Composition du jury :

Prof. ***Frédéric Ducancel***

Commissariat à l'Energie Atomique (CEA) Saclay, France

Prof. ***Bernard Hallet***

Université Catholique de Louvain

Prof. ***Pascal Hols***

Université Catholique de Louvain

Prof. ***André Matagne***

Université de Liège, Centre d'Ingénierie des Protéines

Prof. ***Claude Remacle*** (Président du jury)

Université Catholique de Louvain

Prof. ***Patrice Soumillion*** (Promoteur)

Université Catholique de Louvain

L'important c'est le voyage, pas la destination!

Remerciements

Bon voilà, le moment tant attendu mais néanmoins redouté est arrivé : écrire les remerciements qui clôtureront la rédaction de ma thèse et qui mettront en même temps presque fin à une belle aventure de plus de cinq ans (C'est un euphémisme pour dire presque 6 ans). Et c'est là que l'inspiration me lâche… Je pourrais bien sur vous dire que la thèse c'est comme un marathon et faire une analogie au sport. Mais malheureusement pour moi cette accroche a déjà été utilisée plus d'une fois au laboratoire[1,2]. D'ailleurs je ne suis pas tout à fait d'accord avec ce rapprochement ; la preuve étant que je serais incapable de terminer une course longue de 42,195 km. Et là j'entends les mauvaises langues dire que je pensais la même chose à propos de ma thèse, que je n'arrête pas de me sous-estimer, que je suis pessimiste,… Bon d'accord vous avez raison… je le ferai ce marathon. Bref, cette thèse a été une expérience unique, faite de hauts et de bas (et de hauts) mais surtout agrémentée de rencontres exceptionnelles qui m'ont toutes apporté quelque chose. Le temps est maintenant venu pour moi de remercier les personnes qui ont fait ces rencontres :

En premier lieu je voudrais remercier le professeur Patrice Soumillion pour m'avoir offert l'opportunité de travailler au sein de son laboratoire et pour m'avoir supervisé tout au long de ce travail. Je le remercie pour la confiance qu'il m'a témoignée du début à la fin, pour sa disponibilité, pour m'avoir communiqué sa passion pour les sciences et pour ses grandes qualités humaines.

Je remercie aussi le professeur Jacques Fastrez pour les conseils et sa grande expérience qu'il m'a fait partager.

Je tiens également à remercier les professeurs Frédéric Ducancel, Bernard Hallet, Pascal Hols, André Matagne et Claude Remarcle d'avoir accepté de faire partie de mon jury de thèse. Ils ont également contribué par leurs remarques et suggestions à améliorer la qualité de cette thèse, et je leur en suis très reconnaissant.

Merci à mes collègues et amis qui m'ont apporté beaucoup au niveau scientifique mais encore plus du côté humain :

Merci à toi Véro, pour ta gentillesse, tes encouragements et ton aide pour venir à bout de la paperasse.

Laurence, mon amie et ex belle-maman, je ne pourrais jamais assez te remercier pour tout ce que tu as fait pour moi. Honnêtement, sans toi je n'aurais pu venir à bout de cette thèse. Merci d'avoir toujours été là pour moi, dans les bons comme dans les mauvais moments. Merci pour ton amitié, merci d'être la personne que tu es, juste MERCI!!!

Merci Val, un ami depuis près de six ans sur lequel j'ai pu souvent compter (excepté lorsque je suis tombé dans le lac gelé de LLN). Que de bons souvenirs!… Les soirées labo,

le voyage en Normandie avec Heykel, le séjour en Suisse chez JM, le tennis, les soupers de Noël et notre traditionnel apéro,...

Merci JM d'avoir grandement participé à mes meilleures années passées au laboratoire. Que de soirées mémorables avec Val, Heykel, Vaidas. La soirée "surprends moi", la soirée "je te vole ton coup à l'insu de mon plein gré", la soirée "je sors avec ma coloc et je l'épouse",... Je suis heureux d'avoir pu partager ces moments avec un vrai ami comme toi!

Merci Gabi pour ta perpétuelle bonne humeur (je pense être le seul au labo à avoir jamais réussi à t'énerver et mon tibia s'en souvient encore ☺), pour ton incroyable positive attitude ainsi que pour ton soutien et ton aide dans les corrections. J'ai finalement explosé ton temps passé en thèse de 80 jours (le temps nécessaire pour faire un tour du monde[3]).

Merci Gene pour ton amitié et pour m'avoir transmis ta passion pour la course à pied, passion qui m'a aidé à évacuer le stress lors de la rédaction. Si seulement je pouvais un jour atteindre ton niveau... A très vite lors d'un entraînement!

Merci Anast pour m'avoir soutenu lors des moments difficiles de la thèse. Je n'oublierai jamais ce que tu as fait pour moi. Le labo te doit beaucoup et tu mérites la magnifique carrière qui s'ouvre à toi.

Merci Steph pour avoir mis l'ambiance au labo, en particulier tous les vendredis! J'ai été content de relever le défi des 20km et d'avoir lancer la mode course à pied au labo avec toi. J'espère que l'on fera encore de nombreuses courses ensemble (mais plus faciles que les 5-10 miles de LLN).

Merci Heykel, compagnon d'infortune durant la thèse, compagnon d'aventure lors des 20km de Bruxelles, compagnon de débauche (surtout toi) pendant les sorties labo. Bon courage pour la fin de ta thèse, bientôt la délivrance!

Merci Mary pour ton enthousiasme, pour ton rire communicatif, pour m'avoir fait découvrir le Liban à travers ta cuisine. J'espère avoir la chance de découvrir ton beau pays de mes propres yeux.

Merci Philippe, premier Geek officiel du labo, initiateur des soirées jeux et du programme spatial du labo avec Thomas. J'espère que tu trouveras la thèse de tes rêves car la Science a besoin de personnes passionnées comme toi.

Merci Nadia ou Poupette comme Laurence aime à t'appeler. Merci pour le vent de fraîcheur que tu as apporté au laboratoire par ta gentillesse, ta spontanéité et ton adorable naïveté (dans le bon sens du terme bien sûr),... Merci pour ton aide, tes précieux conseils, tes mots d'encouragement,...

Merci Julia pour tous ces bons moments passés en ta compagnie et pour ton soutient. Il y a des rencontres qui surgissent dans notre vie quand on ne les attend pas et que l'on ne peut oublier. Tu en fais partie!

Un grand merci aussi à S'Olivier, en permanence de bonne humeur et toujours prêt à donner un coup de main (merci pour la magnifique figure de la p.107), à Virg qui a

courageusement supporté mon côté stressant pendant 2 semaines, à Gilles qui est passé de chouette mémorant à super collègue, à Pierre et ses phrases mémorables que je n'oserais retranscrire ici, à Julie, à Benjamin, à Thomas, à Vaidas, à Pierre-Yves, à Gaby, à Samira,…

Merci à mon équipe de choc de correction de fautes de "franssè", ma sœur Gaëlle et mon grand-père Louis. Votre aide m'a été très précieuse ! Promis c'était le dernier manuscrit que je vous faisais corriger!

Je remercie aussi tous mes amis, d'ici et de là-bas. Merci pour m'avoir accompagné et aidé pendant ces dernières années.

Pour terminer je voudrais remercier ma maman Brigitte, mes sœurs Gaëlle et Emilie, mon frère Julien, mon beauf Christophe, ma belle-sœur Stéphanie et le reste de ma famille d'avoir toujours été là pour moi. Merci d'avoir cru en moi, de m'avoir supporté, encouragé,… J'ai une famille formidable qui n'arrête pas de s'agrandir. Spéciale dédicace à mes 4 plus jeunes supporters (Romane, Rodrigue, Noah et Julie).

Un livre de ma vie se referme à nouveau. Mais celui-là fera à coup sûr partie des best-sellers de ma bibliothèque.

Encore Merci à toutes et à tous!!!

1. Deschuyteneer, G. Vers l'évolution dirigée de nouveaux composés antibiotiques : collections de peptides cycliques et sélection « anti-darwinienne ». Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2010).
2. Mathieu, V. Engineering allosteric regulation into a circulalry permuted TEM-1 beta-lactamase. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2010).
3. Verne, J. Le Tour du monde en quatre-vingt jours. Folio. (2004).

Abréviations

A	Ampere
Å	Angström
Aa	Amino acid – Acide aminé
ADN	Acide désoxyribonucléique
Amp	Ampicillin
ASA	Accessible Surface Area
BCA	Bicinchoninic Acid
BEt	Ethidium bromide
bla	β-lactamase
blaS	Stabilized β-lactamase
Bp	Basepair
BSA	Bovine Serum Albumin
°C	Celsius degree
cAMP	Adénosine Monophosphate Cyclique
CAP	Catabolite Activator Protein
CDRs	Complementary Determining Region
cfu	Colony forming unit
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DsbA	Disulfide oxidoreductase
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
ε	Molar extinction coefficient
et al.	*et alii*
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
e.g.	*exempli gratia* (par exemple)
ELISA	Enzyme-Linked Immunosorbent Assay
ESBL	Extended spectrum β-lactamases
EtBr	Ethidium Bromide
Fwd	Forward
g	Gram
GFP	Green Fluorescent Protein
h	Hour
HIV	Human immunodeficiency virus
HRP	Horseradish Peroxidase
ID	Intrinsically Disordered
i.e.	*id est* (c'est à dire)
IPTG	Isopropyl-β-thiogalactoside
Kan	Kanamycin
KanA	Kanamycin A

kb	Kilobase
kcat	Catalytic constant
Kd	Constante de dissociation – Dissociation constant
kDa	KiloDalton
kJ	Kilo joule
Km	Michaelis-Menten constant
l	Litre
KNF	Modèle Koshland-Némethy-Filmer
LB	Luria Bertani (medium)
m	Meter
M	Molar
mA	Milli ampere
mAb	Monoclonal antibody
mAc	Anticorps monoclonal
MBP	Maltose Binding Protein
MCS	Multiple cloning site
MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligram
MIC	Minimum inhibitory concentration
min	Minute
ml	Millilitre
mM	Millimolar
mol	Mole
MWC	Modèle Monod-Winand-Changeux
MWCO	Molecular Weight Cut Off
ng	Nanogram
Ni	Nickel
nm	Nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic Acid
O.D.	Optical density
O.N.	Overnight
PAGE	Polyacrylamid gel
PBP	Penicillin binding protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein DataBank
PEG	Polyethylen glycol
PenG	Penicillin G or benzylpenicillin
Pfu	Plaque forming unit
pH	Potential hydrogen
Rd	Round
Rev	Reverse
RMN	Résonnance magnétique nucléaire
rpm	Rotation per minute
RT	Room temperature
s	Second
SDS	Sodium dodecyl sulphate

Sec	Secretory pathway
S.P.	Signal Peptide
SOC	Super Optimal broth with Catabolite repression
SRP	Signal Recognition Particle
ssDNA	Single-stranded DNA
TAE	Tris Acetate EDTA
TBS	Tris Buffered Saline
TEMED	N,N,N'N'-tetramethylethylenediamine
Tet	Tetracyclin
T_m	Melting Temperature
µg	Microgram
µl	Microlitre
µM	Micromolar
UV	Ultraviolet
V	Volt
V_{max}	Maximum velocity
v/v	Volume/volume ratio
w/v	Weight/volume ratio
WT	Wild type
Zn	Zinc

Amino acids table :

One lettre code	Three lettre code	Full name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine

Summary

In this work, we studied the potential of a prestabilized TEM-1 β-lactamase as an evolvable scaffold to engineer allostery. Allostery refers to the regulation of a protein activity in modifying the properties of a site (*e.g.* conformation, flexibility, electrostatic properties,…) through the binding of a regulatory molecule at a distal site, called allosteric. The design and study of allosteric β-lactamases would allow a better understanding of this activity regulation phenomenon. These enzymes could also serve as molecular sensors for the detection and dosage of various ligands.

The strategy consisted in randomizing by insertional and/or substitutional mutagenesis three contiguous surface loops (L1, L2 and L3) in the vicinity of the catalytic center of the enzyme in order to create *de novo* a large collection of mutants. In the starting scaffold, an extra thermodynamic stability of 27 kJ/mol provided by four stabilizing mutations (H153R, M182T, T195S et A224V) highly increased the enzyme tolerance towards those structural perturbations. A library of 6.6×10^7 active mutants with the three loops simultaneously mutated was obtained in a phage vector (fd-tet), which is 70 times more compared to an analogous collection built in a wt TEM-1 β-lactamase. Prestabilizing the enzyme has also largely increased the potential diversity of the degenerated loop L2 in removing a disulfide bridge constraint otherwise required for clones to grow on ampicillin 20 μg/ml.

Libraries built in the stabilized or the wt β-lactamases were subjected to phage display selections in order to isolate various clones with new ligand binding abilities. Twelve clones displaying several histidines within their inserts emerged from a selection on nickel beads. Three of them (25%), including two with the stabilizing mutations, showed a regulation of activity upon nickel binding. The most interesting regulation profiles were observed for the two prestabilized clones. One was inhibited by 80% and the second was activated by 90%. Selection of anti-idiotypic β-lactamases recognizing a monoclonal antibody specific of the lipopolysaccharide outer core of *Coxiella burnetii* was however not successful. Interestingly, the proportion of clones with cysteines at both ends of L2 has

increased along the phage display selections, despite the elimination of this structural constraint upon libraries construction.

The stabilization effect on the activity regulation of a β-lactamase activated by aminoglycosides and previously isolated in the laboratory without the stabilizing mutations was then investigated. The insertion of the mutations had no impact on the regulation profile of the enzyme.

We also studied the potential benefit of using a phagemid vector (pHDi-Ex) in obtaining DNA libraries of high diversity and in selecting mutants of interest by phage-display. The easier handling of the phagemid compared to the one of the fd-tet phage vector facilitated the construction of the highly diverse libraries. Yet, the high expression rate of the membrane anchored β-lactamase (for phage display application) induced fitness cost to the bacteria leading to loss of diversity in the libraries. During the phage display selections on streptavidin, clones not expressing any fusion protein were *in vivo* selected. Nevertheless, some clones whose affinity for streptavidin seemed to be confirmed by ELISA were isolated from this background noise. However, none of them showed regulation of activity in the presence of the ligand.

Résumé

Le potentiel évolutif d'une β-lactamase TEM-1 stabilisée en terme de propension à l'émergence d'une allostérie a été étudié dans ce travail. L'allostérie est un mode de régulation de l'activité d'une protéine par lequel la liaison d'un effecteur en un site dit allostérique modifie les propriétés (*e.g.* conformation, flexibilité, propriétés électrostatiques...) d'un site distant. L'ingénierie génétique et l'étude de β-lactamases allostériques permettraient une meilleure compréhension de ce phénomène de régulation d'activité. Ces enzymes pourraient aussi servir de senseurs moléculaires pour la détection, voire le dosage de ligands divers.

La stratégie a consisté à randomiser par mutagenèse insertionnelle/substitutionnelle trois boucles de surface contiguës (L1, L2 et L3) et proches du centre catalytique de l'enzyme à fin de créer *de novo* une large collection de mutants. Un gain de stabilité thermodynamique de 27 kJ/mol apporté par quatre mutations stabilisantes (H153R, M182T, T195S et A224V) a fortement augmenté la tolérance de l'enzyme vis-à-vis de ces perturbations structurelles. Une banque de 6,6 x 10^7 mutants actifs et présentant les 3 boucles simultanément mutées a ainsi été obtenue dans un vecteur phagique (fd-tet), soit 70 fois plus que pour une banque analogue préalablement construite dans une β-lactamase TEM-1 sauvage. La stabilisation de l'enzyme a en outre permis d'augmenter grandement la diversité potentielle au sein de la boucle L2 en autorisant l'élimination d'un pont disulfure autrement requis pour obtenir des mutants actifs sur ampicilline 20 µg/ml.

Les banques obtenues dans une β-lactamase préstabilisée ou sauvage ont été soumises à des sélections par phage display afin d'isoler des clones ayant acquis une affinité pour divers ligands cibles. Douze clones présentant plusieurs histidines dans leurs inserts ont émergé d'une sélection sur billes de nickel. Trois d'entre eux (25 %), dont deux possédant les mutations stabilisantes, ont montré une régulation d'activité par la liaison d'ions Ni(II). Les deux clones préstabilisés montraient les profils de régulation les plus intéressant, à savoir une inhibition et une activation d'activité de respectivement 80% et 90%. Une sélection sur un anticorps monoclonal reconnaissant un lipopolysaccharide de la membrane externe de *Coxiella burnetii* n'a par contre pas permis d'isoler de β-lactamases anti-

idiotypiques. A noter qu'au cours des différentes sélections à partir de la banque de β-lactamases préstabilisées, la proportion de clones présentant des cystéines aux deux extrémités de L2 a augmenté malgré l'élimination de cette contrainte structurale lors de la construction de la banque.

L'impact des mutations stabilisantes sur la régulation allostérique d'un clone activé par des aminoglycosides et précédemment sélectionné sans ces mutations a par la suite été exploré. L'insertion des mutations au niveau de l'enzyme n'a eu aucune influence sur son profil d'activation.

Nous avons aussi étudié l'avantage éventuel de l'utilisation d'un vecteur de type phagemidique (pHDi-Ex) dans l'obtention de banques d'ADN de haute diversité et dans la sélection par phage-display de mutants d'intérêt. La construction des banques a été facilitée par la manipulation plus aisée du phagemide par rapport au vecteur phage fd-tet. Cependant, le haut taux d'expression de β-lactamases ancrées a induit d'importants coûts de fitness pour les bactéries provoquant des pertes de diversité au niveau des banques. Les sélections sur streptavidine à partir de la banque phagemidique se sont trouvées perturbées par l'émergence de clones n'exprimant pas de protéine de fusion et sélectionnés via leur meilleur fitness *in vivo*. Quelques clones dont l'affinité pour la streptavidine a semble-t-il été confirmée par ELISA ont néanmoins pu être isolés de ce bruit de fond mais aucun n'a montré de régulation d'activité en présence du ligand.

Table des matières

Summary **1**

Résumé **3**

Introduction générale **9**

I. Préambule **11**

II. Ingénierie de la régulation allostérique **13**

II.1. Définition de l'allostérie 13

II.2. Modèles et mécanismes de l'allostérie naturelle 14

II.2.A. Premiers modèles 14

II.2.B. Modèles actuels 16

II.2.B.a. Modèle d' "Ajustement Induit (Induced Fit)" 16

II.2.B.b. Modèle de "Sélection conformationnelle (Population shift)" 17

II.2.B.c. L'allostérie sans changement conformationnel 18

II.3. Ingénierie de l'allostérie 20

II.3.A. Design rationnel 21

II.3.A.a. Ingénierie d'un site de fixation à proximité du site actif 21

II.3.A.b. Ingénierie d'une régulation par intégration de réseaux d'acides aminés 22

II.3.A.c. Ingénierie d'une régulation par sélection conformationnelle 23

II.3.A.d. Ingénierie d'une coopérativité positive par modulation d'une dynamique moléculaire 26

II.3.B. Approche aléatoire 27

II.3.B.a. Ingénierie d'une régulation par fusion aléatoire de domaines protéiques 28

II.3.B.b. Insertions de peptides aléatoires dans des boucles de surface 30

III. La TEM-1 β-lactamase **37**

III.1.Description générale, structurale et fonctionnelle 37

III.1.A. Généralités 37

III.1.B. Structure, repliement et stabilité : informations générales 38

III.1.C Mécanisme catalytique 40

III.2. La β-lactamase comme plateforme d'ingénierie 42

III.2.A. Tolérance aux mutations 42

III.2.B. Ingénierie de la spécificité 45

III.2.C. Ingénierie d'une stabilité 47

III.2.C.a Perturbation structurale et compensation par évolution dirigée 47

III.2.C.b. Stabilisation in vitro par la méthode "Proside" 49

III.2.C.c. Caractérisation des mutations stabilisantes 51

IV. Stabilité et potentiel d'évolution 55

IV.1. Mutations et stabilité protéique 55
IV.2. Stabilité protéique et robustesse 57
IV.3. Robustesse et potentiel d'évolution 60
IV.3.A. Un gain de stabilité favorise le potentiel d'évolution 60
IV.3.B Dérive neutre et stabilité 63
IV.3.C. Stabilité et contraintes évolutives 65

V. Références 69

Objectifs et stratégie 79

I. Objectifs et stratégie générale 81
II. Références 85

Résultats 87

Protein stability promotes tolerance to insertional mutagenesis 89

Abstract 91
Introduction 93
Results and discussion 97
Insertions in the C-terminal loop are fully tolerated 97
Highly increased tolerance in loops L1 and L2 mutagenesis 99
Combinatorial library construction 101
Final libraries analysis 102
Conclusions 107
Materials and methods 109
Acknowledgements 113
Funding 114
Reference 115
Supplementary data 119
Phagemid description and construction 119
..... 126
Supplementary materials and methods 129
Supplementary references 129

Phage display selections of allosteric β-lactamases 131

Introduction 133
Results 135
Selection from the phagemid library 135
Construction and characterization of the library 135
Biopanning against streptavidin 137
Selection from the fd-tet libraries 144
Libraries display and fitness assessment 144
Biopanning against nickel(II) ion 147
Biopanning against anti-Coxiella burnetii antibody 151

Discussion and conclusion 153
Vector influence on libraries quality and diversity 153
Selection of streptavidin binders 154
Fitness cost disturbed phage display selections 155
Selection of allosteric clones regulated by nickel 159
Selection against an anti-Coxiella burnetii antibody 160
Relationship between TEM-1 β-lactamase stabilization and library fitness cost 161
Conclusion 162
Materials and methods 163
Acknowledgements 174
References 175
Supplementary data 179

Stabilization impact on BlaKr activity regulation 189

Introduction 191
Results & discussion 193
Materials and methods 199
References 201

Conclusion et perspectives 203

I. Discussion et conclusion générales 205
Ingénierie allostérique de β-lactamases TEM-1 allostériques 205
Coût de fitness du couple phagemide/β-lactamase TEM-1 207
II. Perspectives 211
Poursuite des sélections sur diverses cibles d'intérêt 211
Impact de la stabilisation sur l'émergence d'une régulation allostérique 213
Investigation du coût de fitness 214
III. Références 215

Annexes 219

Introduction générale

I. Préambule

Les protéines sont des macromolécules biologiques essentielles aux êtres vivants, participant à pratiquement toutes les fonctions cellulaires. Encodées par l'ADN et assemblées à partir des 20 acides aminés naturels, leur diversité naturelle tant séquentielle que structurale et fonctionnelle est immense bien qu'elle ne représente qu'une infime partie de la diversité possible. Ainsi, pour une petite protéine de 100 résidus, 20^{100} séquences sont possibles à partir de la succession aléatoire de 20 acides aminés naturels, soit plus que le nombre d'atomes dans l'Univers[1]. En raison notamment de leur rôle essentiel et de leur grande variabilité, les scientifiques se sont rapidement intéressés au potentiel offert par l'étude, l'utilisation et la manipulation de ces biomolécules. L'ingénierie des protéines est cependant une discipline relativement récente liée à l'avènement des techniques de biologie moléculaire permettant de manipuler à volonté le matériel génétique et donc les protéines. Par "design" rationnel ou par approche aléatoire, de nombreuses équipes de scientifiques sont parvenues à modifier ou à améliorer les propriétés physicochimiques de protéines diverses. Les études d'ingénierie portent essentiellement sur les propriétés catalytiques des enzymes, sur le repliement et la stabilité des protéines ainsi que sur la reconnaissance moléculaire non covalente.

L'étude est centrée ici sur l'ingénierie d'un site de régulation allostérique par insertion peptidique dans une β-lactamase préalablement stabilisée et aborde de ce fait les trois aspects : catalyse, reconnaissance moléculaire et stabilité. Les résultats de ce travail sont précédés par une introduction divisée en trois parties. La première partie définira l'allostérie et s'intéressera, via des exemples concrets, à son origine naturelle et à son ingénierie. La deuxième partie caractérisera l'enzyme modèle utilisée dans ce travail comme plateforme d'ingénierie. Enfin, la dernière section traitera de l'importance de la stabilité protéique dans le potentiel d'évolution des protéines.

II. Ingénierie de la régulation allostérique

II.1. Définition de l'allostérie

L'environnement dans lequel les organismes vivants évoluent est un système dit "ouvert" et est donc soumis à des variations diverses. Pour survivre, les différents êtres vivants ont du développer différentes stratégies leur permettant de contrôler ces fluctuations extérieures afin de maintenir un équilibre interne, appelé homéostasie. L'activité d'une protéine peut être ajustée à différents niveaux. Ce contrôle peut s'exercer via la régulation de l'expression du gène, de la dégradation de la protéine, mais également au travers d'une modulation directe de l'activité protéique via des modifications covalentes ou l'interaction non covalente avec des inhibiteurs ou activateurs. La régulation dite allostérique agit à ce dernier niveau. Le terme vient du grec *ἄλλος*, allos : autre et *στερεός*, stereós : solide. L'allostérie implique donc un changement des propriétés (*e.g.* conformation, flexibilité, propriétés électrostatiques...) d'un site d'une protéine suite à un évènement déclencheur (*e.g.* liaison d'un ligand, phosphorylation, glycosylation,...) se produisant au niveau d'un site distal[2]. Ce travail s'intéresse à la création d'une allostérie au sein d'une enzyme mais le phénomène est également observable au niveau de protéines de transport[3,4], de protéines de régulation[5,6], de protéines de structure[7],… L'allostérie est dite positive si l'évènement déclencheur induit une augmentation de l'activité et/ou de l'affinité pour un substrat **(Figure 1.a)**. A l'inverse, une diminution de l'activité et/ou de l'affinité sera le résultat d'une régulation allostérique dite négative **(Figure 1.b)**. De même, l'allostérie sera qualifiée d'homotropique si l'effecteur et le substrat sont identiques **(Figure 1.c)**. Le terme hétérotropique implique donc que l'effecteur et le substrat soient de nature différente **(Figure 1.a et 1.b)**. Bien qu'initialement mise en évidence au niveau de protéines oligomériques, l'allostérie tant hétérotropique qu'homotropique est également observée pour des protéines monomériques[8–10] **(Figure 1)**.

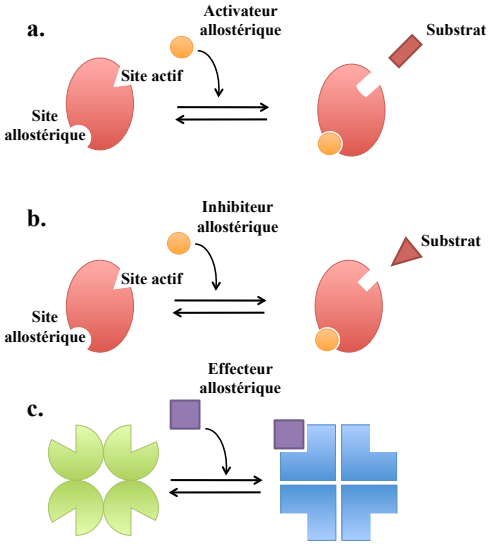


Figure 1 : (a). Allostérie hétérotropique positive : la liaison d'un activateur au niveau du site allostérique induit un changement conformationnel au niveau du site actif, se traduisant par une augmentation d'affinité et/ou d'activité. (b). Allostérie hétérotropique négative : la liaison d'un inhibiteur au niveau du site allostérique induit un changement conformationnel au niveau du site actif, se traduisant par une diminution d'affinité et/ou d'activité. (c). Allostérie homotropique positive : la liaison d'un ligand au niveau d'un monomère d'une protéine multimérique induit un changement conformationnel au niveau des autres monomères, qui voient leur affinité pour ce même ligand augmentée.

II.2. Modèles et mécanismes de l'allostérie naturelle

Depuis la découverte dans les années soixante de la régulation allostérique au sein des protéines, différents modèles se sont succédés pour tenter d'expliquer ce phénomène du point de vue moléculaire. Les avancées dans les techniques d'analyse structurale telles que la cristallographie et la RMN ont permis une étude plus précise des mécanismes moléculaires médiant la transmission de la perturbation du site allostérique au site actif. Après un bref historique, ce chapitre détaillera les différents modèles théoriques actuels. Leur compréhension permettra de mieux appréhender les différentes stratégies abordées au point suivant et développées afin de créer *de novo* une régulation allostérique au niveau de protéines modèles.

II.2.A. Premiers modèles

C'est en 1965 que le premier modèle permettant d'expliquer l'allostérie a été proposé. Le modèle MWC, des initiales des noms de ses découvreurs Jacques Monod, Jeffries

Winand et Jean-Pierre Changeux, décrit les transitions allostériques pour des protéines multimériques présentant un axe de symétrie[11] **(Figure 2.a)**. Dans ce modèle, chaque sous-unité (ou protomère) de la protéine peut exister sous deux états différents : l'état dit relaxé (R), ayant une plus grande affinité pour le ligand que l'état dit tendu (T). Les différentes sous-unités peuvent passer d'un état à l'autre, mais toujours de manière concertée de sorte que la symétrie de la protéine multimérique soit conservée. La liaison d'un activateur allostérique au niveau d'un protomère induit un déplacement de l'équilibre vers la population "R", l'ensemble du système devenant davantage réceptif à la liaison d'autres ligands. Inversement, l'interaction avec un inhibiteur ayant une plus forte affinité pour l'état "T", induit un déplacement d'équilibre vers ce dernier et donc une diminution de l'affinité du système pour le ligand. Le phénomène allostérique de la liaison de l'oxygène au niveau de l'hémoglobine sert souvent à illustrer ce modèle bien qu'une plus grande complexité du système fut démontrée par la suite [4,12].

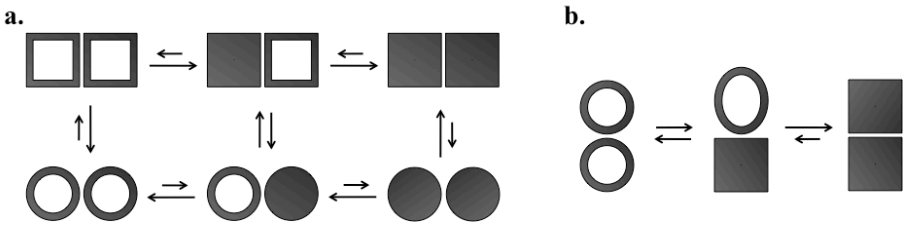


Figure 2: Représentation schématique des modèles MWC et KNF au niveau d'une protéine dimérique. Les états "R" et "T" sont représentés respectivement par un carré ou un cercle. (a.) Modèle MWC : en l'absence de ligand (forme vide), un équilibre existe entre les 2 seules conformations permettant une symétrie. La liaison d'un ligand au niveau d'une sous-unité (forme pleine) induit un déplacement de l'équilibre vers l'état "R". L'affinité du système pour le ligand s'en trouve augmentée. (b.) Modèle KNF : le changement d'état de T vers R n'est induit que par la liaison du ligand. Le passage vers l'état "R" induit une modification structurale au niveau de la sous-unité adjacente. Dans ce cas-ci, cette dernière voit son affinité pour le ligand augmentée (= coopérativité positive). L'inverse est également possible (= coopérativité négative).

Daniel Koshland, George Némethy et David Filmer proposèrent leur propre modèle[13], dit "séquentiel" (ou KNF des initiales des noms de ses découvreurs), pour lequel le changement d'état de "T" vers "R" n'est induit que par la liaison du ligand (= théorie de l'ajustement induit ou "Induced Fit" [14]) **(Figure 2.b)**. Ce changement d'état affecte la structure des sous-unités adjacentes ainsi que leur affinité pour le

ligand. La coopérativité est dite positive si l'affinité est augmentée et négative dans la situation inverse.

II.2.B. Modèles actuels

Les précédents modèles ne permettent d'expliquer l'allostérie que dans le cas de protéines multimériques. Or depuis les années 1960, plusieurs cas de régulation allostérique ont été décrits au niveau de protéines monomériques[2,3,15–19]. Les progrès dans les techniques d'analyse structurale des protéines ont rendu possible l'élaboration de nouveaux modèles.

II.2.B.a. Modèle d' "Ajustement Induit (Induced Fit)"

Ce modèle permet d'expliquer l'allostérie d'un point de vue enthalpique[20] et reprend les bases du modèle KNF. La théorie prévoit que la liaison du ligand au niveau du site allostérique se fasse par ajustement induit et que le changement conformationnel en résultant soit transmis à travers la protéine jusqu'au site actif par propagation d'interactions moléculaires[12,14,21] **(Figure 3)**. Il existerait dès lors au sein de la structure tertiaire de la protéine, un "réseau" de résidus physiquement ou énergétiquement couplés et reliant les deux sites[2,12,21,22] **(Figure 3)**.

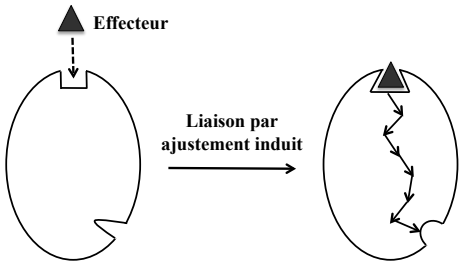


Figure 3 : Représentation schématique du modèle allostérique par "Ajustement Induit". La modification conformationnelle amorcée par la liaison d'un effecteur au niveau du site allostérique est transmise via un réseau d'acides aminés (→) jusqu'au site actif. Le changement structural induit au niveau de ce dernier augmente ou diminue l'affinité pour le substrat. Figure copiée de Whitley *et al*[21].

De tels réseaux ont été mis en évidence par les travaux de Ranganathan et son équipe par analyse bioinformatique des résidus co-évolués au sein de familles de protéines allostériques (Récepteurs couplés aux protéines G, sérine-protéases de type chymotrypsine, hémoglobines, familles à domaine PDZ)[12,22]. De manière

intéressante, l'analyse de la propagation de l'effet à partir du site de liaison du domaine PDZ de PSD-95 définit un réseau se superposant à celui obtenu par analyse bioinformatique[23].

II.2.B.b. Modèle de "Sélection conformationnelle (Population shift)"

Le modèle d' "Ajustement Induit" ne permet pas toujours d'expliquer la transmission des signaux allostériques sur de longues distances (> 100 Å) comme observé au niveau du récepteur aspartate[24]. Le modèle de "Sélection conformationnelle", se basant davantage sur le modèle MWC, apporte une nouvelle interprétation de l'allostérie[25,26]. La vision des protéines comme des structures statiques a laissé place depuis longtemps à la théorie de paysage énergétique[27]. Les protéines existeraient à l'équilibre en un ensemble de conformères de niveaux énergétiques proches **(Figure 4)**.

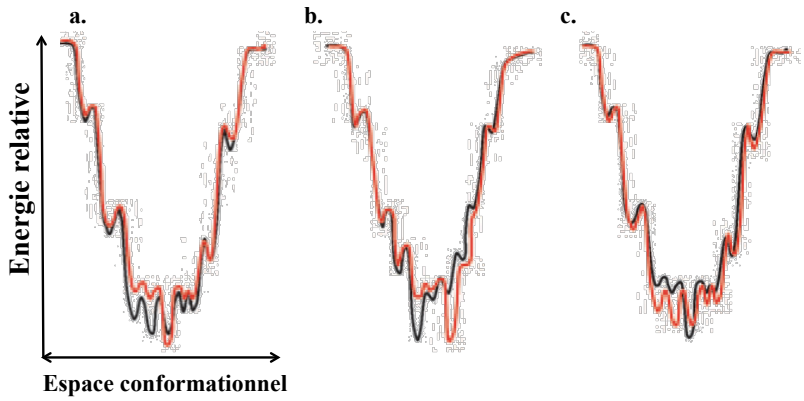


Figure 4 : Redistribution allostérique du paysage énergétique d'une protéine. Les tracés en noir et en rouge représentent respectivement les paysages en absence ou en présence du ligand. (a.) La liaison du ligand stabilise un état énergétique au détriment de l'hétérogénéité conformationnelle du système. (b.) La protéine existe essentiellement sous 2 états différents. La liaison du ligand stabilise l'un de ces états = Modèle MWC (c.) L'état natif, avec une distribution conformationnelle limitée, peut être perturbé par la liaison du ligand et mener à une plus grande hétérogénéité structurale. Figure de Swain *et al.*[28].

L'allostérie reflèterait la redistribution du paysage énergétique ("population shift") d'une protéine suite à la liaison d'un ligand (ou autre évènement déclencheur) au niveau du site allostérique[28] **(Figure 4)**. Ainsi par exemple, le déplacement de

l'équilibre vers des conformères ayant un site catalytique plus effectif ou une plus grande affinité pour le substrat mènerait à une activation de l'enzyme.
L'étude de plusieurs cas de régulation allostérique naturelle supporte ce modèle[2,16,29–31]. Un cas intéressant est celui de NtrC (protéine C régulatrice de l'azote) qui, une fois phosphorylée, active l'expression du gène de la glutamine synthétase. Une analyse RMN démontre que les conformations actives et inactives coexistent au niveau de la protéine non-phosphorylée, l'équilibre étant biaisé en faveur de la seconde[16]. La phosphorylation ne crée donc pas de nouvelle structure mais induit plutôt un déplacement de l'équilibre vers la forme active de la protéine[16].

Récemment, une vision plus étendue du modèle de "Déplacement conformationnel", où la sélection des conformères serait suivie d'une modification structurale par ajustement induit, a été proposée[32,33]. La régulation allostérique de la protéine CheY (protéine "Chemotaxis Y ") est également mieux expliquée par l'existence d'une synergie entre les modèles de "Sélection conformationnelle" et d' "Ajustement Induit"[2]. Chez cette protéine, la phosphorylation du résidu Asp57 provoque un déplacement d'une boucle de surface (β4-α4) permettant une rotation du résidu Tyr106 qui, dans sa position native, ne permet par l'interaction de CheY avec sa protéine flagellaire cible[2]. Des résultats RMN et une approche par simulation ont démontré que la conformation active est déjà présente au niveau de la protéine non-phoshorylée[2]. Cependant la forme active serait stabilisée par la présence d'un réseau couplant les résidus Asp57 et Tyr 106 via la boucle β4-α4[2].

II.2.B.c. L'allostérie sans changement conformationnel

Les exemples de régulation allostérique cités ci-avant impliquent tous une modification conformationnelle de la protéine. Cependant, une altération de la dynamique intrinsèque des protéines peut être à la base d'une allostérie sans changement structural apparent[34,35]. La dynamique des protéines est un terme générique englobant, dans ce cas-ci, un spectre de fluctuations (nanosecondes à picosecondes) allant des vibrations atomiques, aux rotations des groupements chimiques, en passant par les translations moléculaires. Un changement au niveau de la fréquence et de l'amplitude de ces fluctuations, suite à la liaison d'un ligand, permet d'expliquer le phénomène de coopérativité en termes d'entropie **(Figure 5)**.

Par modélisation mathématique, A. Cooper and D. Dryden, ont démontré que des variations d'énergie libre de plusieurs kJ/mole pouvaient être générées par une rigidification modeste des protéines[35]. Si le coût entropique payé suite à la liaison du 1er ligand est supérieur à celui engendré par la liaison du 2ème ligand, la coopérativité sera positive **(Figure 5.a)**. La situation inverse mènera à une coopérativité dite négative **(Figure 5.b)**.

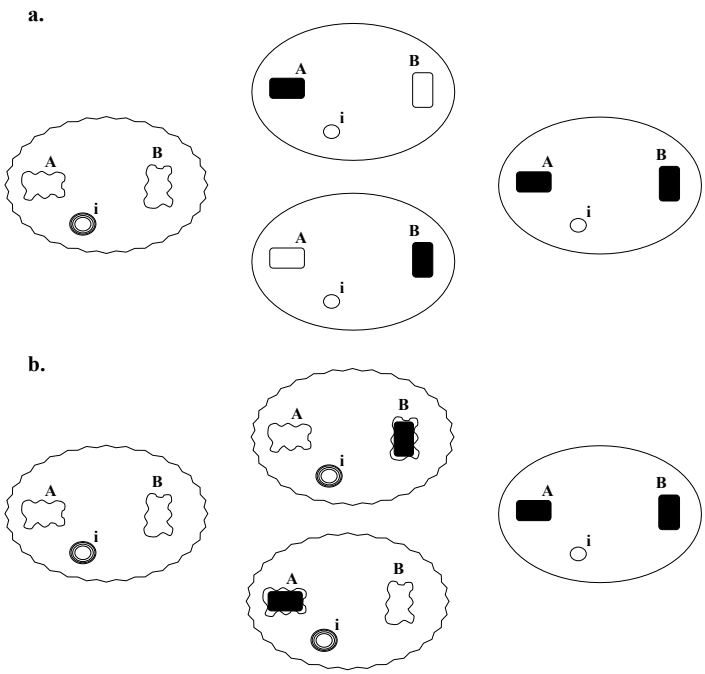


Figure 5 : Représentation schématique du phénomène de coopérativité induit par un changement de dynamique. (a.) Coopérativité positive : en l'absence de ligand, la protéine présente une certaine dynamique interne (contours ondulés). La liaison du ligand au niveau du site A ou B "rigidifie" la protéine (symbolisé par des contours nets) et s'accompagne d'une perte de mobilité et donc d'entropie au niveau de l'atome "i" (symbolisé par la diminution des cercles). L'atome "i" est représentatif de nombreux atomes au sein de la protéine. La liaison du second ligand se fait sans perte additionnelle de mobilité, d'où un coût entropique moindre et une liaison favorisée. (b.) Coopérativité négative : la liaison du 1er ligand n'atténue pas (ou peu) les fluctuations internes. La perte majeure d'entropie provient de la liaison du 2ème ligand. Figure copiée de Kern *et al.*[34].

La liaison de cAMP au niveau de la protéine dimérique CAP (catabolite activator protein) permet d'illustrer cette dernière situation[36]. La liaison de la première molécule de cAMP n'induit pas de changement conformationnel au niveau de la seconde unité, comme en atteste les résultats de RMN. La dynamique interne du

système est cependant modulée par la liaison séquentielle des molécules de cAMP, la première molécule induisant une augmentation partielle et la seconde une inhibition marquée des mouvements moléculaires[36]. La liaison du 2ème ligand induit donc un coût entropique plus élevé et la coopérativité apparaît négative.

II.3. Ingénierie de l'allostérie

L'ingénierie *de novo* d'une régulation allostérique au niveau d'une enzyme présente de nombreux intérêts. D'un point de vue fondamental, créer une allostérie "artificielle" au sein d'une protéine de son choix (facilement étudiable) offre de grandes perspectives dans la compréhension de ce phénomène tant du côté fonctionnel qu'évolutif. Disposer d'une enzyme naturelle régulable est également un outil puissant dans l'étude *in situ* de sa fonction biologique[37]. Contrairement à la régulation transcriptionnelle, la régulation allostérique est immédiate et facilement réversible[37].

En recherche appliquée la possibilité de disposer d'enzymes spécifiquement régulables par des molécules d'intérêt permet d'envisager des nouvelles méthodes de détection de ces molécules simplement basées sur la mesure de la variation de l'activité enzymatique en présence d'un échantillon à doser. Cette technique est à la base du concept de certains biosenseurs. Un biosenseur est un dispositif analytique hybride constitué d'un récepteur de nature biologique couplé à un transducteur physicochimique (optique, électrochimique, calorimétrique,…) capable de convertir un changement de propriété lié à la fixation du ligand en un signal électrique quantifiable[38,39].

Mieux cerner les mécanismes sous-jacents à l'allostérie, via son ingénierie, est également intéressant pour la recherche de nouveaux médicaments ciblant les sites allostériques. Ces derniers ne subissent généralement pas les mêmes contraintes évolutives que les sites actifs et sont donc souvent plus diversifiés au sein d'une même famille protéique[9,40]. Cette variabilité permet de cibler une protéine, appartenant à une famille protéique au site actif fortement conservé, de manière plus spécifique. La famille des récepteurs couplés aux protéines G (RCPG) illustre bien cette situation[41]. Un relargage ciblé d'un médicament pourrait aussi être envisagé par la régulation allostérique de l'affinité d'un "transporteur" sensible à la signature moléculaire (*e.g.* concentration anormale en métabolites, en molécules de

signalisation,…) de la maladie visée[42]. En reconnaissant la signature, le transporteur verrait son affinité pour le médicament réduite et ce dernier serait relargué uniquement au niveau de la zone ciblée.

Ce chapitre qui ne se veut pas exhaustif présentera à l'aide d'exemples concrets diverses stratégies ayant mené à la création *de novo* d'une régulation allostérique au sein de protéines variées. Les différentes stratégies peuvent être divisées en deux groupes selon l'approche générale utilisée. Soit une approche rationnelle, on parlera alors de design rationnel, soit une approche partiellement aléatoire et on parlera dans ce cas d'évolution dirigée. La subdivision de ce chapitre sera calquée sur cette même répartition.

II.3.A. Design rationnel

Le design rationnel se base sur les connaissances ou modélisations structurales de la protéine étudiée afin de prédire les modifications (mutations) à apporter pour générer la ou les propriétés recherchées. Dans le cas qui nous intéresse cette propriété est la régulation allostérique.

II.3.A.a. Ingénierie d'un site de fixation à proximité du site actif

L'une des premières approches rationnelles développée consiste en l'insertion d'un site de liaison avéré à plus ou moins proche distance du centre actif de la protéine étudiée. La stratégie part du postulat que les perturbations structurales liées à la liaison d'un ligand aux environs immédiats du site actif ont une chance certaine d'induire une modulation d'activité.

L'un des exemples régulièrement cité concerne la régulation de l'activité de la trypsine de rat par des ions métalliques[43]. Les auteurs de ce travail ont substitué le résidu Arg_{96} par une histidine afin qu'il forme en association avec le résidu catalytique His_{57} un site de coordination d'ions métalliques **(Figure 6)**. L'His_{57} fait partie avec l'Asp_{102} et la Ser_{195} de la triade catalytique de cette protéase à sérine. La liaison de Cu^{2+}, Ni^{2+} ou Zn^{2+} au niveau du site de coordination créé provoque une rotation de l'His_{57}, empêchant dès lors ce résidu de jouer son rôle de base dans la catalyse. Une diminution de l'efficience catalytique (k_{cat}/K_m) de près de 90 % a pu

être mesurée. L'activité de l'enzyme est restaurée par l'ajout d'EDTA complexant les ions métalliques.

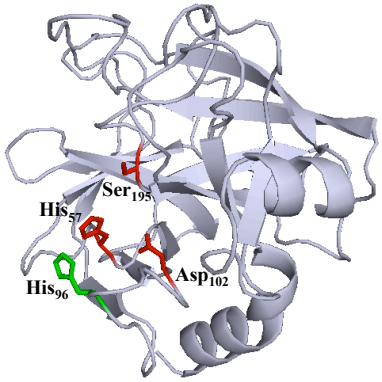


Figure 6 : Strucutre du variant R96H de la trypsine de rat (code PDB : 1AND). Les résidus His_{57}, Asp_{102} et Ser_{195} de la triade catalytique sont représentés en rouge. Le résidu His_{96} formant un site de complexation d'ions métalliques avec l'His_{57} est coloré en vert. Structure réalisée avec MacPymol.

La création de nouveaux sites de liaison peut aussi se faire par l'ingénierie de boucles contiguës connectant les brins β ou hélices α d'une protéine[44]. Ainsi, une seconde approche rationnelle d'ingénierie de l'allostérie consiste en l'insertion d'épitopes en surface de la protéine à réguler. La reconnaissance de ces séquences antigéniques par des anticorps monoclonaux (mAc) spécifiques peut induire dans certains cas une modulation d'activité liée à un changement conformationnel. Ainsi, la fixation de mAc à la surface de la phosphatase alcaline présentant les épitopes de la protéine gp120 (protéine du virus de l'immunodéficience de type 1 'HIV-1') ou de la protéine structurale de la capside du virus de l'hépatite C a mené à des activations (jusqu'à 300 %) ou inhibitions (jusqu'à 40 %) d'activité[45]. Ces stratégies insertionnelles trouvent leur limite dans la disponibilité des domaines et des épitopes possédant les fonctions désirées[46].

II.3.A.b. Ingénierie d'une régulation par intégration de réseaux d'acides aminés

La dihydrofolate réductase d'*E. coli* a été rendue allostériquement régulable par la lumière en se basant sur le modèle d' "ajustement induit" et l'existence potentielle de réseaux d'interactions entre acides aminés couplant deux sites distants[47]. Le postulat de départ prévoyait que l'intégration des réseaux de deux domaines protéiques

différents, via des sites de surface, pourrait mener à un couplage d'activité **(Figure 7)**. Le domaine photosensible de la phototropine (LOV2) a été choisi pour jouer le rôle de senseur ("input"). L'absorption de lumière par un chromophore lié à ce domaine induit un changement conformationnel au niveau des extrémités C- et N-terminales du domaine. L'analyse bioinformatique des résidus co-évolués au sein de la famille des phototropines a permis d'identifier une voie de communication entre le chromophore et les extrémités terminales. La dihydrofolate réductase (DHFR) a été utilisée comme transducteur ("output"). Un réseau d'acides aminés co-évolués connecte le site actif de l'enzyme à une boucle de surface (βF-βG). Différentes protéines chimériques ont été créées en insérant LOV2, via ses deux extrémités terminales, en différents endroits de la boucle βF-βG. L'exposition d'un des mutants à la lumière a permis d'enregistrer une modulation de l'activité dihydrofolate réductase d'un facteur 2.

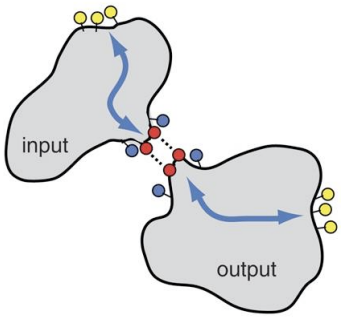


Figure 7 : Représentation schématique de l'intégration de deux réseaux d'acides aminés par design rationnel. Les sites fonctionnels (en jaune) de deux protéines (input et output) sont interconnectés via leurs réseaux internes d'acides aminés (flèches bleues) et deux positions de surface (en rouge). Les sphères bleues sont représentatives de l'ensemble des résidus structuraux. Un signal allostérique (*e.g.* changement conformationnel) peut théoriquement être transmis du site fonctionnel de la protéine "input" vers celui de la protéine "output" via l'interconnexion de leurs réseaux internes d'acides aminés . Figure de Lee *et al.*[47]

II.3.A.c. Ingénierie d'une régulation par sélection conformationnelle

S. Loh et son équipe sont parvenus à créer une régulation allostérique en se basant sur les modèles "MWC" et de "sélection conformationnelle". Le système consiste en une protéine hybride pour laquelle l'énergie libre de repliement d'un des deux domaines est utilisée pour induire le dépliement du second[48]. L'ubiquitine humaine (Ub) a été insérée, via ses extrémités C- et N- terminales, entre les résidus Pro64 et Thr70 d'une boucle de surface de la ribonucléase barnase (Bn) de *Bacillus*

amyloliquefaciens **(Figure 8)**. La différence de distance entre les extrémités de l'ubiquitine et celle mesurée pour les *termini* de la boucle de la barnase provoque une contrainte topologique ne permettant pas un co-repliement des domaines au sein de la même structure. Deux états sont donc possibles : "Bn foldé/Ub non-foldé" et "Bn non-foldé/Ub foldé". La position de l'équilibre entre ces deux conformères peut être ajustée par une variation de la température ou par liaison de l'inhibiteur barstar au niveau de la barnase. A basse température, le domaine barnase est davantage stable et l'équilibre est dirigé vers l'état Bn foldé/Ub non-foldé. Une augmentation progressive de la température jusqu'à 55° C déplace l'équilibre vers la conformation Bn non-foldé/Ub foldé, avec un T_m ("Melting Temperature" = température à laquelle 50 % des protéines sont dénaturées) équivalent à 30°C. Le déplacement conformationnel est réversible. Cette transition peut être suivie par dichroïsme circulaire mais également *in vivo* de par la toxicité intrinsèque de l'état natif de la barnase pour *E. coli*. A 37°C, la toxicité de la protéine chimérique Bn/Ub est moindre qu'à 15°C, ce qui permet aux bactéries de survivre à un taux d'expression de cette même protéine plus élevé. La liaison de l'inhibiteur barstar au niveau de la barnase est capable de promouvoir le repliement de ce domaine et de déplacer l'équilibre vers l'état Bn foldé/Ub non-foldé.

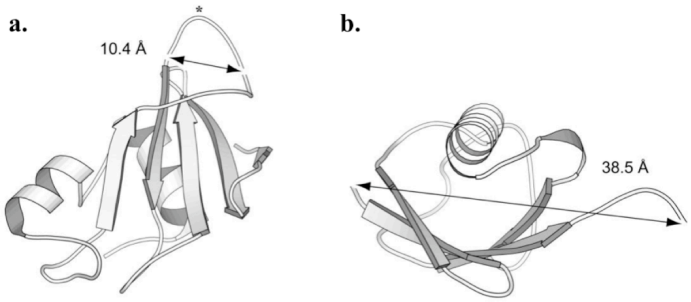


Figure 8 : Strucutre tridimensionnelle de la ribonucléase barnase de chez Bacillus amyloliquefaciens (a.) et de l'ubiquitine humaine (b.). Les distances Cα-Cα entre les résidus Pro64 et Thr70 de la barnase et entre les extrémités C- et N-terminales de l'ubiquitine sont représentées. Figure de Radley *et al.*[48]

Un deuxième exemple d'ingénierie d'une régulation allostérique par sélection conformationnelle tire profit des propriétés structurales de la Maltose Binding Protein d'*E. coli* (MBP)[49,50]. La MBP fait partie du système maltose/maltodextrine responsable, chez *E. coli*, de l'absorption et du catabolisme des maltodextrines. Cette

protéine est constituée de deux domaines reliés par une région charnière, le site de liaison du maltose se trouvant à leur interface. La protéine adopte deux conformations, une dite ouverte et correspondant à la configuration *apo* (sans ligand), l'autre dite fermée et correspondant à la configuration *holo* (avec ligand) **(Figure 9)**. Le changement conformationnel se fait par un mouvement de torsion et de flexion au niveau de la charnière. Il a été démontré par une approche RMN que, en l'absence de ligand, la forme "fermée" existe déjà et que la conversion structurale correspond dès lors à un déplacement d'équilibre[51]. Hellinga et son équipe ont identifié au sein de la MBP, quatre régions subissant un changement conformationnel conséquent suite à la liaison du maltose et situées à bonne distance (> 11 Å) du site de liaison[49] **(Figure 9)**. Le couplage d'un fluorophore IANBD (Dérivé 7-nitrobenz-2-oxa-1,3-diazole) au niveau du site "B" a permis de mesurer des variations de fluorescence (augmentation de 4,4 x) allostériquement couplées à la liaison du maltose.

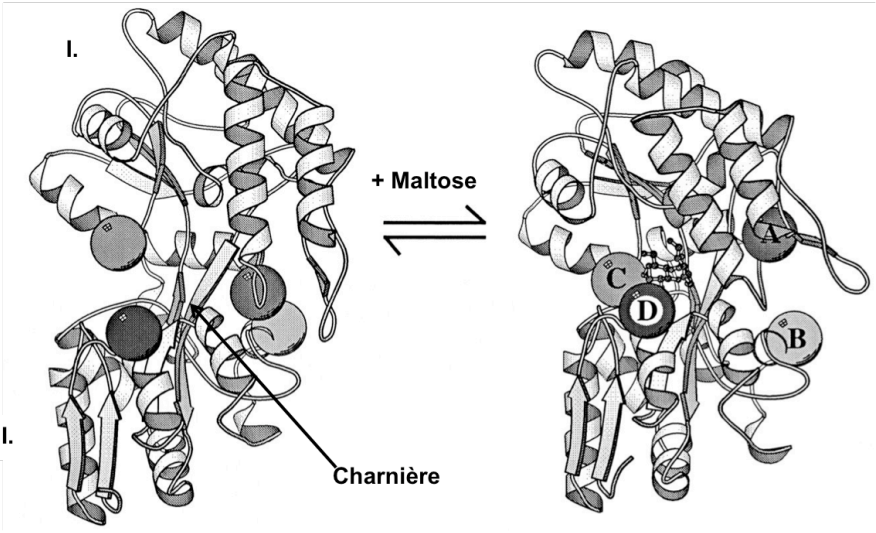


Figure 9 : Structures 3D des configurations *apo* (à gauche) et *holo* (à droite) de la MBP. La liaison du maltose induit un réarrangement spatial des domaines I et II induit par une torsion au niveau de la région charnière (flèche). Le site de liaison du maltose (molécule représentée au niveau de la forme holo) se trouve à l'interface des 2 domaines. Les sphères (A, B, C et D) indiquent 4 positions allostériquement liées au site de liaison du maltose. Figure de Marvin *et al.*[50]

En créant une sphère de coordination du zinc par mutation de quatre résidus du site de liaison en histidines et glutamates (A63H ; R66H ; Y155E ; W340E), il a

également été possible d'engendrer une spécificité au zinc (K_d = 5,1 µM)[49]. Tout comme le maltose, la liaison du zinc induit une augmentation de la fluorescence. Par la suite, l'affinité pour le zinc a été augmentée via l'insertion de mutations stabilisant la forme fermée et déstabilisant la forme ouverte. L'affinité du système pour l'ion métallique a été augmentée d'un facteur +/- 15 (K_d = 350 nM). Une étude structurale par cristallographie et par diffraction des rayons X à de petits angles (SAXS pour « Small-angle X-ray scattering ») a néanmoins mis en évidence que la liaison du zinc n'induisait pas de déplacement d'équilibre de la forme protéique ouverte à la forme fermée[52]. Un changement conformationnel plus restreint est cependant induit et serait suffisant pour perturber le réseau de liaisons hydrogène établi entre le fluorophore et les résidus tyrosine opposés. Cette perturbation pourrait expliquer la modulation de fluorescence observée.

II.3.A.d. Ingénierie d'une coopérativité positive par modulation d'une dynamique moléculaire.

Une allostérie peut également être créée au niveau de molécules autres que des protéines. Le positionnement d'un métal de grande taille, tel le cérium, entre deux molécules de porphyrine forme un complexe de type sandwich **(Figure 10)**. A température ambiante, les deux macrocycles sont animés d'une dynamique rotationnelle l'un par rapport à l'autre[53,54] **(Figure 10)**.

