

Université de Montréal

Mise au point d'une approche de sélection génétique de peptides inhibiteurs
d'interactions protéiques fonctionnelles en cellules de mammifères

Par
Alexandre Ostiguy

Département de Microbiologie et Immunologie
Faculté de Médecine

Mémoire présenté à la Faculté des Études supérieures
En vue de l'obtention du grade de M.Sc.
En Microbiologie et Immunologie

Octobre 2003

© Alexandre Ostiguy, 2003



W

4

U58

2004

V. 089

Direction des bibliothèques

AVIS

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

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

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

NOTICE

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

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

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

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Mise au point d'une approche de sélection génétique de peptides inhibiteurs
d'interactions protéiques fonctionnelles en cellules de mammifères

Présenté par :
Alexandre Ostiguy

A été évalué par un jury composé des personnes suivantes :

George Szatmari
Président rapporteur

Éric A. Cohen
Directeur de recherche

Pierre Belhumeur
Codirecteur

France Daigle
Membre du jury

RÉSUMÉ

Les interactions protéiques sont au cœur des voies de signalisation intracellulaires qui régissent les processus internes de la cellule. À ce jour, plusieurs méthodes, comme le «phage display» et le système deux hybrides en levures, ont tenté d'identifier ces interactions et de leur trouver une signification fonctionnelle dans la régulation de la cellule. Malgré leur apport indéniable à notre connaissance des mécanismes régissant les interactions protéiques intracellulaires, ces méthodes sont limitées soit par le fait que, comme le «phage display», elles soient effectuées *in vitro*, ou encore par le fait qu'elles ne permettent pas de travailler directement en cellules de mammifères, ce qui empêche de faire des corrélations fonctionnelles directes entre les interactions identifiées et le fonctionnement de cellules eucaryotes supérieures. Face à cette situation, nous avons décidé de mettre au point un système de sélection génétique de nonapeptides inhibiteurs de relations fonctionnelles protéiques pouvant être exprimé en cellules de mammifères. Cette approche possède des avantages indéniables par rapport aux méthodes existantes : elle permet d'identifier des interactions entre un peptide et une protéine dans son milieu physiologique normal, donc avec toutes les modifications post-traductionnelles, la compartimentalisation intracellulaire et les conditions physiologiques lui permettant d'effectuer sa fonction. De plus, notre système possède une diversité de l'ordre de 2.5×10^{11} nonapeptides différents fusionnés à la glutathione-S-transférase, ce qui augmente les chances d'identifier un partenaire fonctionnel de la protéine étudiée en plus de permettre sa copurification par chromatographie d'affinité. Nous avons donc tenté de mettre sur pied cette librairie de nonapeptides dans le vecteur épisomal pCEP4, pour ensuite transposer le système dans le vecteur rétroviral pBABE HAGIZ. Nous avons également mis sur pied une lignée cellulaire exprimant la thymidine kinase (TK) du virus *Herpes simplex* (HSV) qui allait servir de système de sélection de peptides inhibiteurs de la TK. Le système est basé sur l'utilisation du GCV, un agent antiviral tuant les cellules exprimant la TK de HSV, le but étant d'isoler des peptides conférant une résistance au GCV par l'inhibition de la TK. Ce mémoire fait donc état des stratégies employées pour construire cette librairie de nonapeptides dans

deux systèmes différents ainsi que de la mise au point d'un système de validation de cette librairie dans des cellules de mammifères en culture.

Mots clés : HSV TK, librairie de nonapeptides, sélection génétique, cellules de mammifères, interactions protéiques.

ABSTRACT

Protein-protein interactions take part in transduction pathways to regulate the processes that make a cell function. To date, several approaches were set up to study these interactions and their implication in a cell's intimate mechanisms, «phage display» and yeast two-hybrid assay being two good examples of such methods. Despite their valuable contribution to the study of inter-protein interactions, these methods are limited by several factors : as it is the case with «phage display», they often take place in a strictly *in vitro* environment, or, like with the yeast two hybrid assay, they don't take place directly in mammalian cells, preventing us from making direct functional correlations between inhibitory peptides or cDNAs and the protein studied. We thus decided to set up a genetic selection system that would allow the identification of inhibitory peptides of protein interactions directly in mammalian cells. This approach has many advantages over the widespread methods actually used, the main one being the possibility of identifying interactions between a peptide and a protein directly in its physiological environment, with all the post-translational modifications, intracellular localisation and physiological conditions involved in the mechanism of action of this protein. Moreover, the great theoretical diversity that can be achieved with this library (2.5×10^{11} different nonapeptides fused to inert carrier protein glutathione-S-transferase (GST)) allows for a more probable identification of inhibitory peptides and their possible copurification by affinity chromatography. This library was first constructed in the episomal vector pCEP4 and then further developed in a retroviral system using the pBABE HAGIZ vector. We also developed a screening system based on the HSVTK/GCV suicide gene approach to select for peptides inhibiting HSVTK. This thesis reports the steps taken in the setup of this approach in two different expression systems as well as the setup of a genetic screening system built to validate the concept of our library.

Keywords : HSV TK, genetic selection, nonapeptide library, protein interactions.

SECTION INTRODUCTION :	1
1. CONTEXTE	1
2. THÉORIE RELATIVE AUX LIBRAIRIES	2
2.1 Définition d'une librairie.....	2
2.2 Diversité de la librairie.....	4
2.3 Système de criblage et identification d'homologie.....	6
2.4 Bruit de fond	8
2.5 Présentation	8
3. SURVOL DES MÉTHODES EXISTENTES	10
3.1 «phage display».....	10
3.2 «ribosome display»	12
3.3 Système deux hybrides en levure.....	14
4. GST	16
5. VECTEURS	18
5.1 Vecteur pCEP4.....	18
5.2 Principe et diversité des vecteurs rétroviraux	19
5.2.1 Rétrovirus.....	19
5.2.2 Vecteur rétroviral	22
6. NOTRE LIBRAIRIE	23
7. SYSTÈME DE VALIDATION DE LA LIBRAIRIE	24
7.1 Thymidine kinase du virus <i>Herpes simplex</i>	24
7.2 Systèmes de sélection HSV TK/GCV	26
8. OBJECTIFS DU PROJET	28
SECTION MATÉRIEL ET MÉTHODES	29
1. BIOLOGIE MOLÉCULAIRE :	29
Considérations générales.....	29
Clonage :	29
1.1 Constructions pCEP4	31

1.1.1 Construction pCEP GST	31
1.1.2 Construction pCEP GST GFP	31
1.2 Constructions pBABE	32
1.2.1 pBABE HAZ GST	32
1.2.2 pBABE HAGIZ	32
1.2.3 pBABE HAGIZ GFP	34
1.3 Synthèse du 2 ^{icme} brin de la librairie	34
1.4 Clonage de la librairie dans pCEP GST/pBABE HAGIZ	35
1.5 Évaluation de la diversité de la librairie (digestion, séquençage) dans système pCEP	38
1.6 Récupération du vecteur pCEP GST transfecté dans des cellules 293T	39
2. BIOLOGIE CELLULAIRE :	40
2.1 Lignées cellulaires utilisées	40
2.2 Lignée cellulaire 293TK+	40
2.2.1 Production des virus HSVTK+	40
2.2.2 Infection des cellules 293 par le virus HSVTK+	41
2.2.4 Test toxicité du ganciclovir dans 293TK+	42
2.3 Sélection zéocyne de cellules transduites ou transfectées avec le vecteur pBABE HAGIZ	42
3. BIOCHIMIE :	43
3.1 Expression de GST dans les cellules 293T à partir de pCEP GST	43
3.2 Expression long terme de GST dans 293TK+ sous sélection hygromycine	44
3.3 Expression de la fusion GFP/TK+ dans 293TK+	44
3.4 Expression de GST à partir du vecteur rétroviral	45
4. PRODUCTION ET INFECTION VIRALE :	46
4.1 Production de particules virales à partir de pBABE HAGIZ dans les cellules 293T	46
4.2 Standardisation de l'infection par essai RT	46
4.3 Infection de cellules 293T pour montrer la résistance à la zéocyne à partir du vecteur rétroviral pBABE	47
4.4 Détection de l'intégration du gène de la zéocyne dans les cellules 293T exprimant la cassette GST-Zéo de façon stable	47
SECTION RÉSULTATS :	49
1. CONSTRUCTION D'UNE LIBRAIRIE DANS UN VECTEUR ÉPISOMAL	49
1.1 Analyse de l'expression de GST à partir du vecteur épisomal pCEP GST	49

1.2 Analyse de l'expression à long terme de GST à partir du vecteur pCEP GST dans des cellules 293TK+.....	50
1.3 Construction d'une librairie de nonapeptides dans le vecteur pCEP GST et évaluation de la diversité réelle de cette librairie.....	51
1.4 Récupération du vecteur pCEP GST transfecté dans des cellules 293T.....	51
2. CONSTRUCTION D'UNE LIBRAIRIE DE NONAPEPTIDES DANS UN VECTEUR RÉTROVIRAL	52
2.1 Construction du vecteur rétroviral pBABE HAGIZ	53
2.2 Analyse de l'aptitude du vecteur rétroviral pBABE HAGIZ à produire des particules virales infectieuses.....	53
2.3 Analyse de l'expression transitoire de GST à partir du vecteur rétroviral pBABE HAGIZ dans des cellules 293T	54
2.4 Évaluation du vecteur pBABE HAGIZ pour son utilisation dans le transfert de la librairie de nonapeptides et du gène de résistance à la zéocyne dans les cellules 293T	55
2.5 Clonage d'une librairie de nonapeptides dans le vecteur rétroviral pBABE HAGIZ	58
2.6 Évaluation de la diversité des séquences peptidiques clonées dans les vecteurs pCEP GST et pBABE HAGIZ.....	58
3. DÉVELOPPEMENT DU SYSTÈME DE CRIBLAGE DE LA LIBRAIRIE EN CELLULES DE MAMMIFÈRES.....	60
3.1 Analyse de l'expression du gène de la TK de HSV dans les cellules de la lignée cellulaire 293TK+	61
3.2 Évaluation de la capacité du système HSVTK/GCV à induire la mort cellulaire des cellules 293TK+.....	61
SECTION DISCUSSION	63
1. Clonage de la librairie dans le vecteur épisomal pCEP4	63
2. Clonage de la librairie dans le vecteur rétroviral pBABE HAZ	70
3. Mise au point d'un système de criblage de la librairie en cellules de mammifères	78
4. Perspectives, applications futures et conclusion	79
BIBLIOGRAPHIE.....	83
ANNEXE I : VECTEURS UTILISÉS.....	XIV

LISTE DES FIGURES ET TABLEAUX

FIGURE		PAGE
1	Principe du «phage display».....	12
2	Principe du «ribosome display».....	14
3	Système deux-hybrides en levures.....	15
4	Organisation de la librairie dans pCEP4.....	19
5	Cassette d'expression bicistronique dans le vecteur pBABE HAGIZ.....	22
6	Structure du GCV.....	27
7	Méthode de clonage de la cassette bicistronique dans pBABE HAGIZ.....	33
8	Organigramme du clonage de la librairie de nonapeptides.....	38
9	Expression de GST à partir du vecteur pCEP GST.....	50
10	Expression à long terme de GST à partir du vecteur pCEP GST.....	51
11	Récupération du vecteur pCEP GST transfecté dans des cellules 293T.....	52
12	Structure de la cassette bicistronique dans pBABE HAGIZ et stratégie de séquençage.....	53
13	Quantification de la RT incorporée dans les particules virales produites à partir de pBABE HAGIZ.....	54
14	Expression transitoire de la GST à partir de pBABE HAGIZ.....	55
15	Intégration du gène de résistance à la zéocine dans des cellules 293T transduites avec pBABEHAGIZ.....	56
16	Expression de GST dans des clones résistants à la zéocine.....	57
17	Expression de la fusion GFPTK dans les cellules 293TK+.....	61
18	Toxicité du GCV dans les cellules 293TK+.....	62

TABLEAU

I	Séquence et fréquence des peptides clonés dans pCEP GST et pBABE HAGIZ.....	60
---	---	----

Liste des abréviations :

ADN :	Acide désoxyribonucléique
Amp:	Ampicilline
ARN :	Acide ribonucléique
ARNm :	ARN messenger
BSA :	Bovine Serum Albumin
CA :	Protéine de capsid
Cdk2 :	cyclin dependent kinase
dATP:	désoxyadéninetriphosphate
dCTP:	désoxycytosinetriphosphate
dGTP:	désoxyguaninetriphosphate
dTTP:	désoxythymidinetriphosphate
DMEM :	Dulbecco's Modified Eagle Medium
DMSO :	Dimethyl sulfoxide
DO :	Densité optique
DTT :	Dithiothréitol
EDTA :	(Ethylenedinitrilo)tetraacetic Acid
FACS :	Fluorescence Activated Cell Sorting
FBS :	Foetal Bovine Serum
GCV :	Ganciclovir
GFP :	Green Fluorescent Protein
GST :	Glutathione-S-tranferase
h:	heure
HA :	Hémagglutinine
IN :	Intégrase
IC ₅₀ :	Inhibitory concentration 50%
IRES :	Internal Ribosome Entry Site
Kb :	kilobase
LB :	Luria-Bertani
MA :	protéine de matrice
LMP :	Low Melting Point

LTR :	Long Terminal Repeat
MAPK :	Mitogen Activated Protein Kinase
mM :	millimolaire
MuLV :	Murine Leukemia Virus
NC :	Nucléocapside
O/N :	Overnight (toute la nuit)
pBABE HAZ :	HA-Zéocyne
pBABE HAGIZ :	HA-GST-IRES-ZÉOCYNE
PBS :	Phosphate Buffered Saline
PCR :	Polymerase Chain Reaction
PM :	Poids moléculaire
PPT :	Polypurine Tract
PR :	Protéase
RT :	Transcriptase inverse
Rpm :	rotation par minute
SH3 :	Src homology domain 3
SDS :	Sodium Dodecyl Sulfate
SV40 :	simian virus 40
TCA :	Acide trichloroacétique
TE :	Tris EDTA
TM :	Domaine transmembranaire de la glycoprotéine d'enveloppe
UFC :	Unité formant des colonies
VSV :	Vesicular Stomatitis virus

«L'expérience n'est pas expérimentale. On ne la
provoque pas. On la subit. Plutôt patience
qu'expérience [...] Au sortir de l'expérience,
on n'est pas savant, on est expert. Mais en quoi? »

-Albert Camus
Carnets

REMERCIEMENTS

Voilà, c'est fini. Deux ans et demi de laboratoire, quatre-vingt-six fins de semaine au labo, quelques litres de café, un nuage inquantifiable de fumée de cigarette, plusieurs hectolitres de cervoise tiède, beaucoup moins de cheveux et une relation privilégiée avec des cellules en culture plus tard, tout se résume en cent petites pages à double interligne... et en plusieurs personnes sans qui tout cela eut été impensable et qui ne sauraient se réduire à quelques lignes en préface d'un mémoire. Mais tout de même.

Je tiens donc à remercier le Dr Éric Cohen pour m'avoir donné l'occasion de travailler dans son laboratoire, et pour avoir pris le temps de discuter quand les choses n'allaient pas aussi bien qu'on l'eût espéré. Merci également au Dr Pierre Belhumeur pour sa collaboration et ses conseils judicieux quand les expériences semblaient vouloir n'en faire qu'à leur tête.

Un chaleureux et bien senti « MERCI » à tous mes collègues du laboratoire qui ont partagé ces deux années et qui ont rempli le quotidien de leur présence toute particulière. Un gracias muy especial a Andrés, por su amistad, sus consejos preciosos y su vision tan especial que le hacen un amigo que no olvidaré. Merci aux deux Europes de chez Louis, qui m'ont appris à parler et qui ont été des partenaires festifs aussi infatigables qu'appréciés, ainsi qu'à Charles S. Fortin pour les mêmes raisons. Merci à la Sherbrooke connection : Dom pour ton amitié, JF pour la folie, François pour l'organisation, Julie pour les cigarettes et le gâteau concave.

Et finalement, dans la catégorie je ne sais pas ce que j'aurais fait sans vous, et dans le désordre : les boys, mon professeur de croissance personnelle, guru Longchen Nyingtik Dilgo Khyentse Rinpoche, l'Intermed et ses pizzas, le Yer'mad, la Brasserie du Cheval Blanc, ma famille, Tylenol, et en tout dernier lieu, mais c'est pour que ça paraisse plus, Mademoiselle K, pour tout et surtout pour rien, c'est gratuit.

SECTION INTRODUCTION :

1. CONTEXTE

L'étude des voies de signalisation intracellulaires a été bouleversée par l'avènement de la génomique, et plus particulièrement par l'utilisation à grande échelle des systèmes de puces à ADN. Des réseaux complexes de signalisation ont ainsi été mis au jour et de nombreuses équipes se sont attelées à l'élucidation des mécanismes fins qui permettent à la cellule de transformer les signaux provenant de l'environnement en réactions biochimiques lui permettant de s'adapter à cet environnement.

Plusieurs outils ont été développés au cours des deux dernières décennies dans le but d'identifier des interactions protéines-protéines ayant cours dans les réseaux de signalisation cellulaire. Plusieurs d'entre eux, comme le «phage display» et le système deux hybrides en levure, ont permis et permettent encore de mieux définir les interactions entre les protéines, que ce soit au niveau des domaines impliqués ou tout simplement au niveau de l'identification de nouveaux partenaires fonctionnels de protéines déjà connues. Malgré leur puissance et leurs utilisations variées, ces systèmes comportent un inconvénient majeur : ils ne permettent pas d'étudier des interactions protéiques directement à l'intérieur d'une cellule de mammifère en culture. En effet, le «phage display» fonctionne *in vitro* alors que le système deux hybrides, même s'il est utilisé *in vivo* dans des cellules de levures, ne permet pas toujours de faire la transposition dans des systèmes supérieurs comme les cellules de mammifères en culture. À un moment où la génomique croise la protéomique et où la communauté scientifique cherche à mieux définir les interactions protéiques qui sont au cœur des processus cellulaires, il devient très pertinent de développer un outil qui permettrait d'étudier ces interactions *in vivo*. Nous nous sommes donc penchés sur le développement

d'une approche de sélection génétique de peptides inhibiteurs de certaines interactions protéiques ayant cours au sein des cellules de mammifères.

Ce mémoire comporte 3 objectifs particuliers : premièrement, cloner une librairie aléatoire de nonapeptides dans un vecteur épisomal, deuxièmement, appliquer cette approche à la technologie des vecteurs rétroviraux et, finalement, mettre sur pied un système fiable et reproductible de validation de ces deux types de librairies.

2. THÉORIE RELATIVE AUX LIBRAIRIES

2.1 Définition d'une librairie

Qu'est-ce qu'une librairie? Une librairie est un système qui contient l'information génétique permettant l'expression d'un très grand nombre d'éléments protéiques (protéines, domaines fonctionnels, peptides) dans un système biologique. Il existe plusieurs types de librairies : des librairies d'ADNc pouvant être exprimées en levures et en cellules de mammifères, des librairies d'ADN génomique, des librairies de peptides exprimés à la surface d'un phage, dans des levures, dans des cellules de mamifères, etc. Le principe de base qui sous-tend l'utilisation des librairies, peu importe leur forme, est le suivant : dans une multitude de molécules différentes (peptides, ADNc, protéines, etc.), les probabilités existent pour qu'un petit nombre de celles-ci puissent avoir un effet au sein d'un système biologique donné. L'utilisation d'une librairie se fait en conjonction avec un système de criblage qui permet l'identification et la récupération de la molécule responsable de cet effet. On se base sur une amplification d'un événement rare pour identifier les molécules qui agissent sur le mécanisme étudié dans le système de criblage.

L'histoire des bibliothèques commence avec le clonage de banques d'ADN complémentaires (ADNc). En clonant ces banques, on espérait pouvoir identifier des séquences complètes de gènes ayant un effet, un phénotype, lorsqu'exprimées dans un organisme. Cependant, une fois la séquence identifiée, beaucoup de travail restait à faire au niveau biochimique : quels étaient les domaines de la protéine impliqués dans son effet, quels sont les partenaires fonctionnels de cette protéine, quels sont les acides aminés importants pour son rôle dans la cellule? Pour répondre à cette question, on a tenté d'utiliser des bibliothèques d'ADNc fragmenté. On pensait ainsi identifier non pas des gènes complets mais des domaines fonctionnels impliqués dans divers mécanismes intracellulaires. On s'est ensuite rendu compte qu'il était possible d'exposer de petits fragments de protéines (des peptides) générés aléatoirement à la surface de phages et qu'on pouvait ainsi identifier des motifs d'interactions entre protéines. On a ensuite étendu le raisonnement à l'utilisation de peptides intracellulaires. Ces peptides, lorsqu'exprimés dans des cellules, pouvaient avoir un effet sur certaines voies biochimiques et être sélectionnés comme tel (83). Les peptides ont cet avantage majeur qu'ils sont l'expression la plus simple de leur effet, c'est-à-dire qu'ils peuvent être utilisés directement comme un inhibiteur biochimique sans avoir à délimiter plus précisément de sous-domaines responsables de leur activité biologique. Les peptides peuvent également servir directement d'épitopes pour isoler et purifier leurs partenaires par affinité. Par le fait qu'ils perturbent des voies de signalisation intracellulaires, les peptides peuvent servir à démontrer la fonction d'une interaction protéique, et c'est là que réside toute la puissance de l'approche.

Il existe déjà des bibliothèques de peptides pouvant être exprimées dans des bactéries et des levures, mais les bibliothèques pouvant être utilisées dans des cellules de mammifères sont beaucoup plus rares et permettent des études qui seraient impossibles dans des micro-organismes plus simples. Certains processus propres aux cellules eucaryotes supérieures (apoptose, biologie des ré-

cepteurs de surfaces comme le CD4 ou les protéines G, interactions virus-hôtes) ne peuvent en effet pas être étudiés comme tels dans les bactéries ou les levures. Pour toutes ces raisons, nous avons décidé de mettre sur pied un système de sélection génétique basé sur l'utilisation d'une librairie de nonapeptides exprimée en cellules de mammifères.

2.2 Diversité de la librairie

Le point le plus important à considérer lors de la construction d'une librairie est sa diversité. La diversité d'une librairie représente le nombre total de molécules différentes provenant de l'information génétique codée par cette librairie. Chaque codon pouvant coder pour 20 acides aminés, la diversité théorique d'une librairie peut se calculer comme en utilisant la formule 20^x , où x est le nombre d'acides aminés dans le peptide. Par exemple, une librairie comme la nôtre, qui possède 9 codons aléatoires, contient l'information nécessaire à la synthèse de 5×10^{11} peptides différents. Une librairie codant pour 8 acides aminés contiendrait 1×10^{10} peptides différents et une librairie codant pour 12 acides aminés en contiendrait 10×10^{15} . Plus un peptide contient d'acides aminés, plus l'interaction avec un partenaire potentiel sera forte. Par contre, plus le peptide est long, plus l'énergie de liaison est grande et moins on a de chances d'isoler une interaction fonctionnelle. Nous avons choisi d'utiliser une librairie codant pour des nonapeptides, ce qui représente un bon compromis entre spécificité et diversité. Théoriquement, plus le nombre de peptides différents est grand, plus on a de chances de pouvoir isoler un peptide bioactif (54). Comme les techniques de clonages actuelles ne permettent pas d'obtenir plus de 10^9 clones différents (63) en raison des limitations pratiques inhérentes aux différentes étapes, l'augmentation de la diversité théorique n'entraîne pas nécessairement l'augmentation de la diversité réelle de la librairie. Il est important de faire la différence entre la *diversité théorique* d'une librairie, qui représente le nombre potentiel de molécules

différentes pouvant être codées par cette librairie, la *diversité clonale* d'une librairie, qui représente le nombre total de clones obtenus après le clonage de la librairie, et la *diversité réelle* d'une librairie, i.e. le nombre réel de clones contenant un insert codant pour une molécule de la librairie.

Comme on l'a dit au point 2.1, le principe d'une librairie est basé sur la probabilité d'isoler un événement rare à partir d'un très grand nombre d'événements qui ne devraient pas avoir, *a priori*, d'effet sur le système étudié. Pour qu'un événement rare puisse se produire et, à plus forte raison être détecté, il faut que la diversité réelle de la librairie soit assez grande. Prenons en exemple l'expérience menée par Xu et ses collaborateurs (83). Le but de cette expérience était d'identifier des peptides conférant une sensibilité au Taxol, un médicament anticancéreux, dans des cellules résistantes au Taxol. Dans cet article, on a identifié 4 peptides actifs à partir d'une banque de plus de 10 millions de peptides, pour en arriver à une fréquence de sélection d'environ un peptide biologiquement actif sur un million, fréquence pouvant être retrouvée dans plusieurs autres expériences du même type (4, 27, 55).

Si on espère isoler une molécule active à partir du criblage d'un système rapporteur avec une librairie, il est impératif de viser une diversité réelle d'au moins 10^6 peptides différents pour voir l'apparition d'un événement positif. Des données statistiques montrent de plus que pour espérer cribler l'ensemble de la diversité réelle d'une librairie, il faut en réalité la cribler au moins 3 fois pour être certain que tous les événements possibles ont eu lieu (53). Dans le cas qui nous concerne, nous visons l'atteinte d'une diversité finale de 1×10^8 nonapeptides différents dans le but d'isoler un nombre relativement important de séquences peptidiques biologiquement actives. La différence entre le clonage conventionnel, où l'obtention d'un seul clone positif est suffisante à la poursuite de l'expérience, et le clonage d'une librairie qui nécessite des millions de clones positifs différents, est immense. Il devient

donc indispensable de développer des méthodes de clonage aussi efficaces que possible, tant au niveau de la synthèse de la librairie, de la digestion des vecteurs et inserts, de la ligation et de la transformation des produits de ligation. Ces méthodes seront passées en revue plus loin dans ce mémoire.

