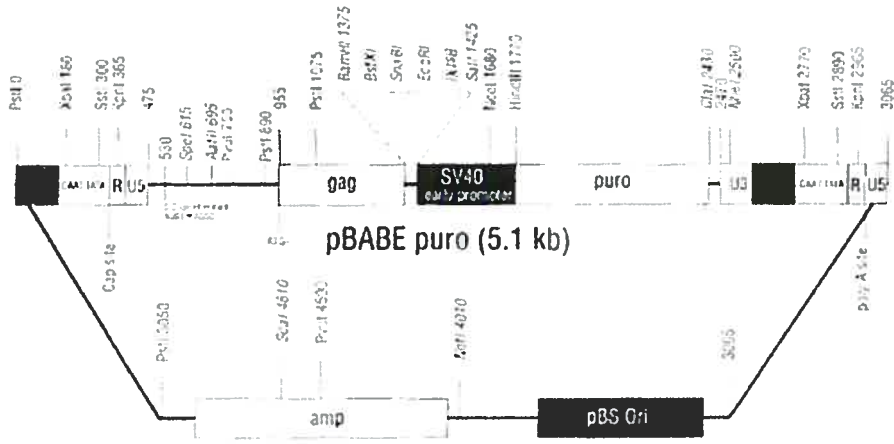
-
50. XU X., LEO C., JANG Y., CHAN E., PADILLA D., HUANG BC., LIN T., GURURAJA T., HITOSHI Y., LORENS JB., ANDERSON DC., SIKIC B., LUO Y., PAYAN DG., NOLAN GP. (2001) Dominant effector genetics in mammalian cells. *Nat. Genet.*, **27**(1) : 23-9.
51. KITAMURA T., ONISHI M., KINOSHITA S., SHIBUYA A., MIYAJIMA A., NOLAN GP. (1995) Efficient screening of retroviral cDNA expression libraries. *Proc. Natl. Acad. Sci. USA*, **92**(20) : 9146-50.
52. WOLKOWICZ R., JAGER GC., NOLAN GP. (2005) A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. *J. Biol. Chem.*, **280**(15) : 15195-201.
53. KINSELLA TM., OHASHI CT., HARDER AG., YAM GC., LI W., PEELLE B., PALI ES., BENNETT MK., MOLINEAUX SM., ANDERSON DA., MASUDA ES., PAYAN DG. (2002) Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J. Biol. Chem.*, **277**(40) : 37512-8.
54. ROEPE PD. (2001) A peptide needle in a signaling haystack. *Nat. Genet.*, **27**(1) : 6-8.
55. WOLKOWICZ R., NOLAN GP. (2003) Retroviral technology--applications for expressed peptide libraries. *Front. Biosci.*, **8** : d603-19. Review.
56. HALE MB., NOLAN GP., WOLKOWICZ R. (2004) Oligonucleotide-directed site-specific integration of high complexity libraries into ssDNA templates. *Nucleic Acids Res.*, **32**(2) : e22.
57. OLSEN DB., ECKSTEIN F. (1990) High-efficiency oligonucleotide-directed plasmid mutagenesis. *Proc. Natl. Acad. Sci. USA*, **87**(4) : 1451-5.
58. MARTI TM., KUNZ C., FLECK O. (2002) DNA mismatch repair and mutation avoidance pathways. *J. Cell Physiol.*, **191**(1) : 28-41. Review.

