

CHAPITRE III :  
INGÉNIERIE DES PROTÉINES ;  
MÉTHODES DE CRIBLAGE ET DE SÉLECTION

## 1. Méthodes de criblage et de sélection

La deuxième étape du processus d'évolution dirigée est l'étape de sélection des variants ayant incorporé la propriété d'intérêt. De la même manière, selon le mode opératoire, il existe des méthodes de sélection *in vitro* et d'autre *in vivo*. Ces méthodes sont essentiellement basées sur les propriétés d'interaction protéine-ligand.

## 2. Méthodes de sélection *in vivo*

Les méthodes *in vivo* nécessitent une étape préalable de clonage des gènes mutés dans des plasmides ou des phagemides appropriés pour la transformation cellulaire (bactéries, levures). En effet si la sélection est basée sur une interaction intracellulaire, les techniques utilisables reposent sur les approches de doubles hybrides chez la levure et les approches de complémentation de fragments protéique ou PCA (Protein-fragment Complementation Assay). Dans le cas où la sélection est basée sur une simple interaction protéine-ligand les techniques développées sont l'exposition sur cellules (bactérie ou levure), l'exposition sur phages et sur d'autres types de virus

### 2.1. Technique du double hybride

La technique de double hybride est un système artificiel de mise en évidence d'interactions protéiques. Par mutagenèse dirigée, il est possible d'identifier des domaines minimums et des résidus critiques pour l'interaction. Le double hybride et ses techniques dérivées permettent donc facilement l'identification, l'optimisation et la modification des interactions protéine-ADN ou protéine-protéine.

C'est une technique développée par Fields et Song en 1989. C'est un système génétique visant à étudier des interactions protéines- protéines se basant sur les propriétés de la protéine GAL4 de la levure *Saccharomyces cerevisiae* (Fig. 1). Cette protéine est un activateur de la transcription, essentielle pour l'expression des gènes codant les enzymes de dégradation du galactose. Elle se compose de deux domaines fonctionnels séparables : un domaine N-terminal qui se lie spécifiquement à la séquence d'ADN 'UASG', et un domaine C-terminal contenant des résidus nécessaires pour activer la transcription. Le système généré comporte deux protéines hybrides. La première contient le domaine d'interaction avec l'ADN fusionné à une protéine X. La deuxième comprend le domaine d'activation de la traduction de la GAL4 fusionné à une protéine 'Y'. Par la suite, si X et Y peuvent former un complexe et reconstituer ainsi la protéine GAL4, il y aura induction de la transcription d'un gène rapporteur qui sera régulé par la séquence UASG. Le système peut être appliqué à une grande variété de protéines, plus particulièrement pour identifier une protéine interagissant avec une protéine connue par simple sélection sur milieu galactose.