2.3 Système de criblage et identification d'homologie

Si une grande diversité constitue la base fondamentale de toute librairie, le système de criblage, ou système rapporteur, n'en demeure pas moins primordial. Il serait en effet inutile de développer un système de vecteur et de clonage ultra-efficace s'il n'y avait pas de possibilité d'isoler l'événement rare tant recherché. L'utilisation d'une librairie pour étudier des interactions protéiques propres à un phénomène biologique est donc étroitement liée à la conception d'un système rapporteur (de criblage) efficace. Pour ce faire, le système rapporteur doit posséder plusieurs qualités. Premièrement, être assez discriminatoire pour éliminer la majeure partie des faux-positifs, c'est-à-dire les molécules agissant sur le système par le biais de mécanismes stochastiques ou non reliés au sujet de l'expérience. Deuxièmement, permettre la récupération de la molécule responsable de l'effet observé (que ce soit l'information génétique codant pour un peptide, la séquence d'un ADNc ou la séquence en acides aminés d'une protéine). La récupération de la séquence d'un peptide permet, d'une part, d'identifier des homologies avec d'autres protéines connues et, d'autre part, de réintroduire ce peptide dans le système afin de vérifier la reproductibilité de son effet.

Le système de criblage est d'autant plus important qu'il permet l'identification de partenaires potentiels des protéines qui y sont étudiées. La procédure générale consiste à isoler des clones provenant de peptides actifs puis à récupérer la séquence d'ADN codant pour ce peptide. Cette séquence permet ensuite de déterminer la séquence en acides aminés du peptide isolé. On peut ainsi identifier des homologies entre la séquence du peptide et des

protéines et supputer des interactions possibles entre la protéine étudiée et les partenaires identifiés. Jusqu'à récemment, l'approche classique d'identification de partenaires potentiels à la protéine étudiée passait par le criblage d'une banque de données comme BLAST (51) ou SWISS-PROT(1), qui permettent d'identifier des protéines à partir d'une séquence peptidique partielle. Cependant, le BLAST a ses limites : il ne permet en général d'identifier que des homologies parfaites entre deux séquences ou encore des homologies comportant des disparités entre certains acides aminés tout en conservant des propriétés biochimiques semblables. De plus, il ne permet pas d'identifier des homologies conformationnelles ou structurales entre deux protéines. Avec l'avènement de systèmes de criblage à grand volume (74), certains groupes ont développé des méthodes bioinformatiques puissantes permettant de mettre en évidence les interactions pertinentes à un système donné. Par exemple, le groupe de Sidhu (65) a croisé les données provenant du criblage d'une librairie de «phage display» avec celles provenant d'un système deux-hybrides en levures. Cette approche a permis de définir un réseau d'interactions fonctionnelles entre plusieurs partenaires protéiques des domaines SH3. Ces interactions théoriques ont ensuite été validées par les données de la littérature, ce qui a permis d'identifier 39 interactions très fortement susceptibles d'être vraies. Certaines de ces interactions ont ensuite été validées expérimentalement. Un autre groupe a développé un moteur de recherche permettant de cribler des bases de données de protéines en créant des matrices provenant des données d'un criblage de librairie de peptides (84). Ces matrices fournissent des critères de recherche basés sur un système de pointage attribué à chaque acide aminé d'une séquence peptidique. Les interactions les plus susceptibles d'être vraies sont ensuite signalées. Ces approches permettent de rendre la recherche de partenaires beaucoup plus efficace, moins aléatoire et font maintenant partie intégrante de toute expérience de criblage d'une librairie.

2.4 Bruit de fond

Une diversité maximale et un système de criblage sélectif sont les deux clés de voûte d'une expérience de criblage réussie. Cependant, il est impossible, malgré toutes les précautions prises dans la conception de notre système de criblage, d'éviter l'apparition de bruit de fond, i.e. d'événements aléatoires non-reliés au phénomène que l'on étudie mais qui vont quand même donner des résultats positifs lors du criblage préliminaire. Ces événements peuvent être divisés en deux catégories (53). La première regroupe les événements isolés en tant qu'événements positifs mais qui sont en réalité des manifestations transitoires qui ne sont pas vraiment dues à l'effet sur le système d'une molécule provenant de notre librairie. Ces événements aléatoires ne devraient théoriquement pas se reproduire si la population d'événements positifs est soumise à une nouvelle ronde de sélection. Ainsi, chaque ronde de sélection supplémentaire devrait enrichir la population en vrais événements positifs. Il se peut par contre que les événements isolés, bien qu'ils ne soient pas dus à l'effet d'une molécule provenant de la librairie, ne se produisent pas de façon aléatoire et soient dus soit à l'introduction de la librairie dans le système de criblage, soit au mode de sélection lui-même. Dans ce cas, il est peu probable que des rondes de sélection supplémentaires amènent une augmentation du taux de vrais événements positifs. Il devient alors très difficile d'isoler des molécules ayant un effet réel sur le système à l'étude. Cela nous ramène à l'importance primordiale de mettre sur pied un système de criblage qui considère les possibilités d'isoler des événements non-significatifs et qui vise à les éliminer.

2.5 Présentation

Bien qu'il ne soit pas aussi critique que la diversité ou le système de criblage, le contexte dans lequel est présenté le peptide influence considérablement le type d'interactions qui seront détectées. De façon simple, on peut dire qu'il y

a deux façons de présenter un peptide dans une librairie : il peut être exprimé seul et agir librement au sein de la cellule ou encore être exprimé dans le cadre d'une fusion avec d'autres éléments protéiques (82). Dans le cas où le peptide est exprimé librement, une cassette d'expression typique contiendra un codon d'initiation de transcription, la séquence dégénérée codant pour les peptides aléatoires ainsi qu'un codon de terminaison de la traduction permettant de cesser la synthèse des peptides dans la cellule. Certains peptides exprimés librement sont aussi fusionnés à de courts domaines terminaux ayant la séquence consensus EFLIVIKS leur permettant d'adopter une conformation en forme de « U » suite à la dimérisation de ces domaines (5, 83). Ces peptides, puisqu'ils sont exprimés librement, ont généralement accès à tous les compartiments intracellulaires et peuvent jusqu'à un certain point avoir les caractéristiques de petites molécules organiques (83). On peut de plus prétendre que l'énergie de liaison d'un peptide libre (qui peut donc adopter une multitude de conformations tridimensionnelles) est plus grande que celle d'un peptide qui serait restreint dans une structure tridimensionnelle fixe, ce qui pourrait limiter la détection d'interactions de faible affinité (82).

Dans le cas où le peptide est exprimé dans le cadre d'une fusion avec d'autres éléments protéiques, la personne qui fait la conception de la librairie fait face à plusieurs choix. La fusion des peptides avec une protéine peut restreindre la localisation cellulaire, l'accessibilité du peptide, le taux d'expression ainsi que la stabilité de la librairie (43). Il a également été montré que le type de protéine supportant le peptide peut influencer le type d'interactions et de partenaires de ce peptide (43). Il faut être très consciencieux dans le choix de la protéine de fusion pour éviter l'introduction d'un biais de sélection non-désiré. Si on choisit la fusion, celle-ci peut être faite soit à une extrémité d'une protéine inerte, comme par exemple la glutathione-S-transférase (GST) (86), soit au sein d'une structure secondaire externe d'une protéine bactérienne comme la thiorédoxine A (4, 43), la nucléase staphylococcique S1 (55), la GFP (green fluorescent protein) (58) ou encore en

fusion avec un marqueur de sélection (24) . Le choix de la protéine de fusion est laissé à l'imagination du concepteur de la librairie. Il est cependant important de considérer les points suivants : elle doit le plus possible être inerte dans le système rapporteur utilisé, pour éviter de créer du bruit de fond non aléatoire dû au « transporteur » du peptide et non au peptide lui-même. La protéine de fusion doit permettre une exposition de la séquence peptidique afin de favoriser l'interférence de ce peptide avec les protéines impliquées dans le mécanisme étudié. Elle doit également être le plus stable possible, être exprimée fortement sans être toxique, et, dans le cas d'une protéine de fusion qui possède une activité intrinsèque autre que celle conférée par le peptide de la librairie (par exemple la fluorescence de GFP ou encore la résistance à un antibiotique conférée par un marqueur de résistance), la protéine de fusion doit pouvoir conserver ses propriétés originales.

3. SURVOL DES MÉTHODES EXISTENTES

3.1 «phage display»

Le «phage display», i.e. l'exposition d'un peptide généré aléatoirement fusionné à une protéine de surface d'un phage filamenteux, a été développé au milieu des années 80 par G.P. Smith (67). Le principe du «phage display» consiste à construire une banque de vecteurs permettant l'expression d'un peptide à l'extérieur d'un phage en modifiant son génome pour y introduire la séquence codant pour ce peptide. Celui-ci sera exprimé en fusion avec une protéine d'enveloppe du phage. On peut alors étudier les interactions entre les peptides exposés à la surface du phage et une protéine cible fixée à un support plastique au cours d'expériences de criblage. Ces expériences consistent à mettre en présence une banque de phages exprimant différents peptides à leurs surface et une protéine pour laquelle on veut identifier un ou plusieurs motifs de liaison (figure 1). Après une étape d'exposition des phages à la protéine cible et un lavage, seuls les phages exprimant des

peptides liant la protéine cible (avec des affinités variables) sont conservés puis amplifiés. On effectue ainsi plusieurs rondes de sélection pour garder seulement les phages liant spécifiquement la protéine d'intérêt. La séquence d'ADN codant pour chaque motif peptidique est analysée et on détermine des motifs consensus de liaison à la protéine étudiée (54). Cette approche expérimentale a permis d'identifier un grand nombre d'interactions entre différentes protéines ainsi que des motifs de reconnaissance d'anticorps. La littérature comporte une pléthore d'études ayant validé et utilisé le concept de «phage display». Une étude menée en 1990 a permis d'identifier des épitopes reconnus par un anticorps dirigé contre la portion N-terminale de la beta-endorphine (13). Une autre étude du début des années 90 a permis d'identifier des motifs de liaisons peptidiques spécifiques à la streptavidine (15). Il a été possible de déterminer des motifs de liaison à la protéine accessoire Vpr du VIH-1 (6). Plus récemment, le «phage display» a été utilisé pour déterminer la séquence consensus optimale pour la phosphorylation du proto-oncogène AKT (56). Ces études ne représentent qu'une infime partie des applications du «phage display». Même si on est présentement en train de développer des systèmes permettant l'expression des vecteurs de «phage display» à l'intérieur de cellules de mammifères (46), la méthode a encore le défaut majeur d'utiliser une méthode de criblage *in vitro*, ce qui pose des problèmes non négligeables lorsque vient le temps de transposer les données obtenues dans un système physiologique. Les peptides sélectionnés par «phage display» ne seront pas nécessairement stables lorsqu'exprimés dans des cellules. De plus, il devient nécessaire de changer le contexte dans lequel seront exprimés les phages si on veut les introduire dans des cellules de mammifères. L'affinité des peptides pour leur cible ne sera pas forcément la même dans un contexte différent, d'autant plus que toutes les modifications post-traductionnelles propres aux cellules de mammifères ne sont pas effectuées dans les bactéries où sont produits les phages. De plus, avec le «phage display», le peptide reconnaît la protéine cible dans un contexte isolé. *In vivo*, les protéines fonctionnelles font souvent

partie de complexes multiprotéiques. Certaines interfaces ne sont donc pas exposées *in vivo*, et certains épitopes ne sont formés que dans le contexte de ces complexes multiprotéiques. La transposition d'interactions détectées par «phage display» dans un système *in vivo* ne se fait donc pas sans heurts, et l'identification *in vitro* d'interactions entre une protéine et un peptide n'est pas une garantie qu'il y aura une implication fonctionnelle dans un système *in vivo*.

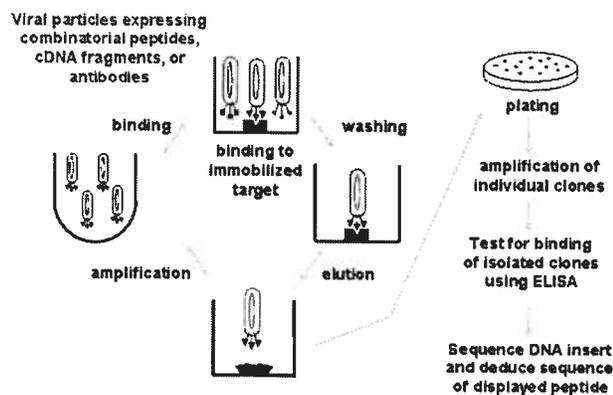


Figure 1. Principe du «phage display». Tiré de (41)

3.2 «ribosome display»

Le «ribosome display» est une méthode permettant de sélectionner des peptides ayant de très fortes affinités pour une protéine cible. Cette approche est totalement effectuée *in vitro* et constitue un raffinement des méthodes existantes de production d'anticorps monoclonaux. Le principe du «ribosome display» a été adapté d'un type d'expérience faites par Tuerk et Gold au milieu des années 1990, le SELEX (systematic evolution of ligands by exponential enrichment) (76). Cette méthode cherchait à sélectionner des ARN liant des protéines, chaque cycle de sélection permettant d'amplifier ces ARN par RT-PCR. Le «ribosome display» utilise ce concept pour isoler des fragment variables d'anticorps (provenant d'une banque aléatoire)

possédant une très grande affinité (de l'ordre du picomolaire) avec une protéine fixe ou en solution. La figure 2 schématise le fonctionnement du «ribosome display». Un ADN codant pour la portion variable d'un anticorps est transcrit *in vitro*. Les ARNm comprennent une partie C-terminale codant pour une trentaine d'acides aminés inertes et dénuée de codon stop. Lors de l'étape de traduction *in vitro*, le ribosome s'arrête à la fin de l'ARNm au lieu de s'en détacher (puisque'il n'y a pas de codon stop), liant ainsi la protéine néo-synthétisée à son ARNm. Ce complexe est mis en présence de la protéine cible. Suite à une étape de lavage où sont éliminés les complexes qui ne lient pas la protéine cible, les ARNm sont élués du complexe, purifiés et amplifiés par RT-PCR. On peut ainsi procéder à plusieurs rondes d'amplification consécutives et isoler des fragments d'anticorps possédant une très grande affinité pour la protéine cible.

Cette méthode pourrait aussi être appliquée à l'identification d'interactions entre des peptides et une protéine cible comme on le fait avec le «phage display». L'avantage majeur du «ribosome display» réside dans le fait que la diversité de la librairie n'est en aucun cas affectée par les limites imposées par l'étape de transformation. La banque de vecteurs ligés peut être transcrite immédiatement, ce qui laisse envisager l'utilisation de banques extrêmement diversifiées. Le «ribosome display» possède toutefois tous les défauts du «phage display» en plus d'être très difficilement transposable dans un contexte *in vivo*. Tout ce qui concerne l'identification de partenaires fonctionnels d'une protéine peut difficilement être fait à l'aide de cette méthode.

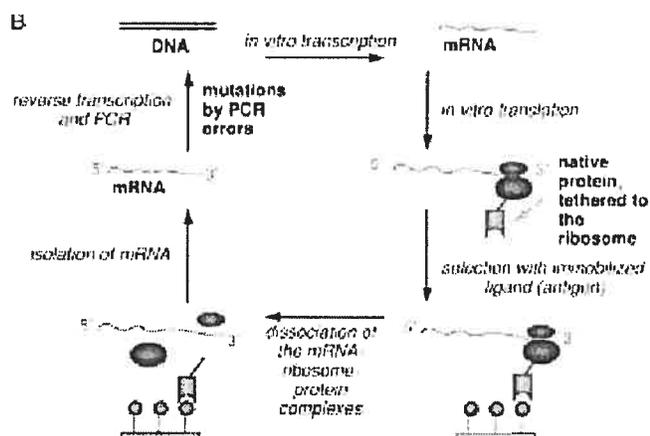


Figure 2. «ribosome display», fonctionnement. Tiré de (63)

3.3 Système deux hybrides en levure

Le système deux hybrides en levure a été développé pour identifier des interactions spécifiques entre protéines, et, plus particulièrement, pour cribler des banques d'ADNc directement dans des levures (10, 20). Le principe du système deux hybrides est le suivant : lorsqu'il y a interaction entre une protéine cible et un membre d'une librairie d'ADNc ou de peptides, un gène rapporteur est exprimé, ce qui permet de récupérer la séquence du partenaire et de caractériser l'interaction plus en profondeur. Pour ce faire, la protéine cible est fusionnée à un domaine de liaison à l'ADN provenant d'un facteur de transcription, l'un des plus utilisés étant le facteur de transcription Gal4. Les membres de la librairie sont pour leur part fusionnés à un domaine d'activation de la transcription. Lorsqu'il y a interaction entre la protéine cible et un partenaire de la banque, le complexe ainsi formé permet d'amener le facteur d'activation de la transcription à proximité du promoteur d'un gène rapporteur. Ce gène est alors exprimé et permet d'isoler les levures dans lesquelles l'interaction s'est produite (figure 3). Ce système a été adapté avec succès pour pouvoir exprimer des librairies de peptides au lieu de banques d'ADNc et quelques groupes ont réussi à obtenir des résultats intéressants. Par exemple, le groupe de Cohen et ses collègues ont identifié

un inhibiteur de la protéine de régulation du cycle cellulaire cdk2 (12). D'autres ont réussi à identifier des séquences consensus de liaison à la protéine Rb (retinoblastoma protein) (85).

Même si le système deux-hybrides en levure adapté pour l'expression de bibliothèques de peptides a montré qu'il pouvait être utilisé pour identifier des interactions protéine-protéine intéressantes, il a toutefois ses limites. Premièrement, il permet d'identifier des interactions physiques entre un peptide et une protéine d'intérêt, mais il ne permet pas dans un premier temps d'étendre cette interaction à une implication fonctionnelle. Cela est dû au fait que le système de criblage utilisé est indépendant de la fonction réelle de la protéine étudiée. Deuxièmement, l'identification de partenaires en levure peut ne pas être transposable en cellules de mammifères puisque les modifications post-traductionnelles n'y sont pas les mêmes. Les voies de signalisation et de transport sont également différentes. De ce point de vue, notre bibliothèque permet de répondre plus directement et avec plus de certitude à des questions concernant l'interaction fonctionnelle d'une protéine avec ses partenaires cellulaires.

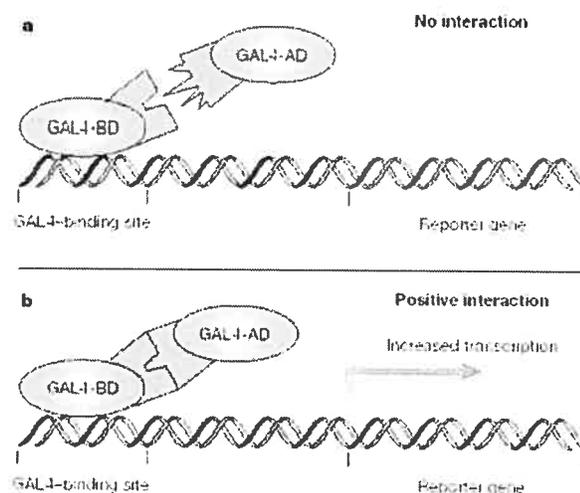


Figure 3. Fonctionnement du système deux hybrides en levure. Tiré de (34)

Les approches de librairies utilisées actuellement ont toutes des applications spécifiques pour lesquelles elles sont appropriées. Cependant, pour identifier des interactions fonctionnelles entre protéines directement dans des cellules de mammifères, nous croyons qu'une approche de librairie de nonapeptides pourra permettre de pallier aux désavantages inhérents à chacune de ces méthodes.

4. GST

La GST fait partie d'un groupe d'enzymes (E.C. 2.5.1.18) (PM entre 24 et 28 kDa) qui protègent les cellules des dommages causés par les produits chimiques toxiques et les métabolites dus au stress environnemental (16). Ces enzymes détoxifient les produits toxiques hydrophobes par l'ajout d'un résidu glutathione, les rendant ainsi plus solubles et facilitant leur excrétion sous la forme d'un résidu merthiolate (81). Les GST sont des protéines ubiquitaires soit associées aux membranes, soit exprimées dans le cytosol de la plupart des organismes aérobiques (66). On les retrouve chez les plantes, les poissons, les mycètes, les insectes, les levures ainsi que chez certaines bactéries (voir réf (81) pour une revue). Il existe 5 classes de GST différentes : alpha, mu, pi, theta ainsi qu'une classe de GST microsomaux (81). Ces enzymes sont connues pour former des homodimères ainsi que des hétérodimères entre molécules différentes appartenant à la même classe. La GST est très abondante dans les tissus de rat, de souris et de l'humain. Plus particulièrement, les GST de classe alpha sont retrouvées abondamment dans les tissus de reins humains d'où proviennent les lignées cellulaires (293, 293T) utilisées dans ce mémoire (70, 71).

La fin des années 1980 a vu naître les applications de purification de protéines à partir de fusion avec la GST. Nous avons décidé d'utiliser ce système,

développé par la compagnie Amersham, afin d'exprimer notre librairie dans des cellules de mammifères. Le principe de la méthode consiste à fusionner une protéine ou un peptide en C-terminal de GST (le plus souvent dans un vecteur de la série pGEX pour l'expression en bactéries ou dans un vecteur d'expression eucaryote) et à l'exprimer soit dans la bactérie *Escherichia coli*, soit en levures ou en cellules de mammifères. Les cellules exprimant la protéine chimérique sont ensuite lysées et la fusion GST récupérée par chromatographie d'affinité en utilisant des billes de sépharose auxquelles ont été liées covalamment des molécules de glutathione. Cette chromatographie permet d'une part de récupérer la protéine de fusion exprimée dans les cellules, et, d'autre part, d'identifier certaines interactions protéines-protéines en copurifiant des partenaires de la protéine de fusion. Dès 1991, le groupe de Kaelin a réussi à identifier des partenaires liant la protéine Rb en faisant des « GST pull-down » (40). Plus récemment, Vodicka et al ont purifié la protéine accessoire Vpr fusionnée à GST à partir de lysats de levures (77). Dans notre laboratoire, il a été possible d'identifier des peptides fusionnés à GST qui bloquaient la cytotoxicité causée par la protéine accessoire du virus de l'immunodéficience humaine (VIH-1) Vpr dans *Sacharomyces cerevisiae* (86).

La fusion de protéines avec GST comporte des avantages certains : il existe des anticorps commerciaux très efficaces permettant d'analyser l'expression de la protéine de fusion par immunobuvardage. Comme on l'a dit précédemment, les systèmes de purification par chromatographie d'affinité permettent de purifier les protéines fusionnées avec GST en plus de rendre possible l'identification d'interactions protéine-protéine par copurification. De plus, étant donné son faible poids moléculaire, la GST a accès à tous les compartiments cellulaires par simple diffusion, ce qui permet aux peptides qui y sont fusionnés d'avoir accès à un plus grand éventail de cibles possibles.

5. VECTEURS

Le vecteur dans lequel sera cloné la librairie est un facteur important qui influencera sa diversité finale (à cause de son influence sur l'efficacité de transformation), la facilité du transfert de cette librairie dans le système rapporteur (et donc la possibilité d'utiliser la diversité de la librairie au maximum), la méthode de récupération de la séquence du peptide ainsi que la versatilité de l'utilisation de cette librairie. La section suivante décrit les principales caractéristiques des vecteurs ayant été utilisés au cours de cette étude.

5.1 Vecteur pCEP4

Le vecteur pCEP4 est un vecteur épisomal disponible commercialement chez la compagnie INVITROGEN (Annexe 1). Toute l'approche du clonage de la librairie dans le vecteur épisomal pCEP4 est basée sur 3 qualités principales de ce vecteur. Premièrement, ce vecteur possède une origine de réplication du virus Epstein-Barr qui lui permet de se répliquer de façon épisomale, i.e. à l'extérieur du chromosome cellulaire. Cette caractéristique permet l'isolement direct du vecteur et la récupération de la séquence du peptide sans passer par des étapes d'extraction d'ADN génomique et de PCR. La réplication épisomale confère un avantage de plus à ce vecteur puisque son expression ne sera pas influencée par le site d'intégration lors d'une expression stable. Deuxièmement, la présence d'un gène de résistance à l'hygromycine permet d'établir une lignée cellulaire exprimant de façon stable la librairie de peptides, ce qui permet d'effectuer les étapes de sélection à partir d'une population exprimant uniformément la librairie. Finalement, ce vecteur est conçu pour exprimer fortement le gène qui y est cloné, en raison du promoteur viral CMV. La structure de la librairie telle que clonée dans le vecteur épisomal pCEP4 est schématisée dans la figure 4.