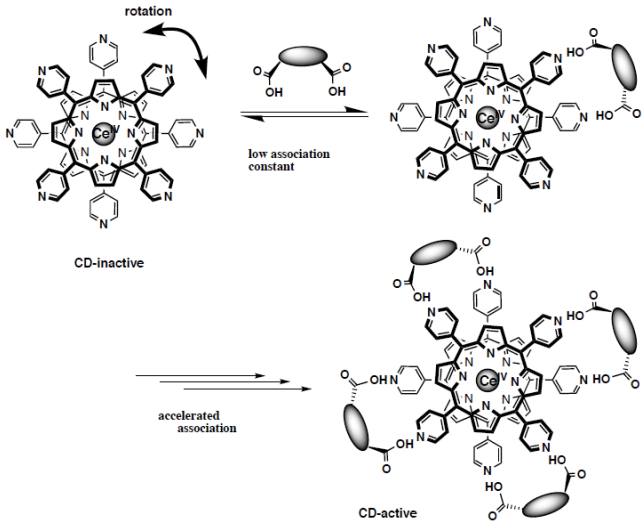


Figure 10 : Représentation schématique de la coordination de diacides au niveau d'un complexe porphyrinique. Figure de Takeuchy *et al.*[53]

Cette propriété a permis à Shinkai et son équipe de développer un système chimique de coopérativité positive[53,54]. Une substitution des quatre positions *méso* de chaque porphyrine par des groupements 4-pyridyle permet au complexe de lier quatre acides dicarboxyliques. La liaison de ces acides se fait de manière coopérative. La coordination du premier acide est relativement difficile mais immobilise les deux plans des porphyrines dans une configuration chirale, observable par dichroïsme circulaire. Une fois la première molécule liée, le coût entropique de liaison est réduit car les sites d'interaction sont déjà idéalement positionnés. Cette réduction permet aux trois autres acides de se lier plus facilement. Un coefficient de Hill de 4, signe d'une coopérativité extrême, a été obtenu.

II.3.B. Approche aléatoire

L'ingénierie d'une régulation allostérique par approche rationnelle est une méthodologie de travail relativement puissante nécessitant une connaissance structurale et fonctionnelle approfondie de la plateforme protéique utilisée. Mais malgré un flux entrant continu d'informations dans les bases de données des protéines, il n'est pas encore possible de prédire avec certitude les modifications à apporter à une protéine en vue de la rendre allostérique. Une deuxième approche d'ingénierie de l'allostérie, dite aléatoire, lève partiellement cette incertitude en permettant de tester simultanément une multitude de variants protéiques. Cette approche, aussi connue sous le nom d'évolution dirigée, applique les mécanismes de base de l'évolution pour créer ou améliorer des fonctions et des caractéristiques protéiques d'intérêt. Deux étapes, souvent itératives, peuvent être reconnues dans cette stratégie. La première consiste à créer une diversité génétique en mutant le gène encodant la protéine ciblée. Dans le cas de l'ingénierie d'une allostérie, cette étape consiste généralement à former un grand nombre de sites allostériques potentiels par mutation ou insertion aléatoire dans la structure protéique. Ensuite, une étape de sélection et/ou de criblage efficace doit permettre d'isoler les mutants intéressants au sein de la diversité créée. Les clones d'intérêt sont, dans le cas présent, ceux régulés par la liaison d'une molécule cible. Le fait de disposer d'une grande diversité de départ ($> 10^6$ clones) permet à partir d'une seule et même collection de sélectionner des clones régulés par des cibles variées ou via des mécanismes moléculaires différents.

II.3.B.a. Ingénierie d'une régulation par fusion aléatoire de domaines protéiques

Les insertions de domaines sont considérées, au même titre que les mutations ponctuelles, comme des évènements importants dans l'évolution des protéines en permettant à celles-ci d'acquérir des fonctions ou caractéristiques nouvelles[55]. Dans les exemples d'ingénierie suivants, les scientifiques sont partis de ce postulat pour créer une régulation allostérique par fusion de domaine à l'image de ce qui a été fait de manière rationnelle avec LOV2 et DHFR (point II.3.A.b.). Stratégiquement parlant, deux protéines existantes, l'une avec la fonction de reconnaissance du signal, l'autre avec l'activité à moduler sont fusionnées de telle sorte que la capture du signal au niveau du 1er domaine affecte la fonction du second[42].

L'une des premières constructions de protéines chimériques allostériques est à mettre à l'actif de Doi et Yanagawa[56]. Ces deux auteurs sont parvenus à développer un senseur spécifique de l'inhibiteur de β-lactamases BLIP (β-lactamase Inhibitory Protein) en insérant une TEM-1 β-lactamase dans une boucle de surface d'une GFP (Green Fluorescent Protein). Le site muté tolère bien les grandes insertions et est proche du fluorophore. L'insertion n'affecte pas les activités de la β-lactamase et de la GFP et la liaison de BLIP n'induit pas de modulation de la fluorescence. Plusieurs tours d' "*error prone PCR*", suivis d'un criblage visuel de l'effet de BLIP sur la fluorescence de bactéries *E. coli* exprimant les différents clones, ont permis d'isoler un clone régulé allostériquement par la liaison du ligand. Après deux tours de sélection un clone dont la fluorescence est activée par BLIP a été sélectionné. Ce mutant possède deux mutations au niveau du gène de la β-lactamase connues pour déstabiliser l'enzyme. La fluorescence de la GFP, affectée par cette déstabilisation, est rétablie par la liaison de BLIP stabilisant la β-lactamase. Cet exemple d'ingénierie est le pendant aléatoire de l'approche utilisée pour la création de la protéine chimérique régulable Ub/Bn (point II.3.A.c.).

La TEM-1 β-lactamase a également été fusionnée avec la MBP (Maltose Binding Protein) afin de développer une protéine chimérique dont l'activité enzymatique est régulée par la liaison du maltose[42,46,57–59] **(Figure 11)**. Le choix des enzymes n'est pas anodin. La MBP, comme vu précédemment, subit un important changement conformationnel suite à la liaison du maltose **(Figure 11.a)**. L'activité β-lactamase

est quant à elle facilement sélectionnable *in vivo,* l'enzyme étant capable d'apporter une résistance à l'ampicilline aux bactéries. Le premier essai a consisté à insérer aléatoirement la β-lactamase au niveau de différents sites de la MBP[42]. Les clones bi-fonctionnels ont été sélectionnés en faisant pousser les banques bactériennes sur un milieu minimum contenant de l'ampicilline et du maltose comme seule source de carbone. La souche bactérienne ayant été utilisée est une souche auxotrophe, capable de pousser sur ce milieu minimum uniquement en présence d'une MBP fonctionnelle. Un criblage par mesure de l'activité β-lactamase en présence et en absence de maltose a permis d'isoler parmi ces clones bi-fonctionnels un mutant allostériquement régulé. L'activité enzymatique de ce mutant est augmentée d'un facteur 2 en présence de maltose. Dans le but d'obtenir une régulation davantage marquée, une diversité conformationelle des protéines chimériques plus importante a été explorée. Le gène de la β-lactamase a ainsi subi, préalablement à son insertion aléatoire au niveau du gène de la MBP, une permutation circulaire également aléatoire[59]. Un mutant (RG13) a été sélectionné comme précédemment et a montré une efficience catalytique (k_{cat}/K_m) augmentée d'un facteur 25 en présence de maltose, la régulation étant réversible. Dans un souci de continuité, la même équipe a développé une stratégie itérative lui permettant de sélectionner les meilleurs sites d'insertion et de permutation au niveau respectivement de la MBP et la β-lactamase[58]. Un des mutants (MBP317-347), présentant une permutation circulaire au niveau de la boucle Ω de la β-lactamase a vu son activité enzymatique améliorée d'un facteur 600 x en présence de maltose **(Figure 11.b)**. *In vivo*, ce phénomène de régulation permet d'augmenter le MIC des bactéries pour l'ampicilline d'un facteur 16 en présence du ligand.

De façon inattendue, les mutants RG13 et MBP317-347 présentent tous deux une régulation allostérique (inhibition) par la liaison d'ions zinc avec une affinité micromolaire[57]. Cette régulation est d'autant plus surprenante que ni la β-lactamase ni la MBP ne disposent d'une telle affinité pour les ions Zn^{2+}. Cette inactivation, complète et réversible, affecte le k_{cat} de l'enzyme mais pas son K_m éliminant de ce fait la possibilité de toute inhibition autre que la non-compétitive pure. Au vu de ces résultats, Ostermeier et son équipe ont posé l'hypothèse que la reconnaissance d'un ligand et l'allostérie étaient deux phénomènes pouvant émerger simultanément[57,60]. Les changements inhérents à l'apparition d'une affinité pour un nouveau ligand

pourraient prédisposer le mutant à présenter une régulation allostérique[60]. Nussinov *et al.*[9] vont encore plus loin dans ce raisonnement en arguant que l'existence systématique d'un équilibre conformationnel ferait de chaque protéine un système allostérique potentiel. Le tout serait de trouver le bon ligand, les bonnes conditions ou les mutations ponctuelles appropriées capables de déplacer cet équilibre vers des conformères plus ou moins actifs[9].

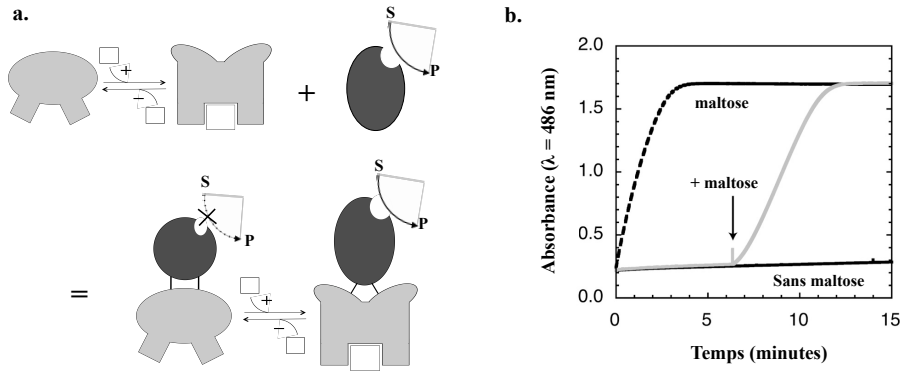


Figure 11 : (a.) Création d'une régulation allostérique par fusion d'une MBP (gris clair) et d'une β-lactamase (gris foncé). L'activité β-lactamase est modulée (activation) par l'ajout de maltose (carré blanc). Figure adaptée de Ostermeier[47]. (b.) Hydrolyse de la nitrocéfine (80 µM) par la protéine de fusion MBP317-347 en présence et en absence de maltose 5 mM à 25°C. Figure de Guntas *et al.*[58]

II.3.B.b. Insertions de peptides aléatoires dans des boucles de surface

De manière analogue à la technique d'insertion d'épitopes (point II.3.A.a.), l'équipe de Fastrez et Soumillion a développé une certaine expertise dans la création *de novo* de sites allostériques par randomisation de boucles de surface au niveau de la TEM-1 β-lactamase. La stratégie consiste en la construction d'une banque de mutants de haute diversité suivie d'une sélection de clones reconnaissant une molécule d'intérêt par la technique du phage display (voir aparté, p.34). Un criblage par analyse cinétique permet ensuite d'isoler, parmi les clones "liants", ceux régulés allostériquement par la fixation du ligand. Une des forces du système réside dans la possibilité de cibler différents ligands à partir d'une seule et même banque. Dans la première banque construite, des peptides aléatoires (3 à 7 résidus) ont été insérés au niveau de deux boucles de surface situées de part et d'autre du site actif de l'enzyme[61] **(Figure 12)**.

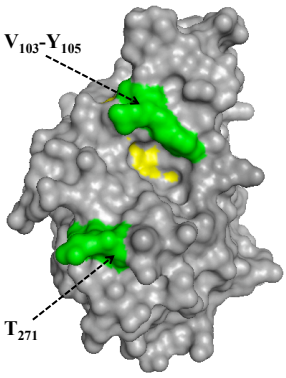


Figure 12 : Localisation des boucles de surface randomisées (en vert) sur la structure 3D de la TEM-1 β-lactamase (PDB 1ZG4). L'acide aminé T_{271} a été remplacé par des inserts aléatoires de 5 à 6 résidus. Les acides aminés V_{103} à Y_{105} ont été remplacés par des inserts aléatoires de 3 à 7 résidus. Le site actif est représenté en jaune. La structure a été générée à partir du programme MacPymol.

Une série de clones reconnus par divers anticorps monoclonaux dirigés contre l'antigène prostatique spécifique (PSA) ont pu être sélectionnés à partir de cette collection. Parmi ces mutants, plusieurs ont présenté des inhibitions d'activité marquées (de 50 à 90 %) suite à la liaison du mAc. Un des clones a montré un comportement intéressant, la liaison du mAc menant à une inhibition ou à une activation en fonction du substrat β-lactame utilisé[61]. Des régulations d'activité modérées au niveau d'autres β-lactamases mutantes, liant la streptavidine ou la β-galactosidase, ont pu être obtenues en appliquant la même méthodologie (résultats non publiés).

Fort de ce succès, le laboratoire s'est intéressé à la création de β-lactamases régulées allostériquement par la liaison de ligands de petite taille (molécules organiques ou ions). La reconnaissance de petites molécules nécessite de créer des sites complexes, les ligands de faible poids moléculaire étant souvent logés dans des cavités afin de maximiser les contacts intermoléculaires[62]. De même, les sites allostériques sont régulièrement situés dans des régions peu structurées, constituées de boucles de surface[62,63]. La transition structurelle de ces sites, induite par la liaison du ligand, jouerait un rôle important dans la transmission du signal allostérique. En accord avec ces constatations, trois boucles de surface contiguës de la TEM-1 β-lactamase, L1, L2 et L3 ont été randomisées par l'insertion de peptides aléatoires ou par mutation ponctuelle[64,65] **(Figure 13)**. Le choix des boucles a principalement été dicté par leur

tolérance aux mutations et insertions ainsi que par leur proximité spatiale les unes envers les autres et envers le site actif. Considérées comme un ensemble, la position de ces trois boucles mime assez bien l'ancrage des trois CDR (Complementary Determining Region) des anticorps monochaînes de camélidés. La banque construite dans un vecteur phagique fd-tet a été exprimée à la surface de phages et sélectionnée pour affinité envers des ions métalliques. Cinq tours de sélection ont suffi pour isoler une série de clones actifs présentant plusieurs histidines réparties au sein des trois boucles. Un tiers de ces mutants étaient soit activés soit inhibés en présence d'ions métalliques (Ni^{2+}, Zn^{2+}ou Cu^{2+}). Un clone particulier présentait les deux types de régulation en fonction de l'ion bivalent utilisé : activation avec Ni^{2+} et inhibition avec Cu^{2+}. Les affinités de l'ordre du micromolaire obtenues impliquent la participation de plus d'une histidine dans la liaison et donc une grande probabilité de changements conformationnels intra et/ou inter-boucle.

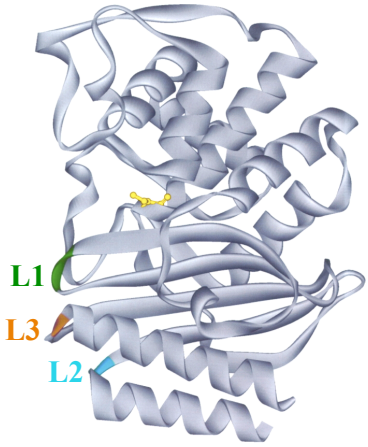


Figure 13 : Localisation des boucles randomisées L_1, L_2 et L_3 sur la structure 3D de la TEM-1 β-lactamase. Les résidus 238 à 240 ont été mutés aléatoirement au niveau de la boucle L_1. Des inserts de 6 à 9 résidus aléatoires ont été insérés à la place des acides aminés G_{41}-A_{42} de la boucle L_2. Des hexapeptides dégénérés ont été insérés à la place de la T_{271} pour former la boucle L_3. La sérine 70 du site actif est représentée en jaune. Figure de Mathonet *et al.*[64]

La kanamycine, une molécule hydrophile avec laquelle il est difficile d'interagir fortement dans l'eau, a par la suite été ciblée à partir de la même collection afin de valider la stratégie d'ingénierie[66]. La banque a subi plusieurs tours de sélection par capture d'affinité à l'aide de la kanamycine A (KanA) biotinylée et immobilisée sur billes de streptavidine. Un clone présentant une affinité pour le ligand a été isolé et soumis à un cycle d'*error prone PCR* suivi de cinq nouveaux tours de sélection en

vue d'une maturation de l'affinité. Un mutant (BlaKr) fortement régulé (activation) par KanA en présence d'aminosulfonates a été sélectionné et davantage caractérisé[67] **(Figure 14)**. Ce mutant diffère de l'enzyme sauvage par le remplacement des résidus $G_{41}A_{42}$ de la boucle L2 par un insert de 8 acides aminés ($C_{41}R_{41a}T_{41b}S_{41c}H_{41d}R_{41e}P_{41f}C_{42}$). Au niveau de la boucle L_1 le résidu E_{240} est muté en arginine et la séquence de la boucle L_3 $Q_{269}A_{270}T_{271}M_{272}$ est remplacée par $R_{269}K_{270}K_{271}T_{272}$. La mutation $E_{104}K$ est également présente. Par analyse cinétique il a été montré que l'activation est indirecte et consiste très probablement en l'expulsion, par KanA, d'un aminosulfonate constituant du tampon et agissant comme un inhibiteur quasi compétitif. L'enzyme possède à ce titre deux sites se chevauchant (physiquement ou fonctionnellement) : un site de faible affinité pour les acides aminosulfoniques et un site allostérique de haute affinité (nM) pour les aminglycosides. Une étude par RMN montre qu'en présence de MES (acide 2-(N-morpholino)éthane-sulfonique), l'ajout de KanA à BlaKr provoque des perturbations des déplacements chimiques dans la région des boucles randomisées, site de liaison probable **(Figure 14.b)**. La région perturbée inclut également les résidus N_{132} et E_{166} importants pour l'activité β-lactamase. L'absence de déplacements chimiques en tampon phosphate suggère que, soit les changements conformationnels sont davantage dus à l'expulsion du MES qu'à la liaison de la kanamycine, soit le phosphate se lie fortement à l'enzyme et empêche la liaison de KanA. A noter également que l'inhibiteur ne se lie pas au site actif, une certaine activité sur nitrocéfine étant maintenue à saturation en MES. La prise en compte de l'ensemble des résultats suggère que l'effet d'inhibition ne soit pas uniquement dû à un phénomène d'expulsion stérique du substrat. L'effet pourrait également provenir d'un déplacement conformationnel de BlaKr.

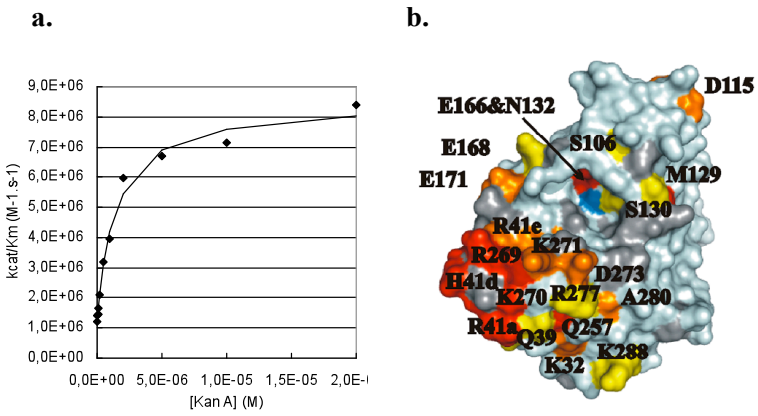


Figure 14 : (a.) Effet de l'augmentation progressive de la concentration en Kanamycine A sur l'efficience catalytique (k_{cat}/K_m) de BlaKr à pH 6,3 dans tampon 24mM MES/ 24 mM Bistris (25°C). (b.) Cartographie des déplacements chimiques de BlaKr en présence de KanA. Les résidus sont colorés en fonction de l'importance du déplacement chimique subi (δ) : cyan (Résidus non affectés $\Delta\delta_{moy}$ < moyenne + 1/2 σ) ; jaune ($\Delta\delta_{moy}$ > moyenne + 1/2 σ) ; orange ($\Delta\delta_{moy}$ > moyenne + σ) ; rouge ($\Delta\delta_{moy}$ > moyenne + 2σ). La sérine 70 est en bleu. Figures de Volkov *et al.*[67]

Aparté : Phage display[68–71]

Le phage display est une technique de sélection *in vitro* utilisant le plus souvent des phages filamenteux de la famille Ff (fd, f1, M13) pour établir un lien entre la protéine (ou peptide) d'intérêt et l'information génétique l'encodant **(Figure 15)**. La connexion est établie en fusionnant le gène de la protéine à celui d'un constituant de la capside du bactériophage. Une banque de "phage display" est donc un pool hétérogène de phages, chacun portant un insert ADN différent et présentant à sa surface une protéine encodée par cet ADN. Grâce aux techniques de biologie moléculaire actuelles, la diversité de telles collections peut atteindre 10^{11} variants différents d'une même protéine[70].

Un phage filamenteux Ff est constitué d'un vecteur ADN_{ss} (≈ 6,4 kb) encapsidé dans un "cylindre" formé de +/- 2700 copies (en fonction de la taille du génome) de la protéine p8 et fermé à ses extrémités par quatre types de protéines (p3, p6, p7 et p9) présentes en quelques copies (~ 5) **(Figure 15.a)**. La protéine p3, outre son implication dans la structure du virus, joue un rôle important dans l'infection de la bactérie hôte. C'est en effet via son extrémité N-terminale que le phage se lie aux *pili*

"F" des bactéries et que la translocation de l'ADN viral est initiée. Une fois à l'intérieur de la cellule, l'ADN simple brin est converti en sa forme réplicative double brin. La biosynthèse de nouvelles particules phagiques peut alors commencer, le cycle étant non-lytique. Les séquences peptidiques "étrangères" sont souvent fusionnées au niveau des extrémités N-terminales de la p8 ou de la p3, cette dernière tolérant des insertants de grande taille telle la pénicilline acylase (86 kDa). Dans notre cas, nous nous intéresserons uniquement au système "p3". Trois types de "display" sont possibles[71] **(Figure 15.a)** : Dans le "type 3", la protéine d'intérêt est fusionnée à l'unique p3 du génome. La protéine étrangère est donc en théorie présentée à la surface du phage en cinq exemplaires. Ce nombre n'est toutefois que rarement atteint, l'activité protéolytique de la bactérie clivant une partie de ces protéines. Dans le "type 33", la protéine p3 recombinante est accompagnée d'une p3 sauvage (wt). Les particules virales obtenues sont de type mosaïque, leur capside présentant à la fois des p3 wt et des p3 recombinantes. Le "type 3+3" diffère du précédent de par le fait que les deux gènes de la p3 se trouvent sur deux génomes distincts. La p3 wt est sur un phage dit "helper" alors que la version recombinante se trouve sur un phagemide. Le phagemide est un plasmide présentant un marqueur de résistance, deux origines de réplication (double et simple brin), le gène de la protéine de fusion sous contrôle d'un promoteur inductible ainsi qu'une séquence signal d'encapsidation. Le phagemide doit être secouru par le phage helper encodant le reste des protéines phagiques afin de produire des particules virales. Le phagemide sera préférentiellement encapsidé, le vecteur phagique étant déficient au niveau de la séquence d'encapsidation. A noter également que la p3 recombinante est souvent délétée de son domaine N-terminal, celui-ci protégeant les bactéries de la surinfection (nécessaire dans ce cas-ci). L'intérêt du système porte notamment sur la plus petite taille et le "copy number" plus élevé du phagemide le rendant plus génétiquement manipulable. L'induction contrôlée de la protéine de fusion permet également d'exprimer plus facilement des molécules toxiques[68].

Un cycle de sélection pour affinité par "phage display" comporte différentes étapes **(Figure 15.b)** : (1) - *Encapsidation* - La collection de phages est produite à partir de bactéries transformées par la banque ADN. (2) - *Liaison* - Les phages sont incubés en présence d'une molécule cible immobilisée sur un support. Les clones reconnaissant la cible via leur protéine recombinante sont capturés. (3) - *Lavages* -

Les phages non-liés ou liés aspécifiquement sont éliminés par des lavages. (4) - *Elution* - Les clones spécifiques sont récupérés par élution spécifique (Elution compétitive,...) ou non spécifique (Elution acide, clivage protéolytique,...). (5) - *Amplification* - Les phages récupérés sont amplifiés par infection de bactéries. (6) - *Criblage ou second tour* - Les clones amplifiés sont soit criblés pour la fonction d'affinité recherchée soit soumis à un second tour de sélection.

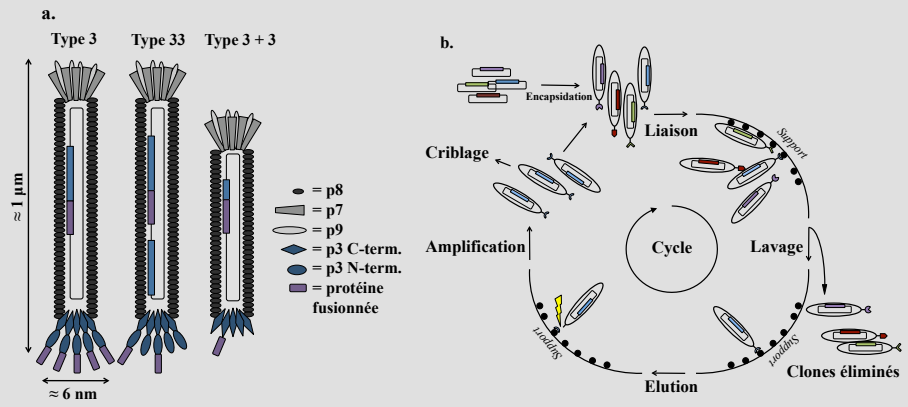


Figure 15 : (a.) Représentation schématique d'un phage filamenteux et de trois types de phage display. Le phage du type 3+3 porte le phagemide d'où sa plus petite taille. Les détails sont abordés dans le texte. (b.) Expérience type de sélection pour affinité par phage display. Les différentes étapes sont décrites dans le texte.

Outre l'affinité, la banque peut être sélectionnée pour d'autres caractéristiques en adaptant adéquatement le protocole : Sélection pour activité catalytique[72], sélection pour stabilité (résistance protéolytique)[73], sélection de séquences substrats de protéases[74],...

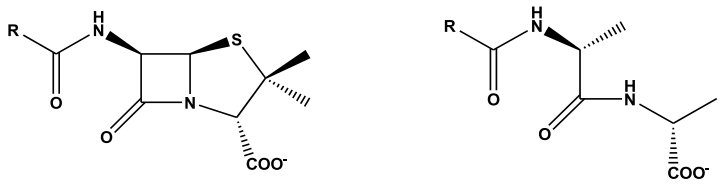
III. La TEM-1 β-lactamase

Comme précédemment mentionné, la TEM-1 β-lactamase est l'enzyme modèle utilisée par le laboratoire pour l'ingénierie d'une régulation allostérique. Cette seconde partie de l'introduction s'attardera sur la description générale de cette enzyme ainsi que sur son utilisation comme plateforme d'ingénierie et les connaissances en découlant.

III.1.Description générale, structurale et fonctionnelle

III.1.A. Généralités

Les β-lactamases sont une famille d'enzymes (EC 3.5.2.6) produites par certaines bactéries et responsables de leur résistance aux pénicillines, céphalosporines et autres dérivés β-lactame. Ces dernières molécules présentent une similarité structurale avec l'extrémité D-alanyl-D-alanine des précurseurs du peptidoglycane, substrat naturel des DD-peptidases[75,76] **(Figure 16)**. L'activité antibiotique des β-lactames s'explique par cette ressemblance qui leur permet d'être reconnus par les DD-peptidases et de former un acyle-enzyme stable incapable de jouer son rôle dans la biosynthèse de la paroi bactérienne[76,77].



Pénicillines **D-alanyl-D-alanine**

Figure 16 : Similarités structurales entre les pénicillines et l'extrémité D-alanyl-D-alanine du peptidoglycane.

Les β-lactamases exportées dans le périplasme chez les $Gram^-$ ou secrétées dans le milieu extérieur chez les $Gram^+$ agissent en inactivant le noyau β-lactame de ces antibiotiques par hydrolyse. Ces enzymes sont encodées par un gène chromosomique ou plasmidique et leur expression est induite ou constitutive. Les β-lactamases se voient classiquement réparties en quatre groupes sur base de leur structure primaire[78]. Les classes A, C et D ont comme caractéristique commune la présence d'une sérine catalytique au niveau de leur site actif. Les enzymes de classe

B ou métallo-β-lactamases nécessitent pour leur part la présence d'un ion zinc pour être actives. A noter qu'il existe également une classification basée sur les propriétés fonctionnelles (spécificité de substrats) des β-lactamases[79]. La TEM-1 β-lactamase utilisée comme enzyme modèle dans ce travail fait partie de la classe A et est responsable de 90 % des phénomènes de résistance à l'ampicilline observés chez *E. coli* [80].

Le mécanisme d'action général des β-lactamases à sérine comprend deux étapes principales **(Figure 17)**. Durant la première étape dite d'acylation l'enzyme et le substrat forment d'abord un complexe non covalent qui sera converti par la suite en un acyle-enzyme instable. La seconde étape dite de déacylation correspond à une hydrolyse de l'ester de l'acyle-enzyme permettant la régénération de l'enzyme et la libération du substrat clivé.

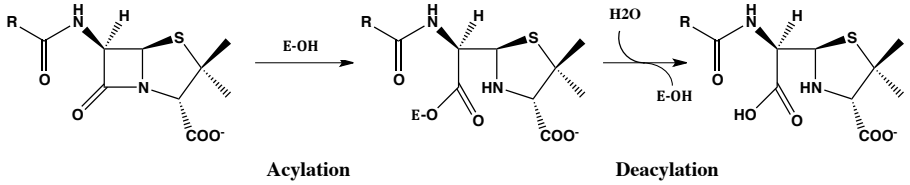


Figure 17 : Mécanisme d'action général de l'hydrolyse d'un antibiotique de type β-lactam par une β-lactamase (E-OH).

Les étapes d'acylation et de déacylation sont en général très rapides faisant des β-lactamases des enzymes quasi parfaites avec un taux d'acylation (k_{cat}/K_m) de 10^7 - 10^8 $M^{-1}.s^{-1}$, valeur proche de la constante de vitesse de diffusion des petites molécules[81]. L'inhibition des DD-peptidases par les antibiotiques à noyau β-lactame procède par le même mécanisme mais la constante de déacylation est très faible (10^{-4}-10^{-3} s^{-1})[81,82].

III.1.B. Structure, repliement et stabilité : informations générales

La TEM-1 β-lactamase est une enzyme de +/- 30 kDa (263 acides aminés) composée de deux domaines connectés entre eux par deux charnières **(Figure 18)**. Le premier domaine comprend uniquement des hélices α (= domaine α) tandis que le second est formé à la fois d'hélices α et de feuillets β (= domaine α/β). Le site actif de l'enzyme se trouve à l'interface de ces deux domaines et est caractérisé par plusieurs motifs conservés. Le premier motif (Ser_{70}-Xaa-Xaa-Lys_{73}) est situé à l'extrémité N-

terminale d'une hélice α et comprend la sérine catalytique suivie à un tour d'hélice par une lysine essentielle dont la chaîne latérale pointe vers le site actif [77]. Le second élément est formé par la triade "Ser_{130}-Asp-Asn_{132}", les chaînes latérales du premier et du dernier résidu étant dirigées vers la cavité catalytique[82]. Dans certaines β-lactamases homologues, l'asparagine est remplacée par une sérine ou une glycine. L'asparte ne joue pas de rôle direct dans la catalyse mais est important pour la stabilité de l'enzyme[83]. L'asparagine permet de positionner adéquatement le substrat via la formation d'un pont hydrogène avec le groupe carbonyle du noyau β-lactame.

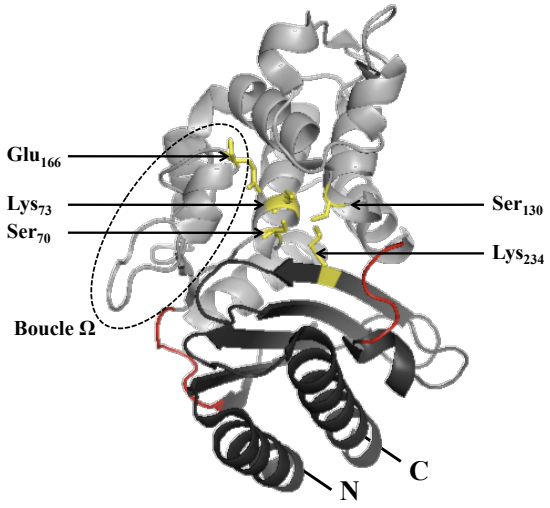


Figure 18 : Structure 3D de la TEM-1 β-lactamase (PDB 1ZG4). Le domaine α est représenté en gris et le domaine α/β en noir. Les charnières sont colorées en rouge et le site actif en jaune. Certains résidus catalytiques importants sont annotés. Figure réalisée avec MacPymol.

Le troisième motif (Lys_{234}-(Thr ou Ser)-Gly_{236}) est situé sur la paroi opposée du site actif, la lysine formant un pont hydrogène avec la serine du second motif[82]. La conservation de la glycine est expliquée par l'entrave aux substrats que provoquerait la présence de toute chaîne latérale à cet endroit[82]. Un quatrième élément, la boucle oméga « Ω », propre aux β-lactamases de classe A complète le site catalytique. Cette boucle se situe à l'entrée du site catalytique et jouit d'une certaine flexibilité. Elle contient généralement la séquence Glu_{166}-Xaa-Glu-Leu-Asn_{170}. Le glutamate et l'asparagine semblent essentiels pour le positionnement d'une des deux molécules d'eau conservées (W1) près de la sérine catalytique[77]. Le glutamate joue également un rôle important dans les étapes d'acylation et de déacylation (voir point suivant).

Au niveau de la TEM-1, une proline occupe la position 167 et adopte une conformation *cis* nécessaire au bon positionnement des résidus Glu_{166} et Asn_{170} [84]. L'isomérisation de cette proline serait un événement limitant dans la cinétique de repliement de la TEM-1 β-lactamase. L'équilibre de dénaturation de l'enzyme induite par un agent chaotropique (e.g. chlorure de guanidium) se déroule en deux étapes, ce qui témoigne de l'existence d'un intermédiaire de repliement thermodynamiquement stable[84,85]. Cet état intermédiaire est dépourvu d'une organisation tertiaire bien définie mais présente un contenu élevé en structures secondaires. Ces caractéristiques correspondent à la définition du terme « molten globule » régulièrement employé pour référer à certains intermédiaires de repliement. La seconde phase de repliement serait cinétiquement contrôlée par le passage de la proline d'une position *trans* au niveau de l'intermédiaire en une position *cis* au niveau de la forme native. L'énergie libre de dénaturation de la TEM-1 β-lactamase a une valeur estimée de 44 ± 3 kJ/mol [85], ce qui en fait une protéine relativement stable. L'état intermédiaire est lui-même 23,9 ± 0,8 kJ/mol plus stable que l'état complètement dénaturé.

III.1.C Mécanisme catalytique

Au cours de l'étape d'acylation, un intermédiaire acyle-enzyme est formé par l'attaque nucléophile de la sérine catalytique (Ser70) sur le groupement carbonyle du noyau β-lactame[86] **(Figure 19)**. Si cette action est unanimement acceptée il n'en est pas de même de l'identité de la base catalytique augmentant le pouvoir nucléophile de la sérine. Les résidus Glu166 et Lys73 ont tous deux été proposés comme candidat possible[86–89]. Ce débat sort du cadre de cette thèse et il est probable que les deux résidus agissent de concert. Seul le mécanisme faisant intervenir le glutamate sera abordé dans cette partie, celui-ci étant le plus couramment cité. La base catalytique active la sérine 70 via l'intermédiaire d'une molécule d'eau (W1), la distance séparant les deux résidus ne permettant pas une interaction directe[81] **(Figure 19.a)**. Il a également été proposé que la flexibilité de la boucle Ω puisse intervenir dans le rapprochement des deux résidus[81]. Le proton libéré au cours de l'attaque nucléophile serait transmis à l'atome d'azote du cycle β-lactame via un réseau de ponts hydrogènes impliquant une seconde molécule d'eau (W2), les groupes aminés des lysines 73 et 234 ainsi que le groupe hydroxyle de la sérine 130

(Figure 19.b). Les groupements amines (-NH) de la Ser70 et de l'Ala237 formeraient un trou oxyanion stabilisant l'état de transition (intermédiaire tétraédrique) chargé négativement via des liaisons hydrogène[81] **(Figure 19)**. L'étape de déacylation est moins sujette à controverse et serait également catalysée par le glutamate 166. Celui-ci activerait la première molécule d'eau (W1) lui permettant d'attaquer le carbonyle de l'acyle-enzyme[90]. Une nouvelle molécule d'eau (W1') se substituerait à la première afin que le résidu Glu166 puisse rendre un proton à l'oxygène de la sérine 70.

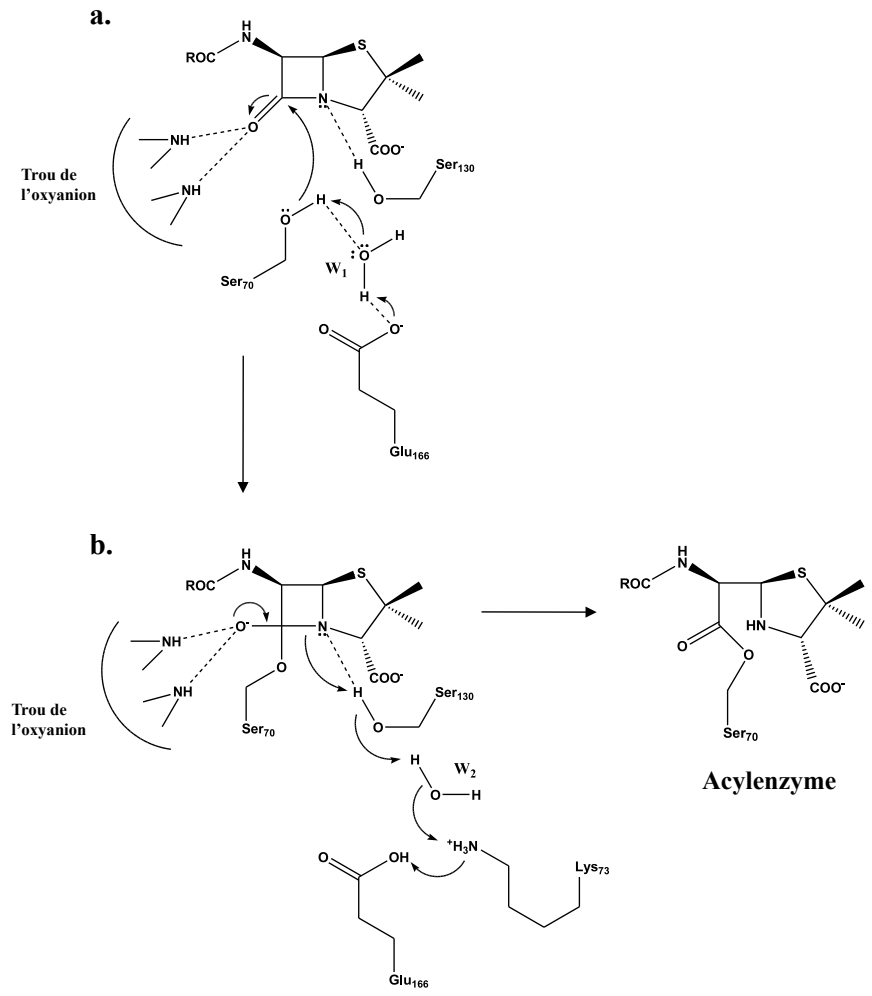


Figure 19 : Mécanisme d'action de l'acylation d'un antibiotique à noyau β-lactam par une β-lactamase de classe A. Les lignes discontinues correspondent à des ponts hydrogène. Les flèches indiquent le déplacement des électrons. Pour plus de clarté, la lysine 234 n'est pas représentée.

III.2. La β-lactamase comme plateforme d'ingénierie

L'un des premiers choix qui s'impose lors de la mise en place d'une étude d'ingénierie est la sélection de l'enzyme de départ. La TEM-1 β-lactamase présente plusieurs avantages pratiques et fait partie des protéines régulièrement utilisées comme plateforme d'ingénierie. Comme précédemment mentionné, il s'agit d'une protéine monomérique bien caractérisée et de haute activité. Sa tolérance aux mutations et insertions est bien documentée et sera abordée dans ce chapitre. Protéine d'origine bactérienne, sa production s'en trouve facilitée et son activité est facilement sélectionnable *in vivo*, les bactéries hôtes devenant résistantes à l'ampicilline. L'enzyme est de plus aisément exprimée sur phage et des substrats chromogéniques et fluorogéniques existent pour la mesure de son activité[91]. Divers exemples d'ingénierie d'une régulation allostérique utilisant la TEM-1 β-lactamase ont déjà été abordés au point "II.3". Dans la partie suivante, d'autres études d'ingénierie (activité catalytique, stabilisation protéique) faisant intervenir cette enzyme seront décrites.

III.2.A. Tolérance aux mutations

Des 263 résidus d'acides aminés constituant la TEM-1 β-lactamase, tous ne contribuent pas de manière équivalente à la structure et à la fonction de la protéine[92]. Certains résidus sont davantage importants et ne tolèrent que peu les mutations substitutionnelles et/ou insertionnelles. Au contraire, certaines positions au sein de la protéine sont plus sujettes à variation. La tolérance générale d'une protéine vis-à-vis des événements mutationnels est, comme nous le verrons au point IV, un facteur clé de son potentiel d'évolution.

Palzkill et son équipe ont étudié la tolérance aux substitutions de chaque résidu de la TEM-1 β-lactamase par mutagenèse saturante[92]. La stratégie consistait à muter aléatoirement trois codons contigus et à créer une banque de mutants contenant toutes les substitutions en acide aminé possibles pour la région mutagénisée. La construction de 88 collections a permis d'explorer l'entièreté de la séquence protéique. Pour chaque collection, une moyenne de neuf mutants capables de pousser sur ampicilline (1 mg/ml) ont été sélectionnés *in vivo* et séquencés afin de définir l'identité des substitutions permises pour chaque résidu. Il est ressorti de cette

étude que 43 positions sur 263 (16 %) ne tolèraient aucune substitution **(Table 1 ; Figure 20.a)**. D'une facon plus positive, cela signifie que 220 résidus tolèrent des substitutions ce qui fait de la TEM-1 β-lactamase une enzyme relativement robuste aux mutations. Parmi les 43 résidus "essentiels", 17 participent de manière directe ou indirecte (stabilisation du substrat) à l'activité catalytique de l'enzyme. Onze résidus se trouvent dans le coeur hydrophobe de la protéine ou à l'interface des deux domaines mais ne représentent que 16 % des résidus enfouis. D'autres résidus sont essentiels de par la formation de ponts hydrogène ou d'interactions ioniques maintenant l'intégrité de la protéine. Enfin l'étude relève une certaine corrélation entre l'accessibilité au solvant et la variabilité séquentielle, les résidus enfouis étant plus sensibles aux substitutions.

Résidus de la TEM-1 β-lactamase ne tolérant pas de substitutions

Leu30	**Thr71**	**Ala125**	**Arg164**	Thr181	**Leu220**	**Asp233**	**Arg244**	**Asn276**
Tyr46	**Lys73**	**Ser130**	**Glu166**	Trp210	**Arg222**	**Lys234**	**Gly245**	**Ile282**
Phe66	Leu76	**Asp131**	Asp176	**Asp214**	Pro226	**Ser235**	**Leu250**	Trp290
Pro67	Pro107	**Asn132**	**Asp179**	**Val216**	Trp229	**Gly236**	Gly251	
Ser70	**Leu122**	**Ala134**	**Thr180**	**Ala217**	Ala232	Gly242	**Met272**	

Table 1 : Positions au niveau de la TEM-1 β-lactamase qui ne tolèrent pas de substitutions. Les résidus intervenant directement ou indirectement dans la catalyse sont en rouge. Les résidus enfouis dans la structure protéique sont en bleu. Sont mis en gras les résidus formant des interactions (ponts hydrogène, liaisons ioniques) essentielles à l'intégrité de l'enzyme. Tableau reproduit de Huang *et al.*[92]

L'insertion aléatoire de pentapeptides au sein de la TEM-1 β-lactamase a également permis de dresser une cartographie de la tolérance de l'enzyme aux mutations insertionnelles[93]. La stratégie s'est basée sur l'insertion aléatoire du transposon TN4430 dont la déletion *in vitro* résulte en l'intégration d'une séquence aléatoire de 15 bp au niveau du site d'entrée. L'effet de 23 insertions sur l'activité β-lactamase a été évalué en déterminant la concentration maximale en ampicilline tolérée par les bactéries exprimant ces mutants. Ces derniers ont été répartis dans trois classes : Cinq insertants confèraient une résistance à l'ampicilline élevée (> 2500 µg/ml) ; Huit autres ne possèdaient pas de résistance détectable ; enfin les 10 mutants restants affichaient des niveaux de résistance intermédiaires (50-1500 µg/ml). Les insertions délétères pour la résistance à l'ampicilline sont localisées dans le coeur protéique ou aux alentours du site actif (**Figure 20.b)**. A l'inverse, les insertions maintenant une

haute activé β-lactamase sont situées au niveau de boucles de surfaces distantes du site actif (entre le brin β5 et l'hélice α8 et entre les hélices α6 et α7) (**Figure 20.b)**. Les autres insertions sont dispersées au sein de la protéine, dans les structures secondaires ne participant pas au site actif ainsi que dans une des régions charnières (**Figure 20.b)**.

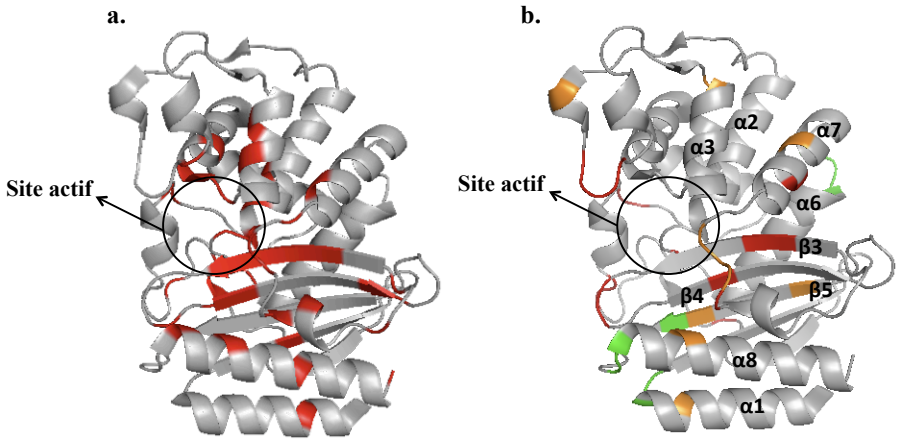


Figure 20 : Cartographie de la tolérance aux substitutions (a.) et aux insertions (b.) de la TEM-1 β-lactamase. Les sites en rouge correspondent aux résidus ne tolérant pas de substitution ou d'insertion. Les régions en orange et en vert indiquent respectivement les positions où l'insertion est moyennement ou bien tolérée. Représentations basées sur les travaux de Huang *et al.*[92], Hallet *et al.*[93], Legendre *et al.*[94], Mathonet *et al.*[64]

En application de ces résultats, Galleni et son équipe se sont servi de la TEM-1 β-lactamase comme protéine "porteuse" (carrier) afin d'induire une réponse immunitaire à l'encontre d'épitopes faiblement immunogènes[95]. Pour ce faire, des peptides dont l'entérotoxine thermostable STa d'*E. coli* entérotoxique ont été insérés au niveau de huit sites définis comme permissifs, semi-permissifs ou non-permissifs par la pentapeptidogenèse. L'effet insertionnel sur l'activité et la production de la β-lactamase a été mesuré. Bien qu'une concordance générale existe entre les résultats des deux études, quelques divergences sont apparues. Ainsi, une insertion au niveau de la position 260 (β5) est davantage tolérée dans ce cas-ci même si elle induit une sensibilité aux protéases accrue par rapport à l'enzyme sauvage. A l'inverse, une insertion au niveau du site 198 (entre hélices α6 et α7) défini comme permissif par la pentapeptidogenèse induit une importante diminution de la résistance à l'ampicilline.

Des insertions au niveau des résidus 216 (charnière), 195 (entre hélices α6 et α7) et même 232 (β3) sont bien tolérées et ont permis une immunisation envers STa.

La tolérance aux insertions de différentes boucles de surface de la TEM-1 β-lactamase a également été explorée au laboratoire lors de l'ingénierie d'une régulation allostérique[61,64,94] (voir point II.3.B.b). La tolérance a été évaluée par mesure du pourcentage d'insertants conférant aux bactéries une résistance à l'ampicilline supérieure à 2,5 % de la résistance d'origine. Contrairement aux deux études précédentes, ce travail a permis d'étudier la tolérance aux insertions dans un contexte de variation aléatoire de la séquence et de la taille de l'insert, deux facteurs pouvant influencer fortement les résultats. Les boucles connectant les hélices α2 et α3 ainsi que les brins β3 et β4 sont sensibles aux insertions, même d'un seul acide aminé **(Figure 20.b)**. Cette intolérance est certainement à mettre en relation avec leur proximité au site actif. Les boucles reliant l'hélice N-terminale α1 et le brin β1 ainsi que le brin β5 et l'hélice C-terminale α8 sont davantage tolérantes bien que la première nécéssite la présence d'un pont disulfure rigidifiant l'insert et diminuant de ce fait le coût entropique de repliement **(Figure 20.b)**.

III.2.B. Ingénierie de la spécificité

Bon nombre d'études ont soumis la TEM-1 β-lactamase à une pression mutationnelle variable et sélectionné des mutants présentant des propriétés catalytiques modifiées[96–100]. L'une des fonctionnalités les plus classiquement améliorées concerne l'activité catalytique envers les céphalosporines de 3ème génération (e.g. céfotaxime), antibiotiques faiblement hydrolysés par l'enzyme TEM-1[96–98,100,101].

Par DNA shuffling, une enzyme mutante présentant un MIC pour la céfotaxime 32.000 fois plus élevé (640 µg/ml) que celui de l'enzyme sauvage a été sélectionnée[96]. La création de diversité par DNA shuffling implique une recombinaison homologue d'un pool de gènes parentaux (ou comme dans le cas présent randomisés par *error prone PCR*) fragmentés aléatoirement et réassemblés par PCR. Les mutants ainsi créés ont été soumis à plusieurs tours de sélection *in vitro* sur des concentrations croissantes en céfotaxime. Trois cycles de shuffling suivis de deux cycles de backcross avec le gène sauvage ont suffi pour isoler un clone présentant une efficience catalytique 3585 fois plus élevée que celle de

l'enzyme sauvage. Parmi les mutations présentes au niveau du gène de la β-lactamase évoluée, quatre (E104K, M182T, G238S et R241H) sont régulièrement observées chez des variants naturels à large spectre antibiotique de la TEM-1[102]. Les mutations des résidus 104 et 238 permettent de remodeler le site actif afin qu'il puisse accueillir la chaîne latérale R volumineuse des céphalosporines de 3ème génération[102]. La mutation M182T n'intervient pas directement dans la catalyse mais complémente la perte de stabilité introduite par les autres mutations[103] (cf. points III.2.C et IV).

Les trois mutations E104K, M182T, G238S ont également été décrites au niveau d'un second mutant sélectionné *in vitro* et présentant une résistance à la céfotaxime 20.000 fois supérieure à celle de l'enzyme sauvage de départ[98]. Une banque de TEM-1 β-lactamases présentant un haut taux de mutations (≈ 27 mutations/gène) a été obtenue par une technique de PCR mutagénisante utilisant des analogues de nucléosides triphosphates[104]. A noter que la robustesse aux mutations de la β-lactamase a permis à 2,5 % des clones d'encore pousser sur une concentration en ampicilline de 1 mg/ml malgré une fréquence de randomisation élevée. Trois tours de mutagenèse alternant avec des sélections sur des concentrations croissantes en céfotaxime ont permis d'isoler un clone présentant une efficience catalytique 2328 fois supérieure à celle de la TEM-1 sauvage. Un tel gain d'activité nécessite une action synergique des 3 mutations précédemment citées et n'a pu être obtenu en seulement trois tours de sélection que grâce à un taux de mutagenèse important. Une "hypermutagenèse" *in vitro* permet en effet d'explorer des mutations multiples possédant des effets compensatoires, additifs ou coopératifs sans passer par une sélection phénotypique des simples mutants intermédiaires[98]. On évite ainsi le risque de perdre des mutations ne devenant phénotypiquement intéressantes qu'une fois combinées à d'autres substitutions (e.g. M182T).

Tawfik et son équipe ont étudié l'avantage conféré par la mutation M182T en terme de potentiel d'évolution de l'activité céphotaximase[105]. Les niveaux de résistance à la céfotaxime atteints par la banque dérivée du mutant TEM-1-M182T étaient ≈ 2,5 fois plus élevés que ceux obtenus avec la banque dérivée de l'enzyme TEM-1 sauvage. Les niveaux de résistance les plus hauts (≥ 5 µg/ml) ont à nouveau été enregistrés pour le triple mutant E104K, M182T, G238S. Le rôle joué par la

mutation M182T dans le potentiel d'évolution des TEM-1 β-lactamases a été précisé par diverses études de stabilisation de l'enzyme sauvage[105–108]. Ces études sont abordées dans le point suivant.

III.2.C. Ingénierie d'une stabilité

La TEM-1 β-lactamase a servi d'enzyme modèle pour la compréhension et la mise au point de techniques d'ingénierie de la stabilité protéique.

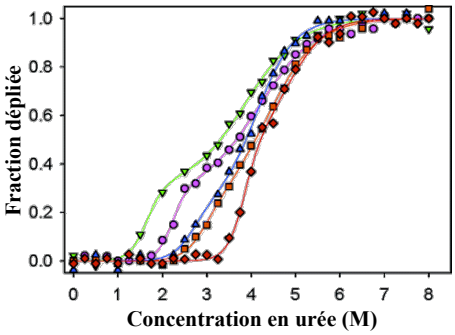
III.2.C.a Perturbation structurale et compensation par évolution dirigée

Deux de ces techniques sont basées sur une étape de déstabilisation de l'enzyme de départ suivie d'une sélection de mutations permettant de complémenter la perte de stabilité et de restaurer l'activité catalytique de départ[106,107]. Dans la première étude, la déstabilisation de la TEM-1 β-lactamase a été induite par une permutation circulaire de l'enzyme[107]. Pour ce faire, de nouvelles extrémités N- et C- terminales ont été introduites en fusionnant les anciennes avec un peptide GGS et en clivant l'enzyme au niveau d'une des charnières (position 216). La résistance à l'ampicilline conférée par l'enzyme mutante était 160 fois inférieure à celle de l'enzyme non permutée. Plusieurs cycles d'évolution dirigée associant mutagenèse aléatoire et DNA shuffling ont permis d'isoler une série de clones capables de pousser sur des concentrations en ampicilline de 600 µg/ml. Le meilleur mutant présentant une efficience catalytique équivalente à 8 % de celle de l'enzyme sauvage possédait les mutations suivantes : V31A, I47V, L51F, D115E, H153R, I208M, P219S, N276D, A280T. Bien qu'il soit difficile de déterminer le rôle précis joué par chaque mutation, certaines (H153R, I208M) sont connues pour leur effet stabilisant[108]. D'autres mutations dont l'effet stabilisant est avéré ou qui sont mentionnées dans d'autres études d'ingénierie de stabilité[105,106,108] sont également apparues dans d'autres mutants : E104A, R120G, M182T, A224V, R275L et K288R.

Dans la seconde étude, la déstabilisation de la TEM-1 a été induite par une délétion de cinq acides aminés au niveau de son extrémité N-terminale[106]. Une perte de stabilité de 10 kJ/mol a été enregistrée pour le mutant délété (NΔ5) par rapport à l'enzyme sauvage. Le K_m de l'enzyme n'a pas été affecté par cette perturbation structurale mais son k_{cat} était 6 fois moins élevé. Trois tours de sélection combinant mutagenèse aléatoire, DNA shuffling et sélection métabolique sur ampicilline ont

permis d'identifier des mutations restaurant l'activité sauvage. Les substitutions prévalant au 3ème tour de sélection étaient les suivantes : I84V, H153R, M182T, T195S et A224V. Les mutations V31A, R120G et I208M sont également apparues chez certains clones. Un clone unique a présenté jusqu'à 12 substitutions simultanées, preuve à nouveau de la grande tolérance aux mutations de la TEM-1 β-lactamase. Une efficience catalytique et une stabilité thermodynamique supérieures à celles de l'enzyme sauvage ont été enregistrées pour le quadruple mutant "S3-6" (H153R, M182T, T195S, A224V) **(Figure 21)**. Les profils de dénaturation de l'enzyme sauvage et du mutant délété (NΔ5) par l'urée s'ajustent parfaitement à un modèle à trois états : N ⇌ I ⇌ U où N représente l'état natif, I un intermédiaire de type molten globule et U (pour "unfolded") l'état dénaturé de la protéine **(Figure 21.a)**. Les mutations stabilisantes sélectionnées affectent les deux transitions. La stabilité du clone S3-6 a été davantage améliorée par la "restauration" de son extrémité N-terminale (clone FL-S3/6) **(Figure 21)**. Il en ressort que les mutations compensant la perte de stabilité induite par délétion fournissent également une stabilisation accrue dans le contexte d'une enzyme non-délétée.

a.



b.

Variants	k_{cat}/K_m (M^{-1} s^{-1})	$\Delta G°_{NI,H2O}$ (kJ/mol)	$\Delta G°_{IU,H2O}$ (kJ/mol)
TEM-1 wt	11.5 x 10^6	30.4 ± 4.5	15.8 ± 1.1
NΔ5	2.02 x 10^6	20.2 ± 5.5	15.3 ± 1.7
NΔ5-S3/6	12.4 x 10^6	26.2 ± 10.1	23.4 ± 3.2
FL-S3/6	8.97 x 10^6	45.8 ± 7.7	27.2 ± 7.0

Figure 21 : (a.) Dénaturation induite par l'urée de différents variants β-lactamases. La dénaturation a été suivie en mesurant le déplacement du maximum d'émission de la fluorescence intrinsèque vers les rouges en fonction de la concentration en urée. Les données ont été ajustées assumant un mécanisme de dénaturation à trois états (Natif (N), Intermédiaire (I), Non-foldé (U)). (vert ▼) : clone NΔ5 ; (magenta ●) : TEM-1 wt ; (bleu ▲) : clone NΔ5-S3/6 ; (rouge ◆) : clone FL-S3/6 ; (orange ■) : clone NΔ5-S3/7, variant non mentionné dans le texte. (b.) Paramètres cinétiques et thermodynamiques des différents variants. Mesures réalisées à 25°C dans tampon phosphate 50 mM, pH 7. $\Delta G°_{NI,H2O}$ est la différence d'énergie libre de Gibbs entre l'intermédiaire de repliemement et la protéine native à une concentration nulle en dénaturant. $\Delta G°_{IU,H2O}$ est la différence d'énergie libre de Gibbs entre la protéine dénaturée et l'intermédiaire de repliement à une concentration nulle en dénaturant. Figure et données de Hecky & Müller[106]

III.2.C.b. Stabilisation in vitro par la méthode "Proside"

Durant l'évolution, les mutations stabilisantes ne sont en général directement maintenues que si une pression de sélection pour la stabilité protéique existe. La méthode Proside est une technique de sélection *in vitro* appliquant une telle pression de sélection en liant l'infectivité des phages filamenteux à la résistance aux protéases d'une protéine qui est elle-même dépendante de sa stabilité[109] **(Figure 22)**. Lors de l'infection, la reconnaissance des bactéries hôtes est médiée par la partie N-terminale de la p3, protéine de la capside phagique ancrée par son extrémité C-terminale au niveau de la particule virale. Une collection de mutants de la protéine cible est exprimée entre les parties N- et C-terminales de la p3 avant d'être incubée

en présence de protéases. Le lien physique entre les domaines, et par conséquent, le pouvoir infectieux du phage n'est maintenu qu'à condition que la protéine insérée soit bien repliée et résistante aux protéases **(Figure 22)**.

