

Avant-propos	1
Introduction	5
Principes d'évolution naturelle des protéines	7
Définitions	7
L'évolution de nouvelles fonctions protéiques	8
Les réarrangements majeurs responsables de la diversité des protéines	9
Evolution divergente et fonction enzymatique	12
La plasticité des sites actifs	15
Evolution convergente ou divergente ?	18
Comment reconnaître des protéines homologues ?	19
L'évolution divergente des protéines, concept de diversité conformationnelle	21
L'ingénierie des protéines : approches stratégiques et applications, création de banques d'enzymes, criblage et sélections	25
Ingénierie des protéines : Approches stratégiques et applications	25
Design rationnel ou évolution dirigée ?	25
Choix d'une protéine matrice-Concept "d'évolvabilité"	32
Création de banques d'enzymes : comment générer de la diversité	34
Mutagenèse dirigée par oligonucléotide	34
PCR mutagénisante	37
Recombinaison homologue d'ADN in vitro	38
Recombinaison non homologue d'ADN in vitro	40
Insertions et délétions de séquences	41
Lien génotype – protéine	43
Liaison par compartimentalisation	43
"Phage display"	45
Méthodes basées sur des systèmes d'IVE	51
Criblage et sélections	54
Criblage	54
Sélection	55
Les DD-peptidases et les β-lactamases	63
Les DD-peptidases	63
Le peptidoglycane	63
Les PBPs	65
Les β -lactamases	69
Origines des résistances bactériennes aux β -lactames	69
Mécanisme catalytique général	72
Les différentes classes de β -lactamases	74
Bases de la phylogénie entre les PBPs et les β -lactamases	81

Comment une PBP a pu devenir une β -lactamase ?	84
Evolution d'une PBP en β -lactamase	85
Objectifs et Stratégie	89
Bibliographie	95
Résultats	105
Identification d'une nouvelle famille de PBPs chez des cyanobactéries, une famille de protéines homologues aux β -lactamases de classe A :	107
Abstract	111
Introduction	113
Results	117
Comparison of PBP-A with β -lactamases.	117
Cloning and expression of PBP-A protein and mutants.	119
β -lactamase activity of purified PBP-A.	122
PBP-A-D ₁₆₀ A PBP-A-S ₆₁ A and PBP-A- Δ mutants.	124
PBP-A-L ₁₅₈ D and PBP-A-L ₁₅₈ E mutants.	125
Discussion	129
Physiological role of PBP-A.	129
Place of PBP-A in the PRPs	130
Materials and methods	133
Annexes	139
Bibliography	145
PBP-A de <i>Thermosynechococcus elongatus</i> , une PBP proche des β -lactamases de classe A. Analyse structurale de l'enzyme sauvage et d'un mutant désacylant plus rapidement	149
Abstract	153
Introduction	155
Results	159
Structure	159
Overall structure.	159
Active site	160
Structure of covalent adduct	163
Comparison to ASPRE proteins: TEM-1, Per-1 and PBP-5	165
Mutants	172

Discussion	175
Interaction of PBP-A with peptide substrates	175
Interaction of PBP-A with penicillins-Proposition for a catalytic mechanism	176
Improving PBP-A deacylation	178
Phylogenetic link	179
Materials and methods	181
Bibliography	185
Tentatives d'évolution de PBP-A en β -lactamase. Différentes stratégies associant mutagenèse et sélections	191
Abstract	195
Introduction	197
Results	201
Display of PBP-A and PBP-A-L ₁₅₈ E enzymes on phage	201
Libraries design and construction.	203
In vivo selections.	204
In vitro selections-Model experiment	205
Mutants expression, purification and kinetics.	216
Discussion	219
Materials and Methods	225
Complementary study: selection for phages acylating 100 to 1000 times faster	231
Introduction	231
Results	232
Discussion	238
Material and methods	239
Annexes	241
Bibliography	249
Conclusions et Perspectives	253
Conclusions	255
Approche rationnelle	255
Etudes réalisées sur base des alignements	255
Etudes réalisées sur base de la structure	256
Approche par évolution dirigée	258
Sélection in vivo	258
Sélection pour activité catalytique	258
Sélection pour acylation	259

Perspectives	261
Perspectives à court terme	261
Perspectives à long terme	262
Bibliography	267

Avant-propos

« Les espèces qui survivent ne sont pas les espèces les plus fortes, ni les plus intelligentes, mais celles qui s'adaptent le mieux aux changements »

Charles Darwin.

La théorie de l'évolution de Charles Darwin au XIX^{ème} siècle, établit que la plupart des individus d'une population sont différents les uns des autres. Certains d'entre eux sont mieux adaptés à leur environnement que les autres et ont, de ce fait, de meilleures chances de survivre et de se reproduire. Ces caractéristiques avantageuses sont héritées par les générations suivantes et, avec le temps, deviennent dominantes dans la population. Ce processus progressif et continu dont la force majeure est la sélection naturelle, résulte en l'apparition de variations phénotypiques et ultimement, de nouvelles espèces. Sur base de cette théorie est née celle du néodarwinisme qui enrichit le concept original de Darwin en attirant l'attention sur la façon dont la variation biologique est générée. Dans ce cas est exclu l'héritage génétique comme un mécanisme d'évolution, et la reproduction sexuelle (recombinaison) est proposée comme le mécanisme permettant de créer, à chaque génération, une population d'individus différents. Enfin, la théorie synthétique incorpore aux théories précédentes les mécanismes liés aux organismes procaryotes pour qui la reproduction sexuelle est remplacée par une reproduction asexuelle par fission binaire, et la recombinaison se fait par des mécanismes de transferts horizontaux (conjugaison, transformation...). Aussi, les transitions dans l'évolution se font de manière graduelle et les nouvelles espèces évoluent à partir d'espèces préexistantes par un processus lent et se maintiennent à chaque étape de leur adaptation spécifique [1].

Cette évolution de la théorie de Darwin est liée aux découvertes successives, réalisées en parallèle, de l'existence d'un lien entre un gène et une protéine, de l'ADN comme support de l'information génétique, de la génétique bactérienne et de la structure de l'ADN (Watson et Crick, 1953), qui ont permis l'émergence de la biologie moléculaire.

Linus Pauling est à l'origine de l'utilisation des séquences de protéines pour la reconstruction d'arbres phylogéniques, et il propose avec Emile Zuckerkandl, en 1965, le concept d'horloge moléculaire. Des comparaisons de séquences d'acides aminés conduisent à la constatation que

les mesures de la divergence des séquences sont corrélées au temps et de la même manière que l'on peut retracer la phylogénie des espèces vivantes, on peut retracer des chemins évolutifs de protéines. Ces observations à l'échelle moléculaire montrent que la variabilité des protéines serait apparue selon un processus Darwinien d'évolution.

Les protéines remplissent des fonctions très diverses de catalyse, de transport, de communication, de signalisation, de reconnaissance et de structure, c'est pourquoi elles peuvent être considérées comme les outils majeurs du monde vivant à l'échelle moléculaire. Au cours de l'évolution, les accumulations de mutations ont fait diverger les gènes entre les espèces. De là provient la diversité des protéines codées par ces gènes.

De par l'importance des protéines dans le monde vivant, elles occupent une place importante dans le domaine industriel, agricole, médical ou biotechnologique, ou dans le domaine académique, pour la compréhension des liens existants entre la structure et la fonction des protéines ou la compréhension des mécanismes évolutifs qui ont permis la spécialisation et l'optimisation de leurs fonctions. Ainsi, les techniques visant à changer les propriétés des protéines, principalement basées sur le principe Darwinien de mutation et de sélection ont été grandement développées.

Le cadre de cette thèse s'inscrit dans la compréhension des liens existants entre deux familles de protéines, les PBPs (Penicillin Binding Proteins) et les β -lactamases, en essayant de recréer artificiellement une activité de type β -lactamase à partir d'une PBP.

Dans cette introduction, seront présentés dans un premier temps quelques principes généraux de l'évolution naturelle des protéines. Un chapitre concernant l'ingénierie des protéines nous permettra d'avoir une meilleure vue des applications et stratégies élaborées pour la création d'enzymes ayant une propriété recherchée, des technologies permettant de créer un ou plusieurs mutants d'une même protéine (banques ou collections), et enfin des techniques qui vont permettre de sélectionner un mutant ou cribler une collection de mutants. Dans la dernière partie, les PBPs et les β -lactamases seront présentées, ainsi que les liens phylogénétiques existant entre elles.

Introduction

Principes d'évolution naturelle des protéines

Les protéines sont des éléments primordiaux des êtres vivants car elles sont impliquées dans pratiquement tous leurs aspects fonctionnels et structuraux. Afin d'accomplir leurs différentes fonctions au sein d'un être vivant, les protéines constituent une large gamme de molécules spécifiques et efficaces. Leur diversité est issue de leur versatilité structurale et fonctionnelle, réalisée à travers la diversité des séquences obtenues par la polymérisation d'acides aminés distincts.

Cependant, 150 génomes comptent un million de séquences protéiques distinctes, dont 85% sont classées en 50 000 familles. Chacune de ces familles comporte la même composition en domaines structuraux. 80 % des séquences de ces domaines sont réparties en seulement 5000 types de domaines [2]. Dans ce chapitre, nous allons tenter d'expliquer brièvement comment le large répertoire d'enzymes existant a pu être créé à partir d'un nombre limité de domaines structuraux, et les implications pour les relations entre les protéines d'une même famille ainsi que les différents modèles d'évolution possibles.

Définitions

Pour plus de clarté, quelques définitions de termes dont la signification peut varier suivant le contexte et d'un auteur à l'autre, sont données.

Similarité : la similarité est généralement utilisée pour donner une vue quantitative sur le degré de ressemblance entre deux protéines. Pour quantifier et qualifier cette similarité, un score est généralement calculé. Trois types d'opérateurs de transformation de séquence peuvent être différenciés :

- la substitution : un acide aminé est remplacé par un autre.
- l'insertion : un ou des acides aminés sont insérés dans une séquence.
- la délétion : un ou des acides aminés sont éliminés d'une séquence.

On regroupe les insertions et délétions sous le terme générique d'indels ou encore gaps. Un score peut être calculé pour chacun des alignements possibles entre les deux séquences en utilisant un coût déterminé pour chaque événement évolutif. Les coûts de substitution sont variables en fonction de la fréquence d'observation des mutations d'un acide aminé donné vers un autre.

Homologie : des enzymes homologues dérivent d'un ancêtre commun et sont structurellement semblables.

Domaine : Un domaine structural est l'unité la plus petite qui possède les caractéristiques que l'on peut associer à une protéine : il est compact, il a un cœur hydrophobe, il est capable de se replier

indépendamment du reste de la protéine et peut parfois fonctionner indépendamment du reste de la structure.

Famille : une famille d'enzymes est un groupe d'enzymes homologues qui catalysent la même réaction.

Superfamille : d'un point de vue structural, une superfamille est composée de protéines qui partagent le même repliement tridimensionnel et de faibles identités de séquences, mais dont la structure et dans de nombreux cas, la fonction, suggèrent qu'une origine évolutive commune est probable. Gerlt & Babbitt donnent une dimension fonctionnelle à la superfamille et la définissent comme un groupe d'enzymes homologues qui catalysent i) la même réaction avec des spécificités de substrat différentes, ii) des réactions différentes qui partagent un même attribut mécanistique, permis par la conservation de résidus du site actif qui réalisent les mêmes fonctions dans tous les membres de la superfamille [3].

L'évolution de nouvelles fonctions protéiques

L'évolution du répertoire de protéines a requis l'invention *ab initio* de nouvelles protéines. Un des mécanismes possibles pour l'invention de nouvelles protéines est l'arrangement *de novo* de petits peptides structurés. Un tel mécanisme peut être représenté dans la nature par la famille des glycoprotéines antigel, composées en partie de la répétition d'une unité fondamentale constituée d'un tripeptide (Thr-Ala/Pro-Ala)_n. Les structures polymériques constituées de la répétition du tripeptide auraient évolué à partir de l'amplification de fragments de 9 nucléotides à partir d'un élément codant pour le trypsinogène [4]. Cependant, ce type de création est inefficace et des créations d'enzymes *ab initio* utilisant des mécanismes de ce type sont peut être encore possibles, mais il est clair que les mécanismes dominants pour l'expansion de l'éventail protéique sont la duplication de gènes, la divergence et la recombinaison.

Deux causes possibles pour la création d'enzymes par ces mécanismes sont : i) un ensemble de domaines dont les fonctions sont assez variées pour supporter une forme de vie basique a été créé, il était alors plus rapide de produire de nouvelles protéines avec différentes fonctions par duplication, divergence et recombinaison, ii) une fois que les procédures de correction d'erreurs lors de la réplication de l'ADN et de la synthèse des protéines ont été développés, le processus d'invention *ab initio* des protéines est devenu trop difficile pour être efficace. En conséquence, un petit nombre de protéines a été créé au départ, et même les génomes les plus simples sont le produit de duplication et recombinaison extensives [5].

Le domaine comme unité de base

La manière la plus efficace d'évoluer de nouvelles fonctions et propriétés est le remodelage d'une protéine existante. Les protéines sont constituées de domaines d'une taille variant généralement de 100 à 250 résidus qui sont des unités d'évolution dont la séquence peut être

dupliquée et/ou subir des phénomènes de recombinaison. Alors que les petites protéines sont constituées d'un seul domaine, les grosses sont constituées d'une combinaison de domaines, de cette manière, 2/3 des protéines procaryotes sont constituées de 2 domaines ou plus. Une famille de domaines est alors composée de petites protéines, et/ou de parties des plus grosses, issues d'un ancêtre commun. Les mécanismes dominants qui produisent l'augmentation du large éventail protéique sont : i) la duplication de séquences qui codent pour un ou plusieurs domaines, ii) la divergence de séquences dupliquées par des mutations, des délétions, et des insertions, pour produire des structures modifiées qui ont de nouvelles propriétés adaptatives et peuvent être sélectionnées, et dans certains cas, iii) la recombinaison de gènes qui résulte en un nouvel arrangement de domaines [5]. D'autres mécanismes existent et sont détaillés dans le paragraphe suivant.

Ces phénomènes donnent naissance à des familles de protéines dont les relations entre les membres les plus distants ne peuvent être détectées par la comparaison directe des séquences et seule la résolution de la structure tridimensionnelle permet de déterminer leurs liens de parenté [2].

Les réarrangements majeurs responsables de la diversité des protéines

Des changements mineurs pour explorer l'espace de séquence local par mutagenèse de positions individuelles sur une structure établie ont prouvé leur efficacité en permettant une adaptation aux changements environnementaux et en optimisant des fonctions existantes (Figure 1-a). Ainsi, deux protéines différant de quelques acides aminés pourront par exemple avoir des propriétés de thermostabilité différentes, la pression de sélection environnementale (milieu chaud) ayant permis à l'une de fixer des mutations lui permettant de s'adapter à ce milieu particulier (voir page 26). Cependant, le large éventail de structures protéiques est issu de réarrangements majeurs par duplication de gènes, permutation circulaire, fusion, délétion et insertion de grands fragments, créant ainsi de nouvelles combinaisons structurales avec de nouvelles fonctions potentielles.

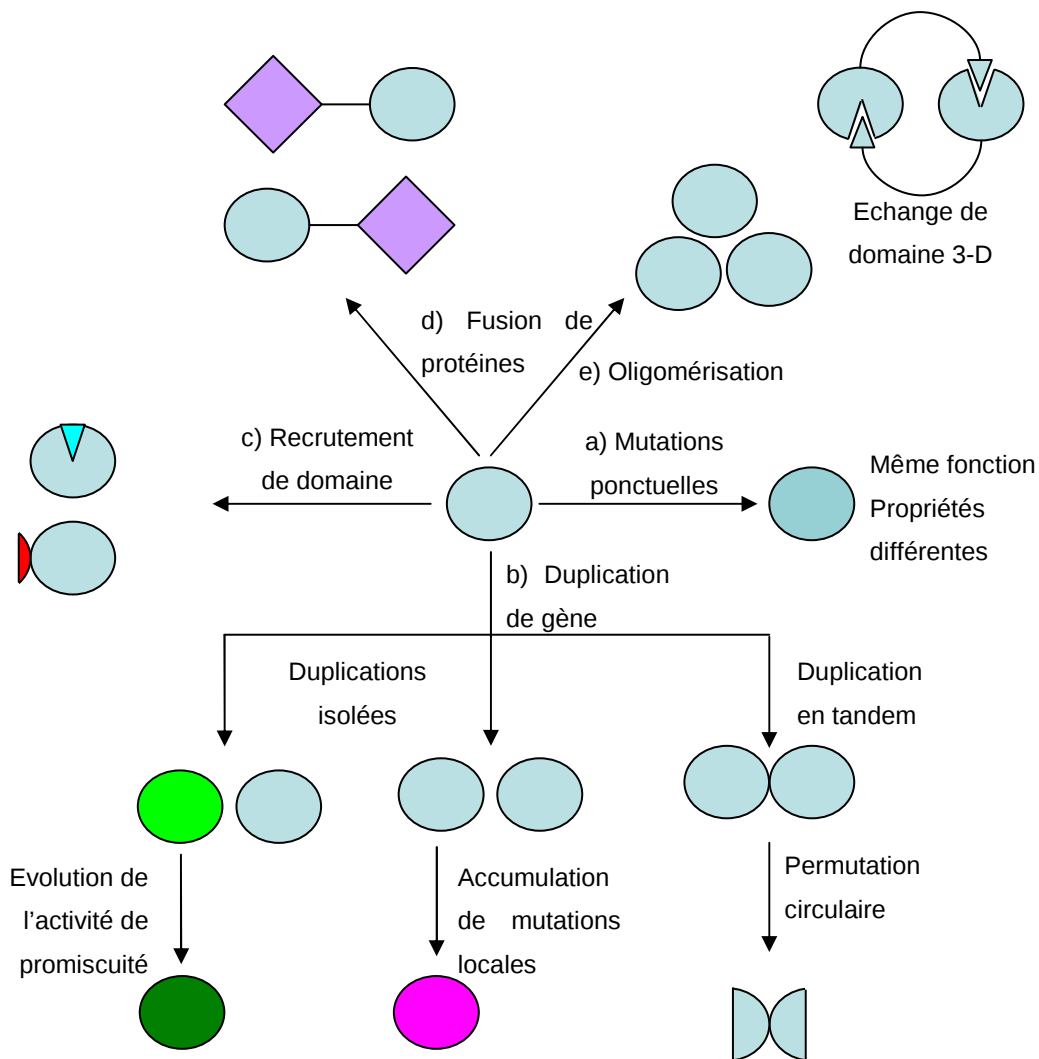


Figure 1 : origine de la diversité des protéines : les différents mécanismes d'évolution divergente d'une protéine. Les couleurs différentes signifient une fonction différente. Schéma adapté de [6].

La réorganisation, l'optimisation et le développement d'une nouvelle fonction par duplication du matériel génétique est un des mécanismes fondamentaux de l'évolution des protéines. Mécanistiquement, la duplication de gènes peut être réalisée durant la réplication de l'ADN et les phénomènes de recombinaison durant la méiose ou lors de transferts horizontaux de matériel génétique. En fonction du mécanisme, 2 copies en tandem ou isolées de la séquence parentale sont générées (Figure 1, b). Dans les deux cas, trois schémas sont envisageables : i) la forme dupliquée est inactivée par accumulation de mutations, ii) l'absence de contraintes fonctionnelles facilite l'accumulation de mutations locales et conduit à une différenciation fonctionnelle, iii) les produits des deux gènes maintiennent l'activité sauvage. La seconde issue mène à l'évolution d'une nouvelle fonction, et est illustrée par ce qui est réalisé en évolution dirigée. On recherche généralement une nouvelle fonction dans une protéine en la soumettant à une certaine pression de sélection, sans chercher à garder intacte son activité

naturelle (voir page 30). Dans ce modèle, une période de bifonctionnalité d'un même gène peut précéder l'événement de duplication où une copie maintient la forme ancestrale et l'autre adopte la nouvelle fonction. Ceci mène au concept de promiscuité, important en évolution des protéines. On parle de promiscuité lorsque deux activités catalysées par le même site actif, l'une principale et l'autre secondaire, coexistent. Dans un contexte évolutif, l'efficacité de la réaction secondaire est plus faible par rapport à la réaction principale, mais assez élevée pour fournir un avantage sélectif. Par rapport aux enzymes actuelles, l'activité de promiscuité pourrait être le vestige de la fonction à partir de laquelle est issue une nouvelle enzyme ou son plus proche homologue. L'hypothèse posant que les enzymes possédant un spectre large de spécificité ou une promiscuité dans leurs activités fournissent ou ont fourni de bons points de départ pour l'évolution de nouvelles fonctions a été formulée par Jensen. Il propose que, contrairement aux enzymes modernes, les enzymes primitives possèdent des spectres de spécificité très larges. La versatilité catalytique a permis à quelques enzymes d'atteindre une multitude de fonctions qui étaient nécessaires pour la survie des organismes ancestraux [7]. De cette manière, l'évolution de nouvelles fonctions précède la duplication des gènes. Cette hypothèse est toujours d'actualité et ses mécanismes ainsi que les issues de ce modèle pour l'évolution des protéines aussi bien de manière naturelle que par ingénierie sont étudiés. Il s'avère que la théorie d'évolution divergente par recrutement d'activités de promiscuité est aussi bien illustrée par différentes familles d'enzymes, comme la famille des enzymes à tonneau (α/β)₈ [8], que par l'ingénierie d'enzymes, générant des protéines généralistes à partir de protéines montrant une certaine promiscuité [9].

La duplication en tandem résulte de l'expression de 2 copies ou plus du gène sur une même chaîne polypeptidique. Une telle multiplicité présente différents avantages comme l'augmentation de la stabilité, l'apparition de nouvelles fonctions coopératives, l'apparition de sites de liaison dans la nouvelle cavité formée par le contact des deux domaines, l'émergence de longues structures répétitives comme les protéines fibreuses et enfin, la production de sites de liaison multiples en série augmentant la spécificité et l'efficacité de la protéine. Cette duplication en tandem peut ensuite être permutée circulairement, réorganisant ainsi l'intégrité structurale et fonctionnelle de la séquence parentale.

La permutation circulaire d'une protéine résulte en la relocation de ses extrémités N et C terminales à l'intérieur de la structure existante (Figure 1, b et Figure 4, e page 16). Les avantages de ce mécanisme au niveau de l'évolution ne sont pas clairs, des expériences *in vitro* ont montré qu'elle affecte la stabilité conformationnelle de la protéine.

L'oligomérisation est l'association non covalente de protéines homo ou hétérologues et joue un rôle important dans la fonction enzymatique (Figure 1, e). L'oligomérisation d'une protéine peut d'une part changer son activité. D'autre part, l'agglomération de structures avec des interfaces hydrophobes exposées augmente la stabilité thermodynamique et peut créer des cavités à l'interface protéine-protéine qui peuvent accommoder le substrat ou des sites de liaison et de régulation. Il existe des cas spéciaux d'oligomérisation où des homodimères sont formés par échange de domaines 3-D ("swapping"). Ceci résulte de la réorganisation structurale entre protéines multi-domaines identiques sans rupture de lien covalent. La structure échangée peut aussi bien être une structure secondaire comme un domaine entier. Les avantages de ce type d'échange sont probablement la possibilité de créer des phénomènes de régulation et d'augmenter la stabilité structurale.

La fusion de gènes va permettre entre autres la coopérativité d'enzymes individuelles, la stabilisation des protéines et des fonctions de régulation avancée (Figure 1, d).

Le recrutement de domaines est la réorganisation de sous-unités protéiques ou domaines individuels, générant une versatilité fonctionnelle des protéines (Figure 1, c).

Evolution divergente et fonction enzymatique

Dans la plupart des cas, des enzymes divergentes conservent une même fonction, mais il arrive que la fonction de l'enzyme diverge elle aussi. Il existe trois mécanismes principaux pour l'évolution divergente d'une fonction protéique dont deux reposent sur des hypothèses en concurrence.

Les deux premiers modèles répondent à la question suivante : "Comment les processus de duplication, de divergence et de recombinaison se placent-ils dans la formation ou l'extension de complexes physiques de protéines ou de voies métaboliques ? " Dans la voie métabolique d'une petite molécule par exemple, de nombreuses protéines ont les mêmes contraintes en termes de mécanisme fonctionnel ou reconnaissance du substrat. Ainsi, on s'attendrait à ce qu'une fraction de ces protéines soient liées les unes aux autres. Deux modèles de chemins évolutifs sont développés à partir de cette idée.

Le premier modèle élaboré par Horowitz en 1945 propose que, parce que les intermédiaires d'une voie métabolique gardent des similarités de structure, l'enzyme évolue par duplication et divergence de protéines dans lesquelles le mécanisme catalytique a été changé et quelques propriétés de liaison au substrat gardées. Il a ainsi été suggéré que les chemins métaboliques évoluent de manière rétrograde : lorsque le substrat d'une enzyme d'une voie métabolique est

épuisé, une nouvelle enzyme évolue pour fournir ce substrat à partir d'un précurseur disponible, par évolution de l'enzyme qui utilise ce substrat [10].

De cette manière, l'évolution est contrainte de retenir une spécificité de liaison, car la nouvelle et l'ancienne enzyme doivent partager la capacité à lier des substrats similaires. Les mécanismes des réactions catalysées par les enzymes originale et évoluée ne sont pas liés, et la nouvelle et l'ancienne enzyme appartiendraient à des membres de suprafamilles¹ fonctionnellement distinctes, c'est-à-dire des groupes d'enzymes homologues qui catalysent des réactions totalement différentes et dont les résidus du site actif, même s'ils sont conservés, n'ont pas les même fonctions. Ce type de mécanisme est très bien illustré par les enzymes PRAI (PhosphoRibosylAnthranilate Isomérase) et InGPS (Indoleglycerol Phosphate Synthase) impliquées dans la voie de synthèse du tryptophane. Ces enzymes partagent un même repliement en tonneau (α/β)₈, présentent 22% d'identité de séquence, et catalysent les réactions présentées en Figure 2. Le produit de synthèse de la PRAI est le substrat de l'InGPS, et la PRAI pourrait avoir évolué à partir de l'InGPS.

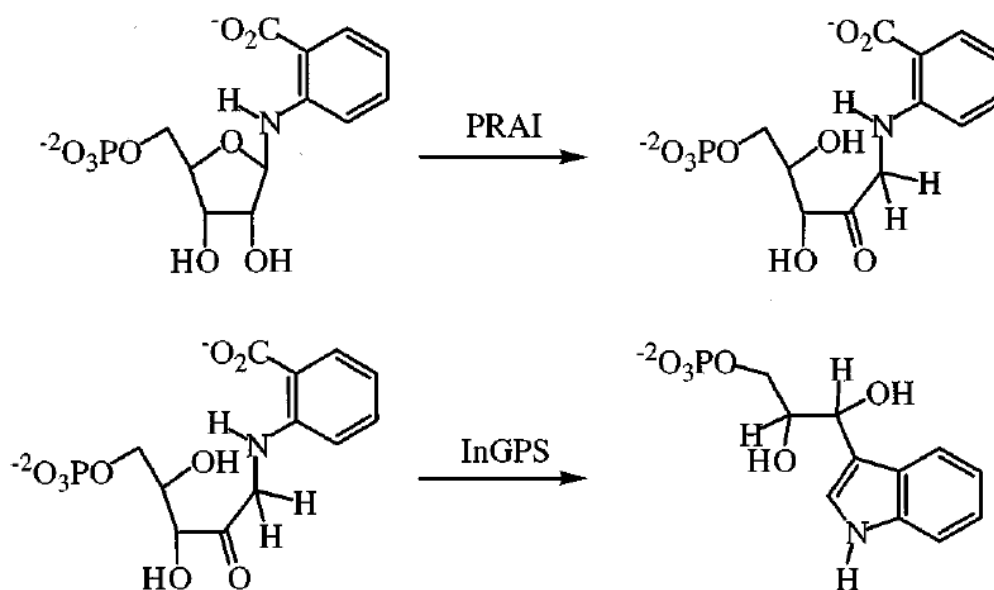


Figure 2 : réactions catalysées par la PRAI et la InGPS deux réactions consécutives dans la voie de biosynthèse du tryptophane, illustrant le concept de suprafamilles fonctionnellement distinctes [3].

¹ Dans un contexte structure-fonction une suprafamille selon la définition de Gerlt & Babbitt est un groupe d'enzymes homologues qui catalysent des réactions différentes et ne partagent aucun attribut mécanistique, et des résidus conservés dans le site actif assureraient des fonctions différentes.

Le second modèle, selon Jensen, propose que les enzymes soient recrutées le long de voies métaboliques, avec les enzymes dupliquées conservant leur fonction catalytique mais évoluant vers différentes spécificités de substrat [7]. La nature sélectionne une protéine à partir d'un pool d'enzymes dont le mécanisme fournit la réaction partielle ou la stratégie requise pour la stabilisation d'intermédiaires énergétiquement non favorables ou d'états de transition. L'évolution altère alors les autres caractéristiques mécanistiques et/ou la spécificité de substrat pour que la réaction soit catalysée avec une compétence améliorée. L'ancienne et la nouvelle enzyme seraient membres de superfamilles mécanistiquement diverses, c'est-à-dire que des groupes d'enzymes homologues vont catalyser des réactions différentes mais partager un même attribut mécanistique permis par la conservation des résidus du site actif qui ont les mêmes fonctions chez tous les membres d'une même famille. De la même manière, la nature sélectionne peut-être un ancêtre qui a la capacité de catalyser la nouvelle réaction, à un niveau relativement bas mais suffisant pour fournir un avantage sélectif à l'organisme (promiscuité). L'évolution ultérieure de la protéine permet d'améliorer les compétences de l'enzyme et la voie métabolique est optimisée. Ce mécanisme évolutif a pu être recréé *in vitro*. Les voies métaboliques de l'histidine et du tryptophane requièrent des enzymes catalysant un réarrangement d'Amadori (Figure 3).

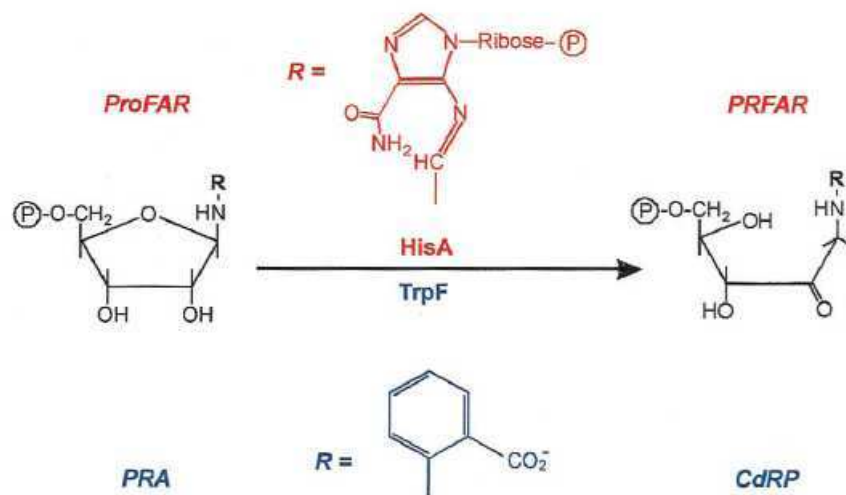


Figure 3 : réactions catalysées par la ProFAR isomérise dans la voie de biosynthèse de l'histidine et la PRA isomérise dans la voie de biosynthèse du tryptophane [11], illustrant le concept de superfamilles mécanistiquement diverses.

Un "ancêtre commun" à ces enzymes, catalysant cette réaction mais avec une spécificité plus large en acceptant les substrats des voies de synthèse de l'histidine et du tryptophane, a pu être créé et caractérisé (voir page 26) [11].

En examinant les fonctions des membres des différentes familles de domaines pour des enzymes impliquées dans des voies métaboliques de petites molécules chez *E. coli*, on observe que dans presque tous les cas, c'est le mécanisme catalytique ou les propriétés de liaison au cofacteur qui sont conservées et la spécificité de substrat changée. Ceci suggère qu'il est plus facile d'évoluer de nouveaux sites de liaison à des substrats que de nouveaux mécanismes catalytiques [5]. Les travaux d'évolution dirigée mènent aux mêmes conclusions, il est plus facile de changer une spécificité de substrat que de créer une nouvelle activité sur un support protéique (voir page 25).

La troisième hypothèse pour l'évolution de nouvelles fonctions reposerait sur la conservation de l'architecture du site actif. L'évolution serait opportuniste et capable de recruter un site actif capable de supporter des réactions alternatives et pouvant utiliser des groupes fonctionnels du site actif dans des contextes mécanistiques et métaboliques différents. L'ancienne et la nouvelle enzyme seraient par définition membres de suprafamilles fonctionnellement distinctes, comme le type d'enzymes ayant évolué selon le système proposé par Horowitz [3]. Ceci peut être illustré par deux enzymes, la KGPDC (3-Kéto-L-Gluconate 6-Phosphate DéCarboxylase) et l'OMPDC (Orotidine 5'-MonoPhosphate DéCarboxylase) sont des enzymes qui catalysent des réactions différentes dans des contextes métaboliques différents. Malgré une faible identité de séquence (< 25%), ces deux enzymes partagent une même structure dimérique de tonneau (α/β)₈ dont l'interface est conservée, en particulier, un motif Asp-X-Lys-X-X-Asp est retrouvé dans les deux enzymes, et d'autres résidus du site actif sont conservés, même s'ils n'assurent pas un rôle similaire dans les catalyses respectives de ces deux enzymes. Sur base de ces données, il est proposé que ces deux enzymes sont un exemple typique d'évolution par recrutement de l'architecture du site actif [12, 13].

La plasticité des sites actifs

Nous venons de voir que si une divergence de fonction a accompagné une divergence de séquence, dans beaucoup de cas, les membres d'une famille de protéines partagent une caractéristique fonctionnelle comme une propriété de liaison à un substrat ou une stratégie mécanistique de catalyse. Mais des enzymes homologues peuvent aussi catalyser des réactions tout à fait différentes, avec un mécanisme catalytique parfois différent. Ceci est dû à des modifications fonctionnelles réalisées à travers des changements de caractéristiques structurales au voisinage du site actif, ou par des changements de structure qui peuvent affecter les interactions entre domaines ou sous-unités. La situation inverse où de mêmes

réactions chimiques sont réalisées de différentes manières est plus rare. Dans ce paragraphe, nous verrons des exemples d'enzymes homologues pour lesquelles des groupes fonctionnels différents, non conservés par rapport à leur position dans la séquence primaire ont le même rôle mécanistique dans la catalyse. Ainsi, deux solutions ou plus au même problème ont évolué en une seule et même superfamille d'enzymes. Les origines possibles de la variabilité des résidus dans le site actif d'enzymes homologues sont illustrées dans la Figure 4 [14].