-
59. NOLAN, GP. (2005) *Peptide Library Mini-Tutorial* [En ligne] http://www.stanford.edu/group/nolan/library_systems/peptide_systems.html, site consulté en nov. 2006.
60. OSTIGUY, A. (2003) Mise au point d'une approche de sélection génétique de peptides inhibiteurs de fonctions protéiques fonctionnelles en cellules de mammifères. *Mémoire de maîtrise*, Université de Montréal, 87 pages.
61. GAO G., GUO X., GOFF SP. (2002) Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science*, **297**(5587) : 1703-6.
62. MORGENSTERN JP., LAND H. (1990) Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line., *Nucleic Acids Res.*, **18**(12) : 3587-96.
63. SAMBROOK J., FRITSCH E.F., MANIATIS T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY, Vol. 1, 2, 3.
64. GRAHAM FL., SMILEY J., RUSSELL WC., NAIRN R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.*, **36**(1) : 59-74.
65. ROYER Y., MENU C., LIU X., CONSTANTINESCU SN. (2004) High-throughput gateway bicistronic retroviral vectors for stable expression in mammalian cells: exploring the biologic effects of STAT5 overexpression. *DNA Cell Biol.*, **23**(6) : 355-65.
66. HARRIES M., PHILLIPPS N., ANDERSON R., PRENTICE G. (2000) Collins M. Comparison of bicistronic retroviral vectors containing internal ribosome entry sites (IRES) using expression of human interleukin-12 (IL-12) as a readout., *J. Gene Med.*, **2**(4) : 243-9.
67. PIERCE BIOTECHNOLOGY INC. (2002) *Co-Immunoprecipitation* , <http://www.piercenet.com/Proteomics/browse.cfm?fldID=9C471132-0F72-4F39-8DF0-455FB515718F>, site consulté en nov. 2006.

-
68. LABEAN TH., KAUFFMAN SA. (1993) Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics. *Protein Sci.*, **2(8)** : 1249-54.
69. GENE LINK INC. (2005) *Synthesis of Long Oligos*, <http://www.genelink.com/Literature/ps/longoligos-cat.pdf>, site consulté en nov. 2006.
70. BERKHOUT B., DE RONDE A., VAN DER HOEK L. (2005) Aggressive HIV-1?, *Retrovirology*, **2(1)** : 13.
71. MarketResearch.com, (2005) *Protein-Protein Interactions as Drug Targets: Focusing on Cell Circuitry*, Drug and Market Development Publishing, 229 Pages.
72. HATZIOANNOU T, COWAN S, BIENIASZ PD. (2004) Capsid-dependent and -independent postentry restriction of primate lentivirus tropism in rodent cells. *J. Virol.*, **78(2)** : 1006-11

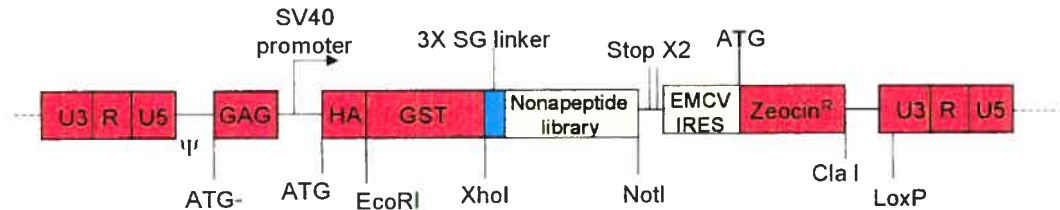
Annexes

CARTES DES VECTEURS UTILISÉS

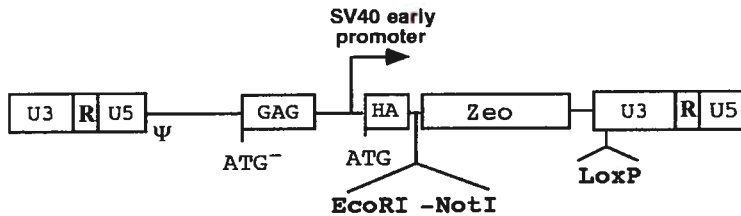


Unique sites are shown in italics.