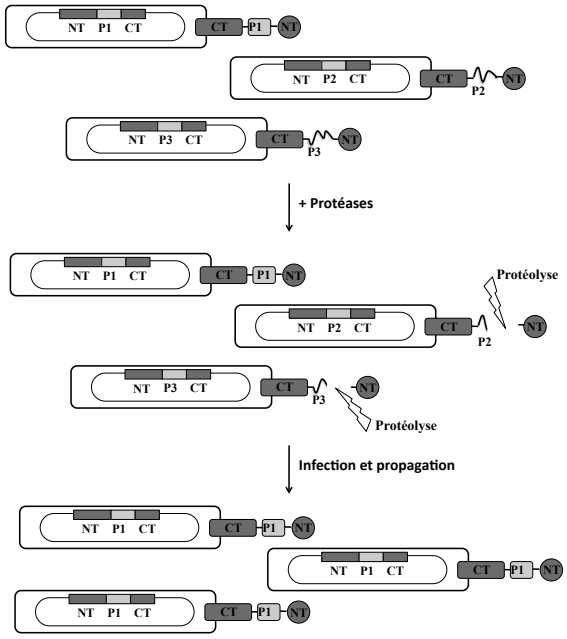


Figure 22 : Représentation schématique de la technique Proside. Voir explication dans le texte. NT = Domaine N-terminal de la p3 ; P = Protéine cible ; CT = Domaine C-terminal de la p3.

Les clones possédant une stabilité conformationnelle et donc une résistance accrue aux protéases peuvent ainsi être amplifiés et enrichis au cours de plusieurs tours de sélection *in vitro*. Une collection de TEM-1 β-lactamases randomisées par *error prone PCR* a ainsi été soumise à une dizaine de tours de sélection de type Proside en présence de chymotrypsine (0,25 µM) à 40-45°C [108]. La majorité des clones isolés présentait soit la mutation M182T soit la mutation A224V, cette dernière étant parfois accompagnée de la mutation I208M. Ces trois substitutions augmentent respectivement le T_m de l'enzyme de 5.7°C, 3.1°C et 1°C. Dans le but de sélectionner des mutations davantage stabilisantes, une TEM-1 β-lactamase déstabilisée par la suppression du pont disulfure connectant les résidus Cys77 et Cys123 a été utilisée comme protéine modèle. Tous les clones isolés après 11 tours de sélection présentaient la mutation M182T associée à d'autres substitutions dont

R275L, L201P, V80I et E147G. Les 9 meilleures substitutions (P62S, V80I, E147G, M182T, L201P, I208M, A224V, I247V et R275L) en terme de stabilité apportée ont été simultanément introduites au niveau de la TEM-1 β-lactamase. Le T_m du mutant ainsi créé est augmenté de 18.4°C ce qui en fait le variant le plus stable connu de la TEM-1 β-lactamase. Son activité catalytique envers la nitrocéfine est également supérieure à celle de l'enzyme sauvage. A noter que toutes les mutations sélectionnées dans cette étude stabilisent l'état natif de la protéine par rapport à l'état intermédiaire de type molten globule déplaçant ainsi le "midpoint" de la première transition vers des concentrations en dénaturant plus élevées.

III.2.C.c. Caractérisation des mutations stabilisantes

Un diagramme de Venn reprend les principales substitutions mentionnées dans les trois études précédentes **(Figure 23)**. Quelques unes de ces mutations sont davantage caractérisées dans ce chapitre. Les paramètres catalytiques et biophysiques des différents variants ont été compilés à partir des études de Bershtein *et al.*[105] et de Kather *et al.*[108] **(Table 2)**. Les interprétations structurales des effets stabilisants apportés par les mutations individuelles ont été proposées par Kather *et al.*[108].

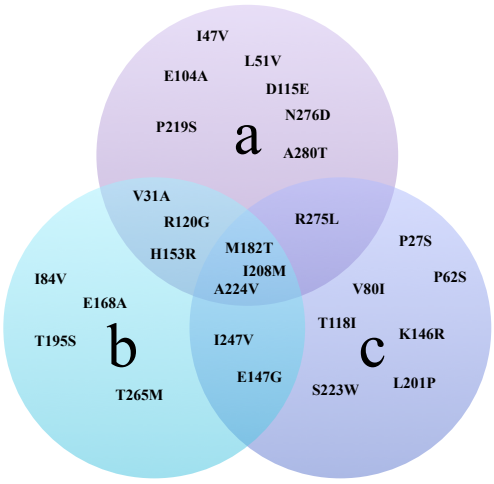


Figure 23 : Diagramme de Venn reprenant les principales substitutions mentionnées dans : a) L'étude de Osuna *et al.*[107] portant sur la stabilisation d'une TEM-1 β-lactamase permutée circulairement. b) L'étude de Hecky & Müller[106] portant sur la stabilisation d'une TEM-1 β-lactamase tronquée en extrémité N-terminale. c) L'étude de Kather *et al.*[108] portant sur la stabilisation de la TEM-1 β-lactamase par la méthode 'Proside'.

H153R : La mutation H153R est une substitution "back-to-consensus" stabilisant l'état intermédiaire molten globule et affectant ainsi la stabilité cinétique de la TEM-1 β-lactamase. Le T_m de l'enzyme est augmenté de 3,3°C et son efficience catalytique est légèrement améliorée.

M182T (Figure 24.a): La mutation M182T stabilise l'enzyme à hauteur de 9,0 kJ/mol et est régulièrement retrouvée au niveau de variants naturels de la TEM-1 β-lactamase actifs envers les céphalosporines de 3ème génération. Kather *et al.* [108] ont postulé que au contraire de la méthionine la thréonine serait capable de jouer le rôle d'un résidu N-cap au niveau de l'hélice α6. L'O^{γ} participerait à un pont hydrogène avec le NH de l'alanine 185.

L201P (Figure 24.b) : La mutation L201P apporte une stabilisation à la β-lactamase de 1,9 kJ/mol. La leucine occupe la première position de l'hélice α7 et présente des angles Φ et Ψ de respectivement de -55° et -47°. Cette position et la conformation du "backbone" sont idéales pour une proline. La mutation en proline devrait ainsi réduire le coût entropique lors du repliement.

I208M (Figure 24.c) : Ile208 occupe la base d'un sillon délimité par le résidu Leu194 d'un côté et les résidus Pro257 et Phe230 de l'autre. Ile208 ne peut interagir qu'avec Leu194 alors que Met208, possédant une chaîne latérale plus large, connecte aussi les résidus Pro257 et Phe230. La stabilisation de 1,7 kJ/mol viendrait d'un meilleur "packing" moléculaire.

A224V (Figure 24.d) : Le groupement méthyl de l'Ala224 pointe vers une poche délimitée par les chaînes latérales des résidus Ala280, Ala284 et Ile287 de l'hélice α8. La chaîne latérale de la valine remplirait mieux l'espace créé par ces différents résidus et le packing de l'enzyme en serait amélioré. Le gain de stabilité est estimé à 5,0 kJ/mol.

R275L (Figure 24.e) : L'effet stabilisant de la mutation R275L proviendrait du remplacement du groupement chargé guanidine partiellement enfoui et proche de l'extrémité N-terminale de l'hélice α8. La substitution par une leucine et sa chaîne latérale hydrophobe augmente fortement la stabilité protéique (7,9 kJ/mol).

Mutants	k_{cat} (s^{-1})	K_m (x 10^{-6} M)	k_{cat}/K_m (x 10^{-6} M^{-1} s^{-1})	T_m (°C)	$\Delta G_{D(50°C)}$ (kJ mol^{-1})
wt	199 ± 2	45.5 ± 1.4	4.4	50.8 (54.2)	1.4
R120G	226 ± 3	30.7 ± 1	7.4	(56)	/
E147G	248 ± 0.3	50.8 ± 1.2	4.9	53.2 (56.8)	6.0
H153R	222 ± 2.1	38.3 ± 1.3	5.8	(57.5)	/
M182T	169 ± 5	35.8 ± 0.9	4.7	56.5 (59.2)	10.4
T195S	/	/	/	/	/
L201P	214 ± 3.2	38.7 ± 1.8	5.5	52.2 (55.6)	3.3
I208M	/	/	/	51.8	3.1
A224V	/	/	/	53.9	6.4
R275L	/	/	/	55.8	9.3

Table 2 : Paramètres catalytiques et biophysiques de différents mutants de la TEM-1 β-lactamase. Les données catalytiques ainsi que les valeurs de T_m entre parenthèses proviennent de Bershtein, S. *et al.*[105]. Les autres données sont tirées de Kather *et al.*[108]. $\Delta G_{D(50°C)}$ correspond à l'énergie libre de dénaturation à 50°C. Les lecteurs sont invités à consulter les fichiers sources pour toutes informations complémentaires.

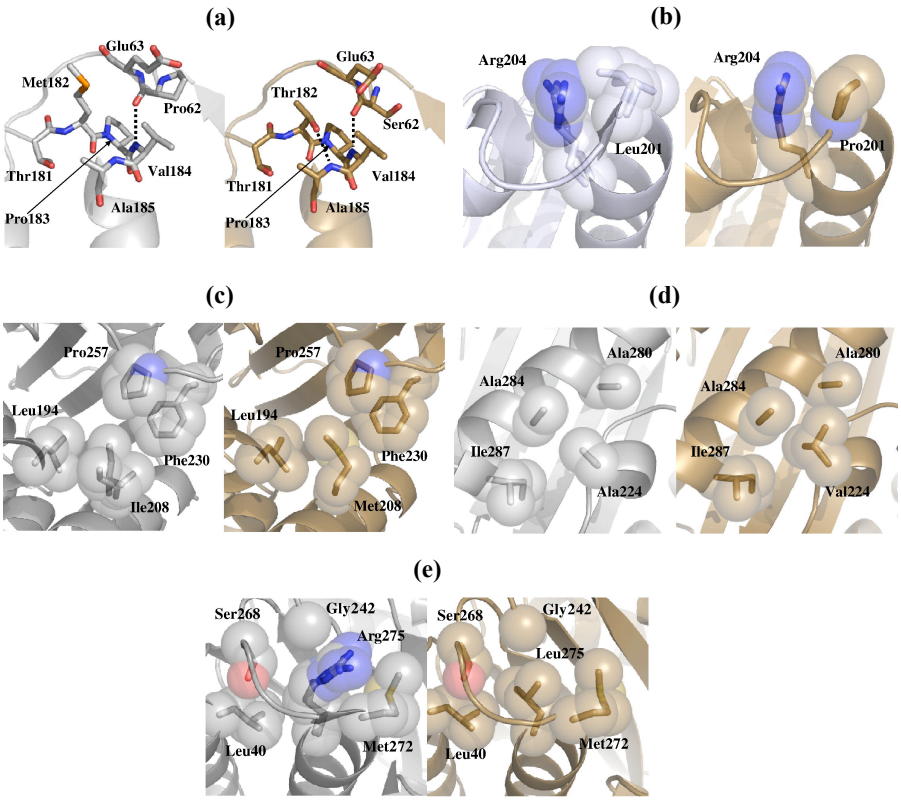


Figure 24 : Comparaisons entre les structures locales de la TEM-1 sauvage et de différents variants à proximité des mutation stabilisantes M182T (a), L201P (b), I208M (c), A224V (d) et R275L (e). Figure de Kather *et al.*[108]

IV. Stabilité et potentiel d'évolution

Les mutations apportent la diversité nécessaire aux phénomènes d'évolution naturelle et dirigée. Le potentiel d'évolution d'une protéine est d'ailleurs défini comme sa capacité à acquérir de nouvelles fonctions ou caractéristiques en réponse à l'introduction de mutations au sein de sa séquence et à une pression de sélection. L'intégrité structurale et fonctionnelle d'une protéine étant largement dépendante de sa stabilité thermodynamique, on comprend vite que cette dernière puisse jouer un rôle prépondérant dans le concept de potentiel d'évolution. Cette partie de l'introduction tentera de faire le lien entre la stabilité protéique, la tolérance aux mutations et le potentiel d'évolution.

IV.1. Mutations et stabilité protéique

La stabilité thermodynamique d'une protéine est caractérisée par la différence d'énergie libre de Gibbs ($\Delta G°$) entre ses états natif (N) et dénaturé (D) : $\Delta G° = G°_N - G°_D$. Plus la valeur est négative, plus la protéine est dite stable et plus son ratio natif/dénaturé est élevé. Contrairement à l'idée que l'on pourrait se faire, les protéines ne sont que marginalement stables avec des $\Delta G°$ variant de -20,9 à -62,8 kJ/mol. Cette stabilité marginale est en partie expliquée par le fait que la sélection naturelle cible davantage la fonction biochimique d'une protéine plutôt que sa stabilité[110]. Seule la stabilité minimale requise pour le repliement correct de la protéine et donc pour sa fonctionnalité est contrainte par l'évolution[110]. De surcroît, il s'avère que la majorité des mutations ont un effet déstabilisant sur les protéines **(Figure 25)**. Le changement de stabilité provoqué par une substitution ($\Delta\Delta G°$) est défini comme la différence de stabilité entre le mutant (mut) et la protéine sauvage (wt) [111] : $\Delta\Delta G° = \Delta G_f^{°mut} - \Delta G_f^{°wt}$. Selon les estimations fournies par l'algorithme FoldX[112], ≈ 70 % des substitutions seraient déstabilisantes ($\Delta\Delta G° > 0$ kJ/mol), ≈ 20 % seraient significativement déstabilisantes ($\Delta\Delta G° > 8,4$ kJ/mol) et seulement une minorité (5 %) seraient significativement stabilisantes ($\Delta\Delta G° < -4,2$ kJ/mol)[113,114]. Les substitutions restantes seraient neutres en terme d'effet sur la stabilité protéique. La valeur moyenne de $\Delta\Delta G°$ tournerait autour de 3,8 kJ/mol et les effets de deux substitutions sur la stabilité protéique seraient généralement additifs[111]. Il est par ailleurs souvent observé que les substitutions menant à de

nouvelles fonctionnalités en modifiant le site actif sont davantage déstabilisantes[110,114,115]. Cela s'explique par le fait que les résidus catalytiques, souvent chargés ou polaires, sont généralement confinés dans une cavité réduite à l'écart de l'eau et que cette organisation est thermodynamiquement défavorable. La perte d'activité de par la mutation d'un résidu catalytique est de ce fait souvent accompagnée d'un gain de stabilité[116,117]. En revanche, les substitutions promouvant une nouvelle fonction sans pour autant agir au niveau du site actif ne semblent pas plus déstabilisantes que toute autre mutation **(Figure 25)**[114].

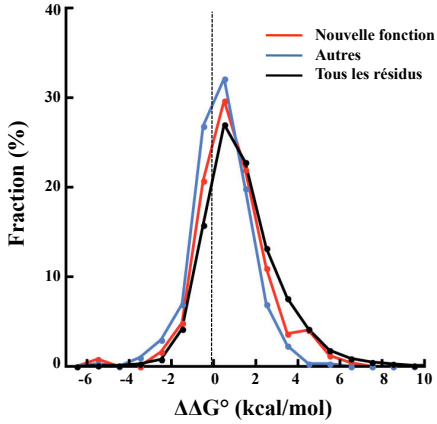


Figure 25 : Distribution des ΔΔG° de diverses mutations calculées par FoldX pour 22 enzymes. "Nouvelle fonction" fait référence aux substitutions menant à de nouvelles fonctions protéiques. Le terme "Autres" regroupe les mutations observées au niveau des ces différentes enzymes mais ne menant pas à de nouvelles fonctions. La catégorie "Tous les résidus" prend en compte toutes les substitutions possibles au sein de ces enzymes par la mutation d'un seul nucléotide. Figure de Tokuriki *et al.*[114]

A noter que d'autres paramètres peuvent également entrer en compte pour expliquer la stabilité marginale des protéines. Une trop grande stabilité pourrait nuire à l'activité enzymatique en rendant l'enzyme trop rigide[113,115]. L'étude structurale et enzymatique de protéines extrêmophiles (psychrophiles et thermophiles) et de leurs homologues mésophiles a en effet mis en évidence une certaine relation entre stabilité, flexibilité et activité[118–120]. Ainsi, les enzymes thermophiles présentent généralement une grande stabilité afin de maintenir leur architecture intacte à de hautes températures (> 50°C). Il en résulterait une importante rigidité protéique menant à de faibles activités enzymatiques à température ambiante. Inversement, les enzymes psychrophiles doivent compenser les plus faibles vitesses des réactions

chimiques observées aux basses températures (< 15°C) par des activités spécifiques (en terme de k_{cat}) souvent élevées dans cette zone de température. Ces enzymes semblent privilégier une flexibilité structurale favorable à la catalyse au détriment de la stabilité thermodynamique. La stabilisation d'une α-amylase psychrophile par substitution d'un ou plusieurs acides aminés (chacun introduisant une ou des interactions faibles mises en évidence au sein d'une enzyme mésophile homologue) induit une diminution du k_{cat} de l'enzyme et conforte cette notion de relation entre stabilité, flexibilité et activité protéiques[121]. A noter qu'une trop grande stabilité et rigidité pourrait également affecter le potentiel d'évolution des protéines en restreignant le nombre de conformations alternatives pouvant mener à de nouvelles fonctions[115,113]. De plus, une protéine trop stable se soustrairait de façon prononcée aux activités protéasiques de la cellule rendant sa régulation plus délicate[115].

IV.2. Stabilité protéique et robustesse

On l'a vu précédemment, la TEM-1 β-lactamase ainsi que la grande majorité des protéines sont relativement tolérantes face aux mutations. Plusieurs études se sont penchées sur l'origine de cette tolérance et ont tenté de déchiffrer les propriétés protéiques sous-jacentes à cette robustesse. Le concept de m-neutralité proposé par Bloom est une mesure fondamentale de cette robustesse[111]. Il est défini pour une protéine donnée comme la fraction $P_f(m)$ de variants conservant la structure native parmi toutes les séquences divergeant de "m" substitutions par rapport à la séquence sauvage[111] :

$$P_f(m) = \int_{-\infty}^{-\Delta G_f^{\circ extra}} p_m(\Delta\Delta G^{\circ m}) d(\Delta\Delta G^{\circ m})$$

Le terme $\Delta G_f^{\circ extra}$ définit l'excédent de stabilité d'une protéine par rapport au seuil minimal de stabilité en dessus duquel la protéine considérée ne conserve pas sa structure native. Le terme $P_m(\Delta\Delta G^{\circ m})$ est la probabilité que "m" substitutions induisent un changement de stabilité de $\Delta\Delta G^{\circ m}$. Par simulation sur des modèles

protéiques simplifiés de type lattice*, il a été déterminé que l'importance du déclin exponentiel de la m-neutralité en fonction du nombre de substitutions était dépendante de la structure protéique. Toutes structures confondues, certaines représentent en effet l'état énergétique le plus bas pour davantage de séquences et ont de ce fait une flexibilité séquentielle plus importante[122]. Cependant, toute protéine pourrait retarder le déclin de cette m-neutralité en augmentant sa stabilité thermodynamique. L'excédent de stabilité gagné permettrait en quelque sorte à la protéine d'encaisser les premières substitutions avant d'inévitablement amorcer la chute de m-neutralité. Ces prédictions ont été validées en comparant les fractions de protéines fonctionnelles (par criblage sur ampicilline) dans des banques *error prone PCR* de la TEM-1 β-lactamase sauvage et du variant stabilisé M182T **(Figure 26)**. Les valeurs obtenues se superposent parfaitement aux prédictions théoriques se basant sur des $\Delta\Delta G^{\circ}$ estimés à partir des algorithmes POPMUSIC[123] et FOLDEF[124] et sur un $\Delta G_f^{\circ extra}$ pour le mutant M182T supérieur de 11,3 kJ/mol à l'enzyme sauvage **(Figure 26)**.

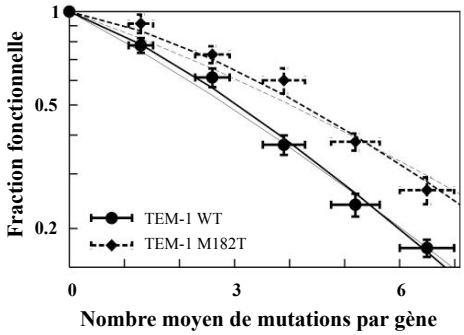


Figure 26 : Tolérances aux mutations des TEM-1 β-lactamases sauvage (WT) et mutante (M182T). Les lignes épaisses montrent les prédictions obtenues avec POPMUSIC[123] et les lignes fines montrent celles obtenues avec FOLDEF[124]. Figure de Bloom *et al.*[111]

---

* *Protéines lattice : modélisation informatique simplifiée des protéines, généralement utilisée pour des études de repliement. Dans le modèle le plus simple, dit HP, les acides aminés y sont représentés par des sphères hydrophobes ou hydrophiles connectées par des traits (liaisons peptidiques) et s'agençant en un réseau (lattice) bi ou tridimensionnel de forme généralement cubique. Les énergies d'interaction entre les différents résidus sont spécifiées par une série de conditions.*

Le mutant M182T de par son excédent de stabilité est davantage robuste face aux premières mutations. Ce gain en robustesse donne lieu à un phénomène d'épistasie* négative s'exprimant par un effet délétère des mutations sur le fitness (ou l'activité) de la protéine plus grand que ce que prévoit la somme de leurs effets individuels[113,125]. Les premières mutations n'ont qu'un effet limité sur le fitness de la protéine mais une fois le seuil de tolérance outrepassé, celle-ci décroit rapidement **(Figure 27)**. Cette épistasie est d'autant plus marquée que le seuil de robustesse est élevé[125] **(Figure 27)**.

Les résultats précédents font référence à une robustesse envers des substitutions ponctuelles mais s'appliquent probablement aussi à la mutation insertionnelle. Il a notamment été démontré au niveau d'une protéine fluorescente (eCGP123) qu'un gain de stabilité thermodynamique permettait à la protéine de tolérer davantage des insertions de peptides (20 résidus) au niveau de trois boucles de surface[126].

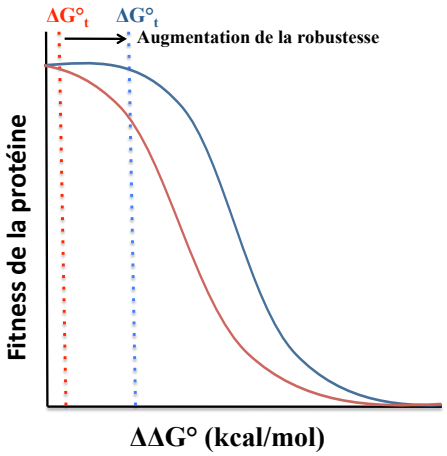


Figure 27 : Allure sigmoïdale typique de la chute de fitness (ou d'activité) d'une protéine en fonction de la perte de stabilité (ΔΔG°) engendrée par l'accumulation de mutations. La perte de fitness est minime tant que la stabilité de la protéine reste au-dessus d'un certain seuil (ΔG°t ; courbe rouge). Une augmentation de ce seuil résulte en une plus grande robustesse face aux mutations (courbe bleue). Une fois l'excédent de stabilité consommé (seuil de robustesse dépassé), la chute de fitness s'intensifie. La présence de ce seuil est responsable du phénomène d'épistasie négative observé : l'effet combiné des mutations est supérieur à ce qui est attendu de leur effet individuel. Figure de Tokuriki & Tawfik[113]

---

* *Epistasie : Toute interaction non-additive entre deux ou plusieurs mutations de sorte que leur effet combiné sur la fitness d'une protéine donnée s'écarte de la somme de leurs effets individuels.*

IV.3. Robustesse et potentiel d'évolution

IV.3.A. Un gain de stabilité favorise le potentiel d'évolution

En considérant le potentiel d'évolution d'une protéine comme sa capacité à acquérir une nouvelle fonction suite à une ou plusieurs mutations, une protéine tolérant une gamme plus étendue de mutations devrait en tout état de cause être plus à même d'évoluer. Un gain de stabilité permettant en outre d'augmenter cette tolérance, on s'attend à voir le potentiel d'évolution suivre la même tendance. Ce lien entre stabilité et pouvoir évolutif d'une protéine a été mis en évidence par l'utilisation d'un système protéique simplifié de type lattice[110]. Deux variants virtuels d'un même modèle protéique **(Figure 28.a)**, tous deux liant avec la même affinité un ligand rigide mais présentant des stabilités différentes (ΔG_f = -0.5 et -1.5), ont été obtenus par modélisation informatique. Ces deux variants ont subi une simulation d'évolution dirigée à la recherche de mutants possédant une affinité pour de nouveaux ligands plus grande que la protéine parentale tout en satisfaisant un seuil de stabilité minimale imposé. Cette contrainte requérait que les mutants simulés se replient selon la structure native avec une stabilité de $\Delta G_f \leq 0$ **(Figure 28.b)**. Un criblage virtuel de 1500 clones dans chaque banque a donné près de quatre fois plus de mutants améliorés pour la protéine stabilisée (ΔG_f = -1.5) par rapport à la seconde. Cette différence de potentiel d'évolution est expliquée par l'excès de stabilité de la première protéine permettant à davantage de mutants déstabilisés mais fonctionnellement améliorés de satisfaire le seuil minimal de stabilité imposé **(Figure 28.b)**. Ce modèle informatique a été validé expérimentalement en faisant évoluer simultanément deux variants du cytochrome P450 BM3. Les deux protéines (21B3 et 5H6) ayant la même activité initiale mais des thermostabilités différentes (T_m respectifs = 47°C et 56°C) ont été randomisées par *error prone PCR* (≈ 4,5 mutations/gène). Sans surprise, près de deux fois plus de mutants bien repliés et conservant l'activité parentale ont été obtenus à partir de la protéine stabilisée. Le potentiel d'évolution des deux protéines a été évalué par criblage de clones montrant une activité accrue sur cinq nouveaux substrats. Treize mutants améliorés ont été obtenus pour la protéine stabilisée contre seulement quatre pour la seconde. De l'analyse des différents variants sélectionnés, il en est ressorti que la perte de stabilité (T_m) ne dépassait pas 3°C pour les clones provenant de 21B3 alors que des

déstabilisations de 14°C ont pu être mesurées pour des mutants de 5H6. Seul l'excédent de stabilité de 5H6 permet de tolérer des déstabilisations aussi prononcées. Ainsi, les mutations "activantes" sélectionnées chez 5H6 ne sont généralement pas tolérées par 21B3 vu leur coût élevé en terme de stabilité.

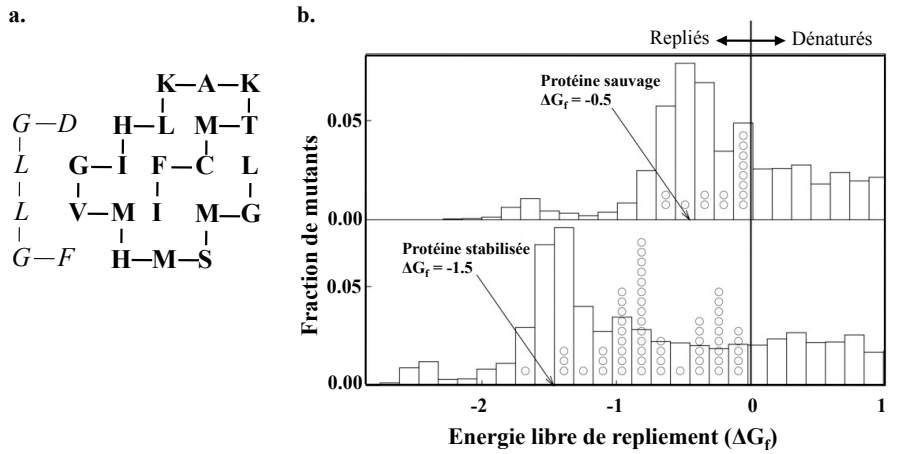


Figure 28 : Relation entre stabilité et potentiel d'évolution au niveau d'un modèle protéique de type lattice. (a.) Modèle protéique de type lattice utilisé (en gras) et son ligand (en italique) (b.) Le modèle stabilisé (cadre du dessous) a un potentiel évolutif plus grand que le modèle non-stabilisé (cadre du dessus) car il possède davantage de mutants déstabilisés mais fonctionnellement améliorés (cercles) satisfaisant le seuil minimal de stabilité imposé ($\Delta G_f = 0$). Les barres représentent la distribution des stabilités parmi l'ensemble des variants. Figure de Bloom *et al.*[110]

La Figure 29 proposée par Romero *et al.*[1] résume parfaitement la relation pouvant exister entre stabilité et potentiel d'évolution en mettant en scène les notions de seuil de stabilité, de robustesse et la nature généralement déstabilisante des mutations.

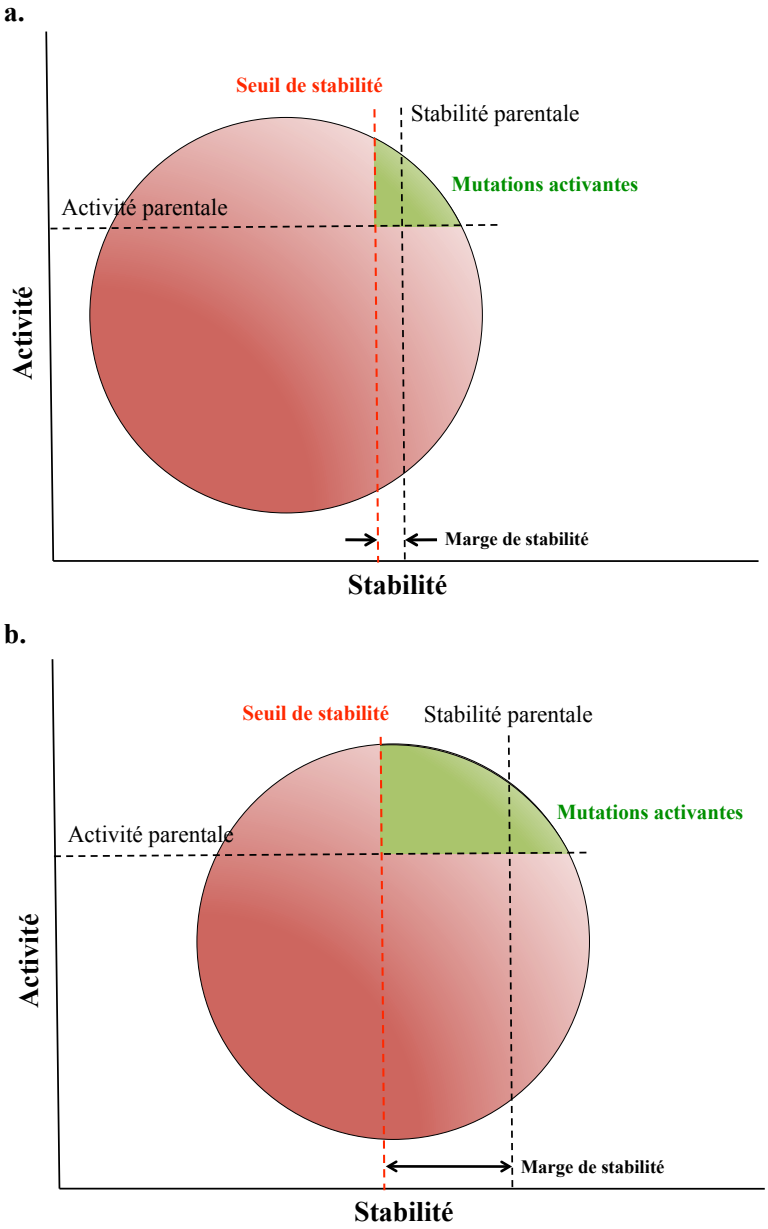


Figure 29 : Représentation schématique de la relation entre robustesse et potentiel d'évolution. (a.) Considérant que la plupart des mutations sont déstabilisantes, beaucoup de variants d'une protéine dont la stabilité est proche du seuil minimal requis seront instables et donc contre-sélectionnés (zone rouge). Réciproquement, peu de clones stables présenteront des mutations activantes (zone verte). (b.) Un gain de stabilité rend la protéine plus robuste face aux mutations, augmentant la probabilité d'obtenir des mutations activantes (zone verte plus vaste). Figure inspirée de Romero & Arnold[1]

IV.3.B Dérive neutre et stabilité

La dérive neutre est grandement responsable de la diversité génétique des êtres vivants et correspond à une modification graduelle et aléatoire des séquences géniques sous une pression de sélection maintenant les fonctions et structures protéiques existantes[105]. Tawfik et son équipe ont reproduit *in vitro* cette dérive en accéléré au niveau de la TEM-1 β-lactamase et ont observé une tendance à l'enrichissement et au maintien des résidus consensus et/ou ancestraux, en particulier ceux augmentant la stabilité de l'enzyme[105]. Le gène de la protéine a été soumis à une mutagenèse aléatoire par *error prone* PCR introduisant en moyenne deux mutations par gène et par tour de sélection, soit un taux de dérive $\approx 10^6$ fois plus élevé que *in natura*. Les clones générés ont par la suite été soumis à diverses pressions de sélection par étalement sur différentes concentrations en ampicilline : 0 µg/ml (Lib0), 12,5 µg/ml (Lib12.5) et 250 µg/ml (Lib250). Dix-huit tours de sélection ont ainsi été réalisés. De manière attendue, la fraction des mutations étant non-synonymes était inversement proportionnel à la pression de sélection appliquée : 71 % en l'absence d'ampicilline et 50 % en présence de l'antibiotique. Il en était de même pour le nombre de sites tolérant les substitutions : 84,3 % pour Lib0 et 62,4 % pour Lib250. Les positions mutées ont été classées par les auteurs en deux catégories. Le premier groupe inclut les positions pour lesquelles une diminution de fréquence de substitutions est observée au niveau de Lib12.5 et Lib250 par rapport à Lib0. Ce phénomène reflèterait une sélection négative "purgeant" les mutations délétères. Le second ensemble regroupe les positions où un enrichissement en substitutions est observé au niveau de Lib12.5 et Lib250 par rapport à Lib0, enrichissement qui serait le reflet d'une sélection positive. De manière intéressante, la majorité des mutations purgées s'éloignent de la séquence consensus et /ou ancestrale observée chez les β-lactamases de classe A **(Table 3)**. Inversement, la moitié des mutations enrichies rétablissent cette séquence consensus et/ou ancestrale **(Table 3)**.

		Purged mutations																	Enriched mutations												
Position		27	46	50	57	66	68	102	105	106	138	152	163	179	186	211	263	264	7	15	19	52	63	111	120	147	153	182	201	208	288
TEM-1 *(Klebsiella pneumoniae)*	100%	P	Y	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	H	F	F	N	E	K	R	E	H	M	L	I	K
Most frequent mutation		S	H	G	P	L	V	W	C	P	V	S	A	G	T	V	S	H	R	C	L	D	A	E	G	G	R	T	P	L	R
Pla-6 *(Raoultella planticola)*	72%	Q	Y	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	Q	L	L	A	D	K	G	G	R	T	S	Q	R
ORN-5 *(Raoultella ornithinolytica)*	71%	E	Y	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	Q	L	L	A	G	K	G	G	R	T	P	Q	R
AAB *(Klebsiella pneumoniae)*	68%	P	M	D	L	F	L	L	Y	S	L	L	D	D	M	M	I	Y	Y	L	L	A	S	K	G	G	R	T	A	L	E
SHV-1 *(Klebsiella pneumoniae)*	65%	P	M	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	Y	L	L	A	D	K	G	G	R	T	A	L	E
Len-1 *(Klebsiella pneumoniae)*	64%	P	I	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	Y	L	L	A	D	K	G	G	R	T	A	L	E
Ohio-1 *(Enterobacter cloacae)*	61%	L	M	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	Y	L	A	A	D	K	G	G	R	T	A	L	E
CKO-1 *(Citrobacter koseri)*	48%		L	D	T	F	L	L	Y	S	A	L	D	D	I	M	I	Y	K		L	A	E	K	A	A	R	T	P	M	S
CARB-5 *(Acinetobacter calcoaceticus)*	47%	D		D	W	F	L	L	Y	S	L	L	D	D	M	L	L	Y	S	F	A	E	N	K	A	A	R	T	I	E	K
HERA-2 *(Escherichia hermannii)*	45%	P	V	N	V	F	L	L	Y	A	L	M	D	D	M	M	I	F	L	L		Q	D	N	R	A	R	S	P	A	A
AST-1 *(Nocardia asteroides)*	45%		L	D	V	F	F	L	Y	S	L	L	D	D	L	L	V	L				G	D	T	R	G	R	T	A	K	D
AmpC *(Vibrio fischeri)*	44%	G	F	N	W	F	M	I	W	S	L	L	D	D	M	M	I	Y	L	N	G	A	D	P	K	S	R	T	P	K	N
blaF *(Mycobacterium fortuitum)*	43%	A	V	N	I	F	L	L	N	S	L	A	D	D	L	M	T	R	L		P	Q	D	A	A	A	R	T	P	E	P
ABPS-1 *(Burkholderia pseudomallei)*	42%	V	V	D	I	F	M	L	Y	S	I	A	D	D	M	M	V	Y			A	A	D	Q	A	A	R	T	A	N	R
L2 *(Stenotrophomonas maltophilia)*	42%	I	I	D	I	F	M	L	H	A	L	L	D	D	M	L	A	Y	F	S	V	A	D	R	R	A	R	T	P	A	R
PSE-4 *(Pseudomonas aeruginosa)*	42%	Q	V	D	W	F	M	L	Y	S	L	L	D	D	I	M	I	Y	L			Q	N	K	D	G	R	T	E	E	D
NB64 *(Xanthomonas axonopodis)*	41%	D	V	D	L	F	M	M	H	A	F	L	D	D	M	M	T	Y	A	A	A	A	H	R	G	G	R	T	P	T	K
JN56		I	V	D	L	F	M	L	Y	S	L	L	D	D	M	M	I	Y	L	L	A	E	D	K	A	A	R	T	P	T	E

Table 3 : Tableau des mutations "purgées" ou "enrichies" au cours des sélections. Différentes séquences de β-lactamases de classe A (les chiffres à droites représentent le pourcentage d'identité avec la TEM-1) ont été alignées par TCoffee. Les résidus conservés sont marqués en rouge. JN56 est l'ancêtre inféré du sous-groupe de β-lactamases de classe A auquel TEM-1 appartient. "Most frequent mutation" fait référence aux mutations les plus fréquemment observées dans Lib0 (pour les mutations purgées), ou dans lib250 (pour les mutations enrichies). Tableau tiré de Bershtein *et al.*[105]

La nature des mutations enrichies suggère qu'elles permettent à l'enzyme de tolérer, par une stabilité protéique accrue, le haut taux de mutagenèse imposé. Sept de ces substitutions ont en effet déjà été mentionnées dans des sélections *in vitro* pour stabilité ou sont connues pour stabiliser la TEM-1 : R120G, E147G, H153R, M182T, L201P, I208M et K288R. Les mesures des paramètres biophysiques des simples mutants ont confirmé le gain de stabilité apporté par ces différentes substitutions **(Table 2, p.53)**. Les mutations influencent soit la stabilité thermodynamique (M182T), soit la stabilité cinétique (R120G, H153R) ou les deux (E147G, L201P, I208M et K288R). La nature stabilisante des substitutions enrichies a également été appuyée par leur capacité à compenser l'effet délétère de certaines mutations. Ainsi, 1% des clones sélectionnés au tour 5 à partir de la banque Lib250 étaient capables de contrebalancer l'effet déstabilisant imposé par l'introduction de la mutation L76N. Ce pourcentage était inférieur à 0,1 pour les clones provenant de la fraction active de la collection Lib0. Les gènes "compensés" présentaient exactement les mêmes mutations consensus observées lors de la dérive neutre, la majorité des clones (91 %) portant la mutation M182T. Du fait de ces effets stabilisants et "compensatoires", Tawfik *et al.*[105] suggèrent que les gènes portant des

mutations rétablissant la séquence consensus seraient un meilleur point de départ pour des études d'évolution dirigée.

L'effet compensatoire ou suppresseur de la mutation M182T a également été confirmé dans une seconde étude[103,127]. L'insertion de la mutation L76N altère le repliement et la stabilité de la TEM-1 β-lactamase augmentant à tel point sa sensibilité aux protéases que sa concentration périplasmique ne représente plus que 20 % de celle de l'enzyme sauvage. L'activité β-lactamase du mutant est également diminuée de 70 % par rapport à l'enzyme sauvage. La mutation M182T n'a, à elle seule, aucun impact sur l'activité enzymatique. Néanmoins, combinée à L76N, elle permet de rétablir l'efficience catalytique à 53 % de celle de la TEM-1 sauvage. La concentration périplasmique est également restaurée à des niveaux proches de la protéine non mutante. A la différence de l'étude précédente, les auteurs suggèrent que la substitution M182T n'augmente pas la stabilité thermodynamique de l'enzyme mais agit plutôt en tant que suppresseur de phénomènes d'agrégation et de mauvais repliement.

IV.3.C. Stabilité et contraintes évolutives

La relation établie entre le potentiel d'évolution et la stabilité protéique a mené à un modèle théorique décrivant une marche évolutive des protéines contrainte par la stabilité. Par mutations successives, la protéine explorerait l'espace de séquences selon une trajectoire définie par la pression de sélection et canalisée de telle sorte à ce que la stabilité reste en dedans ou proche d'une zone neutre en terme de fitness **(Figure 30)**[115]. En d'autres termes, la stabilité protéique et l'acquisition de nouvelles fonctions s'arbitreraient (trade-off) au cours de l'évolution[114,128]. Ainsi, les mutations fonctionnellement bénéfiques mais altérant la stabilité au delà de la zone de neutralité sont généralement purgées au cours de l'évolution. Sous certaines conditions, ces mutations "fonctionnelles" peuvent être maintenues par l'acquisition subséquente d'une mutation stabilisante compensatoire rétablissant le fitness (divergence uphill) **(Figure 30)**. Cette nécessité de compensation restreint la trajectoire évolutive et ralentit le taux d'adaptation[113]. De manière réciproque, l'acquisition d'une mutation compensatoire avant ou parallèlement à la divergence d'une nouvelle fonction peut grandement améliorer le potentiel d'évolution en

ouvrant des trajectoires adaptatives autrement non-accessibles (divergence downhill)[113] **(Figure 30)**. A noter que les chaperonnes peuvent également augmenter le potentiel d'évolution d'une protéine en étendant la zone neutrale de stabilité[113]. Le modèle théorique permet également d'expliquer le lien existant entre la stabilité et le phénomène d'épistasie observé, où l'effet d'une mutation est conditionné par la stabilité de la protéine parentale[1] (**Figure 30**).

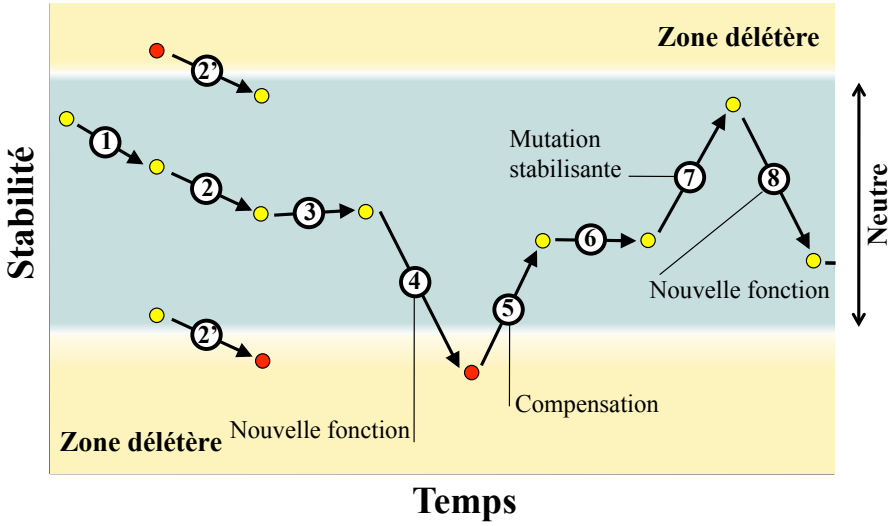


Figure 30 : Marche évolutive d'une protéine contrainte par sa stabilité. La zone bleue représente une gamme de stabilités neutres en terme de fitness. Les zones jaunes sont des régions "délétères" : chutes de fitness liées à une stabilité trop faible ou trop importante. Les cercles de couleurs représentent différentes séquences protéiques (jaunes : séquences neutres en terme de stabilité ; rouge : séquences délétères). Les flèches (mutations non silencieuses) numérotées de (1) à (8) représentent un parcours théorique à travers le paysage de séquences. Les mutations (1) à (3) représentent une dérive neutre. Les mutations (2') mettent en évidence le phénomène d'épistasie (positive en haut et négative en bas) en terme de fitness, où une même mutation peut être bénéfique ou délétère en fonction de la stabilité de la protéine parentale. L'acquisition d'une nouvelle fonction est souvent déstabilisante (4) et induit une pression de sélection forte pour la sélection d'une mutation stabilisante (5) compensant la perte de stabilité = divergence "uphill". La fixation d'une mutation stabilisante (7) avant ou en parallèle avec une substitution "fonctionnelle" (8) peut augmenter grandement le potentiel d'évolution de la protéine = divergence downhill. Figure de DePristo *et al.*[115] modifiée selon Tokuriki *et al.*[113]

D'autres études font état du phénomène de compromis (trade-off) évolutif entre stabilité et acquisition de nouvelles fonctions[128,129]. Un bel exemple provient de l'analyse des propriétés physicochimiques des intermédiaires évolutifs sélectionnés

au cours de l'évolution dirigée d'une cytochrome P450 hydroxylase en cytochrome P450 propane monooxygénase[1,130]. L'augmentation de l'activité sur le propane au cours des différents tours de sélection est accompagnée d'une perte de thermostabilité (T_m) de la protéine **(Figure 31)**. La perte de stabilité est telle que l'incorporation d'une mutation stabilisante mais fonctionnellement neutre (voir légèrement délétère) est nécessaire à la 9ème génération **(Figure 31)**. Cette mutation permet à la protéine de sortir d'une "impasse évolutive" et d'explorer de nouvelles routes à travers l'espace de séquences.

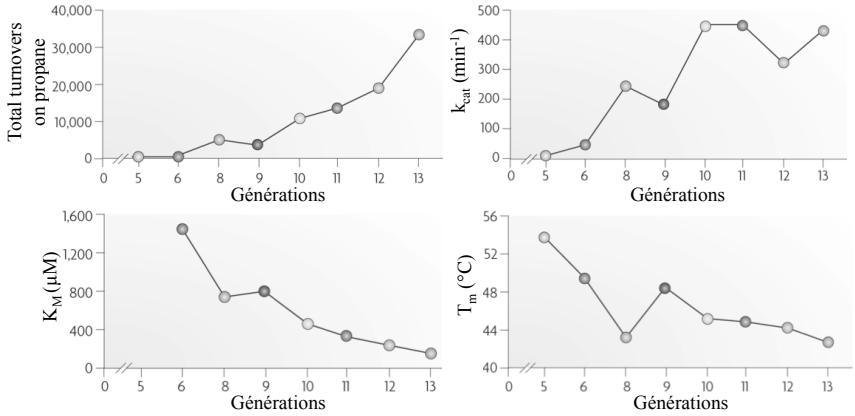


Figure 31 : Propriétés physicochimiques des intermédiaires évolutifs sélectionnés au cours de l'évolution dirigée d'une cytochrome P450 hydroxylase en cytochrome P450 propane monooxygénase. Figure de Fasan *et al.*[129]

L'évolution de l'activité de certaines TEM β-lactamases envers les céphalosporines de 3ème générations illustrerait aussi une évolution de type "trade-off" avec l'intégration de la mutation suppressive M182T permettant de compenser la perte de stabilité consécutive à l'insertion de certaines mutations activantes (e.g. : E104K, R164S, G238S)[128]. La mutation M182T est fonctionnellement neutre mais restaure le fitness et la stabilité des mutants actifs présentant un site actif élargi capable d'accommoder les chaînes latérales volumineuses des céphalosporines de 3ème génération[128]. Sur base de l'émergence tardive de la mutation M182T tant *in natura* qu' *in vitro*, cet exemple évolutif correspondrait à une divergence qualifiée de uphill **(Figure 30)** par Tawfik *et al*[113]. L'apparition tardive de la mutation M182T a été prédite dans l'étude de Weinreich *et al.*[131] Les auteurs ont calculé une probabilité d'occurrence pour chacune des 120 trajectoires mutationnelles pouvant mener au

quintuple mutant « A42G, E104K, M182T, G238S, g4205a (mutation non-codante)», 100.000 fois plus actif sur céfotaxime. Parmi les 120 trajectoires possibles, 102 ne seraient pas compatibles avec une évolution de type darwinienne. Cette incompatibilité proviendrait du fait que 4 des 5 mutations n'apportent un avantage sélectif (résistance à la céfotaxime accrue) qu'une fois combinées à certaines mutations. Cette observation est le reflet d'un phénomène d'épistasie positive. Des 18 trajectoires restantes, la plus probable verrait les mutations se succéder selon l'ordre suivant : G238S, E104K, A42G, M182T et g4205a. Cet ordre de succession, avec notamment une arrivée tardive de M182T, serait également expliqué par l'épistasie. Seule, la mutation M182T est neutre voir légèrement délétère pour l'enzyme. Son effet bénéfique (dans ce cas-ci, augmentation de la résistance à la céfotaxime) n'apparaît que grâce à son action stabilisatrice lui permettant de compenser les pertes de stabilité induites par les mutations G238S et/ou E104K et/ou A42G.

V. Références

1. Romero, P.A. & Arnold, F.H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866-876 (2009).
2. Formaneck, M.S., Ma, L. & Cui, Q. Reconciling the « old » and « new » views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* **63**, 846-867 (2006).
3. Janssen, L.H., Van Wilgenburg, M.T. & Wilting, J. Human serum albumin as an allosteric two-state protein. Evidence from effects of calcium and warfarin on proton binding behaviour. *Biochim. Biophys. Acta* **669**, 244-250 (1981).
4. Ackers, G.K. & Holt, J.M. Asymmetric cooperativity in a symmetric tetramer: human hemoglobin. *J. Biol. Chem.* **281**, 11441-11443 (2006).
5. Soisson, S.M., MacDougall-Shackleton, B., Schleif, R. & Wolberger, C. Structural basis for ligand-regulated oligomerization of AraC. *Science* **276**, 421-425 (1997).
6. Streaker, E.D. & Beckett, D. Ligand-linked Structural Changes in the Escherichia coli Biotin Repressor: The Significance of Surface Loops for Binding and Allostery,. *J. Mol. Biol.* **292**, 619-632 (1999).
7. Chu, J.-W. & Voth, G.A. Allostery of actin filaments: molecular dynamics simulations and coarse-grained analysis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13111-13116 (2005).
8. Ascenzi, P., Bocedi, A., Bolli, A., Fasano, M., Notari, S. & Polticelli, F. Allosteric modulation of monomeric proteins. *Biochem. Mol. Biol. Educ.* **33**, 169-176 (2005).
9. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433-443 (2004).
10. Deponte, M., Sturm, N., Mittler, S., Harner, M., Mack, H. & Becker, K. Allosteric coupling of two different functional active sites in monomeric Plasmodium falciparum glyoxalase I. *J. Biol. Chem.* **282**, 28419-28430 (2007).
11. Monod, J., Wyman, J. & Changeux, J.P. On the nature of allosteric transitions : A plausible model. *J. Mol. Biol.* **12**, 88-118 (1965).
12. Süel, G.M., Lockless, S.W., Wall, M.A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59-69 (2003).
13. Koshland, D.E., Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365-385 (1966).
14. Koshland, D.E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 98-104 (1958).

15. Wlodarski, T. & Zagrovic, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19346-19351 (2009).

16. Volkman, B.F., Lipson, D., Wemmer, D.E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429-2433 (2001).

17. Panagou, D., Orr, M.D., Dunstone, J.R. & Blakley, R.L. A monomeric, allosteric enzyme with a single polypeptide chain. Ribonucleotide reductase of Lactobacillus leichmannii. *Biochemistry* **11**, 2378-2388 (1972).

18. Giardina, B., Ascenzi, P., Clementi, M.E., De Sanctis, G., Rizzi, M. & Coletta, M. Functional modulation by lactate of myoglobin. A monomeric allosteric hemoprotein. *J. Biol. Chem.* **271**, 16999-17001 (1996).

19. Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J.-I. & Nagata, Y. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* **12**, 429-438 (2004).

20. Tsai, C.-J., del Sol, A. & Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **378**, 1-11 (2008).

21. Whitley, M.J. & Lee, A.L. Frameworks for understanding long-range intra-protein communication. *Curr. Protein Pept. Sci.* **10**, 116-127 (2009).

22. Lockless, S.W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-299 (1999).

23. Ota, N. & Agard, D.A. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* **351**, 345-354 (2005).

24. Yu, E.W. & Koshland, D.E., Jr Propagating conformational changes over long (and short) distances in proteins. *Proc. Natl. Acad. Sci. U.S.A* **98**, 9517-9520 (2001).

25. Ma, B., Kumar, S., Tsai, C.J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **12**, 713-720 (1999).

26. Tsai, C.J., Kumar, S., Ma, B. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **8**, 1181-1190 (1999).

27. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **254**, 1598-1603 (1991).

28. Swain, J.F. & Gierasch, L.M. The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* **16**, 102-108 (2006).

29. Lee, J.C. Modulation of allostery of pyruvate kinase by shifting of an ensemble of microstates. *Acta Biochim. Biophys. Sin.* **40**, 663-669 (2008).

30. Nelson, S.W., Iancu, C.V., Choe, J.Y., Honzatko, R.B. & Fromm, H.J. Tryptophan fluorescence reveals the conformational state of a dynamic loop in recombinant porcine fructose-1,6-bisphosphatase. *Biochemistry* **39**, 11100-11106 (2000).

31. Malmendal, A., Evenäs, J., Forsén, S. & Akke, M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. *J. Mol. Biol.* **293**, 883-899 (1999).

32. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35**, 539-546 (2010).

33. Bahar, I., Chennubhotla, C. & Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **17**, 633-640 (2007).

34. Kern, D. & Zuiderweg, E.R.P. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748-757 (2003).

35. Cooper, A. & Dryden, D.T. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **11**, 103-109 (1984).

36. Popovych, N., Sun, S., Ebright, R.H. & Kalodimos, C.G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **13**, 831-838 (2006).

37. Buskirk, A.R. & Liu, D.R. Creating Small-Molecule-Dependent Switches to Modulate Biological Functions. *Chemistry & Biology* **12**, 151-161 (2005).

38. Ferraz, R.M., Vera, A., Arís, A. & Villaverde, A. Insertional protein engineering for analytical molecular sensing. *Microb. Cell Fact.* **5**, 15 (2006).

39. Villaverde, A. Allosteric enzymes as biosensors for molecular diagnosis. *FEBS Lett.* **554**, 169-172 (2003).

40. Laskowski, R.A., Gerick, F. & Thornton, J.M. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **583**, 1692-1698 (2009).

41. Christopoulos, A. & Kenakin, T. G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev.* **54**, 323-374 (2002).

42. Guntas, G. & Ostermeier, M. Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.* **336**, 263-273 (2004).

43. Higaki, J.N., Haymore, B.L., Chen, S., Fletterick, R.J. & Craik, C.S. Regulation of serine protease activity by an engineered metal switch. *Biochemistry* **29**, 8582-8586 (1990).

44. Mathonet, P. & Fastrez, J. Engineering of non-natural receptors. *Curr. Opin. Struct. Biol.* **14**, 505-511 (2004).

45. Brennan, C.A., Christianson, K., La Fleur, M.A. & Mandecki, W. A molecular sensor system based on genetically engineered alkaline phosphatase. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5783-5787 (1995).

46. Ostermeier, M. Engineering allosteric protein switches by domain insertion. *Prot. Eng. Des. Sel.* **18**, 359-364 (2005).

47. Lee, J., Natarajan, M., Nashine, V.C., Socolich, M., Vo, T., Russ, W.P., Benkovic, S.J. & Ranganathan, R. Surface Sites for Engineering Allosteric Control in Proteins. *Science* **322**, 438-442 (2008).

48. Radley, T.L., Markowska, A.I., Bettinger, B.T., Ha, J.-H. & Loh, S.N. Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol.* **332**, 529-536 (2003).

49. Marvin, J.S. & Hellinga, H.W. Conversion of a maltose receptor into a zinc biosensor by computational design. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4955-4960 (2001).

50. Marvin, J.S., Corcoran, E.E., Hattangadi, N.A., Zhang, J.V., Gere, S.A. & Hellinga, H.W. The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4366-4371 (1997).

51. Tang, C., Schwieters, C.D. & Clore, G.M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078-1082 (2007).

52. Telmer, P.G. & Shilton, B.H. Structural studies of an engineered zinc biosensor reveal an unanticipated mode of zinc binding. *J. Mol. Biol.* **354**, 829-840 (2005).

53. Takeuchi, M., Ikeda, M., Sugasaki, A. & Shinkai, S. Molecular design of artificial molecular and ion recognition systems with allosteric guest responses. *Acc. Chem. Res.* **34**, 865-873 (2001).

54. Takeuchi, M., Imada, T. & Shinkai, S. A Strong Positive Allosteric Effect in the Molecular Recognition of Dicarboxylic Acids by a Cerium(IV) Bis[tetrakis(4-pyridyl)porphyrinate] Double Decker. *Angew. Chem. Intern. Ed.* **37**, 2096-2099 (1998).

55. Ostermeier, M. & Benkovic, S.J. Evolution of protein function by domain swapping. *Adv. Protein Chem.* **55**, 29-77 (2000).

56. Doi, N. & Yanagawa, H. Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett.* **453**, 305-307 (1999).

57. Liang, J., Kim, J.R., Boock, J.T., Mansell, T.J. & Ostermeier, M. Ligand binding and allostery can emerge simultaneously. *Protein Sci.* **16**, 929-37 (2007).

58. Guntas, G., Mansell, T.J., Kim, J.R. & Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11224-11229 (2005).

59. Guntas, G., Mitchell, S.F. & Ostermeier, M. A molecular switch created by in vitro recombination of nonhomologous genes. *Chem. Biol.* **11**, 1483-1487 (2004).

60. Wright, C.M., Heins, R.A. & Ostermeier, M. As easy as flipping a switch? *Curr. Opin. in Chem. Biol.* **11**, 342-346 (2007).