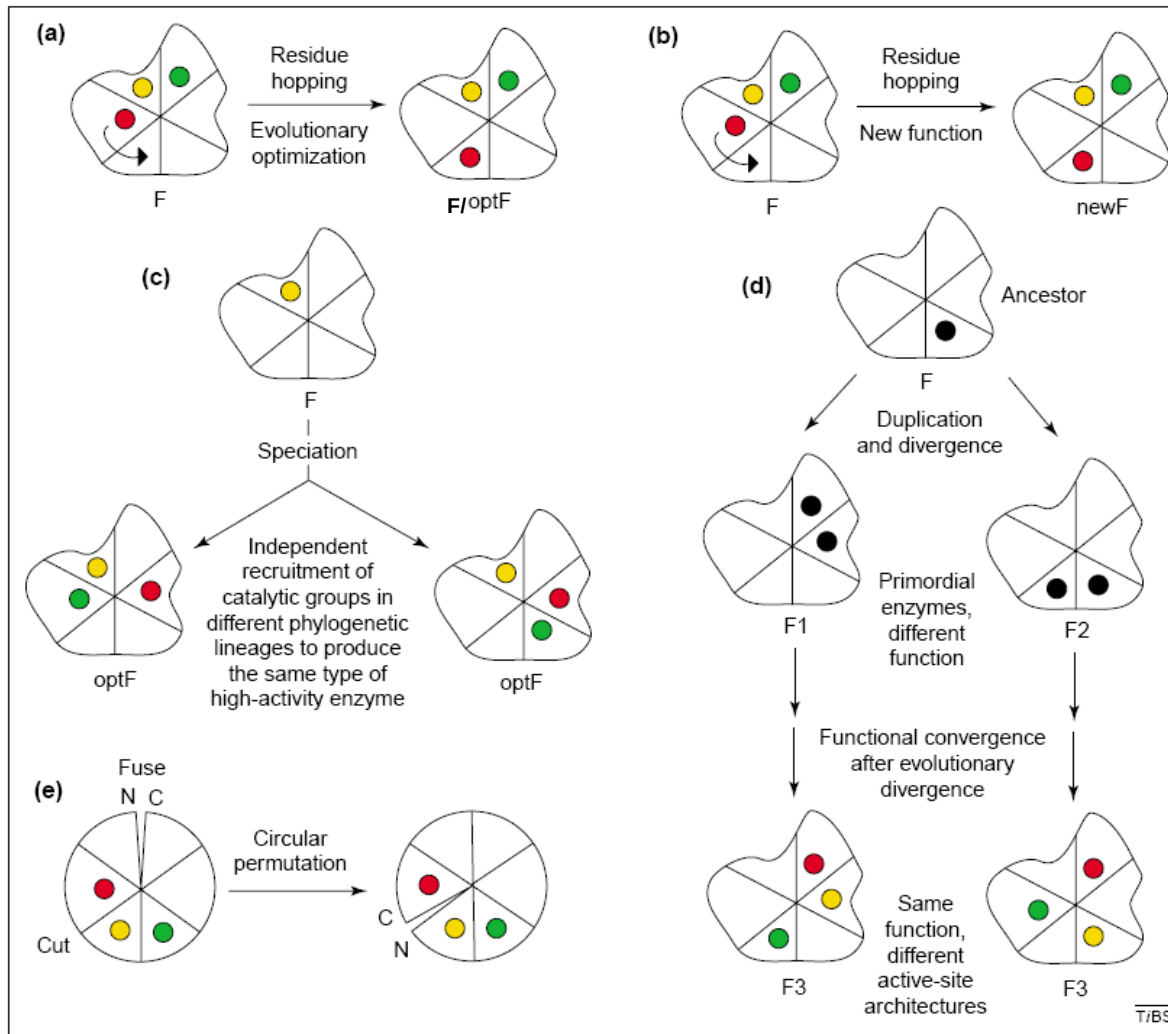


Figure 4 : hypothèses pour la non-équivalence des résidus fonctionnels qui jouent le même rôle catalytique dans des enzymes homologues, reproduit de [14]. Chaque forme représente une structure protéique, et chaque "division" de la structure protéique représente une position de la structure où un résidu fonctionnel peut être situé. Les cercles noirs et colorés représentent les résidus fonctionnels, et ceux qui ont la même couleur ont un rôle identique dans la catalyse. "F" ou "F1" dénotent des fonctions particulières, et "optF" la même fonction mais optimisée. "newF" dénote une nouvelle fonction et "N" et "C" représentent les extrémités N et C-terminales.

La migration de résidu fonctionnel (hopping) consiste en l'évolution dans le site actif de nouveaux groupes qui prennent le rôle d'un ancien résidu. Ceci peut se faire dans le but d'optimiser une fonction. Le "hopping" pourrait aussi résulter de la mutation du résidu catalytique et du recrutement d'un nouveau résidu assurant une fonction équivalente, sans but d'optimiser la catalyse ou la stabilité (Figure 4, a). Dans les deux cas, il sera possible de retrouver des vestiges de l'ancien résidu ou la présence du futur résidu catalytique chez des enzymes homologues. Enfin, l'apparition de nouveaux résidus catalytiques alternatifs pourrait faire évoluer l'enzyme vers de nouvelles fonctions (Figure 4, b).

Le deuxième mécanisme pouvant être à l'origine de la plasticité du site actif est l'optimisation indépendante de lignées phylogénétiquement différentes. Durant l'évolution, différents acides aminés peuvent être recrutés dans des familles distinctes pour néanmoins évoluer vers le même type d'enzyme (Figure 4, c). C'est un cas typique d'évolution convergente, qui peut être illustré par les familles des protéases à sérine. Ces familles d'enzymes partagent une même triade catalytique dont les chaînes latérales donnent au site actif une géométrie similaire, alors que les repliements tridimensionnels et par conséquent, les séquences, ne sont pas homologues [15].

La divergence d'une enzyme suivie par un phénomène de convergence fonctionnelle pourrait être une des explications de la plasticité du site actif. Une activité enzymatique particulière aurait évolué deux ou plusieurs fois pendant l'évolution à l'intérieur d'une même superfamille par des moyens indépendants (Figure 4, d). La famille des β -lactamases illustre bien ce phénomène, les β -lactamases de classe A et D par exemple ont évolué de manière divergente et parallèle à partir d'un ancêtre commun. Les premières ont acquis leur mécanisme catalytique par recrutement d'un résidu acide situé sur une boucle, les secondes ont acquis la capacité de carboxyler un résidu du site catalytique pour assurer la même fonction (voir page 74).

Enfin, la permutation circulaire d'une enzyme due à des réarrangements génétiques pourrait être à l'origine de la migration apparente des résidus catalytiques (Figure 4, e).

Dans la plupart des cas, des analyses de structure tridimensionnelle des protéines homologues permettent d'identifier les vestiges de résidus catalytiques ou les nouveaux résidus, et des mutations ponctuelles peuvent générer des activités ancestrales ou nouvelles selon le cas. Ce genre de migration peut aussi être généré de manière artificielle. La KDGP aldolase (2-Kéto-3-Déoxy-6-PhosphoGluconate aldolase) est une enzyme appartenant à la famille des enzymes à tonneau (α/β)₈. Le résidu clé de la catalyse est une lysine, la création d'un double mutant

repositionnant cette lysine sur un brin β différent et de manière symétrique par rapport à sa position initiale permet d'élargir la spécificité de substrat de cette enzyme [16].

Evolution convergente ou divergente ?

Les structures protéiques similaires sont issues d'ancêtres communs qui ont été dupliqués et ont évolué de manière divergente (protéines homologues) ou sont apparues plusieurs fois à partir d'ancêtres différents par évolution convergente (protéines analogues). Ces structures ont alors pu se recombiner entre elles de différentes manières pour fournir la diversité fonctionnelle représentée à travers l'univers protéique.

La procédure généralement acceptée pour distinguer entre divergence et convergence est de compter le nombre de caractéristiques communes : la séquence en ADN des gènes, la séquence en acides aminés, la structure tridimensionnelle, les interactions enzyme-substrat, les mécanismes catalytiques et enfin, la position des résidus catalytiques dans les segments de chaînes polypeptidiques (non transposés). Si les critères de similarité de séquence sont acceptables, le reste suit dans la plupart des cas, et les protéines sont des homologues.

La démarcation entre divergence et convergence pour les protéines dont les séquences sont moins de 25-40 % identiques ("twilight zone") reste matière à débat [17]. Les protéines qui ont divergé à partir d'un ancêtre commun vont généralement préserver les éléments de cet héritage en termes de fonction, structure et séquence. Parce que les structures tertiaires sont relativement robustes aux perturbations de séquence, les repliements protéiques sont bien conservés et évoluent plus lentement que les séquences en acides aminés, et il est généralement accepté que des protéines adoptant un même repliement sont homologues. Cependant, des séquences convergentes vont aussi avoir tendance à adopter un même repliement protéique, qui sera alors convergent, puisque le nombre de repliements possibles à partir d'une séquence donnée est limité. De même, des contraintes extérieures (physiques, chimiques...) vont forcer la convergence du repliement. Ceci a été confirmée récemment par la détermination de la structure du premier domaine protéique créé *de novo* [18-20]. Une banque de chaînes polypeptidiques composées de 80 acides aminés aléatoires, comprenant 6.10^{12} clones a été soumise à des sélections par ARNm display (voir page 51) pour une fonction de liaison à l'ATP. De cette manière, quatre protéines avec des constantes de dissociation (k_d) variant de 100 nM à 10 μ M ont été sélectionnées et caractérisées [19]. La cristallisation de l'une d'entre elles, l'ANBP (Artificial Nucleotide Binding Protein) montre que la protéine comporte un site de liaison à un ion Zn^{2+} , et bien que la structure générale de la protéine ne ressemble à aucune autre connue, la région du site de liaison présente une

grande similarité avec la région d'une autre protéine liant le zinc [18, 20]. Dans ce cas, les contraintes chimiques et géométriques de liaison au Zn mènent à une structure convergente du site de liaison.

Ainsi, lorsque les séquences de deux domaines protéiques sont comparées, une similarité significative est une forte preuve de l'existence d'un ancêtre commun. En l'absence de similarité de séquence, il était jusqu'ici accepté par défaut que des conservations de structure et de fonction sont la preuve d'évolution divergente de part l'improbabilité de la convergence de structure et de fonction. Cependant, des domaines protéiques non liés pourraient adopter le même repliement à cause de contraintes physiques, de sélection pour une même fonction ou par hasard [21]. Ainsi, il est nécessaire d'être très prudents dans cette zone où les séquences sont peu semblables avant de parler de convergence ou de divergence.

Comment reconnaître des protéines homologues ?

Au vu des mécanismes d'évolution, on peut considérer que des protéines issues d'un même ancêtre commun ont pu diverger au point que leur séquences ne partagent plus une identité suffisante pour être reconnues comme telles. Ainsi, différents niveaux sont à étudier pour trouver les homologues de protéines.

Pour des protéines dont l'identité de séquence est supérieure ou égale à 20%, des programmes d'alignements de séquences deux à deux qui recherchent des fragments homologues en ignorant les insertions, comme FASTA et BLAST sont tout à fait appropriés. L'étude des structures 3-D des protéines a cependant montré que certains homologues ont divergé au point que la similarité de séquence devient non significative. Les programmes d'alignements deux à deux ne sont plus alors capables de reconnaître ces homologues et la reconnaissance de ces homologues n'est possible qu'en construisant des profils de séquences qui utilisent des jeux d'homologues connus d'une famille de protéines pour déterminer les contraintes sur les changements évolutifs. Des programmes tels que psiBLAST construisent des motifs ou profils de résidus spécifiques en analysant automatiquement des alignements multiples de séquences et caractérisant la préférence de tel ou tel résidu à chaque position.

Cependant, même ce type de programmes ne permet parfois pas de retrouver les homologues d'une protéine et on doit se tourner vers la structure 3-D. La plupart des méthodes se basant sur la comparaison des propriétés structurales des résidus pour trouver des homologues sont lentes et ne sont pas adaptées à des comparaisons à grande échelle. Alternativement, on peut alors utiliser des protocoles comparant les structures secondaires entre protéines pour lesquelles le nombre est d'un ordre de grandeur plus petit que celui des résidus. Ces

protocoles ont l'avantage de permettre la réalisation d'un grand nombre de comparaisons entre des protéines non liées pour tester statistiquement si la similarité est significative [2]. Le nombre croissant d'entrées dans les banques de données de structure 3-D (PDB) nécessite une gestion de ces données afin de classer les protéines homologues en rapport avec leur structure.

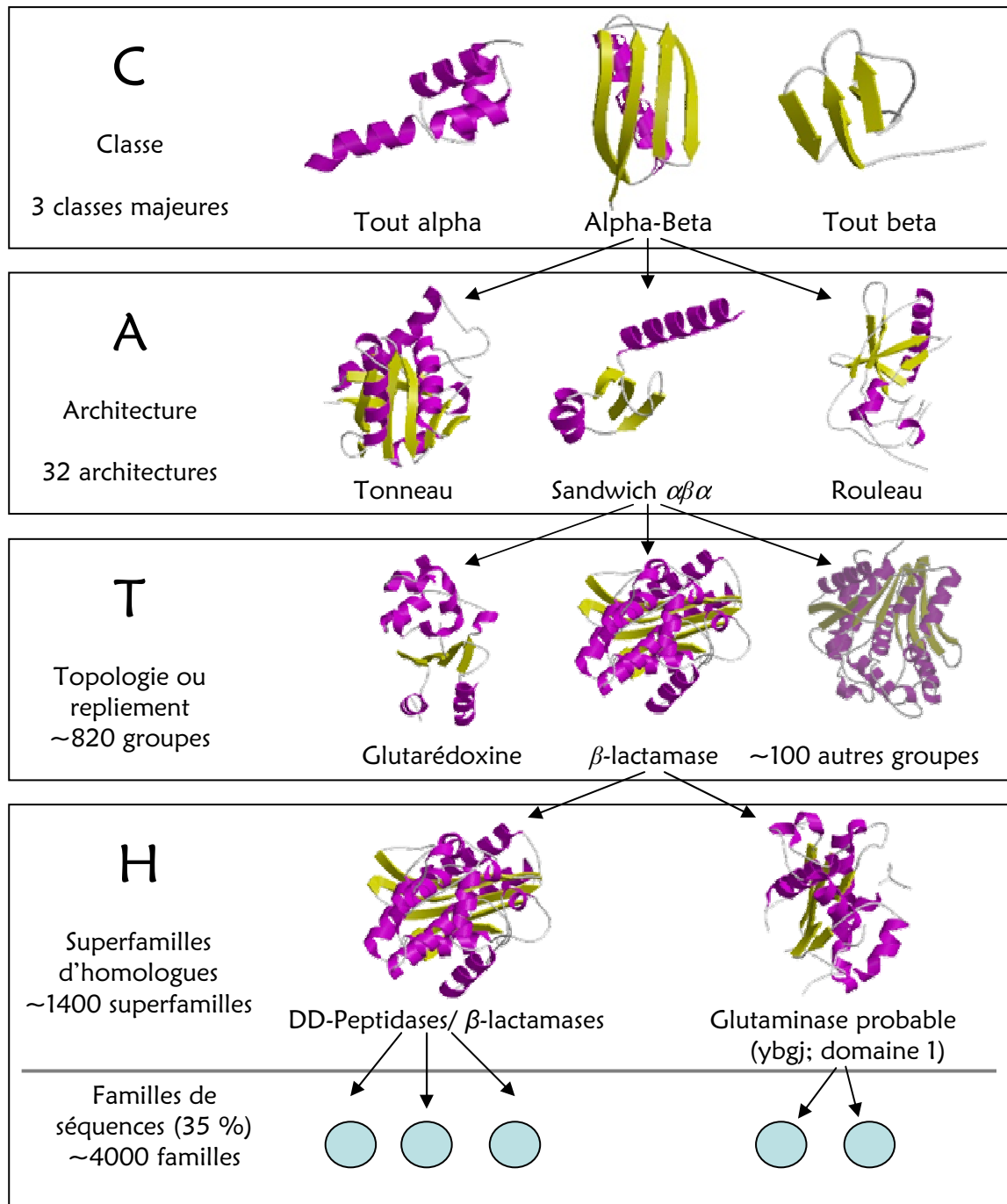


Figure 5 : les quatre niveaux hiérarchiques majeurs dans la classification structurale du CATH : Class, Architecture, Topology or fold level, and Homologous superfamily. L'exemple est pris pour la topologie de type β -lactamase.

C'est ce qui est réalisé par deux principaux systèmes de classification, CATH (Class, Architecture, Topology or fold, Homologous family) et SCOP (Structural Classification Of Proteins). Afin de classer des protéines en familles de structures homologues, les protéines doivent d'abord être découpées en domaines, puis à un domaine est attribué un repliement. Alors que pour la classification SCOP, les domaines protéiques ne montrant pas d'homologie de séquence évidente à d'autres domaines, sont définis et classés manuellement, la classification CATH utilise une combinaison de procédures manuelles et automatiques afin de définir et classer les domaines protéiques. Un exemple est donné pour la classification CATH dans la Figure 5.

Il existe aussi des systèmes de classification structure-fonction [22]. L'un d'entre eux, appelé le Structure-Function Linkage Database (SFLD) a été développé pour capturer l'information sur les attributs mécanistiques partagés (comme des réactions partielles), les résidus catalytiques conservés, les structures et les fonctions, pour des superfamilles d'enzymes mécanistiquement diverses [23]. La classification des familles de protéines en fonction du mécanisme de divergence des fonctions serait un outil intéressant pour les recherches de protéines homologues.

Malgré les avancées dans les bases de données concernant la classification des familles de protéines, et l'augmentation incessante des entrées PDB, de nombreuses protéines restent à découvrir, et avec elles probablement des données sur les mécanismes d'évolution naturelle des protéines. Cependant, alors que ces mécanismes sont encore mal connus, de nombreux chercheurs tentent de "dépasser" ce que l'évolution naturelle a accompli en quelques dizaines, centaines, milliers voire millions d'années, en créant en quelques semaines de nouvelles fonctions protéiques.

L'évolution divergente des protéines, concept de diversité conformationnelle

Outre les mécanismes d'évolution étudiés précédemment, d'autres éléments sont susceptibles d'entrer en jeu. Un élément en particulier attire notre attention, c'est la diversité conformationnelle des protéines.

Les modèles d'évolution vus jusqu'ici présentent la structure tridimensionnelle des protéines comme les éléments les mieux conservés. Généralement, la structure primaire d'une protéine dicte sa topologie et sa fonction, et des changements cumulatifs de la séquence mènent à l'évolution de la structure et des fonctions de la protéine. Cependant, les boucles dans les protéines adoptent des conformations différentes et généralement seule la structure du cœur

protéique est conservée. De plus, certaines protéines peuvent adopter des conformations différentes, l'exemple le plus connu étant le phénomène d'allostérie où la liaison de la protéine à son ligand va entraîner un changement de sa conformation, présupposant d'une certaine adaptabilité de la conformation des protéines.

Il existe aussi des exemples pour lesquels des séquences très similaires ou même identiques vont adopter des structures partielles différentes. La structure du double mutant $N_{11}L/L_{12}N$ de la protéine répresseur Arc est un exemple de plasticité de la structure des protéines où le mutant adopte un nouveau repliement en hélice 3_{10} et remplace le feuillet β antiparallèle de la structure native. Le simple mutant comportant la mutation $N_{11}L$ peut adopter l'un ou l'autre repliement et existe sous forme d'équilibre dynamique en solution (Figure 6) [24].

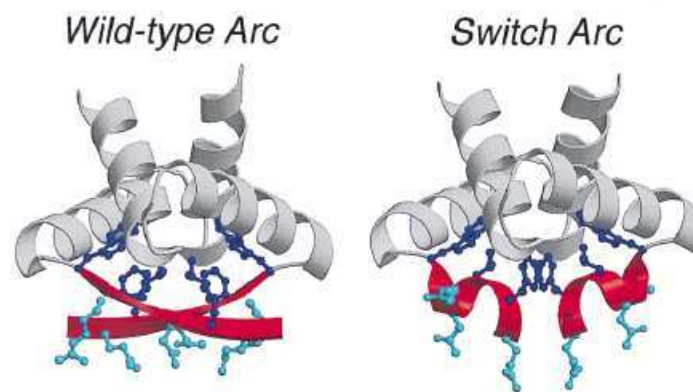


Figure 6 : répartition des résidus hydrophobes et polaires entre la surface et l'intérieur de homodimères de la protéine Arc sauvage et mutante (switch Arc). Une représentation MOLSCRIPT des résidus 8-46 est montrée, avec la plus grande différence structurale (résidus 8-14) colorée en rouge, et les chaînes latérales à l'intérieur et en surface colorées en bleu et cyan respectivement [24].

De même, la protéine Mad2, impliquée dans la régulation du cycle cellulaire comporte une séquence "caméléon", c'est-à-dire qu'elle adopte deux conformations en équilibre à l'état natif, N1-Mad2 et N2-Mad2. Un cœur compact commun représente environ 70% de la protéine et le reste subit des changements conformationnels majeurs comme le repliement et la translocation de la structure C-terminale (Figure 7) [25].

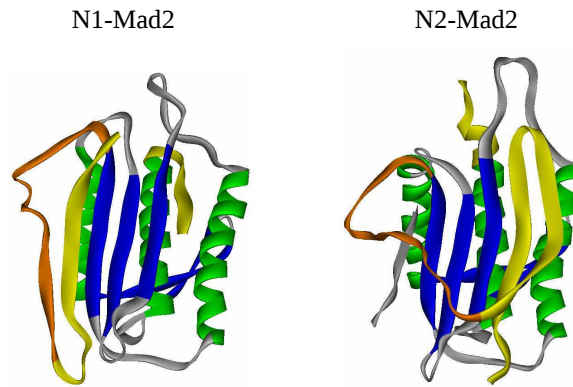


Figure 7 : structure en rubans de N1-Mad1 et N2-Mad2, deux états natifs de la protéine Mad en équilibre. Les brins β sont représentés en bleu, les hélices α en vert et les boucles en gris. Les éléments structuraux qui subissent des changements majeurs entre les conformères N1 et N2 sont indiqués en jaune et orange [25].

La manifestation extrême de la diversité de conformation des protéines se retrouve dans les protéines "moonlight" ou "protéines multi-tâches". Elles sont par définition des protéines qui ont la capacité d'exécuter plus d'une fonction, apparemment non liées. Dans les cas classiques de "moonlighting", ces phénomènes de fonction distinctes peuvent être interprétés dans le cadre classique structure-fonction, en attribuant des fonctions additionnelles à des surfaces de liaisons de protéines globulaires qui sont distinctes du site responsable de leur fonction primaire. Cette capacité à lier différents partenaires pourrait être initiée par différents mécanismes comme l'expression de la protéine dans différents types cellulaires ou dans des compartiments cellulaires différents [26]. La glucose phosphate isomérase par exemple présente une activité enzymatique de conversion du glucose-6-phosphate en fructose-6-phosphate dans le cytosol et une fonction extracellulaire de cytokine en se liant à un récepteur [27].

Cependant, il existe aussi des protéines qui apparaissent en partie ou complètement déstructurées dans les conditions natives, mais adoptent une structure parfaitement ordonnée en présence de ligand, et la fonction même de ces protéines peut varier suivant qu'elles sont à l'état natif ou structuré [26]. De cette manière, la protéine CFTR (Cystic Fibrosis Transmembrane Regulator) est un canal chlore impliqué dans la fibrose kystique qui est bloqué et inhibé quand son domaine désordonné R de régulation est déphosphorylé. La phosphorylation de CFTR par la protéine kinase A induit un changement conformationnel de R qui stimule l'ouverture du canal [28]. La protéine p21^{Cip1} est un inhibiteur de kinase cycline dépendante de 164 acides aminés dont les 94 acides aminés du côté N-terminal n'adoptent pas de structure secondaire régulière, on considère donc que cette protéine à l'état natif est très désordonnée. Cette protéine est capable de se lier à la cytokine Cdk₂ (Cyclin-dependent

kinase) et adopte dans ce cas une structure stable et définie [29]. La protéine p21Cip1 est capable de reconnaître différentes Cdk, elle se lie aux sous-unités cycline et kinase de la cytokine activée et inhibe l'action kinase. Cependant, l'assemblage du complexe cycline-kinase est aussi facilité par p21Cip1, et dans ce cas, la protéine active la cytokine. Ainsi, la protéine p21Cip1 va inhiber ou activer la Cdk suivant le cas [26]. Il apparaît que ces protéines "intrinsèquement déstructurées" sont des éléments centraux dans les ensembles complexes d'interactions protéine-protéine, en étant capables d'interagir avec des dizaines, voire des centaines de partenaires [30].

Ces exemples signifient qu'il existe une alternative au concept de conservation de structure supérieur aux séquences et montre que des structures protéiques peuvent évoluer et changer [31, 32]. De cette manière, une fonction pourrait coévoluer en même temps que le repliement tridimensionnel d'une protéine.

La diversité conformationnelle des protéines lève comme hypothèse qu'une protéine peut exister sous forme d'un ensemble d'isomères pouvant lier un ligand différent, la diversité fonctionnelle pouvant alors aller au-delà de la diversité de séquence.

L'ingénierie des protéines : approches stratégiques et applications, création de banques d'enzymes, criblage et sélections

La nature a évolué des enzymes pour catalyser un éventail étonnant de réactions. Cette diversité suscite l'intérêt des chercheurs car ils y voient la possibilité de répondre à différentes demandes en élaborant des enzymes sur mesure, par design rationnel ou évolution dirigée. En biotechnologie, le design d'enzymes et d'autres biomolécules avec des propriétés améliorées ou altérées a de nombreuses applications en médecine, dans l'industrie, l'agriculture, les sciences environnementales, pour des revues, voir [33, 34]... Outre les applications pratiques, les avancées réalisées dans le domaine de l'évolution dirigée sont les moyens par lesquels les chercheurs peuvent prouver qu'une structure protéique présente une certaine plasticité fonctionnelle, importante pour la compréhension de l'évolution naturelle de nouvelles activités protéiques. Ainsi, l'ingénierie des protéines va mimer la stratégie que la nature utilise pour faire évoluer des protéines. Celles-ci ont évolué à travers des processus de duplication de gènes, de recrutements de domaines, d'insertions, de délétions et de fixation de mutations ponctuelles multiples. Ces mêmes processus recréés en laboratoire cherchent à recruter des fonctions établies et des propriétés de protéines existantes pour les incorporer dans de nouvelles protéines par ingénierie.

Ingénierie des protéines : Approches stratégiques et applications

Design rationnel ou évolution dirigée ?

Malgré les avancées récentes en techniques d'ingénierie des protéines, la conversion d'une protéine X (protéine mère) en protéine Y ayant acquis une nouvelle propriété, activité enzymatique ou une capacité nouvelle à lier un ligand reste une tâche difficile. Il existe deux stratégies principales pour générer de nouvelles protéines. Le design rationnel permet de générer de nouvelles protéines sur base de connaissances structurales ou mécanistiques de la protéine mère. Dans ce cas, un nombre limité de variants (quelques dizaines) de la protéine vont être construits et chaque variant sera testé pour l'acquisition de la propriété recherchée. L'évolution dirigée repose sur le principe Darwinien d'évolution. Des collections de mutants pouvant aller de plusieurs milliers à plusieurs milliards de clones vont être générées par

différentes techniques puis les protéines comportant les propriétés recherchées seront identifiées par sélection ou criblage. Dans ce paragraphe, nous allons étudier comment ces deux approches ont pu donner lieu à des protéines ayant acquis de nouvelles propriétés ou fonctions, mais aussi quelles sont leurs limites. Une troisième approche semi-rationnelle, combinant les deux types de stratégie, est plus propice à donner des résultats intéressants en terme d'évolution d'activités enzymatiques.

Design rationnel

La stratégie de design rationnel est une approche moins stochastique que l'évolution dirigée, car elle consiste en l'utilisation de propriétés existant dans les enzymes et des connaissances de séquence et de structure pour construire une enzyme ayant acquis de nouvelles propriétés. L'introduction principalement de mutations ponctuelles et l'échange de structures secondaires sur base d'études des relations structure-fonction de protéines ont permis le changement de propriétés non catalytiques de protéines comme leur thermostabilité, une résistance à un certain pH ou la régulation d'une activité. Ces méthodes ont aussi permis d'agir sur l'activité d'enzymes en modifiant leur spécificité, ou en créant ou améliorant des activités catalytiques [35, 36]. Faire une revue exhaustive de toutes les enzymes générées par cette stratégie est hors des limites de cette introduction. Nous donnerons donc quelques exemples récents et représentatifs d'ingénierie illustrant l'activité dans ce domaine.

Une des approches du design rationnel est d'utiliser les fonctionnalités d'enzymes issues de différentes espèces et de les combiner entre elles afin d'obtenir une nouvelle enzyme ayant acquis les fonctionnalités optimales des deux protéines. Récemment, une forme hyperthermostable d'avidine, une protéine liant la biotine avec une grande affinité et largement utilisée comme outil biotechnologique, a pu être générée. Des comparaisons structurales entre l'avidine de poulet et AVR4 (Avidine-Related), une protéine de poulet plus stable, homologue à l'avidine, ont permis d'identifier des régions de la protéine AVR4 semblant impliquées dans la stabilité. Une région de l'avidine a été remplacée par 21 acides aminés d'AVR4 et une mutation ponctuelle a été introduite. La protéine chimère résultante a les mêmes propriétés de liaison à la biotine mais un T_m plus grand de 27,6°C par rapport à l'avidine sauvage [37].

Le design rationnel d'enzymes peut donner lieu à des améliorations d'activités catalytiques, ou des changements de spécificité. *Jurgens et al.* ont montré qu'il était possible de "recréer" un ancêtre commun à la N'-[5'-phosphoribosyl-formimino]-5-aminoimidazole-4-carboxamide ribonucléotide (ProFAR) isomérase (HisA) et à la phosphoribosylanthranilate (PRA)

isomérase (TrpF) [11]. Ces deux enzymes appartiennent à la famille des protéines à tonneau ($\beta\alpha$)₈ et catalysent un réarrangement d'Amadori dans les voies de synthèse de l'histidine et du tryptophane respectivement. Ceci suggère un modèle d'évolution comprenant l'existence d'un ancêtre commun généraliste catalysant la même chimie mais avec une large spécificité de substrat (voir page 12). A partir de la ProFAR isomérase, une mutation ponctuelle au niveau d'une des boucles liant un brin β à une hélice α a permis de générer une enzyme portant les deux types de spécificité, suggérant le "retour" à cet ancêtre commun. On note cependant une perte d'efficacité de la catalyse pour le substrat naturel de l'enzyme. Dans ce cas, on a pu prouver la plasticité des enzymes pour la spécificité de substrat, tout en recréant expérimentalement un modèle d'évolution possible pour l'évolution naturelle des enzymes.

Plusieurs aspects d'une protéine peuvent parfois être améliorés en même temps. La flavoenzyme pyranose 2-oxidase (P2Ox) est impliquée dans la transformation de carbohydrates et a de nombreuses applications en chimie analytique clinique, le suivi de bioprocédés et la chimie de carbohydrates synthétiques. Sur base d'études précédentes sur des homologues de la P2Ox de *Peniophora* sp., et la détermination de la structure de cette protéine, l'introduction d'une mutation ponctuelle a permis d'augmenter la stabilité à la température et au pH, et d'augmenter l'efficacité catalytique de l'enzyme pour les substrats principaux. L'introduction d'une deuxième mutation et d'une étiquette histidine en C-terminal a entraîné la modulation de l'activité pour différents substrats. Ainsi, l'analyse des relations structure-fonction pour cette protéine a pu être examinée plus en détails [38].

Enfin, le plus grand défi du design rationnel est l'ingénierie de nouveaux mécanismes catalytiques. On pense généralement que la chimie est le facteur dominant de l'évolution de nouvelles activités enzymatiques, par conséquent, des protéines avec des structures similaires peuvent réaliser des réactions chimiques très différentes après l'incorporation de nouveaux groupes catalytiques. Ainsi, 5 mutations ponctuelles dans le site actif de la glutathione transférase A2-2 caractérisée par une activité glutathione peroxydase ont été suffisantes pour la transformer en glutathione transférase A3-3, caractérisée par une activité isomérase [39], démontrant une certaine plasticité fonctionnelle de la protéine. Plus récemment, la chimie du résidu His dans des réactions d'hydrolyse et de transestérification dans une structure protéique modèle a été explorée. Sur base de la structure de la glutathione humaine GST A1-1 en complexe avec son substrat, cinq mutants comportant des substitutions uniques ou multiples pour l'incorporation d'histidines ont été préparés afin d'introduire une activité hydrolytique. Un des simples mutants catalyse l'hydrolyse du substrat de la glutathione, et le résidu histidine incorporé agit comme une base générale pour la catalyse [40].

Un autre moyen de générer de nouvelles fonctions enzymatiques est d'assembler les éléments nécessaires, comme la machinerie catalytique ou le site de liaison au substrat, sur une matrice macromoléculaire choisie. Le groupe de *Benkovic* a exploré la possibilité de créer une activité inexistante sur une structure protéique homologue à une protéine comportant une activité catalytique. Malgré une faible similarité de séquence et aucune fonction commune, la scytalone déshydratase et le facteur de transport nucléaire 2 (NTF2) partagent des similarités structurales, en particulier une poche hydrophobe importante pour la liaison au substrat de la scytalone déshydratase est située dans une région similaire de la NTF2, lui permettant de lier faiblement ce substrat. L'introduction de quatre mutations et de l'hélice C-terminale de la scytalone déshydratase dans le NTF2 a permis de générer une faible activité scytalone déshydratase [41].

En plus des données sur les structures tridimensionnelles des protéines, le design rationnel d'enzymes a pu tirer profit de méthodes informatiques pour générer des protéines ayant acquis une propriété recherchée. Certaines propriétés comme la stabilité d'une protéine peuvent facilement être améliorées par l'utilisation d'algorithmes de "packing" et sont capables de prédire des séquences avec une certaine précision. Dans quelques cas, les méthodes informatiques ont permis le design de protéines récepteurs, ou de nouvelles activités. Ceci est particulièrement bien illustré par les travaux de *Hellingsa* [42, 43]. L'utilisation de différents algorithmes a ainsi permis de prédire des mutations capables d'introduire une activité triose phosphate isomérase dans une RBP (ribose-binding protein), un récepteur ne comportant pas d'activité catalytique [43]. Enfin, les chercheurs tentent de faire du design de protéines *de novo* en créant de nouvelles structures tridimensionnelles [44]. Une protéine de 97 résidus adoptant une nouvelle topologie a ainsi pu être créée, sa structure tridimensionnelle étant en adéquation avec la prédiction de structure utilisée [45].

