

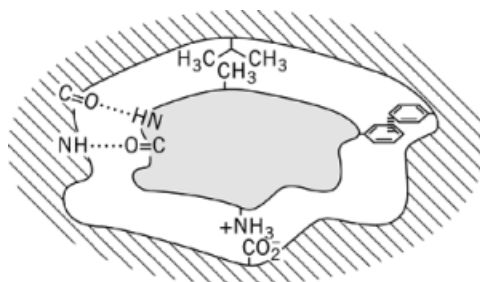
Les organismes vivants s'adaptent aux changements environnementaux en subissant des mutations dans leurs gènes ; de cette façon, les protéines acquièrent de nouvelles propriétés leur permettant de survivre à la sélection naturelle. L'évolution accélérée de protéines peut être effectuée *in vitro* par mutagenèse dirigée ou aléatoire pour créer des mutants possédant de nouvelles caractéristiques.

L'objectif de ce travail est d'introduire des sites allostériques dans une enzyme pour moduler son activité grâce à la liaison d'une molécule étrangère. Cette enzyme chimérique pourrait alors être utilisée pour la détection et le dosage de ce nouveau ligand.

La première section de ce chapitre introductif est consacrée à la description des interactions moléculaires en général. Les différentes techniques permettant de changer la spécificité d'une protéine seront ensuite décrites dans la partie 'Ingénierie des protéines'. Nous y rapporterons également des exemples d'application de ces techniques à la création de nouvelles interactions moléculaires et nous verrons comment l'ingénierie des interactions peut être appliquée à la création de sites de régulation. Une autre section sera consacrée à l'enzyme choisie pour ce travail, la  $\beta$ -lactamase TEM-1. Finalement, les objectifs et la stratégie générale de ce travail seront décrits plus en détail.

## I. Reconnaissance moléculaire

La reconnaissance moléculaire est un terme utilisé pour décrire les interactions non covalentes entre des molécules. Ces processus de reconnaissance sont présents dans chaque système biologique et vont de la liaison de petites molécules aux protéines plasmatiques jusqu'à la liaison de protéines à l'ADN, en passant par l'interaction entre une enzyme et son substrat, entre un antigène et un anticorps, entre un récepteur et un ligand... La compréhension de ces principes reste problématique même si le premier éclaircissement des règles gouvernant la reconnaissance moléculaire date de la fin du 19<sup>ème</sup> siècle. En 1894, Emil Fisher proposa qu'une enzyme et son substrat puissent être comparés à une serrure et une clef pour expliquer la sélectivité entre les molécules impliquées [1]. Dans ce modèle, la structure tridimensionnelle du ligand est complémentaire à celle du récepteur de la même manière qu'une clef est complémentaire à une serrure. Cette complémentarité se passe à deux niveaux. Premièrement, la forme de la cavité du récepteur doit être complémentaire à la forme du ligand. Deuxièmement, une complémentarité chimique est nécessaire entre les groupes de liaison formant l'intérieur de la cavité et les caractéristiques chimiques du ligand (cf. Figure 1).

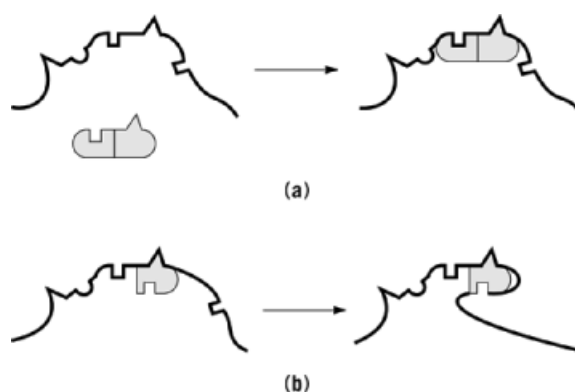


**Figure 1** Représentation schématique d'un complexe hôte–ligand impliquant l'inclusion du ligand dans la cavité et la formation de différents types d'interactions entre l'hôte (hachuré) et le ligand (gris) (reproduit de Weber [2]).

Les forces d'interaction principalement impliquées dans les processus de reconnaissance moléculaire sont les ponts hydrogènes entre centres basiques et acides, les interactions électrostatiques entre centres anioniques et cationiques, les interactions métal–ligand, les interactions dipôle–dipôle, les transferts de charge entre résidus aromatiques, les attractions de van der Waals entre régions hydrophobes et les liaisons covalentes réversibles [3].

Le récepteur peut adopter deux modes de comportement différents. L'un d'entre eux, évoqué précédemment, est représenté par l'image clef–et–serrure, impliquant la complémentarité entre un hôte rigide et un ligand rigide (cf. Figure 2 (a)). L'avantage de cette pré-organisation est la minimisation de la perte d'entropie lors de la liaison.

Cependant, dans la plupart des systèmes biologiques, le récepteur possède un certain degré de flexibilité. L'approche du ligand induit des changements conformationnels et une organisation du site de liaison autour de lui (cf. Figure 2 (b)). Avec ce mécanisme de liaison induit [4], la perte entropique est plus importante mais ce mécanisme possède quelques avantages. Un récepteur flexible permet un meilleur enveloppement du ligand ou même l'encapsulation complète de celui-ci. De cette façon, plus d'interactions seront impliquées, ce qui permettra une haute sélectivité de liaison.



**Figure 2** Mécanismes principaux de la formation d'un complexe récepteur–ligand : (a) modèle rigide clef–et–serrure ; (b) modèle induit, montrant les changements conformationnels du récepteur suite à la liaison du ligand (reproduit de Weber [2]).

Les atomes d'un hôte ne rentrent pas tous en contact avec le ligand cible mais ils imposent la géométrie adéquate pour que les atomes de contact puissent présenter une surface possédant la forme correcte. Ces atomes forment l'ossature de l'hôte. Celle-ci peut être très simple, comme une chaîne peptidique qui maintient les chaînes latérales à l'espacement approprié, ou beaucoup plus compliquée, comme le domaine variable (V) d'un anticorps qui maintient les régions hypervariables (CDRs ou 'Complementary-Determining Region') dans une juxtaposition correcte par rapport à la cible.

L'affinité d'un hôte pour son ligand peut être mesurée grâce à la constante de dissociation ( $K_d$ ) ou l'inverse, la constante d'association ( $K_a$ ). Pour l'équilibre de liaison :



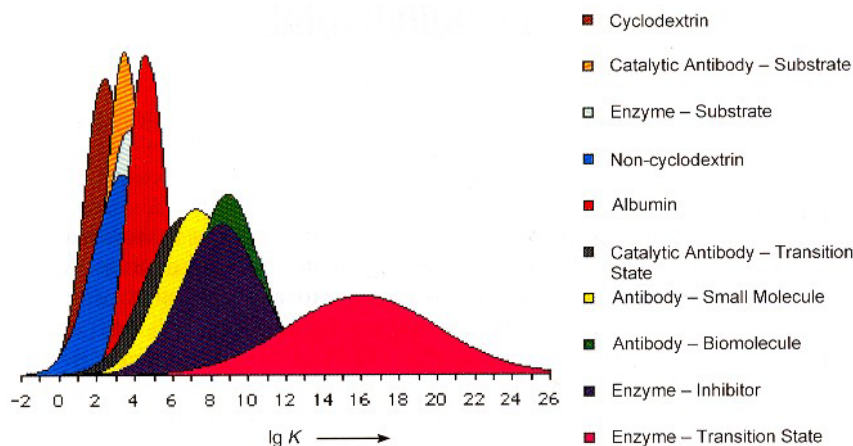
(où A est le récepteur, B est le ligand et AB est le complexe récepteur–ligand), la constante de dissociation vaut :

$$K_d = \frac{[A] \cdot [B]}{[AB]} = \frac{1}{K_a}$$

Donc, plus la constante de dissociation est faible, ou plus la constante d'association est élevée, plus le complexe AB est stable.

Nous pouvons nous poser la question de l'affinité maximale atteignable entre un récepteur donné et son ligand. Dans les biomolécules, une échelle d'affinité très variée est observée. Celle-ci est fonction de la complexité et de la rigidité de l'hôte, de la surface d'interaction, de la nature du ligand. Houk *et al.* ont comparé les affinités de différents hôtes pour leurs ligands [5]. Les classes examinées dans cet article sont les petites cavités synthétiques, les albumines et d'autres protéines liant des molécules organiques, les anticorps avec des biomolécules et des molécules organiques, et les enzymes avec leurs substrats. Par contre, les complexes avec les ions

métalliques n'ont pas été pris en considération dans cette étude. La Figure 3 représente la distribution des affinités obtenues pour chaque classe d'hôte.

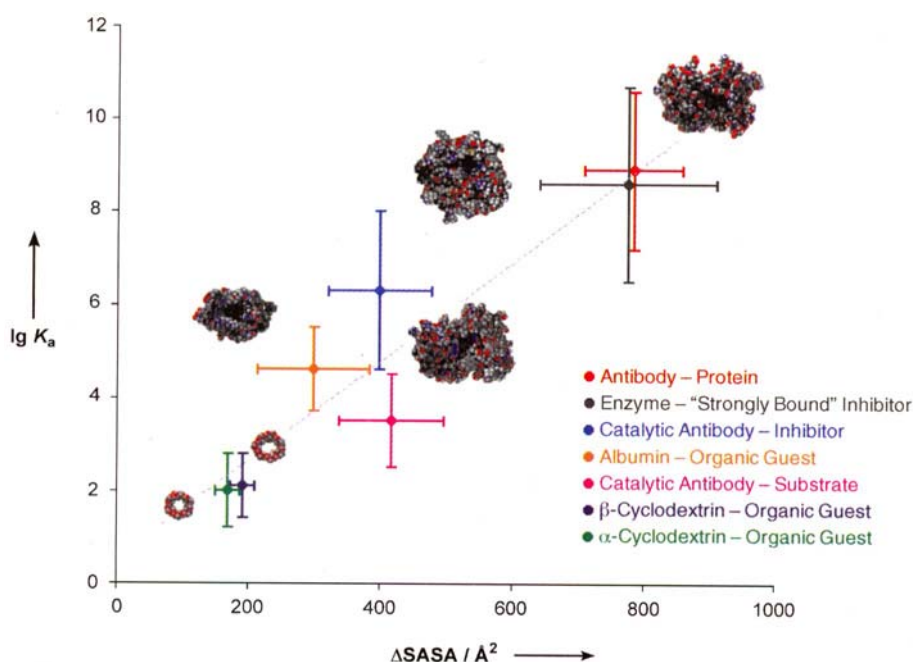


**Figure 3** Constantes d'association moyennes des différents complexes hôte–ligand (reproduit de Houk *et al.* [5]).

L'affinité des différents types d'hôtes se retrouve dans quelques régions bien déterminées. Les complexes impliquant les cyclodextrines, les hôtes synthétiques, les anticorps catalytiques, les enzymes et leur substrat ainsi que les albumines ont des constantes de dissociation comprises entre  $10^{-2,4}$  et  $10^{-4,6}$  M. Le second groupe de complexes se trouve dans l'échelle  $10^{-6,6} - 10^{-8,6}$  M, il comprend les complexes entre des anticorps catalytiques et des états de transition, ou entre des anticorps et des petites molécules organiques et des biomolécules, les complexes récepteur–ligand et les complexes enzymes–inhibiteur. Finalement, les liaisons les plus fortes impliquent les enzymes et les états de transition, avec des constantes de dissociation comprises entre  $10^{-12}$  et  $10^{-21}$  M. Notons ici que les états de transition n'existent pas de manière stable et que les constantes de dissociation ont été calculées sur base des paramètres cinétiques.

Les hôtes, en général, ont une forte préférence pour une série de molécules relativement restreinte, molécules qui s'accordent avec leur cavité de complexation. La plupart des classes d'hôtes possèdent une fourchette de valeurs de  $K_a$  d'au moins quatre ordres de grandeur, ce qui permet une très grande sélectivité lorsque les complexes les plus stables sont comparés aux plus faibles dans une classe particulière.

Les auteurs de cette étude [5] ont également tenté d'établir une corrélation entre la surface de l'hôte intervenant lors de la complexation et les constantes d'affinité pour les différentes classes de complexes. Comme le montre la Figure 4, il y a effectivement une corrélation entre les surfaces enfouies des hôtes et les constantes d'affinité, dont la pente correspond à environ  $90 \text{ Å}^2$  par unité de  $\log(K_a)$ .



**Figure 4** Graphique des  $\log(K_a)$  moyens en fonction de la surface enfouie moyenne (reproduit de Houk *et al.* [5]).

La corrélation entre le  $\Delta\text{SASA}$  (surface enfouie) et la constante d'affinité suggère que l'ordre de grandeur des constantes d'affinités est prévisible à partir de la taille du site de complexation de l'hôte et la surface du ligand qui entre en contact avec l'hôte lors de la complexation. Ceci nous permet donc d'avoir une idée de l'affinité maximale atteignable par ingénierie.