61. Legendre, D., Soumillion, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67-72 (1999).

62. Luque, I. & Freire, E. Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins* **41**, 63-71 (2000).

63. Freire, E. The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10118-10122 (1999).

64. Mathonet, P., Deherve, J., Soumillion, P. & Fastrez, J. Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. *Protein Sci.* **15**, 2323-2334 (2006).

65. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).

66. Mathonet, P. Création d'un site allostérique dans la beta-lactamase TEM-1 par évolution dirigée. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2004).

67. Volkov, A.N., Humberto, B., Mathonet, P., Evrard, C., Ubbink, M., Declercq, J.P., Soumillion, P. & Fastrez, J. Engineering an Allosteric Binding Site for Aminoglycosides into TEM1-β-Lactamase. *Chembiochem* **12**, 904-913 (2011).

68. Fernandez-Gacio, A., Uguen, M. & Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **21**, 408-14 (2003).

69. Kehoe, J.W. & Kay, B.K. Filamentous phage display in the new millennium. *Chem. Rev.* **105**, 4056-72 (2005).

70. Azzazy, H.M.E. & Highsmith, W.E. Phage display technology: clinical applications and recent innovations. *Clin. Biochem.* **35**, 425-45 (2002).

71. Smith, G. & Petrenko, V. Phage Display. *Chem. Rev.* **97**, 391-410 (1997).

72. Soumillion, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G. & Fastrez, J. Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J. Mol. Biol.* **237**, 415-422 (1994).

73. Kristensen, P. & Winter, G. Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* **3**, 321-328 (1998).

74. Deperthes, D. Phage display substrate: a blind method for determining protease specificity. *Biol. Chem.* **383**, 1107-1112 (2002).

75. Waxman, D.J. & Strominger, J.L. Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. *Annu. Rev. Biochem.* **52**, 825-869 (1983).

76. Tipper, D.J. & Strominger, J.L. 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.* **54**, 1133-1141 (1965).

77. Frère, J.-M. Beta-lactamases and bacterial resistance to antibiotics. *Mol. Microbiol.* **16**, 385-395 (1995).

78. Ambler, R.P. The structure of beta-lactamases. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **289**, 321-331 (1980).

79. Bush, K., Jacoby, G.A. & Medeiros, A.A. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob. Agents Chemother.* **39**, 1211-1233 (1995).

80. Livermore, D.M. beta-Lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* **8**, 557-584 (1995).

81. Matagne, A., Lamotte-Brasseur, J. & Frère, J.M. Catalytic properties of class A beta-lactamases: efficiency and diversity. *Biochem. J.* **330 (Pt 2)**, 581-598 (1998).

82. Philippon, A., Dusart, J., Joris, B. & Frère, J.M. The diversity, structure and regulation of beta-lactamases. *Cell. Mol. Life Sci.* **54**, 341-346 (1998).

83. Jacob, F., Joris, B., Lepage, S., Dusart, J. & Frère, J.M. Role of the conserved amino acids of the « SDN » loop (Ser130, Asp131 and Asn132) in a class A beta-lactamase studied by site-directed mutagenesis. *Biochem. J.* **271**, 399-406 (1990).

84. Vanhove, M., Raquet, X., Palzkill, T., Pain, R.H. & Frère, J.M. The rate-limiting step in the folding of the cis-Pro167Thr mutant of TEM-1 beta-lactamase is the trans to cis isomerization of a non-proline peptide bond. *Proteins* **25**, 104-111 (1996).

85. Vanhove, M., Raquet, X. & Frère, J.M. Investigation of the folding pathway of the TEM-1 beta-lactamase. *Proteins* **22**, 110-118 (1995).

86. Hermann, J.C., Ridder, L., Mulholland, A.J. & Höltje, H.-D. Identification of Glu166 as the general base in the acylation reaction of class A beta-lactamases through QM/MM modeling. *J. Am. Chem. Soc.* **125**, 9590-9591 (2003).

87. Guillaume, G., Vanhove, M., Lamotte-Brasseur, J., Ledent, P., Jamin, M., Joris, B. & Frère, J.M. Site-directed mutagenesis of glutamate 166 in two beta-lactamases. Kinetic and molecular modeling studies. *J. Biol. Chem.* **272**, 5438-5444 (1997).

88. Adachi, H., Ohta, T. & Matsuzawa, H. Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin. *J. Biol. Chem.* **266**, 3186-3191 (1991).

89. Strynadka, N.C., Adachi, H., Jensen, S.E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K. & James, M.N. Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution. *Nature* **359**, 700-705 (1992).

90. Damblon, C., Raquet, X., Lian, L.Y., Lamotte-Brasseur, J., Fonze, E., Charlier, P., Roberts, G.C. & Frère, J.M. 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.* **93**, 1747-1752 (1996).

91. Legendre, D., Fastrez, J. & Soumillion, P. L'ingénierie génétique des enzymes pour créer des nouveaux outils diagnostiques. *Bioforum Intern.* **1**, 40-42 (2002).

92. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S. & Palzkill, T. Amino acid sequence determinants of beta-lactamase structure and activity. *J. Mol. Biol.* **258**, 688-703 (1996).

93. Hallet, B., Sherratt, D.J. & Hayes, F. Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Res.* **25**, 1866-1867 (1997).

94. Legendre, D., Vucic, B., Hougardy, V., Girboux, A.-L., Henrioul, C., Haute, J.V., Soumillion, P. & Fastrez, J. TEM-1 β-lactamase as a scaffold for protein recognition and assay. *Protein Sci.* **11**, 1506–1518 (2002).

95. Ruth, N., Quinting, B., Mainil, J., Hallet, B., Frère, J.-M., Huygen, K. & Galleni, M. Creating hybrid proteins by insertion of exogenous peptides into permissive sites of a class A beta-lactamase. *FEBS J.* **275**, 5150-5160 (2008).

96. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391 (1994).

97. Long-McGie, J., Liu, A.D. & Schellenberger, V. Rapid in vivo evolution of a beta-lactamase using phagemids. *Biotechn. Bioeng.* **68**, 121-125 (2000).

98. Zaccolo, M. & Gherardi, E. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775-783 (1999).

99. Ripoll, A., Baquero, F., Novais, A., Rodríguez-Domínguez, M.J., Turrientes, M.-C., Cantón, R. & Galán, J.-C. In vitro selection of variants resistant to beta-lactams plus beta-lactamase inhibitors in CTX-M beta-lactamases: predicting the in vivo scenario? *Antimicrob. Agents Chemother.* **55**, 4530-4536 (2011).

100. Hayes, F., Hallet, B. & Cao, Y. Insertion mutagenesis as a tool in the modification of protein function. Extended substrate specificity conferred by pentapeptide insertions in the omega-loop of TEM-1 beta-lactamase. *J. Biol. Chem.* **272**, 28833-28836 (1997).

101. Barlow, M. & Hall, B.G. Predicting evolutionary potential: in vitro evolution accurately reproduces natural evolution of the tem beta-lactamase. *Genetics* **160**, 823-832 (2002).

102. Babic, M., Hujer, A.M. & Bonomo, R.A. What's new in antibiotic resistance? Focus on beta-lactamases. *Drug Resist. Updat.* **9**, 142-156 (2006).

103. Sideraki, V., Huang, W., Palzkill, T. & Gilbert, H.F. A secondary drug resistance mutation of TEM-1 β-lactamase that suppresses misfolding and aggregation. *Proc. Natl. Acad. Sc. U.S.A.* **98**, 283-288 (2001).

104. Zaccolo, M., Williams, D.M., Brown, D.M. & Gherardi, E. An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues. *J. Mol. Biol.* **255**, 589-603 (1996).

105. Bershtein, S., Goldin, K. & Tawfik, D.S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029-1044 (2008).

106. Hecky, J. & Muller, K.M. Structural Perturbation and Compensation by Directed Evolution at Physiological Temperature Leads to Thermostabilization of β-Lactamase. *Biochemistry* **44**, 12640-12654 (2005).

107. Osuna, J., Pérez-Blancas, A. & Soberón, X. Improving a circularly permuted TEM-1 beta-lactamase by directed evolution. *Protein. Eng.* **15**, 463-70 (2002).

108. Kather, I., Jakob, R.P., Dobbek, H. & Schmid, F.X. Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J. Mol. Biol.* **383**, 238-51 (2008).

109. Sieber, V., Plückthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-60 (1998).

110. Bloom, J.D., Labthavikul, S.T., Otey, C.R. & Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869-5874 (2006).

111. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C. & Arnold, F.H. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 606-611 (2005).

112. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382-388 (2005).

113. Tokuriki, N. & Tawfik, D.S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596-604 (2009).

114. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D.S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002 (2008).

115. DePristo, M.A., Weinreich, D.M. & Hartl, D.L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678-687 (2005).

116. Meiering, E.M., Serrano, L. & Fersht, A.R. Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* **225**, 585-589 (1992).

117. Shoichet, B.K., Baase, W.A., Kuroki, R. & Matthews, B.W. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 452-456 (1995).

118. Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys.: Condens. Matter* **22**, 323101 (2010).

119. Zavodszky, P., Kardos, J., Svingor, A. & Petsko, G.A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7406-7411 (1998).

120. Collins, T., Meuwis, M.-A., Gerday, C. & Feller, G. Activity, Stability and Flexibility in Glycosidases Adapted to Extreme Thermal Environments. *J. Mol. Biol.* **328**, 419-428 (2003).

121. D'Amico, S., Gerday, C. & Feller, G. Temperature Adaptation of Proteins: Engineering Mesophilic-like Activity and Stability in a Cold-adapted α-Amylase. *J. Mol. Biol.* **332**, 981-988 (2003).

122. Wingreen, N.S., Li, H. & Tang, C. Designability and thermal stability of protein structures : Protein Conformational Computation. *Polymer* **45**, 699-705

123. Gilis, D. & Rooman, M. PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Eng.* **13**, 849-856 (2000).

124. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369-387 (2002).

125. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D.S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929-932 (2006).

126. Kiss, C., Temirov, J., Chasteen, L., Waldo, G.S. & Bradbury, A.R.M. Directed evolution of an extremely stable fluorescent protein. *Protein Eng. Des. Sel.* **22**, 313-323 (2009).

127. Huang, W. & Palzkill, T. A natural polymorphism in β-lactamase is a global suppressor. *Proc. Natl. Aca. Sc. U.S.A.* **94**, 8801-8806 (1997).

128. Wang, X., Minasov, G. & Shoichet, B.K. Evolution of an Antibiotic Resistance Enzyme Constrained by Stability and Activity Trade-offs. *J. Mol. Biol.* **320**, 85-95 (2002).

129. Fasan, R., Chen, M.M., Crook, N.C. & Arnold, F.H. Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew. Chem. Int. Ed. Engl.* **46**, 8414-8418 (2007).

130. Fasan, R., Meharenna, Y.T., Snow, C.D., Poulos, T.L. & Arnold, F.H. Evolutionary history of a specialized p450 propane monooxygenase. *J. Mol. Biol.* **383**, 1069-1080 (2008).

131. Weinreich, D.M., Delaney, N.F., Depristo, M.A. & Hartl, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114 (2006).

Objectifs et stratégie

I. Objectifs et stratégie générale

Cette thèse portant sur l'ingénierie génétique de β-lactamases allostériques s'inscrit dans la continuité d'un projet commencé il y a de cela plusieurs années au laboratoire BGMB. L'originalité de ce projet porte sur la création *de novo* d'un site de régulation allostérique au niveau d'une enzyme modèle, la β-lactamase TEM-1. L'ingénierie d'enzymes allostériques par les techniques modernes du génie génétique représente un défi actuel important du fait de ces nombreux intérêts tant du point de vue fondamental qu'appliqué (voir Introduction point II.3).

De par la complexité et relative méconnaissance du phénomène d'allostérie, différentes stratégies rationnelles ou aléatoires ont été développées dans le cadre de son ingénierie. La plupart de ces stratégies partagent néanmoins une base commune dans l'insertion d'un site de liaison à plus ou moins proche distance du centre actif de la protéine étudiée. Au laboratoire, une approche aléatoire a été élaborée afin de créer un tel site au sein de la β-lactamase TEM-1.

La stratégie mise en place auparavant par P. Mathonet consiste à construire une banque de phages présentant à leur surface des β-lactamases mutées au niveau de trois de leurs boucles de surface[1,2] **(Figure 1)**. Ces trois boucles (L1, L2 et L3) considérées ensemble créent des sites de liaison potentiels hautement aléatoires possédant une surface accessible au solvent (ASA) de plus de 1000 $Å^2$. Ainsi, une valeur d'ASA de ± 1249,4 $Å^2$ a été calculée (en utilisant le programme PyMOL, "probe radius" 1,4 Å ; "dot density", 4) pour le clone BlaKr précédemment cristallisé (voir Introduction) et présentant un insert de 8 résidus au niveau de sa boucle L2 et les résidus 240 et 269-272 randomisés respectivement au niveau de ses boucles L1 et L3. De tels sites devraient pouvoir reconnaître des ligands d'intérêt avec une haute affinité. En effet, selon une étude de Day *et al.*, une interaction protéine-protéine engendrant un changement d'ASA (ΔASA) de ± 1100 $Å^2$ à l'interface de liaison serait succeptible de générer une affinité de l'ordre du nM [3]. Par ailleurs, nous avons calculé selon la même méthode appliquée à BlaKr une valeur d'ASA de ± 2017,2 $Å^2$ pour le domaine variable d'un anticorps monochaine de camélidé reconnaissant la Rnase A avec une affinité de 35 nM (cAb-RN05 ; PDB file : 1BZQ)[4]. Tout comme les sites de liaison potentiels créés au sein des β-lactamases, le domaine variable de cet anticorps monochaine n'est constitué que de

3 boucles (ou CDR pour "Complementary Determining Region") par l'absence de la chaine légère. Des régions variables possédant une valeur d'ASA de plus de 2000 $Å^2$ devraient également pouvoir être observées au niveau des clones de β-lactamases possédant des inserts simultanément en L2 et L3. A noter que, concernant l'anticorps cAB-RN05, la reconnaissance de l'antigène ne fait intervenir que les CDR 1 et 3, soit une surface d'interaction de 570 $Å^2$ seulement[4].

A l'image des CDR des anticorps, il est légitime de penser que les sites de liaison construits dans la β-lactamase TEM-1 puissent adopter différentes conformations en fonction de la nature, de la structure et de la taille de la molécule reconnue. Ainsi, la formation de cavités, de sillons ou des surfaces de contact étendues pourrait permettre respectivement la reconnaissance de petites molécules organiques, de peptides ou d'ADN et de grosses molécules de type protéique.

La banque est construite de manière hiérarchique **(Figure 1)**. Deux banques indépendantes sont initialement construites avant d'être recombinées pour donner la banque finale contenant une large diversité de β-lactamases mutées simultanément au niveau des trois boucles. Une sélection *in vivo* pour activité des banques initiales permet éventuellement de ne recombiner que des clones actifs et donc bien repliés, ce qui devrait améliorer la qualité de la banque combinatoire finale.

Les insertions/mutations multiples sont généralement fortement perturbantes pour l'architecture protéique. De ce fait, la collection combinatoire précédemment construite par P. Mathonet ne contenait qu'un million de mutants actifs et mutés dans les trois boucles, soit environ 0,2 % de la diversité totale[1]. Par ailleurs, le maintien d'une activité enzymatique minimale imposait la présence démontrée d'un pont disulfure au niveau de l'insert dans la boucle N-terminale (L2) [1]. Cette fixation induit une rigidification de la boucle pouvant être préjudiciable dans la recherche d'une régulation allostérique. Autre limitation, dans la boucle C-terminale (L3) cette fois, la fixation préférentielle d'une thréonine ou d'une sérine aux extrémités de l'insert qui favorise le maintien de deux liaisons hydrogène entre les résidus sauvages T_{271} et Glu_{274}.

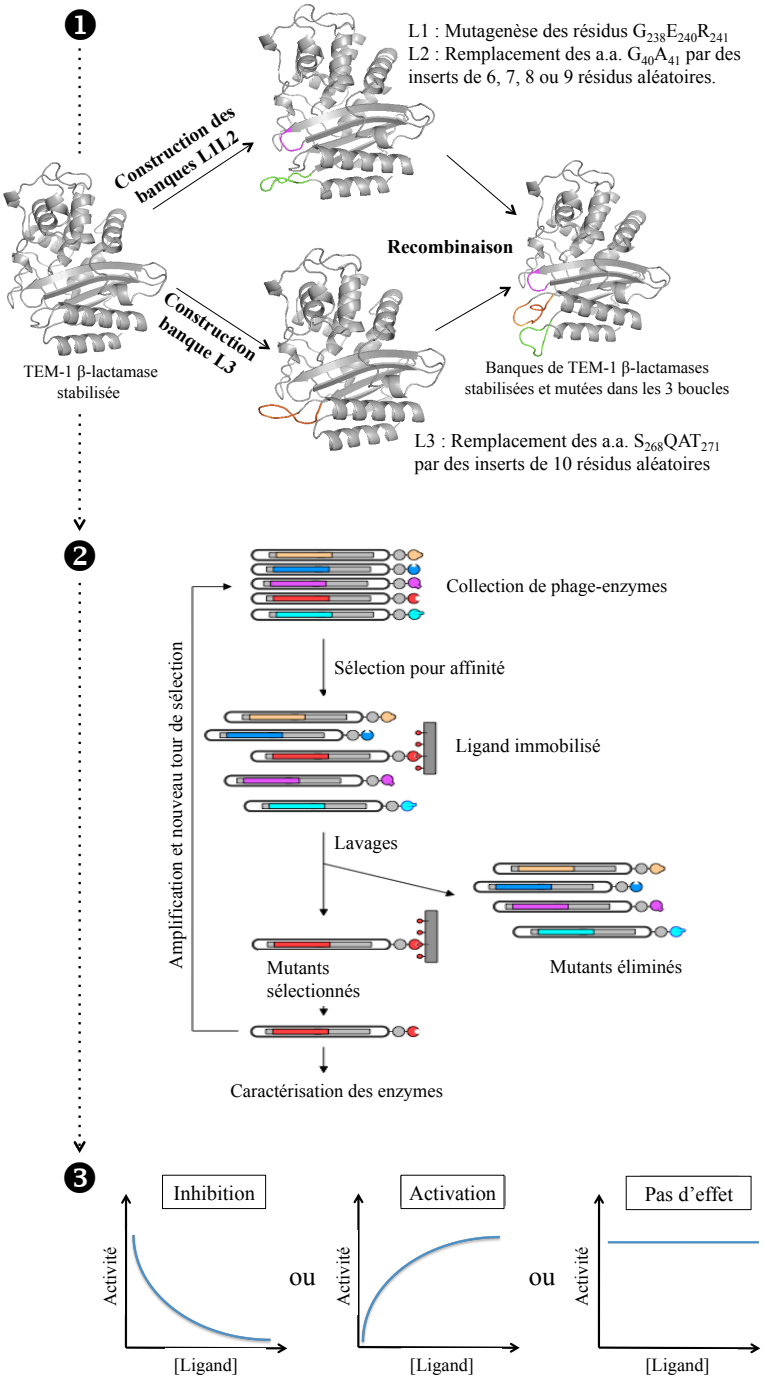


Figure 1 : Stratégie d'ingénierie de β-lactamases allostériques. (1) Construction hiérarchique d'une collection de sites de liaison potentiels au sein de la β-lactamase. (2) Sélection par phage display de clones reconnaissant une cible d'intérêt. (3) Criblage pour régulation allostérique par mesures cinétiques.

Dans ce projet, nous proposons de construire une collection complexe de mutants dont la diversité et la qualité seront largement augmentées, ce qui devrait multiplier les chances de sélectionner des clones d'intérêt. Pour cela, nous introduirons initialement des mutations stabilisantes (H153R, M182T, T195S et A224V) qui ont été récemment décrites pour la β-lactamase TEM-1[5–9]. Ces résidus se situent en dehors du site d'ingénierie et leurs mutations devraient permettre à l'enzyme de mieux supporter les perturbations ultérieures. Il a en effet été démontré qu'un gain de stabilité thermodynamique augmentait la tolérance des protéines face aux mutations et insertions[10–12]. D'autre part, une analyse structurale attentive laisse penser que la dégénérescence au niveau de la boucle C-terminale (L3) pourrait être augmentée sans nuire au repliement de l'enzyme. Antérieurement, la collection avait été construite en remplaçant la thréonine 271 par une séquence dégénérée de six résidus[1]. Pour ce projet, les résidus $S_{268}QAT_{271}$ seront remplacés par des inserts de dix résidus aléatoires. Cette modification devrait complexifier les sites de liaison et augmenter la surface de contact avec les ligands potentiels dans la perspective de sélectionner davantage d'interactions de haute affinité. D'autre part, le vecteur phage précédemment employé sera remplacé par un vecteur phagemidique afin de faciliter la construction des banques ADN. La possibilité offerte par ce vecteur de contrôler l'expression des protéines mutantes devrait également permettre de limiter les biais éventuels liés à l'expression de certains variants. Enfin, la voie de translocation SRP sera privilégiée à la voie SEC pour le ciblage de la protéine de fusion β-lactamase-p3 vers le périplasme. L'utilisation de la translocation co-traductionelle peut en effet augmenter la présentation de certaines protéines sur phages filamenteux[13].

Une sélection *in vitro* de clones reconnaissant une cible d'intérêt sera réalisée par phage display **(Figure 1)**. La collection de phages, chacun présentant à sa surface un mutant β-lactamase unique, sera incubée avec la cible immobilisée. Les phages sélectivement capturés pourront ensuite être amplifiés. Des tours de sélection successifs doivent ainsi permettre d'isoler des enzymes ayant une forte affinité pour le ligand choisi. L'activité β-lactamase des clones sélectionnés sera par la suite analysée en présence et en absence du ligand cible en vue d'isoler des β-lactamases régulées allostériquement. On s'intéressera aussi bien aux phénomènes d'activation que d'inhibition **(Figure 1)**. Cette stratégie a précédemment permis de sélectionner

des clones régulés par des anticorps[14], des aminoglycosides (clone Bla-Kr)[15] et des ions métalliques[2]. Les sélections à partir de la nouvelle banque seront validées en utilisant des cibles modèles telles que la streptavidine et l'ion nickel (II). D'autres cibles plus intéressantes pourront être choisies par la suite, tel des molécules de petites tailles ou des anticorps.

Les enzymes régulées seront éventuellement caractérisées davantage au niveau du phénomène de régulation, de leur stabilité et de leur affinité pour le ligand. De nouveaux tours de mutations/sélection pourront être envisagés afin d'améliorer l'effet de régulation et/ou l'affinité pour le ligand. Enfin, les enzymes les plus intéressantes pourront être purifiées et étudiées en détail par diverses méthodes allant de la RMN à la diffraction des rayons X.

Nous nous intéresserons également à l'éventuel impact des mutations stabilisantes sur la régulation allostérique d'un mutant sélectionné précédemment sans ces mutations (Bla-Kr). Le but étant de savoir si l'éventuelle rigidification de la structure résultant de la stabilisation affecte les propriétés de régulation.

II. Références

1. Mathonet, P., Deherve, J., Soumillion, P. & Fastrez, J. Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. *Protein Sci.* **15**, 2323-2334 (2006).
2. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).
3. Day, E.S., Cote, S.M. & Whitty, A. Binding Efficiency of Protein–Protein Complexes. *Biochemistry* **51**, 9124-9136 (2012).
4. Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M.A., Muyldermans, S. & Wyns, L. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **7**, 361-370 (1999).
5. Hecky, J. & Muller, K.M. Structural Perturbation and Compensation by Directed Evolution at Physiological Temperature Leads to Thermostabilization of β-Lactamase. *Biochemistry* **44**, 12640-12654 (2005).
6. Osuna, J., Pérez-Blancas, A. & Soberón, X. Improving a circularly permuted TEM-1 beta-lactamase by directed evolution. *Protein. Eng.* **15**, 463-70 (2002).
7. Sieber, V., Plückthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-60 (1998).

8. Bershtein, S., Goldin, K. & Tawfik, D.S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029-1044 (2008).

9. Kather, I., Jakob, R.P., Dobbek, H. & Schmid, F.X. Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J. Mol. Biol.* **383**, 238-51 (2008).

10. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C. & Arnold, F.H. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 606-611 (2005).

11. Bloom, J.D., Labthavikul, S.T., Otey, C.R. & Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869-5874 (2006).

12. Kiss, C., Temirov, J., Chasteen, L., Waldo, G.S. & Bradbury, A.R.M. Directed evolution of an extremely stable fluorescent protein. *Protein Eng. Des. Sel.* **22**, 313-323 (2009).

13. Steiner, D., Forrer, P., Stumpp, M.T. & Pluckthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat. Biotechnol.* **24**, 823-831 (2006).

14. Legendre, D., Soumillion, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67-72 (1999).

15. Volkov, A.N., Humberto, B., Mathonet, P., Evrard, C., Ubbink, M., Declercq, J.P., Soumillion, P. & Fastrez, J. Engineering an Allosteric Binding Site for Aminoglycosides into TEM1-β-Lactamase. *Chembiochem* **12**, 904-913 (2011).

Résultats

Protein stability promotes tolerance to insertional mutagenesis

Bruno Baudoux, Patrice Soumillion*

Laboratoire de Biochimie et de Génétique Moléculaire Bactérienne (BGMB)

Institut des Sciences de la Vie, Université Catholique de Louvain,

Place Croix du Sud, 4-5 boîte 3, 1348 Louvain-La-Neuve, Belgium

*To whom correspondence should be addressed:

E-mail: patrice.soumillion@uclouvain.be

Tel.: +32 10 47 30 75, fax +32 10 47 28 20

This paper has been submitted in Protein Engineering, Design and Selection (PEDS)

Keywords: TEM-1 β-lactamase, insertional tolerance, stability, loops, protein engineering.

Abstract

The mutational engineering of a protein can be highly elaborated but is generally limited by the mutations tolerance of the starting scaffold. Past studies essentially highlighted a positive correlation between protein stability and tolerance to substitutional mutagenesis. With the aim of generating allosteric enzymes selectable by phage display, we managed to highly increase the TEM-1 β-lactamase tolerance to random insertions by scaffold stabilization. Starting from a prestabilized tetramutant, we inserted random peptides in two loops after and before the N- and C-terminal helices of the protein and assessed the tolerance by measuring the percentage of active mutants. Stabilization increased the number of active clones by respectively around 30 and more than a thousand fold for the C-terminal and N-terminal insertions. Notably, at the N-terminal insertion site, the requirement for a stabilizing disulfide bridge is abolished. Our results show that a prestabilized protein is a better starting point for insertional and/or complex mutagenesis.

Introduction

Protein engineering is becoming a mature discipline that has developed through the emergence of genetic engineering techniques in the 1970s[1]. It aims to improve, modify, create, as well as to investigate protein structure function relationships. The endless opportunities rapidly interested the industry and this research area has expanded well beyond the academic world. We have lost count of the number of proteins that have been improved or modified with fundamental or applied objectives. Still, some protein scaffolds are more amenable to engineering than others thanks to their intrinsic characteristics. So is the TEM-1 β-lactamase, one of the first model enzyme to have been engineered by oligonucleotide-directed mutagenesis[1–4]. TEM-1 β-lactamase is a well-known monomeric enzyme, well characterized and of good stability with a free energy of unfolding of 44.0 ± 3 kJ.mol^{-1} [5]. Its high activity ($k_{cat}/K_m = 8.10^7$ $M^{-1}.s^{-1}$) [6] that approaches the diffusion rate of small molecules in solution is easily measurable and active enzymes are *in-vivo* selectable on ampicillin. Its catalytic activity is also maintained once displayed on phage[7]. Various engineering studies succeeded in improving its catalytic activity towards third generation β-lactam antibiotics[8–11] or in increasing its thermodynamic stability[12–15]. Along time, engineering became eventually more complex in response to increasingly audacious challenges. Regulatable β-lactamases were created by circular permutation and domain insertion[16,17] or by random insertions within surface loops[18]. The engineering potential of a given protein is, however, limited by its intrinsic tolerance to mutations.

A project aimed to generate allosteric binding sites in phage-displayed β-lactamases faced such a limit. Allosteric regulation by small molecules of interest often requires designing complex binding sites as low molecular weight ligands are generally buried in cavities, often covered by loops[19]. Based on their tolerance to insertions and mutations, surface exposure and vicinity of the active site, three contiguous loops of the TEM-1 β-lactamase were selected by P. Mathonet for random insertional and substitutional mutagenesis[20]. The combination of these three loops could mimic the structure of a variable domain of a camel monochain antibody consisting of only three CDRs (Complementary Determining Regions)[21] **(Figure 1)**. A library of mutants containing 5.10^8 clones where two insertions and substitutions

were simultaneously introduced in the three loops was constructed but only 0.2% of the mutants were active. This low percentage probably reflects a negative epistasis associated with the additivity of destabilizing effects. Insertion of degenerated peptides within a loop increases the entropy cost of the folding and can be quite destabilizing for a protein[22]. This may explain the fact that a disulfide bridge in the engineered N-terminal loop (L2) was required. Mutational tolerance of the two residues located at the ends of the C-terminal loop (L3) insert was also restricted. A threonine or a serine was favored, possibly keeping two stabilizing hydrogen bonds found in the wild-type enzyme[20]. Despite the rather poor diversity of this so-called "mimotopes" insertant library, several clones regulated by metal ions[23] or aminoglycosides[24] were selected from it.

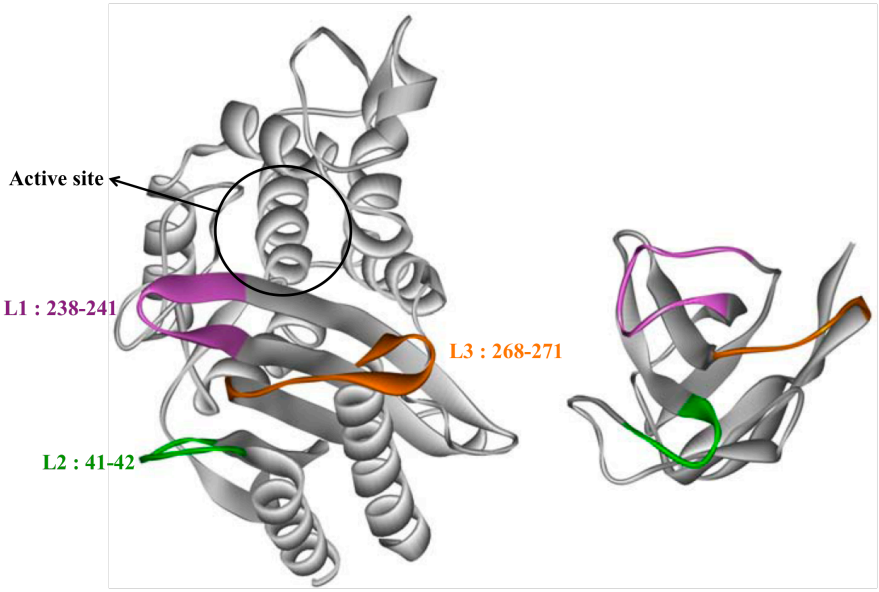


Figure 1: Model structure of a TEM-1 β-lactamase (left) featuring mutations/insertions in loops L1 (purple), L2 (green) and L3 (orange) and highlighting the ressemblance with the structure of a variable domain of a camel monochain antibody anti-RNase A (right; CDRs are coloured). The two structures are shown on the same scale. Figure from P.Mathonet *et al*[25].

In this work, we aimed to extend the insertional tolerance of the TEM-1 β-lactamase in order to improve the quality and diversity of the mimotope library. An improved diversity should promote the selection of clones of interest. Past studies highlighted a relation between protein stability and tolerance to mutations[26–28]. For example, Bloom *et al.* showed that stabilization of the TEM-1 β-lactamase allowed the protein

to gain extra robustness to substitutions, extra thermodynamic stability acting in buffering the destabilizing effect of the first few substitutions[27]. The effect of stabilization on the tolerance to insertional mutagenesis is however much less documented. By directed evolution, Kiss *et al.* found stabilizing mutations that could thwart eCGP123 fluorescent protein destabilization upon insertion of an antibody's CDR3 in three loops[28].We hypothesized that a stabilized β-lactamase could be a better starting scaffold for the construction of complex insertant libraries. Stabilization of the TEM-1 β-lactamase has been achieved several times in directed evolution studies[12,14,15,29]. Different stabilizing mutations are shared between these studies. We chose four of them (H153R, M182T, T195S & A224V) as their additive effects in the TEM-1 β-lactamase have already been studied in details by Hecky *et al*[12].

A combinatorial library in which the three loops were engineered was created in a hierarchical way. Two initial libraries were firstly constructed before being combined to create a large second-generation library. This strategy allowed assessing the individual impacts on protein stability of the insertions in L3 and of the insertions/mutations in L1-L2. The tolerance of the stabilized TEM-1 β-lactamase to insertions/mutations was phenotypically assessed in respect to the percentage of active insertants on ampicillin. This ensures that the selected enzymes are properly folded. A statistical analysis of the amino acids repartition in the inserts of the combinatorial library was also performed. The results were compared with those obtained with an unstabilized enzyme.

Results and discussion

A stabilized tetramutant of the TEM-1 β-lactamase, called blaS, was used as an engineering scaffold to assess the gain in robustness to insertional mutagenesis brought by the extra thermodynamic stability. The four stabilizing mutations, H153R, M182T, T195S and A224V are remote from the engineering sites. Compared to the wild type enzyme, the chemical stability of the tetramutant is increased by 15.4 kJ/mol and 11.4 kJ/mol respectively for the first and the second transition, assuming a three states unfolding scheme (native, intermediate, unfolded)[12]. Insertant libraries were constructed in the phagemid pHDi-Ex[30] so as to allow, later on, phage display selections of mutants that have acquired an affinity for target ligands. The stabilized β-lactamase gene was cloned in fusion with the C-terminal part of the phage coat protein gene p3 **(See supplementary data, Figure S1)**. A suppressive amber stop codon (TAG) was placed between the two genes. Working with a non-suppressive or a suppressive *E. coli* strain, the β-lactamase will be expressed respectively as a free enzyme or as a membrane anchored fusion protein.

Insertions in the C-terminal loop are fully tolerated

In the first library, the C-terminal loop that we called L3 was chosen for insertional mutagenesis. This loop is connecting the β5 strand to the C-terminal helix α11 **(Figure 1)**. In the wild type enzyme, the replacement of Thr_{271} by six random residues in this loop afforded a library containing around 7% of active insertants[18]. In this work, we tried to increase the degeneracy in L3 by replacing the residues Ser_{268} to Thr_{271} by 10 random residues. The collection was constructed by cassette insertion **(See supplementary data, Figure S2)**. *E. coli* suppressive strain TG1 was transformed with the DNA library and spread on LB agar plates supplemented with kanamycin and 1% glucose. Glucose represses the expression of the β-lactamases and limits biases that might result from the overexpression of some variants. This library, called p/blaS-L_3[10], contained approximately 3.5 x 10^7 individual clones **(Table 1)**. The sequences of seventeen clones picked at random were all different and well constructed **(Table 2)**. The percentage of active clones under inductive condition was approximated by plating the library simultaneously on LB media

supplemented with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and with or without ampicillin (20 µg/ml). Under induction without ampicillin, the number of colonies oddly dropped to 25% of that measured under repression condition. However, all the induced clones were active on ampicillin (20 µg/ml). With a similar library built from the unstabilized wild-type β-lactamase (p/bla-L_3^{10}), the percentage of active clones was only of 3.3%, highlighting the stabilization effect **(Table 1)**.

Library	Mutations introduced	Vector	Library size	% Active mutants
p/blaS-L_3^{10}	S_{268}QAT ➔$(X)_{10}$	pHDi-Ex	3.5 x 10^7 clones	100%
p/bla-L_3^{10}	S_{268}QAT ➔$(X)_{10}$	pHDi-Ex	2.0 x 10^7 clones	3.3%
fd/blaS-$L_1^RL_2^6$ fd/blaS-$L_1^RL_2^7$ fd/blaS-$L_1^RL_2^8$	G_{41}A ➔$(X)_{6,7\ or\ 8}$ & G_{238}ER ➔$(X)_3$	fd-tet ″ ″	1.0 x 10^7 clones 9.6 x 10^6 clones 9.7 x 10^6 clones	16.3% 35.0% 19.5%
fd/bla-$L_1^RL_2^6$ fd/bla-$L_1^RL_2^7$ fd/bla-$L_1^RL_2^8$	G_{41}A ➔$(X)_{6,7\ or\ 8}$ & G_{238}ER ➔$(X)_3$	fd-tet ″ ″	1.6 x 10^6 clones 2.2 x 10^6 clones 1.2 x 10^6 clones	0.08% (*) 0.07% (*) 0.08% (*)

Table 1: Libraries descriptions, diversities and percentages of active clones on ampicillin (20 µg/ml) in the amber suppressive strain TG1. (*) Results from P. Mathonet *et al.*[20].

Some clones from the activity-selected library were isolated and sequenced. The induction of the fusion protein expression induced a bias towards clones presenting at least one amber stop codon in their insert sequence **(Table 2)**. The requirement of an amber stop codon within the L3 insert would reduce the diversity by 70% considering a theoretical occurrence of 3% for this codon (1/32 with a NNK codon degeneracy). This estimation fits with the diversity loss observed upon induction. Being in an amber suppressive strain, amber stop codons are partially decoded as glutamines with an efficiency ranging from 0.7 to 25% depending notably on the 3' codon context[31]. The presence of an amber stop codon within the L3 insert thus decreases the expression level of the full-length fusion protein. Interestingly, when the library was transferred in the non-suppressive *E. coli* strain DH12S, the deleterious effect of IPTG induction disappeared and 80% of the induced clones were active on ampicillin. Moreover, the amber codons were forbidden in the active clones **(Table 2)**. Indeed, the 20% of inactive clones corresponds to the estimated

amount of clones with at least one amber stop codon within their L3 insert. These clones, active in TG1, are expressed as truncated enzymes in DH12S and are therefore expected to encode inactive β-lactamases. The induction problem observed in TG1 and not in DH12S is probably resulting from a toxicity of the membrane anchored β-lactamase-p3c fusion protein that is only expressed in the suppressive host. Regarding insertional mutagenesis, the results indicate a full tolerance in L3 loop.

Non-selected for activity (under repression)	Selected for activity in TG1 suppressive strain (under induction)	Selected for activity in DH12S non-suppressive strain (under induction)
A G W V K F C R L S	F K R V G A G H * A	I V A V S E S S S G
Q G M D C V H R V G	R L V P P T R L F *	G G Q W A N G T V L
Y R S G L R R N Q V	L R G * K G K E A G	G L S A Q P V R S T
M G T A A S A R V Y	A L M C S R A Q C *	R S P V A V F A G T
M M G P W R R T W R	V G * V L L Q S S T	W V V P K V Q V F N
C H W W A G H L F I	W S K Q H F L R P *	L A G A S R V P K P
L G L Y S D N A V A	I H Z R W A * G V T	R P I A T R G D G V
G P S F M S W A R L	F * A R T L T E N V	D C H V V G V W C V
M N S F L W C G G R	S R G W L H * G S L	R A C W D G G M C L
V M R V L A C L R V	W * R S L F N * P G	
S G C S R S D H R A	M A G A W V * S N T	
G D D P L R R F V S	R G * F L W V Q T D	
I R V A R F G * G A	Q R G A W P E Q P *	
W V H E R A A P C Y	R * H R G L L * E T	
G P T S L V D W R L	M L G G R W * G N G	
A P A P T A G G R V	G A * * G H P G I S	
G C R R L E V * G E	E W R V Q L V S * A	

Table 2: Insert sequences of mutants randomly isolated from the p/blaS-L_3^{10} library selected or not for activity on ampicillin (20 µg/ml). ✶: Amber stop read as a glutamine in the TG1 suppressive strain.

Highly increased tolerance in loops L1 and L2 mutagenesis

In the second libraries, we simultaneously mutated the loops called L1 and L2 **(Figure 1)**. The loop L1 is located between the strands β3 and β4 and comprises the residues Gly_{238}, Glu_{240}, and Arg_{241} that were randomized. The residue 239 is missing in the Ambler numbering scheme[32]. This loop is known to be tolerant to substitutional mutagenesis but intolerant to even short insertion of one amino acid[20]. Alanine, cysteine, glycine and serine were only allowed at position 238 because of limited mutational tolerance[33]. The second loop, L2, makes the junction between the N-terminal helix α1 and the strand β1 **(Figure 1)**. Inserts of 6, 7 or 8 random residues replaced the residues 41 and 42. In order to reduce the toxicity issue highlighted with the L3 library, these L1-L2 collections were built in a phage vector, fd-tet **(See annexes 1 & 3; p.219-221)**, which affords a lower expression level

through its constitutive promoter[34]. Indeed, similar libraries were previously built in fd-tet without detecting any fitness or toxicity problem[20]. These libraries, called fd/blaS-$L_1^R L_2^{6-8}$, were constructed by PCR with degenerated oligonucleotides **(See supplementary data, Figure S3).**

The DNA libraries were transformed in *E. coli* TG1 strain by electroporation. We obtained around 1.0 x 10^7 individual transformants, with an average of active clones on ampicillin (20 µg/ml) reaching 23.6% **(Table 1)**. This percentage represents an increase by around 300 fold compared to the results obtained with an unstabilized wild type β-lactamase[20]. Some clones randomly picked from the naïve or the activity selected libraries were sequenced **(See supplementary data, Table S1)**. All the clones were different and well constructed. No correlation was observed between the insert length in L2 and the percentage of active clones although an increased loss of stability upon loop elongation has been reported for two circular permutants of the alpha-spectrin SH3 domain[35].

In the unstabilized wild type β-lactamase libraries (fd/bla-$L_1^R L_2^{C5-7C}$), every activity selected clone presented a couple of cysteines at the extremities of the randomized L2 sequence[20]. These cysteines make a disulfide bridge that stabilizes the protein and is required for proper folding. The increased stability of blaS made that disulfide bridge unnecessary. This partly explains the large increased tolerance of L2 in the stabilized β-lactamase. Indeed, given the experimental codon distribution **(See supplementary data, Figure S5)** in the inserts of L2, the frequency of cysteine residue should be around 6.0% at each position. Imposing two cysteines at both extremities of the insert automatically reduces the diversity by approximately a factor 300 (= $1/(0.06)^2$).

In contrast to the bias observed during the loop L3 randomization, active clones analysis revealed a negative selection of the amber stop codon partially decoded as a glutamine in TG1. The lower expression level of the membrane anchored β-lactamase-p3 fusion protein together with its continuous assembly into secreted phage particles was apparently sufficient to decrease the fitness cost. On the contrary, the presence of such a codon might not allow enough β-lactamase expression for providing resistance to the bacterium against ampicillin.

Combinatorial library construction

First generation libraries were used for building combinatorial libraries of enzyme mutants with inserts in L2 and L3 and substitutions in L1. The quantitative and qualitative characteristics of these libraries and of a similar collection previously built starting from the wild type enzyme is shown in Table 3.

Given the initial results on loops tolerance to insertional mutagenesis, we initially recovered the inserts in L3 from the unselected library and cloned them in the fd/blaS-L_1L_2 libraries preselected on ampicillin (20 µg/ml) **(See supplementary data, Figure S4)**. This preselection allowed maximizing the functional diversity in the combinatorial library (fd/blaS-$L_1^{R}L_2^{6\text{-}8}L_3^{10}$). Highly electrocompetent *E. coli* TG1 cells were transformed with the recombined libraries DNA and 3.0 x 10^8 individual transformants were obtained.

By sequencing, 27 of 30 (90%) randomly picked clones were found carrying the three simultaneously engineered loops. Seven percents of mutant enzymes (2.12 x 10^7 clones) conferred resistance to ampicillin (20 µg/ml). Whereas all of them presented the expected modifications in loops L1 and L2, only a small fraction of these active clones (5%) carried the insertion in L3; this fraction represents only 10^6 clones or 0.35% of the total diversity **(Table 3)**. This numeric diversity is similar to the previous library constructed in the unstabilized β-lactamase[20] (fd/bla-$L_1^{R}L_2^{C5\text{-}7C}L_3^{6}$) **(Table 3)**. Yet, 10 random residues were inserted in L3 instead of 6. This qualitative difference can be seen as an increased diversity as five different hexapeptides can be distinguished in a decapeptide. Moreover, the need of a disulfide bridge in L2 was overcome. The lack of this constraint in the stabilized β-lactamase mutants increased the degeneracy of the insert and probably also the conformational diversity occupied by the engineered loops. This should promote the selection of clones with new binding activities and possibly the emergence of allosteric regulation.

Two non exclusive hypotheses may be proposed for explaining the weak insertional tolerance of L3 in the combinatorial phage library compared to the full tolerance observed in the single-insert phagemid library: the effect may result from the lower expression level in the phage vector and/or from the additivity of the destabilizing effects of multiple loops engineering[36].

In order to reduce the deleterious effect of the insertion in L3, a smaller six residues insert was introduced in loop as previously reported for libraries constructed by P. Mathonet in the wild-type β-lactamase[20]. The inserts were recovered by PCR from the fd/bla-$L_1^R L_2^{C6C} L_3^6$ library and cloned back in the fd/blaS-$L_1^R L_2^{6-8}$ libraries **(See supplementary data, Figure S4)**. It is worth noting that the recovered L3 inserts initially came from a single loop insertants library selected for activity. Three libraries totalizing 4.4 x 10^8 individual clones were obtained in TG1, 90% (3.1 x 10^8) of them being simultaneously mutated within the three loops **(Table 3)**. Half of mutants (46.8%) were active on ampicillin (20 µg/ml) and 15.7% (6.9 x 10^7 clones) were active and simultaneously mutated in the three loops. This last diversity is 70 times bigger than for the libraries fd/bla-$L_1^R L_2^{C5-7C} L_3^6$ that were built similarly from the wild type enzyme and with an imposed disulfide bridge in loop L2.

Library	Library size	% triple mutants	% Active triple mutants	Library characteristics
fd/bla-$L_1^R L_2^{C5-7C} L_3^6$	2.9 x 10^8 clones (*)	95.7%(*)	0.34%(*)	- Cysteine imposed at both extremities of the loop L2 - Residue T_{271} replaced by 6 NNK residues in L3
fd/blaS-$L_1^R L_2^{6-8} L_3^{10}$	3.0 x 10^8 clones	90.0%	0.35%	- No cysteine imposed at the extremities of the loop L2 - Residues $S_{268}QAT_{271}$ replaced by 10 NNK residues in L3
fd/blaS-$L_1^R L_2^{6-8} L_3^6$	4.4 x 10^8 clones	90.0%	15.7%	- No cysteine imposed at the extremities of the loop L2 - Residue T_{271} replaced by 6 NNK residues in L3

Table 3: Total and triple mutants diversities and qualitative characteristics of three combinatorial libraries selected or not for activity on ampicillin (20 µg/ml) in TG1. The term « triple mutants » refers to the β-lactamase variants simultaneously mutated in the three loops. (*) Results from Mathonet *et al*[20].

Final libraries analysis

Sequences of more than a hundred clones selected or not for activity in libraries fd/blaS-$L_1^R L_2^{6-8} L_3^6$ were determined **(See supplementary data, Tables S2-S4)**. A statistical analysis of those from the active combinatorial libraries (on 20 µg/ml ampicillin) was performed for comparing the observed amino acids distribution within the L1 & L2 inserts with the expected distribution that was extrapolated from the codon sequences from the unselected clones **(See supplementary data, Figure S5)**. The ratios of the observed occurrence (x_{obs}) to that of the expected one (x_{exp}) for each amino acid are represented in Figure 2. The sequences analysis was not performed for the loop L3 inserts because the number of sequences was not

sufficient for the 10 residues inserts and because the 6 residues inserts were recovered from a non stabilized β-lactamase library which had already undergone a selection for activity after the single loop library construction[18].

Loop L1 (residues 238-241):

Randomization of the position 238 was restricted to four residues: glycine, serine, cysteine and alanine. The over-representation of the wild-type glycine in active clones is consistent with the mutagenesis tolerance of this position[33]. Alanine and serine are tolerated but clearly disfavored as in the wild type β-lactamase[20]. The mutation G238S is widely observed in extended spectrum β-lactamases (ESBL) of the TEM family and is known to decrease the catalytic efficiency of the enzymes on ampicillin by 20 times[37] and to induce a protein stability loss of 8.4 $kJ.mol^{-1}$ [38]. Although the mutation M182T is able to counterbalance the destabilization effect, it has no impact on the loss of activity[38]. This would explain the low occurrence of the G238S substitution in our libraries. As in the wild type β-lactamase library[20], cysteine is never observed in position 238. Single cysteines are usually badly tolerated in phage-displayed protein. Unpaired cysteines could form a disulfide bond with a p3 cysteine residue and impair the infection process[39,40]. Illegitimate disulfide bridges with the p3 or other bacterial proteins could also disrupt the β-lactamase activity. Surprisingly, single cysteines were not rare (~ 12%) in positions 240 and 241 as in loops L2, a phenomenon rarely observed in the wild type β-lactamase library[20]. This increased tolerance may result from a faster folding rate linked to the stabilizing mutations that can potentially reduce the formation of illegitimate disulfide bridges.

Besides the cysteine difference, amino acids distribution in position 240 and 241 fits well with that previously observed in the wild type β-lactamase collection[20]. Nearly all residues are allowed but hydrophobic ones are less frequent. Selection of the stabilizing mutation E240K is however not favored unlike in the wild-type β-lactamase library suggesting that the associated gain in stability may become marginal in the pre-stabilized scaffold, resulting in an additional increased tolerance to mutagenesis of this position.

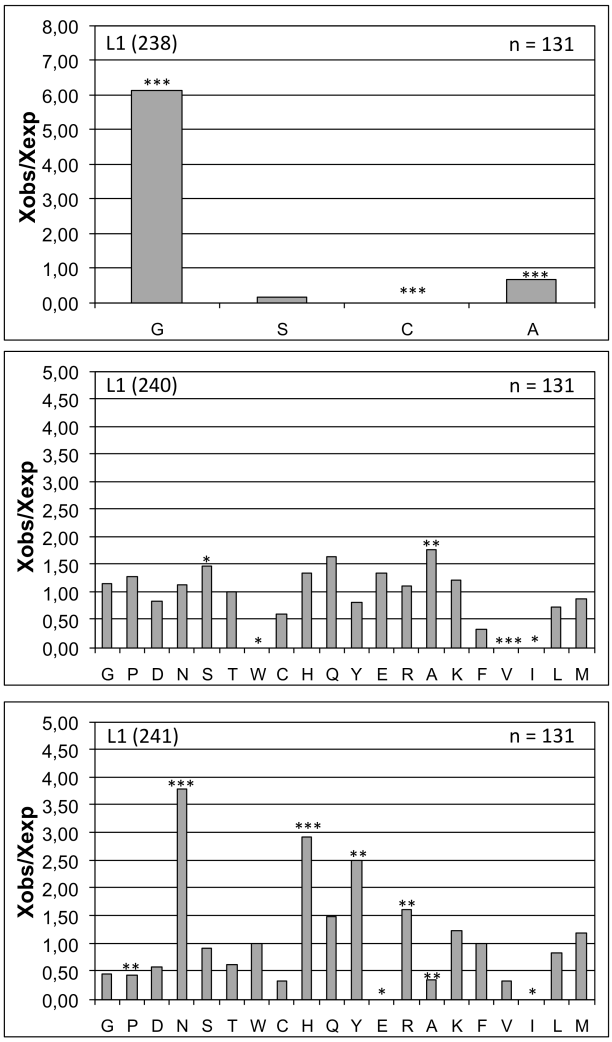


Figure 2: Relative amino acid occurrences in degenerated loops L1 and L2. The relative occurrence was defined by P. Mathonet as the ratio of the observed occurrence (X_{obs}) to that expected from the total number of observation (n = number of clones analysed), given the number of codons coding for a particular residue (X_{exp}). The deviation of X_{obs} relative to X_{exp} is shown as significant (*), highly significant (**) or very highly significant (***), with the probability of a chance observation being respectively 5%, 1% and 0.1%. The amino acids are arranged in order of decreasing propensity of presence in Ω-loop of high accessibility.

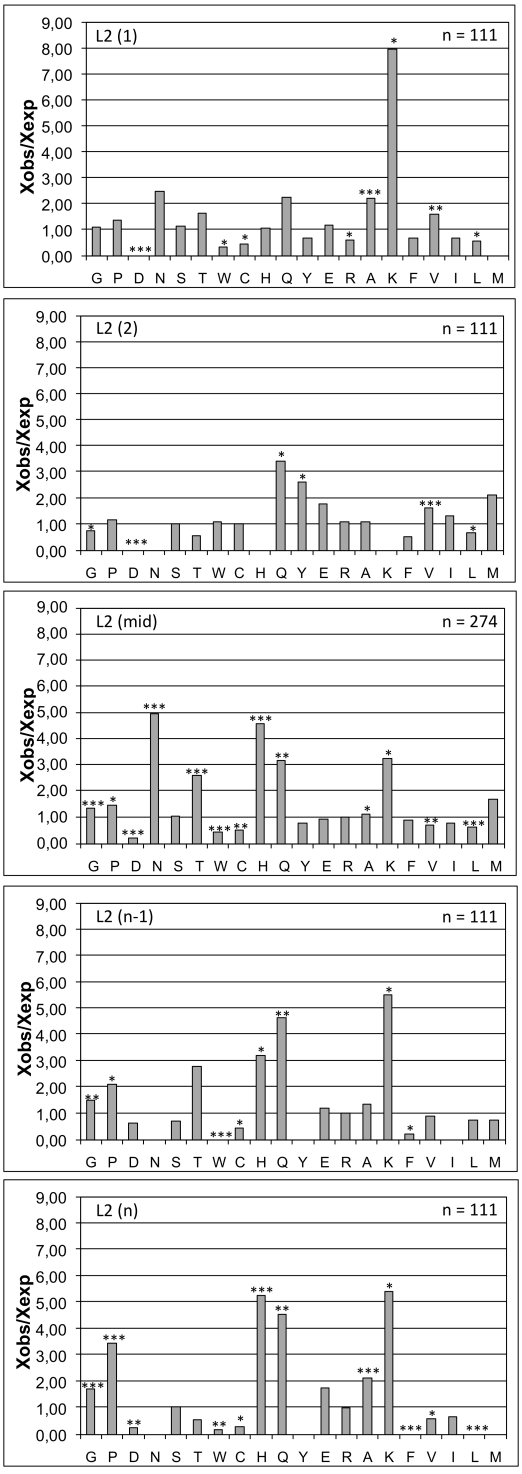


Figure 2: See legend on previous page.

Loop L2 (residues 41-42):

For the statistical analysis, the residues positions in the loop L2 were defined as follows:

$$Leu_{39}\text{-}Xaa\text{-}Xaa\text{-}(Xaa)_{2,\ 3\ or\ 4}\text{-}Xaa\text{-}Xaa\text{-}Arg_{42}$$

Positions:	1	2	Mid	n-1	n

The comparison of residue distribution in the stabilized versus wild-type enzyme in L2 is not meaningful since, unlike with the wild-type, the disulfide bridge was not imposed in the prestabilized scaffold, accounting alone for a large increase in insertional tolerance of this loop. Nevertheless, nearly all residues are observed at each position but large hydrophobic amino acids and aspartate are mostly disfavored. On the other hand and contrary to what is observed with the disulfide bridge constrained L2 inserts in the wild type β-lactamase[20], some large polar residues (Gln, His & Lys) are favored. These differences could be related to a higher mobility of the loops in absence of the disulfide bridge, pointing to an increased solvent accessibility in the context of a stabilized scaffold. Position 2 differs from the others by the absence of histidine and lysine. Proline is favored in the two last positions. This residue frequently occurs in turns and loops where it brings rigidity by imposing certain torsion angles on the segment of the polypeptide chain. It could stabilize the protein by reducing the loss in conformational entropy upon loop closure[41].

Conclusions

Gain in thermostability is known to increase the robustness of proteins toward substitutions14,26,27. Stabilization effect on tolerance to insertion is however not well documented. This study establishes that a stabilized TEM-1 β-lactamase is a better starting scaffold than the wild type enzyme for loop insertionnal mutagenesis. The tolerance of the enzyme toward single or multiple random insertions and mutations in three contiguous surface loops was indeed highly increased upon stabilization. The amplitude of the prestabilization advantage is however difficult to predict. It is most probably correlated with the amplitude of the stability loss associated with the mutagenic event. Assuming a cooperative thermodynamics of unfolding as schematized in Figure 3, the advantage of the prestabilization will be modest when mutagenesis is associated with a small stability loss (left arrow). This situation is probably what we observed when building the single L3 loop library. On the other hand, a large advantage of the prestabilization will be observed for mutagenic events associated with intermediate stability loss (middle arrow) as during the simultaneous randomization of the loops L1 and L2. Finally, a significant part of the prestabilization advantage will be lost upon an extensive mutagenesis that causes an important loss of stability (right arrow) as we observed when constructing the three loops combinatorial libraries.

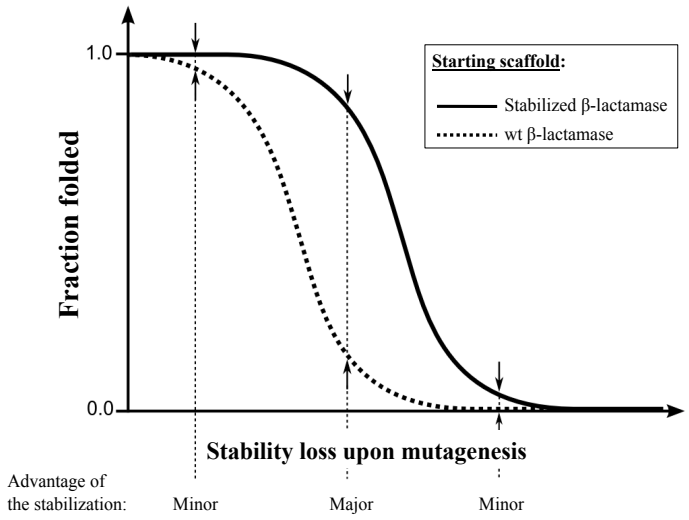


Figure 3: Schematic comparison of the tolerance of the wt (in dotted line) and stabilized (in solid line) TEM-1 β-lactamases as a function of the stability loss associated with insertionnal or substitutional mutagenesis. The tolerance is correlated to the fraction of folded protein remaining after the mutational event.

Overall, our results open new perspectives and challenges in the field of protein design and evolution, pushing forward the limits of engineering imposed by the intrinsic stability of the natural scaffolds.

Materials and methods

General

Expand High Fidelity, Fast Start High Fidelity & Pwo DNA polymerases were purchased from Roche. The T4 DNA ligase was from Promega. Restriction enzymes were purchased from New England Biolabs. Ampicillin and tetracycline were from AppliChem, kanamycin from Gibco.

Oligonucleotides

The oligonucleotides (Eurogentec) used in the different experiments are listed in Supplementary materials.