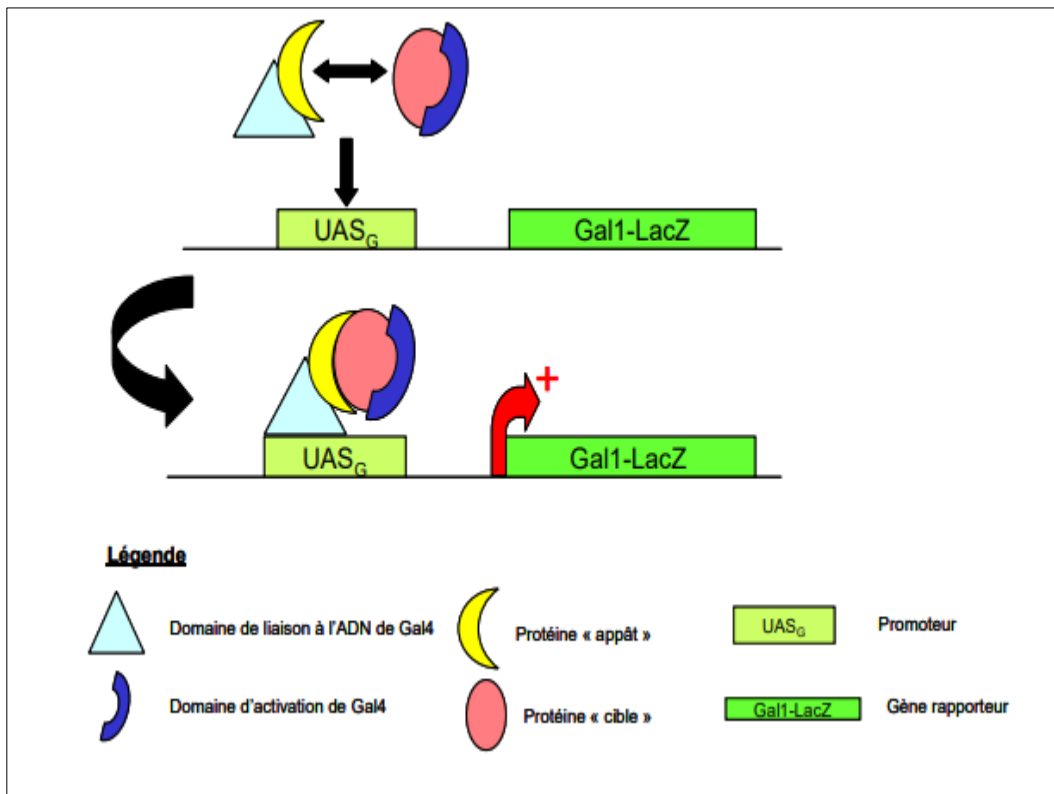


Fig. 1 : : Principe du double hybride de levure mis au point par Fields and Song.

## 2.2. PCA « Protein Fragment complementary assay »

La PCA est une technique développée par l'équipe de S. Michnick, dans le but d'établir une approche expérimentale permettant de répondre à la série de questions suivantes : comment, quand, où et dans quelles conditions les protéines sont activées/désactivées ou encore interagissent ? (Michnick, 2003)

Le principe de la PCA est le suivant : 2 protéines à tester sont fusionnées chacune à un fragment d'une protéine rapporteuse et sont exprimées dans une même cellule. Si ces deux protéines interagissent la protéine rapporteuse est alors reconstituée rétablissant ainsi sa fonction naturelle (Michnick, 2001) et permettant la détection des cellules ayant les deux partenaires d'interaction. La protéine rapporteuse peut être, soit une enzyme soit une protéine fluorescente. En effet, plusieurs protéines ont été utilisées pour cette application : on peut citer la dihydrofolate réductase (DHFR), l'aminoglycoside kinase, la TEM  $\beta$ -lactamase et la protéine à fluorescence verte (Green Fluorescent Protein : GFP). La particularité de ces protéines est que leurs fragments sont incapables d'interagir spontanément et reformer la protéine entière (Fig. 2), le but est qu'ils le deviennent par la reconnaissance des partenaires auxquels ils sont fusionnés. Cette technique permet de détecter directement des interactions protéines – protéines. Elle a l'avantage d'exprimer les protéines sous leur forme native, avec les modifications post-traductionnelles si le système d'expression le permet, et de les localiser dans les cellules. Par ailleurs, cette approche peut être réalisée dans tous types

cellulaires, dans les bactéries comme dans les cellules eucaryotes. Dans les bactéries, cette technique permet de créer des bibliothèques plus diverses que celles obtenues par double hybride chez *S. Cerevisiae*.