## II. Ingénierie des protéines

Les enzymes peuvent catalyser une large gamme de transformations chimiques avec d'impressionnantes augmentations de la vitesse de réaction (jusqu'à  $10^{19}$  fois) [6]. Mais la performance optimale d'une enzyme n'est pas seulement une question de vitesse maximale. Elle implique un compromis entre spécificité, vitesse, stabilité et d'autres contraintes chimiques ou biologiques. Les enzymes étant capables de catalyser des réactions dans des conditions 'douces' et avec une grande spécificité de substrat (souvent accompagnée d'énantio- ou régio-sélectivité), il n'est pas surprenant que de plus en plus d'enzymes soient utilisées en chimie de synthèse ainsi que dans des applications biochimiques et biomédicales.

L'évolution naturelle a optimisé et spécialisé les enzymes pour des fonctions biologiques spécifiques dans le contexte d'un organisme vivant dans un environnement particulier. Par conséquent, peu d'enzymes sont adaptées aux demandes de la synthèse moderne et des applications biotechnologiques : besoin d'enzymes stables et actives longtemps, enzymes actives dans des solvants non aqueux, enzymes acceptant des substrats non naturels,... Pour leur conférer

ces propriétés, il est donc nécessaire d'en modifier et/ou d'en améliorer leurs capacités naturelles. Dans cette ingénierie d'enzymes, deux stratégies sont possibles : le design rationnel (mutagenèse dirigée) et l'évolution dirigée.

## 1. Le design rationnel

Le design rationnel consiste à tenter de prédire les changements à effectuer dans une protéine pour lui conférer les propriétés recherchées. Ces modifications sont ensuite introduites dans la séquence en acides aminés par mutagenèse dirigée. Cette stratégie nécessite donc une connaissance détaillée de la structure de la protéine et des relations entre séquence, structure et mécanisme/fonction. De plus en plus de données sont accessibles grâce aux progrès réalisés dans la résolution de structures de protéines par spectroscopie RMN ou par diffraction des rayons X sur des cristaux mais les connaissances dans ce domaine restent malgré tout limitées. La modélisation moléculaire permet également de prédire, grâce à des algorithmes informatiques itératifs, les mutations à introduire en l'absence de données structurales ; la structure d'une protéine homologue sert alors de modèle. Les substitutions en acides aminés sont souvent choisies en comparant les séquences de protéines homologues.

En dépit de quelques exemples spectaculaires, beaucoup d'essais de design de protéines ont échoué car les effets de mutations multiples sur la fonction d'une protéine sont difficilement prévisibles, des changements mineurs au niveau de la séquence (même une mutation ponctuelle) pouvant provoquer des perturbations structurales considérables.

### 1.1. Ingénierie des interactions par design rationnel

La création de protéines possédant de nouvelles propriétés de liaison est un objectif important dans l'ingénierie des protéines. Ces nouvelles protéines permettent de fournir des éclaircissements supplémentaires dans les processus de la reconnaissance moléculaire. De plus, elles peuvent également avoir une variété d'applications commerciales comme agents thérapeutiques, comme outils de diagnostic ou comme ligands d'affinité. L'ingénierie des propriétés de liaison peut être réalisée suivant deux approches différentes : soit un site de liaison existant est modifié afin de lui conférer de nouvelles propriétés de liaison, soit un site de liaison est créé *de novo*.

Le design rationnel des protéines a permis la construction de protéines se liant à des ligands simples, comme les ions métalliques. Cependant, la conception de sites de liaison pour des antigènes plus complexes est une tâche ardue. En effet, il est difficile de prédire la combinaison de résidus nécessaire pour promouvoir une interaction particulière.

Les premières expériences de création de site de complexation métallique ont été réalisées suivant une approche empirique. Cette approche implique soit la mutation de résidus en acides aminés coordonnant le métal à des positions spécifiques, soit le remplacement d'une boucle située entre les structures secondaires de la protéine. Comme exemple d'insertion d'acides aminés coordonnant les métaux, nous pouvons citer l'incorporation d'une série de résidus histidine sur une boucle de surface ou aux extrémités d'une protéine. Ceci permet la création d'un site métallique non spécifique de relativement haute affinité (de  $10^{-5}$  à  $10^{-7}$  M) [7]. Ces répétitions d'histidines sont largement utilisées lors de la purification de protéines [8]. Comme exemple de remplacement de boucle, citons le greffage, dans une protéase neutre, d'une boucle liant le calcium provenant d'une thermolysine [9]. L'enzyme chimérique possède une activité comparable à celle de l'enzyme sauvage et lie le calcium avec une constante de dissociation de l'ordre du millimolaire. En présence de forte concentration en calcium (0,1 M), la protéine mutante est deux fois plus stable que la protéine native.

Depuis lors, des algorithmes informatiques itératifs ont vu le jour et ont pu être appliqués au design de métalloprotéines. Les sites de complexation du zinc ont été les premiers à être introduits dans les protéines grâce à de tels programmes informatiques. En utilisant un de ces programmes, Klembe *et al.* ont choisi le domaine B1 d'une protéine G pour y introduire un site de complexation His<sub>3</sub>Cys [10]. En mutant quatre résidus, les auteurs ont créé un site de complexation pour le Co(II) avec une constante de dissociation de 4  $\mu$ M. Des études de compétition avec le Zn(II) suggèrent que ce mutant complexe ce métal environ 1.000 fois plus fortement.

Les techniques informatisées ont également permis le changement de la spécificité d'une protéine liant le maltose (MBP) [11]. Des sites de complexation du zinc ont été créés à l'interface des deux domaines constituant la protéine. Une affinité de 350 nM a été obtenue en combinant deux sites de liaison du zinc, en optimisant la sphère de coordination primaire de l'ion métallique et en manipulant l'équilibre entre les deux formes conformationnelles de la protéine pour stabiliser l'état liant le ligand. Dans ce nouveau senseur moléculaire, la complexation du zinc est détectée par le signal fluorescent d'un groupe rapporteur qui réagit aux changements de conformation dus à la complexation du ligand.

Ces exemples ne représentent qu'une infime partie de ce qui a été réalisé dans la création de sites métalliques (pour une revue sur le sujet, cf. référence [12]).

Récemment, la méthode informatisée a également permis la création de sites de complexation pour des petites molécules [13]. Grâce à la connaissance de la structure tridimensionnelle de la protéine, l'algorithme identifie les séquences en acides aminés qui forment une surface

complémentaire entre la protéine et le ligand cible. Cette méthode a été utilisée pour l'ingénierie de sites de liaison pour le trinitrotoluène (TNT), le L-lactate ou la sérotonine dans cinq membres de la famille des protéines liantes périplasmiques (PBP). L'affinité des meilleurs clones testés expérimentalement est de l'ordre du nanomolaire pour le TNT et du micromolaire pour L-lactate et la sérotonine. De plus, les clones montrent une grande sélectivité : un clone possède une affinité 100 fois supérieure pour le L-lactate par rapport au D-lactate.

## **1.2. Ingénierie de la régulation par design rationnel**

Partant du principe qu'une activité enzymatique peut être modulée par la liaison d'un cofacteur ou d'une protéine différente de l'enzyme, l'ingénierie de la régulation a permis d'accéder à des molécules permettant la détection et/ou le dosage de petites molécules organiques telles que des drogues ou des hormones.

Beaucoup d'enzymes ont été modifiées chimiquement par couplage avec des haptènes divers et leur activité a pu être régulée positivement ou négativement par liaison d'un anticorps anti-haptène. Ces enzymes peuvent donc fonctionner comme senseurs moléculaires en présence de l'haptène libre. En effet, lorsque l'haptène libre se lie à l'anticorps, l'activité initiale est restaurée et ceci en fonction de sa concentration. Comme exemple, nous pouvons citer la conjugaison de la malate déshydrogénase couplée à la morphine [14]. L'activité enzymatique du conjugué enzyme-morphine est inhibée en présence d'anticorps anti-morphine. Dans des expériences de compétition, des concentrations de  $2 \cdot 10^{-9}$  M en morphine et  $2 \cdot 10^{-10}$  M en codéine engendrent une diminution détectable de cette inhibition.

Par la suite, les applications de ce système ont été étendues à la détection de protéines antigéniques : l'insertion génétique d'épitopes linéaires viraux a permis de créer des enzymes hybrides dont l'activité est modulée par liaison aux anticorps monoclonaux correspondants. A nouveau, la liaison compétitive des ligands viraux restaure l'activité. De cette manière, une série de phosphatases alcalines (AP) hybrides a été créée par insertion d'un peptide antigénique de la protéine gp120 du virus HIV ('Human Immunodeficiency Virus') à proximité du site actif de l'enzyme. La présence d'anticorps anti-peptide inhibe de 40–50 % la vitesse catalytique de l'enzyme hybride [15]. L'introduction de deux mutations ponctuelles indépendantes dans l'enzyme inverse la réponse en activant l'enzyme suite à la liaison de l'anticorps. Par la suite, la performance de ces deux senseurs (activé ou inhibé en présence d'anticorps) a été confirmée en utilisant le peptide antigénique de l'hépatite C. Les constructions résultantes répondent à l'anticorps anti-hépatite C de la même façon que les enzymes portant l'épitope HIV [16]. Ces résultats démontrent que la modulation de l'activité enzymatique des protéines hybrides AP–



épitope n'est spécifique ni d'une séquence d'épitope particulière, ni d'une combinaison anticorps-épitope particulière.

La  $\beta$ -galactosidase a également été utilisée pour l'insertion d'épitopes viraux. Un peptide de la protéine VP1 du virus de la fièvre aphteuse dans différentes boucles de l'enzyme [17, 18]. Deux des mutants montrent une activation significative suite à la liaison des anticorps anti-peptide. Dans ces deux constructions, le peptide viral a été inséré dans des boucles proches du site actif et leur activité spécifique est plus faible que celle de l'enzyme sauvage. L'activation allostérique de la vitesse d'hydrolyse est due à une réduction du  $K_m$  et une augmentation du  $k_{cat}$  [19]. L'activation par les anticorps n'était que de 1,8 à 2 fois, mais celle-ci a été augmentée à plus de 4 fois en optimisant les conditions de réaction et en réalisant des insertions multiples du peptide antigénique dans l'enzyme [20]. D'autres senseurs ont été construits avec la  $\beta$ -galactosidase en insérant des peptides antigéniques de la protéine gp41 du virus HIV-1. Ces enzymes-hybrides fournissent également de très bonnes réponses aux sérums humains par rapport aux tests immunologiques commerciaux [21].

Un autre exemple de design rationnel s'inspire du principe de la régulation intrastérique présente dans quelques enzymes naturelles [22]. Dans une construction partiellement synthétique, une protéase neurale a été couplée à son inhibiteur phosphoramidite via une molécule d'ADNsb, charnière courte et flexible, ce qui conduit au blocage de l'activité enzymatique [23]. La présence de brins d'ADN complémentaires, s'hybridant au segment d'ADN charnière et restreignant sa flexibilité, permet le décrochage de l'inhibiteur et restaure dès lors l'activité enzymatique. Celle-ci est donc dépendante de la concentration de molécules d'ADN possédant une séquence spécifique.

La création de sites de complexation métallique a également permis la régulation de l'activité enzymatique. Ceci a notamment été réalisé en mutant un résidu de la trypsine de rat en histidine, à proximité du résidu histidine du site actif [24]. L'ajout de  $Zn(II)$  à ce mutant inactive l'enzyme tandis que l'EDTA rétablit l'activité.

La création de sites de régulation n'a pas été réalisée uniquement dans les enzymes, mais aussi dans les ribozymes (ARNs catalytiques). Comme exemple, citons la construction de quelques ribozymes dont l'activité est contrôlée par l'ATP [25]. Ces constructions ont été réalisées en unissant un domaine 'hammerhead' d'un ribozyme à un domaine d'ARN qui lie l'ATP. Cette fusion crée un conflit stérique entre la structure active du ribozyme et le domaine liant l'ATP, mais uniquement lorsque l'ATP est lié. Par la suite, d'autres constructions ont été réalisées dans lesquelles les acides nucléiques montrent des variations d'activité catalytique suite à la liaison de différents ligands, spécialement les petits effecteurs organiques [26]. Les constantes de

dissociation de quelques ribozymes allostériques ont été déterminées et la plupart montrent des valeurs de l'ordre du micro- ou du millimolaire pour de petits effecteurs organiques [27].