Bacterial Strains

TG1: *sup*E *thi-1 Δ(lac-pro*AB*) Δ(mcr*B-*hsd*SM*) 5(*$r^-_k m^-_k$*)* [F' *tra*D36 *pro*AB *lac*I^qZΔM15]

DH12S: 80d*lac*ZΔM15 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *ara*D139 Δ(*ara, leu*)7697 Δ(*lac*X74 *gal*U *gal*K *rps*L(Str^R) *nup*G *rec*A1/F' *pro*AB^+ *lac*I^qZΔM15

Genetic constructions

Construction of the cloning vectors

The pHDi-Ex.Kan^R vector used in this work is a derivative of the pHDi-Ex phagemid[30] (a gift from Pr. P. Minard, Université de Paris-Sud, France) where the ampicillin resistance gene was replaced by the kanamycin one[42].

pHDi-Ex/blaS vector (See supplementary data, Figure S1): The bla gene from the pGEM-T/bla vector[20] (TEM-1 β-lactamase gene with an *Apa*I restriction site at codons 251-252) was PCR amplified without the sequence encoding the signal peptide using the Pwo polymerase and the primers P3 and P4. The signal peptide sequence of the protein DsbA (DsbA S.P.) was amplified from the vector pBAD/DsbA received from G. Woronoff[43] using the Pwo polymerase and the primers P1 and P2. Both PCR products were purified with the "High Pure PCR Product Purification kit" from Roche. An overlap PCR was carried out using the purified PCR products as templates in a molecular ratio of 1:1, the primers pair P1 + P4 and the Pwo polymerase. Unlike the natural signal sequence of the β-lactamase,

the sequence utilizes a co-translational secretion pathway preventing the folding of the protein prior to secretion[44], a potential problem of the stabilized β-lactamase. The overlap PCR product, purified on 1% (w/v) agarose gel with the "QIAquick Gel Extraction kit" (Qiagen), was cloned between the *Nhe*I and *Nde*I sites of the pHDi-Ex.Kan^R vector. The ligation was done at 16°C for 1 h using the T4 DNA ligase and a molecular ratio insert/vector of 3:1.

The four stabilizing mutations were introduced in the bla gene using the oligonucleotides H153R, M182T, T195S, A224V and the "QuickChange® Multi Site-Directed Mutagenesis kit" from Strategene. The resulting gene is called blaS.

pHDi-Ex/blaS-L3 vector: This vector was constructed on the basis of the pHDi-Ex/blaS vector. Silent mutations were introduced at positions 266-267 of the blaS gene to create an *Age*I site. A *Bbs*I site was also introduced mutating the residues 270-271. The mutations were done using the "QuickChange® Site-Directed Mutagenesis Kit" from Stratagene and the oligonucleotide "*Age*I-*Bbs*I mut". *Bbs*I recognizes a non-palindromic sequence allowing the mutations to be removed upon library construction.

Construction of the initial libraries

p/blaS-L_3^{10} library (See supplementary data, Figure S2): Synthesis of the double stranded cassette was realized by annealing the primer cassette-L3 (24 µg) to the degenerated oligonucleotide cassette-L3 (8 µg) in 60 µl of annealing buffer (20 mM Tris Base, 2 mM $MgCl_2$, 50 mM NaCl, adjusted to pH 8.0 with NaOH). The mix was heated 10 min at 70°C, slowly cooled down to 37°C and stored on ice; 1.25 mM each dNTP in 1 mM DTT, 8.8 mM $MgCl_2$, 20 mM NaCl and 6 units of T4 DNA polymerase (Roche) were added to the annealing reaction product (60 µl) to obtain a final volume of 100 µl. The sample was subjected to the following cycle: 10 min on ice, 10 min at room temperature, 40 min at 37°C. The polymerization was stopped by cooling down to 4°C. The DNA cassette was purified on a 15% acrylamide gel. Soaking the gel in an ethidium bromide bath and exposing it under U.V. light revealed the DNA. The band of interest was cut off and the DNA was eluted from the gel by overnight incubation in milliQ water (500 µl) on a rotating wheel at 4°C. The DNA was restricted with *Age*I and *Bbs*I enzymes and the product was purified again on a 15% acrylamide gel as described above. The eluted DNA was

concentrated 20 times using a speedvac (Christ Alpha I-A) and then ligated in the pHDi-Ex/blaS-L3 vector restricted the same way and purified with the "High Pure PCR Product Purification kit" (Roche). The molecular insert/vector ratio of the ligation was 10:1 and the ligation was carried out at 4°C for 48 h using the T4 DNA ligase. *E. coli* TG1 strain was electroporated with the ligation product and plated on large LB agar plates supplemented with kanamycin (50 µg/ml) and 1% (w/v) glucose. For the unstabilized p/bla-L_3^{10} library, the same protocol was used but with the pHDi-Ex/bla-L3 vector.

fd/blaS-$L_1^R L_2^{6\text{-}8}$ libraries (See supplementary data, Figure S3): Fragment of the blaS gene was amplified from the pHDi-Ex/blaS vector using the degenerated primers pairs L1 & L2 (6), (7) or (8) and the Fast Start High Fidelity polymerase. The PCR products were purified with the "High Pure PCR Product Purification kit" from Roche, digested with *Bbs*I and purified again. These restricted fragment were cloned between the *Bbs*I sites of the fd*/blaΔ2* vector digested the same way and purified on "chromaspin 1000" (Clontech). The ligation was carried out overnight at 4°C using the T4 DNA ligase and a molecular ratio insert/vector of 5:1. TG1 *E. coli* strain was electroporated with the ligation product and plated on large LB agar plates supplemented with tetracycline (50 µg/ml) and ampicillin (20 µg/ml).

fd/blaS-$L_1^R L_2^{6\text{-}8} L3^{6,10}$ combinatorial libraries (See supplementary data, Figure S4)

(1) Vectors DNA of the fd/blaS-$L_1^R L_2^{6\text{-}8}$ and fd/blaS-$L_3^{6 \text{ or } 10}$ libraries were recovered from bacteria using the "Plasmid Maxi kit" (Qiagen). For each library, the initial culture was started with a number of bacteria at least 10 times higher than the library diversity to recover it all. (2) Inserts in L3 were recovered from the naïve fd/blaS-L_3^{10} library (or from the naïve library fd/bla-$L_1^R L_2^{C6C} L_3^6$ constructed by P. Mathonet[20]) by PCR using the P5 and P6 primers pair and the Expand High fidelity polymerase. The PCR was stopped before reaching the reaction plateau to avoid useless cycles of DNA denaturation, renaturation that may reduce transformation efficiency and generate biases. PCR products were incubated 30 min à 37°C with the proteinase K from Biolabs (0.2 mg/ml final) and 0.1% (w/v) SDS to remove the polymerase and then purified with the "High Pure PCR Product Purification Kit" (Roche). (3) The purified PCR product and the fd/blaS-$L_1^R L_2^{6\text{-}8}$ libraries DNA were

restricted with the endonucleases *Apa*I and *Not*I and respectively purified with the "High Pure PCR Product Purification Kit" (Roche) and a Chromaspin 1000 column (Clontech). (4) Restricted vector and PCR product were mixed in a molecular ratio of 1:10 and ligated overnight at 4°C with the T4 DNA ligase. (5) TG1 *E. coli* strain was electroporated with the ligation product and plated on LB agar plates supplemented with tetracycline (50 µg/ml) and ampicillin (20 µg/ml).

DNA libraries electroporation and titration

DNA libraries were electroporated by aliquots of 100 ng in 100 µl of *E. coli* electrocompetent cells (TG1 or DH12S) and suspended in 900 µl of SOC at 37°C. The cells were incubated for 1 h at 37°C, 180 rpm. The tubes were pooled together and 10 µl were sampled for titration. The rest was centrifuged for 20 min at 3500g, 4°C. The bacterial pellet was resuspended in 1 ml of LB and spread on several large LB agar plates (250 ml) supplemented with the adequate antibiotics and chemicals. Plates were incubated overnight at 37°C and colonies were harvested with 5 ml/plate of fresh LB medium and stored at -80°C in 20% (v/v) of glycerol. Libraries diversity was estimated by titration using the 10 µl sampled. Ten-fold serial dilutions were made in a final volume of 1 ml LB. For each dilution, 100 µl were plated on LB agar plates (25 ml) supplemented with the adequate antibiotics and chemicals. The plates were incubated overnight at 37°C. Colonies were counted and the total diversity was estimated using the following formula: CFU x dilution factor x (volume total/volume plated).

Preparation of electrocompetent cells

A preculture of bacteria was started in 10 ml of LB medium supplemented with the appropriate antibiotics and incubated overnight at 37°C under 180 rpm shaking. One liter of fresh LB medium was inoculated with the preculture and the whole was incubated at 37°C, 180 rpm shaking. Once OD_{600} reached 0.7, the culture was centrifuged at 3000g for 10 min, 4°C. The bacterial pellet was washed 2 times with 500 ml of sterile milliQ water and 1 time with 20 ml of 10% glycerol by resuspending each time the pellet and centrifugating at 3000g for 10 min, 4°C. After the last wash, the pellet was recovered in 1 ml of 10% (v/v) glycerol and aliquots of

100 µl were directly frozen in liquid nitrogen and stored at -80°C. All steps following the culture were carried out at 4°C and with solutions kept cold.
The competence of the cells was estimated by electroporating 2 ng of pUC18 amp^R (New England Biolabs) and plating 100 µl of several dilutions on LB Petri dishes supplemented with ampicillin 100 µg/ml. The Petri dishes were incubated overnight at 37°C. The competence was defined as the number of colonies obtained/µg of electroporated DNA.

Sequence analysis and statistics

DNA molecules were sequenced using the DYEnamic ET Terminator cycle Sequencing Kit (GE Healthcare) and the ABI PRISM 377 DNA sequencer (Applied Biosystems). Beta-lactamase genes cloned in the pHDi-Ex vector were sequenced using the pHDi-Ex Fwd and Rev primers. Beta-lactamase genes cloned in the fd-tet vector were sequenced using the fd-tet Fwd and Rev primers.

Analysis of DNA sequences was carried out using the ContigExpress tool from the Vector NTI Advance 9.0 program (Invitrogen).

To determine whether the observed number of a residue (x_{obs}) at a specific position significantly deviated from the expected one (x_{exp}), we used the binomial law to calculate the probability (*P*) of this occurring by chance as:

$$P_{x_{ob}} = C_N^{x_{obs}} \cdot p^{x_{obs}} \cdot (1-p)^{N-x_{obs}} \quad with \quad C_N^{x_{obs}} = \frac{N!}{x_{obs}! \cdot (N-x_{obs})!}$$

Where N corresponds to the total number of observations and p is the ratio between the number of codons coding for the specific amino acid and the total number of codons. A confidence index was assigned to each amino acids for which the gap between the observed and the expected occurrences is significant (*; $1\% < P \leq 5\%$), highly significant (**; $0{,}1\% < P \leq 1\%$) or very highly significant (***; $P \leq 0{,}1\%$)[20].

Acknowledgements

We thank Prof. P. Minard from the Université de Paris-Sud, France for providing the phagemid vector pHDi-Ex.

Funding

This work was supported by the Fonds pour la Recherche dans l'Industrie et l'Agroalimentaire (FRIA).

Reference

1. Leatherbarrow, R.J. & Fersht, A.R. Protein engineering. *Protein Eng.* **1**, 7-16 (1986).
2. Sigal, I.S., Harwood, B.G. & Arentzen, R. Thiol-beta-lactamase: replacement of the active-site serine of RTEM beta-lactamase by a cysteine residue. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7157-7160 (1982).
3. Charles, A.D., Gautier, A.E., Edge, M.D. & Knowles, J.R. Targeted point mutation that creates a unique Eco RI site within the signal codons of the beta-lactamase gene without altering enzyme secretion or processing. *J. Biol. Chem.* **257**, 7930-7932 (1982).
4. Dalbadie-McFarland, G., Cohen, L.W., Riggs, A.D., Morin, C., Itakura, K. & Richards, J.H. Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6409-6413 (1982).
5. Vanhove, M., Raquet, X. & Frère, J.M. Investigation of the folding pathway of the TEM-1 beta-lactamase. *Proteins* **22**, 110-118 (1995).
6. Raquet, X., Lamotte-Brasseur, J., Fonze, E., Goussard, S., Courvalin, P. & Frère, J.M. TEM beta-lactamase mutants hydrolysing third-generation cephalosporins. A kinetic and molecular modelling analysis. *J. Mol. Biol.* **244**, 625-639
7. Soumillion, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G. & Fastrez, J. Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J. Mol. Biol.* **237**, 415-422 (1994).
8. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391 (1994).
9. Long-McGie, J., Liu, A.D. & Schellenberger, V. Rapid in vivo evolution of a beta-lactamase using phagemids. *Biotechn. Bioeng.* **68**, 121-125 (2000).
10. Zaccolo, M. & Gherardi, E. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775-783 (1999).
11. Hayes, F., Hallet, B. & Cao, Y. Insertion mutagenesis as a tool in the modification of protein function. Extended substrate specificity conferred by pentapeptide insertions in the omega-loop of TEM-1 beta-lactamase. *J. Biol. Chem.* **272**, 28833-28836 (1997).
12. Hecky, J. & Muller, K.M. Structural Perturbation and Compensation by Directed Evolution at Physiological Temperature Leads to Thermostabilization of β-Lactamase. *Biochemistry* **44**, 12640-12654 (2005).
13. Sieber, V., Plückthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-60 (1998).
14. Bershtein, S., Goldin, K. & Tawfik, D.S. Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029-1044 (2008).

15. Kather, I., Jakob, R.P., Dobbek, H. & Schmid, F.X. Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J. Mol. Biol.* **383**, 238-51 (2008).

16. Guntas, G., Mitchell, S.F. & Ostermeier, M. A molecular switch created by in vitro recombination of nonhomologous genes. *Chem. Biol.* **11**, 1483-1487 (2004).

17. Guntas, G., Mansell, T.J., Kim, J.R. & Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11224-11229 (2005).

18. Legendre, D., Soumillion, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67-72 (1999).

19. Luque, I. & Freire, E. Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins* **41**, 63-71 (2000).

20. Mathonet, P., Deherve, J., Soumillion, P. & Fastrez, J. Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. *Protein Sci.* **15**, 2323-2334 (2006).

21. Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M.A., Muyldermans, S. & Wyns, L. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **7**, 361-370 (1999).

22. Nagi, A.D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* **2**, 67-75 (1997).

23. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).

24. Volkov, A.N., Humberto, B., Mathonet, P., Evrard, C., Ubbink, M., Declercq, J.P., Soumillion, P. & Fastrez, J. Engineering an Allosteric Binding Site for Aminoglycosides into TEM1-β-Lactamase. *Chembiochem* **12**, 904-913 (2011).

25. Mathonet, P. Création d'un site allostérique dans la beta-lactamase TEM-1 par évolution dirigée. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2004).

26. Bloom, J.D., Labthavikul, S.T., Otey, C.R. & Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869-5874 (2006).

27. Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C. & Arnold, F.H. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 606-611 (2005).

28. Kiss, C., Temirov, J., Chasteen, L., Waldo, G.S. & Bradbury, A.R.M. Directed evolution of an extremely stable fluorescent protein. *Protein Eng. Des. Sel.* **22**, 313-323 (2009).

29. Osuna, J., Pérez-Blancas, A. & Soberón, X. Improving a circularly permuted TEM-1 beta-lactamase by directed evolution. *Protein. Eng.* **15**, 463-70 (2002).

30. Heyd, B., Pecorari, F., Collinet, B., Adjadj, E., Desmadril, M. & Minard, P. In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry* **42**, 5674-5683 (2003).

31. Miller, J.H. & Albertini, A.M. Effects of surrounding sequence on the suppression of nonsense codons. *J. Mol. Biol.* **164**, 59-71 (1983).

32. Ambler, R.P., Coulson, A.F., Frère, J.M., Ghuysen, J.M., Joris, B., Forsman, M., Levesque, R.C., Tiraby, G. & Waley, S.G. A standard numbering scheme for the class A beta-lactamases. *Biochem J.* **276**, 269-270 (1991).

33. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S. & Palzkill, T. Amino acid sequence determinants of beta-lactamase structure and activity. *J. Mol. Biol.* **258**, 688-703 (1996).

34. Zacher, A.N., Stock, C.A., Golden, J.W. & Smith, G.P. A new filamentous phage cloning vector: fd-tet. *Gene* **9**, 127-140 (1980).

35. Viguera, A.-R. & Serrano, L. Loop length, intramolecular diffusion and protein folding. *Nat. Struct. Biol.* **4**, 939-946 (1997).

36. Wells, J.A. Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517 (1990).

37. Venkatachalam, K.V., Huang, W., LaRocco, M. & Palzkill, T. Characterization of TEM-1 beta-lactamase mutants from positions 238 to 241 with increased catalytic efficiency for ceftazidime. *Journal of Biological Chemistry* **269**, 23444 -23450 (1994).

38. Wang, X., Minasov, G. & Shoichet, B.K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85-95 (2002).

39. McConnell, S.J., Uveges, A.J., Fowlkes, D.M. & Spinella, D.G. Construction and screening of M13 phage libraries displaying long random peptides. *Mol. Divers* **1**, 165-176 (1996).

40. Lowman, H.B. & Wells, J.A. Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234**, 564-578 (1993).

41. Krieger, F., Möglich, A. & Kiefhaber, T. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J. Am. Chem. Soc.* **127**, 3346-3352 (2005).

42. Labarbe, C. Towards the evolution of a DD-peptidase into a beta-lactamase. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2006).

43. Woronoff, G. Développement de méthodes pour l'évolution dirigée de la pénicilline G acylase par compartimentalisation in vitro. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2012).

44. Steiner, D., Forrer, P., Stumpp, M.T. & Pluckthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat. Biotechnol.* **24**, 823-831 (2006).

Supplementary data

Phagemid description and construction

The β-lactamase gene (bla) was cloned in the phagemid pHDi-Ex.Kan^{R} [1] in fusion with the C-terminal part of the phage coat protein gene p3 (p3 c-term.) **(Figure S1; Annexes 1 & 2, p219-220**). An His-tag was added at the N-terminal extremity of the β-lactamase as well as the signal peptide of the periplasmic disulfide oxidoreductase protein (DsbA). We chose to target the fusion protein to the co-translational exportation pathway (SRP) instead of the Sec one as it has been shown that it could help in the display of the fusion proteins[2]. The construction is under control of two promoters: pT7 for protein overexpression (in DE3 containing strain) and purification and pLac for phage display. An amber stop (TAG) was added between the β-lactamase and the p3 genes. Working with a suppressive or a non-suppressive strain, we can either decide to express the fusion protein for phage display or the free β-lactamase for purification. The vector possesses a kanamycin resistance gene (KanaR) and two origins of replication; for double- or single-stranded replication.

The β-lactamase gene, the His-tag and the DsbA signal peptide (DsbA S.P.) sequences were assembled using the overlapping PCR method and cloned in the pHDi-Ex.Kan^{R} phagemid **(Figure S1)**. The four stabilizing mutations (H153R, M182T, T195S & A224V) were introduced by site directed mutagenesis to obtain the stabilized TEM-1 β-lactamase gene (blaS). The construction strategy is described in detail in the materials and methods section.

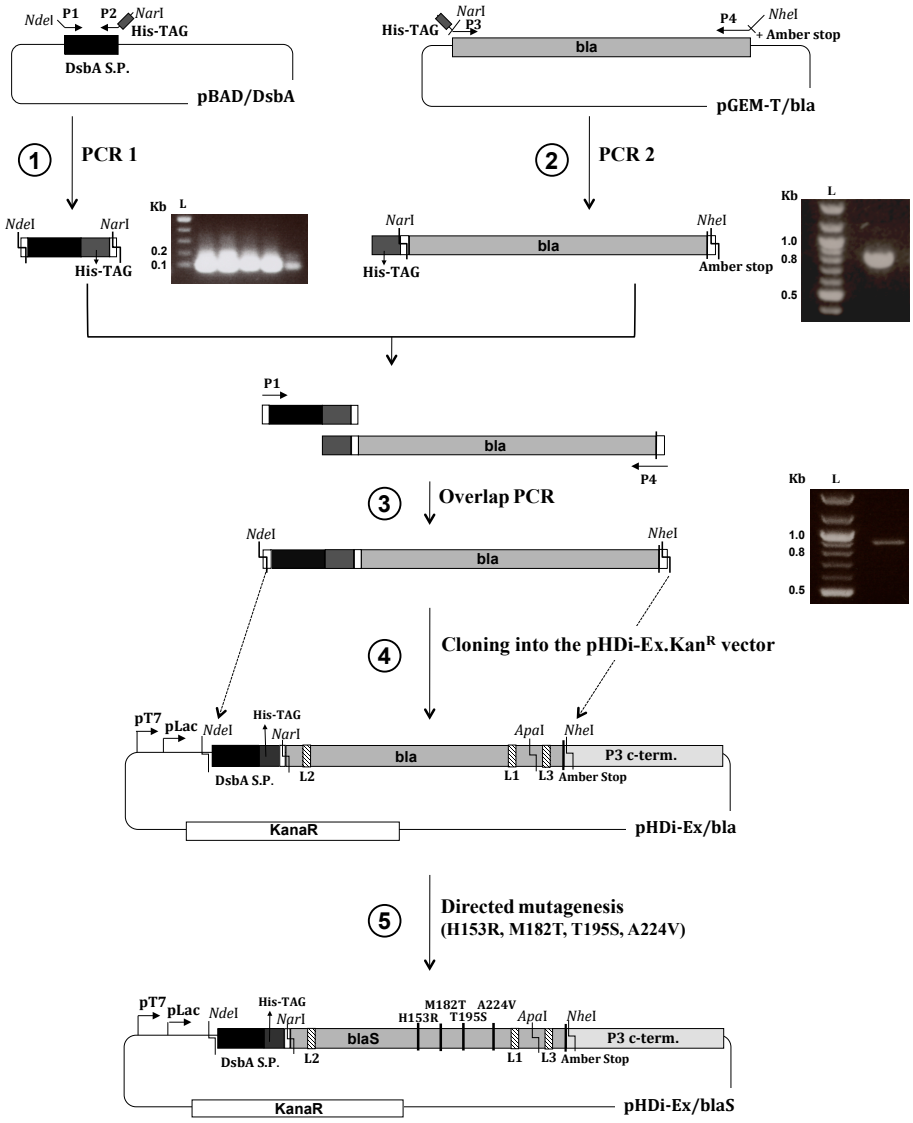


Figure S1: Construction strategy of the pHDi-Ex/blaS vector. ①-② : The DsbA signal peptide sequence (DsbA S.P.) was amplified by PCR using the primers pair P1+P2. The TEM-1 β-lactamase gene (bla) was amplified by PCR using the primers pair P3+P4. The P1 primer tail encodes the *Nde*I restriction site sequence. The P4 primer tail encodes the *Nhe*I restriction site sequence and an amber stop codon. The P2 and P3 primers tails are complementary and encode the His-Tag and a *Nar*I restriction site. ③: The P2-P3 complementarity allowed the PCR products to overlap and to serve as template for an overlapping PCR using primers pair P1+P4. ④: The overlapping PCR product was cloned between the *Nde*I and *Nhe*I sites of the pHDi-Ex vector. ⑤: The stabilized TEM-1 β-lactamase gene (blaS) was obtained by inserting the stabilizing mutations by site directed mutagenesis.
PCR products were analysed on agarose gel 1%. L = Molecular weight ladder (2-log from New England Biolabs).

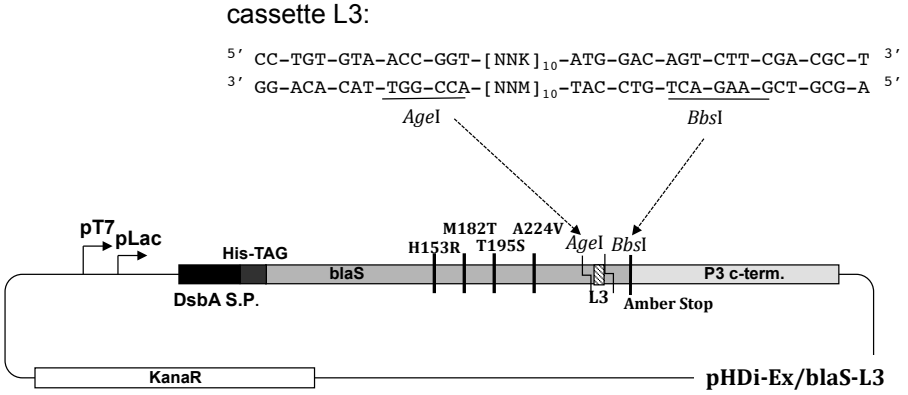


Figure S2: Randomization of the C-terminal loop, L3. A cassette (cassette L3) encoding for 10 random residues of type NNK (where N stands for any of the four deoxynucleotides G, A, T, and C and K stands for G or T) was cloned between the *Age*I and *Bbs*I sites of the pHDi-Ex/blaS-L3 vector. The NNK sequence represents 32 codons and encodes all 20 amino acids but only one stop codon. The stop codon being the amber one, suppressive *E. coli* strains can read through it and replace it by a glutamine. Moreover, the sequence bias due to the over-representation of some amino acids in the genetic code is reduced with the use of the NNK degeneration. The *Bbs*I site is nonpalindromic and disappears upon cassette ligation. Being nonpalindromic, it also prevents self-ligation of inserts and vectors. Experimental details are described in the materials and methods section.

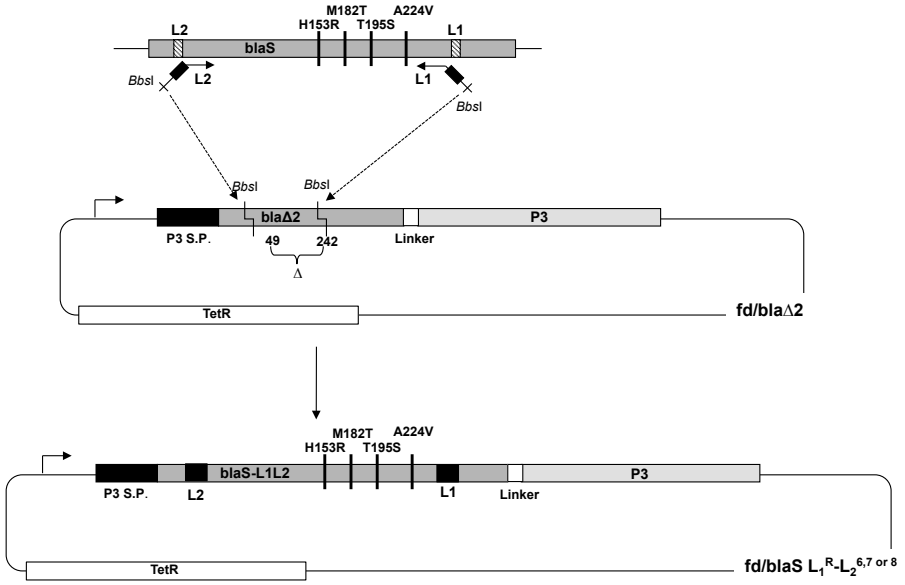


Figure S3: Randomization of loops L1 and L2. A fragment of the stabilized β-lactamase gene (blaS) was amplified using degenerated primers. The forward primer (L2) encodes 6, 7 or 8 degenerated codons of type NNK and a *Bbs*I restriction site at its 5' extremity. The reverse primer (L1) encodes three degenerated codons (MMN-MMN-RSM) and a *Bbs*I restriction site at its 5' extremity. The coding strand gives KSY-NNK-NNK for the degeneration. The KSY sequence only codes for four amino acids (alanine, cysteine, glycine and serine). The forward and reverse primers hybridize respectively just downstream of the residues 41-42 of the loop 2 and upstream of the residue 238 of the loop 1. The PCR product was cloned between the *Bbs*I sites of the fd/blaΔ2 vector. This vector encodes a truncated β-lactamase (blaΔ2) lacking the residues 49 to 242. The truncated gene sequence is replaced by a *Bbs*I restriction site. A second *Bbs*I site is located just upstream the L2 position. There is no frameshift but the expressed truncated β-lactamase is inactive. This phenotype allows creating libraries without any contamination of active wt β-lactamase. Experimental details are described in materials and methods section.

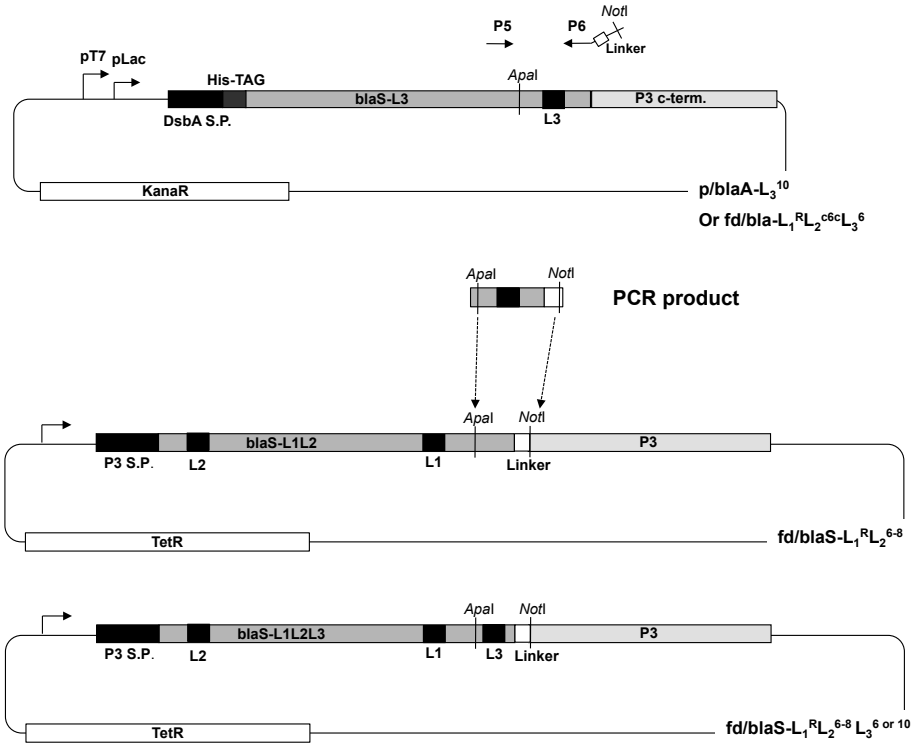


Figure S4: combinatorial library construction. Inserts in L3 were recovered from the initial library by PCR using the primers pair P5 and P6 hybridizing on either side of the randomized area . The 5′ tail of the P6 primer encodes the linker and a *Not*I restriction site (See annexe 3, p.221). The PCR product was cloned between the *Apa*I and *Not*I sites of the initial L1-L2 libraries DNA. The combitanorial libraries own mutations within the three loops. Experimental details are described in the materials and methods section.

	Non selected for activity		Selected for activity	
	L2	L1	L2	L1
$L_1^R L_2^6$	GVGLLL	SES	THSMGQ	SGN
	ASHVWP	C*C	VLIRLQ	GSR
	SAGSWG	nd	LVAPSF	GDP
	VLVGGG	nd	LTGGPH	GHH
	SFLVGC	A*R	SPWRAS	GKC
			LYLVSG	GPH
			RKMTSG	AEF
			HKKSKA	GFA
			ARVGRS	GEA
			HGAGPT	APH
$L_1^R L_2^7$	SGCGVSS	CIP	GEPWAPV	GSP
	GGGVGCV	SFF	GRDIAMS	GQQ
	nd	GAG	VGGCNMG	GSP
	APEILRV	AGN	VGRAGSG	GLP
	CGLGGNC	AGH	SSCDPRG	GGR
			GGVSSAG	GLY
			VLVGAGA	STD
			RSCRVGG	GPR
			GGGRVRS	GGM
			WGLRALG	GEP
$L_1^R L_2^8$	WRAGYAVG	SPP	IGLLGAVA	GGH
	MTRLLCVA	SAP	VAVGGWSP	SPR
	LFCPG*EA	ATV	MLAKVPRV	AGM
	GSLSRGAE	CHD	ATNTNCGA	GTN
	ADQREYSI	SES	GSGGGGVV	AAQ
			VGASPA	GNC
			CVGCGGGC	AYI
			SGGLSP	GLK
			WDGGHGAG	GVG
			QRAVTSYR	GSN

* : amber stop codon
nd : not determined

Table S1: Randomized loops sequences of few clones from the fd/blaS-$L_1^R L_2^{6-8}$ libraries selected or not for activity on 20 µg/ml ampicillin in TG1 *E. coli* strain.

	Non selected for activity			Selected for activity		
	L2	L1	L3	L2	L1	L3
$L_1^R L_2^6 L_3^6$	VVDRHQ	GLY	RSSYPT	LSVSQH	GNF	/
	SAASAA	APN	SKTAVT	WVRYVG	APH	SKPEPT
	GDPYQA	GKS	KRAQAS	AQTKQH	VRL	nd
	LSVLGA	GPS	SAPDIT	AWCAAG	GTH	SENQPS
	LVVERG	AHE	/	VCHLGG	GHN	nd
	RASWKT	GTN	KEVRST	LLSVAG	GRQ	/
	GGVSLG	GQQ	DQASTS	PMRMHR	GSR	/
	FLRFRG	GPV	ANDHGS	KLKTTS	SEP	/
	IYTRQH	GPS	/	SRHNAQ	GRV	/
	ELSWVG	GAS	TSKYST	TYDSAT	ASN	/
				QSTPAG	GSR	nd
				nd	ASG	/
				GWTGER	GSM	nd
				SARCDG	GPS	/
				PQNSTR	GPT	/
				SPVADP	GRY	/
				nd	ARR	nd
				VVAAGH	GAK	/
				ASGAKG	VRR	nd
				FFPVRE	GGS	/
				CWHRLP	GQQ	/
				AVDGHQ	GPV	/
				VGGLAG	GQT	/
				AVALLP	GRS	/
				SVRADG	ATQ	/
				ALPGTP	AHF	PCAAAC
				ASTWPR	GAL	GQTMPS
				LEWKSS	GRS	GNTSIM
				QQGGVP	GTL	/
				SVWALP	GQR	TAHRVT
				CAFRLC	GTL	/
				RVALRV	GHY	/
				PPMTTP	AAH	/
				KMGLGA	GPN	/
				ARTAGG	GLR	AGVAAG
				VCMAAP	GPR	/

* : amber stop codon
nd : not determined
/ : no insert

Table S2: Randomized loops sequences of few clones from the fd/blaS-$L_1^R L_2^6 L_3^6$ libraries selected or not for activity on 20 µg/ml ampicillin in TG1 *E. coli* strain.

	Non selected for activity			Selected for activity		
	L2	L1	L3	L2	L1	L3
$L_1^R L_2^7 L_3^6$	VRVQSGA	SHS	HWENRT	LSVHRDV	GMG	/
	ARSDVSA	GAT	TRGWPS	AACGGSG	GTN	VGNAVT
	YFVPVP	GSK	SEKNAL	VVWAQLE	GNY	/
	TVSACPR	GTN	KNKGAN	GGSFLVS	GGY	?
	GLQHAPR	GSR	TVNEST	FWFGGGG	GAF	/
	GGGMWWP	GLN	RTGGPS	PRYLVPG	GRY	/
	ATRYGSF	GSM	DALTPG	nd	GSR	/
	LMGNWRA	LRG	/	GVPVEAR	GHP	/
	SWRVGGG	AAY	*TTVPS	nd	STS	/
	GSSGTWA	ATN	QSKQPT	nd	ASH	nd
				GGVWGRG	GFS	/
				APALSLG	GCH	APTTGL
				VIGAGW	GSN	AAGQVN
				GCGAACR	GKQ	/
				RWAEQRR	GGH	ESIIPS
				GGRAGGA	ARN	/
				AIWGGLP	SSY	/
				GVLHGRV	ARH	/
				VRVSRCA	GYY	/
				VGSERPG	GQK	SSSQPK
				RYDGGRG	GKS	ESSMST
				RCVVACV	GAN	K
				VPGFRAG	ASY	/
				VCGCGGC	GAH	QGKFAG
				GRVGGGV	GAS	ANSAVA
				SLGGVLG	AQA	/
				GFRVCGA	GGK	/
				EVPTQR	GSN	MTSIPT
				nd	GQR	/
				AVVVSGA	ALQ	/
				FGGGCGA	AKH	/
				GVHRGGV	GLK	/
				GGFFGGG	GYW	/
				SGVNGGA	GSN	GWERRT
				WFFGGPP	GRS	/

* : amber stop codon
nd : not determined
/ : no insert

Table S3: Randomized loops sequences of few clones from the fd/blaS-$L_1^R L_2^7 L_3^6$ libraries selected or not for activity on 20 µg/ml ampicillin in TG1 *E. coli* strain.

	Non selected for activity			Selected for activity		
	L2	L1	L3	L2	L1	L3
$L_1^R L_2^8 L_3^6$	HQSFLG	GPG	DSKAPD	GPGRVLGQ	GMR	EGFSVS
	RAPPHD	GQR	TTNHPS	RRSPVG	GAR	nd
	VLGWGGGG	GSK	TEKDVD	SVFRRSVA	GCQ	nd
	GVGPAARA	ATG	EGDRVS	VRPVGD	GER	nd
	VAGGGSRP	ASS	NAVDVS	VYLRGS	GAH	nd
	VRSRSCCG	GPS	IGEWPT	VGSPLFGP	SPS	/
	VLGGHGPR	GTP	RSKVVN	AWEARAGV	GKR	GHRDGN
	GVSVGA	GLG	EDSYTK	VGGCDGGG	GLN	ERGEHT
	SGLSQP	GSQ	CNAASC	GTQGSA	AAA	/
	ACGGTGVV	GPT	RKPMTT	GRPVA	GQN	/
				HSIGSG	GMN	/
				TVNMER	GSW	TYGYAS
				AVSWGG	GAD	ENTGGN
				GCMLGGPV	AKH	NGSEVP
				EAPVVP	ANP	/
				AVGPFLDA	GLY	/
				NAISGVRG	GER	/
				RSQGQA	GSL	nd
				AMRVPP	GTN	EGQSLT
				PVHQKK	GGT	/
				PSRPMD	VRR	/
				VELHLP	GAK	/
				CCSFGSSG	AEP	GTMQPT
				VAHMVS	SPQ	TQQEGW
				VGCCGGGS	AGN	S
				VLRGAA	SPG	/
				YAGIVG	GDH	GSSYPT
				VVRPRH	ASF	/
				LWGGQYRS	GGN	/
				ARTGFA	GAL	/
				VVSAVP	AGM	V
				TAHNAH	GAR	TSNHAT
				LGARTFGA	nd	nd
				SYGSTA	GAT	RGHVPS
				GALSSS	GET	/
				KLVPVK	GDN	/
				AVRGLSGA	GDN	SSRTPT

* : amber stop codon
nd : not determined
/ : no insert

Table S4: Randomized loops sequences of few clones from the fd/blaS-$L_1^R L_2^8 L_3^6$ libraries selected or not for activity on 20 µg/ml ampicillin in TG1 *E. coli* strain. The library is contaminated with clones from the fd/blaS-$L_1^R L_2^6 L_3^6$ library representing up to 50% of the sequenced clones.

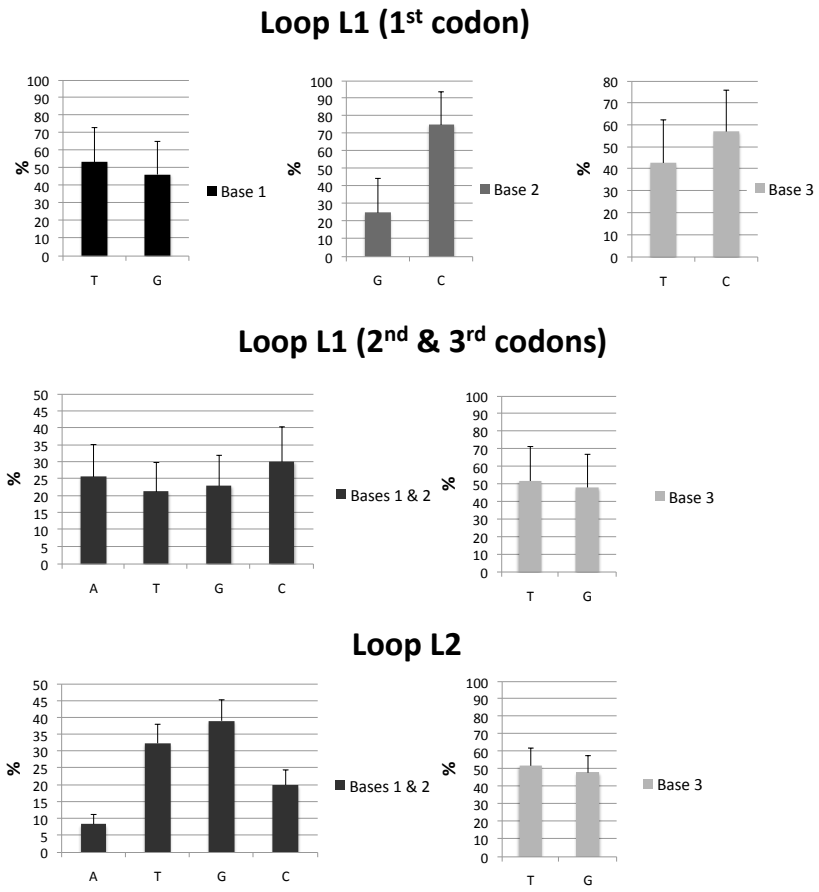


Figure S5: Observed ratio of the different bases within the codons of the degenerated loops L1 and L2 from clones unselected for activity. Error = $(2\sqrt{n})/n$, with n being the number of observations.
The observed ratio differs from the theoretical one as a bias already occurs during oligonucleotides synthesis. Not all nucleotides are equally incorporated during the synthesis, guanine and thymine being usually over-represented due to the faster coupling times of G and T phosphoramidites[3]. Bias might also occurred during libraries construction as some sequences could be preferentially amplified by PCR or better incorporated in vector and electroporated in *E. coli*. As expected, bases G and T prevailed in loops L2 and bases A and C were under-represented. The loop L1 being encoded by a reverse primer upon libraries construction logically showed a reverse, although less pronounced, trend.

Supplementary materials and methods

Oligonucleotides

(restriction sites are underlined, mutations are in bold)

(Fwd & Rev are respectively standing for forward and reverse primer)

P1: 5' ACTACTCATATGAAAAAGATTTGGCTGGC 3'

P2: 5'GGGCGCCATGATGATGATGATGCGCCGATGCGCTAAACG 3'

P3: 5'CATCATCATCATCATCATGGCGCCCACCCAGAAACGCTGGTG 3'

P4: 5' ATCATCGCTAGCCTACCAATGCTTAATCAGTGAGGC 3'

P5: 5' GTATCATTGCAGCACTGG 3'

P6: 5' TCTATGCTAGTCCATGAATAGCGGCCGCACGCCCCTCAATCCCCCAATGCTT AATCAGTGAGGC 3'

A224V: 5' GGACCACTTCTGCGCTCA**GTA**CTTCCGGCTGGCTGG 3'

H153R: 5' CTAACCGCTTTTTTG**CGC**AACATGGGGGATCATG 3'

M182T: 5' CGTGACACCACG**ACA**CCGGTAGCAATGGCAACAACG 3'

T195S: 5' CGCAAACTATTA**TCC**GGAGAACTACTTACTCTAGCTTCCCGG 3'

*Age*I-*Bbs*I mut: 5' CGTAGTTATCTACACGACCGGTAGTCAGAAGACTATGGATGAACG 3'

cassette-L3: 5' CCTGTGTAACCGGT**(NNK)$_{10}$**ATGGACA-GTCTTCGACGCT 3'

Primer cassette-L3: 5' AGCGTCGAAGACTGTCC 3'

L1: 5' ATCCGTCGAAACTCGAAGACCTCGAAGACCGAGACCC**MNNMNNRSM**GGCTCCAGAT TTATCAGCAATAAACC 3'

L2(6),(7),(8): 5' ACAAGTCCAGCTTTGAGGAAGACAAGATCAGTTG**(NNK)$_{6,7,8or9}$**CGAG TGGGTTACATCGAACTGGATCTC 3'

pHDi-Ex Fwd: 5' TTGTGTGGAATTGTGAGC 3'

pHDi-Ex Rev: 5' CCCCTTATTAGCGTTTGC 3'

fd-tet Fwd: 5' CACCTCGAAAGCAAGCTG 3'

fd-tet Rev: 5' TTCTGTATGAGGTTTTGCTAAAC 3'

With: K = (G, T); N = (A, C, G, T); M = (A, C) and R = (A, G)

Supplementary references

1. Labarbe, C. Towards the evolution of a DD-peptidase into a beta-lactamase. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2006).

2. Steiner, D., Forrer, P., Stumpp, M.T. & Pluckthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat. Biotechnol.* **24**, 823-831 (2006).

3. Ellington, A. & Pollard, J.D. Synthesis and purification of oligonucleotides. *Curr. Protoc. Mol. Biol.* **Chapter 2**, Unit2.11 (2001).

Phage display selections of allosteric β-lactamases

Vector influence on the selection output

Bruno Baudoux, Patrice Soumillion

Laboratoire de Biochimie et de Génétique Moléculaire Bactérienne (BGMB)

Institut des Sciences de la Vie, Université Catholique de Louvain,

Place Croix du Sud, 4-5 boîte 3, 1348 Louvain-La-Neuve, Belgium

Introduction

One of the current views about allostery considers that proteins exist under several conformational states and that allosteric regulation is the consequence of a ligand-induced shift in the equilibrium between those states[1]. Assuming that all proteins, except perhaps the fibrous ones, exist as a population of states, Gunaserkaran postulated that allostery is likely to be an intrinsic property of all proteins[2]. Wright suggested that, considering the complexity of the protein surface and the marginal stability of the native state, allostery establishment would only be a matter of finding the right protein's effector[3]. In other words, the emergence of a new ligand binding ability might predispose a protein to exhibit allosteric effects[4]. Liang cited the work realized in the laboratory by P. Mathonet to support that theory: one-third out of 30 mutant β-lactamases selected for nickel(II) binding exhibited allosteric regulation by the metallic ion[5].

We previously discussed the construction of a highly diverse phage display library that encodes stabilized TEM-1 β-lactamase mutants displaying complex potential binding sites (See results part I). In this second part of the project, that collection was biopanned against various targets using phage display technology, a powerful selection tool for ligand binders. Selected clones were then kinetically screened for allosteric regulation upon ligand binding.

The library was initially biopanned against nickel ions for validating the selection protocol and the library itself. Selection efficiency was compared with the one obtained with an analogous but less diverse library constructed in an unstabilized TEM-1 β-lactamase[6]. Since toxicity problems were detected with β-lactamases libraries, *in vivo* fitness was also analyzed. Subsequently, biopanning against an anti-*Coxiella burnetii* antibody[7] was carried out. Besides the fundamental interest of finding allosterically regulated clones, such engineered enzymes could be used as a mimic of the pathogenic bacteria for the development of new serology-based assays for Q fever diagnosis.

Upon libraries construction, we highlighted the importance of the starting scaffold stability in obtaining a highly diverse insertants library (See results part I). The choice of the vector could also impact the diversity since the bacterial fitness cost associated with the expression of the membrane anchored β-lactamase was stronger

with the phagemid pHDi-Ex[8] than with the phage vector fd-tet. As discussed thereafter, vector characteristics can also strongly influence the phage display selection output. The first part of this chapter addresses issues and problems encountered during phage display selection from a collection of mutant β-lactamases cloned in the phagemid pHDi-Ex. The second part discusses how we managed to overcome some of them by switching to a phage vector.

Results

Selection from the phagemid library

Construction and characterization of the library

A collection of stabilized β-lactamases randomized within the C-terminal loop L3 (p/blaS-L_3^{10}) was previously constructed in the phagemid pHDi-Ex (See results part I). Fitness cost issues encountered upon library construction made us switch to the phage vector fd-tet. Yet, phagemid library specificities could be advantageous during phage display selections. Genetic stability of phagemids under successive propagations is usually higher compared to recombinant phages[9]. Moreover, the fact that a phagemid library has to be rescued by a helper phage to produce virions lowers the display of the fusion protein. This could promote the selection of higher affinity binders by decreasing potential avidity effects that could arise in case of multivalent display[10]. The promoter pT7 coupled with the presence of an amber stop codon between the foreign gene and the gp3c gene also enables to overexpress the soluble recombinant proteins in a non-suppressive *E. coli* strain whereas the phage vector requires recloning after the selection.

A collection of stabilized TEM-1 β-lactamases randomized within the three loops was therefore also constructed in the pHDi-Ex vector.

First, the loops L1 and L2 were simultaneously mutated by PCR as for the construction of the fd-tet library **(See supplementary data, Figure S2)**. Large L1-L2 libraries (p/blaS-$L_1^R L_2^{6-9}$) were obtained in the non-suppressive strain DH12S **(Table 1)**. The percentages of active clones on ampicillin (20 µg/ml) under induction (1 mM IPTG) ranged from 4% to 45% depending on the size of the L2 insert **(Table 1)**. Forty-five clones randomly picked from the naïve or the activity selected libraries were sequenced **(See supplementary data Table S1)**. A high proportion (± 30%) of clones co-transformed with two phagemids were observed in the activity selected ones. In these clones, a predominant vector encodes a truncated β-lactamase with a stop codon or a frameshift observed in L2. The second vector encodes a full-length enzyme and most probably brings the resistance toward ampicillin.

Library	Mutations introduced	Library size	% Active mutants
p/blaS-$L_1^R L_2^6$	$G_{41}A \rightarrow (X)_{6,7,8 \text{ or } 9}$ & $G_{238}ER \rightarrow (X)_3$	2.0 x 10^8 clones	33.3%
p/blaS-$L_1^R L_2^7$		1.0 x 10^8 clones	45.0%
p/blaS-$L_1^R L_2^8$		2.0 x 10^7 clones	10.0%
p/blaS-$L_1^R L_2^9$		6.4 x 10^7 clones	3.75%

Table 1: Description of the initial libraries p/blaS-$L_1^R L_2^{6-9}$ obtained in a non-suppressive *E. coli* strain (DH12S). The percentages of active clones were estimated on ampicillin (20 µg/ml) under induction condition (1mM IPTG).

Secondly, large combinatorial L1-L2-L3 libraries (p/blaS- $L_1^R L_2^{6-9} L_3^{10}$;**Table 2**) were obtained in recombining the activity selected L1-L2 libraries (p/blaS-$L_1^R L_2^{6-8}$) with the already described L3 naïve library (p/blaS-L_3^{10}, see results part I). The DNA libraries were electroporated in a suppressive *E. coli* strain compatible with the phage display technology. The percentage of active clones on 20 µg/ml ampicillin **(Table 2)** was estimated without IPTG induction to limit the deleterious effect of the fusion protein overexpression highlighted upon the construction of the p/blaS-L_3^{10} library (See results part I).

Library	Mutations introduced	Library size	% triple mutants	% active mutants	% active triple mutants
p/blaS-$L_1^R L_2^6 L_3^{10}$	$G_{41}A \rightarrow (X)_{6,7,8 \text{ or } 9}$ $G_{238}ER \rightarrow (X)_3$ $S_{268}QAT \rightarrow (X)_{10}$	1.4 x 10^8 clones	62.1%	7.9%	3.9%
p/blaS-$L_1^R L_2^7 L_3^{10}$		3.6 x 10^8 clones	73.3%	8.1%	3.6%
p/blaS-$L_1^R L_2^8 L_3^{10}$		1.8 x 10^8 clones	78.6%	9.8%	4.6%
p/blaS-$L_1^R L_2^9 L_3^{10}$		1.4 x 10^7 clones	65.0%	10.7%	6.2%

Table 2: Descriptions of the combinatorial libraries obtained in a suppressive *E. coli* strain (XL1 blue MRF') compatible with the phage display technology. The percentages of active mutants were estimated on ampicillin (20 µg/ml) without induction. The term "triple mutants" refers to the mutants presenting the three loops simultaneously mutated. The percentages of triple mutants in the naïve and active populations were estimated through a restriction and sequencing analysis (See material & methods for details).

Clones picked at random before and after selection for activity were sequenced **(See supplementary data Table S2)**. Oddly, 16 out of 19 (84%) naïve clones displayed an amber stop codon within their L2 loop sequence. As the initial libraries (p/blaS-$L_1^R L_2^{6-9}$) were selected for activity in a non-suppressive strain, these sequences should have been discarded. They certainly were maintained because of the co-

transformed bacteria. On the contrary, only 14 active clones out of 40 (35%) contained at least one amber stop codon in L2 and/or L3. This is less compared to the 100% observed with the p/blaS-L_3^{10} library (see Results part I) but suggests that a reduced expression level due to stop suppression can still confer resistance to a significant proportion of the clones, even without IPTG induction. Besides, redundancies were observed in the L1 and L2 loops sequences of the active clones, indicative of a reduced diversity in these loops.

Biopanning against streptavidin

Phagemid library was biopanned against streptavidin, a homotetrameric protein consisting of four identical monomers (16490 Dalton) that bind biotin with one of the highest known affinities ($K_d \approx 10^{-14}$ M)[11]. The biopanning of peptides libraries against streptavidin often isolated sequences presenting a HPQ motif[12–14]. In the laboratory, the same motif was observed within the randomized loops of mutant β-lactamases recognizing streptavidin[15]. The selection of "HPQ clones" in our libraries would therefore be an easy way to check the phage display potential.

Phages from the naïve p/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library were produced for 4h under catabolic repression to minimize clone dependent bias. Inducing condition (IPTG) increased bacterial stress leading to instability of the phagemid library by providing a growth advantage to clones not expressing a g3p-fusion protein (data not shown). Since the promoter pLac is leaky[16,17], fusion proteins are produced even under catabolic repression. The minimum level of display on phages was assessed by measuring the β-lactamase activity of a phage-enzyme solution purified from soluble enzymes on an exclusion column and assuming a maximal activity of the mutants equal to the k_{cat} of the wild type enzyme (See material and methods for more details). A minimum percentage of 0.5% (1 enzyme to 200 phages) was obtained for the naïve p/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library. The value is close to the usual range of display (1 to 5%) reported in the literature for phagemid systems[18].

Purified phages were incubated for 4 h in the presence of magnetic beads coated with streptavidin. Magnetic beads were immobilized on a magnet and several washings discarded the unbound or weakly bound phages. The number of washings was gradually increased at each round of selection. Phages specifically retained on the beads were eluted by trypsin cleavage and amplified through a co-infection of

E. coli bacteria with helper phages. The trypsin degrades the β-lactamase but has no effect on the filamentous phage integrity[19,20]. Amplified phages could then enter another round of selection. Five successive rounds were carried out. The selection yield, defined as the ratio between the output and the input phages[21], increased during the first three rounds of selection with a clear jump between the second and the third one **(Figure 1)**. This is usually an indication of an effective selection. Surprisingly, the percentage of captured phages dropped at round 4 before increasing again at round 5. The last three rounds were repeated and confirmed that behavior.

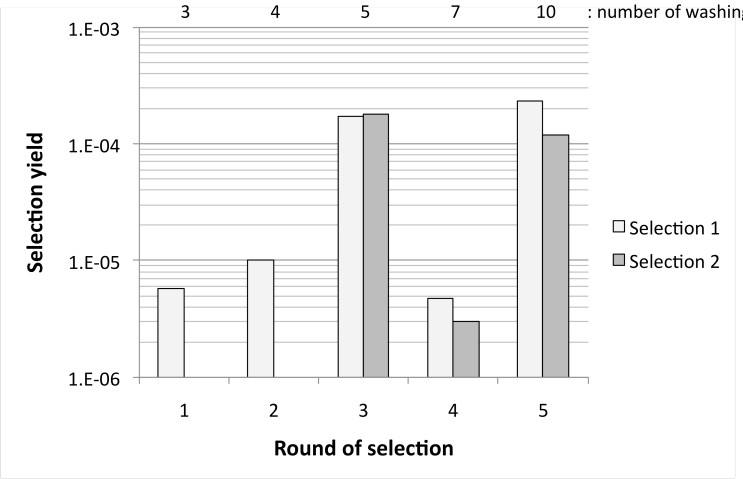


Figure 1: Selection yields obtained from the biopanning of the p/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library against streptavidin. The rounds 3 to 5 were realized in duplicate.

DNA vectors carried by the output and input phages of respectively round 3 and 4 were analyzed by DNA restriction **(Figure 2)**. Two bands at ± 3100bp and ± 900bp (the exact size depends on the length of the insert in L2) were expected. It appeared that some truncated clones took over the overall population during the amplification phase between the rounds 3 and 4. Indeed, at the end of round 3, only one clone out of ten (clone n°2 on figure 2.a) did not show the right restriction profile. In contrast, over ten analyzed clones entering round 4, at least nine did not present the right profile. One clone (n°3 on figure 2.b) had the two expected bands but also an extra band at around 4000 bp that could have arisen from a partial restriction. The others displayed large deletions within the fusion protein gene area. Every analyzed clone isolated after the fifth round of selection presented similar deletions. Sequences

analysis revealed that the deletions encompassed the promoters region and a more or less important part of the β-lactamase gene **(Figure 3)**. Two main types of deletion were observed: one ending just downstream of the loop L2 (type 1) and the other one ending just upstream of the degenerated loop L3 (type 2).

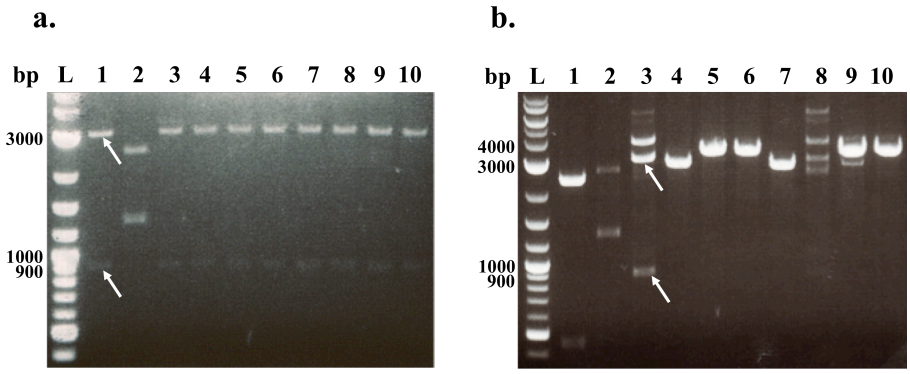


Figure 2: DNA profiles of output and input clones of respectively round 3 (a.) and round 4 (b.) of selection. DNA vectors were restricted by *Nhe*I and *Nde*I respectively cutting upstream of the signal peptide and at the junction of the β-lactamase and gp3c genes (see annexe 2, p.220). Fragment sizes expected: +/- 3100 bp and +/- 900 bp (See white arrows on both gels). L = Molecular weight ladder (2-log from Biolabs)

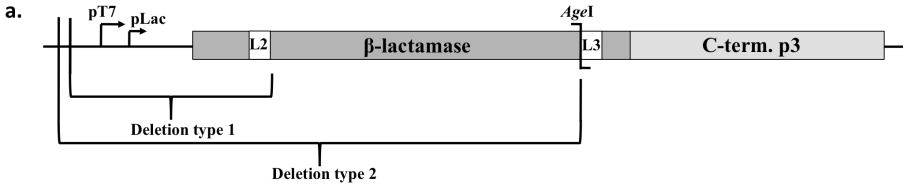


Figure 3: Schematic representation of the deletions occurring in the fusion protein construct

Few clones isolated from the third round of selection were sequenced looking for HPQ motifs. None of them presented the expected HPQ motif and consensus sequences were neither observed **(Table 3)**. Deletion clones from the fifth round were also sequenced **(Table 4)**. Because of the promoter deletion, these clones should not express any fusion protein unless cryptic promoters exist. Again, no HPQ or consensus motifs were found in these sequences.

