Université de Montréal

Shifting the boundaries of experimental studies in engineering enzymatic functions: Combining the benefits of computational and experimental methods

par Maximilian Ebert

Département de biochimie et médecine moléculaire
Faculté de médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de *Philosophiae Doctor*
en biochimie

Décembre, 2016

Résumé

L'industrie chimique mondiale est en pleine mutation, cherchant des solutions pour rendre la synthèse organique classique plus durable. Une telle solution consiste à passer de la catalyse chimique classique à la biocatalyse. Bien que les avantages des enzymes incluent leur stéréo, régio et chimiosélectivité, cette sélectivité réduit souvent leur promiscuité. Les efforts requis pour adapter la fonction enzymatique aux réactions désirées se sont révélés d'une efficacité modérée, de sorte que des méthodes rapides et rentables sont nécessaires pour générer des biocatalyseurs qui rendront la production chimique plus efficace.

Dans l'ère de la bioinformatique et des outils de calcul pour soutenir l'ingénierie des enzymes, le développement rapide de nouvelles fonctions enzymatiques devient une réalité. Cette thèse commence par un examen des développements récents sur les outils de calcul pour l'ingénierie des enzymes. Ceci est suivi par un exemple de l'ingénierie des enzymes purement expérimental ainsi que de l'évolution des protéines. Nous avons exploré l'espace mutationnel d'une enzyme primitive, la dihydrofolate réductase R67 (DHFR R67), en utilisant l'ingénierie semi-rationnelle des protéines. La conception rationnelle d'une librarie de mutants, ou «*Smart library design*», impliquait l'association covalente de monomères de l'homotétramère DHFR R67 en dimères afin d'augmenter la diversité de la librairie d'enzymes mutées. Le criblage par activité enzymatique a révélé un fort biais pour le maintien de la séquence native dans un des protomères tout en tolérant une variation de séquence élevée pour le deuxième. Il est plausible que les protomères natifs procurent l'activité observée, de sorte que nos efforts pour modifier le site actif de la DHFR R67 peuvent n'avoir été que modérément fructueux.

Les limites des méthodes expérimentales sont ensuite abordées par le développement d'outils qui facilitent la prédiction des points chauds mutationnels, c'est-à-dire les sites privilégiés à muter afin de moduler la fonction. Le développement de ces techniques est intensif en termes de calcul, car les protéines sont de grandes molécules complexes dans un environnement à base d'eau, l'un des solvants les plus difficiles à modéliser. Nous présentons l'identification rapide des points chauds mutationnels spécifiques au substrat en utilisant l'exemple d'une enzyme cytochrome P450 industriellement pertinente, la CYP102A1. En appliquant la technique de simulation de la dynamique moléculaire par la force de polarisation

adaptative, ou «ABF», nous confirmons les points chauds mutationnels connus pour l'hydroxylation des acides gras tout en identifiant de nouveaux points chauds mutationnels. Nous prédisons également la conformation du substrat naturel, l'acide palmitique, dans le site actif et nous appliquons ces connaissances pour effectuer un criblage virtuel d'autres substrats de cette enzyme.

Nous effectuons ensuite des simulations de dynamique moléculaire pour traiter l'impact potentiel de la dynamique des protéines sur la catalyse enzymatique, qui est le sujet de discussions animées entre les experts du domaine. Avec la disponibilité accrue de structures cristallines dans la banque de données de protéines (PDB), il devient clair qu'une seule structure de protéine n'est pas suffisante pour élucider la fonction enzymatique. Nous le démontrons en analysant quatre structures cristallines que nous avons obtenues d'une enzyme β-lactamase, parmi lesquelles un réarrangement important des résidus clés du site actif est observable. Nous avons réalisé de longues simulations de dynamique moléculaire pour générer un ensemble de structures suggérant que les structures cristallines ne reflètent pas nécessairement la conformation de plus basse énergie.

Enfin, nous étudions la nécessité de compléter de manière informatisée un hémisphère où l'expérimental n'est actuellement pas possible, à savoir la prédiction de la migration des gaz dans les enzymes. À titre d'exemple, la réactivité des enzymes cytochrome P450 dépend de la disponibilité des molécules d'oxygène envers l'hème du site actif. Par le biais de simulations de la dynamique moléculaire de type Simulation Implicite du Ligand (ILS), nous dérivons le paysage de l'énergie libre de petites molécules neutres de gaz pour cartographier les canaux potentiels empruntés par les gaz dans les cytochromes P450 : CYP102A1 et CYP102A5. La comparaison pour les gaz CO, N_2 et O_2 suggère que ces enzymes évoluent vers l'exclusion du CO inhibiteur. De plus, nous prédisons que les canaux empruntés par les gaz sont distincts des canaux empruntés par le substrat connu et que ces canaux peuvent donc être modifiés indépendamment les uns des autres.

Mots clés: β-lactamase, cytochrome P450, CYP102A1, dihydrofolate réductase, évolution dirigée, ingénierie des protéines, dynamique des protéines, cinétique enzymatique, simulation de dynamique moléculaire *in silico*, migration gazeuse, ensembles structuraux, force de polarisation adaptative, échantillonnage implicite de ligand.

Abstract

The chemical industry worldwide is at a turning point, seeking solutions to make classical organic synthesis more sustainable. One such solution is to shift from classical catalysis to biocatalysis. Although the advantages of enzymes include their stereo-, regio-, and chemoselectivity, their selectivity often reduces versatility. Past efforts to tailor enzymatic function towards desired reactions have met with moderate effectiveness, such that fast and cost-effective methods are in demand to generate biocatalysts that will render the production of fine and bulk chemical production more benign.

In the wake of bioinformatics and computational tools to support enzyme engineering, the fast development of new enzyme functions is becoming a reality. This thesis begins with a review of recent developments on computational tools for enzyme engineering. This is followed by an example of purely experimental enzyme engineering and protein evolution. We explored the mutational space of a primitive enzyme, the R67 dihydrofolate reductase (DHFR), using semi-rational protein engineering. 'Smart library design' involved fusing monomers of the homotetrameric R67 DHFR into dimers, to increase the diversity in the resulting mutated enzyme libraries. Activity-based screening revealed a strong bias for maintenance of the native sequence in one protomer with tolerance for high sequence variation in the second. It is plausible that the native protomers procure the observed activity, such that our efforts to modify the enzyme active site may have been only moderately fruitful.

The limitations of experimental methods are then addressed by developing tools that facilitate computational mutational hotspot prediction. Developing these techniques is computationally intensive, as proteins are large molecular objects and work in aqueous media, one of the most complex solvents to model. We present the rapid, substrate-specific identification of mutational hotspots using the example of the industrially relevant P450 cytochrome CYP102A1. Applying the adaptive biasing force (ABF) molecular dynamics simulation technique, we confirm the known mutational hotspots for fatty acid hydroxylation and identify a new one. We also predict a catalytic binding pose for the natural substrate, palmitic acid, and apply that knowledge to perform virtual screening for further substrates for this enzyme.

We then perform molecular dynamics simulations to address the potential impact of protein dynamics on enzyme catalysis, which is the topic of heated discussions among experts in the field. With the availability of more crystal structures in the Protein Data Bank, it is becoming clear that a single protein structure is not sufficient to elucidate enzyme function. We demonstrate this by analyzing four crystal structures we obtained of a β-lactamase enzyme, among which a striking rearrangement of key active site residues was observed. We performed long molecular dynamics simulations to generate a structural ensemble that suggests that crystal structures do not necessarily reflect the conformation of lowest energy.

Finally, we address the need to computationally complement an area where experimentation is not currently possible, namely the prediction of gas migration into enzymes. As an example, the reactivity of P450 cytochrome enzymes depends on the availability of molecular oxygen at the active-site heme. Using the Implicit Ligand Sampling (ILS) molecular dynamics simulation technique, we derive the free energy landscape of small neutral gas molecules to map potential gas channels in cytochrome P450 CYP102A1 and CYP102A5. Comparison of CO, N_2 and O_2 suggests that those enzymes evolved towards exclusion of the inhibiting CO. In addition, we predict that gas channels are distinct from known substrate channels and therefore can be engineered independently from one another.

Keywords: β-lactamase, cytochrome P450, CYP102A1, dihydrofolate reductase, directed evolution, protein engineering, protein dynamics, enzyme kinetics, *in-silico* molecular dynamics simulations, gas migration, structural ensembles, adaptive biasing force, implicit ligand sampling.

Table of contents

List of tables

List of figures

List of videos

List of abbreviations

A

ABF ..... Adaptive biasing force
AMBER ..... Assisted model building with energy refinement
Amp ..... Ampicillin
ATP ..... Adenosine triphosphate

B

BLAST ..... Basic local alignment search tool
BM3 ..... CYP102A1
BTL2 ..... *Bacillus thermocatenulatus* lipase
BZ ..... Penicillins benzylpenicillin

C

CB ..... Carbenicillin
CCD ..... Charge-coupled device
CD ..... Circular dichroism
CF ..... Cephalosporins cephalothin
CgKR1 ..... *Candida glabrata* ketoreductase 1
Chl ..... Chloramphenicol
CPMG ..... Carr-Purcell-Meiboom-Gill
CSD ..... Cambridge Structural Database
CTX ..... Cefotaxime
CV ..... Column volumes
CYP ..... Cytochromes P450
CZ ..... Cefazolin

D

DDM ..... Doubly-mutated
DHF ..... Dihydrofolate
DHFR ..... Dihydrofolate reductase
DM ..... Dimer mutated
DMF ..... Dimethylformamide
DNA ..... Deoxyribonucleic acid
DNM ..... Non-mutated dimer
DSC ..... Differential scanning calorimetry
DSM ..... Singly mutated
DTT ..... Dithiothreitol

E

EM ..... Electron microscopy
ESBL ..... Extended-spectrum β-lactamases

G

GAFF ..... General AMBER force field
GB/VI ..... Generalized born/volume integral
GROMACS ..... Groningen machine for chemical simulations

H

HPLC ..... High performance liquid chromatography
HSL ..... N-3-oxohexanoyl-L-homoserine lactone
H-WT ..... Historical wild-type

I

ILS ..... Implicit ligand sampling
IPTG ..... Isopropyl β-D-1-thiogalactopyranoside
IRT ..... Inhibitor resistant TEM-like

K

Kan ..... Kanamycin
k_{cat} ..... Catalytic rate constant
k_{ex} ..... Exchange rate constant
K_i ..... Inhibitory constant
K_M ..... Michaelis constant

L

LES ..... Locally enhanced sampling
LINCS ..... LINear constraint solver

M

MD ..... Molecular dynamics
MnP ..... Manganese peroxidase
MOE ..... Molecular operating environment
MS ..... Mass spectrometry

N

NADPH ..... Nicotinamide adenine dinucleotide phosphate
NAMD ..... Nanoscale Molecular dynamics
NMR ..... Nuclear magnetic resonance

O

OD ..... Optical density

P

PAGE ..... Polyacrylamide gel electrophoresis, Polyacrylamide gel electrophoresis
PCR ..... Polymerase chain reaction
PDB ..... Protein Data Bank
PELE ..... Protein energy landscape exploration
PME ..... Particle mesh ewald
PMF ..... Potential of mean force
PMSF ..... Phenylmethane sulfonyl fluoride

Q

Q-PCR ..... Real-time polymerase chain reaction

R

R_2 ..... Rransverse relaxation rate
RAMD .... Random accelerated molecular dynamics
RMSD ..... Root-mean-square deviation
RMSF ..... Root mean square fluctuation
RNA ..... Ribonucleic acid
RT ..... Room temperature

S

S^2 ..... Order parameter
SAXS ..... Small angle X-ray scattering
SDS ..... Sodium dodecyl sulfate
SMD ..... Steered molecular dynamics
SP ..... Sucrose phosphorylase
SvGL ..... Non-heme chloroperoxidase

T

TB ..... Terrific broth
THF ..... Tetrahydrofolate
T_m ..... Melting temperature
TMP ..... Trimethoprim

U

US ..... Umbrella sampling
UV ..... Ultraviolet

V

VdW ..... Van der Waals
VMD ..... Visual molecular dynamics

W

WHAM ..... Weighted histogram analysis method
WT ..... Wild-type

Δ

ΔG ..... Gibbs free energy
ΔG^0_{gas} ..... Solvation energy of gas

Für Inga, Julie und _____

Acknowledgements

My friends and family know that I can express myself much better verbally than I can in writing. I would rather speak for two hours and hear my voice than typing a single paragraph. But occasionally you need to force yourself, right?

Wow was that a journey! Five years of Ph.D. who what have thought that? No other decision had such a huge impact on my life than picking Joelle's offer to come back to Montréal to start working with her. I am more than thankful for the opportunities she gave me to freely explore the new and emerging technologies discussed in this thesis. She always supported me and my work, even though together we had to explore many new fields of science. I also want to thank her for giving me the opportunity to work on grant proposals and conference organisations. Although getting an inside into what it takes to be a professor finally let me abandon my initial carrier goal of becoming one, I am thankful to have gotten this experience. It probably prevented me from making a big mistake.

Only through collaborations with other experts in bioinformatics, structural biology and NMR spectroscopy, I was able to study the systems I presented and ask the questions I asked. Therefore, I also want to thank Prof. Jürgen Pleiss and Prof. Albert Berghuis for giving me the opportunity to conduct research in their labs and for sharing their knowledge. I also want to thank the DAAD, PROTEO, CCVC, FRQ-NT and NSERC for their financial support during my Ph.D. studies.

Prof. Guillaume Lamoureux really marked my scientific career. He is such an inspiring teacher and I am really glad for having had the chance to take his class at the beginning of my Ph.D. I did not only learn the basics of computational biochemistry but also had great scientific debates with him. I am more than happy that we could share our passion for bioinformatics with students all over Quebec by organising two summer schools.

Since I am writing this part without really thinking about it, I just realized how many people I met and have worked with in the past years. I not only had the luck to have great colleagues which always supported me, I was also able to make friends I hope to keep forever. When I first came to Joelle's lab in 2009, I was only a diploma student in engineering and in desperate need for an internship. I was really proud of my C and D grades from the German engineering school but in Canada this looked like being a failure. I am deeply thankful that

Joelle and Jordan took a leap of faith and accepted me as an intern. Jordan was a great teacher and he reinforced my decision to start a Ph.D. in biochemistry instead of seeking a career in engineering. He is a great mentor and I hope he is not annoying too many med-students with his love for the Italian soccer team (how I hate Italy when it comes to soccer). Another intern at the same time was Dominic, with whom I share the passion for computer games. We had a great time together and I am happy he found his purpose later as nutritionist. Thanks for all the food we were allowed to taste during your studies!

Then there was Sophie ☺ What a funny girl! She is not at all streamlined and that is what makes her special. I will never forget how happy she was when I decided to come back to Joelle and start my Ph.D. We are not only sharing our love for science but also developed a real friendship over the years (and I do not call everybody a friend right Jacynthe?). It was so much fun to annoy you with naughty French words or criticise you for your constant use of English words while speaking French. I am happy for you that you found this great post-doc in the US and hope to stay in contact with you in future.

I am the loud guy, the guy who laughs a lot and is always open for some stupid ideas. Maria was a nice catalyst to amplify this and I love how we convinced students to join as interns using Freddy the amazing robotic shark. Yeah the interns! How I enjoyed being the mentor of Armande, Olivier and Simon (and not enjoyed others). Thank you for helping in my research and working on the more risky projects! After Maria left my new buddy for stupid ideas became Olivier. We tried challenging us basically to do anything. I am really sorry though that he cannot eat apples in public anymore. Eating only liquid food for a week was maybe more challenging than the entire Ph.D. and how much fun it was to train for this half-Ironman (viruses are such bastards!). I met him as an intern and I am very proud that he is performing so well in his Master's now.

However, often we were not alone in doing the challenges. Which brings me to Jacynthe. Calling her a friend took a while and maybe even a bit longer since she was so shocked not being called my friend after knowing her for 24h. She makes great beer and took Soylent like a boss ☺. I will never forget our movie nights, the Jack Box parties and the love you have for my little girl Julie. Thank you for being in our life and stay the way you are. I cannot wait to show you Germany and I promise, you will take on some weight since we will eat a lot! And I also promise you I will be likeable!

I really love showing my home country to my friends. Natalie came for a surprise visit when Julie was born. What a great moment it was seeing her! While being in many ways extremely different, I love hanging out with you and am glad that we can share funny moments with your family during Easter and Thanksgiving. We share the same enthusiasm for food, drinks and computer games, which we could unfortunately not live to the fullest during your stay in our house. However, you got to know Julie better and I will keep the secret that you like at least one kid ☺.

I was also fortunate to work with Lorea, Sarah and Daniela, whom I want to thank for their support as well. Just in the past months, David and I finally discovered that we share the same love for maths. Who would have guessed that after working together for three years, we realise that just weeks before you left? I wish you all the best for your future! I also want to thank Marc for the bromance we had at university and continue to have now that he works. I often miss his chemist mind in our biochemistry lab. Finally, I want to thank Fabio for always listening to my tiring stories from work. You are a great friend and I know I can always count on you!

Und nun zu meiner Familie. Auch wenn mein lieber Onkel Thomas mein Ph.D. niemals als Doktor anerkennen wird, bin ich sehr dankbar für die Unterstützung aller, für meine Entscheidung im Ausland meine Doktorarbeit zu schreiben. Es war nicht immer leicht für alle uns beide gehen zu sehen. Nach mehreren Auslandsaufenthalten in Südafrika, Kanada und Japan nach der Hochzeit wieder nach Kanada zu gehen und dann gleich für eine so lange Zeit war sicher schwer zu verstehen. Vor allem nachdem wir Julie bekommen haben ist das ganze sicher nicht einfacher geworden. Vielen, vielen Dank, dass ihr trotzdem immer zu uns gehalten habt. Man soll immer auf sein Leben zurück schauen können und ohne Bedauern und ruhigen Gewissens alles getan zu haben, was man wollte. Natürlich geht das nur mit einer wundervollen Partnerin wie Inga, die mir so viel Kraft und Mut gibt für all meine Ideen im Leben. Manchmal stellt sie sich vielleicht zu sehr in den Hintergrund, jedoch bin ich einfach überglücklich damals Quake 3 gespielt zu haben und im IRC am chatten gewesen zu sein. Ich liebe dich über alles und wünsche mir noch viele glückliche Jahre mit dir und wie Oma immer so schön sagt, unserer kleinen Familie. Und wenn Julie dies vielleicht mal lesen wird: vielen Dank für dein Verständnis als Zweijährige und für die Gläser Wasser die du mir als Unterstützung gebracht hast! Du bist der ganze Stolz von Mama und mir! Wir lieben es, dich

aufwachsen zu sehen und hoffen, dass du dir deine gutmütige Art für immer bewahrst. Je t'adore ma puce!

Ich wollte auch Danke an meine Mama sagen, die so viel gearbeitet hat, damit ich meine Ausbildung machen konnte und es bis hierhin geschafft habe. Du bist eine so starke Frau und ich hoffe, dass ich dich stolz mache mit dem was ich tue. Auch wenn wir oft nicht einer Meinung sind, weiß ich ganz genau was für einen großen Anteil du daran hast was ich heute für ein Mensch bin. Leider ist mein Papa nicht mehr hier um meine Doktorarbeit zu erleben, aber ich bin mir sicher, dass er besonders stolz darauf wäre, dass ich sie in Französisch-Kanada gemacht habe. Dafür ist Okko noch da, der den Geist meines Vaters weiter trägt.

Gleich zwei so tolle Omis zu haben wie ich ist ein besonderes Glück. Vielen Dank, dass ihr immer an mich geglaubt und mir das 1x1 beigebracht habt. Der Marterstuhl war hart und unerbittlich, aber er war die Grundlage meiner akademischen Karriere. Ohne euch hätte ich nie das Abitur geschafft. Ich werde auch nie die Reisen nach Mechow oder auf der Zuckerstange vergessen.

Auch meinen Schwiegereltern möchte ich danken. Ich erinnere mich als wäre es gestern, wie stinkig Jogi war als ich Inga zum ersten Mal nach Afrika entführt habe. Das war der Beginn unserer Reise um die Welt und vielleicht schließt sich der Kreis nun. Vielen Dank Marianne und Jogi, dass ihr mich damals bei euch aufgenommen habt und ich bei euch wohnen durfte. Auch wenn meine Schwimmsachen sehr oft in der Wäsche waren habe ich doch noch das Abitur geschafft, sicher auch weil ihr mich persönlich vor dem Schultor abgeliefert habt. Danke für all euer Verständnis für diesen lauten Menschen Max.

Ach Thomas wie lustig es doch immer ist mit dir! Du bist so schön kauzig und ich bin froh, dass du immer für mich da bist. Zusammen mit Heike bildet ihr einen sehr wichtigen Pol in meinem Leben, den ich nicht missen möchte. Julian, Fabian und Sebastian ihr seid ein bisschen wie die Brüder, die ich nie hatte. Wir sehen uns nicht oft, aber ihr begleitet mich nun schon mein ganzes Leben. Vielleicht macht ihr ja auch mal einen Doktor? Einen echten wie Thomas sagen würde! Ich möchte mich auch bei dem Rest meiner Familie bedanken einfach dafür, dass ihr immer da seid, wenn man euch braucht. Gabie, Timo, Sven, Steffi, Rafael, Lina, Sonja, Stefan, Emma, Lena und Dominic ich bin einfach froh, dass es euch gibt.

Thank you, merci und danke!

Dedication

Over the course of history, several landmark discoveries changed the way researchers were studying proteins. At the time of their publication, they were often criticized, rejected or even fought against. One example is the transition from purely inorganic and organic chemistry towards biotransformation. The greatest minds of the time fervently rejected the fact that fermentation was a result of the metabolism of living organisms. Looking back now, it seems unbelievable and maybe even amusing to consider the energy with which the protagonists defended their beliefs. Throughout history, humans encountered many of such turning points not only limited to events in the scientific community. Today, with the increasing availability of computational tools to study fundamental principles of protein science and protein engineering our field is, yet again, facing again such a critical point.

In the past, bioinformatics was essentially used as proof-of-concept and was often not considered on par with experimentation. Thanks to the advancement of technology and algorithms, today's computational tools can help, augment and in some cases replace many of the experiments carried out daily in the laboratory. With a degree in bioengineering, I only had basic knowledge in biochemistry when I started my Ph.D., and no bioinformatics skills at all. Thus, my research began as a pure experimentalist. But thanks to an extraordinary lecturer, I was introduced to the world of computational biochemistry. I am not sure if Prof. Guillaume Lamoureux is aware of the impact he had on my personal development but ever since I attended his class, my curiosity on computational algorithms has been growing. I redefined my research project to include computational tools whenever possible to benefit from its advantages while at the same time relying on thoroughly executed wet-lab experiments. Throughout my studies in Canada, I promoted the use of computational tools as a regular strategy in laboratory experiments, giving rise to many heated discussions among my colleagues. It is my firm belief that, in order to make better enzymes and to learn more about our proteins of interest, we need to exploit all sources of knowledge by combining the strengths of experimental and computational methodologies. This thesis is dedicated to the goal of overcoming our fears and our disregard for these emerging new technologies in scientific research.

Chapter 1 - Introduction

1.1. Overview

The overarching goal of this thesis is to promote the use of computational tools as a routine complement to laboratory experiments in the field of enzymology. The introduction will review main historic events of protein science leading to the development of computational tools to study proteins. I will discuss challenges that we, as a scientific community, have faced in the area of enzyme engineering and present technological milestones that were necessary to overcome them. The most recent advances in computational tools for protein engineering are presented in Chapter 2, under the form of a review article to be published in April 2017 in *Current Opinion in Chemical Biology*. Finally, we briefly visit the highly controversial topic of the impact of protein dynamics on the structure and function of proteins.

1.2. The science of proteins

Following the scientific literature on proteins might give the impression that the ability and the interest in studying proteins are relatively recent. Having been trained in process engineering, I was left with the impression that the making of any significant discovery in the field of protein science requires a multitude of instruments, computers, methods and databases. I discovered that could not have been further from the truth, yet I was puzzled as to how the pioneers could study such a complex macromolecules without today's technology at hand. To honor those men and women who laid the foundation of what we take for granted, I will first revisit some historic milestones. Early work can be too easily lost in the noise of the huge increase of publications on the topic of protein science that followed the discovery of the polymerase chain reaction (PCR) by Kary Mullis in 1983 (Figure 1-1) [1]. Despite being fewer in number, it was those early discoveries that shaped modern protein science and, surprisingly, continue to do so.

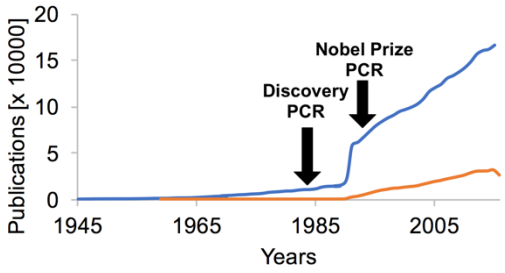


Figure 1-1 Publication history on proteins.
Occurrence of the word "protein" in blue and "PCR" in scientific literature since 1945. Source: Web of Science, visited on December 2016.

1.2.1. First discovery of proteins and protein engineering

Apart from the usual examples of beer, wine and cheese manufacturing, little to nothing is taught about historical use or research on proteins. Without a doubt, humans have applied biotransformations using whole cell catalysis mostly unconsciously or thanks to accidental discovery. This empirical knowledge was used to ferment different types of foods either to preserve their shelf life or in food processing for centuries. Nonetheless, researchers have been conducting research on these remarkable bio-polymers for a long time. Interestingly, researchers started to observe proteins in wheat seeds and gluten, juice or milk as early as 1747 without knowing what they were looking at. A first definition was formalized in 1789, when Antoine Fourcroy published his finding of a molecule with the ability to coagulate in presence of an acid or heat [2]. The molecule remained unnamed for another 49 years, until these polymers were named in 1838 by Gerardus Mulder and Jöns Berzelius [3, 4]. But simply being called protein, from the greek word *proteios* which means "standing in front", was not enough to grasp the importance and chemical makeup of these molecules. Another 64 years went by until, finally, Franz Hofmeister and Emil Fischer recognized the peptide bond as the fundamental feature of proteins that links their amino acid building blocks to form the polypeptide [5, 6].

Ironically, enzymes, which are mainly responsible for all biotransformations, were not considered to be proteins until 1926 when James Sumner discovered that urease is indeed a protein [7]. Nonetheless, even in the absence of this knowledge, naturally occurring enzymes were known and used as early as 1836, with the identification of pepsin by Theodor Schwann

[8]. Efficient methods for the purification, over-expression and isolation of enzymes in large quantities were still elusive at the time, limiting not only the possibility to study them but also their industrial application. Early enzymes were found in large quantities in very specific sources such as blood, egg white, toxins secreted by animals or digestive/metabolic enzymes obtained in offal from slaughterhouses [9]. The first kilogram-scale isolation and purification of an enzyme was accomplished by the hot-dog company Armour Inc. in the 1950s [10]. Bovine pancreatic ribonuclease A was thus made freely available to the scientific community by the company and, hence, became the main protein for biophysical studies for the years to come.

With the advent of defined chromatography matrices in the late 1950s and early 1960s, proteins that were less stable and less abundant became suddenly available [11-13]. Small- and large-scale purification enabled researchers to study a broad variety of protein families. These new possibilities are also reflected in the first large increase in the number of publications on the topic of proteins, around 1965 (Figure 1-1). However, the field was tied to naturally occurring protein sequences and organisms, which dramatically limited the application to industrial processes. Synthetic substrates and products often differ importantly from their natural counterparts. Furthermore, the billions of years of evolution towards the specialisation of enzymes has reduced their versatility. Indeed, certain properties of enzymes that are advantageous for biocatalytic reactions, such as high stereo-, regio-, and chemoselectivity, are disadvantageous where versatility is required. Thus, highly specialized natural enzymes catalyze specific reaction types but perform poorly, with some exceptions, on substrates of industrial interest or in industrial conditions. Catalysts with high thermostability, tolerance to a wide range of organic solvents and pH, high product or substrate concentrations are non-natural reaction conditions but necessary in large scale production processes [14]. The necessity to run enzymatic reactions in aqueous solutions and the fact that most proteins of interest are not naturally over-expressed results in dramatically increased costs for industrial product purification and isolation [15]. Solutions were required in order for enzyme catalysis to become a viable industrial tool.

While methods for molecular cloning were known as early as the 1970s, efforts to optimise an enzymatic reaction relied on naturally-occurring mutations, non-specific mutagenesis (i.e. radiation) or functional homologues from different sources [16, 17].

Researchers soon realized that they would need a method to recombinantly over-express proteins and tailor the properties of an enzyme towards the desired application. Altering the DNA code of the enzyme was the solution, and was pioneered by DNA insertions and deletions in the early 1970s. Facilitated by the technical breakthrough of accessible and inexpensive synthetic oligonucleotides in the late 1970s and early 1980s, the technology was ripe for the development of directed enzyme evolution for industrial applications [18].

1.2.2. First use of enzymes

Early successes in the use of enzymes in industrial processes inspired researchers and process engineers alike to develop new synthetic routes that are more benign, economical and environmentally friendly than conventional total synthesis. However, they were limited to naturally-occurring sequences or mutations until the early 1980s. Besides the classical examples of beer, wine, vinegar and cheese manufacturing using bacteria, yeast and other fungi, it took until 1858 when Louis Pasteur performed the first enzymatic-catalyzed racematic resolution of tartaric acid ammonium salt by *Penicillium glaucum* [19]. On the path to modern biocatalysis an important milestone was reached in 1897 with the finding that cell-free yeast extract is sufficient to achieve the biotransformation of sugar [20]. Even though fiercely criticized by the finest chemists of the time, in particular Justus von Liebig, Jöns Jacob Berzelius, and Friedrich Woehler, fermentation as a process of living organisms became increasingly interesting for researchers and companies alike.

The first large-scale production of fine chemicals was the synthesis of citric acid using fermentation, which was successfully marketed by Pfizer Inc. in 1919 [21]. Soon after, the discovery of Penicillin G and its anti-microbial activity in 1928 by Sir Alexander Fleming raised the question as to how to mass-produce complex natural organic compounds. Fleming quickly realized that *P. notatum* could not produce Penicillin G in large enough quantities for clinical use [22]. Howard Florey and researchers from Oxford University addressed the challenge of formulating the drug to increase its stability and worked to demonstrate its potential life-saving properties. Relying on natural organisms, the worldwide search for a better *Penicillium* strain began in the wake of the second world war [23]. An improved strain was in 1943 in a moldy cantaloupe fruit in a Peoria, Illinois market. This made Penicillin G accessible for mass production by using the corn steep liquor process. The increasing demand

for antibiotics required an optimization of the fermentation process and an increase in product yield. The search for better strains continued and inspired what was perhaps of the first large-scale mutagenesis studies. Indeed, in the late 1930s, researchers discovered the possibility of altering the properties of natural organisms using physical and chemical treatments to generate faster growing strains giving rise to higher product yields and exhibiting tolerance of higher or lower oxygen environments and reduced product or substrate inhibition. In the absence of any molecular biology tools, metabolic engineering by UV-radiation was notably applied by Backus *et al.* who reported, in 1955, the *P. chrysogenum* Q-176 mutant strain which increased production of penicillin nearly ten times. This early example of metabolic engineering represents the ancestor of all strains used in industry today [22, 24].

The commercially important anti-inflammatory drug prednisone (**2**) was produced by Schering and Upjohn by dehydrogenation of cortisone (**1**) using *Corynebacterium simplex* as early as 1955 (Figure 1-2 A) [25]. Another milestone was the preparation of L-phenylacetylcarbinol (**5**) by yeast such as *Saccharomyces cerevisiae* and *Candida utilis* [26]. This synthesis of the precursor for the important drug Ephedrine was industrially established in the 1930s by Knoll AG (Figure 1-2 B) [27].

Whole cell catalysis for fine chemical production using wild-type and genetically modified strains is still an important contributor to industrial biotransformations today. As discussed above, since the 1960s, researchers and industry had the capacity to purify proteins and enzymes on a large scale. Henceforth, purified enzymes emerged as biocatalysts for use in biotransformations as a complement to whole cell catalysis [27, 28]. One early example is the enzymatic preparation of L-amino acids by free or immobilized aminoacylase (Figure 1-2 C) [29, 30]. Research in the Wandrey group and at Degussa resulted in a scaled-up process for the industrial production of L-methionine (**7**). Further industrially important enzymatic reactions are the manipulation of sugars, nucleic acids and lipids [28]. Most notably in terms of annual production worldwide are the production of high-fructose corn syrup, aspartic acid and malic acid.

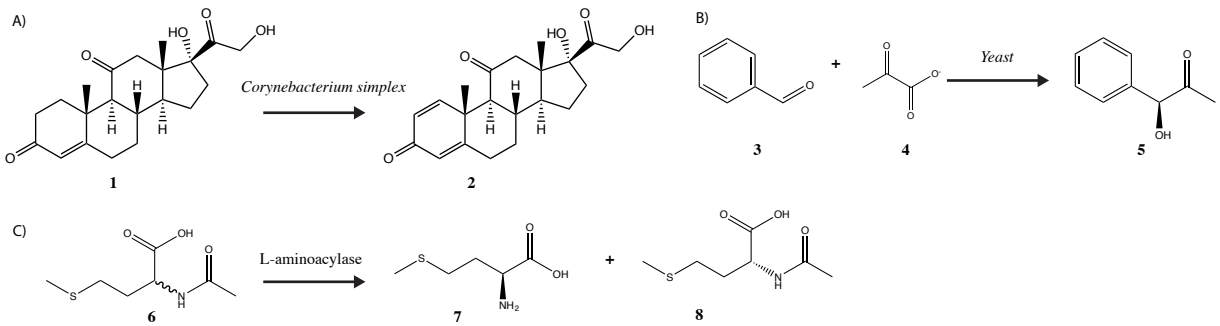


Figure 1-2 Historically important biotransformations.
*A) Dehydrogenation of cortisone (**1**) to prepare prednisone (**2**)*
*B) Condensation of benzaldehyde (**3**) and pyruvate (**4**) to prepare L-phenylacetylcarbinol (**5**)*
*C) Hydrolysis of N-acetyl-rac-methionine (**6**) for the production of L-methionine (**7**) and N-acetyl D-methionine (**8**)*

1.2.3. Beyond natural sequences

Despite the success stories regarding whole cell and enzymatic catalysis using natural sequences or limited mutagenesis, in 1983 only a dozen enzymes had sales exceeding over 10 million dollars per year. These few proteins had a market share of 90% [15]. New technologies were needed to advance the field. As briefly introduced under sub-heading 1.2.1, a methodology was necessary to tailor protein expression and enzymatic function to industrial applications. In the 1970s, researchers started to experiment with recombinant DNA in order to express protein sequences that were non-native to the host [31]. This technology enabled the overexpression of proteins usually found in organisms with slow growth rates. The industrial application of genetically engineered proteins started with the development of recombinant expression of human insulin in *Escherichia coli* by Genentech, later licensed and brought to market by Eli Lilly and Company [32]. Protein overexpression technology enabled the application of proteins and the exploitation of their unique properties in everyday products. It became possible to add amylases, lipases or proteases to washing detergents, rendering them more efficient and active at both higher and at lower temperatures [33, 34].

During that period, researchers and industry were still either using wild-type proteins or randomly mutated strains. The advent of protein crystallography gave researchers insight not only into the composition of proteins but also into their complex three-dimensional structures. The first observation of protein crystal formation was achieved over 150 years ago by Friedrich Hünefeld [35]. He serendipitously found, in 1840, crystals of hemoglobin from

the blood of earthworms simply by slowly drying it between two plates of glass. Slow evaporation remains the foundation of protein crystallization today. Crystals of many other of proteins were grown in the following years, such as urease, insulin, trypsin and lysozyme [36]. Those crystals served to study protein purification or to determine sample purity. X-ray diffractionists started to show interest in protein crystals by the mid-1930s to obtain structural information about the assembly of macromolecules. First experimenting with dry crystals and later with crystals in their wet form, the first structural reconstruction of myoglobin was achieved in 1958 by Kendrew *et al.* (Figure 1-3 A) [37].

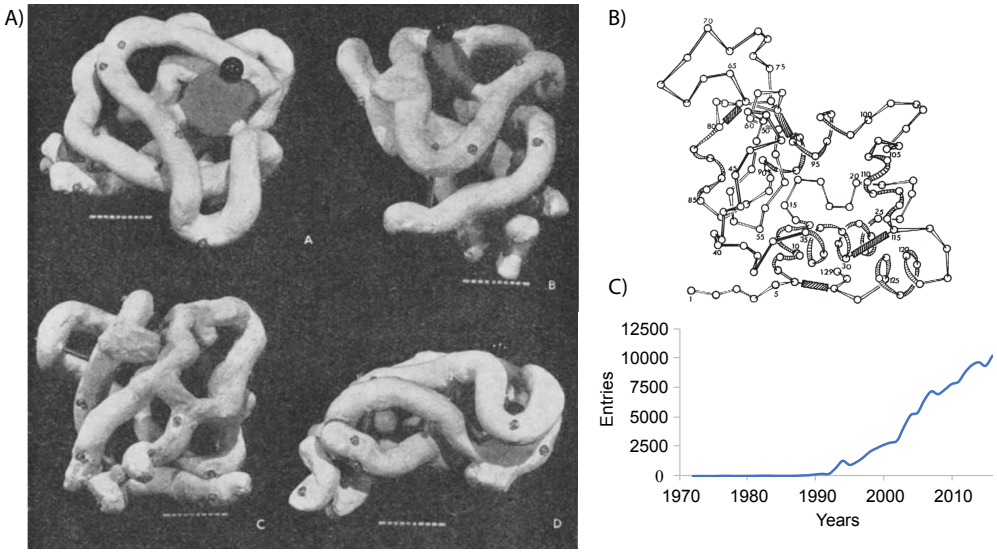


Figure 1-3 The wake of protein crystallography.
A) Representation of the first crystal structure of a protein, myoglobin with heme. Illustration reproduced from Kendrew et al. [37]
B) Atomistic representation of lysozyme, the second structure published. Illustration reproduced from Blake et al. [38]
C) Evolution of the number of protein structures in the PDB database over time

Crystal structures of lysozyme (Figure 1-3 B), ribonuclease S and A, papain, hemoglobin and insulin followed, to be eventually included in the Protein Data Bank (PDB) that was established in 1971 [38-44]. Containing only two structures at first, it has grown to ca. 120,000 structures today, including structures derived from X-ray diffraction, NMR and electron microscopy (EM) (Figure 1-3 C). In the early 1980s, Jane Richardson developed early representations of protein structures, known today as cartoon or ribbon representations [45].

The higher X-ray flux from synchrotron sources and better detectors in the 1980s reduced collection time of the diffraction pattern and simplified subsequence analysis [46]. Computational power became available at the same time, as computing became more affordable and graphic processing was developed [47]. Better algorithms to solve the phase problem and especially the method of molecular replacement helped to reduce the time and effort required to solve a protein structure [48]. Another milestone was presented by Christian Anfinsen who was the first to show a direct link between protein sequence and structure. His book "The Molecular Basis of Evolution" essentially laid the foundation for modern protein science by connecting and integrating, for the first time in a systematic manner, protein chemistry with classical concepts of genetics [49]. The protein amino acid sequence in a polypeptide chain, also called primary sequence, forms a locally folded structure due to hydrogen bonds between atoms of the backbone. The two most common types of this secondary structure are β-sheets and α-helices. To achieve maximum stability and to reach the lowest energy state the secondary structure elements reorient themselves to form the overall three-dimensional shape of the protein or tertiary structure. Finally, a quaternary structure can be reached through protein-protein interactions of multiple protein subunits.

All the basic tools were in place to engineer proteins and it was thought that perfectly tailored proteins were within reach. The idea to create enzymes with novel non-natural functions were already in the minds of the scientific community and it seemed that the solution was right around the corner [15].

1.2.4. Directed evolution and rational design

Among the first efforts in directed mutagenesis was a study on tyrosyl tRNA synthetase by Winter *et al.* in 1982 [50]. Interestingly, while considered as one of the first success stories in the field, the result was mitigated by the reduction in catalytic activity due to an increased K_M for ATP. Nonetheless, the concept of rational design and directed evolution was born and rapidly proved successful for enhancing biophysical properties such as thermostability [51]. The breakthrough technology of PCR, invented in the mid-1980s, led to the widescale development of protein engineering techniques such that many more challenges could be addressed (Figure 1-1) [1]. The most common strategies for mutagenesis include error-prone PCR (random mutagenesis), side-directed saturation mutagenesis and DNA

shuffling. Early studies showed that only a small subset of the amino acid residues of the composing protein is essential for function, folding and stability [52, 53]. Therefore, multiple mutations in the same gene are tolerated, enabling the application of mutagenic rounds until no further improvement of a desired property is observed (Figure 1-4 A and B).

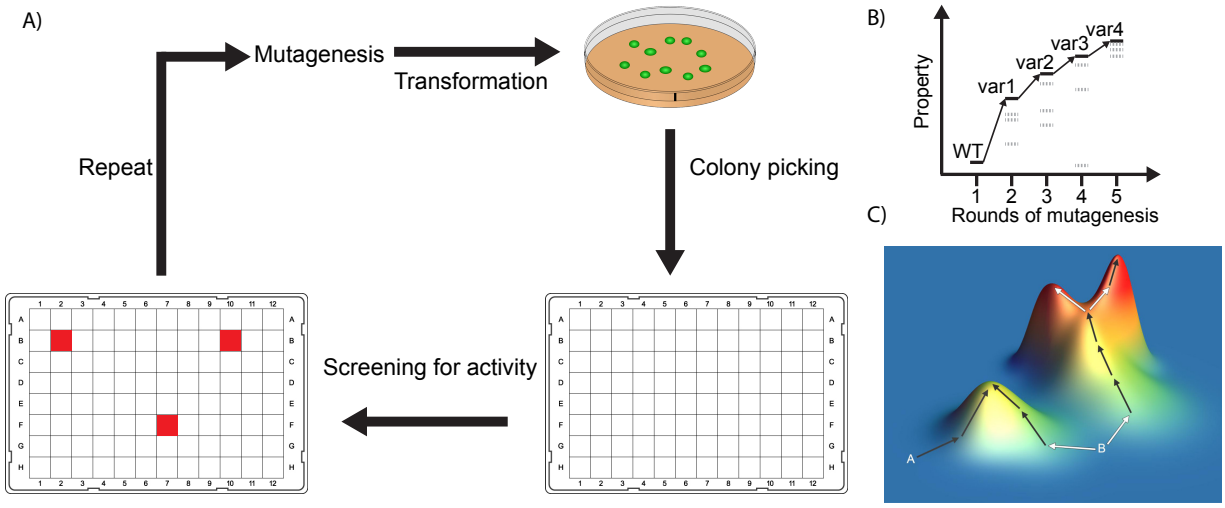


Figure 1-4 The concept of directed evolution.

A) Scheme showing the concept of directed evolution, adapted from [54]. Red squares indicate positive hits of the desired property.

B) Screening for improved property in rounds of evolution until a plateau is reached. Inspired by [55].

C) Protein fitness landscape during directed evolution, where the X and Y axes represent protein sequence space and the Z axis is a desired property, or protein fitness. Variants A and B might both lead to the same local maximum fitness peak after several generations, but only variant B can cover the sequence space necessary to reach the global maximum. Representation reproduced from [56].

In the early 1990s, a study using this paradigm focused on the evolution of subtilisin E to enable catalysis in high concentrations of the industrially-relevant organic solvent dimethylformamide (DMF) [57]. Using three rounds of random mutagenesis, a variant containing six new mutations with 265-fold higher activity in DMF was generated. Numerous other studies using random mutagenesis in the context of directed evolution followed and are still widely practised today. Nonetheless, a major drawback of protein engineering using random mutagenesis is that orders of magnitude more variants can generally be generated in each round than can physically be screened, such that only a fleetingly small fraction of sequence space is actually sampled (Figure 1-4 B and C). When an important path towards the

global maximum fitness of the protein is missed, even several consecutive rounds of evolution might be insufficient to reach the key area of sequence space. As mutations are carried over from one generation to another, one may become trapped in a functional 'island' (or local maximum) without the possibility to further enhance the desired property (Figure 1-4 C: see the smaller of the two peaks).

To that effect, high-throughput screening and selection methods were developed to more effectively cover the generated sequence space. Many easy targets and low-hanging fruits have been studied in the past two decades [58]. The numerous assays developed for screening and selection seldom went further than the proof-of-principal stage, often being too specific for a given substrate or a subset of proteins [59]. Another challenge is the limited sequence coverage resulting from random mutagenesis. Most proteins are in the range of 100~200 amino acids [60], such that there are at least 20^{100} possible mutagenic combinations for even a small protein. Generating all those possible solutions would need more atoms than are available in the universe. Clearly, random mutagenesis explores only an infinitesimal fraction of all sequences that are theoretically available. In addition, the space that is covered is biased: mutagenesis statistically alters one nucleotide of a codon, thus limiting the set of amino acids that can be reached at each position. Indeed, only three to seven amino acid substitutions per residue can be achieved with random mutagenesis methods [61].

Yet again, new technologies and protocols were necessary to achieve a more productive exploration of protein sequence space. In the 1980s and early 1990s, rational protein engineering was made possible by making use of the growing banks of knowledge on protein sequences from different organisms, crystal structures of wild-type and mutant proteins and biophysical data [62]. An interesting application of rational enzyme design for asymmetric synthesis was performed on L-lactate dehydrogenase of *Bacillus stearothermophilus* [63]. Site-directed mutagenesis was based on positions previously identified from structural information, to partially invert stereospecificity.

Thus, rational design was enabled by our increased understanding of fundamental protein function. Site-directed saturation mutagenesis was then introduced: it consists in substituting a residue by all 19 other amino acids to maximize localized sequence exploration. [64]. In addition to being more labor-intensive than random mutagenesis, rational design often produces variants with lower than WT activities, thus requiring screening of a huge number of

mutants. Thus, a smarter way to create libraries of enzyme variants was needed. Again, new technologies needed to be developed to help guide protein engineers to achieve tailored proteins.

1.3. Smart-library design and computational enzyme engineering

Smart-library design was the next milestone in protein engineering. With the goal of further reducing screening efforts, the design of smart libraries depends heavily on the availability of knowledge about the system of interest, as success correlates strongly with their use. Knowledge availability can, however, become our main enemy: genome sequencing, robotic high-throughput screening and the ever-increasing number of publications are so data-rich that it becomes impossible to see the forest for the trees.

To that effect, computational tools have been developed to identify mutational hot-spots, that is to say residues whose mutation is most likely to alter the desired properties. A review of recent advances in computational protein engineering and smart-library generation is provided in Chapter 2. That chapter summarizes computational tools and methods that help to enhance experimental rational protein engineering in the context of biocatalysis. Computational data mining of published studies, protein sequence and structure databases as well as molecular dynamics (MD) simulations provide rich and indispensable knowledge on the system of interest. Even though successful computer-aided engineering has been reported and has been used in industrial applications, many researchers do not apply computational methods in their studies (See Chapter 2) [65]. The reasons are manifold, including the complexity of the methods, the potential unreliability of the predictions and mistrust of a technology that is hard to understand.

In recent years, web-services have become available to simplify access to bioinformatics tools and algorithms have increased reliability. Most of the ready-to-use platforms for the generation of smart libraries or for rational design of proteins consider only protein properties, and are thus independent of a specific substrate of interest. However, enzyme engineering should, ideally, take account of the substrate. The question of how a ligand binds to its target is paramount for drug development and also for the creation of smart

libraries in the context of rational enzyme design. Early 'lock-and-key' and 'induced fit' models of substrate binding [66, 67] were later extended to the 'conformational selection' hypothesis, where the protein is thought to be in constant flux between many conformational sub-states, some among which potentially can bind a specific ligand [68]. This is supported by many docking studies on apo-crystal structures or holo-structures with ligands that differ significantly from the molecule of interest, which fail to find the correct binding pose because the crystallized conformation is not compatible with binding the new ligand [69, 70]. Thus, ligand binding is clearly a dynamic process and can't be viewed as static problem.

For years, MD simulations have been used extensively to study protein dynamics and many methods and protocols were developed to gain detailed information about protein folding, aggregation, catalysis and ligand binding in transmembrane channels and globular proteins. The microscopic nature of proteins would require a quantum mechanically description of the potential energy when studying protein dynamics. These calculations are not feasible to compute a complete macromolecule. Force fields are used to approximate the potential energy using a functional form based on classical mechanics and parameter sets derived from experiments or quantum mechanics calculations [71]. Still computational costs to correctly model protein-ligand binding are high. The main difficulty is the sheer size of the conformational ensemble of a protein structure that result from its numerous degrees of freedom. High energy barriers separating conformational subsets within the ensemble create the need for extremely long simulations, even without considering the presence of a ligand. In Chapter 2, some techniques to predict the substrate binding mode will be discussed. Because most methods are extremely complex to set up and interpret, Chapter 4 presents a new, straightforward protocol to predict the substrate binding mode and hence enable substrate-specific enzyme engineering.

While the support of computer-aided smart-library design, rational protein engineering and directed evolution have brought us closer to the ultimate dream of Ulmer's 1983 review to create perfectly tailored enzymes, some reactions types are still not within reach [15]. Computationally designing a protein from scratch for a specific purpose would fulfil this dream. Recent advances in computational protein design have led to the *de novo* creation of inspiring new enzymes such as the first known Diels-Alderase [72-74]. Thus, 30 years after Ulmer's review, this gives us a glance of what might be possible to achieve in the future,

although the low activity of such artificially created enzymes leaves many questions open for future research.

1.4. Protein dynamics and impact on structure-function relationship

In sub-heading 1.3, it was mentioned that researchers have been aware of the dynamic nature of proteins since the late 1950s [67]. Indeed, proteins are highly dynamic and undergo dynamic processes on different timescales (Figure 1-5). Nonetheless, the dynamic nature of proteins is greatly neglected in all fields of protein research. In the example of the refinement of the protein-ligand interaction model, we can see that ultimately the functions of proteins are governed by their dynamic character.

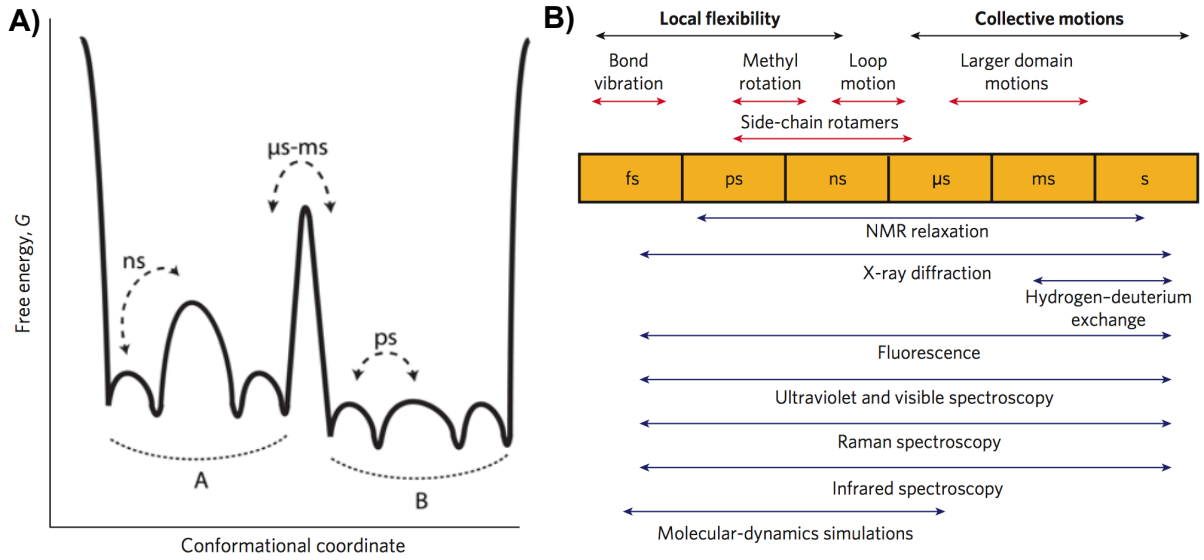


Figure 1-5 Timescale of dynamic processes in proteins.

A) One-dimensional cross-section through the energy landscape of a protein from conformation A to B showing the hierarchy of protein dynamics and the energy barriers.

B) Timescale of dynamic processes in proteins and the experimental methods that can describe them on each timescale. Each timescale describes a certain aspect of function. No experimental technique is able to probe the entire timescale. The timescale covered by each technique does not necessarily deliver uniform data quality since the method of data acquisition is not the same. Inspired by or reproduced from [75] .

Small amplitude atomic movements such as bond-vibrations, side-chain rotations or loop motions, occur on fast timescales, from fs-ns. These movements create sub-states that do not explore real conformational space (Figure 1-5 A) [75]. To transition from a closed to an open conformation of an enzyme active site, protein dynamics on the μs-ms timescale are necessary to allow large amplitude motions such as domain displacement. On the same timescale, we also find that catalytic reactions take place [76]. Hence, one might foresee that protein motions on the μs-ms timescale have a direct impact not only on the structure but also on the function of the protein. A heated contemporary debate on the topic has created two different schools of thought. The first does not believe that protein dynamics play any major role in catalysis. The motions can influence the layout and pre-arrangement of the active site but do not contribute to catalysis by energy transfer [77]. The second defends a direct implication of protein dynamics on enzymatic function, supported by studies showing a correlation between slower catalytic activity and altered protein dynamics by point mutants far from the active site [10, 78]. Part of what separates the two schools is semantics, but there remains a pressing need to study the entire timescale of motions, from fs-ms, to elucidate the impact of dynamics on protein function. This is especially important since motions could be partly responsible for the mitigated success of enzyme engineering and computational protein design [79, 80]. Protein dynamics is not only important for biocatalysis but also plays a major role in protein allostery, switching between active-inactive conformations or flexible ligand binding. For example, transmembrane protein motion is responsible for many gating mechanisms to channel ions or small molecules across cell membranes [81, 82].

While NMR provides useful data on the ps-ns and μs-ms timescale, highly artificial experimental conditions are necessary to study the ns-μs timescale. Other techniques either lack the atomistic description of the system (e.g. spectroscopy) or provide no notion of exact timescale (e.g. crystallography). MD simulations can theoretically cover the entire timescale of all protein properties while providing detailed atomistic information, but are currently limited to the μs by computational resources [83-85]. Another challenging aspect of long-timescale MD simulations is data storage and data analysis. High precision simulation trajectories in the range of low μs can require more than 10 TB of data. To reach the longer simulation times, one can store atomistic snapshots of the system only for increasingly large

time steps, resulting in a loss of information at the fs-ns timescale. We thus need new computational approaches that allow more complete information gathering.

Similarly to protein engineering, where poor results of computational protein design results were improved by laboratory evolution [86, 87], the combination of computational tools and wet-lab experiments ultimately results in a better resolution of protein dynamics. We and others previously showed that MD simulations are in good agreement with dynamics data on the ps-ns timescale. That study is currently under review and is provided in Annex 4. In that example, computational simulations of dynamics on the fs-μs timescale was complemented with CPMG NMR experiments for the μs-ms, producing reliable continuous data spanning the entire protein dynamics timescale (see Figure 1-5 B) [83, 88].

Aside from their impact on function, protein dynamics undoubtably have a major influence on protein structure. The protein structure is a conformational ensemble of which the crystal structure represents one possible energetically favorable sub-state. In addition, ligand binding is highly influenced by, and indisputably linked to, dynamic motions within the protein. Many examples of protein gating mechanisms are known in all types of proteins, including trans-membrane protein channels, protein transporters and globular proteins. Studying those mechanisms require tools similar to those in Figure 1-5 B.

1.5. Conclusion

Interdisciplinary research is required in many fields, including studying proteins. Rather than focusing on either computational or the wet-lab experiments, we should embrace both worlds and use the most appropriate technique or a combination of both, as discussed in a review article presented in Chapter 2. As shown over the course of history, changes in the field have always created resentment and scepticism. To accept emerging techniques requires courage but is worthwhile for the scientific community. In the following chapters, we will visit different aspects of protein research that have highly benefited from computationally-assisted methods.

The research component of this thesis will start, in Chapter 3, with an example that would have highly benefited from computational smart-library design. It is followed, in

Chapter 4, by an example illustrating how substrate-specific prediction of mutational hotspots can help to characterize an enzyme.

The thesis will further demonstrate that the field of protein engineering stands to benefit from increased synergy between experimental data such as that obtained by X-ray crystallography, nuclear magnetic resonance spectroscopy or small angle X-ray scattering, and molecular dynamics simulations. In Chapter 5, we illustrate that the simple combination of experimental and computational work enables the study of protein motions over most of the timescales relevant to protein dynamics. We computationally address protein dynamics and its impact on the structure-function relationship; this topic has been a heated debate in recent years as researchers attempt to find patterns and rules for how protein dynamics affects reactivity, substrate recognition or protein stability.

Finally, computational tools can nowadays answer questions that are not within reach of experimental methods or would need to create a system far from the natural environment such as deficient mutants, non-natural solvents or extremely high protein concentrations. Chapter 6 will present such an example, which helped to correlate and explain experimental data acquired on O_2 binding in P450 enzymes with computational predictions on gas transport.

1.6. References

1. Bartlett, J.M. and D. Stirling, *A short history of the polymerase chain reaction.* Methods Mol Biol, 2003. **226**: p. 3-6.
2. Osborne, T.B., *The vegetable proteins.* 1924.
3. Hartley, H., *Origin of the Word Protein.* Nature, 1951. **168**(4267): p. 244-244.
4. Miquel, F., G. Mulder, and W. Wenckebach, *Bulletin des sciences physiques et naturelles en Néerlande*. Vol. 2. 1838: Chez PH van der Heuvell.
5. Hofmeister, F., *Über Bau und Gruppierung der Eiweisskörper.* Ergebnisse der Physiologie, 1902. **1**(1): p. 759-802.
6. Fischer, E., *Über eine neue Aminosäure aus Leim.* Berichte der deutschen chemischen Gesellschaft, 1902. **35**(3): p. 2660-2665.
7. Sumner, J.B., *The isolation and crystallization of the enzyme urease preliminary paper.* J Biol Chem, 1926. **69**(2): p. 435-441.

8. Fruton, J.S., *A history of pepsin and related enzymes.* Q Rev Biol, 2002. **77**(2): p. 127-147.
9. Perrett, D., *From 'protein' to the beginnings of clinical proteomics.* Proteom Clin Appl, 2007. **1**(8): p. 720-738.
10. Gagne, D. and N. Doucet, *Structural and functional importance of local and global conformational fluctuations in the RNase A superfamily.* Febs J, 2013. **280**(22): p. 5596-5607.
11. Hjerten, S., *The Preparation of Agarose Spheres for Chromatography of Molecules and Particles.* Biochim Biophys Acta, 1964. **79**: p. 393-398.
12. Porath, J. and P. Flodin, *Gel filtration: a method for desalting and group separation.* Nature, 1959. **183**(4676): p. 1657-1659.
13. Peterson, E.A. and H.A. Sober, *Chromatography of Proteins .1. Cellulose Ion-Exchange Adsorbents.* J Am Chem Soc, 1956. **78**(4): p. 751-755.
14. Arnold, F.H., *Directed evolution: Creating biocatalysts for the future.* Chem Eng Sci, 1996. **51**(23): p. 5091-5102.
15. Ulmer, K.M., *Protein Engineering.* Science, 1983. **219**(4585): p. 666-671.
16. Cohen, S.N., et al., *Construction of biologically functional bacterial plasmids in vitro.* Proc Natl Acad Sci USA, 1973. **70**(11): p. 3240-3244.
17. Nathans, D. and H.O. Smith, *Restriction endonucleases in the analysis and restructuring of dna molecules.* Annu Rev Biochem, 1975. **44**: p. 273-293.
18. Shortle, D., D. Dimaio, and D. Nathans, *Directed Mutagenesis.* Annu Rev Genet, 1981. **15**: p. 265-294.
19. Pasteur, L., *Mémoire sur la fermentation applée lactique. Mémoire sur la fermentation alcoolique*. 1857: Mallet-Bachelier.
20. Buchner, E. and R. Rapp, *Alkoholische gährung ohne hefezellen.* Berichte der deutschen chemischen Gesellschaft, 1897. **30**(3): p. 2668-2678.
21. Max, B., et al., *Biotechnological Production of Citric Acid.* Braz J Microbiol, 2010. **41**(4): p. 862-875.
22. Kardos, N. and A.L. Demain, *Penicillin: the medicine with the greatest impact on therapeutic outcomes.* Appl Microbiol Biot, 2011. **92**(4): p. 677-687.

23. Raper, K.B., *The Development of Improved Penicillin-Producing Molds.* Ann Ny Acad Sci, 1946. **48**(2): p. 41-56.
24. Backus, M.P. and J.F. Stauffer, *The Production and Selection of a Family of Strains in Penicillium Chrysogenum.* Mycologia, 1955. **47**(4): p. 429-463.
25. Nobile, A., et al., *Microbiological Transformation of Steroids .1. Delta-1,4-Diene-3-Ketosteroids.* J Am Chem Soc, 1955. **77**(15): p. 4184-4184.
26. Neuberg, C. and J. Hirsch, *Über ein kohlenstoffketten knüpfendes ferment (carboligase).* Biochemische Zeitschrift, 1921. **115**: p. 282-311.
27. Straathof, A.J., S. Panke, and A. Schmid, *The production of fine chemicals by biotransformations.* Curr Opin Biotechnol, 2002. **13**(6): p. 548-556.
28. Whitesides, G.M. and C.H. Wong, *Enzymes as Catalysts in Synthetic Organic-Chemistry.* Angew Chem Int Edit, 1985. **24**(8): p. 617-638.
29. Chibata, I., et al., *Production of L-amino acids by aminoacylase adsorbed on DEAE-Sephadex.* Methods Enzymol, 1976. **44**: p. 746-759.
30. Wandrey, C. and E. Flaschel, *Process development and economic aspects in enzyme engineering. Acylase L-methionine system*, in *Advances in Biomedical Engineering*. 1979, Springer. p. 147-218.
31. Jackson, D.A., R.H. Symons, and P. Berg, *Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli.* Proc Natl Acad Sci USA, 1972. **69**(10): p. 2904-2909.
32. Johnson, I.S., *Human Insulin from Recombinant DNA Technology.* Science, 1983. **219**(4585): p. 632-637.
33. Andree, H., W.-R. Muller, and R.D. Schmid, *Lipases as detergent components.* J Appl Biochem, 1980. **2**(3): p. 218-229.
34. Falbe, J. and R.D. Schmid, *Biotechnology and the Surfactants and Detergents Industry.* Fett Wiss Technol, 1986. **88**(6): p. 203-212.
35. Hünefeld, F.L., *Der Chemismus in der thierschen Organisation... Ein Beitrag zur Physiologie und Heilmittellehre, etc*. 1840.
36. Mcpherson, A., *A Brief-History of Protein Crystal-Growth.* J Cryst Growth, 1991. **110**(1-2): p. 1-10.

37. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-666.
38. Blake, C., et al., *Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution.* Nature, 1965. **206**(4986): p. 757-761.
39. Kartha, G., J. Bello, and D. Harker, *Tertiary structure of ribonuclease.* 1967.
40. Wyckoff, H., et al., *The structure of ribonuclease-S at 3.5 A resolution.* J Biol Chem, 1967. **242**(17): p. 3984-3988.
41. Drenth, J., et al., *Structure of papain.* 1968.
42. Perutz, M., et al., *Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model.* Nature, 1968. **219**: p. 131-139.
43. Blundell, T., et al., *Atomic positions in rhombohedral 2-zinc insulin crystals.* Nature, 1971. **231**: p. 506-511.
44. Bernstein, F.C., et al., *The protein data bank.* Eur J Biochem, 1977. **80**(2): p. 319-324.
45. Richardson, J.S., *Early ribbon drawings of proteins.* Nat Struct Biol, 2000. **7**(8): p. 624-625.
46. Phizackerley, R.P., et al., *Progress Report on the Development of an Area Detector Data Acquisition-System for X-Ray Crystallography and Other X-Ray-Diffraction Experiments.* Nucl Instrum Methods, 1980. **172**(1-2): p. 393-395.
47. Pottle, C., et al., *Conformational-Analysis of Proteins - Algorithms and Data-Structures for Array-Processing.* J Comput Chem, 1980. **1**(1): p. 46-58.
48. Rossmann, M.G., *The Molecular Replacement Method.* Acta Crystallogr A, 1990. **46**: p. 73-82.
49. Anfinsen, C.B., *The molecular basis of evolution.* John Wiley & Sons, 1959.
50. Winter, G., et al., *Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding.* Nature, 1982. **299**(5885): p. 756-758.
51. Pantoliano, M.W., et al., *Large Increases in General Stability for Subtilisin Bpn' through Incremental Changes in the Free-Energy of Unfolding.* Biochemistry-Us, 1989. **28**(18): p. 7205-7213.
52. Heinz, D.W., W.A. Baase, and B.W. Matthews, *Folding and Function of a T4 Lysozyme Containing 10 Consecutive Alanines Illustrate the Redundancy of*

Information in an Amino-Acid-Sequence. P Natl Acad Sci USA, 1992. **89**(9): p. 3751-3755.

53. Poteete, A.R., D. Rennell, and S.E. Bouvier, *Functional-Significance of Conserved Amino-Acid-Residues.* Proteins, 1992. **13**(1): p. 38-40.
54. Reetz, M.T., *Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions.* Angew Chem Int Ed, 2011. **50**(1): p. 138-174.
55. Reetz, M.T., *Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future.* J Am Chem Soc, 2013. **135**(34): p. 12480-12496.
56. Palmer, A.C. and R. Kishony, *Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance.* Nat Rev Genet, 2013. **14**(4): p. 243-248.
57. Chen, K.Q. and F.H. Arnold, *Tuning the Activity of an Enzyme for Unusual Environments - Sequential Random Mutagenesis of Subtilisin-E for Catalysis in Dimethylformamide.* P Natl Acad Sci USA, 1993. **90**(12): p. 5618-5622.
58. Cherry, J.R. and A.L. Fidantsef, *Directed evolution of industrial enzymes: an update.* Curr Opin Biotech, 2003. **14**(4): p. 438-443.
59. Goldsmith, M. and D.S. Tawfik, *Directed enzyme evolution: beyond the low-hanging fruit.* Curr Opin Struc Biol, 2012. **22**(4): p. 406-412.
60. Trifonov, E.N. and I.N. Berezovsky, *Evolutionary aspects of protein structure and folding.* Curr Opin Struc Biol, 2003. **13**(1): p. 110-114.
61. Wong, T.S., et al., *A statistical analysis of random mutagenesis methods used for directed protein evolution.* J Mol Biol, 2006. **355**(4): p. 858-871.
62. Lutz, S., *Beyond directed evolution-semi-rational protein engineering and design.* Curr Opin Biotech, 2010. **21**(6): p. 734-743.
63. Sakowicz, R., M. Gold, and J.B. Jones, *Partial Reversal of the Substrate Stereospecificity of an L-Lactate Dehydrogenase by Site-Directed Mutagenesis.* J Am Chem Soc, 1995. **117**(9): p. 2387-2394.
64. Jones, J.B. and G. Desantis, *Toward understanding and tailoring the specificity of synthetically useful enzymes.* Accounts Chem Res, 1999. **32**(2): p. 99-107.
65. Savile, C.K., et al., *Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture.* Science, 2010. **329**(5989): p. 305-309.

66. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme.* Berichte der deutschen chemischen Gesellschaft, 1894. **27**(3): p. 2985-2993.
67. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* P Natl Acad Sci USA, 1958. **44**(2): p. 98-104.
68. Ma, B.Y., et al., *Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations.* Protein Sci, 2002. **11**(2): p. 184-197.
69. Grebner, C., et al., *Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design.* J Chem Inf Model, 2016. **56**(4): p. 774-787.
70. Lape, M., C. Elam, and S. Paula, *Comparison of current docking tools for the simulation of inhibitor binding by the transmembrane domain of the sarco/endoplasmic reticulum calcium ATPase.* Biophys Chem, 2010. **150**(1-3): p. 88-97.
71. Leach, A.R., *Molecular modelling: principles and applications*. 2001: Pearson education.
72. Siegel, J.B., et al., *Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction.* Science, 2010. **329**(5989): p. 309-313.
73. Kaushik, M., et al., *Protein engineering and de novo designing of a biocatalyst.* J Mol Recognit, 2016. **29**(10): p. 499-503.
74. Huang, P.S., S.E. Boyken, and D. Baker, *The coming of age of de novo protein design.* Nature, 2016. **537**(7620): p. 320-327.
75. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-972.
76. Wolfenden, R. and M.J. Snider, *The depth of chemical time and the power of enzymes as catalysts.* Accounts Chem Res, 2001. **34**(12): p. 938-945.
77. Pisliakov, A.V., et al., *Enzyme millisecond conformational dynamics do not catalyze the chemical step.* P Natl Acad Sci USA, 2009. **106**(41): p. 17359-17364.
78. Bhabha, G., et al., *A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis.* Science, 2011. **332**(6026): p. 234-238.
79. Lippow, S.M. and B. Tidor, *Progress in computational protein design.* Curr Opin Biotechnol, 2007. **18**(4): p. 305-311.

80. Klinman, J.P. and A. Kohen, *Hydrogen tunneling links protein dynamics to enzyme catalysis.* Annu Rev Biochem, 2013. **82**: p. 471-496.
81. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Biol, 2002. **9**(9): p. 646-652.
82. Lindahl, E. and M.S. Sansom, *Membrane proteins: molecular dynamics simulations.* Curr Opin Struct Biol, 2008. **18**(4): p. 425-431.
83. Klepeis, J.L., et al., *Long-timescale molecular dynamics simulations of protein structure and function.* Curr Opin Struct Biol, 2009. **19**(2): p. 120-127.
84. Lindorff-Larsen, K., et al., *Picosecond to Millisecond Structural Dynamics in Human Ubiquitin.* J Phys Chem B, 2016. **120**(33): p. 8313-8320.
85. Lane, T.J., et al., *To milliseconds and beyond: challenges in the simulation of protein folding.* Curr Opin Struct Biol, 2013. **23**(1): p. 58-65.
86. Rothlisberger, D., et al., *Kemp elimination catalysts by computational enzyme design.* Nature, 2008. **453**(7192): p. 190-195.
87. Karanicolas, J., et al., *A de novo protein binding pair by computational design and directed evolution.* Mol Cell, 2011. **42**(2): p. 250-260.
88. Huang, J. and A.D. MacKerell, Jr., *CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data.* J Comput Chem, 2013. **34**(25): p. 2135-2145.

Chapter 2 - Recent developments in computational enzyme engineering

2.1. Context

This chapter is a "pre-print" version of an invited review for the journal *Current Opinion in Chemical Biology*, accepted in December 2016. Entitled *'Computational tools for enzyme improvement: why everyone can – and should – use them.'*, it covers recent advances in biocatalyst development using computational methodologies in conjunction with laboratory experiments. The review was written from the perspective of a novice user of computational tools, with a strong emphasis on tested, ready-to-use algorithms and web-based services. All tools presented can readily be used by computationally-inexperienced experimentalists to optimize screening efforts in the context of enzyme engineering. The review also introduces a new area in computation for enzyme engineering: prediction of the enzyme-substrate complex. Considering that this is a variant of protein-ligand complex prediction, the presented tools are also useful for the characterization of drug targets. In the wake of numerous reviews on the topic of computational protein engineering, we also stress the emerging problem of ensuring continuous availability and maintenance of computational tools and services.

As discussed in Chapter 1, the massive datasets that protein engineers now generate need to be efficiently analyzed to be integrated in our rational protein design strategies, and also should be stored for next generations. Computational tools for data mining are continuously being created, and journals such as *Bioinformatics* and *Nucleic Acids Research* present many publications on this topic. However, maintaining and updating the created computational tools, protocols and algorithms to ensure that they remain active and functional is becoming a major concern. Scientific funding is not adapted to maintaining the *status quo*: support favors emerging technologies that may be unstable rather than perpetuating and expanding validated technologies. This monetarization strategy has successfully supported scientific development of the past centuries but fails when it comes to computational services, which are provided free of charge by and for the scientific community. If we want to continue

to benefit from the work of bioinformatics groups, we have to rethink how we fund research in a sustainable manner.

My contribution to this review was the conceptualization, literature search and writing, which was supported by Prof. Joelle Pelletier.

Computational tools for enzyme improvement: why everyone can – and should – use them.

Maximilian C.C.J.C. Ebert[1,2] and Joelle N. Pelletier[1,2,3]

[1] Département de biochimie and Center for Green Chemistry and Catalysis (CGCC), Université de Montréal, Montréal, QC, Canada H3T 1J4

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, QC, Canada G1V 0A6

[3] Département de chimie, Université de Montréal, Montréal, QC, Canada H3T 1J4

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

2.2. Abstract

This review presents computational methods that experimentalists can readily use to create smart libraries for enzyme engineering and to obtain insights into protein-substrate complexes. Computational tools have the reputation of being hard to use and inaccurate compared to experimental methods in enzyme engineering, yet they are essential to probe datasets of ever-increasing size and complexity. In recent years, bioinformatics groups have made a huge leap forward in providing user-friendly interfaces and accurate algorithms for experimentalists. These methods guide efficient experimental planning and allow the enzyme engineer to rationalize time and resources. Computational tools nevertheless face challenges in the realm of transient modern technology.

2.3. Introduction

The evolution of environmental regulations to reward eco-responsible practises is rapidly shifting the focus of the chemical and pharmaceutical industries towards applying enzymes as 'greener' catalysts in complex organic syntheses. Early successes in the biocatalytic production of fine chemicals have shown that investments in developing such environmentally-friendly process chemistry can be more than offset by increased cost-effectiveness [1]. Target reactions range from those that benefit from the exquisite regio-, enantio- and chemoselectivity of enzymes to less complex reactions where the high reactivity or eco-friendliness of enzymes are advantageous relative to conventional catalysis.

Enzyme engineering is required because the high specificity of enzymes comes with a trade-off. For instance, while possessing the chemical reactivity of interest, the active site of a natural biocatalyst may be unable to bind and transform a desired industrial substrate [2]. In those cases, extensive active-site engineering is necessary to achieve product formation. In other cases, low substrate loads, moderate catalytic efficiency under process conditions or poor operational stability can require extensive, labor-intensive and expensive engineering to tailor biocatalysts to industrial processes. In contrast to conventional catalysis, the lack of a toolbox, or standard workflow, facilitating the development of biocatalyzed processes is delaying the development and adoption of industrial biocatalysts. Indeed, biocatalysts must be optimized for genuine industrial reaction conditions so as to not overtax downstream processing and thus maintain their advantage over traditional synthesis [3].

Early directed evolution experiments on the basis of random mutagenesis gave promising results [4, 5]. However, attempts to sample a meaningful fraction of the vast number of possible mutations soon gave way to the recognition that the chemical space covered was limited: the substitution of two nucleotides within one codon is seldom achieved [6]. For this reason, random mutagenesis can only reach three to seven amino acid substitutions per residue [7]. Even if high-quality, unbiased random libraries are obtained, there remains the often insurmountable problem of screening: tailoring high-throughput screening methods to specific catalytic reactions is not trivial, and is often limited to model systems or to low numbers [8]. The generation of 'smart libraries' that are of smaller size because they target specific residues – typically, those involved in function – has been shown

to cover functional space more effectively [9-11]. However, generating smart libraries typically requires structural and functional knowledge including information on ligand binding, such that for many systems the choice of residues to include in such a library might not be evident.

Here, we present recent advances in computational tools for enzyme engineering with a focus on the generation of substrate-specific smart libraries and finding the active-site pocket. We address non-experts in the field and mainly discuss tools that are readily accessible to would-be users, rather than developer-specific material. We will not cover *de novo* computational protein design methods, directing the reader to a thorough and excellent review by Huang *et al*. [12]. We also refer the reader to a recent review on the computational design of protein-protein interfaces [13].

2.4. What knowledge is available at the outset, and what computational methods does this knowledge enable?

The computational tools available to identify mutational hotspots for the generation of smart libraries can be grouped into two large fields: knowledge based, and simulations. We first discuss knowledge-based methods, which entail mining databases for information on the system of interest (Figure 2-1). However, some enzymes of industrial interest do not belong to a well-populated dataset of homologs. Reliable information on these is sparsely available or elusive. In other cases, substrate specificity is very broad or the active-site conformation is strongly altered in presence of a substrate. For the latter, knowledge-based approaches fail to provide the information required to generate a meaningful smart library. Those shortcomings can be addressed with computational tools and methods that use molecular dynamics simulations and docking, which we present in the second part of the review (Figure 2-1).

2.5. Beyond multiple sequence alignment

The alignment of multiple genes is a computationally-solved problem. It allows the identification of consensus residues that are of particular importance for protein stability and function. Increased computational capacity and availability of structural data now also enables structural alignment of more distantly-related proteins. In parallel, increased sequencing

capacity means that genomic databases hold more information of uncharacterized enzymes with potential different substrate profiles. As the costs of gene synthesis drops, these natural sequences become an alternative to mutagenesis efforts.

The consensus sequence approach identifies conserved residues using multiple sequence alignments (Figure 2-1 A) [14]. High conservation of a residue correlates mainly with protein stability, but is also relevant to properties such as reactivity or substrate selectivity. Consensus residues thus constitute targets for diversification or may be kept strictly constant to prevent loss of function. This straightforward method has been successfully used, mainly to tailor the thermostability of various enzymes for industrial applications, as nicely reviewed [15-18]. An often-overlooked application is the capacity to modulate catalytic efficiency and activity. An automated web-based protocol to identify conserved residues is provided by Hot-Spot Wizard 2.0. Requiring only the sequence of interest as an input, it provides the user with a list of stability and functional hot-spots, their mutability, a list of suitable mutations, a design for appropriate degenerate codons and, if applicable, a protein tunnel analysis [19]. We also invite the reader to try ConSurf, MAP2.0 3D or Selecton [20-22], which offer a similar palette of automated analyses.

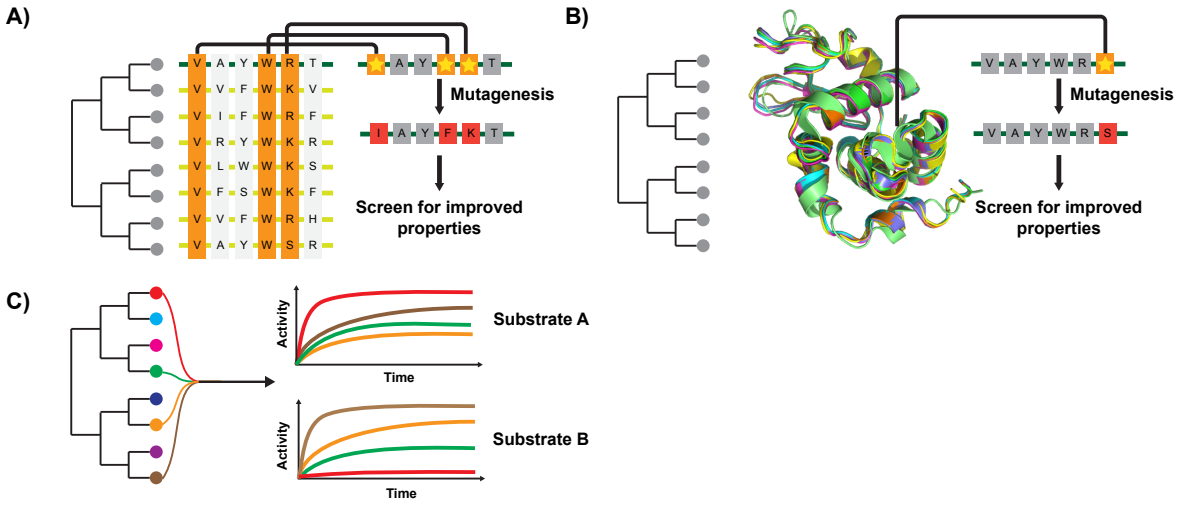


Figure 2-1 Overview of different strategies to generate new protein sequences.
A,B: Finding mutational hotspots by identification of conserved residues using (A) multiple sequence alignment or (B) structural alignment. The latter can successfully align protein folds even when sequence identity is low. C) Using naturally-occurring homologs identified by multiple sequence or structural alignment as a source of diversity; here, substrate specificity is tested.

Using these tools in conjunction with site-specific saturation mutagenesis, researchers enhanced the methanol stability of *Geobacillus stearothermophilus* Lipase T6 66-fold [23]. Thermostability increased by 4.3°C and the yield of fatty acid methyl ester from soybean oil increased two-fold. These results are significantly better than a random mutagenesis approach performed in parallel, where methanol stability was improved only 23-fold while reducing methanolysis activity by 50%. In another study, the catalytic efficiency of α-1,3-fucosyltransferase for the production of 3-fucosyllactose, a potent inhibitor of bacterial or viral adhesion to human epithelial cells, was enhanced by generating a smart library using Hot-Spot Wizard [24]. The best variant gave a 96% yield after 1h, a 40-fold increase in yield and 560-fold greater productivity than the native enzyme.