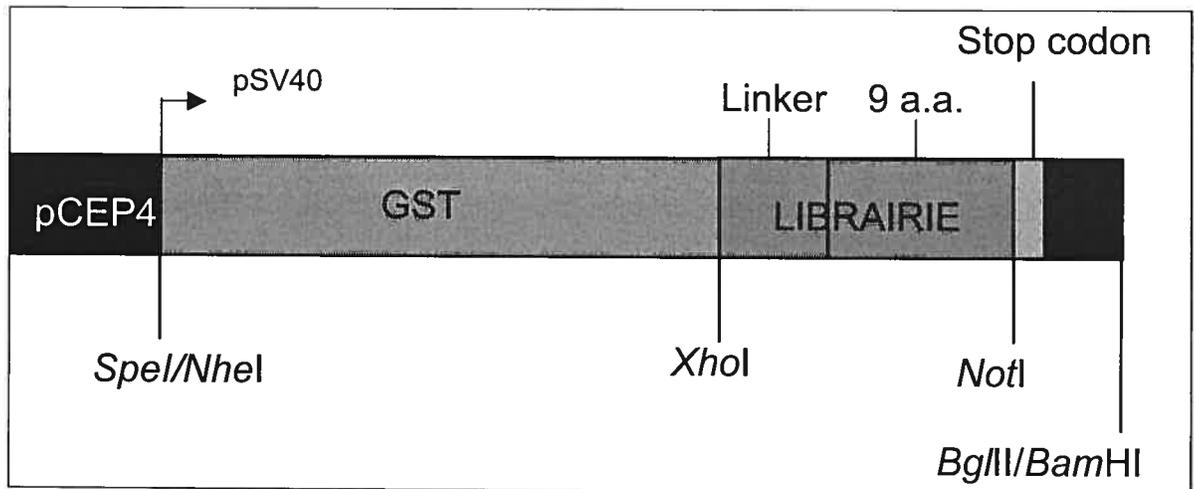


Figure 4. Organisation de la librairie de nonapeptides dans le vecteur pCEP4

5.2 Principe et diversité des vecteurs rétroviraux

Un des facteurs clé dans la réussite d'une expérience de criblage de librairie réside dans l'obtention de la plus grande diversité réelle possible. Afin de prendre avantage d'une librairie très diversifiée, il devient important d'utiliser un système qui permettra l'introduction et l'expression de la librairie dans le plus grand nombre possible de cellules. L'utilisation du vecteur pCEP4 peut suffire à montrer la faisabilité d'un système de librairie de nonapeptides en cellules de mammifères mais comme il doit être transfecté, le spectre des cellules qu'il peut infecter est réduit, et par le fait même ses applications futures. La transposition de la librairie dans un système de vecteur rétroviral devient donc nécessaire pour étendre son application à d'autres études.

5.2.1 Rétrovirus

Les rétrovirus sont des virus à ARN à polarité positive, dont le cycle de répllication passe par un intermédiaire d'ADN double-brin (7). Le génome des

rétrovirus comporte trois gènes essentiels : *gag*, *pol* et *env*. Ces trois gènes sont communs à tous les rétrovirus, bien que les lentivirus possèdent en plus des gènes dits accessoires. Comme nous travaillons dans un système basé sur le virus de la leucémie murine (MuLV), nous restreindrons notre description aux rétrovirus simples. Chez MuLV, les gènes *gag* et *pol* sont dans le même cadre de lecture mais la traduction du gène *pol* ne se fait que dans 5 à 10% des cas lorsqu'il y a continuation de la traduction après le codon stop placé à la fin de *gag* (87). Cette particularité permet de maintenir un ratio molaire idéal entre les protéines issues du gène *gag* et celles provenant de *pol*. Le gène *gag* du MuLV code pour les protéines structurales du virus, la matrice (MA), la capsid (CA) et la nucléocapsid (NC). Le gène *pol* pour sa part, code pour les enzymes nécessaires à la réplication du virus : la protéase (PR), qui clive certaines polyprotéines (notamment la polyprotéine *gag-pol* issue d'une continuation de la traduction après le codon de terminaison de *gag*) et assure la maturation des protéines d'enveloppe, la reverse-transcriptase (RT) qui assure le transfert de l'information génétique contenue dans l'ARN simple-brin à un ADNc double-brin (provirus) et l'intégrase (IN), qui permet l'intégration aléatoire du provirus dans le génome de la cellule infectée. Ce sont ces protéines qui permettent aux rétrovirus de s'intégrer de façon stable dans le génome cellulaire et d'être transmis aux générations suivantes, ce qui en fait d'excellents vecteurs pour une librairie comme la nôtre. On peut ainsi compter sur une population cellulaire qui exprime la librairie de façon stable au cours des divisions cellulaires. Finalement, le gène *env* code pour la polyprotéine d'enveloppe, Env, qui est clivée par la protéase au cours de la maturation du virus pour donner les sous-unités transmembranaire (TM) et de surface (SU). Ces protéines jouent un rôle important dans l'attachement du virus à la cellule hôte. L'enveloppe naturelle des différents virus MuLV leur confère une spécificité différente selon les variantes. On parle de virus amphotropiques lorsque l'enveloppe interagit avec des récepteurs membranaires de phosphate inorganique et de virus écotropiques lorsque l'enveloppe interagit avec plusieurs récepteurs d'acides aminés cationiques.

Les vecteurs rétroviraux basés sur MuLV ont retenu l'organisation génomique du virus sauvage, à l'exception que tous les gènes essentiels ont été remplacés par un gène d'intérêt que l'on cherche à exprimer dans les cellules infectées. La délétion des gènes essentiels empêche le virus de se répliquer plus d'une fois, ce qui permet d'infecter des cellules sans rester aux prises avec des virus se répliquant activement. Dans le cas qui nous concerne, le vecteur pBABE PURO (7), sur la base duquel a été construit le vecteur pBABE HAGIZ, n'a conservé du provirus original que les éléments nécessaires à l'encapsidation de l'ARN, à une reverse-transcription et une intégration optimale (44): les 2 LTRs (long terminal repeats), le PPT (polypurine tract) et le PBS (primer binding site) essentiels à l'étape de la reverse-transcription, et la séquence d'encapsidation de l'ARN psi (qui comprend le gène *gag* tronqué et muté à l'ATG). Une partie de la séquence du gène *gag* a été conservée puisqu'elle semble favoriser l'encapsidation du génome ARN dans la particule virale.(3)

Comme ils ont perdu leurs gènes essentiels, les vecteurs rétroviraux ne sont plus capables de se répliquer de façon autonome. Il faut leur procurer les protéines nécessaires à la formation de la particule virale et à sa réplication dans la cellule permissive. Afin de créer des particules virales qui contiendront l'information génétique qu'on a introduite dans le vecteur rétroviral, on transfecte le vecteur rétroviral dans des cellules qui expriment les gènes essentiels manquants. Pour ce faire, il existe des lignées cellulaires dites d'encapsidation (48) qui ont été modifiées pour exprimer de façon stable les produits des gènes essentiels. De cette manière, il devient possible de produire des particules virales infectieuses contenant l'information génétique de la librairie. Après transfection du vecteur rétroviral dans une lignée d'encapsidation, le surnageant contenant les virus codant pour la librairie est récolté et utilisé pour infecter les cellules cibles du système rapporteur. Il est possible de remplacer l'expresseur d'enveloppe de MuLV par un expresseur

d'enveloppe provenant d'un autre type de virus (par exemple la glycoprotéine G de l'enveloppe du virus de la stomatite vésiculaire (VSV)), par un procédé appelé pseudotypage (8, 38). Ce changement d'enveloppe permet de modifier le tropisme des particules virales produites, augmente la stabilité des particules virales et élargit la diversité des cellules que l'on peut infecter et, par la même occasion, des mécanismes qu'on peut étudier.

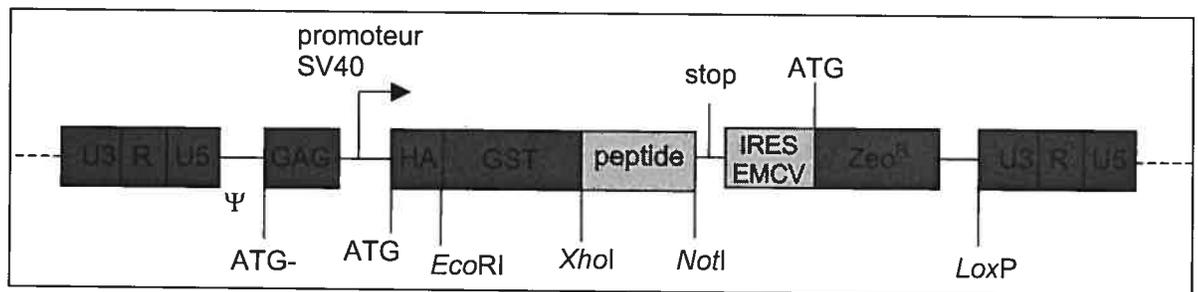


Figure 5. Cassette d'expression de la librairie de nonapeptides dans le vecteur pBABE HAGIZ. HA : épitope HA (hémagglutinine du virus de l'influenza), IRES EMCV : Internal Ribosome Entry Site du virus de l'encéphalomyocardite.

5.2.2 Vecteur rétroviral

La cassette d'expression de la librairie dans le vecteur rétroviral est légèrement différente de celle utilisée dans pCEP4, puisqu'elle comprend des éléments nouveaux qui rendent l'utilisation de la librairie plus facile et plus polyvalente. La figure 5 schématise la librairie telle que clonée dans le vecteur pBABE HAGIZ. Le vecteur rétroviral utilisé pour le clonage de notre librairie de nonapeptides provient du virus de la leucémie murine (MuLV). Ce vecteur (pBABE HAZ) gracieusement fourni par Stephen Goff (24) est une

évolution d'un vecteur largement utilisé, pBABE PURO, qui a été modifié de façon à conférer la résistance à la zéocyne aux cellules infectées. Une séquence LoxP comprise dans la partie U3 du LTR 3' du provirus permet, lorsque mise en présence de la recombinase Cre de *E. coli*, de retirer la séquence du provirus du génome des cellules infectées (45). La transcription inverse du génome viral engendre en effet une duplication de la séquence LoxP, qui encadre la séquence codant pour la librairie dans le provirus. Il devient alors possible de confirmer que l'effet observé lors de la sélection du peptide est bel et bien dû à ce peptide et non à un autre mécanisme inhibiteur puisque la fusion GST-peptide ne sera plus exprimée. Ce vecteur comporte les éléments essentiels à un vecteur rétroviral : 2 LTR, une séquence Ψ (psi) comprenant un gène *gag* tronqué et muté au niveau de l'ATG qui favorise l'encapsidation du transcrit viral, un promoteur précoce de SV40 permettant l'expression forte de la séquence de la librairie. Comme le vecteur original était conçu pour faire une fusion entre la librairie et le gène de résistance à la zéocyne, nous l'avons modifié en ajoutant une séquence d'entrée interne des ribosomes (IRES) du virus EMCV (19) qui rend possible l'expression bicistronique de la librairie et du gène de résistance à la zéocyne. Cette approche a été adoptée afin d'éviter de limiter l'accès du nonapeptide à ses partenaires cellulaires en le « coinçant » entre deux éléments protéiques beaucoup plus gros. Ce vecteur permettra de transduire une grande variété de cellules différentes et donc, d'étudier plusieurs types d'interactions différentes. Malgré ce net avantage, l'expression de la librairie à partir de ce vecteur est plus sujette à subir les effets de la position d'intégration de la cassette rétrovirale, et certains aspects de la méthode de sélection de peptides inhibiteurs pourraient être compromis.

6. NOTRE LIBRAIRIE

Notre librairie se veut une évolution de la librairie d'hexapeptides exprimée en levures utilisée dans notre laboratoire par Xiao Jian Yao et al (86). Elle se

présente sous la forme d'un gène de fusion encodant plusieurs éléments structuraux et régulateurs importants. La protéine structurale, le « squelette » de notre librairie, est la GST. La librairie proprement dite est exprimée sous la forme de 9 acides aminés générés aléatoirement en fusion C-terminale de GST. Pour assurer une certaine flexibilité au nonapeptide, nous avons inséré un linker de 6 acides aminés (SGSGSG) entre GST et la librairie. La traduction est stoppée par deux codons stop en 3' de la séquence codant pour les 9 acides aminés dégénérés. Les sites *XhoI* et *NotI* qu'on voit sur la figure 4 ont servi au clonage de la librairie.

La première approche que nous avons utilisée consiste à cloner un oligonucléotide codant pour 9 acides aminés dans le vecteur pCEP4 décrit dans la section précédente. Le gène de la glutathione-S-transférase avait d'abord été cloné dans le vecteur afin de servir de protéine de support à notre librairie de nonapeptides. La deuxième approche consiste à cloner la librairie en fusion C-terminale de GST dans le vecteur rétroviral pBABE HAGIZ, qui exprime le gène de résistance à la zéocyne à partir d'une séquence IRES contenue dans le même ARNm.

7. SYSTÈME DE VALIDATION DE LA LIBRAIRIE

Le système de validation développé pour valider le concept de sélection génétique de nonapeptides en cellules de mammifères repose sur la capacité de la prodrogue ganciclovir à tuer les cellules exprimant le gène de la thymidine kinase du virus *Herpes simplex*.

7.1 Thymidine kinase du virus *Herpes simplex*

La thymidine kinase (TK) est une enzyme codée par le gène TK du virus *Herpes simplex*. Cette enzyme, identifiée par Kit et ses collaborateurs en 1965 (42), possède une spécificité assez large et peut phosphoryler autant les

purines, les pyrimidines que les analogues de bases azotées (22). C'est cette propriété de la TK qui permet son utilisation en combinaison avec le ganciclovir (GCV), un analogue de base, dans des approches de thérapie génique pour le cancer. Le principe de cette approche consiste à transduire, i.e. infecter des cellules cancéreuses à l'aide d'un vecteur rétroviral pour y introduire le gène de la TK et à maintenir son expression stable à l'aide d'un marqueur de sélection. Ces cellules, qui expriment dorénavant la TK, deviennent sensibles au GCV et peuvent ainsi être éliminées (50) par un traitement au GCV. Notre objectif est d'utiliser ce système de gène suicide pour montrer qu'il est possible d'identifier des peptides inhibiteurs de la TK virale à partir de cellules 293TK+ exprimant notre librairie de nonapeptides de façon stable. Plusieurs groupes ont utilisé le système HSVTK/GCV pour tenter d'enrayer la progression de certains types de cancer. Les groupes de Qasim et Garin ont appliqué la stratégie au phénomène du rejet du greffon (graft versus host disease) dans les cas de transfert de moelle osseuse (25, 60). Ils ont montré que les cellules T immunocompétentes causant le GVHD (graft versus host disease) transférées avec la greffe de cellules souches pouvaient être transduites avec la TK pour permettre leur élimination en présence de GCV une fois l'opération terminée, ce qui éliminait la GVHD chez la plupart des patients. D'autres groupes ont montré le potentiel de ce système dans la régression de carcinomes humains (57) et de métastases hépatiques (9).

Notre système de sélection utilise des cellules 293 (ATCC # CRL-1573) exprimant de façon stable une fusion entre la protéine GFP et la protéine TK de HSV. Un article paru récemment a analysé l'influence des fusions de protéines fluorescentes à la thymidine kinase sur la localisation intracellulaire (68). Il s'est avéré que, dans le cas d'une fusion entre la GFP et la TK dans une lignée de cellules 293, la localisation type intranucléaire de la TK n'était pas affectée par la fusion. Ces considérations seront importantes dans l'analyse des homologues identifiées après le criblage de la librairie dans le système 293TK+.

7.2 Systèmes de sélection HSV TK/GCV

Le ganciclovir (9-[(1,3-dihydroxy-2-propoxy)méthyl]guanine) est un analogue de la guanine utilisé comme agent antiviral (61). Bien qu'il soit utilisé depuis longtemps en combinaison avec la TK de HSV dans des stratégies de thérapie génique antitumorale, le mécanisme d'action du GCV reste mal défini. Plusieurs études contradictoires ont été publiées sur le sujet. De façon générale, on semble s'entendre pour dire que le GCV subit une première phosphorylation dans la cellule par la TK virale et est ensuite di et tri-phosphorylé par les kinases cellulaires (figure 6). L'analogue de base tri-phosphate serait ensuite intégré dans les chaînes d'ADN en élongation. La première manifestation de l'administration de GCV à des cellules TK+ semble être une activation de la voie de signalisation des MAPK (79) et ce, dès 1 heure post-exposition. Le mécanisme par lequel le GCV tue les cellules ne semble pas être relié à la terminaison de l'élongation des chaînes d'ADN. En effet, le GCV n'est pas un terminateur d'élongation au sens strict puisqu'il possède un groupement 3'OH libre (80). Il semblerait donc que l'activation des voies de signalisation des MAPK mène d'une façon encore inconnue mais indépendante de la terminaison de l'élongation des chaînes d'ADN à la mort cellulaire. Différents groupes ont également démontré l'apparition de dommages à l'ADN sous forme de cassures double-brin (73, 78). Ces groupes attribuent l'effet toxique du GCV soit à des mécanismes apoptotiques indépendants de p53 faisant appel à la voie mitochondriale (clivage de Bcl-2 par les caspases 3 et 9) (73), soit à des mécanismes dépendants de p53 et faisant appel à la voie apoptotique impliquant la translocation à la membrane de Fas/CD95 (78). Ces études, bien qu'elles soient contradictoires, vont aider à déterminer si les peptides identifiés présentent des homologies avec des protéines intervenant dans une des voies décrites plus haut.

La toxicité du GCV n'affecte pas seulement les cellules l'ayant absorbé mais également un bon nombre de cellules voisines. Cet effet, appelé « effet bystander », a été décrit chez de multiples lignées cellulaires et constitue un élément important de l'approche de thérapie génique basée sur la combinaison HSVTK/GCV (17, 21, 75). Il semblerait que le transfert des métabolites modifiés par la TK se fasse par l'intermédiaire des « gap junction » (17). La littérature rapporte que, à cause de cet effet, seulement 10% des cellules d'une tumeur doivent exprimer la TK pour qu'il y ait diminution effective de la masse tumorale (9, 21). L'effet « bystander » doit être considéré dans notre système de criblage puisqu'il met en jeu des mécanismes différents de cytotoxicité reliés au système HSVTK/GCV.

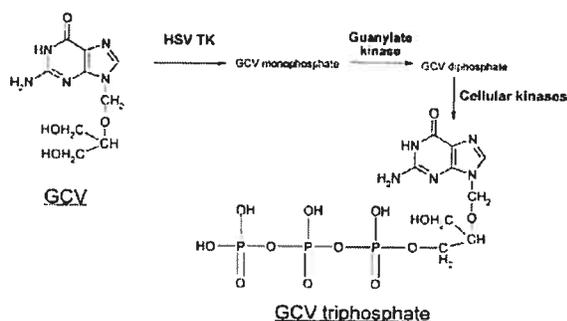


Figure 6. Structure et métabolisme du ganciclovir (GCV) tiré de (32).

8. OBJECTIFS DU PROJET

L'objectif des expériences dont fait état ce mémoire était de mettre sur pied une librairie de nonapeptides aléatoire pouvant être exprimée dans des cellules de mammifères. Un système de criblage génétique a été conçu et testé afin de pouvoir valider le principe selon lequel la librairie peut être utilisée pour identifier des interactions fonctionnelles protéine-protéine *in vivo*. Deux approches ont été tentées pour construire la librairie de nonapeptides : l'une consistait à cloner la librairie dans le vecteur épisomal pCEP4 et l'autre à la cloner dans le vecteur pBABE HAGIZ, la dernière approche étant une solution aux problèmes rencontrés au cours de la première. Ce mémoire rapporte les expériences qui ont été menées dans le but de construire une librairie de peptides pouvant être exprimée en cellules de mammifère. On y discutera des différents problèmes et avantages inhérents à chaque méthode en plus de suggérer des améliorations et des possibilités d'application de cette technologie.

SECTION MATÉRIEL ET MÉTHODES

1. BIOLOGIE MOLÉCULAIRE :

Considérations générales

Clonage :

À moins d'indication contraire, les digestions de vecteurs décrites dans cette section sont faites avec 10 unités d'enzyme dans un volume de 20 µl à 37°C pendant 3 heures et suivies d'une déphosphorylation par une unité de SAP (shrimp alkaline phosphatase) (USB Scientific) à 37°C, 1 heure. Les produits de PCR sont amplifiés sur une machine PJ Cyler (MJ Research) selon le programme suivant : 1) 94°C 2 min, 2) 94°C 1 min, 3) 58°C 1:30 min, 4) 72°C 1:30 min, 5) retour à l'étape #2 29 fois, 6) 72°C 5 min, 7) conservation à 4°C. Les produits de PCR sont toujours digérés toute la nuit à 37°C dans le même volume et avec les mêmes quantités d'enzyme que les vecteurs. Les produits de digestion sont purifiés sur gel d'agarose LMP (Low Melting Point Agarose, INVITROGEN) 1 %, extraits avec le kit QIAquick Gel Extraction Kit (QIAGEN). La ligation s'effectue dans un volume de 20 µl dans le tampon de ligation T4 DNA Ligase Buffer de NEB (New England Biolabs), en présence de 20 U de T4 DNA Ligase (NEB), toute la nuit à 16°C, puis la réaction est inactivée à 70°C pendant 20 minutes. Les produits de ligation sont transformés dans des bactéries chimiocompétentes *E. coli* AG-1. Les colonies issues de la sélection sur pétris LB ampicilline 100 µg/ml sont cultivées dans du milieu LB, O/N à 37°C, lysées selon la méthode standard de lyse alcaline (62) et analysées par digestion enzymatique et électrophorèse sur gel d'agarose. Les clones positifs sont amplifiés dans 1 litre de milieu LB pendant 18-24h à 37°C, incubés en présence de chloramphénicol (17µg/ml) O/N, lysés par la méthode de lyse alcaline modifiée, et purifiés sur gradient de chlorure de césium.

Milieux :**LB (Luria-Bertani)**

10 g Tryptone
5 g Extrait de levure
10g NaCl

Ajuster le pH à 7.5. Compléter à 1 litre. Autoclaver 20 min.

LB-Amp

Même recette que LB. Quand le milieu est refroidi, ajouter 1 ml d'ampiciline 100 mg/ml à 1 litre de LB (concentration finale 100 µg/ml).

LB agar

Même recette que LB. Ajouter 15 g d'agar avant de stériliser. Lorsque le milieu est refroidi à ± 50 °C, verser aseptiquement dans des pétris de 100 mm stériles.

DMEM

Ajouter 50 ml de FBS (concentration finale 10%) et 5 ml de pénicilline/streptomycine à 500 ml de milieu DMEM (Dulbecco Modified Eagle Medium, MultiScience). On ajoute différents agents de sélection selon l'usage voulu.

1.1 Constructions pCEP4

1.1.1 Construction pCEP GST

Le vecteur pCEP GST a été construit à partir du vecteur pCEP4 (Invitrogen). Le gène de la glutathione-S-transférase (GST) de *Schistosoma japonicum* a été cloné dans les sites Asp718 et BamHI du site de clonage multiple de pCEP4. Le fragment GST a été amplifié par PCR à partir du vecteur pGEX-4T3 (AMERSHAM) avec les amorces 5'GST Asp718 (5'-GGG GTA CCC AGG AAA CAG TAT TCA TGT CC-3') et 3'GST polylinker *XhoI/NotI* (les sites *XhoI* et *NotI* sont en gras dans la séquence de l'oligo, ainsi que les 2 codons stops) contenant le site *BglII* (en italique) 5'-GAA *GAT CTT CAT CAT CCT TTT GCG GCC GCT CCT GAT CCT GAT CCT GAT CTC GAG CGC AGA TCC GAT TTT GGA GG*-3' (Hukabel Scientifique). On a digéré 1 µg de pCEP4 avec les enzymes *Asp718* (Roche) et *BamHI* (NEB) dans le tampon de digestion B (ROCHE). Le fragment de PCR a été digéré avec *Asp718* et *BglII*. Les fragments ont été ligués tel que décrit précédemment. L'intégrité de la séquence du gène de la protéine GST a été vérifiée par séquençage automatisé sur un séquenceur automatique ABI 310 (ABI PRISM 310 Genetic Analyser).

1.1.2 Construction pCEP GST GFP

Le gène codant pour eGFP (enhanced green fluorescent protein) a été cloné dans les sites *XhoI* et *NotI* du vecteur pCEP GST afin de servir de contrôle de digestion dans le clonage subséquent de la librairie. Brièvement, un insert eGFP *XhoI/NotI* a été sous-cloné dans pCEP GST à partir du vecteur pEGFP-N1 (Clontech) avec les enzymes *XhoI/NotI* pour donner le nouveau vecteur pCEP GG. La séquence des jonctions *XhoI* et *NotI* a été vérifiée par séquençage pour s'assurer que le sous-clonage n'avait pas affecté les séquences en

amont et en aval des sites *XhoI* et *NotI*. Une fois digéré *XhoI/NotI*, le vecteur pCEP GG sert de squelette pour l'introduction de la librairie en fusion avec le gène de la GST.

1.2 Constructions pBABE

1.2.1 pBABE HAZ GST

Comme nous l'avons fait précédemment pour le vecteur pCEP GST, nous avons introduit le gène codant pour GST dans le vecteur rétroviral pBABE HAZ (24). Le gène a été amplifié à partir du vecteur pGEX 4T3 avec les amorces 5'-GST *EcoRI* ΔATG 5'-TTC ACG AAT TCT CCC CTA TAC TAG GTT ATT G-3' et l'amorce 3'GST polylinker *XhoI/NotI* (voir section 1.1.1). Le fragment de PCR digéré *EcoRI/NotI* a été introduit dans le vecteur pBABE HAZ digéré *EcoRI/NotI*, ce qui a créé une fusion HA-GST-Zéocyne^R. Il est à noter que le gène de résistance à la zéocyne demeure dans un autre cadre de lecture que celui de la GST, ce qui empêche le vecteur pBABE HAZ de conférer la résistance à la zéocyne.