## 2. L'évolution dirigée

L'évolution dirigée (également nommée biotechnologie évolutive ou évolution moléculaire) s'inspire de l'évolution Darwinienne qui attribue toute la complexité du monde vivant à une succession de mutations et de sélection naturelle. Son principe est simple et ne nécessite pas *a priori* de connaissance de structure, fonction ou mécanisme de l'enzyme. L'évolution dirigée se fait en trois étapes : la création d'une banque de gènes mutés, son expression fonctionnelle et la récolte des mutants ayant les propriétés désirées, le(s) gène(s) codant pour ces enzymes améliorées pouvant rentrer dans un autre cycle d'évolution dirigée.

### 2.1. Techniques

#### 2.1.1. Création de la diversité

La création de la diversité génétique peut être réalisée selon deux procédés différents : soit le gène entier subit les mutations, soit un nombre limité de sites sont concernés. Dans le premier cas, les mutations sont introduites au hasard tout le long du gène, en général grâce à des techniques de PCR. Dans le second cas, une mutagenèse exhaustive est réalisée dans des régions très localisées.

La taille des banques ainsi créées est très importante. En effet, plus le nombre de mutants différents est élevé, plus les chances de trouver un mutant intéressant sont grandes. Cependant, une banque de grande taille implique également que ce mutant sera plus difficile à isoler. Des tentatives ont donc été réalisées pour limiter la taille des banques ; soit en combinant l'approche dirigée avec le design rationnel, soit grâce à l'aide d'outils informatiques (cf. Approche informatisée, page 25)

##### (a) Mutagenèse dirigée par oligonucléotide

Les méthodes permettant de rendre aléatoires de petits fragments de gène font partie des premières techniques utilisées dans l'évolution *in vitro*. Les fragments de gènes sauvages sont remplacés par des oligonucléotides synthétiques dégénérés contenant des positions ou des régions aléatoires (cassettes aléatoires) [28, 29] ou des régions semi-aléatoires [30]. Ces méthodes de mutagenèse par cassette ne permettent de rendre aléatoire qu'un petit nombre de résidus proches et ne conviennent que lorsque de courtes régions de la protéine sont visées.

Pour limiter la taille des banques de mutants par insertion de séquences aléatoires, des banques ‘dopées’ ont été imaginées [31]. Les acides aminés ne sont pas mutés de façon complètement aléatoire, seules certaines combinaisons de résidus sont générées à chaque position et ce dans des proportions contrôlées.

(b) Mutagenèse chimique, souches mutagènes et irradiation UV

La mutagenèse peut être réalisée en traitant l’ADN avec des composés mutagènes [32], par irradiation aux UV ou en clonant le gène dans une souche bactérienne mutagène qui ne possède pas les fonctions de réparation de l’ADN [33]. Ces méthodes figurent parmi les premières utilisées dans l’ingénierie évolutive mais elles ont été largement supplantées par des techniques plus modernes. En effet, le principal désavantage de ces techniques est qu’elles ne limitent pas l’effet mutagène au gène d’intérêt mais touchent tout le génome. La sélection pour de nouvelles fonctions peut donc être influencée par des effets non spécifiques, comme des vitesses d’expression altérées, ou même la modification spontanée d’autres enzymes cellulaires ou de voies métaboliques capables de réaliser la tâche pour laquelle la sélection sera réalisée.

(c) La PCR mutagénisante

La PCR mutagénisante ou ‘error-prone PCR’ (epPCR) [34] est une technique simple, efficace et bien caractérisée, qui introduit des mutations ponctuelles aléatoirement dans le gène. En pratique, la fidélité de la polymérase est diminuée par ajout d’ions manganèse et par un déséquilibre des proportions en nucléotides. La fréquence des mutations peut être ajustée en changeant ces conditions. Cette technique est considérée comme ‘évolution asexuée’ car la diversité génétique apparaît dans un seul gène de départ. Elle convient pour de petits fragments génétiques (< 800 pb) mais le nombre de substitutions des acides aminés accessibles est limité par l’improbabilité d’avoir deux ou trois mutations dans le même codon et par les biais importants de l’epPCR (les différents types de changements de nucléotides ne se font pas à la même fréquence). Plusieurs cycles de mutagenèse consécutifs peuvent être réalisés pour contourner ces limitations.

(d) Le ‘DNA shuffling’

La méthode du ‘DNA shuffling’ [35], parfois appelée PCR sexuée, permet de créer des banques de gènes chimériques en recombinant *in vitro* l’ADN de gènes homologues. Comparé à la mutagenèse aléatoire, ce ‘mélange’ de matériel génétique, peut être avantageux en combinant les mutations bénéfiques apparues indépendamment et dont les effets peuvent être additifs ; ou, au

contraire, en combinant des mutations nuisibles qui seront éliminées plus efficacement lors de la sélection [36]. En pratique, de courtes séquences d'ADN sont générées par fragmentation enzymatique aléatoire à partir de gènes parentaux différents. Ce mélange de fragments est ensuite recombiné en gènes entiers par PCR d'assemblage.

La diversité des gènes peut être obtenue de différentes façons :

- par l'utilisation, comme mélange de gènes initial, de gènes issus d'une banque créée par PCR mutagénisante,
- ou de gènes homologues provenant d'organismes différents [37],
- en exploitant le système d'extension et recombinaison lui-même, qui génère un taux d'erreurs ajustable [38],
- en utilisant des amorces mutées [39],
- en engageant les mutants sélectionnés dans un deuxième cycle d'évolution.

(e) Recombinaison *in vivo* et *in vitro*

La recombinaison homologue peut également être réalisée *in vivo*. La méthode la plus utilisée est basée sur la recombinaison intermoléculaire de fragments d'ADN linéaires homologues dans la levure [40]. Des gènes mutants flanqués de séquences homologues au vecteur sont co-transformés avec le vecteur linéarisé [41]. La recombinaison intermoléculaire va permettre la recombinaison des gènes entre eux et avec le plasmide linéaire qui va ainsi être circularisé.

Une technique plus récente de recombinaison, le CLERY ('Combinatorial Libraries Enhanced by Recombination in Yeast') combine le 'DNA shuffling' *in vitro* et le 'shuffling' *in vivo* dans la levure [42]. Cette procédure convient donc particulièrement bien au 'shuffling' de gènes eucaryotes qui ne peuvent pas être exprimés dans des bactéries.

La recombinaison naturelle, elle-même, a été utilisée pour créer de grandes banques de mutants. L'«exon shuffling», par exemple, permet d'assembler des exons indépendants en gènes codant pour de nouvelles protéines. Ceci est réalisé par recombinaison au sein des introns. De grandes banques combinatoires ont été générées par «exon shuffling» [43] en utilisant deux exons artificiels aléatoires et le site de recombinaison lox cloné dans un intron auto-excisant. Par la suite, ce concept de «exon shuffling» a été adapté pour générer la diversité génétique *in vitro* [44]. L'«exon shuffling» permet de recombinaison de très longues séquences d'ADN et particulièrement les gènes des organismes eucaryotes.

#### (f) Approche informatisée

Une approche récente utilise des programmes informatiques de design de protéine. Ces méthodes informatiques ont été imaginées dans le but d'optimiser l'évolution *in vitro* des protéines. En simulant l'évolution avec un modèle statistique, les résidus dans une protéine qui sont tolérants à la mutagenèse sont identifiés et, très souvent, ces sites correspondent aux mutations bénéfiques trouvées lors d'expériences d'évolution dirigée [45]. Si un nombre limité de sites 'mutagénisables' peuvent être identifiés par cette méthode, toutes les substitutions possibles des acides aminés sont réalisables (ce qui n'est pas le cas pour la mutagenèse aléatoire) à l'aide d'oligonucléotides dégénérés. Cependant, cette méthode nécessite une connaissance de la structure tridimensionnelle ou un bon modèle homologue de la protéine et ne permet pas de déterminer *a priori* les sites où les mutations bénéfiques vont se situer.

Ce concept de pré-criblage des banques *in silico* a été étendu grâce à la technologie PDA ('Protein Design Automation') qui élimine les séquences qui sont incompatibles avec le repliement de la protéine [46].

#### **2.1.2. Lien génotype – phénotype**

Lors de l'évolution dirigée d'enzyme, la diversité est créée au niveau de l'ADN tandis que la sélection ou le criblage se passe au niveau de la protéine encodée. Par conséquent, l'expression fonctionnelle de l'information contenue dans les banques d'ADN est essentielle pour la détection des enzymes mutantes améliorées. L'approche la plus répandue pour l'expression de protéines recombinantes utilise les machineries de transcription et de traduction d'organismes tels qu'*Escherichia coli*, *Saccharomyces cerevisiae* ou *Bacillus subtilis*. Ces systèmes d'expression cellulaire garantissent également l'association d'une protéine mutante spécifique avec le gène qui l'encode (un mutant par cellule). Ceci est essentiel pour l'identification et l'amplification des mutants désirés après sélection ou criblage, de même que pour d'autres cycles d'évolution. Alternativement, un lien physique entre le génotype et le phénotype peut être établi en fusionnant la protéine d'intérêt à son gène. Contrairement aux systèmes cellulaires, les conditions de sélection ou de criblage, comme la température ou la composition du milieu, peuvent être mieux contrôlées dans un système *in vitro*. Plusieurs méthodes ont été imaginées dans ce sens et sont décrites ci-dessous. La technique du 'phage display' y est exposée plus en détail car celle-ci est utilisée dans ce travail.

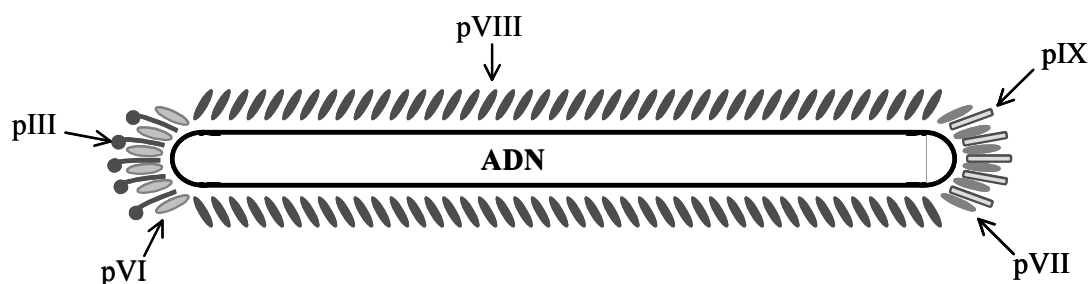
(a) 'Phage display'

Le système le plus répandu actuellement pour la sélection de banques est l'expression en surface de phages, ou 'phage display' [47, 48]. Cette méthode est basée sur le fait que si un fragment de gène ou un gène entier codant pour un polypeptide est fusionné à un gène d'une protéine du manteau du phage, ce gène 'fusionné' peut être encapsulé dans les particules phagiques qui exprimeront le polypeptide encodé à leur surface. De cette façon, un lien physique est établi entre le phénotype et le génotype.

Grâce aux techniques de biologie moléculaire, de grandes banques de polypeptides exprimées en surface de phages peuvent être générées. Les phages exprimant les polypeptides comportant la propriété désirée peuvent ensuite être sélectionnés (cf. Sélection, page 31) et les séquences de ces polypeptides peuvent être déduites à partir de la séquence de l'ADN encapsulé. Le succès du 'phage display' dépend donc de la taille de la banque initiale. Les méthodes actuelles permettent la construction de banques exprimées sur phages contenant jusqu'à  $10^{11}$  séquences d'ADN différentes [49]. L'utilité des banques sur phages dépend également de la qualité de l'expression. En effet, un membre de la banque ne sera sélectionnable que si le polypeptide est suffisamment exprimé à la surface du phage. Les taux d'expression varient en fonction de la longueur et de la séquence du polypeptide [50] : des centaines, voire des milliers de copies de petits peptides (< 8 résidus) peuvent être exprimés par particule phagique [51], tandis que les protéines plus importantes (> 100 résidus) sont généralement exprimées à un taux d'une copie ou moins par phage [52-54]. Actuellement, l'enzyme la plus volumineuse exprimée sur phage est une pénicilline acylase (86 kDa), avec un taux d'expression d'une copie par phage [55].

(i) *Structure des phages filamenteux*

Les bactériophages filamenteux sont des virus qui infectent les souches d'*Escherichia coli* contenant l'épisome F'. La particule virale est composée d'une molécule d'ADN simple-brin enveloppée dans une longue capside cylindrique d'environ 1  $\mu\text{m}$  de long sur moins de 10 nm de diamètre. Celle-ci contient cinq protéines différentes (Figure 5) mais la grande majorité consiste en quelques milliers de copies de la protéine pVIII, codée par le gène 8 (g8p), qui couvrent la longueur de la particule. Les quatre protéines mineures sont présentes à raison d'environ cinq copies par particule : les protéines pVII et pIX ferment une extrémité, tandis que les protéines pIII et pVI sont situées à l'autre extrémité [56].