Because protein structure is more highly conserved than protein sequence throughout evolution, sequence alignment may fail to identify functional homologs [25, 26]. It takes only the conservation of a few residues, for instance a catalytic triad, to conserve activity even in the context of sequence identity as low as 20% (Figure 2-1 B). Strikingly, over 3% of protein structures deposited in the PDB are annotated as having unknown function [27], illustrating that structural space is a rich place for discovery. However, conducting a structural search to identify distant homologs among the > 120000 structures in the PDB is vastly more complex and computationally demanding than is a one-dimensional sequence alignment.

Many different web services are available for structural comparison of a structure of interest against the PDB database, including Dali, FATCAT, MATRAS and PDBeFold [28-31]. Similar to BLAST, these enable the researcher to identify close homologs by comparing the protein fold instead of the sequence. Using the structural search tool Dali, a dataset of proteins sharing no significant sequence similarity with proteins of known function yielded a 17% hit-rate upon structural alignment with structural homologs of known function [27]. For example, the hypothetical protein PA2260 was identified as being a putative carbohydrate isomerase. As a result of being a distant homolog, it may exhibit a distinct substrate scope, which may be difficult to engineer by directed evolution. The Dali server also helped to identify structural homologs of the NahF salicylaldehyde dehydrogenase, despite sharing only 35% sequence identity with other crystallized aldehyde dehydrogenases [32].

A powerful alternative to generating new enzymatic function by mutagenesis consists in mining the natural sequences available in sequence databases. The function of less than 1% of protein sequences has been characterized [33], illustrating the magnitude of this opportunity (Figure 2-1 C). In a recent study, genomic data-mining using UniProtKB and BLAST was successfully applied to identify 26 new biocatalysts from various species, expanding the chemical toolbox to include new aldolase, retroaldolase and transferase activities [34]. In an attempt to gain green access to chiral Vince lactam, a non-heme chloroperoxidase (SvGL) with promiscuous (−)-γ-lactamase activity was identified from *Streptomyces viridochromogenes* [35]. As gene synthesis becomes fast and affordable, cloning is unnecessary making these buried sequences easily accessible [36, 37].

Gleaning information from distantly related natural diversity can also serve to identify hotspots for the generation of smart libraries, as a substitute for building a consensus between close homologs. The subsequent sequence alignment of the above-mentioned aldehyde dehydrogenase structures as well as binding-site predictions generated by the CASTp server [38] helped to interpret the function of active-site residues and shed light on the catalytic mechanism [32]. In another study, asymmetric reduction of α-halo-ketones with *Candida glabrata* ketoreductase 1 (CgKR1) was achieved using prior knowledge of the better characterized yeast methylglyoxal/isovaleraldehyde reductase Gre2 [39]. Their similarity was established by using the crystal structure of the open form of CgKR1 as the template for a structural search against the PDB. They identified a high similarity with the open form of Gre2. Then, based on the crystalized closed form of Gre2, the closed form of CgKR1 was modeled and used in a docking study with Autodock Vina [40] to identify mutational hotspots in the active site. Single and double mutants exhibiting excellent stereoselectivity (> 99% ee) towards 2-chloroacetophenone, 2-chloro-4-fluoroacetophenone and 2-bromoacetophenone were thus obtained.

2.6. Substrate-specific hotspot prediction

The computational methods reviewed above can provide a good estimate of the shape and location of an active site as well as giving good guidance in identifying mutational hotspots. However, the above methods do not directly consider the substrate of interest.

Indeed, alignment methods provide valuable information even in absence of prior knowledge on the location of substrate binding, yet their utility is greatly augmented by such knowledge. There are many systems where substrate-specific knowledge is required. In the case of enzymes with a naturally broad substrate scope and enzymes that are being engineered to bind a new substrate or inhibitor, it is of prime importance to correctly pinpoint the residues holding the greatest potential to modulate binding of specific ligands. To this effect, virtual substrate screening using molecular docking tools such as Autodock VINA, Autodock, Glide, Gold etc. have shown great potential [40-44](Figure 2-2 A). By providing a structure of a protein (receptor) and ligand of interest, molecular docking will predict possible binding poses and help to identify mutational hot-spots in a ligand-specific manner. These tools are either free or commercially available, provide good documentation and are user-friendly, and as a result are seeing their use increase rapidly. They are best used in conjunction with ready-to-dock 3D ligand structures from databases such as ZINC, the Cambridge Structural Database (CSD) or PDB [45-47].

Based on the positions predicted in a docking study, mutational hotspots for long-chain fatty acids were determined in the cytochrome P450 CYP153A$_{M.aq.}$ [48]. Double and triple mutants with 2-14 times higher catalytic activity and altered substrate specificity were found. Using a similar approach, another study focused on increasing the solubility of polyphenols to make them more industrially tractable. Glycolysation using sucrose phosphorylase (SP) offers a possible solution yet shows no significant activity toward the substrates of interest. A combined docking and simulation study helped to identify mutational hotspots on an active-site loop of a thermostable SP from *Thermoanaerobacterium thermosaccharolyticum* [49]. A point mutation then increased the active-site cavity volume to procure the desired activity.

An interesting combination of computational techniques and modern experimental methods was applied to render an alcohol dehydrogenase from *Thermoethanolicus brockii* active towards the asymmetric reduction of difficult-to-reduce ketones such as tetrahydrofuran-3-one and tetrahydrothiofuran-3-one [50]. In that study, molecular docking of tetrahydrofuran-3-one was used to select residues to mutate. Upon mutating the desired positions using only the top codons at each position (following saturation mutagenesis), the best (*R*)-selective double mutant showed 99% ee towards tetrahydrofuran-3-one and the best (*S*)-selective triple mutant 95% ee, compared to 23% ee for the native enzyme.

Despite its advantages, molecular docking has evident shortcomings. For reliable results, the receptor conformation needs to be resolved with high confidence in the presence of a ligand that is homologous to ligands of interest. A factor contributing to the failure of docking approaches is that enzymes may undergo significant structural modification upon substrate binding. Despite recent advances in flexible receptor docking, these methods introduce a strong computational burden [51]. As a result, docking often fails to identify the correct binding mode and hence the false-positive rate of identified mutational hotspots is high [52]. Recently, several successes in structural ensemble docking have been reported (Figure 2-2 B) [53]. The rationale is to build an ensemble of conformational variants of the target protein, against which to dock ligands. The ensemble can be generated either by using multiple available crystal structures, SAXS, EM, molecular dynamics (MD) simulations or NMR structures. Those ensembles can be either obtained from databases such as PDB and SASBDB, or can be generated from a MD trajectory or using web services such as ENCoM [54, 55].

A comparative study has demonstrated an improvement of docking results when using a MD-generated conformational ensemble relative to docking against a simple structure. The model system was an epoxide hydrolase containing two different substrate-binding sites. While docking to the ensemble allowed identification of the correct binding pose using the generated MD ensemble for some substrates, other substrates showed worse performance; furthermore, using the holoenzyme structure as a template gave better results than using the apoenzyme structure[56]. Thus, current docking methods that attempt to capture conformational flexibility of the active site are progressing, but may not yet reliably recreate the relevant active-site conformations for specific substrates.

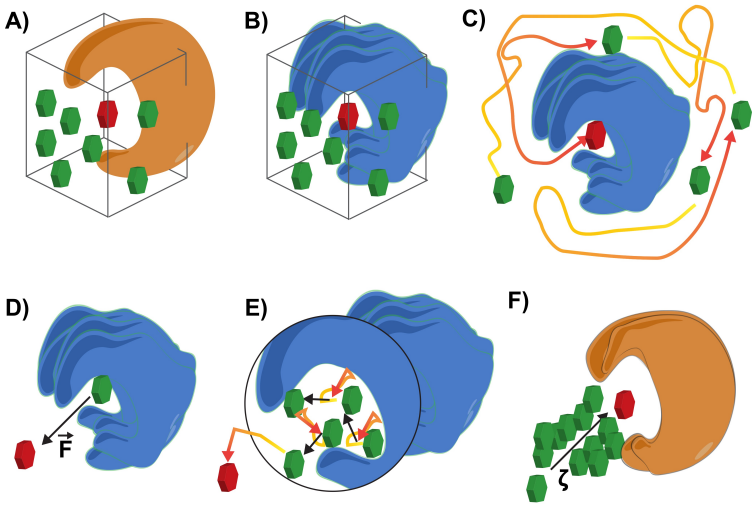


Figure 2-2 Strategies to identify protein-ligand complexes.
A) Molecular docking of a substrate into a static receptor conformation to predict the binding pose and identify promising residues for mutagenesis. B) Same as A) but using a structural ensemble of conformations generated by NMR, MD or using several crystal structures. C) Binding pose search using a free MD simulation of receptor-ligand complex. D) Introduction of bias into a MD simulation in the form of a force along a predefined reaction coordinate. The egress path contains information about initial binding residues and ligand access tunnel topology. Methods include adaptive-biasing force, umbrella sampling and steered MD. E) RAMD is a combination of C) and D) to predict a putative path of ligand egress. An unbiased simulation is carried out until the displacement of the ligand becomes trapped in a local energy minimum. An external biasing force with random direction is applied and an unbiased MD continues until completion of ligand egress. F) The PELE algorithm works along a reaction coordinate by placing a ligand of interest in every possible orientation, while at the same time locally perturbing residue orientation in response to the ligand. This introduces limited unbiased dynamics in the protein.

Several groups have attempted to fully simulate a ligand-receptor complex. During ligand binding to or egress from the active site, the ligand encounters barriers as well as favorable interactions. Extremely long MD simulations are necessary to cover those events, which generally occur on the timescale of the µs to ms (Figure 2-2 C) [57]. Those simulations require both the expertise and a computational infrastructure that are not accessible to most researchers. To overcome those limitations, methods have been developed that introduce a 'computational bias', where energy is applied to cross the encountered barriers and to leave regions of favorable interactions where the ligand would otherwise remain 'stuck'. Those

methods include steered molecular dynamics (SMD) (Figure 2-2 D) or random accelerated (expulsion) molecular dynamics (RAMD) (Figure 2-2 E), among others, and are used to characterize buried active sites or to dock a ligand [58-61].

In SMD, a ligand is placed either in the bulk media or in the active site and is pulled with a constant force or velocity along a reaction coordinate until it reaches the final orientation (Figure 2-2 D). In RAMD, the ligand is allowed to wander out of the active site during a free MD simulation; whenever the ligand does not significantly move over a predefined period, a force of constant energy and random direction is applied to set it back in motion. Cycles of free MD and applied force continue until egress of the ligand (Figure 2-2 E). Both techniques overcome the energy barriers encountered during ligand translation. Mainly applied to studies of protein drug targets, SMD and RAMD can be directly extended to biocatalytic applications. For example, SMD and RAMD were applied to characterize the tunnel network of human microsomal cytochrome P450 2E1 and to identify the main substrate release channels. In addition, tunnel gating residues were identified, which are potential mutational hotspots for altering substrate specificity [62]. In another study, rational design of *Bacillus thermocatenulatus* lipase (BTL2) was carried out using a combined MD, SMD and docking approach [63]. A SMD simulation of the protein-ligand complex identified potential mutations sites to modify triglyceride chain-length selectivity. The most promising single mutant showed a 2-fold higher preference for short-chain triglycerides.

A more accessible approach is Protein Energy Landscape Exploration (PELE), which does not rely on full protein simulation and is hence a computationally faster approach for tunnel exploration and binding site identification (Figure 2-2 F) [64]. This web-based protocol was used to engineer a pH-stable manganese peroxidase (MnP): based on the PELE hotspot identification, residues of the active but pH-sensitive MnP4 from *Pleurotus ostreatus* were introduced into the scaffold of a pH-stable but poorly active MnP6 from *Ceriporiopsis subvermispora.* The resulting pH-stable double mutant displayed higher affinity but lower turnover rate than MnP4 [65].

For all above-mentioned calculations of ligand-protein interactions, an accurate force field definition of the ligand is the key to success. Until recently, these definitions were complex and required extensive quantum mechanical calculations. Recent advances in general force field definitions (for example GAFF and CGenFF) and the availability of web-based

services to automatically derive atom partial charges (RESP ESP charge Derive Server) have simplified the parameterization of ligands for use in simulations [66-69].

2.7. The transient nature of computational tools

In contrast to conventional enzyme engineering where publication of a detailed protocol suffices to ensure longstanding reproducibility by the community, computational methods for protein engineering rely on code of a complexity too great to be modified, or even directly used, by the non-computationally experienced lab biochemist. Code needs to be maintained and updated in order to remain functional in the context of constant updates of operating systems, server environments or database connectivity, thus requiring dedicated resources. Whether standalone tools or web servers, operating costs are a major challenge for the community of developers and of users. To illustrate this point, we refer to comprehensive 2014 volume entitled *Directed Evolution Library Creation Methods and Protocols* that listed 38 computational tools to facilitate the design of smart libraries [70]. At the moment of writing this review, nine of those tools (nearly 25%) are partially or completely broken. Most of those algorithms are heavily interfaced with large, existing databases such as UniProtKB, PDB, or alignment algorithms such as BLAST. They must thus be constantly updated to ensure continued compatibility even as those central databases evolve and adapt to increasing volume and variety of datasets. Hot-Spot Wizard 2.0, Caver 3.0, CASTp, or RESP ESP charge Derive Server, are examples of well-maintained tool [19, 38, 67, 71].

If we, as a community of enzyme engineers, want a stable access to these increasingly powerful and user-friendly computational tools that augment the capacity to improve enzymes, we need to find a monetarization strategy to support the programming efforts of computational chemists as well as the hardware required to enable the use of these tools (disk space, computational time, etc.). Indeed, users are unlikely to have the programming skills, time and/or access to appropriately configured calculation clusters to reproduce all the methods they want to run. We must make our common need known to federal funding agencies to ensure long-term support of these enabling technologies.

2.8. Conclusion

While the scientific literature is rife with reviews on the topic of computational tools for protein engineering, it is remarkable that the referenced use of those tools in research articles is slim. Success stories where computational methods have accelerated enzyme engineering can be found, yet these tools have clearly not been routinely adopted to the extent of PCR, for example. In fact, today's twist on Westheimer's ironic quote would be: 'A month in the laboratory can often save an hour using computational methods and simulations'. Many accurate computational tools today are user-friendly and accessible; biocatalyst engineering will benefit from better exploiting the advantages provided by computational methods in conjunction with experimental methods.

2.9. Acknowledgement

We thank Daniela Quaglia for fruitful discussions. MCCJCE is the recipient of a Vanier Doctoral Scholarship.

2.10. Annotations

••[1]

A model for the substitution of traditional organic chemistry with biocatalysis in the synthesis of a blockbuster drug. Computationally-guided enzyme engineering lead to an efficient, economical, and environmentally benign process at an industrial scale.

••[3]

Academic development of biocatalysts is seldom closely linked to industrial process conditions, such that the resulting biocatalysts cannot be industrially implemented. This review examines parameters to be engineered in the biocatalyst and the modifications to be undertaken in the associated processes.

•[12]

Although beyond the scope of this review, is an interesting read on the next level of protein engineering. Includes the most recent updates on the generation of de novo designed enzymes and functions.

••[19]

The most recent version of a user-friendly and well-maintained web service for the generation of smart libraries. A good starting point requiring little knowledge of computational tools. Also of high value to those seeking to try rational design.

••[23]

This study beautifully compares 'smart-library' design with random mutagenesis. It illustrates the limited chemical space covered by random mutagenesis. A perfect reminder of why the use of computational tools helps to accelerate protein engineering.

•[38]

This intuitive web-service helps to select residues in the active site walls for mutagenesis.

•[41]

Nice comparison between different docking methods and the use of PELE to predict the ligand-binding mode and thus guide 'smart-library' design. The study shows what works and what doesn't in the context of docking in structural ensembles, single receptors or receptor conformations obtained from PELE simulations.

•[45]

Outstanding database with ready-to-dock ligands either for the generation of small targeted libraries or pre-defined subsets for virtual screening. Also provides a list of vendors to purchase hit compounds.

•[50]

Interesting combination of molecular docking and the concept of triple-code saturation mutagenesis to be used in conjunction with CASTing.

••[63]
A model on how to implement several computational tools in concert with laboratory experiments, illustrated on an important biocatalyst. The authors advantageously apply a wealth of available tools to steer their mutagenesis approach and reduce sequence space towards a reduced number of smart targets.

•[70]
Well-written book that covers core aspects on the creation of mutant libraries. Handy tool for every researcher to get up-to-date, from laboratory experiments to computational methods.

2.11. References

1. Savile, C.K., et al., *Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture.* Science, 2010. **329**(5989): p. 305-309.
2. Woodley, J.M., *Protein engineering of enzymes for process applications.* Curr Opin Chem Biol, 2013. **17**(2): p. 310-316.
3. Lima-Ramos, J., W. Neto, and J.M. Woodley, *Engineering of Biocatalysts and Biocatalytic Processes.* Top Catal, 2014. **57**(5): p. 301-320.
4. Chen, K.Q. and F.H. Arnold, *Enzyme Engineering for Nonaqueous Solvents - Random Mutagenesis to Enhance Activity of Subtilisin-E in Polar Organic Media.* Bio-Technol, 1991. **9**(11): p. 1073-1077.
5. You, L. and F.H. Arnold, *Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide.* Protein Eng, 1996. **9**(1): p. 77-83.
6. Wong, T.S., et al., *A statistical analysis of random mutagenesis methods used for directed protein evolution.* J Mol Biol, 2006. **355**(4): p. 858-871.
7. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. **13**(1): p. 19-25.

8. Goldsmith, M. and D.S. Tawfik, *Directed enzyme evolution: beyond the low-hanging fruit.* Curr Opin Struc Biol, 2012. **22**(4): p. 406-412.
9. Verma, R., U. Schwaneberg, and D. Roccatano, *Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering.* Comput Struct Biotechnol J, 2012. **2**: p. e201209008.
10. Nobili, A., et al., *Use of 'small but smart' libraries to enhance the enantioselectivity of an esterase from Bacillus stearothermophilus towards tetrahydrofuran-3-yl acetate.* Febs J, 2013. **280**(13): p. 3084-3093.
11. Lutz, S. and U.T. Bornscheuer, *Protein engineering handbook*. 2012: John Wiley & Sons.
12. Huang, P.S., S.E. Boyken, and D. Baker, *The coming of age of de novo protein design.* Nature, 2016. **537**(7620): p. 320-327.
13. Rosenfeld, L., et al., *Protein Engineering by Combined Computational and In Vitro Evolution Approaches.* Trends Biochem Sci, 2016. **41**(5): p. 421-433.
14. Lehmann, M., et al., *The consensus concept for thermostability engineering of proteins: further proof of concept.* Protein Eng, 2002. **15**(5): p. 403-411.
15. Jackel, C., et al., *Consensus Protein Design without Phylogenetic Bias.* J Mol Biol, 2010. **399**(4): p. 541-546.
16. Porebski, B.T. and A.M. Buckle, *Consensus protein design.* Protein Eng Des Sel, 2016. **29**(7): p. 245-251.
17. Rodriguez-Larrea, D., et al., *Role of conservative mutations in protein multi-property adaptation.* Biochem J, 2010. **429**: p. 243-249.
18. Vazquez-Figueroa, E., J. Chaparro-Riggers, and A.S. Bommarius, *Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept.* ChemBioChem, 2007. **8**(18): p. 2295-2301.
19. Bendl, J., et al., *HotSpot Wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering.* Nucleic Acids Res, 2016. **44**(W1): p. W479-487.
20. Stern, A., et al., *Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach.* Nucleic Acids Res, 2007. **35**: p. W506-W511.

21. Verma, R., U. Schwaneberg, and D. Roccatano, *MAP(2.0)3D: A Sequence/Structure Based Server for Protein Engineering.* Acs Synth Biol, 2012. **1**(4): p. 139-150.
22. Ashkenazy, H., et al., *ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.* Nucleic Acids Res, 2010. **38**: p. W529-W533.
23. Dror, A., et al., *Protein engineering by random mutagenesis and structure-guided consensus of Geobacillus stearothermophilus Lipase T6 for enhanced stability in methanol.* Appl Environ Microbiol, 2014. **80**(4): p. 1515-1527.
24. Choi, Y.H., et al., *Solubilization and Iterative Saturation Mutagenesis of alpha1,3-fucosyltransferase from Helicobacter pylori to enhance its catalytic efficiency.* Biotechnol Bioeng, 2016. **113**(8): p. 1666-1675.
25. Illergard, K., D.H. Ardell, and A. Elofison, *Structure is three to ten times more conserved than sequence-A study of structural response in protein cores.* Proteins: Struct, Funct, Bioinf, 2009. **77**(3): p. 499-508.
26. Bonneau, R., et al., *Functional inferences from blind ab initio protein structure predictions.* J Struct Biol, 2001. **134**(2-3): p. 186-190.
27. Nadzirin, N. and M. Firdaus-Raih, *Proteins of Unknown Function in the Protein Data Bank (PDB): An Inventory of True Uncharacterized Proteins and Computational Tools for Their Analysis.* Int J Mol Sci, 2012. **13**(10): p. 12761-12772.
28. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**: p. W545-W549.
29. Ye, Y.Z. and A. Godzik, *FATCAT: a web server for flexible structure comparison and structure similarity searching.* Nucleic Acids Res, 2004. **32**: p. W582-W585.
30. Kawabata, T., *MATRAS: a program for protein 3D structure comparison.* Nucleic Acids Res, 2003. **31**(13): p. 3367-3369.
31. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-2268.
32. Coitinho, J.B., et al., *Structural and Kinetic Properties of the Aldehyde Dehydrogenase NahF, a Broad Substrate Specificity Enzyme for Aldehyde Oxidation.* Biochemistry, 2016. **55**(38): p. 5453-5463.

33. Apweiler, R., et al., *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Res, 2012. **40**(D1): p. D71-D75.
34. Guerard-Helaine, C., et al., *Genome Mining for Innovative Biocatalysts: New Dihydroxyacetone Aldolases for the Chemist's Toolbox.* Chemcatchem, 2015. **7**(12): p. 1871-1879.
35. Yin, J.G., et al., *Green access to chiral Vince lactam in a buffer-free aqueous system using a newly identified substrate-tolerant (-)-gamma-lactamase.* Catal Sci Technol, 2016. **6**(16): p. 6305-6310.
36. Cameron, D.E., C.J. Bashor, and J.J. Collins, *A brief history of synthetic biology.* Nat Rev Microbiol, 2014. **12**(5): p. 381-390.
37. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**(7397): p. 185-194.
38. Dundas, J., et al., *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.* Nucleic Acids Res, 2006. **34**: p. W116-W118.
39. Qin, F.Y., et al., *Engineering of Candida glabrata Ketoreductase 1 for Asymmetric Reduction of alpha-Halo Ketones.* ACS Catal, 2016. **6**(9): p. 6135-6140.
40. Trott, O. and A.J. Olson, *Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading.* J Comput Chem, 2010. **31**(2): p. 455-461.
41. Grebner, C., et al., *Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design.* J Chem Inf Model, 2016. **56**(4): p. 774-787.
42. Verdonk, M.L., et al., *Improved protein-ligand docking using GOLD.* Proteins: Struct, Funct, Genet, 2003. **52**(4): p. 609-623.
43. Friesner, R.A., et al., *Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* J Med Chem, 2004. **47**(7): p. 1739-1749.
44. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-2791.
45. Irwin, J.J. and B.K. Shoichet, *ZINC - A free database of commercially available compounds for virtual screening.* J Chem Inf Model, 2005. **45**(1): p. 177-182.

46. Allen, F.H., *The Cambridge Structural Database: a quarter of a million crystal structures and rising.* Acta Crystallogr B, 2002. **58**: p. 380-388.
47. Berman, H.M., et al., *The protein data bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-242.
48. Notonier, S., et al., *Semirational Protein Engineering of CYP153A(M.aq.)-CPRBM3 for Efficient Terminal Hydroxylation of Short- to Long-Chain Fatty Acids.* ChemBioChem, 2016. **17**(16): p. 1550-1557.
49. Dirks-Hofmeister, M.E., et al., *Creating Space for Large Acceptors: Rational Biocatalyst Design for Resveratrol Glycosylation in an Aqueous System.* Angew Chem Int Edit, 2015. **54**(32): p. 9289-9292.
50. Sun, Z.T., et al., *Catalytic Asymmetric Reduction of Difficult-to-Reduce Ketones: Triple-Code Saturation Mutagenesis of an Alcohol Dehydrogenase.* ACS Catal, 2016. **6**(3): p. 1598-1605.
51. Fischer, M., et al., *Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery.* Nat Chem, 2014. **6**(7): p. 575-583.
52. Lape, M., C. Elam, and S. Paula, *Comparison of current docking tools for the simulation of inhibitor binding by the transmembrane domain of the sarco/endoplasmic reticulum calcium ATPase.* Biophys Chem, 2010. **150**(1-3): p. 88-97.
53. Totrov, M. and R. Abagyan, *Flexible ligand docking to multiple receptor conformations: a practical alternative.* Curr Opin Struc Biol, 2008. **18**(2): p. 178-184.
54. Frappier, V., M. Chartier, and R.J. Najmanovich, *ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability.* Nucleic Acids Res, 2015. **43**(W1): p. W395-400.
55. Valentini, E., et al., *SASBDB, a repository for biological small-angle scattering data.* Nucleic Acids Res, 2015. **43**(D1): p. D357-D363.
56. Kotev, M., R. Soliva, and M. Orozco, *Challenges of docking in large, flexible and promiscuous binding sites.* Bioorg Med Chem, 2016. **24**(20): p. 4961-4969.
57. Shan, Y.B., et al., *How Does a Drug Molecule Find Its Target Binding Site?* J Am Chem Soc, 2011. **133**(24): p. 9181-9183.

58. Li, W., et al., *Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations.* Proteins, 2011. **79**(1): p. 271-281.

59. Xu, Y.C., et al., *How does Huperzine A enter and leave the binding gorge of acetylcholinesterase? Steered molecular dynamics simulations.* J Am Chem Soc, 2003. **125**(37): p. 11340-11349.

60. Li, W.H., et al., *Possible pathway(s) of testosterone egress from the active site of cytochrome P4502B1: A steered molecular dynamics simulation.* Drug Metab Dispos, 2005. **33**(7): p. 910-919.

61. Klvana, M., et al., *Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations.* J Mol Biol, 2009. **392**(5): p. 1339-1356.

62. Shen, Z.H., et al., *Investigation of Indazole Unbinding Pathways in CYP2E1 by Molecular Dynamics Simulations.* Plos One, 2012. **7**(3).

63. Durmaz, E., S. Kuyucak, and U.O. Sezerman, *Modifying the catalytic preference of tributyrin in Bacillus thermocatenulatus lipase through in-silico modeling of enzyme-substrate complex.* PEDS, 2013. **26**(5): p. 325-333.

64. Borrelli, K.W., et al., *PELE: Protein energy landscape exploration. A novel Monte Carlo based technique.* J Chem Theory Comput, 2005. **1**(6): p. 1304-1311.

65. Acebes, S., et al., *Rational Enzyme Engineering Through Biophysical and Biochemical Modeling.* ACS Catal, 2016. **6**(3): p. 1624-1629.

66. Wang, J.M., et al., *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-1174.

67. Vanquelef, E., et al., *RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Res, 2011. **39**: p. W511-W517.

68. Vanommeslaeghe, K. and A.D. MacKerell, *Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing.* J Chem Inf Model, 2012. **52**(12): p. 3144-3154.

69. Vanommeslaeghe, K., E.P. Raman, and A.D. MacKerell, *Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges.* J Chem Inf Model, 2012. **52**(12): p. 3155-3168.
70. Gillam, E.M., J.N. Copp, and D.F. Ackerley, *Directed Evolution Library Creation.* Methods Mol Biol, 2014.
71. Chovancova, E., et al., *CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures.* PLoS Comput Biol, 2012. **8**(10).

Chapter 3 - Semi-rational exploration of sequence space of the primitive R67 dihydrofolate reductase enzyme

3.1. Context

The Pelletier research group has long been interested in protein engineering and directed evolution towards new enzymatic functions. Examples from literature demonstrated the potential to build function into simple folds and primitive proteins [1-3]. This motivated the group to work on R67 dihydrofolate reductase (DHFR), a primitive enzyme catalyzing the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF). Because THF is an essential DNA precursor, a selection based on bacterial proliferation enabled the semi-rational design of alternative active sites with native-like or improved activity for the natural substrate [4].

In the course of that study, limited knowledge on the R67 DHFR protein family had precluded attempting rational protein engineering. Only three active variants were identified in that semi-rational study out of the 1536 multiply-mutated active-site variants that were generated. This suggests that a relatively small sequence space of simultaneous mutations is compatible with activity for that homotetrameric enzyme. However, none of the targeted active-site residues were indispensable for folding, stability or function.

In an attempt to break the symmetric nature of the protein and potentially increase the available sequence space for mutagenesis, the publication that follows describes the creation of a chimeric enzyme where two of the four protomers were fused, forming a tetramer of two dimers. By way of this protein engineering approach, we show that the available sequence space for mutagenesis was substantially enlarged, increasing the tolerance to active-site substitutions.

This study was an early example of combinatorial active-site mutagenesis, and also represented an important milestone in my personal development. We had focused mutagenesis on residues situated directly in the active site, using a previously published combinatorial library of monomer mutants [4]. This work relied exclusively on experimental methods, inspection of crystal structures and literature research, and has since led me to wonder if the question we asked would have benefited from recently-developed computational methods such

as ‘smart-library’ design or prediction of the ligand binding mode. While characterizing the functional variants for this publication I also worked with my colleague Dominic Bastien on the discovery of inhibitors for the same enzyme [5]. Docking studies of these inhibitors, mimicking the natural substrates, revealed new anchoring points outside of the main active site cavity. While being beyond the scope of the study presented in this chapter, this encounter with computational prediction of binding residues marked the beginning of my interest in bioinformatics. It triggered my curiosity about the workings of computational tools, what they can provide and how to make them an integrated part of my research.

The article of this chapter is a “pre-print” version of the 2015 publication in *Protein Science* entitled *‘Asymmetric mutations in the tetrameric R67 dihydrofolate reductase reveal high tolerance to active-site substitutions’* [6]. My contribution to this research paper was the characterization of the mutants, undertaken in part under the supervision of Dr. Jordan Volpato. All authors contributed to the conceptualization. Library design was done by Prof. Andreea Schmitzer, library creation by Dr. Jordan Volpato and Prof. Andreea Schmitzer, and library selection by Dr. Krista Morley. All research was performed in the laboratory of Prof. Joelle Pelletier. The manuscript was drafted mainly by myself and by Dr. Morley with assistance from Prof. Pelletier and revision by all authors.

Asymmetric Mutations in the Tetrameric R67 Dihydrofolate Reductase Reveal High Tolerance to Active-Site Substitutions

Maximilian C. C. J. C. Ebert[1,2,3], Krista L. Morley[2,4,5], Jordan P. Volpato[1,2], Andreea R. Schmitzer[1,3] and Joelle N. Pelletier[1,2,3,4]

[1] Département de biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada

[2] PROTEO, the Québec Network for Protein Function, Structure and Engineering, Québec, Canada

[3] CGCC, the Center for Green Chemistry and Catalysis, Montréal, Canada

[4] Département de chimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada

[5] Current address: National Research Council Canada, Montréal, Québec H4P 2R2 Canada

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

3.2. Abstract

Type II R67 dihydrofolate reductase (DHFR) is a bacterial plasmid-encoded enzyme that is intrinsically resistant to the widely-administered antibiotic trimethoprim. R67 DHFR is genetically and structurally unrelated to *E. coli* chromosomal DHFR and has an unusual architecture, in that four identical protomers form a single symmetrical active site tunnel that allows only one substrate binding/catalytic event at any given time. As a result, substitution of an active-site residue has as many as four distinct consequences on catalysis, constituting an atypical model of enzyme evolution. Although we previously demonstrated that no single residue of the native active site is indispensable for function, library selection here revealed a strong bias toward maintenance of two native protomers per mutated tetramer. A variety of such 'half-native' tetramers were shown to procure native-like catalytic activity, with similar K_M values but k_{cat} values 5- to 33-fold lower, illustrating a high tolerance for active-site substitutions. The selected variants showed a reduced thermal stability (T_m approximately 12ºC lower), which appears to result from looser association of the protomers, but generally showed a marked increase in resilience to heat denaturation, recovering activity to a significantly greater extent than the variant with no active-site substitutions. Our results suggest that the presence of two native protomers in the R67 DHFR tetramer are sufficient to provide native-like catalytic rate and thus ensure cellular proliferation.

Keywords: trimethoprim resistance, combinatorial mutations, thermostability, active site engineering, homotetramer

3.3. Introduction

Dihydrofolate reductases (DHFRs) play a critical role in proliferation of all living cells. They catalyze the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), using NADPH as the hydride-donating cofactor. Among its metabolic roles, THF is required for the biosynthesis of purines and thymidylate. As a result, inhibition of DHFR leads to cell death, making this enzyme a main target of antibiotic and anticancer drugs over the past six decades.

Trimethoprim (TMP) is a low-cost and effective bacteriostatic antibiotic that is used worldwide in the prophylaxis and treatment of bacterial infections, both in humans and livestock [7]. It is an effective competitive inhibitor of the chromosomally-encoded DHFR in many bacteria and parasites; for example, it inhibits *E. coli* DHFR with K_i = 0.4 nM [8]. Its widespread use has resulted in the development of microbial resistance to TMP, in the form of highly transmissible plasmid-encoded DHFRs. TMP-resistant DHFRs (type I-XVII) are classified into two gene families: *dfrA* and *dfrB* [9, 10]. The *dfrA* family DHFRs are close genetic variants of the monomeric chromosomally-encoded DHFRs that are 160-190 amino acids long. In contrast, *dfrB* genes encode shorter sequences, on the order of 78 amino acids, that homotetramerize to form the active DHFR species. They are genetically and structurally unrelated to chromosomal DHFRs. Type II R67 DHFR belongs to this second class. Its single, symmetrical active-site pore characterized by 222 symmetry lies at the heart of a channel formed by the assembly of four protomers that are largely constituted of a SH3-like β-sandwich fold (Figure 3-1).

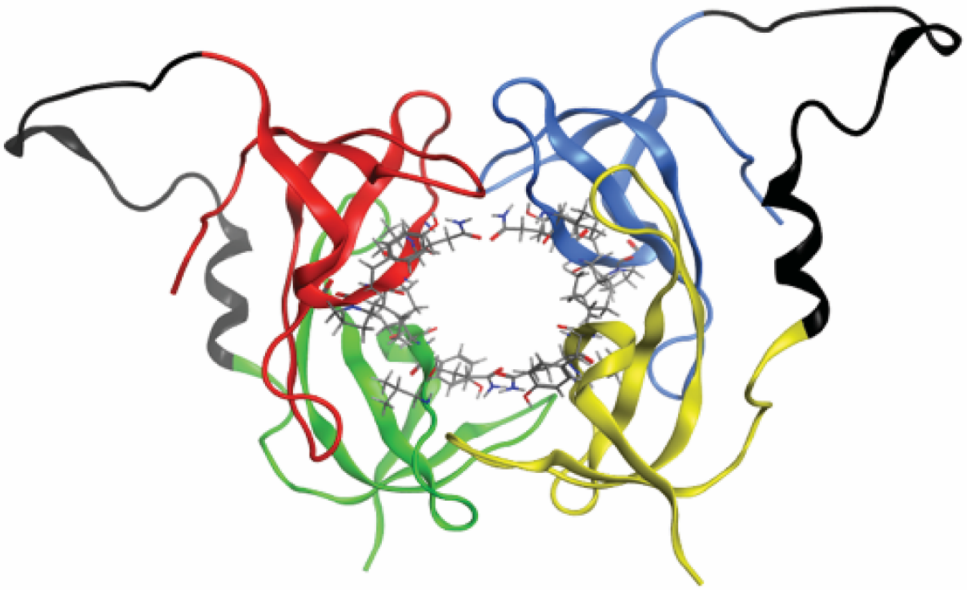


Figure 3-1 Structure of R67 DHFR.
The four protomers (A-D) are illustrated in red, green, blue and yellow, respectively. The active site residues 66-69 are shown in sticks representation. The linker introduced for the dimer construct and the N-terminal tail of R67 connecting protomers A-B and C-D are shown in black. The linker conformation was predicted using the loop-modelling tool of the MOE software [11].

Short loops and both termini extend out of each β-sandwich, forming the inter-protomer interfaces [12-14]. As R67 DHFR is unrelated to its chromosomal counterparts, it has a different catalytic mechanism as well as a different response to antifolate inhibitors [4, 10, 15-18]. It has been qualified as being 'primitive', in that its catalytic strategy lacks in binding specificity, suggesting that it may not yet be evolutionarily optimized [19]. Importantly, as a result of being unrelated to dfrA family DHFRs, it confers resistance to high concentrations of TMP (K_i = 0.15 mM) [20].

Here, we investigate the tolerance to active-site modification of this compact reductase. One β-strand per protomer (V66, Q67, I68, Y69) constitutes the major portion of the active site channel (Figure 3-1). The DHF substrate and NADPH cofactor have been proposed to slip into the channel from opposite ends to juxtapose their reactive atoms at the center of the channel, each forming contacts mainly with one protomer, though additional stabilizing contacts are contributed by the three remaining protomers [21, 22]. For this reason, residues involved in ligand binding can have up to four different consequences within a single

binding/catalytic event [15]. This makes the active site poorly tolerant to point mutations because of a cumulative effect of each four-fold repeated residue. Despite this observation, we have previously demonstrated that the homotetrameric active site can be heavily modified, including up to four simultaneous mutations per protomer (16 in the tetrameric active site) [4]. In that selection experiment, the vast majority of modified homotetrameric active sites (99.8% of the variants tested) did not give rise to native-like function, yet this and other studies showed that the simultaneous variation of more than one residue in R67 DHFR can provide compensatory effects, thus allowing multiple mutations to complement each other [4, 23-25]. This study also revealed that no single residue of the native active site is indispensable for function.

Because our previous active-site mutation efforts using the monomeric species resulted in a low yield of functional variants, here we investigate the tolerance of the active site to more conservative modifications. To reduce the complicating effect of the active-site symmetry, R67 DHFR has previously been genetically linked to provide a covalent dimeric or tetrameric species, using the unstructured *N*- and *C*-termini to provide an inter-domain linker (Figure 3-1) [24, 26]. This breaks the active-site symmetry by introduction of substitutions on only one (for the covalent tetramer) or two (for the covalent dimer) of the four protomers [23, 24, 27]. In this work, we have modified the active site in the context of the dimeric species so that any substitution is present twice rather than four times. The *N*-terminal protomer carried multiple active-site substitutions and the *C*-terminal protomer was either native or carried an independent set of active-site mutations. Screening the library for the native catalytic activity allowed us to further evaluate the tolerance of this antibiotic-resistance enzyme to active-site variation, providing further insight into the potential for this antibiotic-resistance enzyme to adapt its function in response to selective pressure.

3.4. Results

3.4.1. Selection of TMP-resistant dimeric, doubly-mutated (DDM) R67 DHFR variants yields singly-mutated dimers

To introduce asymmetric mutations in the R67 DHFR active site, a library of R67 DHFR protomers mutated at positions 66-69 was linked to a second, similarly mutated library of R67 DHFR protomers to form a library of covalent dimers, doubly-mutated (DDM). The DDM dimer library encoded mutated protomer 1 with an *N*-terminal 6-histidine tag for protein purification and a *C*-terminal Glu-Leu linker (encoded by an inserted, unique *Sac*I restriction site) connecting to the *N*-terminus of mutated protomer 2. Residues 66-69 of both protomers were combinatorially mutated according to a previously described strategy for partial randomization [4]. Briefly, mutations were designed to participate in reactivity by (1) contributing a proton donor to catalysis, (2) altering the dielectric environment to promote hydride transfer, or (3) favoring productive substrate binding. The library was restricted to eight amino acids encoded at position 66, four at position 67, twelve (and the stop codon) at position 68, and four at position 69 (Table A 1-1). This encoded a maximum of 2.4×10^6 unique, combinatorially mutated dimer variants. The native sequence is not encoded, because no Gln is allowed at codon 67. A silent mutation at codon 74 (CTT (Leu) to TTG (Leu)) was also encoded to detect any potential contamination by native sequence. To confirm that the library construction was successful, 25 clones were picked for DNA sequencing from the obtained library of 10^6 unselected clones. Slight biases above the expected 1/8 ratio were observed for both protomers at position 66 for Asn and in position 67 a slightly bias below the expected 1/4 ratio was observed for Glu. Similarly, position 68 was biased slightly above or below the expected 1/12 ratio for Ser, Asn and Trp; no Leu or Trp was observed among those 25 unselected clones and for protomer 2 no Met or stop codon was observed (Table A 1-2). At position 69 a slight below the expected 1/4 ratio was observed for Phe. No contamination with the native sequence was observed. In two instances, the *C*-terminal protomer included unforeseen residues (Table A 1-2: I67 and S69 in DDM clone 14NS; K66 in DDM clone 19NS; compare to amino acids encoded, Table A 1-1), which may result from primer impurity

or errors introduced by the polymerase during PCR. Overall, library quality was deemed to be good.

The 10^6 DDM clones of R67 DHFR were selected for TMP resistance by plating equal aliquots on non-selective and on selective minimal medium plates containing TMP. DNA sequencing of 20 TMP-resistant clones revealed a disparity of the DNA sequence compared to unselected constructs: 25% of the selected clones had only one mutated protomer (Table A 1-3). The Leu74 silent mutation was present in both protomers, demonstrating that this was not the result of contamination with native monomer or with DSM library members. Observation of a non-mutated protomer was unexpected, because the native Gln67 was not encoded in the library; its presence may result from contamination of the mutagenic primers with undesired nucleotides at a level that is too low to be observed in the unselected library, or may result from bacterial mutational activity as a response to strong selective pressure. We note that a single base change can suffice to alter the encoded RAK, to the native CAG.

In order to evaluate the genetic stability of the selected clones, we re-cloned the 20 individual dimeric inserts into fresh vector and transformed anew. Following two more selection rounds, the majority of the selected variants (14/20) held one protomer with native sequence at residues 66-69 and one mutated protomer; again, the Leu74 silent mutation was present in both protomers (Table A 1-3). Thus, clones with two mutated protomers after the first round of selection acquired a non-mutated protomer in subsequent selection rounds. We note that, in all cases but one, it is the *C*-terminal protomer that reverted to native sequence. Previous studies of a heavily mutated, covalent tetrameric species of R67 DHFR, incorporating up to 55 mutations, showed long-lasting genetic stability even after seven days of subculturing under similar conditions [23]. In that case, all mutations were positioned in surface or disordered regions of the construct rather than within the central active-site channel. Here, mutagenesis within the active site appears to have triggered molecular mechanisms to 'correct' the sequence into a native form, preserving the bacteria from loss of activity of R67 DHFR under selective conditions.

Notwithstanding the mode of its appearance during selection, this indicates that the presence of one protomer with native sequence at residues 66-69 provided a significant survival advantage. This is in good agreement with our previous study of mutated monomers, where only three variants with native-like activity were identified among to 1536 possible

variants screened [4]. The selection of the DDM library was repeated in different *E. coli* strains under similar conditions. Results consistently showed the tendency to enrich clones with one native protomer (data not shown).

Among the non-native protomer sequences that were observed following the rounds of selection, an important bias was observed at positions 66 and 67, which held the non-native Asn66/Lys67 pair in 9/20 variants, one occurrence of which was in a dimer where both protomers remained mutated. We note that the DDM sequences that held no native protomer following 3 rounds of selection had generally varied throughout the selection rounds.

3.4.2. Selection of TMP-resistant singly-mutated dimers (DSM) R67 DHFR variants

On the basis of that preliminary observation, we sought to determine the breadth of sequence diversity that can be functionally accommodated in dimeric variants forming a tetramer with two native and two mutated protomers. To this effect, a library of mutated protomers was genetically linked to a native protomer to form a covalent dimer, singly mutated (DSM) library. As above, residues 66-69 of the *N*-terminal protomer were combinatorially mutated. To confirm that the library construction was successful, 32 clones were picked for DNA sequencing from the total library of 1.7×10^4 clones obtained. No deviations from the encoded codons were detected and no contamination with the native sequence was observed (Table A 1-4).

Some biases above the expected 1/8 ratio were observed at position 66 for Ile (9/32), Asn (7/32), and Asp (6/32), and no Ala was observed. Similarly, position 68 was slightly biased above the expected 1/12 ratio for Thr (5/32) and the stop codon (4/32); no Phe or Arg were observed among those 32 unselected clones, though they were identified following selection. Position 67 and 69 showed no strong bias towards any residue. Overall, library quality was deemed to be good.

The DSM dimer library was then selected for TMP resistance as described above. A 79% selection rate (number of selected clones/number of unselected) was obtained. DNA sequencing of 32 TMP-resistant clones confirmed that no native sequence had arisen in the mutated *N*-terminal protomer (Table 3-1).

Table 3-1 Amino acid sequence of R67 DHFR dimer mutants.
Deduced amino acid sequence of the active site residues 66-69 of the mutated N-terminal protomer within the dimeric DSM library following selection for TMP resistance.

Selected clone	Residues 66-69	Selected clone	Residues 66-69	Selected clone	Residues 66-69
1S	NKSH	12S	TEFY	23S	IETY
2S	TENH	13S	IKYH	24S	TEFL
3S	TENH	14S	DESH	25S	AEFL
4S	ANTF	15S	NDTY	26S	DDNY
5S	TEFL	16S	SDWL	27S	HEYF
6S	NDTF	17S	TEFL	28S	NKNY
7S	TNTY	18S	SDCF	29S	VDNH
8S	SKKY	19S	GNNL	30S	NKNY
9S	INSF	20S	DDNH	31S	DDNY
10S	DETL	21S	GNNF	32S	INNF
11S	NNRH	22S	SEYF		

Furthermore, the native *C*-terminal protomer showed no unexpected mutations. We note that the observed 79% selection rate of the DSM library does not necessarily indicate that nearly 4 out of 5 of the possible sequence variations in the *N*-terminal protomer give rise to an active enzyme species. Bacteria may be multiply transformed, thus harbour more than one variant at a time, which increases the apparent selection rate. Nonetheless, this selection rate is much greater than the 0.2% observed upon selection of the same mutated monomeric species [4]. As all transformations and selections were undertaken similarly, this indicates that there is a greater fraction of functional variants when the mutated protomer is combined with a native one. This is further supported by the fact that only three variants among that mutated monomeric library had provided TMP resistance upon selection in *E. coli*, while a broad variety of variants were functionally selected here [4]. Among the 32 clones sequenced, 27 unique sequences were identified in the *N*-terminal protomer (Table 3-1). This represents roughly 2% of all possible combinations. The distribution of codons/amino acids observed for each mutated residue (66-69) was compared for unselected (Table A 1-4) and selected clones (Table 3-1). For the selected clones, a slight (2-fold) enrichment of Thr was observed at position 66 and a similar enrichment of Glu was observed at position 67. Only three occurrences of the non-native Asn66/Lys67 pair, which was observed in nearly 50% of the

selected DDM variants, were recorded. Interestingly, a new pair emerged at the same position in this selection: the combination of Thr66/Glu67 was observed in nearly 20% of selected DSM variants (Table 3-1). At position 68, a 9-fold enrichment of Asn was observed as well as a 6-fold enrichment of Phe. The distribution of codons for position 69 was similar for unselected and selected clones. The only unexpected mutation was His66 in mutant 27S, which was not designed into the mutagenic primers. We note that neither the exact sequences of the three functional monomeric variants (SKIH, INRY and GELH), nor any closely related homolog, were observed among the 32 selected dimers [4]. For those variants, all four protomers were identical such that the mutated sequences needed to be four-fold tolerated in the context of function. Here, the sequence at residues 66-69 for two of the protomers may be complementary, or compensatory, for mutations on the two other protomers. Thus a greater sequence variation can be expected and, indeed, was observed.

The selected mutated dimeric DSM variants 4S, 9S and 12S were expressed at levels similar to the non-mutated dimer (DNM), whereas variants 3S, 13S and 30S were expressed about 50% less; variants 5S, 25S and 27S expressed only in the range of 25% of the DNM level. The molecular weight of DNM and a subset of four selected DSM variants (3S, 5S, 12S and13S) were verified by mass spectrometry (Table A 1-5), confirming that the dimeric species were expressed with no unexpected modifications.

3.4.3. Kinetic characterization of the selected DSM

The kinetic parameters of nine selected DSM were determined and compared to the native-like DNM (Table 3-2). The k_{cat} obtained for DNM was similar to that previously reported for a similar dimer construct (0.28 s^{-1} vs 1.2 s^{-1}) [26]. The average K_M values obtained for DHF and NADPH (12.9 μM and 13.3 μM, respectively) were also similar to values reported for the similar dimer construct (6.3 μM and 2.7 μM, respectively) [26]. The differences may be attributed to experimental conditions (here, measured at ambient temperature in absence of β-mercaptoethanol rather than at 30°C with β-mercaptoethanol) and the different inter-dimer linker.

Table 3-2 Kinetic parameters for the non-mutated DNM dimer and selected mutated DSM dimer variants.

Residues 66-69 in *N*-terminal protomer	k_{cat} (s^{-1})	K_M^{DHF} (μM)	K_M^{NADPH} (μM)	k_{cat}/K_M^{NADPH} (s^{-1} μM^{-1})	k_{cat}/K_M^{DHF} (s^{-1} μM^{-1})
DNM:					
VQIY (native)	0.28 ± 0.03	13 ± 4	13 ± 5	0.02	0.02
DSM:					
TENH (3S)	0.18 ± 0.01	28 ± 4	19 ± 4	0.01	0.007
ANTF (4S)	0.28 ± 0.02	51 ± 8	44 ± 6	0.006	0.005
TEFL (5S)	0.03 ± 0.01	35 ± 23	ND	-	0.001
INSF (9S)	0.06 ± 0.01	16 ± 4	> 200	< 0.0003	0.004
TEFY (12S)	0.12 ± 0.01	18 ± 5	84 ± 20	0.001	0.007
IKYH (13S)	0.050± 0.004	15 ± 4	13 ± 2	0.004	0.003
AEFL (25S)	ND	ND	ND	-	-
HEYF (27S)	0.10 ± 0.01	8.5 ± 3.0	23 ± 7	0.005	0.012
NKNY (30S)	0.11 ± 0.01	13 ± 4	16 ± 5	0.007	0.009

For the selected DSM variants, the k_{cat} values were similar or 2- to 28-fold lower than for the DNM dimer, except for variant 25S where no activity was detected; this variant was not further analyzed. In contrast, the K_M values for DHF and NADPH were generally similar to those of the DNM dimer, except for K_M^{NADPH} of variants 5S and 9S, which were significantly increased.

3.4.4. Functional DSM variants assemble predominantly as 'dimers of dimers'

For all subsequent experiments we focused on the selected DSM variants 3S, 9S, 12S, 13S and 30S, based on their diversity of sequence and kinetic activity. Size exclusion chromatography was carried out to determine the oligomeric state of the native-like DNM dimer as well as the *N*-terminally mutated DSM variants. For DNM, a single species was observed at approximately 36 kDa, confirming the formation of a dimer of the dimeric species, equivalent to the native tetrameric state of R67 DHFR (Table 3-3).

Table 3-3 Molecular weight of R67 DHFR WT and mutants.
Native molecular weight determined by size exclusion chromatography for the non-mutated DNM dimer and selected mutated DSM dimer variants.

Residues 66-69 in *N*-terminal protomer	Main peak (kDa)	Second peak (kDa)	Peak ratio (% main/second)
DNM:			
VQIY (native)	34.0	N/O[a]	100
DSM:			
TENH (3S)	33.4	52.9	61.5
INSF (9S)	33.3	45.5	92.6
TEFY (12S)	31.1	45.8	87.5
IKYH (13S)	33.3	N/O[a]	100
NKNY (30S)	35.3	52.9	74.1

[a] N/O: none observed.

Similar results were observed for the variants 9S, 12S and 13S, where the 'dimer of dimers' species was predominant (Figure 3-2 B).

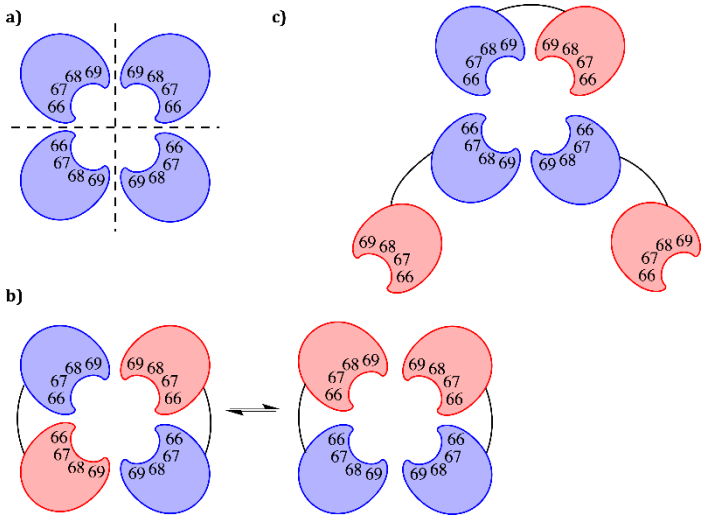


Figure 3-2 Schematic representation of protomer assembly for R67 DHFR variants.
A) Assembly of the native, monomeric R67 DHFR into a tetramer with a central active-site pore. B) Assembly of a dimer of dimers, where the red protomers are mutated. C) Putative assembly of a trimer of dimers to yield an active-site where only one protomer is mutated.

For variants 3S and 30S, the main peak on the chromatogram is consistent with the formation of a 'dimer of dimers' yet a shoulder observed at approximately 53 kDa is

consistent with the formation of a 'trimer of dimers'. The formation of 'trimers of dimers' opens new possibilities for assembly, which are illustrated in Figure 3-2. This model is imperfect because the 'pending' monomers may not be adequately stabilized; we currently have no direct insight into the mode of assembly. We note that higher-order assembly of R67 DHFR multimers has previously been suggested in relation to the intermolecular association of a covalent tetrameric construct [23].

The potential for the dimers to assemble in various configurations to form a functional active site may underlie the high selection rate observed. In a trimeric arrangement, the active site may be constituted of three non-mutated and one mutated protomer, reducing the deleterious effect of some mutations (Figure 3-2 C). The presence of a non-polar residue at position 66 may favor the formation of a dimer-dimer assembly (9S and 13S). A polar residue at that position may increase the prevalence of the trimeric arrangement up to 37.5% in the most distant variant from the WT sequence, 3S.

3.4.5. Secondary structure is native-like

Circular dichroism (CD) allowed an assessment of the extent of secondary structure formation in functional dimers. The CD spectrum of the native DNM dimer is consistent with that reported for a previously constructed dimer, and is also very similar to that of the native homotetrameric R67 DHFR (see Zhuang *et al*., Figure 4B) [26]. The CD scans for DSM variants 3S, 9S, 12S and 30S all yield a CD spectrum that is similar to the non-mutated DNM, and are characteristic of high β-sheet content (Figure 3-3 A). This is consistent with the native structure of R67 DHFR and indicates that the overall secondary structure was not significantly disrupted by the introduction of the mutations. Only slight differences in steepness were observed among variants. The thermal stability of the dimeric variants was evaluated with CD between 25°C and 95°C, to observe any perturbation of the secondary structure (Figure 3-3 B). Surprisingly, monitoring the secondary structure showed no clear temperature denaturation for any R67 DHFR variant. Only variants 12S and 30S showed a small increase in ellipticity in the range of 90 to 95°C. This result was in apparent contradiction with the previously published T_m value for native R67 DHFR and a dimer construction, determined using differential scanning calorimetry (DSC) [26].

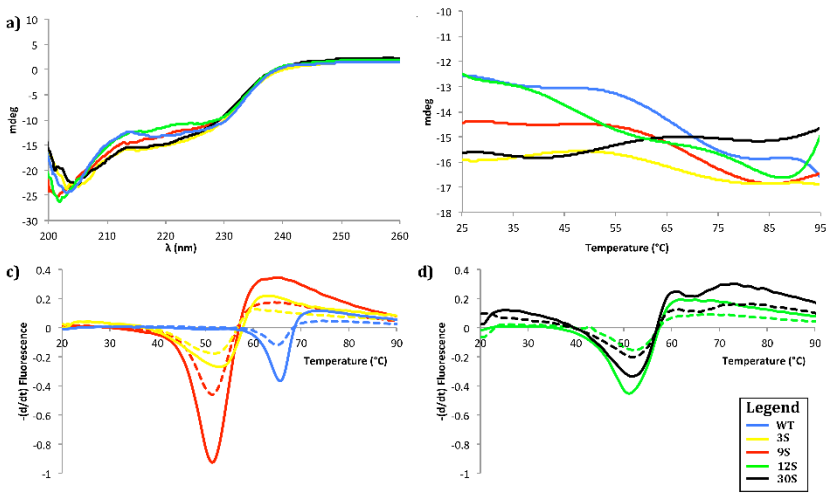


Figure 3-3 Thermal denaturation of R67 DHFR dimer mutants monitored by CD spectrometry and thermal scanning fluorimetry.

A) Far-UV CD spectra at 25°C of WT, 3S, 9S, 12S and 30S. B) Thermal denaturation measured at 215 nm at a rate of 25°C/h for WT, 3S, 9S, 12S and 30S. C, D) First derivative analysis of representative thermal melting curves observed by fluorescence of SYPRO Orange. Melting curves are shown for a ratio of 3.33 x SYPRO to 8 µM (full line) or 4 µM (dashed line) protein, for c) WT, 3S, 9S and d) 12S and 30S.

DSC reveals not only changes to secondary structure, but also changes to higher order structure. Therefore, we speculated that the previously reported T_m may have reported on changes to the tertiary and quaternary structure, while the secondary structure elements appear to be thermally stable up to at least 90°C.

3.4.6. Protomer assembly into the active structure is weaker in the selected DSM variants

To investigate this possibility we performed a thermal scanning fluorimetry shift assay. We used SYPRO orange as the hydrophobic fluorophore to bind hydrophobic patches of the protein as they become increasingly solvent-exposed upon thermal denaturation [28]. Similarly to DSC, this approach can reveal changes in secondary, tertiary and quaternary structure, as observed by an increase in fluorescence. The resulting melting curve for the native DNM dimer yielded a T_m of 64.5°C, consistent with the value of 61.3°C reported by DSC for a similar native dimer construct (Figure 3-3 C and D) [26]. A difference of 1-3°C in T_m between the two methods is consistent with previously published data using other proteins

[29, 30]. These results indicate that the secondary structure elements of the native DNM dimer – which is likely to be stabilized as a β-barrel spanning both protomers 6 – remain essentially native-like up to ≥ 90°C, while the quaternary subunit assembly (association of 2 covalent dimers) breaks down above 60°C. It remains to be investigated whether the high thermostability at the level of secondary structure is an intrinsic property of the native R67 DHFR protomers, or whether it has been increased by linkage of the protomers into a covalent dimer.

The melting curves for the four DSM variants yielded T_m values approximately 10°C lower than for the DNM dimer, indicative of a perturbation to the quaternary subunit assembly (Table 3-4).

Table 3-4 Melting temperatures of R67 DHFR WT and mutants.
T_m values obtained by thermal scanning fluorimetry for the non-mutated DNM dimer and selected mutated DSM dimer variants.

Residues 66-69 in *N*-terminal protomer	T_m (°C)
DNM:	
VQIY (native)	64.5
DSM:	
TENH (3S)	53.0
INSF (9S)	51.7
TEFY (12S)	51.8
NKNY (30S)	52.8

Residues 66-69 are contained in the inner-most β-strand of the active site and do not participate directly in the interface of protein-protein interaction between the two dimers. Nonetheless, their modification, even in only two of the four protomers of the assembled R67 DHFR variants, appears to destabilize the tetrameric core assembly. Interestingly, the T_m value does not differ significantly among the DSM variants, whereas the introduced mutations differ greatly in physical and electrostatic properties.

To further investigate the high thermal stability of the secondary structure elements, we analysed the stability of the protomers with respect to controlled proteolysis. Digestion with α-chymostrypsin over 24h revealed a similar digestion pattern for the native DNM dimer and all DSM variants; variant 3S is shown as a representative example (Figure 3-4).

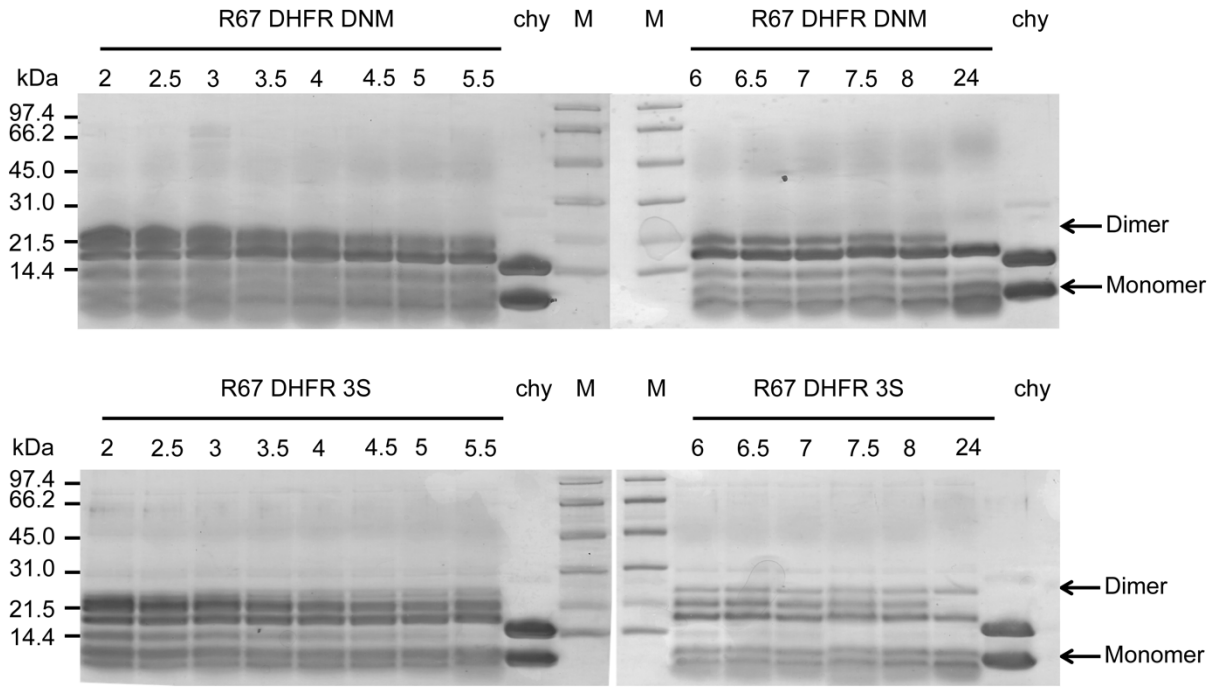


Figure 3-4 Limited chymotryptic digestion of R67 DHFR DNM.
R67 DHFR DNM (upper panel) and DSM variant 3S (lower panel), observed by tricine SDS PAGE R67 DHFR DNM or DSM 3S were digested with chymotrypsin (1:100 chymotrypsin : R67 DHFR) for 2 to 24h at 4°C, as indicated above each lane. Chy: chymotrypsin reference; M: molecular weight markers. Arrows on the right indicate the bands corresponding to the monomeric and dimeric species.

We also verified that no significant difference in digestion pattern was observed during the first 2h (data not shown). α-Chymotrypsin can theoretically recognize up to 22 cleavage sites in the R67 DHFR dimers studied. In all cases, the predominant proteolysis bands observed upon analysis by SDS PAGE following 24 h digestion were equivalent to the molecular weight of a single protomer (monomer) and a slightly shortened dimer, though additional bands indicate further preferred proteolysis sites. Cleavage is most likely to occur in the unstructured *N*- and *C*-terminal regions and in the inter-protomer linker as a result of accessibility. We note that the band of highest molecular weight (see 'Dimer' arrow, Figure 3-4) was digested significantly more rapidly, and to completion, in the DNM sample than for the 3S variant, where a significant undigested band remained even after 24 h digestion.

For all variants, the core of the R67 DHFR protomer remained intact and inaccessible to α-chymotrypsin even following 24 h digestion. This indicates that the reduction in T_m by approximately 10ºC for the selected variants did not destabilize packing to the extent where the core becomes accessible for proteolysis. To verify this, MS/MS analysis was performed on the main band of the DSM 3S dimer remaining following overnight proteolysis, corresponding to a single protomer. Both the non-mutated and the mutated protomers were observed by MS in that sample. Since this method is not quantitative it was not possible to determine the ratio of non-mutated:mutated protomer resulting from digestion. Nonetheless, this result indicates that the tight fold of each protomer is conserved upon introduction of mutations at the active site, consistent with the maintenance of secondary structure observed by CD upon thermal denaturation.

3.4.7. Recovery of activity following thermal denaturation is enhanced in the selected DSM variants

We further investigated the high thermal stability of the secondary structure elements and the multimeric assembly by verifying the activity of the variants following incubation at 95°C, under conditions mimicking the CD thermal denaturation experiment. After incubating the samples at 95°C for 90 minutes, slow cooling for 45 min allowed partial recovery of the activity (Table 3-5, compare conditions A and B).

Table 3-5 Recovery of enzymatic activity following high-temperature incubation.
Samples were incubated at 95°C then at room temperature for the periods indicated, then activity was measured. Activity is expressed as percent of activity for that variant prior to incubation.

Condition	DNM [%]	3S [%]	9S [%]	12S [%]	30S [%]
A: 95°C, 90 min; RT, 1 min	11	18	18	26	17
B: 95°C, 90 min; RT, 45 min	23	72	72	46	47
C: 95°C, 4 min; RT, 1 min	78	95	90	50	11
D: 95°C, 4 min; RT, 45 min	71	> 99	90	53	70

All DSM variants showed a high recovery (≥ 45%) of the initial activity, where the variants 3S and 9S showed the highest recovery (> 70%). Interestingly, the recovery of DNM

was significantly lower (< 25%), consistent with a previous report of 15% activity recovered following heating to 75°C for a similar dimeric construct [26]. Shorter incubation at 95°C (conditions C and D) was generally less disruptive to activity. Thus the selected DSM variants have acquired an increased resilience to heat denaturation, despite having reduced T_m values relative to the non-mutated DNM.

3.5. Discussion

Overall, we have found that the active, tetrameric form of R67 DHFR is very tolerant to active site variation, particularly in a context where two among the four protomers are native. In most multimeric enzymes, this would not be surprising, since a single protomer can often independently provide catalytic function. This is not the case in R67 DHFR, where the relatively small active site tunnel formed at the center of the tetramer can accommodate a single pair of ligands (DHF substrate + NADPH cofactor) at a time; all four protomers contribute simultaneously to ligand binding.

Through the process of selection, we have obtained DSM variants of R67 DHFR that exhibit a lower T_m. This illustrates that assembly of the dimers into an active structure (generally consisting of 4 protomers, though 6 protomers were also observed according to size-exclusion chromatography) was not as thermally stable in the selected DSM variants as in the non-mutated DNM dimer. Nonetheless, those selected DSM variants showed enhanced recovery of activity following thermal denaturation, suggesting an increased rate of dimer assembly into the active enzyme form. We note that the selection strategy (growth on agar plates at 37ºC) was not designed to elicit this trait. We hypothesize that, as 37ºC is largely inferior to the T_m of the selected variants (near 52ºC), reduced thermal stability was not a selective disadvantage, while the improved dimer assembly into the active tetrameric form appears to have conferred a selective advantage.

Our observation of some selected DDM sequences that contain no native active-site sequence supports our earlier report of active-site modification in R67 DHFR, and confirms that no specific residue in the active-site tunnel formed by residues 66-69 is indispensable for catalytic activity [4]. Importantly, we observed no partial enrichment toward native residues with retention of accompanying mutations: the residue 66-69 stretch either became entirely

native, or else remained highly variable with only slight sequence biases. There may be a synergy between residues 66-69 that provides one robust solution for catalysis and structural integrity. We do not know if there is a functional or structural significance to the predominance of *C*-terminal reversion to the native sequence during the selection process; this may reflect different constraints on the two protomers as a result of their non-native fusion and/or the presence of the *N*-terminal 6-His tag.

The non-native Asn66/Lys67 residue pair was frequently observed in replacing the native Val66/Gln67 residue pair when the starting point was the doubly-mutated dimer library DDM (9/20), but was less frequent when the starting point was the DSM singly mutated dimer library (3/32). In the DSM library the Thr66/Glu67 residue pair was predominant (6/32). This suggests that the Asn66/Lys67 or the Thr66/Glu67 substitution pair confers a selective advantage under the more stringent conditions of selecting for activity in a fully-mutated active-site environment (DDM library) relative to the active sites that are part native, part mutated (DSM library). It is tempting to speculate on functional relevance of those pairwise substitutions. However, uncertainty with respect to the 'dimer-dimer' arrangement (do the likewise mutated monomers face each other in the tetramer or not?) and hence the resulting environment of substrate/co-factor binding opens a number of possible interaction profiles [24]. Further investigation of this question would require the structural resolution of a ligand-bound dimer complex; earlier attempts to resolve the structure of a covalently-linked R67 DHFR dimer in our laboratory have not been successful [14].

The observation of high sequence diversity at the active site of one protomer in the presence of the second protomer with native sequence suggests that the enzymatic activity of R67 DHFR can be readily supported by an active site that is 50% native. This is supported by the observation of native-like enzyme kinetics for a number of DSM variants. Assembly of the four protomers to form a central active-site channel allows DHF and NADPH to slip in from opposite mouths of the tunnel; their reactive atoms are juxtaposed near the center of symmetry of the tetramer [22]. Catalysis is proposed to result from proper positioning of the reactive atoms to allow hydride transfer to occur on the pre-protonated DHF molecule. Again, no residue of the enzyme is indispensable for catalysis. Our results suggest that the native 66-69 residues of two protomers are sufficient to provide a native-like catalytic rate enhancement in R67 DHFR.

We also observed an additional structural dimension with the formation of 'trimer-dimer'. This evidence for a functional assembly other than the 'dimer-dimer' assembly increases the complexity of the system and is likely to reduce the impact of the introduced mutations. Along with the observed reversion of one protomer to the native sequence in the DDM dimers, assembly into trimers can provide an additional route for rapid recovery of trimethoprim resistance in the bacterial cell during selection.

In conclusion, we have found that R67 DHFR can explore a large active-site sequence space and various quaternary structures while maintaining native-like activity. It thus appears that this enzyme, postulated to have recently acquired DHFR activity as a response to selective pressure, is not in a highly specialized, evolutionary dead-end; instead, it appears to have a great potential for continued functional evolution. This finding is highly interesting because newly emerged and still evolving proteins might behave similarly to ancient proteins, and may thus serve as models with which to conduct studies on ancestral proteins.

Our results have direct implications on the threat that R67 DHFR holds with respect to antibiotic resistance. Already natively resistant to trimethoprim as a result of its fold being unrelated to that of DfrA DHFRs, R67 DHFR may continue to evolve even as we direct inhibitors to it, and will need to be closely monitored [31].

3.6. Materials and methods

3.6.1. Materials

All reagents were of the highest available purity. Restriction enzymes and DNA-modifying enzymes were from New England Biolabs and MBI Fermentas. Synthetic oligonucleotides were obtained from Integrated DNA Technologies (Coralville, USA). The infrared dye-labeled sequencing primers were from Li-Cor Biotechnology Division. The Sequenase 2.0 DNA sequencing kit was purchased from GE Healthcare. All aqueous solutions were prepared using water purified with a Millipore BioCell system. 7,8-Dihydrofolate (DHF) was prepared as described previously [32].

3.6.2. Bacterial strains and plasmids

E. coli SS320 was used to verify library quality by DNA sequencing. *E. coli* BL21 containing plasmid pRep4 (Qiagen, Mississauga) was used for expression. Chloramphenicol (Chl) was used at 10 μg/mL and kanamycin (Kan) at 50 μg/mL for maintenance of plasmids derived from pQE32Chl and for pRep4, respectively [4]. Ampicillin (Amp) was used at 100 μg/mL for maintenance of plasmids derived from pQE32 (Qiagen). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used at 1 mM for expression and TMP was used at 70 μg/mL during selection.

3.6.3. Construction of the 'dimer, doubly-mutated' (DDM) and 'dimer, singly-mutated' (DSM) libraries

Starting from the construction described in [4], a covalent R67 DHFR dimer mutated at residues 66-69 was created by inserting a second mutated R67 gene between *Sac*I and *Hind*III restriction sites on pQE32Chl (library DM, where DM stands for 'dimer, mutated'). A covalent dimer containing two native (WT) R67 DHFR genes was also created by ligating the *Ban*II-*Xba*I digested products of the native construct described in [4] and a construct containing the native R67 DHFR gene in pQE32 (Amp resistant) between *Sac*I and *Hind*III sites, yielding construct DNM (where DNM stands for 'dimer, non-mutated'). Library DM and construct DNM were digested with *Sac*I and *Xba*I. This generated two fragments of 2500 and 1500 bp for each construct. The 2500 bp fragment of library DM (encoding the *N*-terminal mutated protomer) and the 1500 bp fragment from the DNM or the DM construct (encoding the *C*-terminal non-mutated or mutated protomer, respectively) were extracted using a QiaexII extraction kit (Qiagen). The fragments were ligated yielding the dimer libraries, where protomer 1 is mutated at residues 66-69 and protomer 2 is either non-mutated ('dimer, singly-mutated', DSM) or mutated ('dimer, doubly-mutated', DDM). Apart from the mutations, this dimer construct differs slightly from a previously reported dimer, where protomer 1 was immediately connected to protomer 2; the Glu-Leu linker was adopted here because it introduces a *Sac*I restriction site for recombination of protomers into dimers [26].

The libraries were transformed into *E. coli* SS320 and were plated under non-selective conditions, on 1.2 % agar Terrific Broth (TB) containing Chl. Over 2.4×10^6 colonies were

obtained for the DDM library and 1.7×10^4 colonies for the DSM library. DNA sequencing of individual clones served to evaluate library quality prior to selection. Sequencing was performed using the Sequenase 2.0 DNA sequencing kit (GE Healthcare) with a Li-Cor IR2 automated system, using dye-labeled primers. Sequences were analyzed using the AlignIR V2.0 software (Li-Cor).

3.6.4. Library selection

The DDM and DSM dimer libraries in *E. coli* SS320 were harvested, propagated in Terrific Broth (TB), and the DNA retransformed into *E. coli* BL21/pRep4 for selection. Equal aliquots of transformation mixture were plated on TB plates (+ Chl and Kan) and on M9c minimal medium plates (+ Chl and Kan) containing IPTG and TMP. Colonies were grown at 37°C for 18h. The plasmid inserts from selected clones were sequenced. Those clones were retransformed into *E. coli* BL21/pRep4 and subjected to two further rounds of selection and sequencing to confirm that activity was not due to cellular acquisition of TMP resistance or unforeseen changes in their DNA sequences.

3.6.5. Small-scale protein overexpression and purification

An overnight culture (0.2 mL) was used to inoculate 20 mL TB (+ Kan, and Chl or Amp for DNM). Cultures were propagated at 37°C with shaking until OD_{600} reached 0.7. Protein expression was induced for 3 h following the addition of IPTG (1mM final concentration). The cells were harvested by centrifugation (30 min, 3000 × *g*, 4°C). The cell pellet was resuspended in 4 mL of lysis buffer (100 mM potassium phosphate, pH adjusted to 8.0) and was disrupted by sonication on ice for 3 × 30 s with 1 min intervals. Following centrifugation at 3000 × *g* for 30 min at 4°C, 1 mL of Ni-NTA agarose resin was added to the supernatant. The mixture was gently shaken for 30 min at 4 °C and transferred into a gravitation column. The resin was washed twice with 5 mL lysis buffer, and twice with the same containing 10 and 20 mM imidazole, respectively. The enzyme was eluted with elution buffer 100 mM potassium phosphate buffer, pH 8 containing 250 mM imidazole.

3.6.6. Large-scale protein over-expression and purification

Induction of protein expression was undertaken as above, in 500 mL of TB. The cell pellet was resuspended in 100 mM Tris-Cl buffer pH 8, and lysed by one passage through a cell disrupter (Constant Systems) adjusted to 27 kpsi. Following centrifugation (30 min, 3800 × *g*, 4°C), the supernatant was filtered over a 0.2 μm filter. Purification was performed following a 2-step purification protocol on an Äkta FPLC (GE Healthcare, Piscataway, NJ) at 4°C. The supernatant was applied to a 5 mL His-trap column pre-equilibrated with 100 mM Tris-Cl, pH 8. The protein was eluted with a gradient of imidazole (0-300 mM). Fractions were analyzed for DHFR activity and by tricine SDS-PAGE [4, 33]. Those containing the most target protein were pooled and concentrated using an Ultra Amicon Ultracel 3.5K (Millipore). The concentrate was applied to a 1.5 × 28 cm Superose 12 column equilibrated in 100 mM TrisCl, pH 8 and eluted in the same buffer. The fractions were collected and analyzed for DHFR activity, by tricine SDS-PAGE and using the Bio-Rad protein assay.

3.6.7. Mass spectral characterization of R67 DHFR variants

Mass spectral analysis of R67 variants was achieved by HPLC-MS. Separations were performed on an 1100 LC system coupled to an ESI-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA). The chromatographic column was a Poroshell 300SB-C8, 75 x 2.1 mm, 5μ particle size (Agilent Technologies) operated at 0.4 mL/min. The eluents used were 0.1% formic acid in water and 0.1% formic acid in acetonitrile. A linear acetonitrile gradient from 20 to 95 % was used for elution. The mass spectrometer was operated in positive electrospray mode and mass spectra were acquired from m/z 110 to 3200. Deconvolution on the multiply charged species mass spectrum was performed using BioConfirm, an integrated program from MassHunter Qualitative Analysis.

3.6.8. Determination of multimerization state

Size exclusion chromatography was carried out at 4°C using a 1.5 × 28 cm Superose 12 column equilibrated in 100 mM Tris-Cl buffer pH 8. Ribonuclease A (13.7 kDa ± 15%, 10 mg/mL), chymotrypsinogen A (25.0 kDa ± 25%; 3 mg/mL), ovalbumin (43.0 kDa ± 15%, 7 mg/mL) and albumin (67.0 kDa ± 10%; 7 mg/mL) were used as standards to determine the

oligomeric state of the R67 DHFR variants. The flow rate was 1.5 mL/min. Standard curves were produced by plotting the logarithm of the molecular mass of standards versus K_{av}. $K_{av} = (V_E - Vv)/(V_B - Vv)$ where V_E is the elution volume, Vv is the void volume and V_B is the bed volume of the column matrix.

3.6.9. Kinetic characterization

Spectrophotometric enzyme activity assays were carried using a Varian Cary 100 Bio spectrophotometer. The assay buffer was MTA (50 mM MES, 100 mM Tris-Cl, 50 mM acetic acid), pH 7. Assays were performed in the presence of 50 nM TMP to inhibit any residual *E. coli* chromosomal DHFR. The extinction coefficients used were 28,400 $M^{-1}\cdot cm^{-1}$ at 282 nm for DHF, 6,230 $M^{-1}\cdot cm^{-1}$ at 340 nm for NADPH and 12,300 $M^{-1}\cdot cm^{-1}$ at 340 nm to determine product formation [4]. To determine K_M^{DHF}, DHF was held at 100 μM while NADPH was varied from 0 – 100 μM. Similarly, NADPH was fixed at 100 μM to determine K_M^{NADPH} while DHF was varied from 0 - 100 μM. The data were fitted to the Michaelis-Menten equation using non-linear regression analysis with the GraphPad software. Protein concentration was determined with the BioRad assay.

3.6.10. Circular dichroism

Circular dichroism (CD) spectra were recorded on a Jasco J-815 spectropolarimeter with a Peltier element for temperature control. Protein solutions were prepared at a final concentration of 0.25 μg/μL in 50 mM potassium phosphate, pH 7. The path length of the quartz cell was 0.2 cm. Spectra were acquired at 20°C over the range 190–280 nm at a scan rate of 20 nm/min. Three scans were averaged. Buffer background recorded under the same conditions was subtracted. A single protein preparation was tested for all proteins except for DNM, the non-mutated R67 DHFR dimer (two preparations). Thermal denaturation experiments were performed by monitoring changes in ellipticity at 215 nm, at a scan rate of 25°C/hour. The Savitzky-Golay method was used for curve smoothing [34]. The T_m values were determined using the derivatives function of the Jasco Spectra Manager software using the Savitzky-Golay algorithm for determination of first order derivative, where T_m is given as the maximum of this derivative.

3.6.11. Thermal Scanning Fluorimetry Shift Assays

T_m values were also determined by thermally-induced incorporation of SYPRO Orange into the unfolding protein, with analysis using a Q-PCR thermal cycler as previously described [28]. Briefly, each combination of 10 ×, 5 × and 3.33 × SYPRO Orange solution (Invitrogen) with 4 μM or 8 μM test protein was probed in a 96-well LightCycler plate (Sarstedt). SYPRO Orange and the protein were diluted with 50 mM potassium phosphate, pH 7, to a final volume of 20 μL per well. Controls contained SYPRO Orange in buffer. The plates were sealed using Optically Clear Sealing Tape (Sarstedt) and heated from 20°C to 95°C in a LightCycler 480 apparatus (Roche) with a ramp speed of 0.04°C/sec and 10 acquisitions/°C. Fluorescence was monitored with a CCD camera, using λ_{exc} = 483 nm and λ_{em} = 568 nm and a 1 s exposure time. Any curve showing a maximum fluorescence plateau during denaturation was excluded from the T_m calculation.