1.2.2 pBABE HAGIZ

Dans le but de créer un vecteur avec une cassette d'expression bicistronique qui coderait pour une fusion HA-GST-librairie (sous le contrôle d'un promoteur SV40) et pour le gène de résistance à la zéocyne (sous le contrôle de l'IRES de EMCV), nous avons cloné une séquence IRES du virus EMCV en fusion avec le gène de résistance à la zéocyne entre les sites *NotI* et *ClaI* du vecteur pBABE HAZ GST. Pour ce faire, la séquence de l'IRES de EMCV a été amplifiée par PCR avec les amorces 5'IRES *NotI* Stop WT (qui contient deux codons d'arrêt de la transcription (en italique) pour arrêter la transcription de la fusion HA-GST-LIB et un site *NotI* (en gras) pour le clonage dans le site *NotI* de GST) 5'-AGT CCT **GCG GCC GCT TGA TGA** GCT TGG

AAT AAG GCC GGT G-3' et 3'-IRES Zéocyne 5'-CAT ATT ATC ATC GTG TTT TTC AA-3' (qui permet d'amplifier l'IRES de EMCV jusqu'au codon d'initiation 834 qui est le codon naturel d'initiation de la traduction de EMCV). Le gène de résistance à la zéocyne a pour sa part été amplifié avec les amorces 5' Zéo ATG IRES (qui contient une séquence s'appariant à la partie 3' de l'IRES de EMCV dans le but de faire un PCR double et de créer une fusion IRES/Zéo) 5'-AAA CAC GAT GAT AAT ATG GCC AAG TTG ACC AGT GCC-3' et 3'Zéo *Clal* (qui contient le site de restriction *Clal* (en gras)) 5'-ATA TCG ATT CAG TCC TGC TCC TCG GC-3'. Les deux fragments ainsi amplifiés ont été purifiés sur gel, extraits et réamplifiés par PCR en présence des amorces 5'IRES *NotI* Stop et 3' Zéo *Clal* afin de produire une fusion de l'IRES de EMCV avec le gène de résistance à la zéocyne. Ce fragment, une fois digéré *NotI/Clal*, a été inséré dans le vecteur pBABE HAZ GST digéré lui aussi *NotI/Clal*, pour donner le vecteur final pBABE HAGIZ (figure 7).

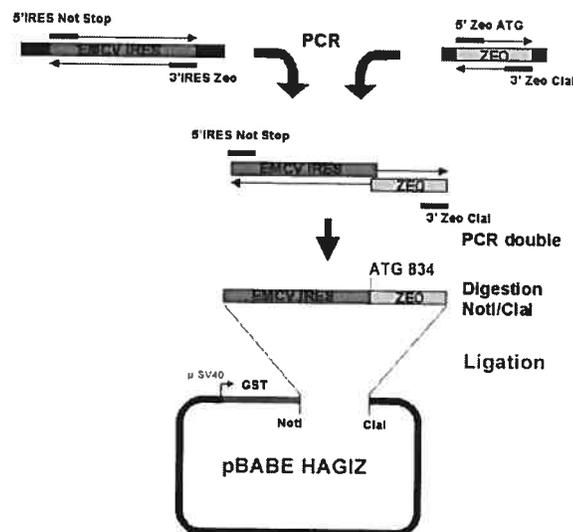


Figure 7. Construction du vecteur pBABE HAGIZ. L'IRES de EMCV amplifié à partir du provirus de EMCV est fusionné par PCR double au niveau de l'ATG 834 avec le gène de résistance à la zéocyne. La fusion IRES/Zéo^R est clonée dans le vecteur pBABE GST dans les sites NotI et Clal.

1.2.3 pBABE HAGIZ GFP

Dans le but d'avoir un contrôle de digestion lors du clonage de la librairie dans pBABE HAGIZ, nous avons cloné un insert contenant le gène codant pour eGFP dans les sites *XhoI* et *NotI* situés en 3' du GST cloné dans pBABE HAGIZ. Brièvement, le fragment eGFP *XhoI/NotI* provenant de pEGFP-N1 a été sous-cloné dans pBABE HAGIZ digéré *XhoI/NotI*. Le vecteur résultant, pBABE HAGIZ GFP, servira une fois digéré *XhoI/NotI* de squelette pour le clonage de la librairie en fusion avec le gène de la GST.

1.3 Synthèse du 2^{ième} brin de la librairie

La librairie de nonapeptides dont fait l'objet ce mémoire a été synthétisée sous la forme d'un oligonucléotide simple-brin (HUKABEL SCIENTIFIQUE) ayant la séquence suivante : 5'-CGC TCG AGA TCA GGA TCA GGA TCA GGA NNB NNB NNB NNB NNB NNB NNB NNB GCG GCC GCA AAA GGA AAA-3' (où N=A,T,C,G et B=T,C,G). Le deuxième brin de cet oligonucléotide a été synthétisé selon un protocole modifié à partir de celui de Noren *et al.* (54). 3 µg d'oligonucléotide simple-brin sont mis en présence de 7 µg de l'amorce 3' Klenow complémentaire à la partie 3' de l'oligonucléotide de la librairie (amorce 3'Klenow : 5'-TTT TCC TTT TGC GGC CGC-3') dans un volume final de 50 µl d'H₂O dd (ratio molaire librairie:amorce de 1:10), dénaturés à 95°C pendant 3 minutes puis le mélange est refroidi à raison de 0,1 °C/s jusqu'à 30°C, puis conservé à 4°C. Cette étape se fait dans une machine PJ-Cycler (MJ Research). Cette étape a pour but de dénaturer les oligonucléotides et de les faire s'apparier pour pouvoir ensuite initier la synthèse du deuxième brin d'ADN. On prépare ensuite le mélange de synthèse comme tel en ajoutant, dans un autre tube et dans l'ordre : 67 µl de tampon Klenow 3X (30 mM TRIS-Cl pH 7.5, 150 mM NaCl, 30 mM MgCl₂, 15 mM DTT, 0.1 mg/ml BSA), 50 µl de librairie provenant de l'étape

précédente, 57 μ l d' H_2O dd, 8 μ l d'un mélange de nucléotides (dATP,dCTP,dGTP, dTTP) 10mM, 4 μ l de Klenow 5U/ μ l (NEB) dans un volume total de 200 μ l. Le mélange est agité et incubé à 37°C 15 minutes, inactivé à 65°C pendant 15 minutes et conservé à 4°C. On conserve un aliquot non-digéré de la librairie (4 μ l) afin d'évaluer l'efficacité de digestion. La librairie néosynthétisée est ensuite digérée O/N à 37 °C dans un volume final de 400 μ l de tampon #3 (NEB) avec 50 unités d'enzyme *NotI* et 50 unités d'enzyme *XhoI*. On purifie ensuite le produit de digestion par extraction au phénol-chloroforme, chloroforme, précipitation à l'éthanol 100%, lavage à l'éthanol 70% et on resuspend dans 40 μ l de TE pH 8.0 (10 mM TRIS-Cl pH8.0, 1mM EDTA pH 8.0) . On purifie sur gel d'agarose LMP 1% en mettant 1 μ l de la librairie digérée dans un puit différent qu'on comparera à la librairie non-digérée pour s'assurer que la digestion s'est bien faite. La bande de librairie digérée obtenue après migration est extraite avec le kit QIAquick Gel Extraction Kit (QIAGEN) et resuspendue dans 75 μ l de TE pH 8.0. Il est important de porter une attention toute particulière à la synthèse et à la digestion du deuxième brin de la librairie puisqu'il s'agit là de l'étape à la fois la plus délicate et la plus importante dans la construction d'une librairie de peptide de grande diversité.

1.4 Clonage de la librairie dans pCEP GST/pBABE HAGIZ

Note : Comme la librairie été insérée de la même façon dans les vecteurs pCEP GST et pBABE HAGIZ, la technique de clonage est décrite une seule fois pour les deux vecteurs.

Le clonage de la librairie dans pCEP GST ou dans pBABE HAGIZ se fait en trois étapes. Premièrement, le vecteur (pCEP GG ou pBABE HAGIZ GFP) est digéré avec les enzymes *XhoI* et *NotI* O/N à 37°C, déphosphorylé, purifié sur gel, extrait du gel et dosé sur gel et par DO_{260} . Deuxièmement, on effectue une étape de ligation/transformation pilote dont le but est de déterminer

le ratio vecteur : insert optimal pour l'obtention du plus haut pourcentage de clones positifs possible afin de conserver le maximum de diversité à la librairie. La troisième étape, la ligation totale, consiste à reproduire les conditions de la ligation pilote à grande échelle et de transformer le produit de ligation dans des bactéries électrocompétentes pour bâtir la librairie de base. Pour les vecteurs pCEP GST et pBABE HAGIZ, la ligation pilote se fait en mettant sur pied une réaction de ligation mettant en présence une faible quantité de vecteur (100 ng) avec une quantité variable d'insert. Le but est de travailler avec des ratios molaires insert:vecteur comportant un léger excès d'insert. Pour chaque réaction de synthèse du deuxième brin de la librairie, la compétence de ladite librairie pour la ligation dans pCEP GST ou dans pBABE HAGIZ est évaluée (parce que le rendement varie avec chaque réaction) et les librairies donnant le plus grand nombre de transformants tout en conservant une proportion élevée de clones positifs sont utilisées pour la ligation totale. La ligation totale consiste à reproduire la même réaction de ligation que lors de la réaction pilote, mais en multipliant le nombre de tubes pour obtenir le maximum de produit de ligation avec les réactifs testés précédemment. Par exemple, si le ratio optimal de ligation est de 1 μ l de vecteur digéré (100ng) pour 4 μ l de librairie digérée, et que nous disposons de 40 μ l de librairie digérée, la ligation totale comportera 10 tubes, soit un total de 1 μ g de vecteur ligué. Après une ligation O/N à 16 °C, les réactions sont inactivées à 70°C pour 30 minutes, réunies en aliquots de 100 μ l et purifiées sur colonne de Sephadex G-50 (AMERSHAM) pour se débarrasser des sels, enzymes et autres éléments de la réaction pouvant interférer avec la transformation subséquente de la librairie dans des bactéries électrocompétentes. L'électroporation se fait dans des bactéries Gene Hogs électrocompétentes (INVITROGEN) à l'aide d'un électroporateur Gene Pulser II (BIORAD). Les réglages de l'électroporateur sont les suivants : 1.6 kV, 200 Ω , 25 μ F. L'électroporation se fait selon les indications du fabricant de bactéries. Brièvement, on ajoute 3 μ l de réaction de ligation purifiée (\pm 50 ng d'ADN) à 50 μ l de bactéries dégelées sur glace. Après une légère agitation, les bacté-

ries sont transférées dans une cuvette d'électroporation stérile (BIORAD) refroidie sur glace. Immédiatement après l'électroporation, on resuspend les bactéries dans 250 μ l de milieu SOC (62) et on incube 1h à 37°C avec forte agitation afin d'exprimer le gène de résistance à l'ampicilline. On regroupe toutes les bactéries transformées dans un seul tube. Elles sont ensuite étalées à raison de 50 μ l sur des pétris de 100 mm par pétris contenant du milieu LB-Amp et incubées O/N à 37°C. On procède en parallèle au titrage des transformants viables afin de déterminer la diversité clonale de la librairie. Le lendemain, toutes les colonies sont récupérées et resuspendues dans du bouillon LB-Amp, culottées par centrifugation à 4000 rpm dans un rotor Heraeus 6445 dans une centrifugeuse Sorvall T-Legend à 4 °C et resuspendues dans un volume final de 10 ml de LB-Amp. Un ml de cette suspension est amplifié dans 1 l de LB-amp, lysé et purifié sur gradient de CsCl afin de produire un stock d'ADN pouvant être transfecté. Le reste est dilué 1:1 dans du glycérol stérile et congelé à -80 °C.

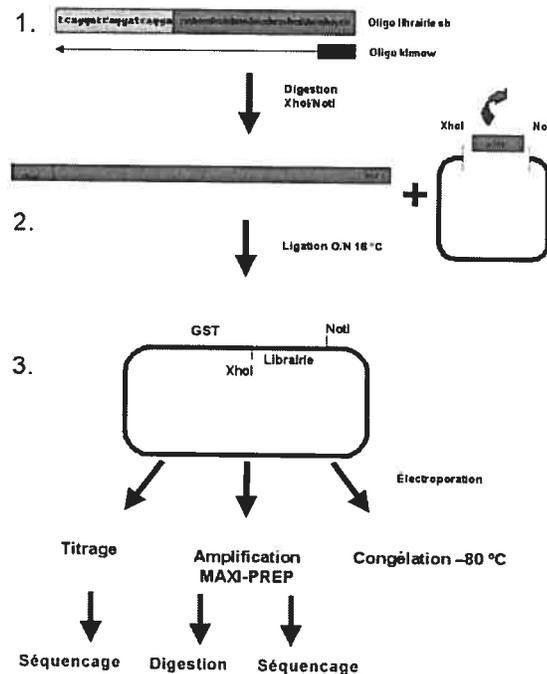


Figure 8. Organigramme de la méthode de construction de la librairie. 1) Le deuxième brin de la librairie est synthétisé par remplissage à la Klenow. 2) La librairie double-brin est digérée *XhoI/NotI* et liguée dans le vecteur cible (soit pCEP GST ou pBABE HAGIZ). 3) Les vecteurs ligés sont transformés par électroporation dans des bactéries *E. coli* électrocompétentes. On titre la diversité clonale de la librairie, un aliquot est amplifié pour extraire l'ADN plasmidique codant pour la librairie et le reste des bactéries sont congelées à -80°C .

1.5 Évaluation de la diversité de la librairie (digestion, séquençage) dans système pCEP

La diversité finale de la librairie clonée dans pCEP GST est évaluée de deux façons. Premièrement, la diversité clonale, c'est-à-dire le nombre total de clones issus de la transformation dans les bactéries électrocompétentes, est évaluée par titrage. Brièvement, on dilue un aliquot du bouillon contenant toutes les bactéries électroporées par dilutions sérielles, de 1×10^{-1} jusqu'à 1

$\times 10^{-7}$, et on compte le nombre de UFC. On obtient le titre en multipliant le nombre de UFC (unité formant des colonies) par le facteur de dilution. Pour connaître la diversité réelle de la librairie, on doit également déterminer le nombre de clones qui ont réellement intégré la librairie. Pour ce faire, on met en culture 48 colonies provenant du titrage et on lyse les bactéries obtenues par la méthode de lyse alcaline. L'ADN plasmidique est ensuite digéré par les enzymes de restriction *AflIII* et *NotI* (NEB). Le patron de restriction est analysé sur gel d'agarose 2% par rapport à celui d'un vecteur vide et le pourcentage de clones positifs est déterminé. La diversité réelle de la librairie peut ensuite être calculée en multipliant la diversité clonale par le pourcentage de clones positifs. Afin de s'assurer que les séquences peptidiques sont vraiment aléatoires, on séquence quelques-uns des clones positifs et on vérifie qu'il n'y a pas de biais au point de vue de la fréquence de chacun des acides aminés (54). Cette façon de déterminer la diversité de la librairie ne tient cependant pas compte du fait que l'on amplifie les clones afin d'extraire suffisamment d'ADN pour pouvoir ensuite transfecter des cellules. Dans le but de déterminer la diversité réelle de la librairie suite à l'amplification, on transforme 1 μg d'ADN plasmidique tel qu'obtenu à l'étape précédente et on cultive 48 colonies. Ces 48 colonies seront digérées afin de vérifier la proportion de clones positifs. Ces clones positifs sont ensuite séquencés pour s'assurer qu'il n'y a pas de biais au niveau de l'insertion des acides aminées dans les nonapeptides.

1.6 Récupération du vecteur pCEP GST tranfecté dans des cellules 293T

Lorsqu'on a introduit une librairie de peptides dans une lignée cellulaire, il est important d'avoir une méthode qui nous permette de récupérer la séquence du peptide biologiquement actif que l'on a isolé à l'aide de notre méthode de criblage. Pour ce faire, nous avons utilisé la méthode d'extraction d'ADN plasmidique à partir de cellules en culture développée par Bernard Hirt en 1967 (37). Il s'agit de lyser les cellules transfectées avec le vecteur

pCEP GST à l'aide d'une solution de 0.6% de SDS dans du EDTA 100 mM pH 7.5. On ajoute 1 ml de cette solution sur 5×10^6 cellules en culture dans une plaque de 6 puits. Après 10 à 20 minutes d'incubation à température pièce, les cellules sont récupérées à l'aide d'un « policeman » et transférées dans un tube eppendorf. Les protéines et l'ADN génomique contenus dans le lysat sont précipités en ajoutant 0.25 ml de NaCl 5M. Cette solution est stockée à 4 °C toute la nuit puis centrifugée à 17 000 rpm pendant 30 min, toujours à 4 °C. Un μ l de surnageant est ensuite transformé dans des bactéries compétentes AG-1, 20 colonies sont cultivées O/N dans du LB amp, et on vérifie par digestion avec *Asp718* et *NotI* si le profil de digestion correspond au vecteur d'origine.

2. BIOLOGIE CELLULAIRE :

2.1 Lignées cellulaires utilisées

Les lignées cellulaires citées au cours de ce mémoire sont les suivantes : HEK 293 (30), 293T (18), 293TK+ (voir ci-dessous).

2.2 Lignée cellulaire 293TK+

2.2.1 Production des virus HSVTK+

La lignée cellulaire 293TK+ a été construite à partir de la lignée cellulaire 293 par transduction à l'aide du vecteur rétroviral HPV 434 (GENETIX). Ce vecteur rétroviral contient une cassette d'expression bicistronique codant pour le gène de résistance à la néomycine (G418) et pour une protéine de fusion entre la thymidine kinase du virus Herpes simplex (HSV-TK) et la GFP. Les particules virales utilisées pour transduire les cellules 293 ont été produites en transfectant 20×10^6 cellules 293T par la méthode du CaPO_4 avec 12

μg du vecteur VPR 275 codant pour les protéines Gag et Pol du HIV-1, 8 μg du vecteur HPV 434, 1 μg d'un expresseur de Rev (p663), 1 μg d'un expresseur de la protéine d'enveloppe G du VSV (pYN15) et 1 μg d'un expresseur de la protéine Tat du HIV (HPV 17). Le milieu est changé 16 heures plus tard. Quarante-huit et 72 heures post-transfection, le surnageant est récolté et éclairci par centrifugation à 3000 rpm dans une centrifugeuse Sorvall T-Legend (Sorvall) pendant 30 minutes pour se débarrasser des débris cellulaires. Le surnageant est ensuite filtré à travers une membrane de 0.45 μm puis ultracentrifugé pour culotter les virus pendant 2 heures à 35 000 rpm à 4 °C dans un rotor 70.1 Ti (Beckman) dans une ultracentrifugeuse Beckman L8-M. Le culot est resuspendu dans 1 ml de DMEM FBS 20%, aliquoté en fractions de 50 μl et congelé à -80 °C.

2.2.2 Infection des cellules 293 par le virus HSVTK+

L'infection des cellules 293 a été effectuée par spinoculation. La spinoculation consiste à recouvrir 2×10^5 cellules adhérentes au fond d'un puits d'une plaque de 6 puits avec 1 ml de suspension virale concentrée produite à partir du vecteur HPV 434 et à centrifuger la plaque à 2600 rpm pendant 2 heures à température pièce pour forcer l'adsorption des virus à la surface cellulaire. Le milieu est remplacé et les cellules incubées à 37 °C pendant 48 heures. Les cellules sont ensuite décollées et resuspendues dans 1 ml de DMEM frais dans un tube à FACS (Fluorescence Activated Cell Sorting) (Falcon #2058). On trie les cellules par FACS, (FACSVantage SE, Becton-Dickinson) sur la base de l'expression de la protéine fluorescente GFP. Les cellules GFP positives sont récupérées et remises en culture en présence de DMEM supplémenté de 400 $\mu\text{g}/\text{ml}$ de G418 (GIBCO) pour sélectionner l'expression stable de la fusion HSVTK/GFP. Les cellules sont amplifiées puis congelées dans l'azote liquide dans du DMEM DMSO 10% FBS 20%. L'expression de la fusion HSVTK/GFP est vérifiée par FACS avant la congélation pour s'assurer qu'au moins 90% des cellules congelées soient positives.

2.2.4 Test toxicité du ganciclovir dans 293TK+

L'expression de la TK de HSV dans des cellules 293 les rend susceptibles à l'effet cytotoxique du ganciclovir (GCV). Pour déterminer si les cellules 293 TK+ sont vraiment sensibles au GCV, nous avons utilisé un protocole adapté de Wei et al (78). Brièvement, 2×10^4 cellules sont étalées dans chaque puits d'une plaque de 24 puits. On ajoute en triplicata 1 ml de milieu contenant différentes dilutions de GCV (0.5, 10, 20, 50 et 100 μ M). 48 heures plus tard, le milieu est changé et renouvelé avec les mêmes concentrations de GCV. 92 heures après le début de l'exposition au GCV, la viabilité est estimée par une méthode colorimétrique basée sur la transformation du MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] (SIGMA) en formazan par les cellules viables (33). Le milieu est donc enlevé et remplacé par 500 μ l d'une solution de 0.5% MTT dans du PBS et la plaque est incubée pendant 4 heures à 37 °C. Le MTT est ensuite remplacé par 500 μ l de DMSO (dimethylsulfoxyde) (ICN). La plaque est ensuite légèrement agitée pour bien dissoudre les sels de formazan. On lit la DO à 570 nm à l'aide d'un lecteur de plaque (EL_x800, BIO-TEK INSTRUMENTS). La viabilité est calculée en comparant la DO des puits contenant du GCV à celle des puits contrôle sans GCV.

2.3 Sélection zéocyne de cellules transduites ou transfectées avec le vecteur pBABE HAGIZ

Les cellules transduites ou transfectées par le vecteur pBABE HAGIZ sont laissées en culture pendant 48 heures post-transfection ou post-infection. Le milieu DMEM est alors remplacé par du milieu DMEM supplémenté de 50 μ g/ml de zéocyne (Invitrogen). Les cellules sont laissées en culture et passées lorsque confluentes. Le milieu de sélection est renouvelé aux 4 jours pour s'assurer que la zéocyne est toujours active et éviter l'apparition de clones spontanés. Chaque expérience de sélection comporte un pétri contrôle de

cellules non transfectées mises en présence de zéocycine pour s'assurer de l'efficacité de l'agent de sélection. Les cellules sont gardées en culture jusqu'à ce qu'il y ait apparition de foyers de croissance bien définis. Ces foyers sont récupérés et remis en culture pour obtenir une population plus abondante de cellules.

3. BIOCHIMIE :

3.1 Expression de GST dans les cellules 293T à partir de pCEP GST

L'expression de GST à partir du vecteur pCEP GST a été vérifiée par immunobuvardage anti-GST. 5×10^6 cellules 293T ont été transfectées avec, respectivement, 5 μg de pCEP GST, de SVCMV GST (expresseur contrôle) et de pCEP4 (contrôle négatif). Quarante-huit heures post-transfection, les cellules sont récoltées, culottées et lysées dans 400 μl de digitonine (200 μl de digitonine et 200 μl de tampon digitonine 2X) 30 minutes sur glace. On centrifuge ensuite les échantillons 30 minutes à 4°C à 14 000 rpm sur une centrifugeuse Eppendorf 5402 (Eppendorf) et on récolte le surnageant. Les surnageants sont dosés pour normaliser la quantité de protéines totale avec le BIORAD Protein Assay Kit (BIORAD) en lisant la DO à 595 nm. La GST contenue dans chaque échantillon est purifiée par affinité en utilisant des billes de sépharose conjuguées à la glutathione (Amersham). Dans un volume de 500 μl de lysat cellulaire (chaque lysat ayant été dilué pour contenir la même quantité de protéines totales), on ajoute 30 μl d'une suspension de billes préalablement lavées 2 fois au PBS et resuspendues dans 100 μl de PBS et on incube 2 heures à 4 °C avec agitation. Les billes sont ensuite culottées à 6000 rpm à 4 °C et lavées 5 fois avec 1 ml de PBS. L'élution se fait en incubant les billes en présence de 40 μl de glutathione réduite 100 mM 20 minutes sur glace. Les billes sont ensuite centrifugées à 6000 rpm et le surnageant transféré dans un autre tube. 20 μl de surnageant sont mélangés avec 5 μl de solution de chargement, bouillis 5 minutes et chargés sur gel de

SDS-PAGE 10%. Les protéines ainsi séparées sont transférées sur une membrane de nitrocellulose (BIORAD) en utilisant un appareil de transfert ElectroBlot SemiDry (15V, 200 mA, 30 min, T° pièce) (BIORAD). L'immunobuvardage se fait dans les conditions suivantes : 45 minutes de blocage dans TBS (Tris 20 mM, NaCl 137 mM, pH 7.6) Tween 0.1%-lait 3%, anticorps primaire 1 :2500 (chèvre anti-GST, AMERSHAM) O/N 4°C dans TBS Tween 0.1%-lait 3%, lavage 1 heure avec changement de solution de lavage (TBS Tween 0.1%), anticorps secondaire 1:8000 (anti-chèvre HRP, Jackson ImmunoResearch) 40 minutes dans TBS Tween 0.1%, lavage 5 heures avec changements réguliers de solution de lavage. La membrane est ensuite séchée, puis les bandes de protéines sont révélées par ECL selon les indications du fabricant.

3.2 Expression long terme de GST dans 293TK+ sous sélection hygromycine

2×10^5 cellules 293TK+ ont été transfectées avec $1 \mu\text{g}$ du plasmide pCEP GST avec le réactif FUGENE selon les instructions du fabricant (ROCHE). Quarante-huit heures post-transfection, les cellules ont été transférées dans du milieu DMEM supplémenté de $600 \mu\text{g/ml}$ d'hygromycine B (CALBIOCHEM). 5, 10, 15, 20 et 25 jours après le début de la sélection, un pétri a été récolté et 2×10^6 cellules lysées dans de la digitonine tel que décrit précédemment. Chaque surnageant a été congelé à $-80 \text{ }^\circ\text{C}$. Chaque échantillon a ensuite été normalisé pour la quantité de protéines totale et la GST purifiée par affinité comme décrit ci-haut. La présence de GST dans chacun des échantillons a été évaluée par immunobuvardage comme décrit au point 3.1.