**Figure 5** Structure schématique d'un phage filamentueux.

Les cinq protéines de manteau contribuent à la stabilité structurale du phage mais la pIII est également nécessaire à la reconnaissance et l'infection de la cellule hôte. Par conséquent, elle est plus volumineuse et plus complexe que les autres protéines du manteau et contient trois domaines distincts [57]. Le domaine amino-terminal initie la translocation de l'ADN viral dans *E. coli* pendant l'infection, tandis que le deuxième domaine confère la reconnaissance de la cellule hôte en se liant au pilus sur la surface d'*E. coli* [58]. Quant au domaine carboxy-terminal, il interagit avec d'autres protéines du manteau phagique et est responsable de l'intégration de la pIII dans le manteau [59].

#### (ii) Assemblage du phage filamentueux

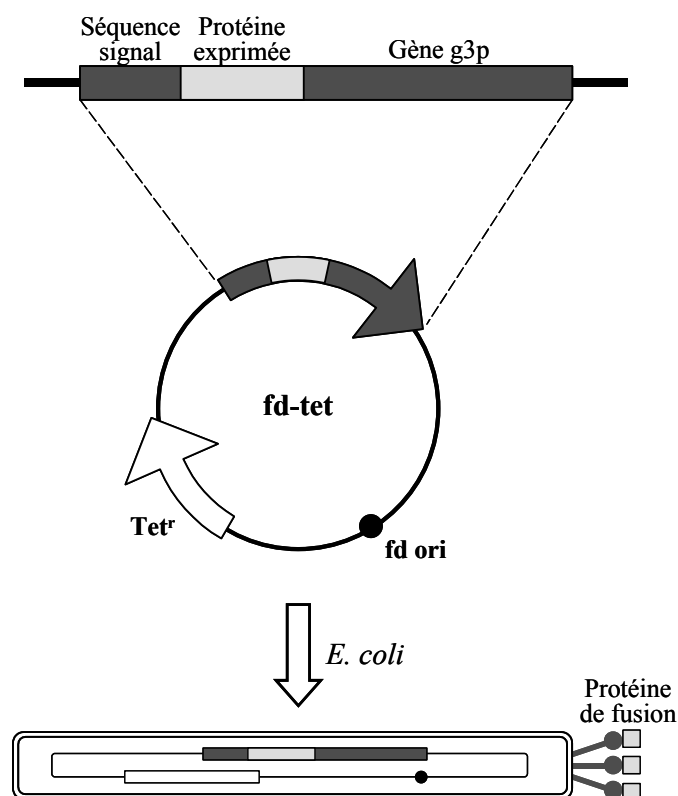
Le génome des phages filamentueux contient onze gènes : cinq de ces gènes encodent les protéines du manteau décrites plus haut, tandis que les autres encodent les protéines nécessaires à la réplication et l'assemblage du virus. L'infection est initiée lorsque la pIII se lie au pilus présent à la surface d'une cellule hôte permettant la translocation de l'ADN viral dans le cytoplasme de la cellule. A l'intérieur de celle-ci, l'ADN simple-brin (brin +) est converti en une forme répliquative double-brin par la synthèse du brin complémentaire (brin -). Celle-ci est ensuite dupliquée par un mécanisme de réplication de type 'rolling-circle' et utilisée comme matrice pour la production de toutes les protéines phagiques. De cette façon, les composants des nouvelles particules phagiques sont accumulés dans la cellule hôte.

Avant l'assemblage des particules, les protéines du manteau sont stockées au niveau de la membrane interne, avec leur extrémité amino-terminale dans le périplasme et la partie carboxy-terminale dans le cytoplasme. D'autres protéines virales forment un pore qui traverse les membranes interne et externe. L'ADN simple-brin est extrudé par ce pore et est simultanément enveloppé par les protéines du manteau qui glissent de la membrane interne vers la particule en formation. L'assemblage est initié par l'incorporation des protéines pVII et pIX à une extrémité de la particule. L'élongation se poursuit et les milliers de molécules pVIII sont assemblées tout le long de la particule. Le processus se termine par l'addition des pVI et pIII à l'autre extrémité du virus. La particule ainsi assemblée est finalement libérée dans l'environnement

extracellulaire. L'assemblage des phages est un processus non lytique ; les nouvelles particules assemblées sont continuellement extrudées des cellules infectées qui continuent à croître et à se diviser, bien qu'à une vitesse réduite.

(iii) *L'expression en surface de phage*

Les vecteurs les plus utilisés pour l'expression en surface de phage sont les bactériophages filamenteux fd, M13 ou les phagemides apparentés [60]. Les premiers systèmes d'expression étaient basés sur des fusions à l'extrémité amino-terminale de la pIII (Figure 6) ou de la pVIII dans le génome viral [61-63]. Ces systèmes sont limités par le fait que les polypeptides plus longs (> 10 résidus pour l'expression en pVIII) risquent de compromettre la fonction de la protéine de manteau et diminuent de ce fait l'efficacité de l'expression.



**Figure 6** Expression en surface de phage sous forme de fusion amino-terminale avec la protéine pIII. Le vecteur fd contient un marqueur sélectionnable (résistance à la tétracycline, Tet<sup>r</sup>) pour le maintenir dans *E. coli* et une origine de réplication simple-brin (fd ori). Le gène de la protéine étrangère est inséré entre les régions codant pour la séquence signal de la pIII et la séquence de la pIII mature. Lors de la morphogénèse du phage, la protéine étrangère est exprimée sous forme de protéine de fusion avec la pIII.

Par la suite, l'utilisation de systèmes hybrides a permis l'expression de protéines plus volumineuses. Il persiste néanmoins une limite à la taille des protéines exprimables car celles-ci



doivent subir l'extrusion à travers un pore pour permettre leur exportation vers le périplasme. Dans le phage hybride, la protéine de fusion ne représente qu'un composant additionnel du manteau qui contient des copies sauvages des cinq protéines qui le composent. Ceci peut être réalisé en additionnant le gène de fusion à un génome viral complet, de sorte que deux versions d'une des protéines de manteau soient exprimées : la sauvage et une copie fusionnée à la protéine à exprimer [64]. Alternativement, l'expression hybride peut être réalisée grâce à un système de phagemide [52]. Le phagemide se réplique dans *E. coli* sous forme de plasmide double-brin, mais la co-infection avec un phage assistant ('helper') permet la production de l'ADN du phagemide sous forme simple-brin, qui sera empaqueté dans des particules phagiques. Le phage assistant fournit toutes les protéines nécessaires à l'assemblage du phage, y compris des copies sauvages des protéines de manteau. Pour l'expression en surface de phage, le phagemide contient un gène codant pour une des protéines de manteau fusionnée au polypeptide à exprimer [52]. La particule phagique résultante contient les protéines de manteau sauvages fournies par le phage assistant ainsi que la protéine fusionnée provenant du phagemide. Grâce aux systèmes hybrides, de grosses protéines peuvent être aisément exprimées sous forme de fusion N-terminale avec la pIII ou la pVIII [52, 54]. Ces systèmes ont également permis le développement d'autres formes d'expression. Des fusions amino-terminales ont été réalisées avec la pVII et la pIX [65]. De plus, des expressions carboxy-terminales ont été effectuées avec la pVI, la pIII et la pVIII [66-68]. Donc, grâce aux systèmes de phagemide, l'expression de polypeptide fonctionnel a été démontrée avec les cinq protéines de manteau du phage. Les phagemides permettent également l'expression de protéines toxiques car, sous le contrôle d'un promoteur inductible, leur expression peut être activée simultanément à l'infection du phage assistant [69].

#### (b) Fusions ARN-peptide

La fusion covalente entre un peptide et l'ARNm qui l'encode est réalisée en couplant une puromycine (antibiotique analogue à un ARNt chargé) à l'extrémité 3' de l'ARNm. Lors de la traduction de l'ARNm *in vitro*, l'extrémité 3' puromycine est transférée sur l'extrémité carboxy-terminale de la chaîne peptidique [70]. Suite à cela, l'ARNm est lié de façon covalente au peptide et le conjugué ARN-peptide se dissocie du ribosome. Il peut ensuite être sélectionné pour la fonction désirée ; l'ARN des séquences retenues est transcrit par transcription inverse, amplifié et cloné pour être analysé [71]. Cette technique a énormément de potentiel car les banques créées atteignent des tailles jamais obtenues par d'autres méthodes, c'est-à-dire plus de

$10^{13}$  molécules, avec comme seule limite la nécessité d'avoir un système de détection assez sensible (il doit pouvoir détecter une seule molécule).

(c) 'Ribosome display'

Une alternative au couplage chimique entre l'ARNm et le peptide est de maintenir leur association sur le ribosome. Cette méthode, développée pour des banques de peptide [72] et de protéines fonctionnelles [73], permet également l'obtention de grandes banques de clones. Les étapes de transformation ne sont pas nécessaires et l'amplification par PCR à chaque cycle d'enrichissement rend possible l'introduction de mutations aléatoires.

(d) Micro-compartmentalisation

Tawfik et Griffiths ont décrit un système ingénieux dans lequel la compartimentalisation cellulaire est mimée grâce à une émulsion eau-huile [74]. Dans ce système, les gènes sont compartimentalisés dans des gouttelettes d'eau dispersées dans de l'huile. Ces émulsions sont stables, faciles à préparer et permettent de travailler uniquement *in vitro*. Cette technique a permis l'évolution dirigée d'une méthyltransférase [75] et de la polymérase Taq [76]. Elle a également été adaptée à la sélection pour affinité [77] et, plus récemment, cette méthode a fait ses preuves dans la sélection pour catalyse avec un substrat non-ADN, qui a rendu possible l'ingénierie d'une phosphotriestérase possédant une activité catalytique très élevée [78].

(e) Peptide sur plasmide

La méthode la plus simple pour coupler l'information et la structure *in vivo* est l'association physique directe de la protéine à son gène par un domaine liant l'ADN. Pour des banques de peptides, cela a été réalisé en les fusionnant à l'extrémité carboxy-terminale du répresseur *lac* [79]. Cette technique nécessite cependant une lyse cellulaire (pour isoler le complexe récepteur-plasmide) douce et quantitative. De plus, le domaine liant l'ADN doit avoir une vitesse de dissociation très lente afin d'éviter les échanges entre plasmides.

(f) Expression en surface de cellule

Plusieurs milliers de copies de banques de protéines ou de peptides peuvent être exprimées en surface de cellules par la fusion d'un peptide hétérologue ou d'une protéine d'intérêt à une protéine ancrée à la surface cellulaire. Ces banques peuvent ensuite être criblées avec un ligand fluorescent ; les cellules étant ensuite triées par FACS ('Fluorescence Assisted Cell Sorting') [80]. Cette technique très prometteuse est de plus en plus utilisée [81].

### 2.1.3. Criblage et sélection

Le criblage nécessite l'analyse systématique des mutants ayant la propriété désirée en les comparant avec tous les autres éléments de la banque alors que la sélection signifie la capture des mutants sans analyse systématique. Dans le cas du criblage, la taille des banques doit être limitée car, en principe, chaque mutant doit être testé individuellement ; la vitesse du test de criblage va donc déterminer la taille analysable. Les méthodes de sélection peuvent quant à elles s'appliquer à des banques beaucoup plus grandes et leur sensibilité est bien meilleure.

#### (a) Criblage

Cette technique requiert l'observation directe de la propriété désirée par des analyses physiques ou biochimiques. Si le criblage permet le contrôle des contraintes soumise et est applicable dans des environnements non naturels ou avec des substrats non naturels, le nombre de mutants pouvant être testés est néanmoins limité par la réaction enzymatique et la sensibilité de la méthode de détection. En général, les techniques de criblage sont utilisées pour rechercher des propriétés catalytiques détectées visuellement ou spectroscopiquement, comme par exemple l'apparition, le changement ou la disparition de fluorescence.

Le criblage se fait traditionnellement par analyse des colonies isolées de bactéries. Celles-ci sont soit visualisées directement, soit lysées ; dans ce cas, des méthodes immunochimiques sont utilisées. Pour les réactions qu'il est nécessaire de quantifier, des plaques multi-puits sont utilisées. Beaucoup d'efforts sont faits pour augmenter le nombre de puits et réduire le volume nécessaire. Cette étape de criblage est également facilitée grâce aux méthodes HTS ('High Throughput Screening') qui permettent le criblage à haut débit [82].