Malgré tous les apports de la stratégie de design rationnel aussi bien pour la création de protéines ayant acquis des propriétés recherchées que pour la compréhension des relations structure-fonction des protéines, ou encore la compréhension des stratégies d'évolution naturelle des protéines, cette approche comporte des limites. Le principal facteur limitant des applications du design rationnel est que cette approche suppose généralement la compréhension détaillée mais souvent inconnue des relations entre une séquence, sa structure et sa fonction. De plus, la plupart des enzymes générées par design rationnel n'atteignent pas les performances catalytiques attendues à cause de problèmes de repliement de la structure, de dynamique ou de stabilité des protéines. La stratégie qui s'impose alors pour dépasser ces problèmes est celle de l'évolution dirigée.

Evolution dirigée

La stratégie d'évolution dirigée consiste en la création de collections ou banques de variants d'une enzyme dans lesquelles des clones ayant des propriétés recherchées vont être identifiées. Le but de ceci est l'accumulation d'amélioration des propriétés recherchées par itération des étapes de mutation et criblage ou sélection. Cette approche est devenue la plus utilisée pour des modifications de propriétés de protéines ou d'activités enzymatiques. De la même manière que pour le design rationnel, cette stratégie a permis entre autres d'améliorer l'activité, la spécificité (énantiosélectivité incluse) ou la stabilité de nombreuses protéines [46-48].

La stratégie la plus naïve pour la construction d'une banque de protéines est la mutagenèse aléatoire de tout le gène, car elle ne requiert pas d'information structurale ou mécanistique de la protéine. Différents cycles de mutagenèse aléatoire et sélection sont généralement appliqués pour obtenir la fonction désirée. La fréquence des mutations bénéfiques est souvent faible par rapport aux mutations délétères. Seules les mutations bénéfiques uniques sont additionnées à chaque tour de mutagenèse et la probabilité d'amélioration de la propriété diminue rapidement quand de multiples mutations sont faites. Ainsi, les approches basées sur l'introduction de mutations ponctuelles aléatoires sont souvent limitées à des améliorations par petits sauts. Une amélioration de la stratégie est la recombinaison de gènes comportant des mutations intéressantes, permettant de créer de nouvelles enzymes avec une combinaison de mutations bénéfiques. On espère alors introduire des mutations qui vont pouvoir agir en synergie, cependant, ces effets synergistiques sont rarement observés pour des enzymes construites sur des méthodes complètement aléatoires [48].

Ainsi, l'évolution dirigée des protéines est efficace pour une large gamme d'applications et présente le double avantage de fournir aussi bien des protéines ayant des fonctions améliorées que des informations importantes pour la compréhension des relations structure-fonction. De cette manière, il est possible d'identifier des positions influençant l'activité ou la stabilité de la protéine, loin du site actif et dont la fonction aurait été difficile à prédire.

Cette méthode d'évolution dirigée présente cependant deux limitations principales. La première est que des méthodes de sélection robustes ne sont pas toujours disponibles pour rechercher les clones les plus intéressants dans les banques, et l'espace de séquence à explorer est souvent très grand. La deuxième limitation de cette stratégie vient de la nature aléatoire des mutations et du nombre limité de mutants qu'il est possible de créer et/ou de cribler si un test individuel doit être appliqué. En effet, si on considère une protéine de 100 acides aminés, le nombre de combinaisons si on a la possibilité de muter chaque résidu vers les 19 autres

acides aminés est de 20^{100} , soient 2.10^{21} fois plus que le nombre d'atomes dans l'univers (10^{80}). Sachant que la meilleure méthode de criblage permet d'analyser 10^7 clones par heure, il faudrait 2.10^{94} années pour cribler toutes les combinaisons de variants de la protéine. Ainsi, dans le cas où une information fonctionnelle ou structurale existe, il est avantageux de l'exploiter en concentrant les mutations où elles sont les plus efficaces.

Approche semi-rationnelle : des banques "intelligentes"

Une approche semi-rationnelle permet de construire des banques plus intelligentes, avec un nombre de clones générés identique à celui d'une approche totalement aléatoire [49]. Il existe plusieurs stratégies de design semi-rationnel, ayant donné comme les stratégies précédentes des résultats convaincants pour la création de protéines avec des propriétés améliorées, et pouvant avoir des applications biochimiques, industrielles ou tout simplement fondamentales pour la compréhension d'un mécanisme catalytique par exemple [33, 50].

La faible fréquence avec laquelle des mutations bénéfiques sont introduites dans ou près du site actif d'enzymes créées par mutagenèse aléatoire et le succès relatif du design rationnel pour l'amélioration de la spécificité de substrat suggèrent que des mutations aléatoires limitées à des sections pertinentes de l'enzyme pourrait produire des gains substantiels, on parle alors de mutagenèse ciblée. De cette manière, les sites actifs des protéines sont une cible de premier choix pour l'introduction de mutations aléatoires, cependant ce type de mutations est souvent déstabilisant pour la protéine.

Une méthode de design semi-rationnel peut comprendre la combinaison de mutations ciblées, choisies sur base de connaissances structurales par exemple, et de mutagenèse aléatoire. Une autre approche est d'alterner de manière cyclique des mutations aléatoires et ciblées. Les expériences d'évolution dirigée aléatoire vont permettre de mettre en évidence des résidus à muter. La mutagenèse de ces résidus de manière ciblée et simultanée va dans la plupart des cas, résulter en des effets synergistiques [50, 51]. Afin de cibler quelques résidus à muter sans passer par une étape préalable d'évolution dirigée, et sans informations structurales ou fonctionnelles d'une protéine, on peut corréler les variations d'un gène d'intérêt au niveau de son évolution avec les divergences phylogénétiques des séquences des membres d'une famille de gènes.

Enfin, une dernière approche serait de combiner les méthodes d'évolution avec des approches informatiques du design de protéines. Même si les méthodes décrites ci-dessus vont permettre de limiter le nombre de variants générés par rapport à l'espace de séquence disponible, et que l'on va cibler une aire de cet espace où les chances de réussite sont plus grandes, le nombre de

variants que l'on peut générer est encore plus grand que ce qu'on est capable de passer au criblage. Les méthodes de construction de banques les plus efficaces permettent de créer jusqu'à 10^{12} variants d'une enzyme, le criblage de cette banque nécessiterait 11 ans à un taux d'analyse de 10^7 variants/heure. Il est donc important d'explorer un espace de séquence utile plus grand. Pour cela, *Voigt et al.* ont décrit des outils informatiques permettant d'identifier les acides aminés tolérants aux mutations dans une protéine [52]. Ce concept de précriblage de banques *in silico* a été développé en utilisant la technologie PDA (Protein Design Automation) qui élimine les séquences incompatibles avec un repliement particulier. De cette manière, la région du site actif de la β -lactamase TEM-1 a pu être redessiné et un tour de sélection sur une banque d'environ 200 000 mutants a permis d'isoler un clone avec un changement de spécificité de substrat [53].

A partir du large éventail de protéines ayant acquis de nouvelles propriétés par évolution dirigée seule ou en combinaison avec du design rationnel, il est difficile d'en tirer les exemples les plus pertinents. Cependant, un exemple attire particulièrement l'attention pour l'utilisation qui est faite de tous les outils d'évolution dirigée pour le design et l'évolution d'une nouvelle activité catalytique sur une structure protéique existante. Ce travail, réalisé par *Park et al.* utilise une approche appelée SIAFE (Simultaneous Incorporation and Adjustment of Functional Elements) qui consiste en l'insertion, la délétion et la substitution de segments de gènes, en conjonction avec l'évolution dirigée. Les éléments fonctionnels qui constituent le site actif sont dessinés sur base de la séquence, d'information structurale et mécanistique et incorporés dans une structure matrice de manière programmée et combinatoire. De cette manière, une activité β -lactamase a pu être générée sur une structure glyoxalase II (GlyII) en trois étapes de reconstruction. La première étape est la délétion du domaine C-terminal de GlyII empêchant l'entrée du substrat dans le site actif. La deuxième étape est l'insertion des résidus catalytiques par mutations ponctuelles. Enfin, la poche catalytique est formée par l'addition de quatre boucles dégénérées, insérées par "PCR" overlap mutagénisante [54].

Malgré les grandes avancées réalisées dans le domaine de l'évolution dirigée, utilisant l'une ou l'autre des stratégies décrites, le succès de cette méthode dépend de manière critique du jeu délicat entre le degré de diversité biologique présente dans la banque, la taille de la banque et la capacité d'un test à détecter les améliorations de l'activité. La force de l'évolution dirigée réside dans sa fonction à mimer l'évolution naturelle des protéines sans une connaissance profonde de la structure/fonction de la protéine à évoluer, alors que sa faiblesse réside dans le petit nombre de séquences qui peuvent être évaluées expérimentalement (10^3 à 10^6 pour des méthodes HTS ou High Throughput Screening et supérieur à 10^{12} pour des méthodes de

display). Avant de décrire les méthodes permettant de créer de la diversité biologique et de sélectionner de manière efficace les variants intéressants d'une protéine nous aborderons la question du choix de la protéine à évoluer.

Choix d'une protéine matrice-Concept "d'évolvabilité"

Le choix d'une protéine de départ ou matrice pour une expérience d'évolution dirigée est important, c'est un sujet cependant peu discuté. Deux concepts pour le choix d'une protéine de départ ont récemment été proposés.

D'abord, Glasner *et al.* mettent en évidence deux idées utilisées en ingénierie des protéines, celle de conservation des mécanismes catalytiques et celle de promiscuité fonctionnelle (publication à paraître [55]). Une des approches pour identifier une protéine de départ serait d'exploiter la conservation d'attributs mécanistiques similaires chez les protéines. En effet, les protéines de superfamilles mécanistiquement distinctes sont capables de catalyser des réactions différentes en conservant une réaction chimique partielle, un intermédiaire ou un état de transition (voir page12). Ceci est typiquement illustré par la superfamille des énolases, qui partagent une structure en tonneau (α/β)₈ et dont tous les membres comportent trois résidus permettant de lier un ion métallique, et un catalyseur général acide/base permettant d'arracher un proton au substrat. Ces enzymes catalysent au moins 12 réactions différentes, présentant un certain degré de ressemblance (Figure 8) [3].

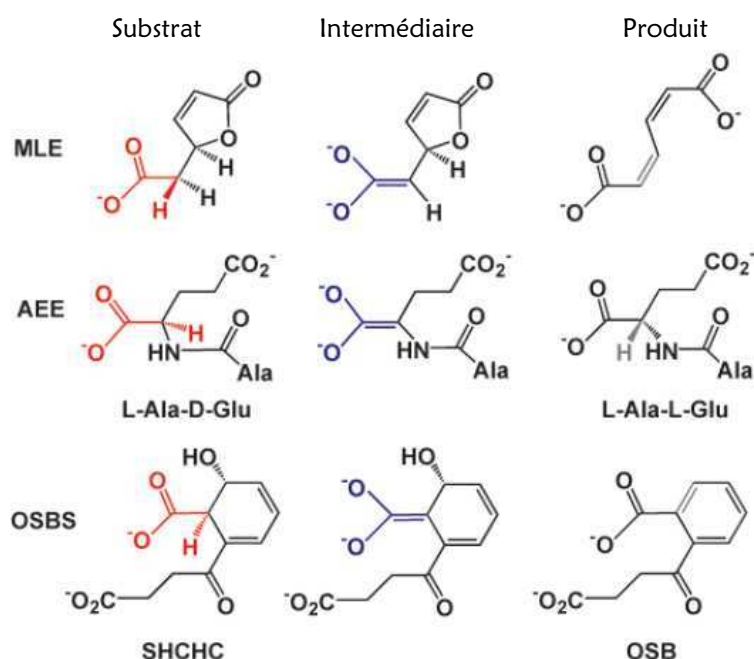


Figure 8 : réactions catalysées par les membres du sous-groupe MLE de la superfamille des énolases : cycloisomérisation (MLEII), transfert de proton 1,1 (AEE) et déshydratation (OSBS) [56].

Ces protéines paraissent donc de bons points de départ pour évoluer de nouvelles activités. Les MLE (Muconate Lactonizing Enzymes) sont un sous-groupe de la superfamille des énolases qui catalysent trois réactions différentes : des cycloisomérisations (MLE), des transferts de protons en 1,1 (D-Ala-D/L-Glu épimérase, AEE) et des β -éliminations (o-succinylbenzoate synthase, OSBS). Ces enzymes partagent les mêmes résidus catalytiques acide/base, deux lysines, dont la position est similaire dans la structure, et une même triade pour la liaison d'un ion Mg^{2+} . Chaque réaction est initiée par l'arrachement d'un proton α d'un carboxylate du substrat par la même lysine pour générer un énoate stabilisé. La seconde partie de la réaction va différer chez les trois enzymes. Des résidus placés sur les boucles liant les brins β de la structure vont conférer la spécificité de substrat et donc l'identité de chaque réaction. Par mutagenèse dirigée, la substitution d'un résidu situé sur une de ces boucles dans l'AEE d'*E.coli* par le résidu équivalent trouvé dans l'OSBS, a permis de créer une activité OSBS dans l'AEE. Indétectable dans l'enzyme de départ, l'activité générée est suffisante ($k_{cat} = 1,3 \cdot 10^{-2} \text{ s}^{-1}$ et $k_{cat}/K_m < 5,2 \cdot 10^{-3} \text{ M}^{-1} \cdot \text{s}^{-1}$) pour conférer un avantage de croissance à une bactérie hôte déficiente en OSBS ($k_{cat} = 28$ et $k_{cat}/K_m = 1,8 \cdot 10^6$ pour l'OSBS de *E. coli*) [57]. Un second résidu identifié par mutagenèse aléatoire et lui aussi placé sur une boucle reliant deux brins β permet lorsqu'il est substitué, d'améliorer la vitesse de croissance en relation avec des constantes cinétiques améliorées de l'AEE pour la réaction OSBS ($k_{cat} = 0,25$ et $k_{cat}/K_m = 340$ pour le meilleur mutant) [56]. Ainsi, le choix d'une protéine de départ pourrait se baser sur l'identification d'une protéine qui catalyse une réaction partielle de celle qu'on veut évoluer. De la même manière, on pourrait se baser sur des suprafamilles d'enzymes, ayant une même architecture du site actif ou une même spécificité de substrat. Cependant, des exemples d'enzymes ayant évolué sur base de capacité à agir sur un substrat similaire sont rares dans la nature (voir page 12) et des expériences d'évolution dirigée basées sur ce point commun auraient probablement moins de chances de réussir [55].

La deuxième approche commentée par Glasner & *al.*, propose de tirer avantage des promiscuités enzymatiques. L'idée qu'il serait plus facile d'évoluer une activité existante plutôt que de partir de zéro a déjà prouvé son efficacité [9]. Les glutathion S-transférases (GSTs) sont des enzymes impliquées dans les processus de détoxification cellulaire dont le mode d'action passe par la conjugaison de composés nucléophiles réactifs sur une glutathione (GSH) endogène. Les GSTs thêta de classe 1-1 humaine (hGSTT1-1) et thêta de classe 2-2 de rat (rGSTT2-2) présentent des spécificités de substrat très différentes, en particulier, la rGSTT2-2 catalyse la conjugaison de la 7-amino-4-chlorométhyl coumarine (CMAC, un composé fluorescent) plus de 400 fois plus efficacement que la hGSTT1-1. Sur base de cette

promiscuité, plusieurs tours de mutagenèse aléatoire et criblage pour l'activité CMAC ont été réalisés. Ceci a permis d'isoler plusieurs mutants de la hGSTT1-1 avec une efficacité catalytique améliorée, et jusqu'à 2.10^4 fois meilleure pour l'un d'entre eux par rapport à l'enzyme sauvage [58].

D'autre part, O'Loughlin *et al.* discutent du problème "d'évolvabilité" des protéines, c'est-à-dire de la capacité d'une protéine à évoluer, pour le choix d'une protéine de départ en évolution dirigée [59]. Il existerait une corrélation entre différents paramètres d'une protéine et son évolvabilité. Le premier paramètre est la spécificité de l'enzyme, l'évolvabilité d'une protéine serait fonction de la capacité d'une protéine à catalyser des réactions secondaires et rejoint le concept de promiscuité vu ci-dessus. Le deuxième paramètre est la thermostabilité et la robustesse (capacité d'une protéine à accepter les mutations sans perturbation de ses propriétés catalytiques). Les protéines thermostables sont de meilleurs points de départ pour l'évolution car elles sont plus robustes. Cependant, ces paramètres sont difficiles à mesurer et donc l'évolvabilité d'une protéine difficile à prédire. O'Loughlin *et al.* suggèrent que la nature a sélectionné certaines enzymes particulièrement propices à évoluer et il propose d'utiliser ces enzymes comme points de départ à des expériences d'évolution dirigée. Ainsi, des protéines soumises à de constants changements environnementaux comme les anticorps ou les protéases du virus HIV seraient des exemples d'évolvabilité et donc des protéines matrices de choix.

Création de banques d'enzymes : comment générer de la diversité

De nombreuses méthodes existent pour générer de la diversité au sein de protéines cibles et sont basées sur des outils de biologie moléculaire. Outre par des outils de biologie moléculaire, des mutations peuvent être générées par l'utilisation d'irradiations UV, de traitements chimiques comme des agents alkylants ou l'utilisation de souches mutatoires. Cependant, ces méthodes ne permettent pas d'introduire de la diversité en un point donné du gène cible, et les mutations peuvent se retrouver en dehors du gène cible. Ainsi, les méthodes les plus largement utilisées font appel à des outils de biologie moléculaire et seront développées dans cette partie.

Mutagenèse dirigée par oligonucléotide

La mutagenèse dirigée est l'outil de base de l'ingénierie des protéines. Cette technique permet de modifier un ou plusieurs résidus choisis dans une séquence protéique et a été un outil critique pour la compréhension des rôles de résidus individuels de protéines.

Alors que pour des stratégies de design rationnel, la séquence de l'oligonucléotide comprenant la mutation à introduire est fixe, pour des stratégies d'évolution dirigée, la diversité est amenée en jouant sur la dégénérescence des nucléotides introduits lors de la synthèse de l'oligonucléotide. Ainsi, un codon NNN codera pour n'importe quel acide aminé (voir Table 1). On peut décider de n'accepter que certains acides aminés, le codon GNN par exemple ne codera que pour Ala, Val, Asp, Glu ou Gly. Enfin, on peut "doper" un oligonucléotide en favorisant la présence d'un acide aminé, en choisissant par exemple T(90%)C(90%)N, on encode une Ser à 81%, et le reste des acides aminés sera faiblement représenté dans la population. Même si le choix et la proportion des mutations paraissent contrôlés, des biais sont introduits, le plus évident étant celui généré par la dégénérescence du code génétique. La synthèse elle-même des oligonucléotides va introduire un certain biais, ce problème est discuté par C. Neylon [60].

1er	U	C	A	G	3ème
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	* (Ochre)	* (Umber)	A
	Leu	Ser	* (Amber)	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

X	NNN	Ala	GCN
Cys	TGY	Asp	GAY
Phe	TTY	Glu	GAR
Leu	TTR, CTN, YTR	Gly	GGN
Ser	TCN, AGY	Val	GTN
Trp	TGG	Ile	ATH
Tyr	TAY	Lys	AAR
His	CAY	Met	ATG
Pro	CCN	Asn	AAY
Gln	CAR	Thr	ACN
Arg	CGN, AGR, MGR	*	TAR, TRA

Table 1 : tables des codons, à gauche, représentation classique, à droite, code pour chaque acide aminé avec N = A, U, G, C, Y = base pyrimidique (U ou C), R = base purique (A ou G), H = U, C ou A, et M = A ou C.

Une méthode classique de mutagenèse dirigée est la mutation par PCR. Celle-ci nécessite l'utilisation de deux paires d'amorces, une paire d'amorces complémentaires comportant la mutation à introduire et une paire d'amorces servant à étendre le gène de part et d'autre de la mutation lors de trois cycles PCR, dont une PCR de chevauchement ou "PCR overlap" (Figure 9). Cette méthode nécessite ensuite une étape de restriction du gène d'intérêt et de clonage dans un nouveau vecteur.

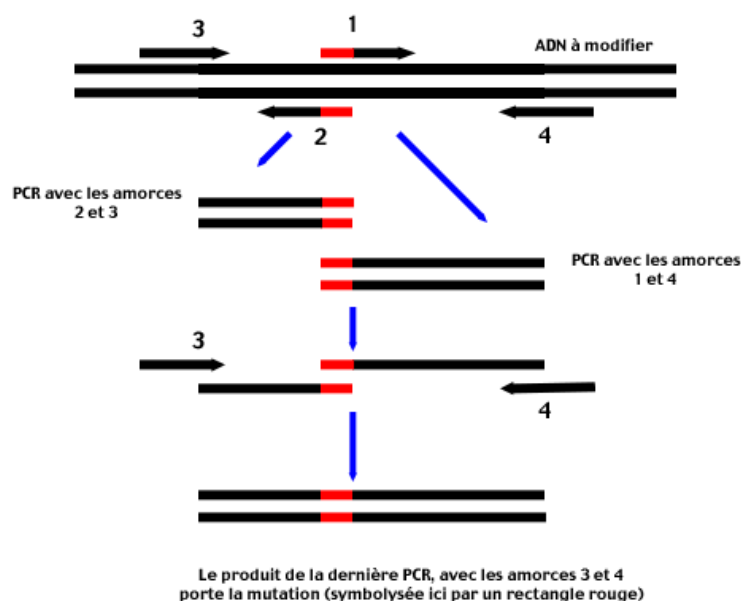


Figure 9 : introduction de mutations ponctuelles par PCR de recouvrement (overlap)

D'autres méthodes utilisées reposent sur l'utilisation d'un oligonucléotide comportant la mutation à introduire et sa polymérisation sur un brin d'ADN simple brin [61] ou double brin [62]. Alors que l'utilisation de la méthode simple brin n'est plus beaucoup utilisée, la méthode utilisant l'ADN double brin, largement améliorée par la firme Stratagène (système QuickChange), est la plus utilisée (Figure 10). Après polymérisation d'ADN à partir des amorces (ou oligonucléotides) comprenant la mutation, l'ADN parental, méthylé s'il a été préparé dans une souche comportant le phénotype Dam^+ , est éliminé par digestion d'une endonucléase *DpnI* reconnaissant uniquement l'ADN méthylé. Une étape de transformation permet de récupérer les diverses formes de l'ADN mutant sans passer par une étape de sous-clonage. Cette méthode peut être utilisée avec de multiples oligonucléotides synthétiques afin de muter plusieurs positions à la fois.

Ces techniques sont limitées par le nombre de résidus à muter : si un nombre de résidus adjacents à muter est trop grand, l'hybridation se fait mal et l'efficacité de polymérisation diminue. Pour palier à ce problème, on peut utiliser la méthode d'insertion de cassette. L'oligonucléotide est conçu de manière à ce que la région à muter soit encadrée par deux sites de restriction, permettant le clonage de l'oligonucléotide double brin directement dans la région du gène qui nécessite de la diversité.

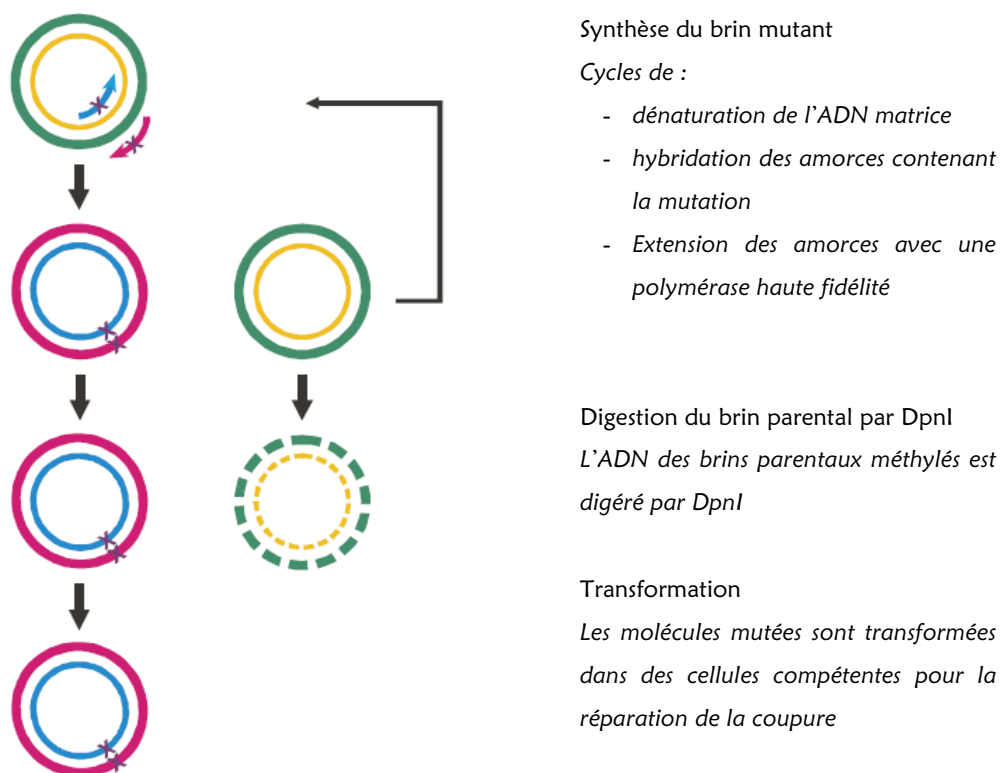


Figure 10 : introduction de mutations ponctuelles au sein d'un gène, par la méthode de l'ADN double brin.

PCR mutagénisante

Lorsque aucune information structurale ou mécanistique sur la protéine à muter n'est connue les banques sont générées par l'introduction de mutations aléatoires le long du gène.

La méthode de PCR mutagénisante ou "error-prone PCR" (ep-PCR) est basée sur la propriété de la Taq polymérase à générer des erreurs au cours de cycles PCR. Alors que le taux basal de mutations est d'environ 1 pour 1000 pb, ce taux d'erreur peut être augmenté avec de plus grandes concentrations en $MgCl_2$, en ajoutant du $MnCl_2$ et en jouant sur des concentrations non équivalentes des quatre dNTPs [63]. De cette manière, l'amplification du gène cible par PCR va générer de la diversité à chaque tour par l'introduction d'erreurs, dont le taux est généralement d'environ 7 pour 1000 pb. On peut faire varier ce taux d'erreurs en jouant sur les concentrations en Mg^{2+} , Mn^{2+} ou sur le nombre de cycles PCR. Cependant, cette méthode comporte des biais. Par exemple, la probabilité d'apparition de deux ou trois substitutions adjacentes est très faible, et un acide aminé ne pourra être muté qu'en 6 ou 7 autres acides aminés par une seule substitution [60]. D'autres méthodes d'insertion de mutations aléatoires le long d'un gène existent (basées sur la méthode PCR ou non), et une analyse statistique utilisant trois indicateurs pour évaluer les conséquences du biais des mutations pour différentes techniques a été réalisée. Cette analyse montre que toutes les méthodes introduisent un certain biais et Wong *et al.* proposent d'utiliser la méthode MAP

(Mutagenesis Assistant Program) qui permet de comparer le biais introduit par 19 techniques différentes pour une séquence donnée, permettant ainsi de choisir la technique la plus adaptée pour un objectif donné [64]. Malgré cela, la technique la plus utilisée est celle de la PCR mutagénéisante et l'une de ses limitations comme le biais introduit au niveau de la diversité peut être dépassé par des combinaisons de techniques. Ainsi, l'ep-PCR est souvent associée à des techniques de recombinaison (shuffling).

Recombinaison homologue d'ADN in vitro

Les méthodes de recombinaison d'ADN permettent la combinaison directe de mutations bénéfiques portées par des gènes multiples tout en ôtant les mutations délétères, d'une manière analogue à la recombinaison sexuelle.

La première méthode de recombinaison appelée shuffling d'ADN a été développée par Stemmer en 1994. Des fragments d'ADN générés par la digestion aléatoire des gènes parents avec la *DNaseI* sont combinés et réassemblés en des gènes entiers chimériques par PCR (Figure 11) [65]. Cette méthode utilisant des gènes parents créés par ep-PCR partageant généralement 95-99% d'identité, est aussi appelée "family shuffling" lorsque les gènes d'une même famille sont recombinaisonnés. Les régions d'identité de séquence vont réaliser des crossing-overs qui recombinaisonnent les séquences des gènes parents. Ce protocole est ainsi limité à des protéines partageant au moins 70-75 % d'identité, les banques créées à partir de séquences plus divergentes tendant à donner pour la plupart des séquences mères. Il reste cependant la technique la plus populaire de recombinaison. De nombreuses méthodes ont été développées pour contourner ce problème comme le StEP (Staggered Extension Process), le RPR (Random Priming Recombination), le RACHITT (RANdom CHimeragenesis on Transient Templates) ou d'autres encore [33, 66]. Le RACHITT par exemple produit un taux de recombinaison plus élevé que les autres méthodes à partir de séquences de plus faible similarité [67]. Dans cette méthode, des fragments d'ADN simple brin sont hybridés à une séquence parente, synthétisée de manière à incorporer des uraciles et qui sert de matrice. Les séquences ne s'hybridant pas sont éliminées et les trous entre les fragments sont remplis, puis la matrice est détruite par une enzyme reconnaissant les uraciles (Figure 11). Ainsi, la nouvelle génération n'est pas contaminée par la séquence de départ et la diversité est plus grande.

Enfin, outre des gènes uniques, des opérons ou même des génomes entiers peuvent être recombinaisonnés pour certaines applications [33].

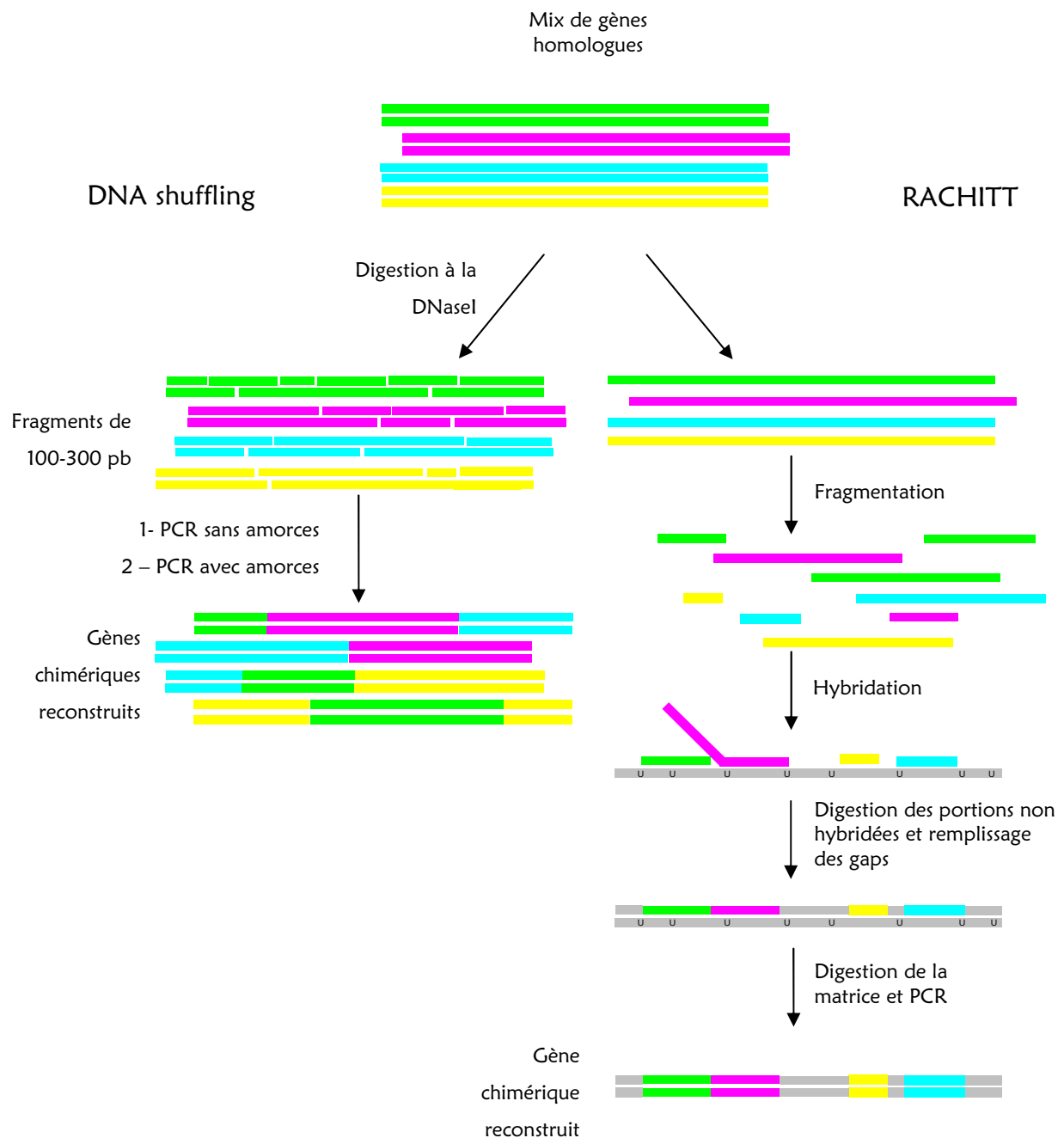


Figure 11 : shuffling de gènes homologues d'une même famille ou générés par PCR mutagénéisante. A gauche, principe du DNA shuffling, comme proposé par Stemmer [65]. Les gènes parents sont fragmentés à la DNaseI et les gènes reconstruits par deux PCR. A droite, principe du RACHITT. L'ADN simple brin des gènes homologues est fragmenté par la DNaseI et les fragments sont hybridés à un brin complémentaire comportant des uraciles. Les fragments chevauchants (non hybridés) sont éliminés et les trous entre les fragments remplis. Une PCR reconstruit l'ADN double brin.

Recombinaison non homologue d'ADN *in vitro*

La difficulté de générer des événements de recombinaison pour des séquences de faible similarité a mené au développement de techniques ne requérant pas d'hybridation à une matrice. En conséquence, alors que les variants protéiques générés par recombinaison homologue ou des mutations aléatoires vont maintenir une similarité structurale avec les protéines parentes avec une plus grande probabilité, ce type de méthode permet la création de nouveaux repliements protéiques.

Différentes techniques ont été développées, ITCHY (Incremental Truncation for the Creation of Hybrid Enzyme), NHR (Non Homologous Recombination), et plusieurs autres encore [33]. La première description d'une telle technique a été celle des banques ITCHY, créées par le clonage de deux gènes (ou fragments de gènes) en tandem dans un vecteur d'expression contenant deux sites uniques de restriction. À partir du vecteur linéarisé, des fragments tronqués sont générés du côté 5'. La génération de bouts francs et une restriction du côté 3' avec la deuxième enzyme de restriction génère des fragments de taille variable, et des chimères peuvent être générées par ligature pour circulariser le vecteur [68]. Le challenge majeur de cette stratégie est que de nombreuses protéines non fonctionnelles vont être générées, et une méthode de sélection à grand rendement doit pouvoir être utilisée.