3.3 Expression de la fusion GFP/TK+ dans 293TK+

5×10^6 cellules 293TK+ ainsi que 5×10^6 cellules transfectées avec $5 \mu\text{g}$ du vecteur pEGFP-N1 par la méthode du CaPO_4 en guise de contrôle positif ont

été lysées dans 800 μ l de RIPA-DOC 30 minutes sur glace puis centrifugées 30 minutes à 4 °C, 14000 rpm dans une centrifugeuse Eppendorf 5402. Le surnageant a été dosé pour la quantité de protéines totales tel que décrit précédemment et chaque échantillon dilué pour charger la même quantité de protéines totales sur gel. Après séparation par SDS-PAGE, les protéines sont transférées sur membrane de nitrocellulose comme décrit précédemment. L'immunobuvardage a été effectué dans les conditions suivantes : blocage 40 minutes dans TBS Tween 0.1%-lait 3%, incubation avec l'anticorps primaire 1 : 500 (lapin anti-GFP) O/N dans TBS Tween 0.1%-lait 3%, lavage 1 heure avec TBS Tween 0.1%, incubation 40 minutes avec anticorps secondaire 1 :5000 (anti-lapin-HRP, AMERSHAM), lavage 5 heures dans TBS Tween 0.1% avec changement fréquent de la solution de lavage. La membrane est ensuite séchée et les bandes révélées par ECL selon les indications du fabricant.

3.4 Expression de GST à partir du vecteur rétroviral

5×10^6 cellules ont été transfectées par la méthode du calcium phosphate avec, respectivement, 5 μ g de pBABE HAZ, 5 μ g de pBABE HAGIZ, et 5 μ g de SVCMV GST. 16 heures post-transfection le milieu a été changé et 48 heures post-transfection, les cellules culottées et lysées dans la digitonine tel que décrit précédemment. L'expression de la GST dans les divers échantillons a été détectée par immunobuvardage anti-GST tel que décrit dans la section 3.1, à l'exception que les billes de glutathione-sépharose ont été bouillies au lieu d'être éluées.

4. PRODUCTION ET INFECTION VIRALE :

4.1 Production de particules virales à partir de pBABE HAGIZ dans les cellules 293T

4 X 10⁷ cellules 293T ont été transfectées par la méthode du calcium phosphate avec un mélange des vecteurs suivants : 10 µg de pBABE HAGIZ, 10 µg gag-pol MuLV (pBS gag-pol) et 2 µg VSVG. 16 heures plus tard, le milieu de transfection a été changé et le milieu contenant les virus récolté à 48 h et 72 h post-transfection. Le milieu contenant les virus a été éclairci par centrifugation à 3000 rpm pendant 30 minutes et filtration sur membrane de 0.45 µm. Les virus ont ensuite été culottés par ultracentrifugation à 35 000 rpm dans un rotor 70.1 Ti (BECKMAN), 4 °C pendant 2 heures. Le culot viral a été resuspendu dans 1 ml de milieu DMEM + 20% FBS, aliquoté en fractions de 50 µl et congelé à -80 °C.

4.2 Standardisation de l'infection par essai RT

Afin de standardiser l'infection de cellules 293T par les virus produits au point 4.1, nous avons déterminé l'activité de la reverse-transcriptase virale selon le protocole suivant. On prépare un tampon de réaction en mélangeant dans l'ordre : 250 µl de Tris HCl pH 8.3, 100 µl 1M DTT, 50 µl 5M NaCl, 250 µl 1% NP40 et 350 µl H₂O. On prépare ensuite un mélange pour 5 réactions en mélangeant : 100 µl de tampon de réaction, 25 µl de dTTP 0.2 mM, 20 µl de polyA dT 15, 50 µl de MnCl₂ 6mM , 20µl dTTP H³ et 35 µl H₂O. Après ajout de 50 µl de ce mélange à 50 µl de suspension virale aliquotée, chaque réaction est incubée 2 heures à 37 °C. Après incubation, on ajoute à chaque tube 1 ml de TCA 10% et on incube sur glace 2 heures. Chaque réaction est récupérée par aspiration sur un filtre Glass Fiber Circle G4 (FISHERBRAND). Chaque filtre est rincé avec du TCA 5% puis du EtOH 70%. Les filtres sont récupérés, placés dans un flacon à scintillation et séchés 1h à

65 °C. On ajoute 50 ml de liquide à scintillation CytoScint (ICN) et on lit sur un compteur à scintillation LS-6000 SC (Beckman-Coulter) avec le programme #3 (pour lire le nombre de comptes associés au tritium (H^3)). L'infection est effectuée en utilisant un ratio fixe de cpm/cellule.

4.3 Infection de cellules 293T pour montrer la résistance à la zéocyne à partir du vecteur rétroviral pBABE

2×10^5 cellules ont été infectées en resuspendant 2×10^5 cellules 293T dans 2 ml de DMEM contenant assez de virus pour obtenir un ratio virus/cellules de 10 cpm/cellule tel que déterminé par la standardisation de la RT contenue dans les particules virales. L'infection s'est faite en présence de 10 μ g/ml de Polybrene, O/N à 37 °C. 48 heures post-infection, les cellules ont été mises sous sélection zéocyne 50 μ g/ml tel que décrit au point 2.3. Après 3 semaines de sélection, les clones résistants ont été isolés et mis en culture dans les puits d'une plaque 24 puits dans 500 μ l de milieu de sélection. Les clones positifs ont ainsi été amplifiés jusqu'à l'obtention de 1×10^7 cellules. La moitié de celles-ci ont été lysées pour détecter l'expression de GST tel que décrit précédemment alors que l'autre moitié a été utilisée pour extraire l'ADN génomique et détecter l'intégration du vecteur rétroviral.

4.4 Détection de l'intégration du gène de la zéocyne dans les cellules 293T exprimant la cassette GST-Zéo de façon stable

5×10^6 cellules 293T résistantes à la zéocyne ont été lysées dans 1 ml de réactif DNAzol (INVITROGEN). L'ADN génomique a été isolé selon les instructions du fabricant. Afin de détecter l'intégration du gène de résistance à la zéocyne, un PCR a été fait sur 200 ng d'ADN génomique isolé de 4 clones résistants à la zéocyne différents. Le PCR utilise la polymérase Pfx (INVITROGEN) dans les conditions suivantes : 1) 94 °C 2 minutes, 2) 94 °C 45 secondes, 3) 62 °C 45 secondes, 4) 68 °C 45 secondes, 5) retour à l'étape

#2 29 fois, 6) élongation finale à 68 °C pendant 5 minutes. 10 µl de chaque réaction sont ensuite analysés sur gel d'agarose 1%.

SECTION RÉSULTATS :

1. CONSTRUCTION D'UNE LIBRAIRIE DANS UN VECTEUR ÉPI-SOMAL

Cette section rapporte les étapes de mise au point du clonage d'une librairie de nonapeptides, tant au niveau de son expression que de sa diversité, dans le vecteur épisomal pCEP GST, un expresseur de GST dérivé du vecteur commercial pCEP4.

1.1 Analyse de l'expression de GST à partir du vecteur épisomal pCEP GST

Afin de vérifier que le vecteur pCEP GST peut exprimer la protéine GST clonée dans les sites *Asp718* et *BamHI*, des cellules 293T ont été transfectées de façon transitoire avec pCEP GST, SVCMV GST (expresseur contrôle de GST) et pCEP4 en guise de contrôle négatif. 48 heures post-transfection, il est possible de détecter par immunobuvardage anti-GST l'expression de GST qui se traduit par l'apparition d'une bande à environ 25 kDa dans le puits de pCEP GST et environ 23 kDa dans le puits SVCMV GST. Cette différence s'explique par la présence d'un linker de 6 acides aminés en C-terminal du gène de GST cloné dans pCEP GST. Les cellules non transfectées ainsi que celles transfectées avec le vecteur vide ne permettent pas la détection d'une bande spécifique à GST.

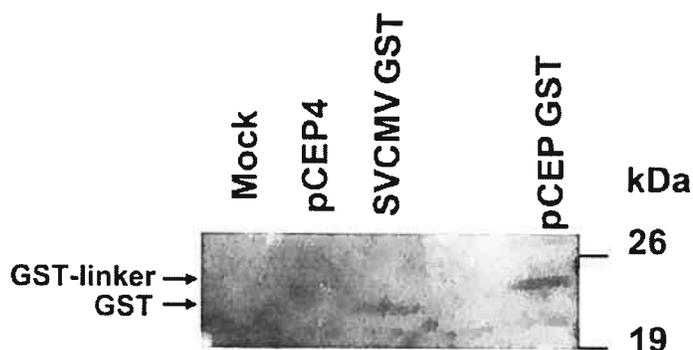


Figure 9. Expression de GST à partir du vecteur pCEP GST, telle que détectée par immunobuvardage anti-GST. Mock : cellules 293T non transfectées, pCEP4 : vecteur contrôle négatif, SVCMV GST : exprèsseur de GST contrôle. Il est à noter que le marqueur de poids moléculaire utilisé ici ne correspond pas à celui utilisé dans toutes les autres figures. Les variations apparentes du poids moléculaire de GST entre cette figure et les autres sont donc dues à la différence de migration entre ce marqueur et celui utilisé dans les autres figures.

1.2 Analyse de l'expression à long terme de GST à partir du vecteur pCEP GST dans des cellules 293TK+

Afin de vérifier que la protéine GST pouvait être exprimée sur une période assez longue pour sélectionner des peptides inhibiteurs, et pour déterminer la fenêtre idéale d'application de la sélection GCV, nous avons étudié la durée d'expression de GST à partir de pCEP GST sous sélection hygromycine dans des cellules 293TK+. L'hygromycine a été appliquée 2 jours post-transfection et l'expression de GST analysée par immunobuvardage anti-GST tous les 5 jours post-sélection, jusqu'à concurrence de 25 jours. On peut constater un plateau d'expression entre le 10^{ième} et le 20^{ième} jours. Avant et après cette période, GST est exprimée à des niveaux moindres. Il est important de noter que chaque échantillon correspond au même nombre de cellules lysées (2×10^6) et que chaque échantillon a été normalisé au niveau de la quantité de protéines totale avant d'effectuer le « GST pull-down ».

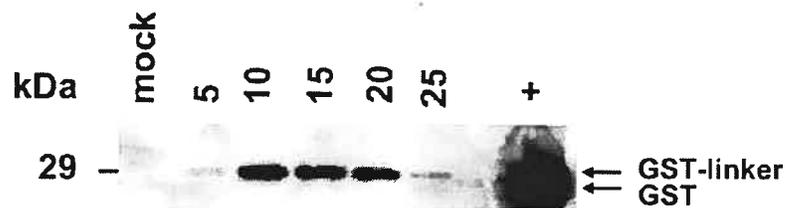


Figure 10. Immunobuvardage anti-GST montrant l'expression à long terme de GST à partir du vecteur pCEP GST dans des cellules 293TK+. **Mock :** cellules 293TK+ non-transfectées. **5-25 :** nombre de jours post-sélection. **+** : GST recombinante produite dans *E. coli* à partir de pGEX-4T3.

1.3 Construction d'une librairie de nonapeptides dans le vecteur pCEP GST et évaluation de la diversité réelle de cette librairie

Un microgramme de la librairie liguée dans le vecteur pCEP GST a été transformé par électroporation dans des bactéries électrocompétentes Gene Hogs (INVITROGEN) (n=2). La diversité clonale de la librairie après l'étape de transformation a été déterminée par titrage des transformants à 1×10^5 transformants. Suite à cette étape de transformation, un aliquot de tous les transformants recueillis (1 ml sur 10 ml) a été amplifié et l'ADN plasmidique extrait. Un microgramme de cet ADN codant pour la librairie a été transformé et 48 clones ont été analysés pour déterminer la diversité réelle de la librairie. Sur 48 clones, 6 avaient effectivement intégré l'insert, pour un pourcentage de clones positifs de 12.5 %. Si on applique ce pourcentage à la diversité clonale de la librairie, on obtient une diversité réelle de 1.25×10^4 , ce qui représente 2 ordres de grandeur sous la fréquence théorique de sélection de clones tel que rapporté dans l'introduction (4, 27, 55).

1.4 Récupération du vecteur pCEP GST transfecté dans des cellules 293T.

Afin de vérifier si le vecteur pCEP GST pouvait être récupéré après sa transfection dans des cellules de mammifères, et ainsi confirmer que la séquence

d'un peptide inhibiteur pourrait être identifiée, nous avons transfecté le vecteur dans 5×10^6 cellules 293T. Quarante-huit heures post-transfection, les cellules ont été lysées par la méthode de HIRT et l'ADN plasmidique épisomal isolé. Cet ADN, une fois transformé dans des bactéries compétentes, a été réisolé et digéré pour voir si le patron de digestion était le même que celui d'un vecteur pCEP GST contrôle. La figure 11 montre que le profil de digestion de 12 colonies issues de la transformation du vecteur isolé à partir des cellules 293T est identique à celui de pCEP GST, confirmant ainsi qu'il est possible de récupérer un vecteur transfecté dans des cellules.

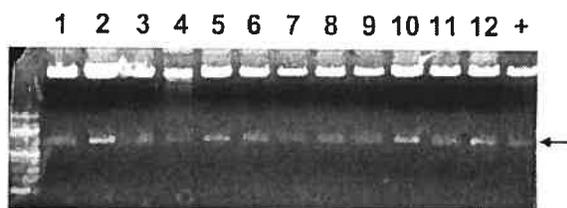


Figure 11. Récupération du vecteur pCEP GST transfecté dans des cellules 293T. Le profil de digestion correspond à celui d'un vecteur pCEP GST contrôle digéré *Asp718/NotI*. + : pCEP GST *Asp718/NotI*. 1-12 : colonies digérées *Asp718/NotI* issues de la transformation de l'ADN isolé à partir de cellules 293T transfectées avec pCEP GST.

2. CONSTRUCTION D'UNE LIBRAIRIE DE NONAPEPTIDES DANS UN VECTEUR RÉTROVIRAL

Dans le but d'améliorer le concept de librairie de peptides clonée dans le vecteur épisomal, et face aux problèmes d'expression de GST mis en lumière dans les sections précédentes, nous avons décidé de mettre au point un système d'expression rétroviral de la librairie dans le vecteur rétroviral pBABE HAGIZ provenant du virus MuLV. Cette approche devait permettre d'augmenter l'efficacité de clonage de la librairie, son introduction dans les cellules du système de criblage ainsi que sa versatilité pour des applications ultérieures.

2.1 Construction du vecteur rétroviral pBABE HAGIZ

Une cassette d'expression bicistronique a été clonée dans le vecteur pBABE HAZ pour créer le vecteur pBABE HAGIZ. La GST y est exprimée sous le contrôle du promoteur précoce du SV40 tandis que le gène de résistance à la zéocyne y est exprimé sous le contrôle de la séquence IRES de EMCV. La construction pBABE HAGIZ a été séquencée à partir de la partie C-terminale de GST (avant le site d'insertion de la librairie, voir figure 8 dans la section matériel et méthode, figure 12) jusqu'à la fin du gène de résistance à la zéocyne. La figure 12 montre l'étendue de la zone séquencée ainsi que les oligonucléotides utilisés pour ce faire. Toute la séquence du vecteur correspond à celles publiées pour l'IRES de EMCV et pour le gène de résistance à la zéocyne.

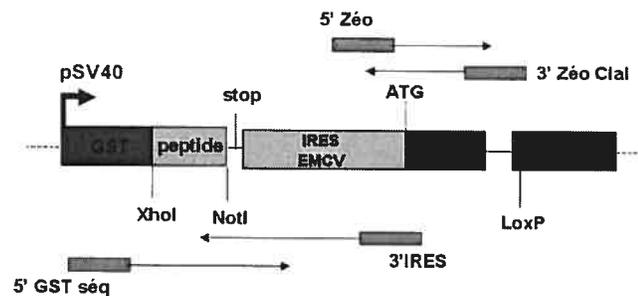


Figure 12. Stratégie de séquençage et organisation de la cassette d'expression de pBABE HAGIZ. Les flèches représentent les oligonucléotides utilisés.

2.2 Analyse de l'aptitude du vecteur rétroviral pBABE HAGIZ à produire des particules virales infectieuses

Le vecteur pBABE HAGIZ a servi à la production de particules virales infectieuses dans des cellules 293T par cotransfection avec un expresseur de *gag-pol* de MuLV et un expresseur de la glycoprotéine d'enveloppe G du VSV. Pour mesurer la production de particules virales à partir de pBABE HAGIZ et pour infecter des cellules 293T avec un ratio constant virus/cellules, nous avons quantifié par essai enzymatique spécifique au virus MuLV l'activité de

la RT incorporée dans les particules virales. La figure 13 montre clairement qu'il y a eu production de particules virales ayant incorporé la RT .

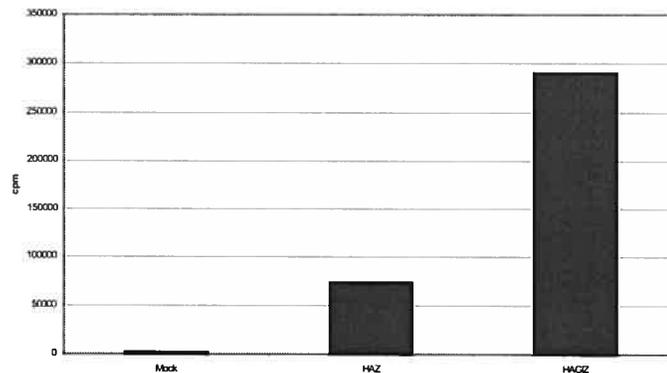


Figure 13. Quantification de la RT incorporée dans les particules virales produites à partir de pBABE HAGIZ. Mock : Milieu DMEM. HAZ : Suspension virale produite à partir du vecteur pBABE HAZ. HAGIZ : Suspension virale produite à partir de pBABE HAGIZ.

2.3 Analyse de l'expression transitoire de GST à partir du vecteur rétroviral pBABE HAGIZ dans des cellules 293T

Afin de vérifier que le vecteur pBABE HAGIZ pouvait exprimer la GST, le vecteur a été transfecté dans des cellules 293T et l'expression de GST détectée par immunobuvardage anti-GST 48 heures post-transfection. La figure 14 montre que la protéine GST peut être exprimée de façon transitoire dans des cellules 293T à partir du vecteur rétroviral. La bande de GST dans le puits du vecteur rétroviral est légèrement plus haute (32 kDa) que le témoin (29 kDa) en raison de la présence d'un épitope HA dans la partie N-terminale et d'un linker de 6 acides aminés dans la partie C-terminale de GST. On peut remarquer dans les puits de pBABE HAGIZ, Mock et SVCMV GST des bandes qui ne sont pas présentes dans les autres figures montrant l'expression de GST par immunobuvardage anti-GST. Ces bandes sont vraisemblablement des protéines ayant été copurifiées avec la GST par les billes de glutathione-sépharose et pour lesquelles l'anticorps possède une réactivité croisée. Leur présence s'expliquerait par le fait que, dans cette expérience, les billes

ont été bouillies au lieu d'être éluées, ce qui représente un traitement plus extrême pouvant faire se détacher des protéines liées de façon non-spécifique aux billes de glutathione-sépharose.

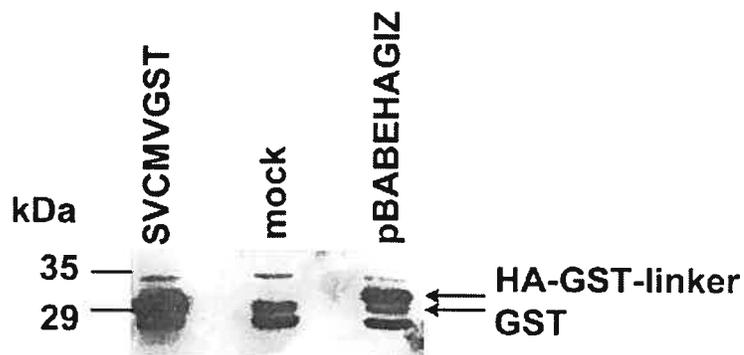


Figure 14. Expression transitoire de GST après transfection dans des cellules 293T détectée par immunobuvardage anti-GST. SVCMV GST : expresseur de GST. Mock : cellules 293T non-transfectées. pBABE HAGIZ : vecteur rétroviral.

2.4 Évaluation du vecteur pBABE HAGIZ pour son utilisation dans le transfert de la librairie de nonapeptides et du gène de résistance à la zéocyne dans les cellules 293T

Pour déterminer la capacité de notre vecteur rétroviral à transférer efficacement le gène de résistance à la zéocyne ainsi qu'à exprimer la librairie fusionnée à GST, nous avons effectué deux types d'expériences. Premièrement, nous avons transfecté le vecteur rétroviral dans des cellules 293T (en utilisant un vecteur pBUDCE4-zéo comme contrôle positif de résistance à la zéocyne). Deuxièmement, nous avons transduit des cellules 293T avec des particules virales contenant le vecteur pBABE HAGIZ. Chaque expérience de sélection a été effectuée en parallèle avec des cellules 293T n'exprimant pas le gène de résistance à la zéocyne, pour lesquelles il n'y a pas eu apparition de clones spontanés, toutes ces cellules étant mortes au cours de la sélection. L'expérience de transfection de pBABE HAGIZ a permis d'isoler 7

clones résistants à la zéocyne (alors que le contrôle positif pBUDCE4-zéo en avait généré une centaine) tandis que l'expérience de transduction de pBABE HAGIZ a permis d'isoler 4 clones résistants à la zéocyne. Nous avons étudié ces derniers au niveau de l'expression de GST (par immunobuvardage anti-GST et anti-HA) puisqu'ils ont été isolés dans des conditions similaires à celles qu'on utiliserait pour sélectionner une population exprimant la librairie. L'ADN génomique des 4 clones résistants à la zéocyne obtenus par transduction (51-53,101) a donc été isolé et l'intégration du gène de la zéocyne vérifiée par un PCR génomique utilisant des amorces permettant d'amplifier le gène complet de résistance à la zéocyne. La figure 15 montre que 3 clones sur 4 ont effectivement intégré le gène de résistance, tandis que le clone 51 ne présente qu'une bande de très faible intensité.

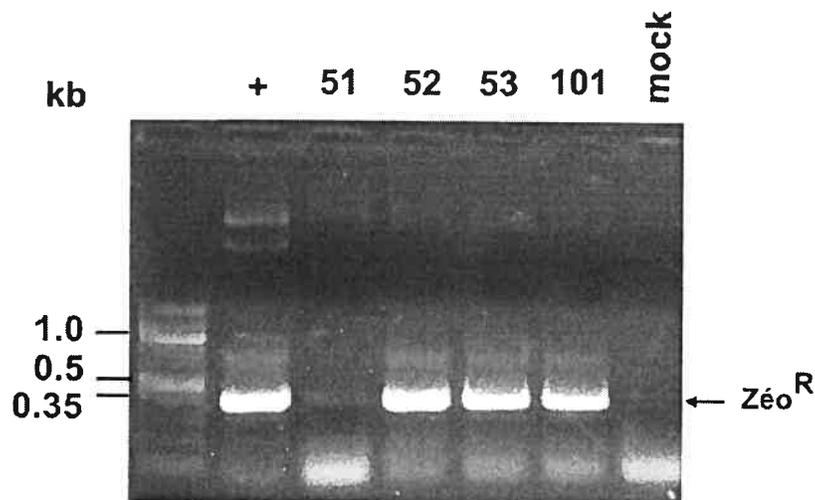


Figure 15. Intégration du gène de résistance à la zéocyne dans des cellules 293T exprimant le vecteur pBABE HAGIZ de façon stable. PCR génomique effectué avec des amorces amplifiant le gène de résistance à la zéocyne. + : vecteur pBABE HAGIZ. 51-53,101 : clones résistants à la zéocyne. Mock : cellules 293T non-transduites. Zéo^R : gène de résistance à la zéocyne.

Ces 4 mêmes clones ont été analysés pour l'expression de GST par immunobuvardage anti-GST. Afin de contrevérifier l'information, on a également tenté de détecter l'expression de GST en faisant un immunobuvardage anti-

HA (la GST exprimée à partir du vecteur pBABE HAGIZ est fusionnée avec un épitope HA). L'immunobuvardage anti-GST n'a pas permis de détecter de bandes spécifiques à la GST dans les clones résistants à la zéocyne (figure 16). Cette absence d'expression a été vérifiée par immunobuvardage anti-HA. On voit bien dans la figure 16 que l'anticorps anti-HA permet de reconnaître la bande de GST-HA produite à partir du vecteur pBABE HAGIZ (présente également dans le panneau du bas), mais ne permet pas de détecter la moindre expression de cette fusion dans les clones résistants à la zéocyne. Ces deux immunobuvardages se confirment mutuellement et établissent avec certitude que les clones résistants à la zéocyne que nous avons isolés n'expriment pas la GST mais ne sont vraisemblablement pas des clones spontanés en raison de la présence du gène de résistance à la zéocyne dans le génome de ces cellules.

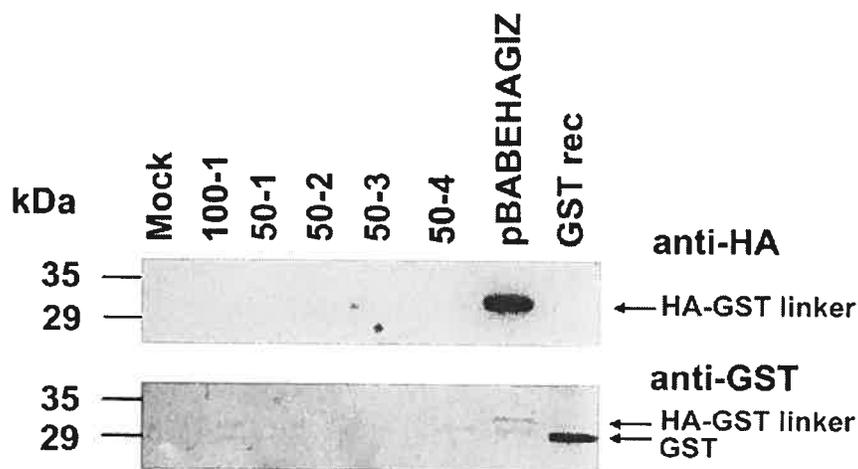


Figure 16. Le gène de la GST n'est pas exprimé à partir du vecteur rétroviral pBABE HAGIZ dans les cellules 293T après sélection à la zéocyne, tel que démontré par immunobuvardage anti-HA (en haut) et anti-GST (en bas). Mock : cellules 293T non-transduites. 100-1, 50-1 à 50-3 : clones résistants à la zéocyne. pBABE HAGIZ : témoin positif (expression transitoire dans cellules 293T). GST rec : GST recombinante produite dans *E. coli* sans épitope HA.