For the T_m calculations, both temperature and fluorescence data were smoothed [34]. The first derivatives dFluo or dTemp were calculated using the cubic spline interpolation. The preliminary maximum was determined to obtain the half-values to the left and right of it. The linear fit for the curve outside the half-values was calculated, followed by the calculation of the average deviation from the fit. If the maximum was below the detection limit (fit value + 3 × deviation), the T_m determination was considered uncertain. The quadratic fit around the maximum was then calculated as follows to obtain T_m. The first derivative of the quadratic fit function (y-value) was set to 0 and the x-axis value (temperature) was resolved. Then, the average deviation of the curve points around the maximum from the quadratic fit was calculated. If the relative deviation was greater than 5%, the T_m values were rejected if the corresponding maximum was below the detection limit. However, T_m values with a maximum above the detection but a relative deviation greater than 5% were defined as uncertain.

3.6.12. Thermal denaturation and refolding assay

To simulate the thermal denaturation conditions of the CD experiment, the same temperature ramp was programmed into a BioRad PTC-200 PCR cycler. The total volume in each PCR tube was 70 μL with a protein concentration at 1 μg/μL in 50 mM potassium phosphate, pH 7. The activity of the R67 DHFR dimer variants was measured as described

above for the different conditions. Condition A: Ramp to 95°C, hold for 90 min, keep at RT for 1 min and determine activity. B: Ramp to 95°C, hold for 90 min, keep at RT for 45 min and determine activity. C: Ramp to 95°C, hold for 4 min, keep at RT for 1 min and determine activity. D: Ramp to 95°C, hold for 4 min, keep at RT for 45 min and determine activity. Control: immediately determine activity. Specific activity was determined as described above.

3.6.13. Proteolysis by chymotrypsin

The packing of the mutated protomers was assessed by digesting with α-chymotrypsin (MP Biomedicals). Each R67 DHFR species was purified and quantitated as described above. The protein concentration was 3 μg/μL and chymotrypsin was added in a mass ratio of 1%. Incubation was at 4°C and samples were taken at various time points up to 24 h. The progress of the proteolysis was analysed by tricine SDS-PAGE.

3.7. Acknowledgments

We thank the Regional Center for Mass Spectrometry at Université de Montréal for their help with the MS/MS experiments. We also thank Professors Joanne Turnbull and Judith Kornblatt from Concordia University for access and training for CD experiments. M.C.C.J.C.E. was a recipient of a scholarship from the German Academic Exchange Service (DAAD), then a NSERC-Vanier scholar. This research was supported by funds from the Natural Sciences and Engineering Research Council of Canada (NSERC).

3.8. References

1. Nanda, V. and R.L. Koder, *Designing artificial enzymes by intuition and computation.* Nat Chem, 2010. **2**(1): p. 15-24.
2. Smith, B.A. and M.H. Hecht, *Novel proteins: from fold to function.* Curr Opin Chem Biol, 2011. **15**(3): p. 421-426.
3. Jensen, R.A., *Enzyme recruitment in evolution of new function.* Annu Rev Microbiol, 1976. **30**: p. 409-425.

4. Schmitzer, A.R., F. Lepine, and J.N. Pelletier, *Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations.* Protein Eng Des Sel, 2004. **17**(11): p. 809-819.
5. Bastien, D., et al., *Fragment-based design of symmetrical bis-benzimidazoles as selective inhibitors of the trimethoprim-resistant, type II R67 dihydrofolate reductase.* J Med Chem, 2012. **55**(7): p. 3182-3192.
6. Ebert, M.C., et al., *Asymmetric mutations in the tetrameric R67 dihydrofolate reductase reveal high tolerance to active-site substitutions.* Protein Sci, 2015. **24**(4): p. 495-507.
7. Threlfall, E.J., et al., *The emergence and spread of antibiotic resistance in food-borne bacteria.* Int J Food Microbiol, 2000. **62**(1-2): p. 1-5.
8. Matthews, D.A., et al., *Crystal-Structure of a Novel Trimethoprim-Resistant Dihydrofolate-Reductase Specified in Escherichia-Coli by R-Plasmid R67.* Biochemistry-Us, 1986. **25**(15): p. 4194-4204.
9. White, P.A. and W.D. Rawlinson, *Current status of the aadA and dfr gene cassette families.* J Antimicrob Chemoth, 2001. **47**(4): p. 495-496.
10. Howell, E.E., *Searching sequence space: Two different approaches to dihydrofolate reductase catalysis.* Chembiochem, 2005. **6**(4): p. 591-600.
11. Inc., C.C.G., *Molecular Operating Environment (MOE)*. 2016: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.
12. Stone, D. and S.L. Smith, *Amino-Acid Sequence of the Trimethoprim-Resistant Dihydrofolate-Reductase Specified in Escherichia-Coli by R-Plasmid R67.* J Biol Chem, 1979. **254**(21): p. 857-861.
13. Narayana, N., et al., *A Plasmid-Encoded Dihydrofolate-Reductase from Trimethoprim-Resistant Bacteria Has a Novel D-2-Symmetrical Active-Site.* Nat Struct Biol, 1995. **2**(11): p. 1018-1025.
14. Yachnin, B.J., et al., *Novel crystallization conditions for tandem variant R67 DHFR yield a wild-type crystal structure.* Acta Crystallogr F, 2011. **67**: p. 1316-1322.
15. Strader, M.B., et al., *Role of S65, Q67, I68, and I69 residues in homotetrameric R67 dihydrofolate reductase.* Biochemistry-Us, 2001. **40**(38): p. 11344-11352.

16. Nichols, R., et al., *Titration of Histidine-62 in R67 Dihydrofolate-Reductase Is Linked to a Tetramer[--] 2-Dimer Equilibrium.* Biochemistry-Us, 1993. **32**(7): p. 1695-1706.
17. Park, H., T.D. Bradrick, and E.E. Howell, *A glutamine 67 -> histidine mutation in homotetrameric R67 dihydrofolate reductase results in four mutations per single active site pore and causes substantial substrate and cofactor inhibition.* Protein Eng, 1997. **10**(12): p. 1415-1424.
18. Bradrick, T.D., J.M. Beechem, and E.E. Howell, *Unusual binding stoichiometries and cooperativity are observed during binary and ternary complex formation in the single active pore of R67 dihydrofolate reductase, a D-2 symmetric protein.* Biochemistry-Us, 1996. **35**(35): p. 11414-11424.
19. Strader, M.B., et al., *Defining the binding site of homotetrameric R67 dihydrofolate reductase and correlating binding enthalpy with catalysis.* Biochemistry-Us, 2004. **43**(23): p. 7403-7412.
20. Amyes, S.G. and J.T. Smith, *The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R388.* Eur J Biochem, 1976. **61**(2): p. 597-603.
21. Deng, H., R. Callender, and E. Howell, *Vibrational structure of dihydrofolate bound to R67 dihydrofolate reductase.* J Biol Chem, 2001. **276**(52): p. 48956-48960.
22. Krahn, J.M., et al., *Crystal structure of a type II dihydrofolate reductase catalytic ternary complex.* Biochemistry-Us, 2007. **46**(51): p. 14878-14888.
23. Feng, J., et al., *Radical redesign of a tandem array of four R67 dihydrofolate reductase genes yields a functional, folded protein possessing 45 substitutions.* Biochemistry-Us, 2010. **49**(34): p. 7384-7392.
24. Feng, J., S. Goswami, and E.E. Howell, *R67, the other dihydrofolate reductase: rational design of an alternate active site configuration.* Biochemistry-Us, 2008. **47**(2): p. 555-565.
25. Martinez, M.A., et al., *Exploring the functional robustness of an enzyme by in vitro evolution.* EMBO J, 1996. **15**(6): p. 1203-1210.
26. Zhuang, P., et al., *Artificial duplication of the R67 dihydrofolate reductase gene to create protein asymmetry. Effects on protein activity and folding.* J Biol Chem, 1993. **268**(30): p. 22672-22679.

27. Dam, J., et al., *Complementation between dimeric mutants as a probe of dimer-dimer interactions in tetrameric dihydrofolate reductase encoded by R67 plasmid of E-coli.* J Mol Biol, 2000. **302**(1): p. 235-250.
28. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-2221.
29. Vedadi, M., et al., *Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination.* P Natl Acad Sci USA, 2006. **103**(43): p. 15835-15840.
30. Rodrigues, J.V., et al., *Protein stability in an ionic liquid milieu: on the use of differential scanning fluorimetry.* Phys Chem Chem Phys, 2011. **13**(30): p. 13614-13616.
31. Bastien, D., et al., *Fragment-Based Design of Symmetrical Bis-benzimidazoles as Selective Inhibitors of the Trimethoprim-Resistant, Type II R67 Dihydrofolate Reductase.* J Med Chem, 2012. **55**(7): p. 3182-3192.
32. Blakley, R.L., *Crystalline Dihydropteroylglutamic Acid.* Nature, 1960. **188**(4746): p. 231-232.
33. Schagger, H., *Tricine-SDS-PAGE.* Nat Protoc, 2006. **1**(1): p. 16-22.
34. Savitzky, A. and M.J.E. Golay, *Smoothing + Differentiation of Data by Simplified Least Squares Procedures.* Anal Chem, 1964. **36**(8): p. 1627-&.

Chapter 4 - Prediction of ligand binding to steer rational protein engineering

4.1. Context

This chapter was directly inspired by the research presented in Chapter 3. Enzyme engineering is highly complex and comprises more variables than we could ever cover using a sequential search of the available sequence space for mutagenesis. Thus, creating small and efficient 'smart libraries' specifically tailored for a specific substrate is of prime importance to reduce the experimental labour and screening efforts for enzyme engineering. Many heuristic methods exist, many of which were introduced and discussed in Chapter 2. The approach I present in this chapter combines MD simulations, molecular docking and experimental techniques to characterize the industrially relevant cytochrome P450 enzyme from *Bacillus megaterium*, CYP102A1 (BM3). Based on this proof-of-concept system, we were able to identify, without prior knowledge, all the main mutational hot-spots that have been characterized experimentally over the past two decades. This represents a particularly advantageous technique for enzymes that do not benefit from many reports of mutagenic data. Experimental verification remains necessary and should not be neglected in any computational study. It is important to combine experimental and computational studies to increase their impact.

This chapter is a "pre-print" version of the article entitled '*Rapid substrate-specific screening for mutational hotspots using biased molecular dynamics simulations*', currently under review in *ACS Catalysis*. My contribution was the conceptualization and realization of laboratory and computational experiments. The computational work was performed under the supervision of Prof. Guillaume Lamoureux, while the experimental work was performed in the laboratory of Prof. Joelle Pelletier. The manuscript was drafted by myself with assistance from Prof. Pelletier and revision by Prof. Lamoureux.

Rapid substrate-specific screening for mutational hotspots using biased molecular dynamics simulations

Maximilian C.C.J.C. Ebert[1,2] , Guillaume Lamoureux[2,3] and Joelle N. Pelletier[1,2,4]

[1] Département de biochimie and Center for Green Chemistry and Catalysis (CGCC), Université de Montréal, Montréal, QC, Canada H3T 1J4

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, QC, Canada G1V 0A6

[3] Department of Chemistry and Biochemistry and Centre for Research in Molecular Modeling (CERMM), Concordia University, Montreal, QC, Canada H4B 1R6

[4] Département de chimie, Université de Montréal, Montréal, QC, Canada H3T 1J4

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

4.2. Abstract

Prediction of substrate-specific mutational hotspots for enzyme engineering is a complex and computationally intensive task. This becomes particularly challenging when the available crystal structures have no ligand, bind a distant homolog of the desired substrate or hold the ligand in a non-productive conformation. To address that shortcoming, we present a combined molecular dynamics simulation and molecular docking protocol to predict the conformation of catalytically-relevant enzyme-ligand complexes even in the absence of a ligand-bound structure. We applied the Adaptive Biasing Force (ABF) method to predict the ligand-specific path of diffusion of a fatty acid substrate from the bulk media into the active site of cytochrome P450 CYP102A1 (BM3). Starting with a ligand-free crystal structure, we successfully identified all residues known to be involved in palmitic acid binding to BM3. The binding trajectory also revealed a new binding residue, Q73, which we confirmed experimentally. Building the free energy landscape illustrates that, similar to human cytochrome P450s, binding is multi-step and does not follow simple Michaelis-Menten binding kinetics. We then applied the predicted BM3:palmitate complex to molecular docking of a library of 29 palmitate analogues. This produced catalytically-relevant binding poses for the entire library, in contrast with the use of ligand-free and ligand-bound crystal structures. This fast and simple computational method is broadly applicable for predicting binding hotspots in a substrate-specific manner and has the potential to drastically reduce the experimental screening effort to tailor an enzyme to new substrates of interest.

Keywords: Biocatalysis, protein engineering, adaptive biasing force, binding mode prediction, virtual screening, energy landscape

4.3. Introduction

In the context of biocatalysis, high substrate selectivity of enzyme-catalyzed reactions can be both beneficial and limiting. Biocatalysts often provide incomparable stereo-, regio- and chemoselectivities, thus reducing the expense of separating undesired side-products. Nonetheless, high selectivity can constitute a barrier to catalyzing a desired, non-native reaction. Enzyme engineering strives to address this limitation. Broadening or altering the substrate scope of enzymes has been accomplished by approaches ranging from random mutagenesis [1-3], to semi-rational [4-6] and rational design [7]. Identifying mutational hotspots for the generation of 'smart' libraries for rational enzyme design is of prime importance to reduce the search space and thereby increase the efficiency with which new or improved catalytic activities are obtained.

Successful approaches to modulate enzyme reactivity increasingly include computational methodologies as a means to reduce the laboratory screening effort to those enzyme variants predicted to have the highest likelihood of success. Such approaches include structure-guided, iterative library screening [8] as well as computationally-guided library screening [9]. Computational information relevant to enzyme engineering is based on molecular dynamics (MD) simulations [10], alignment of primary, secondary and tertiary structure [11] and docking studies [12, 13] and may focus solely on the active site region [14, 15]. Docking algorithms typically take no account of the path of entry of a compound into the active site, thus neglecting aspects that are essential to the binding process: only the final, bound pose of the desired substrate inside the active site is considered. The approach of the substrate to/from the active site and the binding path can only be studied using biased MD simulations [16], because unbiased MD simulations are too computationally intensive to overcome high energy barriers that prevent the substrate from approaching the enzyme surface or entering into the active site.

Several biased computational approaches can be used to study substrate binding and to generate the potential of mean force (PMF) of the protein/ligand complex formation. Among the most widely applied methods are steered molecular dynamics (SMD) and umbrella sampling (US) [17-22]. Those methods investigate substrate egress and binding, protein folding or protein-protein/protein-membrane interactions. In SMD, the substrate or protein is

pulled with a constant force or constant velocity along a reaction coordinate. Those pulling methods, however, require careful tuning. A constant pulling force should be strong enough to overcome the steepest energy barrier along the pathway, but such a strong force may be excessive in other sections of the pathway. A constant pulling velocity should be slow enough to accommodate the long time-scale movements that may be required from the protein, but moving at such a slow velocity over the entire pathway may be unproductive. US is a widely used technique to extract the PMF requiring a complex multistep trial and error setup. Once a reaction coordinate is defined, the reaction pathway is explored over multiple "windows", each corresponding to a separate simulation biased such that the ligand remains within a narrow region of the reaction coordinate. Each new window is typically started from the conformation of a nearby window, to ensure a significant conformational overlap from one window to the next and to allow a continuous free energy landscape to be calculated. Finally, the PMF is extracted by subjecting the simulation of each window to the Weighted Histogram Analysis Method (WHAM) [23].

It would be desirable to have a guided approach to model substrate binding where artefacts due to bias are minimized and where the simulation setup is simpler than umbrella sampling and related techniques. To that end, we applied the Adaptive Biasing Force (ABF) computational method to evaluate the substrate-binding path and the energies for binding a substrate into an enzyme active site [24-27]. The ABF method calculates the mean force along a reaction coordinate describing ligand binding and, at each point of that reaction coordinate, applies an equal and opposite force. This allows the ligand to escape local energy minima and cross energy barriers in the free energy landscape of the reaction coordinate. Because the opposing force is known, revisiting each position of the reaction coordinate provides a summation of the mean force required to overcome each barrier. The process converges when binding of the ligand is governed by Brownian motion described by a flat free energy landscape, where all barriers and wells have been mapped and cancelled.

The advantages of ABF over other biased MD techniques are two-fold: first, unlike SMD, which applies a constant force (or enforces a constant velocity), ABF applies an adaptive force, which allows the protein to relax its conformation at each step along the ligand-binding pathway; second, ABF consists of a single MD simulation with a continuous trajectory, rather than separate MD simulations that need to be combined as in umbrella

sampling. Early applications of ABF to protein systems have reported on transmembrane channels [28-31]. ABF is rapidly gaining in importance, and has recently been applied to study anion binding to the porin channel OprP [32], interaction between CCR5 and the HIV-1 entry inhibitor maraviroc [33] and the regulation of CD44 transmembrane glycoprotein [34]. Importantly, ABF does not require a ligand-bound crystal structure as the starting input, as opposed to methods like molecular docking. Because the input is independent of the ligand under investigation, this method has the potential to be applied to a wide range of binding partners.

We introduce for the first time, to the best of our knowledge, the application of the ABF computational method to guide enzyme engineering. Our target of interest is the cytochrome P450 enzyme CYP102A1 (BM3) from *Bacillus megaterium*. As is the case for a number of enzymes, cytochrome P450s present a tunnel-like entrance to their deeply buried active site, and thus bear a resemblance to the transmembrane channels successfully modeled using ABF [35, 36]. P450 enzymes catalyze the oxidation of non-activated carbon atoms, which is chemically demanding but of the utmost industrial and pharmacological importance [37-41]. They are among the most valuable industrial enzymes because of their ability to perform highly controlled oxidations without the strong oxidants and harsh conditions required in classical organic chemistry. They obviate the need for specific activation of the target reaction site and require no protective groups, thus eliminating reaction steps. Thousands of P450s have been identified, catalyzing oxidative reactions including alkene epoxidation, aromatic hydroxylation, dehalogenation and deamination [40]. P450s are the target of numerous enzyme engineering efforts to broaden their substrate scope and direct their regioselectivity, yet these efforts stand to gain much in efficiency [40, 42]. ABF offers the potential to investigate the process of substrate approach and entry into the active-site tunnel of P450 enzymes, thereby potentially improving structure-based engineering.

Using the ABF method, a single simulation allowed the identification of all residues known to be implicated in fatty acid binding, confirming the predictive accuracy of the method as well as its utility. In addition, we identify a new residue involved in binding and confirm its role by mutagenesis and activity assays. Furthermore, by mapping the 2-dimensional energy landscape describing substrate entry into the active site, we support experimental data concerning the stepwise binding mode of fatty acid substrates. Finally, the

predicted catalytic conformation of BM3 was directly used to dock a variety of analogs of the substrate, palmitic acid, providing a rapid method to attain catalytically-viable binding poses where the use of crystalized conformations fails.

4.4. Results and discussion

4.4.1. Interaction between palmitic acid and BM3

We and others have previously studied the transient nature of P450 substrate channels [35, 36, 40]. There are at least six distinct substrate channels leading in and out of the active site cavity and the main tunnel of BM3 (termed 2a, 2b, 2d, 2f, S, and W). Those channels are in constant flux between the open and closed states. The substrate preference of each channel strongly depends on the nature of the substrate. In addition, fatty acid substrate binding is mainly steered by a distinctive 'fetching mechanism' in BM3 situated at the mouth of the channel 2 regions, promoting entry into channels 2a, 2b and 2f [43, 44]. This fetching mechanism comprises residues R47 and Y51, which form weak initial interactions with the fatty acid substrate carboxylate to lure the alkyl chain towards the substrate channels. Once bound, the fetching residues tether the carboxylate, constraining motion of the substrate to ensure regiospecific hydroxylation of the ω-1-to-3 carbons. Second-shell residues such as S72 and A74 (among others) contribute to the tether and their mutation can alter substrate specificity [45-47].

In a first attempt to undertake substrate-specific prediction of mutational hotspots, we performed computational docking. While both apoenzyme and ligand-bound crystal structures of BM3 have been resolved, no crystal structure of BM3 exists with a fatty acid or a substrate analogue bound in a catalytically relevant orientation. For example, the BM3 structure 1ZO9 is resolved to 1.7 Å resolution; its co-crystallized substrate analogue *N*-palmitoylmethionine has its ω-1-to-3 carbons ≥ 6 Å from the heme iron. This orientation is typical of all BM3 structures in the PDB harboring an aliphatic, long-chain ligand in the active site (more than a dozen), yet the distance is too great to describe reactivity.

We attempted to identify a catalytically active binding pose for the natural substrate palmitic acid by docking into BM3 using Autodock Vina [48]. The coordinates used for the

rigid receptor were either of the BM3 apoenzyme (PDB file 2HPD) or a structure from which the co-crystallized substrate analogue *N*-palmitoylmethionine (PDB file 1ZO9) had been removed. The latter was used to ensure that there was sufficient space for substrate binding inside the active-site cavity; this strategy can be helpful when the docking protocol does not allow flexibility in the receptor protein. The search grid comprised the entire main access tunnel including the active-site cavity and the surface residues at the mouth of the tunnel 2 region.

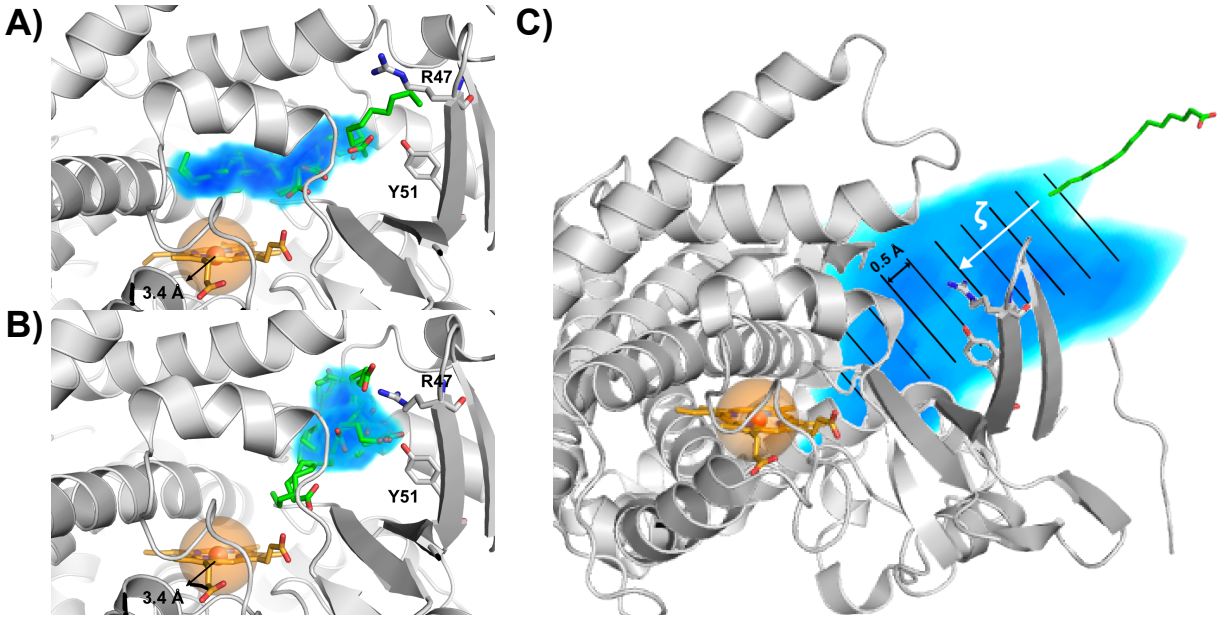


Figure 4-1 Molecular docking of palmitic acid to BM3 and ABF principle.
Molecular docking of palmitic acid in A) ligand-bound 1ZO9 and B) ligand-free 2HPD coordinates. Poses with the carboxylate in proximity to the heme iron were removed. Individual poses resulting from docking are all overlaid and shown in sticks. The blue volume represents the space occupied by an atom in at least 50% of all poses. C) The ABF method. The substrate situated in the bulk solution is pulled along the reaction coordinate (white arrow) towards the heme iron. The path is divided into windows (bins) separated by 0.5 Å, inside each of which the force of the system is probed before applying the ABF. The blue volume is a schematic representation of the channel 2 region. R47 and Y51 are shown as sticks for orientation. In all panels, the orange sphere represents a volume with a 3.4 Å radius around the heme iron.

The entirely flexible palmitic acid substrate molecule was docked and the 20 best poses were ranked based on predicted binding affinity. Poses with the carboxylate close to heme iron

were removed, as the carboxylate cannot travel through the hydrophobic tunnels. The remaining poses docked in 1ZO9 (BM3 with substrate analogue removed) clustered around the crystallized position of *N*-palmitoylmethionine (Figure 4-1 A). Accordingly, the ω-1-to-3 carbons of palmitic acid were not at a catalytically relevant distance of the heme iron. Of yet greater concern, none of the 20 best poses was within the active-site cavity when the ligand-free crystal structure 2HPD was used. The palmitic acid molecules clustered at the entrance of the tunnel, suggesting that the tunnel in that structure was crystallized in an obstructed conformation, preventing ligand docking. This resulted in a predicted binding affinity for palmitic acid of -4.9 kcal/mol in 2HPD; the poses closer to the heme in 1ZO9 may represent a binding intermediate, with a binding affinity of -6.5 kcal/mol.

4.4.2. Steering palmitic acid towards a catalytically relevant orientation

While molecular docking can provide information about key residues involved in substrate or inhibitor binding [16, 49], here it failed to provide satisfactory insights into productive substrate binding due the transient nature of the substrate tunnels in P450 enzymes. The success rate can be enhanced by docking against an ensemble of protein conformations, yet ensemble docking requires further development to increase its success rate [50]. One can also perform a free molecular dynamics simulation with the ligand present, yet full simulations of receptor-ligand complexes are lengthy and computationally expensive [16, 51].

To overcome those limitations, biased molecular dynamic simulations have been used to study receptor-ligand interactions [21, 22, 52, 53]. Here, we reduce simulation time while limiting the error resulting from a too great bias, by applying the adaptive biasing force (ABF) method to the problem of ligand binding. Because the applied adaptive force overcomes local energy barriers, the substrate can progress along the reaction coordinate (Figure 4-1 C). Ideally, all conformations along the reaction coordinate are sampled with an equal probability to improve the accuracy of the free energy calculation. To study the mode of binding of palmitic acid to BM3 (2HPD, open form), including its binding path, the substrate was placed in the bulk solvent, 6 Å away from the protein surface and above the cluster of tunnels purported to transport the substrate (channel 2 region) [35, 36]. The reaction coordinate during

the ABF simulation was defined as the distance between the ω-1 carbon and heme iron (Figure 4-1 C). The force experienced by the substrate travelling along the z-axis was evaluated for every 0.5 Å-wide window (or bin) and energy of the opposite sign was applied to balance the force to zero.

Upon leaving the bulk solvent, the substrate conformations converged toward the mouth of active site tunnel 2b. The aliphatic chain of the palmitate then slipped into the hydrophobic tunnel and gradually moved into a catalytically relevant orientation up to 4 Å away from the heme iron, generally within the first 2 ns of the trajectory. During the remainder of the 10 ns trajectory, the aliphatic chain probed a variety of orientations within the active site cavity.

Out of the 1000 conformations analyzed, we extracted the three most highly populated clusters of orientations, based on the orientation and distance of the ω-1-carbon relative to the heme iron. Cluster 1 (83 out of 1000 conformations) represents the catalytically relevant conformation, with the ω-1-carbon held within 5 Å of the heme iron. The substrate carboxylate was anchored to the mouth of the active-site tunnel mainly by H-bonding with R47, and further oriented by either N70 or Q73. This resulted in two main populations of conformations (Figure 4-2 A,B). This observation lends weight to previous reports suggesting that those residues form a hydrophilic binding region to tether fatty acid carboxylates [43, 44].

Inspection of all conformations in the ensemble revealed that the ω-1 carbon of palmitate was within 5 Å of the following residues: F87, I263, A264, G265, T268, A328, L437 and T438. Interestingly, mutagenesis studies have identified each of those residues as directly affecting the hydroxylation site of palmitate, where they define substrate specificity and regioselectivity [54-59]. Our computational analysis is thus consistent with the experimental reports on the mode of fatty acid binding to BM3.

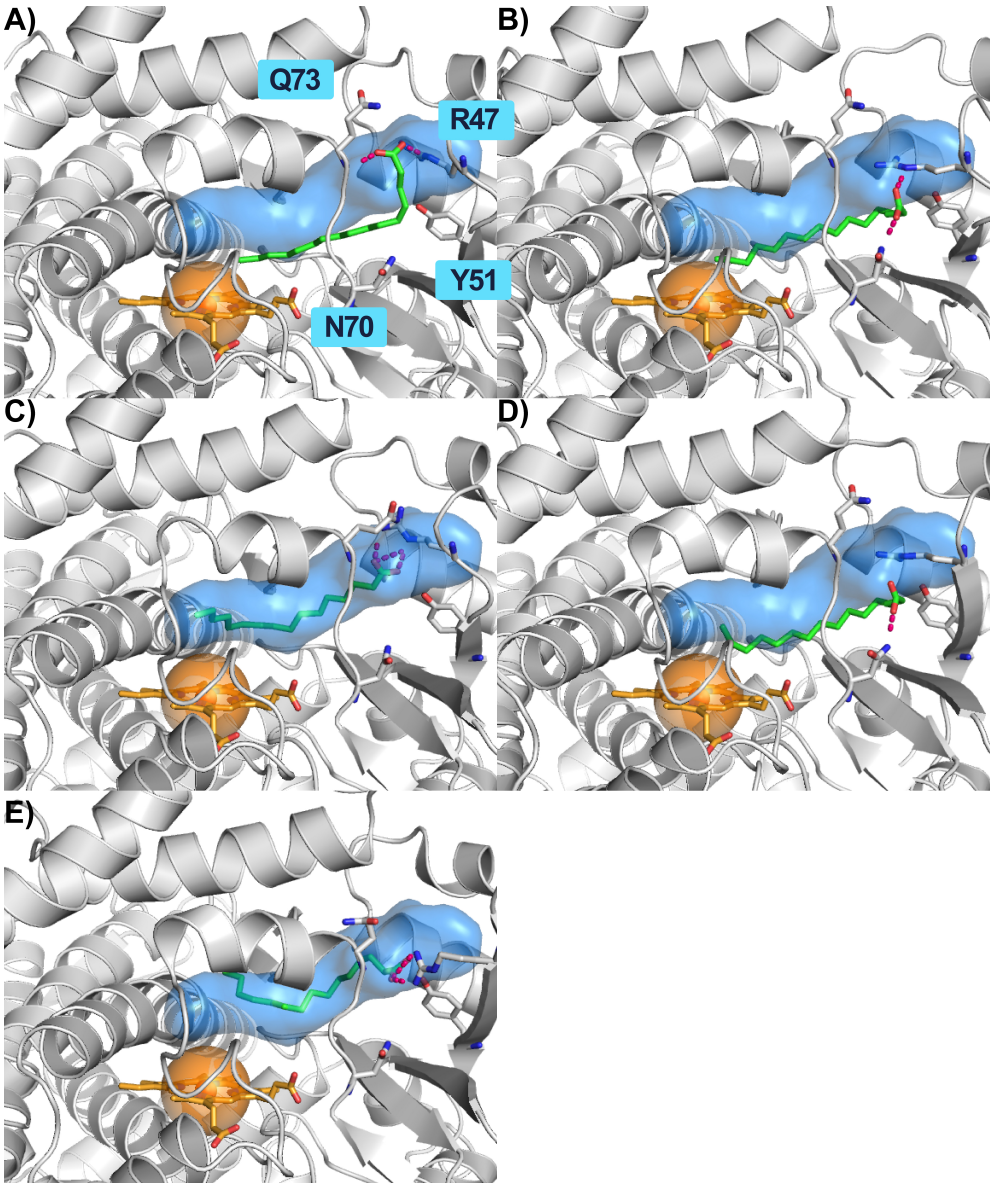


Figure 4-2 Result of ABF simulation using one reaction coordinate.
A, B) The two main populations of conformations in Cluster 1, defined as the catalytically-relevant orientation. C, D) The two main populations of conformations in Cluster 2, which overlays with the crystallized conformation for ligands with long aliphatic chains. E) The main conformation in Cluster 3, a new non catalytic conformation. For orientation, a volume was created (in blue) that surrounds all ligands with long aliphatic chains crystallized in BM3 and reported in the PDB (PDB codes 3BEN 3CBD 2UWH 1ZO9 1ZOA 1SMJ 1JPZ 3KX3 4KPA 3WSP 5DYZ 4ZFB 4ZFA 4ZF6 5JQ2 1FAG 3EKD). The orange sphere represents a volume with a 3.4 Å radius around the heme iron and key residues R47, Y51, Q73 and N70 are shown in sticks

Cluster 2 (241 out of 1000 conformations) kept the palmitate inside the same tunnel yet with the ω-1-to-3 carbons held too far from the heme iron to be reactive. Those conformations correspond to the ensemble of conformations adopted by long-chain aliphatic substrates in crystal structures. Indeed, some of the conformations of palmitate found in this cluster are exact mimics of the crystallographic pose (Figure 4-2 C). Nonetheless, the conformational space within the cluster is not limited to the crystallized substrate orientations: while the ω-1-to-3 carbons explore a restricted region deep in the active site pocket, the remainder of the aliphatic chain and the terminal carboxylate explore space more broadly than in the crystal structures (Figure 4-2 D). In contrast to the catalytic conformation (cluster 1), R47 is the dominant H-bonding partner while N70 and S72 are also plausible interaction partners for the carboxylate of palmitate. Indeed, S72 has been proposed to extend the fatty acid carboxylate tethering mechanism and has been shown to bias selectivity in lauric acid oxidation [45]. The environment of the ω-1 carbon includes the eight residues identified in cluster 1, in addition to L75, V78, A82, T88, L181 and T260. Docking studies have shown the importance of T260/T268 in orientating coumarin, while mutation of A82 with bulky residues appears to restructure the active-site pocket to procure productive binding and oxidation of indole [13, 58]. One such mutation, A82F, also drastically increases the oxidation of long-chain fatty acids suggesting an enrichment of the catalytically relevant cluster by filling up the cavity with its bulky side-chain in which ω-1-to-3 carbons of the substrate were found in the crystal structures. This might eliminate this unreactive intermediate conformation. This hypothesis is also supported by the orientation of palmitic acid in the 2UWH crystal incorporating the A82F mutation. While the distance between the ω-1 carbon and the heme iron is still over 7 Å, it lines up perfectly above the heme iron like found in cluster 1.

Cluster 3 (341 conformations) is a further ensemble of non-catalytic conformations similar to cluster 2 for the position of both termini of the palmitate, with the alkyl chain mirrored in the plane of the substrate tunnel (Figure 4-2 C). No similar conformation has been crystallized. H-bonds with R47 are predominant (78% of the frames), consistent with its proposed role in substrate fetching and also to stabilize the substrate within the active site pocket. T49, N70 and Q73 were minor H-bonding contributors (1%, 2% and 15% of the frames, respectively). New residues interacting with the ω-1 carbon were M177, V178, M185, L262, H266, E267, T436, K434.

The remaining 335 of 1000 conformations represent early events during the simulation while slipping into the active site, or later transitions between the clusters. R47 dominates as the H-bonding partner of palmitate over the course of the entire trajectory (Figure 4-3 A). The second-shell residues of the fetching mechanism, S72 and Y51, were observed with low occupancy. In contrast, the little-known residues N70 and Q73 contributed significantly to the predicted substrate fetching.

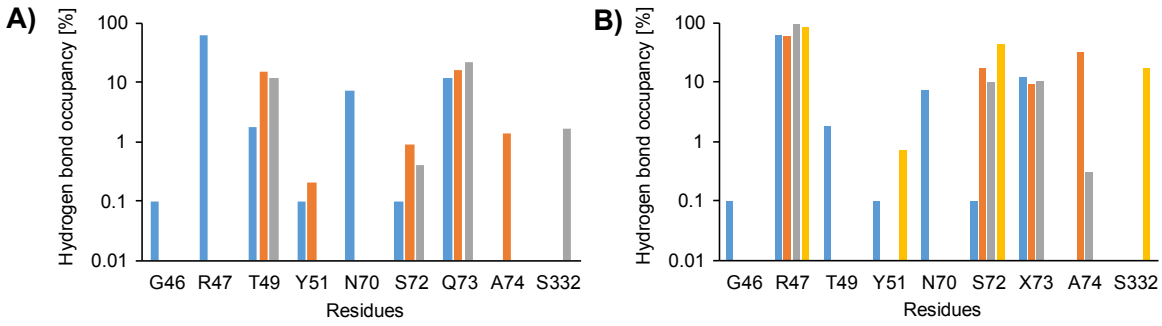


Figure 4-3 Hydrogen bonding between the palmitic acid carboxylate and BM3 variants. *H-bond occupancy during the entire 10 ns 1D ABF simulation is shown for A) WT BM3 (blue) and variants R47A (orange) and R47A/Y51F (grey); B) WT BM3 (blue) and variants Q73N (orange), Q73M (grey) and Q73E (yellow), where X73 is varied accordingly. The Y-axis is on a logarithmic scale.*

In order to gain a better appreciation of the impact of second-shell binding residues in substrate tethering, we analyzed the binding trajectories to the R47A and the R47A/Y51F mutants, thus removing the side-chains of the main fetching mechanism. In the absence of the R47 anchor, the R47A mutant H-bonds with the substrate in only 32.7% of the frames during the trajectory, compared to 59.3% for the WT. We could not identify cluster 2 and 3 in these trajectories, which indicates a much higher mobility of the palmitate carboxyl group. The substrate is poorly stabilized, moving up and down the tunnel. As observed in the WT, Y51 H-bonded with the substrate infrequently and thus appears to have a greater impact in fetching the substrate out of solution than in tethering the substrate once inside the active site, where R47 dominates. No new second shell binding residues were identified in the double mutant, except for a low occupancy for S332 and hence showing a similar destabilized substrate in the active site tunnel.

It was previously speculated that Q73 is at least implicated in binding of amino acid derivatives of fatty acids [60]. Interestingly, this residue was a significant H-bonding partner of palmitic acid in all trajectories (Figure 4-3 A). To estimate its impact on long chain fatty acid binding, we recorded trajectories for variants Q73E, Q73M and Q73N. The neutral substitutions Q73M and Q73N were similar to WT in H-bonding between the main- and side-chain of R47 or position 73 and the substrate carboxylate. We observed a higher percentage of interaction with S72 and no interaction with T49 for all Q73X mutants. However, the additional negative charge at the tunnel mouth in the Q73E mutant shifted the substrate carboxylate toward other H-bonding partners. The negatively charged side-chain introduced in Q73E results in the substrate forming more H-bonds with Y51. This increases the contribution of Y51 to the fetching mechanism, relative to R47. Introduction of Q73E also increases the contribution of residue S332 to H-bonding with the substrate, which had previously been observed only in the trajectory of the R47A/Y51F double mutant. Experimental verification of the kinetic parameters of Q73X mutants supports the computational results, where the greatest impact on substrate binding was a 5-fold increased K_M for Q73E (Table 4-1). The impact of this mutant has the same effect on K_M as the R47A mutations, which abolishes the fetching mechanism [43]. The mutations had no significant impact on k_{cat}, further supporting Q73 as being involved in productive binding rather than turnover.

Table 4-1. Kinetic parameters for the substrate lauric acid and BM3 variants.

Mutant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (s^{-1} μM^{-1})
WT	17 ± 0.4	190 ± 11	90 ± 5.5
Q73N	17 ± 0.5	390 ± 23	43 ± 2.8
Q73M	24 ± 0.5	200 ± 11	110 ± 6.4
Q73E	19 ± 1.4	920 ± 110	21 ± 3.0
Q73A	24 ± 0.4	160 ± 7	150 ± 7.2
Q73S	24 ± 0.5	210 ± 12	120 ± 7.0
R47A [a]	66 ± 3.0	860 ± 79	77 ± 7.9

[a] Data from [43]

4.4.3. Free energy landscape of palmitic acid

None of the above ABF trajectories converged over the course of the 10 ns simulation and hence not free energy landscape could be calculated. Achieving convergence would have required more extensive sampling of the binding space and release of the substrate from the active site. However, the energy bias applied to the ω-1 carbon could not overcome the palmitate binding energy: the strong tethering of the substrate carboxylate by R47 combined with the acyl chain flexibility prevented its egress. Extending the simulation for the WT to 50 ns did not result into the release of the ligand but rather in the distortion of secondary structure elements of the protein. This distortion occurs as response to the ω-1 carbon pushing out the strongly tethered carboxyl group. The only solution for the ABF method, given the reaction coordinate defined, is to push the ω-1 carbon through the active site walls into the bulk media. To overcome this limitation in the sampling efficiency, we introduced a second reaction coordinate corresponding to the distance between the centers of mass of the R47 guanidinium and of the palmitate carboxylate. After 100 ns of the total 190 ns simulation time, no further change in the 2-dimensional free energy landscape was observed and an equilibrium was achieved (Video A 2-1). The palmitate samples the entrance, tunnel and active site before being released and sampling the tunnel mouth and surrounding surface (Video A 2-2). Three energy minima were identified (M1-3), all located within 10 Å of the second reaction coordinate, emphasising the importance of the R47 residue for fatty acid binding (Figure 4-4 A). All energies are expressed with respect to the lowest free energy state in M2.

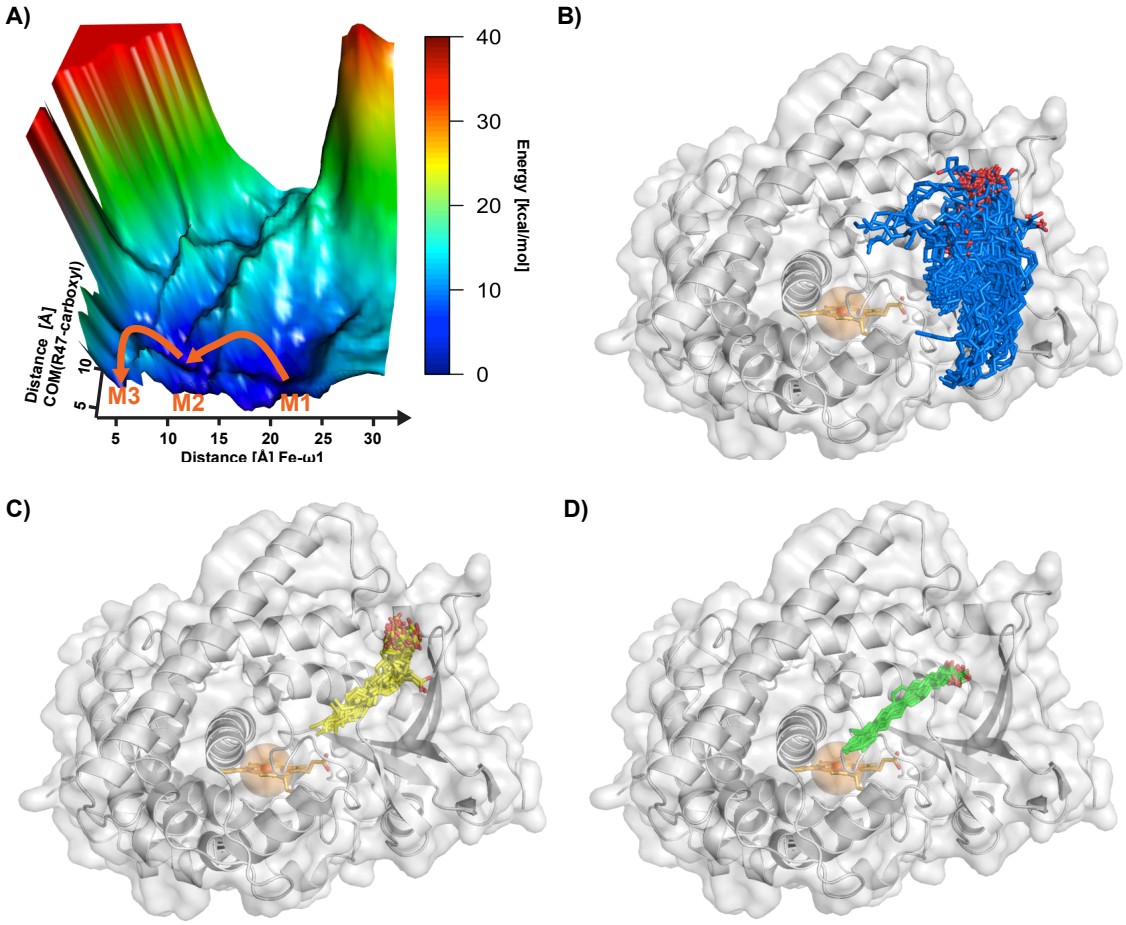


Figure 4-4. Result of ABF simulation using one reaction coordinate.
Result of Two-dimensional ABF simulation of palmitic acid binding to CYP102A1. A) Free energy landscape defined according to the two reaction coordinates: 1) distance between the heme iron and ω-1 of palmitic acid; 2) distance between the centers of mass of the R47 guanidinium and the substrate carboxylate. Energy minima M1-3 are labelled in orange. B-D) Representation of the conformational ensemble of palmitic acid bound in the energy minima B) M1, C) M2 and D) M3. The orange sphere represents a volume with a 3.4 Å radius around the heme iron.

While the palmitic acid at the beginning of the simulation was positioned proximal to the channel most likely to conduct the fatty acid, namely channel 2b, a number of channels were probed during the course of the simulation. The energy landscape confirms that the mouths of channels 2a and 2f were visited, as well as channel 2e that had never yet been observed in BM3. Minimum M1 represents the initial binding or 'fetching' of palmitic acid, where the carboxylate was anchored to the protein surface while the ω-1 carbon remained at

the protein surface. M1 was situated exclusively at the mouth of the channel 2 region, thus supporting the dominant contribution of this entrance for fatty acid 'fetching' and binding (Figure 4-4 B). The H-bonding interactions in this ensemble identify residues of the β-1 1-2 strand and B'-helix as being important for substrate fetching (Table 4-2). The energy found in M1 is 0.2 kcal/mol higher than in M2. The highest energy barrier on the direct path from M1 to M2 is 2.2 kcal/mol with respect to M2.

Minimum M2 represents an intermediate binding step towards the catalytically relevant orientation and displays the lowest energy of the three energy minima (Figure 4-4 C). Backbone and side-chain H-bonds were established between the substrate and the 1-1 β-sheet and 1-2 strand, including strong interactions with F42 and A44 in 1-1 and T49 and Y51 in 1-2 (Table 4-2). The importance of F42 in fatty acid binding was previously demonstrated [43]. It was speculated that F42 acts as a lid to protect the hydrophobic tunnel region. Our results indicate that F42 may also stabilize the substrate carboxylate by backbone H-bonding, while providing the hydrophobic character propitious to stabilizing the alkyl chain. Although little to nothing is known for A44 and T49, Y51 was previously identified as playing a key role in substrate binding and stabilization [61]. The importance of the B'-helix for substrate recognition and binding is further supported by residues Q73 and F77 forming substrate interactions in M2. Residues K187-R190 in the F-G loop also contributed strong interactions. Among these, we note that mutation of L188 in BM3 was essential for oxidation of propane, a gaseous substrate [62]. M354 was extensively exploited for enhanced activity mutants of BM3 using a F87A template[63, 64]. Before reaching the final orientation in M3 palmitic acid has to cross an energy barrier from M2 to M3 of 5.4 kcal/mol with respect to M2 (highest energy barrier on the path of lowest energy from M2 to M3).

Finally, minimum M3 represents the catalytically-relevant orientation corresponding to the catalytic conformation of the binding experiments using a single reaction coordinate. The ω-1 carbon is between 3-4 Å of the heme iron and the carboxylate nicely anchored similar to the crystallized conformation (Figure 4-4 D). We observed similar H-bonding partners, with the striking exception of N70 (Table 4-2 and Figure 4-3 A). The importance of N70 remains inconclusive since this interaction is also absent in the crystal structures with long aliphatic chain substrates. The energy difference from M3 to M2 is 1.1 kcal/mol, which is higher than the energy difference between M1 and M2 (0.2 kcal/mol). The higher energy found in M1

could be in part coupled to the product release mechanism. H-bonding partners in M1-3 correlated with the residues identified in the single reaction experiment, however the 2D experiments appears to cover better weak interaction with second shell residues (Table 4-2 and Figure 4-3 A).

Table 4-2. Residues forming hydrogen bonds with palmitic acid in the energy minima M1 to M3.

Minimum	Hydrogen bonds with palmitate
M1 mouth	P45, G46, R47, Q73, K76, F77
M2 tunnel	L17, F42, A44, T49, Y51, Q73, F77, K187, L188, R190, M354
M3 reaction	L20, F42, R47, Y51, S72, Q73, L188, M354

4.4.4. Palmitic acid binds to BM3 according to a stepwise mechanism

In addition to its application in biocatalysis, BM3 is increasingly applied to the production of pharmaceutical drug metabolites to accelerate the process of drug approval. Approximately 50% of drugs on the market are metabolized by CYP3A4, a human P450, yet the low activity and stability of human P450 variants render them impractical for industrial applications [40]. The successful substitution of the human CYP3A4 variant with the microbial BM3 rests on its higher similarity with eukaryotic P450s compared to bacterial counterparts [65-67]. Human CYP1A2 also shares a similar oxidation profile and high catalytic activity with BM3 [68]. CYP1A2 and CYP3A4 have been well characterized. Their substrate binding is a complex three-step process beginning with interaction at a peripheral site, translocation to the interior of the channel and final positioning in the catalytic orientation. The sequential binding in BM3 via the M1, M2 and M3 minima that we observe here describes a strikingly similar mode of binding (Figure 4-4). In the absence of any report on the mode of substrate binding to BM3 or other microbial CYP [69, 70], we propose an evolutionary conservation of stepwise, non-Michaelis Menten binding of fatty acids in distant P450 homologs.

4.4.5. Expanding chemical space

In the absence of a catalytically-relevant orientation of a ligand in any BM3 crystal structure, the conformations in the energy minimum M3 could serve as the receptor in a docking study. Re-docking palmitic acid into of a common representative of the M3 conformations recreated the simulated binding pose in M3 with a RMSD of 2.6 Å, indicative of a high quality docking prediction. This motivated us to compare long-chain fatty acid derivatives of crystallized ligands in BM3 or kinetically characterized substrates, upon docking either into the open-state conformation of the ligand-free crystal 2HPD, the so-called closed state of the ligand-bound crystal structure 1ZO9 or a conformation of the M3 ensemble.

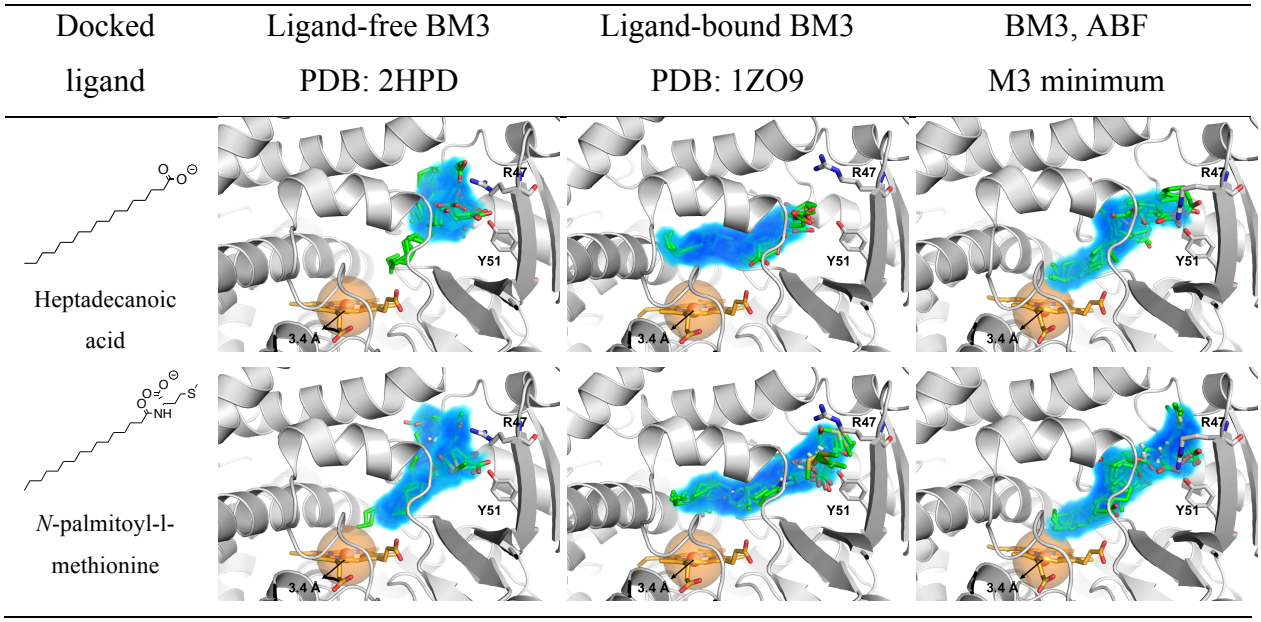


Figure 4-5 Docking result using the predicted catalytically relevant conformation and selected substrates.

Representative results of docking into the open form of BM3 (2HPD), the closed form of BM3 (1ZO9) and a representative conformation of the M3 ensemble. Heptadecanoic acid and N-palmitoyl-l-Met (ligand crystallized in 1ZO9) are shown. Individual poses resulting from docking each molecule are overlaid and shown in sticks. The blue volume represents the space occupied by an atom in at least 50% of all poses. The orange sphere represents a volume with a 3.4 Å radius around the heme iron. Key residues R47 and Y51 are displayed in sticks for orientation. The full dataset of 29 docked molecules is provided in Figure A 2-1.

After excluding non-native binding poses (carboxylate close to the heme and terminal carbons close to the surface), we calculated the volume populated by ligand atoms in at least 25% of all poses (Figure 4-5 and Figure A 2-1). One can clearly see a characteristic imprint left by each conformation. Docking into the ligand-free crystal 2HPD shows an agglomeration of poses close to the surface, similar to the binding mode found in the M2 energy minimum (Figure 4-4 C). None of the 29 ligands docked was found in a catalytically relevant pose, except for 1-octene. Hence, no information can be obtained from docking to the open state for long chain aliphatic compounds. Docking into the ligand-bound crystal structure 1ZO9 shows occupancy volumes for all ligands similar to the originally co-crystallized *N*-palmitoyl-L-methionine (Figure 4-5 and Figure A 2-1). The ligand portion close to the entrance of the active site tunnel shows generally a reasonable binding pose. Nonetheless, the terminal carbons of the bound ligands are not in a catalytically-relevant orientation. Because all long-chain fatty acid derivatives are crystallized in the same non-catalytic binding mode, docking into them will yield poor predictions for mutagenesis or mechanistic studies.

In contrast, docking into the conformation of minimum M3 predominantly generated binding modes where the ω-1 carbon is in a catalytically relevant distance to the heme iron. In addition, the carboxylate portion and the adjacent carbons of the binding poses are generally within 1-2 Å RMSD of the crystalized ligand in BM3. For example, the carboxylate and the α-2-to-8 carbons of palmitoleic acid in the BM3 WT structure 1FAG shows a RMSD of 1.9 Å compared to the best pose found for the M3 conformation. The remaining carbons of the aliphatic chain then drift into the catalytic binding pose. Therefore, in contrast to 1ZO9 this conformation could also steer the terminal carbons in an orientation in which we can get information about the beginning of the catalytic cycle.

Interestingly, the conformation extracted from the ABF energy minimum could also be used for ligands that are known to bind further from the heme iron. The binding mode of the decoy molecule n-perfluorononanoyl-l-tryptophan, which promotes the hydroxylation of gaseous alkanes, was reproduced (RMSD 3.4 Å) [60]. The ligand bound form also elicited the correct pose but the ligand-free crystal structure generated no pose with the correct binding mode, mostly due to steric clashes with the R47 side-chain.

4.5. Conclusion

Using recently-introduced MD techniques, we have directly identified most of the mutational hotspots that had been previously identified by the joint efforts of many research groups, in addition to identifying Q73 as a new residue involved in substrate binding. In contrast to many computational methods for the identification of mutational hotspots (HotSpot Wizard, ConSurf and similar), ABF does not rely on pre-existing information of sequence-structure-function relationships and can thus be applied to any enzyme structure of interest.

We further demonstrated that the substrate can be varied independently of the crystal structure. As a result, ABF is widely applicable to the generation of substrate-specific smart libraries for biocatalytic applications. ABF has the potential to serve as a universal tool for the prediction of the interactions between enzyme and bound ligands, or to identify residues located in binding pockets or access tunnels. In cases where access tunnels are unknown, there is also a potential to identify them using the free energy landscape. When used in conjunction with tunnel mapping tools such as CAVER or implicit ligand sampling, to quickly identify starting positions for the ABF pull experiment one can probe those channels for different substrates using ABF [35].

The ABF methodology should also correctly sample proteins undergoing large conformational changes on the nanosecond timescale. Other approaches such as PELE can potentially yield results similar to those presented here but not designed to model receptor-ligand complexes, which undergo conformational change upon binding [71]. The transient nature of the tunnel system in P450 enzyme are such a case in which a full MD trajectory should be able to model with much higher accuracy the conformational space in the active site.

Finally, docking close structural homologues of palmitate into the conformation of the putative catalytically-relevant conformation could also serve to predict the correct binding mode of inhibitors. The importance of the receptor conformation for the success of a docking study was recently reviewed by Kotev *et al*. [50]. To reduce computational costs to find a viable binding pose for a library of ligands, we propose a two-step protocol: first ABF using a representative ligand homologue of the ligand library to extract the predicted binding pose. This is followed by virtual screening of the library into this conformation.

Computational capacity remains a limiting factor for *in-silico* experiments. Since mapping the free energy landscape is computationally expensive, but not required for all applications, a single reaction coordinate was shown to be sufficient to identify at once the crystallized pose of the substrate well as a many valid mutational hot spots (Figure 4-2 C). That relatively short 10 ns trajectory served as an excellent first approximation, providing enough information to recognize the BM3 fetching mechanism and other determinants of substrate specificity and regioselectivity. The computationally more expensive ABF with two reaction coordinates was necessary, in this case, to map the free energy landscape. It provided more detailed information about the substrate binding mode and residues along the binding path, further supporting the pertinence of applying BM3 for pharmacokinetic analyses in pre-clinical trials.

Overall, ABF is a promising and broadly applicable method for predicting binding residues in a substrate-specific manner. This fast and simple method has the potential to drastically reduce time-consuming characterization of binding and experimental hotspot mapping of biocatalysts and thus become a generalized aid for engineering enzymes.

4.6. Materials and methods

4.6.1. Materials

All reagents were of the highest available purity. Restriction enzymes and DNA-modifying enzymes were from MBI Fermentas. Synthetic oligonucleotides were obtained from Sigma (USA). All aqueous solutions were prepared using deionized water purified with a Millipore BioCell system.

4.6.2. Mutagenesis

Starting from the construct p450 BM3 WT in pcWORI [72], all mutants were obtained using the megaprimer approach [73]. Following digestion with *BamH*I and *Sac*I, the resulting DNA fragments were religated into pcWORI and transformed into *E. coli* DH5α for protein expression. Ampicillin (Amp) was used at 100 µg/mL for plasmid maintenance. Sequences

were confirmed by DNA sequencing (ABI 3730 DNA sequencer, IRIC Genomic Platform at Université de Montréal).

4.6.3. Protein expression

An overnight culture (5 mL) was used to inoculate 500 mL of the following expression medium: Terrific Broth, ampicillin (100 µg/mL), thiamine (5 µg/mL), δ-aminolevulinic acid (12.5 µg/mL) and 125 µL trace elements (0.5 g $MgCl_2$, 50 g $FeCl_2 \cdot 6H_2O$, 1 g $ZnCl_2 \cdot 4\ H_2O$, 0.2 g $CoCl_2 \cdot 6H_2O$), 1 g $Na_2MoO_4 \cdot 2\ H_2O$, 0.5 g $CaCl_2 \cdot 2H_2O$, 1 g $CuCl_2$, 0.2 g H_2BO_5 in 1 L of 10% HCl. Cultures were propagated at 37°C with shaking until OD_{600} reached 1.0. Protein expression was induced for 16 h following the addition of IPTG (0.5 mM final concentration). The cells were harvested by centrifugation (30 min, 3000 × *g*, 4°C). The cell pellet was resuspended in 20 mL of lysis buffer (50 mM Tris-Cl, 1 mM EDTA, pH 8) and was disrupted by sonicating over ice for 3 × 30 s with 1 min intervals. The protease inhibitors benzamidine and PMSF were added to a final concentration of 1 mM. Following centrifugation at 3000 × *g* for 30 min at 4°C, the supernatant was filtered over a 0.22 µm pore size filter and kept on ice.

4.6.4. CO binding

Heme proteins absorb light at 450 nm, corresponding to the CO stretch frequency when the Fe-heme is bound to CO. This Soret band reports on the amount of folded protein [74]. Supernatant from bacterial lysis (100 µl) was mixed with 900 µl of 100 mM potassium phosphate pH 8 and transferred to a 1 ml cuvette. The absorbance was measured at 450 and 490 nm before proteins were reduced by adding a few grains of sodium dithionite and the CO-complex formed by slow bubbling the solutions with CO gas. The absorbance was read again at 450 and 490 nm.

4.6.5. Kinetic characterization

Spectrophotometric enzyme activity assays were carried out by following NADPH consumption using a Varian Cary 100 Bio spectrophotometer. The assay buffer was 100 mM potassium phosphate, pH 8. Reaction mixtures contained 0.1–0.5 nM P450 BM3 and varying concentrations of substrates. Reactions were initiated by the addition of 250 mM NADPH, and

the decrease in absorbance at 340 nm was monitored (ε = 6210 M^{-1} cm^{-1}). Initial rates were obtained from the first 20 s of the reaction. To determine K_M, lauric acid concentration was varied from 0 – 1000 μM, keeping all other reagents constant at 200 μM NADPH and 10 nM BM3. The data were fitted to the Michaelis-Menten equation using non-linear regression analysis with the GraphPad 6 software. Protein concentration was determined with the BioRad assay.

4.6.6. Computational docking

Computational docking was performed using AutoDock Vina and the ligands, protonated at pH 7, were obtained from the ZINC database [48, 75]. The structural coordinates of the receptor were either a ligand-free form of native BM3 (PDB code 2HPD), a ligand-bound form (PDB code 1ZO9) or the predicted catalytic binding pose derived from free energy calculations. While the ligands were completely flexible, all protein residues were kept rigid during docking. The 20 highest-scoring poses of each docking run were kept. Poses where the aliphatic chain was predicted to be at the entrance of the tunnel 2 region while at the same time placing the charged carboxylate close to heme iron were removed. The volume occupied by atoms of ligands in at least 25% of the predicted poses was visualized with the VolMap tool of VMD 1.9.2 [76].

4.6.7. Free energy calculations

Ligand-free BM3 (PDB code 2HPD) was used as initial coordinates. The simulation was performed using NAMD 2.9 with the AMBER99SB-ILDN and TIP3P force field definitions for protein and water, respectively [77-79]. The ferric penta-coordinate high-spin (IC6) heme coordinates and atomic charges from Shahrokh *et al.* were used [80]. Following 1000 steps of conjugate gradient energy minimization, the ABF simulation was performed in an NPT ensemble under periodic boundary conditions. Long-range electrostatic interactions were calculated using the PME method. The electrostatic and van der Waals interactions were cut off after 16 Å.

Using the collective variable module of NAMD, the reaction coordinate for the 1D ABF experiment was set to the distance between the ω-1 carbon of the palmitic acid substrate

and the active-site heme iron along the z axis [25, 27]. The maximum distance was set to 35 Å and the minimum distance to 4 Å. For the 2D ABF experiment, the same reaction coordinate was applied and the second was the distance between the center of mass of the palmitic acid carboxylate and the guanidinium group of Arg47 along the z axis. The width of the local free energy bin was chosen to be 0.5 Å and the adaptive biasing force was applied after sampling each bin 500 times. The wall force constant for boundary potentials was 10 kcal mol^{-1} $Å^{-2}$. To generate the potential of mean force (PMF), the resulting data were integrated using abf_integrate script provided by NAMD.

4.7. Acknowledgements

Computational resources were provided by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants RGPIN 227853 (to JNP) and RGPIN 355789 (to GL). MCCJCE is the recipient of a Vanier Doctoral Scholarship. SD held a DAAD award and AAH held an internship from CGCC.

4.8. References

1. Giver, L., et al., *Directed evolution of a thermostable esterase.* Proc Natl Acad Sci U S A, 1998. **95**(22): p. 12809-12813.
2. Reetz, M.T., et al., *Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution.* Angew Chem Int Edit, 1997. **36**(24): p. 2830-2832.
3. Dror, A., et al., *Protein engineering by random mutagenesis and structure-guided consensus of Geobacillus stearothermophilus Lipase T6 for enhanced stability in methanol.* Appl Environ Microbiol, 2014. **80**(4): p. 1515-1527.
4. Chen, M.M., et al., *Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes.* Protein Eng Des Sel, 2012. **25**(4): p. 171-178.

5. Otten, L.G., F. Hollmann, and I.W. Arends, *Enzyme engineering for enantioselectivity: from trial-and-error to rational design?* Trends Biotechnol, 2010. **28**(1): p. 46-54.
6. Chica, R.A., N. Doucet, and J.N. Pelletier, *Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design.* Curr Opin Biotech, 2005. **16**(4): p. 378-384.
7. Reetz, M.T., et al., *Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods.* J Am Chem Soc, 2010. **132**(26): p. 9144-9152.
8. Savile, C.K., et al., *Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture.* Science, 2010. **329**(5989): p. 305-309.
9. Cheng, T., et al., *Structure-based virtual screening for drug discovery: a problem-centric review.* AAPS J, 2012. **14**(1): p. 133-141.
10. Roccatano, D., *Structure, dynamics, and function of the monooxygenase P450 BM-3: insights from computer simulations studies.* J Phys Condens Matter, 2015. **27**(27): p. 273102.
11. Pleiss, J., *Systematic Analysis of Large Enzyme Families: Identification of Specificity- and Selectivity-Determining Hotspots.* Chemcatchem, 2014. **6**(4): p. 944-950.
12. Huang, W.C., et al., *Control of the stereo-selectivity of styrene epoxidation by cytochrome P450 BM3 using structure-based mutagenesis.* Metallomics, 2011. **3**(4): p. 410-416.
13. Park, S.H., et al., *Engineering bacterial cytochrome P450 (P450) BM3 into a prototype with human P450 enzyme activity using indigo formation.* Drug Metab Dispos, 2010. **38**(5): p. 732-739.
14. Verma, R., U. Schwaneberg, and D. Roccatano, *Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering.* Comput Struct Biotechnol J, 2012. **2**: p. e201209008.
15. Wong, T.S., D. Roccatano, and U. Schwaneberg, *Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries.* Environ Microbiol, 2007. **9**(11): p. 2645-2659.
16. Grebner, C., et al., *Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design.* J Chem Inf Model, 2016. **56**(4): p. 774-787.

17. Gordon, S.E., et al., *Dynamic Modelling Reveals 'Hotspots' on the Pathway to Enzyme-Substrate Complex Formation.* PLoS Comput Biol, 2016. **12**(3).
18. Li, W.H., et al., *Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations.* Proteins, 2011. **79**(1): p. 271-281.
19. Klvana, M., et al., *Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored Using Classical and Random Acceleration Molecular Dynamics Simulations.* J Mol Biol, 2009. **392**(5): p. 1339-1356.
20. Kingsley, L.J. and M.A. Lill, *Including Ligand-Induced Protein Flexibility into Protein Tunnel Prediction.* J Comput Chem, 2014. **35**(24): p. 1748-1756.
21. Park, S., et al., *Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality.* J Chem Phys, 2003. **119**(6): p. 3559-3566.
22. Park, S. and K. Schulten, *Calculating potentials of mean force from steered molecular dynamics simulations.* J Chem Phys, 2004. **120**(13): p. 5946-5961.
23. Souaille, M. and B.t. Roux, *Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations.* Comput Phys Commun, 2001. **135**(1): p. 40-57.
24. Darve, E. and A. Pohorille, *Calculating free energies using average force.* J Chem Phys, 2001. **115**(20): p. 9169-9183.
25. Darve, E., D. Rodriguez-Gomez, and A. Pohorille, *Adaptive biasing force method for scalar and vector free energy calculations.* J Chem Phys, 2008. **128**(14): p. 144120.
26. Henin, J. and C. Chipot, *Overcoming free energy barriers using unconstrained molecular dynamics simulations.* J Chem Phys, 2004. **121**(7): p. 2904-2914.
27. Henin, J., et al., *Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables.* J Chem Theory Comput, 2010. **6**(1): p. 35-47.
28. Dehez, F., E. Pebay-Peyroula, and C. Chipot, *Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel.* J Am Chem Soc, 2008. **130**(38): p. 12725-12733.
29. Henin, J., et al., *Diffusion of glycerol through Escherichia coli aquaglyceroporin GlpF.* Biophys J, 2008. **94**(3): p. 832-839.

30. Lamoureux, G., M.L. Klein, and S. Berneche, *A stable water chain in the hydrophobic pore of the AmtB ammonium transporter.* Biophys J, 2007. **92**(9): p. L82-84.
31. Wang, S., et al., *Ammonium transporters achieve charge transfer by fragmenting their substrate.* J Am Chem Soc, 2012. **134**(25): p. 10419-10427.
32. Modi, N., et al., *Tuning the affinity of anion binding sites in porin channels with negatively charged residues: molecular details for OprP.* ACS Chem Biol, 2015. **10**(2): p. 441-451.
33. Bai, Q., et al., *Computational study on the interaction between CCR5 and HIV-1 entry inhibitor maraviroc: insight from accelerated molecular dynamics simulation and free energy calculation.* Phys Chem Chem Phys, 2014. **16**(44): p. 24332-24338.
34. Faller, C.E. and O. Guvench, *Terminal sialic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine sidechains.* Proteins, 2014. **82**(11): p. 3079-3089.
35. Ebert, M.C.C.J.C., et al., *Evolution of P450 Monooxygenases toward Formation of Transient Channels and Exclusion of Nonproductive Gases.* ACS Catal, 2016: p. 7426-7437.
36. Cojocaru, V., P.J. Winn, and R.C. Wade, *The ins and outs of cytochrome P450s.* Biochim Biophys Acta, 2007. **1770**(3): p. 390-401.
37. Damsten, M.C., et al., *Application of drug metabolising mutants of cytochrome P450 BM3 (CYP102A1) as biocatalysts for the generation of reactive metabolites.* Chemico-biological interactions, 2008. **171**(1): p. 96-107.
38. Di Nardo, G. and G. Gilardi, *Optimization of the bacterial cytochrome P450 BM3 system for the production of human drug metabolites.* Int J Mol Sci, 2012. **13**(12): p. 15901-15924.
39. van Vugt-Lussenburg, B.M., et al., *Heterotropic and homotropic cooperativity by a drug-metabolising mutant of cytochrome P450 BM3.* Biochem Biophys Res Commun, 2006. **346**(3): p. 810-818.
40. Whitehouse, C.J., S.G. Bell, and L.L. Wong, *P450(BM3) (CYP102A1): connecting the dots.* Chem Soc Rev, 2012. **41**(3): p. 1218-1260.
41. Yun, C.H., et al., *The bacterial P450 BM3: a prototype for a biocatalyst with human P450 activities.* Trends Biotechnol, 2007. **25**(7): p. 289-298.

42. Behrendorff, J.B.Y.H., W.L. Huang, and E.M.J. Gillam, *Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity.* Biochem J, 2015. **467**: p. 1-15.
43. Noble, M.A., et al., *Roles of key active-site residues in flavocytochrome P450 BM3.* Biochem J, 1999. **339**: p. 371-379.
44. Cowart, L.A., J.R. Falck, and J.H. Capdevila, *Structural determinants of active site binding affinity and metabolism by cytochrome P450 BM-3.* Arch Biochem Biophys, 2001. **387**(1): p. 117-124.
45. Zilly, F.E., et al., *Tuning a p450 enzyme for methane oxidation.* Angew Chem Int Ed Engl, 2011. **50**(12): p. 2720-2724.
46. Otey, C.R., et al., *Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450.* Biotechnol Bioeng, 2006. **93**(3): p. 494-499.
47. Dietrich, M., et al., *Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards gamma- and delta-positions.* J Biotechnol, 2009. **139**(1): p. 115-117.
48. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-461.
49. Kua, J., Y.K. Zhang, and J.A. McCammon, *Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach.* J Am Chem Soc, 2002. **124**(28): p. 8260-8267.
50. Kotev, M., R. Soliva, and M. Orozco, *Challenges of docking in large, flexible and promiscuous binding sites.* Bioorg Med Chem, 2016. **24**(20): p. 4961-4969.
51. Osuna, S., et al., *Molecular dynamics explorations of active site structure in designed and evolved enzymes.* Acc Chem Res, 2015. **48**(4): p. 1080-1089.
52. Li, W., et al., *Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations.* Proteins, 2011. **79**(1): p. 271-281.
53. Ludemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways.* J Mol Biol, 2000. **303**(5): p. 813-830.

54. Yeom, H., et al., *The role of Thr268 in oxygen activation of cytochrome P450BM-3.* Biochemistry, 1995. **34**(45): p. 14733-14740.
55. Haines, D.C., et al., *Pivotal role of water in the mechanism of P450BM-3.* Biochemistry, 2001. **40**(45): p. 13456-13465.
56. Haines, D.C., et al., *A single active-site mutation of P450BM-3 dramatically enhances substrate binding and rate of product formation.* Biochemistry, 2011. **50**(39): p. 8333-8341.
57. Whitehouse, C.J.C., et al., *Structure, electronic properties and catalytic behaviour of an activity-enhancing CYP102A1 (P450(BM3)) variant.* Dalton T, 2011. **40**(40): p. 10383-10396.
58. Huang, W.C., et al., *Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency.* J Mol Biol, 2007. **373**(3): p. 633-651.
59. Clark, J.P., et al., *The role of Thr268 and Phe393 in cytochrome P450 BM3.* J Inorg Biochem, 2006. **100**(5-6): p. 1075-1090.
60. Cong, Z., et al., *Activation of Wild-Type Cytochrome P450BM3 by the Next Generation of Decoy Molecules: Enhanced Hydroxylation of Gaseous Alkanes and Crystallographic Evidence.* ACS Catal, 2014. **5**(1): p. 150-156.
61. Li, H. and T.L. Poulos, *The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid.* Nat Struct Biol, 1997. **4**(2): p. 140-146.
62. Fasan, R., et al., *Evolutionary History of a Specialized P450 Propane Monooxygenase.* J Mol Biol, 2008. **383**(5): p. 1069-1080.
63. Nazor, J. and U. Schwaneberg, *Laboratory evolution of P450 BM-3 for mediated electron transfer.* Chembiochem, 2006. **7**(4): p. 638-644.
64. Nazor, J., et al., *Laboratory evolution of P450 BM3 for mediated electron transfer yielding an activity-improved and reductase-independent variant.* Protein Eng Des Sel, 2008. **21**(1): p. 29-35.
65. Lewis, D.F., E. Watson, and B.G. Lake, *Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics.* Mutat Res, 1998. **410**(3): p. 245-270.

66. Di Nardo, G., et al., *Wild-type CYP102A1 as a biocatalyst: turnover of drugs usually metabolised by human liver enzymes.* J Biol Inorg Chem, 2007. **12**(3): p. 313-323.
67. Damsten, M.C., et al., *Application of drug metabolising mutants of cytochrome P450 BM3 (CYP102A1) as biocatalysts for the generation of reactive metabolites.* Chem Biol Interact, 2008. **171**(1): p. 96-107.
68. Kim, D.H., et al., *Oxidation of human cytochrome P450 1A2 substrates by Bacillus megaterium cytochrome P450 BM3.* J Mol Catal B-Enzym, 2010. **63**(3-4): p. 179-187.
69. Isin, E.M. and F.P. Guengerich, *Substrate binding to cytochromes P450.* Anal Bioanal Chem, 2008. **392**(6): p. 1019-1030.
70. Conner, K.P., C.M. Woods, and W.M. Atkins, *Interactions of cytochrome P450s with their ligands.* Arch Biochem Biophys, 2011. **507**(1): p. 56-65.
71. Borrelli, K.W., et al., *PELE: Protein energy landscape exploration. A novel Monte Carlo based technique.* J Chem Theory Comput, 2005. **1**(6): p. 1304-1311.
72. Barnes, H.J., M.P. Arlotto, and M.R. Waterman, *Expression and Enzymatic-Activity of Recombinant Cytochrome-P450 17-Alpha-Hydroxylase in Escherichia-Coli.* P Natl Acad Sci USA, 1991. **88**(13): p. 5597-5601.
73. Sarkar, G. and S.S. Sommer, *The Megaprimer Method of Site-Directed Mutagenesis.* Biotechniques, 1990. **8**(4): p. 404-407.
74. Omura, T. and R. Sato, *The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature.* J Biol Chem, 1964. **239**: p. 2370-2378.
75. Irwin, J.J. and B.K. Shoichet, *ZINC - A free database of commercially available compounds for virtual screening.* J Chem Inf Model, 2005. **45**(1): p. 177-182.
76. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-38, 27-38.
77. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-1958.
78. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* J Chem Phys, 1983. **79**(2): p. 926-935.
79. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-1802.

80. Shahrokh, K., et al., *Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle.* J Comput Chem, 2012. **33**(2): p. 119-133.

Chapter 5 - Pushing the boundaries of NMR: Studying protein dynamics in concert with molecular dynamics

5.1. Context

As discussed in Chapter 1, the impact of protein dynamics on catalysis is still a heated discussion in the scientific community. I became interested in this topic while discussing intrinsic limitations of NMR with respect to covering the entire timescale of protein dynamics with my colleague Dr. Sophie Gobeil. She studied protein dynamics during her doctoral work using the model system of β-lactamases [1-4]. Two research questions originated from those discussions: first, whether protein dynamics are conserved throughout evolution and this is true of different timescales of motions; and second, whether we can generate an atomistic conformational ensemble for β-lactamases that could provide answers where experimental and crystallography datasets are incomplete or inconclusive. Earlier studies have shown great success in the generation of such ensembles of protein conformations that represent at the same time the native structure and its associated dynamics [5-7].

While reliable protein dynamics data for the μs-ms timescale for proteins systems of moderate size such as β-lactamases is most readily accessible through NMR CPMG experiments, the entire fs-μs timescale can be easily covered by MD simulations. Using existing NMR data on the ps-ns timescale as reference, we evaluated the quality of our MD simulations. In agreement with other publications, we obtained an excellent correlation between NMR observables and values derived from MD simulation (see article in Annex 4 and Figure A 5-1 and Table A 5-3 and Table A 5-5). Motivated by this result, we extended the simulated trajectory to the μs timescale. That study, entitled '*Conserved Fast Dynamics Accompanied by Altered Slow Conformational Dynamics in Evolutionarily Related β-Lactamases*', is currently under review in *ACS Chemical Biology* and addresses our first question. The "pre-print" version of the article is provided in Annex 4. I am the second author of this article for which I contributed all computational experiments, data analysis, creation of figures and drafting the paragraphs related to that work.

To address the second question, we revisited the trajectories from that study to generate an atomistic conformational ensemble of all proteins. Dr. Gobeil had obtained several crystal structures of the same protein, under the same crystallization conditions. Intriguingly, an unexpected discrepancy among the crystal structures of the residue side-chain orientations was observed. There was no experimental explanation for this observation. Therefore, we combined the strengths of computational and experimental methodologies to better understand the system of interest. The article presented in this chapter, entitled '*Crystallographic evidence for dynamic conformational exchange in β-lactamases*' is a "pre-print" version of a manuscript currently in preparation. Dr. Sophie Gobeil and I share the first authorship on this work. We both contributed to conceptualization, data analysis and writing. While Dr. Gobeil's focus was on the crystallography and knowledge of β-lactamases, I contributed the computational data analysis. The trajectories were partially acquired in Prof. Jürgen Pleiss's laboratory and crystallography was performed in collaboration with Dr. Jaeok Park and Prof. Albert Berghuis. The work was done in Prof. Joelle Pelletier's laboratory. The manuscript was drafted mainly by Dr. Gobeil and myself with assistance from Prof. Pelletier and revisions by all authors.

Crystallographic evidence for dynamic conformational exchange in β-lactamases.

Maximillian Ebert[1,2,†], Sophie Gobeil[1,2,†], Jaeok Park[2,3,4], Albert Berghuis[2,3,4], Jürgen Pleiss[5] and Joelle N. Pelletier[1,2].