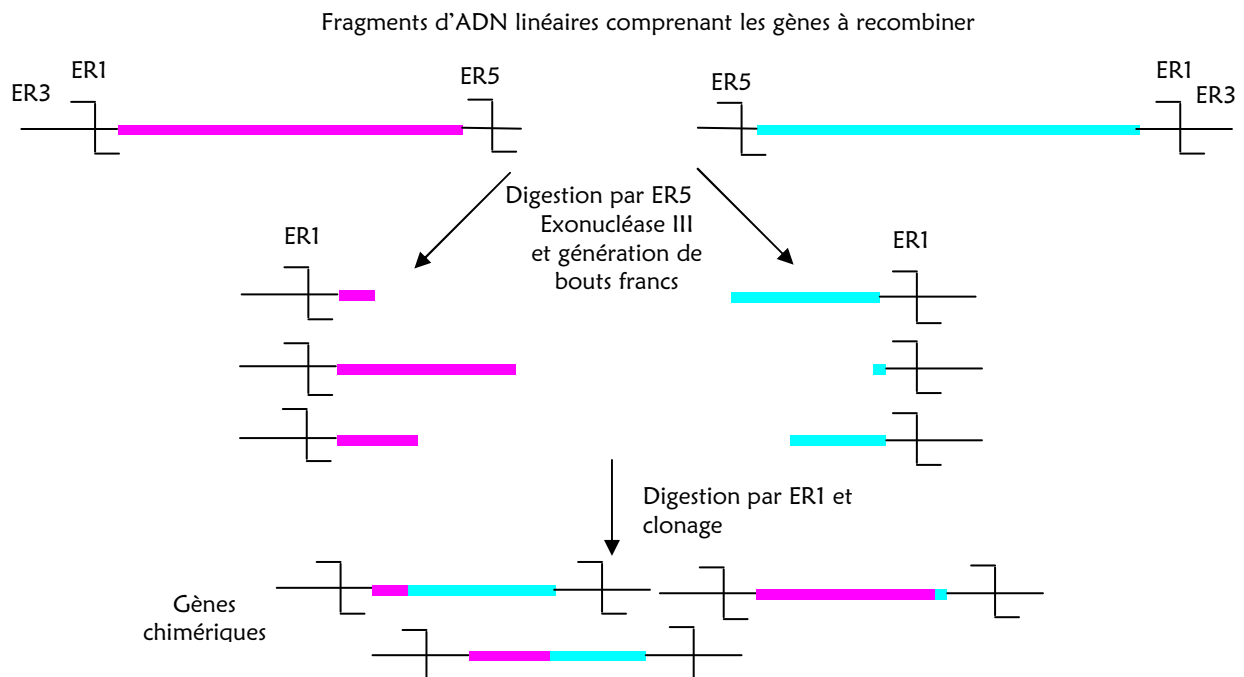


Figure 12 : technique de recombinaison non homologue ITCHY. Les gènes à recombiner sont portés par deux vecteurs. Après linéarisation, les gènes sont fragmentés de manière à obtenir toutes les troncations de 1 acide aminé possible. Les gènes sont ensuite clonés en tandem dans un vecteur d'expression.

Insertions et délétions de séquences

Une manière d'explorer l'espace de séquence peut se faire par des insertions et délétions de séquences. Des insertions/délétions (indels) sont faciles à générer par les techniques utilisant des oligonucléotides (voir page 34), cependant, la place de l'indel sera prédéterminée. La technique de PCR mutagénisante va introduire à faible fréquence des indels généralement d'un nucléotide, induisant des décalages du cadre de lecture. L'exploration de l'espace de séquences est donc limité et des techniques qui permettent l'introduction d'indels de différentes tailles à haute fréquence et de manière aléatoire sont nécessaires et malheureusement peu développées. Une de ces méthodes consiste en l'utilisation de transposons pour l'insertion aléatoire de séquences [69]. Cette technique requiert des constructions particulières d'ADN et est limitée au fait que seules des séquences spécifiques sont introduites. Il existe d'autres méthodes revues par Neylon,C, 2004 [60]. Une technique plus récente est celle du RAISE (RANdom Insertional-deletional Strand Exchange mutagenesis), basée sur la recombinaison de gènes. Le protocole du RAISE consiste en trois étapes. L'ADN est d'abord fragmenté par la *DNaseI*, puis, des nucléotides sont introduits aléatoirement du côté 3' des fragments par une "terminal deoxynucleotidyl transferase" (TdT), une ADN polymérase clé de l'introduction d'insertions aléatoires dans les régions hypervariables des anticorps mammifères. Finalement, ces fragments comportant une queue nucléotidique sont utilisés pour la reconstruction du gène entier, comme pour un shuffling classique. L'incorporation d'indels ou de substitutions serait réalisée suivant le schéma présenté en Figure 13.

Cette technique appliquée à l'évolution dirigée de la β -lactamase TEM-1 a permis après mutagenèse et sélection de trois générations de mutant sur des concentrations croissantes en ceftazidime, d'isoler des variants avec un MIC augmenté d'un facteur 10^4 , contenant des insertions, et des substitutions à des endroits connus pour changer la spécificité de substrat de la TEM-1, et seulement une délétion [70]. Cette méthode est prometteuse mais le taux de mutations efficaces reste faible : 0,2 acides aminés/kb d'ADN subissent des indels, et 0,6 acides aminé/kb d'ADN sont substitués.

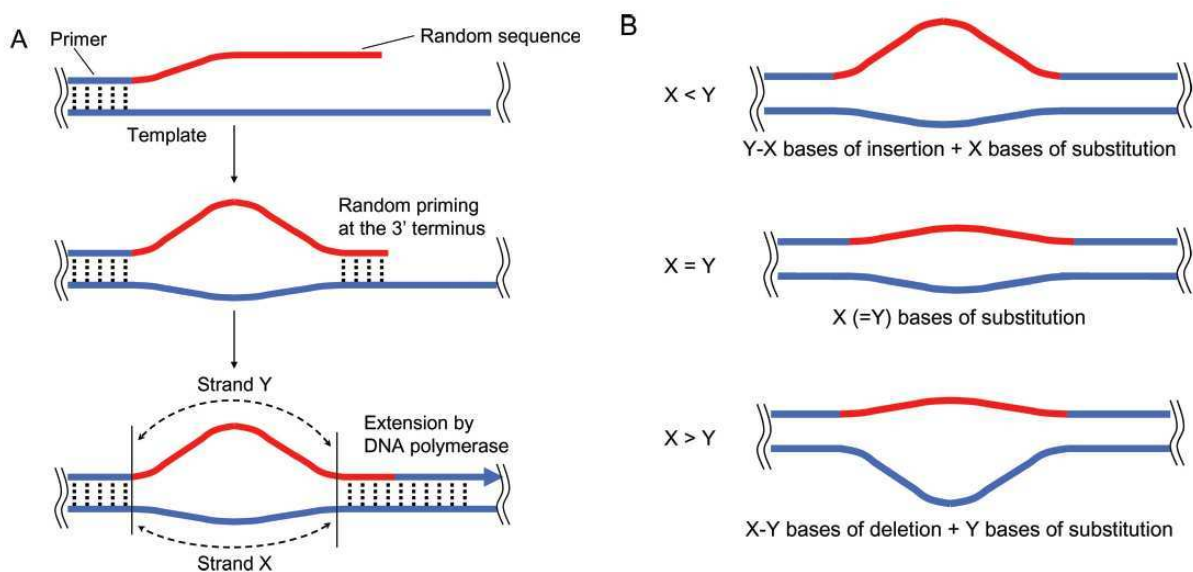


Figure 13 : explication pour le mécanisme de RAISE. A. Mécanisme possible pour l'introduction de mutations durant l'étape de reconstruction du gène. La région de séquence non mutée du fragment va s'hybrider à sa séquence complémentaire et une région de la séquence aléatoire introduite par la TdT va s'hybrider du côté 3' de manière aléatoire. L'ADN polymérase va alors étendre la séquence à partir de la région complémentaire, introduisant la séquence aléatoire dans le gène. B. Types de mutations introduites. Si le fragment muté s'hybride de manière à ce que la région aléatoire soit plus longue que la région complémentaire ($X < Y$), une insertion est introduite. Si le fragment muté s'hybride de manière à ce que la région aléatoire soit de la même taille que le fragment complémentaire ($X = Y$), une substitution est introduite. Si le fragment muté s'hybride de manière à ce que la région aléatoire soit plus courte ($X > Y$), une délétion est introduite [70].

Nous avons donné ici un petit aperçu des thèmes généraux pour les techniques existantes afin de générer des banques de mutants, la tendance à créer des techniques toujours plus compétitives très rapidement fait qu'il est difficile d'en faire une liste exhaustive. De plus, le choix de la technique la plus appropriée pour une application donnée est parfois difficile. Une revue récente par *Wong et al.* est intéressante pour comparer les différents choix possibles pour des expériences de mutagenèse aléatoire [71].

Enfin, la plupart des techniques introduisent un biais dans les mutations (comme le problème dû à la dégénérescence du code génétique, ou les problèmes inhérents à l'ep-PCR), ce point est revu et discuté par *Neylon, C (2004)* et *Lutz & Patrick, (2004)*, et quelques avancées technologiques capables de contourner certains problèmes sont données [60, 66].

Lien génotype – protéine

L'évolution dirigée consiste en des tours successifs de mutagenèse, criblage ou sélection et amplification afin d'isoler un variant d'une protéine, comportant des caractéristiques recherchées. Une considération pratique clé dans l'évolution dirigée de protéines fonctionnelles va être de pouvoir réaliser un lien entre le génotype et une biomolécule (protéine, peptide ou produit de réaction de l'enzyme active correspondante). Le génotype représente l'information génétique qui peut être répliquée. Un lien physique direct ou indirect entre la protéine et son gène est donc requis pour permettre l'identification de variants de cette protéine qui ont les caractéristiques fonctionnelles désirées. Les méthodes de liaison du génotype et des biomolécules peuvent être de différentes natures, les principaux modes sont présentés ici.

Liaison par compartimentalisation

Liaison de type cellulaire

La nature lie un gène et sa protéine en compartimentalisant les gènes dans des cellules. Ce compartiment dans lequel la fonction protéique désirée est exprimée est typiquement sélectionné *in vivo*. Le gène peut alors être récupéré du compartiment et analysé par séquençage de l'ADN. Ainsi, la manière la plus évidente de lier un gène à une protéine est le clonage des gènes d'une banque directement dans un microorganisme (bactérie ou levure) dans lequel chaque gène est transcrit et traduit *in vivo* (Figure 14, A). De cette manière, des variants intéressants de la biomolécule conférant par exemple un avantage de survie à la cellule vont pouvoir facilement être sélectionnés *in vivo* [72].

La présentation en surface de cellules est une autre méthode utilisant des cellules vivantes pour la liaison gène-protéine. Les protéines d'intérêt sont généralement fusionnées à une protéine membranaire de manière à ce qu'elles soient exposées en surface de la cellule (Figure 14, B). De cette manière, des microorganismes tels *E. coli* ont été utilisés, mais le plus intéressant est la levure *S. cerevisiae* qui est GRAS (Generally Recognised As Safe) et capable de réaliser des modifications post-traductionnelles comme des glycosylations, trouvant ainsi plus d'applications dans l'industrie [73]. Cette méthode est en réalité apparue après la méthode du phage display décrite ci-après. Outre la possibilité d'avoir des modifications post-traductionnelles, la présentation sur cellule

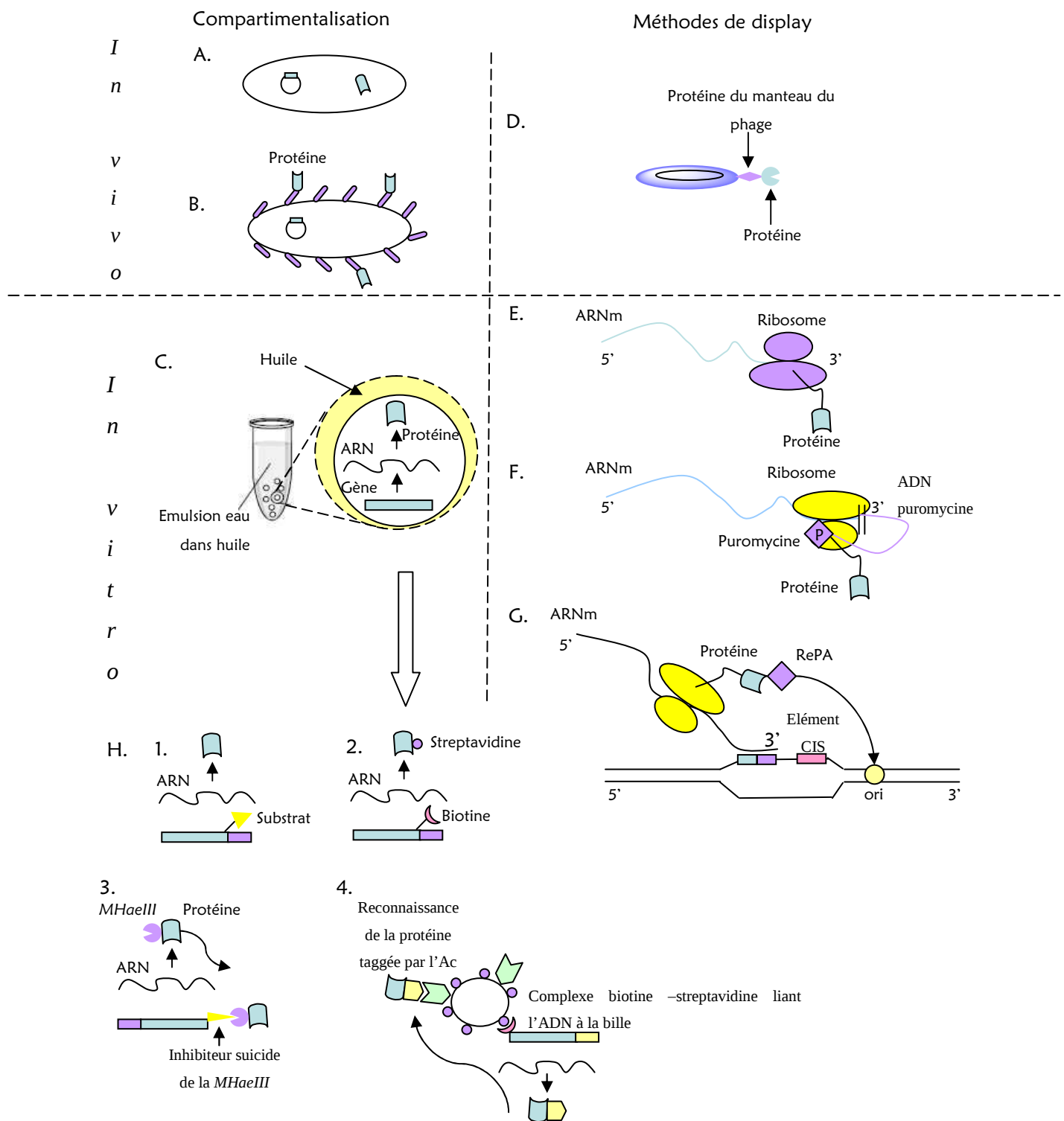


Figure 14 : stratégies pour la liaison gène-protéine par compartimentalisation : A. compartimentalisation naturelle, B. présentation en surface de cellule, C. Compartimentalisation *in vitro*. Liaison gène-protéine par display : D. phage display, E. présentation par ARNm, F. présentation par des ribosomes, G. par CIS "display". Différents types de lien gène-protéine sont représentés pour l'IVC (H), le substrat de la protéine est lié à l'ADN (1), le gène encode une protéine avec un tag streptavidine qui va lier l'ADN biotinylé qui l'encode (2), une liaison covalente est formée entre l'ADN et son produit d'expression via l'inhibiteur suicide la *mHaeIII* qui lui est fusionnée (3) et des billes recouvertes de streptavidine et d'anticorps vont permettre l'accrochage de l'ADN biotinylé qui encode la protéine fusionnée à un tag reconnu par l'Ac (4).

a l'avantage de pouvoir présenter de grosses protéines plus facilement, et en plus grand nombre. La limite de cette méthode se retrouve principalement dans la faible efficacité de transformation de *S. cerevisiae* qui atteint difficilement les 10^6 transformants par μg d'ADN [74].

Ce type de liaison va cependant rapidement rencontrer des limites comme l'impossibilité de réaliser des sélections dans des conditions différentes de l'environnement cellulaire, et des problèmes liés à la sélection de protéines toxiques. Les chercheurs vont alors se tourner vers des méthodes *in vitro*, contournant ces problèmes inhérents à la survie d'une cellule.

Compartmentalisation *in vitro*

La compartmentalisation *in vitro* (IVC : *in vitro* compartmentalisation) mime les compartiments cellulaires naturels en créant des microgouttelettes. Des émulsions d'eau dans l'huile avec des gouttelettes de $2\ \mu\text{m}$ de diamètre peuvent facilement être générées avec de l'huile, des détergents et un émulsifiant. Dans une procédure typique, les molécules d'acide nucléique d'une banque sont réparties dans des compartiments microscopiques de manière telle qu'en moyenne, il en existe une copie par goutte (Figure 14, C). En compartimentalisant également des mélanges de transcription-traduction *in vitro* (IVE : *in vitro* Expression), de multiples copies de la protéine d'intérêt sont synthétisées dans le compartiment. La petite taille des gouttelettes permet de traiter des banques de grande taille, par exemple, un mélange réactif de $50\ \mu\text{L}$ sera dispersé en 10^{10} compartiments [75].

"Phage display"

La présentation de biomolécules en surface de phages ou "phage display" est une méthode où des peptides ou des protéines sont exprimés en fusion avec une des protéines du manteau du virus (appelé aussi capsid). L'ADN encodant cette protéine de fusion étant encapsidé dans la structure protéique du manteau, le lien physique gène-protéine est assuré (Figure 14, D). Le phage display, depuis sa description en 1985 par Smith [76], est une des méthodes les plus largement utilisées pour effectuer des sélections de mutants intéressants parmi des banques. Cette technique se trouve à mi-chemin de l'*in vivo* et de l'*in vitro*. En effet, l'amplification de phages permettant la production de la protéine de fusion se fera *in vivo* alors que les sélections se feront avec des solutions de phages *in vitro*. Cette technique étant celle utilisée dans ce travail, elle est exposée plus en détails. Différentes particules virales peuvent être utilisées pour le phage display, comme les phages filamenteux (Ff), le bactériophage lambda ou encore

des virus eucaryotes. Le système le plus représenté et que nous avons utilisé est celui du phage filamenteux M13, capable d'infecter *E. coli*.

Structure du bactériophage M13

Un phage filamenteux est composé d'une molécule d'ADN simple brin circulaire contenue dans une structure protéique, et a une longueur de 900 nm pour environ 7 nm de large. La capside protéique est constituée de 2700 copies de la protéine majeure du manteau, pVIII. Chaque extrémité du phage est coiffée par deux protéines mineures du manteau : d'un côté, on trouve 5 copies de chaque protéine pVII et pIX et de l'autre, 3 à 5 copies de chaque protéine pIII et pVI. La longueur de la particule virale peut s'adapter à des manipulations de la taille du génome : plus le génome est grand, plus le phage est grand. Jusqu'à 12 000 bases peuvent être ajoutées au génome du phage sauvage (6407 bases) sans perturbation de son empaquetage (Figure 15) [77].

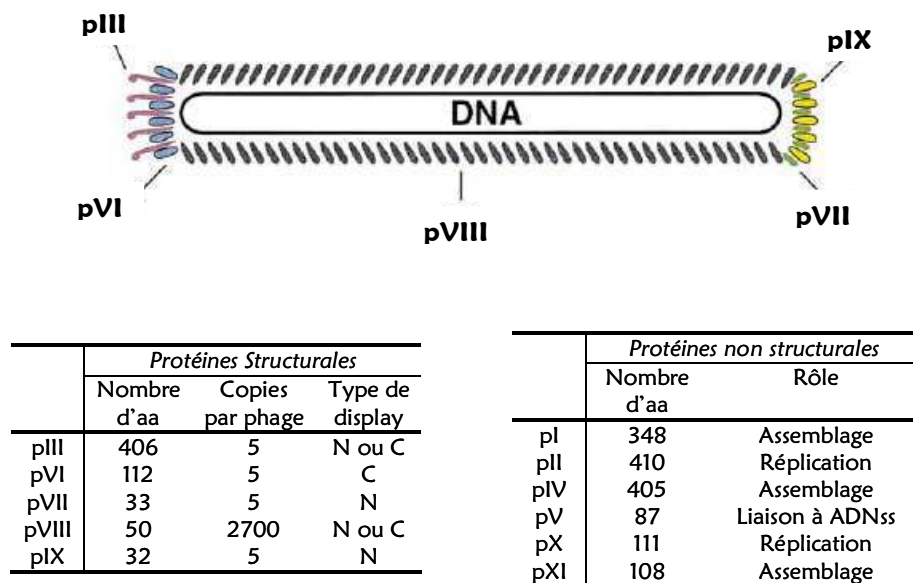


Figure 15 : représentation schématique d'un phage filamenteux. M13 produit 5 protéines de structure et 6 protéines qui ne font pas partie de la particule virale. La partie C-terminale de la pVI est exposée en surface du phage, et les autres protéines exposent leur partie N-terminale.

Les cinq protéines du phage contribuent à la stabilité structurale de la particule phagique, mais la protéine pIII est aussi nécessaire pour la reconnaissance de la cellule hôte et pour l'infection. Cette protéine est constituée de trois domaines distincts. Le domaine N-terminal (N1) initie la translocation de l'ADN dans *E. coli* pendant l'infection alors que le second domaine (N2) confère la reconnaissance de l'hôte en se liant au pilus F à la surface de *E. coli*.

Le domaine C-terminal (C) interagit avec les autres protéines du manteau et est ainsi responsable de l'intégration de la pIII dans la capsid [78].

Assemblage du phage

L'infection d'*E. coli* par un phage filamenteux est initiée par l'interaction spécifique de l'extrémité du phage où est située la pIII avec la pointe du pilus F. Le pilus consiste en un tube protéique assemblé et désassemblé dans un processus de polymérisation et dépolymérisation à partir de sous-unités de piline présentes dans la membrane interne de la bactérie. Dans le cas d'infection par un phage, il est suggéré que l'extrémité du phage attachée au pilus est amenée à la surface de la membrane par dépolymérisation du pilus F. Les protéines majeures du manteau protéique sont alors intégrées dans la membrane et l'ADN du phage est transloqué dans le cytoplasme. Dans le cytoplasme, les enzymes bactériennes synthétisent le brin complémentaire (brin-) et convertissent l'ADN simple brin (brin+) du phage en une molécule répliquative double brin superenroulée. Cette molécule sert de matrice pour la transcription et la traduction à partir desquelles les 11 protéines du phage (structurales, impliquées dans la répliquaison de l'ADN ou l'assemblage), sont synthétisées (Figure 16). Certains produits du phage (pII et pX) dirigent la synthèse d'ADN simple brin selon un processus de "rolling circle", de concert avec des enzymes bactériennes. Les protéines de la capsid et les autres protéines du phage impliquées dans l'assemblage de la particule (pI, pIV et pXI) sont intégrées dans la membrane avec leur portion N-terminale exposée vers le périplasme et leur portion C-terminale exposée vers le cytoplasme, alors que les protéines impliquées dans la répliquaison de l'ADN restent dans le cytoplasme. Lorsque la protéine pV spécifique de la liaison à l'ADN simple brin atteint une concentration donnée, elle séquestre l'ADN simple brin nouvellement synthétisé dans un complexe, et les nouvelles particules phagiques sont assemblées et extrudées de la bactérie via un pore formé par certaines protéines du phage. L'assemblage est initié par l'incorporation des pVII et pIX à une extrémité de la particule. Puis vient un processus d'élongation où les molécules pVIII sont assemblées le long de la particule et le processus se termine avec l'addition des pVI et pIII à l'autre extrémité du phage et enfin, le relargage du phage. Le phage M13 est un phage non lytique, il bourgeonne de la bactérie hôte qui n'est pas tuée par l'infection. Les bactéries infectées continuent à se multiplier et se diviser indéfiniment, cependant à une vitesse inférieure à celle d'une bactérie non infectée [77-79].

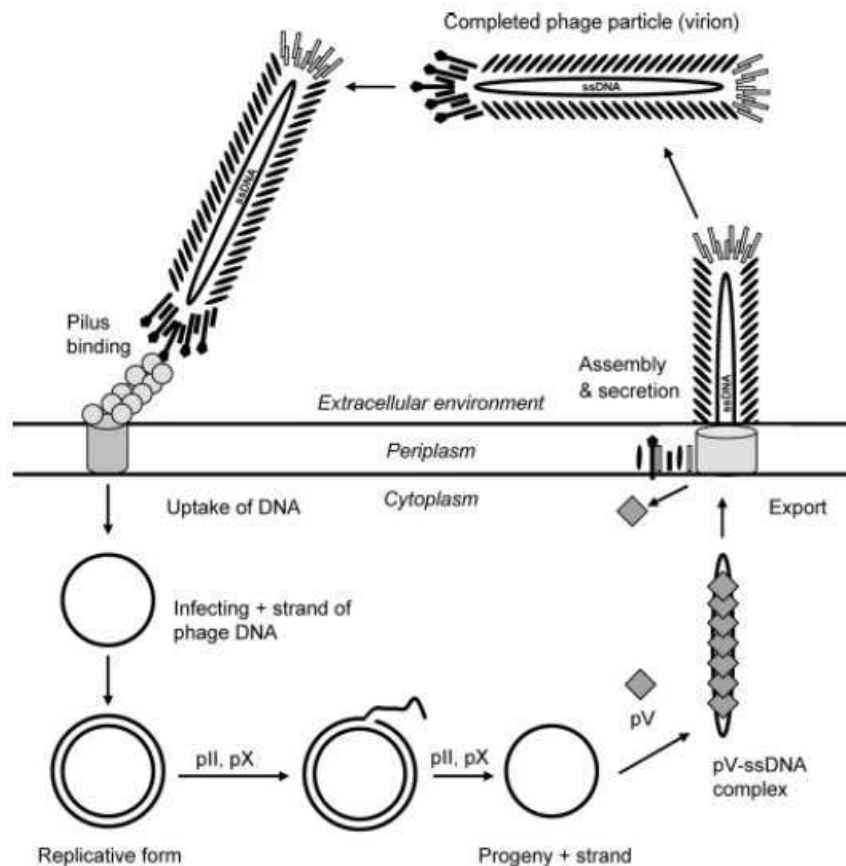


Figure 16 : assemblage du phage M13. Les particules du phage se lient à *E. coli* par le pilus encodé par les gènes portés sur le facteur F. Après liaison au phage, le pilus se rétracte, amenant la particule phagique en surface de la cellule. L'ADN simple brin (+) entre dans la bactérie par un processus mal compris, et est converti en ADN double brin (forme répliquative) par les protéines d'*E. coli*. Les protéines pII et pX sont impliquées dans l'amplification en "rolling circle" de l'ADN double brin, qui mène à la production de brins+. La protéine pV recouvre ces molécules et ce complexe protéine-ADN interagit avec la machinerie d'exportation (pI, pIV, pXI, thiorédoxine d'*E. coli*) qui assemble et secrète simultanément les particules virales en utilisant un pool de protéines du manteau préalablement insérées dans la membrane interne. Les particules virales sont finalement relarguées des cellules sans les tuer et les pV sont recyclées. D'après [79].

Présentation en surface de phage

Les cinq protéines de la capside du phage ont été utilisées pour présenter des protéines. L'approche la plus commune pour le "display" de peptides est de fusionner les séquences peptidiques du côté N-terminal des protéines pIII ou pVIII, alors que des protéines vont plutôt être présentées sur la pIII. Des fusions de peptides et de protéines ont aussi été réalisées du côté N-terminal de la pVII et la pIX, et des fusions du côté C-terminal des pVI, pVIII et pIII. D'une manière générale, les protéines mineures du manteau vont présenter les grosses protéines de manière plus efficace que la pVIII. Cependant l'efficacité de "display" de protéines est très imprévisible et, alors que certains petits peptides ne pourront être présentés, de grosses protéines telles la pénicilline G acylase (un hétérodimère de 86 kD) peuvent être présentées aussi bien sur la pVIII que la pIII. Certaines séquences ne sont pas bien présentées sur phages à cause de problèmes d'assemblage de la particule virale, de stabilité et/ou

d'infectivité. Cette intolérance n'est pas bien comprise, et il existe deux systèmes complémentaires au phage display qui utilisent des virus hybrides et des phagemides, et pour lesquels les problèmes d'intolérance sont plus limités (Figure 17).

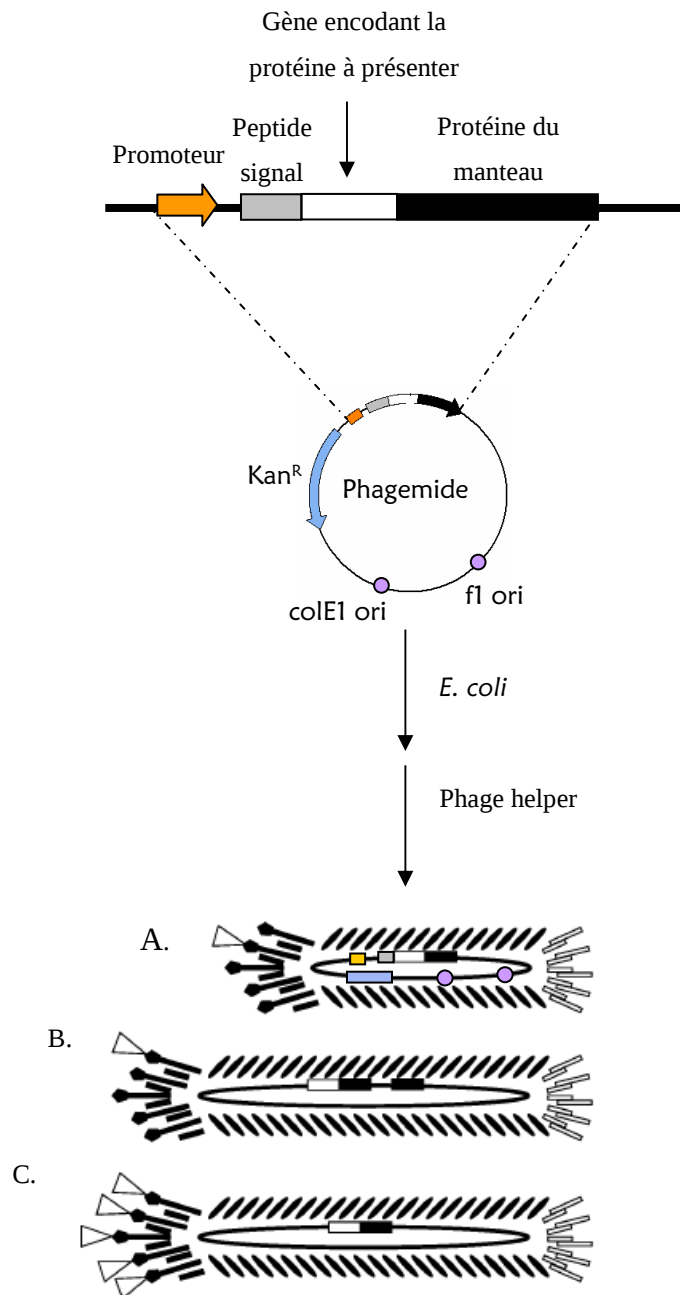


Figure 17 : représentation schématique des trois types de phages présentant une protéine sur la pIII. A. Phagemide, les éléments le composant et les différentes étapes nécessaires à l'infection sont représentés. On trouve en moyenne une copie de la protéine de fusion. f1 = origine de réplication phagique, colE1 = origine de réplication bactérienne. B. Phage hybride comportant deux copies de la protéine de manteau, dont l'une est fusionnée à la protéine d'intérêt. C. Phage classique, présentant jusqu'à 5 copies de la protéine de fusion.

Les systèmes phagiques sont construits de telle manière que la protéine ou le peptide en fusion est présent sur chaque copie de son partenaire. Ainsi, si une banque est présentée sur la

pVIII, le virus présentera 2700 copies de la protéine de fusion. Ceci est accompli en fusionnant la banque d'ADN avec la seule copie de pVIII du génome phagique. Des banques de peptides peuvent être présentées sur la pVIII avec un système phage, mais la taille du peptide est limitée et il existe un phénomène d'avidité² qui peut poser des problèmes lors de sélection pour affinité³. Ce type de sélection va en effet utiliser la capacité de la protéine sur phage de se lier à un ligand d'intérêt immobilisé sur un support solide (Figure 18).

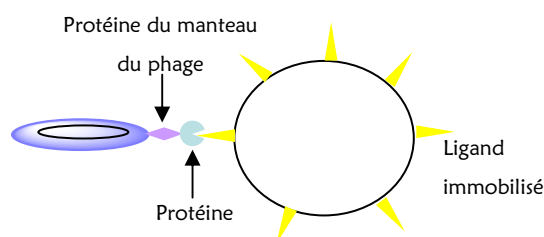


Figure 18 : sélection d'un phage par affinité pour un ligand immobilisé.

Les phages accrochés sont généralement élués par compétition avec le ligand en solution. De nombreuses copies de la protéine sur phage vont ainsi pouvoir mener à la sélection de protéines en fonction de leur avidité. Cette même approche a été réalisée avec la pIII, diminuant l'effet d'avidité. Cependant, même la présentation sur la pIII est limitée, des altérations significatives de la pIII empêchent l'assemblage du phage ou bloquent son infectivité. Enfin, toutes les protéines ne sont pas des protéines de fusion à cause de problèmes de protéolyse de la protéine présentée.

Le deuxième système utilisé est le système phagemide. Un phagemide est un plasmide qui porte un gène de résistance, deux origines de répllication l'une phagique et l'autre bactérienne, et un signal d'emballage du phage. Dans le cadre du phage display, la construction comportera le gène de la protéine d'intérêt fusionné à un gène d'une protéine de la capsid. La banque est clonée dans le phagemide de manière à ce qu'elle soit exprimée en fusion avec la protéine du phage (Figure 17). L'ADN du phagemide est alors utilisé pour transformer des bactéries et peut être amplifié et isolé de la même manière que n'importe quel plasmide. Afin de présenter des peptides ou des protéines en surface de phagemides, la bactérie portant l'ADN du phagemide doit être infectée par un phage assistant ou phage "helper" pour

² On parle d'avidité pour exprimer la force de liaison entre une protéine possédant plusieurs sites de liaison à un ligand multivalent. Dans le cas du phage display, l'expression de plusieurs protéines en surface du même phage pourra donner un effet d'avidité vis-à-vis du substrat qui est généralement représenté de nombreuses fois sur le support de capture.