2.5 Clonage d'une librairie de nonapeptides dans le vecteur rétroviral pBABE HAGIZ

Tout au long de l'analyse du fonctionnement du vecteur rétroviral pBABE HAGIZ, nous avons mis au point le clonage de la librairie dans ce vecteur. Comme les résultats d'expression de GST et de résistance à la zéocyne ont montré que le vecteur ne pourrait pas être utilisé dans sa version actuelle, les essais de clonage de la librairie n'ont pas été poussés plus loin que la ligation pilote. Les résultats tirés de ces expériences sont cependant intéressants. En effet, l'analyse des clones positifs issus des réactions pilotes de ligation a permis de montrer que l'efficacité de clonage dans ce vecteur était beaucoup plus grande que dans le vecteur épisomal, le pourcentage de clones positifs obtenus par digestion se rapprochant du 70% d'efficacité. L'observation de ces données laissait donc présager un gain au niveau de la diversité finale de la librairie par rapport à la librairie clonée dans le vecteur épisomal. Certains des clones positifs ont été séquencés afin de s'assurer que le clonage permettait de générer une librairie exempte de mutations, de changements de cadre de lecture, etc. Les données de séquençage sont présentées dans la section suivante, de pair avec ceux des peptides provenant du vecteur pCEP GST, pour montrer la diversité des peptides obtenus et discuter de la fréquence d'intégration des différents acides aminés.

2.6 Évaluation de la diversité des séquences peptidiques clonées dans les vecteurs pCEP GST et pBABE HAGIZ

Sept clones isolés à partir de la librairie clonée soit dans pCEP GST ou dans pBABE HAGIZ ont été séquencés afin de déterminer la fréquence de chaque acide aminé et de s'assurer qu'il n'y avait pas de biais au niveau de l'introduction de certains acides aminés par rapport à d'autres. Le tableau I montre la fréquence relative de chacun des acides aminés retrouvés dans la librairie. Les séquences de chaque peptide sont reproduites ci-dessous. On peut comparer dans le tableau I la fréquence relative de chaque acide aminé

(y compris ceux qui ne sont pas représentés dans notre échantillonnage de peptides) retrouvés dans nos 7 peptides avec la fréquence théorique attendue selon la représentation de chaque codon dans le code génétique . La fréquence théorique attendue des peptides est calculée selon la proportion de codons du code génétique codant pour chaque acide aminé.

Peptide 1.	TTT GCT CGT TGT CTG GGT TGT TTG TCG Phe ala arg cys leu gly cys leu ser
Peptide 2.	GGT TGC AAT ACT TTG CTT TCT CTC GGT Gly cys asn thr leu leu ser leu gly
Peptide 3.	CCC GTT TCT TCC ATT ATT GCT TTG GAT Pro Val Ser Ser Ile Ile Ala Leu Asp
Peptide 4.	TTG GCT TAG AAC GAG ATT AAC TGG TTT Leu Ala Ter* Asn Glu Ile Thr Trp Phe
Peptide 5.	GCC TGG TTT TGT ATT ATC CTC TTT GCG Ala Trp Phe Cys Ile Ile Leu Phe Ala
Peptide 6.	TTG CTG TGT CCT CTG GGT TGT ATG GGG Leu Leu Cys Pro Leu Gly Cys Met Gly
Peptide 7.	TTG GGT TGT GTT TCC CTT CCT GCT GCT Leu Gly Cys Val Ser Leu Pro Ala Ala

* : **Codon stop.**

Le nombre restreint de peptides présentés ici ne permet une analyse statistique représentative de la fréquence des acides aminés présents dans la librairie. Cependant, on peut conclure qu'il n'y a pas de biais majeur d'intégration d'un acide aminé par rapport à un autre qui serait du à une erreur de synthèse, les fréquences relatives de chaque acide aminé ne dépassant en effet pas beaucoup la fréquence attendue. On remarque également la présence d'un codon stop présent dans 1 peptide sur 7 (14.3%). Cette présence de codon stop pourrait devenir un facteur affectant à la baisse la diversité réelle de la librairie.

Acide aminé	Fréquence	% théorique	% relatif
Ala	7	6.2	11%
Arg	1	9.4	2%
Asn	2	3.1	3%
Asp	1	3.1	2%
Cys	7	3.1	11%
Gln	0	6.2	0%
Glu	1	3.1	2%
Gly	6	6.2	10%
His	0	3.1	0%
Ile	5	3.1	8%
Leu	13	9.4	21%
Lys	0	3.1	0%
Met	1	3.1	2%
Phe	4	3.1	6%
Pro	3	6.2	5%
Ser	5	9.4	8%
Stop	1	-	2%
Thr	2	6.2	3%
Trp	2	3.1	3%
Tyr	0	3.1	0%
Val	2	6.2	3%
total	63		

Tableau I. Séquence de 7 peptides de la librairie clonée dans pCEP GST et pBAGE GST et fréquence relative des différents acides aminés présents dans ces peptides par rapport à la fréquence théorique attendue.

3. DÉVELOPPEMENT DU SYSTÈME DE CRIBLAGE DE LA LIBRAIRIE EN CELLULES DE MAMMIFÈRES

Dans le but de mettre au point un système de criblage efficace qui nous permettrait de valider le principe de notre librairie de nonapeptides, nous avons établi une lignée cellulaire exprimant de façon stable la TK de HSV dans des cellules 293. Ce système permettra une sélection de peptides inhibiteurs de la TK en présence de GCV, un agent antiviral tuant les cellules HSVTK+.

3.1 Analyse de l'expression du gène de la TK de HSV dans les cellules de la lignée cellulaire 293TK+

L'expression de la fusion GFPTK+ transduite dans les cellules 293TK+ a été déterminée par immunobuvardage anti-GFP et par cytométrie en flux. L'immunobuvardage anti-GFP révèle une bande autour de 83 kDa qui correspond au poids attendu de la fusion GFPTK (GFP=28kDa + TK=55 kDa). On peut remarquer dans le même puits une bande de GFP seule correspondant probablement à un produit de clivage de la fusion. La proportion de cellules de la lignée 293TK+ exprimant la fusion GFPTK a été déterminée par cytométrie en flux. Comme le montre le panneau B de la figure 17, plus de 90% des cellules 293TK+ expriment la fusion GFPTK.

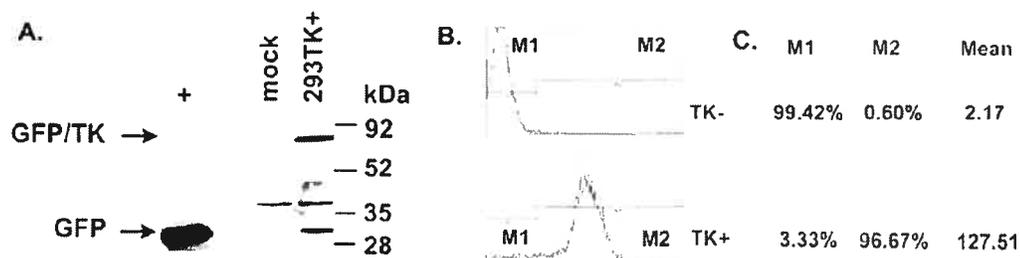


Figure 17. A. Expression de la fusion GFP/TK dans la lignée cellulaire 293TK+ détectée par immunobuvardage anti-GFP. + : pEGFP-N1. Mock : cellules 293. B. Analyse de l'expression de la fusion GFP/TK+ dans les cellules par cytométrie de flux. TK- : cellules 293. TK+ : cellules 293TK+. C. Statistiques d'expression de la TK dans les cellules 293TK+ Mean : moyenne de fluorescence.

3.2 Évaluation de la capacité du système HSVTK/GCV à induire la mort cellulaire des cellules 293TK+

Deux expériences indépendantes, chacune effectuée en triplicata, ont permis de valider le système de sélection HSVTK+/GCV dans les cellules 293TK+ et de déterminer l'IC₅₀ du GCV. Après 4 jours d'exposition à différentes quantités de GCV, on observe une mortalité de 50% dans les cellules TK+

dès 5 μM alors que les cellules contrôles ne sont aucunement affectées et ce, jusqu'à des concentrations de 100 μM . Afin de travailler dans des conditions optimales de sélection et selon les protocoles utilisés dans la littérature (qui fait état de concentrations utilisées équivalant à 10X la IC_{50}) (78), ces données de toxicité du GCV indiquent qu'il faudrait utiliser des concentrations de 50 μM de GCV pour la sélection de peptides inhibiteurs de la TK et pour s'assurer d'un taux de mortalité se rapprochant de 100% des cellules 293TK+.

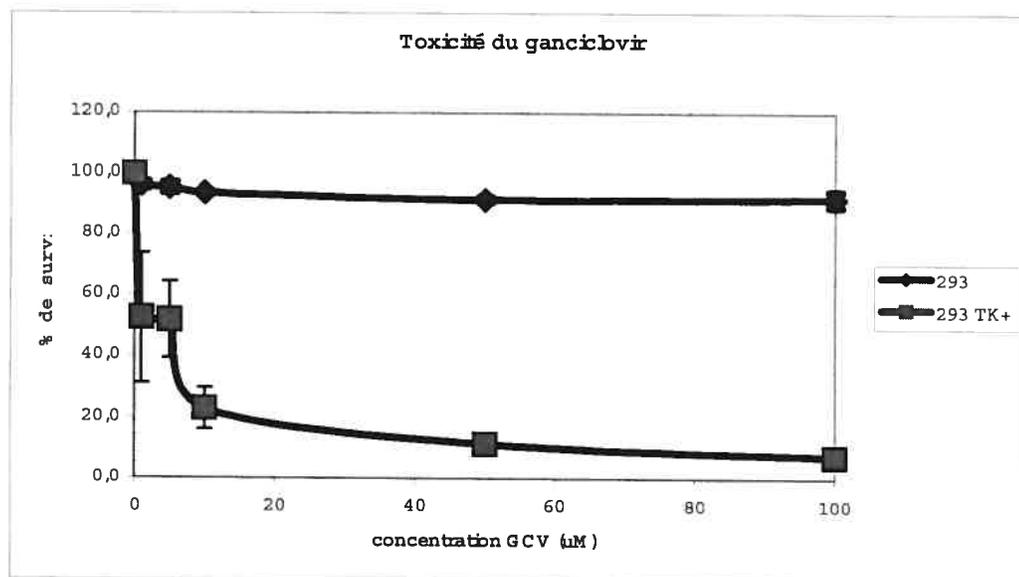


Figure 18. Toxicité du ganciclovir dans des cellules 293 exprimant la TK.

SECTION DISCUSSION

Avec l'avènement de l'ère post-génomique, de plus en plus d'équipes de recherche tentent de développer des moyens permettant de caractériser les relations complexes intervenant entre les protéines des multiples voies de signalisation intracellulaires. Différentes méthodes, tant *in vitro* que *in vivo*, permettent aujourd'hui de caractériser plus en profondeur les interactions inter-protéines régissant les mécanismes intimes de la cellule. Bien qu'elles possèdent leurs avantages, ces méthodes ne permettent pas encore d'étudier directement les interactions protéiques fonctionnelles à l'intérieur de cellules de mammifères, et ne prennent pas avantage de la possibilité de travailler dans ce contexte beaucoup plus proche de la réalité.

Afin de pallier aux différentes lacunes des méthodes existantes, nous nous sommes fixés comme objectif de construire et de mettre au point une librairie de nonapeptides pouvant être exprimée de façon stable dans des cellules de mammifères afin de montrer qu'il est possible d'isoler des peptides inhibiteurs d'interactions fonctionnelles protéine-protéine. Deux approches ont été chronologiquement abordées au cours de ce projet : dans un premier temps, nous avons décidé de construire la librairie dans un vecteur épisomal (pCEP4). Dans un deuxième temps, et suite aux difficultés rencontrées avec le premier vecteur utilisé, nous avons décidé de développer la librairie dans le cadre plus versatile d'un vecteur rétroviral, suite à quoi nous nous sommes encore une fois butés à des difficultés propres à ce système. Cette discussion fait état du cheminement rationnel suivi dans le développement d'une approche de librairie de peptides pouvant être exprimée en cellules de mammifères et surtout, des voies de développement futures pouvant être entreprises à la suite des difficultés rencontrées au cours de ce projet.

1. Clonage de la librairie dans le vecteur épisomal pCEP4

Nous avons choisi d'utiliser le vecteur pCEP4 (voir carte en annexe) comme base de notre librairie pour différentes raisons. Comme pCEP4 est un vecteur épisomal, il se

réplique à l'extérieur du chromosome et son expression n'est pas affectée par la position de son intégration dans le génome lors d'une expression à long terme. La répllication épisomale de ce vecteur rend plus simple la récupération de la séquence du vecteur en éliminant une étape de PCR génomique. Ce vecteur possède également un gène de résistance à l'hygromycine permettant de sélectionner à long terme des cellules exprimant la librairie. La présence d'un promoteur fort (CMV) régissant l'expression de la librairie a aussi penché dans la balance au moment du choix initial du vecteur.

L'approche de clonage de la librairie dans le vecteur épisomal pCEP4 consistait à introduire dans le site de clonage multiple du vecteur le gène de la GST, modifié dans sa partie C-terminale pour comprendre deux sites de restriction permettant le clonage de la librairie et un linker de 6 acides aminés. On pouvait ensuite y cloner l'oligonucléotide double-brin codant pour la librairie de 9 acides aminés aléatoires.

Nous avons donc commencé par cloner dans pCEP4 le gène modifié en C-terminal de la GST pour accueillir la librairie. Un immunobuvardage anti-GST (figure 9) a permis de montrer que le gène modifié était bien exprimé, du moins de façon transitoire, dans des cellules 293T à partir de pCEP GST (voir carte en annexe) et qu'on pouvait passer à la prochaine étape, soit le clonage de la librairie dans le vecteur pCEP GST. Le paragraphe 1.3 de la section résultats fait état d'une diversité finale de 1.25×10^4 clones positifs différents. Si on se rappelle les données de la littérature mentionnées dans l'introduction, qui établissent la fréquence de sélection d'un peptide bioactif dans une expérience de criblage à un sur un million (1×10^{-6}), on constate que nous sommes à presque deux ordres de grandeur sous la norme établie par la littérature. Cela signifie que nos chances d'isoler un peptide bioactif sont très minces puisqu'il se peut que dans les 12 500 peptides de la librairie, aucun ne soit apte à interagir avec un des composants de notre système de criblage. Il pourrait cependant être intéressant de comparer l'efficacité de notre librairie dans différents systèmes et de voir si on ne pouvait pas y isoler un peptide actif. Des expériences menées dans notre laboratoire ont montré que malgré le fait qu'une librairie n'ait été

criblée que dans une fraction de sa diversité, le nombre de peptides identifiés dépassait largement les valeurs attendues (86). Il se pourrait également qu'une certaine catégorie de peptides comportant des propriétés biochimiques communes et pertinentes à notre système puissent être surreprésentés dans la librairie par un effet stochastique et qu'il demeure possible d'en identifier quelques-uns malgré la faible diversité de notre librairie.

Même si l'on peut espérer isoler quelques peptides à partir d'une librairie de faible diversité, la question fondamentale à se poser est plutôt la suivante : pourquoi n'avons-nous pas réussi à cloner notre librairie avec une efficacité assez grande pour pouvoir la cribler et espérer isoler plusieurs peptides bioactifs? Dans le cas qui nous concerne, c'est-à-dire un clonage à grande échelle dans le vecteur pCEP GST, plusieurs facteurs ont été déterminants dans l'obtention d'un faible nombre de transformants. Premièrement, nous faisons face à un plateau pratique au niveau de l'efficacité de transformation dans des bactéries compétentes, les bactéries électrocompétentes disponibles sur le marché ne pouvant en effet pas dépasser le seuil des 10^{10} transformants par μg d'ADN transformé (il est également à noter qu'il est inutile de faire des transformations avec plus de 50 ng d'ADN à la fois, la réaction devenant saturée et le rendement ne s'en trouvant pas augmenté). Ce plateau fixe la limite supérieure de transformants pouvant être obtenus par des expériences d'une ampleur technique raisonnable. Imaginons un instant que nous voulions obtenir un rendement 10 fois supérieur à celui que nous avons présentement en augmentant la quantité d'ADN transformé. On devrait transformer 10 μg d'ADN, c'est-à-dire effectuer plus de 200 électroporations de 50 ng d'ADN dans des bactéries électrocompétentes. Chaque ordre de grandeur gagné au niveau de la diversité l'est au prix d'une hausse disproportionnée de la complexité des procédures expérimentales. Il devient ainsi pratiquement difficile d'obtenir des librairies de diversités plus grandes que 1×10^8 .

Un autre facteur ayant fait baisser la diversité finale de la librairie découle d'une caractéristique intrinsèque de la méthode de clonage classique, qui comprend une

étape de digestion enzymatique, une étape de ligation et une étape de transformation. Comme nous l'avons appris par expérience au cours du clonage de cette librairie, les vecteurs ligués se transforment dans des bactéries compétentes avec une efficacité inférieure de deux logs à celle d'un vecteur surenroulé. Ainsi, en supposant que les bactéries électrocompétentes utilisées donnent le rendement promis (10^{10} transformants/ μg d'ADN), on tombe à une possibilité maximale de 10^8 transformants différents pour chaque microgramme d'ADN ligué. Il faut ensuite compter sur une perte de diversité d'environ 50% si on tient compte du nombre de transformants ayant réellement intégré la librairie (le nombre de 50% est une moyenne approximative de multiples essais de ligation dans les vecteurs pCEP GST et pBAGE HAGIZ avec la librairie de nonapeptides digérée *XhoI/NotI*). On tombe dès lors à une diversité de 5×10^7 peptides différents, ce qui n'est plus qu'à un log de la fréquence théorique.

Finalement, et nous pouvons dire *a posteriori* que c'est là un des facteurs les plus importants à considérer lors de l'ébauche d'un projet de clonage à grande échelle, la taille du vecteur influence énormément son efficacité de transformation. En effet, tel qu'il a été montré par la pratique et comme le soulignent plusieurs fabricants de bactéries compétentes, il devient de plus en plus difficile de transformer des vecteurs de plus en plus gros. Ainsi, les bactéries que nous avons utilisées pour cloner la librairie dans pCEP GST (un vecteur de 11 kb) ne peuvent donner qu'un rendement de 10^7 transformants par microgramme d'ADN (d'un vecteur surenroulé de taille supérieure à 10 kb) transformé, et ce malgré le fait qu'elles aient été optimisées pour la transformation de ces gros vecteurs (>10 kb). Si on applique cette baisse d'efficacité à notre diversité maximale telle que calculée plus haut, on en arrive à un total de 10^4 transformants par μg d'ADN ligué transformé. Toute cette démonstration met en lumière les difficultés inhérentes à l'utilisation du vecteur pCEP4 pour le clonage d'une librairie de grande diversité.

Cela dit, il pourrait être possible d'utiliser le vecteur pCEP GST, et donc de profiter des ses caractéristiques particulières, en modifiant l'approche traditionnelle que nous

avons utilisée. La taille du vecteur, qui constitue un obstacle majeur à l'obtention d'une librairie de grande diversité, pourrait être réduite de façon sensible en enlevant le gène codant pour la protéine EBNA-1 qui permet au vecteur de se répliquer de façon épisomale. Il serait possible de conserver cette propriété du vecteur en co-transfectant la librairie avec un expresseur d'EBNA-1. Bien sûr, dans une cellule transfectée par pCEP GST qui ne le serait pas par l'expresseur d'EBNA, le vecteur serait soumis aux effets régulateurs dus à la position d'intégration, ce qui pourrait diminuer la proportion de cellules exprimant effectivement un peptide. Dans un autre ordre d'idées, il existe maintenant des techniques de clonage par recombinaison homologue (INVITROGEN) qui permettent des efficacités de clonage se rapprochant de 100%, ce qui permettrait de gagner encore une puissance 10 et d'approcher la diversité théorique minimale de 10^6 peptides différents avec le vecteur pCEP GST.

Au cours des procédures de mise au point du clonage de la librairie dans pCEP GST, nous avons étudié la cinétique d'expression de GST lorsqu'exprimée de façon stable dans des cellules 293TK+, cellules dont nous voulions nous servir pour valider notre concept de librairie. Des cellules 293TK+ ont donc été transfectées avec le vecteur pCEP GST et mises sous sélection hygromycine pour une durée de 25 jours. Chaque 5 jours post-sélection, un aliquot de 2×10^6 cellules résistantes était récolté, lysé et congelé. Lorsque tous les échantillons ont été récoltés, la quantité de protéines totale de chaque échantillon était normalisée par rapport à l'échantillon le moins concentré, la GST purifiée par affinité et l'expression de GST analysée par immunobuvardage anti-GST (voir figure 10). On s'est rendu compte que l'expression de GST, bien qu'elle soit relativement durable, n'était pas stable dans le temps mais avait plutôt tendance à plafonner après une vingtaine de jour sous sélection puis à décroître assez rapidement. Cela nous met face à deux constatations : d'une part, il faudra déterminer une fenêtre d'application de la sélection dans le système de criblage à un moment où l'expression de la GST est stable. D'autre part, il doit exister un facteur relié soit au vecteur, soit au système cellulaire, qui rend l'expression de GST instable.

Face à une expression de GST variable dans le temps (et par extension des peptides de la librairie), il devient nécessaire de déterminer des critères qui permettront de déterminer le moment judicieux pour appliquer la sélection GCV dans notre système de criblage et ainsi augmenter nos chances d'isoler un ou plusieurs peptides biologiquement actifs. Quelqu'un qui met au point un système de criblage part en effet un peu dans le noir, i.e. il ne sait pas *a priori* quelles seront les conditions idéales de sélection des peptides. Devrait-on exercer la sélection au moment où la librairie est exprimée à son plus haut niveau, ce qui permettrait de travailler dans des conditions maximales d'expression des peptides et de favoriser des interactions de faible affinité avec des partenaires rares? Devrait-on au contraire appliquer la sélection au même moment que le début de l'expression de la librairie pour avoir une coévolution des deux protagonistes du système et ainsi se donner la possibilité d'isoler des peptides agissant sur des partenaires ayant des rôles précis dans le déroulement temporel du phénomène étudié? Est-ce que l'avantage sélectif conféré à une cellule par un peptide bioactif pourra permettre d'allonger la durée de son expression et faciliter son amplification et la récupération de sa séquence? De façon plus particulière, si on utilise les données d'expression à long terme de GST à partir de pCEP4, on peut se demander si 10-12 jours de sélection en présence de la librairie seront suffisants pour isoler des clones résistants au GCV. Nous ne croyons pas qu'il y ait de réponses parfaites à toutes ces questions. L'utilisation d'une librairie dont l'expression est circonscrite dans un laps de temps précis demande une adaptation, une expérimentation qui est propre à chaque système de criblage utilisé. Toutes ces questions sont cependant importantes et font référence à des facteurs que l'on doit considérer lorsqu'on veut utiliser une librairie dans un système de criblage.

Si on voulait utiliser la librairie telle que clonée dans le vecteur pCEP GST, on pourrait soit se demander dans quelle fenêtre l'utiliser afin d'en tirer le meilleur parti possible, soit essayer d'élucider la cause de la chute d'expression de GST et tenter d'y remédier afin de se donner la possibilité de travailler dans des conditions plus flexibles par la suite. Dans le cas qui nous concerne, la perte d'expression peut être

causée à 3 niveaux : 1) au niveau de la réplication du vecteur, 2) au niveau de la transcription du gène codant pour GST et 3) au niveau de la protéine GST (traduction, stabilité). Les effets de position de l'intégration d'une cassette d'expression lors d'une expression stable sont ici négligés puisque le vecteur se réplique à l'extérieur du chromosome et n'est donc pas susceptible à l'inhibition de la transcription due à la topologie de la chromatine. Comme les cellules demeurent résistantes à l'hygromycine, il semble que le vecteur soit toujours présent dans les cellules sous sélection. La perte d'expression de GST n'est donc vraisemblablement pas due à une perte du vecteur ou encore à une diminution du nombre de copies présentes dans la cellule. D'autre part, la stabilité de la protéine elle-même ne semble pas affectée puisqu'elle est détectable à des niveaux relativement élevés pendant plusieurs jours post-transfection. Un contrôle de la traduction n'est donc pas très probable puisque les niveaux de GST demeurent constants pendant une période de temps assez longue. Il serait plutôt inusité de voir un mécanisme de régulation de la traduction de GST émerger 20 jours après le début de l'expression stable de la protéine dans les cellules. Ce mécanisme paraît d'autant plus improbable que la protéine GST n'est pas toxique pour la cellule. Il n'y a donc pas de raisons de croire à un contrôle de la traduction de la protéine.

L'alternative restante est un contrôle de l'expression de GST au niveau de la transcription du gène. Le vecteur pCEP4 sert à exprimer des protéines sous le contrôle du promoteur viral CMV, un promoteur fort utilisé dans plusieurs approches expérimentales. Comme le promoteur CMV est un promoteur fort, plusieurs groupes ont tenté de développer des approches de thérapie génique l'utilisant afin d'obtenir une expression forte de leur transgène (26, 47, 72). Ces groupes se sont cependant rendu compte que l'expression de protéines à partir du CMV, bien qu'elle soit beaucoup plus forte qu'à partir de la majorité des promoteurs couramment utilisés dans les vecteurs d'expression, n'était pas maintenue de façon stable dans le temps. Ainsi, le groupe de Löser et al. rapporte que, peu importe le vecteur ou le système cellulaire utilisé, l'effet du promoteur CMV chute dramatiquement dans un délai de 2 à 4 semaines post-transfection (47). Un autre groupe rapporte ce type de régulation tout

en excluant par des expériences de PCR que cette perte d'expression soit due à la perte du vecteur, qui reste détectable pour plusieurs semaines (28). Il semblerait donc que le promoteur CMV, par le fait qu'il contienne des îlots CpG (séquences favorisant la méthylation de l'ADN (52)), soit tôt ou tard réprimé par des mécanismes de contrôle cellulaires (31, 59). D'autres groupes ont comparé la stabilité de l'expression de protéines à partir de différents promoteurs et semblent faire consensus pour dire que le promoteur du facteur d'élongation EIF- α est beaucoup plus stable pour des utilisations à long terme dans des lignées cellulaires (28, 72). Ces articles permettent d'expliquer la chute d'expression de GST à partir du promoteur CMV de notre vecteur et mettent en évidence l'importance du choix du promoteur dans la construction d'une librairie pouvant être exprimée de façon stable.