#### (b) Sélection

La sélection peut se faire *in vivo* si un avantage biologique significatif est conféré par les clones possédant l'amélioration souhaitée. Cette méthode permet des enrichissements très efficaces. Cependant, la sélection *in vivo* n'est applicable que dans une minorité de cas. De plus, la pression de sélection sur les cellules vivantes provoque une variété de réponses qui peuvent biaiser la sélection. L'avantage sélectif peut être apporté par complémentation fonctionnelle (en apportant une certaine fonction à une souche ne la possédant pas) ou en couplant la fonction d'intérêt à l'expression d'un gène rapporteur sélectionnable (une résistance à un antibiotique par exemple).

On peut aussi appeler sélection les procédés d'enrichissement *in vitro*, qui ne dépendent donc pas de la survie de cellules. Ces techniques (appelées aussi 'panning') sont le plus fréquemment

utilisées pour la sélection de molécules liant un ligand (immobilisé sur une matrice). Plus récemment, l'approche a également été appliquée à l'enrichissement sélectif de banques d'enzymes fonctionnelles (exprimées sur phage). Donc l'idée de rechercher les meilleurs partenaires de liaison doit être élargie à la recherche de meilleurs catalyseurs en utilisant des analogues de l'état de transition [83], des substrats suicides immobilisés [84] ou des substrats liés de façon covalente au même phage ou à un autre phage via une seconde fusion [85-87]. Récemment, l'enrichissement d'enzymes améliorées par 'biopanning' a été réalisé sur base de l'activité catalytique, c'est-à-dire sur base des paramètres cinétiques [78, 88]. Dans ces expériences, le produit de la catalyse permet la sélection des mutants actifs.

## **2.2. Ingénierie des interactions par évolution dirigée**

Les technologies de création de répertoires et de sélection ont fourni une variété de méthodes puissantes pour générer des polypeptides possédant de nouvelles propriétés de liaison. Ces techniques impliquent la création de grands répertoires de polypeptides et la sélection à partir de ces banques des quelques membres qui possèdent la propriété de liaison recherchée ; les phages filamenteux [47] ont largement été utilisés dans ce but. Comme nous allons le voir plus loin, ces polypeptides n'étaient au départ que de simples peptides linéaires. Par la suite, des efforts considérables ont été réalisés dans le développement de stratégies visant à présenter ces séquences aléatoires dans une conformation restreinte, en greffant les positions aléatoires dans une ossature protéique.

Plusieurs arguments plaident en faveur de l'utilisation de protéines ordonnées de structure connue comme ossature de présentation. Premièrement, de plus grandes affinités peuvent être théoriquement atteintes car la perte d'entropie due à la liaison est plus faible [89]. Deuxièmement, nous pouvons nous attendre à une plus grande stabilité protéolytique dans le cas d'une protéine reployée que dans le cas d'un peptide non structuré. De plus, une ossature protéique permet la présentation effective de résidus hydrophobes, qui seraient enfouis dans un format peptidique. Finalement, les propriétés de liaison d'un domaine seront plus facilement maintenues si celui-ci est transféré dans un autre domaine ou une autre protéine.

Plutôt que de créer ces ossatures de présentation *de novo*, la plupart des expériences d'ingénierie par insertion ont été réalisées sur des protéines ou des domaines existants. Le choix de l'ossature se base sur plusieurs paramètres et aucune protéine n'est idéale pour toutes les situations. Pour l'expression comme protéine de fusion sur phage filamenteux, une condition préalable est que la protéine originale soit effectivement exprimée et sécrétée par les systèmes d'expression d'*Escherichia coli*. La protéine doit également comporter à sa surface des régions qui tolèrent les

variations de séquence ou de longueur, sans affecter la structure tridimensionnelle de l'ossature protéique [90].

Dans les cas où la protéine native possède déjà une fonction de complexation, les résidus impliqués dans l'interaction peuvent être le premier choix pour la mutagenèse aléatoire. Si elle est disponible, la structure de tels complexes peut être très utile pour le choix des régions à muter. Quelques exemples décrivent l'utilisation de telles stratégies pour améliorer la fonction de liaison native [91-93], mais la sélection de protéines possédant des spécificités de liaison entièrement différentes est un défi d'une autre ampleur.

Contrairement aux banques peptidiques dans lesquelles la variation est introduite dans une portion continue de la séquence, dans les banques de protéines, les résidus peuvent être rendus aléatoires dans plusieurs régions discontinues de la chaîne polypeptidique de manière simultanée. Ces résidus aléatoires sont choisis de façon à ce que, dans la protéine reployée, ils forment un secteur continu à la surface de la protéine. Par exemple, dans les banques d'anticorps, la variation est introduite dans six régions différentes de la chaîne polypeptidique. Lors du repliement de ces anticorps, ces régions aléatoires se rassemblent à une face de la protéine pour former un répertoire de sites liant l'antigène.

### **2.2.1. Peptides**

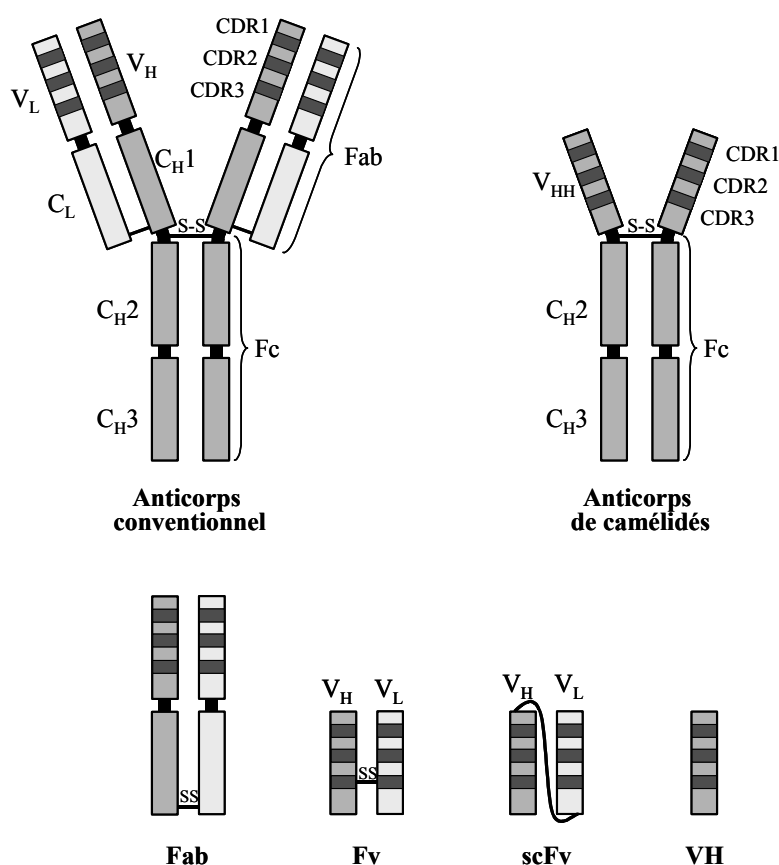
De nombreuses séquences peptidiques intéressantes ont été isolées en utilisant différentes techniques d'expression pour la sélection ou le criblage de banques peptidiques. Ces techniques comprennent les banques de peptides solubles [94], les banques immobilisées sur billes [95], les fusions avec des protéines liant l'ADN [96] et les banques exprimées sur bactéries [97] ou sur phages [98]. Les premières banques de phages étaient constituées de peptides de six résidus aléatoires fusionnés à l'extrémité N-terminale de la protéine pIII du phage [61]. Par la suite, beaucoup de banques de peptides différentes ont été générées, exprimant des peptides plus longs (jusqu'à 40 résidus [99]) ou des peptides contraints (dans lesquels une séquence aléatoire est flanquée de deux résidus cystéine [100]). Ces banques ont été une source riche de ligands pour de nombreuses cibles, notamment les domaines SH3 ('Src Homology 3) [101], le récepteur de l'urokinase [102], la protéine p53 [103], le récepteur de la trombine [104], une variété d'intégrines incluant la fibronectine et des récepteurs de fibrinogène [100, 105] ainsi qu'un récepteur de surface d'une cellule B du lymphome [106]. De plus, ces banques de peptides ont permis d'identifier les épitopes de différents anticorps [107], les ligands d'oligonucléotides simple-brin [108] et les substrats de protéases [109]. Finalement, elles ont également permis l'identification de peptides linaires qui miment des petites molécules comme la biotine [61] ou

des antigènes carbohydrates [110]. Même si les peptides ont beaucoup d'avantages (facilité de synthèse, faible poids moléculaire, faible risque de toxicité, etc.), ils présentent également des désavantages sérieux, comme l'instabilité et de faibles affinités.

### 2.2.2. Anticorps

Initialement, des banques de séquences naturelles d'immunoglobuline ont été utilisées pour la sélection *in vitro* de fragments d'anticorps possédant de l'affinité pour les antigènes ou les haptènes d'intérêt. Ces banques provenaient d'organismes donneurs, préalablement immunisés ou non, (banque naïve) et étaient exprimées soit sous forme de fragments Fab (fragment liant l'antigène), soit comme chaîne unique Fv (ou scFv) [111, 112] (Figure 7). Les fragments d'anticorps obtenus de cette façon étaient encore largement naturels, malgré l'introduction de nouvelles liaisons entre les chaînes lourdes et légères. Par la suite, le potentiel de ces fragments d'immunoglobuline a été étendu grâce à la construction de banques synthétiques partielles ou complètes. Dans ce cas, une collection d'immunoglobulines est combinée avec une ou plusieurs régions hypervariables synthétiques et aléatoires (pour une revue, cf. référence [113]).

Rapidement, l'utilisation d'un domaine variable unique a été prise en considération pour la génération de petites protéines de liaison [114], l'avantage étant d'éviter le clonage de deux régions codantes séparées. Cependant, même si un domaine  $V_H$  ne contenant que trois CDRs (régions hypervariables) paraît suffisant pour la reconnaissance d'antigènes macromoléculaires, il pourrait perdre la capacité de former la cavité qui est nécessaire pour la liaison des haptènes. De plus, ces domaines  $V_H$  ont tendance à s'agglomérer, compte tenu de leur faible solubilité. La situation a changé lors de la découverte d'une classe particulière d'anticorps ne possédant pas de chaîne légère (Figure 7) [115]. Ces anticorps à chaîne lourde (HCAb ou 'Heavy Chain Antibody') font partie d'une sous-classe naturelle des IgG de camélidés. Les domaines  $V_H$  de ces anticorps possèdent des surfaces plus hydrophiles, spécialement au niveau de la région qui fait face au domaine  $V_L$  dans les Ig normaux. Ils possèdent également un CDR3 beaucoup plus long, qui contient souvent un pont disulfure supplémentaire stabilisant la structure [116]. Ces découvertes ont été exploitées afin de générer des banques de domaines  $V_H$  solubles [117, 118]. Un domaine  $V_H$  'camélisé' a été construit comme matrice pour la préparation de banques dans lesquelles le CDR3 est rendu aléatoire. Ces banques exprimées sur phage ont permis la sélection de variants possédant de l'affinité pour deux haptènes : le phényloxazolone ou le nitrophénacétyle, avec des valeurs de  $K_D$  comprises entre 100 et 400 nM.



**Figure 7** Représentation schématique d'un anticorps conventionnel, d'un anticorps de camélidés et des différents fragments d'anticorps utilisés pour l'ingénierie des interactions. Les chaînes lourdes (H) sont indiquées en gris moyen, les chaînes légères (L) en gris clair et les régions hypervariables (CDR) en gris foncé.

Les plus petits domaines d'immunoglobuline construits sont les 'minibody'. Basés sur le domaine V<sub>H</sub> d'un anticorps de souris, ils ne sont constitués que de 61 résidus et ne possèdent que deux boucles hypervariables, correspondant aux CDR1 et 2 [119]. Des acides aminés y ont été substitués pour augmenter la solubilité de l'ossature (substitutions hydrophiles aux positions qui font normalement face au domaine V<sub>L</sub>), pour stabiliser les boucles, pour éviter la présence de cystéines solitaires et pour y introduire un site de liaison métallique (par l'introduction d'histidines). La complexation de l'ion métallique permet d'indiquer un repliement correct. Une ou les deux boucles hypervariables ont ensuite été rendues aléatoires et la banque résultante a permis la sélection d'un clone possédant de l'affinité pour l'interleukine-6 humaine ( $K_D = 0,23 \mu M$ ) [120].

### 2.2.3. Inhibiteurs de protéase

Par la suite, d'autres protéines que des anticorps ont été utilisées pour la création de répertoires de sites de liaison. C'est notamment le cas des inhibiteurs de protéases, qui sont des petites

protéines remarquablement stables. Le premier exemple rapporté est la création d'une banque de mutants dans laquelle une boucle de l'inhibiteur de la trypsine pancréatique de bovin (BPTI) a été rendue aléatoire [121]. La banque a été soumise à une sélection pour affinité pour une élastase et a permis d'isoler un clone dont le  $K_D$  pour la protéine cible était de 1 pM. De plus, ce mutant de BPTI semble être spécifique pour l'élastase.