³ L'affinité va exprimer la force de liaison simple entre une protéine et son ligand.

produire les particules virales utiles. L'approche la plus simple est d'utiliser un phage "helper" défectueux pour sa fonction d'empaquetage. Durant l'infection de la bactérie portant le phagemide, le génome du phage helper initie la synthèse de toutes les protéines sauvages. Comme l'ADN du phagemide porte un promoteur en amont de la séquence encodant la protéine de fusion, la synthèse de la protéine de fusion est elle aussi initiée et elle est ajoutée au pool cellulaire de protéines phagiques. Lors de l'assemblage du phage, les particules phagiques seront préférentiellement assemblées en utilisant l'ADN du phagemide qui diffère du phage "helper" en ce qu'il comporte un signal d'empaquetage totalement fonctionnel. Les phages sécrétés seront composés de protéines sauvages et de fusion et porteront l'ADN phagemidique encodant cette séquence. D'autres types de phage "helper" ont été utilisés afin de réduire les copies sauvages de protéine de la capside par rapport aux protéines de fusion. Enfin, il a été observé que les domaines protéiques sont plus accessibles lorsqu'ils sont fusionnés au domaine N2 de la pIII. Ainsi, une pIII tronquée est parfois utilisée, son domaine N1 (aa 1-198) a été délété et la protéine d'intérêt est fusionnée à la partie de la pIII composée des acides aminés 198 à 406.

Les systèmes hybrides ne requièrent pas de phage helper mais sont aussi composés de protéines de manteau sauvage et de fusion. Pour cela, une deuxième copie du gène pVIII ou pIII est placée dans le génome du phage [77-79].

Le phage display repose sur l'expression des gènes *in vivo*, et la taille de la banque est limitée par l'efficacité de transformation et les contraintes biologiques de la cellule hôte. Cette limitation de diversité et de complexité des banques peut être dépassée en utilisant des systèmes totalement *in vitro* basés sur un système transcription-traduction ne nécessitant pas de passer par un microorganisme.

Méthodes basées sur des systèmes d'IVE

Les méthodes de display

Le ribosome "display", l'ARNm "display" et le cis "display" sont quelques unes des méthodes développées qui ne nécessitent aucune étape *in vivo*. Dans ces méthodes, un polypeptide transcrit et traduit *in vitro* est capable de se lier à l'ARNm ou ADN l'encodant.

Le ribosome display assure une liaison de type non covalente entre un gène et sa protéine via le ribosome, en conservant le complexe ternaire formé par l'ARNm, le ribosome et la chaîne polypeptidique naissante pendant la traduction (Figure 14, E). En l'absence de codon stop, le ribosome s'arrête à la fin de la matrice ARNm. De basses températures et de hautes

concentrations en Mg^{2+} stabilisent ce complexe qui peut directement être utilisé pour des sélections [80].

L'ARNm display va permettre d'établir un lien covalent entre le génotype et le phénotype en liant l'ARNm et le polypeptide encodé par cet ARNm par une petite molécule adaptatrice, la puromycine (Figure 14, F). La transcription de l'ADN est suivie d'une étape chimique de couplage d'un complexe ADN-puromycine, à l'extrémité 3' de l'ARNm dans le ribosome, via une liaison peptidique. Comme le complexe ARNm-protéine se fait entièrement *in vitro*, de grandes banques peuvent être construites, et comprennent jusqu'à 10^{12} - 10^{13} séquences uniques [80]. Cependant, ces techniques présentent plusieurs limitations. L'ARNm utilisé est généralement sujet à une dégradation rapide, et l'ARNm sélectionné doit être soumis à une transcription inverse avant d'être amplifié. Enfin, un autre problème est lié au ribosome. *In vitro*, la traduction d'une protéine entière est difficile à réaliser et une certaine fraction des molécules seront tronquées. Certains de ces problèmes peuvent être évités en liant la protéine d'intérêt à l'ADN.

Le CIS display comme le ribosome display est une technique qui va permettre de lier un gène à sa protéine de façon non covalente. Ce système exploite l'activité CIS haute fidélité d'un groupe de protéines bactériennes d'initiation de la réplication d'ADN, les RepA du plasmide R1. Dans ce contexte, l'activité CIS fait référence à la propriété des RepA de se lier exclusivement à l'ADN à partir duquel elles ont été exprimées, dans un contexte de transcription/traduction couplées. La réplication du plasmide R1 est initiée par la liaison de RepA à l'origine de réplication du plasmide (ori). L'ori est séparée de la séquence encodant la protéine RepA par un élément d'ADN appelé CIS. Le modèle consensus pour l'activité de RepA est que l'élément CIS contient un terminateur de transcription ρ -dépendant qui provoque l'arrêt de l'ARN polymérase. Cet arrêt laisse l'opportunité au polypeptide naissant de RepA qui émerge du ribosome la traduisant de se lier transitoirement à l'élément CIS, qui en retour, dirige RepA vers l'ori. Ainsi, en fusionnant des banques de peptides du côté N-terminal de la protéine RepA, un lien direct peut se faire entre les peptides et la molécule d'ADN qui les encode (Figure 14, G) [81].

Stratégies basées sur l'IVC

Ces stratégies sont généralement utilisées pour sélectionner une activité enzymatique. Il existe différentes méthodes qui vont différer dans la manière dont l'ADN et la protéine sont connectés à l'intérieur d'une microgouttelette. Ces méthodes vont dépendre des conditions de sélection ou criblage. Typiquement, l'IVC va permettre la traduction et transcription d'une protéine dans une microgouttelette. Il va alors s'agir de capturer la protéine d'intérêt catalysant une réaction donnée, et la liaison gène protéine devra être maintenue de manière directe ou indirecte pendant tout le mode opératoire.

Une gouttelette unique peut contenir un gène encodant une protéine qui agit sur un substrat connecté ou identique à l'ADN. Les émulsions sont détruites et le produit couplé à (identique à) l'ADN encodant la protéine d'intérêt va assurer un lien indirect entre la protéine et le génotype (Figure 14, C-1) [75]. La capture spécifique du produit permet de récupérer l'information génétique intéressante.

Si les gouttelettes sont utilisées dans une telle procédure en une étape, les sélections ultérieures sont limitées aux conditions imposées par l'IVE. Si la synthèse protéique est découplée de la sélection, la gamme de conditions va pouvoir varier. Dans ce cas, une première émulsion va permettre l'IVE, cette émulsion est détruite, et une deuxième émulsion est constituée pour l'étape de sélection ou criblage. Le point délicat est alors de garder la protéine connectée à l'ADN après l'IVE, dans des microgouttelettes. Trois approches ont été mises en place. La première utilise un ADN biotinylé encodant une protéine fusionnée à la streptavidine (Figure 14, C-2). Durant l'IVE, la protéine se connecte à l'ADN qui l'encode, la destruction de cette émulsion pour les étapes ultérieures n'altère pas le lien entre le gène et sa protéine. La deuxième méthode repose sur la formation d'un lien covalent entre l'ADN et la protéine en fusionnant l'ADN de la protéine d'intérêt au gène codant pour l'ADN méthyltransférase *M.HaeIII*, via un inhibiteur suicide de la *M. HaeIII* placé du côté 3' de la molécule d'ADN (Figure 14, C-3). Dans ce cas aussi, l'émulsion peut être détruite tout en gardant le lien direct génotype-protéine. Enfin, dans le système de "display" de microbilles, l'émulsion pour l'IVE contiendra tout le mélange transcription traduction et des billes recouvertes de streptavidine auxquelles l'ADN biotinylé va se lier. L'IVE est réalisée dans des gouttelettes et un anticorps biotinylé retient la protéine sur la même bille via un tag ayant de l'affinité pour l'anticorps (Figure 14, C-4) [82]. L'émulsion peut alors être détruite, le lien gène protéine étant assuré via la bille.

Criblage et sélections

Les expériences d'évolution dirigée s'effectuent généralement à partir de grandes collections de variants d'une protéine ou d'un peptide. Afin d'isoler et d'identifier/caractériser le peptide/la protéine d'intérêt, des méthodes de sélection ou de criblage qui dépendent du système de lien génotype-protéine créé vont être utilisées. Alors qu'une méthode de criblage nécessite l'analyse de chacun des clones ou gènes individuels, une méthode de sélection va permettre d'isoler les clones d'intérêt parmi un pool entier de variants, sans devoir tous les analyser. Malgré cette différence d'approche, ces méthodes suivent des principes similaires. Le premier est qu'il est nécessaire dans les limites du possible de sélectionner/cribler directement pour la propriété d'intérêt, l'adage de l'évolution dirigée étant : "on obtient ce pour quoi on sélectionne" ("You get what you select for") [83]. Ainsi, dans l'évolution de catalyseurs enzymatiques par exemple, le substrat doit être identique ou le plus proche possible du substrat cible, et le produit doit pouvoir être détecté sous des conditions de multiples "turnovers" afin d'assurer la sélection de bons catalyseurs. Deuxièmement, le test doit être assez sensible. Les premiers tours d'évolution demandent d'isoler tous les variants améliorés, y compris ceux qui ne montrent qu'une amélioration quelques fois supérieure à celle de la protéine sauvage. Les tours les plus avancés se feront dans des conditions plus strictes afin d'assurer l'isolement des meilleurs variants. Enfin, la procédure doit être si possible applicable à un grand nombre de mutants (10^4 à 10^{13} variants), on parle alors de système "High Throughput" (HT). Cependant, alors que les méthodes de création de diversité sont devenues communes, les méthodes de criblage ou de sélection doivent souvent être ajustées pour chaque propriété recherchée.

Criblage

Les méthodes couramment utilisées à des fins de criblage sont l'utilisation de tests mettant en jeu des substrats fluorogéniques ou chromogéniques sur boîtes de Pétri, plaques de microtitrage, puces ou en cytométrie de flux associée au FACS (Fluorescence Activated Cell Sorting) [84-86]. En l'absence de tels substrats, il est possible de suivre une activité enzymatique par des tests radiométriques, les activités de type kinase par exemple sont facilement suivies en utilisant de l'ATP radioactif. Cependant, pour les applications qui demandent une caractérisation du substrat, des techniques analytiques classiques de chromatographie en phase gazeuse, spectrométrie de masse ou RMN (Résonance Magnétique Nucléaire) pourront être utilisées. De manière plus sophistiquée, des techniques de chromatographie en phase liquide et spectrométrie de masse (LC/MS) sont développées en

format HT [84]. Le choix de la technique utilisée dépendra de l'activité à cribler mais aussi des limites de temps pour analyser une collection de mutants. Dans une gamme limite de 10^4 à 10^7 mutants par exemple, le matériel et temps requis pour le criblage en utilisant différentes méthodes sont les suivants :

Nombre de variants criblé	No de plaques 96 puits	No de plaques 384 puits	No de boîtes de Petri ^a	Heures de tri au FACS
10^4	104	26	50	0,001
10^5	1042	260	500	0,01
10^6	10 417	2 604	5 000	0,1
10^7	104 167	26 042	50 000	1

^a Estimé pour une densité de 200 colonies par boîte. No = Nombre. Informations tirées de [87].

Ainsi, la méthode la plus classique d'étalement sur boîtes de Pétri va être limitée au criblage de 10^4 - 10^5 colonies, et par la sensibilité du test. De même, à moins d'être automatisées, les méthodes utilisant des plaques de microtitrage souvent associées à de l'immunodétection vont être généralement limitées autour de 10^3 variants. Enfin, même si les technologies HT basées sur la robotique pour l'analyse de microplaques ou de puces vont permettre d'étendre la gamme de variants criblés, elle dépasse rarement 10^6 clones pour le criblage de biocatalyseurs, et les chercheurs se tournent de plus en plus vers les techniques utilisant la cytométrie de flux associée au FACS, qui permettent de cribler 10^7 variants en 1 heure.

Sélection

Des méthodes de sélection ne sont pas limitées par la taille de la collection et vont d'une manière générale permettre l'analyse simultanée d'un plus grand nombre de variants que les méthodes classiques de criblage (hors FACS). Les méthodes de sélection *in vivo* d'étalement de bactéries vont être limitées car elles reposent sur l'existence d'un avantage sélectif donné par la protéine d'intérêt. De plus, les conditions de sélection vont être restreintes aux conditions de croissance du microorganisme utilisé. Aussi, des protéines toxiques pour le microorganisme ne pourront pas être testées [86]. Ainsi, les chercheurs ont développé des méthodes de sélection *in vitro* basées sur les liens gène-protéine décrits ci-dessus. Les avantages potentiels de ces techniques sont que de plus grandes collections vont pouvoir être analysées, moins de biais dû à l'expression cellulaire est introduit, et pour certaines d'entre elles, l'absence d'une étape de transformation pour la construction de la banque va permettre la création de plus grandes collections de variants. Enfin, les méthodes de criblage

automatisées nécessitent un matériel coûteux alors que les techniques de sélection ne requièrent généralement pas de matériel particulier. Sélection/Criblage pour activité catalytique.

D'une manière générale, des sélections *in vitro* vont être appropriées pour isoler des protéines ayant des propriétés de liaison à un ligand particulier ou pour améliorer des propriétés protéiques telle leur stabilité [88]. Dans ce cadre, le phage display ou l'ARNm display ont notamment prouvé leur efficacité pour sélectionner des anticorps [89]. Pour isoler une enzyme sur base de son activité catalytique, des méthodes de sélection *in vivo* ou de criblage *in vitro* vont être utilisées, ceci reste cependant une tâche difficile et les améliorations catalytiques apportées restent souvent limitées.

In vivo

Le criblage sur boîtes de Pétri utilisant des substrats chromogéniques ou fluorogéniques, les sélections sur boîte contenant des concentrations croissantes en antibiotiques et des expériences de complémentations avec des bactéries auxotrophes ont permis d'isoler dans de nombreux cas des variants intéressants d'enzymes [90]. Cependant, toutes les activités enzymatiques ne peuvent pas être couplées à un avantage sélectif ou des substrats chromogéniques/fluorogéniques ne sont pas toujours disponibles ou enfin, les tests enzymatiques ne sont pas assez sensibles.

Une méthode appelée complémentation chimique a récemment été développée, permettant de détecter une activité enzymatique via un gène rapporteur. Le produit de catalyse de l'enzyme est lié à la transcription d'un gène essentiel via un système de levure triple hybride. Le substrat de l'enzyme agit comme inducteur de la dimérisation pour la reconstruction des domaines de liaison à l'ADN et d'activation de la transcription d'un facteur de transcription artificiel. La réaction enzymatique qui peut-être d'hydrolyse ou de synthèse peut ainsi résulter en l'activation ou la répression d'un gène rapporteur essentiel ou toxique [91]. Ce système, est cependant contrôlé par un lien chimique entre deux petites molécules ligand pour lequel une chimie plutôt complexe est requise. Un autre groupe a développé une technique plus simple, basée sur le même type de schéma. Van Sint Fiet *et al.* ont développé des méthodes où le produit de catalyse peut se lier à une protéine qui à travers la formation de ce complexe va pouvoir constituer un activateur de transcription pour un gène rapporteur ou de sélection [92]. De cette manière, des protéines ne comportant pas de caractéristique phénotypique leur permettant d'être sélectionnées ou criblées, vont pouvoir l'être par l'intermédiaire du

rapporteur. Cependant, ces techniques ne résolvent pas les problèmes liés à l'expression de protéines toxiques, et la taille de la banque pouvant être analysée est limitée.

In vitro

Des méthodes de liaison gène-protéine citées ci-dessus, seules les techniques de phage display et les techniques associées à la cytométrie de flux et au FACS, comme l'IVC ont permis d'isoler des variants d'enzymes pour leur activité.

Le phage display est une des premières techniques *in vitro* ayant permis de sélectionner de meilleurs catalyseurs enzymatiques. Il existe des méthodes de sélection directes qui utilisent les substrats et des méthodes de sélection indirectes qui utilisent des inhibiteurs ou des analogues d'état de transition (TSA pour Transition State Analog) [93]. Dans tous les cas, ce type de sélection passe par une étape de capture où les phages présentant l'activité recherchée sont ou deviennent capables de se lier à une molécule particulière. Ils peuvent ainsi être capturés spécifiquement sur un support où cette molécule particulière est immobilisée, puis élués et amplifiés par infection bactérienne pour entrer dans un nouveau tour de sélection (Figure 19, A).

Des méthodes de sélections indirectes vont utiliser des TSAs qui miment la géométrie et la distribution de charge des états de transition. Parce que le pouvoir catalytique des enzymes peut être défini comme une affinité pour un état de transition, les sélections pour une liaison à un TSA devraient mener à la sélection de catalyseurs. Des sélections pour liaison à des TSAs ont cependant donné des résultats plutôt mitigés et dans la plupart des cas, l'enzyme sélectionnée a une meilleure affinité pour le TSA mais l'activité spécifique est diminuée. Bien que certains anticorps catalytiques aient pu être sélectionnés de cette manière, l'activité est généralement bien plus faible que celle d'une enzyme naturelle [93].

Des inhibiteurs suicides qui bloquent irréversiblement une enzyme en utilisant le mécanisme catalytique peuvent aussi être utilisés lors de sélections indirectes. L'incubation d'une banque de variants de la β -lactamase sur phage mutée dans une boucle proche du site actif, avec le substrat naturel de l'enzyme, bloquant les enzymes inactives puis avec un inhibiteur suicide biotinylé, a permis la capture des phages-enzymes ayant gardé une certaine activité [94]. Une des limites de cette méthode est que son succès dépend de la vitesse de marquage avec l'inhibiteur et de la vitesse de turnover du substrat. De plus, il n'existe pas d'inhibiteurs suicides pour toutes les classes d'enzymes.

Des sélections directes peuvent se faire si le substrat est accroché au phage et que la réaction produit un élément capable d'être capturé sélectivement. Cette stratégie a été appliquée à une

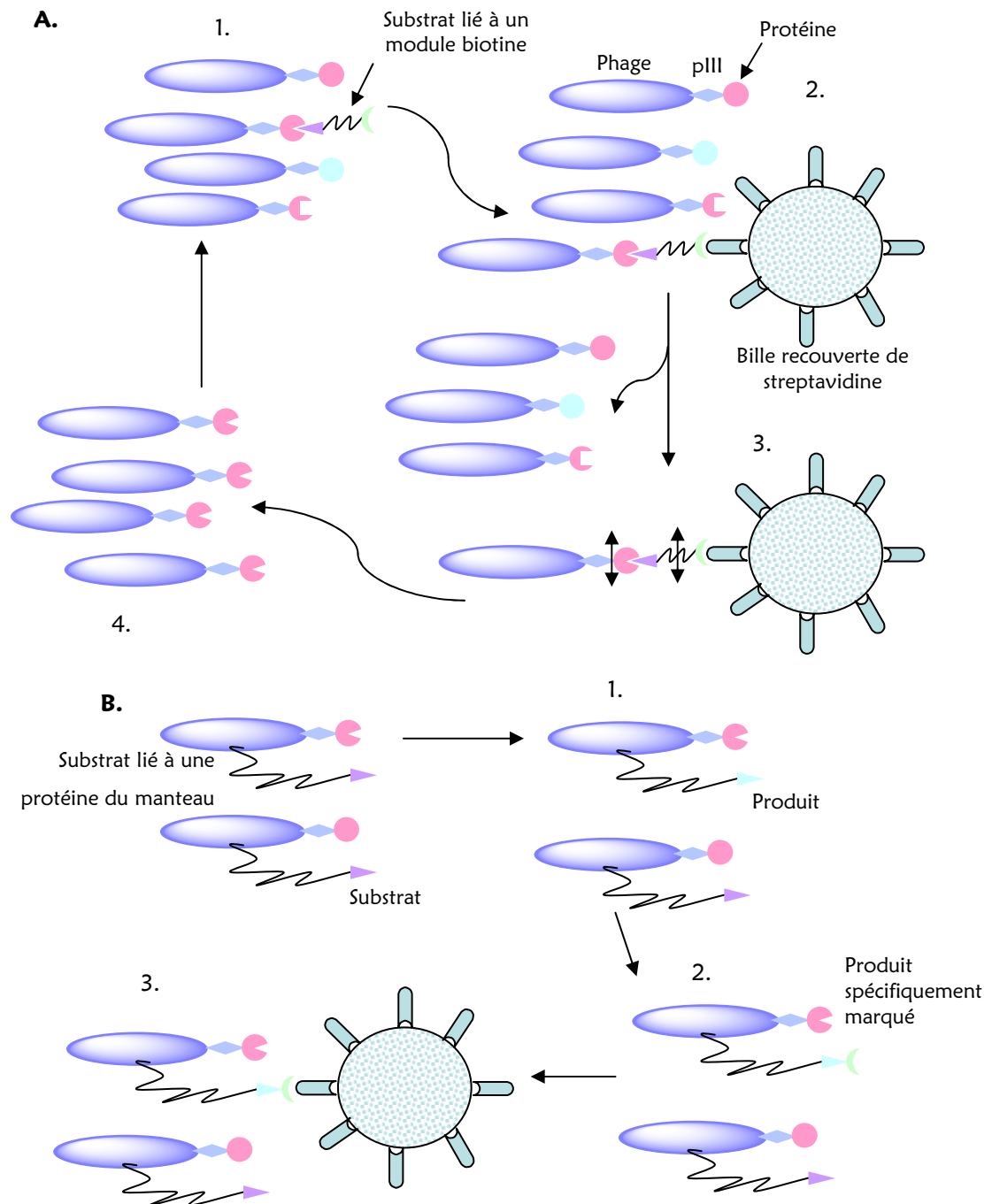


Figure 19 : Phage display. A. Sélection type d'une enzyme présentée en surface de phages. Le pool de phage-enzymes réagit avec le substrat biotinyllé (1), le phage enzyme marqué est capturé par des billes recouvertes de streptavidine (2), des lavages permettent d'éliminer les phages non liés ou liés non spécifiquement, le phage capturé est élué par rupture du lien entre le substrat et la biotine ou la pIII et l'enzyme (3), le phage d'intérêt est amplifié et peut entrer dans un nouveau tour de sélection ou être caractérisé (4). B. Sélection pour activité catalytique intraphage. Les phage-enzymes actifs réagissent avec le substrat qui leur est lié pour donner un produit (1), seul le produit de réaction du substrat et du phage-enzyme peut subir une modification chimique (2) cette modification chimique permet au phage d'être capturé sur un support (3).

sélection modèle de phages présentant une adénylate cyclase. La conversion enzymatique de l'ATP couplé chimiquement au phage en AMPc a permis la sélection des phages enzymes actifs en utilisant des anticorps anti-AMPc, et un facteur d'enrichissement de 70 fois a pu être observé [95]. On peut aussi coupler orthogonalement les réactions de l'enzyme avec son substrat et la réaction chimique menant au marquage du phage par le substrat ou le produit (Figure 19, B). La technique du phage display a été utilisée plusieurs fois pour l'évolution de spécificité d'ADN polymérases [96]. La sélection la plus récente consiste en l'attachement covalent du substrat (une amorce ADN) aux protéines du manteau via une longue flexible. Les variants sont sélectionnés sur base de leur capacité à polymériser l'amorce liée, d'abord en incorporant les nucléotides cible (dans ce cas, des 2'-O-methylribonucleosides), et finalement, un nucléotide biotinylé permettant la capture du phage sur des billes recouvertes de streptavidine [97]. Cependant, dans ces transformations intraphage, un seul événement catalytique va conduire à la sélection du phage, les enzymes ayant une très faible activité vont donc aussi être sélectionnées.

Une stratégie de sélection appelée élution catalytique peut être appliquée pour des enzymes dont l'activité catalytique dépend de la liaison à un cofacteur. Après expression de la protéine en surface du phage, le cofacteur (un ion métallique par exemple) est enlevé, et les phages catalytiquement inactifs sont capturés à l'aide d'un substrat immobilisé. Le cofacteur est alors ajouté et les phages présentant une enzyme active sont élués par conversion du substrat en produit pour lequel l'enzyme présente une moindre affinité. Cette méthode a été appliquée à la sélection d'une métallo- β -lactamase permettant d'augmenter l'activité de la banque de 60 fois pour un substrat particulier après deux tours de sélection [98].

Enfin, une sélection directe basée sur le cycle catalytique complet de l'hydrolyse d'un monoester phosphate a récemment été montrée [99]. Dans ce système, des anticorps catalytiques capables d'hydrolyser un aryl phosphate ont été identifiés. Le produit d'hydrolyse se réarrange spontanément en quinone-méthide qui en retour, réagit avec les résidus nucléophiles dans ou proches du site actif du catalyseur, capturant ainsi la particule phagique. Malgré l'identification d'anticorps avec une activité phosphatase, celle-ci reste bien inférieure à une activité phosphatase naturelle avec une différence d'activité catalytique supérieure à 6.10^3 . Bien que le phage display reste la méthode la plus commune pour les sélections *in vitro*, l'émergence de techniques de criblage par FACS a permis d'isoler des enzymes actives.

Le développement de protéines fluorescentes comme la GFP (Green Fluorescent Protein) comme protéines rapporteurs de l'expression de gène, et des criblages par FACS de cellules

entières ont fourni de nouvelles méthodes pour le suivi d'activités enzymatiques [100]. Cependant, cette technique est applicable seulement si le substrat est capable de traverser la membrane bactérienne. La présentation de protéines en surface de cellules peut permettre la sélection d'enzymes, mais ce type de sélection est limité et des exemples de réussite avec ce type de méthode sont rares. Varadarajan *et al.* ont mis au point un système de criblage de variants d'une protéase exprimée en surface de cellules. L'utilisation de substrats fluorescents de différentes couleurs qui marquent la surface de la bactérie après clivage par l'enzyme permet de cribler l'activité et la spécificité enzymatiques en fonction de profil de fluorescence exposé par chaque cellule [101]. Bien que ces techniques et celle du phage display requièrent une étape de criblage ou de sélection *in vitro*, l'expression de la protéine passe par une étape *in vivo*, et ces méthodes vont être limitées si la protéine à analyser est toxique.

L'IVC a émergé comme étant une technique alternative robuste pour la sélection ou le criblage de fonctions catalytiques, en contournant les problèmes d'expression des variants liés à leur toxicité et ne requérant pas de substrats lipophiles. Cette technique, associée à l'utilisation de substrats fluorogéniques et du FACS s'annonce très prometteuse. A partir d'une banque présentée sur billes (Figure 14, C-4), Griffiths & Tawfik ont pu identifier des variants fonctionnels d'une phosphotriestérase (PTE). Le k_{cat} de la PTE pour l'hydrolyse du paraoxon est 2280 s^{-1} et le k_{cat}/K_M de $6.23 \cdot 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ est proche de la vitesse de diffusion de petites molécules. Malgré une catalyse déjà très efficace, un mutant sélectionné a un k_{cat} de $1.43 \cdot 10^5 \text{ s}^{-1}$, soit 63 fois plus rapide que la PTE sauvage, et un k_{cat}/K_M de $1.83 \cdot 10^8 \text{ M}^{-1} \cdot \text{s}^{-1}$ [102]. Pour arriver à cela, une émulsion est créée dans un premier temps pour permettre l'expression *in vitro* des variants de l'enzyme. Les microbilles présentant le gène et son produit d'expression sont recombpartimentalisées dans un deuxième temps pour la sélection avec un substrat fluorogénique biotinylé. Le produit de catalyse se lie par affinité sur les billes recouvertes de streptavidine et les billes fluorescentes sont alors triées par FACS ou sélectionnées par purification d'affinité avec un anticorps anti-produit. Cette technique permet de discriminer les enzymes sur base de l'efficacité catalytique, le nombre de turnovers étant proportionnel à la fluorescence émise par une bille.

Une adaptation récente de l'IVC avec la formation de doubles émulsions a permis de sélectionner une enzyme Ebg ayant acquis une activité β -galactosidase. Alors que les émulsions simples eau dans huile ne peuvent pas être soumises directement au FACS et qu'une étape de rupture de l'émulsion et sélection de billes ou de l'ADN est requise (Figure 14, C), les doubles émulsions (eau dans huile dans eau) permettent la sélection directe au FACS avec un substrat fluorogénique de l'enzyme à tester [103].

Il existe un large éventail d'outils pour l'ingénierie des protéines, de la construction de banques aux méthodes de sélection ou de criblage. Malgré les récentes avancées, aussi bien pour la mise en place de stratégies pour créer des enzymes sur mesure et des méthodes permettant d'identifier des variants sur base de multiples turnovers, la génération d'une enzyme artificielle aussi active qu'une enzyme naturelle n'a pas encore été réalisée. Dans l'objectif ambitieux d'arriver à un tel résultat, nous avons choisi de travailler sur le modèle de l'évolution des PBPs et des β -lactamases.

Les DD-peptidases et les β -lactamases

La famille des protéines à sérine reconnaissant les pénicillines ou PREs ("Penicillin Recognizing Enzymes") est une grande famille qui englobe un grand nombre d'enzymes avec des activités différentes. Parmi elles, on retrouve des activités de type β -lactamase, DD-peptidase, LD-endopeptidase [104], endopeptidase D-stéréospécifique [105], D-amidase [106], D-estérase [107], D-aminopeptidase [108], et de régulation de l'expression des β -lactamases [109]. Alors que pour certaines de ces activités, le lien avec les PREs est ponctuel car ces protéines ont été relativement récemment découvertes, les familles d'enzymes comprenant des activités DD-peptidases et β -lactamases sont les principales et ont été largement étudiées depuis plus de 50 ans. Dans cette partie, nous étudierons plus en détails les connaissances accumulées sur ces deux classes d'enzymes, et les évidences d'une origine ancestrale commune.

Les DD-peptidases

Les D-alanyl-D-alanine peptidases ou DD-peptidases sont des enzymes pour la plupart liées à la membrane, impliquées dans la synthèse et le remodelage du peptidoglycane. Elles sont aussi appelées PBPs pour "Penicillin Binding Proteins" à cause de leur capacité à former des complexes covalents stables avec les antibiotiques de type pénicilline.

Le peptidoglycane

Le peptidoglycane est une structure macromoléculaire composée de chaînes linéaires de glycanes pontées entre elles par de courts peptides. Il entoure la membrane cytoplasmique des bactéries et est responsable de leur forme et de leur résistance mécanique aux différences de pression osmotique.

De manière plus détaillée, l'ossature du peptidoglycane est composée d'unités répétées du disaccharide N-acétylglucosamine (NAG)-acide-N-acétylmuramique (NAM) liées en β -1,4, avec un pentapeptide lié à la portion D-lactyl de chaque NAM. Ces liens pentapeptidiques participent à des réactions de pontage interglycanes, créant ainsi le polymère de la paroi bactérienne. De manière générale, chez les bactéries Gram⁺, le pentapeptide est composé de L-Ala-D- γ -Glu-L-Lys-D-Ala-D-Ala, et chez les bactéries Gram⁻, la L-Lysine est remplacée par un acide méso-diaminopimélique (*m*-DAP) (Figure 20). La principale réaction de pontage

se fait entre la D-Alanine en position 4 d'une chaîne de glycane et le résidu en position 3 de la chaîne voisine, (pontage 4→3).

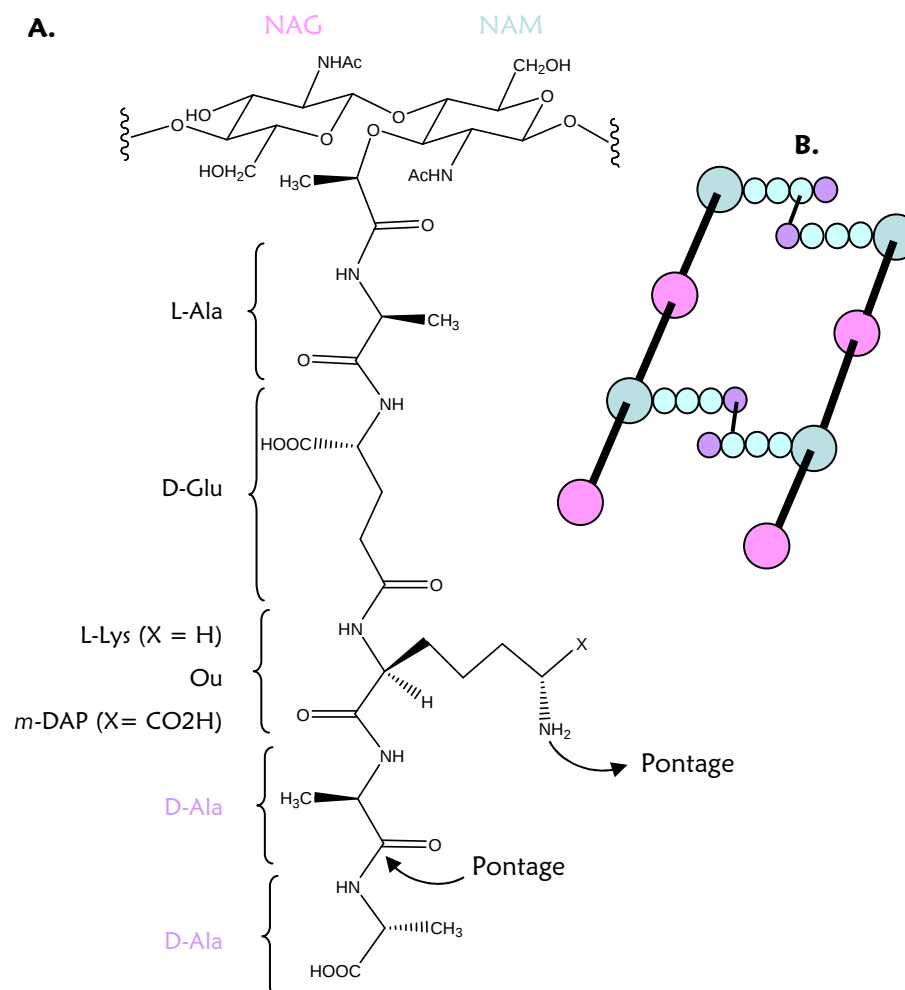


Figure 20 : Structure du peptidoglycane. A. Structure primaire non pontée. B. Représentation schématique du polymère. Les NAG sont représentées en rose, les NAM en bleu et les D-Alanine en mauve.

Des variations vont exister dans la nature chimique des pontages (autres que la variation L-Lys ou *m*-DAP), et il existe un autre type de pontage en 3→3 faiblement représenté chez *E. coli* mais qui peut se retrouver dans des proportions 4→3:3→3 de 2:1 chez certaines bactéries [110, 111]. De nombreuses études ont visé à étudier la structure du peptidoglycane, ces recherches sont revues par Goffin & Ghuyssen pour les bactéries Gram⁺ [111]. a récemment été suggéré que la structure tridimensionnelle de ce peptidoglycane adopte une conformation en hélice droite correspondant à trois paires de NAG-NAM par tour d'hélice [112].