L'approche de clonage d'une librairie dans pCEP GST a donc quelques lacunes importantes. En conjonction avec la perte de diversité inhérente à chaque étape du clonage, la taille du vecteur a empêché l'obtention d'une librairie de grande diversité. De plus, l'utilisation du promoteur fort CMV mène vraisemblablement à une régulation négative de l'expression à long terme de la librairie. Cette régulation négative soulève beaucoup de questions quant à la possibilité de sélectionner des peptides bioactifs à partir d'une population exprimant la librairie de façon stable. Les problèmes rencontrés au cours de la construction de la librairie dans le vecteur pCEP GST nous ont amené à considérer une autre approche de clonage qui contournerait les difficultés propres au système pCEP GST et nous permettrait d'aller de l'avant avec notre approche. Nous nous sommes donc tournés vers un système de vecteur rétroviral.

2. Clonage de la librairie dans le vecteur rétroviral pBABE HAZ

Les vecteurs rétroviraux sont utilisés depuis plusieurs années pour transférer des gènes d'intérêt dans une cellule ou un organisme, particulièrement dans des expériences de thérapie génique. Ces vecteurs servent à produire des particules virales infectieuses pouvant infecter différents types de cellules et à y transférer un gène

d'intérêt. Par un procédé appelé pseudotypage, il est possible de changer l'enveloppe de ces particules virales et de modifier leur tropisme afin d'infecter une gamme plus large de cellules. Les vecteurs rétroviraux présentent plusieurs avantages par rapport aux vecteurs conventionnels pour notre approche de clonage de librairie. Leur petite taille permet d'atteindre une efficacité de transformation plus élevée. Leur tropisme variable permet d'adapter la librairie à toutes sortes de systèmes rapporteurs, la rendant ainsi plus versatile et donc, plus polyvalente pour l'étude de multiples phénomènes biologiques. Le vecteur que nous avons utilisé (pBABE HAGIZ) provient d'un rétrovirus simple, le MuLV, et est dérivé d'un vecteur (pBABE HAZ, voir carte en annexe) ayant déjà été utilisé pour cribler une librairie d'ADNc (24).

La première étape dans la construction de la librairie a été d'en concevoir la structure de la façon la plus efficace et polyvalente possible. Le vecteur original pBABE HAZ permettait de cloner une librairie en fusion N-terminale d'un gène de résistance à la zéocyne. Si nous avons utilisé le vecteur tel quel, notre librairie aurait été une fusion HA-GST-peptide-Zéo^R, ce qui aurait entraîné l'expression des peptides entre deux éléments protéiques de taille beaucoup plus importante, et aurait possiblement réduit l'exposition de notre peptide à ses partenaires potentiels. Afin de favoriser une exposition maximale de notre peptide, nous avons décidé de modifier le vecteur pBABE HAZ pour créer le vecteur pBABE HAGIZ (voir carte en annexe) en y clonant une cassette d'expression bicistronique qui allait permettre l'expression simultanée de la librairie (sous le contrôle du promoteur précoce de SV40) et du gène de résistance à la zéocyne (sous le contrôle d'un élément IRES provenant du virus EMCV). L'élément IRES de EMCV permet une traduction d'un gène de façon indépendante de la coiffe à partir d'un ARNm. L'IRES est en fait une structure secondaire de l'ARN qui permet l'entrée du ribosome au milieu d'un transcrit et l'initiation de la traduction d'un second cistron. Cette séquence comporte plusieurs codons d'initiation de la traduction, mais nous avons choisi plus particulièrement d'utiliser le 11^{ième}, identifié comme étant le codon d'initiation naturel du virus (19). Le gène de résistance à la zéocyne allait donc utiliser l'ATG de l'IRES et être traduit

sous son contrôle. L'expression bicistronique de notre librairie et du marqueur de sélection semblait à prime abord receler beaucoup d'avantages : les cellules résistantes à la zéocyne allaient nécessairement exprimer la librairie puisque la transcription des deux gènes est reliée, seule la traduction de chaque gène se faisant de façon indépendante. Les peptides exprimés en C-terminal de GST allaient être librement exposés, ce qui augmente les chances d'identifier des interactions fonctionnelles protéine-protéine.

Afin de vérifier si notre vecteur avait vraiment toutes les caractéristiques que nous voulions lui conférer, nous avons complètement séquencé toute la partie du vecteur codant pour l'IRES, pour le gène de résistance à la zéocyne et la portion C-terminale de GST (pour s'assurer de la présence d'un codon stop qui allait stopper la traduction de la librairie fusionnée à GST). Les séquences ont montré que notre cassette d'expression bicistronique ne comportait pas de mutation par rapport aux séquences publiées, ce qui indiquait qu'on pouvait aller de l'avant. L'étape suivante a été naturellement de montrer que le vecteur pouvait exprimer la protéine GST correctement. La figure 14 montre que le vecteur pBABE HAGIZ peut exprimer GST de façon transitoire. L'immunobuvardage montre clairement une bande de taille supérieure à celle de la protéine GST contrôle. Cette augmentation s'explique par la présence d'un épitope HA en N-terminal de GST qui augmente la taille de la protéine. Tous les doutes persistant quant à l'identité réelle de la bande de GST-HA se dissipent lorsqu'on observe l'immunobuvardage anti-HA fait sur un échantillon provenant de cellules transfectées par pBABE HAGIZ (figure 16). On peut en effet observer exactement la même bande que dans l'immunobuvardage anti-GST, à la même hauteur. La présence de bandes supplémentaires dans les immunobuvardages effectués avec le vecteur pBABE HAZ s'explique par le fait que les billes de glutathione sépharose ont été bouillies, ce qui engendre parfois la relâche de bandes réagissant avec l'anticorps anti-GST. Nous avons conclu à partir de ces données que le vecteur rétroviral pouvait exprimer la protéine GST de façon transitoire à partir du promoteur SV40.

Nous devons ensuite montrer que le vecteur était fonctionnel et qu'il pouvait être incorporé dans des particules virales sous forme d'ARN, être transduit dans des cellules, s'incorporer à leur génome et leur conférer une résistance à la zéocyne, d'une part et, d'autre part, permettre l'expression de la librairie. Les particules virales ont été produites selon le protocole habituel, qui consiste à cotransfecter un expresseur de gag-pol de MuLV, un expresseur d'enveloppe VSV-G (qui confère un tropisme universel et permet d'atteindre des efficacité de transduction approchant 100%) et le vecteur rétroviral dans des cellules 293T hautement transfectables. Les virus récoltés et concentrés sont ensuite resuspendus dans un faible volume de milieu en aliquots de 50 μ l. Nous avons fait un essai enzymatique mesurant l'activité de la RT incorporée dans la particule virale afin de s'assurer que la production virale avait bien fonctionné. Comme on peut le voir à la figure 13, l'activité de la RT dans l'échantillon de pBABE HAGIZ dépasse de plusieurs dizaines de fois celle du bruit de fond. Nous avons utilisé ces données de RT afin d'utiliser un ratio constant de virus par cellule pour effectuer l'infection.

Nous avons approché le problème de la démonstration du fonctionnement de notre vecteur de deux façons différentes. Nous avons d'une part transfecté le vecteur dans des cellules 293T et procédé à des expériences de sélection à la zéocyne sur les cellules transfectées. Comme nous travaillions dans des cellules hautement transfectables, l'hypothèse était que les niveaux de transfection et d'infection seraient comparables. Nous avons donc essayé de sélectionner avec la zéocyne des cellules transfectées avec pBABE HAGIZ, des cellules transfectées avec pBABE HAZ (dans lequel le gène de la zéocyne est dans un mauvais cadre de lecture et est inactif), des cellules transfectées avec pBUD-CE4-zéo (expresseur du gène de résistance à la zéocyne) et des cellules non-transfectées. À la suite de plusieurs jours de sélection (entre 14 et 21), nous avons observé que les cellules transfectées avec pBUD-CE4-Zéo avaient recommencé à croître de façon normale, le pétri étant recouvert de clones résistants. Au contraire, les pétris contenant les cellules contrôle négative ne présentaient aucun foyer de croissance, indiquant que la sélection avait été efficace et n'avait pas permis la formation de clones spontanés. Le pétri de cellules transfec-

tées avec pBABE HAGIZ ne comportait pour sa part que 7 clones résistants, énormément moins que le pétri contrôle positif. Il semblait donc y avoir un problème avec l'expression du gène de résistance à la zéocyne.

Pendant que nous sélectionnions les cellules transfectées, nous avons également fait un essai de transduction avec les particules virales produites à partir de pBABE HAGIZ. Dans ce cas, nous n'avions pas de contrôle positif mais des cellules non transduites ont été utilisées pour confirmer l'efficacité de la sélection à la zéocyne. Nous avons procédé à la transduction de cellules 293T et sélectionné les cellules de la même façon que dans l'expérience de transfection du vecteur rétroviral. Après un laps de temps comparable (2 à 3 semaines), les cellules contrôles non transduites étaient toutes mortes alors qu'il y avait eu apparition de 4 clones résistants dans le pétri de cellules transduites. Comme on s'attendait à avoir une efficacité de transduction frôlant les 100%, ce nombre très restreint de clones résistants venait confirmer les soupçons éveillés par les résultats de l'expérience de transfection, à savoir qu'il y avait réellement un problème d'expression du gène de résistance à la zéocyne. Les résultats de sélection de clones résistants à la zéocyne obtenus par transfection et par transduction sont très semblables si on les compare au témoin positif, ce qui laisse présumer une efficacité de transfert du gène de résistance semblable pour les deux méthodes. Cela nous a conforté dans la décision de poursuivre les expériences avec les clones provenant de l'expérience de transduction.

Comme nous devions déterminer si le vecteur exprimait la GST dans les clones résistants à la zéocyne, nous avons décidé de continuer l'amplification des clones résistants obtenus par transduction parce qu'ils correspondaient exactement à la méthode que nous allions utiliser avec la librairie. Une fois les clones résistants amplifiés de façon à avoir assez de cellules pour faire un immunobuvardage et pour extraire l'ADN génomique, nous avons analysé l'expression de la GST par immunobuvardage anti-GST et anti-HA (figure 16). Les immunobuvardages anti-HA et anti-GST de la figure 16 montrent l'absence totale d'expression de GST dans les clones résistants à la zéocyne. Cette absence d'expression de la protéine GST pou-

vait être due au fait que les clones résistants isolés n'étaient en fait que des clones spontanés dans lesquels la cassette d'expression de la librairie et du gène de résistance à la zéocyne ne s'étaient pas vraiment intégrés. Nous avons donc voulu vérifier que le gène codant pour la résistance à la zéocyne était bel et bien intégré au génome, ce qui fut confirmé par un PCR génomique utilisant des amorces permettant l'amplification du gène de la zéocyne (figure 15). Ces résultats de PCR montraient que le gène de résistance à la zéocyne était présent dans le génome des cellules résistantes et permettaient de croire que les clones étaient de vrais clones résistants à la zéocyne. La situation posait alors deux problèmes : le rendement de sélection de cellules résistantes à la zéocyne était très faible et ne correspondait absolument pas aux nombres attendus et nécessaires pour l'utilisation de notre librairie. Qui plus est, le peu de clones résistants n'exprimaient pas le gène de la librairie, ce qui excluait l'utilisation du vecteur dans ces conditions pour le clonage de la librairie. De nombreuses questions pouvaient être soulevées face à ces résultats : pourquoi le gène de résistance à la zéocyne n'est-il pas exprimé de façon optimale? Pourquoi la GST n'est-elle pas exprimée dans les clones résistants? Pourquoi la GST est-elle exprimée de façon transitoire mais pas de façon stable? Y-a-t-il un lien entre l'expression bicistronique de notre librairie et du gène de résistance et ce comportement inattendu du vecteur?

Il existe un bon nombre d'évidences tant publiées que non publiées qui expliquent les difficultés rencontrées dans la mise au point d'un système d'expression bicistronique (communication personnelle, Stephen P. Goff). Il semblerait en effet que ce type d'expression soit sujet à de nombreux types de régulation qui empêchent une expression simultanée efficace et stable des deux gènes, ce qui pose particulièrement problème dans une approche comme la nôtre. À cet effet, le groupe du Dr Goff a opté pour l'option de la fusion (de la librairie avec le gène de résistance) au détriment de celle de la cassette bicistronique. Ce choix était principalement basé sur l'impérative nécessité d'avoir une expression simultanée du marqueur de sélection et de la librairie pour s'assurer le plus de chances possible d'identifier des interactions intéressantes. La façon la plus sûre et la plus directe d'y arriver consiste à fusionner

les deux gènes. On s'assure ainsi qu'un clone résistant est aussi un clone qui exprime un membre de notre librairie.

À la lumière de ces évidences nouvelles, nous avons pu corréler les résultats de notre expérience de transduction du vecteur pBABE HAGIZ avec ceux que nous avons obtenus auparavant lors de la création de la lignée cellulaire exprimant la TK de HSV. Cette lignée, créée par la transduction de cellules 293 avec le vecteur rétroviral bicistronique HPV 434, (le gène de résistance à la néomycine étant sous le contrôle d'un promoteur EIF- α et la fusion GFPTK sous le contrôle de l'IRES de EMCV, voir carte en annexe), avait été obtenue par triage par FACS des cellules exprimant la fusion GFPTK. Les cellules positives avaient ensuite été mises sous sélection au G418 et la lignée avait pu être établie. Une chose qui était passée inaperçue alors devenait maintenant très significative. Lors du tri des cellules 293, seulement un faible pourcentage (autour de 1%) des cellules exprimaient la fusion GFPTK, ce qui était surprenant étant donné le fait que nous avons utilisé un vecteur rétroviral pseudotypé avec VSVG. Le phénomène était resté inexpliqué puisque l'objectif final, i.e. l'établissement d'une lignée TK+, avait été atteint. En comparant ces données avec nos derniers résultats, et en considérant la communication avec le Dr Goff, on se rend compte que toutes les données concordent : l'expression d'un gène à partir d'un IRES lors d'une expérience de transduction ne permet pas d'atteindre les hauts pourcentages d'expression auxquels on s'attendrait avec un vecteur rétroviral pseudotypé VSVG. De fait, le faible niveau d'expression à partir de l'IRES réduit énormément le nombre de clones positifs résistants à la zéocyne.

Ces constatations nous ont amené à faire une recherche exhaustive de la littérature et, encore une fois, les données concordent : l'expression d'un gène à partir d'un IRES est généralement beaucoup plus faible que l'expression du gène placé en amont de cet IRES sous le contrôle d'un promoteur « cap-dépendant ». Hennecke et ses collègues ont démontré en 2001 que la position relative de deux gènes par rapport à un élément IRES pouvait grandement influencer le niveau d'expression de ces gènes (36). De plus, ils ont montré que non seulement la position des gènes influen-

çait leur expression, mais que la nature même de ces gènes pouvait également moduler leur expression. D'autres rapportent également des niveaux d'expression pouvant diminuer jusqu'à 10% de ceux du gène situé en amont de l'IRES (14, 24, 49, 88). C'est donc dire que chaque expérience, chaque vecteur dans lequel on veut exprimer deux gènes simultanément à l'aide d'un IRES, sont soumis à des variations inévitables du niveau d'expression et qu'il n'existe pas de règles pour déterminer l'importance de cette régulation. Une solution à ce problème serait d'inverser l'ordre du gène de résistance et de la librairie. On pourrait alors compter sur une expression plus forte du gène de résistance et isoler plus de clones résistants. Cette façon de faire pourrait toutefois causer une perte d'expression de la librairie exprimée sous le contrôle de l'IRES, ce qui réduirait le nombre de clones exprimant des peptides et ainsi l'efficacité totale de la librairie. Certains groupes (dont nous) ont utilisé l'approche d'expression bicistronique à partir d'un IRES de façon concluante (23), mais ont dû procéder à une double sélection des cellules positives pour en arriver à établir une population de cellules exprimant correctement la cassette bicistronique. Cela met en relief une différence majeure entre ces expériences et la nôtre : alors qu'eux n'ont théoriquement besoin que d'un clone pour établir une lignée stable, nous cherchons à sélectionner des populations entières exprimant deux gènes simultanément. Il semble donc que l'utilisation d'une cassette d'expression bicistronique ne soit pas assez fiable pour qu'on puisse l'appliquer à notre librairie.

À la lumière de ces observations, il est devenu clair que notre vecteur rétroviral, du moins dans sa configuration actuelle, ne pouvait pas être utilisé pour cloner la librairie et qu'il fallait retourner à la planche à dessin pour concevoir un système pouvant être exprimé de la façon la plus uniforme possible, dans le plus de cellules possible et le plus stablement possible. Pour ce faire, et en se basant sur l'apprentissage que nous avons fait en mettant au point les deux approches de clonage de notre librairie de nonapeptides, il faudrait trouver une façon d'exprimer la librairie en fusion avec une protéine (idéalement un marqueur de résistance), sous le contrôle d'un promoteur fort mais surtout stable (comme le promoteur de E1F- α). Cette fusion serait clonée dans un vecteur rétroviral pouvant avoir un large spectre d'hôtes et

pourrait être utilisée pour transduire des populations de cellules. Le vecteur serait idéalement petit afin de favoriser l'efficacité de transformation lors du clonage de la librairie. En réunissant toutes ces conditions, il deviendrait possible de travailler avec une librairie puissante pouvant être exprimée dans une multitude de contextes. Cette librairie pourrait permettre l'identification de partenaires nouveaux de protéines impliquées dans la pathogénèse virale, les maladies génétiques, etc.

3. Mise au point d'un système de criblage de la librairie en cellules de mammi-fères

Le complément naturel d'une librairie bien conçue et bien construite, son pendant cellulaire, est un système de criblage efficace, assez discriminatoire sans l'être trop, permettant de retrouver la séquence des peptides isolés pour leur activité perturbatrice au sein de ce système. Pour ce faire, nous avons choisi d'utiliser un système cellulaire exprimant la thymidine kinase du virus *Herpes simplex* en combinaison avec la prodrogue ganciclovir (GCV). Dans ce système, le GCV est phosphorylé par la TK de HSV puis est pris en charge par les kinases cellulaires, incorporé à l'ADN cellulaire, entraînant la mort de la cellule par des mécanismes encore mal connus. Nous voulions utiliser ce système comme modèle pour identifier des peptides qui inhiberaient l'activité de la TK et ainsi démontrer la validité du concept de librairie de nonapeptides.

La lignée cellulaire 293TK+ a été créée en transduisant des cellules 293 avec un vecteur rétroviral codant pour une cassette d'expression bicistronique. La cassette comprend un gène de résistance à la néomycine (G418) exprimé sous le contrôle du promoteur du facteur de transcription E1F- α ainsi qu'une protéine de fusion GFPTK sous le contrôle d'un IRES de EMCV (exactement comme la cassette de notre vecteur rétroviral pBABE HAGIZ sauf pour le promoteur E1F- α). Après la transduction, les cellules ont été triées par FACS selon leur niveau d'expression de GFP et les cellules GFPTK+ ont été remises en culture sous sélection G418. Elles ont été amplifiées afin de constituer des stocks congelés. Avant la congélation, l'expression de la fusion GFPTK+ a été montrée par immunobuvardage anti-GFP (figure 17A).

L'analyse de l'expression protéique par immunobuvardage est utile en ce qu'elle permettra éventuellement de vérifier que les peptides isolés ont une action au niveau des interactions interprotéiques et non pas au niveau de l'expression de cette protéine. On pourra ainsi s'assurer que l'on a isolé des inhibiteurs fonctionnels de la protéine et ainsi valider le concept de notre librairie. Afin de s'assurer que l'expression de la fusion GFPTK+ était stable dans le temps, nous avons procédé à des analyses par FACS avant la congélation (figure 17B) et après chaque prélèvement des échantillons dans l'expression de GST à long terme dans les cellules 293TK+ (données non présentées). Ces analyses montrent que l'expression de la fusion est stable dans une proportion de 90% des cellules sur une période d'au moins 25 jours et qu'elle ne semble pas diminuer avec le temps. Ces données sont en accord avec celles des groupes ayant rapporté la stabilité supérieure du promoteur EIF- α par rapport à d'autres promoteurs couramment utilisés.

La stabilité de la lignée cellulaire 293TK+ ayant été démontrée, il fallait aussi montrer l'action cytotoxique du GCV. Pour ce faire, nous avons utilisé un test de prolifération cellulaire basée sur la capacité du MTT à être transformé en formazan (composé soluble violet) par les enzymes mitochondriales des cellules viables. Les résultats de la figure 18 montrent l'efficacité du GCV sur le système, de même que sa spécificité. Les cellules contrôle n'exprimant pas la fusion GFPTK ne sont en effet pas affectées par l'addition de GCV à des concentrations qui tuent par ailleurs plus de 90 % des cellules 293TK+ sur une période de 4 jours. Ces résultats confirment que le système pourra être utilisé afin de valider une librairie de nonapeptides. Il pourrait s'avérer utile de tester la sélection de peptides en présence de différentes concentrations de GCV pour évaluer l'incidence de la concentration sur la récupération de peptides actifs. L'essentiel est de déterminer des conditions qui sont assez stringentes pour éviter l'apparition de clones spontanés (de bruit de fond) mais qui permettent tout de même la sélection de quelques peptides.

4. Perspectives, applications futures et conclusion

Comme nous venons de le voir, le développement d'une librairie de nonapeptides en cellules de mammifère est une tâche complexe comportant de multiples écueils. Même si nous n'avons pas réussi à montrer que le concept pouvait fonctionner en cellules de mammifères, nous avons appris beaucoup sur la conception de ce type de librairie et nous croyons que cette technologie pourra être appliquée à plusieurs systèmes.

Le plus gros avantage des librairies de peptides, par rapport à d'autres types de librairie, réside dans la possibilité de faire des corrélations structure-fonction directement à partir des séquences du peptide et de la protéine cible. Comme on l'a dit dans l'introduction, les peptides sont l'expression la plus simple de leur effet, ils en sont le plus petit dénominateur commun. On pourra ainsi identifier des domaines de liaison, et la détermination des acides aminés importants pour cette liaison, et à plus forte raison pour l'interaction fonctionnelle, sera facilitée.

La mise au point d'une librairie contenant une grande diversité de peptides est selon nous une approche extrêmement intéressante de par sa versatilité et sa puissance. Il existe déjà dans l'industrie des systèmes de criblage de molécules organiques à grande échelle (high throughput) utilisés pour identifier des composés thérapeutiques pour une foule de maladies, notamment le cancer. Il pourrait être utile de développer une librairie pouvant s'adapter à ces différents systèmes de criblage et permettant d'identifier des peptides inhibiteurs des mécanismes cellulaires sous-jacents aux grands maux de notre époque. La technologie est actuellement en développement pour faire le lien entre la séquence d'un peptide de quelques acides aminés et la structure de molécules organiques pouvant reproduire l'action de ce peptide (technique appelée « peptidomimetics ») (2, 11, 39, 69). Le développement de méthodes à grande échelle d'élucidation des voies de signalisation intracellulaires n'en est qu'à ses débuts et l'utilisation d'une librairie comme celle que nous tentons de développer constitue une des facettes importantes de ce travail de défrichage. Dans le domaine plus particulier de la biologie du virus d'immunodéficience humaine (VIH-1), l'identification de facteurs de résistance à l'infection virale dans des cellu-

les non-permissives est actuellement en pleine effervescence (29, 35, 64). L'utilisation d'une librairie dans ces cellules pourrait permettre de sélectionner des peptides inhibiteurs du ou des facteurs de résistance et ainsi ouvrir une toute nouvelle voie de recherche dans le domaine du VIH-1.

Bien sûr, l'utilisation d'une approche de librairie n'est pas une panacée et comporte des limites qui sont importantes à considérer lors de la conception d'une expérience de criblage. Ainsi, les interactions protéiques que l'on veut étudier à l'aide des librairies font souvent partie de complexes multiprotéiques (protein scaffold) qui peuvent présenter des épitopes mixtes, i.e. des surfaces d'interaction appartenant non pas à une mais à plusieurs protéines. Ce genre de situation rend l'identification de partenaires par les méthodes bio-informatiques conventionnelles beaucoup plus difficile, voire impossible. Un des avantages de travailler avec une librairie *in vivo* est de pouvoir mettre en présence des protéines ayant subi toutes les modifications post-traductionnelles propres aux cellules de mammifères avec des peptides. Que faire si ces peptides interagissent avec des composants des voies de modifications post-traductionnelles et interfèrent avec la maturation de la protéine étudiée? Même si on identifie un peptide ayant un effet inhibiteur réel, il pourrait être difficile d'en expliquer le mécanisme. Cela met un bémol à l'utilisation d'une librairie pour identifier des mécanismes intracellulaires. Comme toute méthode, notre librairie a des désavantages. Cependant, le travail effectué et rapporté au cours de ce mémoire nous permettra de mettre en pratique les principes fondamentaux de la construction de librairies aléatoires de grande diversité et de tirer parti des multiples avantages de cette méthode. Cette approche combine les avantages des autres types de librairie (diversité, versatilité, efficacité) sans en avoir les désavantages (changement de contexte du peptide, transposition des homologues de la levure aux mammifères, manipulations *in vitro*). En travaillant directement dans des cellules de mammifères, on ouvre la porte à la caractérisation de protéines modifiées post-traductionnellement, directement dans leur environnement physiologique et au sein des voies de signalisation dans lesquelles elles sont impliquées. Cette façon de faire nous permettra non seulement d'identifier des peptides inhibiteurs de la protéine que

l'on étudie, mais également des composants en amont et en aval de la voie de signalisation dans laquelle agit cette protéine. Maintenant que la majorité des problèmes conceptuels reliés à l'utilisation de notre librairie semblent réglés, il ne reste plus qu'à la construire et à profiter des possibilités qu'elle nous offre.