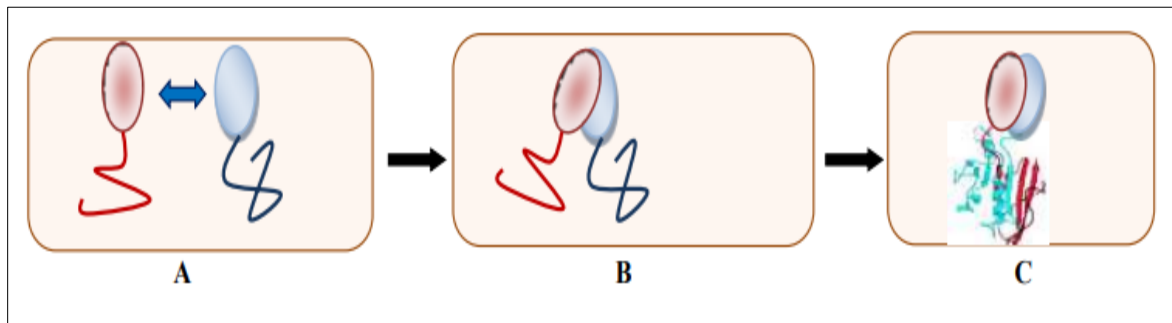


Fig. 2 : Schéma général de la PCA : A : Les 2 protéines d'intérêt sont exprimées dans la même cellule, fusionnées chacune à un domaine de la protéine rapporteuse (enzyme ayant une activité vitale pour la cellule ou un effet fluorescent). B : Si les protéines d'intérêt interagissent les 2 domaines de la protéine porteuse s'approchent. C : la protéine porteuse (ici DHFR) retrouve sa conformation naturelle à partir des 2 domaines et sa fonction est rétablie.

### 2.3. Exposition sur cellules : Cell surface Display

Grâce au développement des systèmes d'expression, des polypeptides sont efficacement exposés à la surface des cellules. Selon la nature des protéines à faire évoluer et du système d'expression mis en œuvre, les cellules support utilisées peuvent être des bactéries, des levures ou encore des cellules de mammifères (Fig. 3).

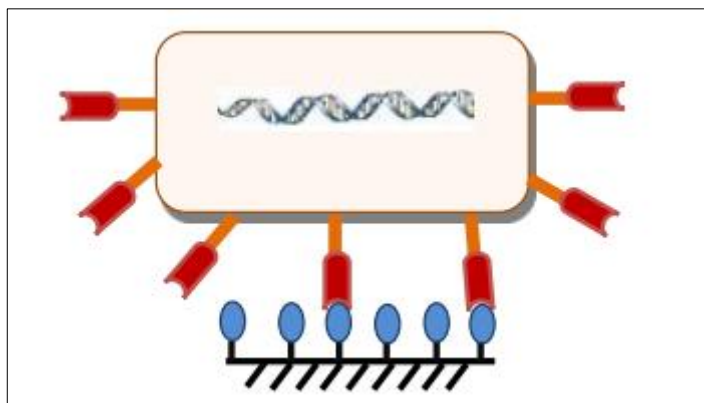


Fig. 3 : Exposition sur cellule.

A titre d'exemple, ce système a permis de cribler, une bibliothèque de variants de protéase de membrane externe OmpT, exposée à la surface d'*E.coli* et d'isoler des enzymes qui sont plus actives que les clones de départ. En effet, l'équipe est partie d'une bibliothèque de  $6 \cdot 10^5$  clones exposant à leur surface l'enzyme OmpT préalablement mutée et, par tri de cellules par fluorescence (FACS) et un substrat peptidique fluorescent, il a été possible de sélectionner des variants 60 fois plus actifs. Les bactéries sont isolées grâce au clivage du substrat synthétique avec lequel elles sont incubées. Le substrat contient un colorant fluorescent, une séquence de clivage (Arg-Val) et un partenaire qui éteint

le signal de fluorescence avant clivage. Les levures ont été utilisées en tant que support, la première fois par l'équipe de Witthrup, pour l'exposition d'une bibliothèque relativement petite ( $3 \times 10^5$  clones) de scFV aléatoirement mutés. L'exposition à la surface de la levure a aussi l'avantage de pouvoir cribler des protéines eucaryotes de haut poids moléculaire, qui peuvent être glycosilées ou présenter des ponts disulfures, ce qui est impossible à obtenir avec les bactéries. Bien que plusieurs cellules hôtes soient disponibles, *Saccharomyces cerevisiae* reste l'organisme le plus communément utilisé.