[1] Département de biochimie and 6Département de chimie, Université de Montréal, Montréal, QC, Canada

[2] PROTEO, the Québec Network for Research on Protein Structure, Function and Engineering

[3] Department of Biochemistry, McGill University, Montréal, QC, Canada

[4] GRASP, Groupe de Recherche Axé sur la Structure des Protéines

[5] Institute of Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

[†] M.E and S.G contributed equally to the work

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

5.2. Introduction

β-Lactamase production is the predominant mechanism, among many used by Gram-negative bacteria, conferring resistance toward the most clinically prescribed class of antimicrobial agents, the β-lactams [8, 9]. More than 1800 unique and naturally occurring β-lactamases, possessing different selectivity toward the recognized β-lactams, have to this day been reported [10]. The class A is the most prevalent of the four subgroups of β-lactamases. Its more than 550 members incorporate single or few amino acid variations relative to the native enzyme, resulting in hundreds of Extended-Spectrum β-Lactamases (ESBL) and Inhibitor Resistant TEM-like (IRT) β-lactamase variants that confer broader recognition of β-lactam substrates and inhibitors [11, 12].

Because slight sequence alterations suffice to significantly modulate the recognition pattern of β-lactamases, understanding the subtle effects produced by mutations has been challenging. Indeed, while over 50 crystal structures of TEM-1 β-lactamases have been reported, understanding their differences has been stymied by the fact that all structures are (virtually indistinguishable save for the mutations), and that half of the structures include a bound inhibitor. While it appears clear that modulation of active site volume alters ligand recognition [13], the exact mode of substrate binding remains unclear. Class A β-lactamases are constituted of an all-α and an α/β domain with the active site cavity located at their interface (Figure 5-1). Rearrangements affecting the active-site walls as well as shifting water molecules have been reported [14-20]. Among these, a triple mutant in the Ω-loop of TEM-1 procuring altered substrate specificity and increased activity toward a bulky cephalosporin substrate displayed important structural rearrangements of the active-site side-chains and water molecules [14]. Furthermore, residual activity has been observed in a TEM-1 variant where the catalytic nucleophile, Ser70, was mutated to the non-nucleophilic Gly [15]. Residual activity was explained by the introduction of a new water molecule in the crystal structure, substituting for the missing nucleophilic hydroxyl group.

The question we address here is whether these crystal structures suffice to understand how the sequence variation translates into functional alterations, or whether additional description of dynamic populations may be required. It is thought that the consensus, or lowest energy conformation, is generally frozen in the crystal state of the protein [21]. Although X-

ray crystallography [20, 22-27] studies have been occasionally complemented with NMR [28-31] and computational studies [32-41], all those reports have considered the proteins in crystal-like conformations. Indeed, previous work using NMR and *in silico* methodologies performed by us and others has shown that the naturally evolved Class A β-lactamases TEM-1 and PSE-4 generally maintain a rigid backbone over the entire ps to ms timescale regime [2, 3, 28-30, 42], which seemingly justifies direct interpretation of crystal structure information.

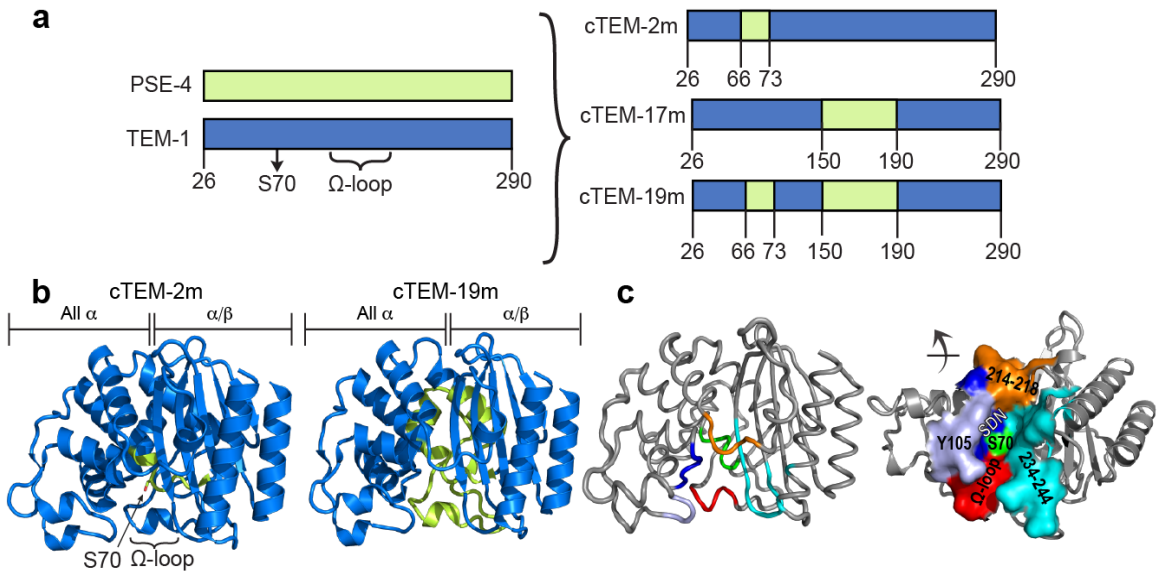


Figure 5-1 Relation of chimeras to the parental class A β-lactamases TEM-1 and PSE-4. *A) Numbering of the sequence blocks originating from TEM-1 (blue) and PSE-4 (green) in cTEM-2m, cTEM-17m and cTEM-19m. Numbering according to Ambler [43]. The nomenclature of the chimeras is as follows: 'c' indicates chimera, followed by the number of substitutions (or mutations: 'm') in each chimera relative to the TEM-1 parental sequence, which shares 40% sequence identity with PSE-4. The catalytic nucleophile, Ser70, and the catalytically relevant Ω-loop, are indicated in the TEM-1 sequence. B) Structural representation of the parental origin in chimeras cTEM-2m (PDB 4MEZ) and cTEM-19m (PDB 4QY5), colored as in panel A. The catalytic nucleophile, Ser70 (in sticks) and the catalytically relevant Ω-loop are indicated on the cTEM-2m structure to highlight the hybrid composition of the active-site region located at the interface of the all-α domain (left half) and the α/β domain (right half). C) The active-site walls, set in TEM-1 (PDB 1XPB). Green: S70 wall (Met69-Lys73); lilac, Y105 wall (Val 103-Ser106); dark blue, SDN wall (Met129-Asn132); red, Ω-loop wall (Glu166-Asn170); orange, 214-218 wall (Val214-Gly218); and cyan, 234-244 wall (Lys234-Arg244 + Met272 and Asn276). (Right) The structure is reoriented as indicated by the arrow and the active-site walls are represented by their solvent-accessible surface.*

Nonetheless, proteins experience a wide range of motions occurring over multiple timescales, both on the backbone and the side-chains, and those motions may underlie function [44, 45]. We recently examined a model system that mimics the natural process of homologous recombination and thus simulates potential evolutionary recombinants of TEM-1 and PSE-4 β-lactamases, [46, 47]. The resulting chimeric β-lactamases incorporate one or more segments of the PSE-4 active site into the TEM-1 framework and are functional (Figure 5-1 A,B and Annex 4).Their backbone motions over the ps to μs timescale were maintained as in the native TEM-1 and PSE-4. However, striking modifications on the μs to ms timescale of backbone motions were observed.

In the process of elucidating the potential structural and functional implication of the altered dynamics in the chimeric β-lactamases, we attempted to crystalize the chimera cTEM-19m (Figure 5-1 A) in different catalytic states. We obtained two new crystal structures with strikingly different side-chain conformations in the immediate active-site environment. The crystallization conditions were almost identical, which raises the question as to whether the analysis of a single crystal structure of a variant in this particular system is sufficient to fully comprehend and interpret the functional results.

We thus hypothesized that crystal structures do not always represent the consensus conformation, but rather may be subject to crystallographic biases. Using μs-long MD trajectories, we used the crystal structures as starting points to generate a large conformational ensemble, allowing the determination of a conformational consensus. This ensemble better reconciles the geometric and dynamic implications observed by kinetic data than does the set of crystal structures we have obtained. By these means, we demonstrate that computational methodologies constitute a powerful tool to extract the most information from crystallographic data.

5.3. Results and discussion

5.3.1. The active-site environment of class A β-lactamases

The chimeric β-lactamase cTEM-2m results from inserting the S70 active-site wall from PSE-4 into the TEM-1 framework. That active-site wall constitutes one of the two

connectors between the all-α and the α/β-domains (Figure 5-1). This recombination resulted in only two substitutions to the TEM-1 sequence, Met68Leu and Met69Thr, which are the immediate neighbours of the conserved catalytic nucleophile, Ser70. Chimera cTEM-17m incorporates residues 150 to 190 from PSE-4. This introduced 17 substitutions relative to the TEM-1 sequence in the Ω-loop active-site wall and its adjacent helices. This active-site wall contains the catalytically relevant Glu166, proposed to serve as a general base in the catalytic acylation and deacylation steps [48] (Figure A 3-1). Finally, chimera cTEM-19m combines both of these exchanged regions, resulting in 19 substitutions and a more highly blended active-site cavity.

The active-site walls of Class A β-lactamases are conserved, and serve to orient the substrate and promote catalytic activity. These walls are: the S70 wall (residues $S^{70}xxK^{73}$), the SDN wall (residues $S^{130}D^{131}N^{132}$), the Ω-loop wall (residues $E^{166}xxL^{169}N^{170}$), the 214-218 wall (residues 214-218) and the 234-244 wall (residues $K^{234}T^{235}G^{236}$) (Figure 5-1) [29]. Residues within those walls that are not directly implicated in the catalytic mechanism tolerate sequence variations (Figure 5-1 C).

The two-step acylation-deacylation catalytic mechanism is performed by five residues located on four different active-site walls, and a water molecule (Figure A 3-1). These are Lys73, proposed to act as a general base to activate the catalytic nucleophile Ser70 [49] [33], Ser130, the proton donor in the β-lactam ring opening [50, 51], Glu166, proposed to act as the general base during the acylation and deacylation steps [27, 51], Arg234, important for the carbenicillinase specificity of PSE-4 [52] and the deacylating water molecule coordinated by Glu166 and Asn170 [51, 53, 54]. In addition, the Y105 wall appears to be implicated in a gate-keeping mechanism using Tyr105 side chain [29, 55].

5.3.2. Analysis of the backbone conformations in a crystallographic environment

Crystal formation of cTEM-19m is rapid (within hours). Because the turnover rate of β-lactamases is high (k_{cat} for benzylpenicillin of TEM-1 and cTEM-19m is 450 s^{-1} and 230 s^{-1}, respectively [3]), it is generally not possible to maintain the enzyme in the presence of substrate for biophysical measurements made over days; here, fast crystallization led us to

propose that inclusion of a substrate during nucleation events could result in alternative crystal conformations, particularly in the active-site area. Among many conditions tested, the presence of the native substrate benzylpenicillin gave rise to two new high-resolution structures (PDB 4QY5, 1.50 Å resolution and PDB 4QY6, 1.15 Å resolution) (Table A 3-1).

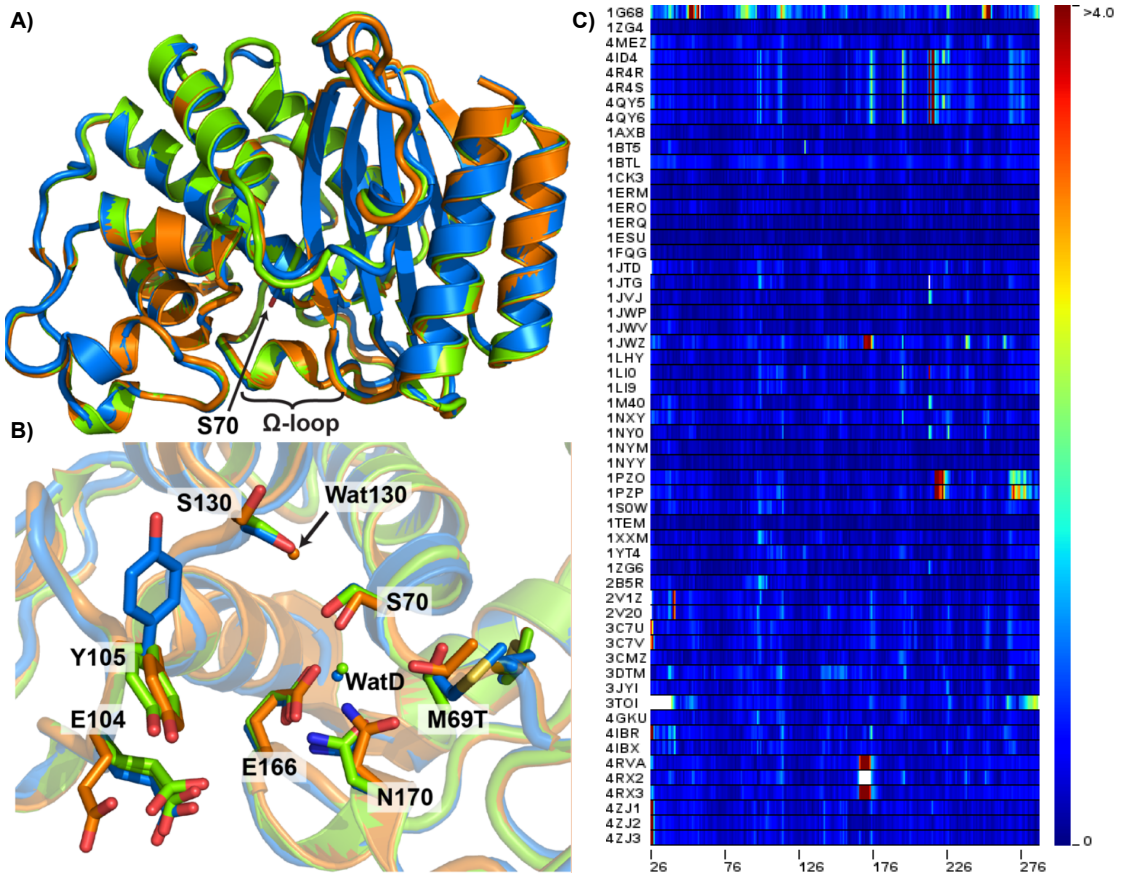


Figure 5-2 Crystal structures of cTEM-19m and Cα-RMSD to TEM-1 (PDB 1XPB) of all TEM-1 and PSE-4 variants available in the PDB.

A) Backbone overlay of the crystal structures of TEM-1 (blue, PDB 1XPB) and the chimera cTEM-19m (green; PDB 4QY6 and orange PDB 4QY5 B) Active-site view of the overlay of cTEM-19m with TEM-1. Wat130 is the new water molecule in cTEM-19m at the position of Ser130 hydroxyl moiety; WatD is the conserved deacylating water molecule: Wat292 in TEM-1 PDB 1XPB and Wat439 in cTEM-19m PDB 4QY6, shown here, and Wat1046 in TEM-1 PDB 1ZG4, Wat411 in PSE-4 PDB 1G68, Wat454 in cTEM-2m PDB 4MEZ, Wat427 in cTEM-17m PDB 4ID4; Wat430 in cTEM-19m PDB 4R4R and Wat411 in cTEM-19m PDB 4R4S. C) Cα-RMSD to TEM-1 (PDB 1XPB) of PSE-4 (1G68) and 55 variants of TEM-1 available in the PDB based on Uniprot entry P62593. The RMSD values of each residue are shown from 0 Å, in deep blue, to 4 Å or above, in red.

Among many conditions tested, the presence of the native substrate benzylpenicillin gave rise to two new high-resolution structures (PDB 4QY5, 1.50 Å resolution and PDB 4QY6, 1.15 Å resolution) (Table A 3-1). Similar to other TEM-1 variants the backbone is virtually identical compared to wild-type (WT) (Figure 5-2 A). Although no electron density consistent with the presence of benzylpenicillin was detected, significant structural variations were observed for active site side chain Glu104, Ser130, Asn170 and Thr69 in one of these models (PDB 4QY5) relative to the three other structures obtained for chimera cTEM-19m (4QY6 and the previously reported PDB 4R4R and 4R4S (Annex 4)) (Figure 5-2 B). The structural differences were analyzed on the basis of the Cα-RMSD of TEM-1 (PDB 1XPB) to all TEM-1 and PSE-4 variants reported in the PDB. This revealed that the most extensively recombined chimeras, *i.e.* cTEM-17m and cTEM-19m, contrast with the other structures as a result of their elevated Cα-RMSD (> 1.2 Å) in three regions: 174-176, 196-197 and 214-227 (Figure 5-2 C). While Gly196/Glu197 are surface-exposed residues of no known catalytic relevance, residues 174-176 form the tip of the catalytically important Ω-loop while residues 214-227 constitute the 214-218 active-site wall and part of the second domain connector (Figure 5-1). Considering that only region 174-176 was modified by recombination, this reveals a structural relation between the Ω-loop and the two other altered regions. We note that the two active-site regions are also those that diverge the most among the TEM-1 PDB structures; the altered Cα-RMSD results either from presence of a core-disrupting allosteric inhibitor (1PZO, 1PZP – affected in the 218-224 region), a boronate analog (1JWZ – affected in the 174-176 region) or mutations in the Ω-loop region (4RVA and 4RX3) that introduce large conformational changes altering the substrate profile.

Based on the Cα-RMSD profile of region 214-227, the structures of the chimeras were further differentiated. Similar to the structure of cTEM-17m PDB 4ID4, the new crystal structure of cTEM-19m 4QY5 shows an elevated Cα-RMSD over the entire 214-227 region, whereas the three other cTEM-19m structures, PDB 4R4R, 4R4S and 4QY6, differ from TEM-1 only in the 214-220 stretch.

We analyzed the b-factors of the 4QY5 Cα-atoms to rule out experimental uncertainties regarding backbone atoms being at the root of the elevated Cα-RMSD in the 220-227 segment. B-factors are a measure to describe fluctuations of atoms around their

experimentally determined coordinates. They hence provide information on the dynamic nature of the protein in general, however without any notion of time-scale [56]. That crystal structure displays elevated b-factors across all residues compared to the other structures of cTEM-19m (Figure 5-3 A, red line) yet shows a basal b-factor value similar to TEM-1 (PDB 1ZG4) and PSE-4 (PDB 1G68) (Figure 5-3 B), consistent with its basal b-factor value being within error.

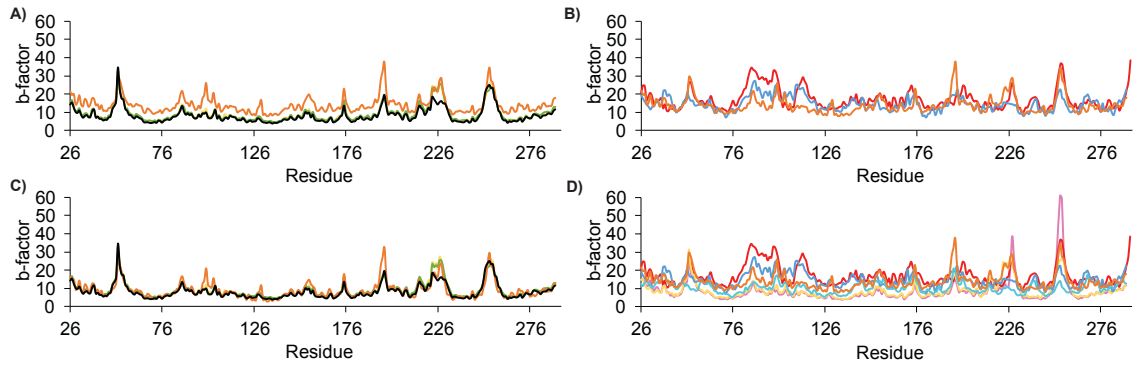


Figure 5-3. Cα b-factors for PSE-4, TEM-1 and chimeras.
Cα b-factors for PSE-4, TEM-1 and chimeras extracted from their respective PDB structures. A) The four cTEM-19m crystal structures : 4R4R – yellow, 4R4S – black, 4QY5 – orange and 4QY6 green B) Comparison of 4QY5 in orange with PSE-4 (1G68) in red and TEM-1 1XPB in blue. The b-factors of 4QY5 do not show higher dynamics than WT PSE-4 or TEM-1 C) Superposition of the b-factor profile of 4QY5 on the profile of the other three structures of cTEM-19m. 4R4R – yellow, 4R4S – black, 4QY5 – orange and 4QY6 green D) b-factor profile of TEM-1, PSE-4 and all chimeras not superimposed including cTEM-17m – pink.

After superimposing the b-factor profile of all four cTEM-19m crystals, only surface-exposed residues Asn100, Gly156 and Glu197, and active site residues Glu104 and Ser130 exhibited higher b-factors in 4QY5 (Figure 5-3 C). We note an apparent lower b-factor profile for cTEM-17m and cTEM-19m compared to their 'parents', TEM-1 and PSE-4 (Figure 5-3 D). Thus, the significantly increased slow protein dynamics we have reported for cTEM-17m and cTEM-19m relative to the parental enzymes are not reflected in the b-factors, despite the fact that crystallographic b-factors have been used to report on protein dynamics [45, 57, 58]. Indeed, we find no correlation between protein dynamics from NMR or MD studies and b-factors among the proteins under study.

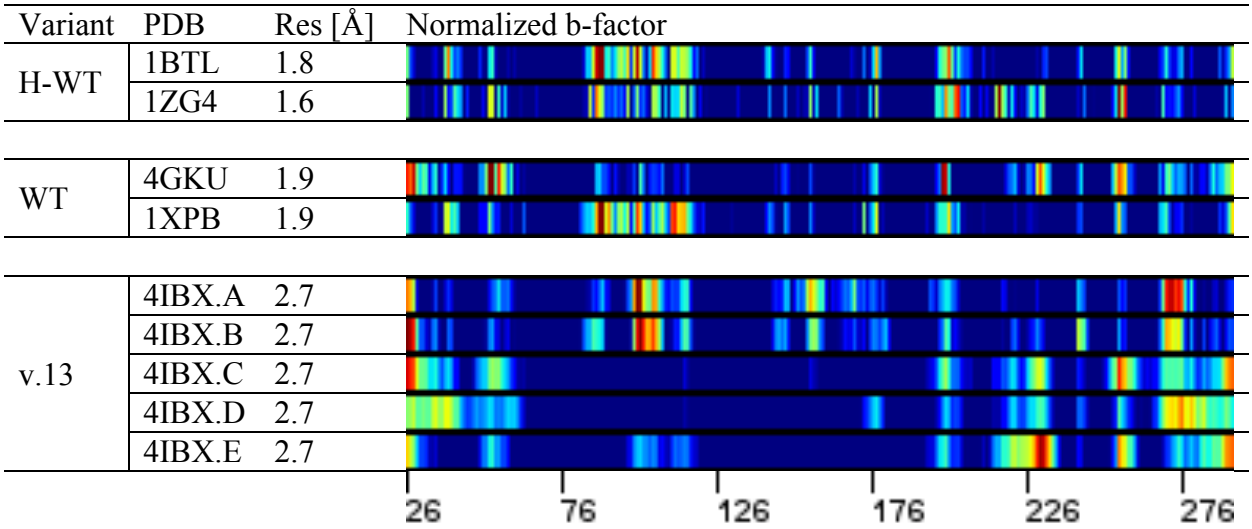


Figure 5-4. Normalized b-Factors of selected TEM-1 β-lactamases.
Normalized b-Factors of the ligand free crystal structures of the historical TEM-1 WT (Val84Ile/Ala184Val), WT and all individual chains of stabilized TEM-1 variant v.13. Normalized b-factors are represented in a gradient from low to high in blue to red.

Comparison of the normalized b-factors for crystal structures of TEM-1 revealed specific instances of striking deviations: see Figure 5-4. Of particular interest are crystal structures historically named WT (H-WT) yet carrying the double mutation I84V/V184A to eliminate restriction sites: while regions with elevated b-factor remains the same (residues 80-120 and 180-225), the amplitude and individual residues with elevated b-factors differ significantly among the two H-WT crystals shown in Figure 5-4. The same is true for the two ligand free WT crystal structures 1XPB and 4GKU. Perhaps most telling is the variation in b-factors among different chains of a single unit cell of 4IBX a stabilized TEM-1 variant, for example (Figure 5-4). Finally, there is only a weak correlation between the crystal structure resolution and the average b-factor ($R^2 = 0.34$) or the maximum b-factor ($R^2 = 0.45$) (Figure A 3-2 A and B), and there is a no significant correlation between the resolution and the minimum b-factor ($R^2 = 0.15$) or the standard deviation on the b-factor ($R^2 = 0.19$) (Figure A 3-2 C and D). These observations lead us to conclude that direct comparison of crystallographic b-factors of β-lactamase variants cannot directly inform on the relative dynamism of residues.

We further assessed the quality of the 4QY5 crystal by comparing the number alternate states for each residue in that structure to all structures reported in the PDB for TEM-1, PSE-4 and variants (Figure A 3-3). As expected, we find a strong correlation between the resolution

and the number of alternate states *i.e.* better resolution results in a higher number of alternate states. It has been suggested that dynamic information is hidden in the additional electron density used to define alternate states [59-61]. However, no correlation was found between the number of dynamic residues previously observed on the ps-ms timescale (Annex 4) and the number of crystallographic alternate conformations of side-chains for the β-lactamase system, but rather between the crystal resolution and the number of alternative states (Figure A 3-3). The three crystal structures of cTEM-19m 4R4R, 4R4S and 4QY5 with a resolution of ~1.2 Å show about 30 alternative states, while 4QY6 at the lower resolution of 1.5 Å has only 20. The highly dynamic chimeras cTEM-17m and cTEM-19m also have less alternative states than the H-WT structure 1M40, which was determined at a resolution of 0.85 Å and has 175 alternative states (Figure A 3-3). We previously demonstrated the absence of strong protein dynamics for TEM-1 WT at the μs-ms timescale (Annex 4). In addition, many structures of the analyzed β-lactamases have no alternative states in their crystal structures independent from their resolution.

5.3.3. Backbone structural ensembles derived from long molecular dynamic simulations

We could not find any rationale for the observed differences in Cα-RMSD profile of 4QY5 in the 214-227 region based solely on experimental error. Furthermore, we could not rely on static crystal structures to rule out any dynamic cause for divergence in profile among the cTEM-19m crystals. To verify whether the newly observed conformations are an essential part of the structural ensemble of the respective proteins or if we are observing crystallographic artefacts, we revisited our previously published unconstrained MD simulations of TEM-1, PSE-4 and the three chimeras cTEM-2m, cTEM-17m and cTEM-19m β-lactamases (Annex 4).

In the relevant crystal structures, only region 214-218 showed significant variation of the φ and ψ angles, coherent with the Cα-RMSD profiles described above (Figure 5-4 and Figure A 3-4). The φ and ψ angles were extracted over the course of the entire trajectories for every residue, thus revealing the ensemble of possible backbone conformations within 6 μs; an example for selected residues is shown in Figure A 3-4. Using a clustering algorithm, we

analyzed the conformation of region 214-218 over the course of the trajectories. In PSE-4, the 214-218 active-site wall remained in its crystalized conformation, while TEM-1 showed a significant shift from the crystallographic orientation (Figure A 3-5). A similar shift was observed for chimera cTEM-2m, with evidence for more degrees of conformational freedom. This is coherent with the previously reported S^2 values and the RMSF calculated from the MD trajectories, where this active-site wall was rigid in PSE-4 and moderately mobile in the other proteins. Strikingly, in the chimeras cTEM-17m and cTEM-19m the loop also explored a conformation similar to that crystallized, yet with a larger degree of freedom resulting in greater structural diversity than for the other proteins (Figure A 3-5). It also correlates with the failure to assign this loop in these chimeras using NMR methodologies (Annex 4) [2]. Despite a large conformational diversity found in this loop, the orientation observed in this region in TEM-1, PSE-4 and cTEM-2m was not found in the ensemble of cTEM-17m and cTEM-19m. Since the amino acid sequence of this loop is identical in all proteins, long distance interactions with the exchanged Ω-loop region must be responsible for the new orientation.

We also verified whether the simulations reflected the elevated Cα-RMSD of 4QY5 in the 222-227 segment. All the β-lactamases displayed a strong fluctuation in the simulated ensemble for this region. It appears that this loop is not highly structured and can easily be disturbed as a result of small energy barriers between those conformations.

There was no difference in the φ and ψ angles between the crystallized conformation and the simulated ensemble for the active-site walls S70, SDN, Ω-loop and 234-244, with the significant exception of residue 69. In all crystal structures of class A β-lactamases – including the chimeras – the residue at position 69 adopts a high-energy conformation outside of the favoured Ramachandran areas [22, 62, 63]. However, after 5 ns of simulation time, cTEM-19m Thr69 adopted and remained in a favoured, low-energy conformation (Figure 5-5 A and B), resulting in an extensive water influx behind that active-site wall.

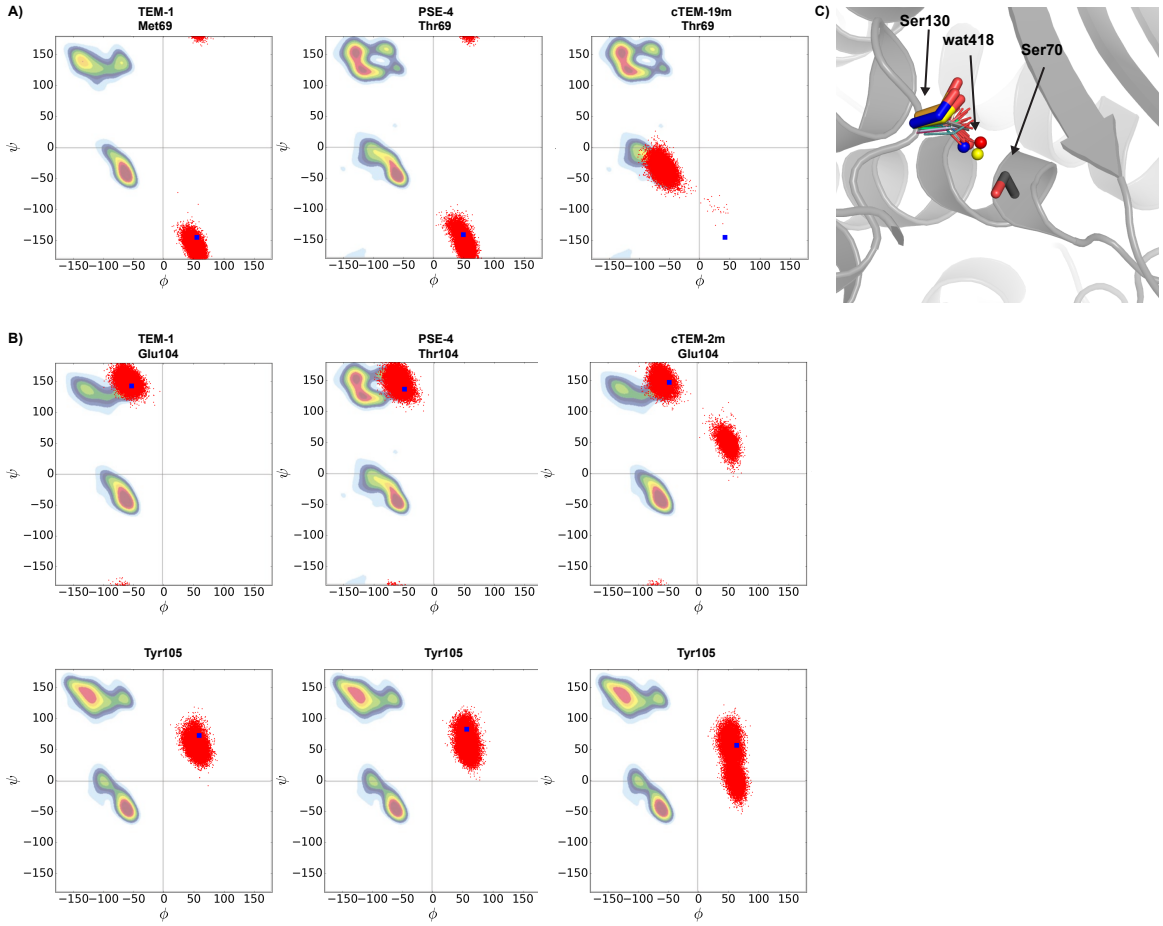


Figure 5-5 Ramachandran plot and side chain orientation of TEM-1, PSE-4 and the chimeras cTEM-19m or cTEM-2m for selected residues.

A) Met/Thr69 and B). Glu/Thr104 and Tyr105 over the 6μs of simulation. The orientation observed in the crystal structures of TEM-1 (PDB 1XPB), PSE-4 (PDB 1G68), cTEM-19m (PDB 4R4R) and cTEM-2m (PDB 4MEZ) are indicated by a blue square. The optimal conformation per residue type are indicated according to [64]. C) Superposition of all structures of TEM-1, PSE-4 and variants found in the PDB. Residue Ser130 is shown in sticks for cTEM-19m (4QY5; orange), TEM-32 (1LI0; blue) and TEM-1 mutant Ser70Gly (1ZG6; yellow). The newly introduced water molecules are shown as spheres in their respective color, with the addition of the water molecule found in TEM-1 Ser130Gly mutant (1YT4; red). All other structures are shown in a line representation. The cartoon representation is based on TEM-1 (1BTL).

Two further significant differences were observed, both in the active-site wall Y105, gate-keeper to the active-site cavity. Here, Glu104 and Tyr105 explored additional high energy conformations only in the cTEM-2m ensemble (Figure 5-5 A and B) [55, 65]. This suggests motions specific to this chimera, therefore resulting from the Met68Leu and

Met69Thr substitution and compensated for in chimera cTEM-19m by its 17 further substitutions. Despite the potential for this difference to affect gate-keeping, there was no change in productive binding as reflected by unaltered K_M (Annex 4). The higher mobility is also consistent with the above-determined elevated RMSF of this wall in cTEM-2m (0.14 relative to 0.08-0.10 in the native proteins).

5.3.4. A crystal structure showing rearrangement of catalytically important side-chains

Within the active-site region, the crystal structure main-chain conformations were in general agreement with the findings from the conformational ensemble, consistent with previous NMR studies [3, 28, 30, 42]. However, the active-site side-chains and water molecules of cTEM-19m PDB 4QY5 revealed striking differences relative to the three other crystal structures of cTEM-19m, while its electron density map showed no residual evidence of side-chain conformations observed in all other structures.

On the SDN active-site wall, the Ser130 hydroxyl moiety, which is the proton donor during the β-lactam ring opening, was rotated by 95° around its Cβ, displacing the hydroxyl moiety 3 Å farther from the active-site center (Figure 5-2 B). This corresponds to the adoption of the ['g-'] conformer (nomenclature based on the Dynameomics rotamer library [66]) rather than the ['t'] orientation found in TEM-1, in the chimeras cTEM-2m, cTEM-17m and the three other structures of cTEM-19m; in PSE-4, both ['t'] and ['g-'] orientations were observed with equal occupancy. Interestingly, this rearrangement of Ser130 has been suggested, in concert with Arg234, to be essential for the carbenicillinase activity of PSE-4, to prevent steric clashes [24]. While being able to adopt the ['g-'] conformation, cTEM-19m harbors a Lys at position 234. It was previously shown that cTEM-19m has no increased activity towards carbenicillin, the absence of Arg234 is further supporting the synergetic effect of the Ser130/Arg234 residue pair on the carbenicillinase activity (Annex 4). Displacement of the hydroxyl moiety in the ['g-'] conformer in cTEM-19m introduced a new water molecule (PDB 4QY5: wat418) (Figure 5-2 B). Interestingly, that water molecule has been observed only in the structures of the TEM-1 mutants Ser70Gly (PDB 1ZG6), TEM-32 (TEM-1 Met69Ile/Met182Thr; PDB 1LI0) and TEM-76 (TEM-1 Ser130Gly; PDB 1YT4) (Figure 5-5 C). The TEM-1 Ser70Gly variant,

depleted of a catalytic nucleophile, displays residual activity toward ampicillin (10^5-fold reduction of efficiency) [15]. The new water molecule was proposed to substitute for the missing Ser70 hydroxyl, procuring the residual activity. Interestingly, TEM-32 and TEM76 are inhibitor-resistant variants. In both, distortion of Ser70 or Lys234 displaces Ser130 and introduces the new water molecule [18, 19]. Because cTEM-19m is fully catalytic and the orientation of Ser70 and Lys234 are maintained as in the native proteins, we propose that the observation of this new water molecule results from the dynamism of both Ser70 and Lys234 on a timescale slower than the millisecond [1].

An advantage of examining the simulated ensemble of side-chain orientations rather than the single crystallographic snapshot is that the distribution of rotameric states can be observed over the course of the MD trajectories. Comparison of the rotameric states to the respective crystal structure and expected rotameric states (Table A 3-2) [66] shows that Ser130 converged towards a ratio of approximately 80% ['g-'] and 20% ['t'] in all the proteins examined. Thus, although the ['g-'] conformation was observed only in PSE-4 (1G68) and the 4QY5 structure of cTEM-19m, that conformation can be attained by all proteins and appears to be energetically-favored. This suggests a dynamic exchange between the crystallized ['t'] rotamer and ['g-'] rotamer. This further highlights how any single crystal structure should be interpreted with caution. Indeed, the general assumption that side-chain conformations fixed in crystal structures are always the lowest-energy state can prove incorrect.

On the S70 active-site wall, the catalytic nucleophile Ser70 hydroxyl moiety in 4QY5, compared to the other structures of cTEM-19m, was rotated by 35° around its Cβ and displaced outward from the active-site center by approximatively 1 Å while maintaining the same rotameric conformer. For TEM-1 and the chimeras cTEM-2m and cTEM-17m, the structural ensemble showed that the catalytic nucleophile Ser70 remained in the crystalized ['g-'] conformation. However, in PSE-4 and cTEM-19m, this residue adopted a near-equivalent ratio of the ['g-'] and ['g+'] conformations. This suggests that the catalytic nucleophile is more mobile in these two proteins. We note that in all proteins under study, Ser70 is unassigned by NMR methodologies, consistent with slow dynamism. The neighboring Thr69 side-chain was also rotated by 95° around its Cβ, displacing the side-chain hydroxyl moiety by 1.9 Å (Figure 5-2 B). This led to the adoption of the ['t'] conformer while

the ['g+'] was observed in all other structures containing Thr69: cTEM-19m (apart from 4QY5), PSE-4 and cTEM-2m. Interestingly, the structural ensemble did not reveal the ['t'] conformer for cTEM-19m, which remained in the crystallized ['g+'] conformation over the course of the simulation. It appears that the ['t'] rotameric conformation is a rare event that occurs on a time-scale not covered in the simulation. The slow dynamics observed for Ser70 might influence the dynamic exchange between the two rotamers of Thr69.

Nearby, the Lys73 side-chain of the same active-site wall remained in the crystallized conformation (['g-', 't', 't', 't']) at least 70% of the time for all proteins examined. The sole exception was cTEM-19m, where this residue was shown to be strongly dynamic by NMR (Annex 4). The structural ensemble supports this finding: a second conformer, ['g-', 't', 't', 'g+'], was observed 57% of the time in cTEM-19m.

On the Ω-loop active-site wall, Asn170, coordinating the deacylating water molecule, was rotated by 30° around its Cβ in cTEM-19m PDB 4QY5. This displaced the side-chain nitrogen and carbonyl by 2.1 and 1.5 Å, respectively, toward the exterior of the active site while remaining in the (['g-', 'Og-']) rotameric conformation. Surprisingly, the deacylating water molecule was not observed in cTEM-19m 4QY5 (Figure 5-2 B). This may reflect a less efficient coordination resulting from the increased dynamism of cTEM-19m relative to the native enzymes. This might also be reflected by the important decrease of k_{cat} for cephalosporin hydrolysis (9 to 75-fold) observed for this chimera (Annex 4). Accordingly, motions were previously observed in this chimera on the ns-ms timescales on an extended stretch of residues including the catalytically-relevant Glu166 and Asn170. Although the ensemble showed the ['g-', 'Og-'] rotamer at least 60% of the time for all proteins, cTEM-19m held it only 6% of the time while the ['g-', 'Ng-'] rotamer was found 50% of the time. This new rotamer cannot stabilize the water network with Glu166 and Asn170, which may explain the absence of the water molecule in one of the four crystal structures of cTEM-19m.

The general base during the deacylation step and possibly the acylation step, Glu166, in the Ω-loop active-site wall, adopted two conformers equally over the course of the 6 μs trajectories in TEM-1 and cTEM-2m *i.e.* the crystallized conformer (['g-', 'g-', 't']) and a second one (['g-', 'g-', 'g+']). PSE-4 and cTEM-17m adopted two further conformations (['g-', 't', 'g+'] and (['g-', 't', 'g+']), each at an approximate ratio of 25%, suggesting more flexibility.

Interestingly in cTEM-19m, the crystallized conformation (['g-', 'g-', 't']) was represented only 3.5% of the time. Instead, the ['g-', 't', 'g-'] conformer dominated 60 % of the time though it was almost absent in the other proteins' trajectories, while the ['g-', 't', 'g+'] conformer was observed over 30% of the MD simulation. This further supports the hypothesis that the coordination of the deacylating water molecule is destabilized in cTEM-19m.

Finally, Tyr105, on the Y105 gate-keeping active site wall was crystallized in the "closed" ['t', 'g'] conformation in TEM-1, PSE-4 and cTEM-2m. In cTEM-17m and cTEM-19m, the contrasting "open" ['g-', 'g'] conformation was crystallized. This conformation was only observed within the PDB for TEM-1 and PSE-4 crystal structures bound to either β-lactamase inhibitory proteins or inhibitors. The flexibility of this important side-chain has been previously proposed (Figure 5-2 B) [29, 55, 65]. The ensemble of all proteins revealed that, for all proteins under study here, the "open" ['g-', 'g'] conformation was predominantly populated during the simulations. This confirms the dynamic exchange of this residue between the two conformations, showing that the "open" state is also present in the absence of ligands.

5.4. Conclusion

Altogether, these results demonstrate the importance of cautiously analyzing the functional importance of structural variations gleaned from crystal structures. As demonstrated here, protein motions should also be considered when developing this causal relation. Having obtained four structures of cTEM-19m, where three exhibit similar conformations and one shows striking structural rearrangements, we demonstrate how dynamics can influence the resulting structure not necessarily captured by crystallographic parameters such as b-factors or alternate states. On the example of residue Ser130 in cTEM-19m crystal 4QY5 we showed clearly the power of a structural ensemble over static crystal structure analysis. The ensemble shows first indications of orientation towards a carbencillinase activity similar to PSE-4. Such a shift in sub-states of potential transition state conformations of active site residues was previously described by Tokuriki *et al.* along another artificial evolution path of a phosphotriesterase [45]. The structural ensemble revealed this evolutionary step directly, whereas four crystal structures and luck where necessary to show it experimentally.

These results are of prime interest as they exemplify the importance of interpreting crystallographic data with precaution. In the case of β-lactamases where side-chain orientation is important for function, misinterpretation or over-interpretation of crystallographic artefacts can be detrimental. Where sufficiently long, MD simulations have the particular advantage of covering most of the conformational space explored by side-chains without the necessity of resolving numerous crystal structures of the same protein. Since structural interpretations are usually mainly based on a single structural model, MD simulations on the timescale of interest revealing an ensemble of conformations should be considered as a requisite to correlate the importance of structural rearrangement to function.

5.5. Materials and methods

5.5.1. Materials

Nitrocefin was purchased from Calbiochem (Mississauga, ON). Kanamycin, isopropyl-1-thio-β-D-galactopyranoside (IPTG), components for growth media and purification materials were from Bioshop Canada (Burlington, ON). DEAE-Sepharose Fast Flow, Tris-Cl, PEG 4000, HEPES, ammonium sulfate and $MgCl_2$ used for protein crystallization were purchased from Sigma-Aldrich (Oakville, ON).

5.5.2. β-Lactamase expression and purification

Subcloning of the chimera cTEM-19m into pET-24 with the OmpA signal sequence was performed as previously described [3]. Protein expression and purification for crystallography essay was performed in auto-inducing ZYP-5052 medium as previously described [2, 3]. Briefly, chromatography was undertaken at 4°C using an Äkta FPLC (GE Healthcare). Following sample application, the DEAE-Sepharose Fast Flow column (1.6 cm × 30 cm) was washed with 3 column volumes (CV) of 10 mM Tris-Cl pH 7.0. A linear gradient to 200 mM Tris-Cl pH 7.0 was achieved over 4 CV, and the column was further washed over 3 CV. Fractions containing β-lactamase were identified first using a qualitative nitrocefin hydrolysis essay and then by SDS-polyacrylamide gel electrophoresis (15% (w/v) polyacrylamide gel) with Coomassie Brilliant Blue staining. A second purification step was

added. Fractions containing ≥ 75% β-lactamase following the first purification step were concentrated to a volume of 1.5 mL using an Amicon concentrator (MWCO 10000, Millipore, Billerica, MA) and applied to a Superose 12 column (1.6 cm × 55 cm) that had been pre-equilibrated with 50 mM Tris-HCl pH 7.0. Elution was performed at a flow rate of 1 mL/min. Fractions containing β-lactamase were identified by SDS-polyacrylamide gel electrophoresis (15% (w/v) polyacrylamide gel) with zinc-imidazole staining [67]. Purity and concentration were determined as above. Fractions containing β-lactamase with a purity of ≥ 95% were concentrated to 25 mg⁄mL (0.8 mM).

5.5.3. Crystallization of cTEM-19m and data collection

Two crystal forms of cTEM-19m were identified in this study. Crystals of the first form were obtained from a sitting drop prepared by mixing 1 μL protein solution (25 mg/mL in 50 mM Tris-Cl (pH 7.0) and 4 mM benzylpenicillin) and 1 μL reservoir solution (100 mM Tris-Cl (pH 8.5), 30% PEG 4000, and 0.2 M $MgCl_2$) and incubated at 22°C. Crystals of the second form were obtained from a hanging drop prepared by mixing 1 μL protein solution (20 mg/mL in 50 mM Tris-Cl (pH 7.0) and 3.2 mM benzylpenicillin) and 1 μL reservoir solution (100 mM Tris-Cl (pH 8.5), 24% PEG 4000, and 0.2 M $MgCl_2$), streak-seeded with the crystals of the first form, and incubated at 22°C. A single crystal from each group was isolated and flash-frozen in liquid nitrogen. Diffraction data were collected from each crystal at the Canadian Macromolecular Crystallography Facility Beamline 08ID-1 (Canadian Light Source, Saskatoon, SK).

5.5.4. Data processing and refinement

The collected diffraction images were indexed, integrated, and scaled with the *xia2* package [68]. Initial phases were calculated by molecular replacement with the program PHASER [69], with a previously determined TEM-1 structure (PDB 1ZG4) as a search model. The resultant structure models were further improved through iterative rounds of manual and automated refinement with COOT [15, 70], PHENIX [71], and REFMAC5 [72]. Data collection and refinement statistics are summarized in Table A 3-1. The refined structures of cTEM-19m (PDB IDs 4QY5 and 4QY6) have been deposited to the Protein Data Bank.

5.5.5. Molecular dynamics simulations

Crystal structures with the best ratio of resolution and completeness of the sequence were chosen for the two parents (TEM-1: 1XPB, 1.9 Å resolution and PSE-4: 1G68, 1.95 Å resolution) and the three variants (cTEM-2m: 4MEZ, 2.05 Å resolution, cTEM-17m: 4ID4, 1.05 Å resolution and cTEM-19m: 4R4S, 1.1 Å resolution). These structures were used as a starting point for all molecular dynamics (MD) simulations. Missing side-chains in 4MEZ (LYS32, LYS55, LYS111, ASP115, LYS146, LYS1192 and ARG277), 4ID4 (LYS32, LYS55, LYS111, LYS146, GLN154, LYS158, LYS173, ASP254 and LYS256) and 4R4S (LYS 111, GLN 154, LYS 158, LYS215 and LYS256) were reconstructed using the rotamer explorer in the Structure Preparation plugin of MOE [73]. All following steps were executed using GROMACS 5.0.1. The protonation and orientation of histidines was verified using the high-resolution crystal structure of TEM-1 variant M182T (1M40), MolProbity [74] and PROPKA [75, 76]. For each system, crystallographic water molecules were conserved in a truncated dodecahedron periodic water box of SPC/E water molecules with a minimum distance of 10 Å between the box boundary and the protein. The total number of water molecules and net charge for the different systems were: 1XPB: 12,812, -7; 1G68: 12,475, -5; 4MEZ: 9,952, -7; 4ID4: 10,177, -4; 4R4R: 10,703, -4. Each system was neutralized by adding the respective number of sodium counterions. The total size of the systems, in atoms, was: 1XPB: 42,507; 1G68: 41,501; 4MEZ: 33,926, 4ID4: 34,619; 4R4R: 36,196.

For energy minimization and MD simulations, GROMACS 5.0.1 was used with the AMBER99SB-ILDN force field [77, 78]. The LINear Constraint Solver (LINCS) algorithm was applied to all bonds containing hydrogen atoms. Each 1 ps frame was saved in a compressed XTC trajectory and each 100 ps in a full precision TRR trajectory. The Nose-Hoover thermostat and the Parrinello-Rahman barostat were used to couple the system to a constant temperature of 304.65 K and a pressure of 1 bar. The electrostatic interactions were evaluated by the particle-mesh Ewald method, and Lennard-Jones interactions were evaluated using a 1.2 nm cutoff. Each system was energy minimized using the method of the Steepest Descent to a target F_{max} of no greater than 1,000 kJ mol^{-1} nm^{-1}. The time step was chosen to be 2 fs. For equilibration, all systems were subjected to 2 ns NVT ensemble to gradually heat up the system from 10 K to 304.65 K. Next, a 20 ns NPT ensemble was generated and three

conformations with the highest, average and lowest root mean-square deviation (RMSD) were chosen as starting conformations for three individual replicates. A further 2 ns NVT ensemble was performed for each repetition to heat up the system and used as starting point for the 2 μs production run. This results in a total of three 2 μs simulations with 1 ps compressed and 100 ps full precision trajectories for each of the five proteins.

5.5.6. Analysis of protein structure and dynamics

Analysis was performed using GROMACS 5.0.1 on 2 μs for each MD simulation, unless otherwise stated. All protein figures were created using PyMOL [79] or MOE (Chemical Computing Group, Montréal, QC). The stability of the system was evaluated after 100 ns and 1,000 ns by calculating the RMSD of the backbone atoms between each frame and the initial conformation. In addition, the radius of gyration was calculated for each replicate over 2 μs to ensure overall structural integrity.

The cluster analysis was done using the cluster tool of GROMACS 5.0.1 and the Jarvis-Patrick algorithm. Three identical nearest neighbors where required to form a cluster and 20 nearest neighbors are considered for Jarvis-Patrick algorithm.

5.5.7. Normalized b-factors

The average b-factor of the structure was subtracted from the per residue Cα b-factor and normalized to the three times the standard variation of the b-factor of the structure.

5.6. Acknowledgments

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants RGPIN 227853. M.E. is the recipient of an NSERC Vanier Canada Graduate Scholarship and S.G. is the recipient of a FRQ-NT Graduate Scholarship. Computational resources were provided by Calcul Québec and Compute Canada, funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the FRQ-NT.

5.7. References

1. Gobeil, S.M., et al., *15N, 13C and 1H backbone resonance assignments of an artificially engineered TEM-1/PSE-4 class A beta-lactamase chimera and its deconvoluted mutant.* Biomol NMR Assign, 2016. **10**(1): p. 93-99.
2. Gobeil, S.M., et al., *Maintenance of native-like protein dynamics may not be required for engineering functional proteins.* Chemistry & biology, 2014. **21**(10): p. 1330-1340.
3. Clouthier, C.M., et al., *Chimeric beta-lactamases: global conservation of parental function and fast time-scale dynamics with increased slow motions.* PloS one, 2012. **7**(12): p. e52283.
4. Morin, S., et al., *Backbone resonance assignments of an artificially engineered TEM-1/PSE-4 Class A beta-lactamase chimera.* Biomol NMR Assign, 2010. **4**(2): p. 127-130.
5. Lindorff-Larsen, K., et al., *Simultaneous determination of protein structure and dynamics.* Nature, 2005. **433**(7022): p. 128-132.
6. Bonomi, M., et al., *Principles of protein structural ensemble determination.* Curr Opin Struct Biol, 2017. **42**: p. 106-116.
7. Vendruscolo, M., *Determination of conformationally heterogeneous states of proteins.* Curr Opin Struct Biol, 2007. **17**(1): p. 15-20.
8. Babic, M., A. Hujer, and R. Bonomo, *What's new in antibiotic resistance? Focus on beta-lactamases.* Drug Resist Updat, 2006. **9**(3): p. 142-156.
9. Bush, K. and P.A. Bradford, *β-Lactams and β-Lactamase Inhibitors: An Overview.* Cold Spring Harb Perspect Med, 2016: p. a025247.
10. Bush, K., *A resurgence of β-lactamase inhibitor combinations effective against multidrug-resistant Gram-negative pathogens.* Int J Antimicrob Agents, 2015. **46**(5): p. 483-493.
11. Philippon, A., et al., *A Structure-Based Classification of Class A β-Lactamases, a Broadly Diverse Family of Enzymes.* Clin Microbiol Rev, 2015. **29**(1): p. 29-57.
12. Brolund, A. and L. Sandegren, *Characterization of ESBL disseminating plasmids.* J Infect Dis, 2015. **48**(1): p. 18-25.

13. Stojanoski, V., et al., *A triple mutant in the Omega-loop of TEM-1 beta-lactamase changes the substrate profile via a large conformational change and an altered general base for catalysis.* J Biol Chem, 2015. **290**(16): p. 10382-10394.
14. Stojanoski, V., et al., *A Triple Mutant in the Ω-loop of TEM-1 β-Lactamase Changes the Substrate Profile via a Large Conformational Change and an Altered General Base for Catalysis.* J Biol Chem, 2015. **290**(16): p. 10382-10394.
15. Stec, B., et al., *Structure of the wild-type TEM-1 β-lactamase at 1.55 Å and the mutant enzyme Ser70Ala at 2.1 Å suggest the mode of noncovalent catalysis for the mutant enzyme.* Acta Cryst D, 2005. **61**(8): p. 1072-1079.
16. Swarén, P., et al., *X-ray Structure of the Asn276Asp Variant of theEscherichia coliTEM-1 β-Lactamase: Direct Observation of Electrostatic Modulation in Resistance to Inactivation by Clavulanic Acid†,‡.* Biochemistry, 1999. **38**(30): p. 9570-9576.
17. Orencia, M.C., et al., *Predicting the emergence of antibiotic resistance by directed evolution and structural analysis.* Nat Struct Biol, 2001. **8**(3): p. 238-242.
18. Thomas, V.L., et al., *Structural Consequences of the Inhibitor-Resistant Ser130Gly Substitution in TEM β-Lactamase†,‡.* Biochemistry, 2005. **44**(26): p. 9330-9338.
19. Wang, X., *The Structural Bases of Antibiotic Resistance in the Clinically Derived Mutant beta -Lactamases TEM-30, TEM-32, and TEM-34.* J Biol Chem, 2002. **277**(35): p. 32149-32156.
20. Brown, N.G., et al., *Structural and Biochemical Evidence That a TEM-1 -Lactamase N170G Active Site Mutant Acts via Substrate-assisted Catalysis.* J Biol Chem, 2009. **284**(48): p. 33703-33712.
21. Fraser, J.S., et al., *Accessing protein conformational ensembles using room-temperature X-ray crystallography.* Proc Natl Acad Sci USA, 2011. **108**(39): p. 16247-16252.
22. Jelsch, C., et al., *Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 A resolution.* Proteins, 1993. **16**(4): p. 364-383.
23. Jelsch, C., et al., *Beta-lactamase TEM1 of E. coli. Crystal structure determination at 2.5 A resolution.* FEBS Lett, 1992. **299**(2): p. 135-142.

24. Lim, D., et al., *Insights into the molecular basis for the carbenicillinase activity of PSE-4 beta-lactamase from crystallographic and kinetic studies.* Biochemistry, 2001. **40**(2): p. 395-402.

25. Fonze, E., et al., *TEM1 beta-lactamase structure solved by molecular replacement and refined structure of the S235A mutant.* Acta Crystallogr D Biol Crystallogr, 1995. **51**(Pt 5): p. 682-694.

26. Strynadka, N.C., et al., *Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution.* Nature, 1992. **359**(6397): p. 700-705.

27. Minasov, G., X. Wang, and B.K. Shoichet, *An Ultrahigh Resolution Structure of TEM-1 β-Lactamase Suggests a Role for Glu166 as the General Base in Acylation.* J Am Chem Soc, 2002. **124**(19): p. 5333-5340.

28. Savard, P.-Y. and S.M. Gagné, *Backbone Dynamics of TEM-1 Determined by NMR: Evidence for a Highly Ordered Protein†.* Biochemistry, 2006. **45**(38): p. 11414-11424.

29. Doucet, N., et al., *NMR investigation of Tyr105 mutants in TEM-1 beta-lactamase: dynamics are correlated with function.* J Biol Chem, 2007. **282**(29): p. 21448-21459.

30. Morin, S. and S.M. Gagne, *NMR dynamics of PSE-4 beta-lactamase: an interplay of ps-ns order and mus-ms motions in the active site.* Biophys J, 2009. **96**(11): p. 4681-4691.

31. Damblon, C., et al., *The catalytic mechanism of beta-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme.* Proc Natl Acad Sci U S A, 1996. **93**(5): p. 1747-1752.

32. Hermann, J.C., et al., *Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A beta-lactamase with benzylpenicillin.* J Am Chem Soc, 2005. **127**(12): p. 4454-4465.

33. Meroueh, S.O., et al., *Ab initio QM/MM study of class A beta-lactamase acylation: dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70.* J Am Chem Soc, 2005. **127**(44): p. 15397-15407.

34. Meroueh, S.O., et al., *Molecular Dynamics at the Root of Expansion of Function in the M69L Inhibitor-Resistant TEM β-Lactamase fromEscherichiacoli.* J Am Chem Soc, 2002. **124**(32): p. 9422-9430.

35. Roccatano, D., et al., *Dynamical aspects of TEM-1 beta-lactamase probed by molecular dynamics.* J Comput Aided Mol Des, 2005. **19**(5): p. 329-340.
36. Diaz, N., et al., *Molecular dynamics simulations of the TEM-1 beta-lactamase complexed with cephalothin.* J Med Chem, 2005. **48**(3): p. 780-791.
37. Díaz, N., et al., *Acylation of Class A β-lactamases by Penicillins: A Theoretical Examination of the Role of Serine 130 and the β-lactam Carboxylate Group.* J Phys Chem B, 2001. **105**(45): p. 11302-11313.
38. Lamotte-Brasseur, J., et al., *pKa calculations for class A beta-lactamases: influence of substrate binding.* Protein Sci, 1999. **8**(2): p. 404-409.
39. Diaz, N., et al., *Insights into the acylation mechanism of class A beta-lactamases from molecular dynamics simulations of the TEM-1 enzyme complexed with benzylpenicillin.* J Am Chem Soc, 2003. **125**(3): p. 672-684.
40. Golemi-Kotra, D., et al., *The importance of a critical protonation state and the fate of the catalytic steps in class A beta-lactamases and penicillin-binding proteins.* J Biol Chem, 2004. **279**(33): p. 34665-34673.
41. Hermann, J.C., et al., *Identification of Glu166 as the general base in the acylation reaction of class A beta-lactamases through QM/MM modeling.* J Am Chem Soc, 2003. **125**(32): p. 9590-9591.
42. Fisette, O., et al., *TEM-1 backbone dynamics-insights from combined molecular dynamics and nuclear magnetic resonance.* Biophys J, 2010. **98**(4): p. 637-645.
43. Ambler, R.P., et al., *A standard numbering scheme for the class A beta-lactamases.* Biochem J, 1991. **276 (Pt 1)**: p. 269-270.
44. Warshel, A. and R.P. Bora, *Perspective: Defining and quantifying the role of dynamics in enzyme catalysis.* J Chem Phys, 2016. **144**(18): p. 180901.
45. Campbell, E., et al., *The role of protein dynamics in the evolution of new enzyme function.* Nat Chem Biol, 2016. **12**(11): p. 944-950.
46. Meyer, M.M., L. Hochrein, and F.H. Arnold, *Structure-guided SCHEMA recombination of distantly related beta-lactamases.* Protein Eng Des Sel, 2006. **19**(12): p. 563-570.
47. Voigt, C.A., et al., *Protein building blocks preserved by recombination.* Nat Struct Biol, 2002. **9**(7): p. 553-558.

48. Minasov, G., X. Wang, and B.K. Shoichet, *An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation.* J Am Chem Soc, 2002. **124**(19): p. 5333-5340.
49. Waley, S.G., *β-Lactamase: mechanism of action.* 1992: p. 198-228.
50. Wladkowski, B.D., et al., *Acylation of β-Lactams by Class A β-Lactamase: Anab InitioTheoretical Study on the Effects of the Oxy-Anion Hole.* J Am Chem Soc, 1997. **119**(27): p. 6423-6431.
51. Strynadka, N.C.J., et al., *Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution.* Nature, 1992. **359**(6397): p. 700-705.
52. Lim, D., et al., *Insights into the Molecular Basis for the Carbenicillinase Activity of PSE-4 β-Lactamase from Crystallographic and Kinetic Studies.* Biochemistry, 2001. **40**(2): p. 395-402.
53. Herzberg, O. and J. Moult, *Penicillin-binding and degrading enzymes.* Curr Opin Struct Biol, 1991. **1**(6): p. 946-953.
54. Herzberg, O. and J. Moult, *Bacterial resistance to beta-lactam antibiotics: crystal structure of beta-lactamase from Staphylococcus aureus PC1 at 2.5 A resolution.* Science, 1987. **236**(4802): p. 694-701.
55. Doucet, N., P.Y. De Wals, and J.N. Pelletier, *Site-saturation mutagenesis of Tyr-105 reveals its importance in substrate stabilization and discrimination in TEM-1 beta-lactamase.* J Biol Chem, 2004. **279**(44): p. 46295-46303.
56. Wlodawer, A., et al., *Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures.* Febs J, 2008. **275**(1): p. 1-21.
57. Dellus-Gur, E., et al., *What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs.* J Mol Biol, 2013. **425**(14): p. 2609-2621.
58. Yuan, Z., T.L. Bailey, and R.D. Teasdale, *Prediction of protein B-factor profiles.* Proteins, 2005. **58**(4): p. 905-912.
59. Keedy, D.A., J.S. Fraser, and H. van den Bedem, *Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit.* PLoS Comput Biol, 2015. **11**(10): p. e1004507.

60. Woldeyes, R.A., D.A. Sivak, and J.S. Fraser, *E pluribus unum, no more: from one crystal, many conformations.* Curr Opin Struct Biol, 2014. **28**: p. 56-62.
61. Fraser, J.S., et al., *Hidden alternative structures of proline isomerase essential for catalysis.* Nature, 2009. **462**(7273): p. 669-U149.
62. Herzberg, O., *Refined crystal structure of beta-lactamase from Staphylococcus aureus PC1 at 2.0 A resolution.* J Mol Biol, 1991. **217**(4): p. 701-719.
63. Knox, J.R. and P.C. Moews, *Beta-lactamase of Bacillus licheniformis 749/C. Refinement at 2 A resolution and analysis of hydration.* J Mol Biol, 1991. **220**(2): p. 435-455.
64. Hovmoller, S., T. Zhou, and T. Ohlson, *Conformations of amino acids in proteins.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 5): p. 768-776.
65. Doucet, N. and J.N. Pelletier, *Simulated annealing exploration of an active-site tyrosine in TEM-1 beta-lactamase suggests the existence of alternate conformations.* Proteins, 2007. **69**(2): p. 340-348.
66. Scouras, A.D. and V. Daggett, *The Dynameomics rotamer library: amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water.* Protein Sci, 2011. **20**(2): p. 341-352.
67. Fernandezpatron, C., L. Castellanosserra, and P. Rodriguez, *Reverse Staining of Sodium Dodecyl-Sulfate Polyacrylamide Gels by Imidazole-Zinc Salts - Sensitive Detection of Unmodified Proteins.* Biotechniques, 1992. **12**(4): p. 564-&.
68. Winter, G., C.M. Lobley, and S.M. Prince, *Decision making in xia2.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1260-1273.
69. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
70. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-2132.
71. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-221.

72. Vagin, A.A., et al., *REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2184-2195.
73. (MOE), M.O.E., *2016*. Chemical Computing Group Inc.
74. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
75. Sondergaard, C.R., et al., *Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values.* J Chem Theory Comput, 2011. **7**(7): p. 2284-2295.
76. Olsson, M.H., et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J Chem Theory Comput, 2011. **7**(2): p. 525-537.
77. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-1958.
78. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
79. Schrödinger, L., *The PyMOL Molecular Graphics System*.

Chapter 6 - Revealing the hidden information about energy barriers of gas migration in structural ensembles

6.1. Context

As discussed previously, the prediction of ligand migration is a complex and computationally intensive task (see Chapter 1, Chapter 2 and Chapter 4). In particular, the calculation of the free energy landscape can be extremely biased and depends on many variables such as starting point of the ligand, the initial conformation of the protein and even the method of calculation. Nonetheless, the ligand binding mode is of prime importance to better understand the underlying foundation of protein function and ultimately to enable protein engineering. The ligand binding presented thus far in this thesis could, in principle, have been studied exclusively using experimental methods. However, those methods do not scale well and require an enormous investment of labour and time. Examples are known in literature but augmenting experimental methods with computational predictions enables us to study many different ligands [1, 2].

In some cases a ligand cannot be studied using experimental tools or only by using an artificial protein system. O_2 binding to globins or oxygenases is such an example [3-5]. Rather than altering the system under study, existing computational methods could be applied to study O_2 migration in proteins. Indeed, many computational techniques exist to observe gaseous egress in proteins, as presented in the following chapter which is a "pre-print" version of the article entitled '*Evolution of P450 Monooxygenases Toward Formation of Transient Channels and Exclusion of Nonproductive Gases*', published in 2016 in *ACS Catalysis* [6]. The strong correlation between experimental data and free MD simulations, demonstrated in Annex 4 and Chapter 5, encouraged us to rely on methods based on unbiased MD trajectories. Using implicit ligand sampling of CO, N_2 and O_2 as well as free O_2 diffusion enabled us to study the free energy landscape of those gases and egress tunnels, which would not have been possible using experimental methods only. This study is a thought-provoking example where the experimental setup would have been more artificial and highly biased than is the computational prediction.

My contribution to this study was the conceptualization, MD simulation, algorithm development, data analysis, drafting figures and making movies. Simon Dürr performed the implicit ligand sampling simulation and Armande Ang Houle generated the homology model of CYP102A5. The computational work was under the supervision of Prof. Guillaume Lamoureux and Prof. Joelle Pelletier, while the work was performed in the laboratory of Prof. Pelletier. The manuscript was drafted mainly by myself with the contribution of Simon Dürr, assisted by Prof. Pelletier and with revisions by all authors.

The evolution of P450 monooxygenases towards formation of transient channels and exclusion of non-productive gases

Maximilian C.C.J.C. Ebert[1,2,3], Simon L. Dürr[2,3,4], Armande A. Houle[1,2,3], Guillaume Lamoureux[2,5] and Joelle N. Pelletier[1,2,3,4]

[1] Département de biochimie, Université de Montréal, Montréal, H3T 1J4, Canada

[2] PROTEO, The Québec Network for Research on Protein Function, Engineering and Applications, Québec, G1V 0A6, Canada

[3] CGCC, the Center for Green Chemistry and Catalysis, Montréal, H3T 1J4, Canada

[4] Département de chimie, Université de Montréal, Montréal, H3T 1J4, Canada

[5] Department of Chemistry and Biochemistry and Centre for Research in Molecular Modeling (CERMM), Concordia University, Montreal, H4B 1R6, Canada

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

6.2. Abstract

Gaseous molecules are essential for biological processes, yet mapping the migration of gas molecules into and out of proteins represents a significant challenge. Cytochrome P450 enzymes (P450s) contain numerous channels speculated to be populated with their substrates, products, solvents and gases, yet the principles underlying channel preference are unknown. We identified the multiple putative ligand migration channels of two bacterial P450s, CYP102A1 (BM3) from *Bacillus megaterium* and CYP102A5 from *Bacillus cereus* using Implicit Ligand Sampling and free molecular dynamics simulations, and furthermore characterized the energy of gas migration through each. We observed a strong discrimination between preferred gas migration channels, previously identified substrate/product migration channels and water channels, and mapped putative O_2 reservoirs in the enzyme core. The protein backbone dynamics (S^2 order parameter) unexpectedly revealed that some channels are transient in nature, with sub-channels forming and merging, and O_2 molecules hopping between sub-channels. Finally, we present evidence of the evolution towards O_2 binding in conjunction with protection against inhibitory CO and exclusion of N_2. Our results significantly increase our understanding of gas migration in proteins and provide insights into the evolution of gas-utilizing enzymes.

Keywords: P450, implicit ligand sampling, channels, molecular dynamics simulations, oxygen, carbon monoxide, BM3, CYP102A1

6.3. Introduction

The migration of molecules in protein channels is a defining parameter in a number of biological systems. Experimental and theoretical analyses of pores, cavities and channels have revealed their crucial role in biomolecule function, in roles as diverse as membrane transport [7-9], ribosomal protein synthesis [10, 11] and enzyme catalysis [5, 12-17]. Channels inside enzymes provide molecular access to and from the active site. Despite decades of studies on enzyme kinetics and ever-increasing numbers of enzyme structures, detailed knowledge about the stability, lifetime and basis for ligand preference exhibited by enzyme channels remains largely elusive. There is a need to implement facile and reliable tools to examine the molecular nature of channels and the energetics of the molecules that occupy them.

Studying ligand migration in channels experimentally is particularly challenging in the context of gaseous ligands. The migration of gas molecules has been extensively studied in oxygen transporting heme-containing proteins [5, 12, 18] such as globins [19, 20], particularly myoglobin [21-23]. Among the methodologies used to characterize gas migration, experimental methods such time-resolved crystallography [24, 25], spectroscopy [26] and competition kinetics using xenon gas [4] generally provide indirect evidence for migration of the physiologically relevant gases. Furthermore, those methods provide little insight into the preference that gases have for specific channels. This knowledge is required to enable the engineering of these enzymes by modulating ligand entry into the channels.

A further confounding factor is the presence of multiple channels in many enzymes. It is particularly challenging to characterize and to engineer those systems, because of the dearth of tools available to evaluate the preference of each ligand for a given tunnel.

The O_2-dependent heme-containing cytochrome P450 enzymes (P450s) are characterized by a series of interconnected substrate channels, stemming from several mouths on the enzyme surface and joining near the active site. These mono-oxygenases are of prime importance for biocatalysis where they serve to direct O_2 onto non-activated carbons [27, 28]. They are also central to drug development as a result of their natural role in detoxification of xenobiotics and steroid biosynthesis [27, 28]. As observed for other heme-containing proteins, P450s have the particularity of harboring multiple channels [12, 18, 29]. These channels are

the main determinants of substrate specificity [30, 31]. P450s also have a water channel, to evacuate water from the active site upon substrate binding and promote reactivity [32].

P450 mono-oxygenases require molecular oxygen at the heme reaction center during the catalytic cycle and are inhibited by CO. While there is growing knowledge of the mode of binding of substrates to P450s, the diffusion of gas molecules to the reaction center is currently poorly understood. Knowledge about gas migration pathways is of prime importance to better understand reaction dynamics of P450s. The transient nature of the ferrous-O_2 complex in P450s and the fact that free O_2 molecules in proteins are not sufficiently spatially resolved to lend themselves to crystallographic resolution, renders research on gas migration pathways challenging. Previous studies on $P450_{cam}$ using laser flash photolysis of CO revealed gas migration kinetics that were significantly more complex than those of globins. The dynamic events suggested the existence of multiple CO docking sites and CO migration pathways different from those of water in $P450_{cam}$ [33]. The question remains as to how CO migration compares to migration of other gases, particularly the substrate O_2.

While the growing number of P450 structures reported has allowed identification of the main substrate channels, different structures of a single P450 can reveal different channels [12-14]. Using Caver3, Cojocaru *et al.* identified potential substrate migration channels for a number of well-studied P450s [12]. High variability was observed in the region of the main substrate channels (collectively known as the channel 2 region), allowing the definition of many channel subclasses. Supported by computer simulations and experimental data, this region was determined to undergo important conformational changes, indicative of flexibility [5, 13, 14, 34, 35]. That protein dynamics are an intrinsic property of channel formation is a further confounding factor to consider when mapping gas migration on the basis of crystal structures.

In this work we used molecular dynamics (MD) simulation techniques to provide a dynamic framework to elucidate potential gas migration pathways for relevant gases: the catalytic O_2, the highly prevalent N_2 and the inhibitor CO. Over the past years, MD simulations have been used to model the passage of relevant gases in proteins notably Locally Enhanced Sampling (LES) and Implicit Ligand Sampling (ILS) [19, 21, 22, 36-38]. In particular, the ILS methodology has been applied to study the oxygen migration paths in the heme-containing lipoxygenase and hemoglobin [21, 22, 36, 38]. Using ILS methodology, we

derived the 3D free-energy landscape for those molecules during migration inside two members of the important CYP102A subclass of P450s [21]. The first P450 is CYP102A1, also known as BM3, which is among the most studied bacterial P450s. It has numerous biotechnological applications including recent use in drug discovery as a result of its human P450-like substrate scope in drug metabolism [39]. We generated a high-resolution homology model of its high-activity homologue CYP102A5, to extend our insights into channel formation and gas migration to a second member of this class of CYPs [40].

By those means, we predict the pathways for gas migration. We present and rank preference for each gas migration channel based on ILS simulations and on MD simulations that include explicit ligands. Our free MD simulations highlight the protein flexibility and reveal, for the first time, the formation of transient bridges for gas molecules to exchange between known channels. We identify energy barriers that are consistent with the evolutionary protection of the heme from CO. Finally, we analyze the protein backbone dynamics at the fast timescale by deriving the S^2 order parameters, confirming the highly dynamic nature of the channels mouths. Overall, our results indicate that the channels are predominantly stable over time yet can be prone to significant dynamic alterations. This approach will be useful to map gas binding sites or engineer new features in all gas-dependent proteins.

6.4. Results and discussion

6.4.1. CYP102A1 is the template for the homology model of CYP102A5

CYP102A1 (or BM3) is a widely studied bacterial P450. It has been the target of numerous engineering efforts to oxidize new substrates or to perform new chemistry [41-44]. In contrast, CYP102A5 is a relatively new member of the CYP102A subclass of cytochrome P450 enzymes. It exhibits a higher turnover rate for typical substrates such as long-chain fatty acids than most P450s in its subclass, and is characterized by greater regioselectivity, favoring the ω-1 and ω-2 positions to a greater extent than its homologs [40]. It has been speculated that its high rate for the mono-oxygenation reaction is at least partially due to a significantly faster electron transfer rate compared to CYP102A1 [40]. To further elucidate the

underpinnings of those differences in reactivity and to gain new insights into the mode of gas entry and regulation in CYPs, we computationally examined gas migration in the heme-binding domain of CYP102A1 and CYP102A5.

A homology model of CYP102A5 based on a template co-crystalized with a substrate has been reported in a study focused on human drug metabolism [45]. However, our interest lies in the apoenzyme, and thus required the generation of a new model. As there are only two insertions in CYP102A5 (D235 and E355, see also Figure A 6-1) relative to CYP102A1 and as the two homologs share a high sequence identity (66%, Figure A 6-1), the resulting model was expected to be of high quality. The apoenzyme structure of CYP102A1 2HPD was used as a template to generate a high-resolution model of apo-CYP102A5, using the Homology Model module in MOE [46]. Following energy minimization, the best model based on the Generalized Born/Volume Integral (GB/VI) methodology was selected for further analysis [47]. As expected, superposition of the CYP102A5 model with the CYP102A1 template shows a good agreement on the secondary structure elements, with a Cα root mean square deviation (RMSD) of 1.58 Å (Figure 6-1). Three regions exhibit above-average per-residue RMSD, representing major structural differences and hence possible explanations for the functional differences between the two enzymes (Figure 6-1 B). Region I includes the CYP102A1 residues R47 and Y51, recognized as key residues in initial substrate binding; they may also be implicated in the regulation of water and co-solvent access to the active site [48]. Both residues act as hydrophilic binding elements that can tether the carboxylate group of fatty acid substrates to the enzyme surface, thus increasing the local concentration of the substrate and favoring substrate entry in a hydrophobic tunnel. The main driving force is the ionic bond between the guanidinium group of R47 and the carboxylate of the substrate [49]. Interestingly, this ‘fetching mechanism’ is not conserved throughout the CYP102A family: the homologous residues in CYP102A5 are D50 and V54, which cannot support carboxylate binding. While both enzymes share the same substrate scope, the difference in initial substrate recognition at a channel mouth may explain the altered affinity and increased turnover. Furthermore, the mode of substrate anchoring at the mouth may translate into different substrate orientation at the active site and hence modulate specificity of hydroxylation of fatty acids exhibited by CYP102A5 [40].

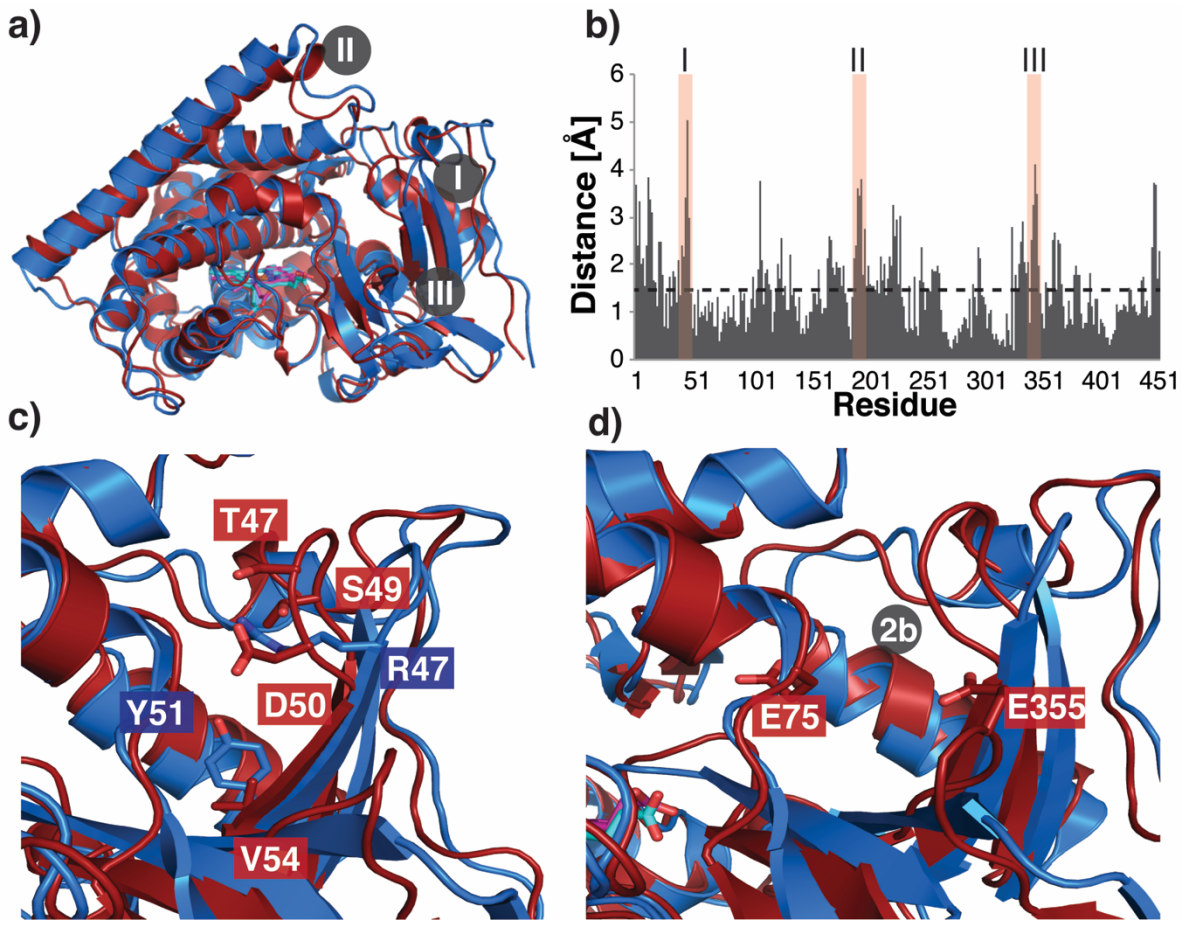


Figure 6-1 Structure of CYP102A1 and CYP102A5.
A) Superposition of the heme-binding domain of CYP102A1 (Protein: blue, Heme: cyan; PDB 2HPD) and the homology model of CYP102A5 (Protein: red, Heme: pink). Regions I-III are defined in panel B. B) Per-residue C-α RMSD of the CYP102A5 homology model and CYP102A1 (2HPD). Regions I-III indicate the largest structural differences. C) Close-up of Region I. Key residues of the fetching mechanism of CYP102A1 (blue) or the proposed fetching mechanism of CYP102A5 (red) are shown in sticks. D) Region III with a marker indicating the position of tunnel 2b, which is obstructed in CYP102A5 (red) as a result the insertion E355 and rearrangement of β-sheet 1.