Par la suite, d'autres inhibiteurs de protéase ont été utilisés : l'inhibiteur du précurseur de la  $\beta$ -protéine amyloïde d'Alzheimer [122], l'inhibiteur de la coagulation associée aux lipoprotéines [123], l'inhibiteur de la trypsine pancréatique humaine [124], un inhibiteur de protéase à sérine bactérien [125], le tendamistat (un inhibiteur de l' $\alpha$ -amylase) [126]. Cependant, seules une ou deux boucles peuvent être modifiées dans ces inhibiteurs. De plus, les cibles utilisées jusqu'à présent sont des protéases et il serait intéressant d'étendre les applications à des macromolécules cibles différentes.

#### **2.2.4. Protéines constituées d'un fuseau d'hélices**

Une autre catégorie de protéines a été employée pour la création de sites de complexation ; il s'agit des protéines formées d'un faisceau d'hélices  $\alpha$ . C'est le cas notamment du cytochrome  $b_{562}$ , qui est composé de quatre hélices  $\alpha$ . Deux boucles de cette protéine ont été rendues aléatoires et une sélection a été réalisée sur un dérivé de la *N*-méthyl-*p*-nitrobenzylamine couplé à la BSA [90]. Les clones sélectionnés possédaient des constantes de dissociation pour l'haptène couplé allant de 5 à 22  $\mu$ M. Malheureusement, l'haptène n'était plus reconnu lorsqu'il était couplé à une autre protéine.

De la même manière, le domaine Z d'une protéine A, composé de trois hélices  $\alpha$ , a servi d'ossature pour la création de nouvelles propriétés de liaison [127]. Trois protéines cibles ont été utilisées : la *Taq* DNA polymérase, l'insuline humaine et l'apolipoprotéine A-1 humaine. Les meilleurs clones sélectionnés possédaient des constantes de dissociation pour la protéine cible de l'ordre du micromolaire.

#### **2.2.5. Lipocalines**

Une nouvelle classe de protéines-récepteur stable, appelée 'anticaline', est basée sur l'ingénierie des lipocalines. Ces petites protéines sont caractérisées par un tonneau  $\beta$  conservé et quatre boucles qui forment l'entrée de la cavité de complexation du ligand. Ces quatre boucles ont subi une mutagenèse aléatoire et trois clones de la banque résultante ont été sélectionnés sur la fluorescéine [128]. Ceux-ci complexent le ligand avec une haute affinité ( $K_D = 35,2$  nM). Cette



banque a également permis la sélection de clones spécifiques à une digoxigénine avec des constantes de dissociation d'environ 30 nM [129].

### 2.2.6. Autres protéines

Une protéine polydactyle composée de six domaines à doigt de zinc a été construite en combinant les techniques de phage display et de design rationnel [130, 131]. Celle-ci reconnaît spécifiquement une séquence de 18 nucléotides dans le gène humain *erbB-2* et lie la séquence cible avec une affinité de 0,5 nM. Lorsqu'elle est exprimée sous forme de protéine de fusion avec un domaine activateur ou répresseur, le facteur de transcription est capable de réguler spécifiquement le promoteur *erbB-2* [132]. De plus, le gène *erbB-3* n'est pas régulé bien que les sites cibles de *erbB-2* et *erbB-3* partagent 15 nucléotides sur les 18 [133].

Un nouveau site de liaison a été créé dans la néocarzinostatine, une protéine utilisée comme agent antitumoral [134]. Une banque a été construite en substituant aléatoirement 13 résidus dont la chaîne latérale pointe vers une crevasse de la protéine. L'utilisation de l'expression en surface de phages a permis la sélection des clones liant un ligand modèle, la testostérone, les constantes de dissociation des protéines sélectionnées pour la testostérone étant comprises entre 7 et 55  $\mu$ M. Mais ces clones possédaient le même niveau d'affinité pour les 15 stéroïdes apparentés à la testostérone testés.

Récemment, Zeytun *et al.* ont greffé des boucles de reconnaissance d'anticorps dans une protéine fluorescente, la GFP ('Green Fluorescent Protein') [135]. Pratiquement, jusqu'à quatre boucles situées à une extrémité de la GFP ont été remplacées par une banque de CDR3 de la chaîne lourde d'un anticorps. Ceci a permis la sélection de mutants spécifiques à différents antigènes avec des affinités de l'ordre du nanomolaire. Ces 'fluorobodies' peuvent remplacer avantageusement les anticorps dans les tests immunologiques traditionnels car la fonction reportrice (la fluorescence) est portée par la même molécule que la fonction de liaison.

## 2.3. Ingénierie de la régulation par évolution dirigée

Dans notre laboratoire, des séquences peptidiques aléatoires ont été insérées dans diverses boucles avoisinant le site actif de la  $\beta$ -lactamase TEM-1 [136]. La résistance aux antibiotiques conférée par la  $\beta$ -lactamase a permis la sélection *in vivo* des clones actifs. Le pourcentage de clones actifs obtenus a également permis de déterminer la tolérance des diverses boucles à l'insertion peptidique. Les clones actifs ont ensuite subi une sélection pour affinité pour un anticorps anti-PSA (antigène spécifique de la prostate). L'activité des enzymes isolées contenant les mimotopes de ces anticorps a été testée en présence d'anticorps. Dans la plupart des cas,

l'activité est fortement inhibée (jusqu'à 95 %) par la liaison de l'anticorps qui provoque un encombrement stérique du site actif. Cependant, un clone réagit à l'anticorps anti-PSA en augmentant de 1,7 fois la vitesse d'hydrolyse sur un substrat et en diminuant la catalyse sur un autre substrat (35 % de l'activité sauvage). Les meilleures constantes de dissociation obtenues pour les complexes anticorps–enzymes étaient de l'ordre du nanomolaire. Lors d'expériences de compétition, la détection du PSA à une concentration de  $10^{-9}$  M a été mesurée. Ces banques d'insertions dans la  $\beta$ -lactamase TEM-1 ont également permis la sélection d'une série d'enzymes hybrides possédant de l'affinité pour des ligands non-anticorps comme la streptavidine, la ferritine et la  $\beta$ -galactosidase [137]. Les constantes de dissociation les plus basses obtenues étaient de l'ordre de  $10^{-7}$  M pour la streptavidine et, après un cycle de maturation d'affinité par PCR mutagénisante, de  $10^{-9}$  M pour la ferritine. Elles n'ont pas été estimées dans le cas des clones sélectionnés sur la  $\beta$ -galactosidase. Cependant, contrairement à ce qui a été observé avec l'anticorps, la complexation par l'antigène des  $\beta$ -lactamases sélectionnées ne provoque qu'un faible effet sur l'activité catalytique, et cela dans les trois cas.

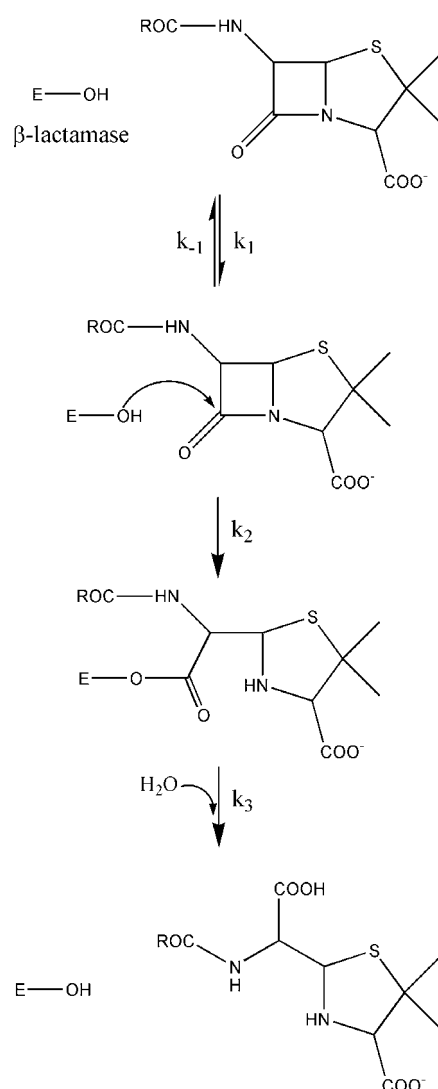
L'approche aléatoire a également été utilisée pour la création de ribozymes allostériques. Un ribozyme ARN ligase dépendant d'un oligonucléotide a été isolé d'une banque aléatoire d'ARN [138]. En présence de l'oligonucléotide effecteur, son activité catalytique est augmentée 10.000 fois. Les ARNs ayant subi la ligation peuvent être amplifiés et quantifiés par PCR inverse. Les ribozymes dépendant de molécules d'ARN répondent à des séquences cibles spécifiques ; la création de nouveaux biochips permettant la détection directe de la présence de certains ARNm dans des échantillons biologiques pourrait donc être envisagée.

### **III. $\beta$ -lactamase TEM-1**

L'enzyme choisie afin d'y introduire un site de régulation est la  $\beta$ -lactamase TEM-1 car elle est relativement petite, monomérique, bien caractérisée structuellement et fonctionnellement, pouvant être exprimée facilement et non toxique pour les cellules procaryotes et eucaryotes [139, 140]. De plus, comme cette enzyme confère une résistance à la pénicilline, il sera aisé de contrôler la qualité des banques de mutants. En effet, seuls les mutants actifs permettront aux bactéries infectées de croître en présence de l'antibiotique.

Les  $\beta$ -lactamases sont des enzymes responsables de nombreux phénomènes de résistance aux antibiotiques de type  $\beta$ -lactame (comme les pénicillines, les céphalosporines,...) car elles hydrolysent le lien amide du noyau  $\beta$ -lactame des antibiotiques [141]. Les  $\beta$ -lactamases sont sécrétées dans le périplasme par les bactéries Gram-négatives ou dans le milieu extérieur par les

Gram-positives. Les gènes codant pour les  $\beta$ -lactamases peuvent se trouver sur le chromosome bactérien ou sur un plasmide [142]. Les  $\beta$ -lactamases sont divisées en quatre classes suivant leur homologie de séquence en acides aminés [143]. Les  $\beta$ -lactamases de classe A, C et D utilisent une sérine dans le site actif [144] tandis que les enzymes de classe B nécessitent la présence d'un ion zinc pour être actives [145]. Le mécanisme d'action des  $\beta$ -lactamases à sérine est décrit à la Figure 8. Dans la première étape, un complexe non covalent est formé entre l'enzyme et le substrat. Dans la deuxième étape, il y a formation d'un acyl-enzyme instable qui sera hydrolysé lors de la troisième étape permettant la régénération de l'enzyme active. Ces enzymes ont un  $k_{\text{cat}}$  élevé et un  $K_M$  faible pour leur substrat naturel, ce qui signifie qu'elles sont très rapides. La  $\beta$ -lactamase TEM-1 a un  $k_{\text{cat}}/K_M$  de  $8.10^8 \text{ M}^{-1} \text{ s}^{-1}$  pour la benzyl-pénicilline.

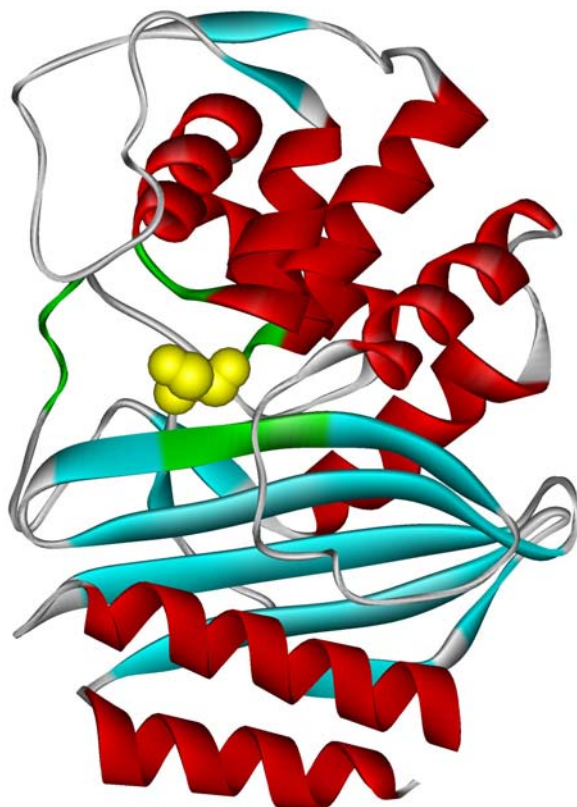


**Figure 8** Mécanisme d'hydrolyse des antibiotiques de type pénicilline par les  $\beta$ -lactamases à sérine.

Les  $\beta$ -lactamases de classe A, dont fait partie la TEM-1, sont présentes aussi bien chez les bactéries Gram-positives que Gram-négatives et sont souvent codées par un plasmide [146]. Leur masse moléculaire est d'environ 30 kDa.