Bien que des zones d'ombres planent sur les mécanismes permettant de générer la morphologie des bactéries en rapport avec la synthèse du peptidoglycane [113], le métabolisme de ce dernier est assez bien compris pour des bactéries Gram⁻ comme *E. coli*

[114, 115]. De manière très schématisée, un précurseur disaccharidique pentapeptide prénylé (lipide II) est synthétisé dans le cytosol puis transféré du côté extérieur de la membrane cytoplasmique où il est pris en charge par la machinerie métabolique du périplasma pour sa polymérisation. Le peptidoglycane est une structure dynamique qui est continuellement remodelée au cours du cycle cellulaire sous le contrôle régulé de deux machineries opposées de synthèse, à travers des activités transglycosylase pour l'élongation et transpeptidase (ou acyl transférase) pour la réticulation, et d'hydrolyse à travers des activités endopeptidase, carboxypeptidase ou glycosydase. Ces réactions sont en majorité catalysées par les PBPs [116].

Les PBPs

Les DD-peptidases ou PBPs catalysent principalement des réactions de transpeptidation et de carboxypeptidation (Figure 21).

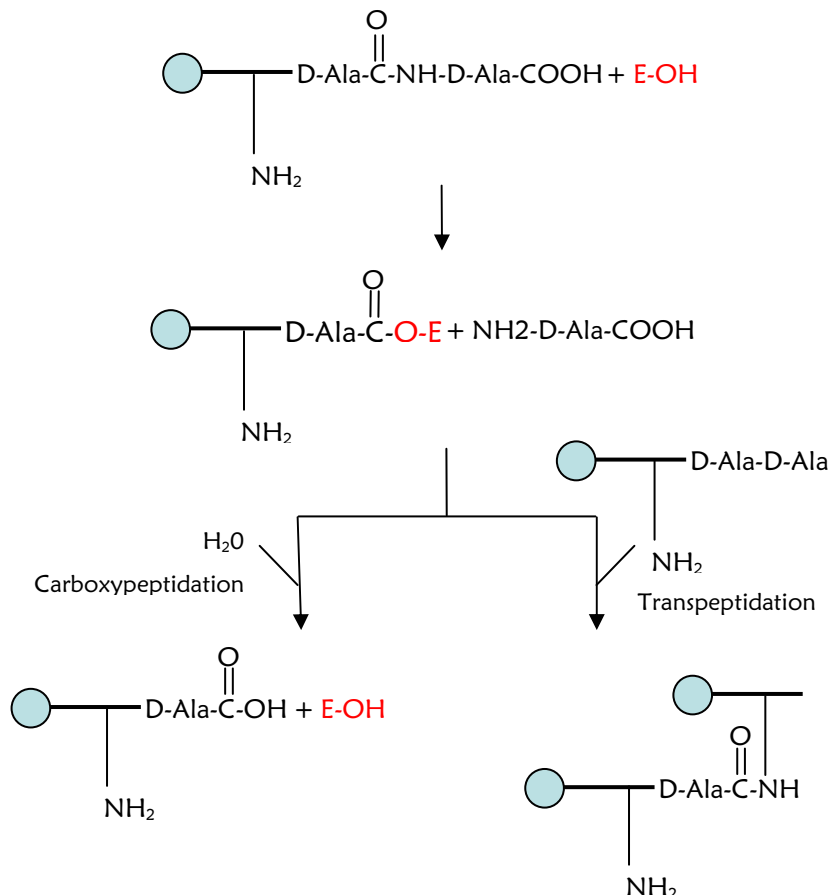


Figure 21 : Réactions catalysées par le domaine PB des DD-peptidases sur le pentapeptide du peptidoglycane. E-OH : enzyme.

Toutes les DD-peptidases découvertes jusqu'à aujourd'hui sont des enzymes comportant une sérine au niveau du site actif, et leur machinerie catalytique implique des intermédiaires acyl-enzymes. Elles sont divisées en deux classes, de haut et de bas poids moléculaire en fonction

de la localisation de la sérine catalytique dans la structure primaire. Les PBPs de bas poids moléculaire ou LMW (Low Molecular Weight) sont des entités catalytiques monofonctionnelles et les PBPs de haut poids moléculaire ou HMW (High Molecular Weight) sont des protéines composées de deux domaines : un domaine de quelques centaines d'acides aminés est fusionné du côté N-terminal à un domaine de reconnaissance des pénicillines (domaine PB, pour "Penicillin Binding"). Des comparaisons de séquences des enzymes des deux classes ont permis d'établir des sous-classes d'enzymes pour chacune des PBPs LMW et HMW. Malgré des divergences de séquences parfois grandes, trois motifs importants pour la catalyse restent très conservés chez les PBPs LMW et dans le module PB des PBPs HMW, ce sont les motifs Ser-Xaa-Xaa-Lys, Ser/Tyr-Xaa-Asn/Cys et Lys/His-Thr/Ser-Gly (Table 2).

	Motif 1	Motif 2	Motif 3
β-lactamases de classe A	70 Ser*-Xaa-Xaa-Lys	130 Ser-Asp-Asn Ser-Asp-Ser Ser-Asp-Gly	234 Lys-Thr-Gly Lys-Ser-Gly Arg-Thr-Gly Arg-Ser-Gly
β-lactamases de classe C	64 Ser*-Xaa-Ser-Lys	150 Tyr-Ala-Asn Tyr-Ser-Asn	314 Lys-Thr-Gly
β-lactamases de classe D	70 Ser*-Xaa-Xaa-Lys	144 Tyr-Gly-Asn	214 Lys-Thr-Gly
DD-peptidases	62 Ser*-Xaa-Xaa-Lys	159 Tyr-Xaa-Asn Ser-Xaa-Asn Ser-Xaa-Cys	298 His-Thr-Gly Lys-Thr-Gly Lys-Ser-Gly

Table 2 : les trois motifs structurellement et fonctionnellement équivalents chez les β -lactamases à sérine et les DD-peptidases. La sérine catalytique est indiquée par *.

Les PBPs LMW

Sur bases de comparaison de séquences, les PBPs LMW sont subdivisées en trois sous-classes, A, B, et C. Cette classe de PBPs est générée avec un peptide signal clivable. Toutes les PBPs LMW sont exprimées sous forme de protéines solubles (PBP7 et PBP8 de *E. coli*), liées à la membrane par une extension C-terminale soit directement ancrée dans la membrane, soit en interaction avec la surface membranaire (PBP4, 5, 6 et 6b de *E. coli*). Bien qu'elles catalysent des réactions de transpeptidation, la plupart d'entre elles ne sont que des DD-carboxypeptidases (PBP4 à PBP7) [117]. Pour tester des activités DD-peptidases, deux types de substrats sont généralement utilisés, des substrats peptidiques synthétiques simples terminant par D-Ala-D-Ala et des thiolesters, des exemples de données cinétiques obtenues avec des substrats types sont donnés en Table 3.

$Ac_2KAA = \text{Diacétyl-L-Lys-D-Ala-D-Ala}$ $S2d = \text{Benzoyl-D-Ala-Thioglycolate}$		Ac_2KAA		$S2d$	
		k_2/K_s ($M^{-1}.s^{-1}$)	K_m (mM)	k_2/K_s ($M^{-1}.s^{-1}$)	K_m (mM)
PBPs LMW	K15 peptidase ^b <i>Streptomyces</i>	50 ⁱ	6,2	75	1,5
	R61 peptidase ^b <i>Streptomyces</i>	4000	14	700 000	0,1
	R39 peptidase ^b <i>Actinomadura</i>	22 500	0,8	400 000	15
	PBP3 <i>Neisseria gonorrhoeae</i> ^f	29 000	19	ND	ND
	PBP4a <i>Bacillus subtilis</i> ^d	ND	ND	61 000	0,38
	PBP5 <i>E. coli</i>	10 400 ^c	14 ^c	ND	ND
		32 ^e	ND	ND	ND
PBPs HMW	PBP1 <i>Enterococcus hirae</i> ^a	0	0	1600	ND
	PBP2b <i>Streptococcus pneumoniae</i> ^a	0	0	300	ND
	PBP2x <i>Streptococcus pneumoniae</i> ^b	0	0	5 000	5,6
	PBP3 <i>Enterococcus hirae</i> ^b	0	0	3 200	1,8
	PBP3 <i>E. coli</i> ^b	0	0	80	3

Table 3 : paramètres cinétiques d'hydrolyse des substrats peptidique Ac_2KAA et thiolester $S2d$ à 37°C. i, dans tampon contenant 1,5 mM d'accepteur Gly-Gly. a, tiré de [118], b, tiré de [119], c, tiré de [120], d, tiré de [121], e tiré de [122], f, tiré de [123]. ND : Non Déterminé.

Le mécanisme catalytique de ces enzymes se fait en deux étapes, une acylation suivie d'une désacylation. L'attaque du carbonyle du pénultième D-Ala des peptides par la sérine catalytique mène à la formation d'un intermédiaire covalent acyl-enzyme lié à la sérine par une liaison ester. Parce que la liaison scissile est acyclique, le groupe partant D-Ala ne fait pas partie de l'acyl-enzyme et celui-ci a un temps d'occupation très court dans le site actif. L'efficacité avec laquelle les LMW PBPs réalisent la réaction de désacylation par l'hydrolyse ou la transpeptidation du carbonyle donneur dépend, de manière simplifiée de l'efficacité avec laquelle l'eau, le groupe partant D-Ala et un peptide exogène accepteur peuvent attaquer l'acyl-enzyme (Figure 21). Les paramètres cinétiques avec les différents substrats sont très variables d'une PBP à l'autre, les PBPs les plus efficaces à ce jour étant les DD-peptidases R61 et R39. Toutes les bactéries possèdent une ou plusieurs PBPs LMW. Bien qu'apparemment non essentielles pour la survie des bactéries, elles semblent impliquées dans le maintien de leur morphologie [124].

Les PBPs HMW

Les PBPs HMW de classe A et B possèdent deux domaines fusionnés qui opèrent de manière concertée. Un domaine amino-terminal de plusieurs centaines d'acides aminés est lié au domaine PB. Dans la plupart des PBPs HMW, le domaine PB porte une extension C-terminale d'une centaine d'acides aminés. Ces PBPs sont impliquées dans les phénomènes d'élongation de la paroi cellulaire, de septation pendant la division cellulaire et de détermination de la forme des bactéries [125].

Ces protéines ne sont pas synthétisées avec un peptide signal mais possèdent une région hydrophobe proche du côté N-terminal qui agit comme un segment signal non clivé pour transloquer la protéine dans le périplasme et ancrer la protéine dans la membrane plasmique. Les PBPs HMW de classe A sont des enzymes bifonctionnelles transglycosylase et transpeptidase, les plus étudiées étant PBP1a et PBP1b de *E. coli*. L'expression et la purification de ces protéines tronquées ou non est difficile et leur activité sur les substrats peptidiques synthétiques simples est indétectable (Table 3). Leur activité en présence de précurseur du peptidoglycane a pu être mise en évidence *in vitro*, le faible nombre de turnovers observés suggère cependant que les conditions ne reflètent pas ce qu'il se passe *in vivo*, ces enzymes faisant probablement partie de complexes enzymatiques. Leur activité a pu être mesurée avec les substrats synthétiques de type thioesters. Contrairement aux PBPs LMW, ces enzymes sont essentielles pour la survie des bactéries [117].

Les PBPs HMW de classe B sont des entités monofonctionnelles qui possèdent une activité transpeptidase, l'activité de leur portion N-terminale reste obscure. Elles sont représentées par les PBP2 et 3 de *E. coli*, qui sont les plus étudiées. Ces PBPs, sont elles aussi essentielles et font partie du complexe de division cellulaire.

Les PBPs HMW de class C constituent la classe de PBPs comprenant les protéines récepteurs des β -lactames (comme BlaR1) impliquées dans l'induction de la synthèse des β -lactamases et ne présentent pas d'activité DD-peptidase ou thiolestérase [126].

Les β -lactamases

Les PBPs sont des enzymes vulnérables car elles sont essentielles et exposées. En mimant la portion D-alanine-D-alanyl du peptidoglycane, les antibiotiques de type β -lactames vont se lier de manière covalente aux PBPs et les inhiber de manière réversible mais durable [127]. Afin de contourner cette inhibition qui entraîne la mort cellulaire, les bactéries ont développé divers mécanismes de résistance. L'émergence des β -lactamases, des enzymes généralement sécrétées dans le milieu extérieur par les bactéries Gram⁺ et dans le périplasme par les bactéries Gram⁻, est un des mécanismes principaux utilisé par les bactéries pour se défendre contre ces antibiotiques.

Origines des résistances bactériennes aux β -lactames

Les infections bactériennes ont toujours posé de graves problèmes de santé publique. Afin de les combattre, il est nécessaire de développer des antibiotiques qui vont empêcher la croissance bactérienne ou mener à leur lyse et ceci, en ciblant des éléments fonctionnels absents chez les mammifères. Ainsi, le peptidoglycane constituant la paroi bactérienne est une cible de choix pour sa sélectivité. Les antibiotiques de type β -lactames ont été découverts officiellement en 1928 par A. Fleming sous la forme de pénicilline G ou benzylpénicilline, et sont des inhibiteurs spécifiques des PBPs, en inhibant les DD-peptidases, les β -lactames bloquent les réactions de carboxypeptidation et surtout de transpeptidation, générant un peptidoglycane faiblement réticulé provoquant la lyse bactérienne. Ces antibiotiques ont donc été largement développés depuis plus de 50 ans afin d'éradiquer les infections de type bactériennes. Cependant, la large utilisation de ces antibiotiques a mené à l'émergence de phénomènes de résistance bactérienne. Parmi les mécanismes de résistance générés par les bactéries, on retrouve celui d'altération de la perméabilité des cellules pour empêcher l'accès de l'antibiotique à sa cible. Ceci se fait par le biais de systèmes de pompes d'efflux qui vont généralement refouler une large gamme d'antibiotiques, le système MexAB-OprM de *P. aeruginosa* est l'un des mieux étudiés [128]. Les bactéries vont aussi se défendre contre les β -lactames par le biais de changements dans la cible elle-même de l'antibiotique. Des PBPs moins sensibles à ces molécules et appelées PBPs de faible affinité ont donc été générées. De nombreuses études tendent à comprendre l'origine de ces PBPs résistantes, comme par exemple l'apparition de PBPs mosaïques ou des PBPs avec des mutations ponctuelles [129]. Enfin, le mécanisme de résistance le plus répandu et le plus étudié est la destruction ou la modification des antibiotiques avant qu'ils n'atteignent leur cible par le biais d'enzymes

appelées β -lactamases. Ces enzymes confèrent une résistance aux antibiotiques significative à leur hôte en hydrolysant la liaison amide du noyau β -lactame. La dissémination des gènes des β -lactamases a été grandement exacerbée par leur intégration dans des éléments génétiques mobiles comme des plasmides ou des transposons qui facilitent le transfert génétique rapide entre deux bactéries [130]. Afin de contrecarrer ces phénomènes de résistance, d'autres antibiotiques dérivés de la pénicilline ou d'autres antibiotiques à noyau β -lactame tels les céphalosporines, les carbapénèmes, les monobactames (Figure 22) ou encore des inhibiteurs suicides des β -lactamases ont été développés. Cependant, le développement de nouveaux antibiotiques est rapidement contrecarré par l'apparition de bactéries résistantes, par le biais notamment de β -lactamases à large spectre.

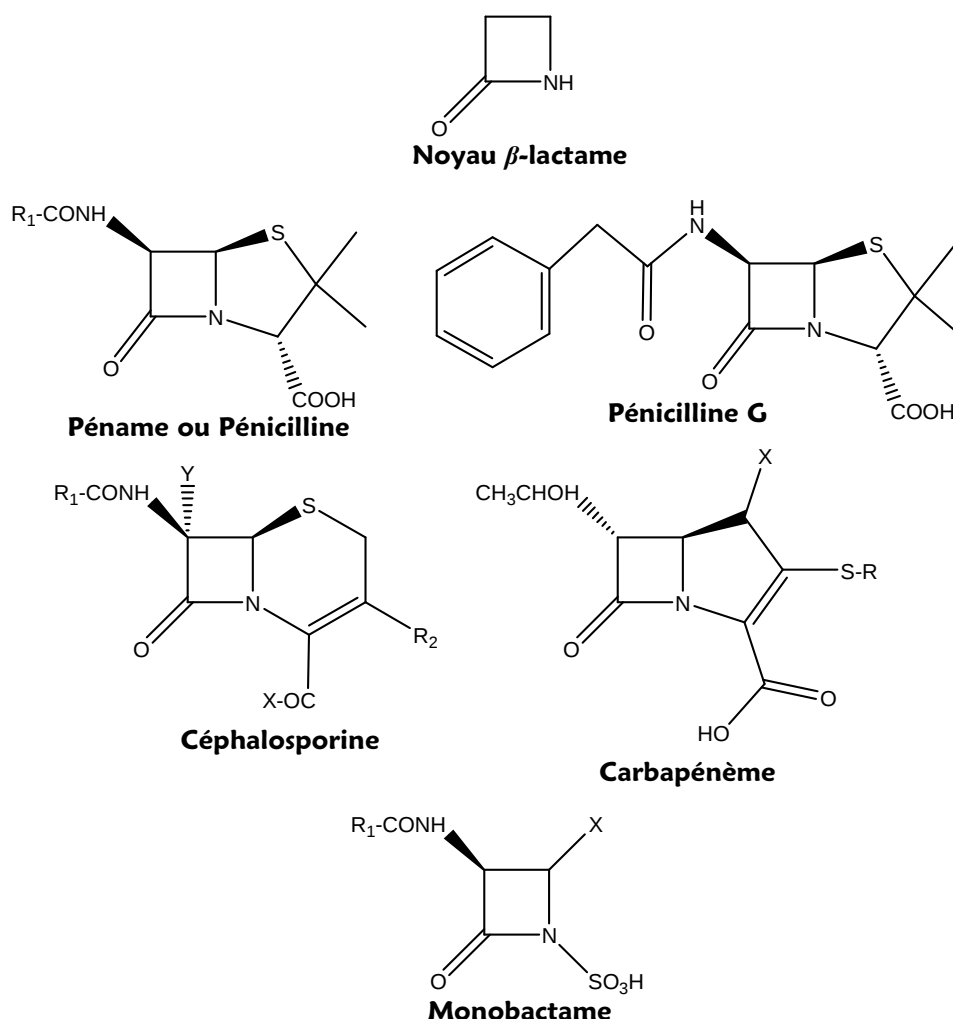


Figure 22 : les différentes classes d'antibiotiques de type β -lactame.

Description du mécanisme

Les β -lactamases sont divisées en quatre classes A, B, C et D sur base de leur similarité de séquence. Alors que les β -lactamases de classe A, C et D possèdent une sérine dans leur site actif pour leur réaction de catalyse et cette sérine fait partie d'un des trois motifs hautement conservés (Table 2), les β -lactamases de classe B sont des métallo-enzymes, requérant un ion métallique, le zinc, dans le site actif. Les β -lactamases de classe A, C et D hydrolysent les β -lactames selon un même mécanisme réactionnel d'acylation/déacylation en trois étapes. La première étape est la formation réversible d'un complexe non covalent entre l'enzyme et le substrat ou complexe de Michaélis. La deuxième étape est la transformation de ce complexe non covalent en complexe covalent acyl-enzyme et nécessite le passage par un intermédiaire tétraédrique. Ces deux étapes constituent la phase d'acylation du substrat qui est généralement très rapide et proche de la vitesse limite de diffusion des petites molécules. Enfin, la troisième étape consiste en l'hydrolyse de l'acyl-enzyme libérant un antibiotique inactif et une enzyme régénérée (Figure 23). Cette étape est également très rapide chez les β -lactamases ce qui explique l'impossibilité d'isoler le complexe covalent acyl-enzyme chez ces protéines.

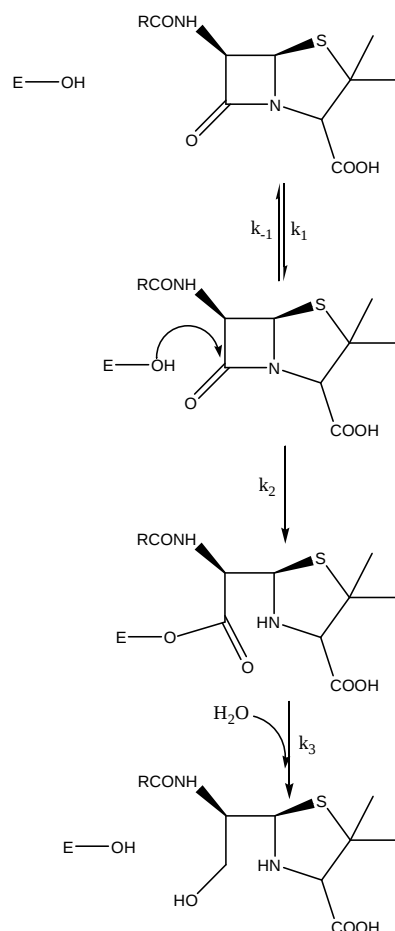


Figure 23 : mécanisme d'hydrolyse des pénicillines par les β -lactamases à sérine.

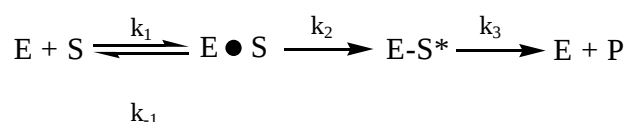
Mécanisme catalytique général

Bien que la fonction des β -lactamases à sérine et des PBPs (par le module PB) soit différente et qu'elles catalysent des réactions distinctes, elles ont en commun un même substrat donneur de carbonyle $R_1\text{-CO-X-R}_2$ et la catalyse se fait par un double transfert de protons similaire. Pendant l'acylation, le proton γOH de la sérine catalytique est transféré au groupe partant $X\text{-R}_2$ avec l'attaque nucléophile concomitante du carbonyle du donneur par le O_γ activé [131]. Dans l'étape de désacylation, le proton de l'accepteur (un groupement amine ou une molécule d'eau) est arraché, l'accepteur activé attaque le carbonyle de l'acyl-enzyme, et le proton arraché est rendu à l' O_γ de la sérine. Ce mécanisme catalytique a été étudié en détails pour ces familles d'enzymes permettant de définir les rôles de différents résidus du site actif. Comme dans toutes les protéases à sérine, le mécanisme catalytique inclut un trou à oxyanion défini par les groupements NH des chaînes principales de la sérine catalytique et du résidu suivant le motif 3 [132].

Ainsi, le site actif des β -lactamases de classe A, C et D se compose de deux ensembles d'acides aminés importants pour la catalyse. Le premier, composé d'un minimum de cinq acides aminés est l'ensemble catalytique, et comprend la sérine catalytique, une base générale pour la sérine, le trou à oxyanion (deux acides aminés) et un site de reconnaissance catalytique pour l'acide carboxylique. On s'attend à ce que ces résidus soient invariables ou très largement conservés parmi les différentes classes de β -lactamases. Le second est l'ensemble de reconnaissance qui complète les segments hydrophobe et hydrophile du reste de la structure du β -lactame. Ces acides aminés sont variables et sont ceux qui vont être mutés lorsque les bactéries sont soumises à des pressions de sélection.

Les équations cinétiques de l'interaction PRP- β -lactame

Le mécanisme en trois étapes décrit ci-dessus peut être schématisé de la façon suivante :



où k_2 et k_3 sont des constantes de premier ordre d'acylation et de désacylation, respectivement. Les paramètres généralement étudiés pour ce genre de modèle sont les paramètres de Michaëlis-Menten (k_{cat} , K_m et k_{cat}/K_m). Le k_{cat} aussi appelée constante catalytique, est dans ce modèle, fonction de toutes les vitesses de premier ordre et ne peut être attribué à un processus particulier, sauf si on peut faire des simplifications (voir ci-dessous).

Un deuxième paramètre de Michaëlis-Menten important pour l'étude du mécanisme est le K_m , appelée la constante d'équilibre et qui équivaut à la concentration en substrat pour laquelle $v = V_{\max}/2$. Le troisième paramètre important pour l'étude du mécanisme est la constante de spécificité ou k_{cat}/K_m . La valeur de k_{cat}/K_m varie selon le substrat et son affinité pour l'enzyme, c'est pourquoi on l'appelle constante de spécificité et peut être considéré comme un indice de l'efficacité de la réaction de l'enzyme libre avec un substrat donné.

La vitesse d'apparition de P, v va pouvoir s'exprimer de la manière suivante :

$$v = \frac{\left(\frac{k_2 \cdot k_3}{k_2 + k_3} \right) \cdot [E]_{tot} \cdot [S]}{[S] + \frac{k_3}{k_2 + k_3} \left(\frac{k_{-1} + k_2}{k_1} \right)}$$

où $[E]_{tot}$ représente la concentration en enzyme totale de la réaction et $[S]$ la concentration en substrat.

Cette vitesse est de la forme :

$$v = \frac{k_{cat} \cdot [E]_{tot} \cdot [S]}{[S] + K_m}$$

On peut donc en tirer les valeurs de k_{cat} et K_m :

$$k_{cat} = \frac{k_2 \cdot k_3}{k_2 + k_3} \quad \text{et} \quad K_m = \frac{k_3}{k_2 + k_3} \left(\frac{k_{-1} + k_2}{k_1} \right)$$

Des simplifications sont possibles, en particulier, si $k_{-1} \gg k_2$:

$$k_{cat} = \frac{k_2 \cdot k_3}{k_2 + k_3} \quad \text{et} \quad K_m = K_s \cdot \frac{k_3}{k_2 + k_3} \quad \text{avec} \quad K_s = \frac{[E] \cdot [S]}{[E \cdot S]} = \frac{k_{-1}}{k_{+1}}$$

où K_s est la constante de dissociation du complexe de Michaëlis.

Dans le cas particulier où $k_{-1} \gg k_2$ et $k_2 \gg k_3$, alors :

$$k_{cat} = k_3 \quad \text{et} \quad K_m = K_s \cdot \frac{k_3}{k_2} \quad (K_m \ll K_s)$$

Nous nous retrouvons dans ce cadre dans le cas où des PBPs vont réagir avec un substrat de type pénicilline. Dans ce cas, k_{cat}/K_m va être égal à k_2/K_s .

Si $k_{-1} \gg k_2$ et $k_2 \ll k_3$, alors :

$$k_{cat} = k_2 \quad \text{et} \quad K_m = K_s$$

Dans ce travail, nous allons retrouver cette situation pour la réaction de la PBP étudiée (PBP-A) avec des substrats de type thioesters.

Enfin, si $k_{-1} \ll k_2$, alors :

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \quad \text{et} \quad K_m = \frac{k_3}{k_2 + k_3} \left(\frac{k_{-1} + k_2}{k_{+1}} \right) \quad \text{alors} \quad \frac{k_{cat}}{K_m} = k_{+1}$$

C'est généralement le cas pour les β -lactamases, dont la valeur de k_{cat}/K_m est proche de la vitesse limite de diffusion des petites molécules.

Les différentes classes de β -lactamases

Les β -lactamases de classe A

Dans ce travail, nous nous intéressons plus particulièrement aux β -lactamases de classe A qui sont les plus répandues et les mieux étudiées. Historiquement, ces protéines étaient décrites comme des "pénicillinases" pour leur meilleure capacité à hydrolyser les pénicillines par rapport aux céphalosporines. Il existe pour ces enzymes un consensus de numérotation proposé par Ambler et que nous utiliserons par la suite [133].

La première structure de β -lactamase de classe A permettant de décrire la topologie générale de ces protéines est apparue en 1987 [134]. Depuis lors, une dizaine d'autres structures ont été résolues et toutes présentent le même repliement tridimensionnel et la même organisation du site actif. Ces protéines sont toutes constituées d'un domaine α et d'un domaine α/β avec le site actif situé à l'interface entre les deux domaines (Figure 24). Ce site actif comprend les trois motifs conservés dans toutes les β -lactamases à sérine et une boucle Ω caractéristique des β -lactamases de classe A (Table 2). Le premier élément, situé sur une hélice du domaine tout α contient le résidu sérine actif (Ser₇₀) et un tour d'hélice plus loin, une lysine (Lys₇₃) dont la chaîne latérale pointe dans le site actif. Le deuxième élément, une boucle dans le domaine tout α , consiste en la séquence S₁₃₀XN, les résidus 1 et 3 pointent dans le site actif. Face à cette triade, le troisième élément, situé sur un brin β , forme le côté opposé de la cavité, c'est la séquence K₂₃₄TG. Les deux premiers résidus pointent dans le site actif, et la glycine, un résidu ne comportant pas de chaîne latérale, est strictement conservée. Un quatrième élément, comportant un résidu chargé négativement (Glu₁₆₆), et jouant un rôle fonctionnel important est une boucle de 16 ou 19 résidus (Arg₁₆₄/Arg₁₆₁-Asp₁₇₉) généralement appelé boucle Ω . Tous ces résidus forment avec des molécules d'eau conservées un dense réseau de liaisons hydrogène important pour la catalyse de la réaction [135].

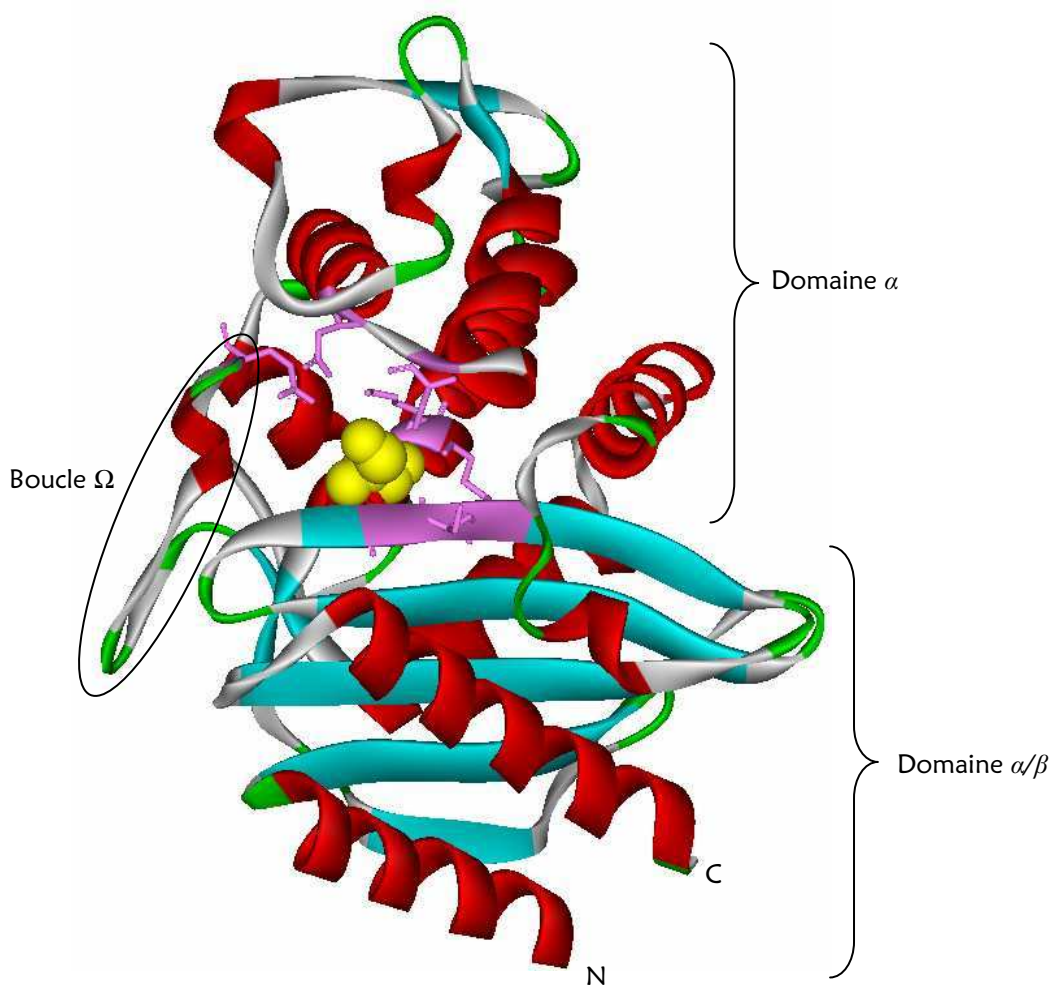


Figure 24 : structure 3-D de la β -lactamase TEM-1. La sérine 70 catalytique est indiquée en jaune, les autres résidus du site catalytique sont indiqués en rose. Les brins β sont couleur cyan et les hélices α rouge.

Le glutamate 166 est un pivot dans la catalyse pour son rôle clairement établi dans la réaction de désacylation, et son rôle qui reste encore plus ou moins controversé dans la réaction d'acylation. La situation critique du Glu₁₆₆ a d'abord été pointée par les études cristallographiques sur base de la structure de la β -lactamase PC1 de *Staphylococcus aureus* [134], puis sur base d'étude du mutant inactif D₁₇₀N de cette même enzyme, dont la boucle Ω déstabilisée ne permet plus l'orientation correcte du Glu₁₆₆ [136]. De manière parallèle, des expériences de mutagenèse dirigée du Glu₁₆₆ ont mis en évidence son importance dans la catalyse aussi bien pour l'étape d'acylation que de désacylation [137]. Les études successives de structure et de mutants de β -lactamases ont permis de proposer un modèle de réaction où une molécule d'eau conservée W1, qui forme un pont salin entre le groupement hydroxyle de la Ser₇₀ et l'acide carboxylique du Glu₁₆₆ servirait de relais pour le transfert de proton (Figure 25, A). L'oxygène γ activé de la Ser₇₀ pourrait alors attaquer le carbonyle du β -lactame sur la

face α de la molécule et le proton serait transféré à l'azote du β -lactame via un réseau de liaisons hydrogène impliquant les groupements amine des résidus Lys₇₃ et Lys₂₃₄, une seconde molécule d'eau W2 et l'hydroxyle de la Ser₁₃₀ qui agit comme le dernier donneur de proton. L'hydrolyse de l'acyl-enzyme se ferait alors selon un mécanisme symétrique où le Glu₁₆₆ active la molécule d'eau hydrolytique W1 pour attaquer le carbonyle de l'acyl-enzyme et rendre un proton à la Ser₇₀, menant à la régénération de l'enzyme libre [138]. Cependant, l'étude d'Adachi *et al.* du mutant E₁₆₆N de la β -lactamase TEM-1 montre que contrairement à la désacylation qui est très perturbée la vitesse d'acylation est diminuée mais l'enzyme est toujours capable d'acyler la pénicilline [139]. L'étude cristallographique de ce mutant a mené à la proposition que le résidu impliqué dans l'acylation est la Lys₇₃ et le mécanisme global d'hydrolyse serait asymétrique [140]. Dans ce mécanisme, le site actif de l'enzyme complexée au substrat fournirait un environnement local permettant de diminuer le pKa de la lysine qui serait alors déprotonée à pH neutre, et pourrait jouer le rôle de base pour activer la Ser₇₀ (Figure 25, B).