#### **2.4. Exposition sur phage : Phage Display**

Cette technique permet l'expression de peptides ou de protéines exogènes à la surface des particules de phages, dans le but de sélectionner et d'amplifier un polypeptide capable d'interagir avec une molécule cible choisie. Cette technique est probablement la plus utilisée dans la découverte de nouvelles interactions spécifiques à partir de répertoires de peptides ou polypeptides. Le concept a été décrit pour la première fois par Smith en 1985 lorsqu'il est apparu que le gène III du virion du phage codait pour une protéine mineure de l'enveloppe virale. Cette protéine est constituée de deux domaines : un domaine N-terminal qui se lie au pilus F lors de l'infection de la bactérie et un domaine C-terminal enfoui dans le virion participant à sa morphogénèse. L'observation initiale était que l'insertion d'un ADN étranger entre les 2 domaines de la protéine III semblait ne pas perturber le fonctionnement de la protéine et que la protéine synthétisée était exposée et accessible à la surface du phage. Cette approche a d'abord été décrite comme une approche d'identification de gènes naturels clonés dans des vecteurs dérivés du phage M13. L'innovation majeure a été de construire, sur le même principe, des bibliothèques de peptides de séquences aléatoires et de montrer qu'il est alors possible de « cloner » des séquences artificielles susceptibles d'interagir efficacement avec des protéines cibles choisies. L'approche permettait, d'une façon très générale, la sélection et l'amplification de séquences, même très rares, au sein de bibliothèques de grande diversité (Fig. 4). Le développement de cette méthodologie a joué un rôle déterminant dans le développement des stratégies combinatoires en ingénierie des protéines.

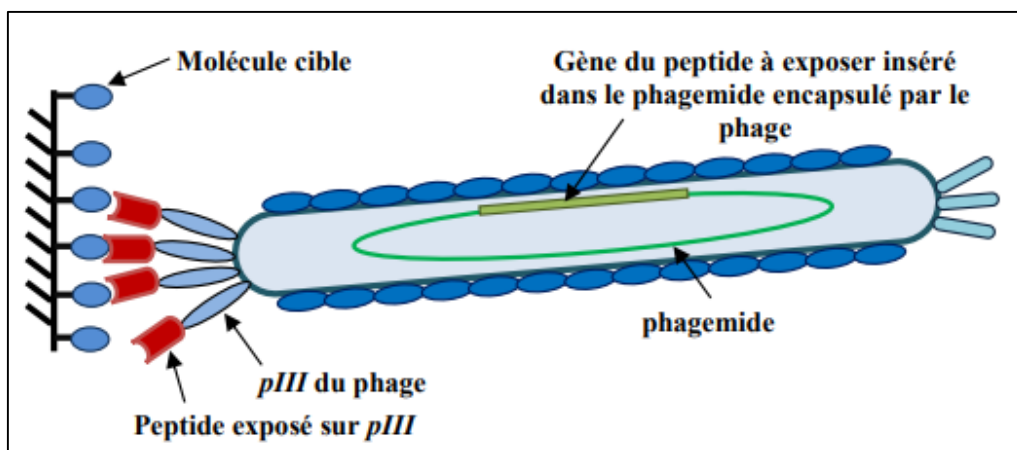


Fig.4 : Exposition de peptides sur le phage M13

### 3. Méthodes de sélection *in vitro*

Bien que les méthodes de sélections *in vitro*, décrites ci-dessous, soient des méthodes de sélection et de criblage de banques de protéines, il est important de mentionner qu'elles sont inspirées par une méthode de sélection d'acides nucléiques. C'est la technique intitulée évolution systématique de ligand par enrichissement exponentiel (Systematic Evolution of Ligands by EXponential Enrichement : SELEX).