Clones	L2	L1	L3
Rd 3.1	GGGCCRV	GQR	ACKFKVCRLI
Rd 3.2	/	/	NRGVLFRVPS
Rd 3.3	VGAFGHR	GQR	QACQSAVCCV
Rd 3.4	/	/	LLGCMMPPKL
Rd 3.5	/	/	VDKPVHRQVD
Rd 3.6	/	/	KNGGVSFLWL
Rd 3.7	/	/	VCTPASGLAL
Rd 3.8	RAVHGLG	CAN	FCLISAPSQA

Table 3: Degenerated loops sequences of clones isolated after 3 rounds of selection on streptavidin. / = No insert.

Clones	L1	L3
Δ 5.1	AN*	FQLATWVVPL
Δ 5.2	AN*	RMCGFLGDTT
Δ 5.3	AN*	FHDAKVAYAE
Δ 5.4	AN*	PCRSPRLRQF
Δ 5.5	AN*	RLSLRSVPHI
Δ 5.6	CVW	RGGALWRVMR
Δ 5.7	CVW	HQVSAYTWSV
Δ 5.8	/	RLSVWGEHEM
Δ 5.9	/	VRGILHELVG

Table 4: Sequences of the randomized loops L1 & L3 in the analyzed truncated clones of round 5. / = No insert. ✱ = Amber stop codon (TAG).

A selection for activity on 20 µg/ml ampicillin was applied to get rid of these truncated clones. Over ten active clones analyzed, only one presented the right DNA profile once restricted with the *Nhe*I and *Nde*I endonucleases. Sequencing of the other clones revealed that they carried two vectors. One vector displayed deletions similar to the previously described ones. The other one encoded a full-length β-lactamase mutant and brought the resistance to ampicillin. Full-length β-lactamase genes were selectively recovered from the round 5 population by PCR on vector DNA preparation. Primers (pHDi-Ex Fwd and Rev) hybridizing on vector sequences just upstream and downstream of the gene were used. PCR products were cloned back in the pHDi-Ex/blaΔ2 vector **(See supplementary data, Figure S3)** between the *Nhe*I and *Nde*I restriction sites. *E. coli* XL1 blue MRF' competent cells were electroporated with the ligation product and plated on agar plates supplemented with kanamycin and 1% glucose. Several clones encoding a complete β-lactamase gene were sequenced **(Table 5)** and further analyzed.

Clones	L2	L1	L3
Rd 5.1	ERMMLRVG	GER	MWLL*VFVEV
Rd 5.2	GVHQLGFSG	GER	ARDIRHPDPF
Rd 5.3	VSMWGK	GER	TG*WDVSKLW
Rd 5.4	GVGSQTT	ASG	/
Rd 5.5	FGGVGYV	CLP	*AYGQSKIG
Rd 5.6	FGGVGYV	STQ	RCVLDFPKWL
Rd 5.7	WD*HTE	GER	RQGGRMPTRI
Rd 5.8	/	GER	PYTRAVTRAT
Rd 5.9	K**QKEQR	GHR	TG*WDVSKLW
Rd 5.10	/	GER	GPRSGVFGSA
Rd 5.11	RVRGPP	GER	nd
Rd 5.12	FGGVGYV	CLP	*AYGQSKIG
Rd 5.13	/	/	ISLEGE*RAF
Rd 5.14	/	STW	KFFIPRETWA
Rd 5.15	YGSEVGGMV	AN*	ESRCSG*RSK
Rd 5.16	/	CVW	RRSMAGVCGV
Rd 5.17	/	AN*	RFKATSQYYS
Rd 5.18	VSMWGK	GER	TG*WDVSKLW
Rd 5.19	FGGVGYV	STQ	RCVLDFPKWL

Table 5: Sequences of the randomized loops of several clones selected after 5 rounds of selection on streptavidin. ✶= Amber stop codon (TAG) read as a glutamine in suppressive *E. coli* strains. nd = not determined. / = no insert. Redundancies are highlighted in color.

No HPQ motif or consensus sequence was observed. Some sequences were however redundant over the three loops or only one loop. These redundancies could arise from a bias over the PCR step used to recover the β-lactamase genes and not be related to a specific selection of these clones.

All sequenced clones were kinetically screened for allosteric regulation in the presence of streptavidin. Hydrolysis of benzylpenicillin (PenG) or nitrocefin catalyzed by the phage-enzymes was measured in the presence and absence of 1 µM streptavidin. No regulation of activity by streptavidin was observed for the screened clones. An ELISA assay for streptavidin-binding activity was then performed on phages produced from the active clones selected after five rounds of selection **(Figure 4)**. Binding activity of the individual clones Rd 5.3, Rd 5.5 and Rd 5.6 **(Table 5)** was also measured. Purified phages were loaded on a microplate coated with streptavidin. Bound phages were detected with a primary anti fd-antibody itself recognized by a secondary anti-rabbit antibody coupled to horseradish peroxidase (HRP). The signal obtained with clones having undergone five rounds of selection was two times stronger than the one of the naïve library **(Figure 4)**. These results indicate that a selection for streptaviding binding activity probably occurred. The

signals displayed by the monoclonal phage-enzymes were even stronger **(Figure 4)**. That difference could be linked to a weaker display of the round 5 phages population due to the presence of bacteria co-transformed with a truncated vector encoding no fusion protein. Similarly, the lower signals of the clones Rd 5.3 and 5.5 compared to the one of the clone Rd 5.6 could rather come from a lower display than from a weaker affinity for streptavidin. Indeed, the presence of an amber stop codon in the L3 of the first two clones is expected to decrease the display level of the β-lactamases.

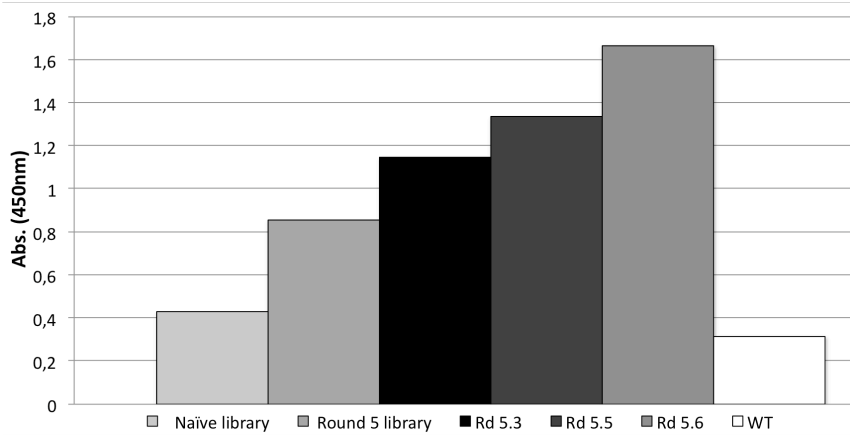


Figure 4: ELISA results of the streptavidin binding activity of round 5 phage-enzymes. Purified phage-enzymes were loaded on a microplate coated with streptavidin. The naïve library (Rd 0) and the wt TEM-1 β-lactamase served as negative controls. Bound phages were detected with a primary anti fd-antibody itself recognized by a secondary anti-rabbit antibody coupled to horseradish peroxidase (HRP). The data are derived from a single experiment with duplicates.

The free β-lactamase encoded by the clone Rd 5.6 was purified to confirm that the streptavidin recognition came from the protein itself and not from the phage-enzyme. The stabilized TEM-1 β-lactamase (blaS) served as a negative control. Both enzymes were separately incubated in the presence of magnetic beads coated with streptavidin for 2 hours. Unbound proteins were discarded by washing. Loading buffer containing SDS and DTT was added to the magnetic beads and the solution was heated for 10 minutes at 100°C for denaturating and releasing the bound proteins from the beads. The magnetic beads were immobilized on a magnet and the supernatants (= captured fraction) were loaded on a 12% acrylamide gel **(Figure 5)**. The same amount of enzymes that was incubated with the beads (= incubated

fraction) was also loaded on gel. It allowed estimating the fraction of bound β-lactamases.

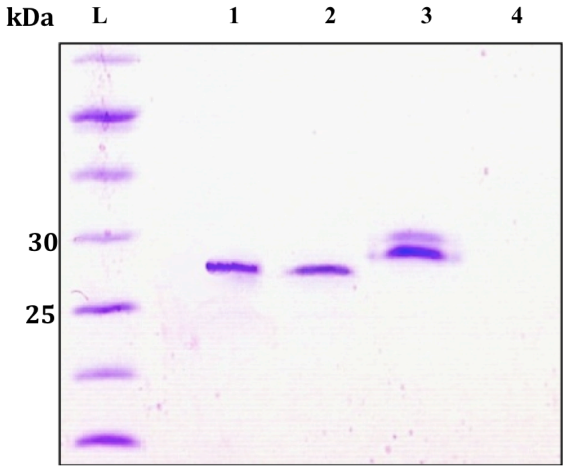


Figure 5: SDS-PAGE showing the incubated & captured fractions of the Rd 5.6 & stabilized β-lactamase (blaS) incubated with streptavidin beads. 1 = Incubated fraction of the Rd 5.6 β-lactamase. 2 = Captured fraction of the Rd 5.6 β-lactamase. 3 = Incubated fraction of the stabilized β-lactamase. 4 = Captured fraction of the stabilized β-lactamase. L = Molecular weight ladder (Unstained PageRuler Fermentas)

Protein bands of similar intensity were observed for the incubated and captured fractions of the clone Rd 5.6. No band was detected in the captured fraction of the blaS enzyme. These results suggest that the Rd 5.6 β-lactamase, unlike the stabilized one (blaS), binds to the streptavidin beads. The size difference observed on gel between the two enzymes could arise from an *in vivo* proteolytic cleavage of the C-terminal extremity (the his-tag is at the N-terminus) of the Rd 5.6 β-lactamase. The degeneration we imposed to the C-terminal loop L3 could have destabilized that region and made it more sensitive to proteases. For example, it has been shown that the insertion of the thermostable STa enterotoxin produced by enterotoxic *E. coli* at position 260 of the TEM-1 β-lactamase modified the structure of the protein so that its susceptibility to proteolysis was increased[22]. It resulted a protein deleted from its C-terminal α-helix but with a catalytic efficiency on cephalosporin only decreased by a factor 10 as compared to the wild-type β-lactamase.

No allosteric clone was isolated but streptavidin binders seem to have been selected despite that HPQ motifs did not emerge from the selection. Nevertheless, in the aim

of reducing fitness cost problems, we performed the next selections from the fd-tet phage libraries.

Selection from the fd-tet libraries

Three different phage libraries encoding TEM-1 β-lactamases mutated within three surface loops were available for phage display selections **(Table 6)**. Two were constructed in a stabilized TEM-1 β-lactamase (fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ or L_3^{10}) and differed in their randomization degree of the C-terminal loop (L3) and in their triple mutants diversity (see Results part I). The third one was constructed in an unstabilized TEM-1 β-lactamase (fd/bla- $L_1^R L_2^{C5\text{-}7C} L_3^6$) by P. Mathonet[6] and presented a disulfide bridge in the randomized N-terminal loop (L2). These three libraries were compared hereafter in term of phage display selection efficiency.

Library	Mutations introduced	Library size
fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$	$G_{41}A \rightarrow (X)_{6,7 \text{ or } 8}$; $G_{238}ER \rightarrow (X)_3$ & $T_{271} \rightarrow (X)_6$	4.4 x 10^8 clones (15.7%)
fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$	$G_{41}A \rightarrow (X)_{6,7 \text{ or } 8}$; $G_{238}ER \rightarrow (X)_3$ & $S_{268}QAT \rightarrow (X)_{10}$	3.0 x 10^8 clones (0.33%)
fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$	$G_{41}A \rightarrow C(X)_{5,6 \text{ or } 7}C$; $G_{238}ER \rightarrow (X)_3$ & $T_{271} \rightarrow (X)_6$	2.9 x 10^8 clones (0.34%)

Table 6: Descriptions of the three different phage libraries. In red are the mutations differing from one library to another. In brackets is the percentage of active mutants simultaneously mutated in the three loops.

Libraries display and fitness assessment

Before performing phage display selections, the libraries were compared in term of phage display level and in terms of *in vivo* fitness, two factors that could potentially affect selection efficiency.

A western blot using an anti-β-lactamase antibody to detect the fusion protein was performed to assess the display levels. The proteins from around 4.0 x 10^{14} phages produced from each library at 23°C for 48h were loaded on a SDS-PAGE gel after a thermal/SDS denaturation. Phages similarly produced from bacteria transformed with a fd-tet vector encoding a p3-fused wt TEM-1 β-lactamase were loaded in an equivalent amount and served as a positive control. A secondary antibody coupled to HRP recognized the anti-β-lactamase antibody. Unfortunately, no signal was obtained for the analyzed samples. A band was however detected for a second

positive control corresponding to a soluble wt TEM-1 β-lactamase down to a concentration of 100 nM. Such a concentration would represent around 1.0 x 10^{12} enzyme on gel. A display of 0.1 to 1 β-lactamase per phage could be expected from previous western blot results obtained by P. Mathonet with the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library and an anti-p3 antibody[23]. That is at least 4.0 x 10^{13} β-lactamases on gel considering the amount of phages loaded. This number is 40 times higher than the lowest amount of free TEM-1 β-lactamase detected. The experiment was repeated in varying the conditions of the phages production (overnight at 30°C or for 4h at 30°C starting from an overnight culture at 37°C) without success. The anti-β-lactamase antibody might not well recognize the β-lactamase in fusion with the p3. It would be interesting to repeat the western blot with an anti-p3 antibody. As an alternative to the western blot, we measured the β-lactamase activity of the phage-enzymes solutions on PenG. The measured initial velocities normalized to the concentration of phages (k_{obs}) are shown on Table 7**.** The values did not differ significantly between libraries, suggesting equivalent display levels. The estimation is only valid if we consider that the average β-lactamase activities (k_{cat}) from each library are similar and that the β-lactamase activity comes from the displayed enzymes and not from a possible contamination by soluble enzymes.

Libraries	k_{obs} (s^{-1})
fd/bla $L_1^R L_2^{C5\text{-}7C} L_3^6$	37.48
fd/blaS $L_1^R L_2^{6\text{-}8} L_3^6$	23.72
fd/blaS $L_1^R L_2^{6\text{-}8} L_3^{10}$	52.09

Table 7: Beta-lactamase activity (v_0) of the different libraries normalized to the concentration of phages = k_{obs}. The initial velocities (v_0) of the phage-enzymes solutions were measured following the hydrolysis of 0.5 mM PenG at 232 nm.

Fitness difference between the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ and the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ libraries was evaluated by growth competition. Both activity selected libraries were mixed in a ratio of 1:1 and co-cultured overnight (16 h) at 37°C in LB medium supplemented with tetracycline and ampicillin (20 µg/ml). Every day, a new culture was started from *E. coli* TG1 bacteria infected with phages purified from the previous one. Five successive cultures were carried out and the evolution of the ratio between the two libraries was followed over time. Vector DNA was purified from

every successive culture and the β-lactamase genes were pool-sequenced. DNA mutations introduced upon β-lactamase stabilization helped to discriminate between the two populations **(Figure 6)**. Populations' ratio was evaluated by measuring the areas of their respective base peaks. The average values were graphed **(Figure 7)**.

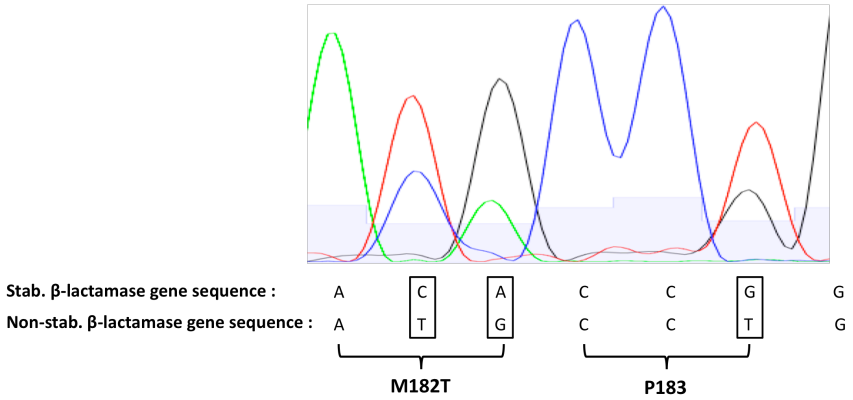


Figure 6: Sequence chromatogram obtained from pool-sequencing of DNA vector after 32h of co-culture. Region of the mutation M182T is represented. Bases modifications between the stabilized and the unstabilized β-lactamase genes are highlighted. A: Green ; T: Red ; C: Blue ; G: Black.

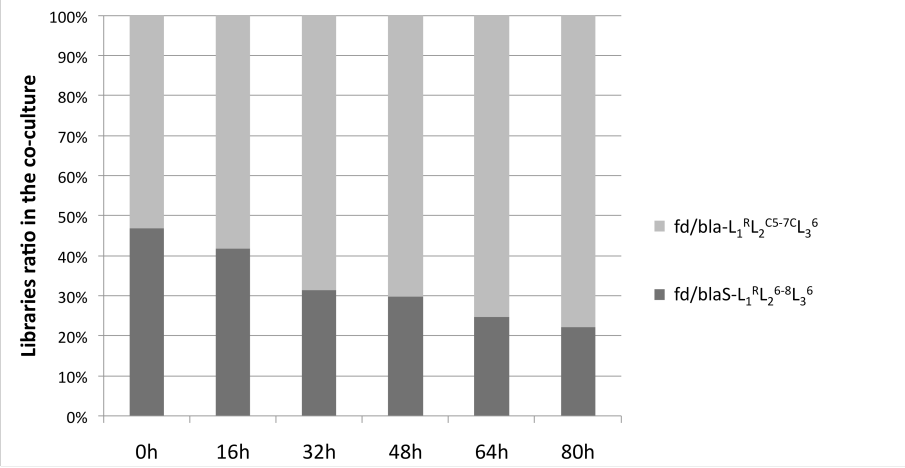


Figure 7: Growth competition assay between the fd/blaS-$L_1^R L_2^{c5\text{-}7c} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ libraries. Ratio between the two co-cultured libraries was followed over five successive cultures of 16 hours each.

A relatively small but significant fitness difference was observed between the two libraries. After five successive cultures, the ratio went to 3:1 in favor of the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library.

Biopanning against nickel(II) ion

The libraries were biopanned against nickel(II) ion as a model selection. The arising of several histidines within the randomized loops of the selected clones would confirm the selection efficiency. Histidines have indeed a good affinity for metal ions[24].

Purified phages were incubated in the presence of magnetic beads coated with nickel via metal-chelating nitrilotriacetic acid (NTA) groups. Phages with no or weak affinity for the target were washed away with a buffer containing a low concentration of imidazole (20 mM) and Tween 20 (0.05%) as a detergent. The number of washing was gradually increased at each round of selection. Increasing the concentration of imidazole to 250 mM eluted bound phages. Five rounds of selection were carried out and a jump in the selection yields was observed between the rounds 1 and 2 as well as between the rounds 3 and 4 for each of the three libraries **(Figure 8)**.

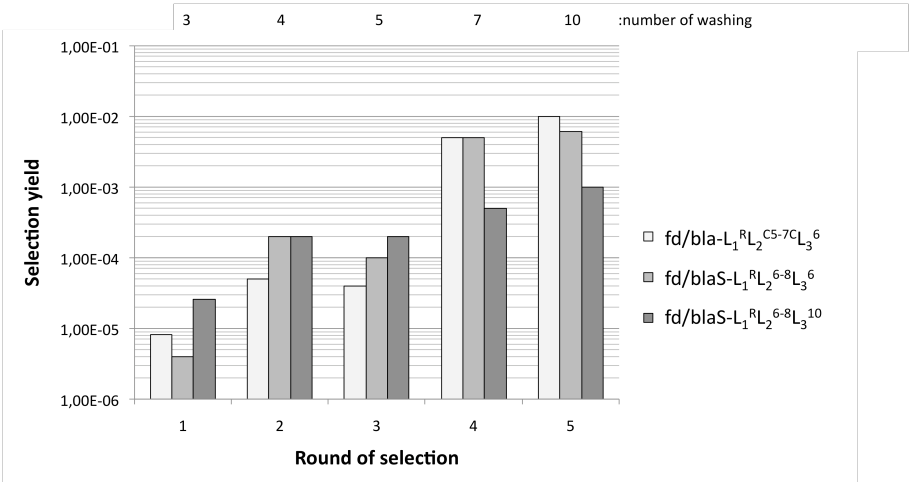


Figure 8: Semi-log graph of the selection yields measured at each round of selection on nickel with the three different fd/bla libraries.

Clones isolated after four rounds of selection were selected on ampicillin (20 µg/ml) and nine of them were sequenced for each library. Only four different clones were identified. Three were well constructed but no histidines were found within their degenerated loops **(Table 8)**. These variants might recognize other parts of the

system (agarose beads, plastic tube, specifically bound phages,…) and/or arose from an *in vivo* selection.

Library fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$			
Clones	**L2**	**L1**	**L3**
PM Tr4.1 (x5)	CWRVTRESC	GNR	/
PM Tr4.2 (x4)	CQSASVGSC	VVL	T261A

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$			
Clones	**L2**	**L1**	**L3**
L3(6) Tr4.1 (x4)	LVVSTS	AQN	/
fd/pHDi-Ex (x5)	/	GST	/

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$			
Clones	**L2**	**L1**	**L3**
fd/pHDi-Ex (x9)	/	GST	/

Table 8: Randomized loops sequences of the clones isolated after 4 rounds of selection on nickel from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{6\text{-}10}$ libraries. In brackets (x…) is the occurrence number.

Every sequenced clone from the biopanned fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library and half from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library was the same unique misconstructed variant (fd/pHDi-Ex). This clone was a contamination from another phage library. It presented a fragment of the pHDi-Ex/blaS vector (see results part I) cloned in frame with the fusion protein gene **(Figure 9)**. The pHDi-Ex fragment encodes a his-tag explaining the *in vitro* selection of the clone.

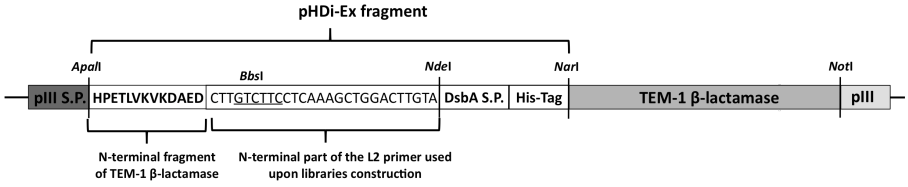


Figure 9: Schematic representation of the misconstructed clone selected on nickel from the fd/blaA stabilized $L_1L_2L_3^6$ and L_3^{10} libraries. Inserted fragment of the pHDi-Ex/blaA vector is in white. It encodes a N-terminal fragment of the TEM-1 β-lactamase, a sequence belonging to the L2 primer used upon libraries construction, the signal peptide of DsbA (DsbA S.P.) as well as a His-tag.

Consequently, clones from the second round of selection were analyzed. First, the misconstructed clones were eliminated from the selected populations by DNA restriction with *Bbs*I and re-electroporation in TG1 *E. coli* strain. This endonuclease recognizes a site in the pHDi-Ex fragment **(Figure 9)** but none in the fd-tet vector.

Forty clones (18 from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library, 16 from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library and 6 from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library) growing on 20 µg/ml ampicillin were sequenced. Half of them displayed several histidines within their randomized loops **(Table 9)**. Interestingly, two clones (12.5%) selected from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library displayed cysteines at both ends of the L2 randomized loop although this constraint was not imposed during library construction.

Library fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$			
Clones	**L2**	**L1**	**L3**
PM Tr2.1 (x6)	CVRESVRDC	G**HH**	/
PM Tr2.2 (x4)	CARVD**HH**MC	G**H**G	/
PM Tr2.3	CVAQPRSGC	GDR	**H**GG**HH**N
PM Tr2.4	CTQVRFASC	SRN	/
PM Tr2.5	CYTSQFSPC	GSS	/
PM Tr2.6	C**HH**F**H**DC	G*H	T261N
PM Tr2.7 (x3)	CWRVTRESC	GNR	/
PM Tr2.8	CVDIG**HH**AC	GG**H**	/

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$			
Clones	**L2**	**L1**	**L3**
L3(6) Tr2.1	**HH**G**HH**Q	GQR	/
L3(6) Tr2.2	CALRGPDC	AKQ	R**H**T**H**S**H**
L3(6) Tr2.3	CWVAPC	G**H**S	K**H**P**H**S**H**
L3(6) Tr2.4 (x6)	LVVSTS	AQN	/
L3(6) Tr2.5	TAISPR	GER	/
L3(6) Tr2.6	SVLSPA	SNK	/
L3(6) Tr2.6 (x2)	VSSASS	GR**H**	TR**HH**VS
L3(6) Tr2.6	QVAYSG	GPR	/
L3(6) Tr2.7	VAVVPP	GMG	/
L3(6) Tr2.7	TIQAPP	GEP	G**HHH**VT

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$			
Clones	**L2**	**L1**	**L3**
L3(10) Tr2.1 (x5)	LSAVCTD	GG**H**	SRT**H**P**H**S**HH**S
L3(10) Tr2.2	**H**T**H**T**H**Q	AD**H**	/

Table 9: Randomized loops sequences of the clones isolated after 2 rounds of selection on nickel from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ & L_3^{10} libraries. ✱: Amber stop codon. The histidines are in boldface type. Clones showing activity regulation upon nickel addition are shaded in light grey. In brackets (x…) is the occurrence number.

All clones featuring histidines in their inserts were kinetically screened for allosteric regulation in the presence of different concentrations of nickel. Initial β-lactamase activities were measured on phage as a function of nickel concentration, using PenG as substrate. One fourth of the screened clones (3/12) showed a significative activity

regulation (> 20% effect) upon nickel addition; one from each library. In a control experiment, Ni^{2+} at 2 mM concentration did not affect the β-lactamase activity of the stabilized TEM-1 β-lactamase. Two clones showed inhibition (PM Tr2.3 ; L3(6) Tr2.3) and one clone activation (L3(10) Tr2.1) **(Figure 10)**. Kinetic parameters given in Table 10 were obtained in fitting the experimental values of the initial velocity ($v_{0(obs)}$) as a function of nickel concentrations ($[Ni^{2+}]$) to equation 1 (See annexe 4, p.222):

$$v_{0(obs)} = v_0^L + \frac{Kd}{Kd + [Ni^{2+}]} \cdot (v_0^\circ - v_0^L),$$

where $v_0°$ and v_0^L are the initial velocities of respectively the free enzyme and the nickel-complexed enzyme and K_d the dissociation constant of the nickel-enzyme complex.

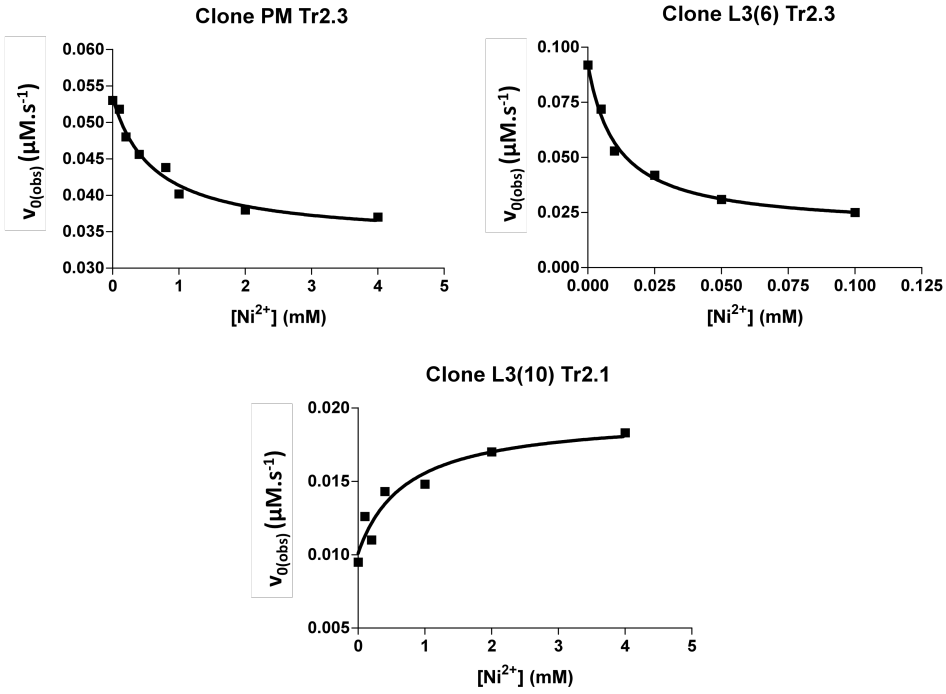


Figure 10: Kinetic analysis of phage-enzymes activity regulation by nickel. Effect of increasing [Ni^{2+}] on the initial velocity (v_0) of PenG hydrolysis by clones « PM Tr2.3 », « L3(6) Tr2.3 » & « L3(10) Tr2.1 » in 50 mM Hepes buffer, pH 7.5. The data are derived from single experiments without replicate.

Clones	$v_0°$ (µM. s^{-1})	v_0^L (µM. s^{-1})	Effect of Ni^{2+} on v_0	K_d (M)
PM Tr2.3	0.053	0.034	**0.64 x**	0.61 x 10^{-3}
L3(6) Tr2.3	0.092	0.018	**0.20 x**	0.11 x 10^{-4}
L3(10) Tr2.1	0.010	0.019	**1.90 x**	0.72 x 10^{-3}

Table 10: Kinetic parameters for the hydrolysis of 0.5 mM PenG catalyzed by clones « PM Tr2.3 », « L3(6) Tr2.3 » or « L3(10) Tr2.1 » in the presence of nickel. $v_0°$ and v_0^L are the initial velocities of respectively the free enzyme and the nickel-complexed enzyme and K_d the dissociation constant of the nickel-enzyme complex.

Estimated values of the dissociation constants were in the micromolar to millimolar range. Inhibition up to 80% and activation up to 90% of the v_0 were observed. More detailed kinetic analyzes would determine whether the regulation effect are affecting the k_{cat}, the K_m or both. Sequence analyzes revealed that all three regulated clones showed their three loops randomized. It is also worth noting that histidines were never found simultaneously in loops L2 and L3.

*Biopanning against anti-*Coxiella burnetii *antibody*

(The work described in this section was conducted in collaboration with Samira Boarbi from the Centre d'Etude et de Recherches Vétérinaires et Agrochimiques.)

Coxiella burnetii is a Gram-negative bacterium responsible for Q fever in humans and other animals. Q fever is a cause of abortions in ruminants (cattle, goat, sheep) that are a potential reservoir for the bacteria. In humans, the disease can lead to complications including pneumonia, hepatitis and myocarditis. Diagnosis of Q fever relies mainly on serology-based assays as the ELISA one[25]. Yet, antigen production is problematic, *C. burnetii* being a very infectious germ. Only qualified personnel can manipulate the pathogen in a biosafety level 3 laboratory. Moreover, *C. burnetii* is an obligate intracellular bacterial pathogen and has to be isolated from inoculated cell cultures (*e.g.* monkey kidney cells), embryonated hen yolk sacs or laboratory animals (*e.g.* mice, guinea pigs)[25–27]. Anti-idiotypic β-lactamases could replace the natural antigen in the actual serology-based assays. Moreover, selection of β-lactamases allosterically regulated by anti-*Coxiella burnetii* antibodies could result in the development of alternative homogeneous immunoassays.

Phages libraries were biopanned against a monoclonal anti-*Coxiella burnetii* antibody (received from the CERVA) recognizing the lipopolysaccharide outer core

of the bacteria[7]. The antibody seems specific for a methylated-6-deoxyhexose (virenose) that is considered as a unique biomarker of *C. burnetii*. The biotinylated antibody was immobilized on magnetic beads coated with streptavidin and incubated with the phage-enzymes collections. Phage-enzymes with no or weak affinity were discarded and bound ones were recovered by an acidic elution. Four rounds of selection were carried out. The selection yield increased between rounds 1 & 2 and reached a plateau at a relatively low value **(Figure 11)**. The observed decrease at round 4 could arise from the higher number of washing.

Several clones from rounds 3 & 4 were sequenced **(See supplementary data Tables S3 & S4)**. No consensus sequences were observed but three clones from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ libraries were redundant. It is worth mentioning that the two redundant clones from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ displayed cysteines at both extremities of the L2 insert. Every sequenced clone from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ was simultaneously mutated in the three loops as expected from the naïve library. The number of clones featuring an insert in L3 was lower for the other two libraries as these were selected for activity before starting the selections. Beta-lactamase activity of the phage-enzymes produced from the three redundant clones was measured on PenG in the presence or absence of 1 µM anti-*Coxiella burnetii* antibody. No regulation of activity was observed. The kinetic screening needs to be extended to other clones.

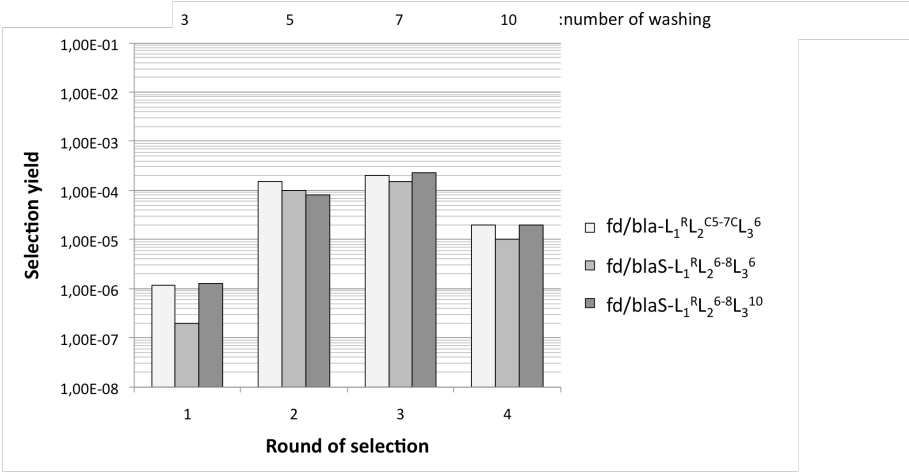


Figure 11: Semi-log graph of the selection yields measured at each round of selection on anti-*Coxiella burnetii* antibody with the three different fd/bla-$L_1L_2L_3$ libraries.

Discussion and conclusion

Phage display is a powerful biotechnology technique in selecting binders recognizing ligands of interest. Its high efficiency allows to select a clone with a frequency not exceeding $1/10^8$ in the initial collection[28]. Yet, many aspects can affect the success of a phage display selection as the type of vector used and the quality of the original library. We discussed hereafter, how the choice between a phagemid and a phage vector impacted on the libraries diversity and on the affinity selection results. We discussed as well how quality and characteristics of a phages collection encoding mutant β-lactamases could influence the properties of recognition and activity regulation of clones selected on nickel.

Vector influence on libraries quality and diversity

A collection of TEM-1 β-lactamases randomized within three contiguous surface loops (p/blaS-$L_1^R L_2^{6\text{-}9} L_3^{10}$) was constructed in the pHDi-Ex vector. This library encodes 40 times more triple mutants active on 20 µg/ml ampicillin than the analogous one constructed in the fd-tet vector (See results part I). The higher expression level of the fusion protein with the pHDi-Ex vector may counterbalance more effectively the loss of β-lactamase activity resulting from the simultaneous structural perturbations. Yet, this higher expression level in the bacteria had a cost that induced clones bias in both p/blaS-L_3^{10} and p/blaS-$L_1^R L_2^{6\text{-}9}$ initial libraries: in a suppressive strain, a bias toward clones with stop codons that should express less fusion protein was previously observed upon randomization of the C-terminal loop (L3) of the enzyme (See results part I). A bias in favor of clones that might express less TEM-1 β-lactamase was also highlighted during mutagenesis of the N-terminal loop (L2) and the loop L1. A high percentage (up to 30%) of clones co-transformed by two vectors was indeed observed under inducing condition in the non-suppressive *E. coli* strain DH12S. The main vector encoded a truncated β-lactamase, an amber stop codon in L1 or L2 or a frameshift interrupting the reading frame. The secondary vector encoded a full-length enzyme and provided resistance to ampicillin. We can fairly suppose that this copresence reduced the respective copy number of both phagemids and so the expression rate of the full-length TEM-1 β-lactamase. These clones could originate from co-electroporation events that occur

with a frequency of up to 10% with *E. coli* cells[29]. Cross-infection events could also explain such clones. Indeed, protection against overinfection process is provided by the N-terminal part of p3 which is not encoded by the pHDi-Ex phagemid[30]. Hence, a phage particle containing a phagemid vector could potentially infect a bacteria already transformed by a phagemid. These biases in the initial libraries induced an impoverishment of the sequences in the combinatorial collection. Redundancies are indeed observed in the inserts of the loops L1 and L2 of clones selected for activity. The global diversity seemed however maintained through variation of the L3 insert sequences.

Selection of streptavidin binders

No HPQ motif, known to recognize streptavidin with submicromolar affinity, could be observed within the inserts of the selected clones. Yet, the selection of streptavidin binders seemed to be confirmed by the ELISA assays. HPQ motifs were mainly described in biopanning experiments involving peptides collections[12–14]. The higher complexity of our engineered binding sites may have driven the selection toward more diverse and heterogeneous binding sites. Moreover, elution through trypsin cleavage instead of competitive elution with biotin may have allowed selecting binders than unlike HPQ clones do not recognize the biotin-binding pocket of streptavidin. On the other hand, non-competitive elution may also favor the selection of unspecific binders. Specific HPQ clones could be lost in a population of false positive clones *in vivo* selected or recognizing other components (blocking reagents, beads, plastic tubes,...) of the biopanning system. These clones could give ELISA signal if recognizing the plastic or the blocking reagent, which was the same as that used during the biopanning experiment. Vodnik reviewed several identified target-unrelated peptides described in the literature[31]. None of those peptides was found in the inserts sequences of the selected clones. Different databases also enable to screen the phage display output data for the presence of target-unrelated sequences: SAROTUP[32], PepBank[33], MimoDB[34]. No hit arose from the screening of our data. A competitive ELISA could provide more information about the clones specificity and affinity for streptavidin. The analyzed β-lactamase is incubated with a range of soluble streptavidin concentrations until the equilibrium is reached. A classical indirect ELISA on a plate coated with streptavidin then measures the

proportion of unsaturated enzyme. The concentration of streptavidin inducing a 50% decrease of the ELISA signal is considered the K_d value of the β-lactamase-streptavidin complex.

Fitness cost disturbed phage display selections

In phage display, besides the *in vitro* selection by affinity capture for a specific ligand, an *in vivo* selection occurs during the bacterial amplification steps. This *in vivo* selection is driven by the faster propagation rate of some clones[31]. Different elements can increase that rate such as a better bacterial generation time, a lower bacterial mortality rate or a higher phage infectivity and producibility. These *in vivo* selected phages can potentially take over the others during the rounds of biopanning and give rise to false positives especially if the displayed protein is capable of non-specific binding during the *in vitro* steps. Non-binders may also be recovered due to their *in vivo* advantages and be referred as propagation-related clones[31]. Such clones disturbed the phage display selection of streptavidin binders form the pHDi-Ex libraries. After three rounds of selection, input clones were essentially mutants with large deletions encompassing the promoters and the β-lactamase gene regions. These clones most likely arose from misconstruction events during libraries recombination. Indeed, several of them shared same deletion and insert in L1 but a different sequence in L3. In any case, they should allow a substantial *in vivo* selective advantage from not expressing any fusion-protein. This is consistent with the fitness cost of the β-lactamase-gp3c fusion protein expression that was highlighted during libraries construction (See Results part I). Indeed, higher growth rates for bacteria that do not encode a protein-gp3c fusion protein were already reported[35]. In the laboratory, V. Mathieu measured that the generation time of bacteria in exponential phase was two times higher for clones not expressing a circularly permuted β-lactamase-gp3c fusion protein[16]. These phages could also be more infectious as they do not display any recombinant p3 protein[36]. It appeared that the *in vitro* selection was not sufficiently stringent or efficient to counteract this *in vivo* bias. The emergence of the truncated clones over the selection would gradually increase the proportion of phages that do not display any fusion protein in the input fraction entering the *in vitro* selection for binding. This may explain the yield loss observed

in the fourth round of the *in vitro* selection during the biopanning experiment against streptavidin **(Figure 1, p.138)**.

A model allowed simulation of the selection yield evolution observed experimentally and validated the hypothesis. For simplicity, the virtual population was restricted to three main types of clones:

- Non-specific clones that do not have affinity for the target and neither an *in vivo* advantage. They predominate in the starting population.
- Specific clones that have affinity for the target but no *in vivo* advantage.
- Propagation-related clones that do not have an affinity for the target but present a strong *in vivo* advantage over the other clones.

An *in vivo* amplification factor (F_A *in vivo*) was assigned to each type of clones. Multiplying the number of clones by their respective F_A allowed simulating the evolution of the clones ratio over every *in vivo* amplification step. An *in vitro* yield factor (F_Y *in vitro*) was also assigned to them. This factor is proportional to the affinity of the clones for the target and allowed simulating the number of clones recovered over every *in vitro* affinity capture step.

We simulated an initial phage *in vivo* amplification by multiplying putative numbers of starting clones by their respective F_A. The obtained distribution was then proportionally adjusted to reach a total population of 10^{12} cfu. This population represented the input phages that were incubated with the target. The numbers of clones recovered from the *in vitro* step were determined by multiplying the numbers of input phages by their respective F_Y. A second round of selection was started from the output phages population. Four rounds were simulated and their respective selection yields, corresponding to the output/input ratio of phages, were calculated. It was possible to find realistic numbers of starting clones and values of F_A and F_Y **(Table 11)** that allowed reproducing quite well the evolution of the selection yields observed experimentally and shown in Figure 1 **(Figure 12)**. The ratio of the different class of phages in the output fraction of round 3 **(Figure 13.a)** and in the input fraction of round 4 **(Figure 13.b)** also matched to the experimental ratios **(Figure 2, p.139)**.

Clones	Number of clones	F_A *in vivo*	F_Y *in vitro*
Non-specific	1,00E+12	1	1,00E-05
Specific	3,00E+07	1	1,00E-03
Propagation-related	2,00E+03	1,00E+03	1,00E-05

Table 11: Numbers of starting clones, *in vivo* amplification factor (F_A *in vivo*) and *in vitro* yield factor (F_Y *in vitro*) of the putative population submitted to a virtual phage display selection.

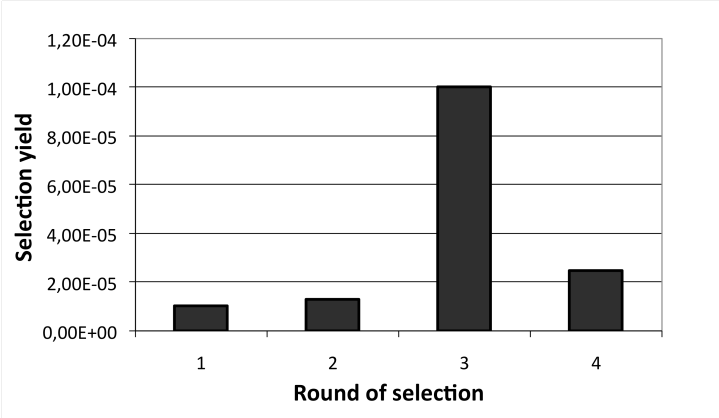


Figure 12: Evolution of the selection yield of the virtual biopanning experiment over the 4 rounds of selection. The selection yield is defined as the output/input ratio of phages.

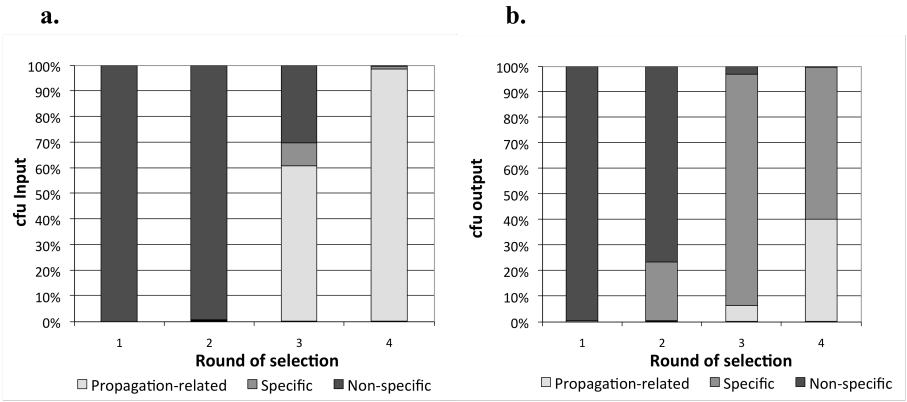


Figure 13: Evolution of the ratio of the different class of clones in the input (a) and output (b) phages populations over the 4 rounds of the virtual selection experiment.

The increase of the selection yield following the loss at round 4 would be explained through the emergence of the co-transformed clones saturating the active clones population recovered at the round 5. These particular clones would have a higher enrichment factor (= F_A x F_Y) than the propagation-related ones from being both *in*

vitro and *in vivo* selected. Indeed, a minority but functional vector would allow the clones to produce phages displaying mutant β-lactamases that possibly recognize streptavidin. On the other hand, the co-presence of a non-functional vector would decrease the copy number of the functional one and so the expression rate of the fusion protein, providing an *in vivo* advantage to these clones.

Propagation-related selection can be overcome either by increasing the stringency and the efficiency of the *in vitro* selection or by reducing the *in vivo* fitness difference between clones. Working under pLac repression condition to decrease the fitness cost of the fusion protein reduced this phenomenon without eliminating it. Catabolic repression of pLac does not prevent from high promoter leakage which represents up to 15% of the total amount of proteins expressed under induction[17,37]. LacI concentration might not be sufficient to repress every copies of the high copy pHDi-Ex vector. Beekwilder successfully reduced the propagation-related problem by decreasing the vector copy number and by replacing the pLac promoter by the *E. coli* phage shock protein (psp) promoter[35]. We obtained a similar result by switching to the phage vector fd-tet that has a lower copy number and a weaker promoter than the pHDi-Ex phagemid. The tetracycline resistance cassette of the phage vector also alters the minus strand origin of replication. This slows down the formation of the double stranded replicative form[31] as well as the phages propagation rate. Fd-tet derived libraries are thereby less prone to propagation-related selections[31]. Moreover, fd-tet mutants are less subject to the "cell killing" phenomenon in which filamentous phages kill the host as the production of viral particles is reduced[38]. The fd-tet vector is, unlike the phagemid pHDi-Ex, a fully functional phage system that does not need to be rescued by a helper phage to produce viral particles. In the absence of helper phage, every p3 produced is fused to the β-lactamase. Theoretical display of five enzymes per phage is however never achieved due to proteolytic cleavage in bacteria[39]. P. Mathonet using the exact same system obtained a display of less than one β-lactamase per phage[23]. This high display should increase the *in vitro* selection efficiency of binders.

Selection of allosteric clones regulated by nickel

Twelve β-lactamase mutants presenting several histidines within their inserts sequences arose from the biopanning of three different fd-tet libraries against nickel. Two libraries (fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ or L_3^{10}) encoded stabilized TEM-1 β-lactamases randomized within three contiguous surface loops. One (fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$) had 70 times more active triple mutants but with smaller random inserts within their C-terminal loop (L3). The third collection (fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$), analogous to the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ one, was constructed in a non-stabilized enzyme[6]. It was 70 times less diverse in term of active triple mutants, these ones displaying a disulfide bridge within their N-terminal loop (L2). Three clones out of the twelve (25*%*), one from each library, had their activity regulated by nickel. This high percentage is similar to what was observed previously by Pascale Mathonet with the unstabilized fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library (one third of nickel binding clones showed a regulation of activity), consistent with the hypothesis that new binding affinities promote allostery emergence[4]. Furthermore, the proportion of regulated clones is significatively not different (2/7) considering only the prestabilized clones. The stabilization mutations within the TEM-1 β-lactamase scaffold do not seem to impact negatively (or positively) on the selection of allosteric properties. On the other hand, the elimination of the disulfide bridge constraint in L2 through the prestabilization of the enzyme does not appear to promote the emergence of allostery. Indeed, over five binders with no disulfide bridge in L2, only one presented a regulation of activity upon nickel binding. Regarding the clones with a disulfide bond in L2, the proportion of regulated clones was 2:7. More clones should be screened to make the data statistically relevant. Clones with the three loops simultaneously mutated seem to have a greater propensity to be regulated. A compilation of data obtained in this study and that by P. Mathonet[5], shows that 8 out of 16 β-lactamases featuring the three loops randomized were allosterically regulated by nickel. In contrast, none of the 9 clones with only the loops L1 and L2 mutated displayed an activity regulation upon nickel binding. TEM-1 β-lactamase stabilization would therefore be beneficial to allostery emergence by increasing the insertional tolerance in L3 in the context of triple mutants (See results part I).

The first clone (PM Tr2.3), isolated from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library, displayed three histidines within its C-terminal loop (L3). The enzyme had an affinity for

nickel of 0.6 mM and was down regulated 1.6 times upon metal ion binding. The measured affinity implies that at least two histidines are involved in the complexation site because dissociation constants of metal ions for imidazole are in the millimolar range (pK_{D1} = 3.36 mM for Ni^{2+})[40]. Binding of nickel would impose a conformational constraint to the C-terminal loop that could be transmitted up to the active site and thereby modulate the enzyme activity.

The second clone (L3(6) Tr2.3), selected from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library, had four histidines distributed within the loops L1 and L3. Its affinity for nickel was much higher with a dissociation constant of 10 µM. It could result from a more complex interloop metal coordination site involving the four histidines. An interaction between the loops L3 and L1 (bordering the active site) upon nickel binding might induce bigger conformational changes in the vicinity of the catalytic region than those arising from intraloop interactions in the first clone. This could explain the greater enzyme activity inhibition (80% at saturation) that was observed.

The third clone, L3(10) Tr2.1, displayed four histidines in the loop L3 and one in the loop L1. It was activated 1.9x by nickel ion binding. At this stage, the available data are not sufficient to propose a clear description of the mechanism underlying this activity up-regulation. The affinity for nickel is not only depending on the number of histidines in the inserts as with 5 of them, this clone showed the lowest dissociation constant for the metallic ion (0.72 mM).

Selection against an anti-Coxiella burnetii antibody

The selection is still ongoing but no clone with good affinity for the antibody seems to have currently been found. The three redundant clones isolated after 4 rounds of selection did not exhibit activity regulation in the presence of the antibody. The binding of large proteins as antibodies in the vicinity of the active site is expected to easily generate activity regulation[41]. These clones have probably emerged as a result of *in vivo* selection or non-specific *in vitro* binding and could therefore have only a weak or no affinity for the antibody. An ELISA assay should nevertheless be performed for checking the binding properties independently of the potential regulation effects. The plateau reached at low selection yield values could be explained among others by the absence of binders in the libraries. The isolation of

mutant β-lactamases that bind an anti-bacterial LPS antibody might be intrinsically difficult.

The two redundant clones selected from the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library displayed cysteines at both extremities of the L2 insert. This characteristic was also observed in 12.5% (2/16) of the clones selected on nickel (at the round 2) from this same library. The proportion of mutant enzymes presenting a pair of cysteines in L2 was only of 3% in the biopanned library (See result part I, Tables S2-4). Although the L2 disulfide bridge is not dispensable for generating active enzymes in the prestabilized scaffold, these results indicate that it is favored in selected clones. This suggests that the stabilization gain of the disulfide bridge could increase the protein display on phage and offer an *in vitro* advantage to these clones. These phages might also have an advantage *in vivo* as discussed hereafter.

Relationship between TEM-1 β-lactamase stabilization and library fitness cost

Growth competition assay revealed a slightly better *in vivo* fitness of the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ library compared to the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ one. Stabilization of the enzyme could increase the amount of fusion protein produced in bacteria that could potentially result in a fitness cost if the anchored β-lactamase is toxic or if the phage assembly is impaired. A direct correlation between proteins half time in bacteria and their thermodynamic stability has already been shown[42]. A higher display on the p3 is also known to decrease the phages infectivity resulting in a loss of *in vivo* fitness. Yet, the similar β-lactamase activities of the different phage-enzymes populations suggest that the levels of display are equivalent or slightly lower in the stabilized libraries. A higher percentage of mutants simultaneously mutated within the three loops could also impact on the fitness of the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ library. The regions encompassing these contiguous loops could form more complex binding sites and be more prone to partial unfolding events. It has been shown that mutations and insertions in the CDR loops of immunoglobulin can generate significant destabilization of the domain[43]. Highly diverse and disordered regions might induce more unspecific interactions within the periplasmic space and interference with, among others, peptidoglycan synthesis. Indeed, extended disordered domains and their large solvent-accessible surfaces are believed to generate large intermolecular surfaces[44,45]. Moreover, intrinsically disordered (ID) regions are frequently observed

in hub proteins that play major roles in the network of protein-protein interactions by their ability to recognize multiple partners. These ID regions would promote these various interactions[45]. Region of low structural stability are also known to have a high aggregation propensity and to promote inclusion bodies formation[46,47]. Entropic constraints imposed by the presence of a disulfide bridge in the loop L2 could minimize these effects. Disulfide bridges are often observed within the aggregation-prone interfaces and would prevent from conformational rearrangements leading to misfolding and aggregation[48]. They are notably known to prevent the human islet amyloid polypeptide presence (hIAPP) and the tau protein to self-assemble into amyloid fibrils[46,49].

Conclusion

Several phage-display libraries encoding mutant β-lactamases were biopanned against different targets seeking enzymes allosterically regulated. The selection from a pHDi-Ex phagemid library was highly disturbed by the emergence of *in vivo* selected clones expressing no fusion protein. Clones bias was most probably generated by the fitness cost associated with the high expression rate of the membrane-anchored β-lactamase. Switching to the phage vector fd-tet attenuated biases likely through decreasing the expression rate of the fusion protein. Stabilized TEM-1 β-lactamase was validated as a scaffold for engineering allosteric enzymes by selecting clones regulated by nickel. The current results do not suggest that the complexification of the engineered sites and the absence of a disulfide bridge in L2, allowed by the enzyme stabilization, promote the emergence of binders and allostery. However, the higher clones diversity achieved in the stabilized TEM-1 β-lactamases library should increase the absolute number of selected clones featuring new binding abilities and allosteric regulation. The collection constructed in the stabilized enzyme seemed to generate more fitness cost to the bacteria than that constructed in the wild-type TEM-1 β-lactamase. The plausible causes of this phenomenon are under analysis. Interestingly the presence of a disulfide bridge within the L2 insert appeared to bring an *in vivo* and/or *in vitro* advantage to the clones. Selections toward targets of more interest, as an anti-*Coxiella burnetii*, are ongoing and should bring more information about the potential of the different libraries in generating binding and possibly allosteric properties.

Materials and methods

General

T4 DNA ligase and Fast Start High Fidelity polymerase were purchased from Roche. GoTaq and Expand High Fidelity polymerases were from Promega.

Restriction enzymes were purchased from New England Biolabs.

Synthetic oligonucleotides were from Eurogentec.

Ampicillin was from AppliChem, kanamycin from Gibco. Stocks of respectively 50 mg/ml and 7.5 mg/ml were prepared in milliQ water, filtered through a 0.45 µm filter (Millipore) and stored at -20°C.

Tetracycline was from AppliChem and a stock of 5 mg/ml was prepared in a solution of 50% ethanol and stored at -20°C.

Glucose (Fluka) was freshly prepared as a 20% (w/v) solution, filtered through a 0.45 µm filter (Millipore) and kept at 4°C.

IPTG (Eurogentec) was solubilized in milliQ water to a final concentration of 100 mM, filtered through a 0.45 µm filter (Millipore) and stored at -20°C.

Streptavidin was furnished by the Dr. T. Ward from the University of Basel (Switzerland).

Bacterial strains

TG1: *F'* [*traD36 proAB*$^{+}$ *lacI*^{q} *lacZΔM15*]*supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5, (*$r_K^-m_K^-$*)*

XL1 blue MRF': Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1* gyrA96 relA1 lac *[F'* proAB lacI^{q}*ZΔ*M15 *Tn*10 *(Tet*^{r}*)]*

BL21(DE3): *F*$^{-}$ *ompT gal dcm lon hsdSB(rB*$^{-}$ *mB*$^{-}$*) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*

Oligonucleotides

(Restriction sites are underlined)

(Fwd & Rev are respectively standing for forward and reverse primer)

blaA Fwd: 5' CATCATCATCAT<u>GGCGCC</u>CACCCAGAAACGCTGGTG 3'

TEM-His Rev: 5' ATCATC<u>GCTAGC</u>CTACCAATGCTTAATCAGTGAGGC 3'

pHDi-Ex Fwd: 5' TTGTGTGGAATTGTGAGC 3'

pHDi-Ex Rev: 5' CCCCTTATTAGCGTTTGC 3'

fd-Fwd: 5' GTGCACACCCAGAAACGCTGGTG 3'

fd-Rev: 5' GCCGCACGCCCCTCAATCCCCCAATGCTTAATCAGTGAGGC 3'

Comp-Fwd: 5’ GGAGGACCGAAGGAGCTAACCGC 3’

Comp-Rev: 5’ CAGCAATAAACCAGCCAGCCG 3’

L1:5’ATCCGTCGAAACTCGAAGACCTCGAAGACCGAGACCC**MNNMNNRS M**GGCTCCAGATTTATCAGCAATAAACC 3’

L2(6),(7),(8):5’ACAAGTCCAGCTTTGAGGAAGACAAGATCAGTTG**(NNK)**$_{6,7,8}$ CGAGTGGGTTACATCGAACTGGATCTC 3’

pHDi-Ex library construction

pHDi-Ex/blaAΔ2 vector construction (Supplementary data Figure S1): The gene blaAΔ2 constructed by P.Mathonet[6] was amplified using the primers blaA Fwd and TEM-His Rev and the Fast Start High Fidelity polymerase. The PCR product was purified with the “High Pure PCR Product Purification” Kit (Roche), restricted with the *Nhe*I and *Nar*I restriction enzymes and cloned in the pHDi-Ex/blaS vector (See results part I) restricted the same way. The ligation was carried out overnight at 4°C using the T4 DNA ligase and an insert/vector molecular ratio of 3:1.

p/blaS-$L_1^R L_2^{6-8}$ libraries construction (Supplementary data Figure S2): Fragment of the blaS gene was amplified from the pHDi-Ex/blaS vector using the degenerated primers pair L1 + L2 (6), (7), (8) and the Fast Start High Fidelity polymerase (See results part I). The PCR products were purified with the “High Pure PCR Product Purification kit” from Roche, restricted with *Bbs*I and purified again. These cassettes were cloned in the pHDi-Ex*/blaAΔ2* vector restricted with *Bbs*I and purified on “chromaspin 1000” (Clontech). The ligation was carried out overnight at 4°C using the T4 DNA ligase and a molecular ratio insert/vector of 5:1. The DH12S *E. coli* strain was electroporated with the ligation products and plated on large LB agar plates supplemented with 1% (w/v) glucose and kanamycin (50 µg/ml). Bacteria were harvested with 5 ml/plate of LB medium and stored at -80°C in 20% (v/v) glycerol.