## 1. Structure

La structure de la  $\beta$ -lactamase TEM-1 [147] révèle que l'enzyme est composée de deux domaines (cf. Figure 9). Le domaine  $\alpha/\beta$  est composé de cinq brins  $\beta$  antiparallèles entourés de deux hélices  $\alpha$ , tandis que le second domaine, le domaine  $\alpha$ , est constitué de six hélices  $\alpha$ . Le site actif de l'enzyme se situe à l'interface entre ces deux domaines et comprend la sérine essentielle qui se trouve à l'extrémité amino-terminale de l'hélice  $\alpha_2$ .



**Figure 9** Structure tridimensionnelle de la  $\beta$ -lactamase TEM-1. La sérine active est indiquée en jaune et les quatre éléments conservés, en vert.

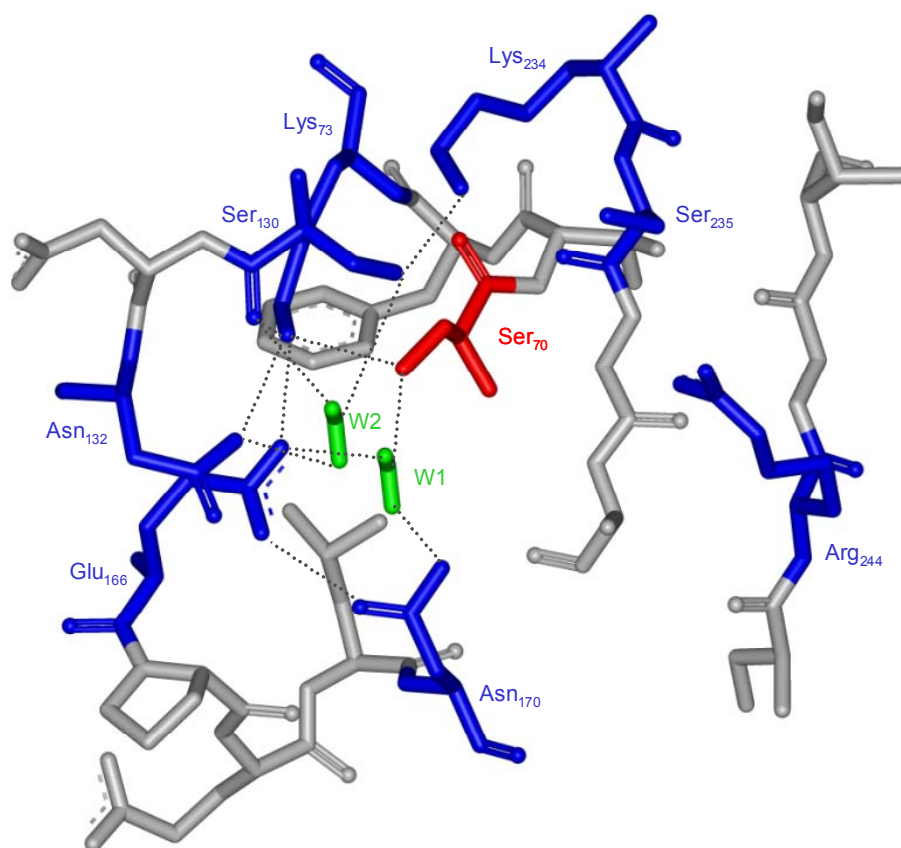
Autour de cette sérine, plusieurs éléments sont impliqués directement ou indirectement dans la reconnaissance du substrat ou dans la catalyse. Ces éléments sont conservés dans les  $\beta$ -lactamases de classe A [148] et sont illustrés à la Figure 10.

Le premier élément conservé (Ser<sub>70</sub>-Xaa-Xaa-Lys<sub>73</sub>) contient la sérine nucléophile et, après un tour d'hélice, une lysine dont la chaîne latérale pointe vers le site actif. La sérine 70 peut former un pont hydrogène avec la lysine 73, ce qui suggère une implication de la lysine dans la catalyse.

Le deuxième élément (Ser<sub>130</sub>-Xaa-Asn<sub>132</sub>) est situé sur la boucle connectant les hélices  $\alpha 3$  et  $\alpha 4$ , dans le domaine  $\alpha$ , et constitue un côté de la cavité catalytique. Les chaînes latérales de la sérine et de l'asparagine pointent vers le site actif et sont impliquées dans la formation de ponts hydrogène avec les lysines 73 et 234. Dans certaines  $\beta$ -lactamases, l'asparagine est remplacée par une sérine ou une glycine ; tandis que la sérine est remplacée par une tyrosine dans les  $\beta$ -lactamases de classe C et D.

Le troisième élément, Lys<sub>234</sub>-(Thr ou Ser)-Gly<sub>236</sub>, se trouve sur le brin  $\beta 3$  et constitue le côté opposé de la cavité catalytique. La chaîne latérale de la lysine forme un pont hydrogène avec le groupe hydroxyle de la sérine 130. Dans quelques  $\beta$ -lactamases, une arginine ou une histidine substitue la lysine.

Le quatrième élément est situé sur une boucle de 16 à 19 résidus (Arg<sub>161</sub>  $\rightarrow$  Asn<sub>179</sub> dans la  $\beta$ -lactamase TEM-1). Cette boucle, appelée boucle  $\Omega$ , est relativement mobile et se trouve à l'entrée du site catalytique. Dans la plupart des cas, elle contient la séquence Glu<sub>166</sub>-Xaa-Glu-Leu-Asn<sub>170</sub>. Le glutamate 166 et l'asparagine 170 sont liés par un pont hydrogène et semblent essentiels pour le positionnement de la molécule d'eau W1 près de la sérine catalytique [149].



**Figure 10** Structure du site actif de la  $\beta$ -lactamase TEM-1. Les résidus impliqués dans la catalyse sont indiqués en bleu, la sérine 70 en rouge et les deux molécules d'eau (W1 et W2) en vert. Le réseau de ponts hydrogène est indiqué en pointillé.

## 2. Mécanisme catalytique

Lors de l'hydrolyse du lien amide du noyau  $\beta$ -lactame, un intermédiaire de type acyl-enzyme est formé. Le groupe hydroxyle de la sérine 70 doit donc être activé par augmentation de son caractère nucléophile. Deux résidus ont été proposés pour jouer ce rôle de base et, à ce jour, le mécanisme catalytique des  $\beta$ -lactamases de classe A reste controversé.

Selon Lamotte-Brasseur *et al.* [149], le rôle de base est joué par le glutamate 166. En effet, des expériences de mutagenèse dirigée suggèrent que le glutamate joue le rôle de base dans l'étape d'acylation [150]. Cependant, il semble trop éloigné de la sérine 70 pour agir de manière directe. Des données cristallographiques ont mis en évidence la présence d'une molécule d'eau (W1) entre la sérine 70 et le glutamate 166 et des études de modélisation moléculaire ont suggéré que cette molécule d'eau puisse servir de relais dans le transfert du proton entre ces deux résidus [149].

Une autre explication possible repose sur la flexibilité de la boucle  $\Omega$  qui permettrait le rapprochement du glutamate et de la sérine pendant l'acylation [151]. Ensuite, l'oxygène  $\gamma$  activé de la sérine 70 attaquerait le carbone de la fonction carbonyle du  $\beta$ -lactame, qui prend une configuration tétraédrique. Suite à la rupture de la liaison amide et au transfert du proton du glutamate 166 vers l'azote du  $\beta$ -lactame (via une deuxième molécule d'eau, la lysine 73 et la sérine 130), il y aurait formation de l'acyl-enzyme. La désacylation (hydrolyse de l'acyl-enzyme) serait également catalysée par le glutamate 166, celui-ci activant la première molécule d'eau pour attaquer le carbonyle de l'acyl-enzyme. Finalement, une troisième molécule d'eau entrerait dans le site catalytique, le glutamate 166 assurant le retour d'un proton sur l'oxygène de la sérine 70.

La seconde hypothèse est basée sur un mécanisme asymétrique dans lequel l'acylation et la désacylation ne seraient pas catalysées par le même résidu. Selon Strynadka *et al.* [152], le groupement  $\epsilon$ -amino de la lysine 73 jouerait le rôle principal dans l'étape d'acylation impliquant le transfert du proton de la lysine sur l'azote du  $\beta$ -lactame via un réseau de ponts hydrogène. Cette hypothèse est basée sur les propriétés du mutant Glu<sub>166</sub>Asn de la  $\beta$ -lactamase TEM-1 qui forme assez rapidement l'intermédiaire acyl-enzyme [153]. Par contre, ce mutant désacyle très lentement. Le glutamate 166 ne s'avèrerait donc pas indispensable à l'étape d'acylation. En ce qui concerne l'étape d'hydrolyse de l'acyl-enzyme, le rôle de base serait joué par le glutamate 166, comme dans le premier modèle.

Des études cinétiques et de modélisation ont été réalisées avec des substrats plus encombrants qui déplacent la molécule d'eau W1 [154]. Si le glutamate 166 est impliqué dans l'acylation et la désacylation, ces deux étapes devraient être défavorisées. Tandis que, dans l'hypothèse d'un

mécanisme asymétrique, l'acylation ne devrait pas être perturbée car les substrats utilisés ne gênent pas le pont hydrogène entre la sérine 70 et la lysine 73. Or il apparaît que ces substrats sont acylés très lentement, ce qui plaide en faveur d'un mécanisme symétrique. Le glutamate agit donc comme base lors de l'acylation et de la désacylation, la lysine 73 jouant le rôle de base alternative lors de l'acylation dans le cas du mutant Glu<sub>166</sub>Asn.

### 3. Tolérance à la mutagenèse

Des études sur la tolérance de la  $\beta$ -lactamase TEM-1 à la mutagenèse ont été réalisées. Huang *et al.* [155] ont déterminé la tolérance de chaque résidu à la substitution. La stratégie consistait à rendre aléatoire trois codons contigus pour créer une banque aléatoire contenant toutes les substitutions en acide aminé possibles pour la région mutagenisée. Les mutants fonctionnels ont ensuite été sélectionnés *in vivo* et séquencés pour déterminer l'identité des substitutions permises à chaque position. De cette façon, les auteurs ont déterminé que 43 résidus sur les 263 composants l'enzyme sont essentiels à la fonction de la  $\beta$ -lactamase TEM-1 (cf. Tableau 1). Parmi ces résidus, certains sont directement impliqués dans la catalyse ou interviennent dans des interactions importantes dans le site actif. D'autres ne sont pas dans la cavité enzymatique mais influencent néanmoins la liaison du substrat et la catalyse. Les résidus essentiels qui ne sont pas dans le voisinage du site actif sont éparpillés dans la structure de la protéine. Certains d'entre eux se trouvent dans des positions enfouies dans les régions hydrophobes des deux domaines ou à l'interface entre ces deux domaines. D'autres résidus essentiels forment des interactions ioniques ou des ponts hydrogène indispensables à la structure de l'enzyme.

|        |         |         |         |         |         |         |         |
|--------|---------|---------|---------|---------|---------|---------|---------|
| Leu 30 | Lys 73  | Asp 131 | Asp 179 | Ala 217 | Asp 233 | Gly 245 | Trp 290 |
| Tyr 46 | Leu 76  | Asn 132 | Thr 180 | Leu 220 | Lys 234 | Leu 250 |         |
| Phe 66 | Pro 107 | Ala 134 | Thr 181 | Arg 222 | Ser 235 | Gly 251 |         |
| Pro 67 | Leu 122 | Arg 164 | Trp 210 | Pro 226 | Gly 236 | Met 272 |         |
| Ser 70 | Ala 125 | Glu 166 | Asp 214 | Trp 229 | Gly 242 | Asn 276 |         |
| Thr 71 | Ser 130 | Asp 176 | Val 216 | Ala 232 | Arg 244 | Ile 282 |         |

**Tableau 1** Résidus de la  $\beta$ -lactamase TEM-1 qui ne tolèrent pas de substitutions (reproduit de Huang *et al.* [155]).

Il reste donc 220 résidus qui tolèrent quelques substitutions tout en gardant la fonctionnalité de l'enzyme.