Examination of the homology model suggests that the polar T47 and S49, located in the vicinity of CYP102A1 R47 and Y51 (Figure 6-1 C), can provide a substitute fetching mechanism. This would be aided by their position on the flexible loop between the 1-1 and 1-2 strands of the β1-sheet (nomenclature according to Hasemann *et al* .[50]), to compensate for their reduced side-chain length compared to R47 of CYP102A1. The loop orientation results in a superposition of the CYP102A5 T47 side-chain hydroxyl with the CYP102A1 R47 side-

chain guanidinium, suggesting similar H-bonding with the substrate. S49 of CYP102A5 offers additional H-bonding possibilities, in a manner similar to Y51 of CYP102A1. The model also confirms that the structurally conserved B'-helix, the dipole of which is thought to stabilize substrate binding in CYP102A1, is positioned to play a similar role in substrate binding in CYP102A5 [40, 50-52].

Region II is the F-G loop, which shows the highest per-residue RMSD. This region is known to be flexible, as illustrated by its displacement in almost all crystal structures of CYP102A1 [53]. The conformation of the CYP102A5 F-G loop modeled here represents a new orientation that is more closed and tilted towards the B'-helix than observed in structures of CYP102A1. The different conformation likely reflects the fact that the template, CYP102A1, holds two Pro in this loop where CYP102A5 has none. The reduced conformational constraints and the different pattern of charged residues appear to promote the different orientation of the F-G loop in CYP102A5.

Region III is of particular interest because it harbors the insertion of E355 relative to CYP102A1. This insertion appears to obstruct the 2b egress channel (Figure 1d; channels are identified according to the nomenclature of Cojocaru *et al* .[12]). This channel belongs to the channel 2 region shared by most members of the P450 superfamily, which harbors the greatest number of transient substrate migration channels identified to date. While Thermal Motion Pathway calculations have revealed the additional channel 1 and channel 3 regions, MD simulations on $P450_{cam}$ have shown that the channel 2 region is energetically favored and is the preferred path of substrate entry [13, 14]. Crystal structures and MD simulations show that sub-channels nested in this region often merge in response to structural reorientations [13, 14, 34, 35]. The conformation of the channel 2 region is dictated by the F-G helix and the β1-sheet; the insertion of E355 in the β1-sheet of CYP102A5 alters the conformation of the channel 2 region to obstruct it.

Chowdhary *et al.* found that all residues 6 Å from ω1-3 of substrates crystallized in CYP102A1 are conserved in CYP102A5, and thus speculated that their difference in regiospecificity is conformationally driven. To investigate this hypothesis, we calculated the active site volume using MOE [46] Site Finder and DoGSiteScorer [54] and found it be approximately 50% smaller for CYP102A5 than in the crystal structure of CYP102A1 (PDB:

2HPD). This is likely to result in important differences in conformational restriction and may form the basis for greater regioselectivity of CYP102A5.

6.4.2. Identification of ligand migration channels in P450s using CAVER3

To investigate the apparent obstruction of channel 2b in the model of CYP102A5, we predicted the substrate migration channels using CAVER3. CAVER defines channels by placing spheres consecutively from a starting point on a grid until it reaches the bulk media. A sphere is accepted on the grid if it causes no steric clash with the protein structure. We used a sphere of radius 1.2 Å, which represents the size of a water molecule in the CAVER algorithm. To provide a reasonable comparison, the procedure was validated with CYP102A1 using the single input PDB file 2HPD; we note once more that different crystal structures of P450s may reveal different subsets of channels, as a result of being static rather than dynamic.

In CYP102A1 (PDB 2HPD), CAVER3 identified channels 2a, 2b, 2f and W, which form a subset of the six channels identified in a report that considered all crystal structures available in 2006 (Figure A 6-2 A) [12]. In contrast, channels 2a and 2b were not identified by CAVER3 in CYP102A5, supporting our hypothesis that E355 causes an obstruction and a reorientation of the channel 2 region. In fact, only channels 2f and W were predicted by CAVER3 using the homology model of CYP102A5 (Figure A 6-2 B). It has been speculated that the substrate carboxylate binds differently in both enzymes due to the negatively charged residues E355 and E75 in CYP102A5, that lie in the vicinity of the channel 2 carboxylate binding site of CYP102A1 [40]. The absence of channels 2a and 2b in our CAVER3 calculations may diminish the relevance of E355 and E75 on the mechanism of substrate binding. Furthermore, the proposed fetching mechanism based on T47 and S49 would shift initial substrate binding away from those negatively charged residues, closely mimicking the binding mode involving the fetching residue Y51 in CYP102A1. Overall, comparison of the channels predicted by CAVER3 using a single input file per protein provides an indication that CYP102A1 and CYP102A5 differ in the distribution of their ligand migration channels.

6.4.3. Global energy minima for O_2 inside the CYPs using ILS

CAVER3 was run using static protein coordinates which may have limited its capacity to identify any transient channels. Indeed, the complete identification of all channels in CYP102A1 required dynamics simulations in addition to several crystal structures [12]. CAVER3 offers the option to probe for channels while allowing for protein motions using MD trajectories and can extract the presence of transient tunnels based on a geometric probe. However, the method cannot extract information about the energy landscape. For that reason, we applied the method of Implicit Ligand Sampling (ILS). Using a gaseous probe equipped with both a geometric and a van der Waals definition, this method not only reveals channels but also allows monitoring the probable sites of high gas occupancy and ultimately ranking the channels based on the ILS Gibbs free energy calculation. The substrate O_2 served as the gaseous probe to calculate the 3D free-energy landscape inside CYP102A1 and CYP102A5. Specifically, we determined the Gibbs free energy (ΔG_{gas}) associated with placing a single gas molecule from vacuum into 1 $Å^3$ increments throughout the entire volume of each 1 ps timeframe of a free molecular dynamics (MD) trajectory obtained in the absence of that gas. Regions with a low ΔG_{gas} have a high probability of accommodating a gas molecule. The ΔG_{gas} of solvation ($\Delta G_{s,gas}$) was used to validate the ILS calculation. Comparison of the theoretical $\Delta G_{s,gas,Henry}$ determined using Henry's law and $\Delta G_{s,gas,ILS}$ calculated by ILS of our MD trajectory showed good agreement (Table A 6-1). All ΔG_{gas} values calculated herein are thus reported in reference to the appropriate $\Delta G_{s,gas}$. For ease of comparison, a ΔG^0_{gas} of 0 is defined as the solvation energy of each gas.

The free energy distribution at the core of CYP102A1 and CYP102A5 show clear differences in their global minima (Figure 6-2). CYP102A1 exhibits two global minima. The first is located on top of the I-Helix, which forms a deep wall inside the active site cavity. The second is inside the hydrophobic substrate tunnel between the F and B'-helices. CYP102A5 shows one global minimum between the E and I-helices, in close proximity to the H-helix. Interestingly, this minimum converges with the region where three out of five xenon atoms crystalized in the homologous cytochrome $P450_{cam}$ (PDB 1UYU; Figure 6-2); Xe had been used as a stable probe of gas binding pockets [5].

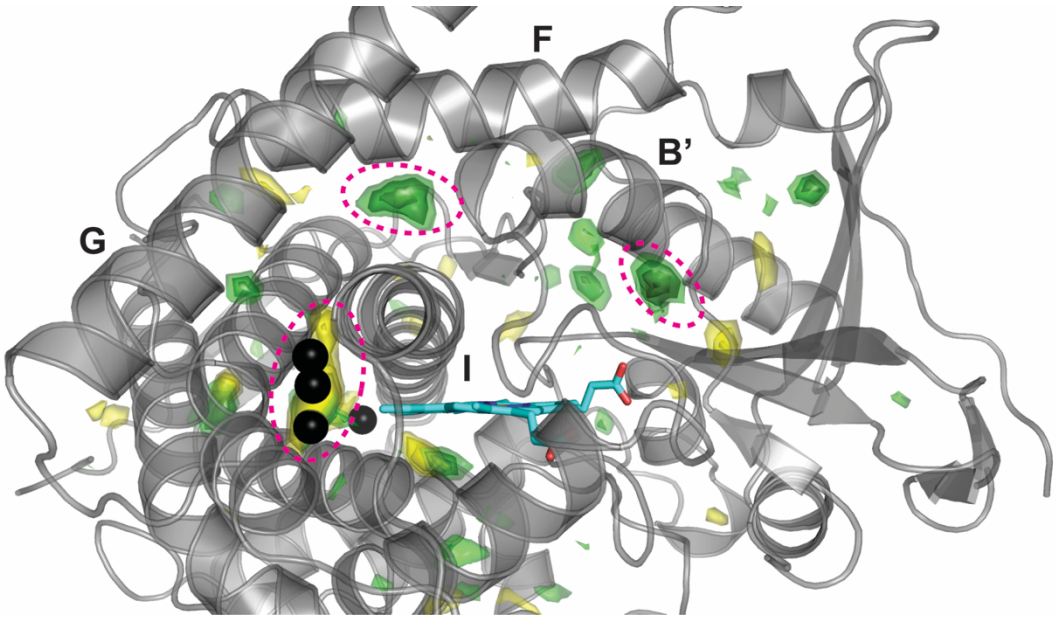


Figure 6-2 Free energy distribution at the core of CYP102A1 and CYP102A5.
Positions of local energy minima for O_2 found by ILS for CYP102A1 (Protein: grey, PDB 2HPD; heme: cyan; isosurface: green) and the overlayed homology model of CYP102A5 (protein not shown; isosurface: yellow). The four shades of isosurfaces represent $\Delta G^0_{,O2}$ of -16.5, -14.0, -11.5 and -9.0 kJ/mol (from darkest to lightest, respectively). The global energy minima, two for CYP102A1 and one for CYP102A5, are indicated by pink dashed outlines. Xe gas molecules crystalized in the overlayed $P450_{cam}$ (PDB 1UYU) are shown in black. O_2, CO and N_2 share the same CYP-specific global energy minima.

CYP102A1 also shows a high probability of finding gas molecules at this position, though its free energy is not as favorable (ΔG^0_{O2} = -12.6 kJ/mol for CYP102A1 and –20.9 kJ/mol for CYP102A5). A further Xe-binding site lies between the ends of the I- and L-helices; O_2 occupies a site at a distance of about 4.5 Å from the Xe position in both enzymes, with similar calculated binding energies (ΔG^0_{O2} = -12.6 kJ/mol for CYP102A1 and –13.9 kJ/mol for CYP102A5). The last crystalized Xe binding site of $P450_{cam}$ superimposes onto secondary structure elements of both homologous enzymes and is hence not further considered in these systems.

Interestingly, CYP102A5 exhibited no local free energy minimum for O_2 binding directly inside the active-site cavity. The nearest are the aforementioned sites proximal to the crystalized Xe of $P450_{cam}$, along a previously speculated O_2 ingress route [5]. The predicted smaller active site of CYP102A5 might favor O_2 storage in this distal pocket. While the

calculated probability of stably binding O_2 close to the reactive heme iron is low, both enzymes exhibit a higher O_2 occupancy inside the protein than in the solvent. The increased local concentration favors transient events such as reactivity. It is thought that the spin state of the heme iron must play a major role in increasing the affinity for gas binding. Overall, our results of O_2 binding to these two microbial CYPs are consistent with the observed binding of Xe to $P450_{cam}$ and with functional requirements of these CYPs.

We confirmed that, in accordance with their different polarity, the O_2 binding sites do not coincide with crystalized water binding sites. To this effect, we aligned all reported crystal structures of CYP102A1 and calculated the mass-weighted water density throughout. This resulted in clear identification of the water channel (W), while no water binding sites were found overlapping with or in proximity to gas binding sites (Figure 6-3).

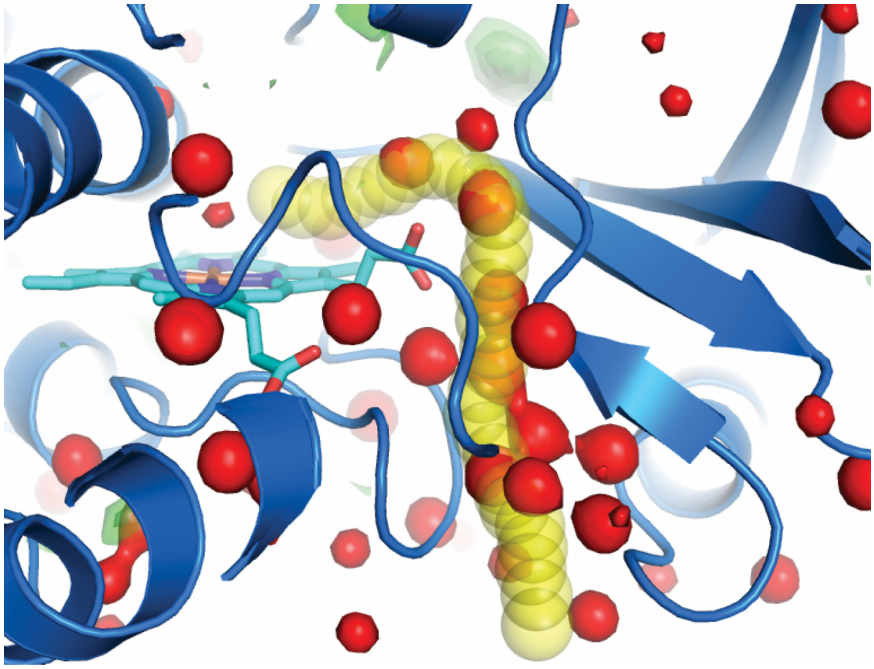


Figure 6-3 Crystallographic water channel of CYP102A1.
Crystallographic water density calculated for the heme-binding domain of all crystal structures of CYP102A1 available in the PDB. The crystallographic water is shown as red spheres and represented on the coordinates of PDB 2HPD (Protein: blue, Heme: cyan). The water channel W identified using CAVER3 is shown in yellow.

This indicates a distinct ordering for water and gas binding within P450 enzymes. A high density of crystallographic water was also found in the active site cavity, consistent with the striking absence of gas binding sites close to the heme iron according to both our ILS result and to Xe co-crystallization.

6.4.4. An exhaustive set of channels revealed by ILS in CYP102A1 and CYP102A5

We developed a 3-D algorithm to identify the lowest energy pathways for gas to travel from the reactive heme center towards the enzyme surface. The start point was above the heme iron and end points were in the bulk solvent, proximal to potential surface outlets. Starting with a high-energy probe, each grid point in the free energy landscape equal to, or below, a predefined energy threshold was visited in an iterative manner until the shortest path was found. Then, the probe energy was reduced until the point where no path to the outlet could be found; the complete paths of the lowest energy were designated as gas migration channels.

CYP102A1 was shown to possess ten potential gas migration channels (Figure 6-4). All but two of the ten channels are known potential substrate or solvent migration channels for P450 enzymes: six were observed in crystal structures of CYP102A and two in structures of other CYPs or by MD simulation yet, importantly, no single crystal structure has allowed identification of all those channels (for a detailed review, refer to [12]). In addition, to those eight, two new channels are proposed: 1a and a putative alternative W channel, W^{alt}. The channels form two major clusters. Channels 1a and 3 lead from the heme to the surface over the I-helix, visiting one of the global free energy minima. All other channels (S, W, W^{alt}, 2a, 2b, 2d, 2f and 5) follow a second path leading from the heme towards the second global free energy minimum identified in CYP102A1. They then separate into further channels at intersections (Figure 6-4).

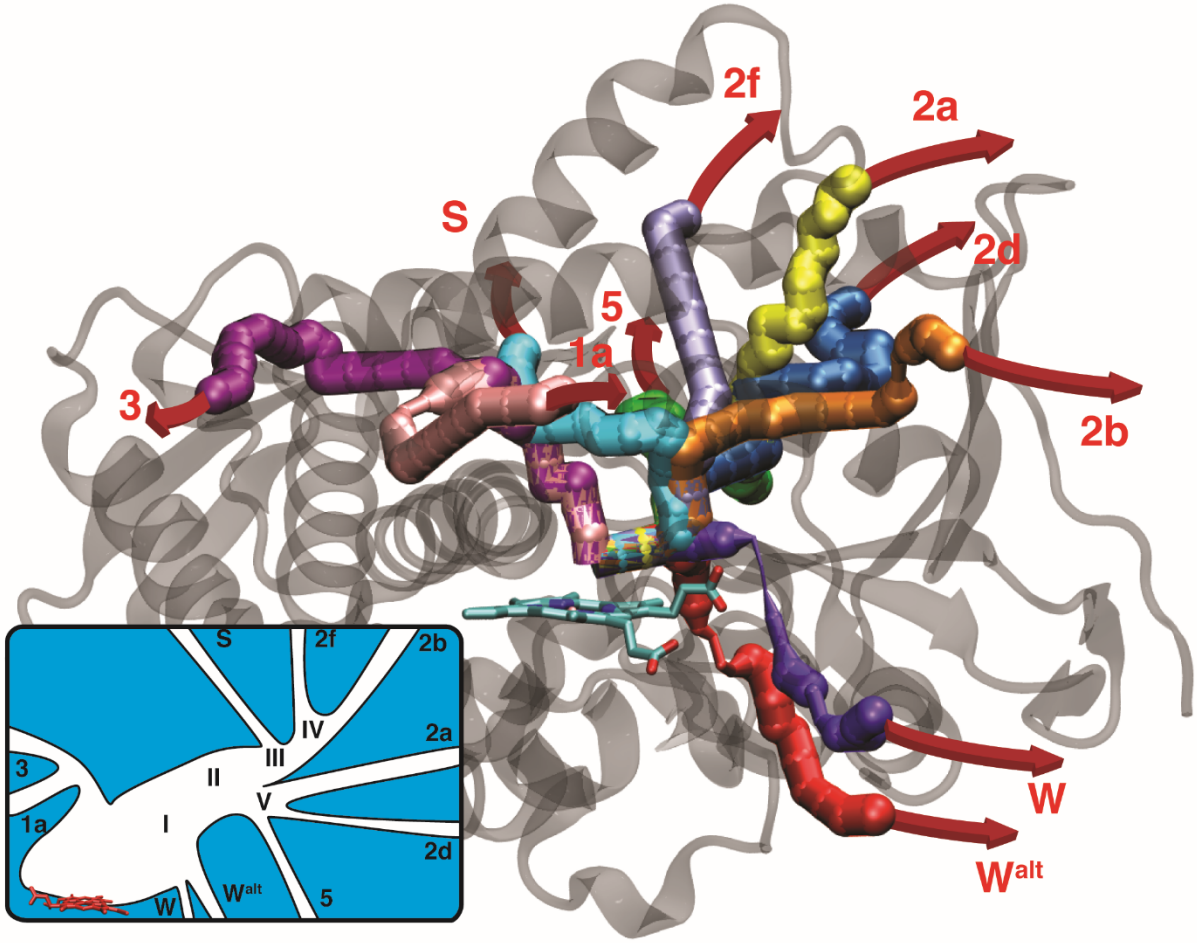


Figure 6-4 Channels mapped in CYP102A1 (PDB 2HPD).
The heme group is in sticks. The nomenclature of channels is based on Cojocaru et al .[12] with our addition of 1a and W^{alt}. The diameter along a tube represents the free energy of O_2: the thinner the tube, the higher the energy barrier for oxygen. Inset: Layout of all channels identified in CYP102A1 using ILS. Roman numerals I-V indicate intersections. Drawing not to scale.

Within any channel, the gas molecules may encounter significant energy barriers ($\Delta G^0_{gas} > 0$ kJ/mol). Mapping the free energy along each channel path thus allows for ranking the channels according to their propensity to conduct gas. This is a function of the channel geometry, including channel length, as well as Lennard-Jones interactions formed with the gas molecule. For CYP102A family members, the broad surface area encompassing channels 2a, 2b, 2d and 2f, known as channel 2 region, is the entrance for a number of substrates. Our calculation of the free energy along each channel shows that the solvent channel S exhibits the

lowest overall free energy landscape for O_2 and is the shortest channel; hence it is the most likely gas channel in CYP102A1 (Figure 6-5).

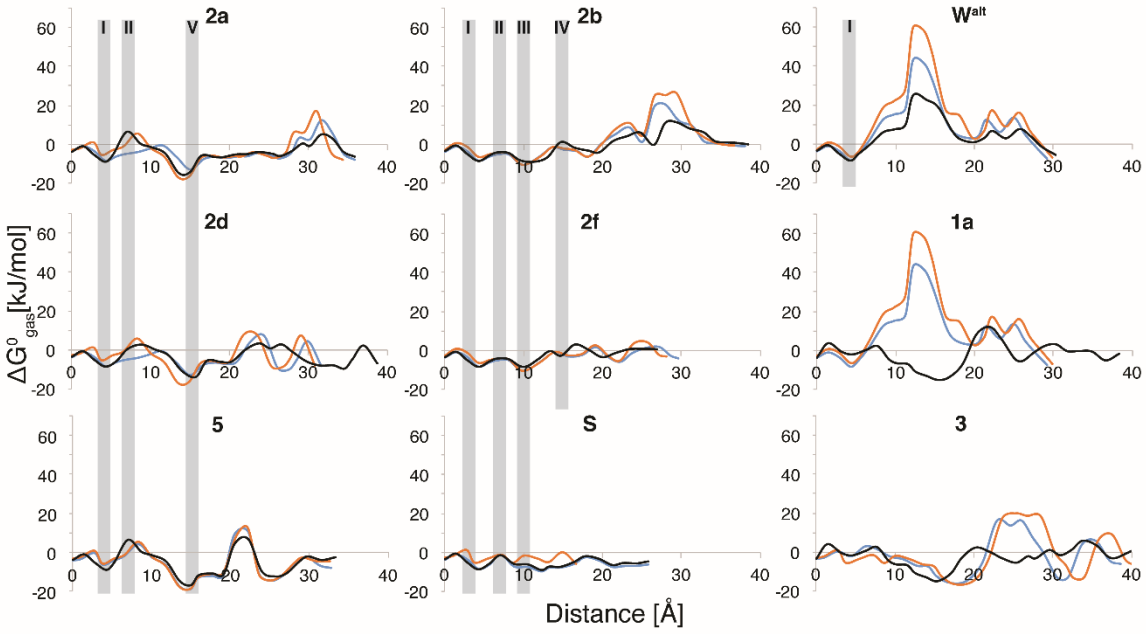


Figure 6-5 Free energy profiles for tunnels in CYP102A1.
Free energy profiles for oxygen (black), carbon monoxide (orange) and nitrogen (blue) movement along the channels of CYP102A1. The distance is the optimal path length with the lowest sum of energies on each grid point excluding grid points outside the cut-off value. Intersections I-V are highlighted in grey and identified by Roman numerals.

Thus despite higher spatial constraints due to the smaller overall diameter of channel S compared to the channel 2 region, it is energetically more favorable for O_2 diffusion. The second most likely channel is 2f, which lies in the highly flexible region of channel 2. All other channels in region 2 have low steric constraints, as evidenced by CAVER3 analysis, but are energetically constrained either as a result of being longer or of dynamic side chains resulting in high-energy barriers. Interestingly, the water channel W has the overall highest energy barrier (ΔG^0_{O2} > 20 kJ/mol) among all channels observed in CYP102A1. This channel was previously identified according to the crystallographic water molecules that populate it. Although CAVER3 analysis suggested low steric constraints in this tunnel, the ILS calculation suggests that access is blocked by K69 participating in a H-bond network with the main-chain oxygen of the specificity-influencing and activity-regulating residue F87, a water molecule

and the heme carboxylates [53]. The location of this network coincides with the high-energy barrier mapped for O_2 diffusion; the passage of O_2 would thus require displacing that tightly-bound water molecule and all others within the W channel.

While the energy barriers through that main W channel (up to 72.2 kJ/mol) were inconsistent with O_2 diffusion, we identified a putative alternative W channel, W^{alt}. It shares the same path as W within the active site core but splits at the last strand of the β1-sheet, running parallel to W on the other side of a β-strand and joining at the outlet. No crystallographic evidence for stabilized water was found for this putative W^{alt} channel in the PDB, hence its identity as a water carrier is uncertain. Nonetheless, its free energy landscape for O_2 migration was more favorable than the main W channel.

The same calculation with CYP102A5 showed similar results, with overall lower energy profiles for channels in the channel 2 region (Figure A 6-3, Figure A 6-4). Although channels 1a and 2b were not predicted, four other channels, the 2ac and newly predicted W^{alt}, 2af1, 2af1 and 3a, were observed. As mentioned above, the channel outlets of region 2 vary more among P450s than do other channels, due to the highly flexibility of that region. This is made evident by the observation of the additional CYP102A5 channels 2af1 and 2af2. Nonetheless, the potential 2b outlet never opened during the three 10-ns MD simulations (including equilibration). This is in agreement with CAVER3 findings of the closure of channel 2b entrance by E355 and E75. Finally, the solvent channel S of CYP102A5 exhibits higher energy barriers for O_2 migration than in CYP102A1, suggesting higher influx of O_2 through substrate migration channels instead.

6.4.5. Accommodation of N_2 and CO in the CYPs

Like all heme-containing proteins, P450s are covalently inhibited by CO gas. CO inhibition can directly influence cellular processes such as drug detoxification, inflammation, respiration, and possibly oxygen sensing [55-58]. Although CO concentration in the air is low, its affinity for heme-iron is high [3]. We hypothesized that P450s have evolved towards exclusion of CO gas. At the opposite end of the spectrum, the abundance of N_2 in the air gives rise to a relatively high concentration in aqueous media despite its low solubility. While it does not react with the heme-iron, N_2 could potentially compete with O_2 and reduce its local concentration. Hence, we also expect evolution towards the exclusion of N_2.

As for O_2, we observe clear differences in the free energy distribution and global minima for N_2 and CO at the core of CYP102A1 and CYP102A5 (Figure 6-5, Figure A 6-4). For the three gases, the solvent channel S in CYP102A1 exhibits the lowest overall free energy landscape and the shortest channel, and hence is the most likely gas migration channel. In CYP102A5, the same trend is seen with an overall lower free energy landscape for O_2 compared to the other gases. Again CO exhibits the lowest affinity, with channels of region 2 being the main gas migration channels.

Since CO gas is produced during metabolism and is thus present in non-negligible local concentrations, CYPs need to prevent CO binding. Strikingly, our results show a high-energy barrier for CO along most channels (Figure 6-5, Figure A 6-4). The free energy for N_2 is generally intermediate, being higher overall than O_2 but mostly lower than CO. Interestingly, all channels except 1a and 3 present the same barriers and local free energy minima for all gases. We note that the energy barrier calculated for N_2 and CO migration in the main W channel was too high to be considered for further analysis of gas diffusion.

6.4.6. Free MD simulation with O_2 reveals gas path usage

In order to provide an independent control for the bias on the free energy calculation introduced by ILS, we performed 20 free MD simulations to observe diffusion of an O_2 molecule from the inside of the protein to the outside. Following the movement of the O_2 molecule over the course of the simulation revealed the preferred channels, and the rate of the gas diffusion through the protein. For the simulation, O_2 was placed proximal to the heme of CYP102A1, at the starting position used in our channel-finding algorithm. Using Fick's second law, we calculated that 4 ns are required for free diffusion of O_2 in water to cover a 40 Å path length, which approximates the length of the longest channel. This allowed us to suppose that a 40 ns simulation time should generally allow egress of the gas. Hence, we accumulated at least 10 ns of simulation time until the molecule reached the protein surface, until a maximum simulation time of 40 ns. Indeed, O_2 reached the bulk solvent in 16 out of 20 simulations, using different channels and egress times (Table A 6-2).

The S channel of CYP102A1 was taken most often and had the fastest egress rate. Importantly, these results of unbiased simulations agree with the ILS calculations, where the S channel was the most likely channel for oxygen migration. When O_2 used other channels, the

gas molecule reached region 2, generally spending a relatively long time (on the order of 10 ns) exploring the channels prior to egress via one of the possible outlets. For channels 2d and 5, the gas molecule resided exclusively within that channel prior to reaching the bulk solvent (Video A 6-1). However, if channel 2f or 2b were used, the gas molecule crossed over to neighboring channels, using transient bridges that formed during the simulation (Video A 6-2). We note that, while channel 2b was populated with O_2 during the simulations, O_2 never exited from that tunnel, consistent with the high-energy barrier observed near its mouth.

A detailed analysis of the trajectories revealed that all channels within region 2 (excluding channel 2d) were dynamically forming and collapsing during the simulation time. In addition, bridges linking channels, or subchannels, were transiently formed. As a result, in some simulations no distinct, static path was followed: O_2 molecules entered one channel and crossed over to another. This is in opposition to all channels outside of region 2, which remained clearly defined throughout the simulations. We identified residues that enable the formation of these transient connections between the main channels of region 2. Residue Q189 is the major gatekeeper between channels 2a and 2f: rearrangement of its backbone and side-chain opened then closed a connecting bridge (Video A 6-2). As a result, we observed the gas molecule switching twice between channels 2a and 2f during three simulations. Another example is residue R47, which transiently trapped O_2 molecules between the entrance of channels 2a and 2b by controlling the bridge between those channels; again, O_2 then chose an exit of lower energy than 2b (Video A 6-2).

We further analyzed the regions of highest occupancy throughout the MD simulation to identify the energy wells where O_2 spent the most time during its egress (Figure 6-6, colored markers). The areas of highest occupancy matched perfectly with free energy wells and barriers in the free energy landscape predicted by ILS. In particular, the channel intersections that are characterized by favorable binding energy and flanked by at least one high-energy barrier were highly populated.

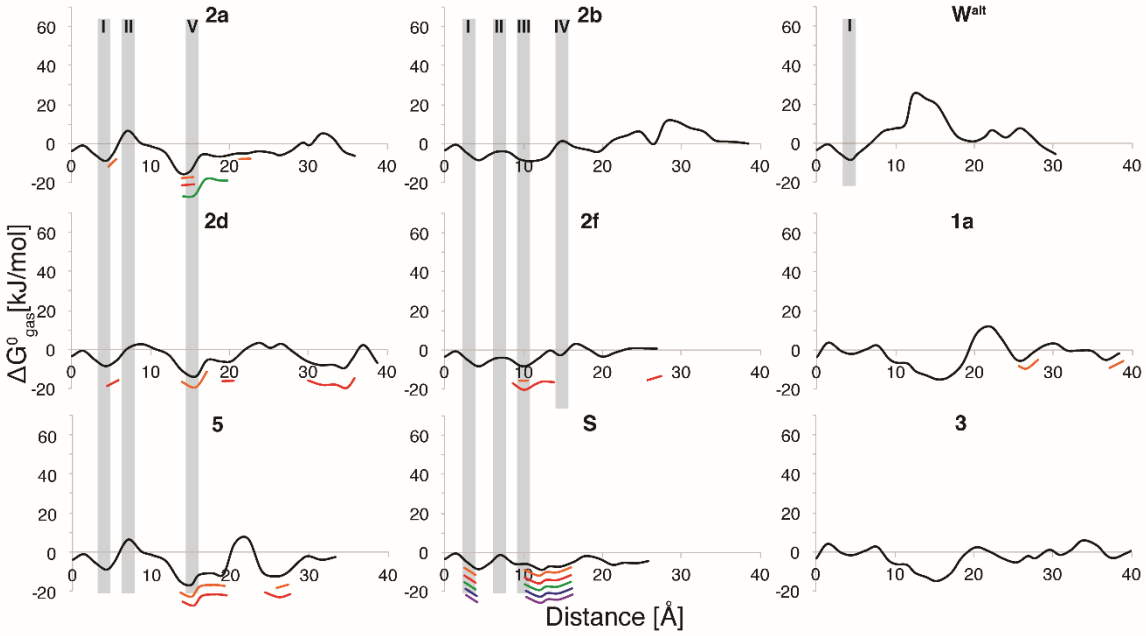


Figure 6-6 O_2 movement along the channels of CYP102A1.
Free energy profiles for O_2 movement along the channels of CYP102A1 (black). The distance is the optimal path length with the lowest sum of energies on each grid point excluding grid points outside the cut-off value. The partial curves represent the regions of highest occupancy of oxygen during the free MD simulation. All colors (except black) stand for an individual MD simulation and represent the region of highest occupancy of O_2 along the path. No energy value is attributed to those curves. Intersections I-V are highlighted in grey and identified by Roman numerals.

Interestingly, we also observed the entry of O_2 from the solvent into CYP102A1 during one simulation. This event mapped the exit of an O_2 molecule from channel 2a followed by re-entry into channel 5, after which it remained in an energy well marked by the high-energy barrier at 23 Å, until the end of the 40 ns simulation. The observation of this event indicates that the simulations were not biased for egress, but allowed free diffusion of the gas.

6.4.7. S^2 order parameters are consistent with flexible channel 2 region

The high plasticity of the CYP102A1 region 2 motivated us to calculate the S^2 order parameter for the CYP102A1 heme domain, thus providing a quantifiable measure of the dynamics of this region. We chose to extract the S^2 order parameter from the auto-correlation

function of the simulated N-H vector as S^2 can be quantitatively compared to NMR dynamics methodologies, and compared for different proteins. We derived the S^2 value for each backbone amide and averaged them over the first 10 ns of the 20 MD simulations (Table A 6-3). The heme domain of CYP102A1 is relatively static, with an average S^2 value of 0.85±0.06 (Figure 6-7).

As expected, unstructured regions and loops show below-average S^2 values and above-average standard deviation. The S^2 order parameter and b-factors describe similar phenomena and are often – but not necessarily – correlated. b-factors are obtained upon crystal structure determination and are subject to crystallization artifacts. Here, the S^2 calculations are mostly in agreement with b-factors in the crystal structure of CYP102A1 heme domain (PDB 2HPD) in the highly dynamic 1-1 and 1-2 strands of the β1-sheet (Figure A 6-5). Both are implicated in the formation of the transient channel 2 region. S^2 values of residues facing the active site tunnel in the B' helix are above average. This helix is known to be important for substrate binding and recognition [50-52]. Similarly, the I-helix is relatively rigid, with S^2 values above 0.9, with the exception of residues I263 and E267 and residues at the beginning of the helix.

The lowest S^2 values were observed in the highly plastic F/G loop and the final turns of its flanking helices (residues 189–200), which define one of the walls of region 2. No S^2 value could be calculated for the loop 427-431, the last residue of the G-helix and the G/H loop, which may indicate that the MD simulation had not converged, or may indicate unusually high dynamics on this timescale. Our calculations show that L437 is the most dynamic residue (S^2_{L437}=0.82±0.04) within a 10 Å radius of the reactive heme, on this fast timescale. This residue is thought to interact with the side-chain of F87 to prevent it from rotating into the substrate-bound conformation [59]. In crystal structures, the L437–F87 distance is too great to allow direct interaction. Our S^2 results suggest a mechanism of dynamic exchange between the crystalline state and a shorter distance between L437 and F87.

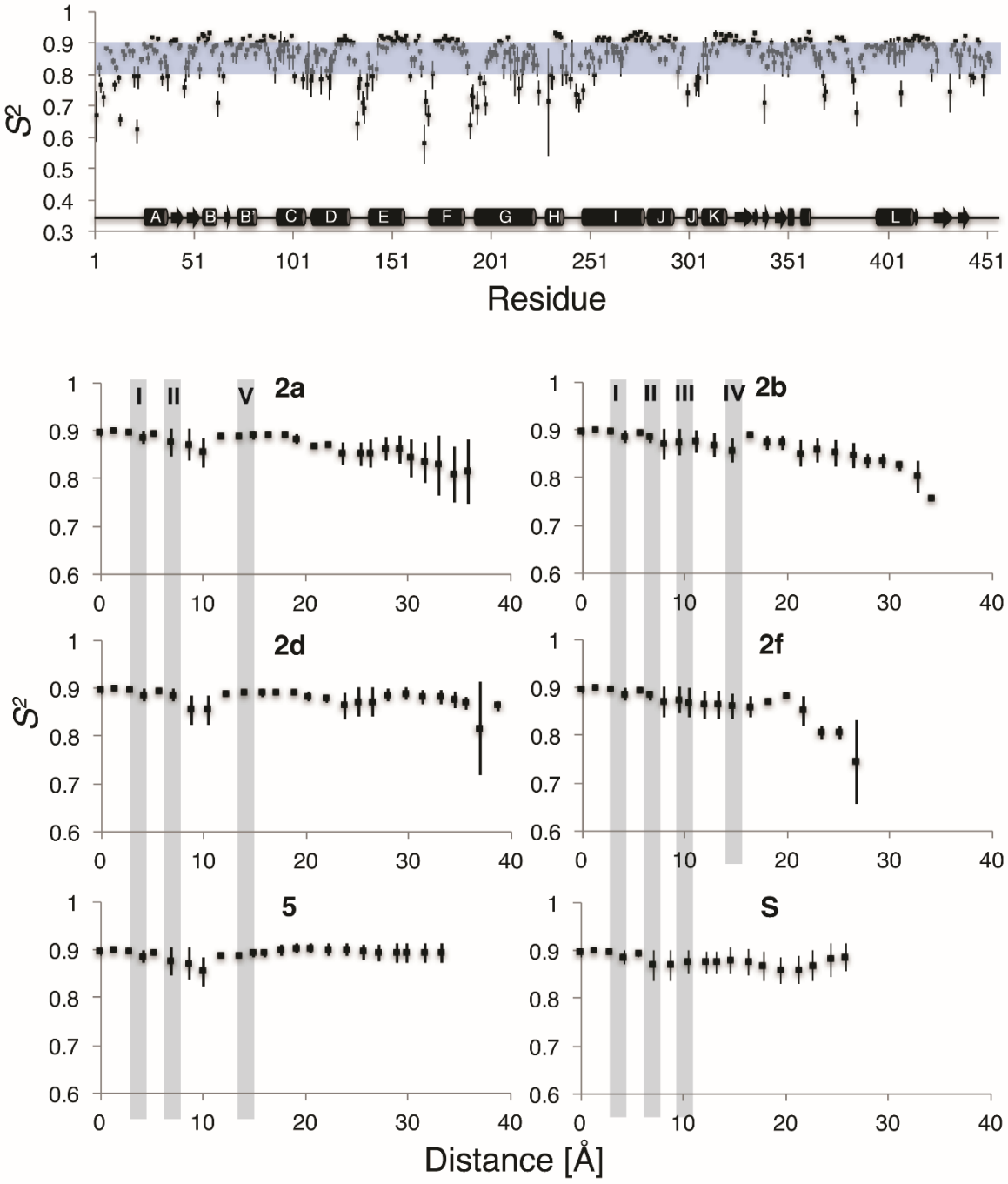


Figure 6-7 S^2 values of tunnels and structure of CYP102A1.

Upper panel: The calculated S^2 values of CYP102A1 according to the amino acid sequence and the secondary structure elements. The blue stripe is centered at the average S^2 and is two standard deviations wide. α-Helices are represented as cylinders and follow the nomenclature of Hasemann et al., whereas the β-sheets are depicted as arrows [50]. Lower panel: Average S^2 values of residues along selected channels of CYP102A1. Intersections I-V are highlighted in grey and identified by Roman numerals.

The aforementioned residues I263 and E267, which are more dynamic than the I-helix they are set in, define the path towards channels 1a and 3, and separate the active site from the global free energy minimum found in the ILS calculation for all three gases. This observation led us to investigate whether the average S^2 of residues that line channels have particularities (Figure 6-7). The solvent channel S, which was predicted by ILS to have the lowest energy barriers to the passage of gas molecules, showed predominantly high S^2 values with small standard deviations (average S^2: 0.88). Similarly, the stable channels 5 and 2d showed uniformly high S^2 values. This is consistent with the fact that, during the course of the simulation, no inter-channel bridges were formed. For all other channels in region 2, a significantly lower average S^2 value was observed near the protein surface, where the transient bridges were observed. For example, bridge-forming residues Q189 and R47 showed an S^2 of 0.82, and residues neighboring Q189 showed S^2 values below 0.8 (as low as 0.73 for residue A191). The backbone dynamics of residues in channels of region 2 further support their importance as the entry port for substrates as opposed to all other channels in CYP102A1. Consistent with previous findings, our S^2 calculations provide further evidence that flexibility is a determinant of the promiscuity for bulky substrates in CYP102A1 [31].

6.5. Conclusion

Comparison of the tunnels in CYP102A1 and its homolog CYP102A5 enhances our understanding of the greater regiospecificity observed during fatty acid hydroxylation by CYP102A5. Our homology model of CYP102A5 is consistent with the hypothesis that the initial binding mode for substrates at the entrance of the channels differs from CYP102A1 [40]: the smaller number and different identity of channel mouths in region 2 of CYP102A5 are consistent with modulation of substrate entry into the active site tunnels. Because the fatty acid carboxylate group remains tethered at the tunnel mouth while the greasy portion slips in for hydroxylation, entry point variation will influence which of the carbons can access the active-site heme for reaction. Furthermore, with a 50% smaller active site cavity, substrate positioning within CYP102A5 is more constrained than in CYP102A1, consistent with the higher regioselectivity of the former. Structural variability at tunnel mouths as an adaptive strategy for substrate binding is not unique to P450s but has previously been determined in a

distinct enzyme system, the haloalkane dehalogenase [15]. We note that the active-site residues (defined as being within 6 Å of the crystalized ω-1 to ω-3 carbons of *N*-palmitoyl-L-Met in CYP102A1 (PDB 1ZO9)) are conserved in both enzymes. Thus, those differences in product formation may result from the altered cavity size due to conformational rearrangement of the F and I-helices.

The calculation of S^2 values for CYP102A1 is, to our knowledge, the first attempt to analyze fast timescale dynamics in this class of enzymes. During the MD simulations and based on the S^2 values, we observed channels opening, merging and oxygen molecules crossing from one channel to the other. The MD simulations, S^2 values and ILS calculations were in agreement, showing that tunnels in region 2 were characterized by greater flexibility than the remainder of the protein. The transient connections observed between region 2 egress channels suggests that, in addition to providing a multitude of interconnecting paths for gas passage, the potential for region 2 channels to merge and collapse renders the region apt to accept substrates with a variety of shapes and chemical properties, consistent with the versatility of P450 enzymes.

In order to experimentally visualize gas-binding sites, co-crystallization of Xe gas and other indirect techniques have been reported[5, 25, 57]. While those models are highly informative, they fail to report directly on potential binding modes of O_2. Using the ILS method, we calculated the Gibbs free energy map in two P450s for three gases using short MD simulations. The consistency of our results illustrates that the ILS method can readily be applied to mapping complex patterns of gas channels in P450 enzymes. The free energy minima we identified for each gas are in agreement with the previously identified Xe gas binding sites, and also reveal additional gas-binding minima. Channels 1a and 3 are the closest to those Xe binding sites and could potentially supply the sites with gas as a result of their favorable gas migration energy profiles. Of particular interest is the clear difference of the energy profile of O_2 compared to N_2 and CO (see Figure 6-5 and Figure A 6-4). In the case of CYP102A1, the major gas-binding site was found within the active site cavity. In contrast, CYP102A5 has a smaller active site and exhibits no high probability gas-binding sites in the active site cavity. The lower abundance of oxygen in the active site of CYP102A5 in addition to spatial constraints may play a role in its higher regiospecificity and may contribute to energetically favor hydroxylation rather than other oxidation reactions such as epoxidation.

The local free energy minima in the core of both enzymes suggest the existence of discrete gas binding sites, or reservoirs. These would increase the local gas concentration compared to bulk gas concentration and thus facilitate the enzymatic reaction. Surprisingly, the minima were not in proximity to the heme iron, yet O_2 must be available for reactivity. This suggests that diffusion of O_2 from the reservoirs towards the reaction center becomes favorable only when the heme-iron attains its activated oxidation state. Analysis of crystallographic water probability reveals that water molecules are highly stabilized within the hydrophilic active-site cavity. Upon binding of a hydrophobic substrate such as a fatty acid, the active site environment becomes more hydrophobic and the heme-iron is reduced to the ferrous state. These events may favor the displacement of heme-coordinated water molecules by O_2 supplied by the gas-binding reservoirs in the tunnel or behind the I-helix, to yield the reactive oxy-complex.

The solvent channel S exhibited a higher affinity for O_2 than all predicted substrate channels in CYP102A1. We speculate that it may provide the main route to re-supply the active site with O_2 after product release. The solvent channel in CYP102A1 is present in substrate-free crystal structures but disappears in the substrate bound structures [60]. The solvent channel could become a target for inhibitor development, to reduce control the O_2 influx to alter the regiospecificity towards the reaction with the lowest activation barrier due to a reduced O_2 concentration.

Little is known about O_2 regulation and its impact on catalysis in P450s. Ost *et al.* found that F393 in CYP102A1 may tune the rate of heme reduction and optimizes the oxygen binding rate [61]. In the ILS calculation, we identified a large local free energy minimum for O_2 in direct proximity of F393, consistent with this area forming an O_2 reservoir. We speculate that mutations of F393 may modify this reservoir and modify the availability of O_2 in the active site.

Strikingly, the ILS free energy landscape of the three gases provided strong evidence that both enzymes evolved towards the preference for O_2 binding. Migration of N_2 and CO exhibit significantly higher energy barriers than O_2. The energy barriers were, in instances such as channel 1a or 3, sufficient to ‘close’ channels to effective gas migration. Most important is the protection against CO, a potent inhibitor of P450 enzymes. In contrast to heme model compounds, where the ratio of binding affinity for CO to O_2 is 20,000, and

myoglobin where it is 25 [62], the ratio for P450s is near 1 as shown by Marchal *et al.* using the F393H mutant of CYP102A1, where CO and O_2 have similar rate constants and the same activation volumes [3]. The increase in CO concentration, for example in the blood of heavy smokers, could potentially surpass the evolved natural protection of P450s against CO and render P450s inactive, which would reduce the efficiency of detoxification [63]. N_2 has lower energy barriers than CO and might additionally protect the active site by competing with CO molecules.

The gas channels identified in this work are clearly distinct from main substrate channels and can hence be engineered orthogonally to the substrate channels. As mentioned above, the availability of oxygen in the active site of P450 enzymes might shift regioselectivity of hydroxylation. The possibility to act on gas molecule supply to the active site without altering substrate discrimination has a great potential for biocatalyst development.

6.6. Materials and methods

6.6.1. Homology modeling of CYP102A5

While the PDB reports over 50 crystal structures of the heme-binding domain of CYP102A1 from *Bacillus megaterium*, there is no structure reported of CYP102A5 from *Bacillus cereus*. A BLAST alignment of the CYP102A5 protein sequence against the PDB allowed identification of the P450 crystal structure with the highest sequence identity: CYP102A1, with 66% identity and 99% coverage (Figure A 6-1). Its crystal structure 2HPD (resolution 2.0 Å) was chosen as a template to generate a homology model of the heme-binding domain of CYP102A5 using the Homology Modeling module of MOE (Molecular Operating Environment version 2013.08) [46]. One thousand models were generated using LowModeMD and ranked using the Generalized Born/Volume Integral (GB/VI) methodology. The intermediate models and the best scoring model were subjected to a fine structure refinement. The final model was verified using MolProbity and PROPKA [64-66].

6.6.2. Identification of tunnels using CAVER3

CAVER3 3.0.1 was used as a PyMOL plugin [67]. Tunnels were calculated using the CYP102A1 structure 2HPD and the homology model of CYP102A5. The minimum radius of the probe spheres was 1.2 Å. The shell depth and radius were set to 4 and 5, respectively. The starting node for the search was 2 Å from the heme iron, perpendicular to the heme in the direction of the channel 2 region. The remaining parameters were set to default values.

6.6.3. CYP102A1 and CYP102A5 MD simulations

The CYP102A1 crystal structure 2HPD and the CYP102A5 homology model were the starting points for simulations using GROMACS 5.0.6 [68]. The protonation states and orientations of histidines were verified using MolProbity and PROPKA [64-66]. For 2HPD, crystallographic water molecules were removed. The proteins were embedded in a TIP3P water box with cubic periodic boundary conditions, keeping a distance of 10 Å between the boundaries and the protein. The net charge of the protein was neutralized with 50 mM NaCl.

For energy minimization and MD simulations, the AMBER99SB-ILDN force field and the heme definition of the non-oxygenated state of Shahrokh *et al.* was used [69, 70]. The heme definition was converted from the native AMBER format to the AMBER format of GROMACS to extend the force field definition. The LINear Constraint Solver (LINCS) algorithm was applied to all bonds containing hydrogen atoms, and a frame was written every 1 ps in a compressed XTC trajectory. In addition, a full precision TRR trajectory was written every 100 ps. The Nose-Hoover thermostat and barostat were used to couple the system to a constant temperature of 304.65 K and a pressure of 1 bar. The electrostatic interactions were evaluated by the particle-mesh Ewald method, and Lennard-Jones interactions were evaluated using a 1.2 nm cutoff. Each system was energy-minimized in triplicate using the Steepest Descent method to a target F_{max} of no greater than 1000 kJ mol^{-1} nm^{-1}. For equilibration, the system was subjected to 2 ns NVT ensemble to gradually heat the system from 10K to 304.65K. Next, a 10 ns NPT ensemble was generated to reach an average pressure of 1 bar. Then, for each enzyme, three 10 ns production trajectories were generated for the ILS calculation.

6.6.4. Implicit ligand sampling (ILS) calculations

The 10 ns MD trajectories of CYP102A1 and CYP102A5 were analyzed with ILS [21, 38]. The methodology calculates a 3D Gibbs free energy grid upon placing a gas molecule from vacuum into a protein. The gas molecule may thus encounter atoms of the system, including protein, ligand or solute. The algorithm assumes that the presence of gas would result in only a weak perturbation to the natural thermal motions of the protein. The ILS analysis tool from VMD was used to calculate the grid [71]. The tool reads force field parameters in the CHARMM format and therefore the non-bonded interaction term of AMBER was converted into the CHARMM file format. To enhance the single core performance of the ILS, calculations were performed on a Tesla K80 GPU cluster (Helios CalculQuebec). During ILS, each frame of the trajectory was divided into 1 $Å^3$ voxels into which O_2, N_2 or CO molecules were translated many times within each grid point. This so-called subsampling of each voxel was performed in a 5×5×5 subgrid and the Boltzmann average calculated over all frames [21, 38]. Three replicates of each enzyme were averaged arithmetically for further analysis. The default O_2 and CO ILS parameters were used. The energy of solvation of each gas was calculated using five randomly chosen positions in the bulk solvent, and compared with the value calculated using Henry's law. The N_2 parameters were fitted for a better agreement with the value calculated using Henry's law. To this effect, the van der Waals (VdW) parameters were iteratively adjusted starting from the default parameters. For each new set of parameters, we performed a new ILS analysis of the CYP102A1 trajectory and calculated the solvation energy at five randomly picked positions in bulk solution. The final VdW parameters used in this study were a VdW radius of 1.85 Å and an epsilon of -0.12 kcal/mol.

6.6.5. Channel search algorithm

In order to identify the tunnel of lowest energy for passage of a given gas, a 3D version of Dijkstra's shortest path algorithm was developed [72]. The 3D Gibbs free energy grid output from ILS was used as input for the algorithm. Each grid point carries the weight of the Gibbs free energy for the respective gas analyzed, reported as arbitrary iso-values. Those iso-values were converted to ΔG for data interpretation.

We converted the ILS grid into a graph by considering all grid points as vertices and by connecting each point to its 6 orthogonal neighbors and its 8 diagonal neighbors. The starting node was approximately 2 Å from the heme iron site. Channel outlet positions were identified based on known channels from CAVER3 analysis, literature and low energy regions at the surface of the enzyme. These outlets were chosen as final nodes. To map the path from the starting node to the final node, any adjacent grid point (in all directions) was an allowed step. The algorithm picked the grid point with the shortest path from the start-point if the weight on the grid was equal to or lower than an iso-value threshold. For the energy grid of O_2, N_2 and CO gas, an initial maximum Gibbs free energy threshold was set at a value equivalent to 40 iso-values. This threshold represents the maximum energy barrier allowed in the shortest path. Similar to the 2D Dijkstra's algorithm, the shortest path was explored in an iterative manner by visiting shortest distance vertices until the total path length became longer than a previous path. This was repeated until the shortest path, with vertices equal to or lower than the threshold, was found. When a path was found, the Gibbs free energy threshold was lowered the value equivalent to 2 iso-value units until no new path was found; this step increased the resolution of the path energy. Each path was further refined by incrementally lowering the iso-value by 0.5 unit, then by 0.1 unit. Each coordinate of the vertices participating in the shortest path found at any maximal threshold for any channel was written as output in a VMD readable format. This method thus identified the Pareto front defined by the shortest low energy tunnels from the starting node (near the heme iron) to various final nodes (channel outlets).

6.6.6. MD simulation of oxygen diffusion

The crystal structure 2HPD (resolution 2.0 Å) was chosen as the starting point for the simulation of O_2 diffusion in CYP102A1. The MD was performed as described above except that a truncated dodecahedron water box was used. One molecule of O_2 was added at the coordinates of the starting node defined in the channel search algorithm. The parameters for O_2 were derived from the GAFF force field [73].

For the minimization and the NVT and NPT ensembles, the O_2 molecule was tethered with a force constant of 1000 kJ mol^{-1} nm^{-2} to prevent it from prematurely diffusing away. Following equilibration, the O_2 restraints in the 3 dimensions were released and 20 individual production simulations were run. A simulation ended if O_2 diffused into the bulk media, or

after a maximum of 40 ns. The trajectory of each run was superimposed with the channels identified in ILS for characterization.

6.6.7. S^2 order parameters

The S^2 order parameters for the peptide bond N-H internuclear vectors were calculated applying the model-free approach of Lipari and Szabo [74] using the toolset of GROMACS 5.0.6 [68]. After removal of translational and rotational motion over each production trajectory obtained for the MD simulation of oxygen diffusion, the autocorrelation function was calculated for each N-H vector over the first 10 ns. The S^2 value was calculated as the average over the second half of the autocorrelation function. Residues that did not converge to a plateau at the second half of this function were excluded from further analysis. The final S^2 value was obtained by averaging over 20 10 ns segments.

Figure legends.

6.7. Acknowledgements

Dedicated to Romas Kazlauskas on the occasion of his 60th birthday. We thank Emilien Huet and Albert Eloyan for help in developing Dijkstra's 3D algorithm. Computational resources were provided by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), NanoQuébec, RMGA and the Fonds de recherche du Québec - Nature et technologies (FRQ-NT). This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants RGPIN 227853 (to JNP) and RGPIN 355789 (to GL). MCCJCE is the recipient of a Vanier Doctoral Scholarship. SD held a DAAD award and AAH held an internship from CGCC.

6.8. References

1. Isin, E.M. and F.P. Guengerich, *Substrate binding to cytochromes P450.* Anal Bioanal Chem, 2008. **392**(6): p. 1019-1030.
2. Conner, K.P., C.M. Woods, and W.M. Atkins, *Interactions of cytochrome P450s with their ligands.* Arch Biochem Biophys, 2011. **507**(1): p. 56-65.

3. Marchal, S., et al., *Formation of transient oxygen complexes of cytochrome p450 BM3 and nitric oxide synthase under high pressure.* Biophys J, 2003. **85**(5): p. 3303-3309.
4. Tetreau, C., et al., *Competition with xenon elicits ligand migration and escape pathways in myoglobin.* Biophys J, 2004. **86**(1 Pt 1): p. 435-447.
5. Wade, R.C., et al., *A survey of active site access channels in cytochromes P450.* J Inorg Biochem, 2004. **98**(7): p. 1175-1182.
6. Ebert, M.C., et al., *Evolution of P450 Monooxygenases toward Formation of Transient Channels and Exclusion of Nonproductive Gases.* ACS Catalysis, 2016. **6**(11): p. 7426-7437.
7. Kozai, D., et al., *Transnitrosylation directs TRPA1 selectivity in N-nitrosamine activators.* Mol Pharmacol, 2014. **85**(1): p. 175-185.
8. Takahashi, N., et al., *Roles of TRPM2 in oxidative stress.* Cell Calcium, 2011. **50**(3): p. 279-287.
9. MacKinnon, R., *Nothing automatic about ion-channel structures.* Nature, 2002. **416**(6878): p. 261-262.
10. Petrone, P.M., et al., *Side-chain recognition and gating in the ribosome exit tunnel.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16549-16554.
11. Yap, M.N. and H.D. Bernstein, *The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel.* Mol Cell, 2009. **34**(2): p. 201-211.
12. Cojocaru, V., P.J. Winn, and R.C. Wade, *The ins and outs of cytochrome P450s.* Biochim Biophys Acta, 2007. **1770**(3): p. 390-401.
13. Ludemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways.* J Mol Biol, 2000. **303**(5): p. 813-830.
14. Ludemann, S.K., V. Lounnas, and R.C. Wade, *How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms.* J Mol Biol, 2000. **303**(5): p. 797-811.
15. Jesenska, A., et al., *Nanosecond time-dependent Stokes shift at the tunnel mouth of haloalkane dehalogenases.* J Am Chem Soc, 2009. **131**(2): p. 494-501.

16. Pavlova, M., et al., *Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate.* Nat Chem Biol, 2009. **5**(10): p. 727-733.
17. Dunn, M.F., et al., *Tryptophan synthase: the workings of a channeling nanomachine.* Trends Biochem Sci, 2008. **33**(6): p. 254-264.
18. Milani, M., et al., *Mycobacterium tuberculosis hemoglobin N displays a protein tunnel suited for O2 diffusion to the heme.* EMBO J, 2001. **20**(15): p. 3902-3909.
19. Orlowski, S. and W. Nowak, *Locally enhanced sampling molecular dynamics study of the dioxygen transport in human cytoglobin.* J Mol Model, 2007. **13**(6-7): p. 715-723.
20. Bocahut, A., et al., *Relating the diffusion of small ligands in human neuroglobin to its structural and mechanical properties.* J Phys Chem B, 2009. **113**(50): p. 16257-16267.
21. Cohen, J., et al., *Imaging the migration pathways for O2, CO, NO, and Xe inside myoglobin.* Biophys J, 2006. **91**(5): p. 1844-1857.
22. Cohen, J. and K. Schulten, *O2 migration pathways are not conserved across proteins of a similar fold.* Biophys J, 2007. **93**(10): p. 3591-3600.
23. Elber, R. and M. Karplus, *Enhanced Sampling in Molecular-Dynamics - Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon-Monoxide Diffusion through Myoglobin.* J Am Chem Soc, 1990. **112**(25): p. 9161-9175.
24. Srajer, V., et al., *Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography.* Science, 1996. **274**(5293): p. 1726-1729.
25. Vallone, B., *Time-resolved crystallography for protein structure: the case of heme proteins.* Rend Lincei-Sci Fis, 2013. **24**: p. S101-S107.
26. Rupenyan, A., J. Commandeur, and M.L. Groot, *CO photodissociation dynamics in cytochrome P450BM3 studied by subpicosecond visible and mid-infrared spectroscopy.* Biochemistry, 2009. **48**(26): p. 6104-6110.
27. Urlacher, V.B., S. Lutz-Wahl, and R.D. Schmid, *Microbial P450 enzymes in biotechnology.* Appl Microbiol Biotechnol, 2004. **64**(3): p. 317-325.
28. Julsing, M.K., et al., *Heme-iron oxygenases: powerful industrial biocatalysts?* Curr Opin Chem Biol, 2008. **12**(2): p. 177-186.
29. Collazo, L. and J.P. Klinman, *Control of the Position of Oxygen Delivery in Soybean Lipoxygenase-1 by Amino Acid Side Chains within a Gas Migration Channel.* J Biol Chem, 2016. **291**(17): p. 9052-9059.

30. Werck-Reichhart, D. and R. Feyereisen, *Cytochromes P450: a success story.* Genome Biol, 2000. **1**(6): p. REVIEWS3003.
31. Otyepka, M., K. Berka, and P. Anzenbacher, *Is There a Relationship Between the Substrate Preferences and Structural Flexibility of Cytochromes P450?* Curr Drug Metab, 2012. **13**(2): p. 130-142.
32. Oprea, T.I., G. Hummer, and A.E. Garcia, *Identification of a functional water channel in cytochrome P450 enzymes.* Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2133-2138.
33. Mouawad, L., et al., *CO migration pathways in cytochrome P450cam studied by molecular dynamics simulations.* Protein Sci, 2007. **16**(5): p. 781-794.
34. Schleinkofer, K., et al., *Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling?* Embo Rep, 2005. **6**(6): p. 584-589.
35. Winn, P.J., et al., *Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5361-5366.
36. Saam, J., et al., *Molecular dioxygen enters the active site of 12/15-lipoxygenase via dynamic oxygen access channels.* Proc Natl Acad Sci U S A, 2007. **104**(33): p. 13319-13324.
37. Shadrina, M.S., A.M. English, and G.H. Peslherbe, *Effective simulations of gas diffusion through kinetically accessible tunnels in multisubunit proteins: O2 pathways and escape routes in T-state deoxyhemoglobin.* J Am Chem Soc, 2012. **134**(27): p. 11177-11184.
38. Cohen, J., K.W. Olsen, and K. Schulten, *Finding gas migration pathways in proteins using implicit ligand sampling.* Methods Enzymol, 2008. **437**: p. 439-457.
39. Yun, C.H., et al., *The bacterial P450 BM3: a prototype for a biocatalyst with human P450 activities.* Trends Biotechnol, 2007. **25**(7): p. 289-298.
40. Chowdhary, P.K., M. Alemseghed, and D.C. Haines, *Cloning, expression and characterization of a fast self-sufficient P450: CYP102A5 from Bacillus cereus.* Arch Biochem Biophys, 2007. **468**(1): p. 32-43.
41. Singh, R., et al., *Enzymatic C(sp3)-H Amination: P450-Catalyzed Conversion of Carbonazidates into Oxazolidinones.* Acs Catal, 2015. **5**(3): p. 1685-1691.

42. Bordeaux, M., V. Tyagi, and R. Fasan, *Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts.* Angew Chem Int Ed Engl, 2015. **54**(6): p. 1744-1748.

43. Dietrich, J.A., et al., *A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3).* Acs Chem Biol, 2009. **4**(4): p. 261-267.

44. Watanabe, Y., et al., *Oxidation of acyclic monoterpenes by P450 BM-3 monooxygenase: influence of the substrate E/Z-isomerism on enzyme chemo- and regioselectivity.* Tetrahedron, 2007. **63**(38): p. 9413-9422.

45. Pendyala, B., et al., *Heterologous expression of CYP102A5 variant from Bacillus cereus CYPPB-1: Validation of model for predicting drug metabolism of human P450 probe substrates.* Appl Microbiol Biotechnol, 2013. **97**(18): p. 8107-8119.

46. Inc., C.C.G., *Molecular Operating Environment (MOE)*. 2016: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

47. Labute, P., *The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area.* J Comput Chem, 2008. **29**(10): p. 1693-1698.

48. Roccatano, D., et al., *Structural and dynamic properties of cytochrome P450 BM-3 in pure water and in a dimethylsulfoxide/water mixture.* Biopolymers, 2005. **78**(5): p. 259-267.

49. Cowart, L.A., J.R. Falck, and J.H. Capdevila, *Structural determinants of active site binding affinity and metabolism by cytochrome P450 BM-3.* Arch Biochem Biophys, 2001. **387**(1): p. 117-124.

50. Hasemann, C.A., et al., *Structure and function of cytochromes P450: a comparative analysis of three crystal structures.* Structure, 1995. **3**(1): p. 41-62.

51. Ravichandran, K.G., et al., *Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's.* Science, 1993. **261**(5122): p. 731-736.

52. Rentmeister, A., et al., *Engineered Bacterial Mimics of Human Drug Metabolizing Enzyme CYP2C9.* Chemcatchem, 2011. **3**(6): p. 1065-1071.

53. Whitehouse, C.J., S.G. Bell, and L.L. Wong, *P450(BM3) (CYP102A1): connecting the dots.* Chem Soc Rev, 2012. **41**(3): p. 1218-1260.

54. Volkamer, A., et al., *DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment.* Bioinformatics, 2012. **28**(15): p. 2074-2075.

55. Volpe, J.A., M.C. O'Toole, and W.S. Caughey, *Quantitative infrared spectroscopy of CO complexes of cytochrome c oxidase, hemoglobin and myoglobin: evidence for one CO per heme.* Biochem Biophys Res Commun, 1975. **62**(1): p. 48-53.

56. Guengerich, F.P., D.P. Ballou, and M.J. Coon, *Purified liver microsomal cytochrome P-450. Electron-accepting properties and oxidation-reduction potential.* J Biol Chem, 1975. **250**(18): p. 7405-7414.

57. Stevenson, T.H., et al., *Kinetics of CO binding to the haem domain of murine inducible nitric oxide synthase: differential effects of haem domain ligands.* Biochem J, 2001. **358**(Pt 1): p. 201-208.

58. Ryter, S.W., et al., *Heme oxygenase/carbon monoxide signaling pathways: regulation and functional significance.* Mol Cell Biochem, 2002. **234-235**(1-2): p. 249-263.

59. Huang, W.C., et al., *Filling a hole in cytochrome P450 BM3 improves substrate binding and catalytic efficiency.* J Mol Biol, 2007. **373**(3): p. 633-651.

60. Haines, D.C., et al., *Pivotal role of water in the mechanism of P450BM-3.* Biochemistry, 2001. **40**(45): p. 13456-13465.

61. Ost, T.W., et al., *Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3.* Biochemistry, 2001. **40**(45): p. 13421-13429.

62. Olson, J.S. and G.N. Phillips, *Myoglobin discriminates between O-2, NO, and CO by electrostatic interactions with the bound ligand.* J Biol Inorg Chem, 1997. **2**(4): p. 544-552.

63. Raub, J.A., et al., *Carbon monoxide poisoning--a public health perspective.* Toxicology, 2000. **145**(1): p. 1-14.

64. Olsson, M.H., et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J Chem Theory Comput, 2011. **7**(2): p. 525-537.

65. Sondergaard, C.R., et al., *Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values.* J Chem Theory Comput, 2011. **7**(7): p. 2284-2295.

66. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
67. Chovancova, E., et al., *CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures.* PLoS Comput Biol, 2012. **8**(10): p. e1002708.
68. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
69. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-1958.
70. Shahrokh, K., et al., *Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle.* J Comput Chem, 2012. **33**(2): p. 119-133.
71. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-38, 27-38.
72. Dijkstra, E.W., *A note on two problems in connexion with graphs.* Numerische mathematik, 1959. **1**(1): p. 269-271.
73. Wang, J., et al., *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-1174.
74. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.* J Am Chem Soc, 1982. **104**(17): p. 4546-4559.

Chapter 7 - Discussion and future work

7.1. Complementary projects

The previous chapters report on the work I have done in the context of my main doctoral project. However, Ph.D. studies give us the unique opportunity during our scientific career to start high-risk side projects, which we would never be able to do in any other context. Here, I briefly highlight two projects that I initiated over the course of my doctoral work.

7.1.1. High-throughput screening assay for cytochrome P450 enzymes

As mentioned in the introduction, one major problem in enzyme engineering is screening variants, for example generated through random mutagenesis, for potential new activities. Many different screening approaches exist, including survival assays and colorimetric assays for the rapid identification of active species. In the case of survival assays, a potential high-throughput selection method could be as simple as plating the mutant library on agar plates that only permit growth of positive hits. However, it is not trivial to couple a specific activity of interest, such as hydroxylation of a fatty acid, with bacterial survival. Colorimetric and fluorimetric assays are more general; they involve following the absorbance or fluorescence of product formation, substrate consumption or cofactor turnover. The primary reaction may be directly monitored or a second, coupled reaction can be monitored. In high-throughput screening assays, each manipulation step between enzyme expression and the actual screening reaction constitutes a potential bottleneck, especially if done manually, which reduces the number of variants that can be screened. Whole-cell screening assays are the method of choice if the substrate can reliably penetrate or diffuse through the bacterial cell wall. After the variants are expressed in 96-well plates, all reactants are added and the reaction can be followed directly through a plate reader or using HPLC-MS equipped with automated sampling. If the reaction cannot take place in the expression media, resting cells or partially lysed cells can be prepared using multiple washing steps. However, these additional steps are

labour intensive and the resulting cells are similar to living cells, such that one needs to ensure that each substrate of interest can reach the enzyme inside the cell. To circumvent this limitation, the cells can be lysed; chemical lysis scales nicely and is highly automatable.

During my doctoral work, I worked mostly with P450 enzymes and was thus highly interested in creating a fully automated high-throughput screening platform to screen P450 enzymes using the Beckman NXP liquid handler available in the Pelletier laboratory. This 8-finger, programmable liquid dispenser is linked to several heating/shaking pods for 96-well plates and is linked to a UV-Vis plate-reader. The major problem to be addressed in order to fully automate this process was cell lysis: the main component of chemical lysis is SDS, which is a substrate for P450 enzymes. Therefore, using chemical lysis buffer containing SDS would mask low-activity mutants and lead to either many false positive or false negative hits. Other chemical lysis buffers include lysozyme, which needs additional washing steps which are not easily achievable using the available infrastructure, as there is no plate-aspiration module and plate centrifugation must be done off-line. To overcome the necessity to lyse the cells, I attempted to identify a system for protein secretion into the media. In a first approach, I used the known secretion leader sequences OmpA and PelB and fused them to the P450 enzyme CYP102A5 [1]. In addition, I fused CYP102A1 (BM3) and CYP102A5 to the *E. coli*-specific transporter gene Ag43, which was previously used to successfully secrete alginate lyases [2] (Figure 7-1 A). The *N*-terminal signal peptide (SP) directs the construct to the cell membrane. The β-subunit integrates into the cell wall and shuttles the P450 passenger to the surface. It is speculated that the signal peptide and the β-subunit are cleaved through autolysis (Figure 7-1 A) [2].

After expression for 24h at 37°C, the OmpA, PelB and Ag43 constructs of CYP102A5 and the Ag43 construct of CYP102A1 all show expression in the lysate but the desired product was not observed in the media (Figure 7-1 B). After concentrating the media of the OmpA and PelB construct 30 ×, CYP102A5 could be identified in the media (Figure 7-1 C). However, the enzyme was too diluted to show any significant activity. The Ag43 construct did not shuttle CYP102A1 or CYP102A5 to the surface. Since the lysate of those constructs showed activity for palmitic acid, it is likely that the P450 enzymes are too large for the β-subunit to carry it through the cell membrane.

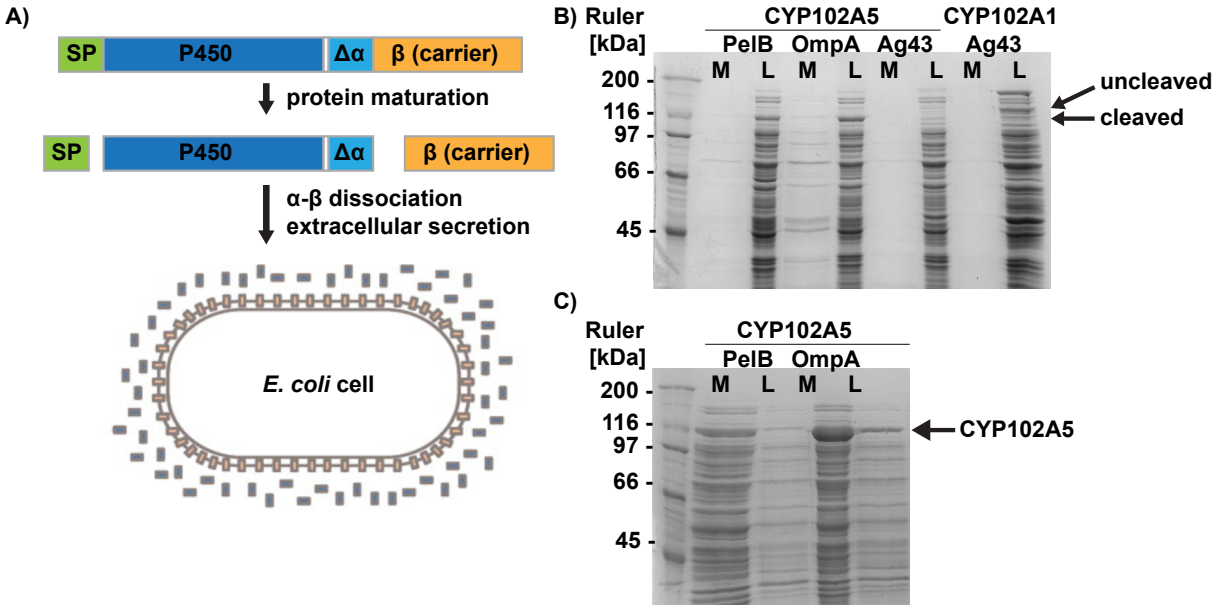


Figure 7-1 *E. coli* secretion system for P450 enzymes.
A) The P450 passenger subunit is shuttled to the cell surface, auto-cleaved during protein maturation and finally released into the media. Adapted from [2]. B) SDS-PAGE gel the different secretion constructs. M: media, L: lysate. The OmpA and PelB constructs of CYP102A5 migrate at 116 kDa. The uncleaved CYP102A1 and CYP102A5 Ag43 constructs in the lysate migrate at 180 kDa and the expected cleaved construct in the media is at 116 kDa. C) SDS-PAGE gel the OmpA and PelB constructs of CYP102A5. The lysate is 30 × concentrated.

Another possibility to lyse cells is programmed cell death through internal lysozyme expression. Pasotti *et al.* developed an autolysis device expressing not only lysozyme to degrade the outer membrane from the inside, but also holins, which integrate into the inner membrane to allow lysozyme to gain access to the peptidoglycan layer [3]. Together with my intern Olivier Rousseau, I co-transformed this autolysis device and a pcWORI vector containing the genes for CYP102A1 or CYP102A5. The enzyme expression of the pcWORI vector is initiated by adding IPTG, while the autolysis device carries a *N*-3-oxohexanoyl-L-homoserine lactone (HSL)-inducible promoter. After overnight enzyme expression induced by IPTG, HSL is added to initiate the programmed cell death. Cell lysis could be observed in most cases over 15 minutes, but never went to completion. Reliability of the lysis device was too low for use in a large library. Pasotti *et al.* described in their publication that internal cell defense mechanisms mutate the autolysis device to protect themselves against the

programmed cell death. Using a low copy vector for the lysis device prevented this phenomenon for Pasotti *et al.* but it was still observed in our laboratory using DH5α and BL21 (DE3) *E. coli* strains. We communicated with Pasotti *et al.* to address this issue without success, and had to abandon this method to release CYP102A1 and CYP102A5 into the media.

The limiting step in a fully automated liquid high-throughput screening assay remains the cell lysis or secretion of the P450 enzyme. Despite a huge time investment, I did not succeed in establishing a protocol to either lyse the cells or secrete the enzyme to have a fully hands-off screening platform. However, our preliminary data using the lysis device provides evidence that theoretically, programmed cell death is a viable option for cell lysis. Ideally a future Ph.D. student can pick up this project to generate a more genetically stable lysis device to complete a robotic high-throughput screening platform.

7.1.2. Quaternary structure determination of CYP102A1

As mentioned in Chapter 4 and Chapter 6, CYP102A1 (BM3) is one of the most widely studied P450s and it has high industrial potential. Despite this fact, the structure of the entire enzyme remains elusive. This self-sufficient P450 enzyme is comprised of three domains: the heme, FMN and reductase domains (Figure 7-2 A). Due to the high molecular weight of the protein complex (120 kDa) and the inter-domain flexibility, self-sufficient P450s are not amenable to NMR structure determination or to X-ray crystallography. Therefore, to determine the active form of self-sufficient P450s and hence, gain a better understanding of the mechanism of electron transfer in those proteins, I started using small-angle X-ray scattering (SAXS) [4]. SAXS is a solution-based technique requiring only minimal sample preparation and material. In crystallography, the samples are exposed to non-physiological environments, which can lead to conformation with no functional relevance. In contrast, SAXS samples are not in crystal form. The native character of this approach is appropriate to characterize systems possessing substantial flexibility. With this technique, it is possible to characterize shape and conformation of proteins in solution at a low resolution. The information content in the scattering curves derived from a SAXS experiment is much lower than in crystallography; however, previously determined structures of the individual domains can be modeled into the scattering data to model the complete conformation of the protein [4].

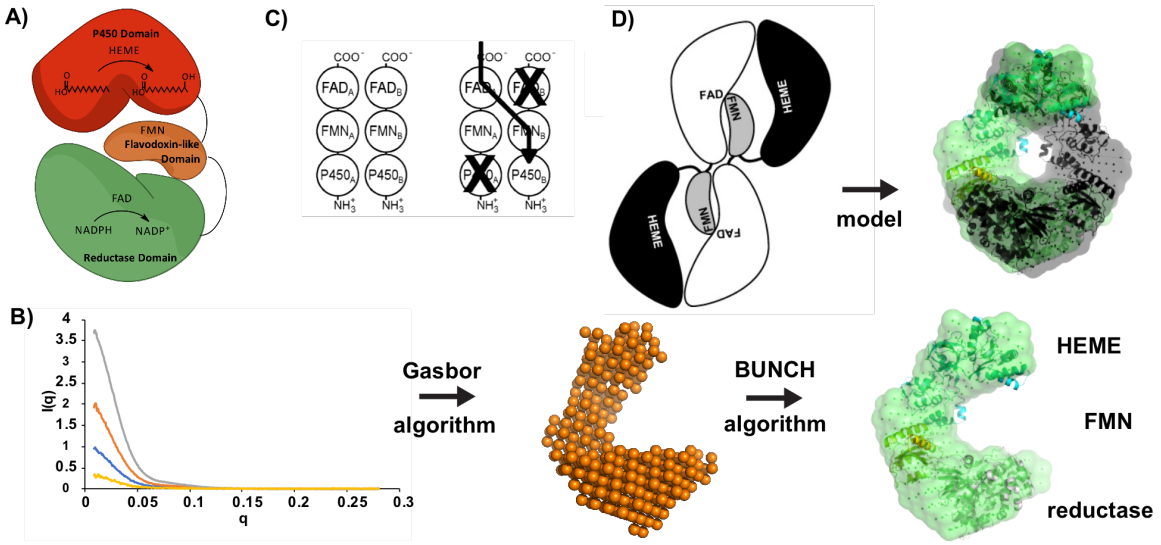


Figure 7-2 Scheme of CYP102A1 and SAXS results.

A) Scheme of the self-sufficient P450 CYP102A1 consituted of the heme, FMN and reductase domains. B) SAXS scattering curves for CYP102A1 at 4 mg/mL (grey), 2 mg/mL (orange), 1 mg/mL (blue) and 0.5 mg/mL (yellow). Using the Gasbor algorithm [5], one can reconstruct the protein envelope based on the SAXS data. The crystal structures of the heme and FMN domain (1BVY) and the reductase domain (4DQK) were used as input for BUNCH [5] to model the multi-domain enzyme against SAXS data. C) The active form of CYP102A1 is a dimer based on the suggested electron pathway. On the left, two intact monomers of CYP102A1 form the active dimer. On the right two individually inactive monomers with either knocked-out heme or reductase domain form an active dimer after being mixed together. Representation reproduced from [6]. D) Scheme of the dimer model reproduced from [6] and our 3D dimer model based on the SAXS data.

The SAXS experiments were done in collaboration with the SAXS and crystallography expert Prof. A. Berghuis with guidance from Dr. Brahm Yachnin in his laboratory (McGill). CYP102A1 is an ideal candidate for SAXS experiments since high-resolution crystal structures of each domain are available and the abundance of prior knowledge of the system. Promising preliminary data acquired at the Cornell University synchrotron led to the first ever reconstruction of the structure of full length CYP102A1 (Figure 7-2 B). The obtained envelope of CYP102A1 using the Gasbor algorithm on the SAXS data was in good agreement of previously published analytical ultracentrifuge data [5, 7]. Therefore, I modeled the multi-domain enzyme using the BUNCH algorithm and published crystal structures of the individual

domains (Figure 7-2 B) [5]. Kitazume *et al.* provided evidence that the dimer of CYP102A1 is the active form of this enzyme. They could show that electron transfer occurs across two individually inactive monomers, where the heme or the reductase domain was knocked-out respectively (Figure 7-2 C) [6]. Based on their assumption, we modeled a 3D structure of the dimer by arranging the monomer derived from SAXS data so that it fits the proposed model of Kitazume *et al.* (Figure 7-2 D). However, some of the atoms from the crystal structures did not fit into the predicted envelope (Figure 7-2 B). Even though the acquisition of the SAXS data was done in the presence of 50 mM dithiothreitol (DTT) to prevent dimerization it appears that some dimer species were still present in solution. Higher molecular weight structures highly influence the SAXS diffraction and hence leave a trace in the envelope. A remaining fraction of the homodimer in the presence of DTT was previously observed and I confirmed it in our sample using an analytical ultracentrifuge [7]. In a further attempt to generate a monomer, I eliminated surface exposed cysteines of CYP102A1 by creating the quadruple mutant C62S/C156S/C773A/C999A. Similar to the wild-type CYP102A1, analytical ultracentrifuge data of the quadruple mutant showed evidence of remaining homodimers in the presence of 50 mM DTT.

Despite promising preliminary results of the reconstruction of the full size CYP102A1, a fully monomeric species is necessary to obtain reliable high quality SAXS data. Further disruption of the protein-protein interactions, for instance by prediction of the protein-protein interfaces using protein-protein docking tools, might allow the creation of a true monomer of CYP102A1. To abandon this project was maybe one of the most disappointing moments of my Ph.D., since I felt very close to determining the structure of this important enzyme. However, the risk and the time investment were becoming too high. This preliminary data hopefully inspires a future Ph.D. student to follow this path.