La technique SELEX consiste à réaliser plusieurs cycles comportant une étape de transcription d'un pool d'acide nucléique aléatoires, suivie d'une étape de sélection par contact avec la molécule cible et finalement une étape de RT-PCR permettant une amplification exponentielle des acides nucléiques sélectionnés. Après les cycles de sélection, les interacteurs obtenus sont clonés, séquencés et caractérisés. Cette technique présente l'avantage d'avoir l'information génotypique et phénotypique sur la même molécule et d'obtenir rapidement des interacteurs hautement spécifiques à partir de très larges bibliothèques initiales ( $10^{15}$  à  $10^{16}$  séquences). Un des inconvénients de cette méthode est la sensibilité des ARN aux RNases ubiquitaires mais l'inconvénient majeur réside dans le fait que l'ARN est une molécule polyanionique interagissant préférentiellement avec des cibles positives ce qui restreint la gamme de cibles que l'on peut choisir. Pour contourner ce problème et élargir le répertoire des interactions chimiques, les ARN ont été remplacés par les peptides et les protéines d'où les méthodes de sélection protéiques *in vitro*.

#### 3.1. Exposition sur ribosome

Elle a été utilisée pour la première fois, pour identifier parmi des séquences aléatoires codant des décapeptides, celles qui reconnaissent un anticorps monoclonal spécifique de la dynorphine B en utilisant le système de transcription S30 d'*E.coli*. L'identification, par Ribosome Display, d'interacteurs spécifiques pour une molécule cible se fait en plusieurs cycles de sélection. Un cycle de sélection comporte différentes étapes dont la première est la formation d'un complexe ribosomal

(ARNm + Ribosome + chaîne polypeptidique) (Fig. 5). La première étape de formation du complexe ribosomique peut se faire de deux façons différentes : soit en couplant transcription-traduction partant d'une bibliothèque d'ADN, soit en préparant l'ARNm par transcription et purification préalables pour passer par la suite à l'étape de traduction et formation du complexe. A ce stade, il y a formation de complexes ribosomiques qui contiennent une protéine repliée exposée mais qui reste liée par sa séquence C-terminal à l'ARNt, qui est lui-même encore complexé au ribosome. La traduction est par la suite arrêtée par refroidissement de la réaction en ajoutant du tampon froid contenant du  $Mg^{2+}$ . Ce traitement favorise la condensation des ribosomes empêchant ainsi la dissociation ou l'hydrolyse des complexes. Une fois la bibliothèque exposée sur les ribosomes, le mélange est incubé en présence de la cible. Cette dernière peut être immobilisée directement sur la surface d'un tube, d'une plaque ou encore biotinylée et fixée à la surface de billes magnétiques couvertes de streptavidine. L'avantage de cette pratique est de s'assurer que la molécule cible biotinylée est efficacement capturée dans sa conformation naturelle et non déformée par d'éventuelles interactions avec le support. Par ailleurs, il est ainsi possible de contrôler la concentration exacte de cible utilisée pour la sélection. Les complexes peu ou pas spécifiques sont éliminés par des lavages avec du tampon contenant du magnésium. Par la suite, l'éluion peut être réalisée soit par ajout d'un excès d'EDTA, ce qui élimine le  $Mg^{2+}$  dissociant ainsi les complexes. Les ARNm libres sont récupérés sans avoir recours à la dissociation des interacteurs protéiques. L'éluion peut aussi être réalisée par compétition avec la cible libre ce qui permet d'isoler les ARNm correspondant aux interacteurs fonctionnels et spécifiques seulement. Enfin, Les ARNm isolés sont amplifiés par RT-PCR et l'ADN résultant est utilisé pour le cycle suivant. Une partie de cet ADN est analysée par clonage, séquençage et tests ELISA.