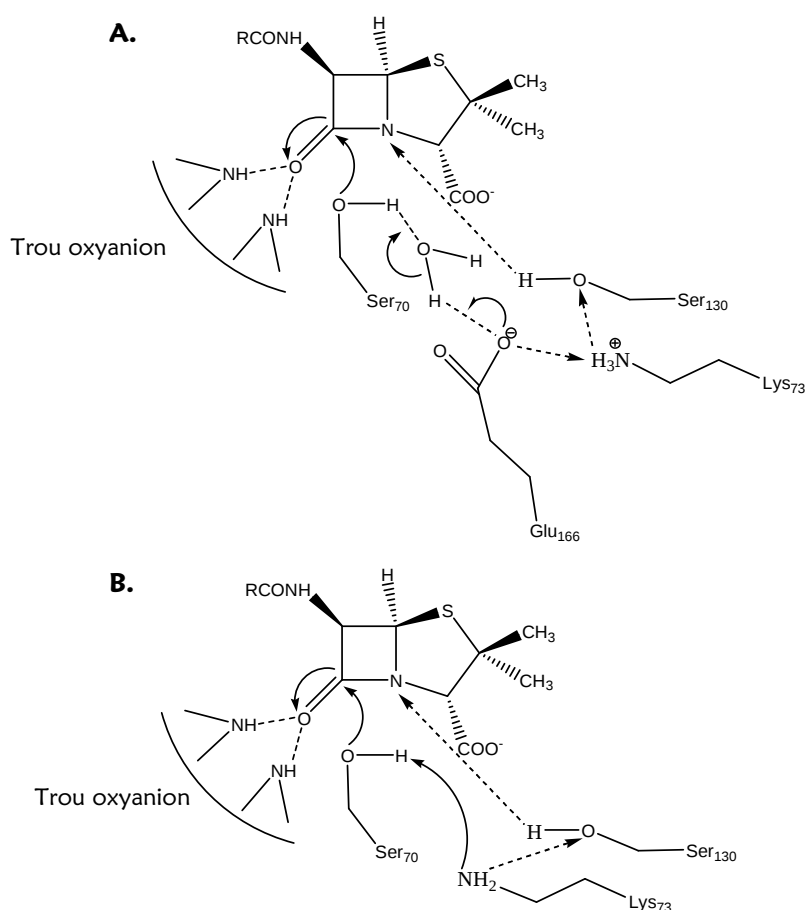


Figure 25 : mécanismes possibles pour la formation de l'intermédiaire tétraédrique dans l'acylation des β -lactamases de classe A. Le rôle de base générale est supporté par le Glu₁₆₆ (A) ou la Lys₇₃ (B). Dans les deux cas, la rupture de l'intermédiaire tétraédrique implique le transfert d'un proton de la base à l'azote du β -lactame (flèches en pointillés) par un chemin où l'hydroxyle de la Ser₁₃₀ est le dernier accepteur/donneur. Pour plus de clarté, les éléments intermédiaires du transfert de proton ne sont pas indiqués (W2, Lys₂₃₄).

Alors que ces deux hypothèses s'affrontent, cette dualité est maintenue par des études de titrage RMN de la lysine qui prouvent que ce résidu est protoné à pH 7, mais en absence de substrat [141] et d'autres qui montrent que le mutant Glu₁₆₆Asn de la TEM-1 voit sa constante d'acylation diminuée seulement d'environ 190 fois alors que la constante de désacylation est diminuée d'un facteur proche de 3.10^8 [142]. Ces derniers résultats suggèrent qu'il existe un mécanisme alternatif à l'acylation catalysée par le Glu₁₆₆. De récentes études *in silico* suggèrent que les deux voies catalytiques peuvent en réalité coexister [143].

Les résidus cités ci-dessus sont impliqués directement dans la catalyse, les résidus des trois motifs ont successivement été mutés afin de déterminer leur rôle [144]. Le motif SXXK discuté précédemment est directement impliqué dans la catalyse. Le motif SDN semblerait jouer un rôle structural pour le maintien de la géométrie du site actif. Outre son rôle dans la catalyse pour le transfert du proton dans la réaction d'acylation, la Ser₁₃₀ fait aussi partie du réseau de liaisons hydrogène impliqué dans la liaison au ligand en maintenant l'acide carboxylique des pénicillines. Le résidu Asp₁₃₁ semble jouer un rôle pour la stabilité de la protéine, et l'Asn₁₃₂ contribue au bon positionnement du substrat par une liaison hydrogène formée avec le carbonyle de la chaîne latérale de la pénicilline. Enfin, le motif KTG comprend la Lys₂₃₄, impliquée dans la liaison au substrat et qui stabilise l'acide carboxylique de l'intermédiaire tétraédrique. Sur base de la structure du site actif, le résidu 235 qui porte un groupement hydroxyle semble être impliqué dans la liaison au substrat. Cependant, des mutants de ce résidu n'ont pu montrer clairement son rôle. Enfin, la glycine 236 est strictement conservée car elle est nécessaire pour garder intacte la géométrie du site actif. Les azotes de la chaîne principale de la Ser₇₀ et du résidu en position 237 (souvent une alanine) forment le trou à oxyanion et vont stabiliser l'oxygène du carbonyle de la pénicilline.

D'autres résidus ne faisant pas partie des motifs conservés semblent importants pour la liaison au substrat, le maintien de la géométrie du site actif, ou sont indirectement impliqués dans la catalyse [135, 144]. Ces résidus ne seront pas revus en détails car certains ne sont pas conservés dans toutes les β -lactamases de classe A, mais quelques exemples sont donnés. L'Arg₂₄₄ est trouvée chez environ 75 % des β -lactamases de classe A. Son groupe guanidinium pointant dans le site actif serait impliqué dans la stabilisation du substrat [145]. Certains résidus de la boucle Ω sont eux aussi importants pour la spécificité de substrat [146]. Le résidu 170 est une asparagine très conservée de la boucle Ω qui en créant une liaison hydrogène avec la molécule d'eau catalytique, jouerait un rôle pour son bon positionnement et donc est importante pour la catalyse [147]. Les résidus situés en aval du motif KTG et les résidus 104 et 105 portés par une boucle proche du site actif, semblent eux aussi impliqués

dans la spécificité de substrat [148-150]. La réaction d'hydrolyse des β -lactames par les β -lactamases de classe A est ainsi un processus complexe qui met en jeu de nombreux résidus, afin de réaliser une catalyse ajustée et atteindre un niveau d'efficacité proche de la perfection.

Les β -lactamases de classe C

Les β -lactamases de classes C étaient originellement appelées "céphalosporinases" pour leur capacité à mieux hydrolyser les céphalosporines. Elles partagent avec les β -lactamases de classe A un mécanisme similaire pour l'hydrolyse des β -lactames. Cependant, il existe des différences au niveau du mécanisme et des résidus impliqués dans la désacylation. La molécule d'eau hydrolytique va, dans ces deux classes d'enzymes, attaquer l'acyl-enzyme par une face différente, par la face β pour les classes C et par la face α pour les classes A. De plus, le mécanisme responsable de l'activation de la molécule d'eau hydrolytique est différent. On a longtemps pensé que la Tyr₁₅₀, selon un processus assisté par l'amine de l'acyl-enzyme, et selon un rôle similaire à celui de l'His des protéases à sérine, possédait cette fonction. Cependant, des études récentes montrent que pour des pH allant jusqu'à 11, cette tyrosine reste neutre et sa mutation en phénylalanine n'altère pas l'activité de l'enzyme, suggérant que la Tyr₁₅₀ ne participe pas directement à la désacylation [151-153]. Il a été proposé que l'azote du noyau thiazoline (après ouverture du β -lactame) est idéalement positionné pour promouvoir l'addition de la molécule d'eau hydrolytique à l'acyl-enzyme et accomplir la désacylation. Ceci serait un exemple de catalyse assistée par le substrat [154]. Enfin, une étude théorique suggère que la Tyr₁₅₀ sous forme déprotonée pourrait être stabilisée dans le site actif par les chaînes latérales des Lys 67 et 315, remettant en jeu le rôle de la Tyr₁₅₀ dans la désacylation [155]. Le résidu impliqué dans le mécanisme d'acylation est lui aussi sujet à controverse [156], une étude récente montre qu'elle serait supportée par un mécanisme similaire à celui des PBPs, à savoir par l'intermédiaire de la Lys₆₇ [157].

La structure tridimensionnelle de ces protéines présente une grande similarité avec celle des β -lactamases de classe A. La superposition de β -lactamases de classe A et C révèle une disposition similaire de nombreux résidus du site actif. Chez les β -lactamases de classe C, les résidus Ser₆₄, Lys₆₇, Lys₃₁₅ et Tyr₁₅₀ correspondent respectivement aux résidus Ser₇₀, Lys₇₃, Lys₂₃₄ et Ser₁₃₀ des β -lactamases de classe A [158].

Les β -lactamases de classe D

Cette classe de β -lactamases était à l'origine appelée "oxacillinases" en référence à leur meilleur substrat, les oxacillines. Dans ces enzymes, l'activation de la sérine pour la catalyse se fait par l'intermédiaire du résidu équivalent à la Lys₇₃, carboxylée par l'addition de CO₂ à l'amine de la chaîne latérale. L'acide carbamique résultant est ionisé pour donner un groupe fonctionnel carbamate en contact par liaison hydrogène avec la sérine catalytique [159].

Le repliement tridimensionnel des β -lactamases de classe D est lui aussi similaire à celui des β -lactamases de classe A [160].

Les métallob- β -lactamases de classe B

Ces enzymes sont des protéines de petite taille, partageant un motif $\alpha\beta\beta\alpha$ avec un sandwich β central et deux hélices α de chaque côté. Ce motif, probablement né d'un événement de duplication est aussi trouvé dans d'autres types d'enzymes, les glyoxalases, et chez certaines flavoenzymes. Ce motif contient un site de liaison intrinsèque aux ions métalliques et est localisé sur le bord du sandwich β . Chez les β -lactamases de classe B, ce site est occupé par un ion zinc divalent, maintenu dans une structure tétraédrique par trois histidines et une molécule d'eau [161]. Il existe un consensus selon lequel la molécule d'eau ligand du Zn²⁺ est le nucléophile qui va permettre l'ouverture du cycle β -lactame par un mécanisme similaire à celui des métalloprotéases à zinc, mais il n'existe pas de preuves directes établissant ce rôle.

L'existence d'un deuxième site de liaison à un ion métallique a été montrée dans toutes les β -lactamases de classe B, leurs rôles respectifs dans la catalyse n'a pas encore été expliqué [162]. Cette classe d'enzymes présente un réel danger de santé publique car elles sont de plus en plus nombreuses et possèdent un large spectre d'activité. Comme leur mécanisme catalytique est complètement différent de celui des autres classes de β -lactamases, notamment, l'intermédiaire covalent acyl-enzyme n'est pas inclus dans le mécanisme, une stratégie complètement différente pour leur inhibition est requise.

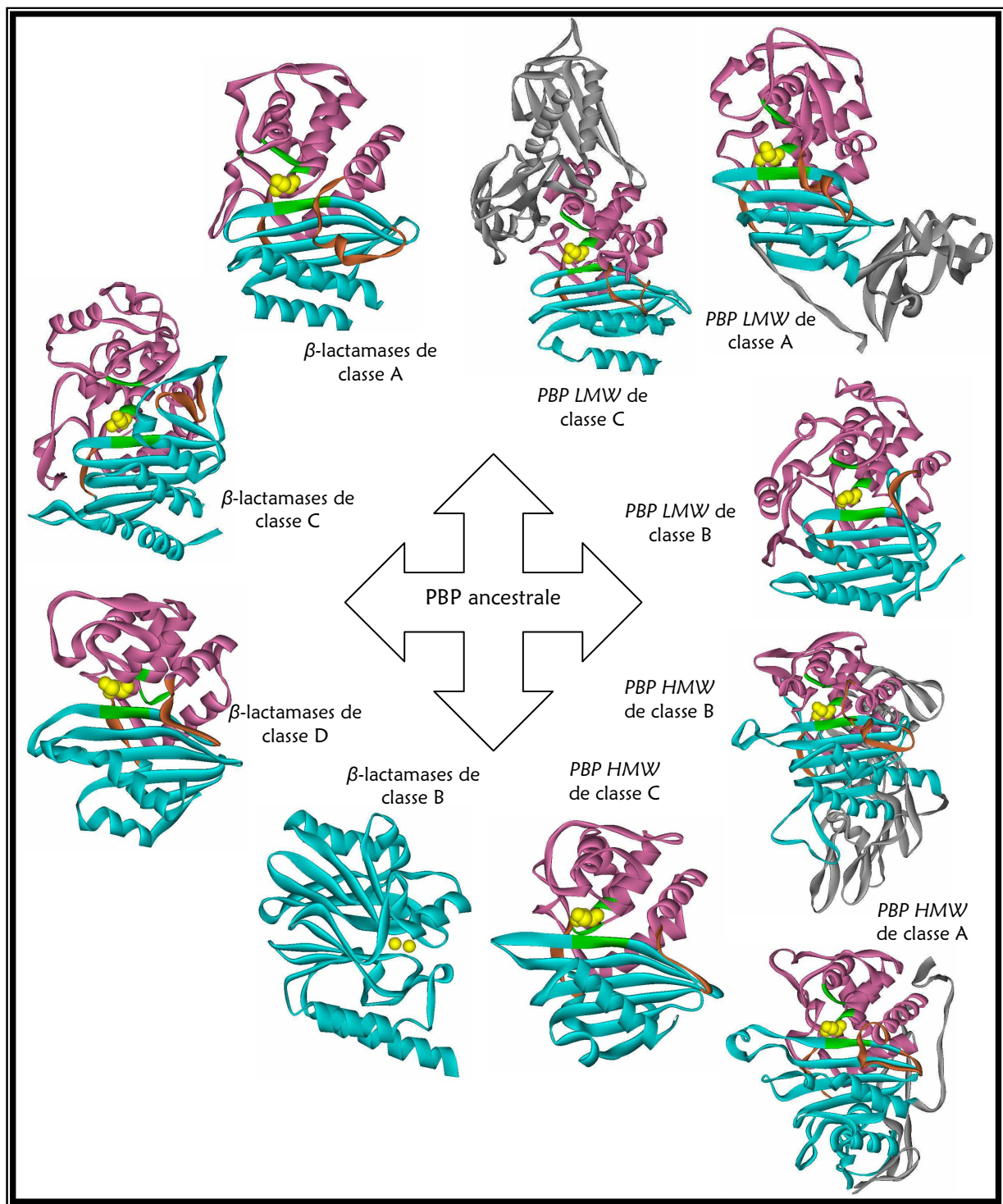


Figure 26 : structure des différentes classes de DD-peptidases et β -lactamases ayant évolué à partir d'une PBP ancestrale. Le domaine tout α est indiqué en rose, le domaine α/β est indiqué en cyan et les connections entre les deux domaines sont indiquées en marron. La serine catalytique est représentée en jaune et les motifs conservés en vert. Les régions ne faisant pas partie du repliement type (insertions ou domaines supplémentaires) sont représentées en gris. Pour les β -lactamases de classe B, les atomes de zinc sont représentés en jaune.

Liens de parenté entre PBP_s et β -lactamases

Il est largement accepté que les PBP_s et les β -lactamases ont évolué de manière divergente à partir d'un même ancêtre commun qui serait une PBP ancestrale. Dans ce paragraphe, nous verrons sur quelles caractéristiques communes des PBP_s et des β -lactamases cette hypothèse est basée, les différentes étapes hypothétiques ayant pu mener d'une PBP à une β -lactamase, et enfin, les tentatives d'évolution d'une PBP en β -lactamase réalisées à ce jour.

Bases de la phylogénie entre les PBP_s et les β -lactamases

Bien que les β -lactamases à sérine et les DD-peptidases forment deux familles d'enzymes, les liens phylogénétiques à l'intérieur même de ces deux familles sont difficiles à établir sur base de leurs séquences qui ont grandement divergé et la similarité est faible. Ainsi, une classification à l'intérieur de ces familles va requérir aussi bien des alignements de structure primaire que des similarités de structure tertiaire, des conservations d'éléments fonctionnels au sein de ces structures (les trois motifs décrits ci-dessus), et aussi des conservations de fonctions et de profils de substrats [117, 163].

Alors qu'à l'intérieur de ces familles d'enzymes l'homologie entre les différentes classes est indétectable, on peut imaginer qu'entre deux familles d'enzymes, le degré de conservation entre classes au niveau des séquences va être aussi faible voire plus. C'est en effet le cas, et il est difficile d'inférer un lien de parenté entre les β -lactamases et les PBP_s sur base de comparaisons de séquences. Néanmoins, pour les différentes classes de ces deux enzymes, un lien de parenté pourrait être révélé par la conservation de la structure tridimensionnelle et des trois motifs fonctionnels. Cette structure conservée est celle décrite pour les β -lactamases de classe A (Figure 26). La deuxième caractéristique importante pour assurer d'une évolution divergente de ces deux familles d'enzymes est la conservation d'une même stratégie pour leur mécanisme catalytique, impliquant une sérine. De plus, le mécanisme par lequel les DD-peptidases sont inhibées par les β -lactames suit le même schéma d'hydrolyse de ces antibiotiques par les β -lactamases (Figure 23), la différence se situant dans l'étape de désacylation qui est lente chez les DD-peptidases et très rapide chez les β -lactamases (Table 4). De cette manière, une β -lactamase issue d'une PBP qui soumise à une forte pression de sélection, aurait acquis la capacité d'hydrolyser les β -lactames. Selon un schéma généralement accepté, une PBP ancestrale aurait divergé pour donner différentes classes de PBP_s ayant elles aussi divergé de manière parallèle pour donner des PBP_s et des β -lactamases apparentées.

Classe de PRP	Protéine	k_2/K_s ou k_{cat}/K_m ($M^{-1}.s^{-1}$)	k_3 ou k_{cat} (s^{-1})
PBPs LMW	<i>Streptomyces</i> R61 ^c	$1,3.10^{4a}$	$1,4.10^{-4a}$
	<i>Actinomadura</i> R39 ^c	3.10^5	3.10^{-6}
	<i>Streptomyces</i> K15 ^c	150	$1,1.10^{-4}$
	PBP4 <i>S. aureus</i> ^c	2.10^3	$1,2.10^{-3}$
	PBP5 <i>B. subtilis</i> ^c	$1,2.10^3$ (4°C)	6.10^{-5}
	PBP4 <i>B. licheniformis</i> ^c	80	$< 6.10^{-5}$
	PBP6 <i>E. coli</i> ^c	175	$2,3.10^{-3}$
	PBP5 <i>E. coli</i> ^b	$1,21.10^4$	$3,3.10^{-3}$
	PBP4 <i>E. coli</i> ^c	7.10^3	$1,8.10^{-4}$
PBPs HMW	PBP3 <i>E. coli</i> ^c	300	$1,8.10^{-4}$
	PBP2 <i>E. coli</i> ^c	130	$1,4.10^{-4}$
	PBP1b <i>E. coli</i> ^c	200	3.10^{-5}
	PBP2x <i>S. pneumoniae</i> ^d	$1,41.10^5$	$2,3.10^{-5}$
	BlaR <i>B. licheniformis</i> ^e	$8,7.10^6$	$1,4.10^{-5}$
β-lactamases de Classe A	<i>B. licheniformis</i>	$2,9.10^7$	$2,2.10^3$
	TEM-1 <i>E. coli</i>	$8,4.10^7$	$1,6.10^4$
	<i>S. aureus</i>	1.10^7	400
β-lactamases de Classe B	<i>B. cereus</i> II	$4,5.10^5$	680
β-lactamases de Classe C	P99 <i>E. cloacae</i>	$3,5.10^7$	15
β-lactamases de Classe D	OXA-10	5.10^6	109

Table 4 : paramètres cinétiques de l'hydrolyse de la pénicilline G par des PBPs et des β -lactamases. Pour les PBPs les paramètres k_2/K_s et k_3 sont donnés à 25°C (a) ou 37°C pour les autres, b tiré de [120], c tirés de [118], d tiré de [164], pour BlaR, l'acylation est donnée par k_2/K' [165]. Pour les β -lactamases, les paramètres k_{cat}/K_m et k_{cat} sont donnés à 30°C [166], pour OXA-10, la réaction a été faite en présence de $NaHCO_3$ [159].

Ceci est supporté par exemple, par des similarités de séquence faibles mais significatives entre la PBP LMW de classe B R61 de *Streptomyces* et la β -lactamase de classe C de *Citrobacter freundii*, entre la PBP LMW de classe A de *Streptomyces* K15 et la β -lactamase de classe A de *S. aureus* ou de manière plus affirmée entre le domaine PB de la PBP HMW de classe C BlaR de *B. licheniformis* et la β -lactamase de classe D Oxa-2 [117]. Selon ce schéma évolutif, chaque classe de β -lactamase serait phylogénétiquement liée à une classe de PBPs. Les cas où les alignements deux à deux entre des PBPs et des β -lactamases se font facilement sont rares et des alignements multiples entre des PBPs et des β -lactamases sont difficiles à

réaliser, voire impossibles, à moins de fixer les motifs consensus conservés. Malgré ces difficultés, un alignement multiple a été réalisé par Massova & Mobashery, donnant une vue possible de la phylogénie entre PBPs et β -lactamases dans la Figure 27 [126].

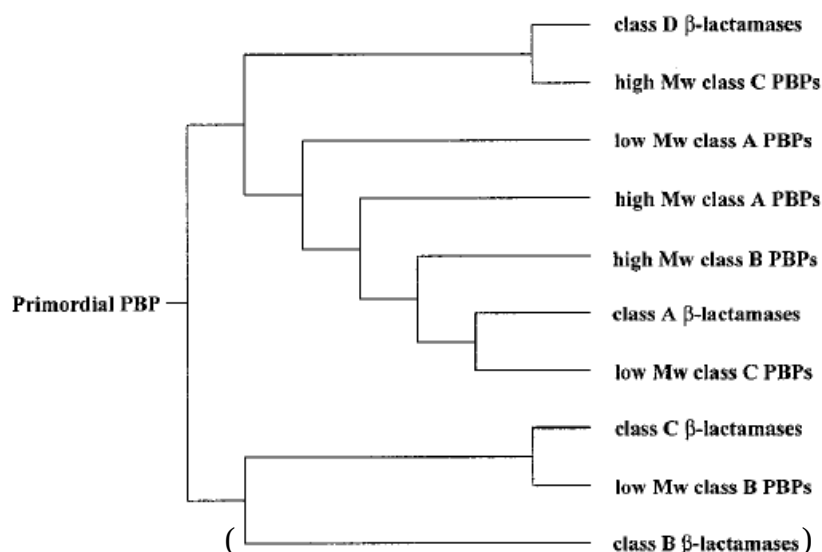


Figure 27 : schéma simplifié de l'alignement multiple de séquences réalisé par Massova & Mobashery. Reproduit de [126].

Le premier point de diversification pour les β -lactamases va donner lieu d'une part aux β -lactamases à sérine et d'autre part aux métallo- β -lactamases. La divergence est grande entre elles, une duplication du domaine α/β d'une PBP ancestrale ayant peut-être évolué vers les β -lactamases de classe B avec une structure tridimensionnelle très différente de celle des autres β -lactamases (Figure 26). Cependant, la relation phylogénique de cette classe d'enzyme avec les glyoxalases en fait probablement un cas d'évolution convergente de la fonction, sans lien phylogénétique avec les autres classes de β -lactamases. Un deuxième point de diversification aurait donné naissance aux branches menant aux β -lactamases de classe C d'un côté et aux β -lactamases de classe D et A de l'autre. Cette hypothèse a été récemment supportée par une étude phylogénique des β -lactamases à sérine en fonction de leur structure, montrant que les β -lactamases de classe A et D sont voisines, avec les β -lactamases de classe C divergeant les premières [167]. Enfin les attributions de liens entre classes de PBPs et de β -lactamases semblent concordantes avec les structures observées en Figure 26, et ce schéma reflète un chemin d'évolution possible pour ces familles d'enzymes. Cependant, alors que dans ce schéma, les PBPs les plus proches des β -lactamases de classe A seraient les PBPs LMW de classe C (comportant certaines PBP4), il a récemment été suggéré que ces β -lactamases auraient directement évolué à partir d'une PBP de type PBP5, appartenant à la classe des

PBPs LMW de classe A. La PBP5 d'*E. coli* présente une des vitesses de désacylation les plus élevées parmi les PBPs et la seule grande différence structurale avec les β -lactamases de classe A se situe au niveau de la boucle Ω plus courte, et dont la séquence et la conformation sont différentes [168]. Ainsi, même si l'allure générale du schéma évolutif proposé par Massova & Mobashery semble correspondre à la réalité, la précision à certains branchements paraît faible, et des comparaisons de structures seraient nécessaires pour affiner ce schéma.

Comment une PBP a pu devenir une β -lactamase ?

Pour qu'une PBP évolue en β -lactamase, elle a dû acquérir la capacité de catalyser la désacylation des acyl-enzymes formés avec des β -lactames. Cette capacité à hydrolyser rapidement les β -lactames est apparue par des mécanismes différents suivant la classe de β -lactamases considérée. Les β -lactamases de classe A utilisent un résidu nucléophile situé sur une boucle oméga optimisée alors que chez les β -lactamases de classe C, la catalyse est assurée par la Tyr du deuxième motif ou est assistée par le substrat. Ceci est une illustration du phénomène de plasticité des sites actifs acquise par duplication et divergence, explicité dans la partie concernant l'évolution naturelle des protéines (voir page 15).

A côté de l'acquisition de l'étape hydrolytique, les β -lactamases naissantes ont dû se distancer des réactions catalysées par les PBPs parentes afin d'être plus efficaces dans leur nouvelle fonction. Ainsi, des éléments structuraux défavorisant les interactions avec le peptidoglycane ont pu être incorporées dans les séquences des β -lactamases. Ces enzymes ont ainsi perdu la capacité de se lier au peptidoglycane [169], et elles catalysent difficilement des réactions d'hydrolyse de petits peptides [170]. L'importance du groupement méthyle de l'avant-dernière D-alanine pour l'évolution des PBPs en β -lactamases a en particulier été mise en évidence. Structurellement, cela se traduit par la présence d'une poche hydrophobe présente chez les PBPs afin d'accepter ce groupement méthyl, et absente chez les β -lactamases [171, 172].

Les β -lactamases ont alors pu, par l'acquisition de mutations fixées par pression de sélection, optimiser leur nouvelle fonction, atteignant pour certaines un niveau de perfection catalytique [126].

Evolution d'une PBP en β -lactamase

Les β -lactamases de classe B sont une classe de β -lactamases particulière puisque ce sont des métallo-enzymes, et elles seraient phylogénétiquement liées à des enzymes de fonctions différentes comme des glyoxalases. Par une stratégie d'incorporation et ajustement d'éléments fonctionnels en conjonction avec de l'évolution dirigée expliquée plus haut (page 30), Park *et al.* ont pu introduire une activité β -lactamase dans une structure de glyoxalase. Bien que l'enzyme générée ait un k_{cat} 150 fois plus faible que la β -lactamase sauvage modèle (IMP-1) et que l'efficacité catalytique soit 3 à 4 ordres de grandeur plus faible, cette enzyme a pu conférer un certain niveau de résistance à la céfotaxime *in vivo* à la bactérie hôte [54].

D'une manière différente, deux groupes ont travaillé sur les relations existant entre une PBP HMW de classe B, la PBP2x de *S. pneumoniae*, et les β -lactamases de classe A. En se basant sur les structures tridimensionnelles des deux types d'enzymes et en particulier des acides aminés présents au niveau du site actif, Chesnel *et al.* ont tenté d'augmenter la vitesse de désacylation de la PBP2x. Pour cela, la Phe₄₅₀ qui occupe spatialement une position similaire au Glu₁₆₆ a été remplacée par divers résidus, incluant deux résidus comportant des acides carboxyliques sur leur chaîne latérale, Asp et Glu. Le variant contenant un acide aspartique désacyle 110 fois plus rapidement que l'enzyme sauvage à pH 8. Cependant, la vitesse de désacylation de la TEM-1 reste 2.10^6 fois plus élevée que celle du mutant Asp. Ainsi, si l'absence de la boucle Ω et la présence de la Phe₄₅₀ à la place du Glu₁₆₆ expliquent certainement pourquoi la PBP2x n'est pas une β -lactamase, il existe d'autres mutations dans la séquence qui doivent moduler la vitesse de désacylation de la protéine [164].

Segovia & Peimbert ont créé une banque combinatoire de PBP2x, c'est à dire une banque composée de variant aléatoires de 10 acides aminés impliqués dans le mécanisme catalytique de l'enzyme ou dans la spécificité de substrat. Puis, la région codant pour le domaine transpeptidase a été amplifiée par PCR mutagénisante, permettant de créer des mutations de manière aléatoire. Il a ainsi été possible de montrer que seulement 3 mutations sont nécessaires pour augmenter le niveau de résistance à la céfotaxime de 10 fois. Aussi, le produit de purification d'un des mutants a une vitesse de désacylation 10^5 fois supérieure à celle de l'enzyme sauvage. Cependant, on n'arrive qu'à un k_{cat} de $10^{-2} s^{-1}$ avec la céfotaxime, et ce mutant ne confère pas de résistance *in vivo* à l'ampicilline ce qui peut s'expliquer par un k_{cat} de seulement $5,22.10^{-2} s^{-1}$, enfin l'efficacité de l'enzyme pour ce substrat est dix fois plus faible qu'avec la céfotaxime. Le mutant généré n'est donc pas une β -lactamase et encore

moins une β -lactamase de classe A dont les meilleurs substrats sont des pénicillines, mais un intermédiaire possible dans l'évolution des PBPs en β -lactamases.

Ces expériences prouvent la difficulté d'évoluer une PBP en β -lactamase, en particulier, si le substrat choisi est une pénicilline, l'adaptation du site actif des PBPs à ces types de substrats semblant requérir de plus grands changements que pour s'adapter aux substrats de type céphalosporines. De nombreux essais ont probablement échoué car un tour de mutagenèse ne permet pas de réaliser des sélections *in vivo* qui requièrent une activité minimale de l'enzyme évoluée 10^5 à 10^6 fois plus élevée qu'une PBP.

Objectifs et Stratégie

Objectifs

Le processus de sélection naturelle d'une protéine permet l'évolution progressive de ses propriétés (stabilité, reconnaissance moléculaire...) et/ou de son activité enzymatique. L'ingénierie de la reconnaissance moléculaire est un domaine dans lequel les chercheurs ont énormément progressé, en particulier avec l'ingénierie des anticorps. Le premier anticorps thérapeutique a été approuvé en 1986 par la "Food and Drug Administration" (FDA) pour traiter les rejets de greffe. Depuis lors, l'ingénierie des anticorps a évolué de manière spectaculaire pour fournir des traitements thérapeutiques de plus en plus adaptés, en particulier en diminuant les réactions immunitaires générées contre ces anticorps, en les "humanisant" [173, 174].

Aux antipodes des avancées de l'ingénierie de la reconnaissance moléculaire, se situe l'ingénierie de la catalyse, qui, malgré des résultats ponctuels encourageants (voir chapitre "Ingénierie des protéines : Approches stratégiques et applications", page 25), reste très difficile à réaliser. Deux raisons sont probablement à l'origine de ces échecs. D'une part, même si les mécanismes catalytiques sont généralement connus, ils comportent souvent des intermédiaires d'états de transition dont la stabilisation demande une certaine dynamique du site actif, liée à la dynamique de toute la protéine. Cette dynamique est mal comprise et donc difficile à recréer de manière artificielle. D'autre part, l'évolution d'activités enzymatiques nécessite, comme vu au point "Criblage et sélections", page 54, des méthodes de sélection ou de criblage robustes qui nous font défaut.

Les objectifs de cette thèse entrent dans le cadre de l'ingénierie d'activités enzymatiques. En prenant comme système modèle celui des DD-peptidases et des β -lactamases, nous allons tenter de créer *in vitro*, une activité β -lactamases à partir d'une DD-peptidase. Malgré le grand nombre d'études structurales et mécanistiques, l'origine de la grande différence de vitesse de désacylation entre ces deux enzymes n'est pas comprise. Plusieurs tentatives de transformation d'une PBP en β -lactamase par mutagenèse aléatoire et sélection *in vivo* (résistance aux pénicillines) ont échoué, y compris dans notre laboratoire. Les raisons de ces échecs sont doubles. D'une part, avec moins de 20 % d'identité de séquence, les DD-peptidases utilisées pour ces tentatives sont très éloignées des β -lactamases. D'autre part, l'activité minimale nécessaire pour permettre une sélection *in vivo* est très élevée. En effet, pour obtenir une résistance à l'ampicilline, la vitesse de désacylation d'une DD-peptidase naturelle devrait augmenter d'au moins quatre ordres de grandeur, ce qui est très peu probable

après seulement un tour de mutagenèse aléatoire. Ainsi, la sélection d'activités enzymatiques faibles, est probablement un passage obligé si l'on souhaite créer des activités catalytiques *de novo*. La sélection de telles activités enzymatiques est actuellement très difficile et demande l'élaboration de nouvelles stratégies. Notre objectif était de simuler l'évolution naturelle en cherchant à contourner les difficultés mentionnées ci-dessus, inhérentes à l'ingénierie de la catalyse.

Stratégie

Il était prévu initialement d'explorer les possibilités d'évolution génétique avec une DD-peptidase modèle, la PBP4 d'*Escherichia coli*, plus proche des β -lactamases de classe A. Avant de commencer la recherche avec la PBP4, nous avons décidé de réévaluer le choix de cette seconde enzyme modèle en examinant les résultats récents des séquençages de génomes divers dans l'espoir de trouver une DD-peptidase dont la séquence était un peu moins éloignée de celle des β -lactamases de classe A que la PBP4. Les alignements de séquences entre celle-ci et les β -lactamases restent très difficiles. La présence de trois courts motifs conservés (SXXK, SDN et KS/TG) et la similitude de la structure tridimensionnelle sont les principaux témoins de la relation phylogénétique.

De cette manière, nous avons identifié, dans des génomes de cyanobactéries récemment séquencés, des gènes encodant des protéines annotées comme des β -lactamases de classe A sur base des similarités de séquence. Les séquences de ces protéines possèdent les trois motifs mentionnés ci-dessus. Cependant, de manière surprenante, ces protéines présentent une région d'environ 15 acides aminés qui est très mal alignée avec ces β -lactamases. Dans la structure, cette région correspond à la boucle Ω des β -lactamases qui porte le glutamate essentiel à la catalyse et plus particulièrement à l'étape de désacylation. Ce glutamate est absent dans les protéines identifiées et de plus, leur boucle Ω est plus courte de six résidus. Ces observations suggèrent que ces protéines ne sont pas des β -lactamases mais des PBPs et que cette famille de PBPs est phylogénétiquement la plus proche connue à ce jour des β -lactamases de classe A.