In vivo selection for activity

Bacterial library glycerol stock (at -80°C) was thawed on ice. A number of bacteria representing at least 10 times the library diversity was sampled and diluted to a final

volume of 4 ml in LB medium supplemented with 1 mM IPTG. The whole was spread on large LB agar plates supplemented with 1 mM IPTG, kanamycin (50 µg/ml) and ampicillin (20 µg/ml). The plates were incubated overnight at 37°C and colonies were harvested with 5 ml/plate of fresh LB medium and stored at -80°C in 20% (v/v) of glycerol.

p/blaS-$L_1^R L_2^{6-8} L_3^{10}$ libraries construction (Supplementary data Figure S3)

(1) Phagemid DNA was recovered from the activity selected p/blaS-$L_1^R L_2^{6-8}$ libraries and from the naïve p/blaS-L_3^{10} bacterial library (See Results part I) using the "Maxiprep Kit" from Qiagen. (2) DNA was restricted with the endonucleases *Xma*I, *Apa*I, *Nhe*I for the p/blaS-$L_1^R L_2^{6-8}$ libraries and *Xma*I, *Apa*I and *Bpm*I for the p/blaS-L_3^{10} library. Restrictions generated three fragments of respectively "2228 bp, 1646 bp & 120 bp" and "2201 bp, 1766 bp & 27 bp". The restriction products were purified on a Chromaspin 1000 column (Clontech) to eliminate the salts and the small DNA fragments generated upon restriction. (3) Restricted phagemids were mixed in a molecular ratio of 1:1 and cross ligated overnight at 4°C using the T4 DNA ligase. Only the ligation of the two DNA fragments presenting the mutations within the loops could restore a full-length and recircularized vector. (4) XL1 blue MRF' *E. coli* strain was transformed with the ligation products and plated on large agar plates supplemented with 1% (w/v) glucose and kanamycin (50 µg/ml). The plates were incubated overnight at 37°C and colonies were harvested with 5 ml/plate of fresh LB medium and stored at -80°C in 20% (v/v) of glycerol. Clones transformed with a recombined vector could be differentiated from clones transformed with one of the initial vectors by analyzing the DNA profile of their purified vector restricted with *Nar*I and *Age*I (See supplementary data Figure S4). Few clones transformed with a recombined vector were sequenced.

DNA libraries electroporation and titration

See Results part I

Electrocompetent cells

See Results part I

Agarose gel and electrophoresis

The agarose gel was of 1% (w/v) agarose (Eurogentec) dissolved in TAE 1x buffer (40 mM TrisBase, 0.1142% (v/v) acetic acid, 1 mM EDTA adjusted to pH 8.0 with NaOH) supplemented with ethidium bromide (EtBr) at a final concentration of 0.5 µg/ml. DNA samples diluted in the 6x orange loading dye (Thermo scientific) were loaded into the wells and the bands were separated by electrophoresis in TAE 1X/EtBr buffer, at 120 V and 70 mA.

DNA sequencing

DNA molecules were sent to the sequencing service of the Macrogen Company.

Beta-lactamase genes in the pHDi-Ex vector were sequenced using the primers pair pHDi-Ex Fwd and Rev.

Beta-lactamase genes in the fd-tet vector were sequenced using the primers pair fd-Fwd and –Rev.

Sequence analysis

See Results part I

Phage-enzymes production and purification from the pHDi-Ex libraries

Phage-enzymes were produced from a culture of *E. coli* bacteria transformed with the DNA library of interest. The culture was started with a number of bacteria (glycerol stock) at least 10 times higher than the diversity of the considered library to recover it all. *E. coli* bacteria were grown to exponential phase at 37°C and 180 rpm shaking in LB medium supplemented with kanamycin (50 µg/ml) and 1% (w/v) glucose. The culture was then infected with R408 helper phages (Promega) at a multiplicity of infection (MOI) of 20. Infection was carried out for 1h at 37°C without shaking. The culture was centrifuged at 3000g for 10 min at 4°C and the bacterial pellet was resuspended in fresh LB medium supplemented with kanamycin and 1% (w/v) glucose. Bacteria were incubated for 4 more hours at 37°C, 180 rpm. The culture was centrifuged at 14.000g for 10 min at 4°C and the supernatant containing the phages was recovered. Phages were precipitate overnight at 4°C adding 1/5 volume of PEG/NaCl solution (20% (w/v) PEG 6000 from AppliChem, 2.5 M NaCl). Centrifugation at 14.000g for 10 min (4°C) allowed pelleting the phages. They were resolubilized in 10 ml of sterile milliQ water and filtered through

a 0.45 µm filter (Millipore). A second precipitation was carried out adding 1/5 volume of PEG/NaCl. The solution was incubated for 30 min at 4°C before being centrifuged for 10 min at 14.000g, 4°C. Phages were recovered with 1 ml of TBS (50 mM Tris, 150 mM NaCl, pH 7.5).

Phage-enzymes production and purification from the fd-tet libraries

A culture in LB medium supplemented with tetracycline (7.5 µg/ml) was started with a number of bacteria (glycerol stock) at least 10 times higher than the diversity of the considered library. The culture was incubated overnight at 37°C and 180 rpm shaking. The bacteria were discarded by centrifugation at 14.000g for 10 min at 4°C and the supernatant containing the phages was recovered. The phages were precipitated adding 1/5 volume of PEG/NaCl and incubating for 1h at 4°C. The phages were pelleted by centrifugation at 14.000g for 10 min at 4°C. The pellet was resolubilized in 10 ml of sterile milliQ water and filtered through a 0.45 µM filter (Millipore). A second precipitation was carried out adding 1/5 volume of PEG/NaCl. The solution was incubated for 30 min at 4°C before being centrifuged for 10 min at 14.000g, 4°C. Phages were recovered with 1 ml of TBS.

Phage-enzymes titration

Absorption spectroscopy at 265 nm was used to estimate the total number of phage-enzymes ($\varepsilon = 3.92 \times 10^7\ M^{-1}.cm^{-1}$).

The infectious titer of a phage-enzymes solution was estimated by infecting 900 µl of *E. coli* bacteria (*F'*) in exponential growing phase with 100 µl of different dilutions of the phages solution. The infection was carried out at 37°C for 1 h without shaking. Drops (10 µl) of each infected solution were spotted on LB plates supplemented with kanamycin and 1% (w/v) glucose for phages produced from the pHDi-Ex library and on LB plates supplemented with tetracycline (7.5 µg/ml) for phages produced from the fd-tet library. The cfu (colony forming units) were counted and the infectious titer of the phage-enzymes solution estimated.

R408 helper phages titration

A mix of 2.7 ml of TA7 medium (for 1 L: 2.5 g NaCl, 7 g agar, 10 g bactotryptone) maintained melted at 48°C and 300 µl of TG1 culture in exponential growing phase was poured on LB agar plates. Plates were dried out at room temperature (RT).

Drops (10 µl) of different dilutions of the phages solution were spotted on the plates and allowed to be absorbed. The plates were then incubated overnight at 37° C. The next morning, pfu (plaque forming units) were counted. The number of infectious phages/ml was estimated using the following formula: pfu x dilution factor x 100.

Biotinylation

The biotinylation of the anti-*Coxiella burnetii* antibody[7] was carried out by reaction with the NHS-PEG_4-Biotin reagent from Thermo Scientific. Twenty microliters of the antibody solution (2 mg/ml) in PBS (10 mM $Na_2HPO_4.2\ H_2O$, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) were added to 2.6 µl of the NHS-PEG_4-Biotin reagent (2 mM) dissolved in milliQ water. The reaction volume was brought to 100 µl with PBS. The reaction solution was incubated for 2 h on ice before being dialyzed overnight at 4°C against 1 liter of PBS using a Slide A Lyser mini dialysis unit (10.000 MWCO) from Thermo Scientific.

In vitro selection for binding affinity

On streptavidin beads

Selections for binding affinity were achieved using magnetic beads coated with streptavidin (Dynabeads MyOne Streptavidn C1, from Invitrogen).

Ten microliters of beads (1 mg/ml) were added to a siliconized 1.5 ml Eppendorf tube and washed 3 times with 1 ml of TBS. The washing procedure involved immobilizing the beads on a magnet before removing the supernatant and resuspending the beads in the washing buffer.

The beads were activated (or not if looking for streptavidin binders) by adding 150 µl of biotinylated anti-*Coxiella burnetii* antibody[7] at a concentration of 10.7 µg/ml in TBS. The activation was carried out for 30 min at RT on a rotating wheel. After placing the tube on the magnet, unbound biotinylated molecules were discarded and the beads were washed 4 times with TBS + 2% (w/v) dry milk (Biorad). Beads (with or without antibody) were transferred to a new tube previously blocked with 1 ml of TBS + 2% dry milk for 1 h on a rotating wheel. The beads were incubated for 4 h at RT (or overnight at 4°C) on a rotating wheel with 1 ml of a phage solution (+/- 5.0 x 10^{12} phages/ml) in TBS + 2% dry milk. Following the incubation, the beads were transferred to a new tube previously blocked as described above. Unbound phages

were discarded and the beads were washed 3 times with TBS + 0.1% (v/v) tween 20. The number of washings was progressively increased along the rounds of selection. At the fifth round, the washing number was of 10. The bound phages were either released by trypsin cleavage of the TEM-1 β-lactamase or by acidic elution. Trypsin cleavage was carried out by adding one milliliter of a 10^{-7} M trypsin (Promega) solution to the beads and incubating the tube for 1 h at RT on a rotating wheel. The beads were immobilized on a magnet and the supernatant containing the phages of interest sampled. Acidic elution was carried out adding 1 ml of 0.2 M glycine-HCl solution at pH 2.2 to the beads and incubating them for 10 min on a rotating wheel. Supernatant was sampled and neutralized with 150 µl of 1 M Tris-HCl, pH 9.1. The recovered phages were amplified by infecting *E. coli* (XL1 blue MRF' for the pHDi-Ex libraries or TG1 for the fd-tet libraries) exponentially growing cells, purified and used for another round of selection.

On nickel beads

The selection on nickel followed the same protocol, the streptavidin beads being replaced by 50 µl (5% suspension) of Ni-NTA agarose magnetic beads (Qiagen). The washing buffer contained 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.05% (v/v) Tween 20 (adjusted to pH 8.0 with NaOH). The bound phages were eluted for 20 min with 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 0.05% (v/v) Tween 20, adjusted to pH 8.0 with NaOH).

Phages purification on an exclusion column

A phages preparation was passed through an exclusion column to isolate the phage-enzymes from the free TEM-1 β-lactamases.

Around 10^{12} phages produced from a culture of the p/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$ library and purified by PEG/NaCl precipitations were applied on a Superdex 200 HR 30/10 column from GE Healthcare. The column was previously washed as described in the furnished protocol and equilibrated with 2 column volumes of equilibration buffer (50 mM Na-phosphate, 150 mM NaCl, pH 7.5) at a flow rate of 0.5 ml/min. A 250 µl volume of the phages solution was injected and elution was carried out at a flow rate of 0.3 ml/min with the equilibration buffer. Fractions of 2 ml were collected and analyzed during the whole elution step. The initial β-lactamase activity and the titer of infectious phages in the different collected fractions were measured

(See corresponding protocols). The kinetic assays were done at a substrate concentration (0.5 mM PenG) far above the K_m (20 µM) of the TEM-1 β-lactamase[50]. Therefore the measured initial velocity was considered as an approximation of the V_{max} and the enzyme concentration $[E]_0$ could be determined applying the following formula: $[E]_0 = V_{max}/k_{cat}$. A maximal k_{cat} of 1500 sec^{-1} was approximated for the library. That value corresponds to the k_{cat} of the wt TEM-1 β-lactamase displayed on phage[6].

Production and purification of the Rd 5.6 and the stabilized blaS TEM-1 β-lactamases

The enzymes were produced in the periplasmic space of *E. coli* BL21 (DE3) strain. The pHDi-Ex vector encoding the β-lactamase gene in fusion with the T7 promoter was electroporated in the bacteria. The electroporation product (1 ml) was inoculated into 500 ml of LB medium supplemented with kanamycin (50 µg/ml) and glucose 1% (w/v). The culture was incubated at 37°C, 180 rpm until the OD_{600} reached a value of 0.6. The culture was centrifuged for 10 min at 4400g, 4°C and the bacterial pellet was resuspended in fresh medium supplemented with IPTG at a final concentration of 1 mM to induce the protein expression. The expression was carried out for 4 h at 30°C, 180 rpm and the protein was recovered via a periplasmic extraction: (1) the culture was centrifuged for 10 min at 4400g and the bacterial pellet was resuspended in 33 ml of resuspension buffer (30 mM Tris-HCl, 20% glucose (w/v), 10 mM EDTA, adjusted to pH 8.0 with NaOH); (2) 10 min incubation step at room temperature under shaking was followed by a centrifugation at 4400g for 10 min, 4°C; (3) the harvested cells were resuspended in 50 ml of cold 5 mM $MgSO_4$ solution and incubated for 10 min at 4°C under shaking; (4) the periplasmic extract (supernatant) was recovered by a final centrifugation at 15.000g for 15 min at 4°C and dialyzed overnight at 4°C against buffer A (20 mM HEPES, 0.5 M NaCl, 10 mM imidazole, pH 8.0) using a Spectra por membrane (Mr cut-off 10.000, Spectrum Laboratories). The dialysate was passed through a HiTrap Chelating column (Amerscham Biosciences) previously activated with nickel as described in the column protocol. The purification was carried out with a ÅKTA prime apparatus (GE HealthCare). The column was washed with 10 column volumes

of buffer A and eluted with a 10-500 mM imidazole gradient in buffer A. The fractions were analyzed on SDS-PAGE.

SDS-PAGE

The separating gel (for 10 ml: 0.375 M Tris-Base adjusted to pH 8.8 with NaOH, 10% (v/v) acrylamide, 0.1% (w/v) SDS, 1% (w/v) ammonium persulfate and 4 ul of Temed added just prior polymerization) was poured between the two glasses of the cast. After polymerization, the stacking gel was poured on the separating gel (for 5 ml: 0.375 M Tris-Base adjusted to pH 6.8 with HCl, 10% (v/v) acrylamide, 1% (w/v) SDS, 0.1% (w/v) ammonium persulfate and 5 µl of Temed added just before the pouring). Protein samples were mixed with the 2 X loading buffer (0.1 M Tris-HCl pH 6.8, 4% (w/v) SDS, 3 mM bromophenol blue, 20% (v/v) glycerol and 0.1 M DTT) and heated for 10 min at 100°C. The samples were loaded on the gel and the migration was carried out in the electrophoresis buffer (10 X: 25 mM Tris-Base, 250 mM glycine, 1% (w/v) SDS, adjusted to pH 8.3 with NaOH) at 200 V, 40 mA. The protein bands were revealed by Coomassie blue staining. The gel was lain in the staining solution (for 500 ml: 1 g Brilliant Blue, 180 ml MeOH, 40 ml AcOH, 380 ml milliQ water) and heated in microwave till the solution started to boil. The gel was washed 2 times with the distaining solution (30 ml MeOH, 25 ml AcOH, milliQ water to 1 L) and heated the same way in that solution. The distaining solution was renewed and the whole was incubated overnight at room temperature under shaking.

ELISA

The wells of a Streptawell plate from Roche were activated (or not if looking for streptavidin binders) adding 100 µl of the biotinylated anti-*Coxiella burnetii* antibody[7] at a concentration of 4 µg/ml in TBS + 2% (w/v) dry milk. The coating was carried out at RT for 1 h under a moderate shaking. Unbound molecules were discarded and the wells were washed 5 times with 300 µl of TBS + 2% dry milk. A 100 µl volume of the phages solution at an OD_{265} 0.15 was loaded into the wells and incubation was carried out for 2 h at RT under moderate shaking. The wells were washed 5 times with TBS + 0.1% (v/v) Tween 20 and one time with TBS. A volume of 100 µl of the anti-fd antibody (Sigma Aldrich) diluted 1000 times in TBS + 0.1% (w/v) BSA was loaded into the wells and an incubation of 1 h at RT was observed.

Unbound antibodies were discarded and the wells were washed 4 times with TBS + 0.2% Tween 20 and 1 time with TBS. Hundred microliters of the secondary antibody anti-rabbit IgG-Peroxidase (Sigma Aldrich), diluted 8000 times in TBS + 0.1% BSA, were added to the wells. An incubation of 1 h at RT under a moderate shaking was observed. The excess of antibodies was discarded and the wells were washed 4 times with TBS + 0.2% Tween 20 and 1 time with TBS. The peroxidase substrate, 3-3', 5-5'-tetramethylbenidine (Sigma Aldrich), was added at a volume of 200 µl and the plate was incubated in the dark for 20 minutes. The reaction was stopped adding 100 µl of H_2SO_4 1 M. The absorbance at 450 nm was measured using the microplate reader Spectracount from Packard Company.

Western blot

Phage-enzymes were produced from the libraries (or from TG1 transformed with the fd/bla-wt vector[23]) for 48 h at 23°C, 180 rpm and purified by PEG/NaCl precipitations. Samples (10 µl) containing around 10^{14} phages were mixed with 10 µl of 2x loading buffer (0.1 M Tris-HCl pH 6.8, 4% (w/v) SDS, 3 mM bromophenol blue, 20% (v/v) glycerol and 0.1 M DTT) and heated for 10 min at 100°C. The samples, the free stabilized TEM-1 β-lactamase (blaS) at different concentrations (= positive control) and the molecular weight ladder (Kaleidoscope Prestained Standards, Bio-Rad) were loaded on a 7.5% Mini-PROTEAN® TGX™ Precast Gel from Bio-Rad. The migration was carried out at 200V, 40mA in Tris-glycine buffer (25 mM Tris, 250 mM glycine pH 8.3, 0.1% (w/v) SDS). The gel was incubated for 10 min in the transfer buffer (39 mM glycine, 48 mM Tris, 20% (v/v) methanol). Proteins were then transferred on a PVDF membrane (Bio-Rad) using the "Trans-Blot SD Semi-Dry Transfer Cell" from Bio-Rad. The membrane was blocked with a MTBS solution (TBS + 5% (w/v) dry milk) for 1h and washed 2 times with TTBS (TBS + 0.05% (v/v) tween 20). It was then incubated for 1h with a polyclonal antibody serum raised against a wild-type TEM-1 β-lactamase in rabbit (Eurogentec) (diluted 500 times) in a MTBS solution and washed 3 x 10 min with TTBS. The membrane was incubated for 1h with a bovine anti-rabbit antibody coupled to horseradish peroxidase (Santa Cruz Biotechnology) diluted 2500 times in a MTBS solution. It was washed 3 x 10 min with TTBS. Revelation was carried out with the Immun-Star™ HRP Protein Detection System kit from Bio-Rad.

Beta-lactamase activity measurement

The β-lactamase activity was measured following the hydrolysis of 0.5 mM benzylpenicillin (Sigma Aldrich) at 232nm ($\Delta\varepsilon_{substrate\text{-}product}$ = 1042 $M^{-1}.cm^{-1}$) in 50 mM Hepes buffer pH 7.5, 25°C in the presence of different concentrations of the ligand of interest. The absorbance decrease as a function of time was followed using the Cary Varian 3-BIO spectrophotometer.

The β-lactamase activity was also measured following the hydrolysis of 30 µM nitrocefin at 482nm ($\Delta\varepsilon_{product\text{-}substrate}$ = 15900 $M^{-1}.cm^{-1}$) in 50 mM Hepes buffer, pH 7.5.

Curves fitting were carried out with the GraphPad Prism software.

Streptavidin binding activity measurement on beads

Eighteen microliters of purified β-lactamase at a concentration of 1.3 µM were added to 20 µl of Dynabeads MyOne Streptavidn C1 from Invitrogen (1 mg/ml). The volume was brought to 200 µl with milliQ water and the whole was incubated for 2 h on a rotating wheel at room temperature. The beads were washed 5 times with TBS + 0.1% tween and 1 time with TBS only. The beads were resuspended in 24 µl of loading buffer 1x (50 mM Tris HCl pH 6.8, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue, 10% (v/v) glycerol, 50 µM DTT) and heated for 10 min at 100°C. The magnetic beads were immobilized on a magnet and the supernatant was recovered. The solution was then loaded on a 12% acrylamide gel. Migration was carried out at 200 V and 40 mA. Protein bands were detected by coomassie blue staining.

Libraries competition assay

Glycerol stocks of the activity-selected libraries were mixed in a ratio of 1:1 in term of living bacteria. About 5.0 x 10^8 bacteria of the mix were inoculated in 250 ml of LB medium supplemented with tetracycline (7.5 µg/ml) and ampicillin (20 µg/ml). The co-culture was incubated for 16h at 37°C, 180 rpm. Phages were purified (see corresponding protocol) and used to infect 10 ml of a TG1 culture in exponential growing phase. Infected bacteria were used to start a new culture. Five successive cultures were carried out on this basis. Every 16 hours, 5 ml of the culture were sampled and vector DNA extracted using the « High Pure Plasmid Isolation Kit »

from Roche. Beta-lactamase gene fragments between loops L2 and L1 were amplified by PCR using the GoTaq polymerase (Promega) and the primers pair Comp-Fwd and Rev. PCR products were sent to sequencing using the same primers. Sequence chromatograms of the analyzed bases were enlarged and printed on an A4 sheet. Peaks were cut off and their respective area estimated by weighing the paper piece on a precision balance (Mettler AJ100).

Acknowledgements

We thank Prof. P. Minard from the Université de Paris-Sud, France for providing the phagemid vector pHDi-Ex and Prof. T. Ward from the University of Basel, Switzerland for providing the streptavidin.

References

1. Swain, J.F. & Gierasch, L.M. The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.* **16**, 102-108 (2006).
2. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433-443 (2004).
3. Wright, C.M., Heins, R.A. & Ostermeier, M. As easy as flipping a switch? *Curr. Opin. in Chem. Biol.* **11**, 342-346 (2007).
4. Liang, J., Kim, J.R., Boock, J.T., Mansell, T.J. & Ostermeier, M. Ligand binding and allostery can emerge simultaneously. *Protein Sci.* **16**, 929-37 (2007).
5. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).
6. Mathonet, P., Deherve, J., Soumillion, P. & Fastrez, J. Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. *Protein Sci.* **15**, 2323-2334 (2006).
7. Palkovicova, K., Ihnatko, R., Vadovic, P., Betinova, E., Skultety, L., Frangoulidis, D. & Toman, R. A monoclonal antibody specific for a unique biomarker, virenose, in a lipopolysaccharide of Coxiella burnetii. *Clin. Microbiol. Infect.* **15**, 183-184 (2009).
8. Heyd, B., Pecorari, F., Collinet, B., Adjadj, E., Desmadril, M. & Minard, P. In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry* **42**, 5674-5683 (2003).
9. Qi, H., Lu, H., Qiu, H.-J., Petrenko, V. & Liu, A. Phagemid Vectors for Phage Display: Properties, Characteristics and Construction. *J. Mol. Biol.* **417**, 129-143 (2012).
10. O'Connell, D., Becerril, B., Roy-Burman, A., Daws, M. & Marks, J.D. Phage versus Phagemid Libraries for Generation of Human Monoclonal Antibodies. *J. Mol. Biol.* **321**, 49-56 (2002).
11. Freitag, S., Le Trong, I., Klumb, L., Stayton, P.S. & Stenkamp, R.E. Structural studies of the streptavidin binding loop. *Protein Sci* **6**, 1157-1166 (1997).
12. Devlin, J., Panganiban, L. & Devlin, P. Random peptide libraries: a source of specific protein binding molecules. *Science* **249**, 404 -406 (1990).
13. Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. & Knapp, R.J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **354**, 82-84 (1991).
14. Giebel, L.B., Cass, R., Milligan, D.L., Young, D., Arze, R. & Johnson, C. Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry* **34**, 15430-15435 (1995).

15. Legendre, D., Vucic, B., Hougardy, V., Girboux, A.-L., Henrioul, C., Haute, J.V., Soumillion, P. & Fastrez, J. TEM-1 β-lactamase as a scaffold for protein recognition and assay. *Protein Sci.* **11**, 1506–1518 (2002).

16. Mathieu, V., Fastrez, J. & Soumillion, P. Engineering allosteric regulation into the hinge region of a circularly permuted TEM-1 β-lactamase. *Prot. Eng. Des. Sel.* **23**, 699 -709 (2010).

17. Krebber, A., Burmester, J. & Plückthun, A. Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p. *Gene* **178**, 71-74 (1996).

18. Fernandez-Gacio, A., Uguen, M. & Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.* **21**, 408-14 (2003).

19. Løset, G.A., Kristinsson, S.G. & Sandlie, I. Reliable titration of filamentous bacteriophages independent of pIII fusion moiety and genome size by using trypsin to restore wild-type pIII phenotype. *BioTechniques* **44**, 551-552, 554 (2008).

20. Goletz, S., Christensen, P.A., Kristensen, P., Blohm, D., Tomlinson, I., Winter, G. & Karsten, U. Selection of large diversities of antiidiotypic antibody fragments by phage display. *J. Mol. Biol.* **315**, 1087-1097 (2002).

21. Smith, G. & Petrenko, V. Phage Display. *Chem. Rev.* **97**, 391-410 (1997).

22. Ruth, N., Quinting, B., Mainil, J., Hallet, B., Frère, J.-M., Huygen, K. & Galleni, M. Creating hybrid proteins by insertion of exogenous peptides into permissive sites of a class A beta-lactamase. *FEBS J.* **275**, 5150-5160 (2008).

23. Mathonet, P. Création d'un site allostérique dans la beta-lactamase TEM-1 par évolution dirigée. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2004).

24. Hengen, P. Purification of His-Tag fusion proteins from Escherichia coli. *Trends Biochem. Sci.* **20**, 285-286 (1995).

25. Fournier, P.-E., Marrie, T.J. & Raoult, D. Diagnosis of Q Fever. *J Clin Microbiol* **36**, 1823-1834 (1998).

26. Williams, J.C., Thomas, L.A. & Peacock, M.G. Humoral immune response to Q fever: enzyme-linked immunosorbent assay antibody response to Coxiella burnetii in experimentally infected guinea pigs. *J Clin Microbiol* **24**, 935-939 (1986).

27. Ormsbee, R.A. The growth of Coxiella burnetii in embryonated eggs. *J Bacteriol* **63**, 73-86 (1952).

28. Souriau, C., Hua, T.D., Lefranc, M.-P. & Weill, M. Présentation à la surface de phages filamenteux : les multiples applications du phage display. *Médecine sciences* **14**, 300-309 (1998).

29. Goldsmith, M., Kiss, C., Bradbury, A.R.M. & Tawfik, D.S. Avoiding and controlling double transformation artifacts. *Prot. Eng. Des. Sel.* **20**, 315-318 (2007).

30. Barbas, C.F., Burton, D.R., Scott, J.K. & Silverman, G.J. *Phage Display : A Laboratory Manual*.

31. Vodnik, M., Zager, U., Strukelj, B. & Lunder, M. Phage Display: Selecting Straws Instead of a Needle from a Haystack. *Molecules* **16**, 790-817 (2011).

32. Huang, J., Ru, B., Li, S., Lin, H. & Guo, F.-B. SAROTUP: Scanner and Reporter of Target-Unrelated Peptides. *J. Biomed. Biotechnol.* **2010**, (2010).

33. Shtatland, T., Guettler, D., Kossodo, M., Pivovarov, M. & Weissleder, R. PepBank--a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics* **8**, 280 (2007).

34. Ru, B., Huang, J., Dai, P., Li, S., Xia, Z., Ding, H., Lin, H., Guo, F. & Wang, X. MimoDB: a new repository for mimotope data derived from phage display technology. *Molecules* **15**, 8279-8288 (2010).

35. Beekwilder, J., Rakonjac, J., Jongsma, M. & Bosch, D. A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. *Gene* **228**, 23-31 (1999).

36. Kehoe, J.W. & Kay, B.K. Filamentous phage display in the new millennium. *Chem. Rev.* **105**, 4056-72 (2005).

37. Mathieu, V. Engineering allosteric regulation into a circulalry permuted TEM-1 beta-lactamase. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2010).

38. Smith, G.P. & Fernández, A.-M. Effect of DNA copy number on genetic stability of phage-displayed peptides. *BioTechniques* **36**, 610-614 (2004).

39. Chasteen, L., Ayriss, J., Pavlik, P. & Bradbury, A.R.M. Eliminating helper phage from phage display. *Nucl. Acids Res.* **34**, e145 (2006).

40. Sillén, L. Stability constants of metal-ion complexes. *J. Chem. Educ.* **42**, 521 (1965).

41. Legendre, D., Soumillion, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67-72 (1999).

42. Kwon, W.S., Da Silva, N.A. & Kellis, J.T. Relationship between thermal stability, degradation rate and expressio yield of barnase variants in the periplasm of Escherichia coli. *Protein Eng.* **9**, 1197-1202 (1996).

43. Helms, L.R. & Wetzel, R. Destabilizing loop swaps in the CDRs of an immunoglobulin VL domain. *Protein Sci* **4**, 2073-2081 (1995).

44. Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D. & Nussinov, R. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* **28**, 81-85 (2003).

45. Keith Dunker, M.S.C. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129-48 (2005).

46. Tartaglia, G.G., Pawar, A.P., Campioni, S., Dobson, C.M., Chiti, F. & Vendruscolo, M. Prediction of Aggregation-Prone Regions in Structured Proteins. *J. Mol. Biol.* **380**, 425-436 (2008).

47. Chan, W., Helms, L.R., Brooks, I., Lee, G., Ngola, S., McNulty, D., Maleeff, B., Hensley, P. & Wetzel, R. Mutational effects on inclusion body formation in the periplasmic expression of the immunoglobulin VL domain REI. *Folding and Design* **1**, 77-89 (1996).

48. Pechmann, S., Levy, E.D., Tartaglia, G.G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc Natl Acad Sci U S A* **106**, 10159-10164 (2009).

49. Walker, S., Ullman, O. & Stultz, C.M. Using Intramolecular Disulfide Bonds in Tau Protein to Deduce Structural Features of Aggregation-resistant Conformations. *J Biol Chem* **287**, 9591-9600 (2012).

50. Raquet, X., Lamotte-Brasseur, J., Fonze, E., Goussard, S., Courvalin, P. & Frère, J.M. TEM beta-lactamase mutants hydrolysing third-generation cephalosporins. A kinetic and molecular modelling analysis. *J. Mol. Biol.* **244**, 625-639

Supplementary data

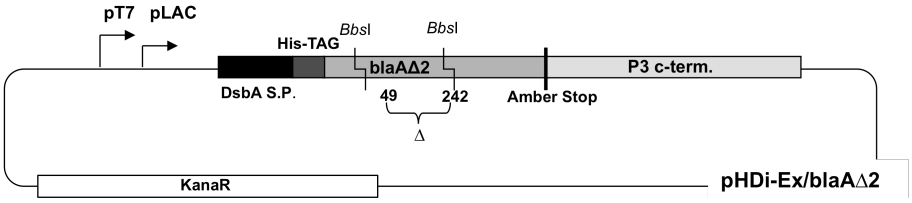


Figure S1: The pHDi-Ex/blAΔ2 vector encodes a truncated β-lactamase (blaAΔ2) lacking the residues 49 to 242. The truncated gene sequence is replaced by a *Bbs*I restriction site. A second *Bbs*I site is located just upstream the L2 position. There is no frameshift but the expressed truncated β-lactamase is inactive. This phenotype allows creating libraries without any contamination of active wt β-lactamase.

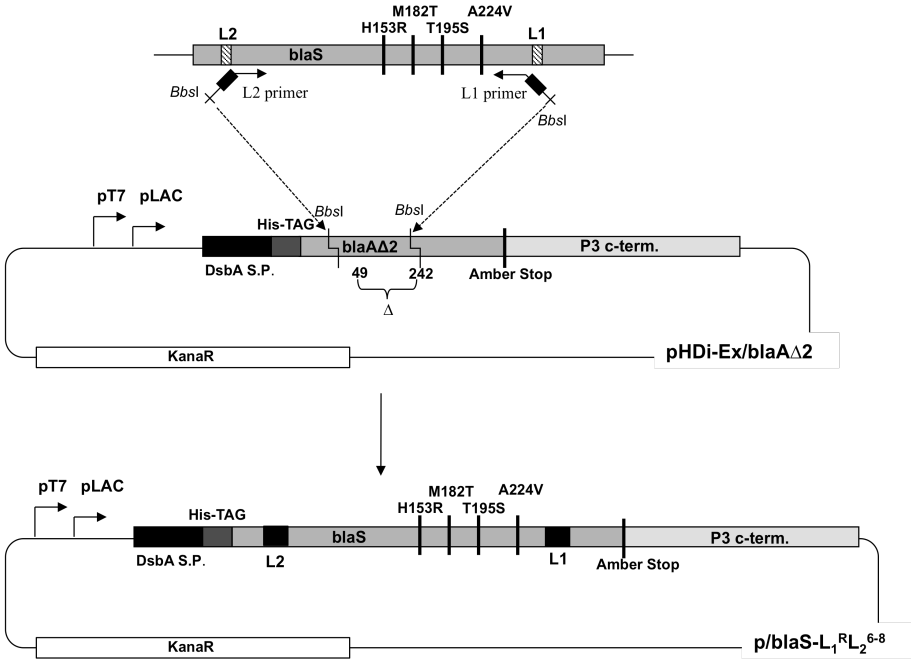


Figure S2: Randomization of loops L1 and L2. A fragment of the stabilized β-lactamase gene (blaS) was amplified using degenerated primers. The forward primer (L2) encodes 6, 7 or 8 degenerated codons of type NNK and a BbsI restriction site at its 5' extremity. The reverse primer (L1) encodes three degenerated codons (MMN-MMN-RSM) and a BbsI restriction site at its 5' extremity. The coding strand gives KSY-NNK-NNK for the degeneration. The KSY sequence only codes for four amino acids (alanine, cysteine, glycine and serine). The forward and reverse primers hybridize respectively just downstream of the residues 41-42 of the loop 2 and upstream of the residue 238 of the loop 1. The PCR product was cloned between the BbsI sites of the pHDi-Ex/blaAΔ2 vector. Experimental details are described in materials and methods section.

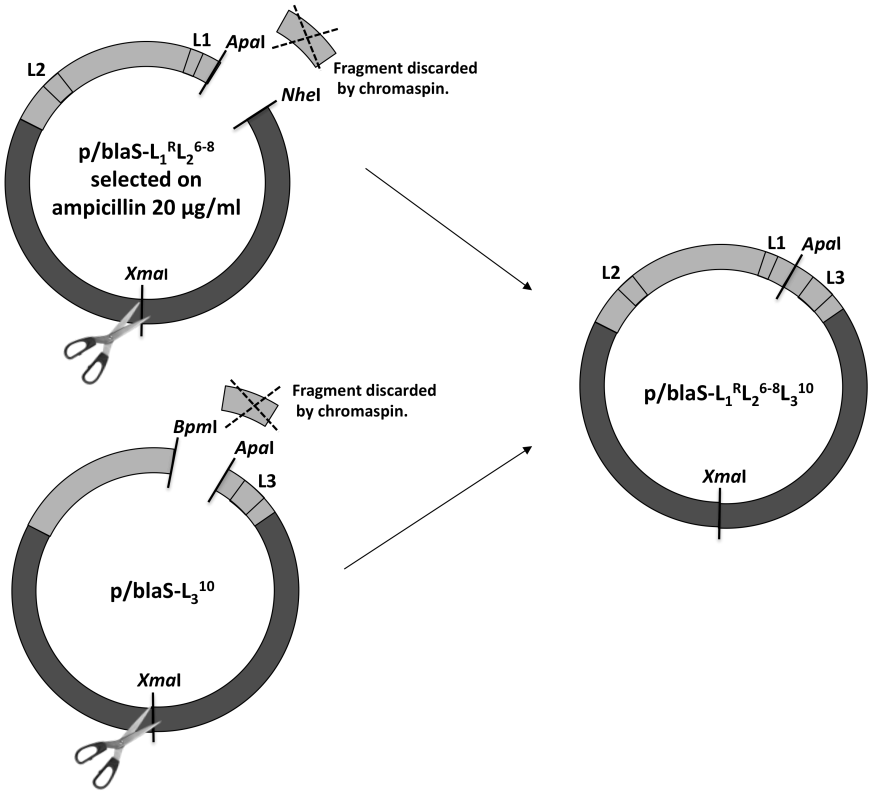


Figure S3: Schematic representation of the strategy used during the combinatorial library construction. DNA vectors of the p/blaS-$L_1^R L_2^{6-8}$ and L_3^{10} libraries were restricited with *Apa*I, *Xma*I and respectively *Nhe*I or *Bpm*I. The small DNA fragments that were generated upon restrictions were discarded through chromaspin. This step prevented from recirculization of the parental vectors. Restricted vectors were cross-ligated. Only the ligation of the two DNA fragments carrying the mutations within the loops could restore a full-length and recircularized vector. Recombined vectors were transformed in a suppressive and F' strain required for the phage display application.

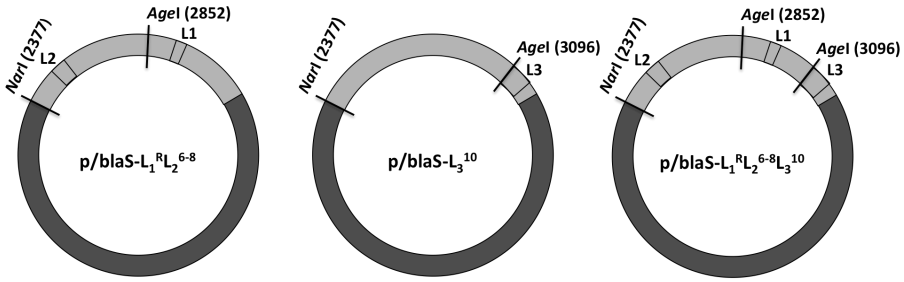


Figure S4: Schematic representation of the initial vectors (p/blaS-$L_1^R L_2^{6-8}$ and p/blaS-L_3^{10}) and of the recombined vector (p/blaS-$L_1^R L_2^{6-8} L_3^{10}$) with the positions of the restriction sites *Nar*I and *Age*I represented. Recombined vectors could be differentiated from the initial vectors by restriction with *Nar*I and *Age*I. Indeed, the initial vectors p/blaS-$L_1^R L_2^{6-8}$ and p/blaS-L_3^{10} present a unique *Age*I site respectively at the position 2852 and 3096. Hence, the recombined vector will own the 2 sites. A restriction with the *Age*I and *Nar*I enzymes would give among others, a band at 475 bp for p/blaS-$L_1^R L_2^{6-8}$, a band at 719 bp for p/blaS-L_3^{10} and two bands at 475 bp and 244 bp for p/blaS-$L_1^R L_2^{6-8} L_3^{10}$.

Table S1: Inserts sequences of clones from the p/blaS-$L_1^R L_2^{6-8}$ libraries selected or not for activity on IPTG-ampicillin 20 µg/ml in DH12S *E. coli* strain. /: no insert ; *: amber stop codon.

	Non selected for activity		Selected for activity	
	L2	L1	L2	L1
$L_1^R L_2^6$	* V W V V L L R H C P F S Q A V V Y	C P I S V Q nd	L G L L S * (E Q L L G P) P W S P S Q F G G T S G A P G H N G Q S S R * V (P M Y D L M)	G P T G Q V G V P C W P nd
$L_1^R L_2^7$	K F P H R M T Y V K M L L V P G R F V V P L W R * V F C	S L V S S M S P T A L G	K F P H R M T D S M L G T S G G G G * R G (G R V P P T L) G G R G G A A L A R G G A A L D C G V W G G F G D G N V	A N F A P G C P S C K S A H T G Y S C N S
$L_1^R L_2^8$	V L V R C V G C G M V L L F S V G A G G G G G K G R V D C T C C H E G W R V S E	T Q R S N N G I P A F V A L R	G V I P V F P A V T G A R T C * (nd) Q R V V I L * R (G A P G S S V G) FRAMESHIFT (Y V F C C G G F) V C F G G G C G R G G G A H L E FRAMESHIFT (nd) G L T C L S G S V I G * S V G L (nd)	nd A P R S R A G E R A R A C A S G E R C C H S T R
$L_1^R L_2^9$	V G V F C G L V V G W G G P F D F Y A L G C G V G G C V V S L G F C G V G C R A G F C G S	S K T C P * A Q M C F T A S S	A L R G A V S F R G S S D G L G V R W G V L G A V P V C G Y A G W L E T W G G T M F Y * C (C F F F W F C G L) V C G F G V G A C R L G V V G G W R	S N P A A L S A P A S S G E S G E P S P T

* : amber stop codon
nd : unreadable sequence
In brackets : Sequence encoded by the minority vector of co-transformed clones

Tables S2: Inserts sequences of clones from the p/blaS-$L_1^R L_2^{6\text{-}9} L_3^{10}$ libraries selected or not for activity on ampicillin 20 µg/ml in XL1 blue MRF' *E. coli* strain. Only clones transformed with a recombined vector were sequenced. /: no insert ; *: amber stop codon. Redundacies are highlighted in color. Note that the p/blaS-$L_1^R L_2^8 L_3^{10}$ and p/blaS-$L_1^R L_2^9 L_3^{10}$ libraries were respectively contaminated with clones of the p/blaS-$L_1^R L_2^6 L_3^{10}$ and p/bla-$L_1^R L_2^7 L_3^{10}$ libraries

Clones	L2	L1	L3
$L_1^R L_2^6 L_3$ – Glucose #1	PAPV*A	AKV	PGRAWKQLTA
$L_1^R L_2^6 L_3$ – Glucose #2	SRHMFC	GVQ	EPWVSGDCGT
$L_1^R L_2^6 L_3$ – Glucose #3	S*ARGR	GSY	ELWRMSRATI
$L_1^R L_2^6 L_3$ – Glucose #4	SGLGGG	CPT	DAMWRRM*QS
$L_1^R L_2^6 L_3$ – Glucose #5	NL*DGF	ASA	LPI*GPVWLT
Clones	**L2**	**L1**	**L3**
$L_1^R L_2^6 L_3$ – Amp_{20} #1	Y*	GRT	VIWNLPV*TV
$L_1^R L_2^6 L_3$ – Amp_{20} #2	/	/	RVKIPRWPYW
$L_1^R L_2^6 L_3$ – Amp_{20} #3	GWGGLY	GSL	VRIGSLTSTA
$L_1^R L_2^6 L_3$ – Amp_{20} #4	PPVFTP	GGG	VNSGAQDNGP
$L_1^R L_2^6 L_3$ – Amp_{20} #5	GWGGLY	GSL	GGTPCFT*MA
$L_1^R L_2^6 L_3$ – Amp_{20} #6	GG*CAG	ATS	WRPWQQSPCS
$L_1^R L_2^6 L_3$ – Amp_{20} #7	GSSGVG	SEL	NGLSHDLPCL
$L_1^R L_2^6 L_3$ – Amp_{20} #8	GWGGLY	GSL	GGEGAELGGW
$L_1^R L_2^6 L_3$ – Amp_{20} #9	GSDGRF	GNT	ADSMSVCPAV
$L_1^R L_2^6 L_3$ – Amp_{20} #10	FPPRHL	GAA	GCSGPRLTTM

Clones	L2	L1	L3
$L_1^R L_2^7 L_3$ – Glucose #1	GGCSGGV	SIT	AEDRGVGDVS
$L_1^R L_2^7 L_3$ – Glucose #2	FGLSRL*	CFN	WVGCRGVCIR
$L_1^R L_2^7 L_3$ – Glucose #3	GWCW	AKL	GRSDKYVLET
$L_1^R L_2^7 L_3$ – Glucose #4	VRVGP*A	GTT	GLFERQAAWG
Clones	**L2**	**L1**	**L3**
$L_1^R L_2^7 L_3$ – Amp_{20} #1	AFVGRFI	SRA	YFQSAGDRNV
$L_1^R L_2^7 L_3$ – Amp_{20} #2	GRGFHGA	AKT	NGPAESTISG
$L_1^R L_2^7 L_3$ – Amp_{20} #3	RGPFWGW	ALH	*MYPGSGYSW
$L_1^R L_2^7 L_3$ – Amp_{20} #4	EDVWG*LS	SIS	SRQFPMLPF*
$L_1^R L_2^7 L_3$ – Amp_{20} #5	GGPRLG*	GLN	VKPGV*AGGL
$L_1^R L_2^7 L_3$ – Amp_{20} #6	/	/	GFLSRRRWEF
$L_1^R L_2^7 L_3$ – Amp_{20} #7	FW	AQR	RACWDGGMCL
$L_1^R L_2^7 L_3$ – Amp_{20} #8	/	/	RECWGTLVEE
$L_1^R L_2^7 L_3$ – Amp_{20} #9	/	/	VGL*WRRQLQ
$L_1^R L_2^7 L_3$ – Amp_{20} #10	RLRYGWG	SNE	CIGSCMVCRP

Clones	L2	L1	L3
$L_1^R L_2^8 L_3$ – Glucose #1	PLGP*P	AHA	LCVRHEQAMT
$L_1^R L_2^8 L_3$ – Glucose #2	G*LASA	APE	CEVSFLGMCY
$L_1^R L_2^8 L_3$ – Glucose #3	WSYLGI*S	GCR	TGHSRQFHWV
$L_1^R L_2^8 L_3$ – Glucose #4	LP*LVG	AGY	NLIRGLGV*R
$L_1^R L_2^8 L_3$ – Glucose #5	DVAPGGST	GGH	MWRVGE*GRL
Clones	**L2**	**L1**	**L3**
$L_1^R L_2^8 L_3$ – Amp_{20} #1	FPPRHL	GAA	ESNLSRNVVR
$L_1^R L_2^8 L_3$ – Amp_{20} #2	CPKAGE	SPR	MWKLLSGPGL
$L_1^R L_2^8 L_3$ – Amp_{20} #3	AITGRA	APN	LVGHSVAATS
$L_1^R L_2^8 L_3$ – Amp_{20} #4	GWGGLY	GSL	L*EWGWR*LV
$L_1^R L_2^8 L_3$ – Amp_{20} #5	*RNANT	SNM	LQPIGTPSAS
$L_1^R L_2^8 L_3$ – Amp_{20} #6	GLSGAG	SRK	/
$L_1^R L_2^8 L_3$ – Amp_{20} #7	GSDGRF	GNT	FCMSGAGAES
$L_1^R L_2^8 L_3$ – Amp_{20} #8	GWGGLY	GSL	GRSDDAVSLH
$L_1^R L_2^8 L_3$ – Amp_{20} #9	GGRGPPG	STT	G*SGSRGSSL
$L_1^R L_2^8 L_3$ – Amp_{20} #10	WVG*PSPG	ALD	ASRV*ELGLV

Clones	L2	L1	L3
$L_1^R L_2^9 L_3$ – Glucose #1	PLPGR*LTR	GQN	CGVSTRSMGV
$L_1^R L_2^9 L_3$ – Glucose #2	SFRAG*L	AAQ	GSMGRSIRRG
$L_1^R L_2^9 L_3$ – Glucose #3	GT*LMGWVT	GLF	EYWVGAVTPS
$L_1^R L_2^9 L_3$ – Glucose #4	GAC*PPG	SPM	VCRSVFLLGS
$L_1^R L_2^9 L_3$ – Glucose #5	WCLSLC*	GNI	RGPSLFWWAI
Clones	**L2**	**L1**	**L3**
$L_1^R L_2^9 L_3$ – Amp_{20} #1	AFVGRFI	SRA	AGSCGSTPAS
$L_1^R L_2^9 L_3$ – Amp_{20} #2	RMRGPPA	AAV	SRSTRSTKHI
$L_1^R L_2^9 L_3$ – Amp_{20} #3	/	/	GSRLAWFCPS
$L_1^R L_2^9 L_3$ – Amp_{20} #4	RPMM*PR*A	SDG	RAPMMGDDAP
$L_1^R L_2^9 L_3$ – Amp_{20} #5	AFVGRFI	SRA	PRVPPYSGG
$L_1^R L_2^9 L_3$ – Amp_{20} #6	FGRGPPL	GTT	PRTAAH*EPE
$L_1^R L_2^9 L_3$ – Amp_{20} #7	/	/	VVGPYRPSGM
$L_1^R L_2^9 L_3$ – Amp_{20} #8	/	/	EVLRTGLTGL
$L_1^R L_2^9 L_3$ – Amp_{20} #9	RGSL*CRAC	SLT	VFRPVPTRLA
$L_1^R L_2^9 L_3$ – Amp_{20} #10	/	/	VYERGVPCAR

Tables S3: Randomized loops sequences of the clones isolated after 3 rounds of selection on the anti-*Coxiella burnetii* antibody from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ & L_3^{10} libraries. Redundant clones in rounds 3 and 4 are highlighted in color. Pairs of cysteine in the L2 inserts of the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ & L_3^{10} libraries are colored in red.

Library fd/bla-$L_1^R L_2^{c5\text{-}7c} L_3^6$			
Clones	**L2**	**L1**	**L3**
PM Rd3.1	CVYAPLTC	AQN	SRQWPT
PM Rd3.2	CLSLSPC	GGW	HSGDRT
PM Rd3.3	CYADSTTC	GRD	HNKANT
PM Rd3.4	CISSMAGC	SAQ	DRKEPT
PM Rd3.5	CWKLPGC	GYG	KSKVSS
PM Rd3.6	CVSEAVC	GHR	QAGQPS
PM Rd3.7	CLLPRSEC	GAR	NTVNDQ
PM Rd3.8	CSAFSGC	GDH	DDGTYD

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$			
Clones	**L2**	**L1**	**L3**
L3(6) Rd3.1	VFCVGQDQ	GAS	/
L3(6) Rd3.2	CGGFVCG	GSN	QRLYPS
L3(6) Rd3.3	QALPAY	GDF	/
L3(6) Rd3.4	TTLPSG	GLD	/
L3(6) Rd3.5	CVPFRGC	GRR	/
L3(6) Rd3.6	VASCAC	GLY	/
L3(6) Rd3.7	STQVGT	AVR	/
L3(6) Rd3.8	VYRGSS	GNA	/

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$			
Clones	**L2**	**L1**	**L3**
L3(10) Rd3.1	VAAVQQ	GRF	/
L3(10) Rd3.2	QAAREY	GLY	/
L3(10) Rd3.3	VSAMVS	GKP	/
L3(10) Rd3.4	LLSSPG	GGV	/
L3(10) Rd3.5	IAHACS	GAR	/
L3(10) Rd3.6	TLSVPG	AAQ	/
L3(10) Rd3.7	GDCLPFC	GNR	/

Tables S4: Randomized loops sequences of the clones isolated after 4 rounds of selection on the anti-*Coxiella burnetii* antibody from the fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ and the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ & L_3^{10} libraries. Redundant clones in rounds 3 and 4 are highlighted in color. Pairs of cysteine in the L2 inserts of the fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ & L_3^{10} libraries are colored in red.

Library fd/bla-$L_1^R L_2^{c5\text{-}7c} L_3^6$			
Clones	**L2**	**L1**	**L3**
PM Rd4.1	CIESDVSC	SKR	TGRDPS
PM Rd4.2	CIKKFDSC	AQH	ASNYAS
PM Rd4.3	CLAHSSVC	GMR	PGVNSN
PM Rd4.4	CLPFDPHC	GSQ	AG
PM Rd4.5	CLYREGSC	AKR	PAGPAS
PM Rd4.6	CLPFDPHC	GSQ	AG
PM Rd4.7	CLPFDPHC	GSQ	AG
PM Rd4.8	CVSTVGGC	SPS	TGTYPT
PM Rd4.9	CRRMDDC	GTH	KGIRAT
PM Rd4.10	CIYRSDEC	GHA	KALIPS

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$			
Clones	**L2**	**L1**	**L3**
L3(6) Rd4.1	CGGFVCG	GSN	QRLYPS
L3(6) Rd4.2	VLAAGPSA	GYM	LAA
L3(6) Rd4.3	CVPFRGC	GRR	/
L3(6) Rd4.4	SVQMLA	GYQ	/
L3(6) Rd4.5	CVPFRGC	GRR	/
L3(6) Rd4.6	CGGFVCG	GSN	QRLYPS
L3(6) Rd4.7	SFAGGA	GAW	SS
L3(6) Rd4.8	CGGFVCG	GSN	QRLYPS
L3(6) Rd4.9	CVPFRGC	GRR	/
L3(6) Rd4.10	CVPFRGC	GRR	/

Library fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^{10}$			
Clones	**L2**	**L1**	**L3**
L3(10) Rd4.1	ELTVE	KPL	/
L3(10) Rd4.2	VCSSLQ	SGA	/
L3(10) Rd4.3	VQHVDN	GQV	PGSARST
L3(10) Rd4.4	MVSTAA	GRY	/
L3(10) Rd4.5	VSAYAG	AQQ	/
L3(10) Rd4.6	SILTAP	ARD	/
L3(10) Rd4.7	CIGLPRG	GTN	/
L3(10) Rd4.8	YQMKAG	GAR	/
L3(10) Rd4.9	VVVERS	GQP	/
L3(10) Rd4.10	STFVHA	ANS	/

Stabilization impact on BlaKr activity regulation

Bruno Baudoux, Patrice Soumillion

Laboratoire de Biochimie et de Génétique Moléculaire Bactérienne (BGMB)

Institut des Sciences de la Vie, Université Catholique de Louvain,

Place Croix du Sud, 4-5 boîte 3, 1348 Louvain-La-Neuve, Belgium

Introduction

We previously showed that the stabilized TEM-1 β-lactamase was a better scaffold than the wild type enzyme to create a highly diverse library of insertants (See results part I). However, the probable stiffening of the protein scaffold arising from the stabilization might reduce the chances of selecting allosteric enzymes from that collection. Indeed, a positive correlation has been highlighted between allostery, protein dynamics and flexibility[1]. Daily and Gray analyzed local motions in a set of allosteric and non-allosteric proteins by comparing pairs of high-resolution structures in different liganded states recovered from the Protein Data Bank. They found that, in average, allosteric enzymes displayed higher flexibility and extent of motion than their non-allosteric counterparts. Moreover, they showed a reverse correlation between the density of constraints such as contacts and hydrogen bonds at a protein site and the propensity for motion at this same site[1]. Therefore, scaffold stiffening that arises from protein stabilization could impair a potential conformational change of the enzyme upon effector binding and chance of allostery to appear. On the other hand, a gain in global stability and scaffold rigidity could promote or allow local motions and perturbations at the engineering site otherwise not tolerated. The stabilizing mutations introduced could also induce a redistribution of the energy of stabilization within the protein structure and modify the allosteric properties. For evaluating these hypothesis, we studied the stabilization effect on the activity regulation of BlaKr, an allosteric β-lactamase previously isolated in the laboratory[2]. This β-lactamase is positively regulated by aminoglycosides (e.g. kanamycin A) in the presence of aminosulfonic acids (*e.g.* MES, 2-(N-morpholino)ethanesulfonic acid). The activation is indirect and results from the expulsion of an inhibitor (aminosulfonic acids) by the aminoglycosides. The aminosulfonates have the same effect as competitive inhibitors although K_m increase of the enzyme is not fully explained by the steric competition with the penicillin substrate. A remnant of activity on nitrocefin at saturation in MES suggests that it does not bind to the active site but close to it. Moreover, NMR studies recorded chemical shift perturbations of buried amino acids likely resulting of backbone motions upon MES binding[2]. These results suggest that inhibition might also be due to a shift in the population states of the enzyme[2]. Kanamycin binding to an

overlapping site causes the MES expulsion and hence enzyme reactivation. A more detailed view of the mechanism is discussed in the thesis introduction.

Results & discussion

The BlaKr gene was cloned in the pHDi-Ex/blaA vector (See results part I) in place of the wt β-lactamase gene. The four previously described stabilizing mutations[3] (H153R, M182T, T195S & A224V) were inserted in the BlaKr gene by site-directed mutagenesis. Both His-tagged BlaKr and stabilized BlaKr enzymes were produced in the periplasmic space of *E. coli* BL21 (DE3), cell extracted and purified on an immobilized metal affinity column (IMAC) **(Figure 1)**. The benzylpenicillin (PenG) hydrolysis catalysed by these two enzymes was followed by kinetic measurements at 232 nm in MES buffer and in the presence of increasing amount of kanamycin A (KanA) **(Figure 2)**.

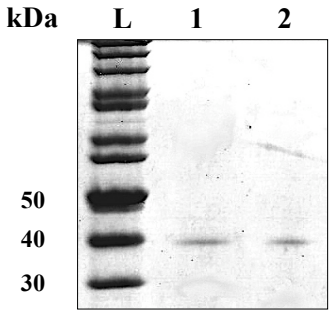


Figure 1: SDS-PAGE showing IMAC purified BlaKr (1) and stabilized BlaKr (2). The collected fractions showing activity on nitrocefin were pooled together and loaded on the gel. L = Molecular Weight Ladder (Fermentas PageRuler unstained protein ladder).