7.2. Co-existence of experimental and computational tools

As mentioned at the beginning of this thesis, my work is dedicated to building bridges between the fields of biochemistry and bioinformatics. I mainly focused on showing that using the right tools can require courage and leaving one's comfort zone. It is more important than ever to keep an open mind towards new and complex technologies and beliefs. We too often

reject what we do not understand or is outside our belief system. We should deal with these new realities like Girolamo Cardano, who discovered complex numbers around the year 1545 while solving cubic equations. He wrote: "Dismissing mental tortures, and multiplying $5 + \sqrt{-15}$ by $5 - \sqrt{-15}$, we obtain $25 - (-15)$. Therefore the product is 40. … and thus far does arithmetical subtlety go, of which this, the extreme, is, as I have said, so subtle that it is useless" [8]. So even though he was not agreeing with the way he had to solve the equation because he was using a field in mathematics that was not yet fully understood, he was farsighted enough to realize its potential and he recognized that mathematicians had better start dealing with it.

I strongly believe that unless Schrödinger's equation will be solved for multi-body systems and bound states, bioinformatics cannot exist without experimental work to constantly question and review their predictions. At the same time, experimentalists cannot continue their work without taking advantage of the achievements of modern computational tools. The knowledge we, as humans, generate each day outpaces our capacity to follow even the development of our own field of interest. Journals like *Science* and *Nature* are published biweekly and some journals have issues every week. I created a Google alert for publications on my protein of interest CYP102A1 (BM3): every two to three days I received an email with three or more papers. This huge amount of input can only be treated in a reasonable manner using computational algorithms to help us study new questions without repeating other people's successes or mistakes. Smart library design and rational protein engineering becomes more predictable and more powerful as more information is included (see Chapter 2, Chapter 3 and Chapter 4). Not only is data-mining an important target for computational tools but it also equips the experimentalists with methods to answer questions for which no experimental protocol exist (see Chapter 5 and Chapter 6).

As boundaries between scientific fields get blurry, interdisciplinary work is becoming increasingly important. We can even see this happening by creating new fields such as biochemistry which joins biology and chemistry, or physical chemistry, which studies the physical nature of chemical reactions. So the strict separation of the scientific fields might only exist in our minds and is created artificially. We are therefore in need of a new generation of renaissance men and women, who are not experts in all subject areas but speak a common

language to bring the right minds together to create new findings and inspire new fields of science.

7.3. Future work

Aside from the fundamental work presented in this thesis, I applied those methodologies to protein systems such as lipases and other P450 enzymes. Preliminary data shows that the protocols developed here work as well on those systems. Thus, I have created proof-of-concept solutions but would like to finish my other projects to value the time I spent to set them up and the thoughts I put in. Another major aspect is to render those methods more accessible to the larger audience of experimentalists. I have started to develop a web-based interface for the ILS calculations, so that researchers with either limited access to computational resources or without the required expertise to set up such a simulation can benefit from our work.

Even though my Ph.D. studies took nearly five years, they passed swiftly and I thought I would be able to do so much more. Just like with computational simulations, I learned simply that the time available during a researcher's life is invaluable and of short supply.

7.4. References

1. Mergulhao, F.J., D.K. Summers, and G.A. Monteiro, *Recombinant protein secretion in Escherichia coli.* Biotechnol Adv, 2005. **23**(3): p. 177-202.
2. Wargacki, A.J., et al., *An engineered microbial platform for direct biofuel production from brown macroalgae.* Science, 2012. **335**(6066): p. 308-313.
3. Pasotti, L., et al., *Characterization of a synthetic bacterial self-destruction device for programmed cell death and for recombinant proteins release.* J Biol Eng, 2011. **5**: p. 8.
4. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40**(3): p. 191-285.
5. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis.* J Appl Crystallogr, 2012. **45**(Pt 2): p. 342-350.

6. Kitazume, T., et al., *Obligatory intermolecular electron-transfer from FAD to FMN in dimeric P450BM-3*. Biochemistry-Us, 2007. **46**(42): p. 11892-11901.
7. Black, S.D. and S.T. Martin, *Evidence for conformational dynamics and molecular aggregation in cytochrome P450 102 (BM-3)*. Biochemistry-Us, 1994. **33**(40): p. 12056-12062.
8. Cardano, G., T.R. Witmer, and O. Ore, *The Rules of Algebra: Ars Magna*. Vol. 685. 2007: Courier Corporation.

Annex

Annex 1 - Semi-rational exploration of sequence space of the primitive R67 dihydrofolate reductase enzyme

1.1. Tables

Table A 1-1 Mutations introduced combinatorially at the active site of the R67 DHFR protomer(s) in the dimeric library constructs DSM and DDM.

Residue	Native codon	Degenerate codon	No. of codons encoded	No. of amino acids encoded	Amino acids encoded
Val66	GTA	RNT	8	8	D, A, G, V, T, I, S, N
Gln67	CAG	RAK	4	4	D, E, K, N
Ile68	ATT	WNB	24	12	S, T, C, I, F, Y, N, K, R, L, W, M, stop
Tyr69	TAT	YWT	4	4	L, H, F, Y
Leu74	CTT	TTG	1	1	silent mutation
	Combinatorial library		3072	1536	

Table A 1-2 Deduced amino acid sequence of the four randomized amino acids (66-69).
Residues at the active site of the N- and C-terminal protomers of the DDM dimeric R67 DHFR clones prior to selection.

DDM clone	Amino acids 66-69	DDM clone	Amino acids 66-69	DDM clone	Amino acids 66-69
1NS[a]	(N) NNML (C) IKTH	10NS	(N) VKIH (C) TKSY	19NS	(N) NKFH (C) KKSL
2NS	(N) SNTL (C) ADTH	11NS	(N) SEIH (C) DKTH	20NS	(N) SDNL (C) NKYV
3NS	(N) GKTL (C) VNFF	12NS	(N) GKSH (C) NKSH	21NS	(N) NKNH (C) NKSL
4NS	(N) VDIL (C) IDNF	13NS	(N) INYH (C) NNYY	22NS	(N) ADKH (C) SEYL
5NS	(N) NNNL (C) GEKY	14NS	(N) IDKH (C) VISS	23NS	(N) DNSF (C) ANRH
6NS	(N) NNYH (C) GKCY	15NS	(N) NKNL (C) TDNF	24NS	(N) IDNH (C) IDTY
7NS	(N) DDSY (C) TKKY	16NS	(N) INYH (C) NNYY	25NS	(N) ANXY (C) NKNY
8NS	(N) GNYH (C) TNSY	17NS	(N) TK**X**H[b] (C) DNNL		
9NS	(N) TKIH	18NS	(N) IKNH		

	(C) NDYY	(C) NNRH

[a] NS: non-selected clone (or, prior to selection)

[b] X: stop codon

Table A 1-3 Stability of amino acid sequence over three rounds of selection.

Deduced amino acid sequence of the active-site amino acids (66-69) of the N- and C-terminal protomers of the DDM dimeric R67 DHFR clones, through three rounds of selection.

DDM clone, selected	Amino acids 66-69		
	1st round	2nd round	3rd round
1S[a]	(N) VENH	(N) VENH	(N) VENH
	(C) TEIS	(C) **VQIY**[b]	(C) **VQIY**
2S	(N) GENH	(N) GENH	(N) GENH
	(C) LVIF	(C) **VQIY**	(C) **VQIY**
3S	(N) NNFH	(N) NNFH	(N) NNFH
	(C) IENF	(C) IENF	(C) IENF
4S	(N) NNKH	(N) NNKH	(N) NNKH
	(C) SSDI	(C) SSDI	(C) SSDI
5S	(N) NKIL	(N) NKIL	(N) NKIL
	(C) ENAL	(C) **VQIY**	(C) **VQIY**
6S	(N) SEVL	(N) SEVL	(N) SEVL
	(C) KDAF	(C) **VQIY**	(C) **VQIY**
7S	(N) **VQIY**	(N) **VQIY**	(N) **VQIY**
	(C) VKYY	(C) VKYY	(C) VKYY
8S	(N) INIL	(N) INMH	(N) INNF
	(C) NNKH	(C) NDTF	(C) NNRL
9S	(N) NKRH	(N) IDNH	(N) NKRH
	(C) NYIL	(C) DEIY	(C) DEKH
10S	(N) TDMY	(N) IDIY	(N) IDIY
	(C) SEKY	(C) SENY	(C) SENY
11S	(N) AKSH	(N) VKSH	(N) VKSH
	(C) IENY	(C) IEIY	(C) IEIY
12S	(N) NKKH	(N) NKKH	N/A[c]
	(C) **VQIY**	(C) **VQIY**	
13S	(N) NKRH	(N) NKRH	N/A
	(C) LNAL	(C) **VQIY**	
14S	(N) NKLL	(N) NKLL	N/A
	(C) INIH	(C) **VQIY**	
15S	(N) NNNH	(N) NNNH	N/A
	(C) IQIH	(C) **VQIY**	
16S	(N) NKYH	(N) NKYH	N/A
	(C) DNIH	(C) **VQIY**	
17S	(N) NKYH	(N) NKYH	N/A

	(C) **VQIY**	(C) **VQIY**	
18S	(N) NKHL	(N) NKHL	N/A
	(C) **VQIY**	(C) **VQIY**	
19S	(N) INNH	(N) INNH	N/A
	(C) TSNH	(C) **VQIY**	
20S	(N) NKSH	(N) NKSH	N/A
	(C) **VQIY**	(C) **VQIY**	

[a] S: selected clone

[b] **VQIY**: native sequence

[c] N/A**:** Not sequenced

Table A 1-4 Amino acid sequence prior to selection.

Deduced amino acid sequence of the four randomized amino acids (66-69) at the active site of the mutated N-terminal protomer of the DSM dimeric R67 DHFR clones prior to selection.

DSM clone	Amino acids 66-69	DSM clone	Amino acids 66-69
1NS[a]	NKSY	17NS	SD**X**F[b]
2NS	DDTL	18NS	IKML
3NS	NEML	19NS	DKTY
4NS	NNNH	20NS	DKSL
5NS	IDIL	21NS	TEIY
6NS	NDKH	22NS	IEMH
7NS	GELL	23NS	IDWL
8NS	IESL	24NS	VNTY
9NS	DNCL	25NS	DDTH
10NS	SK**X**Y[b]	26NS	IN**X**H[b]
11NS	NESF	27NS	SKKH
12NS	SNTY	28NS	IDYF
13NS	TNIY	29NS	NKYF
14NS	VKSH	30NS	IDCF
15NS	IKSL	31NS	TNYH
16NS	NK**X**Y[b]	32NS	DNSY

[a] NS: non-selected clone (or, prior to selection)

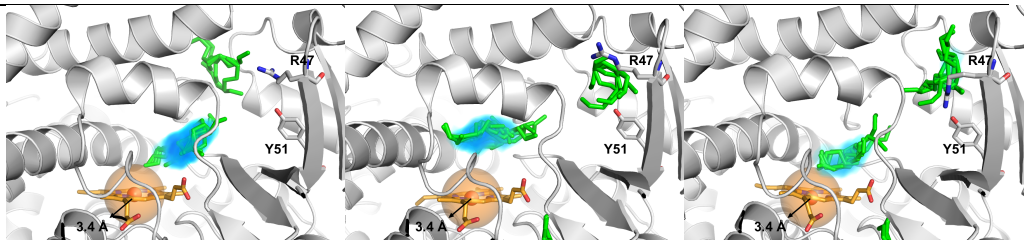
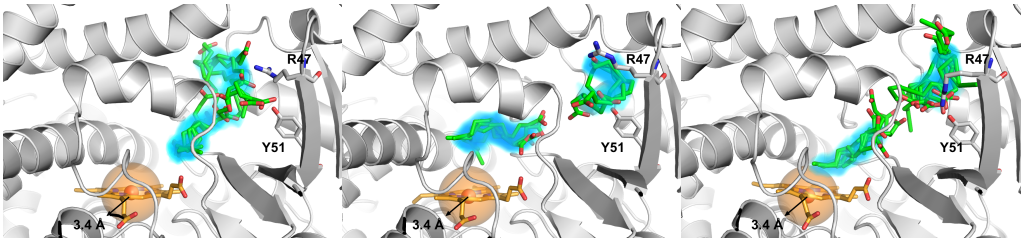
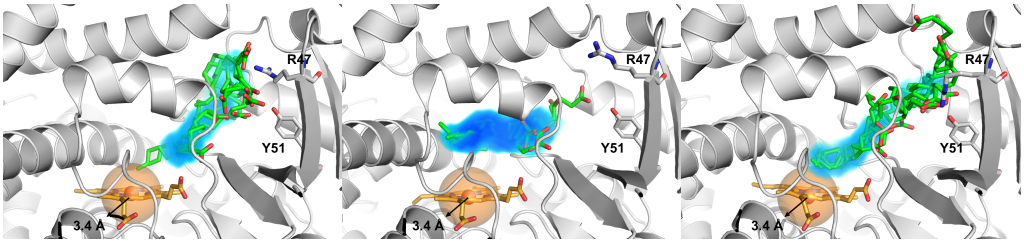
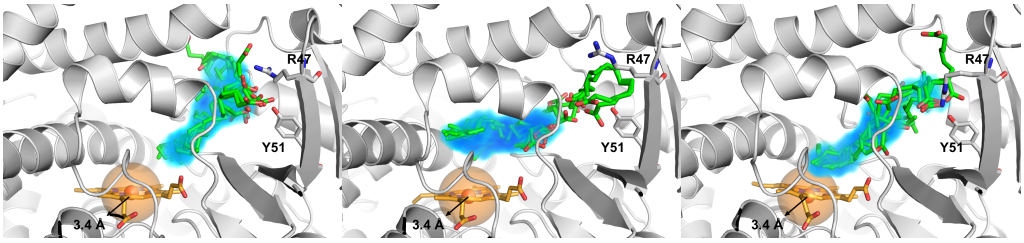
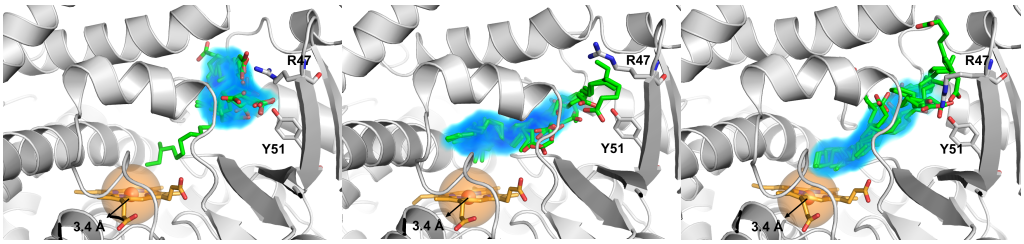
[b] **X**: stop codon

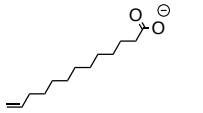
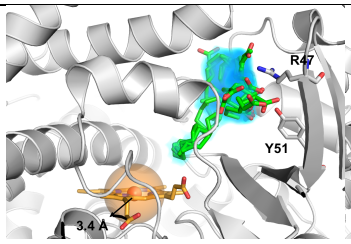
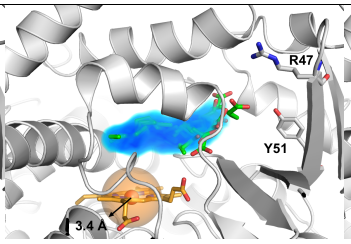
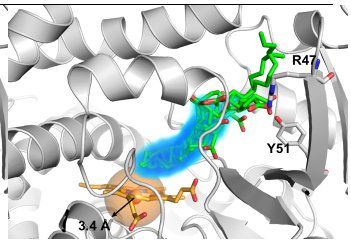
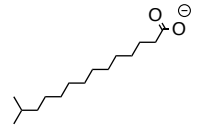
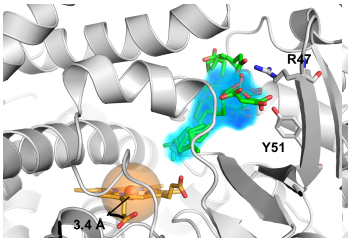
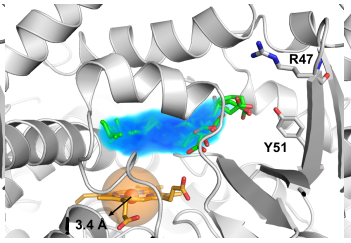
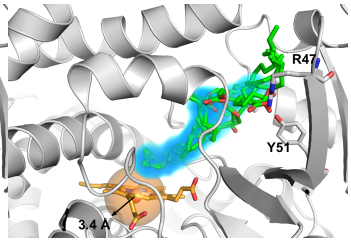
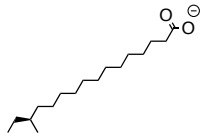
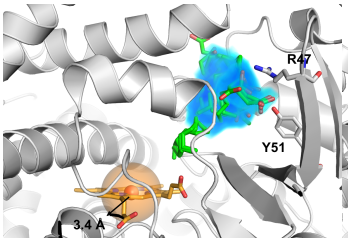
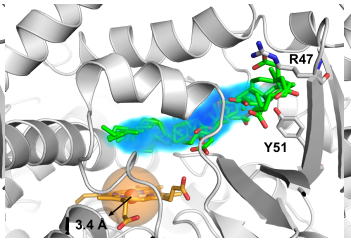
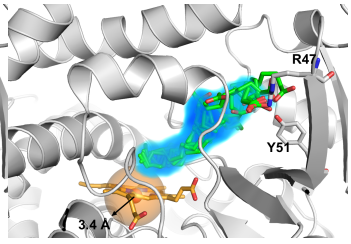
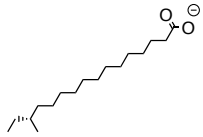
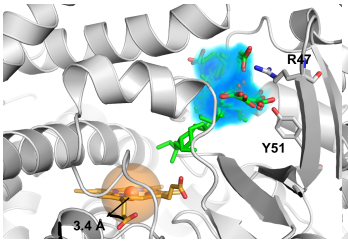
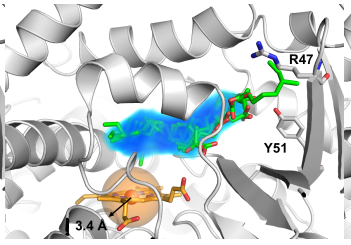
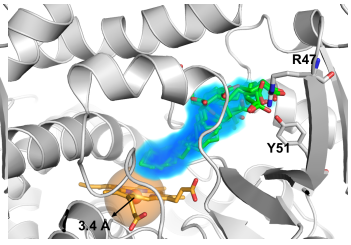
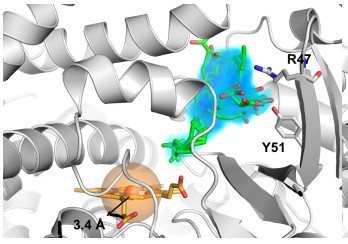
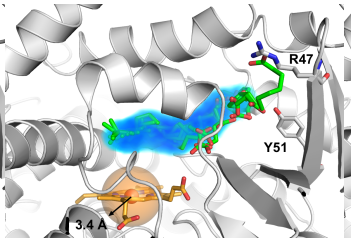
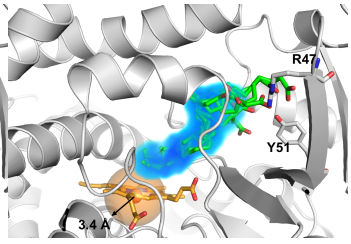
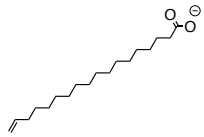
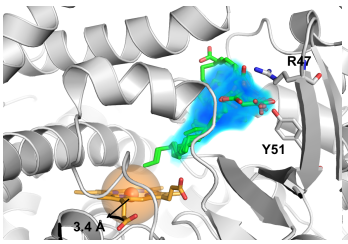
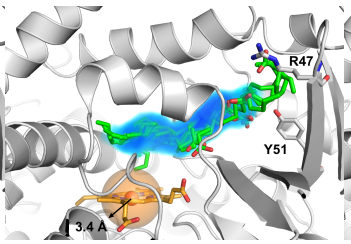
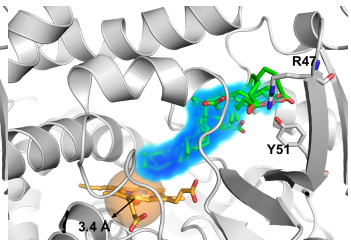
Table A 1-5 Theoretical and observed molecular weight of R67 DHFR and four selected DSM variants.

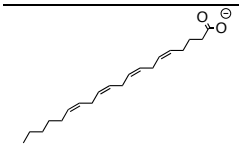
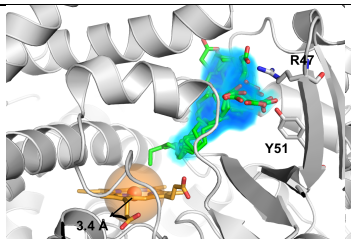
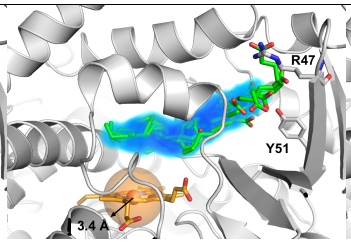
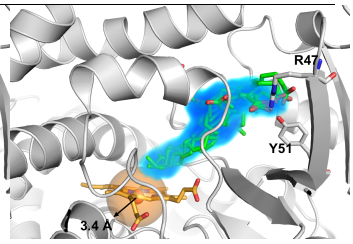
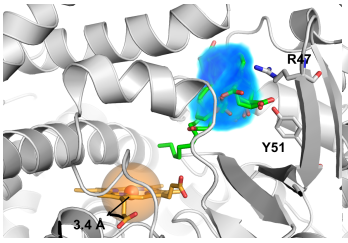
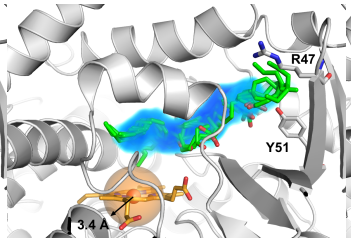
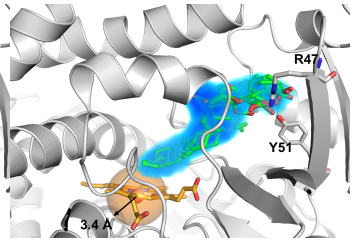
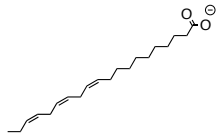
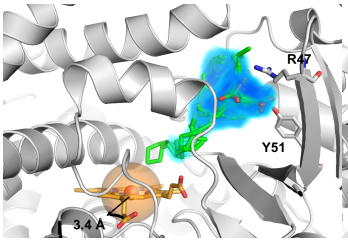
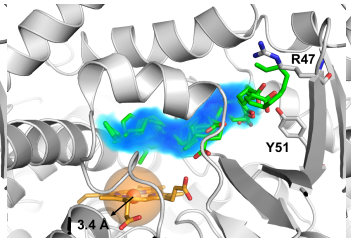
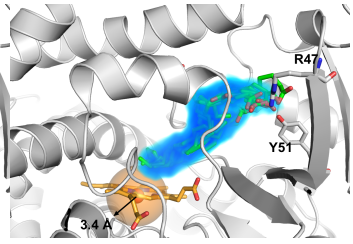
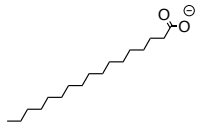
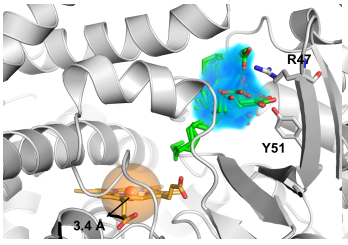
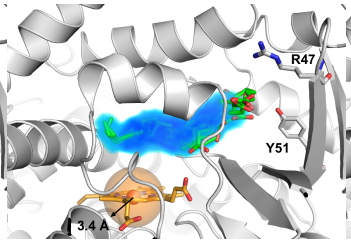
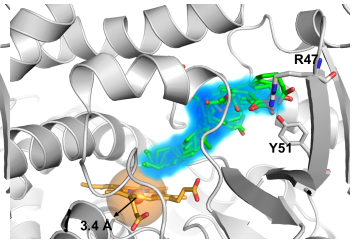
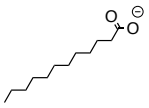
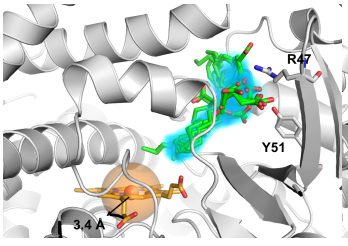
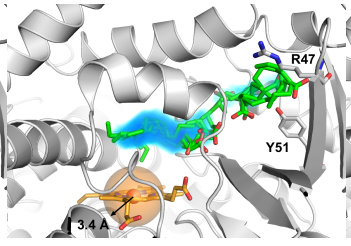
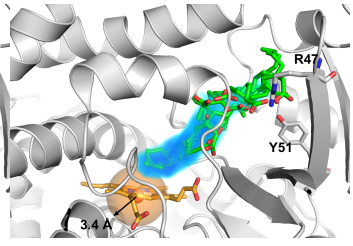
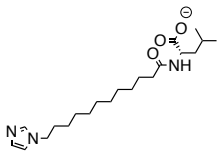
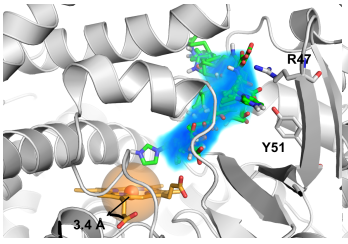
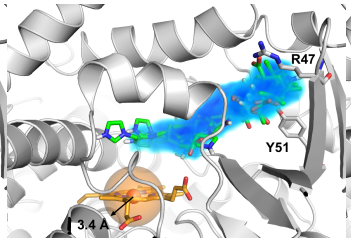
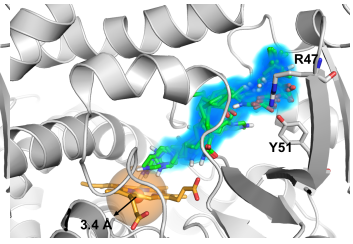
Residues 66-69 in N-terminal protomer	Calculated MW	Observed MW
DNM:		
VQIY (native)	18678.7	18680.8
DSM:		
TENH (3S)	18656.6	18658.9
TEFL (5S)	18665.7	18667.9
TEFY (12S)	18715.7	18717.8
ILYH (13S)	18716.8	18718.3

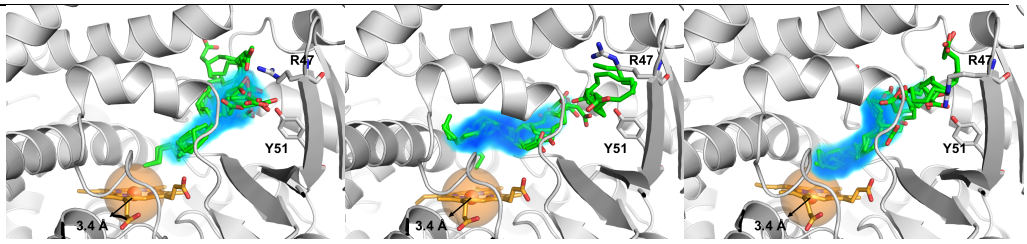
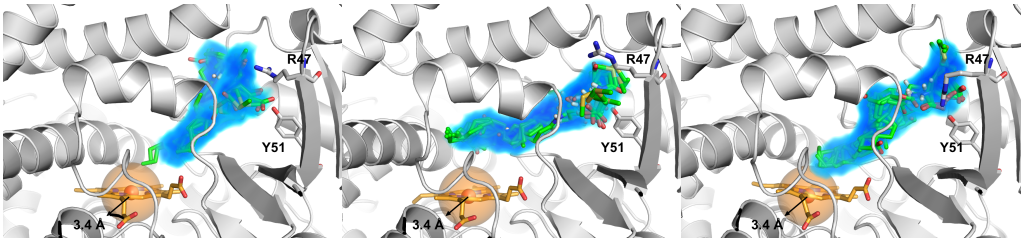
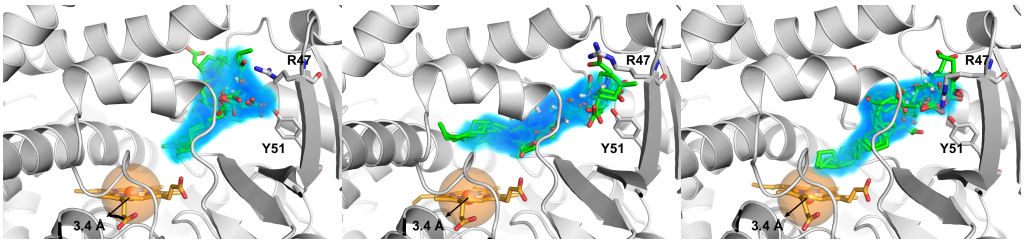
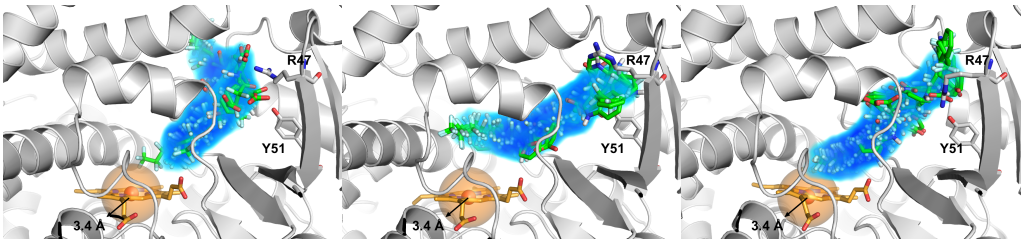
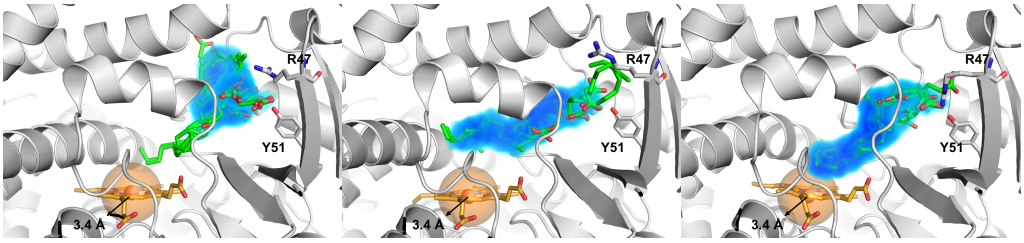
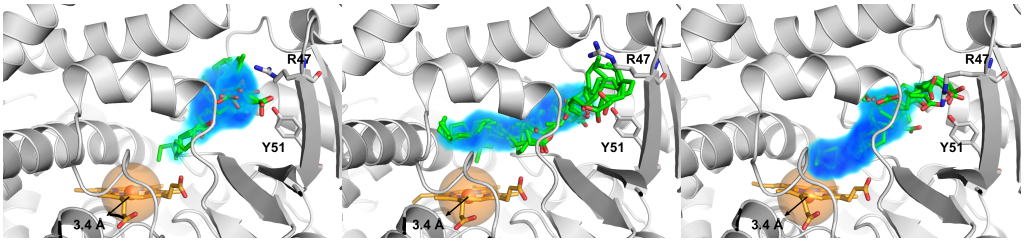
Annex 2 - Prediction of ligand binding to steer rational protein engineering

2.1. Figures

Ligand	BM3 ligand free PDB: 2HPD	BM3 ligand bound PDB: 1ZO9	BM3 ABF M3 minima
1-Octene			
10-Undecynoic acid			
11-Dodecenoic acid			
(+)-12-Methylmyristic acid			
(+)-12-Methylmyristic acid			

Ligand	BM3 ligand free PDB: 2HPD	BM3 ligand bound PDB: 1ZO9	BM3 ABF M3 minima
 12-Tridecenoic acid			
 13-Methylmyristic acid			
 (+)-14-Methylpalmitic acid			
 (+)-14-Methylpalmitic acid			
 15-Methylpalmitic acid			
 17-Octadecenoic acid			

Ligand	BM3 ligand free PDB: 2HPD	BM3 ligand bound PDB: 1ZO9	BM3 ABF M3 minima
 Arachidonic acid			
 Eicosapentaenoic acid			
 Eicosatrienoic acid			
 Heptadecanoic acid			
 Lauric acid			
 *N*-[12-(1H-Imidazol-1-yl)dodecanoyl]-L-leucine			

Ligand	BM3 ligand free PDB: 2HPD	BM3 ligand bound PDB: 1ZO9	BM3 ABF M3 minima
Myristic acid			
N-palmitoyl-L-methionine			
N-palmitoyl glycine			
N-perfluorononanoyl-L-tryptophan			
Stearic acid			
Elaidic acid			

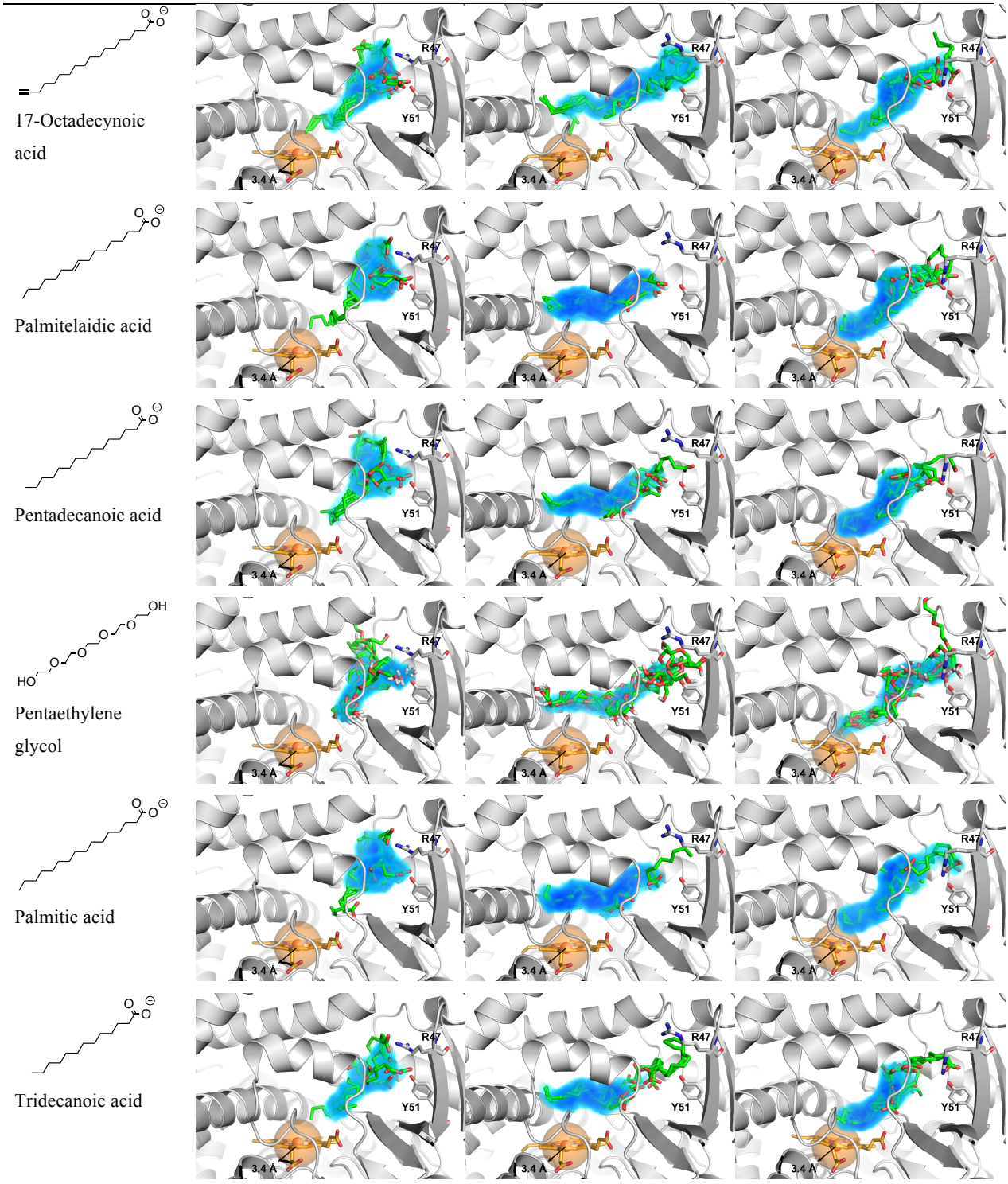


Figure A 2-1 Docking result using the predicted catalytically relevant conformation and remaining substrates.

Docking into the open form of BM3 (2HPD), the closed form of BM3 (1ZO9) and a representative conformation of the M3 ensemble. Individual poses resulting from docking each molecule are overlaid and shown in sticks. The blue volume represents the space occupied by an atom in at least 50% of all poses. The orange sphere represents a volume with a 3.4 Å radius around the heme iron. Key residues R47 and Y51 are displayed in sticks for orientation.

2.2. Videos

Video A 2-1 Free energy landscape of the BM3 / palmitic acid complex formation over the course of the 190 ns ABF simulation.

Free energy landscape defined according to the two reaction coordinates: 1) distance between the heme iron and ω-1 of palmitic acid; 2) distance between the centers of mass of the R47 guanidinium and the substrate carboxylate. Energy minima M1-3 are labelled in orange.

Video A 2-2 Trajectory of the ABF simulation of the BM3 (grey) / palmitic acid (green) complex formation.

The orange sphere represents a volume with a 3.4 Å radius around the heme iron and residue R47 is shown in magenta. Trajectory smoothing window size was set to 5.

Annex 3 - Pushing the boundaries of NMR: Studying protein dynamics in concert with molecular dynamics

3.1. Tables

Table A 3-1 Data collection and refinement statistics for crystal structures of chimera cTEM-19m showing different conformations.

	cTEM-19m	
Data collection	**Crystal 1**	**Crystal 2**
Space group	P 21 21 21	P 21 21 21
Unit-cell parameters (Å)		
a	36.52	36.84
b	58.21	58.76
c	109.50	109.94
Resolution (Å)	1.50	1.15
Completeness (%)	99.9 (99.3)	98.0 (81.4)
Average redundancy	7.2 (7.1)	6.1 (3.9)
I/σ(I)	20.9 (2.2)	23.6 (2.9)
R-merge (%)	6.2 (84.1)	4.1 (37.2)
Refinement	**PDB: 4QY5**	**PDB: 4QY6**
R-work (%)	11.0	10.5
R-free (%)	16.1	13.9
No. of water	317	470
r.m.s. deviations		
Bond lengths (Å)	0.018	0.021
Bond angles (°)	1.8	2.1

Values in the parentheses are for the highest resolution shell.

Table A 3-2 Rotamer adopted by TEM-1, PSE-4 and the chimeras cTEM-2, cTEM-17m and cTEM-19m in the crystal structures and over the 6μs of MD simulation.

The conformers of cTEM-19m crystal struture PDB 4QY5 are indicated for reference.

		TEM-1 (PDB 1XPB)		PSE-4 (PDB1G68)		cTEM-2m (PDB 4MEZ)		cTEM-17m (PDB 4ID4)		cTEM-19m (PDB 4R4R)		cTEM-19m (PDB 4QY5)
residue	rotamer	PDB	MD	PDB	MD	PDB	MD	PDB	MD	PDB	MD	PDB
S70 active site wall												
Met68	['g-', 't', 'g-']	100	20.2	-	-	-	-	100	2.26	-	-	-
	['t', 't', 'g+']	0	**45.61**	-	-	-	-	0	**41.3**	-	-	-
Leu68	['g-', 't']	-	-	100	5.87	100	24.6	-	-	100	23.21	100
	['t', 'g+']	-	-	0	**77.07**	0	**66.77**	-	-	0	**63.76**	0
Met69	['g-', 't', 'g-']	100	**62.92**	-	-	-	-	100	**81.54**	-	-	-
	['g-', 't', 't']	0	19.29	-	-	-	-	0	9.52	-	-	-
Thr69	['g+']	-	-	100	**99.98**	100	**100**	-	-	100	**97.68**	['t'] = 100
Ser70	['g-']	100	**95.8**	100	**57.05**	100	**90.28**	100	**69.41**	100	**51.91**	100
	['g+']	0	2.28	0	40.85	0	5.02	0	21.96	0	47.87	0
Lys73	['g-', 't', 't', 't']	100	**72.59**	100	**87.53**	100	**67.88**	100	**89.1**	100	33.3	100
	['g-', 't', 't', 'g+']	0	2.6	0	5.23	0	3.49	0	7.87	0	**54.98**	0
	['g-', 't', 'g-', 't']	0	23.94	0	2.49	0	19.81	0	0.21	0	0.9	0
Y105 active site wall												
Glu 104	['g+', 't', 'g+']	100	**17.37**	['g-'] = 100	**['g-'] = 47.2**	100	**14.01**	0	**18.65**	40	**9.64**	['g-', 't', 't'] = 100
	['g+', 't', 't']	0	3.26	['g+'] = 0	**['g+'] = 47.35**	0	2.82	100	3.43	60	2.79	0
Tyr105	['t', 'g']	100	21.82	100	39.69	100	18.83	0	16.57	0	22.06	0
	['g-', 'g']	0	**74.16**	0	**55.69**	0	**76.35**	100	**79.73**	100	**70.84**	100
Ser106	['g-']	100	99.99	100	**97.44**	100	**99.78**	100	**100**	100	**97.9**	100
SDN active site wall												
Ser130	['t']	100	21.46	50	22.13	100	6.7	100	15.84	100	14.44	0
	['g-']	0	**78.49**	50	**77.84**	0	**93.04**	0	**83.98**	0	**85.49**	100
Asp131	['t', 'g-']	100	**99.28**	100	**99.59**	100	**98.2**	100	**99.33**	100	**99.37**	100
Asn132	['g-', 'Nt']	100	**64.4**	0	**45.95**	100	**46.84**	100	**74.51**	100	29.39	100
	['g-', 'Og-']	0	17.41	100	36.11	0	27.4	0	21.34	0	**42.42**	0
Ω-loop active site wall												
Glu166	['g-', 'g-', 't']	100	**57.69**	100	**25.96**	100	**37.58**	100	**43.12**	100	3.48	100
	['g-', 'g-', 'g+']	0	26.68	0	11.87	0	31.89	0	23.56	0	1.61	0
	['g-', 't', 'g+']	0	0.25	0	5.26	0	34.8	0	11.85	0	29	0
	['g-', 't', 'g-']	0	2.29	0	14.31	0	24.57	0	14.67	0	**60.32**	0
Leu169	['g+', 'g+']	100	**89.99**	100	39.11	100	**67.67**	100	**64.49**	100	3.31	100
	['t', 'g+']	0	0.19	0	**48.41**	0	22.58	0	19.85	0	**78.18**	0
	['g-', 'Og-']	100	**76.43**	100	**64.23**	100	**67.27**	100	**58.39**	100	6.32	100
Asn170	['g-', 'Ng+']	0	21.92	0	7.14	0	12.58	0	18.77	0	2.08	0
	['g-', 'Ng-']	0	0.12	0	2.73	0	0.69	0	7.08	0	**43.7**	0
214-218 active site wall												
Glu212	['t', 't', 'g-']	100	**23.23**	['t'] = 100	['t'] = 28.15	0	16.84	0	14.07	0	12.25	0
	['t', 't', 'g+']	0	15.81	['g+'] = 0	**['g+'] = 46.92**	0	16.33	0	20.64	0	**26.79**	0
	['t', 't', 't']	0	19.8	['g-'] = 0	['g-'] = 24.93	100	**24.26**	49	**24.66**	55	24.03	100
Ala213	-	-	-	-	-	-	-	-	-	-	-	-
Asp214	['t', 't']	100	3.17	['t', 'Nt'] = 100	**['t', 'Nt'] = 66.23**	100	1.1	0	1.58	0	1.47	0
	['g-', 'g+']	0	1.66	['t', 'Og-'] = 0	['t', 'Og-'] = 33.13	0	2.84	0	11.25	0	12.69	0
	['g-', 't']	0	0.04	-	-	0	0.05	100	3.12	100	2.84	100
Lys215	['g-', 'g-', 't', 'g-']	100	2.86	['g-', 'g-', 'Og-'] =100	['g-', 'g-', 'Og-'] =18.98	0	4.96	0	13.49	0	3.75	0
	['g-', 't', 't', 'g+']	0	6.15			100	7.56	45	**16.64**	0	9.32	0
Val216	['g-']	0	**72.76**	100	**95.98**	100	**78.29**	0	**75.05**	0	**68.7**	0
	['t']	0	17.43	0	3.42	0	12.52	100	12.77	100	12	100
Ala217	-	-	-	['g+'] = 100	**['g+'] = 99.75**	-	-	-	-	-	-	-
Gly218	-	-	-	-	-	-	-	-	-	-	-	-
Pro219	['g+']	100	47.58	['g+', 'Nt'] = 100	['g+', 'Nt'] = 20.36	100	**58.18**	100	**76.12**	100	**77.77**	100
	['g-']	0	**52.42**	['g+', 'Og-'] = 0	**['g+', 'Og-'] = 35.62**	0	41.82	0	23.88	0	22.23	0
234-244 active site wall												
Lys234	['t', 'g+', 't', 'g+']	100	6.44	['t', 'g+', 'g-', 'g-'] = 100	**['t', 'g+', 'g-', 'g-'] = 49.56**	100	13.43	100	34.03	100	43.95	100
	['t', 'g+', 't', 't']	0	8.64	['t', 't', 'g-', 'g-'] = 0	['t', 't', 'g-', 'g-'] = 10.49	0	**25.84**	0	**40.45**	0	**49.84**	0
Ser235	['g+']	100	**73.17**	100	**98.26**	100	**51.24**	100	**46.37**	100	24.37	100
	['g-']	0	22.6	0	1.32	0	45.05	0	44.73	0	**61.93**	0
Gly236	-	-	-	-	-	-	-	-	-	-	-	-
Ala237	-	-	-	-	-	-	-	-	-	-	-	-
Gly238	-	-	-	-	-	-	-	-	-	-	-	-
Glu240	['g-', 't', 't']	100	0.44			0	0.44	0	1.31	32	0.74	0
	['g-', 't', 'g+']	0	0.96		Gly	77.5	0.42	53	4.2	42	2.47	54
	['g-', 'g-', 'g-']	0	**41.15**			22.5	**43.81**	0	**22.19**	25	7.26	46
Arg241	['g-', 't', 't', 'g-']	100	1.96	['g-', 'g'] = 100	**58.94**	0	1.25	0	1.48	0	3.09	0
	['g-', 't', 't', 'g+']	0	4.73	['t', 't'] = 0	23.29	100	1.72	100	3.69	100	2.04	100
Gly242	-	-	-	-	-	-	-	-	-	-	-	-
Ser243	['g-']	100	**98.9**			0	**93.31**	55	**98.81**	57	**90.75**	48
	['g+']	0	1.06		Ala	100	2.64	45	1.13	43	4.79	52
Arg244	['t', 'g+', 'g-', 't']	100	0	0	0.05	0	0	0	0	0	0	0
	['t', 'g+', 't', 'g+']	0	14.72	0	**30.22**	100	2.88	100	9.21	100	1.16	100
	['t', 't', 't', 't']	0	**44.6**	0	5.44	0	**21.21**	0	**14.52**	0	**16.67**	0

3.2. Figures

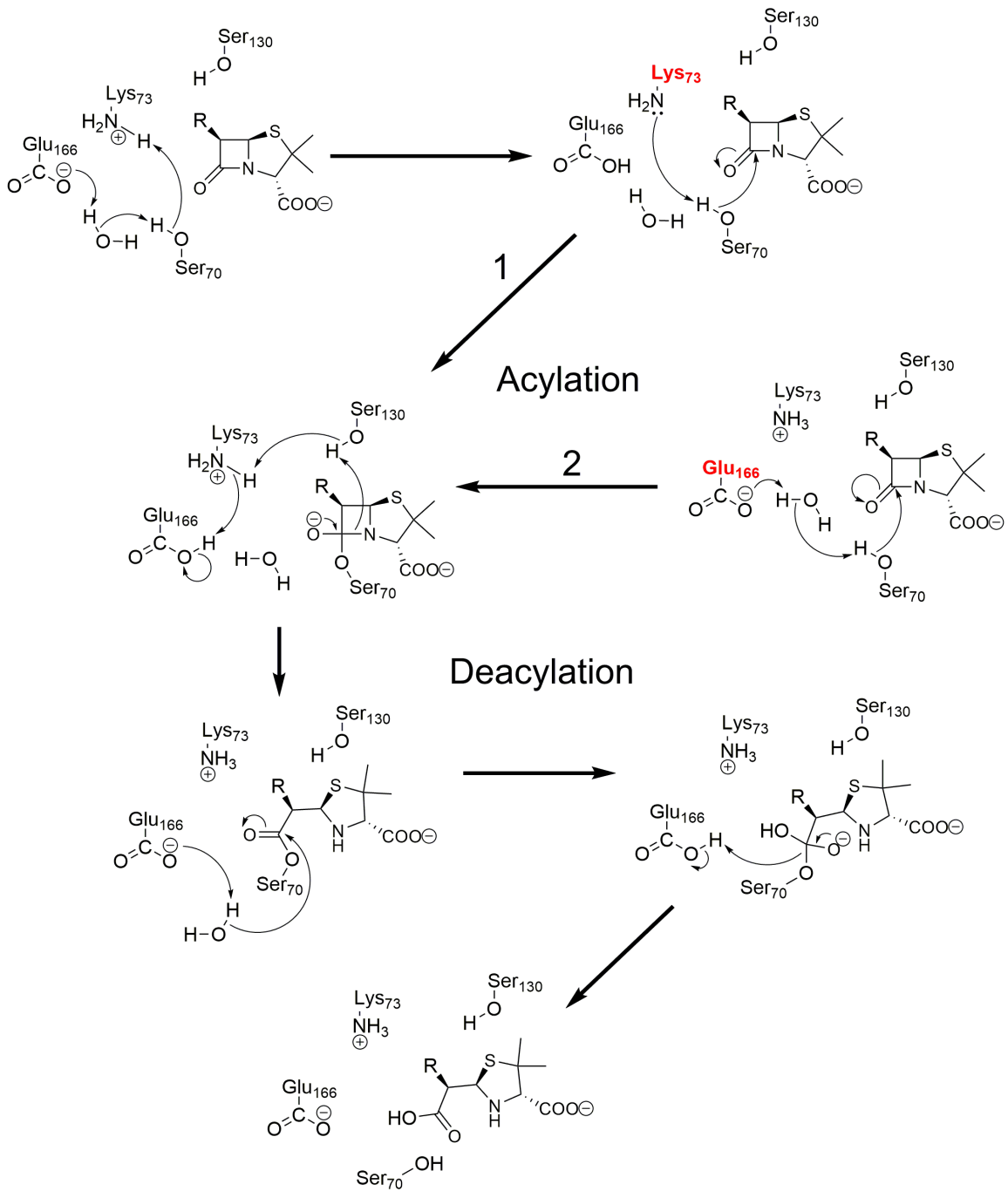


Figure A 3-1 Catalytic mechanism of Class A β-lactamases.
Acylation starts by activation of Ser70 either through Lys73 (1) or Glu166 (2). The tetrahedral intermediate is then reorganized to form the acyl-enzyme intermediate. Deacylation through Glu166 and a conserved water molecule releases the hydrolyzed substrate and regenerates the enzyme [1-3].

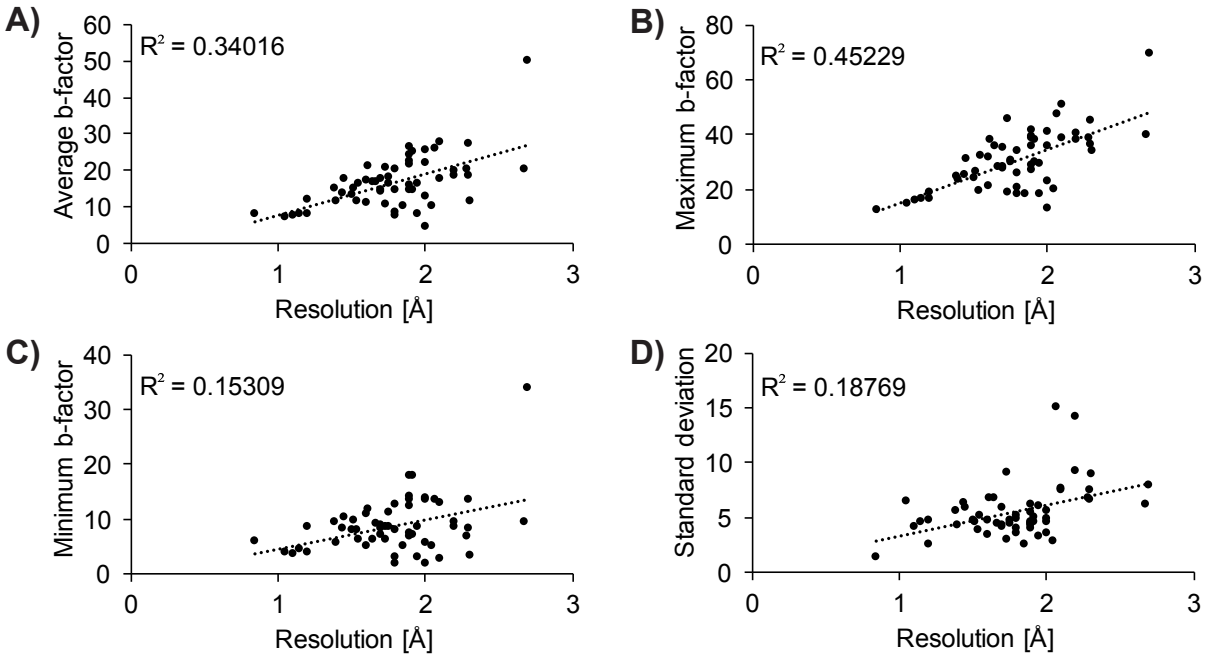


Figure A 3-2 Correlation between the resolution and A) average, B) maximum, C) minimum and D) standard deviation of the b-factors of Uniprot entries P62593 and P16897

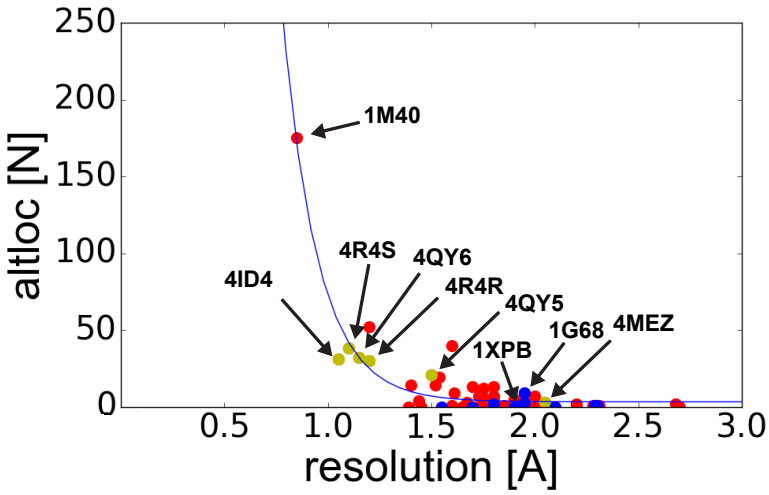
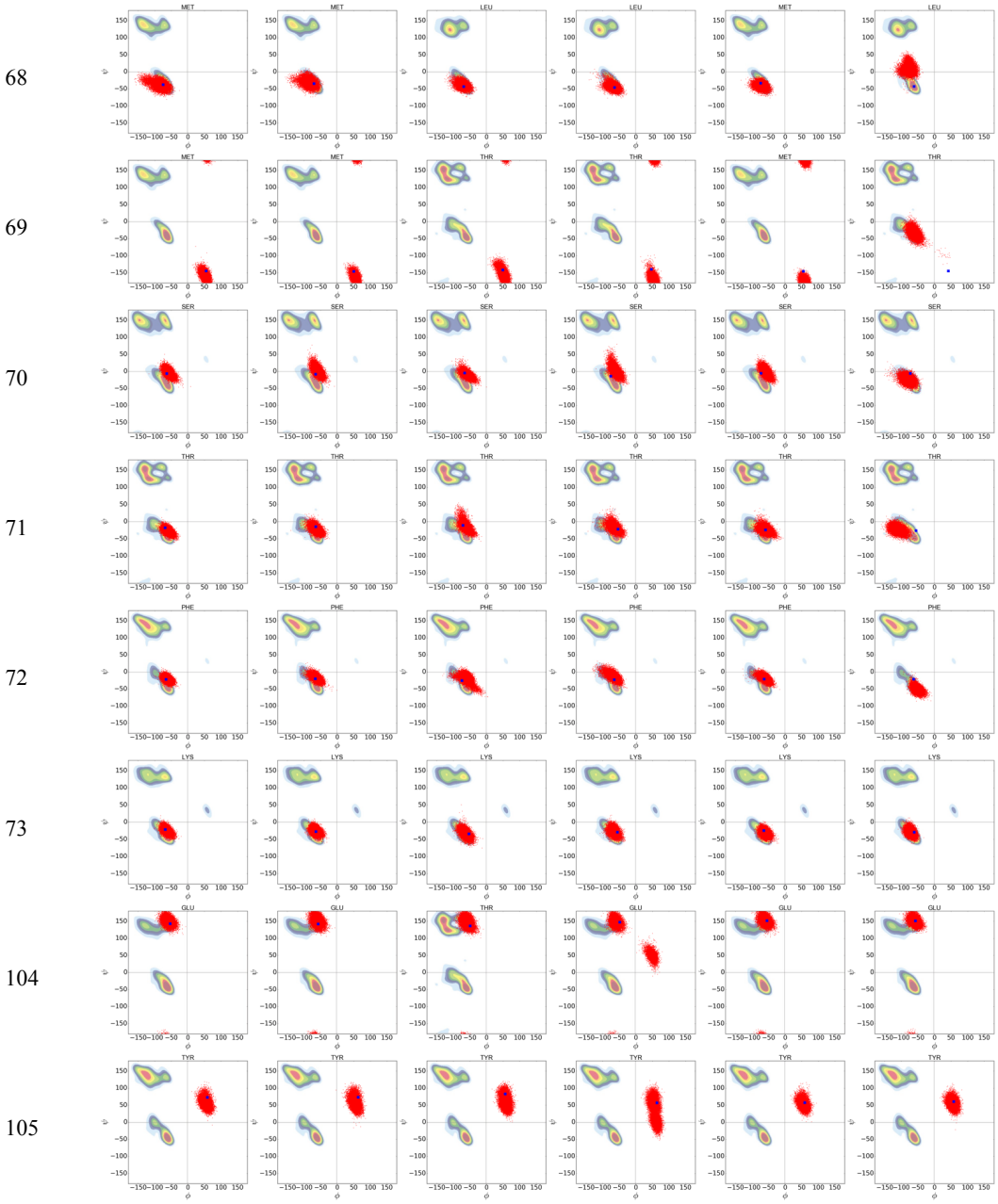
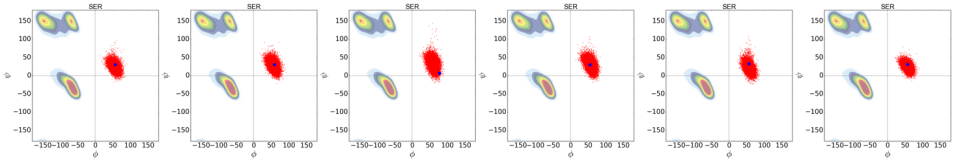
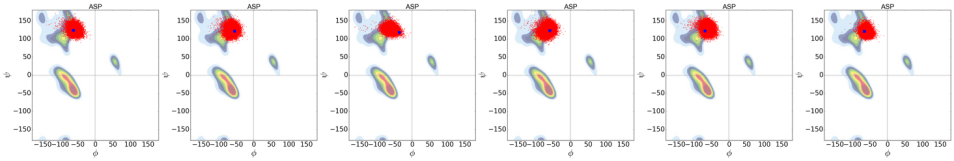
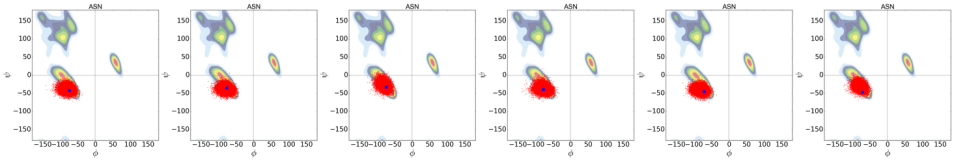
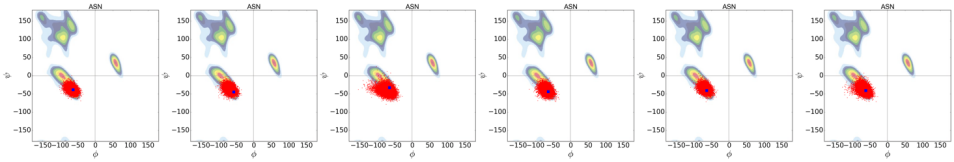
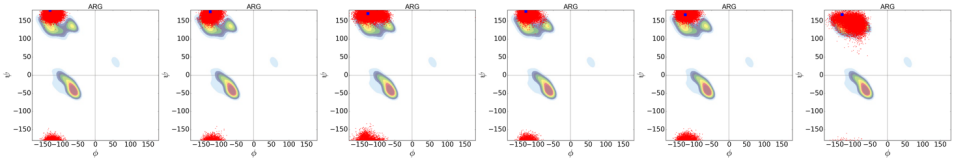
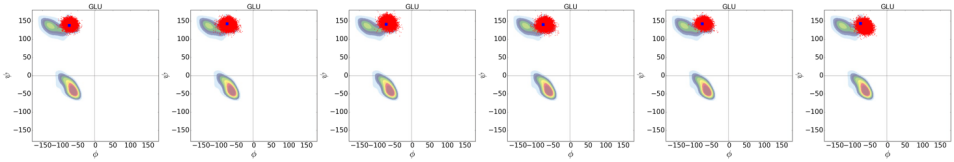
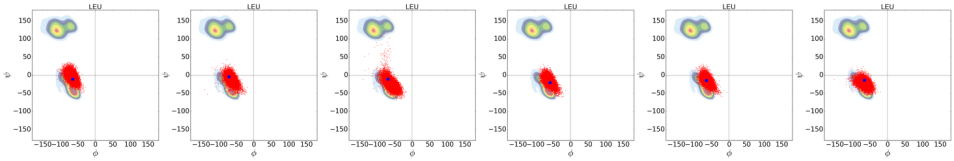
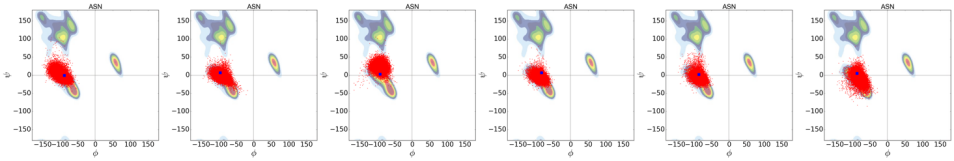
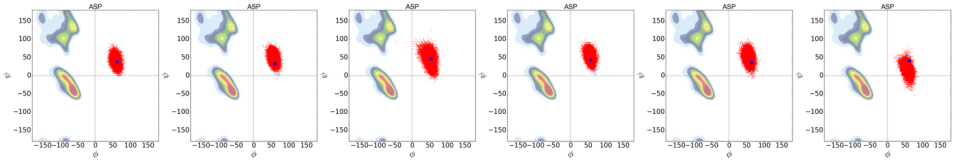
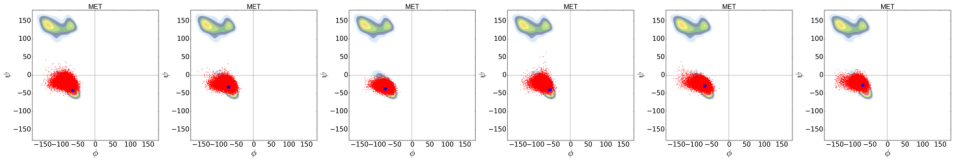
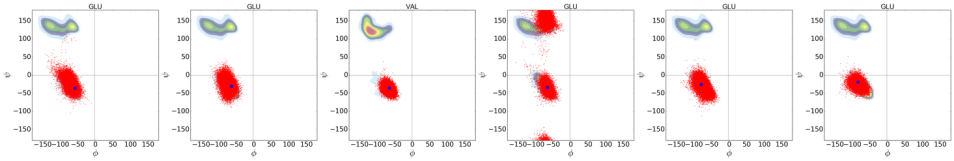
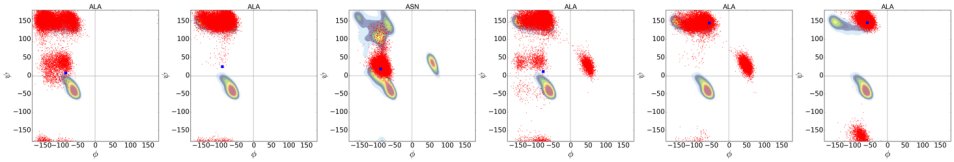
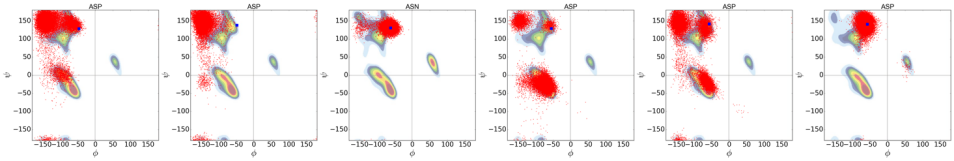
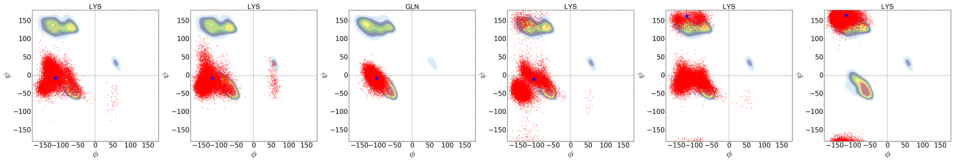
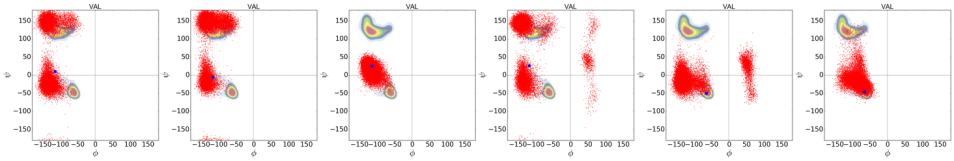
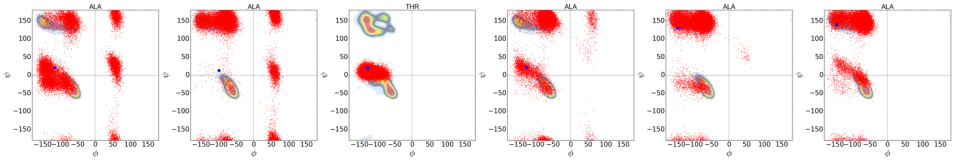


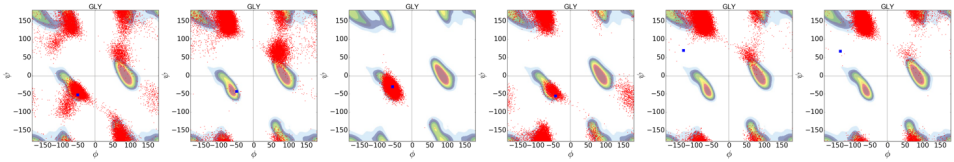
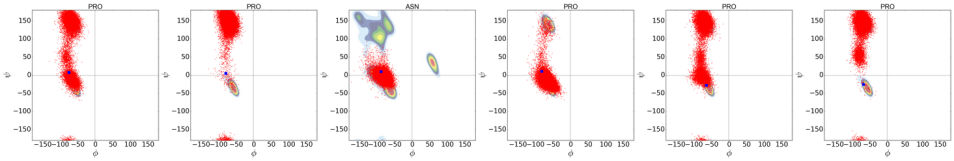
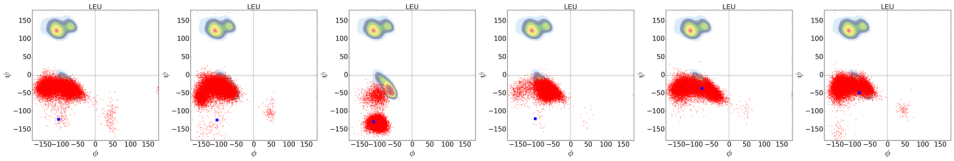
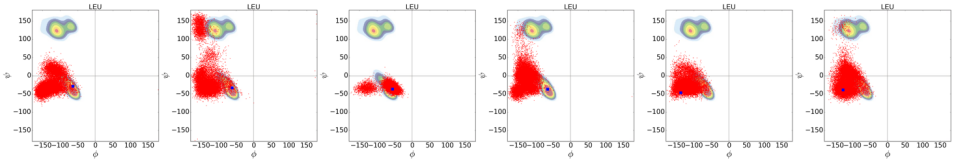
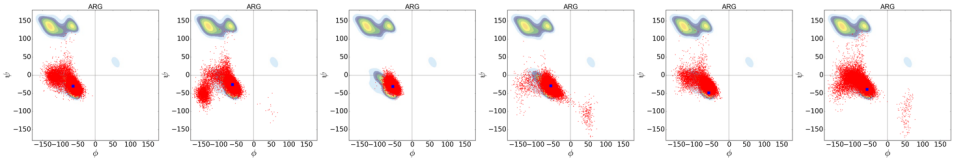
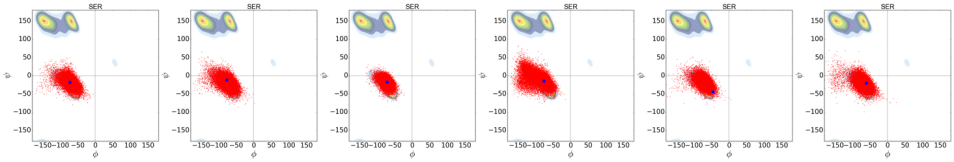
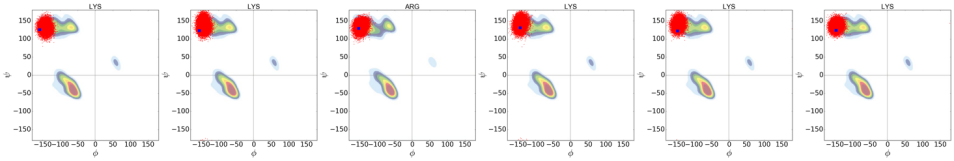
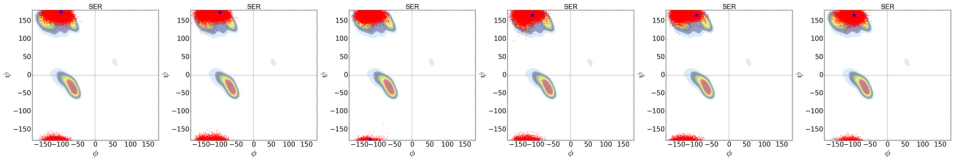
Figure A 3-3 Correlation between the number of side-chain alternate states and the resolution of the crystal structure.

For all PDB entries found for Uniprot entries P62593 (TEM-1) and P16897 (PSE-4). WT crystal structures are shown in blue, variants in red and TEM-1 chimeras in yellow. An exponential regression curve is shown in blue.


Residue
TEM-1
PDB: 1XPB
TEM-1V84I/A184V
PDB: 1ZG4
PSE-4
PDB: 1G68
cTEM-2m
PDB: 4MEZ
cTEM-17m
PDB: 4ID4
cTEM-19m
PDB: 4R4R
68
69
70
71
72
73
104
105

Residue	TEM-1 PDB: 1XPB	TEM-1$_{V84I/A184V}$ PDB: 1ZG4	PSE-4 PDB: 1G68	cTEM-2m PDB: 4MEZ	cTEM-17m PDB: 4ID4	cTEM-19m PDB: 4R4R
130						
131						
132						
136						
164						
166						
169						
170						

Residue	TEM-1 PDB: 1XPB	TEM-1$_{V84I/A184V}$ PDB: 1ZG4	PSE-4 PDB: 1G68	cTEM-2m PDB: 4MEZ	cTEM-17m PDB: 4ID4	cTEM-19m PDB: 4R4R
179						
211						
212						
213						
214						
215						
216						
217						

Residue	TEM-1 PDB: 1XPB	TEM-1$_{V84I/A184V}$ PDB: 1ZG4	PSE-4 PDB: 1G68	cTEM-2m PDB: 4MEZ	cTEM-17m PDB: 4ID4	cTEM-19m PDB: 4R4R
218						
219						
220						
221						
222						
223						
234						
235						

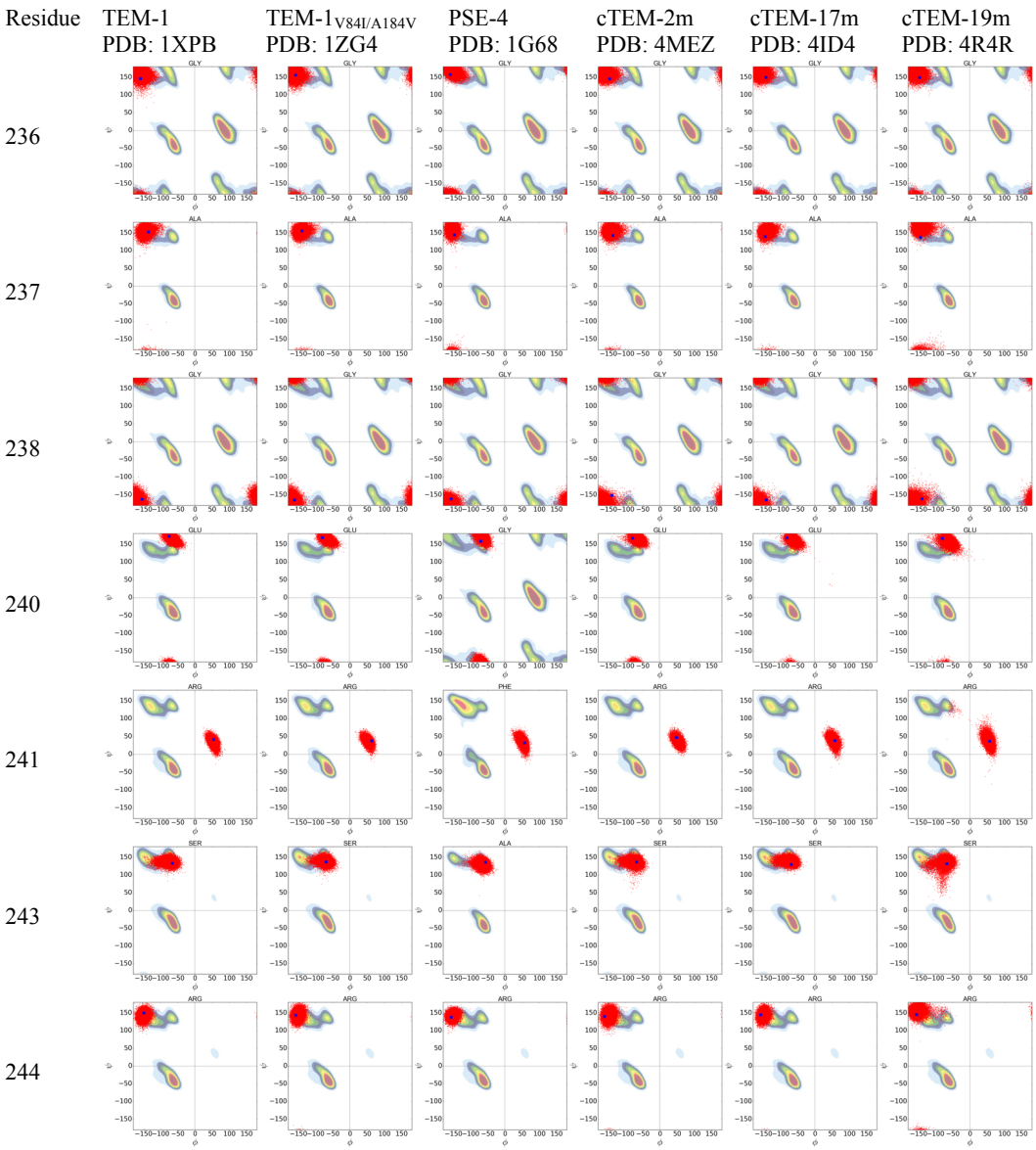


Figure A 3-4 Ramachandran plot of selected residues.

For the MD trajectories, the ψ and Φ angles of each repetition were calculated over the entire 2 µs simulation (total of 6 µs) for every 100th ps and are shown as red dots. The crystallographic ψ and Φ angles are indicated as a blue square.

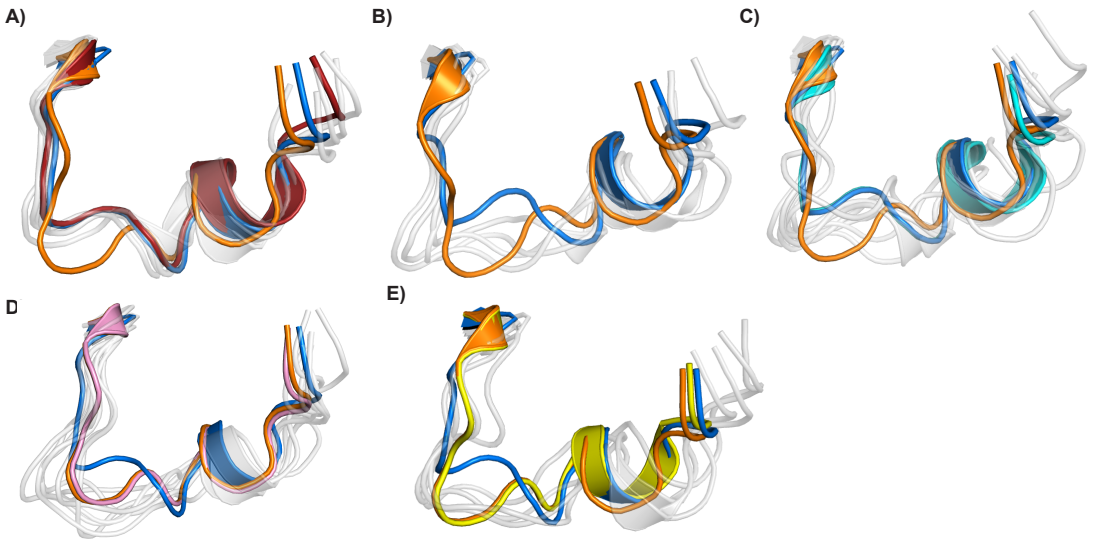


Figure A 3-5 Cluster analysis of the 214-218 active-site loop.
All clusters found in the respective repetitions are shown in grey. For reference, all clusters are overlaid with the loop of TEM-1 (1XPB; blue) and cTEM-19m (4QY5; orange). A) PSE-4 with crystallized loop conformation (1G68; red). B) TEM-1. C) cTEM-2m with crystallized loop conformation (4MEZ; cyan). D) cTEM-17m with crystallized loop conformation (4ID4; pink). E) cTEM-19m with crystallized loop conformation (4R4R; yellow).

3.3. References

1. Knox, J.R., et al., *A catalytically-impaired class A beta-lactamase: 2 A crystal structure and kinetics of the Bacillus licheniformis E166A mutant.* Protein Eng, 1993. **6**(1): p. 11-18.
2. Guillaume, G., et al., *Site-directed mutagenesis of glutamate 166 in two beta-lactamases. Kinetic and molecular modeling studies.* J Biol Chem, 1997. **272**(9): p. 5438-5444.
3. Gibson, R.M., H. Christensen, and S.G. Waley, *Site-directed mutagenesis of beta-lactamase I. Single and double mutants of Glu-166 and Lys-73.* Biochem J, 1990. **272**(3): p. 613-619.

Annex 4 - Conserved Fast Dynamics Accompanied by Altered Slow Conformational Dynamics in Functional Engineered β-Lactamases

4.1. Context

Protein dynamics describe a continuum of motion over the entire timescale from the fs to the ms, or more (see also Chapter 1 sub-heading 1.4). As mentioned earlier, there is no single experimental or computational technique that can currently inform on the full timescale of protein motions. NMR is a well-established experimental technique and delivers reliable and highly detailed information about protein dynamics on a variety of timescales. There are, however, shortcomings to NMR experiments and their interpretation (more details in Chapter 5). MD simulations can provide a faster and cheaper alternative to determination of dynamics on the fs to μs timescale; as opposed to NMR data for large protein systems, which are limited to analysis of the protein backbone, computational methods deliver a fully atomistic representation of protein dynamics. Dr. Sophie Gobeil (former Ph.D. student in the Pelletier laboratory) and I are highly interested in the question of whether protein dynamics represent a trait that has been conserved over the course of natural evolution. In the following article, we first demonstrate the high quality of predicted NMR observables obtained from MD simulations, suggesting that for the ps to ns timescale, NMR experiments are no longer the method of choice as a result of their high cost and lengthy analysis. The setup for the MD simulation is minimal and S^2 observables can be extracted within a single day. Also, there is basically no limit to the size of the protein, which enables the researcher to overcome apparent NMR limitations (see also Chapter 6 for deriving S^2 observables for a P450 enzyme).