Nous avons donc passé en revue la procédure que nous avons utilisé pour tenter de construire une librairie de nonapeptides pouvant être exprimée en cellules de mammifères et évalué les différents facteurs nous ayant empêché de mener à bien cette entreprise. De façon évidente, certains facteurs nous sont apparus *a posteriori* comme essentiels à la réussite d'un tel projet : la taille du vecteur, la force et la stabilité du promoteur de la librairie, la nécessité de fusionner cette librairie avec un marqueur quelconque (de résistance ou phénotypique) de façon à s'assurer que les cellules positives expriment aussi la librairie, l'efficacité du système de criblage. D'un point de vue dynamique, c'est-à-dire en considérant que cette partie du projet ne fait que précéder la construction proprement dite d'une librairie contenant les éléments mentionnés plus haut, l'exercice a été bénéfique et a permis de finalement mettre au point une approche rationnelle de construction d'une librairie de peptides pouvant être exprimée en cellules de mammifères, ouvrant la voie à la construction et l'utilisation de ce nouvel outil biotechnologique.

BIBLIOGRAPHIE

1. <http://us.expasy.org/sprot/>
2. **Baell, J. B., and D. C. Huang.** 2002. Prospects for targeting the Bcl-2 family of proteins to develop novel cytotoxic drugs. *Biochem Pharmacol* **64**:851-63.
3. **Bender, M. A., T. D. Palmer, R. E. Gelinas, and A. D. Miller.** 1987. Evidence that the packaging signal of Moloney murine leukemia virus extends into the gag region. *J Virol* **61**:1639-46.
4. **Blum, J. H., S. L. Dove, A. Hochschild, and J. J. Mekalanos.** 2000. Isolation of peptide aptamers that inhibit intracellular processes. *Proc Natl Acad Sci U S A* **97**:2241-6.
5. **Bodenmuller, H., E. Schilling, B. Zachmann, and H. C. Schaller.** 1986. The neuropeptide head activator loses its biological activity by dimerization. *Embo J* **5**:1825-9.
6. **BouHamdan, M., Y. Xue, Y. Baudat, B. Hu, J. Sire, R. J. Pomerantz, and L. X. Duan.** 1998. Diversity of HIV-1 Vpr interactions involves usage of the WXXF motif of host cell proteins. *J Biol Chem* **273**:8009-16.
7. **Buchschacher, G. L., Jr.** 2001. Introduction to retroviruses and retroviral vectors. *Somat Cell Mol Genet* **26**:1-11.
8. **Burns, J. C., T. Friedmann, W. Driever, M. Burrascano, and J. K. Yee.** 1993. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci U S A* **90**:8033-7.
9. **Caruso, M., Y. Panis, S. Gagandeep, D. Houssin, J. L. Salzmänn, and D. Klatzmänn.** 1993. Regression of established macroscopic liver metastases after in situ transduction of a suicide gene. *Proc Natl Acad Sci U S A* **90**:7024-8.
10. **Chien, C. T., P. L. Bartel, R. Sternglanz, and S. Fields.** 1991. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**:9578-82.
11. **Cochran, A. G.** 2001. Protein-protein interfaces: mimics and inhibitors. *Curr Opin Chem Biol* **5**:654-9.

12. **Cohen, B. A., P. Colas, and R. Brent.** 1998. An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* **95**:14272-7.
13. **Cwirla, S. E., E. A. Peters, R. W. Barrett, and W. J. Dower.** 1990. Peptides on phage: a vast library of peptides for identifying ligands. *Proc Natl Acad Sci U S A* **87**:6378-82.
14. **de Felipe, P.** 2002. Polycistronic viral vectors. *Curr Gene Ther* **2**:355-78.
15. **Devlin, J. J., L. C. Panganiban, and P. E. Devlin.** 1990. Random peptide libraries: a source of specific protein binding molecules. *Science* **249**:404-6.
16. **Dirr, H., P. Reinemer, and R. Huber.** 1994. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur J Biochem* **220**:645-61.
17. **Drake, R. R., K. Pitlyk, R. A. McMasters, K. E. Mercer, H. Young, and M. P. Moyer.** 2000. Connexin-independent ganciclovir-mediated killing conferred on bystander effect-resistant cell lines by a herpes simplex virus-thymidine kinase-expressing colon cell line. *Mol Ther* **2**:515-23.
18. **DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos.** 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* **7**:379-87.
19. **Duke, G. M., M. A. Hoffman, and A. C. Palmenberg.** 1992. Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J Virol* **66**:1602-9.
20. **Fields, S., and O. Song.** 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**:245-6.
21. **Freeman, S. M., C. N. Abboud, K. A. Whartenby, C. H. Packman, D. S. Koeplin, F. L. Moolten, and G. N. Abraham.** 1993. The "bystander effect": tumor regression when a fraction of the tumor mass is genetically modified. *Cancer Res* **53**:5274-83.
22. **Fyfe, J. A., P. M. Keller, P. A. Furman, R. L. Miller, and G. B. Elion.** 1978. Thymidine kinase from herpes simplex virus phosphorylates the new antiviral compound, 9-(2-hydroxyethoxymethyl)guanine. *J Biol Chem* **253**:8721-7.
23. **Gallardo, H. F., C. Tan, and M. Sadelain.** 1997. The internal ribosomal entry site of the encephalomyocarditis virus enables reliable coexpression of two transgenes in human primary T lymphocytes. *Gene Ther* **4**:1115-9.

24. **Gao, G., X. Guo, and S. P. Goff.** 2002. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* **297**:1703-6.
25. **Garin, M. I., E. Garrett, P. Tiberghien, J. F. Apperley, D. Chalmers, J. V. Melo, and C. Ferrand.** 2001. Molecular mechanism for ganciclovir resistance in human T lymphocytes transduced with retroviral vectors carrying the herpes simplex virus thymidine kinase gene. *Blood* **97**:122-9.
26. **Gerolami, R., R. Uch, F. Jordier, S. Chapel, C. Bagnis, C. Brechot, and P. Mannoni.** 2000. Gene transfer to hepatocellular carcinoma: transduction efficacy and transgene expression kinetics by using retroviral and lentiviral vectors. *Cancer Gene Ther* **7**:1286-92.
27. **Geyer, C. R., A. Colman-Lerner, and R. Brent.** 1999. "Mutagenesis" by peptide aptamers identifies genetic network members and pathway connections. *Proc Natl Acad Sci U S A* **96**:8567-72.
28. **Gill, D. R., S. E. Smyth, C. A. Goddard, I. A. Pringle, C. F. Higgins, W. H. Colledge, and S. C. Hyde.** 2001. Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1alpha promoter. *Gene Ther* **8**:1539-46.
29. **Goff, S. P.** 2003. Death by deamination: a novel host restriction system for HIV-1. *Cell* **114**:281-3.
30. **Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* **36**:59-74.
31. **Grassi, G., P. Maccaroni, R. Meyer, H. Kaiser, E. D'Ambrosio, E. Pascale, M. Grassi, A. Kuhn, P. Di Nardo, R. Kandolf, and J. H. Kupper.** 2003. Inhibitors of DNA methylation and histone deacetylation activate cytomegalovirus promoter-controlled reporter gene expression in human glioblastoma cell line U87. *Carcinogenesis*.
32. **Greco, O., and G. U. Dachs.** 2001. Gene directed enzyme/prodrug therapy of cancer: historical appraisal and future prospectives. *J Cell Physiol* **187**:22-36.
33. **Green, L. M., J. L. Reade, and C. F. Ware.** 1984. Rapid colorimetric assay for cell viability: application to the quantitation of cytotoxic and growth inhibitory lymphokines. *J Immunol Methods* **70**:257-68.

34. **Hallett, M. T.** 2002, posting date. Two-hybrid systems and protein-protein interaction. http://www.mcb.mcgill.ca/~hallett/GEP/PLecture6/2Hybrid_files/image002.jpg
35. **Harris, R. S., K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, and M. H. Malim.** 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**:803-9.
36. **Hennecke, M., M. Kwissa, K. Metzger, A. Oumard, A. Kroger, R. Schirmbeck, J. Reimann, and H. Hauser.** 2001. Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. *Nucleic Acids Res* **29**:3327-34.
37. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol* **26**:365-9.
38. **Hopkins, N.** 1993. High titers of retrovirus (vesicular stomatitis virus) pseudotypes, at last. *Proc Natl Acad Sci U S A* **90**:8759-60.
39. **Jones, R., Douglas Boatman, P., Semple, G., Shin, Y-J., Tamura, SY.** 2003. Clinically validated peptides as templates for *de novo* peptidomimetic drug screen at G-protein-coupled receptors. *Curr Opin Pharmacol* **3**:1-14.
40. **Kaelin, W. G., Jr., D. C. Pallas, J. A. DeCaprio, F. J. Kaye, and D. M. Livingston.** 1991. Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell* **64**:521-32.
41. **Kay, B.** 2002, posting date. Molecular affinity reagents. [Online.]
42. **Kit, S., D. R. Dubbs, R. A. DeTorres, and J. L. Melnick.** 1965. Enhanced thymidine kinase activity following infection of green monkey kidney cells by simian adenoviruses, simian papovavirus SV40, and an adenovirus-SV40 "hybrid". *Virology* **27**:453-7.
43. **Klevenz, B., K. Butz, and F. Hoppe-Seyler.** 2002. Peptide aptamers: exchange of the thioredoxin-A scaffold by alternative platform proteins and its influence on target protein binding. *Cell Mol Life Sci* **59**:1993-8.
44. **Knipes, D.** 2001. *Fundamental Virology*, 4th ed. Lippincott, Williams & Wilkins, Philadelphia.
45. **Kuhn, R., and R. M. Torres.** 2002. Cre/loxP recombination system and gene targeting. *Methods Mol Biol* **180**:175-204.

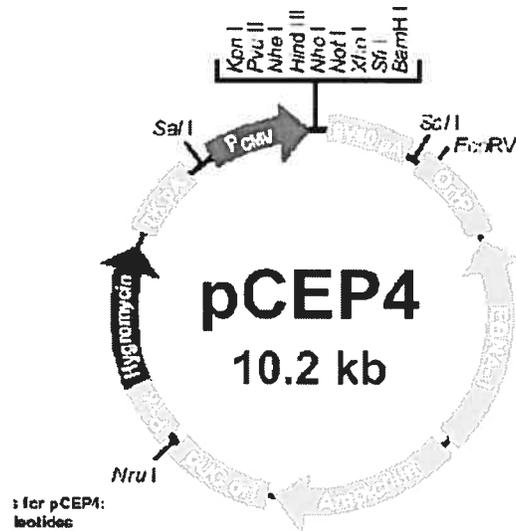
46. **Larocca, D., M. A. Burg, K. Jensen-Pergakes, E. P. Ravey, A. M. Gonzalez, and A. Baird.** 2002. Evolving phage vectors for cell targeted gene delivery. *Curr Pharm Biotechnol* **3**:45-57.
47. **Loser, P., G. S. Jennings, M. Strauss, and V. Sandig.** 1998. Reactivation of the previously silenced cytomegalovirus major immediate-early promoter in the mouse liver: involvement of NF κ B. *J Virol* **72**:180-90.
48. **Miller, A. D.** 1990. Retrovirus packaging cells. *Hum Gene Ther* **1**:5-14.
49. **Mizuguchi, H., Z. Xu, A. Ishii-Watabe, E. Uchida, and T. Hayakawa.** 2000. IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* **1**:376-82.
50. **Moolten, F. L.** 1986. Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy. *Cancer Res* **46**:5276-81.
51. **NCBI** <http://www.ncbi.nlm.nih.gov/BLAST/>
52. **Newell-Price, J., A. J. Clark, and P. King.** 2000. DNA methylation and silencing of gene expression. *Trends Endocrinol Metab* **11**:142-8.
53. **Nolan, G. P.** 2001, posting date. Peptide libraries : a tutorial. Gary P. Nolan Lab. http://www.stanford.edu/group/nolan/library_systems/peptide_systems.html
54. **Noren, K. A., and C. J. Noren.** 2001. Construction of high-complexity combinatorial phage display peptide libraries. *Methods* **23**:169-78.
55. **Norman, T. C., D. L. Smith, P. K. Sorger, B. L. Drees, S. M. O'Rourke, T. R. Hughes, C. J. Roberts, S. H. Friend, S. Fields, and A. W. Murray.** 1999. Genetic selection of peptide inhibitors of biological pathways. *Science* **285**:591-5.
56. **Obata, T., M. B. Yaffe, G. G. Leparc, E. T. Piro, H. Maegawa, A. Kashiwagi, R. Kikkawa, and L. C. Cantley.** 2000. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J Biol Chem* **275**:36108-15.
57. **O'Malley, B. W., Jr., S. H. Chen, M. R. Schwartz, and S. L. Woo.** 1995. Adenovirus-mediated gene therapy for human head and neck squamous cell cancer in a nude mouse model. *Cancer Res* **55**:1080-5.
58. **Peelle, B., J. Lorens, W. Li, J. Bogenberger, D. G. Payan, and D. C. Anderson.** 2001. Intracellular protein scaffold-mediated display of random

- peptide libraries for phenotypic screens in mammalian cells. *Chem Biol* **8**:521-34.
59. **Prosch, S., J. Stein, K. Staak, C. Liebenthal, H. D. Volk, and D. H. Kruger.** 1996. Inactivation of the very strong HCMV immediate early promoter by DNA CpG methylation in vitro. *Biol Chem Hoppe Seyler* **377**:195-201.
 60. **Qasim, W., A. J. Thrasher, J. Buddle, C. Kinnon, M. E. Black, and H. B. Gaspar.** 2002. T cell transduction and suicide with an enhanced mutant thymidine kinase. *Gene Ther* **9**:824-7.
 61. **Reardon, J. E.** 1989. Herpes simplex virus type 1 and human DNA polymerase interactions with 2'-deoxyguanosine 5'-triphosphate analogues. Kinetics of incorporation into DNA and induction of inhibition. *J Biol Chem* **264**:19039-44.
 62. **Sambrook, J., Fritsch, E. F., and Maniatis, T.** 1989. *Molecular Cloning: a Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, New York.
 63. **Schaffitzel, C., J. Hanes, L. Jermutus, and A. Pluckthun.** 1999. Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *J Immunol Methods* **231**:119-35.
 64. **Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**:646-50.
 65. **Sidhu, S. S., G. D. Bader, and C. Boone.** 2003. Functional genomics of intracellular peptide recognition domains with combinatorial biology methods. *Curr Opin Chem Biol* **7**:97-102.
 66. **Sinning, I., G. J. Kleywegt, S. W. Cowan, P. Reinemer, H. W. Dirr, R. Huber, G. L. Gilliland, R. N. Armstrong, X. Ji, P. G. Board, and et al.** 1993. Structure determination and refinement of human alpha class glutathione transferase A1-1, and a comparison with the Mu and Pi class enzymes. *J Mol Biol* **232**:192-212.
 67. **Smith, G. P.** 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**:1315-7.
 68. **Soling, A., A. Simm, and N. Rainov.** 2002. Intracellular localization of Herpes simplex virus type 1 thymidine kinase fused to different fluorescent proteins depends on choice of fluorescent tag. *FEBS Lett* **527**:153-8.

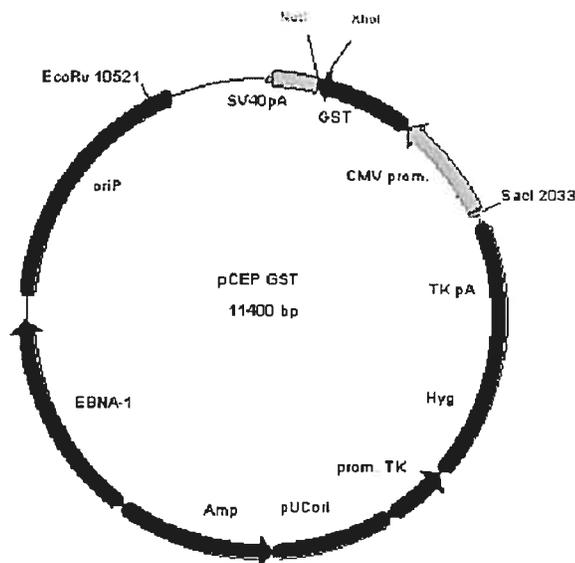
69. **Stigers, K. D., M. J. Soth, and J. S. Nowick.** 1999. Designed molecules that fold to mimic protein secondary structures. *Curr Opin Chem Biol* **3**:714-23.
70. **Strange, R. C., C. G. Faulder, B. A. Davis, R. Hume, J. A. Brown, W. Cotton, and D. A. Hopkinson.** 1984. The human glutathione S-transferases: studies on the tissue distribution and genetic variation of the GST1, GST2 and GST3 isozymes. *Ann Hum Genet* **48 (Pt 1)**:11-20.
71. **Suzuki, T., M. Coggan, D. C. Shaw, and P. G. Board.** 1987. Electrophoretic and immunological analysis of human glutathione S-transferase isozymes. *Ann Hum Genet* **51 (Pt 2)**:95-106.
72. **Teschendorf, C., K. H. Warrington, Jr., D. W. Siemann, and N. Muzyczka.** 2002. Comparison of the EF-1 alpha and the CMV promoter for engineering stable tumor cell lines using recombinant adeno-associated virus. *Anticancer Res* **22**:3325-30.
73. **Tomicic, M. T., R. Thust, and B. Kaina.** 2002. Ganciclovir-induced apoptosis in HSV-1 thymidine kinase expressing cells: critical role of DNA breaks, Bcl-2 decline and caspase-9 activation. *Oncogene* **21**:2141-53.
74. **Tong, A. H., B. Drees, G. Nardelli, G. D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C. W. Hogue, S. Fields, C. Boone, and G. Cesareni.** 2002. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* **295**:321-4.
75. **Touraine, R. L., H. Ishii-Morita, W. J. Ramsey, and R. M. Blaese.** 1998. The bystander effect in the HSVtk/ganciclovir system and its relationship to gap junctional communication. *Gene Ther* **5**:1705-11.
76. **Tuerk, C., and L. Gold.** 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**:505-10.
77. **Vodicka, M. A., D. M. Koepf, P. A. Silver, and M. Emerman.** 1998. HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. *Genes Dev* **12**:175-85.
78. **Wei, S. J., Y. Chao, Y. M. Hung, W. C. Lin, D. M. Yang, Y. L. Shih, L. Y. Ch'ang, J. Whang-Peng, and W. K. Yang.** 1998. S- and G2-phase cell cycle arrests and apoptosis induced by ganciclovir in murine melanoma cells transduced with herpes simplex virus thymidine kinase. *Exp Cell Res* **241**:66-75.

79. **Whartenby, K. A., J. W. Darnowski, S. M. Freeman, and P. Calabresi.** 2002. A role for MAP kinase in the antitumor activity of a nucleoside analog. *Cancer Gene Ther* **9**:37-43.
80. **Whartenby, K. A., J. W. Darnowski, S. M. Freeman, K. Yurasha, and P. Calabresi.** 1999. Recombinant interferon alpha2a synergistically enhances ganciclovir-mediated tumor cell killing in the herpes simplex virus thymidine kinase system. *Cancer Gene Ther* **6**:402-8.
81. **Wilce, M. C., and M. W. Parker.** 1994. Structure and function of glutathione S-transferases. *Biochim Biophys Acta* **1205**:1-18.
82. **Wolkowicz, R., and G. P. Nolan.** 2003. Retroviral technology--applications for expressed peptide libraries. *Front Biosci* **8**:d603-19.
83. **Xu, X., C. Leo, Y. Jang, E. Chan, D. Padilla, B. C. Huang, T. Lin, T. Gururaja, Y. Hitoshi, J. B. Lorens, D. C. Anderson, B. Sikic, Y. Luo, D. G. Payan, and G. P. Nolan.** 2001. Dominant effector genetics in mammalian cells. *Nat Genet* **27**:23-9.
84. **Yaffe, M. B., G. G. Leparc, J. Lai, T. Obata, S. Volinia, and L. C. Cantley.** 2001. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* **19**:348-53.
85. **Yang, M., Z. Wu, and S. Fields.** 1995. Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res* **23**:1152-6.
86. **Yao, X. J., J. Lemay, N. Rougeau, M. Clement, S. Kurtz, P. Belhumeur, and E. A. Cohen.** 2002. Genetic selection of peptide inhibitors of human immunodeficiency virus type 1 Vpr. *J Biol Chem* **277**:48816-26.
87. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci U S A* **82**:1618-22.
88. **Zhu, Y., G. Feuer, S. L. Day, S. Wrzesinski, and V. Planelles.** 2001. Multigene lentiviral vectors based on differential splicing and translational control. *Mol Ther* **4**:375-82.

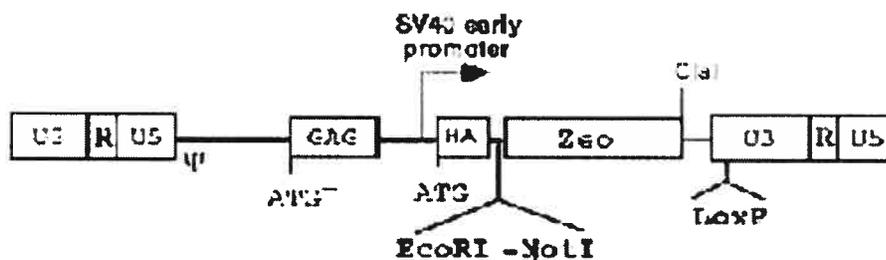
ANNEXE I : VECTEURS UTILISÉS



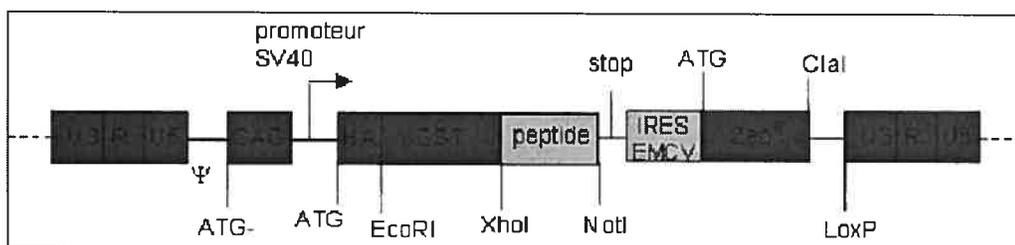
pCEP4. Vecteur épisomal utilisé pour le clonage de la librairie de nonapeptides.



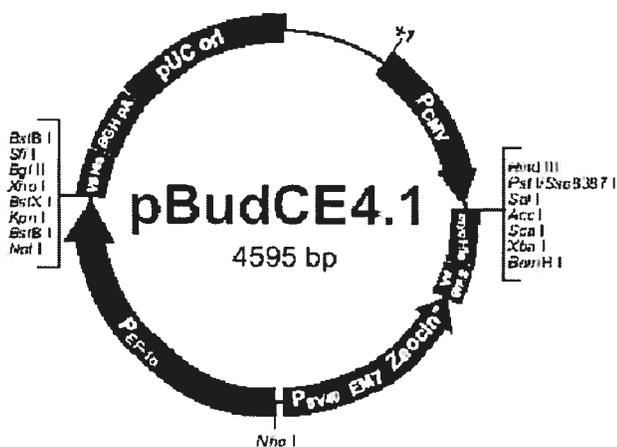
pCEP GST. Version modifiée de pCEP4 permettant le clonage de la librairie en fusion avec le gène de la GST dans les sites XhoI et NotI.



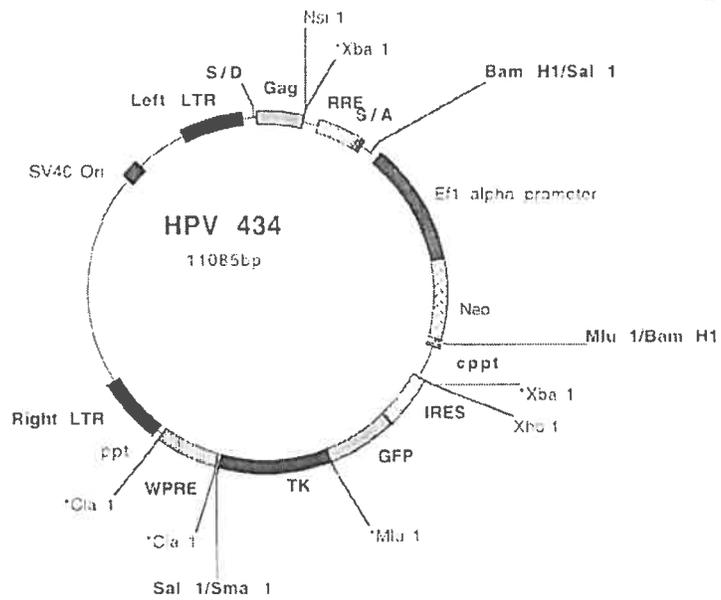
pBabe HAZ. Vecteur de base pour la construction d'une librairie de nonapeptides dans un vecteur rétroviral.



pBabe HAGIZ. Cassette d'expression bicistronique clonée dans les sites EcoRI et NotI du vecteur pBabe HAZ



pBUDCE4-Zéo . Vecteur contrôle positif pour les tests de résistance à la zéocine



HPV 434. Vecteur rétroviral ayant servi à l'établissement de la lignée cellulaire 293TK+.