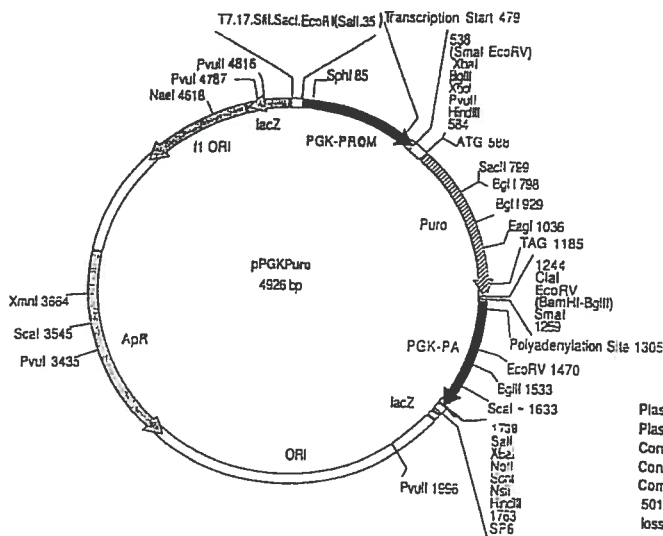
1. pBABE puro. (Représentation modifiée de la publication NIH AIDS Research & Reference Reagent Program. Data Sheet du produit #3321)



2. pBABE HAGIZ (partiel). (Représentation de tirée de Alexandre Ostiguy, 2003)

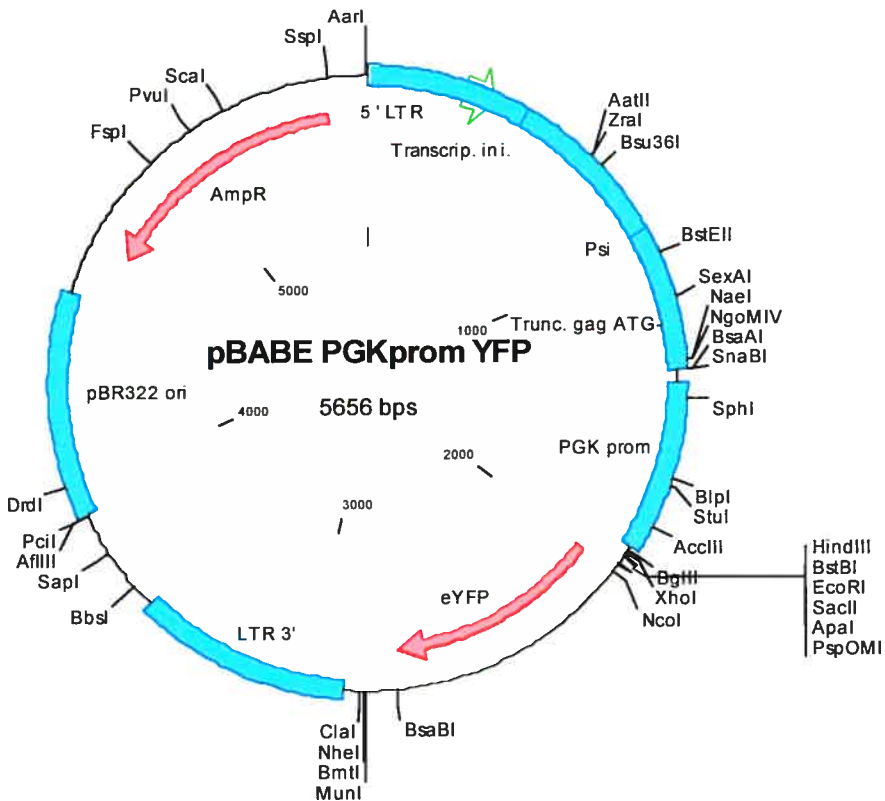


3. pBABE HAZ (partiel) (Représentation tirée de Gao G et al, 2002)