The intermediate timescale between the ns and μs timescale can, in principle, be covered by the NMR technique of residual dipolar couplings. However, it requires a system setup that differs highly from the natural protein environment such that the resulting data may be biased. In this chapter, we analyze the dynamics on the ns-to-μs timescale using MD simulations and link them to CPMG NMR data obtained for the μs-to-ms timescale, which is

not yet attainable in MD simulations without incurring significant trade-offs (see also Chapter 2 for details on long MD simulations). Combining all those techniques enable us to study whether protein evolution is conserved over a model evolutionary path linking TEM-1 to PSE-4 β-lactamase.

This chapter is a "pre-print" version of the article entitled '*Conserved Fast Dynamics Accompanied by Altered Slow Conformational Dynamics in Functional Engineered β-Lactamases*', currently under review in *ACS Chemical Biology*. Dr. Sophie Gobeil is the first author, while I am the second. While we both contributed to conceptualization, data analysis and writing, Dr. Gobeil's focus was on the crystallography, NMR experiments, enzyme kinetics and knowledge of β-lactamases. I contributed the computational data acquisition and analysis. The trajectories were partially acquired in the laboratory of Prof. Jürgen Pleiss at the University of Stuttgart where I undertook a 4-month, FRQ-NT-funded international internship in 2014-2015. Crystallography was performed in collaboration with Dr. Jaeok Park, under the supervision of Prof. Albert Berghuis, at McGill University. The NMR analysis was performed with Dr. Donald Gagné, under the supervision of Prof. Nicolas Doucet, at INRS-Institut Armand Frappier. The work was done under the direction of Prof. Joelle Pelletier. The manuscript was drafted mainly by Dr. Gobeil while I contributed the computational section. The article was written with assistance from Prof. Pelletier and revisions by all authors.

Conserved Fast Dynamics Accompanied by Altered Slow Conformational Dynamics in Functional Engineered β-Lactamases

Sophie M.C. Gobeil[1,2], Maximillian C.C.J.C. Ebert[1,2], Jaeok Park[2,3,4], Donald Gagné[2,4,5,7], Nicolas Doucet[2,4,5], Albert M. Berghuis[2,3,4] Jürgen Pleiss[6] and Joelle N. Pelletier[1,2]

[1] Département de biochimie and [g] Département de chimie, Université de Montréal, Montréal, QC, Canada, H3T 1J4

[2] PROTEO, the Québec Network for Research on Protein Function, Engineering, and Applications, G1V 0A6

[3] Department of Biochemistry, McGill University, Montréal, QC, Canada, H3G 1Y6

[4] GRASP, Groupe de Recherche Axé sur la Structure des Protéines, H3G 0B1

[5] INRS-Institut Armand-Frappier, Université du Québec, Laval, QC, Canada, H7V 1B7

[6] Institute of Technical Biochemistry, University of Stuttgart, Stuttgart, Germany, 70569

[7]Present address: Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY, USA

Corresponding Author: Joelle N. Pelletier <joelle.pelletier@umontreal.ca>

4.2. Abstract

Protein motions range from rapid side-chain fluctuations to slower backbone displacements. Increasing evidence that protein motions contribute to function compels us to examine the evolutionary significance of protein dynamics. Using experimental and *in silico* methodologies, dynamics on the entire picosecond-to-millisecond regime were determined for two naturally-evolved β-lactamases and three active, recombined variants displaying native-like structure. Catalytic efficiency was unaltered for hydrolysis of penicillins, while k_{cat} varied up to 75-fold for bulkier cephalosporins. Fast dynamics in the engineered, evolutionarily-related variants were conserved yet we observed strikingly increased slow dynamics, coinciding with the timescale of turnover. The proliferation of dynamic residues extending outward from the active site was characterized by frequencies ranging from native-like to slower-than-observable. Our results support a model where maintenance of fast motions is associated with conservation of the hydrolytic reactivity, while modified slow dynamics are functionally tolerated to varying extents according to substrate identity, in this highly evolvable antibiotic resistance system.

Keywords: X-ray crystal Structure, Protein dynamics, Enzyme evolution, NMR dynamics, MD simulations, β-Lactamase.

4.3. Introduction

Protein motions occur on various timescales. Rapid bond vibrations and rotations (femto- to nanosecond motions) allow proteins to explore local conformational changes while slower motions (nano- to millisecond or second), including loop and domain rearrangements, favor the exploration of conformational states separated by high energy barriers [1, 2]. Interconversion of conformers over different timescales is thought to underlie functional diversity [3]. While the exact role of enzyme motions in transition state chemistry is debated, correlation between the dynamics (*i.e.* conformational exchange over time) on fast to slow timescales and function have been put forward for various proteins [4-8].

Widely differing systems show evolutionary conservation of dynamics consistent with specific motions being linked to enzyme function [9-15], although there is also evidence of dynamic variation throughout evolution [9, 16]. While being highly informative, none of those studies assessed dynamics on a breadth of continuous timescales. Here, we investigate protein dynamics over the entire span of fast (ps) to slow (ms) timescales for a system of inter-related chimeric β-lactamases originating from segments of the TEM-1 penicillinase and the PSE-4 early-generation cephalosporinase, which share 40% sequence identity [17]. Those chimeras were previously produced by laboratory recombination in a process that mimics natural homologous recombination and were chosen among the 20% of recombinants that allowed bacterial survival against ampicillin [18, 19]. As a virtue of being catalytically active, these chimeras constitute a relevant system to probe the role of dynamics in an engineered model of β-lactmase evolution that links TEM-1 and PSE-4 (Figure A 4-1 A).

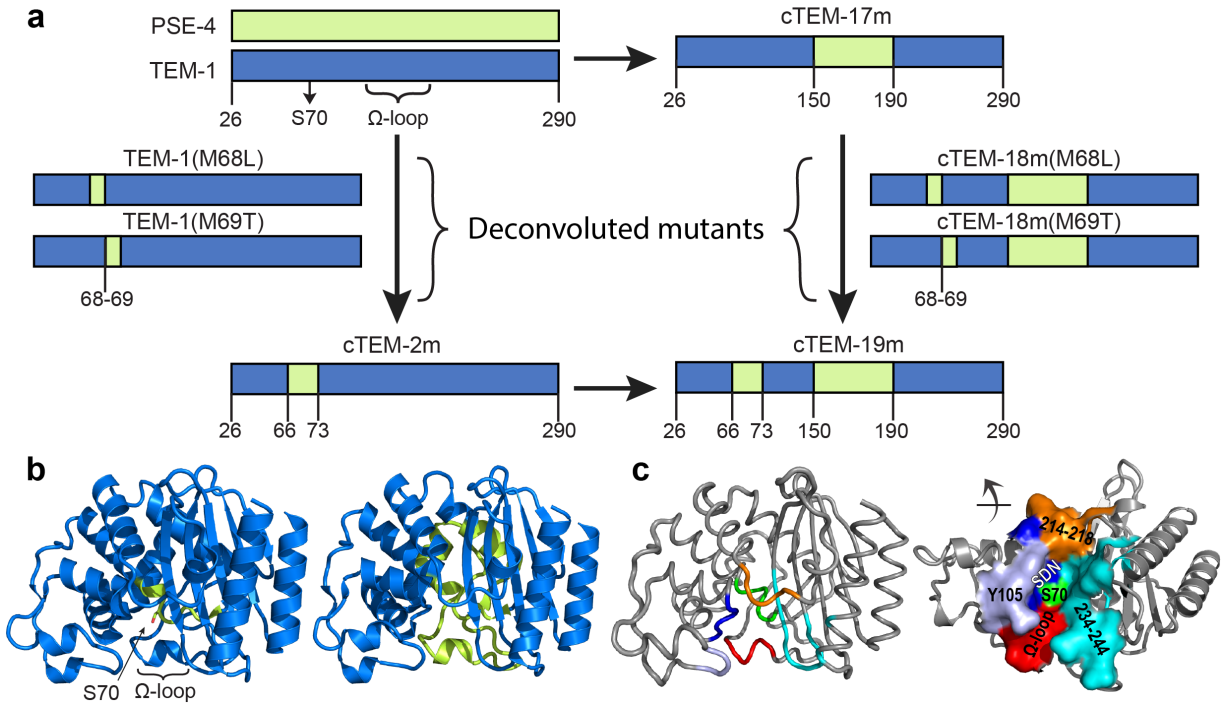


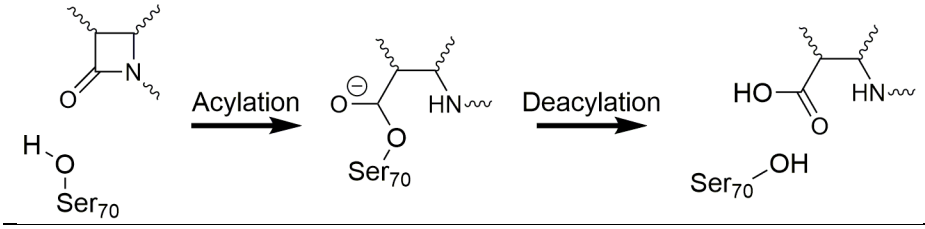
Figure A 4-1 Relation of chimeras and the deconvoluted mutants to the parental class A β-lactamases TEM-1 and PSE-4.

A) Numbering of the sequence blocks originating from TEM-1 (blue) and PSE-4 (green) in cTEM-2m, cTEM-17m, cTEM-19m and the deconvoluted mutants cTEM-18m(M68L), cTEM-18m(M69T), TEM-1(M68L) and TEM-1(M69T). Numbering according to Ambler [20]. The chimeras are named as follows: 'c' indicates chimera, 'm' the number of substitutions (or mutations) relative to TEM-1. The catalytic nucleophile, Ser70, and the catalytic Ω-loop are indicated in the TEM-1 sequence. B) Structural representation of the parental origin in cTEM-2m (PDB 4MEZ) and cTEM-19m (PDB 4R4S), colored as in panel A. The catalytic nucleophile, Ser70 (in sticks) and the Ω-loop are indicated on cTEM-2m to highlight the hybrid composition of the active-site region at the interface of the all-α (left) and the α/β (right) domains. C) The active-site walls, set in TEM-1 (PDB 1XPB). Green: S70 wall (Met69-Lys73); lilac, Y105 wall (Val 103-Ser106); dark blue, SDN wall (Met129-Asn132); red, Ω-loop wall (Glu166-Asn170); orange, 214-218 wall; and cyan, 234-244 wall. (Right) The structure is reoriented as indicated by the arrow and the active-site walls are represented by their solvent-accessible surface.

We previously examined one such chimeric β-lactamase in regards to its fast (ps-ns) and slow (μs-ms) dynamics, leaving a knowledge gap relative to the intermediary ns-μs motions [14]. A core challenge in understanding the dynamics/structure/function relationship throughout evolution is describing dynamics on continuous timescales for all proteins within an evolutionary system [21]. X-ray diffraction data generally provide little insight into

timescales at which motions occur. NMR can probe motions from the picosecond to seconds, yet the NMR residual dipolar coupling method reporting on the nano- to microsecond timescale suffers from inherent technical challenges, precluding its application to large protein systems [22, 23]. Nonetheless, knowledge of ns-μs motions provide insight into the dynamics/function relationship, as shown for an inhibitor-resistant clinical variant of TEM-1 β-lactamase [24]. To access the ns-μs timescale, we include computational methodologies to compare the motions of the related β-lactamases. We consider these in relation to steady-state kinetic parameters that broadly characterize catalytic function. Penicillins and cephalosporins were investigated since the rate-limiting step differs for these two subclasses of β-lactam antibiotics (Table A 4-1).

Table A 4-1 Reaction scheme and turnover rate constants for the hydrolysis of cephalosporins by TEM-1, the chimeric β-lactamases and PSE-4[a].



		k_{cat} (s^{-1})	k_{cat} fold variation relative to TEM-1
Cephalothin	TEM-1[a]	84 ± 12	1.0
	cTEM-17m[a]	120 ± 8.0	1.4 (↑)
	TEM-1(M69T)	7.0 ± 4.0	12 (↓)
	cTEM-18m(M69T)	1.4 ± 0.3	60 (↓)
	cTEM-2m	2.8 ± 0.8	30 (↓)
	cTEM-19m*	6.0 ± 0.5	14 (↓)
	PSE-4[a]	0.80 ± 0.01	105 (↓)
Cefazolin	TEM-1[a]	55 ± 8.0	1.0
	cTEM-17m[a]	75 ± 21	1.4 (↑)
	TEM-1(M69T)	9.0 ± 2.0	6.1 (↓)
	cTEM-18m(M69T)	4.0 ± 0.6	14 (↓)
	cTEM-2m	3.3 ± 0.1	17 (↓)
	cTEM-19m*	6.0 ± 0.3	9.2 (↓)
	PSE-4[a]	1.9 ± 0.3	29 (↓)
Cefotaxime	TEM-1[a]	0.74 ± 0.10	1.0
	cTEM-17m[a]	0.14 ± 0.04	5.3 (↓)
	TEM-1(M69T)	0.03 ± 0.02	25 (↓)
	cTEM-18m(M69T)	0.06 ± 0.02	12 (↓)

	cTEM-2m	0.02 ± 0.01	37 (↓)
	cTEM-19m*	0.01 ± 0.002	74 (↓)
	PSE-4[a]	0.03 ± 0.01	25 (↓)

[a] The Met68Leu variants k_{cat} are presented in Table A 5-9.

[b] Data from [25].

[*] k_{cat} (apparent) due to saturation at a substrate concentration ≥ 5 μM.

Our investigation of related β-lactamases over the entire span of fast (ps) to slow (ms) motion regime unveils striking differences in dynamic conservation throughout the timescales investigated.

4.4. Results

The system of interrelated, functional β-lactamases is based on the 'parental' homologs, TEM-1 and PSE-4 (Figure A 4-1). These were previously recombined block-wise with the SCHEMA algorithm to yield chimeric β-lactamases [18, 19]. Chimera cTEM-2m is based on TEM-1 and includes residues 66-73 from PSE-4 (Figure A 4-1 B and Figure A 5-1). This introduces Met68Leu and Met69Thr relative to TEM-1, at the core of the active site: they are immediate neighbors to the catalytic nucleophile Ser70. Chimera cTEM-17m [14, 25] adopts residues 150-190 from PSE-4, resulting in 17 substitutions in the catalytically relevant Ω-loop (161-179) and its adjacent helices. This active-site wall contains Glu166, proposed to serve as a general base in the catalytic acylation and deacylation steps [26]. Chimera cTEM-19m combines the two sets of recombinations, resulting in 19 substitutions on two active-site walls, relative to TEM-1. Among the 19 positions, 13 have been explored during natural evolution resulting in functional TEM variants (Table A 5-1) [27].

4.4.1. Crystal structure of the chimeras cTEM-2m and cTEM-19m

We report the crystal structures of cTEM-2m (PDB 4MEZ, resolution 2.05 Å) and cTEM-19m (PDB 4R4S, 1.1 Å and 4R4R, 1.2 Å) (Figure A 4-2; Table A 5-2). Because β-lactamases turn over their natural substrates in the absence of any co-factor at turnover rates approaching the diffusion limit, structural and NMR dynamics studies with intact substrate or inhibitors are not readily feasible.

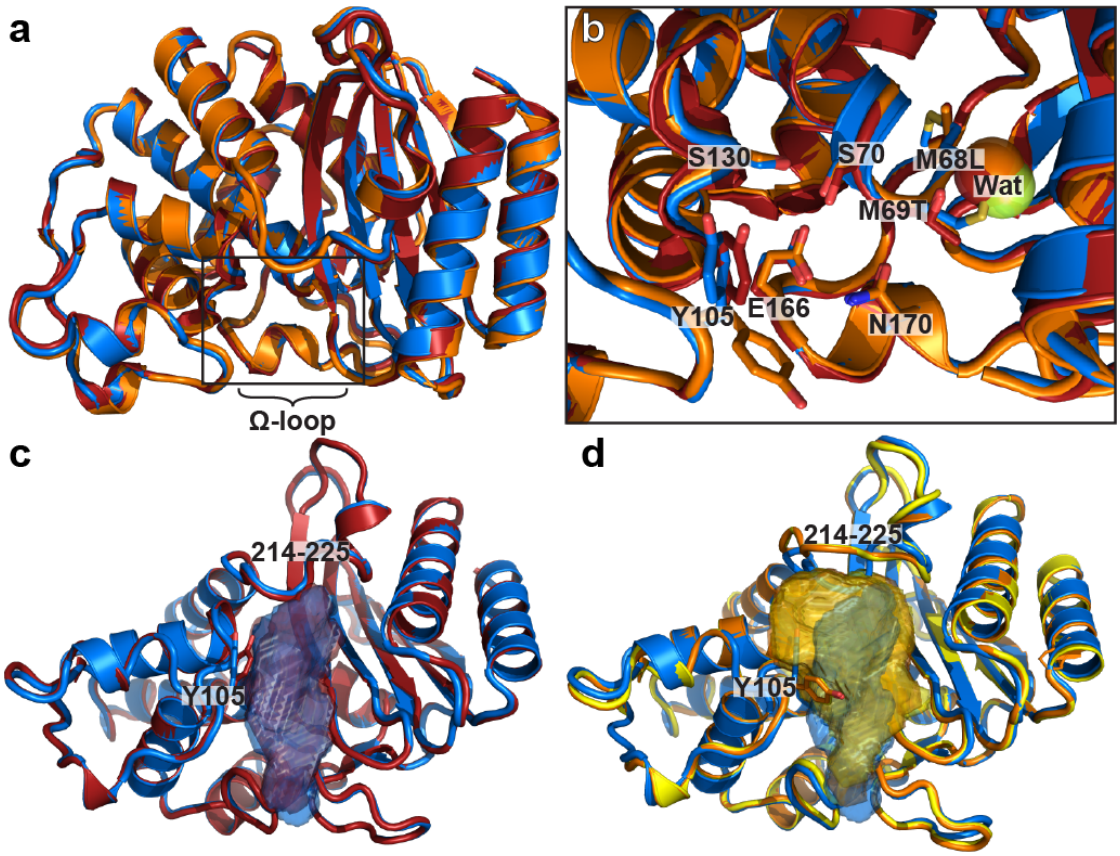


Figure A 4-2 Crystal structures of cTEM-2m and cTEM-19m.
A) Backbone overlay of the crystal structures of chimeras cTEM-2m (red; PDB 4MEZ) and cTEM-19m (orange; PDB 4R4S) over TEM-1 (blue; PDB 1XPB). B) Active-site view of the overlay of cTEM-2m and cTEM-19m with TEM-1. Wat represents the new water molecule in cTEM-2m (Wat519) and cTEM-19m (Wat478) relative to PSE-4 (green; Wat483). Tyr105 of cTEM-19m was modeled to the m-30° conformation contrasting with the t80° conformation observed in TEM-1 and cTEM-2m. C) Overlay of the active-site volume of TEM-1 and cTEM-2m. Orientation as in Figure A 4-1 C. D) Overlay of the active-site volume of TEM-1, cTEM-17m (yellow; PDB 4ID4) and cTEM-19m. For residues with alternate conformations in the crystal structure, the conformer with the highest occupancy was illustrated. The active-site volume was calculated using 3V and probes of 1.5 Å and 8 Å radius. See Table A 5-2 for data collection and refinement statistics.

Chimera cTEM-2m maintained a native-like fold with no major modifications (Figure A 4-2 A). In contrast, cTEM-19m showed significant differences in the active-site area. The SDN wall (containing Ser130, which stabilizes the active site and transition state), the Ω-loop (containing the catalytic Glu166) and the 234-244 wall (which coordinates the catalytic deacylating water jointly with Asn170 [26] and where Arg244 coordinates the substrate

carboxylate [28]) are all native-like (Figure A 4-1 C and Figure A 4-2 B). However, the Y105 gate-keeper wall revealed Tyr105 adopting the *m-30°* conformation [29]. This was previously observed in cTEM-17m, contrasting with the *t80°* conformation found in all other native and mutant apoenzyme structures of TEM-1 and PSE-4. Our previous simulations and NMR relaxation experiments are consistent with interconversion of Tyr105 between the two conformations [14, 30]. This side-chain rearrangement resulted in a small increase in the active-site volume (Figure A 4-2 C and D). Furthermore, the active-site wall 214-218, which belongs to the 214-225 connector linking the all-α domain to the α/β domain, has adopted an open conformation in cTEM-19m. This significantly increased the active-site volume to 741 $Å^3$, comparable to that for cTEM-17m (746 $Å^3$) and contrasting with those of TEM-1 (384 $Å^3$), PSE-4 (427 $Å^3$) and cTEM-2m (393 $Å^3$). This difference is attributed to the recombined segment 150-190 and its 17 substitutions. A new water molecule in cTEM-2m and cTEM-19m, also present in PSE-4 (Wat483), fills the space occupied by Met69 in TEM-1 and cTEM-17m (Figure A 4-2 B) and becomes part of the active-site architecture.

We did not attempt to interpret crystallographic b-factors with respect to the protein dynamics, because the relevant crystal structures were diffracted at significantly different resolutions (1.05 – 2.05 Å) and in different space groups (P 4_12_12; P $2_12_12_1$; P 1); furthermore, analysis of b-factors does not directly inform on the timescale of motions. To obtain insight into the dynamics of this interrelated model of β-lactamase evolution and the timescales at which motions occur, we performed MD simulations and NMR relaxation experiments.

4.4.2. Validation of MD simulation methodology

The crystal structures were energy minimized and subjected to unconstrained MD simulations. We validated our capacity to simulate the S^2 order parameter previously determined by NMR for TEM-1, PSE-4 and cTEM-17m [25, 31] (Table A 5-3 to Table A 5-5). The S^2 verifies if the peptide N-H bonds remain rigid (S^2 near 1) or exhibit fast motions (S^2 tending toward 0.5). The mean S^2 calculated by NMR and MD simulations compared favorably for all three proteins (S^2 values within 0.03) (Table A 5-3). The per-residue S^2 value in all proteins was between the average ± 1 S.D. (*i.e.* 0.81 to 0.94) for 87% of residues, and 73% of all residues were in excellent MD-NMR agreement (difference between the NMR and

MD S^2 values ≤ 0.05). For the most dynamic residues (S^2 < 0.81), the average MD-NMR difference (or Δ) was 0.10. The S^2 values of these residues agree well in most cases, as exemplified by region 192-200 (Δ = 0.02-0.05). We note greater S^2 divergence for the surface-exposed residues Asp115, Gly143, Asp157, Val159, Asp176 and Trp229 (Figure A 5-1). Despite these isolated variations, our results allow us to confidently extract S^2 order parameters from the MD simulations of cTEM-2m and cTEM-19m.

MD S^2 datasets are more complete than NMR S^2 datasets as a result of incomplete NMR assignment as well as overlapping N-H resonances that render model-free S^2 calculations imprecise. While the active-site walls are of particular interest, NMR S^2 data are missing for the S70, SDN, Ω-loop, 214-218 and 234-244 walls in one or more proteins (Table A 5-5), most likely resulting from significant resonance broadening. For this reason, the complete MD S^2 datasets were used for comparisons.

4.4.3. Dynamics of the β-lactamase system on the fast (ps-ns) timescale

Comparison of S^2 values showed that the core of each β-lactamase was rigid while most residues experiencing fast dynamics were on surface loops, including active-site loops (Figure A 4-3). The hinge leading into the Ω-loop (residues 155-160) showed fast dynamics in all proteins (average S^2 0.79 to 0.81), as did the tip of the Ω-loop (residues 171-178; S^2 0.78 – 0.80) (Table A 5-8). The active site wall 214-218 was more dynamic in TEM-1 (S^2 = 0.80 ± 0.04) and in the three chimeras (S^2 = 0.76 to 0.78) than in the more distantly related PSE-4 (S^2 = 0.85 ± 0.01). Elevated standard deviations were observed in certain regions, particularly for the three chimeras (Figure A 5-1). As the S^2 calculation was performed on multiple 10 ns segments over the total 6 μs simulated trajectories, the standard deviation reflects conformational changes that may occur on timescales slower than 10 ns. While the simulated and NMR-derived S^2 values were comparable, the standard deviation derived from simulations reveals that the residue explored states with high and low S^2 over the course of the simulated trajectories, while in NMR it reflects standard deviation on the model-free fitting for the S^2. Indeed, increased motions were confirmed on slower timescales for the regions discussed

below. S^2 in these regions generally could not be characterized by NMR due to overlapping peaks, again justifying the use of computationally-derived S^2.

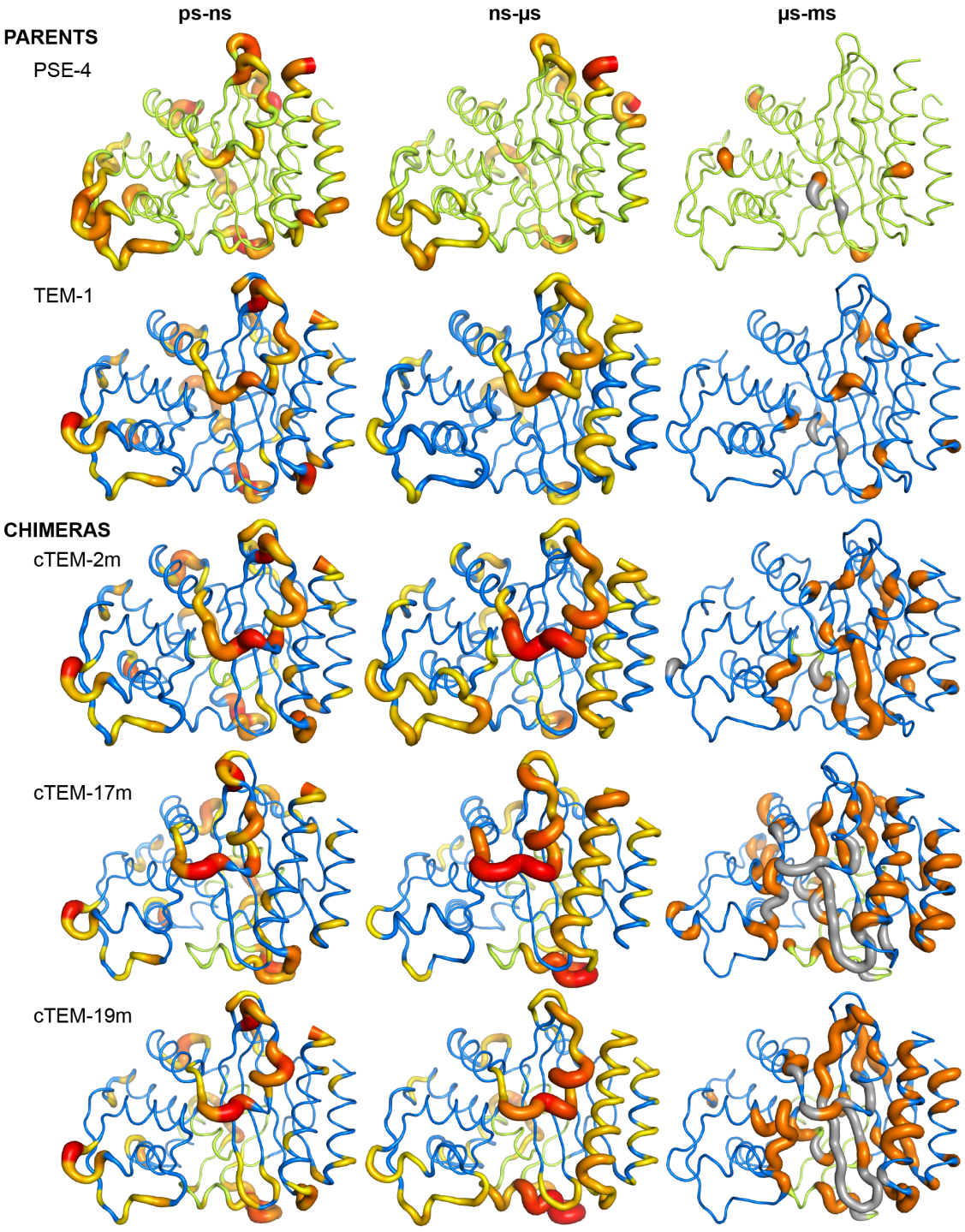


Figure A 4-3 β-Lactamase dynamics on the ps to ms timescales.

Dynamics of the parents TEM-1 (blue, PDB 1XPB), PSE-4 (green, PDB 1G68) and the laboratory-evolved chimeras cTEM-2m (PDB 4MEZ), cTEM-17m (PDB 4ID4) and cTEM-19m (PDB 4R4S) (colored on a blue/green scheme identifying the parental origin of the sequence block as in Figure 1). Left: Dynamic residues on the ps-ns are represented by the MD-derived order parameter S^2 on the amide NH bonds. Residues with S^2 lower than protein average S^2 are colored and thickened on a yellow (< 0.85) to bright red scale (≤ 0.7). Middle: Dynamics on the ns-μs timescale are represented by the Cα-RMSF derived from triplicate 2 μs MD simulations. Residues with Cα-RMSF higher than the protein average Cα-RMSF are colored and thickened on a yellow (> 0.10 nm) to bright red scale (≥ 0.25 nm). Right: Dynamics on the μs-ms timescale were monitored by CPMG-NMR on the ^{1}H-^{15}N vector. Residues showing dispersion curves with $\Delta R_2 \geq 7.0\ s^{-1}$ at 800 MHz are colored orange. NMR unassigned residues are colored gray. See Table A 5-2 to Table A 5-7and Figure A 5-1 to Figure A 5-4 for the raw data.

Finally, the active-site wall 234-244 was rigid in all proteins (S^2 = 0.86-0.88), though somewhat higher per-residue standard deviation in cTEM-19m suggests motions occurring on a slower timescale.

4.4.4. Dynamics on the intermediate (ns-μs) timescale

Three 2 μs MD simulations, probing a total of 6 μs per protein, were run using independent starting conformations to minimize the impact of local energy minima. We extracted the Cα root mean square fluctuation (RMSF), as Cα is the atom that best captures residue motion (Figure A 5-2 and Table A 5-6).

As observed for fast motions, the core of each β-lactamase was rigid while the ns-μs motions were generally located on surface loops (Figure A 4-3). TEM-1 and the chimeras were significantly more dynamic than PSE-4, indicating that the greater homology of the chimeras to TEM-1 determines their intermediate timescale motions. The pattern of active-site wall dynamics was essentially the same for both timescales, where loop motions involved similar active-site walls (Table A 5-8). All proteins exhibited above-average dynamics in the leading hinge and tip of the Ω-loop (RMSF = 0.12 – 0.23 nm, relative to protein averages of 0.07 – 0.10 nm). All but PSE-4 were dynamic in the 214-218 wall (RMSF = 0.15 – 0.25 nm; 0.07 nm for PSE-4). Further instances of active-site dynamics were noted. Chimeras cTEM-17m and cTEM-19m showed above-average RMSF for the entire Ω-loop (RMSF = 0.14 and 0.17 nm, respectively; 0.07 – 0.10 nm for the others). Finally, the Y105 active-site wall of

cTEM-2m had an RMSF of 0.14 ± 0.05 nm (0.8-0.11 nm for the others), indicating ns-μs motions unique to this chimera.

As observed for the ps-ns timescale, above-average per-residue standard deviation of several active-site walls was observed, often linked to above-average RMSF (Figure A 5-2). This was the case for the cTEM-17m and cTEM-19m Ω-loop wall and the 214-218 wall of the three chimeras. The Y105 wall of cTEM-2m showed elevated RMSF and standard deviation, whereas only the standard deviation was elevated in cTEM-19m. Residues 238-242 of the 234-244 wall showed increased RMSF in cTEM-19m and moderately increased standard deviation in cTEM-17m. Again, increased standard deviation may reflect conformational changes slower than probed.

4.4.5. Dynamics on the slow (μs-ms) timescale probed by ^{15}N CPMG NMR

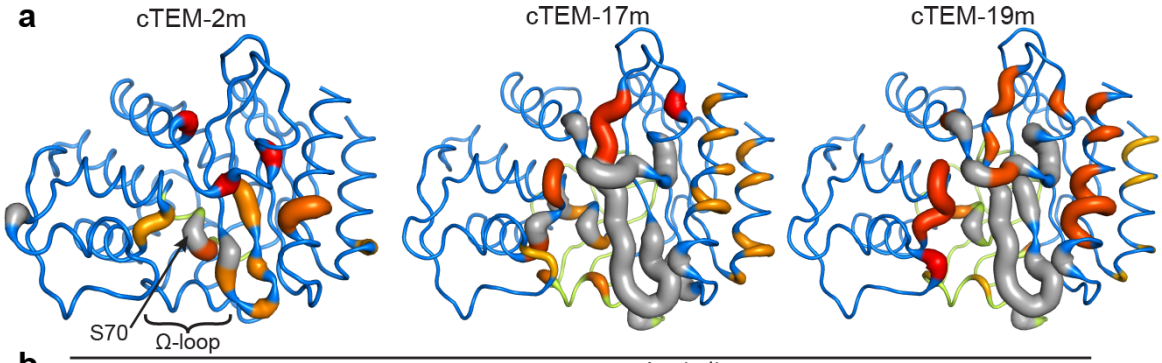
Comparison of NMR assignments [32, 33] revealed minimal backbone chemical shift (σ) differences except where there are sequence differences (Figure A 5-3). Above-average σ differences were observed for residues 236-246 of cTEM-2m relative to TEM-1 ($\Delta\sigma$ = 0.12 ppm; overall $\Delta\sigma$ = 0.05 ppm). They belong to active-site wall 234-244 and were unassigned in chimeras cTEM-17m and cTEM-19m. Indeed, backbone resonance assignments were significantly less complete for cTEM-17m and cTEM-19m (91-92%) than TEM-1, PSE-4 and cTEM-2m (98-99%). This appears to result from NMR resonance broadening, preventing assignment of residues that exhibit motions on a timescale slower than that probed [22, 23, 34] and thus provides further evidence of increased slow timescale dynamics in cTEM-17m and cTEM-19m.

Probing μs-ms backbone amide dynamics by ^{15}N-CPMG relaxation dispersion NMR revealed 26 residues in cTEM-2m and 60 residues in cTEM-19m undergoing conformational exchange (ΔR_2 values $\geq 7 s^{-1}$) on the timescale monitored by this experiment ($\approx 50 - 3{,}000\ s^{-1}$) (Figure A 4-3 and Figure A 5-4; Table A 5-7). This offers a sharp contrast with TEM-1 and PSE-4, which exhibit a conserved pattern of restricted slow dynamics in only 11 and 5 residues, respectively, including core and active-site residues. cTEM-17m is similar to cTEM-

19m, exhibiting slow motions in 60 residues [14]. These results illustrate that various well-folded, functional β-lactamases can tolerate altered dynamics on this timescale.

Among the 26 dynamic residues in cTEM-2m, we find 7 of the 11 dynamic residues of TEM-1. This indicates that the slow dynamic properties of TEM-1 were generally maintained while additional residues acquired ms conformational exchange. In cTEM-2m, most residues exhibiting slow dynamics were located in the α/β domain; the substitutions in cTEM-2m (residues 68-69) are located at the interface between the α/β and the all-α domains, so it is surprising that the new dynamics should extend into only one of the domains (Figure A 4-3). Interestingly, this chimera includes a stretch of dynamic residues bordering the active-site wall 234-244 (Ile231, Lys234, Gly238 and residues 241-246), where TEM-1 showed a single dynamic residue (Arg241). The exchange rate constant (k_{ex}) is a metric for flexibility. It informs on the sampling rate of interconverting conformations. The per-residue k_{ex} was calculated when ^{15}N-CPMG relaxation curves were observed at both NMR fields (Table A 5-7). The per residue k_{ex} were also globally fitted for stretches of residues defining each active site wall, where the per residue k_{ex} values overlapped (Table A 5-8). Although an increased number of dynamic residues in the cTEM-2m 214-225 connector and 234-244 wall was observed, the k_{ex} of these regions were maintained as for TEM-1 (Figure A 4-4; Table A 5-8).

a



b

Active-site wall	k_{ex} (s^{-1})				
	TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
S70	-	-	937 ± 254	480 ± 110	667 ± 141
Y105	-	-	ND	300 ± 90	1,219 ± 225
SDN	ND	-	480 ± 205	775 ± 270	979 ± 203
Ω-loop	-	2,019 ± 813	-	586 ± 134	521 ± 113
214-225	1,387 ± 368	-	1,339 ± 784	1,098 ± 443 *	1,131 ± 423 *
234-244	646 ± 29	-	611 ± 221	910 ± 205 *	796 ± 250 *

Figure A 4-4 Global exchange rates (k_{ex}) of cTEM-2m, cTEM-17m and cTEM-19m.

*A) The k_{ex} of the fitted regions are colored on a yellow (300 s^{-1}) to red (1500 s^{-1}) gradient on the crystal structure coordinates of cTEM-2m (PDB 4MEZ), cTEM-17m (PDB 4ID4) and cTEM-19m (PDB 4R4S). NMR unassigned residues are colored dark gray. The fitted regions were the N-terminal helix (H1, residues 28-41), S70 wall (residue 69-72), Y105 wall (residue 105-106), SDN wall (residue 128-132), Ω-loop wall (residue 166-170), domains connector (214-225), 234-244 wall (234-244) and the C-terminal helix (H11, 272-288). B) Active-site wall calculated global exchange rates (k_{ex}). The * indicates that the region includes multiple unassigned residues. ND: k_{ex} was not determined as a result of large error on the calculated fit. See Table A 5-8.*

Among the 60 assigned dynamic residues in cTEM-19m, 35 were identical to cTEM-17m while 12 in cTEM-17m and 18 in cTEM-19m were also dynamic in cTEM-2m. The remaining dynamic residues mostly clustered nearby. In addition to those assigned, dynamic residues, the majority of residues constituting both active-site walls belonging to the α/β domain (*i.e.* 214-218 and 234-244) of cTEM-17m and cTEM-19m were unassigned. This is consistent with residues experiencing motions on a slower rate of exchange that broaden the NMR resonances. Interestingly, the dynamic residues also extend beyond those two active-site walls. The extension of the dynamic stretches in cTEM-17m and cTEM-19m is such that only four residues (including a Pro) interrupt the now near-continuous 214-244 dynamic region.

These results demonstrate that the increased number of dynamic residues in these walls results from either recombination of region 66-73 or 150-190. Interestingly, the exchange rate k_{ex} for the assigned, dynamic residues in the 214-244 region of cTEM-17m and cTEM-19m are comparable to their homologs TEM-1 and cTEM-2m (Figure A 4-4). However, the frequency of motions is not fully conserved since unassigned residues are indicative of an exchange rate slower than the millisecond. These contiguous walls of the α/β domain are therefore highly tolerant to modifications in the number and rate of the dynamic residues.

In the all-α domain active-site walls (S70, Y105, SDN and Ω-loop), cTEM-17m and cTEM-19m showed extensive ms dynamics, where limited or no dynamics were found in the parents. A new stretch of dynamic residues with a similar k_{ex} was observed in the Ω-loop wall of cTEM-17m and cTEM-19m. Similarly, an increased number of dynamic residues in the S70 and S130 active-site walls of the chimeras differed by less than 2-fold in k_{ex} (Table A 5-8). Thus, recombination of region 150-190 in cTEM-17m and cTEM-19m and to a lesser extent region 66-73 in cTEM-2m increased the number of dynamic residues on the all-α domain side

of the active site, yet maintained a similar k_{ex} regardless of the recombination. In contrast, while a comparable number of dynamic residues was observed in the chimeras' Y105 wall, the k_{ex} of cTEM-17m (300 ± 90 s^{-1}) and cTEM-19m (1,219 ± 225 s^{-1}) differ by 4-fold. This difference may have a functional implication as Tyr105 is a gatekeeping residue involved in substrate stabilization [35, 36].

4.4.6. Kinetic characterization of cTEM-2m and cTEM-19m

To verify any correlation between dynamics and function (*i.e.* hydrolysis of β-lactams), we characterized variations of the macroscopic constants K_M and k_{cat} for hydrolysis of two penicillins (benzylpenicillin and carbenicillin), two first-generation cephalosporins (cephalothin and cefazolin) and a third-generation cephalosporin (cefotaxime). The catalytic mechanism of class A β-lactamases is initiated by formation of a Michaelis complex that proceeds to the acyl-enzyme intermediate and thereafter deacylation (Table A 4-1). Deacylation is rate-limiting for penicillin hydrolysis [37] while for cephalosporins acylation is thought to be rate-limiting [38].

While TEM-1 is considered as a penicillinase and early-generation cephalosporinase, and PSE-4 as a carbenicillinase [17], they differ minimally for kinetics of penicillin hydrolysis. Their distinction comes mostly from a 5-fold greater k_{cat} for carbenicillin by PSE-4 accompanied by a 25- to 105-fold lower k_{cat} for cephalosporin hydrolysis (Table A 4-1). Overall, substitutions Met68Leu, Met69Thr and Met68Leu/Met69Thr (cTEM-2m) had little impact on hydrolysis of penicillins. This confirms that all members of this β-lactamase system are catalytically competent and that these mutations do not influence the carbenicillinase activity of PSE-4. However, Met68Leu/Met69Thr progressively decreased the k_{cat} of cephalosporin hydrolysis (9- to 75-fold) from a TEM-1-like to a PSE-4-like k_{cat}, with a small decrease in K_M (2- to 7-fold), regardless of the identity of residues 150-190 (Table A 5-9: Complete kinetic dataset). Thus, the substitutions Met68Leu/Met69Thr influence cephalosporin hydrolysis and procure PSE-4-like kinetic properties to TEM-1.

Met69Thr was the main contributor to those kinetic changes. By deconvolution (Figure A 4-1), we observed that this substitution reduced k_{cat} up to 80-fold whether in TEM-1 or in cTEM-17m while also reducing K_M up to 40-fold (Table A 5-9**)**. Because k_{cat} and K_M were

generally both reduced, catalytic efficiency (k_{cat}/K_M) changed little. The Met68Leu substitution had little effect on the kinetics of TEM-1 or cTEM-17m (< 6-fold variation, exception: k_{cat}^{CTX} 12-fold decreased).

4.5. Discussion

4.5.1. The continuum of fast-to-slow dynamics in relation to function in a β-lactamase system

Our results shed light on the structure/function/dynamics relation over the entire ps to ms timescale in an engineered model of β-lactamase evolution. The interrelated system of chimeric β-lactamases stems from the recombination of the PSE-4 carbenicillinase and its more recently evolved homolog, TEM-1 penicillinase [39]. Both of the exchanged segments show natural diversity: 13 of the 19 mutated positions cTEM-19m have been exchanged during evolution resulting in functional variants. Mutations M68I, M69I/L/V occur in 32 natural TEM variants and are close to the mutations M68L and M69T in cTEM-2m and cTEM-19m [27]. Mutations H153R or M155I in cTEM-17m and cTEM-19m are found in 6 natural TEM variants.

At the fastest timescale of motions (ps-ns), the native TEM-1 and PSE-4 and the three chimeras exhibited a conservation of core rigidity with modest surface loop dynamics (Figure A 4-3) [14, 25, 36]. The intermediate timescale of motions (ns-μs) held less dynamic residues yet included a "dynamic personality": the mildly dynamic surface loops in TEM-1 were less so in the three chimeras and were virtually rigid in PSE-4. Thus, the chimeras adopted dynamics that are intermediate between those of TEM-1 and PSE-4, and mostly resemble the closer homolog, TEM-1. On the slowest timescale (μs to ms), the two native homologs once again exhibited very few dynamic residues [14], while dramatic changes were observed in the chimeras. First, extensive dynamics appeared within the protein core. Second, the substitution of residues 68-69 and/or substitution of 17 residues in the 150-190 region resulted in blatant and widespread increases in slow dynamics. Finally, the high standard deviations observed on the faster timescales, suggestive of slower motions, were validated by observation of motions on the slower timescales such that a continuum of motions was observed.

The most striking instance of variations in dynamics was in active-site wall 234-244. Although rigid on the fast and intermediate timescales in all proteins under study, its moderately elevated RMSF or above-average standard deviation in cTEM-19m and cTEM-17m, respectively, hinted at new slow dynamics. Consequently, this entire wall was unassigned and slow dynamics were observed in flanking residues. Thus, the 17 substitutions in the Ω-loop region modified motions on this wall. Interestingly, substitutions 68-69 also modified slow dynamics of that wall. Since the spatial relation between wall 234-244 and both sets of substitutions is different (Figure A 4-1 C), we demonstrate that different events near this active-site wall trigger different, new slow motions consistent with native-like function. Similarly, Tyr105 of the gatekeeper wall exhibited slow dynamics only in the three chimeras, again indicating that new dynamics resulted from either of the substitutions. Overall, a variety of slow dynamic patterns were observed throughout all faces of the active site as a result of the recombinations.

Two further observations came to light upon aligning the motions over the 3 observed timescales (Figure A 4-5). First, the greatest dynamic changes occurred outside of the substituted regions. Second, the dynamics on each timescale varied significantly for any given residue. We thus directly observe that one cannot extrapolate residue dynamics to timescales that have not been probed.

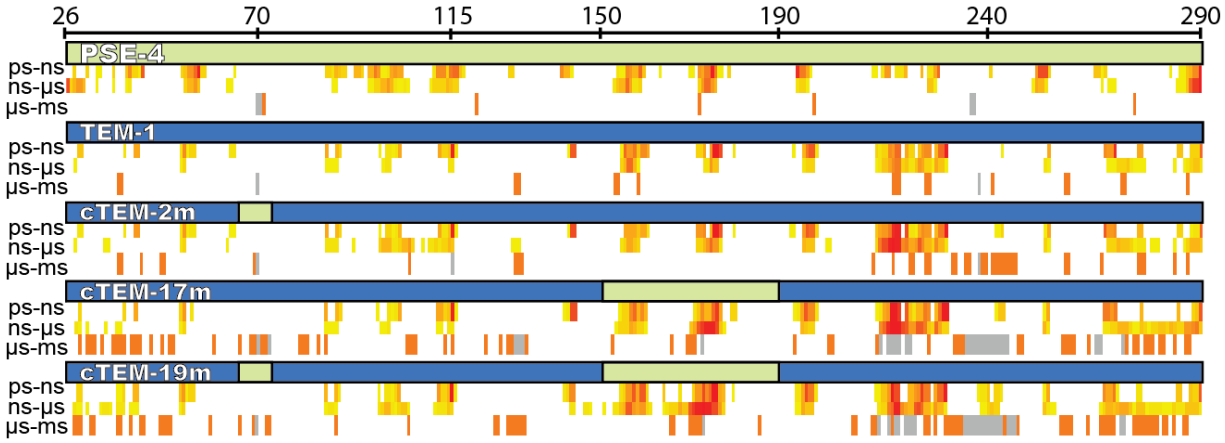


Figure A 4-5 Distribution of motions in the β-lactamases over the 3 observed timescales in relation to the linear sequence.

The sequence blocks originating from TEM-1 (blue) and PSE-4 (green) are colored in cTEM-2m, cTEM-17m and cTEM-19m on a blue/green scheme identifying their parental origin (as in Figure A 4-1). The ps to ms motions are colored as in Figure A 4-3.

Intriguingly, in all but one case where residue assignment and NMR observables allowed, the global exchange rate (k_{ex}) was conserved in each of the active-site walls, regardless of the recombinations. We were surprised that the frequency of slow motions did not vary more in the different sequence contexts, in particular in cTEM-17m and cTEM-19m where neighboring unassigned residues were indicative of new motions slower than the millisecond. The exception was a 4-fold difference in the global k_{ex} observed between Y105 walls of cTEM-17m and cTEM-19m, which may be functionally significant (*vide infra*). We emphasize that, while the slow motions observed by CPMG NMR (all between 300 s^{-1} and 2,000 s^{-1}) occurred at a rate comparable to turnover of the natural substrate benzylpenicillin (TEM-1 k_{cat}^{BZ} = 450 s^{-1}), catalytic efficiency was largely conserved and slow dynamic variations were tolerated; fast dynamics, however, were essentially conserved throughout this evolutionary system. The rate-limiting step for penicillin hydrolysis is the deacylation step [37]. Because hydrolytic efficiency was similar for all enzymes while slow dynamic differs, the late steps of catalysis are independent of slow dynamics. In contrast, the rate-limiting step for hydrolysis of the bulkier cephalosporins is the acylating step [38], therefore we conclude that the slow dynamics may have a functional impact on the rate-limiting acylating step in the catalytic mechanism (Table A 4-1). Rather than negating the relevance of dynamics to enzyme function [40], this underscores the importance of probing motions over multiple timescales to fully comprehend their effects on enzyme function and eventually permit engineering of motions. Ideally, all timescales should be treated as a continuum of motions: as previously observed for adenylate kinase, ps to ns motions may facilitate larger-scale, slower motions occurring on the timescale of function [41].

4.5.2. Protein motions in a model of evolution.

Mapping the dynamics of these interrelated β-lactamases provides a unique insight into patterns of dynamic conservation and tolerance to changes in dynamics, in a model of evolution. We demonstrate that variations in dynamics can be functionally tolerated in class A β-lactamases: all observed slow dynamic variations are compatible with native-like catalytic function toward penicillins yet new slow motions observed near Y105 of the chimeras may modulate hydrolysis of the bulkier cephalosporins. In contrast to the native TEM-1, where the

minor *m* conformer of Tyr105 is thought to stabilize binding of the penicillin substrate BZ but not of the cephalosporin substrate CF [35, 42], the atypical *m-30°* conformation was observed in the structures of cTEM-17m and cTEM-19m. The altered structure change in frequency of motions (k_{ex}) in the gate-keeping Tyr105 region may underlie altered cephalosporin hydrolysis.

As previously asked, does tolerance to changes in dynamics at a rate comparable to turnover provide evolvability? [43] Our system of artificially evolved chimeric enzymes mimics recombination events between TEM-1 and PSE-4. The chimeras displayed native-like rigidity accompanied by a pattern of specific motions on the fast and intermediate timescales, consistent with their importance toward reactivity. In parallel, we speculate that the observed tolerance to a variety of slow active-site dynamics may facilitate evolution and promote functional diversity, particularly substrate promiscuity in the case of this class of highly evolvable antibiotic resistance enzymes. It is striking that all three chimeras inspected exhibited important, different patterns of slow dynamics, which may constitute a mechanism of natural evolution in β-lactamases. Our results are strikingly similar to a recent report of point mutations in *Bacillus cereus* metallo-β-lactamase II: in that model of evolution, fast dynamics were conserved and important variations in slow active-site dynamics were observed [44]. As in natural evolution, features that favor evolvability may be transient. In another model, evolution of a bacterial phosphotriesterase into an arylesterase revealed links between changing protein dynamics and specialization toward one function to the detriment of the other. Crystallographic b-factors and short MD-simulations of variants over the evolutionary path demonstrated that some regions adopted new dynamics to test out conformations, later stabilizing the most productive conformations with further evolution, while other regions became more rigid [45]. Similarly, the new, slow dynamics observed in the chimeras could potentially become attenuated upon further evolution with the introduction of stabilizing mutations, locking in beneficial mutations while regaining a native-like rigidity. Indeed, the modern-day TEM-1 is thought to have a more rigid catalytic pocket than ancestral β-lactamases as a consequence of TEM-1 evolving to become more of a ‘specialist’ than the more ‘generalist’ ancient β-lactamases [46]. We do not yet understand the underlying mechanism relating those motions to function. While this system is advantageous by being

functionally promiscuous and diverse in homologs, current methodologies preclude ready inclusion of substrates in crystal structures and in NMR, limiting the depth of our mechanistic comprehension. Our robust MD simulations pave the way to inclusion of substrates in simulations that will eventually reach the timescale of turnover. Because evolution depends on function, the observation of conserved fast dynamics supports the importance of dynamics in function. The slow dynamic variation we have observed may serve to promote functional diversity: β-lactamases are specialized (and secondary) metabolic enzymes that respond to environmental stress to ensure bacterial survival. In contrast to enzymes of primary metabolism that have a lower requirement for constant evolution, bacterial resistance enzymes require greater adaptability to the consequences of sequence variation.

4.6. Material and methods

Protein expression and purification was performed as previously reported [14, 25, 32, 33]. Crystals were grown at 22°C in hanging drops with a 1:1 ratio of the protein and reservoir solutions. Diffraction data were collected at the Canadian Macromolecular Crystallography Facility Beamline 08ID-1 or with a Rigaku MicroMax 007 HF X-ray generator and a Rigaku Saturn 944 HG CCD detector. Initial phases were calculated by molecular replacement, with PDB 1ZG4. Iterative rounds of manual model building were performed. Table A 5-2 summarizes the data collection and refinement statistics. The active-site volume was estimated using 3V: Voss Volume Voxelator. For energy minimization, MD simulations and analysis, GROMACS 5.0.1 was used with the AMBER99SB-ILDN force field [47, 48]. Three 2 μs simulations with 1 ps compressed and 100 ps full precision trajectories for each of the five proteins was acquired. S^2 order parameters were calculated applying the model-free approach [49]. The Cα-RMSF was calculated for each residue and converted into b-factors. NMR experiments were performed as previously reported [14, 33]. [^{15}N] and [^{2}H,^{15}N]-labeled samples were characterized using ^{15}N TROSY relaxation-compensated Carr-Purcell-Meiboom-Gill (rcCPMG) experiments [50] on Agilent 800 MHz, 600 MHz and 500 MHz NMR spectrometers calibrated to 31.5°C. Residue fits and model analyses were performed using the full single-quantum ^{15}N-CPMG equation [51]. Substrate hydrolysis was monitored

according to initial steady-state velocities. More experimental details can be found in annex 5.3.

4.7. Acknowledgment

The authors thank M. Meyer and F. Arnold for the original cTEM-19m construct; T. Sprules and S. Al-Abdul-Wahid (Québec/Eastern Canada High Field NMR Facility) and M. Osborne (IRIC Biophysics Research Facilities). Computational resources were provided by Calcul Québec, Compute Canada, and High Performance Computing Center Stuttgart. This work was supported by NSERC Discovery Grants 227853 (J.N.P) and 2016-07555 (N.D.), and NIH award R01GM105978 (N.D.). S.G. is a FRQ-NT Graduate Scholar, M.E. a NSERC Vanier Canada Graduate Scholar and D.G. a NSERC Alexander Graham Bell Canada Graduate Scholar. N.D. holds a FRQ-S Research Scholar Junior 2 Career Award.

4.8. References

1. Ramanathan, A., et al., *Protein conformational populations and functionally relevant substates.* Acc Chem Res, 2014. **47**(1): p. 149-156.
2. Henzler-Wildman, K. and D. Kern, *Dynamic personalities of proteins.* Nature, 2007. **450**(7172): p. 964-972.
3. Gibbs, A.C., *Elements and modulation of functional dynamics.* J Med Chem, 2014. **57**(19): p. 7819-7837.
4. Bhabha, G., et al., *A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis.* Science, 2011. **332**(6026): p. 234-238.
5. Boehr, D.D., et al., *Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands.* Proc Natl Acad Sci U S A, 2010. **107**(4): p. 1373-1378.
6. Gagné, D., et al., *Perturbation of the Conformational Dynamics of an Active-Site Loop Alters Enzyme Activity.* Structure, 2015. **23**(12): p. 2256-2266.
7. Eisenmesser, E.Z., et al., *Intrinsic dynamics of an enzyme underlies catalysis.* Nature, 2005. **438**(7064): p. 117-121.

8. Klinman, J.P., *Dynamically achieved active site precision in enzyme catalysis.* Acc Chem Res, 2015. **48**(2): p. 449-456.
9. Keskin, O., R.L. Jernigan, and I. Bahar, *Proteins with similar architecture exhibit similar large-scale dynamic behavior.* Biophys J, 2000. **78**(4): p. 2093-2106.
10. Maguid, S., S. Fernandez-Alberti, and J. Echave, *Evolutionary conservation of protein vibrational dynamics.* Gene, 2008. **422**(1-2): p. 7-13.
11. Ramanathan, A. and P.K. Agarwal, *Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis.* PLoS Biol, 2011. **9**(11): p. e1001193.
12. Francis, K., V. Stojkovic, and A. Kohen, *Preservation of protein dynamics in dihydrofolate reductase evolution.* J Biol Chem, 2013. **288**(50): p. 35961-35968.
13. Gagné, D., et al., *Conservation of flexible residue clusters among structural and functional enzyme homologues.* J Biol Chem, 2012. **287**(53): p. 44289-44300.
14. Gobeil, S.M., et al., *Maintenance of native-like protein dynamics may not be required for engineering functional proteins.* Chemistry & biology, 2014. **21**(10): p. 1330-1340.
15. Liu, Y. and I. Bahar, *Sequence evolution correlates with structural dynamics.* Mol Biol Evol, 2012. **29**(9): p. 2253-2263.
16. Bhabha, G., et al., *Divergent evolution of protein conformational dynamics in dihydrofolate reductase.* Nat Struct Mol Biol, 2013. **20**(11): p. 1243-1249.
17. Bush, K. and G.A. Jacoby, *Updated functional classification of beta-lactamases.* Antimicrob Agents Chemother, 2010. **54**(3): p. 969-976.
18. Voigt, C.A., et al., *Protein building blocks preserved by recombination.* Nat Struct Biol, 2002. **9**(7): p. 553-558.
19. Meyer, M.M., L. Hochrein, and F.H. Arnold, *Structure-guided SCHEMA recombination of distantly related beta-lactamases.* Protein Eng Des Sel, 2006. **19**(12): p. 563-570.
20. Ambler, R.P., et al., *A standard numbering scheme for the class A beta-lactamases.* Biochem J, 1991. **276 (Pt 1)**: p. 269-270.
21. van den Bedem, H. and J.S. Fraser, *Integrative, dynamic structural biology at atomic resolution--it's about time.* Nat Methods, 2015. **12**(4): p. 307-318.
22. Marion, D., *An introduction to biological NMR spectroscopy.* Mol Cell Proteomics, 2013. **12**(11): p. 3006-3025.

23. Kleckner, I.R. and M.P. Foster, *An introduction to NMR-based approaches for measuring protein dynamics.* Biochim Biophys Acta, 2011. **1814**(8): p. 942-968.
24. Meroueh, S.O., et al., *Molecular dynamics at the root of expansion of function in the M69L inhibitor-resistant TEM beta-lactamase from Escherichia coli.* J Am Chem Soc, 2002. **124**(32): p. 9422-9430.
25. Clouthier, C.M., et al., *Chimeric beta-lactamases: global conservation of parental function and fast time-scale dynamics with increased slow motions.* PloS one, 2012. **7**(12): p. e52283.
26. Minasov, G., X. Wang, and B.K. Shoichet, *An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation.* J Am Chem Soc, 2002. **124**(19): p. 5333-5340.
27. Zeil, C., et al., *Network Analysis of Sequence-Function Relationships and Exploration of Sequence Space of TEM beta-Lactamases.* Antimicrob Agents Chemother, 2016. **60**(5): p. 2709-2717.
28. Swaren, P., et al., *X-ray structure of the Asn276Asp variant of the Escherichia coli TEM-1 beta-lactamase: direct observation of electrostatic modulation in resistance to inactivation by clavulanic acid.* Biochemistry, 1999. **38**(30): p. 9570-9576.
29. Lovell, S.C., et al., *The penultimate rotamer library.* Proteins, 2000. **40**(3): p. 389-408.
30. Doucet, N. and J.N. Pelletier, *Simulated annealing exploration of an active-site tyrosine in TEM-1 beta-lactamase suggests the existence of alternate conformations.* Proteins, 2007. **69**(2): p. 340-348.
31. Morin, S. and S.M. Gagne, *NMR dynamics of PSE-4 beta-lactamase: an interplay of ps-ns order and mus-ms motions in the active site.* Biophys J, 2009. **96**(11): p. 4681-4691.
32. Morin, S., et al., *Backbone resonance assignments of an artificially engineered TEM-1/PSE-4 Class A beta-lactamase chimera.* Biomol NMR Assign, 2010. **4**(2): p. 127-130.
33. Gobeil, S.M., et al., *15N, 13C and 1H backbone resonance assignments of an artificially engineered TEM-1/PSE-4 class A beta-lactamase chimera and its deconvoluted mutant.* Biomol NMR Assign, 2016. **10**(1): p. 93-99.

34. Mittermaier, A.K. and L.E. Kay, *Observing biological dynamics at atomic resolution using NMR.* Trends Biochem Sci, 2009. **34**(12): p. 601-611.
35. Diaz, N., et al., *Insights into the acylation mechanism of class A beta-lactamases from molecular dynamics simulations of the TEM-1 enzyme complexed with benzylpenicillin.* J Am Chem Soc, 2003. **125**(3): p. 672-684.
36. Doucet, N., P.Y. De Wals, and J.N. Pelletier, *Site-saturation mutagenesis of Tyr-105 reveals its importance in substrate stabilization and discrimination in TEM-1 beta-lactamase.* J Biol Chem, 2004. **279**(44): p. 46295-46303.
37. Christensen, H., M.T. Martin, and S.G. Waley, *Beta-lactamases as fully efficient enzymes. Determination of all the rate constants in the acyl-enzyme mechanism.* Biochem J, 1990. **266**(3): p. 853-861.
38. Page, M.G., *Extended-spectrum beta-lactamases: structure and kinetic mechanism.* Clin Microbiol Infect, 2008. **14 Suppl 1**: p. 63-74.
39. Choury, D., et al., *Nucleotide Sequence of the blaRTG-2 (CARB-5) Gene and Phylogeny of a New Group of Carbenicillinases.* Antimicrob Agents Chemother, 2000. **44**(4): p. 1070-1074.
40. Warshel, A. and R.P. Bora, *Perspective: Defining and quantifying the role of dynamics in enzyme catalysis.* J Chem Phys, 2016. **144**(18): p. 180901.
41. Henzler-Wildman, K.A., et al., *A hierarchy of timescales in protein dynamics is linked to enzyme catalysis.* Nature, 2007. **450**(7171): p. 913-916.
42. Diaz, N., et al., *Molecular dynamics simulations of the TEM-1 beta-lactamase complexed with cephalothin.* J Med Chem, 2005. **48**(3): p. 780-791.
43. Tokuriki, N. and D.S. Tawfik, *Protein dynamism and evolvability.* Science, 2009. **324**(5924): p. 203-207.
44. Gonzalez, M.M., et al., *Optimization of Conformational Dynamics in an Epistatic Evolutionary Trajectory.* Mol Biol Evol, 2016. **33**(7): p. 1768-1776.
45. Campbell, E., et al., *The role of protein dynamics in the evolution of new enzyme function.* Nat Chem Biol, 2016. **12**(11): p. 944-950.
46. Zou, T., et al., *Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme.* Mol Biol Evol, 2015. **32**(1): p. 132-143.

47. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-1958.
48. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
49. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.* J Am Chem Soc, 1982. **104**(17): p. 4546-4559.
50. Loria, J.P., M. Rance, and A.G. Palmer, 3rd, *A TROSY CPMG sequence for characterizing chemical exchange in large proteins.* Journal of biomolecular NMR, 1999. **15**(2): p. 151-155.
51. Manley, G. and J.P. Loria, *NMR insights into protein allostery.* Arch Biochem Biophys, 2012. **519**(2): p. 223-231.

Annex 5 - Supplemental material for Annex 4

5.1. Tables

Table A 5-1 Natural TEM variants at the 19 positions that differ between TEM-1 and cTEM-19m [1].

TEM-1	cTEM-19m	TEM variant	
M68	L	I	TEM-192
M69	T	I	TEM-32, TEM-37, TEM-40, TEM-159, TEM-182
		L	TEM-33, TEM-35, TEM-39, TEM-45, TEM-50, TEM-77, TEM-80, TEM-81, TEM-83, TEM-109, TEM-125, TEM-154, TEM-158, TEM-169, TEM-185, TEM-189, TEM-190
		V	TEM-34, TEM-97, TEM-36, TEM-38, TEM-78, TEM-82, TEM-151, TEM-152, TEM-160
A150	D	-	
H153	R	R	TEM-21, TEM-56, TEM-112, TEM-215
N154	Q	-	
M155	I	I	TEM-156, TEM-199
H158	K	N	TEM-127
V159	E	-	
W165	I	C	TEM-83
		G	TEM-169, TEM-190
		L	TEM-182
		R	TEM-39, TEM-78, TEM-125, TEM-185
E168	D	K	TEM-193
A172	G	-	
I173	K	V	TEM-132
P174	L	-	
N175	G	H	TEM-193, TEM-194, TEM-195
		I	TEM-138
E177	T	-	
M182	L	T	TEM-20, TEM-32, TEM-43, TEM-52, TEM-63, TEM-72, TEM-87, TEM-88, TEM-91, TEM-92, TEM-93, TEM-94, TEM-106, TEM-107, TEM-113, TEM-124, TEM-126, TEM-131, TEM-135, TEM-149, TEM-153, TEM-159, TEM-161, TEM-165, TEM-177, TEM-184, TEM-205, TEM-211, TEM-220
A184	K	V	TEM-116, TEM-157, TEM-162, TEM-181, TEM-187, TEM-197, TEM-205

M186	I	V	TEM-193, TEM-194, TEM-195
T188	S	K	TEM-148
		N	TEM-193

Table A 5-2 (A) Data collection and refinement statistics and (b) Cα RMSD and secondary structure matching RMSD (in parentheses).

Comparison for crystal structures of the chimeras cTEM-2m (PDB 4MEZ) and cTEM-19m (PDB 4R4R and 4R4S). Although the classical β-lactamase inhibitor tazobactam was included during crystallization for cTEM-19m (PDB 4R4S), no corresponding electron density was observed. The crystal of cTEM-19m grown in the absence of tazobactam (PDB 4R4R) provided a nearly identical structure (Cα-RMSD 0.04 Å).

(A)

	cTEM-19m PDB 4R4R	**PDB 4R4S**	**cTEM-2m PDB 4MEZ**
Data collection			
Space group	P 21 21 21	P 21 21 21	P1
Unit-cell parameters (Å)			
a	36.69	36.71	34.33
b	58.79	58.71	55.07
c	109.83	109.77	77.68
Resolution (Å)	1.20	1.10	2.05
Completeness (%)	98.3 (91.6)	98.3 (84.5)	99.2 (97.2)
Average redundancy	3.0 (2.6)	5.6 (3.2)	7.9 (5.9)
I/σ(I)	16.3 (4.3)	21.3 (3.4)	11.8 (3.5)
R-merge (%)	3.6 (23.6)	4.1 (28.6)	15.2 (39.2)
Refinement			
R-work (%)	10.5	10.4	19.2
R-free (%)	13.8	12.8	24.8
No. of water	408	464	346
r.m.s. deviations			
Bond lengths (Å)	0.021	0.021	0.020
Bond angles (°)	2.0	2.0	2.0

(B)

	TEM-1 *(1XPB - 1.9 Å)*	**cTEM-2m** *(4ID4 - 1.05 Å)*	**cTEM-19m** *(4R4S - 1.12 Å)*
TEM-1 *(1XPB - 1.9 Å)*	--	0.71 (0.42)	0.93 (0.51)
PSE-4 *(1G68 - 1.95 Å)*	1.35 (1.01)	1.42 (1.12)	1.44 (1.11)
cTEM-17m *(4MEZ - 2.05 Å)*	0.93 (0.55)	0.80 (0.59)	0.14 (0.28)

*Crystal structure resolution indicated next to PDB code.

Table A 5-3 Dynamics of the naturally evolved class A β-lactamases TEM-1 and PSE-4 and chimeric β-lactamases cTEM-2m, cTEM-17m and cTEM-19m.

Protein-averaged, squared generalized order parameters (S^2) obtained by MD simulations and NMR relaxation for the amide NH bonds and Cα-RMSF extracted from the MD simulations.

	NMR S^2	MD simulation S^2	RMSF (nm)
TEM-1	0.90 ± 0.05	0.87 ± 0.05	0.07 ± 0.03
PSE-4	0.87 ± 0.05	0.86 ± 0.05	0.08 ± 0.04
cTEM-2m	N.D.	0.87 ± 0.05	0.09 ± 0.04
cTEM-17m	0.89 ± 0.05	0.87 ± 0.05	0.09 ± 0.05
cTEM-19m	N.D.	0.86 ± 0.05	0.10 ± 0.04

N.D.:not determined.

Table A 5-4 Cα-RMSD between the average structure from each 2 μs simulation and the crystal structure used for that simulation.

TEM-1 (PDB 1XPB), PSE-4 (PDB 1G68), cTEM-2m (PDB 4MEZ), cTEM-17m (PDB 4ID4) and cTEM-19m (PDB 4R4S).

	Repetition 1	Repetition 2	Repetition 3	Average
PSE-4	1.21	1.11	1.40	1.24 ± 0.12
TEM-1	1.18	0.84	0.82	0.95 ± 0.20
cTEM-2m	1.14	1.03	1.18	1.12 ± 0.06
cTEM-17m	1.67	0.915	1.01	1.20 ± 0.34
cTEM-19m	1.69	1.624	1.52	1.61 ± 0.07

Table A 5-5 Per residue S^2 extracted from MD simulations.

For TEM-1, PSE-4, cTEM-2m, cTEM-17m and cTEM-19m compared to the NMR-determined S^2 for TEM-1, PSE-4 and cTEM-17m.