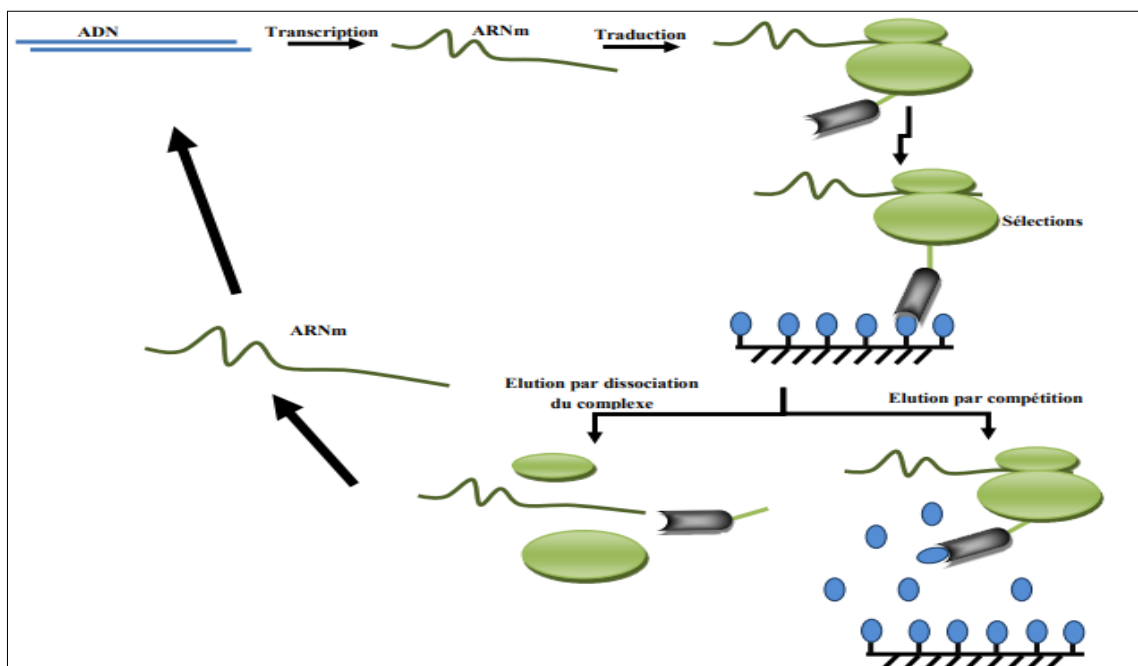


Fig.5 : Principe général de la technique du Ribosome display.

### 3.2. La fusion Peptide –ARN : RNA-Peptide fusion ou In Vitro virus

Cette approche décrit une méthode permettant de lier de manière covalente un peptide à l'ARN qui le code. En effet, l'ARNm est transcrit *in vitro* puis purifié et par la suite lié par son extrémité 3', à une séquence ADN *linker* marquée à la puromycine. Cette construction ADN-ARN-puromycine est purifiée puis traduite *in vitro*. Une fois la région ADN atteinte, les ribosomes se détachent du complexe permettant ainsi à la puromycine d'atteindre le site de la peptidyltransférase et de se lier covalamment au peptide. Les complexes hybrides obtenus sont utilisés dans l'étape suivante de sélection, ceux qui reconnaissent la cible sont élués et le matériel génétique est amplifié par RT-PCR. La seule étape critique de ce processus est l'étape de fixation de l'ADN marqué à l'ARNm mais son avantage majeur est la stabilité du complexe ce qui permet des conditions de sélection contrôlées et contraignantes (Fig. 6).