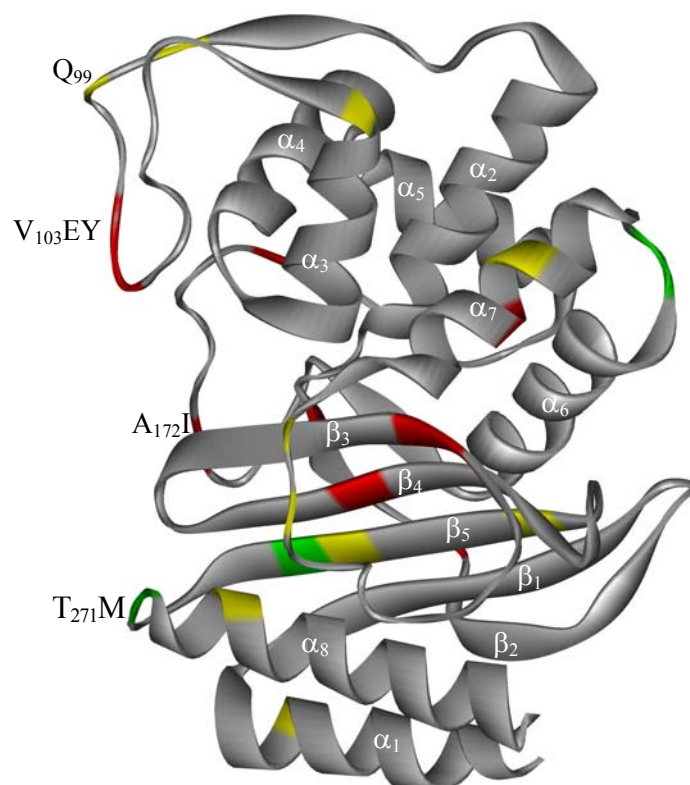
Des études d'insertion aléatoire ont également été réalisées sur la  $\beta$ -lactamase TEM-1 [156]. La stratégie utilisée est basée sur l'insertion aléatoire du transposon TN4430. Après la délétion *in vitro* du transposon, une insertion de 15 bp reste dans le gène cible. De cette manière, 23 insertions d'un pentapeptide ont été créées dans l'enzyme, à des positions différentes. L'effet de

ces insertions sur l'activité catalytique a été testé en déterminant la concentration maximale en ampicilline permettant la croissance bactérienne conférée par ces protéines. Les mutants ont été divisés en trois catégories : cinq mutants présentent de hautes résistances à l'ampicilline (supérieure à la moitié de la résistance conférée par l'enzyme sauvage) ; huit des insertions pentapeptidiques ne possèdent pas de résistance détectable ; et les dix insertions restantes montrent des résistances intermédiaires (cf. Figure 11). Parmi les insertions conférant une haute résistance à l'ampicilline, trois sont situées dans la boucle entre le brin  $\beta 5$  et l'hélice  $\alpha 8$ . Etonnement, cette boucle contient le résidu méthionine 272 qui ne tolère pas la mutagenèse (cf. Tableau 1). Les deux autres insertions tolérées sont situées dans la boucle entre les hélices  $\alpha 6$  et  $\alpha 7$ , sur la face de l'enzyme opposée au site actif. Les dix insertions moyennement tolérées se trouvent soit dans les structures secondaires de l'enzyme (hélice  $\alpha$  ou brin  $\beta$ ), soit dans la grande boucle située entre les hélices  $\alpha 2$  et  $\alpha 3$ , soit sur la boucle entre l'hélice  $\alpha 7$  et le brin  $\beta 3$  (opposé au site actif). Finalement, les insertions non tolérées sont situées soit dans les structures secondaires de l'enzyme, soit sur la partie N-terminale de la boucle  $\Omega$ .

Comme nous l'avons vu précédemment, des peptides aléatoires ont été insérés dans la  $\beta$ -lactamase TEM-1 au laboratoire [136]. Les insertions de séquences dégénérées ont été réalisées dans trois boucles de l'enzyme (cf. Figure 11) : la première banque (lib 1) a été créée dans une boucle située sur le bord du site actif (Val<sub>103</sub>-Tyr<sub>105</sub>), entre les hélices  $\alpha 2$  et  $\alpha 3$  ; la deuxième (lib 2), dans la partie C-terminale de la boucle  $\Omega$  (entre Ala<sub>172</sub> et Ile<sub>173</sub>) ; et la troisième (lib 3), dans une boucle qui se trouve un peu plus éloigné du site actif, en amont de la dernière hélice (Thr<sub>271</sub>). La tolérance de ces sites à l'insertion a été déterminée en calculant le pourcentage de clones conférant aux bactéries infectées une résistance supérieure à 2,5 % de la résistance d'origine. Les banques lib 1 et lib 2 n'ont qu'une très faible activité vis-à-vis d'un substrat naturel de l'enzyme, la pénicilline G, (< 0,1 % de clones actifs), tandis que lib 3 a une activité catalytique importante (7 % de clones actifs).

Les résultats obtenus pour les banques lib 2 et lib 3 sont en accord avec les résultats de la pentapeptidogenèse [156]. En ce qui concerne l'insertion entre les hélices  $\alpha 2$  et  $\alpha 3$  (lib 1), nous constatons qu'elle n'est pas tolérée au niveau de Val<sub>103</sub>-Tyr<sub>105</sub> tandis qu'elle est moyennement tolérée du côté N-terminal (Gln<sub>99</sub>).





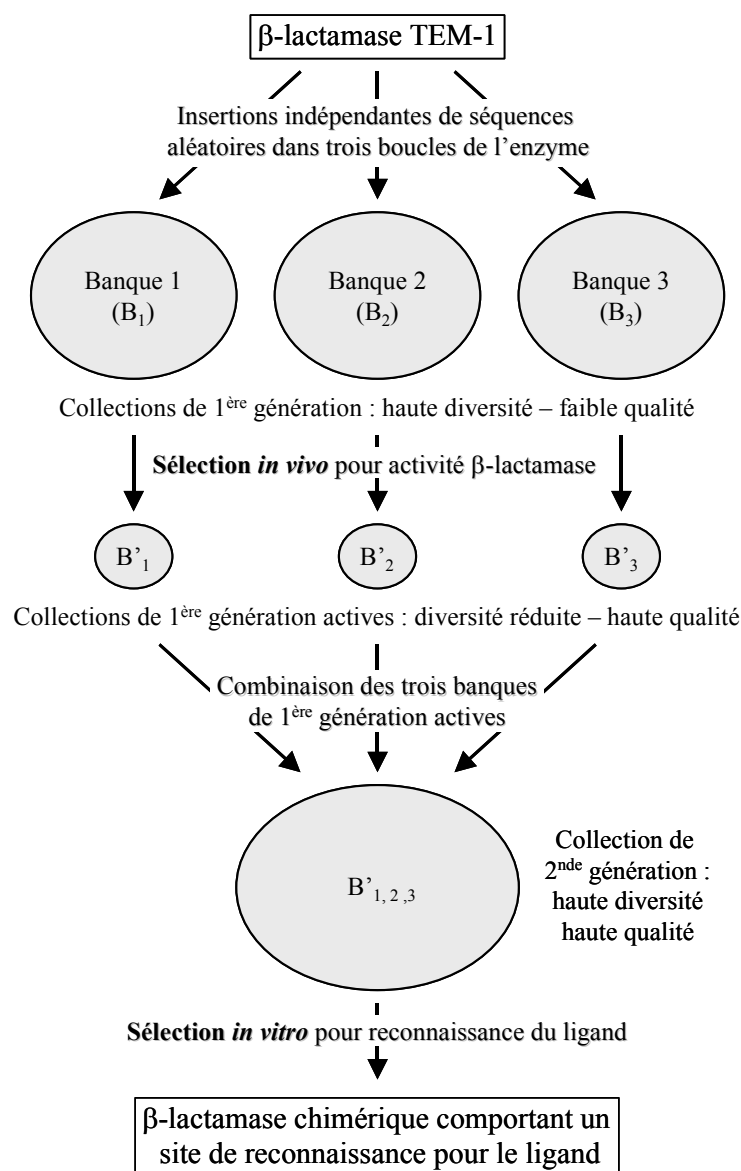
**Figure 11** Sites des insertions pentapeptidiques [156] et des insertions aléatoires [136] dans la  $\beta$ -lactamase TEM-1. Les résidus en vert, jaune et rouge correspondent respectivement aux sites où l'insertion est tolérée, moyennement tolérée et non tolérée.

## IV. Objectifs et stratégie générale

L'objectif de cette thèse est de créer un répertoire de  $\beta$ -lactamases TEM-1 chimériques présentant chacune un site de complexation susceptible de reconnaître un ligand cible et plus particulièrement de petites molécules organiques ou des ions métalliques. Alors que la reconnaissance de protéines semble pouvoir être réalisée par une protubérance sur l'enzyme constituée d'une séquence peptidique linéaire [137], notre hypothèse de travail est que la reconnaissance de petites molécules nécessite la construction d'une cavité complexe. Celle-ci sera construite à proximité du site actif de l'enzyme, en espérant que la liaison du ligand choisi affecte l'activité catalytique. De telles enzymes pourraient être utilisées pour la détection, voire même le dosage du ligand.

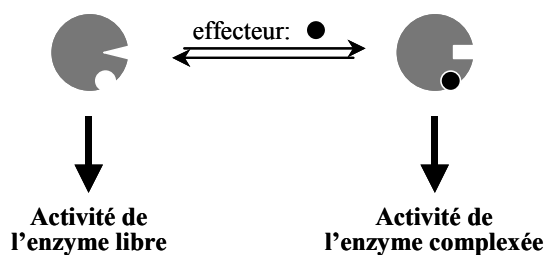
La stratégie envisagée consiste à insérer des peptides dans trois boucles connectant des structures secondaires de l'enzyme et proches l'une de l'autre dans l'espace. Notre hypothèse est que ces trois séquences pourraient constituer une cavité susceptible de complexer une grande variété de molécules. Les enzymes hybrides créées partageront en effet certaines similitudes avec les anticorps de camélidés qui ne comportent qu'une chaîne lourde (cf. Figure 7, page 34) et dont la cavité de complexation est construit à partir de trois CDRs seulement [157].

Comme nous ne connaissons pas les séquences susceptibles de complexer un ligand donné, l'approche utilisée consiste à créer une banque de mutants en insérant des peptides aléatoires dans trois boucles de l'enzyme. Cette collection de mutants sera construite de manière hiérarchique selon la séquence suivante (cf. Figure 12) : (1) les banques d'insertion seront construites indépendamment, par clonage de collections de fragments d'ADN obtenus par PCR à l'aide d'oligonucléotides dégénérés ; (2) chaque banque subira ensuite une sélection *in vivo* pour activité catalytique, afin de ne garder que les clones possédant une structure tertiaire ; (3) les trois banques d'insertion seront finalement combinées pour donner une banque de seconde génération contenant les trois insertions aléatoires. La sélection pour activité nous permettra d'obtenir des banques de qualité contrôlée, ainsi que des informations sur la tolérance des boucles à l'insertion peptidique. De plus, nous supposons que la combinaison de collections de haute qualité devrait fournir une banque de seconde génération de bonne qualité également. Cette stratégie hiérarchique permet également de construire aisément une banque de grande taille. En effet, si chaque collection intermédiaire n'est composée que de 1.000 clones actifs, leur combinaison devrait permettre l'obtention d'une banque de  $10^9$  clones.



**Figure 12** Schéma illustrant la stratégie de construction hiérarchique des collections de  $\beta$ -lactamases chimériques pour la sélection de clones ayant acquis de l'affinité vis-à-vis d'un ligand cible.

Ensuite, la technique de l'expression en surface de phage permettra de sélectionner, par capture d'affinité, les enzymes mutantes ayant acquis la capacité de se lier à des molécules cibles. Cette étape sera réalisée à l'aide de versions immobilisées des cibles choisies. Parmi ces clones, nous espérons en trouver dont l'activité est modulée suite à la complexation de cette molécule à proximité du site actif (cf. Figure 13). Idéalement, nous souhaiterions que ce soit la constante catalytique ( $k_{cat}$ ) plutôt que l'affinité apparente de l'enzyme pour le substrat ( $K_M$ ) qui soit affectée. En effet, ceci permettrait de détecter la présence du ligand modulateur d'activité par des mesures en cinétique initiale.



**Figure 13** Schéma représentant la régulation de l'activité d'une enzyme par la complexation d'un ligand.

A notre connaissance, un tel travail d'ingénierie d'une enzyme régulée par une petite molécule n'a jamais été réalisé et est donc tout à fait innovant.

Le travail présenté ici est divisé en deux parties. Le premier chapitre est consacré à la construction et à la caractérisation des banques d'insertion dans la  $\beta$ -lactamase TEM-1. Cette partie du travail permettra de déterminer la tolérance à l'insertion en fonction des caractéristiques des sites choisis et d'évaluer les possibilités de créer des banques d'enzymes qui gardent leur activité malgré la présence de multiples insertions. Le second chapitre décrit les processus de sélection des clones ayant acquis une affinité pour les ligands choisis et leur criblage pour tenter d'y repérer les clones dont l'activité catalytique est modulée par complexation avec le ligand choisi.