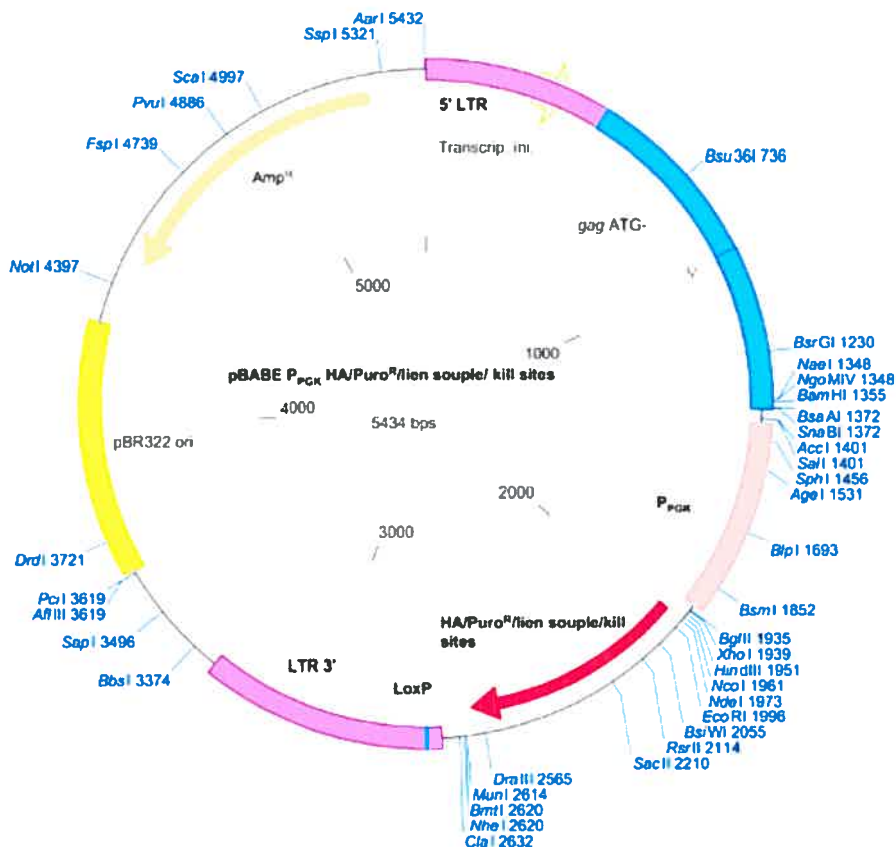


Plasmid name: pPGKPuro
 Plasmid size: 4926 bp
 Constructed by: Peter W. Laird
 Construction date: January 1992
 Comments/References: Three-way ligation of 3741-bp NotI-XhoI fragment of pPWL514, 501-bp BamHI-NotI fragment of pPWL513, and 684-bp XhoI-BglII fragment of pSPPuro with loss of the BamHI and BglII sites.