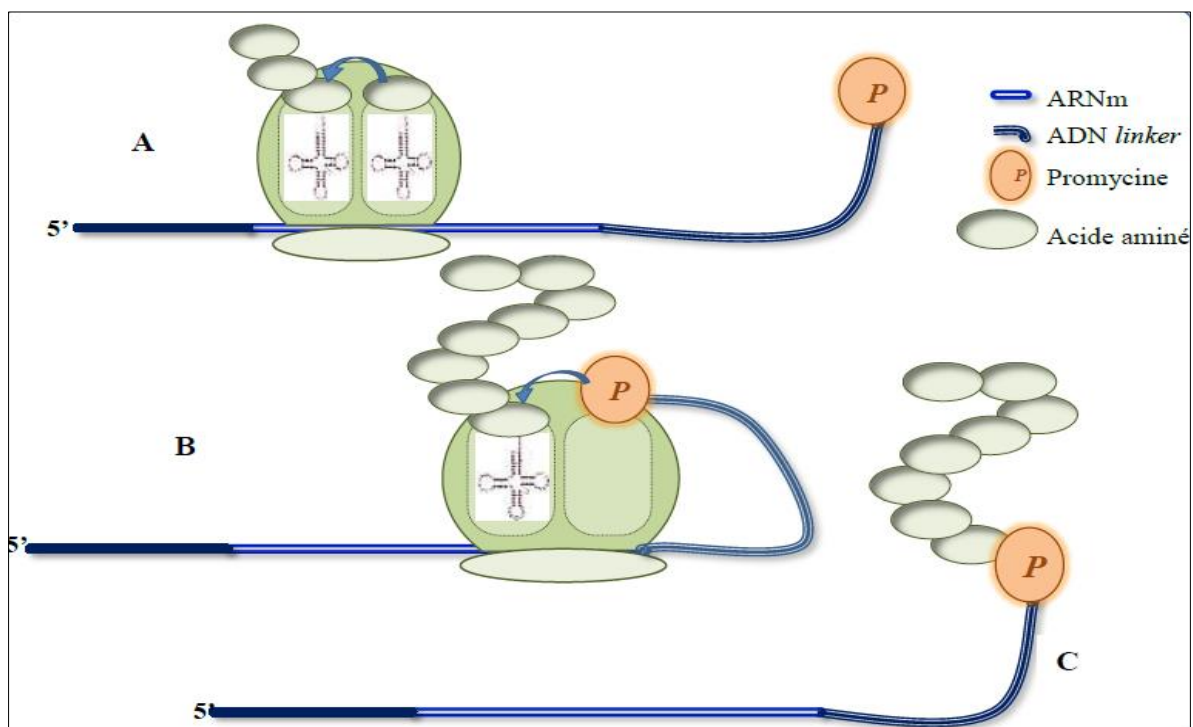


Fig. 6 : Schéma général de la formation du complexe ARNm-peptide : A : Initiation de la traduction et évolution des ribosomes au long de l'ARNm. B : Au niveau de la jonction ADNlinker-ARNm, la puromycine est intégrée au niveau du site A du ribosome où elle est liée à la chaîne polypeptidique. C : Complexe ARNm-peptide obtenu après chromatographie d'affinité.

### 3.3. Compartimentation *in vitro*: Water in Oil Emulsions for Binding Selections (STABLE)

Les méthodes de type *phage*, *ribosome* ou *ARN display* permettent de sélectionner efficacement des peptides ou des protéines ayant de nouvelles capacités d'interaction. Ces méthodes sont néanmoins plus difficiles à exploiter lorsque l'objectif est de faire émerger de nouveaux catalyseurs. Le problème fondamental est que le produit de la réaction enzymatique est généralement détaché de l'enzyme et libéré dans le milieu où il s'accumule d'autant plus rapidement que l'enzyme



est performant. Dans la solution toutes les protéines, les plus comme les moins efficaces, sont mélangées. Ainsi, même s'il est possible d'associer chaque variant d'une bibliothèque d'enzymes à « son » gène, il n'est pas possible de savoir quel variant du catalyseur est le plus efficace si le lien avec les molécules qu'il a contribué à produire s'est évanoui par la diffusion du produit dans la solution. Ce problème peut être contourné de deux manières différentes : en liant l'enzyme à son substrat ou en créant de la compartimentation. Le concept central des approches de microcompartimentation est qu'une population d'enzyme pourra faire l'objet d'un processus de sélection à la condition que le produit accumulé par l'activité de chaque variant du catalyseur soit séquestré dans le compartiment qui contient ce variant ainsi que le gène qui le code.

En 1998, ce concept a été mis en pratique grâce à des travaux visant la sélection du gène codant la *HeaIII* méthyltransférase à partir d'un mélange contenant  $10^7$  gènes différents codants pour d'autres enzymes. Les sélections ont été réalisées par un système de traduction-transcription dans des émulsions d'eau dans l'huile (Fig. 7).