Notre objectif, à ce stade, a été d'identifier la fonction de ces protéines et d'en faire évoluer une vers l'activité β -lactamase. Afin de caractériser ces fonctions, trois choix de protéines ont été envisagés. Des résultats préliminaires montrant que la protéine de *Thermosynechococcus elongatus* est une PBP très stable et donc avec un potentiel évolutif probablement grand [175], nous ont encouragé à poursuivre la caractérisation plus détaillée de cette enzyme. Dans un deuxième temps, nous nous sommes concentrés sur l'évolution de cette protéine en β -

lactamase. Pour cela, nous avons d'abord considéré une approche dirigée en mutant quelques résidus du site actif, nous permettant de définir une orientation possible pour l'évolution de cette enzyme. La mutation d'un résidu permet d'augmenter d'un facteur 90 l'activité d'hydrolyse de l'acyl-enzyme, cette mutation pourrait donc servir de base pour l'évolution dirigée de la protéine en β -lactamase.

En parallèle, des essais de cristallisation de la protéine (C. Evrard et J. P. Declercq) ont été fructueux. Contrairement à nos attentes, l'analyse fine de la structure aussi bien de la protéine sauvage que mutante sous leur forme d'apo-enzymes ou acylées avec un substrat pénicilline ne nous a pas permis de cibler une zone de la protéine rationnellement susceptible de diriger l'activité de l'enzyme vers une activité de type β -lactamase.

Ainsi, deux approches d'évolution dirigée ont été envisagées. La première consiste en la construction d'une banque de mutants par PCR mutagénéisante sur le gène de la protéine mutante. La deuxième consiste à insérer une boucle Ω dégénérée plus longue de 6 acides aminés dans la protéine sauvage. La première banque pourra nous apporter des informations sur les résidus situés ou non dans le site actif, importants pour la catalyse de la réaction. La deuxième banque répondra à la question : « est-ce qu'une boucle Ω complète est nécessaire pour atteindre une activité de type β -lactamase ? ». Pour la construction de ces banques, nous avons choisi de travailler avec un vecteur de type phagemide qui restreint le nombre de copies de la protéine de fusion présentées en surface du phage et permet de limiter d'éventuels problèmes de toxicité de la protéine. Le système phagemide utilisé nous permettra aussi facilement de passer d'un système d'expression en surface de phages à un système d'expression ou surexpression de la protéine non fusionnée à la pIII dans le périplasma, suivant la souche bactérienne utilisée. De cette manière, la stratégie de sélection d'enzymes actives passe dans un premier temps par une méthode de sélection *in vivo*. L'acquisition d'une activité suffisante par les enzymes pour être sélectionnées *in vivo* risque cependant d'être difficile. En effet, ceci demanderait d'augmenter l'activité catalytique de l'enzyme d'environ un facteur 1000, qui est possible mais peu probable en un tour de mutagenèse. C'est pourquoi nous pensons utiliser l'alternative de sélections par "phage display". Le développement de nouvelles méthodes de sélection *in vitro* devait permettre de sélectionner dans une collection de mutants ceux qui ont acquis une activité β -lactamase même relativement faible, autrement dit des mutants qui forment des acyl-enzymes qui s'hydrolysent plus vite. Au départ, nous disposons de techniques développées dans notre laboratoire et basées sur la capture d'enzymes actives par des substrats suicides biotinylés [94, 164, 176]. Nous envisageons

d'utiliser une méthode développée en parallèle de notre projet par F. Chang-Pi-Hin et visant à améliorer ces techniques et éventuellement de développer des techniques alternatives plus performantes. Dans le cas où le développement de ces techniques ne serait pas satisfaisant, nous envisagions de développer une stratégie en deux étapes où des enzymes ayant acquis une meilleure capacité à se lier aux pénicillines seraient sélectionnées par "phage display" via l'utilisation de substrats biotinylés, dans un deuxième temps, la deuxième partie de la réaction serait améliorée par un deuxième tour de mutagenèse.

Bibliographie

1. Orengo, C.A. and J.M. Thornton, *Protein families and their evolution-a structural perspective*. Annu Rev Biochem, 2005. 74: p. 867-900.
2. Chen, L., A.L. DeVries, and C.H. Cheng, *Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish*. Proc Natl Acad Sci U S A, 1997. 94(8): p. 3811-6.
3. Chothia, C., et al., *Evolution of the protein repertoire*. Science, 2003. 300(5626): p. 1701-3.
4. Lutz, S. and S.J. Benkovic, Chapter 9 - *Engineering Protein Evolution*, in *Directed Molecular Evolution of Proteins or how to improve enzymes for biocatalysis*, S. Brakmann and K. Johnsson, Editors. 2002, Wiley-VCH: Weinheim, Germany. p. 177-186.
5. Jensen, R.A., *Enzyme recruitment in evolution of new function*. Annu Rev Microbiol, 1976. 30: p. 409-25.
6. Roodveldt, C. and D.S. Tawfik, *Shared promiscuous activities and evolutionary features in various members of the amidohydrolase superfamily*. Biochemistry, 2005. 44(38): p. 12728-36.
7. Khersonsky, O., C. Roodveldt, and D.S. Tawfik, *Enzyme promiscuity: evolutionary and mechanistic aspects*. Curr Opin Chem Biol, 2006. 10(5): p. 498-508.
8. Gerlt, J.A. and P.C. Babbitt, *Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies*. Annu Rev Biochem, 2001. 70: p. 209-46.
9. Jurgens, C., et al., *Directed evolution of a (beta alpha)₈-barrel enzyme to catalyze related reactions in two different metabolic pathways*. Proc Natl Acad Sci U S A, 2000. 97(18): p. 9925-30.
10. Horowitz, N.H., *On the Evolution of Biochemical Syntheses*. Proc Natl Acad Sci U S A, 1945. 31(6): p. 153-7.
11. Todd, A.E., C.A. Orengo, and J.M. Thornton, *Plasticity of enzyme active sites*. Trends Biochem Sci, 2002. 27(8): p. 419-26.
12. Yew, W.S., et al., *Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase suprafamily: mechanistic evidence for a proton relay system in the active site of 3-keto-L-gulonate 6-phosphate decarboxylase*. Biochemistry, 2004. 43(21): p. 6427-37.
13. Wise, E., et al., *Homologous (beta/alpha)₈-barrel enzymes that catalyze unrelated reactions: orotidine 5'-monophosphate decarboxylase and 3-keto-L-gulonate 6-phosphate decarboxylase*. Biochemistry, 2002. 41(12): p. 3861-9.
14. Wymer, N., et al., *Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase from Escherichia coli*. Structure, 2001. 9(1): p. 1-9.
15. Cordes, M.H., et al., *An evolutionary bridge to a new protein fold*. Nat Struct Biol, 2000. 7(12): p. 1129-32.
16. Luo, X., et al., *The Mad2 spindle checkpoint protein has two distinct natively folded states*. Nat Struct Mol Biol, 2004. 11(4): p. 338-45.
17. Tompa, P., C. Szasz, and L. Buday, *Structural disorder throws new light on moonlighting*. Trends Biochem Sci, 2005. 30(9): p. 484-9.
18. Yanagawa, T., et al., *Novel roles of the autocrine motility factor/phosphoglucose isomerase in tumor malignancy*. Endocr Relat Cancer, 2004. 11(4): p. 749-59.

19. Dulhanty, A.M., X.B. Chang, and J.R. Riordan, *Mutation of potential phosphorylation sites in the recombinant R domain of the cystic fibrosis transmembrane conductance regulator has significant effects on domain conformation*. Biochem Biophys Res Commun, 1995. **206**(1): p. 207-14.
20. Kriwacki, R.W., et al., *Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity*. Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11504-9.
21. Kinch, L.N. and N.V. Grishin, *Evolution of protein structures and functions*. Curr Opin Struct Biol, 2002. **12**(3): p. 400-8.
22. Andreeva, A. and A.G. Murzin, *Evolution of protein fold in the presence of functional constraints*. Curr Opin Struct Biol, 2006. **16**(3): p. 399-408.
23. Krishna, S.S. and N.V. Grishin, *Structurally analogous proteins do exist!* Structure, 2004. **12**(7): p. 1125-7.
24. Keefe, A.D. and J.W. Szostak, *Functional proteins from a random-sequence library*. Nature, 2001. **410**(6829): p. 715-8.
25. Lo Surdo, P., M.A. Walsh, and M. Sollazzo, *A novel ADP- and zinc-binding fold from function-directed in vitro evolution*. Nat Struct Mol Biol, 2004. **11**(4): p. 382-3.
26. Theobald, D.L. and D.S. Wuttke, *Divergent evolution within protein superfolds inferred from profile-based phylogenetics*. J Mol Biol, 2005. **354**(3): p. 722-37.
27. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, *Evolution of enzyme superfamilies*. Curr Opin Chem Biol, 2006. **10**(5): p. 492-7.
28. Pegg, S.C., et al., *Representing structure-function relationships in mechanistically diverse enzyme superfamilies*. Pac Symp Biocomput, 2005: p. 358-69.
29. Yuan, L., et al., *Laboratory-directed protein evolution*. Microbiol Mol Biol Rev, 2005. **69**(3): p. 373-92.
30. Kaur, J. and R. Sharma, *Directed evolution: an approach to engineer enzymes*. Crit Rev Biotechnol, 2006. **26**(3): p. 165-99.
31. Nixon, A.E., M. Ostermeier, and S.J. Benkovic, *Hybrid enzymes: manipulating enzyme design*. Trends Biotechnol, 1998. **16**(6): p. 258-64.
32. Cedrone, F., A. Menez, and E. Quemeneur, *Tailoring new enzyme functions by rational redesign*. Curr Opin Struct Biol, 2000. **10**(4): p. 405-10.
33. Hytonen, V.P., et al., *Design and construction of highly stable, protease-resistant chimeric avidins*. J Biol Chem, 2005. **280**(11): p. 10228-33.
34. Heckmann-Pohl, D.M., et al., *Improvement of the fungal enzyme pyranose 2-oxidase using protein engineering*. J Biotechnol, 2006. **124**(1): p. 26-40.
35. Pettersson, P.L., A.S. Johansson, and B. Mannervik, *Transmutation of human glutathione transferase A2-2 with peroxidase activity into an efficient steroid isomerase*. J Biol Chem, 2002. **277**(33): p. 30019-22.
36. Hederos, S., et al., *Incorporation of a single His residue by rational design enables thiol-ester hydrolysis by human glutathione transferase A1-1*. Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13163-7.
37. Nixon, A.E., et al., *Rational design of a scytalone dehydratase-like enzyme using a structurally homologous protein scaffold*. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3568-71.
38. Looger, L.L., et al., *Computational design of receptor and sensor proteins with novel functions*. Nature, 2003. **423**(6936): p. 185-90.
39. Dwyer, M.A., L.L. Looger, and H.W. Hellinga, *Computational design of a biologically active enzyme*. Science, 2004. **304**(5679): p. 1967-71.
40. Butterfoss, G.L. and B. Kuhlman, *Computer-based design of novel protein structures*. Annu Rev Biophys Biomol Struct, 2006. **35**: p. 49-65.

41. Kuhlman, B., et al., Design of a novel globular protein fold with atomic-level accuracy. *Science*, 2003. **302**(5649): p. 1364-8.
42. Johannes, T.W. and H. Zhao, Directed evolution of enzymes and biosynthetic pathways. *Curr Opin Microbiol*, 2006. **9**(3): p. 261-7.
43. Dalby, P.A., Optimising enzyme function by directed evolution. *Curr Opin Struct Biol*, 2003. **13**(4): p. 500-5.
44. Tao, H. and V.W. Cornish, Milestones in directed enzyme evolution. *Curr Opin Chem Biol*, 2002. **6**(6): p. 858-64.
45. Chica, R.A., N. Doucet, and J.N. Pelletier, Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol*, 2005. **16**(4): p. 378-84.
46. Bloom, J.D., et al., Evolving strategies for enzyme engineering. *Curr Opin Struct Biol*, 2005. **15**(4): p. 447-52.
47. Miyazaki, K. and F.H. Arnold, Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol*, 1999. **49**(6): p. 716-20.
48. Voigt, C.A., et al., Computational method to reduce the search space for directed protein evolution. *Proc Natl Acad Sci U S A*, 2001. **98**(7): p. 3778-83.
49. Hayes, R.J., et al., Combining computational and experimental screening for rapid optimization of protein properties. *Proc Natl Acad Sci U S A*, 2002. **99**(25): p. 15926-31.
50. Park, H.S., et al., Design and evolution of new catalytic activity with an existing protein scaffold. *Science*, 2006. **311**(5760): p. 535-8.
51. Vick, J.E., D.M. Schmidt, and J.A. Gerlt, Evolutionary potential of (beta/alpha)₈-barrels: in vitro enhancement of a "new" reaction in the enolase superfamily. *Biochemistry*, 2005. **44**(35): p. 11722-9.
52. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt, Mechanism of Protein Evolution and Their Application to Protein Engineering. *Advances in Enzymology and Related Areas of Molecular Biology*, ed. E. Toone. Vol. 75. 2007: Wiley & Sons. 193-239.
53. Schmidt, D.M., et al., Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry*, 2003. **42**(28): p. 8387-93.
54. Griswold, K.E., et al., The Evolution of Catalytic Efficiency and Substrate Promiscuity in Human Theta Class 1-1 Glutathione Transferase. *J Mol Biol*, 2006.
55. O'Loughlin T, L., W.M. Patrick, and I. Matsumura, Natural history as a predictor of protein evolvability. *Protein Eng Des Sel*, 2006. **19**(10): p. 439-42.
56. Neylon, C., Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res*, 2004. **32**(4): p. 1448-59.
57. Kunkel, T.A., Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A*, 1985. **82**(2): p. 488-92.
58. Weiner, M.P., et al., A method for the site-directed mono- and multi-mutagenesis of double-stranded DNA. *Gene*, 1993. **126**(1): p. 35-41.
59. Cadwell, R.C. and G.F. Joyce, Randomization of genes by PCR mutagenesis. *PCR Methods Appl*, 1992. **2**(1): p. 28-33.
60. Stemmer, W.P., Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 1994. **370**(6488): p. 389-91.
61. Wong, T.S., et al., A statistical analysis of random mutagenesis methods used for directed protein evolution. *J Mol Biol*, 2006. **355**(4): p. 858-71.

62. Lutz, S. and W.M. Patrick, *Novel methods for directed evolution of enzymes: quality, not quantity*. Curr Opin Biotechnol, 2004. **15**(4): p. 291-7.
63. Pelletier, J.N., *A RACHITT for our toolbox*. Nat Biotechnol, 2001. **19**(4): p. 314-5.
64. Ostermeier, M., et al., *Combinatorial protein engineering by incremental truncation*. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3562-7.
65. Hallet, B., D.J. Sherratt, and F. Hayes, *Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein*. Nucleic Acids Res, 1997. **25**(9): p. 1866-7.
66. Fujii, R., M. Kitaoka, and K. Hayashi, *RAISE: a simple and novel method of generating random insertion and deletion mutations*. Nucleic Acids Res, 2006. **34**(4): p. e30.
67. Wong, T.S., D. Zhurina, and U. Schwaneberg, *The diversity challenge in directed protein evolution*. Comb Chem High Throughput Screen, 2006. **9**(4): p. 271-88.
68. Doi, N. and H. Yanagawa, *Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries*. Comb Chem High Throughput Screen, 2001. **4**(6): p. 497-509.
69. Kondo, A. and M. Ueda, *Yeast cell-surface display--applications of molecular display*. Appl Microbiol Biotechnol, 2004. **64**(1): p. 28-40.
70. Prasanna, G.L., T. Panda, and P.P. Rao, *Transformation of intact cells of Saccharomyces cerevisiae by square wave pulses using castellated microelectrodes*. Bioprocess Engineering, 1997. **16**(5): p. 265-268.
71. Tawfik, D.S. and A.D. Griffiths, *Man-made cell-like compartments for molecular evolution*. Nat Biotechnol, 1998. **16**(7): p. 652-6.
72. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface*. Science, 1985. **228**(4705): p. 1315-7.
73. Barbas III, C.F., et al., *Phage Display: A laboratory manual*. 2001: Cold Spring Harbor Laboratory Press.
74. Sidhu, S.S., *Engineering M13 for phage display*. Biomol Eng, 2001. **18**(2): p. 57-63.
75. Kehoe, J.W. and B.K. Kay, *Filamentous phage display in the new millennium*. Chem Rev, 2005. **105**(11): p. 4056-72.
76. Lipovsek, D. and A. Pluckthun, *In-vitro protein evolution by ribosome display and mRNA display*. J Immunol Methods, 2004. **290**(1-2): p. 51-67.
77. Odegrip, R., et al., *CIS display: In vitro selection of peptides from libraries of protein-DNA complexes*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2806-10.
78. Griffiths, A.D. and D.S. Tawfik, *Miniaturising the laboratory in emulsion droplets*. Trends Biotechnol, 2006. **24**(9): p. 395-402.
79. Schmidt-Dannert, C. and F.H. Arnold, *Directed evolution of industrial enzymes*. Trends Biotechnol, 1999. **17**(4): p. 135-6.
80. Kumar, R.A. and D.S. Clark, *High-throughput screening of biocatalytic activity: applications in drug discovery*. Curr Opin Chem Biol, 2006. **10**(2): p. 162-8.
81. Goddard, J.P. and J.L. Reymond, *Enzyme assays for high-throughput screening*. Curr Opin Biotechnol, 2004. **15**(4): p. 314-22.
82. Fastrez, J., *In vivo versus in vitro screening or selection for catalytic activity in enzymes and abzymes*. Mol Biotechnol, 1997. **7**(1): p. 37-55.
83. Polizzi, K.M., et al., *Pooling for improved screening of combinatorial libraries for directed evolution*. Biotechnol Prog, 2006. **22**(4): p. 961-7.
84. Amstutz, P., et al., *In vitro display technologies: novel developments and applications*. Curr Opin Biotechnol, 2001. **12**(4): p. 400-5.
85. Brekke, O.H. and G.A. Loset, *New technologies in therapeutic antibody development*. Curr Opin Pharmacol, 2003. **3**(5): p. 544-50.

86. Jestin, J.L. and P.A. Kaminski, Directed enzyme evolution and selections for catalysis based on product formation. *J Biotechnol*, 2004. **113**(1-3): p. 85-103.
87. Baker, K., et al., Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A*, 2002. **99**(26): p. 16537-42.
88. van Sint Fiet, S., J.B. van Beilen, and B. Witholt, Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci U S A*, 2006. **103**(6): p. 1693-8.
89. Fernandez-Gacio, A., M. Uguen, and J. Fastrez, Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol*, 2003. **21**(9): p. 408-14.
90. Vanwetswinkel, S., B. Avalle, and J. Fastrez, Selection of beta-lactamases and penicillin binding mutants from a library of phage displayed TEM-1 beta-lactamase randomly mutated in the active site omega-loop. *J Mol Biol*, 2000. **295**(3): p. 527-40.
91. Strobel, H., D. Ladant, and J.L. Jestin, In vitro selection for enzymatic activity: a model study using adenylate cyclase. *J Mol Biol*, 2003. **332**(1): p. 1-7.
92. Henry, A.A. and F.E. Romesberg, The evolution of DNA polymerases with novel activities. *Curr Opin Biotechnol*, 2005. **16**(4): p. 370-7.
93. Fa, M., et al., Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J Am Chem Soc*, 2004. **126**(6): p. 1748-54.
94. Ponsard, I., et al., Selection of metalloenzymes by catalytic activity using phage display and catalytic elution. *Chembiochem*, 2001. **2**(4): p. 253-9.
95. Cesaro-Tadic, S., et al., Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library. *Nat Biotechnol*, 2003. **21**(6): p. 679-85.
96. Becker, S., et al., Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr Opin Biotechnol*, 2004. **15**(4): p. 323-9.
97. Varadarajan, N., et al., Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci U S A*, 2005. **102**(19): p. 6855-60.
98. Griffiths, A.D. and D.S. Tawfik, Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *Embo J*, 2003. **22**(1): p. 24-35.
99. Mastrobattista, E., et al., High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem Biol*, 2005. **12**(12): p. 1291-300.
100. Bourne, D.G., et al., Characterisation of a gene cluster involved in bacterial degradation of the cyanobacterial toxin microcystin LR. *Environ Toxicol*, 2001. **16**(6): p. 523-34.
101. Asano, Y., et al., An alkaline D-stereospecific endopeptidase with beta-lactamase activity from *Bacillus cereus*. *J Biol Chem*, 1996. **271**(47): p. 30256-62.
102. Komeda, H. and Y. Asano, Gene cloning, nucleotide sequencing, and purification and characterization of the D-stereospecific amino-acid amidase from *Ochrobactrum anthropi* SV3. *Eur J Biochem*, 2000. **267**(7): p. 2028-35.
103. Petersen, E.I., et al., A novel esterase from *Burkholderia gladioli* which shows high deacetylation activity on cephalosporins is related to beta-lactamases and DD-peptidases. *J Biotechnol*, 2001. **89**(1): p. 11-25.
104. Asano, Y., et al., Structural similarity of D-aminopeptidase to carboxypeptidase DD and beta-lactamases. *Biochemistry*, 1992. **31**(8): p. 2316-28.
105. Raskin, C., et al., BlaB, a protein involved in the regulation of *Streptomyces cacaoi* beta-lactamases, is a penicillin-binding protein. *Cell Mol Life Sci*, 2003. **60**(7): p. 1460-9.
106. Ghuysen, J.M., Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriol Rev*, 1968. **32**(4 Pt 2): p. 425-64.

107. Goffin, C. and J.M. Ghuysen, Biochemistry and comparative genomics of SxxK superfamily acyltransferases offer a clue to the mycobacterial paradox: presence of penicillin-susceptible target proteins versus lack of efficiency of penicillin as therapeutic agent. *Microbiol Mol Biol Rev*, 2002. **66**(4): p. 702-38, table of contents.
108. Vollmer, W. and J.V. Holtje, *The architecture of the murein (peptidoglycan) in gram-negative bacteria: vertical scaffold or horizontal layer(s)?* *J Bacteriol*, 2004. **186**(18): p. 5978-87.
109. Meroueh, S.O., et al., *Three-dimensional structure of the bacterial cell wall peptidoglycan*. *Proc Natl Acad Sci U S A*, 2006. **103**(12): p. 4404-9.
110. Young, K.D., *Bacterial shape*. *Mol Microbiol*, 2003. **49**(3): p. 571-80.
111. Vollmer, W. and J.V. Holtje, *Morphogenesis of Escherichia coli*. *Curr Opin Microbiol*, 2001. **4**(6): p. 625-33.
112. Nanninga, N., *Morphogenesis of Escherichia coli*. *Microbiol Mol Biol Rev*, 1998. **62**(1): p. 110-29.
113. Popham, D.L. and K.D. Young, *Role of penicillin-binding proteins in bacterial cell morphogenesis*. *Curr Opin Microbiol*, 2003. **6**(6): p. 594-9.
114. J. M. Frere, M.N.-D., J. Coyette and B. Joris, *Mode of action: interaction with the penicillin binding proteins*, in *The chemistry of beta-lactams*, B.A.a. Professional, Editor. 1992. p. 149-97.
115. Granier, B., et al., *Serine-type D-Ala-D-Ala peptidases and penicillin-binding proteins*. *Methods Enzymol*, 1994. **244**: p. 249-66.
116. van der Linden, M.P., et al., *Site-directed mutagenesis of proposed active-site residues of penicillin-binding protein 5 from Escherichia coli*. *Biochem J*, 1994. **303 (Pt 2)**: p. 357-62.
117. Duez, C., et al., *Purification and characterization of PBP4a, a new low-molecular-weight penicillin-binding protein from Bacillus subtilis*. *J Bacteriol*, 2001. **183**(5): p. 1595-9.
118. Nicholas, R.A., et al., *Crystal structure of wild-type penicillin-binding protein 5 from Escherichia coli: implications for deacylation of the acyl-enzyme complex*. *J Biol Chem*, 2003. **278**(52): p. 52826-33.
119. Stefanova, M.E., et al., *Neisseria gonorrhoeae penicillin-binding protein 3 exhibits exceptionally high carboxypeptidase and beta-lactam binding activities*. *Biochemistry*, 2003. **42**(49): p. 14614-25.
120. Ghuysen, J.M., *Serine beta-lactamases and penicillin-binding proteins*. *Annu Rev Microbiol*, 1991. **45**: p. 37-67.
121. Young, K.D., *Approaching the physiological functions of penicillin-binding proteins in Escherichia coli*. *Biochimie*, 2001. **83**(1): p. 99-102.
122. Goffin, C. and J.M. Ghuysen, *Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs*. *Microbiol Mol Biol Rev*, 1998. **62**(4): p. 1079-93.
123. Massova, I. and S. Mobashery, *Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases*. *Antimicrob Agents Chemother*, 1998. **42**(1): p. 1-17.
124. Tipper, D.J. and J.L. Strominger, *Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine*. *Proc Natl Acad Sci U S A*, 1965. **54**(4): p. 1133-41.
125. Wilke, M.S., A.L. Lovering, and N.C. Strynadka, *Beta-lactam antibiotic resistance: a current structural perspective*. *Curr Opin Microbiol*, 2005. **8**(5): p. 525-33.
126. Hakenbeck, R. and J. Coyette, *Resistant penicillin-binding proteins*. *Cell Mol Life Sci*, 1998. **54**(4): p. 332-40.

127. Salyers, A.A. and C.F. Amabile-Cuevas, *Why are antibiotic resistance genes so resistant to elimination?* Antimicrob Agents Chemother, 1997. **41**(11): p. 2321-5.
128. Massova, I. and P.A. Kollman, *pKa, MM, and QM studies of mechanisms of beta-lactamases and penicillin-binding proteins: acylation step.* J Comput Chem, 2002. **23**(16): p. 1559-76.
129. Murphy, B.P. and R.F. Pratt, *Evidence for an oxyanion hole in serine beta-lactamases and DD-peptidases.* Biochem J, 1988. **256**(2): p. 669-72.
130. Ambler, R.P., et al., *A standard numbering scheme for the class A beta-lactamases.* Biochem J, 1991. **276** (Pt 1): p. 269-70.
131. Herzberg, O. and J. Moult, *Bacterial resistance to beta-lactam antibiotics: crystal structure of beta-lactamase from Staphylococcus aureus PC1 at 2.5 Å resolution.* Science, 1987. **236**(4802): p. 694-701.
132. Matagne, A., J. Lamotte-Brasseur, and J.M. Frere, *Catalytic properties of class A beta-lactamases: efficiency and diversity.* Biochem J, 1998. **330** (Pt 2): p. 581-98.
133. Herzberg, O., et al., *Structural basis for the inactivation of the P54 mutant of beta-lactamase from Staphylococcus aureus PC1.* Biochemistry, 1991. **30**(39): p. 9503-9.
134. Gibson, R.M., H. Christensen, and S.G. Waley, *Site-directed mutagenesis of beta-lactamase I. Single and double mutants of Glu-166 and Lys-73.* Biochem J, 1990. **272**(3): p. 613-9.
135. Lamotte-Brasseur, J., et al., *Mechanism of acyl transfer by the class A serine beta-lactamase of Streptomyces albus G.* Biochem J, 1991. **279** (Pt 1): p. 213-21.
136. Adachi, H., T. Ohta, and H. Matsuzawa, *Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin.* J Biol Chem, 1991. **266**(5): p. 3186-91.
137. Strynadka, N.C., et al., *Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 Å resolution.* Nature, 1992. **359**(6397): p. 700-5.
138. Damblon, C., et al., *The catalytic mechanism of beta-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme.* Proc Natl Acad Sci U S A, 1996. **93**(5): p. 1747-52.
139. Guillaume, G., et al., *Site-directed mutagenesis of glutamate 166 in two beta-lactamases. Kinetic and molecular modeling studies.* J Biol Chem, 1997. **272**(9): p. 5438-44.
140. Meroueh, S.O., et al., *Ab initio QM/MM study of class A beta-lactamase acylation: dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70.* J Am Chem Soc, 2005. **127**(44): p. 15397-407.
141. Matagne, A. and J.M. Frère, *Contribution of mutant analysis to the understanding of enzyme catalysis: the case of class A beta-lactamases.* Biochim Biophys Acta, 1995. **1246**(2): p. 109-27.
142. Delaire, M., et al., *Site-directed mutagenesis at the active site of Escherichia coli TEM-1 beta-lactamase. Suicide inhibitor-resistant mutants reveal the role of arginine 244 and methionine 69 in catalysis.* J Biol Chem, 1992. **267**(29): p. 20600-6.
143. Therrien, C., et al., *Roles of amino acids 161 to 179 in the PSE-4 omega loop in substrate specificity and in resistance to ceftazidime.* Antimicrob Agents Chemother, 1998. **42**(10): p. 2576-83.
144. Zawadzke, L.E., et al., *Elimination of the hydrolytic water molecule in a class A beta-lactamase mutant: crystal structure and kinetics.* Biochemistry, 1996. **35**(51): p. 16475-82.
145. Escobar, W.A., J. Miller, and A.L. Fink, *Effects of site-specific mutagenesis of tyrosine 105 in a class A beta-lactamase.* Biochem J, 1994. **303** (Pt 2): p. 555-8.

146. Venkatachalam, K.V., et al., Characterization of TEM-1 beta-lactamase mutants from positions 238 to 241 with increased catalytic efficiency for ceftazidime. *J Biol Chem*, 1994. **269**(38): p. 23444-50.
147. Guo, F., et al., The role of the non-conserved residue at position 104 of class A beta-lactamases in susceptibility to mechanism-based inhibitors. *Biochim Biophys Acta*, 1999. **1431**(1): p. 132-47.
148. Kato-Toma, Y., et al., pKa measurements from nuclear magnetic resonance of tyrosine-150 in class C beta-lactamase. *Biochem J*, 2003. **371**(Pt 1): p. 175-81.
149. Dubus, A., et al., The role of tyrosine 150 in catalysis of beta-lactam hydrolysis by AmpC beta-lactamase from *Escherichia coli* investigated by site-directed mutagenesis. *Biochemistry*, 1994. **33**(28): p. 8577-86.
150. Oefner, C., et al., Refined crystal structure of beta-lactamase from *Citrobacter freundii* indicates a mechanism for beta-lactam hydrolysis. *Nature*, 1990. **343**(6255): p. 284-8.
151. Gherman, B.F., et al., Mixed quantum mechanical/molecular mechanical (QM/MM) study of the deacylation reaction in a penicillin binding protein (PBP) versus in a class C beta-lactamase. *J Am Chem Soc*, 2004. **126**(24): p. 7652-64.
152. Hata, M., et al., A theoretical study on the substrate deacylation mechanism of class C beta-lactamase. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 2005. **109**(33): p. 16153-60.
153. Goldberg, S.D., et al., Identification of residues critical for catalysis in a class C beta-lactamase by combinatorial scanning mutagenesis. *Protein Sci*, 2003. **12**(8): p. 1633-45.
154. Diaz, N., D. Suarez, and T.L. Sordo, Molecular dynamics simulations of class C beta-lactamase from *Citrobacter freundii*: insights into the base catalyst for acylation. *Biochemistry*, 2006. **45**(2): p. 439-51.
155. Lobkovsky, E., et al., Evolution of an enzyme activity: crystallographic structure at 2-Å resolution of cephalosporinase from the ampC gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc Natl Acad Sci U S A*, 1993. **90**(23): p. 11257-61.
156. Golemi, D., et al., Critical involvement of a carbamylated lysine in catalytic function of class D beta-lactamases. *Proc Natl Acad Sci U S A*, 2001. **98**(25): p. 14280-5.
157. Sun, T., et al., Comparison of beta-lactamases of classes A and D: 1.5-Å crystallographic structure of the class D OXA-1 oxacillinase. *Protein Sci*, 2003. **12**(1): p. 82-91.
158. Wang, Z., et al., Metallo-beta-lactamase: structure and mechanism. *Curr Opin Chem Biol*, 1999. **3**(5): p. 614-22.
159. Heinz, U. and H.W. Adolph, Metallo-beta-lactamases: two binding sites for one catalytic metal ion? *Cell Mol Life Sci*, 2004. **61**(22): p. 2827-39.
160. Bush, K., G.A. Jacoby, and A.A. Medeiros, A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother*, 1995. **39**(6): p. 1211-33.
161. Chesnel, L., et al., Increase of the deacylation rate of PBP2x from *Streptococcus pneumoniae* by single point mutations mimicking the class A beta-lactamases. *Eur J Biochem*, 2002. **269**(6): p. 1678-83.
162. Duval, V., et al., The kinetic properties of the carboxy terminal domain of the *Bacillus licheniformis* 749/I BlaR penicillin-receptor shed a new light on the derepression of beta-lactamase synthesis. *Mol Microbiol*, 2003. **48**(6): p. 1553-64.
163. Frère, J.M., Beta-lactamases and bacterial resistance to antibiotics. *Mol Microbiol*, 1995. **16**(3): p. 385-95.
164. Hall, B.G. and M. Barlow, Structure-based phylogenies of the serine beta-lactamases. *J Mol Evol*, 2003. **57**(3): p. 255-60.

165. Davies, C., S.W. White, and R.A. Nicholas, *Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-angstrom resolution*. Journal of Biological Chemistry, 2001. **276**(1): p. 616-623.
166. Meroueh, S.O., et al., *Structural aspects for evolution of beta-lactamases from penicillin-binding proteins*. J Am Chem Soc, 2003. **125**(32): p. 9612-8.
167. Rhazi, N., et al., *Peptidase activity of beta-lactamases*. Biochem J, 1999. **341 (Pt 2)**: p. 409-13.
168. Bernstein, N.J. and R.F. Pratt, *On the importance of a methyl group in beta-lactamase evolution: free energy profiles and molecular modeling*. Biochemistry, 1999. **38**(32): p. 10499-510.
169. Adediran, S.A., et al., *The D-methyl group in beta-lactamase evolution: evidence from the Y221G and GC1 mutants of the class C beta-lactamase of Enterobacter cloacae P99*. Biochemistry, 2005. **44**(20): p. 7543-52.
170. Chesnel, L., et al., *Increase of the deacylation rate of PBP2x from Streptococcus pneumoniae by single point mutations mimicking the class A beta-lactamases*. European Journal of Biochemistry, 2002. **269**(6): p. 1678-1683.
171. Bloom, J.D., et al., *Protein stability promotes evolvability*. Proc Natl Acad Sci U S A, 2006. **103**(15): p. 5869-74.
172. Avalle, B., S. Vanwetswinkel, and J. Fastrez, *In vitro selection for catalytic turnover from a library of beta-lactamase mutants and penicillin-binding proteins*. Bioorganic & Medicinal Chemistry Letters, 1997. **7**(4): p. 479-484.
173. Vanwetswinkel, S., B. Avalle, and J. Fastrez, *Selection of beta-lactamases and penicillin binding mutants from a library of phage displayed TEM-1 beta-lactamase randomly mutated in the active site Omega-loop*. Journal of Molecular Biology, 2000. **295**(3): p. 527-540.