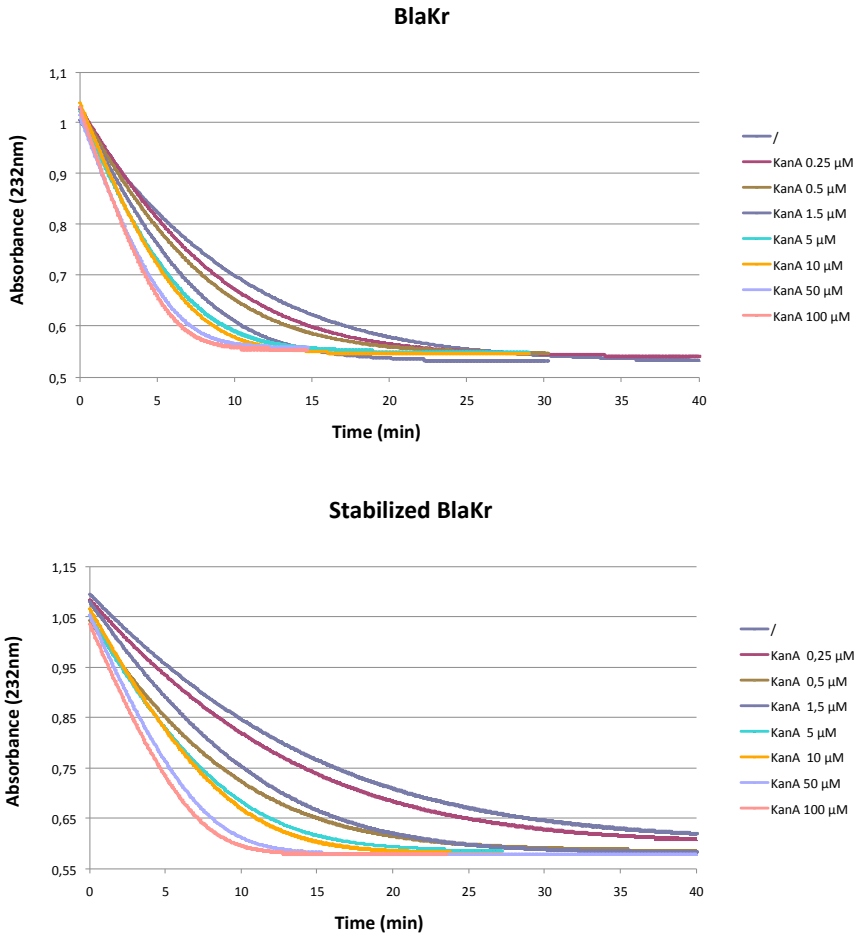


Figure 2: Activity regulation of BlaKr and stabilized BlaKr (2.0×10^{-9} M) by Kanamycin A. The β-lactamase activities were followed by measurement of PenG hydrolysis at 232 nm in 20 mM MES buffer at 25°C, pH 6.6.

For each time course, the experimental kinetic parameters were obtained by fitting the hydrolysis curves with the Equation 1 (see Materials & methods). The values of the three kinetic parameters, K_m, k_{cat} and k_{cat}/K_m were plotted against KanA concentration **(Figure 3)**.

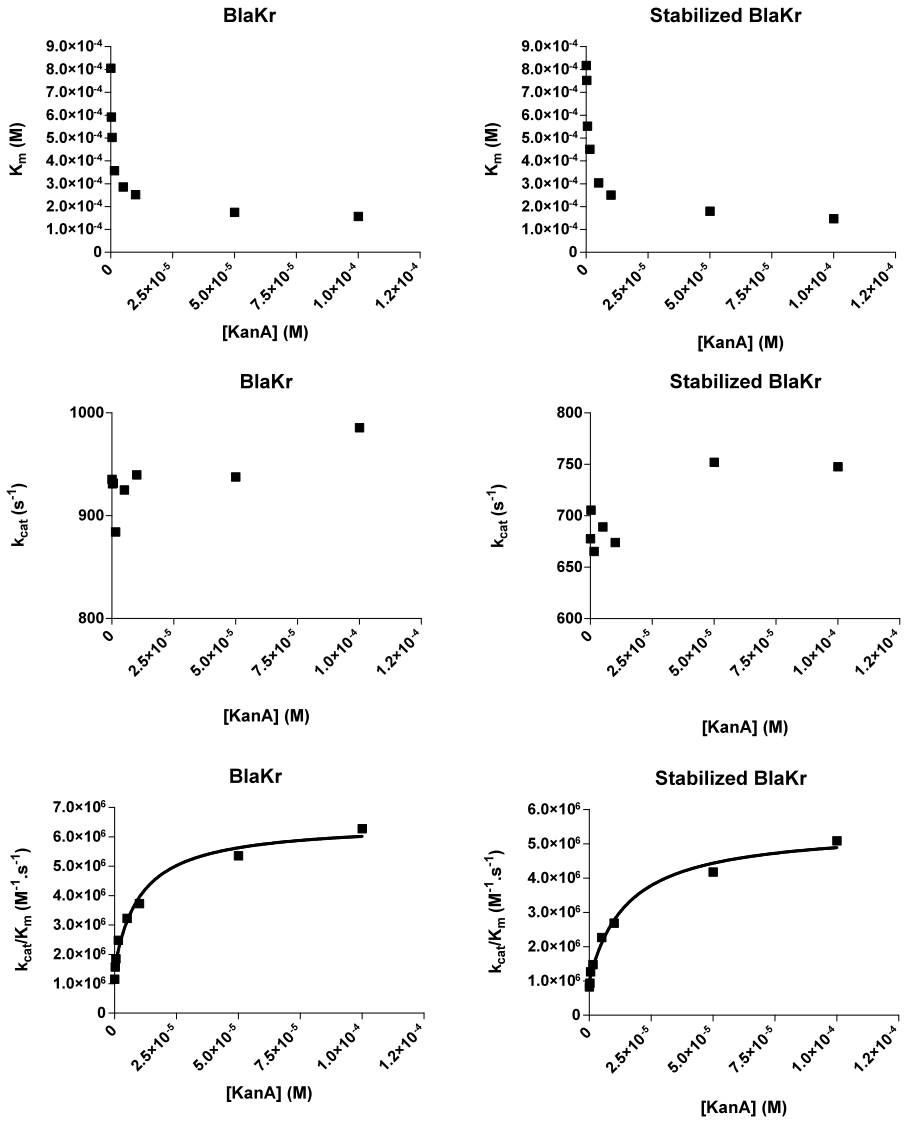


Figure 3: Effect of increasing KanA concentrations on the kinetic parameters of PenG hydrolysis by BlaKr (a.) or stabilized BlaKr (b.) at pH 6.6 in 20 mM MES buffer, 25°C. The curves were fitted with the equation 2 using the GraphPad Prism software. The data are derived from a single experiment without replicate.

Increasing concentrations of KanA induce a similar K_m decrease for both enzymes in MES buffer **(Figure 3)**. The experimental points of k_{cat}/K_m as a function of kanA concentration were fitted to Equation 2 (See Materials & methods). It provided the parameters $(k_{cat}/K_m)°_{app}$ and $(k_{cat}/K_m)^L_{app}$ where the naught and L exponents (L for ligand) correspond to absence or saturation in KanA respectively **(Table 1)**. The affinity of KanA for the free enzyme (Kd) was also obtained **(Table 1)**.

Enzyme	$(k_{cat}/K_m)°_{app}$ $(M^{-1}. s^{-1})$	$(k_{cat}/K_m)^L_{app}$ $(M^{-1}. s^{-1})$	Kd (M)
BlaKr	1.5 (± 0.2) x 10^6	6.5 (± 0.3) x 10^6	1.1 (± 0.3) x 10^{-5}
Stabilized BlaKr	1.0 (± 0.1) x 10^6	5.5 (± 0.3) x 10^6	1.5 (± 0.4) x 10^{-5}

Table 1: Kinetic parameters for the hydrolysis of PenG catalyzed by BlaKr and stabilized BlaKr enzymes in 20 mM MES buffer at pH 6.6, 25°C. The naught and L exponents correspond to absence or saturation in KanA respectively.

It appeared that the increase of the specificity constant (k_{cat}/K_m) in presence of KanA was of same amplitude for both enzymes. The measured affinities of the free enzymes for KanA were comparable too. Therefore, scaffold stabilization of BlaKr seems not to have modified its activity regulation properties. Several interpretations can be proposed from these results. If, as suggested by the previous study[2], BlaKr allosteric regulation is partly linked to a conformational change, stabilizing mutations might simply have no effect on these structural perturbations. The local changes would not correlate with the global protein stability. The enzyme dynamics might not be highly decreased upon stabilization or the resulting stiffening might not extend to the perturbed region. The stabilizing mutations are indeed located outside the chemical shift perturbations area **(Figure 4)**. In particular, the mutations are more than 20 Å apart from two shifted residues of the active site, N132 & E166, which could be involved in the K_m modulation of the enzyme. The asparagine 132 is involved in positioning the substrate via the formation of a hydrogen bond with the carbonyl group of the β-lactam ring[4] and the glutamate 166 is an essential residue located in the omega loop flanking the active site. This loop, playing a major role in the catalysis, folds lately upon cis-isomerization of the E166 P167 peptide bond and is relatively mobile and loosely packed against the main scaffold of the enzyme[5,6]. Therefore, TEM-1 β-lactamase stabilization is unlikely to alter its mobility.

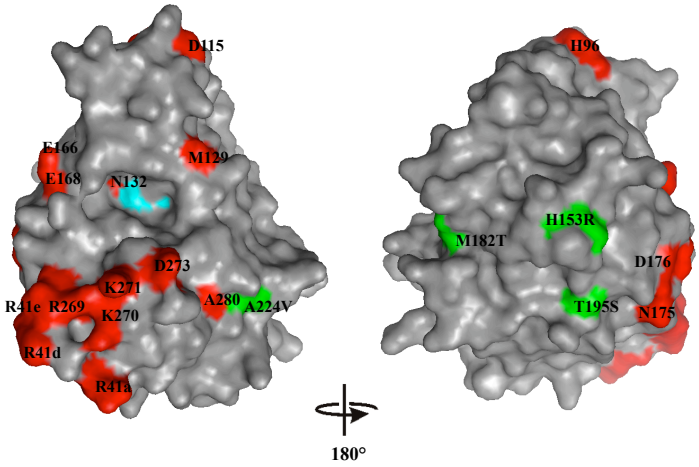


Figure 4: BlaKr mapped with the stabilizing mutations (in green) and with the main chemically shifted residues upon KanA binding in MES buffer (in red). In cyan is the active site. The pictures were made with MacPymol.

Secondly, conformational changes observed by NMR upon KanA binding in MES buffer might not be directly linked to the activity regulation of BlaKr. The observed allostery might not involve conformational changes in the protein if resulting from a strict competitive inhibition by MES suppressed upon KanA binding. A NMR analysis should be performed on the stabilized BlaKr to observe eventual variations of the chemical shifts profile upon KanA binding.

It is also worth noting that the chemical shifts could as well result from a change in the solvation sphere of the residues in question.

Yet, these results reinforce or at least do not invalidate the choice of the stabilized TEM-1 β-lactamase as an engineering scaffold to design allosteric enzymes. However, this study on a single mutant enzyme is not sufficient for drawing a firm conclusion. The same experiment should be repeated on other β-lactamases regulated by ligands such as metallic ions[7,8] and for which the allosteric mechanism is more certain given the small size of the ligand.

Materials and methods

General

T4 DNA ligase and Fast Start High Fidelity DNA polymerases were purchased from Roche.

Restriction enzymes were from New England Biolabs.

Kanamycin (Gibco) was prepared as a stock of 10 mg/ml in milliQ water, filtered through a 0.45 µm filter (Millipore) and stored at -20°C.

Glucose (Fluka) was freshly prepared as a 20% (w/v) solution, filtered through a 0.45 µm filter (Millipore) and kept at 4°C.

IPTG (Eurogentec) was solubilized in milliQ water to a final concentration of 100 mM, filtered through a 0.45 µm filter (Millipore) and stored at -20°C.

Bacterial Strains

BL21(DE3): *F^- ompT gal dcm lon hsdSB(rB^- mB^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*

Oligonucleotides

Synthetic oligonucleotides were purchased from Eurogentec.

Restriction sites are underlined and mutations are in bold.

Fwd & Rev are respectively standing for forward and reverse primer.

H153R: 5' CTAACCGCTTTTTTG**CGC**AACATGGGGGATCATG 3'

M182T: 5' CGTGACACCACG**ACA**CCGGTAGCAATGGCAACAACG 3'

T195S: 5' CGCAAACTATTA**TCC**GGAGAACTACTTACTCTAGCTTCCCGG 3'

A224V: 5' GGACCACTTCTGCGCTCA**GTA**CTTCCGGCTGGCTGG 3'

TEM-His Fwd: 5'CATCATCATCATCATCATGGCGCCCACCCAGAAACGCTGGTG 3'

TEM-His Rev: 5' ATCATCGCTAGCCTACCAATGCTTAATCAGTGAGGC 3'

Vector construction

The BlaKr gene was amplified by PCR from the pBAD/BlaKr vector[2] using the Fast Start High Fidelity polymerase and the primers pair TEM-His Fwd and TEM-His Rev. The PCR product was purified with the "High Pure PCR Product Purification kit" from Roche and restricted with the restriction enzymes *Nar*I and *Nhe*I. The restriction product was purified on a column CHROMASPIN 1000 (Clontech) and

cloned back in the pHDi-Ex/blaS vector (See results part I) restricted and purified the same way. The ligation was done at 16°C for 1 h using the T4 DNA ligase and an insert/vector molecular ratio of 3:1.

Site-directed mutagenesis

Stabilizing mutations were inserted in the BlaKr gene using the "QuickChange® Multi Site-Directed kit" from Roche and the 4 oligonucleotides: H153R, M182T, T195S & A224V. The kit protocol was followed.

Cells transformation

See Results part I

Enzymes production and purification

See Results part II

A BCA assay using the "BCA™ Protein Assay kit" from ThermoFisher was used to determine the protein concentration.

SDS-PAGE

See Results part II

Activity measurements

Enzyme activities were measured following PenG (Sigma Aldrich) hydrolysis at 232 nm ($\Delta\varepsilon$ = 1042 $M^{-1}.cm^{-1}$), 25°C. Twenty microliters of the enzyme solution (2.0 x 10^{-7} M, containing 10 µM BSA) were added to a 2 ml solution of 0.5 mM PenG in MES buffer (20 mM, pH 6.6) supplemented with different concentrations of KanA (Sigma Aldrich). The absorbance decrease as a function of time was followed using the Cary Varian 3-BIO spectrophotometer.

Data analysis

Numeric integration using Equation 1 allowed fitting the experimental time courses (absorbance over time) while optimizing the kinetic parameters k_{cat} & K_m by ordinary least squares (OLS) regression.

Equation 1 (See annexe 5; p.224):

$$A_{t+\Delta t} = \left([S]_t - \frac{k_{cat} \cdot [E]_0 \cdot [S]_t}{K_m + [S]_t} \cdot \Delta t \right) \cdot \Delta\varepsilon \cdot l + A_\infty$$

with,

$$\left([S]_t - \frac{k_{cat} \cdot [E]_0 \cdot [S]_t}{K_m + [S]_t} \cdot \Delta t \right) = [S]_t + \Delta[S] = [S]_{t+\Delta t}$$

and where $\Delta\varepsilon$ is the molar extinction coefficient difference between the PenG and the hydrolyzed PenG at 232 nm (1042 $M^{-1}.cm^{-1}$), l is the path length, A_∞ is the absorbance at infinity time, $[E]_0$ is the initial enzyme concentration, $[S]_t$ is the substrate concentration at time t and Δt is the time laps between two absorbance measurements. For using this numeric integration, Δt must be sufficiently small so that [S] does not vary more than 1%. At time 0, $[S]_0$ and A_0 are known. Then, an iterative process allows calculating, for each increament of Δt, the values of [S] and A at every time point from the previous value of [S]. The k_{cat}, K_m and A_∞ constants are optimized so that the calculated progress curve fits to the experimental one.

The obtained k_{cat}/K_m values were plotted as a function of KanA concentration ([L]). The corresponding curves were fitted to the equation 2 to obtain the parameters $(k_{cat}/K_m)^\circ_{app}$, $(k_{cat}/K_m)^L_{app}$ and Kd which are respectively the specificity constants of the enzyme in absence or saturation in KanA and the dissociation constant of the free enzyme for KanA.

Equation 2 (See annexe 4; p.222):

$$(k_{cat}/K_m)_{obs} = (k_{cat}/K_m)^L + \frac{K_d}{K_d + [L]} \left((k_{cat}/K_m)^\circ - (k_{cat}/K_m)^L \right)$$

Curves fitting were carried out with the GraphPad Prism software.

References

1. Daily, M.D. & Gray, J.J. Local motions in a benchmark of allosteric proteins. *Proteins* **67**, 385-399 (2007).

2. Volkov, A.N., Humberto, B., Mathonet, P., Evrard, C., Ubbink, M., Declercq, J.P., Soumillion, P. & Fastrez, J. Engineering an Allosteric Binding Site for Aminoglycosides into TEM1-β-Lactamase. *Chembiochem* **12**, 904-913 (2011).

3. Hecky, J. & Muller, K.M. Structural Perturbation and Compensation by Directed Evolution at Physiological Temperature Leads to Thermostabilization of β-Lactamase. *Biochemistry* **44**, 12640-12654 (2005).

4. Jacob, F., Joris, B., Lepage, S., Dusart, J. & Frère, J.M. Role of the conserved amino acids of the « SDN » loop (Ser130, Asp131 and Asn132) in a class A beta-lactamase studied by site-directed mutagenesis. *Biochem. J.* **271**, 399-406 (1990).

5. Vandenameele, J., Lejeune, A., Di Paolo, A., Brans, A., Frère, J.-M., Schmid, F.X. & Matagne, A. Folding of class A beta-lactamases is rate-limited by peptide bond isomerization and occurs via parallel pathways. *Biochemistry* **49**, 4264-4275 (2010).

6. Vanhove, M., Lejeune, A. & Pain, R.H. Beta-lactamases as models for protein-folding studies. *Cell Mol Life Sci* **54**, 372 (1998).

7. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).

8. Legendre, D., Soumillion, P. & Fastrez, J. Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.* **17**, 67-72 (1999).

Conclusion et perspectives

I. Discussion et conclusion générales

Ingénierie allostérique de β-lactamases TEM-1 allostériques

Cette thèse s'est inscrite dans la continuité d'un projet visant à créer *de novo* un site de régulation allostérique au sein d'une β-lactamase TEM-1. Une banque de phage display a pour cela été construite par randomisation de trois boucles de surface de l'enzyme, contiguës et proches du site catalytique. Les résidus $G_{238}E_{240}R_{241}$ de la boucle L1, située à proximité du site actif, ont été mutés aléatoirement. Des inserts dégénérés de taille variable ont été insérés au niveau des boucles C-terminale (L2) et N-terminale (L3) de l'enzyme. Des sélections pour affinité suivies de criblages cinétiques ont par la suite été réalisées à partir de cette banque afin d'isoler des clones régulés allostériquement par des cibles variées (nickel, streptavidine, anticorps).

Nous avons montré qu'une β-lactamase TEM-1 stabilisée constituait une meilleure architecture protéique que l'enzyme sauvage pour la création d'une banque d'insertants de haute diversité. Le gain de stabilité thermodynamique apporté par quatre mutations stabilisantes (H153R, M182T, T195S et A224V) a fortement augmenté la tolérance de l'enzyme vis à vis des perturbations structurelles au sein du site d'ingénierie. Un nombre de 6,6 x 10^7 mutants actifs présentant les 3 boucles simultanément mutées a ainsi été obtenu dans un vecteur phagique (fd-tet), soit 70 fois plus qu'avec une β-lactamase TEM-1 sauvage[1]. La stabilisation de l'enzyme a en outre permis d'augmenter grandement la diversité potentielle au sein de la boucle L2 en autorisant l'élimination d'un pont disulfure autrement requis pour obtenir des mutants actifs sur ampicilline 20 µg/ml. L'absence de cette contrainte physique augmente le nombre d'états conformationnels potentiellement accessibles par la boucle et pourrait éventuellement influer sur l'émergence de nouvelles capacités de liaison et de phénomènes d'allostérie.

Le potentiel évolutif de la β-lactamase stabilisée en terme de propension à l'émergence d'une régulation allostérique a été évalué en cherchant des clones régulés par des ions métalliques. Une banque analogue, précédemment construite dans une β-lactamase sauvage par Pascale Mathonet[1], a servi de référence. Douze clones présentant plusieurs histidines dans leurs inserts ont émergé d'une sélection sur billes de nickel. Trois d'entre eux (25 %), dont deux possédant les mutations

stabilisantes, ont montré une régulation d'activité par la liaison d'ions métalliques. Les deux variants préstabilisés présentaient les profils de régulation les plus intéressants. Le premier était significativement inhibé (v_0 x 0,2) en présence de nickel et possédait la plus forte affinité pour l'ion métallique (10 µM). De manière intéressante, un pont disulfure était présent au niveau de la boucle L2 de ce clone. Les premiers résultats suggèrent que la présence de cette contrainte structurelle ne réduit pas les chances de sélectionner des phénomènes d'allostérie. Le second clone, malgré une faible affinité pour le nickel (720 µM), se distinguait par un profil d'activation (v_0 x 1,9). La préstabilisation de l'enzyme accompagnée d'une éventuelle rigidification du squelette protéique ne semble pas avoir eu d'effet négatif sur la sélection de clones allostériques. Au contraire, l'augmentation de la tolérance à l'insertion en L3 dans le contexte de mutagenèse simultanée des trois boucles semble favoriser la sélection de clones allostériques. En effet, tous les clones (8) régulés par le nickel isolés dans ce travail et celui de P. Mathonet[2] étaient des triples mutants malgré la plus faible représentativité de ceux-ci dans les banques. Le gain de stabilité pourrait permettre des perturbations conformationnelles au niveau du site d'ingénierie autrement non tolérées et promouvoir ainsi l'émergence de phénomènes allostériques. En effet, la liaison de ligand au niveau de régions de faible stabilité induit souvent des changements conformationnels à l'origine de phénomènes d'allostérie[3]. A tout le moins, le gain de diversité au niveau des banques suite à la stabilisation de l'enzyme devrait permettre d'augmenter le nombre absolu de clones présentant de nouvelles capacités de liaison et régulés allostériquement. L'insertion des 4 mutations stabilisantes au sein d'un clone allostérique précédemment sélectionné (BlaKr) et régulé par des aminoglycosides[4] n'a d'autre part pas eu d'incidence sur le profil de régulation de l'enzyme. L'ensemble de ces résultats conforte le choix de la β-lactamase stabilisée comme plateforme d'ingénierie pour l'évolution d'une régulation allostérique. Néanmoins, le potentiel évolutif de cette enzyme doit être exploré davantage par diverses expériences discutées ci-après dans la partie « Perspectives ».

Coût de fitness du couple phagemide/β-lactamase TEM-1

A l'inverse de la stabilisation de l'enzyme, d'autres modifications apportées à la stratégie ont généré davantage de problèmes que d'améliorations. Tel a été le cas de l'utilisation du vecteur de type phagemidique, pHDi-Ex[5], et de l'exportation de la protéine de fusion via la voie co-traductionnelle SRP au lieu de la voie SEC.

La construction de banques de haute diversité a pourtant été facilitée par la manipulation plus aisée du phagemide par rapport au vecteur phagique. De même, un taux d'expression de la protéine de fusion plus élevé contrebalancerait plus efficacement les pertes d'activité enzymatique liées aux perturbations structurelles induites au cours de l'ingénierie. Cela a permis d'augmenter la dégénérescence de la boucle L3 sans affecter la diversité numérique de triple-mutants actifs. Cependant, ce haut taux d'expression était accompagné d'un important coût de fitness pour les bactéries. Il en a résulté des biais dans les banques en faveur de clones exprimant moins (ou pas) de protéines de fusion. Ces variants présentaient des codons stop ambre au niveau des inserts ou de larges délétions dans la construction génétique. Les sélections sur streptavidine à partir de la banque phagemidique se sont trouvées perturbées par l'émergence de clones n'exprimant pas de protéine de fusion et sélectionnés via leur meilleure fitness *in vivo*. Quelques clones dont l'affinité pour la streptavidine a semble-t-il été confirmée par ELISA ont néanmoins pu être isolés de ce bruit de fond. Aucun de ceux-ci n'a montré de régulation d'activité en présence de streptavidine.

Les causes du coût de fitness sont difficiles à définir et certainement multifactorielles. Outre la dépense énergétique liée à la surproduction de l'enzyme, la saturation du système d'exportation co-traductionnel SRP pourrait en être à l'origine. Au laboratoire, G. Woronoff a montré un effet toxique de l'utilisation de la voie SRP pour l'exportation périplasmique d'une pénicilline G acylase surexprimée chez *E. coli*[6]. La surproduction de la protéine pourrait également mener à la formation de corps d'inclusion dont l'accumulation dans le périplasme nuirait à la bactérie. Une instabilité de l'enzyme suite à l'introduction des mutations/insertions au niveau des boucles promouvrait éventuellement la formation de tels agrégats. Valéry Mathieu, dans une étude d'ingénierie de la charnière d'une β-lactamase TEM-1 permutée circulairement, a d'ailleurs montré une possible

corrélation entre la stabilité des clones et le coût de fitness associé à leur expression[7].

L'absence de toxicité avérée lors de l'expression de protéines autres que la β-lactamase à partir du vecteur pHDi-Ex[5] laisse néanmoins présumer d'un effet β-lactamase-dépendant. Par ailleurs, certains résultats obtenus au laboratoire avec une β-lactamase TEM-1 permutée circulairement suggèrent une toxicité plus prononcée de l'enzyme active par rapport à son variant inactif. Cela se traduit notamment par un temps de génération en phase exponentielle de croissance 18 % plus élevé pour les bactéries exprimant la β-lactamase active par rapport à celles exprimant le mutant inactif. C'est effet n'a cependant pas été observé avec la TEM-1 sauvage étudiée dans cette thèse. Dans la littérature, plusieurs études font état d'un coût de fitness lié à l'expression de β-lactamases[8–10]. Il a été proposé que certaines de ces enzymes puissent présenter une activité peptidase et/ou reconnaître les extrémités D-alanyl-D-alanine des précurseurs du peptidoglycane et interférer ainsi avec la formation ou le remodelage de ce dernier[8,9,11]. Il est communément accepté que les β-lactamases de classe A et C ont un ancêtre DD-carboxy-peptidase commun impliqué dans la biosynthèse et la maintenance de la paroi bactérienne[12,13]. Il n'est dès lors pas étonnant de voir ces enzymes parentes présenter des activités se chevauchant. Ainsi, certaines β-lactamases sauvages présentent une faible mais significative activité d'hydrolyse de di- ou tri-peptides N-acétylés présentant le groupement D-alanyl-D-alanine[14]. Une β-lactamase de classe C (ampC) semble également jouer un rôle dans la synthèse, le remodelage ou le recyclage du peptidoglycane. La suppression de son expression provoque des modifications morphologiques (augmentation du diamètre, contours et extrémités anormaux) chez une souche d'*E. coli* n'exprimant pas les PBPs 1a et 5 [15]. D'autre part, l'expression de β-lactamases SFO-1 ou de type OXA induit des changements au niveau de la composition du peptidoglycane (diminution du nombre de muropeptides réticulés et augmentation de leur taille) et affecte la fitness des bactéries[11]. Il a aussi été montré dans une étude de Georgiou[16] et au laboratoire que la surpexpression de la β-lactamase TEM-1 provoquait un allongement des bactéries. Ce phénomène a néanmoins également été observé lors de surexpression d'autres protéines. A noter que l'ingénierie des boucles de l'enzyme dans la présente étude pourrait avoir généré des clones ayant une tendance plus grande aux interactions aspécifiques avec

des constituants du peptidoglycane et/ou de l'espace périplasmique et induire dès lors une toxicité plus importante.

Des indications dans ce travail suggèrent que l'ancrage de la β-lactamase TEM-1 au niveau de la membrane interne de la bactérie, via l'extrémité C-terminale de la protéine de phage p3, génère une toxicité accrue. Au cours de la construction de la banque p/blaS-L_3[10], un biais important a été observé en faveur de clones exprimant moins de protéine de fusion du fait de la présence d'un codon stop ambre dans leur insert en L3. Au laboratoire, Valéry Mathieu a également mis en évidence que le temps de génération des bactéries exprimant une β-lactamase permutée circulairement en fusion avec l'extrémité C-terminale de la protéine p3 était doublé par rapport à celles exprimant l'enzyme sous sa forme libre[7]. Suvorov a en outre montré que sur des co-cultures de huit jours, les bactéries exprimant une β-lactamase ancrée étaient complètement éliminées au profit de celles encodant une β-lactamase libre[17]. L'ancrage membranaire pourrait renforcer les phénomènes de perturbation de la paroi bactérienne en localisant l'enzyme à proximité immédiate de la machinerie de biosynthèse du peptidoglycane. La toxicité pourrait aussi provenir d'une déstabilisation de la membrane interne des bactéries[18]. De manière intéressante, il a été montré que la β-lactamase se retrouvait relarguée en grande quantité dans le surnageant lorsque surexprimée dans le périplasme d'*E. coli*[16]. Cet effet, qui pourrait permettre à la bactérie d'échapper à la toxicité de la β-lactamase, ne serait pas autorisé une fois l'enzyme ancrée au niveau de la membrane interne. Cela expliquerait le coût de fitness plus important de la β-lactamase TEM-1 lorsque celle-ci est exprimée en fusion avec l'extrémité C-terminale de la p3.

A noter que la stabilisation de l'enzyme semble accentuer indirectement ces problèmes de toxicité au niveau des banques. Une augmentation de la concentration en β-lactamases, entraînée par une résistance accrue à la protéolyse, serait une cause possible. D'autre part, la stabilisation a permis d'obtenir des sites de liaison potentiels éventuellement plus complexes au niveau de la β-lactamase TEM-1. Cela pourrait générer davantage d'interactions aspécifiques délétères entre l'enzyme et les constituants périplasmiques. Cette complexification des sites pourrait également provoquer une plus grande instabilité structurelle au niveau du site d'ingénierie qui contribuerait à la formation de corps d'inclusion.

II. Perspectives

Poursuite des sélections sur diverses cibles d'intérêt

Les sélections par phage display seront poursuivies afin de sélectionner des β-lactamases TEM-1 régulées allostériquement par diverses cibles d'intérêt. Les résultats des sélections obtenus avec les banques construites dans l'enzyme stabilisée ou sauvage seront comparés en terme quantitatif et qualitatif. Une caractérisation des clones isolés sera menée par une étude approfondie de leurs cinétiques enzymatiques. L'affinité pour le ligand et l'amplitude de l'effet de régulation pourront éventuellement être améliorées par un nouveau tour de mutations/sélection, les sélections étant introduites par PCR mutagénisante. Finalement, les enzymes les plus intéressantes seront produites et purifiées pour une caractérisation structurale par diffraction aux rayons X et/ou résonance magnétique nucléaire (RMN). A plus long terme, des biosenseurs pourront être développés en utilisant les β-lactamases régulées allostériquement par des molécules cibles d'intérêt en tant que biocapteur. Plusieurs de ces cibles sont déjà disponibles au laboratoire sous forme biotinylée afin de permettre facilement leur immobilisation au cours des sélections par phage display.

En premier lieu, la sélection sur l'anticorps anti-*Coxiella burnetii* devrait être poursuivie en criblant davantage de clones isolés aux tours 3 et 4 soit par des mesures cinétiques en présence et en absence de la cible soit par un test ELISA. Ce dernier pourrait mettre en évidence d'éventuels clones ayant une affinité pour la cible sans pour autant présenter de régulation allostérique.

Un autre anticorps, reconnaissant un allergène majeur (Alt-a1) des moisissures appartenant au genre *Alternaria*, représenterait une seconde cible d'intérêt. Cet allergène est responsable de plus en plus de cas d'allergies majeures chez l'homme[19].

Une troisième molécule, le nicotinamide adénine dinucléotide (NAD), serait également une cible de choix tant d'un point de vue fondamental qu'appliqué. La molécule consiste en deux nucléotides reliés par leurs groupements phosphate, l'un présentant une base adénine et l'autre un nicotinamide **(Figure 1.a)**. C'est un coenzyme d'oxydoréduction impliqué dans de nombreuses réactions cellulaires[20]. En tant que transporteur d'électrons, la molécule existe sous deux états différents : la

forme NAD^+ est l'agent oxydant et la forme NADH l'agent réducteur **(Figure 1.b)**. Le NAD^+ joue également le rôle de substrat dans de nombreuses réactions au niveau de la réparation de l'ADN, de l'apoptose, de la signalisation cellulaire,…[21] Le rapport entre les formes NAD^+ et NADH est donc le reflet des activités métaboliques et de la santé des cellules. Il est généralement déterminé par mesure spectrophotométrique. Les deux formes partagent en effet un pic d'absorbance à 260 nm mais seul le NADH présente un second pic à 340 nm. La mesure de l'évolution de l'absorbance à 340 nm permet également de suivre les réactions enzymatiques impliquant ce couple redox comme coenzyme. La sensibilité des mesures est néanmoins réduite par un important bruit de fond généré par de nombreuses molécules biologiques ayant de fortes absorbances dans cette gamme de longueur d'onde. La sélection de β-lactamases capables de discriminer entre les formes oxydée et réduite du NAD et régulées allostériquement ouvrirait de nouvelles perspectives dans le domaine des tests diagnostiques et enzymatiques. Ces enzymes pourraient servir de biocapteur dans le développement de biopuces permettant de doser l'activité de diverses réactions enzymatiques NAD/NADH dépendantes et les molécules servant de substrat à ces dernières.

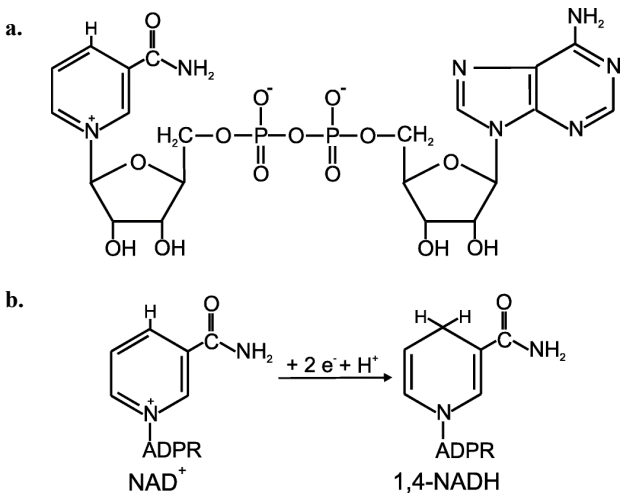


Figure 1: (a.) Forme oxydée du nicotinamide adénine dinucléotide (NAD^+) (b.) Réaction d'oxydo-réduction du nicotinamide adénine dinucléotide.

Impact de la stabilisation sur l'émergence d'une régulation allostérique

Il serait intéressant de poursuivre l'étude de l'impact des mutations stabilisantes sur la propension à l'émergence d'une régulation allostérique au niveau de la β-lactamase TEM-1.

Même si la stabilisation du mutant BlaKr n'a pas eu d'effet sur son profil de régulation, une comparaison des spectres RMN des enzymes non-stabilisée et stabilisée en présence de MES et de kanamycine A pourrait s'avérer informatif sur plusieurs points. Une meilleure compréhension du mécanisme sous jacent à la régulation allostérique et une mesure de l'effet des mutations stabilisantes sur la dynamique de la protéine devraient en découler. Il serait également pertinent d'étudier comment la suppression du pont disulfure en L2 influe sur la régulation allostérique du clone. Les deux cystéines seraient mutées en une glycine et une alanine correspondant respectivement aux résidus sauvages des positions 240 et 241 de la boucle L2. La mesure de FRET (Fluorescence par transfert d'énergie de résonance) entre les résidus tryptophane de l'enzyme et une kanamycine A couplée à un fluorophore (disponible au laboratoire) devrait également apporter des éclaircissements sur le mécanisme de régulation de BlaKr et de son mutant stabilisé. Les affinités de l'enzyme pour la KanA en présence et en absence de MES ainsi que les constantes d'inhibition pour le MES en présence et en absence de KanA pourraient être déterminées avec plus de précision et comparées à celles obtenues par les études cinétiques. Une correspondance entre les mesures conforterait le modèle théorique actuel expliquant la régulation d'activité observée.

Comme précédemment discuté, la stabilisation de β-lactamases régulées par le nickel permettrait d'étudier l'impact des mutations stabilisantes sur des phénomènes d'allostérie conformationnelle davantage avérés. A nouveau, l'influence du pont disulfure en L2 sur la régulation allostérique pourrait être analysée en supprimant cette contrainte une fois les clones stabilisés. De même, il serait intéressant d'analyser le rôle joué par la préstabilisation de l'enzyme dans l'émergence d'une régulation allostérique au niveau des clones « L3.6 Tr3.2 » et « L3.10 Tr2.1 » décrits dans ce travail. Les mutations stabilisantes seront supprimées afin de voir si les enzymes conservent d'une part leur stabilité et d'autre part leur profil de régulation.

L'insertion d'un pont disulfure dans la boucle L2 du mutant « L3.10 Tr2.1 » permettrait également d'étudier l'impact de cette contrainte sur l'allostérie par une approche inverse à celle précédemment mentionnée.

Investigation du coût de fitness

Il serait intéressant d'étudier en détail l'origine des coûts de fitness observés lors de la surexpression de la protéine de fusion. Une meilleure compréhension du phénomène permettrait éventuellement de pouvoir mieux le contrôler au cours des sélections et d'expliquer les différences de fitness mesurées entre les banques fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ et fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$.

Les différentes hypothèses sur l'origine de la toxicité discutées dans les conclusions générales pourraient être investiguées en comparant les coûts de fitness engendrés par l'expression de différents variants de la β-lactamase TEM-1 et de protéines contrôles de nature variée. Ainsi la confrontation des résultats obtenus avec l'enzyme sauvage et son variant inactif renseignerait sur l'incidence de l'activité de l'enzyme sur la cytotoxicité. Une mise en compétition *in vivo* de bactéries exprimant l'un ou l'autre de ces variants n'a pas mis en évidence une différence de fitness. Il conviendrait de poursuivre la comparaison par une analyse de la morphologie des bactéries par microscopie électronique en transmission et de la composition du peptidoglycane par HPLC[11]. Ces études devraient également être menées dans des conditions d'ancrage membranaire en exprimant ces enzymes en fusion avec l'extrémité C-terminale de la p3. De même, il serait intéressant d'étudier l'impact de la surexpression des différents variants de la β-lactamase TEM-1 sur les bactéries au cours des différentes phases de croissance (latence, exponentielle, stationaire). En effet, les modifications transcriptionnelles et métaboliques occurant au niveau des cellules au fil des ces phases pourraient influer sur la nature et l'ampleur de la toxicité induite par les différentes protéines. Ainsi par exemple, le peptidoglycane subit un important remodelage lors du passage de la phase exponentielle de croissance à la phase stationnaire[22]. L'effet toxique de β-lactamases interférant avec la machinerie enzymatique responsable de ce remodelage pourrait dès lors n'être observé que lors de cette phase de transition. Enfin, le ciblage des enzymes vers la voie d'exportation Sec ou Tat au lieu de la voie SRP pourrait également définir le rôle joué par le système d'export dans le coût de fitness.

Les mêmes analyses pourraient être poursuivies avec les deux banques fd/bla-$L_1^R L_2^{C5\text{-}7C} L_3^6$ et fd/blaS-$L_1^R L_2^{6\text{-}8} L_3^6$ ou avec des clones isolés provenant de celles-ci. Cela permettrait éventuellement la mise en évidence d'une corrélation entre la complexification des sites de liaison ou l'absence de pont disulfure en L2 et un coût de fitness plus important. On pourrait également voir si les mutants les plus toxiques génèrent davantage de corps d'inclusion ou induisent plus de modifications au niveau de la structure bactérienne ou de la composition du peptidoglycane.

La fraction de phages présentant une β-lactamase à leur surface sera également déterminée pour les deux banques. Le taux de display peut influer sur l'infectivité de la banque de phage et donc sur le fitness *in vivo*. Cela permettra aussi d'estimer si la stabilisation de l'enzyme augmente notablement la quantité de β-lactamase ancrée au niveau de la membrane interne de la bactérie. Pour ce faire, un western blot sur des préparations de phages sera à nouveau réalisé mais en utilisant cette fois un anticorps anti-p3 à la place de l'anticorps anti-β-lactamase afin de déterminer le rapport protéine de fusion/p3 libre. Ce type de western blot permettrait aussi de déterminer si les clones avec un pont disulfure en L2 présentent un meilleur display sur phage pouvant expliquer leur émergence au cours des sélections.

III. Références

1. Mathonet, P., Deherve, J., Soumillion, P. & Fastrez, J. Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. *Protein Sci.* **15**, 2323-2334 (2006).
2. Mathonet, P., Barrios, H., Soumillion, P. & Fastrez, J. Selection of allosteric beta-lactamase mutants featuring an activity regulation by transition metal ions. *Protein Sci.* **15**, 2335-2343 (2006).
3. Luque, I. & Freire, E. Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins* **41**, 63-71 (2000).
4. Volkov, A.N., Humberto, B., Mathonet, P., Evrard, C., Ubbink, M., Declercq, J.P., Soumillion, P. & Fastrez, J. Engineering an Allosteric Binding Site for Aminoglycosides into TEM1-β-Lactamase. *Chembiochem* **12**, 904-913 (2011).
5. Heyd, B., Pecorari, F., Collinet, B., Adjadj, E., Desmadril, M. & Minard, P. In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry* **42**, 5674-5683 (2003).
6. Woronoff, G. Développement de méthodes pour l'évolution dirigée de la pénicilline G acylase par compartimentalisation in vitro. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2012).

7. Mathieu, V. Engineering allosteric regulation into a circulalry permuted TEM-1 beta-lactamase. Ph.D. Thesis. Université Catholique de Louvain, Belgique. (2010).

8. Morosini, M.I., Ayala, J.A., Baquero, F., Martínez, J.L. & Blázquez, J. Biological cost of AmpC production for Salmonella enterica serotype Typhimurium. *Antimicrob. Agents Chemother.* **44**, 3137-3143 (2000).

9. Bishop, R.E. & Weiner, J.H. Coordinate regulation of murein peptidase activity and AmpC beta-lactamase synthesis in Escherichia coli. *FEBS Lett.* **304**, 103-108 (1992).

10. Marciano, D.C., Karkouti, O.Y. & Palzkill, T. A Fitness Cost Associated With the Antibiotic Resistance Enzyme SME-1 β-Lactamase. *Genetics* **176**, 2381-2392 (2007).

11. Fernández, A., Pérez, A., Ayala, J.A., Mallo, S., Rumbo-Feal, S., Tomás, M., Poza, M. & Bou, G. Expression of OXA-type and SFO-1 β-lactamases induces changes in peptidoglycan composition and affects bacterial fitness. *Antimicrob. Agents Chemother.* **56**, 1877-1884 (2012).

12. Kirby, R. Evolutionary origin of the class A and class C beta-lactamases. *J. Mol. Evol.* **34**, 345-350 (1992).

13. Kelly, J., Dideberg, O., Charlier, P., Wery, J., Libert, M., Moews, P., Knox, Duez, C., Fraipont, C., Joris, B. & et, al. On the origin of bacterial resistance to penicillin: comparison of a beta-lactamase and a penicillin target. *Science* **231**, 1429 -1431 (1986).

14. Rhazi, N., Galleni, M., Page, M.I. & Frere, J.M. Peptidase activity of beta-lactamases. *Biochem. J.* **341**, 409-413 (1999).

15. Henderson, T.A., Young, K.D., Denome, S.A. & Elf, P.K. AmpC and AmpH, proteins related to the class C beta-lactamases, bind penicillin and contribute to the normal morphology of Escherichia coli. *J Bacteriol.* **179**, 6112-6121 (1997).

16. Georgiou, G., Shuler, M.L. & Wilson, D.B. Release of periplasmic enzymes and other physiological effects of β−lactamase overproduction in Escherichia coli. *Biotechnol. Bioeng.* **32**, 741 - 748 (1988).

17. Suvorov, M., Vakulenko, S.B. & Mobashery, S. Cytoplasmic-Membrane Anchoring of a Class A beta-lactamase and Its Capacity in Manifesting Antibiotic Resistance. *Antimicrob. Agents Chemother.* **51**, 2937-2942 (2007).

18. Zinder N.D., Model P. & Boeke J.D. Method for facilitating externalization of proteins synthesized in bacteria. *US patent 4,595,598*

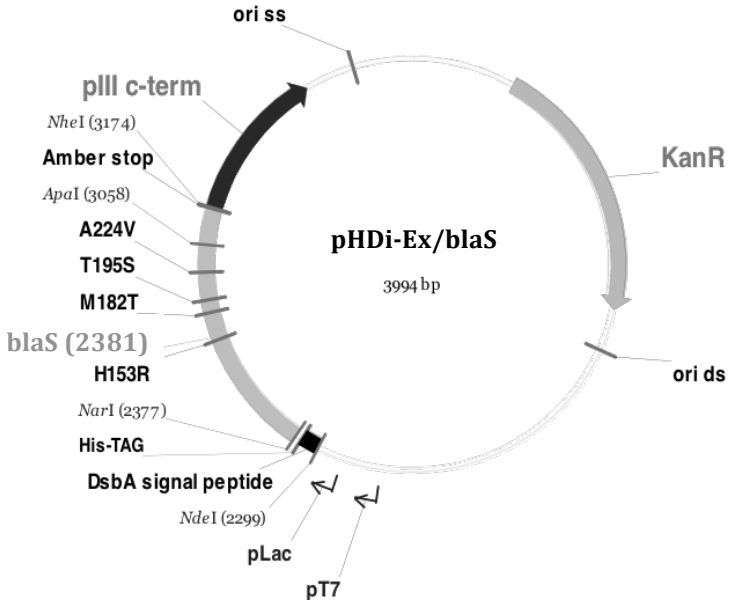
19. Portnoy, J., Brothers, D., Pacheco, F., Landuyt, J. & Barnes, C. Monoclonal antibody-based assay for Alt a1, a major Alternaria allergen. *Ann. Allergy Asthma Immunol.* **81**, 59-64 (1998).

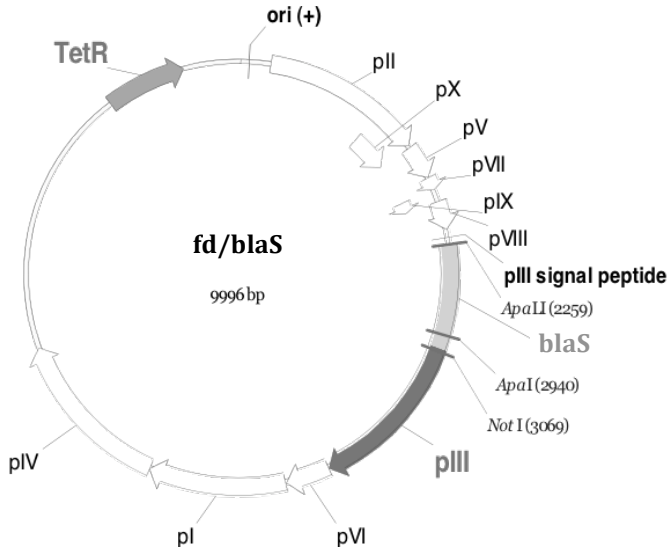
20. Khan, J.A., Forouhar, F., Tao, X. & Tong, L. Nicotinamide adenine dinucleotide metabolism as an attractive target for drug discovery. *Expert Opin. Ther. Targets* **11**, 695-705 (2007).

21. Ando, T., Ohashi, K., Ochiai, A., Mikami, B., Kawai, S. & Murata, K. Structural determinants of discrimination of NAD+ from NADH in yeast mitochondrial NADH kinase Pos5. *J. Biol. Chem.* **286**, 29984-29992 (2011).

22. Typas, A., Banzhaf, M., Gross, C.A. & Vollmer, W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* **10**, 123-137 (2011).

Annexe 1: Schematic representation of the pHDi-Ex/blaS and fd/blaS vectors.





Annexe 2: pHDi-Ex/blaS vector sequence around the β-lactamase gene (in bold).

*Nde*I — *DsbA signal peptide*

2297 cat atg aaa aag att tgg ctg gcg ctg gct ggt tta gtt tta gcg 2341
His Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala

His-Tag — *Nar*I

2342 ttt agc gca tcg gcg cat cat cat cat cat cat ggc gcc cag cca 2386
Phe Ser Ala Ser Ala His His His His His His Gly Ala **Gln Pro**

2387 gaa acg ctg gtg aaa gta aaa gat gct gaa gat cag ttg ggt gct 2431
Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln Leu Gly Ala

L2

2432 cga gtg ggt tac atc gaa ctg gat ctc aac agc ggt aag atc ctt 2476
Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu

2477 gag agt ttt cgc ccc gaa gaa cgt ttt cca atg atg agc act ttt 2521
Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe

2522 aaa gtt ctg cta tgt ggc gcg gta tta tcc cgt att gac gcc ggg 2566
Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Ile Asp Ala Gly

2567 caa gag caa ctc ggt cgc cgc ata cac tat tct cag aat gac ttg 2611
Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu

2612 gtt gag tac tca cca gtc aca gaa aag cat ctt acg gat ggc atg 2656
Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met

2657 aca gta aga gaa tta tgc agt gct gcc ata acc atg agt gat aac 2701
Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn

2702 act gcg gcc aac tta ctt ctg aca acg atc gga gga ccg aag gag 2746
Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys Glu

2747 cta acc gct ttt ttg cgc aac atg ggg gat cat gta act cgc ctt 2791
Leu Thr Ala Phe Leu Arg_{153} Asn Met Gly Asp His Val Thr Arg Leu

2792 gat cgt tgg gaa ccg gag ctg aat gaa gcc ata cca aac gac gag 2836
Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu

2837 cgt gac acc acg aca ccg gta gca atg gca aca acg ttg cgc aaa 2881
Arg Asp Thr Thr Thr_{182} Pro Val Ala Met Ala Thr Thr Leu Arg Lys

2882 cta tta tcc gga gaa cta ctt act cta gct tcc cgg caa caa tta 2926
Leu Leu Ser_{195} Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu

2927 ata gac tgg atg gag gcg gat aaa gtt gca gga cca ctt ctg cgc 2971
Ile Asp Trp Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg

2972 tca gta ctt ccg gct ggc tgg ttt att gct gat aaa tct gga gcc 3016
Ser Val_{224} Leu Pro Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala

*Apa*I

3017 ggt gag cgt ggg tct cgc ggt atc att gca gca ctg ggg ccc gat 3061
Gly Glu Arg Gly Ser Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp

L1

3062 ggt aag ccc tcc cgt atc gta gtt atc tac acg acg ggg agt cag 3106
Gly Lys Pro Ser Arg Ile Val Val Ile Tyr Thr Thr Gly Ser Gln

L3

3107 gca act atg gat gaa cga aat aga cag atc gct gag ata ggt gcc 3151
Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu Ile Gly Ala

L3 — *Nhe*I — C-terminal p3 gene

3152 tca ctg att aag cat tgg tag gct agc tcc ggt ggt ggc tct ggt 3196
Ser Leu Ile Lys His Trp End* Ala Ser Ser Gly Gly Gly Ser Gly

***Suppressible stop codon**

3197 tcc ggt gat ttt gat tat gaa aag atg gca aac gct aat aag ggg 3241
Ser Gly Asp Phe Asp Tyr Glu Lys Met Ala Asn Ala Asn Lys Gly

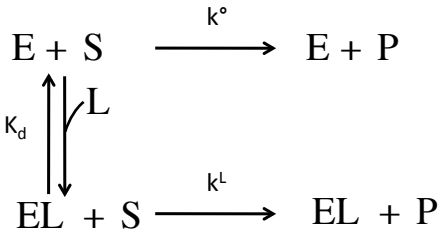
Annexe 3: fd/blaS vector sequence around the β-lactamase gene (in bold).

p3 signal peptide

```
2206  atg aaa aaa tta tta ttc gca att cct tta gtt gtt cct ttc tat   2250
      Met Lys Lys Leu Leu Phe Ala Ile Pro Leu Val Val Pro Phe Tyr
                  ApaLI
2251  tct cac agt gca cac cca gaa acg ctg gtg aaa gta aaa gat gct   2295
      Ser His Ser Ala His Pro Glu Thr Leu Val Lys Val Lys Asp Ala

2296  gaa gat cag ttg ggt gct cga gtg ggt tac atc gaa ctg gat ctc   2340
      Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu
                        L2
2341  aac agc ggt aag atc ctt gag agt ttt cgc ccc gaa gaa cgt ttt   2385
      Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe

2386  cca atg atg agc act ttt aaa gtt ctg cta tgt ggc gcg gta tta   2430
      Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu

2431  tcc cgt att gac gcc ggg caa gag caa ctc ggt cgc cgc ata cac   2475
      Ser Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His

2476  tat tct cag aat gac ttg gtt gag tac tca cca gtc aca gaa aag   2520
      Tyr Ser Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys

2521  cat ctt acg gat ggc atg aca gta aga gaa tta tgc agt gct gcc   2565
      His Leu Thr Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala

2566  ata acc atg agt gat aac act gcg gcc aac tta ctt ctg aca acg   2610
      Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr

2611  atc gga gga ccg aag gag cta acc gct ttt ttg cgc   aac atg ggg 2655
      Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu Arg153 Asn Met Gly

2656  gat cat gta act cgc ctt gat cgt tgg gaa ccg gag ctg aat gaa   2700
      Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu

2701  gcc ata cca aac gac gag cgt gac acc acg aca   ccg gta gca atg 2745
      Ala Ile Pro Asn Asp Glu Arg Asp Thr Thr Thr182 Pro Val Ala Met

2746  gca aca acg ttg cgc aaa cta tta tcc   gga gaa cta ctt act cta 2790
      Ala Thr Thr Leu Arg Lys Leu Leu Ser195 Gly Glu Leu Leu Thr Leu

2791  gct tcc cgg caa caa tta ata gac tgg atg gag gcg gat aaa gtt   2835
      Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys Val

2836  gca gga cca ctt ctg cgc tca gta   ctt ccg gct ggc tgg ttt att 2880
      Ala Gly Pro Leu Leu Arg Ser Val224 Leu Pro Ala Gly Trp Phe Ile

2881  gct gat aaa tct gga gcc ggt gag cgt ggg tct cgc ggt atc att   2925
      Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile
                  ApaI                 L1
2926  gca gca ctg ggg ccc gat ggt aag ccc tcc cgt atc gta gtt atc   2970
      Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile

2971  tac acg acg ggg agt cag gca act atg gat gaa cga aat aga cag   3015
      Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln
                                 L3                    Linker
3016  atc gct gag ata ggt gcc tca ctg att aag cat tgg ggg att gag   3060
      Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp Gly Ile Glu
                NotI                 p3 gene
3061  ggg cgt gcg gcc gca gaa act gtt gaa agt tgt tta gca aaa cct   3105
      Gly Arg Ala Ala Ala Glu Thr Val Glu Ser Cys Leu Ala Lys Pro
```

Annexe 4: Calculation of the dissociation constant K_d and the kinetic parameters k° and k^L

The reaction scheme for an enzyme that is regulated by a ligand can be represented as followed:



where E is the enzyme, S is the substrate, P is the product, L is the ligand and EL is the complex between E and L. K_d is the dissociation constant and k°, k^L the enzymatic activities with the ligand bound or not, respectively.

Following this scheme:

$$K_d = \frac{[E].[L]}{[EL]} \quad (1)$$

$$fE = \frac{[E]}{[E]+[EL]} \quad (2)$$

$$fEL = 1 - fE \quad (3)$$

$$k_{obs} = k^\circ.fE + k^L.fEL \quad (4)$$

Developing the different equations gives:

(1) $\longrightarrow$ $$[EL] = \frac{[E].[L]}{[K_d]} \quad (5)$$

(2) + (5) $\longrightarrow$ $$fE = \frac{[E]}{[E] + \frac{[E][L]}{K_d}}$$

$$\longrightarrow \quad fE = \frac{K_d}{K_d + [L]} \qquad (6)$$

(3) + (4) $$\longrightarrow \quad k_{obs} = k^{\circ}.fE + k^{L}.(1 - fE)$$

$$\longrightarrow \quad k_{obs} = k^{\circ}.fE + k^{L} - k^{L}.fE$$

$$\longrightarrow \quad k_{obs} = k^{L} + fE(k^{\circ} - k^{L}) \qquad (7)$$

Finally, we obtain an equation that contains three unknown constants:

(6) + (7) $$\longrightarrow \quad k_{obs} = k^{L} + \frac{K_d}{K_d + [L]}(k^{\circ} - k^{L})$$

A graph is constructed by plotting the measured experimental activity (k_{obs}) for different concentrations of ligand. K_d, k° and k^{L} are then determined by fitting the experimental data using the final equation.

Annexe 5: Mathematical development to obtain the Equation 1 from « Results part 3 ».

- In a kinetic reaction, the absorbance at time "*t*" (A_t) can be defined as:

$$A_t = A_t^S + A_t^P + A_{cst}$$

where A^S_t is the absorbance of the substrate at time "*t*", A^P_t is the absorbance of the product at time "*t*" and A_{cst} is a constant absorbance due to the buffer,...

- Applying the Beer-Lambert law we obtain:

$$A_t = [S]_t . \varepsilon_s . l + [P]_t . \varepsilon_p . l + A_{cst} \qquad (1)$$

where $[S]_t$ et $[P]_t$ are respectively the concentrations of the substrate and the product at time "*t*", ε_s and ε_p are respectively the molar extinction coefficients of the substrate and the product, and *l* is the path length.

- At any time "*t*", the sum of the concentrations of the substrate and the product equals the initial concentration of substrate $[S]_0$:

$$[S]_t + [P]_t = [S]_0 \qquad (2)$$

(1) + (2)
$$A_t = [S]_t . \varepsilon_s . l + \left([S]_0 - [S]_t\right) . \varepsilon_p . l + A_{cst}$$

$$A_t = [S]_t . \Delta\varepsilon . l + [S]_0 . \varepsilon_p . l + A_{cst} \qquad (3)$$

where $\Delta\varepsilon$ is the difference between the molar extinction coefficients of the substrate and the product (ε_s-ε_p).

- At infinity time, we can estimated that the substrate is completely consumed:

$$[S]_\infty = 0 \text{ and } [P]_\infty = [S]_0 \qquad (4)$$

- Therefore the absorbance at infinity time (A_∞) is:

(1) + (4) $$A_\infty = [S]_0 . \varepsilon_p . l + A_{cst} \quad (5)$$

and,

(1) + (5) $$A_t = [S]_t . \Delta\varepsilon . l + A_\infty \quad (6)$$

- The evolution of [S] overtime is depending on the differential form of the Michaelis-Menten equation:

$$-\frac{d[S]}{dt} = \frac{k_{cat}[E]_0[S]_t}{K_m + [S]_t} \quad or \quad d[S] = -\frac{k_{cat}[E]_0[S]_t}{K_m + [S]_t} . dt$$

- For numeric integration purpose, we consider that:

$$-\frac{d[S]}{dt} \cong -\frac{\Delta[S]}{\Delta t}$$

This is valid if Δt is sufficiently small so that [S] does not vary more than 1%.

- So,

$$\Delta[S] = -\frac{k_{cat}[E]_0[S]_t}{K_m + [S]_t} . \Delta t \quad (7)$$

- The concentration of the substrate at time $t+\Delta t$ ($[S]_{t+\Delta t}$) can be calculated:

$$[S]_{t+\Delta t} = [S]_t + \Delta[S] \quad (8)$$

and the absorbance at time $t+\Delta t$ ($A_{t+\Delta t}$) is therefore (7) + (8) in (6):

$$A_{t+\Delta t} = \left([S]_t - \frac{k_{cat} \cdot [E]_0 \cdot [S]_t}{K_m + [S]_t} \cdot \Delta t \right) \cdot \Delta\varepsilon \cdot l + A_\infty$$