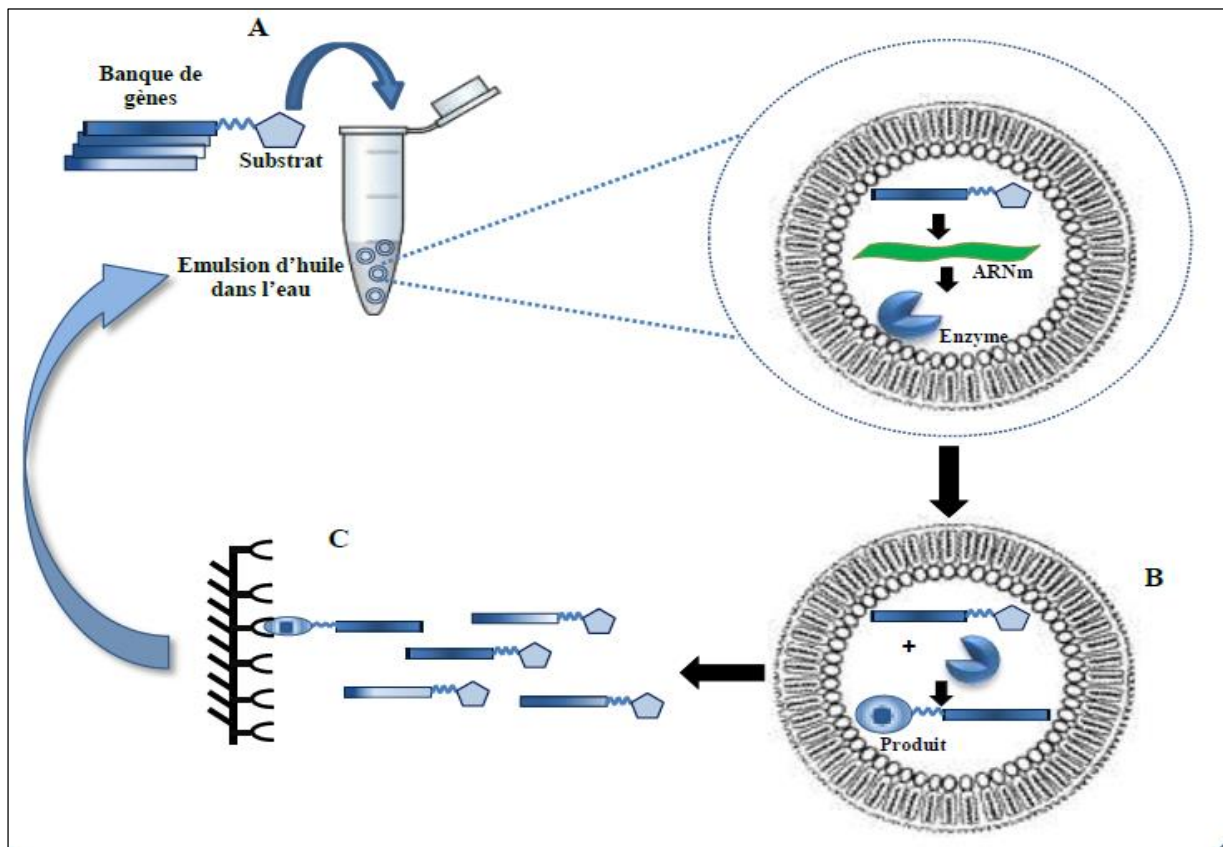


Fig. 7 : Schéma général d'un tour de sélection par compartimentation. A : Réaction de transcription/traduction couplées *in vitro* ayant lieu dans un milieu réactionnel localisé dans des émulsions d'eau dans l'huile (1 émulsion = 1 gène = 1 enzyme). B : Les enzymes synthétisées transforment les substrats en produit qui restent liés aux gènes. C : Les produits obtenus par la réaction enzymatique sont récupérés par affinité permettant ainsi de collecter des gènes correspondant aux enzymes actives. Les gènes récupérés sont utilisés pour une analyse de séquences ou encore amplifiés pour être réinjectés dans un deuxième tour de sélection.

Références :

Cette partie de cours est extraite de la thèse suivante :

Guellouz, A. 2012. *Création par évolution dirigée de protéines artificielles en alternatives aux anticorps*.  
Thèse de doctorat. 'Université Paris Sud 11. <https://tel.archives-ouvertes.fr/tel-00767675/document>