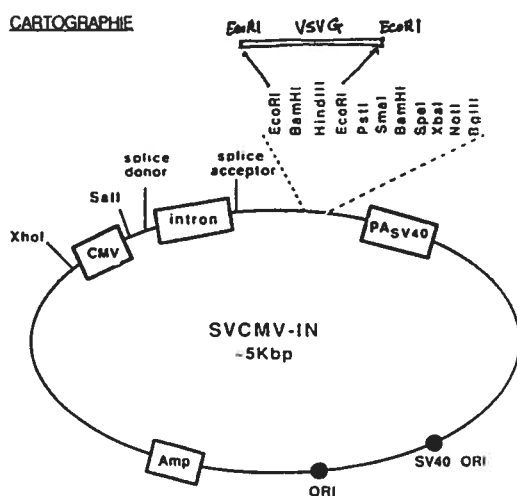
4. pPGK PURO



5. pBABE P_{PGK} eYFP

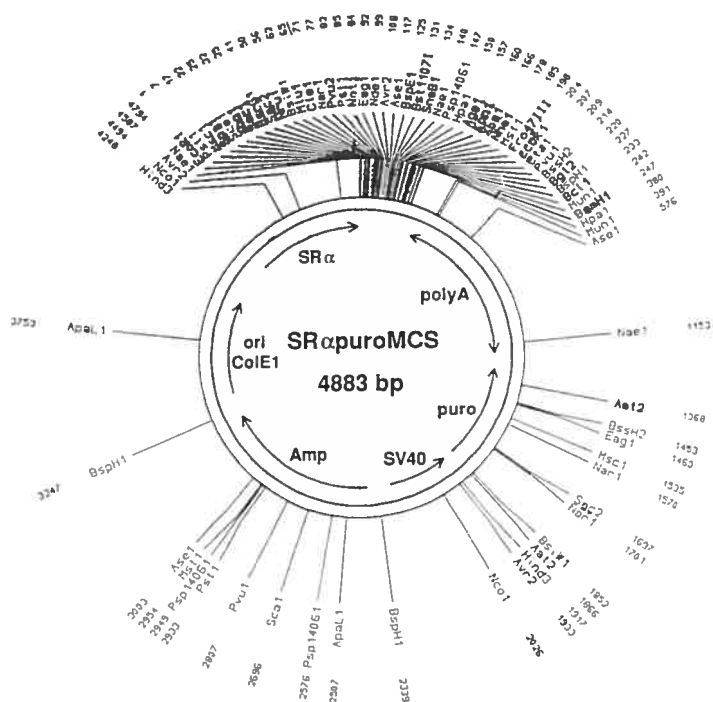


6. pBABE PPGK HA/PuroR/lien souple/kill sites

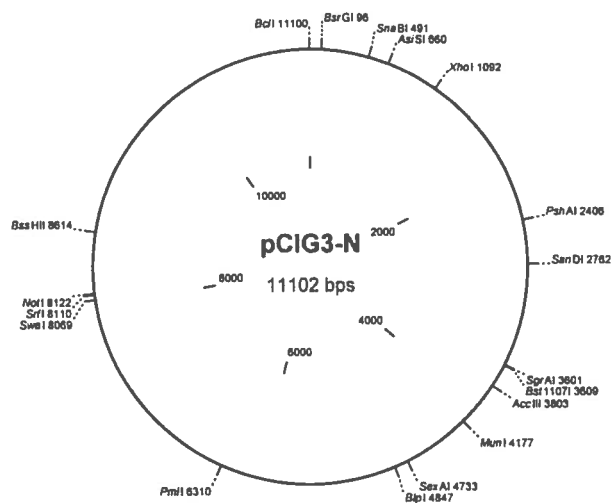


NOM DU PLASMIDE : SVCMV-IN (intron) VSV-G
 GENE INTRODUIT : VSV-G
 VECTEUR : pBluescript backbone
 HOTE : AG1, ou autres souches *E. coli*
 RÉSISTANCE ANTIBIOTIQUE : Amp
 TAILLE DU PLASMIDE : ~5 Kbp
 TAILLE DU GENE INTRODUIT : ---
 LOCALISATION : - Congélateur :
 - boîte :
 RÉFÉRENCES : obtained from Craig Rosen (05/28/92)
 COMMENTAIRES : possède un intron qui augmente de 5-7 fois l'expression d'un gène placé en aval de cet intron
 L'intron provient du gène B-globin

7. pSVCMV-IN. Ce vecteur permet de pseudotyper les virions avec la protéine d'enveloppe VSV-G.



8. Plasmide SRα-puro



Molecule: pCIG3-N, 11102 bps DNA Circular
 File Name: pCIG3-N (gag-pol).cm5, dated 18 Apr 2005

Description: Vecteur gag/pol MLV B-tropic

Notes: S
 quence fournie par G. Towers

Pour details sur la construction to pCIG3N et B, voir Bock et al. JV
 2000 (Stoye est dernier auteur).

Here is pCIG3 N sequence. The sequence of the B capsid is on genbank I
 guess. There is only 3 aa differences between N and B in gag and 1 in
 pro.

9. Expresseur de Gag-Pol pCIG3-N