		TEM-1		PSE-4		cTEM-2m	cTEM-17m		cTEM-19m
TEM-1	Residue	MD	NMR[1]	MD	NMR[2]	MD	MD	NMR[3]	MD
26	His			0.829 ± 0.024	0.761 ± 0.017				
27	Pro			0.868 ± 0.007	0.877 ± 0.012				
28	Glu	0.897 ± 0.002		0.865 ± 0.016	0.838 ± 0.017	0.889 ± 0.014	0.890 ± 0.007	0.840 ± 0.017	0.886 ± 0.000
29	Thr	0.824 ± 0.005		0.826 ± 0.011	0.861 ± 0.010	0.838 ± 0.011	0.833 ± 0.004	0.946 ± 0.014	0.830 ± 0.000
30	Leu	0.903 ± 0.002	0.890 ± 0.020	0.888 ± 0.003	0.895 ± 0.012	0.902 ± 0.003	0.900 ± 0.006	0.891 ± 0.016	0.901 ± 0.000
31	Val	0.894 ± 0.006	0.910 ± 0.020	0.848 ± 0.006	0.857 ± 0.011	0.886 ± 0.020	0.882 ± 0.006	0.862 ± 0.011	0.882 ± 0.000
32	Lys	0.890 ± 0.004	0.910 ± 0.020	0.848 ± 0.002	0.858 ± 0.014	0.885 ± 0.007	0.884 ± 0.006	0.894 ± 0.020	0.884 ± 0.000
33	Val	0.893 ± 0.002	0.870 ± 0.010	0.863 ± 0.004	0.891 ± 0.010	0.890 ± 0.003	0.890 ± 0.005	o.l.	0.890 ± 0.000
34	Lys	0.903 ± 0.003	0.930 ± 0.020	0.891 ± 0.004	0.905 ± 0.012	0.899 ± 0.010	0.894 ± 0.007	0.916 ± 0.009	0.895 ± 0.000
35	Asp	0.884 ± 0.002	±	0.834 ± 0.007	0.873 ± 0.016	0.878 ± 0.009	0.877 ± 0.006	0.926 ± 0.011	0.872 ± 0.000
36	Ala	0.884 ± 0.003	0.910 ± 0.020	0.858 ± 0.006	0.868 ± 0.010	0.878 ± 0.009	0.873 ± 0.007	0.896 ± 0.018	0.868 ± 0.000
37	Glu	0.893 ± 0.002	0.940 ± 0.020	0.894 ± 0.004	o.l.	0.891 ± 0.001	0.887 ± 0.004	0.922 ± 0.011	0.884 ± 0.000
38	Asp	0.867 ± 0.002	0.880 ± 0.020	0.824 ± 0.013	0.844 ± 0.010	0.862 ± 0.008	0.855 ± 0.008	0.907 ± 0.013	0.851 ± 0.000
39	Gln	0.837 ± 0.003	0.890 ± 0.020	0.779 ± 0.007	0.875 ± 0.012	0.840 ± 0.005	0.828 ± 0.011	0.926 ± 0.012	0.829 ± 0.000
40	Leu	0.870 ± 0.002	0.930 ± 0.020	0.833 ± 0.003	o.l.	0.870 ± 0.001	0.862 ± 0.006	0.948 ± 0.021	0.859 ± 0.000
41	Gly	0.882 ± 0.001	0.910 ± 0.020	0.847 ± 0.004	0.849 ± 0.011	0.881 ± 0.001	0.880 ± 0.001		0.872 ± 0.000
42	Ala	0.799 ± 0.006	0.840 ± 0.010	0.752 ± 0.027	o.l.	0.790 ± 0.016	0.801 ± 0.003	o.l.	0.813 ± 0.000
43	Arg	0.876 ± 0.016		0.870 ± 0.004	0.868 ± 0.012	0.876 ± 0.008	0.882 ± 0.003	0.792 ± 0.019	0.863 ± 0.000
44	Val	0.916 ± 0.002	0.870 ± 0.020	0.887 ± 0.009	0.844 ± 0.011	0.917 ± 0.002	0.910 ± 0.005	o.l.	0.904 ± 0.000
45	Gly	0.895 ± 0.004	0.820 ± 0.010	0.895 ± 0.007	0.873 ± 0.011	0.888 ± 0.009	0.897 ± 0.010	o.l.	0.875 ± 0.000
46	Tyr	0.915 ± 0.002	0.890 ± 0.020	0.911 ± 0.002	0.879 ± 0.011	0.909 ± 0.005	0.910 ± 0.001	0.867 ± 0.020	0.911 ± 0.000
47	Ile	0.918 ± 0.004		0.918 ± 0.005	0.902 ± 0.012	0.920 ± 0.001	0.922 ± 0.001	0.940 ± 0.013	0.922 ± 0.000
48	Glu	0.914 ± 0.001	0.880 ± 0.020	0.921 ± 0.001	0.864 ± 0.012	0.914 ± 0.003	0.913 ± 0.002	0.890 ± 0.020	0.914 ± 0.000
49	Leu	0.914 ± 0.001	0.850 ± 0.020	0.908 ± 0.001	0.873 ± 0.012	0.916 ± 0.001	0.913 ± 0.001	0.842 ± 0.022	0.913 ± 0.000
50	Asp	0.906 ± 0.002		0.886 ± 0.008	0.867 ± 0.011	0.908 ± 0.001	0.905 ± 0.001	0.917 ± 0.030	0.907 ± 0.000
51	Leu	0.882 ± 0.001	0.930 ± 0.020	0.826 ± 0.018	0.811 ± 0.019	0.884 ± 0.001	0.885 ± 0.003	0.908 ± 0.021	0.885 ± 0.000
52	Asn	0.835 ± 0.013	0.840 ± 0.020	0.823 ± 0.026	0.829 ± 0.015	0.840 ± 0.006	0.830 ± 0.006	0.784 ± 0.018	0.822 ± 0.000
53	Ser	0.783 ± 0.008		0.792 ± 0.015	0.640 ± 0.015	0.786 ± 0.005	0.787 ± 0.011	0.829 ± 0.014	0.793 ± 0.000
54	Gly	0.850 ± 0.002	0.940 ± 0.020	0.761 ± 0.024	0.759 ± 0.019	0.850 ± 0.004	0.857 ± 0.005	0.924 ± 0.014	0.855 ± 0.000
55	Lys	0.833 ± 0.005	0.900 ± 0.010	0.690 ± 0.030	0.791 ± 0.015	0.839 ± 0.005	0.842 ± 0.007	o.l.	0.834 ± 0.000
56	Ile	0.872 ± 0.003	0.820 ± 0.020	0.831 ± 0.012	0.797 ± 0.017	0.876 ± 0.005	0.876 ± 0.004	0.856 ± 0.008	0.876 ± 0.000
57	Leu	0.880 ± 0.004	0.920 ± 0.020	0.879 ± 0.014	0.864 ± 0.018	0.885 ± 0.001	0.881 ± 0.003	0.908 ± 0.026	0.882 ± 0.000
58	Glu	0.883 ± 0.005	0.860 ± 0.020	Non existant in Ambler numbering		0.881 ± 0.009	0.877 ± 0.003	0.878 ± 0.009	0.879 ± 0.000
59	Ser	0.880 ± 0.003	0.870 ± 0.010	0.888 ± 0.002	0.842 ± 0.011	0.879 ± 0.001	0.883 ± 0.002	0.890 ± 0.012	0.882 ± 0.000
60	Phe	0.913 ± 0.003	0.830 ± 0.010	0.908 ± 0.002	0.848 ± 0.012	0.909 ± 0.008	0.916 ± 0.001	0.900 ± 0.010	0.913 ± 0.000
61	Arg	0.905 ± 0.001	0.870 ± 0.020	0.893 ± 0.004	0.898 ± 0.012	0.905 ± 0.001	0.904 ± 0.001	0.928 ± 0.014	0.905 ± 0.000
62	Pro			0.893 ± 0.007	0.883 ± 0.012				
63	Glu	0.897 ± 0.004	0.870 ± 0.020	0.858 ± 0.011	0.910 ± 0.013	0.897 ± 0.017	0.910 ± 0.001	0.901 ± 0.009	0.907 ± 0.000
64	Glu	0.835 ± 0.012	0.870 ± 0.020	0.850 ± 0.012	0.892 ± 0.012	0.848 ± 0.024	0.878 ± 0.007	0.932 ± 0.011	0.871 ± 0.000
65	Arg	0.885 ± 0.004	0.880 ± 0.020	0.881 ± 0.007	0.826 ± 0.011	0.879 ± 0.009	0.882 ± 0.003	0.818 ± 0.008	0.887 ± 0.000
66	Phe	0.895 ± 0.014	0.930 ± 0.020	0.887 ± 0.004	0.890 ± 0.013	0.893 ± 0.004	0.876 ± 0.027		0.883 ± 0.000
67	Pro								
68	Met	0.885 ± 0.005	0.980 ± 0.020	0.903 ± 0.006	0.940 ± 0.020	0.905 ± 0.003	0.896 ± 0.004	o.l.	0.896 ± 0.000
69	Met	0.922 ± 0.003	0.960 ± 0.020	0.922 ± 0.004	0.944 ± 0.015	0.920 ± 0.007	0.917 ± 0.007	o.l.	0.896 ± 0.000
70	Ser	0.904 ± 0.003	N.A.	0.882 ± 0.012	N.A.	0.874 ± 0.014	0.903 ± 0.003	N.A.	0.867 ± 0.000
71	Thr	0.880 ± 0.016	0.990 ± 0.010	0.898 ± 0.000	0.967 ± 0.013	0.870 ± 0.018	0.874 ± 0.007	o.l.	0.898 ± 0.000
72	Phe	0.908 ± 0.011	0.990 ± 0.020	0.886 ± 0.015	o.l.	0.891 ± 0.010	0.906 ± 0.007	o.l.	0.889 ± 0.000
73	Lys	0.920 ± 0.008	0.950 ± 0.020	0.916 ± 0.008	0.943 ± 0.012	0.910 ± 0.004	0.914 ± 0.008	N.A.	0.908 ± 0.000
74	Val	0.911 ± 0.002	0.910 ± 0.020	0.897 ± 0.004	0.892 ± 0.015	0.903 ± 0.001	0.911 ± 0.003	o.l.	0.891 ± 0.000
75	Leu	0.902 ± 0.002	0.920 ± 0.020	0.908 ± 0.003	0.922 ± 0.014	0.904 ± 0.006	0.901 ± 0.006	0.948 ± 0.023	0.918 ± 0.000
76	Leu	0.920 ± 0.008		0.913 ± 0.003	0.881 ± 0.015	0.918 ± 0.004	0.916 ± 0.011	o.l.	0.925 ± 0.000
77	Cys	0.915 ± 0.004	0.940 ± 0.020	0.917 ± 0.001	0.867 ± 0.012	0.915 ± 0.005	0.913 ± 0.005	0.882 ± 0.022	0.912 ± 0.000
78	Gly	0.908 ± 0.003	0.930 ± 0.020	0.906 ± 0.001	0.894 ± 0.016	0.909 ± 0.001	0.905 ± 0.006	0.935 ± 0.014	0.907 ± 0.000
79	Ala	0.911 ± 0.002	0.940 ± 0.020	0.918 ± 0.001	0.905 ± 0.012	0.912 ± 0.001	0.910 ± 0.002	0.947 ± 0.014	0.911 ± 0.000
80	Val	0.900 ± 0.004	0.900 ± 0.020	0.921 ± 0.001	o.l.	0.890 ± 0.016	0.889 ± 0.017	o.l.	0.888 ± 0.000
81	Leu	0.924 ± 0.001	0.940 ± 0.020	0.903 ± 0.004	0.884 ± 0.011	0.923 ± 0.000	0.922 ± 0.003	0.954 ± 0.010	0.922 ± 0.000
82	Ser	0.909 ± 0.001	0.950 ± 0.020	0.896 ± 0.002	0.915 ± 0.010	0.909 ± 0.003	0.907 ± 0.000	0.919 ± 0.007	0.910 ± 0.000
83	Arg	0.883 ± 0.004	0.910 ± 0.020	0.915 ± 0.001	0.905 ± 0.012	0.881 ± 0.002	0.880 ± 0.003	0.942 ± 0.009	0.881 ± 0.000
84	Val	0.902 ± 0.002	0.900 ± 0.010	0.891 ± 0.003	0.896 ± 0.013	0.900 ± 0.003	0.900 ± 0.003	0.891 ± 0.008	0.900 ± 0.000
85	Asp	0.911 ± 0.002		0.885 ± 0.001	0.880 ± 0.010	0.908 ± 0.004	0.907 ± 0.002	o.l.	0.909 ± 0.000
86	Ala	0.805 ± 0.015	0.890 ± 0.020	0.829 ± 0.003	0.817 ± 0.010	0.799 ± 0.007	0.803 ± 0.006	0.886 ± 0.006	0.810 ± 0.000
87	Gly	0.863 ± 0.003	0.880 ± 0.020	0.837 ± 0.002	0.852 ± 0.010	0.861 ± 0.003	0.862 ± 0.001	0.891 ± 0.006	0.865 ± 0.000
88	Gln	0.813 ± 0.002	0.900 ± 0.020	0.815 ± 0.008	0.878 ± 0.012	0.811 ± 0.002	0.809 ± 0.002	o.l.	0.813 ± 0.000
89	Glu	0.830 ± 0.004	0.830 ± 0.020	0.821 ± 0.003	0.794 ± 0.010	0.830 ± 0.003	0.831 ± 0.001	0.812 ± 0.007	0.837 ± 0.000
90	Gln	0.902 ± 0.000	0.840 ± 0.020	0.828 ± 0.002	o.l.	0.901 ± 0.004	0.902 ± 0.002	0.841 ± 0.008	0.901 ± 0.000
91	Leu	0.867 ± 0.003	0.880 ± 0.020		Pro	0.867 ± 0.001	0.870 ± 0.003	0.883 ± 0.011	0.874 ± 0.000
92	Gly	0.886 ± 0.003	0.910 ± 0.020	0.875 ± 0.000	0.851 ± 0.010	0.880 ± 0.013	0.887 ± 0.001	0.939 ± 0.011	0.886 ± 0.000
93	Arg	0.868 ± 0.012	0.930 ± 0.020	0.812 ± 0.006	0.841 ± 0.015	0.864 ± 0.020	0.877 ± 0.004	0.916 ± 0.010	0.877 ± 0.000
94	Arg	0.882 ± 0.003	0.880 ± 0.010	0.829 ± 0.001	0.847 ± 0.011	0.875 ± 0.011	0.883 ± 0.000	o.l.	0.880 ± 0.000
95	Ile	0.871 ± 0.004	0.900 ± 0.020	0.858 ± 0.002	0.859 ± 0.011	0.870 ± 0.001	0.871 ± 0.003	0.901 ± 0.011	0.871 ± 0.000

		TEM-1		PSE-4		cTEM-2m	cTEM-17m		cTEM-19m
TEM-1	Residue	MD	NMR[1]	MD	NMR[2]	MD	MD	NMR[3]	MD
96	His	0.869 ± 0.003	0.850 ± 0.020	0.814 ± 0.003	0.798 ± 0.010	0.866 ± 0.001	0.870 ± 0.002	0.859 ± 0.007	0.868 ± 0.000
97	Tyr	0.860 ± 0.007	0.780 ± 0.020	0.773 ± 0.013	0.805 ± 0.010	0.860 ± 0.001	0.862 ± 0.006	0.805 ± 0.010	0.857 ± 0.000
98	Ser	0.853 ± 0.003	0.830 ± 0.020	0.794 ± 0.004	0.834 ± 0.015	0.854 ± 0.003	0.851 ± 0.006	0.840 ± 0.008	0.846 ± 0.000
99	Gln	0.864 ± 0.002	0.870 ± 0.020	0.846 ± 0.006	o.l.	0.864 ± 0.002	0.866 ± 0.000	0.930 ± 0.010	0.866 ± 0.000
100	Asn	0.835 ± 0.002	0.920 ± 0.020	0.835 ± 0.003	0.907 ± 0.012	0.835 ± 0.003	0.836 ± 0.002	0.959 ± 0.012	0.836 ± 0.000
101	Asp	0.845 ± 0.005	0.910 ± 0.020	0.802 ± 0.009	0.831 ± 0.011	0.847 ± 0.004	0.846 ± 0.006	0.857 ± 0.009	0.848 ± 0.000
102	Leu	0.838 ± 0.004	0.880 ± 0.020	0.789 ± 0.020	0.854 ± 0.011	0.843 ± 0.006	0.838 ± 0.007	0.876 ± 0.009	0.829 ± 0.000
103	Val	0.801 ± 0.005	0.900 ± 0.020	0.804 ± 0.012	o.l.	0.811 ± 0.008	0.801 ± 0.004	0.837 ± 0.022	0.800 ± 0.000
104	Glu	0.885 ± 0.010	0.860 ± 0.020	0.872 ± 0.014	o.l.	0.883 ± 0.001	0.887 ± 0.005	0.916 ± 0.010	0.869 ± 0.000
105	Tyr	0.873 ± 0.006	0.800 ± 0.020	0.864 ± 0.020	0.802 ± 0.018	0.863 ± 0.008	0.873 ± 0.002	0.891 ± 0.027	0.862 ± 0.000
106	Ser	0.879 ± 0.002	0.810 ± 0.020	0.876 ± 0.008	0.825 ± 0.012	0.863 ± 0.022	0.874 ± 0.002	0.903 ± 0.011	0.870 ± 0.000
107	Pro								
108	Val	0.881 ± 0.003		0.883 ± 0.002	0.899 ± 0.012	0.876 ± 0.003	0.872 ± 0.003	0.915 ± 0.017	0.877 ± 0.000
109	Thr	0.920 ± 0.001	0.870 ± 0.020	0.901 ± 0.001	0.882 ± 0.015	0.917 ± 0.002	0.919 ± 0.001	o.l.	0.918 ± 0.000
110	Glu	0.891 ± 0.002	0.960 ± 0.020	0.904 ± 0.003	0.911 ± 0.014	0.888 ± 0.002	0.891 ± 0.001	0.966 ± 0.008	0.885 ± 0.000
111	Lys	0.865 ± 0.003	0.900 ± 0.020	0.844 ± 0.004	0.845 ± 0.014	0.857 ± 0.010	0.863 ± 0.002	0.897 ± 0.008	0.859 ± 0.000
112	His	0.803 ± 0.002	0.860 ± 0.020	0.780 ± 0.007	0.823 ± 0.012	0.795 ± 0.011	0.800 ± 0.001	0.822 ± 0.010	0.793 ± 0.000
113	Leu	0.834 ± 0.001	0.900 ± 0.020	0.778 ± 0.007	0.810 ± 0.010	0.835 ± 0.010	0.834 ± 0.005	o.l.	0.828 ± 0.000
114	Thr	0.817 ± 0.000	0.890 ± 0.020	0.826 ± 0.003	0.892 ± 0.011	0.818 ± 0.008	0.817 ± 0.005	0.887 ± 0.013	0.812 ± 0.000
115	Asp	0.708 ± 0.005	0.880 ± 0.020	0.768 ± 0.008	0.856 ± 0.012	0.711 ± 0.012	0.714 ± 0.002	o.l.	0.703 ± 0.000
116	Gly	0.842 ± 0.003		0.756 ± 0.016	0.756 ± 0.010	0.845 ± 0.001	0.847 ± 0.004	0.921 ± 0.008	0.838 ± 0.000
117	Met	0.887 ± 0.006	0.890 ± 0.020	0.822 ± 0.008	0.825 ± 0.011	0.883 ± 0.006	0.889 ± 0.002	o.l.	0.886 ± 0.000
118	Thr	0.894 ± 0.002	0.890 ± 0.020	0.836 ± 0.001	0.868 ± 0.120	0.891 ± 0.002	0.894 ± 0.000	0.879 ± 0.011	0.892 ± 0.000
119	Val	0.913 ± 0.000	0.910 ± 0.020	0.900 ± 0.001	0.863 ± 0.016	0.911 ± 0.001	0.912 ± 0.000	0.876 ± 0.010	0.912 ± 0.000
120	Arg	0.912 ± 0.004	0.930 ± 0.020	0.911 ± 0.002	0.873 ± 0.015	0.908 ± 0.004	0.912 ± 0.001	0.962 ± 0.011	0.912 ± 0.000
121	Glu	0.891 ± 0.001	0.940 ± 0.020	0.893 ± 0.001	0.894 ± 0.012	0.886 ± 0.004	0.888 ± 0.003	0.940 ± 0.011	0.895 ± 0.000
122	Leu	0.922 ± 0.001	0.920 ± 0.020	0.913 ± 0.000	0.898 ± 0.012	0.920 ± 0.001	0.921 ± 0.000	0.943 ± 0.013	0.922 ± 0.000
123	Cys	0.918 ± 0.001	0.940 ± 0.010	0.918 ± 0.001	0.875 ± 0.014	0.916 ± 0.002	0.917 ± 0.000	o.l.	0.916 ± 0.000
124	Ser	0.900 ± 0.004		0.913 ± 0.003	0.896 ± 0.014	0.899 ± 0.003	0.900 ± 0.002		0.903 ± 0.000
125	Ala	0.901 ± 0.006	0.990 ± 0.020	0.920 ± 0.002	o.l.	0.885 ± 0.014	0.900 ± 0.003		0.901 ± 0.000
126	Ala	0.927 ± 0.002		0.902 ± 0.006	0.925 ± 0.011	0.923 ± 0.004	0.926 ± 0.001		0.924 ± 0.000
127	Ile	0.896 ± 0.007	0.880 ± 0.020	0.892 ± 0.001	0.943 ± 0.018	0.903 ± 0.002	0.897 ± 0.005		0.888 ± 0.000
128	Thr	0.895 ± 0.008		0.888 ± 0.021	0.917 ± 0.017	0.884 ± 0.013	0.888 ± 0.005	o.l.	0.885 ± 0.000
129	Met	0.881 ± 0.004	0.880 ± 0.010	0.827 ± 0.019	0.924 ± 0.017	0.869 ± 0.016	0.865 ± 0.013	0.944 ± 0.024	0.856 ± 0.000
130	Ser	0.877 ± 0.012	0.990 ± 0.010	0.864 ± 0.005	0.952 ± 0.020	0.871 ± 0.009	0.872 ± 0.003	o.l.	0.868 ± 0.000
131	Asp	0.908 ± 0.006	0.990 ± 0.020	0.887 ± 0.008	0.920 ± 0.012	0.893 ± 0.005	0.895 ± 0.008	N.A.	0.899 ± 0.000
132	Asn	0.884 ± 0.015		0.884 ± 0.015	0.966 ± 0.013	0.875 ± 0.007	0.870 ± 0.016	N.A.	0.876 ± 0.000
133	Thr	0.899 ± 0.002	0.960 ± 0.020	0.890 ± 0.004	0.916 ± 0.020	0.892 ± 0.005	0.893 ± 0.001	0.913 ± 0.016	0.893 ± 0.000
134	Ala	0.937 ± 0.001	0.920 ± 0.010	0.933 ± 0.003	0.897 ± 0.011	0.934 ± 0.005	0.938 ± 0.001	0.923 ± 0.012	0.933 ± 0.000
135	Ala	0.915 ± 0.003	0.900 ± 0.020	0.905 ± 0.003	0.886 ± 0.011	0.909 ± 0.001	0.907 ± 0.004	0.904 ± 0.018	0.910 ± 0.000
136	Asn	0.913 ± 0.007	0.930 ± 0.020	0.877 ± 0.029	0.885 ± 0.014	0.909 ± 0.002	0.909 ± 0.009	o.l.	0.884 ± 0.000
137	Leu	0.922 ± 0.002	0.950 ± 0.020	0.903 ± 0.004	0.896 ± 0.013	0.921 ± 0.000	0.918 ± 0.004		0.911 ± 0.000
138	Leu	0.912 ± 0.002	0.900 ± 0.020	0.896 ± 0.005	0.884 ± 0.016	0.911 ± 0.001	0.907 ± 0.004	0.961 ± 0.012	0.898 ± 0.000
139	Leu	0.926 ± 0.003	0.900 ± 0.020	0.903 ± 0.018	0.905 ± 0.012	0.922 ± 0.001	0.917 ± 0.004	o.l.	0.909 ± 0.000
140	Thr	0.907 ± 0.002	0.890 ± 0.010	0.884 ± 0.002	0.858 ± 0.012	0.907 ± 0.002	0.904 ± 0.005	0.888 ± 0.008	0.892 ± 0.000
141	Thr	0.874 ± 0.007		0.777 ± 0.017	0.814 ± 0.019	0.868 ± 0.005	0.848 ± 0.004		0.839 ± 0.000
142	Ile	0.851 ± 0.002	0.900 ± 0.020	0.808 ± 0.011	0.819 ± 0.011	0.848 ± 0.003	0.842 ± 0.007	0.835 ± 0.018	0.833 ± 0.000
143	Gly	0.746 ± 0.004	0.890 ± 0.020	0.846 ± 0.009	0.907 ± 0.012	0.735 ± 0.009	0.741 ± 0.009	0.872 ± 0.009	0.781 ± 0.000
144	Gly		0.880 ± 0.020	0.894 ± 0.002	0.902 ± 0.012			0.710 ± 0.009	
145	Pro								
146	Lys	0.901 ± 0.001	0.920 ± 0.020	0.904 ± 0.001	o.l.	0.903 ± 0.003	0.895 ± 0.005	o.l.	0.887 ± 0.000
147	Glu	0.893 ± 0.001	0.880 ± 0.020	0.896 ± 0.002	0.897 ± 0.012	0.894 ± 0.002	0.886 ± 0.005	0.889 ± 0.010	0.883 ± 0.000
148	Leu	0.906 ± 0.002	0.930 ± 0.020	0.910 ± 0.003	0.872 ± 0.014	0.905 ± 0.001	0.901 ± 0.005	0.913 ± 0.013	0.891 ± 0.000
149	Thr	0.910 ± 0.001	0.910 ± 0.020	0.903 ± 0.001	0.907 ± 0.011	0.912 ± 0.003	0.907 ± 0.002	o.l.	0.898 ± 0.000
150	Ala	0.904 ± 0.001	0.910 ± 0.020	0.912 ± 0.004	0.920 ± 0.012	0.907 ± 0.001	0.905 ± 0.002	0.942 ± 0.006	0.898 ± 0.000
151	Phe	0.877 ± 0.002	0.880 ± 0.020	0.885 ± 0.010	0.906 ± 0.011	0.892 ± 0.009	0.876 ± 0.016	0.926 ± 0.007	0.877 ± 0.000
152	Leu	0.902 ± 0.000		0.909 ± 0.008	0.900 ± 0.017	0.908 ± 0.006	0.915 ± 0.001	o.l.	0.908 ± 0.000
153	His	0.897 ± 0.003	0.910 ± 0.020	0.891 ± 0.019	0.907 ± 0.015	0.904 ± 0.008	0.903 ± 0.005	o.l.	0.885 ± 0.000
154	Asn	0.849 ± 0.004	0.910 ± 0.020	0.843 ± 0.008	0.889 ± 0.012	0.866 ± 0.019	0.850 ± 0.008	0.889 ± 0.009	0.841 ± 0.000
155	Met	0.762 ± 0.007	0.870 ± 0.020	0.845 ± 0.025	0.722 ± 0.024	0.797 ± 0.041	0.827 ± 0.026	0.864 ± 0.008	0.828 ± 0.000
156	Gly	0.789 ± 0.014	0.920 ± 0.020	0.778 ± 0.035	0.887 ± 0.012	0.796 ± 0.027	0.783 ± 0.019	0.909 ± 0.009	0.794 ± 0.000
157	Asp	0.767 ± 0.012	0.910 ± 0.020	0.790 ± 0.023	0.884 ± 0.013	0.800 ± 0.038	0.762 ± 0.097	0.899 ± 0.005	0.747 ± 0.000
158	His	0.822 ± 0.009	0.780 ± 0.020	0.815 ± 0.024	0.845 ± 0.013	0.822 ± 0.028	0.811 ± 0.027	o.l.	0.797 ± 0.000
159	Val	0.781 ± 0.022	0.890 ± 0.020	0.760 ± 0.051	o.l.	0.800 ± 0.013	0.781 ± 0.014	0.909 ± 0.008	0.751 ± 0.000
160	Thr	0.805 ± 0.026	0.880 ± 0.020	0.816 ± 0.043	0.842 ± 0.011	0.837 ± 0.037	0.810 ± 0.012	0.840 ± 0.017	0.828 ± 0.000
161	Arg	0.898 ± 0.009	0.950 ± 0.020	0.915 ± 0.008	0.909 ± 0.016	0.891 ± 0.016	0.880 ± 0.050		0.875 ± 0.000
162	Leu	0.913 ± 0.001		0.911 ± 0.003	0.902 ± 0.013	0.907 ± 0.011	0.895 ± 0.022	0.914 ± 0.014	0.879 ± 0.000
163	Asp	0.929 ± 0.002	0.850 ± 0.020	0.912 ± 0.015	0.859 ± 0.019	0.924 ± 0.006	0.910 ± 0.026	0.810 ± 0.044	0.887 ± 0.000
164	Arg	0.908 ± 0.001	0.820 ± 0.020	0.854 ± 0.054	0.822 ± 0.014	0.890 ± 0.024	0.898 ± 0.011	0.793 ± 0.033	0.823 ± 0.000
165	Trp	0.898 ± 0.004		0.869 ± 0.008	0.817 ± 0.013	0.896 ± 0.002	0.868 ± 0.010	0.784 ± 0.026	0.859 ± 0.000

		TEM-1		PSE-4		cTEM-2m	cTEM-17m		cTEM-19m
TEM-1	Residue	MD	NMR[1]	MD	NMR[2]	MD	MD	NMR[3]	MD
166	Glu	0.910 ± 0.015	0.890 ± 0.030	0.891 ± 0.017	0.906 ± 0.015	0.903 ± 0.012	0.907 ± 0.013	o.l.	0.863 ± 0.000
167	Pro								
168	Glu	0.876 ± 0.003	0.890 ± 0.020	0.873 ± 0.004	o.l.	0.873 ± 0.003	0.862 ± 0.015	o.l.	0.838 ± 0.000
169	Leu	0.911 ± 0.001	0.950 ± 0.030	0.890 ± 0.014	0.925 ± 0.014	0.913 ± 0.005	0.897 ± 0.017	o.l.	0.869 ± 0.000
170	Asn	0.878 ± 0.009	0.980 ± 0.020	0.862 ± 0.027	o.l.	0.891 ± 0.007	0.864 ± 0.021	o.l.	0.861 ± 0.000
171	Glu	0.892 ± 0.006	0.940 ± 0.020	0.854 ± 0.014	0.912 ± 0.011	0.902 ± 0.003	0.838 ± 0.046		0.826 ± 0.000
172	Ala	0.833 ± 0.023	0.930 ± 0.000	0.887 ± 0.005	0.874 ± 0.014	0.787 ± 0.038	0.819 ± 0.072	o.l.	0.792 ± 0.000
173	Ile	0.801 ± 0.014		0.791 ± 0.014	o.l.	0.784 ± 0.050	0.775 ± 0.011	N.A.	0.736 ± 0.000
174	Pro			0.801 ± 0.001	o.l.		0.815 ± 0.040	o.l.	0.777 ± 0.000
175	Asn	0.793 ± 0.007	0.850 ± 0.020	0.769 ± 0.017	0.927 ± 0.031	0.808 ± 0.061	0.756 ± 0.034	o.l.	0.758 ± 0.000
176	Asp	0.570 ± 0.029	0.830 ± 0.020	0.669 ± 0.064	0.883 ± 0.012	0.567 ± 0.057	0.732 ± 0.055	0.893 ± 0.007	0.705 ± 0.000
177	Glu	0.741 ± 0.016	0.840 ± 0.020	0.766 ± 0.048	0.854 ± 0.014	0.753 ± 0.058	0.795 ± 0.052	0.894 ± 0.019	0.800 ± 0.000
178	Arg	0.895 ± 0.002	0.920 ± 0.020	0.847 ± 0.047	0.918 ± 0.017	0.856 ± 0.044	0.857 ± 0.023	0.943 ± 0.013	0.812 ± 0.000
179	Asp	0.921 ± 0.002	0.940 ± 0.020	0.919 ± 0.006	0.878 ± 0.013	0.911 ± 0.013	0.907 ± 0.021		0.900 ± 0.000
180	Thr	0.836 ± 0.011	0.900 ± 0.020	0.828 ± 0.029	0.867 ± 0.012	0.851 ± 0.020	0.830 ± 0.000	0.894 ± 0.021	0.848 ± 0.000
181	Thr	0.919 ± 0.001	0.900 ± 0.020	0.916 ± 0.007	0.889 ± 0.012	0.919 ± 0.002	0.913 ± 0.012	o.l.	0.914 ± 0.000
182	Met	0.882 ± 0.004	0.900 ± 0.020	0.872 ± 0.010	0.858 ± 0.012	0.883 ± 0.005	0.865 ± 0.015	o.l.	0.876 ± 0.000
183	Pro								
184	Ala	0.923 ± 0.003	0.920 ± 0.020	0.889 ± 0.012	o.l.	0.925 ± 0.002	0.924 ± 0.001	0.916 ± 0.015	0.922 ± 0.000
185	Ala	0.858 ± 0.007	0.830 ± 0.020	0.887 ± 0.014	o.l.	0.858 ± 0.002	0.913 ± 0.003	0.913 ± 0.014	0.896 ± 0.000
186	Met	0.911 ± 0.001	0.920 ± 0.020	0.914 ± 0.008	0.885 ± 0.015	0.910 ± 0.004	0.911 ± 0.006	0.933 ± 0.009	0.916 ± 0.000
187	Ala	0.922 ± 0.002	0.940 ± 0.020	0.890 ± 0.013	0.930 ± 0.012	0.917 ± 0.002	0.916 ± 0.006	o.l.	0.922 ± 0.000
188	Thr	0.917 ± 0.001		0.910 ± 0.002	o.l.	0.915 ± 0.003	0.913 ± 0.000	0.894 ± 0.007	0.907 ± 0.000
189	Thr	0.909 ± 0.001	0.930 ± 0.020	0.899 ± 0.011	o.l.	0.912 ± 0.005	0.914 ± 0.003	0.893 ± 0.012	0.908 ± 0.000
190	Leu	0.923 ± 0.004	0.920 ± 0.020	0.911 ± 0.005	0.929 ± 0.012	0.918 ± 0.002	0.918 ± 0.004	0.965 ± 0.010	0.921 ± 0.000
191	Arg	0.917 ± 0.002	0.950 ± 0.010	0.916 ± 0.003	0.895 ± 0.016	0.918 ± 0.001	0.921 ± 0.002	0.924 ± 0.010	0.916 ± 0.000
192	Lys	0.921 ± 0.003	0.910 ± 0.020	0.866 ± 0.023	0.892 ± 0.017	0.921 ± 0.001	0.918 ± 0.002	0.935 ± 0.008	0.915 ± 0.000
193	Leu	0.880 ± 0.017	0.940 ± 0.020	0.860 ± 0.011	0.862 ± 0.012	0.850 ± 0.018	0.863 ± 0.014	0.906 ± 0.009	0.872 ± 0.000
194	Leu	0.850 ± 0.011	0.880 ± 0.020	0.881 ± 0.013	0.884 ± 0.012	0.850 ± 0.009	0.852 ± 0.018	o.l.	0.865 ± 0.000
195	Thr	0.866 ± 0.025		0.883 ± 0.015	0.853 ± 0.014	0.870 ± 0.019	0.842 ± 0.028	0.918 ± 0.011	0.837 ± 0.000
196	Gly	0.808 ± 0.006	0.770 ± 0.010	0.717 ± 0.032	0.732 ± 0.017	0.795 ± 0.010	0.799 ± 0.014	0.780 ± 0.026	0.799 ± 0.000
197	Glu	0.767 ± 0.009	0.790 ± 0.020	0.766 ± 0.013	0.798 ± 0.025	0.749 ± 0.011	0.748 ± 0.020	0.859 ± 0.009	0.744 ± 0.000
198	Leu	0.751 ± 0.011	0.670 ± 0.020	0.803 ± 0.013	0.800 ± 0.017	0.757 ± 0.002	0.776 ± 0.012	0.720 ± 0.006	0.773 ± 0.000
199	Leu	0.836 ± 0.013	±	0.850 ± 0.012	0.838 ± 0.011	0.834 ± 0.004	0.836 ± 0.005	o.l.	0.838 ± 0.000
200	Thr	0.887 ± 0.014	0.890 ± 0.020	0.888 ± 0.003	0.821 ± 0.011	0.882 ± 0.004	0.881 ± 0.010	0.887 ± 0.011	0.884 ± 0.000
201	Leu	0.887 ± 0.001	0.870 ± 0.020	0.889 ± 0.001	0.868 ± 0.012	0.887 ± 0.004	0.887 ± 0.003	0.870 ± 0.013	0.889 ± 0.000
202	Ala	0.885 ± 0.009	0.910 ± 0.020	0.874 ± 0.002	0.870 ± 0.011	0.888 ± 0.005	0.887 ± 0.005	0.923 ± 0.005	0.885 ± 0.000
203	Ser	0.872 ± 0.004	0.900 ± 0.020	0.870 ± 0.002	0.889 ± 0.011	0.873 ± 0.002	0.873 ± 0.003	0.907 ± 0.005	0.874 ± 0.000
204	Arg	0.911 ± 0.002		0.916 ± 0.001	o.l.	0.910 ± 0.002	0.910 ± 0.001	0.929 ± 0.011	0.911 ± 0.000
205	Gln	0.887 ± 0.007	0.900 ± 0.020	0.909 ± 0.000	0.869 ± 0.011	0.893 ± 0.005	0.882 ± 0.011	0.951 ± 0.011	0.882 ± 0.000
206	Gln	0.866 ± 0.005	0.940 ± 0.020	0.896 ± 0.001	0.871 ± 0.012	0.868 ± 0.004	0.871 ± 0.005	0.921 ± 0.007	0.872 ± 0.000
207	Leu	0.913 ± 0.000	0.930 ± 0.020	0.915 ± 0.001	0.888 ± 0.016	0.911 ± 0.002	0.913 ± 0.001	0.950 ± 0.054	0.912 ± 0.000
208	Ile	0.905 ± 0.002	0.940 ± 0.020	0.920 ± 0.002	0.901 ± 0.012	0.901 ± 0.005	0.892 ± 0.011	0.915 ± 0.021	0.889 ± 0.000
209	Asp	0.880 ± 0.004	0.920 ± 0.020	0.909 ± 0.001	0.884 ± 0.011	0.882 ± 0.003	0.886 ± 0.002	0.959 ± 0.011	0.887 ± 0.000
210	Trp	0.872 ± 0.009	0.930 ± 0.020	0.869 ± 0.005	0.900 ± 0.011	0.875 ± 0.007	0.886 ± 0.011	0.923 ± 0.012	0.892 ± 0.000
211	Met	0.862 ± 0.025	0.950 ± 0.020	0.887 ± 0.005	o.l.	0.870 ± 0.014	0.868 ± 0.017	0.966 ± 0.014	0.874 ± 0.000
212	Glu	0.895 ± 0.010	0.960 ± 0.030	0.925 ± 0.001	0.952 ± 0.014	0.886 ± 0.018	0.896 ± 0.002	o.l.	0.898 ± 0.000
213	Ala	0.828 ± 0.018	0.920 ± 0.020	0.870 ± 0.009	0.894 ± 0.012	0.822 ± 0.009	0.851 ± 0.013		0.854 ± 0.000
214	Asp	0.841 ± 0.034	0.930 ± 0.030	0.854 ± 0.025	0.851 ± 0.010	0.797 ± 0.046	0.844 ± 0.010	N.A.	0.851 ± 0.000
215	Lys	0.831 ± 0.017		0.876 ± 0.016	0.895 ± 0.015	0.787 ± 0.024	0.800 ± 0.010	o.l.	0.819 ± 0.000
216	Val	0.775 ± 0.015		0.843 ± 0.010	0.904 ± 0.011	0.772 ± 0.017	0.789 ± 0.012	N.A.	0.792 ± 0.000
217	Ala	0.790 ± 0.077		0.786 ± 0.017	0.918 ± 0.012	0.722 ± 0.030	0.699 ± 0.012	N.A.	0.694 ± 0.000
218	Gly	0.764 ± 0.065	0.930 ± 0.020	0.882 ± 0.006		0.710 ± 0.086	0.726 ± 0.032	N.A.	0.735 ± 0.000
219	Pro			0.844 ± 0.007	0.896 ± 0.016				
220	Leu	0.804 ± 0.059	0.860 ± 0.020	0.842 ± 0.021	0.866 ± 0.014	0.743 ± 0.066	0.750 ± 0.017	N.A.	
221	Leu	0.834 ± 0.009		0.812 ± 0.017	0.855 ± 0.028	0.842 ± 0.009	0.819 ± 0.015	N.A.	0.808 ± 0.000
222	Arg	0.862 ± 0.032	0.960 ± 0.020	0.914 ± 0.009	0.923 ± 0.020	0.845 ± 0.023	0.838 ± 0.016	o.l.	0.791 ± 0.000
223	Ser	0.823 ± 0.027		0.863 ± 0.010	0.907 ± 0.014	0.803 ± 0.016	0.810 ± 0.010	o.l.	0.776 ± 0.000
224	Ala	0.791 ± 0.034	0.840 ± 0.020	0.857 ± 0.010	0.885 ± 0.012	0.779 ± 0.058	0.767 ± 0.024	0.889 ± 0.011	0.730 ± 0.000
225	Leu	0.779 ± 0.022	0.860 ± 0.020	0.810 ± 0.010	0.814 ± 0.010	0.790 ± 0.024	0.782 ± 0.025	0.829 ± 0.022	0.767 ± 0.000
226	Pro								
227	Ala	0.807 ± 0.002	0.790 ± 0.020	0.844 ± 0.018	0.814 ± 0.012	0.811 ± 0.015	0.808 ± 0.007	0.819 ± 0.010	0.805 ± 0.000
228	Gly	0.763 ± 0.005	0.800 ± 0.020	0.800 ± 0.012	0.777 ± 0.015	0.772 ± 0.009	0.763 ± 0.002	0.839 ± 0.014	0.765 ± 0.000
229	Trp	0.700 ± 0.029	0.900 ± 0.020	0.760 ± 0.037	0.852 ± 0.011	0.702 ± 0.027	0.686 ± 0.016	0.909 ± 0.012	0.686 ± 0.000
230	Phe	0.857 ± 0.004	0.920 ± 0.020	0.882 ± 0.012	0.817 ± 0.015	0.861 ± 0.005	0.862 ± 0.004	0.898 ± 0.020	0.857 ± 0.000
231	Ile	0.890 ± 0.002	0.910 ± 0.020	0.886 ± 0.007	0.846 ± 0.011	0.892 ± 0.002	0.895 ± 0.001	0.879 ± 0.036	0.888 ± 0.000
232	Ala	0.899 ± 0.006	0.880 ± 0.020	0.909 ± 0.001	0.859 ± 0.016	0.900 ± 0.002	0.904 ± 0.002	o.l.	0.898 ± 0.000
233	Asp	0.887 ± 0.011	0.910 ± 0.020	0.882 ± 0.005	0.858 ± 0.016	0.879 ± 0.012	0.883 ± 0.003		0.887 ± 0.000
234	Lys	0.927 ± 0.002	0.990 ± 0.010	0.926 ± 0.004	0.881 ± 0.023	0.921 ± 0.005	0.925 ± 0.003	N.A.	0.925 ± 0.000
235	Ser	0.873 ± 0.016		0.895 ± 0.010		0.881 ± 0.006	0.881 ± 0.007	N.A.	0.874 ± 0.000

		TEM-1		PSE-4		cTEM-2m	cTEM-17m		cTEM-19m
TEM-1	**Residue**	**MD**	**NMR[1]**	**MD**	**NMR[2]**	**MD**	**MD**	**NMR[3]**	**MD**
236	Gly	0.879 ± 0.011	0.900 ± 0.030	0.899 ± 0.004	0.882 ± 0.022	0.872 ± 0.009	0.872 ± 0.018	N.A.	0.874 ± 0.000
237	Ala	0.862 ± 0.006	N.A.	0.889 ± 0.004	N.A.	0.875 ± 0.011	0.867 ± 0.020	N.A.	0.849 ± 0.000
238	Gly	0.887 ± 0.001	0.880 ± 0.020	0.894 ± 0.006	0.848 ± 0.011	0.887 ± 0.004	0.877 ± 0.015	N.A.	0.843 ± 0.000
240	Glu	0.891 ± 0.003	0.970 ± 0.020	0.891 ± 0.004	0.914 ± 0.012	0.888 ± 0.006	0.871 ± 0.016	N.A.	0.822 ± 0.000
241	Arg	0.905 ± 0.002	0.960 ± 0.020	0.891 ± 0.004	0.922 ± 0.013	0.895 ± 0.006	0.895 ± 0.010	N.A.	0.870 ± 0.000
242	Gly	0.867 ± 0.004	0.920 ± 0.020	0.827 ± 0.006	0.928 ± 0.013	0.853 ± 0.009	0.852 ± 0.011	N.A.	0.829 ± 0.000
243	Ser	0.846 ± 0.007	0.970 ± 0.020	0.867 ± 0.010	0.934 ± 0.013	0.829 ± 0.007	0.824 ± 0.018	N.A.	0.840 ± 0.000
244	Arg	0.911 ± 0.003		0.908 ± 0.005	0.888 ± 0.016	0.899 ± 0.003	0.911 ± 0.009	N.A.	0.881 ± 0.000
245	Gly	0.895 ± 0.004		0.901 ± 0.001	o.l.	0.878 ± 0.006	0.896 ± 0.003	o.l.	0.892 ± 0.000
246	Ile	0.918 ± 0.006	0.960 ± 0.020	0.931 ± 0.002	0.906 ± 0.013	0.913 ± 0.003	0.907 ± 0.010	o.l.	0.909 ± 0.000
247	Ile	0.921 ± 0.001	0.920 ± 0.020	0.912 ± 0.002	0.909 ± 0.015	0.921 ± 0.002	0.922 ± 0.001	o.l.	0.918 ± 0.000
248	Ala	0.922 ± 0.001	0.890 ± 0.010	0.909 ± 0.003	0.854 ± 0.012	0.921 ± 0.001	0.922 ± 0.002	0.908 ± 0.020	0.922 ± 0.000
249	Ala	0.908 ± 0.002	0.820 ± 0.020	0.905 ± 0.001	0.835 ± 0.010	0.910 ± 0.002	0.904 ± 0.001	0.911 ± 0.022	0.905 ± 0.000
250	Leu	0.907 ± 0.001	0.890 ± 0.020	0.918 ± 0.001	0.914 ± 0.011	0.910 ± 0.001	0.904 ± 0.002	0.888 ± 0.023	0.906 ± 0.000
251	Gly	0.916 ± 0.002	0.960 ± 0.020	0.920 ± 0.002	o.l.	0.916 ± 0.003	0.913 ± 0.002	0.887 ± 0.019	0.912 ± 0.000
252	Pro			0.876 ± 0.016	0.874 ± 0.012				
254	Asp	0.856 ± 0.006	0.870 ± 0.020	0.792 ± 0.031	0.864 ± 0.015	0.852 ± 0.009	0.855 ± 0.003	0.908 ± 0.018	0.853 ± 0.000
255	Gly	0.802 ± 0.006	0.890 ± 0.030	0.739 ± 0.019	0.747 ± 0.020	0.801 ± 0.010	0.802 ± 0.004		0.796 ± 0.000
256	Lys	0.865 ± 0.001	0.900 ± 0.020	0.737 ± 0.020	0.707 ± 0.015	0.860 ± 0.006	0.866 ± 0.001	0.906 ± 0.005	0.865 ± 0.000
257	Pro			0.835 ± 0.014	0.862 ± 0.012				
258	Ser	0.886 ± 0.010	0.910 ± 0.020		Pro	0.886 ± 0.003	0.884 ± 0.005		0.876 ± 0.000
259	Arg	0.901 ± 0.001	0.940 ± 0.020	0.907 ± 0.004	0.878 ± 0.012	0.903 ± 0.002	0.901 ± 0.003	o.l.	0.902 ± 0.000
260	Ile	0.907 ± 0.002	0.900 ± 0.020	0.901 ± 0.001	0.852 ± 0.012	0.910 ± 0.003	0.908 ± 0.002	0.908 ± 0.014	0.911 ± 0.000
261	Val	0.919 ± 0.004	0.900 ± 0.020	0.908 ± 0.001	0.850 ± 0.011	0.919 ± 0.003	0.918 ± 0.002	0.915 ± 0.019	0.915 ± 0.000
262	Val	0.924 ± 0.002	0.890 ± 0.020	0.909 ± 0.003	0.875 ± 0.012	0.923 ± 0.003	0.923 ± 0.001	0.891 ± 0.018	0.922 ± 0.000
263	Ile	0.925 ± 0.001	0.880 ± 0.010	0.919 ± 0.003	0.874 ± 0.013	0.924 ± 0.001	0.924 ± 0.001	o.l.	0.920 ± 0.000
264	Tyr	0.916 ± 0.003	0.900 ± 0.020	0.923 ± 0.001	0.848 ± 0.013	0.913 ± 0.003	0.915 ± 0.002		0.916 ± 0.000
265	Thr	0.904 ± 0.005	0.870 ± 0.020	0.887 ± 0.007	0.846 ± 0.016	0.909 ± 0.006	0.910 ± 0.002	o.l.	0.905 ± 0.000
266	Thr	0.915 ± 0.004	0.950 ± 0.040	0.900 ± 0.002	0.847 ± 0.011	0.910 ± 0.002	0.917 ± 0.003	N.A.	0.892 ± 0.000
267	Gly	0.872 ± 0.011	0.950 ± 0.020	0.904 ± 0.005	0.861 ± 0.012	0.885 ± 0.006	0.884 ± 0.003	N.A.	0.877 ± 0.000
268	Ser	0.778 ± 0.011	0.880 ± 0.020	0.818 ± 0.045	0.849 ± 0.012	0.748 ± 0.015	0.773 ± 0.013		0.775 ± 0.000
269	Gln	0.794 ± 0.016		0.848 ± 0.007	0.877 ± 0.015	0.793 ± 0.017	0.795 ± 0.007	o.l.	0.801 ± 0.000
270	Ala	0.717 ± 0.038		0.828 ± 0.003	o.l.	0.782 ± 0.018	0.774 ± 0.028	o.l.	0.768 ± 0.000
271	Thr	0.858 ± 0.010	0.890 ± 0.020	0.861 ± 0.011	0.866 ± 0.013	0.858 ± 0.005	0.857 ± 0.010	0.925 ± 0.015	0.850 ± 0.000
272	Met	0.877 ± 0.010		0.877 ± 0.003	0.859 ± 0.047	0.875 ± 0.004	0.874 ± 0.006	N.A.	0.873 ± 0.000
273	Asp	0.871 ± 0.014	0.930 ± 0.020	0.874 ± 0.001	0.895 ± 0.011	0.866 ± 0.006	0.866 ± 0.013	0.950 ± 0.010	0.864 ± 0.000
274	Glu	0.878 ± 0.016	0.930 ± 0.020	0.863 ± 0.007	0.880 ± 0.011	0.876 ± 0.006	0.873 ± 0.010	0.889 ± 0.012	0.869 ± 0.000
275	Arg	0.904 ± 0.010	0.930 ± 0.010	0.896 ± 0.002	0.906 ± 0.013	0.903 ± 0.004	0.900 ± 0.005	o.l.	0.902 ± 0.000
276	Asn	0.796 ± 0.021	0.920 ± 0.020	0.866 ± 0.005	0.867 ± 0.015	0.805 ± 0.010	0.807 ± 0.009	o.l.	0.836 ± 0.000
277	Arg	0.880 ± 0.023	0.930 ± 0.020	0.894 ± 0.006	o.l.	0.875 ± 0.010	0.862 ± 0.002	o.l.	0.872 ± 0.000
278	Gln	0.911 ± 0.010	0.920 ± 0.020	0.903 ± 0.002	0.891 ± 0.011	0.912 ± 0.002	0.907 ± 0.009	0.940 ± 0.024	0.906 ± 0.000
279	Ile	0.862 ± 0.010	0.950 ± 0.020	0.865 ± 0.007	0.856 ± 0.011	0.859 ± 0.007	0.858 ± 0.015	o.l.	0.875 ± 0.000
280	Ala	0.907 ± 0.006	0.930 ± 0.020	0.898 ± 0.007	0.898 ± 0.011	0.904 ± 0.003	0.885 ± 0.023	0.853 ± 0.080	0.903 ± 0.000
281	Glu	0.886 ± 0.003	0.910 ± 0.010	0.907 ± 0.007	0.898 ± 0.015	0.885 ± 0.007	0.868 ± 0.014	o.l.	0.885 ± 0.000
282	Ile	0.897 ± 0.004		0.865 ± 0.035	o.l.	0.895 ± 0.001	0.872 ± 0.034	0.906 ± 0.023	0.895 ± 0.000
283	Gly	0.912 ± 0.004	0.930 ± 0.020	0.903 ± 0.017	0.900 ± 0.011	0.914 ± 0.004	0.899 ± 0.016		0.909 ± 0.000
284	Ala	0.881 ± 0.009	0.920 ± 0.020	0.878 ± 0.049	0.903 ± 0.012	0.891 ± 0.010	0.867 ± 0.020	0.936 ± 0.020	0.883 ± 0.000
285	Ser	0.825 ± 0.010	0.890 ± 0.020	0.848 ± 0.050	0.875 ± 0.011	0.830 ± 0.010	0.814 ± 0.010	0.919 ± 0.013	0.822 ± 0.000
286	Leu	0.896 ± 0.015	0.920 ± 0.020	0.871 ± 0.033	0.882 ± 0.020	0.901 ± 0.004	0.881 ± 0.016	0.894 ± 0.011	0.889 ± 0.000
287	Ile	0.873 ± 0.018	0.960 ± 0.020	0.882 ± 0.037	o.l.	0.883 ± 0.011	0.864 ± 0.013	0.920 ± 0.020	0.861 ± 0.000
288	Lys	0.835 ± 0.020		0.850 ± 0.048	0.921 ± 0.011	0.846 ± 0.006	0.828 ± 0.010	o.l.	0.818 ± 0.000
289	His	0.834 ± 0.011	0.850 ± 0.010	0.820 ± 0.004	0.886 ± 0.011	0.840 ± 0.014	0.820 ± 0.005	0.900 ± 0.013	0.826 ± 0.000
290	Trp	0.763 ± 0.010	0.830 ± 0.020	0.807 ± 0.011	0.860 ± 0.012	0.764 ± 0.031	0.750 ± 0.017	0.853 ± 0.009	0.750 ± 0.000
291				0.780 ± 0.020	0.729 ± 0.015				
292				0.669 ± 0.061	0.524 ± 0.013				
Average	Average	0.871 ± 0.050	0.901 ± 0.046	0.864 ± 0.050	0.869 ± 0.051	0.869 ± 0.051	0.868 ± 0.048	0.895 ± 0.046	0.865 ± 0.049

N.A. stands for non assigned by NMR
o.l. stands for overlaid resonances

Table A 5-6 Per residue Cα-RMSF extracted from triplicate MD simulations
For 2 μs each for TEM-1, PSE-4, cTEM-2m, cTEM-17m and cTEM-19m.

TEM-1 Residue		TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
24			0.289 ± 0.070			
25			0.154 ± 0.009			
26	His	0.094 ± 0.027	0.140 ± 0.021	0.103 ± 0.028	0.120 ± 0.029	0.115 ± 0.041
27	Pro	0.099 ± 0.025	0.167 ± 0.025	0.105 ± 0.021	0.121 ± 0.025	0.117 ± 0.035
28	Glu	0.101 ± 0.026	0.152 ± 0.017	0.103 ± 0.015	0.124 ± 0.026	0.117 ± 0.031
29	Thr	0.079 ± 0.022	0.094 ± 0.009	0.083 ± 0.017	0.096 ± 0.014	0.095 ± 0.030
30	Leu	0.072 ± 0.015	0.083 ± 0.011	0.084 ± 0.022	0.082 ± 0.007	0.092 ± 0.030
31	Val	0.082 ± 0.017	0.104 ± 0.011	0.090 ± 0.018	0.095 ± 0.014	0.102 ± 0.027
32	Lys	0.074 ± 0.014	0.094 ± 0.006	0.081 ± 0.014	0.085 ± 0.010	0.094 ± 0.024
33	Val	0.060 ± 0.010	0.072 ± 0.004	0.073 ± 0.024	0.067 ± 0.004	0.080 ± 0.024
34	Lys	0.074 ± 0.011	0.083 ± 0.005	0.089 ± 0.027	0.086 ± 0.009	0.100 ± 0.026
35	Asp	0.085 ± 0.011	0.095 ± 0.004	0.097 ± 0.018	0.103 ± 0.016	0.115 ± 0.017
36	Ala	0.079 ± 0.007	0.092 ± 0.001	0.085 ± 0.008	0.092 ± 0.013	0.099 ± 0.009
37	Glu	0.075 ± 0.004	0.093 ± 0.004	0.077 ± 0.006	0.084 ± 0.006	0.091 ± 0.011
38	Asp	0.083 ± 0.008	0.098 ± 0.003	0.091 ± 0.004	0.098 ± 0.013	0.112 ± 0.013
39	Gln	0.077 ± 0.012	0.097 ± 0.002	0.093 ± 0.011	0.093 ± 0.016	0.108 ± 0.013
40	Leu	0.070 ± 0.009	0.085 ± 0.002	0.083 ± 0.006	0.079 ± 0.011	0.094 ± 0.006
41	Gly	0.072 ± 0.010	0.084 ± 0.006	0.085 ± 0.006	0.082 ± 0.011	0.104 ± 0.007
42	Ala	0.063 ± 0.006	0.064 ± 0.005	0.068 ± 0.004	0.066 ± 0.009	0.088 ± 0.013
43	Arg	0.059 ± 0.006	0.059 ± 0.009	0.057 ± 0.002	0.058 ± 0.009	0.077 ± 0.006
44	Val	0.050 ± 0.001	0.048 ± 0.005	0.049 ± 0.002	0.051 ± 0.009	0.063 ± 0.011
45	Gly	0.051 ± 0.004	0.050 ± 0.005	0.054 ± 0.005	0.054 ± 0.009	0.064 ± 0.015
46	Tyr	0.044 ± 0.003	0.043 ± 0.004	0.045 ± 0.004	0.048 ± 0.007	0.050 ± 0.008
47	Ile	0.042 ± 0.005	0.047 ± 0.003	0.044 ± 0.003	0.046 ± 0.007	0.048 ± 0.006
48	Glu	0.044 ± 0.005	0.045 ± 0.003	0.048 ± 0.002	0.047 ± 0.004	0.048 ± 0.004
49	Leu	0.052 ± 0.002	0.053 ± 0.003	0.056 ± 0.004	0.056 ± 0.002	0.054 ± 0.001
50	Asp	0.058 ± 0.005	0.065 ± 0.006	0.065 ± 0.003	0.062 ± 0.001	0.061 ± 0.003
51	Leu	0.068 ± 0.009	0.084 ± 0.011	0.071 ± 0.003	0.072 ± 0.002	0.071 ± 0.004
52	Asn	0.102 ± 0.009	0.120 ± 0.015	0.108 ± 0.003	0.102 ± 0.002	0.101 ± 0.001
53	Ser	0.110 ± 0.007	0.137 ± 0.019	0.116 ± 0.005	0.109 ± 0.001	0.107 ± 0.007
54	Gly	0.089 ± 0.007	0.155 ± 0.025	0.092 ± 0.006	0.089 ± 0.001	0.089 ± 0.004
55	Lys	0.083 ± 0.007	0.117 ± 0.027	0.087 ± 0.008	0.085 ± 0.004	0.081 ± 0.006
56	Ile	0.067 ± 0.005	0.081 ± 0.018	0.072 ± 0.006	0.068 ± 0.003	0.068 ± 0.005
57	Leu	0.060 ± 0.005	0.059 ± 0.002	0.068 ± 0.006	0.063 ± 0.003	0.062 ± 0.004
58	Glu	0.052 ± 0.004	existant in Ambler	0.062 ± 0.006	0.056 ± 0.004	0.059 ± 0.006
59	Ser	0.056 ± 0.004	0.058 ± 0.006	0.061 ± 0.002	0.060 ± 0.004	0.061 ± 0.007
60	Phe	0.053 ± 0.004	0.061 ± 0.008	0.057 ± 0.004	0.058 ± 0.008	0.059 ± 0.007
61	Arg	0.055 ± 0.004	0.062 ± 0.004	0.060 ± 0.003	0.060 ± 0.008	0.067 ± 0.006
62	Pro	0.064 ± 0.008	0.082 ± 0.013	0.087 ± 0.032	0.058 ± 0.008	0.071 ± 0.009
63	Glu	0.071 ± 0.012	0.100 ± 0.011	0.106 ± 0.044	0.067 ± 0.011	0.084 ± 0.012
64	Glu	0.068 ± 0.002	0.070 ± 0.005	0.090 ± 0.016	0.072 ± 0.011	0.083 ± 0.006
65	Arg	0.057 ± 0.007	0.059 ± 0.004	0.070 ± 0.019	0.067 ± 0.021	0.076 ± 0.007
66	Phe	0.051 ± 0.008	0.054 ± 0.007	0.065 ± 0.026	0.056 ± 0.015	0.061 ± 0.003
67	Pro	0.056 ± 0.010	0.067 ± 0.017	0.069 ± 0.025	0.059 ± 0.015	0.068 ± 0.003
68	Met	0.065 ± 0.011	0.059 ± 0.010	0.068 ± 0.024	0.061 ± 0.014	0.059 ± 0.005
69	Met	0.060 ± 0.009	0.060 ± 0.016	0.068 ± 0.023	0.060 ± 0.019	0.065 ± 0.007
70	Ser	0.065 ± 0.012	0.071 ± 0.017	0.075 ± 0.024	0.063 ± 0.017	0.060 ± 0.010

TEM-1 Residue		TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
71	Thr	0.055 ± 0.014	0.063 ± 0.008	0.065 ± 0.015	0.071 ± 0.021	0.057 ± 0.009
72	Phe	0.048 ± 0.008	0.058 ± 0.015	0.057 ± 0.007	0.069 ± 0.034	0.060 ± 0.026
73	Lys	0.043 ± 0.007	0.052 ± 0.011	0.050 ± 0.005	0.058 ± 0.024	0.055 ± 0.023
74	Val	0.045 ± 0.008	0.053 ± 0.009	0.057 ± 0.010	0.054 ± 0.012	0.056 ± 0.020
75	Leu	0.045 ± 0.006	0.053 ± 0.012	0.057 ± 0.011	0.051 ± 0.010	0.057 ± 0.018
76	Leu	0.040 ± 0.005	0.048 ± 0.006	0.047 ± 0.005	0.047 ± 0.012	0.049 ± 0.011
77	Cys	0.039 ± 0.004	0.042 ± 0.003	0.048 ± 0.008	0.046 ± 0.012	0.047 ± 0.008
78	Gly	0.049 ± 0.007	0.047 ± 0.008	0.061 ± 0.011	0.055 ± 0.009	0.058 ± 0.010
79	Ala	0.053 ± 0.008	0.049 ± 0.008	0.062 ± 0.009	0.056 ± 0.007	0.060 ± 0.010
80	Val	0.050 ± 0.006	0.045 ± 0.004	0.056 ± 0.006	0.054 ± 0.010	0.058 ± 0.008
81	Leu	0.049 ± 0.007	0.048 ± 0.004	0.060 ± 0.009	0.054 ± 0.010	0.061 ± 0.007
82	Ser	0.060 ± 0.010	0.055 ± 0.007	0.073 ± 0.012	0.066 ± 0.008	0.077 ± 0.007
83	Arg	0.068 ± 0.009	0.058 ± 0.006	0.078 ± 0.010	0.072 ± 0.009	0.079 ± 0.006
84	Val	0.067 ± 0.007	0.062 ± 0.003	0.075 ± 0.009	0.069 ± 0.008	0.074 ± 0.010
85	Asp	0.080 ± 0.011	0.084 ± 0.004	0.093 ± 0.014	0.082 ± 0.005	0.089 ± 0.014
86	Ala	0.094 ± 0.011	0.095 ± 0.005	0.108 ± 0.017	0.098 ± 0.006	0.103 ± 0.014
87	Gly	0.100 ± 0.009	0.097 ± 0.003	0.111 ± 0.013	0.101 ± 0.008	0.106 ± 0.011
88	Gln	0.096 ± 0.009	0.089 ± 0.004	0.108 ± 0.011	0.097 ± 0.008	0.100 ± 0.005
89	Glu	0.078 ± 0.006	0.079 ± 0.002	0.085 ± 0.009	0.079 ± 0.009	0.082 ± 0.004
90	Gln	0.080 ± 0.005	0.084 ± 0.006	0.087 ± 0.013	0.079 ± 0.009	0.083 ± 0.005
91	Leu	0.079 ± 0.010	0.086 ± 0.025	0.090 ± 0.022	0.076 ± 0.008	0.079 ± 0.002
92	Gly	0.083 ± 0.019	0.086 ± 0.008	0.100 ± 0.038	0.081 ± 0.011	0.085 ± 0.006
93	Arg	0.075 ± 0.013	0.083 ± 0.001	0.086 ± 0.019	0.078 ± 0.012	0.083 ± 0.005
94	Arg	0.066 ± 0.008	0.081 ± 0.002	0.074 ± 0.008	0.073 ± 0.011	0.080 ± 0.008
95	Ile	0.063 ± 0.007	0.086 ± 0.001	0.076 ± 0.009	0.070 ± 0.008	0.079 ± 0.007
96	His	0.071 ± 0.010	0.101 ± 0.002	0.093 ± 0.013	0.077 ± 0.008	0.091 ± 0.009
97	Tyr	0.068 ± 0.011	0.106 ± 0.003	0.094 ± 0.016	0.073 ± 0.007	0.091 ± 0.011
98	Ser	0.089 ± 0.015	0.124 ± 0.008	0.121 ± 0.022	0.090 ± 0.007	0.113 ± 0.013
99	Gln	0.100 ± 0.018	0.141 ± 0.018	0.135 ± 0.027	0.098 ± 0.006	0.124 ± 0.013
100	Asn	0.107 ± 0.018	0.166 ± 0.022	0.136 ± 0.026	0.106 ± 0.008	0.137 ± 0.012
101	Asp	0.082 ± 0.016	0.143 ± 0.015	0.107 ± 0.022	0.081 ± 0.008	0.108 ± 0.010
102	Leu	0.076 ± 0.024	0.097 ± 0.022	0.108 ± 0.024	0.072 ± 0.006	0.101 ± 0.021
103	Val	0.082 ± 0.033	0.104 ± 0.032	0.121 ± 0.028	0.077 ± 0.012	0.117 ± 0.040
104	Glu	0.093 ± 0.033	0.114 ± 0.034	0.147 ± 0.037	0.092 ± 0.014	0.131 ± 0.056
105	Tyr	0.085 ± 0.017	0.103 ± 0.023	0.173 ± 0.100	0.092 ± 0.013	0.102 ± 0.026
106	Ser	0.069 ± 0.016	0.077 ± 0.014	0.123 ± 0.050	0.073 ± 0.011	0.080 ± 0.021
107	Pro	0.083 ± 0.016	0.099 ± 0.012	0.119 ± 0.022	0.091 ± 0.010	0.091 ± 0.016
108	Val	0.067 ± 0.012	0.083 ± 0.009	0.098 ± 0.020	0.077 ± 0.010	0.078 ± 0.012
109	Thr	0.057 ± 0.012	0.071 ± 0.006	0.086 ± 0.020	0.062 ± 0.007	0.070 ± 0.016
110	Glu	0.078 ± 0.017	0.097 ± 0.014	0.117 ± 0.028	0.081 ± 0.007	0.091 ± 0.018
111	Lys	0.088 ± 0.016	0.119 ± 0.012	0.125 ± 0.024	0.093 ± 0.009	0.102 ± 0.017
112	His	0.082 ± 0.013	0.109 ± 0.005	0.111 ± 0.019	0.087 ± 0.008	0.097 ± 0.014
113	Leu	0.088 ± 0.013	0.119 ± 0.010	0.120 ± 0.022	0.091 ± 0.007	0.105 ± 0.013
114	Thr	0.109 ± 0.014	0.138 ± 0.011	0.141 ± 0.024	0.112 ± 0.008	0.129 ± 0.012
115	Asp	0.097 ± 0.014	0.119 ± 0.003	0.125 ± 0.022	0.102 ± 0.010	0.115 ± 0.013
116	Gly	0.071 ± 0.011	0.112 ± 0.005	0.094 ± 0.016	0.078 ± 0.008	0.088 ± 0.012
117	Met	0.058 ± 0.008	0.079 ± 0.006	0.074 ± 0.013	0.064 ± 0.008	0.070 ± 0.011

TEM-1 Residue		TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
118	Thr	0.055 ± 0.007	0.060 ± 0.005	0.067 ± 0.014	0.061 ± 0.009	0.064 ± 0.009
119	Val	0.050 ± 0.005	0.051 ± 0.004	0.059 ± 0.009	0.054 ± 0.007	0.055 ± 0.006
120	Arg	0.054 ± 0.004	0.054 ± 0.002	0.062 ± 0.007	0.059 ± 0.008	0.057 ± 0.007
121	Glu	0.051 ± 0.005	0.054 ± 0.004	0.061 ± 0.009	0.058 ± 0.009	0.055 ± 0.009
122	Leu	0.041 ± 0.005	0.044 ± 0.004	0.049 ± 0.008	0.046 ± 0.007	0.044 ± 0.008
123	Cys	0.042 ± 0.005	0.042 ± 0.005	0.052 ± 0.009	0.048 ± 0.008	0.043 ± 0.007
124	Ser	0.052 ± 0.006	0.050 ± 0.006	0.068 ± 0.015	0.060 ± 0.009	0.053 ± 0.008
125	Ala	0.049 ± 0.007	0.052 ± 0.008	0.067 ± 0.016	0.057 ± 0.011	0.053 ± 0.009
126	Ala	0.048 ± 0.009	0.056 ± 0.014	0.064 ± 0.013	0.054 ± 0.009	0.051 ± 0.011
127	Ile	0.051 ± 0.009	0.056 ± 0.014	0.071 ± 0.016	0.059 ± 0.010	0.055 ± 0.009
128	Thr	0.060 ± 0.008	0.062 ± 0.015	0.093 ± 0.027	0.073 ± 0.007	0.066 ± 0.012
129	Met	0.066 ± 0.011	0.072 ± 0.005	0.100 ± 0.032	0.086 ± 0.016	0.076 ± 0.014
130	Ser	0.071 ± 0.016	0.066 ± 0.002	0.102 ± 0.025	0.079 ± 0.011	0.074 ± 0.020
131	Asp	0.056 ± 0.017	0.059 ± 0.007	0.088 ± 0.022	0.060 ± 0.011	0.061 ± 0.020
132	Asn	0.053 ± 0.018	0.063 ± 0.015	0.081 ± 0.020	0.055 ± 0.012	0.069 ± 0.025
133	Thr	0.047 ± 0.013	0.055 ± 0.011	0.070 ± 0.020	0.048 ± 0.007	0.065 ± 0.023
134	Ala	0.041 ± 0.008	0.045 ± 0.004	0.055 ± 0.013	0.044 ± 0.007	0.051 ± 0.015
135	Ala	0.044 ± 0.011	0.052 ± 0.008	0.056 ± 0.011	0.046 ± 0.010	0.056 ± 0.013
136	Asn	0.043 ± 0.011	0.059 ± 0.012	0.058 ± 0.015	0.045 ± 0.012	0.066 ± 0.014
137	Leu	0.040 ± 0.006	0.060 ± 0.006	0.049 ± 0.009	0.045 ± 0.010	0.061 ± 0.013
138	Leu	0.041 ± 0.006	0.059 ± 0.008	0.045 ± 0.004	0.045 ± 0.009	0.052 ± 0.006
139	Leu	0.044 ± 0.007	0.074 ± 0.014	0.050 ± 0.007	0.050 ± 0.010	0.062 ± 0.004
140	Thr	0.052 ± 0.007	0.083 ± 0.013	0.065 ± 0.008	0.069 ± 0.009	0.082 ± 0.001
141	Thr	0.064 ± 0.005	0.070 ± 0.008	0.073 ± 0.009	0.076 ± 0.010	0.090 ± 0.007
142	Ile	0.067 ± 0.006	0.066 ± 0.008	0.073 ± 0.008	0.075 ± 0.013	0.085 ± 0.005
143	Gly	0.063 ± 0.004	0.068 ± 0.006	0.074 ± 0.013	0.077 ± 0.017	0.101 ± 0.016
144	Gly	0.052 ± 0.003	0.057 ± 0.003	0.062 ± 0.013	0.065 ± 0.017	0.087 ± 0.015
145	Pro	0.052 ± 0.004	0.055 ± 0.005	0.059 ± 0.011	0.064 ± 0.022	0.093 ± 0.018
146	Lys	0.065 ± 0.003	0.063 ± 0.002	0.070 ± 0.009	0.080 ± 0.026	0.114 ± 0.012
147	Glu	0.072 ± 0.004	0.065 ± 0.009	0.076 ± 0.010	0.078 ± 0.018	0.098 ± 0.001
148	Leu	0.061 ± 0.002	0.059 ± 0.010	0.065 ± 0.007	0.061 ± 0.010	0.072 ± 0.008
149	Thr	0.053 ± 0.002	0.064 ± 0.012	0.056 ± 0.004	0.062 ± 0.016	0.086 ± 0.015
150	Ala	0.070 ± 0.006	0.081 ± 0.023	0.072 ± 0.008	0.081 ± 0.017	0.099 ± 0.011
151	Phe	0.077 ± 0.008	0.080 ± 0.032	0.077 ± 0.008	0.076 ± 0.010	0.086 ± 0.011
152	Leu	0.076 ± 0.005	0.086 ± 0.039	0.071 ± 0.006	0.071 ± 0.014	0.090 ± 0.009
153	His	0.090 ± 0.012	0.109 ± 0.051	0.083 ± 0.008	0.100 ± 0.022	0.107 ± 0.016
154	Asn	0.115 ± 0.010	0.126 ± 0.065	0.107 ± 0.008	0.124 ± 0.027	0.131 ± 0.023
155	Met	0.131 ± 0.008	0.137 ± 0.081	0.122 ± 0.015	0.117 ± 0.023	0.136 ± 0.022
156	Gly	0.188 ± 0.013	0.187 ± 0.104	0.153 ± 0.041	0.153 ± 0.049	0.204 ± 0.021
157	Asp	0.145 ± 0.033	0.148 ± 0.068	0.128 ± 0.045	0.128 ± 0.039	0.150 ± 0.013
158	His	0.104 ± 0.015	0.107 ± 0.028	0.116 ± 0.026	0.122 ± 0.038	0.146 ± 0.016
159	Val	0.068 ± 0.011	0.107 ± 0.053	0.080 ± 0.005	0.120 ± 0.060	0.158 ± 0.051
160	Thr	0.057 ± 0.012	0.080 ± 0.029	0.070 ± 0.011	0.098 ± 0.046	0.122 ± 0.043
161	Arg	0.052 ± 0.013	0.072 ± 0.013	0.065 ± 0.015	0.084 ± 0.035	0.106 ± 0.018
162	Leu	0.045 ± 0.008	0.060 ± 0.011	0.066 ± 0.024	0.064 ± 0.027	0.086 ± 0.005
163	Asp	0.046 ± 0.003	0.064 ± 0.015	0.070 ± 0.025	0.065 ± 0.024	0.093 ± 0.014
164	Arg	0.045 ± 0.005	0.069 ± 0.023	0.069 ± 0.022	0.061 ± 0.024	0.104 ± 0.023

TEM-1	Residue	TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
165	Trp	0.048 ± 0.004	0.077 ± 0.028	0.073 ± 0.026	0.067 ± 0.032	0.106 ± 0.036
166	Glu	0.055 ± 0.005	0.084 ± 0.031	0.085 ± 0.031	0.071 ± 0.031	0.123 ± 0.071
167	Pro	0.068 ± 0.004	0.094 ± 0.033	0.093 ± 0.030	0.088 ± 0.039	0.146 ± 0.108
168	Glu	0.061 ± 0.004	0.090 ± 0.032	0.085 ± 0.028	0.095 ± 0.058	0.149 ± 0.117
169	Leu	0.048 ± 0.005	0.065 ± 0.017	0.069 ± 0.025	0.085 ± 0.059	0.128 ± 0.095
170	Asn	0.051 ± 0.005	0.060 ± 0.016	0.072 ± 0.030	0.127 ± 0.123	0.206 ± 0.224
171	Glu	0.065 ± 0.003	0.068 ± 0.009	0.084 ± 0.027	0.153 ± 0.149	0.232 ± 0.215
172	Ala	0.090 ± 0.005	0.072 ± 0.002	0.114 ± 0.025	0.213 ± 0.173	0.292 ± 0.188
173	Ile	0.111 ± 0.009	0.109 ± 0.005	0.142 ± 0.024	0.229 ± 0.170	0.252 ± 0.123
174	Pro	0.140 ± 0.012	0.157 ± 0.019	0.193 ± 0.055	0.325 ± 0.256	0.274 ± 0.100
175	Asn	0.160 ± 0.012	0.186 ± 0.022	0.195 ± 0.046	0.285 ± 0.232	0.273 ± 0.085
176	Asp	0.100 ± 0.006	0.148 ± 0.027	0.127 ± 0.027	0.210 ± 0.117	0.214 ± 0.068
177	Glu	0.084 ± 0.006	0.126 ± 0.041	0.104 ± 0.019	0.130 ± 0.064	0.164 ± 0.050
178	Arg	0.060 ± 0.002	0.091 ± 0.028	0.079 ± 0.023	0.102 ± 0.055	0.119 ± 0.014
179	Asp	0.042 ± 0.007	0.070 ± 0.019	0.064 ± 0.023	0.067 ± 0.035	0.088 ± 0.017
180	Thr	0.044 ± 0.010	0.067 ± 0.015	0.057 ± 0.014	0.060 ± 0.023	0.070 ± 0.005
181	Thr	0.043 ± 0.007	0.057 ± 0.011	0.057 ± 0.014	0.058 ± 0.019	0.067 ± 0.003
182	Met	0.045 ± 0.004	0.050 ± 0.005	0.064 ± 0.017	0.056 ± 0.014	0.067 ± 0.002
183	Pro	0.047 ± 0.004	0.069 ± 0.008	0.061 ± 0.006	0.052 ± 0.009	0.064 ± 0.009
184	Ala	0.053 ± 0.004	0.087 ± 0.020	0.066 ± 0.008	0.056 ± 0.010	0.074 ± 0.017
185	Ala	0.054 ± 0.006	0.066 ± 0.023	0.070 ± 0.016	0.062 ± 0.016	0.073 ± 0.011
186	Met	0.048 ± 0.007	0.076 ± 0.030	0.063 ± 0.012	0.057 ± 0.013	0.068 ± 0.011
187	Ala	0.047 ± 0.007	0.062 ± 0.018	0.054 ± 0.003	0.051 ± 0.006	0.061 ± 0.006
188	Thr	0.052 ± 0.006	0.056 ± 0.007	0.058 ± 0.006	0.056 ± 0.007	0.066 ± 0.009
189	Thr	0.054 ± 0.005	0.064 ± 0.012	0.062 ± 0.007	0.058 ± 0.006	0.074 ± 0.014
190	Leu	0.053 ± 0.008	0.065 ± 0.021	0.058 ± 0.003	0.055 ± 0.005	0.069 ± 0.010
191	Arg	0.053 ± 0.007	0.060 ± 0.012	0.055 ± 0.002	0.054 ± 0.004	0.061 ± 0.004
192	Lys	0.061 ± 0.006	0.069 ± 0.014	0.063 ± 0.004	0.064 ± 0.005	0.073 ± 0.010
193	Leu	0.063 ± 0.007	0.078 ± 0.021	0.067 ± 0.001	0.067 ± 0.009	0.083 ± 0.013
194	Leu	0.067 ± 0.011	0.070 ± 0.011	0.070 ± 0.003	0.070 ± 0.008	0.084 ± 0.009
195	Thr	0.085 ± 0.017	0.079 ± 0.013	0.085 ± 0.008	0.087 ± 0.008	0.102 ± 0.007
196	Gly	0.114 ± 0.032	0.101 ± 0.044	0.138 ± 0.025	0.142 ± 0.003	0.149 ± 0.009
197	Glu	0.111 ± 0.020	0.115 ± 0.040	0.123 ± 0.010	0.119 ± 0.009	0.120 ± 0.006
198	Leu	0.106 ± 0.021	0.094 ± 0.034	0.123 ± 0.007	0.113 ± 0.007	0.125 ± 0.009
199	Leu	0.075 ± 0.013	0.071 ± 0.021	0.091 ± 0.006	0.083 ± 0.005	0.088 ± 0.009
200	Thr	0.074 ± 0.008	0.078 ± 0.021	0.093 ± 0.015	0.081 ± 0.001	0.083 ± 0.008
201	Leu	0.077 ± 0.006	0.082 ± 0.021	0.095 ± 0.015	0.083 ± 0.005	0.087 ± 0.007
202	Ala	0.078 ± 0.006	0.079 ± 0.015	0.097 ± 0.015	0.087 ± 0.005	0.088 ± 0.003
203	Ser	0.062 ± 0.005	0.063 ± 0.011	0.076 ± 0.012	0.071 ± 0.002	0.071 ± 0.006
204	Arg	0.063 ± 0.003	0.059 ± 0.013	0.074 ± 0.010	0.067 ± 0.003	0.071 ± 0.008
205	Gln	0.067 ± 0.005	0.062 ± 0.010	0.083 ± 0.012	0.074 ± 0.005	0.078 ± 0.005
206	Gln	0.061 ± 0.006	0.057 ± 0.005	0.076 ± 0.011	0.070 ± 0.003	0.067 ± 0.004
207	Leu	0.053 ± 0.005	0.050 ± 0.005	0.066 ± 0.011	0.063 ± 0.005	0.059 ± 0.004
208	Ile	0.059 ± 0.005	0.052 ± 0.006	0.072 ± 0.012	0.069 ± 0.007	0.071 ± 0.003
209	Asp	0.068 ± 0.006	0.060 ± 0.005	0.081 ± 0.012	0.074 ± 0.002	0.069 ± 0.006
210	Trp	0.060 ± 0.009	0.054 ± 0.004	0.076 ± 0.013	0.067 ± 0.002	0.060 ± 0.006
211	Met	0.057 ± 0.009	0.049 ± 0.003	0.074 ± 0.013	0.065 ± 0.003	0.059 ± 0.002

TEM-1 Residue		TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
212	Glu	0.062 ± 0.013	0.047 ± 0.002	0.091 ± 0.024	0.071 ± 0.002	0.063 ± 0.001
213	Ala	0.097 ± 0.024	0.058 ± 0.011	0.125 ± 0.041	0.092 ± 0.014	0.083 ± 0.010
214	Asp	0.112 ± 0.041	0.059 ± 0.018	0.185 ± 0.085	0.180 ± 0.071	0.113 ± 0.009
215	Lys	0.127 ± 0.023	0.073 ± 0.014	0.220 ± 0.087	0.189 ± 0.092	0.124 ± 0.027
216	Val	0.150 ± 0.035	0.076 ± 0.017	0.287 ± 0.122	0.235 ± 0.055	0.177 ± 0.020
217	Ala	0.167 ± 0.077	0.074 ± 0.028	0.301 ± 0.113	0.280 ± 0.078	0.185 ± 0.052
218	Gly	0.178 ± 0.065	0.077 ± 0.027	0.243 ± 0.100	0.300 ± 0.119	0.236 ± 0.071
219	Pro	0.153 ± 0.048	0.082 ± 0.020	0.210 ± 0.053	0.247 ± 0.089	0.207 ± 0.084
220	Leu	0.125 ± 0.046	0.066 ± 0.017	0.178 ± 0.024	0.204 ± 0.060	0.181 ± 0.075
221	Leu	0.112 ± 0.048	0.057 ± 0.013	0.148 ± 0.034	0.156 ± 0.035	0.143 ± 0.028
222	Arg	0.128 ± 0.049	0.053 ± 0.004	0.173 ± 0.043	0.171 ± 0.055	0.199 ± 0.051
223	Ser	0.157 ± 0.060	0.067 ± 0.010	0.205 ± 0.045	0.210 ± 0.080	0.207 ± 0.020
224	Ala	0.149 ± 0.041	0.068 ± 0.008	0.172 ± 0.026	0.180 ± 0.034	0.185 ± 0.047
225	Leu	0.108 ± 0.023	0.073 ± 0.005	0.139 ± 0.009	0.130 ± 0.017	0.134 ± 0.031
226	Pro	0.127 ± 0.020	0.101 ± 0.020	0.150 ± 0.016	0.154 ± 0.006	0.144 ± 0.031
227	Ala	0.149 ± 0.019	0.128 ± 0.027	0.174 ± 0.005	0.187 ± 0.016	0.161 ± 0.029
228	Gly	0.121 ± 0.019	0.103 ± 0.011	0.131 ± 0.012	0.152 ± 0.013	0.132 ± 0.022
229	Trp	0.084 ± 0.012	0.066 ± 0.001	0.094 ± 0.011	0.104 ± 0.004	0.089 ± 0.017
230	Phe	0.069 ± 0.010	0.060 ± 0.002	0.074 ± 0.007	0.079 ± 0.004	0.072 ± 0.012
231	Ile	0.058 ± 0.011	0.047 ± 0.000	0.069 ± 0.003	0.067 ± 0.007	0.060 ± 0.007
232	Ala	0.054 ± 0.013	0.044 ± 0.001	0.072 ± 0.006	0.062 ± 0.008	0.057 ± 0.001
233	Asp	0.052 ± 0.011	0.040 ± 0.001	0.073 ± 0.007	0.066 ± 0.015	0.054 ± 0.002
234	Lys	0.045 ± 0.007	0.043 ± 0.003	0.066 ± 0.005	0.062 ± 0.018	0.051 ± 0.006
235	Ser	0.053 ± 0.009	0.049 ± 0.004	0.060 ± 0.006	0.064 ± 0.026	0.057 ± 0.007
236	Gly	0.058 ± 0.005	0.054 ± 0.006	0.061 ± 0.009	0.066 ± 0.023	0.067 ± 0.005
237	Ala	0.053 ± 0.007	0.051 ± 0.007	0.056 ± 0.004	0.059 ± 0.019	0.074 ± 0.007
238	Gly	0.068 ± 0.005	0.058 ± 0.007	0.066 ± 0.006	0.072 ± 0.025	0.103 ± 0.018
240	Glu	0.065 ± 0.008	0.061 ± 0.005	0.075 ± 0.010	0.080 ± 0.034	0.120 ± 0.007
241	Arg	0.080 ± 0.011	0.068 ± 0.011	0.087 ± 0.004	0.091 ± 0.026	0.131 ± 0.017
242	Gly	0.072 ± 0.010	0.073 ± 0.009	0.078 ± 0.004	0.083 ± 0.020	0.122 ± 0.021
243	Ser	0.054 ± 0.007	0.050 ± 0.003	0.062 ± 0.003	0.057 ± 0.012	0.074 ± 0.012
244	Arg	0.055 ± 0.009	0.045 ± 0.002	0.061 ± 0.007	0.060 ± 0.008	0.065 ± 0.008
245	Gly	0.064 ± 0.016	0.047 ± 0.003	0.069 ± 0.008	0.072 ± 0.013	0.069 ± 0.014
246	Ile	0.046 ± 0.010	0.047 ± 0.005	0.051 ± 0.002	0.056 ± 0.018	0.051 ± 0.005
247	Ile	0.041 ± 0.007	0.044 ± 0.002	0.048 ± 0.002	0.052 ± 0.015	0.048 ± 0.004
248	Ala	0.045 ± 0.008	0.046 ± 0.002	0.052 ± 0.002	0.054 ± 0.010	0.052 ± 0.004
249	Ala	0.045 ± 0.007	0.045 ± 0.001	0.052 ± 0.003	0.048 ± 0.004	0.052 ± 0.003
250	Leu	0.049 ± 0.008	0.048 ± 0.002	0.055 ± 0.004	0.059 ± 0.004	0.056 ± 0.007
251	Gly	0.059 ± 0.010	0.058 ± 0.002	0.067 ± 0.010	0.074 ± 0.006	0.071 ± 0.015
252	Pro	0.082 ± 0.016	0.101 ± 0.023	0.095 ± 0.015	0.105 ± 0.005	0.099 ± 0.025
254	Asp	0.110 ± 0.016	0.147 ± 0.029	0.122 ± 0.013	0.137 ± 0.006	0.127 ± 0.026
255	Gly	0.095 ± 0.013	0.131 ± 0.008	0.101 ± 0.009	0.109 ± 0.008	0.109 ± 0.023
256	Lys	0.088 ± 0.009	0.099 ± 0.003	0.093 ± 0.004	0.094 ± 0.003	0.097 ± 0.019
257	Pro	0.065 ± 0.007	0.082 ± 0.005	0.070 ± 0.005	0.071 ± 0.002	0.073 ± 0.012
258	Ser	0.058 ± 0.008	0.058 ± 0.003	0.064 ± 0.008	0.067 ± 0.004	0.066 ± 0.007
259	Arg	0.046 ± 0.007	0.048 ± 0.002	0.052 ± 0.006	0.054 ± 0.004	0.053 ± 0.004
260	Ile	0.042 ± 0.006	0.043 ± 0.001	0.046 ± 0.003	0.047 ± 0.003	0.051 ± 0.002

TEM-1 Residue		TEM-1	PSE-4	cTEM-2m	cTEM-17m	cTEM-19m
261	Val	0.042 ± 0.009	0.045 ± 0.003	0.042 ± 0.003	0.047 ± 0.005	0.047 ± 0.004
262	Val	0.040 ± 0.008	0.045 ± 0.004	0.042 ± 0.004	0.048 ± 0.009	0.048 ± 0.006
263	Ile	0.044 ± 0.008	0.043 ± 0.003	0.045 ± 0.000	0.054 ± 0.012	0.047 ± 0.006
264	Tyr	0.046 ± 0.008	0.044 ± 0.003	0.048 ± 0.004	0.051 ± 0.011	0.049 ± 0.005
265	Thr	0.057 ± 0.008	0.050 ± 0.002	0.054 ± 0.005	0.053 ± 0.009	0.055 ± 0.003
266	Thr	0.072 ± 0.024	0.057 ± 0.006	0.058 ± 0.005	0.057 ± 0.010	0.070 ± 0.005
267	Gly	0.089 ± 0.024	0.067 ± 0.010	0.076 ± 0.005	0.077 ± 0.007	0.098 ± 0.014
268	Ser	0.070 ± 0.008	0.081 ± 0.006	0.084 ± 0.005	0.096 ± 0.023	0.110 ± 0.031
269	Gln	0.085 ± 0.015	0.103 ± 0.001	0.111 ± 0.013	0.106 ± 0.020	0.135 ± 0.024
270	Ala	0.090 ± 0.023	0.091 ± 0.001	0.112 ± 0.011	0.109 ± 0.022	0.121 ± 0.011
271	Thr	0.115 ± 0.036	0.099 ± 0.005	0.137 ± 0.010	0.138 ± 0.027	0.143 ± 0.008
272	Met	0.121 ± 0.042	0.095 ± 0.008	0.139 ± 0.010	0.146 ± 0.032	0.157 ± 0.016
273	Asp	0.126 ± 0.056	0.093 ± 0.011	0.142 ± 0.010	0.150 ± 0.038	0.163 ± 0.023
274	Glu	0.109 ± 0.051	0.076 ± 0.006	0.117 ± 0.011	0.118 ± 0.031	0.131 ± 0.021
275	Arg	0.101 ± 0.040	0.068 ± 0.008	0.104 ± 0.011	0.109 ± 0.028	0.112 ± 0.019
276	Asn	0.103 ± 0.047	0.069 ± 0.010	0.114 ± 0.011	0.129 ± 0.034	0.128 ± 0.022
277	Arg	0.101 ± 0.042	0.067 ± 0.006	0.113 ± 0.012	0.123 ± 0.033	0.128 ± 0.023
278	Gln	0.092 ± 0.035	0.067 ± 0.007	0.097 ± 0.010	0.108 ± 0.029	0.112 ± 0.025
279	Ile	0.094 ± 0.037	0.065 ± 0.008	0.099 ± 0.013	0.115 ± 0.032	0.109 ± 0.027
280	Ala	0.092 ± 0.035	0.065 ± 0.007	0.103 ± 0.015	0.134 ± 0.045	0.116 ± 0.022
281	Glu	0.083 ± 0.030	0.064 ± 0.010	0.089 ± 0.012	0.133 ± 0.069	0.110 ± 0.022
282	Ile	0.081 ± 0.037	0.067 ± 0.013	0.077 ± 0.015	0.120 ± 0.057	0.097 ± 0.034
283	Gly	0.097 ± 0.040	0.076 ± 0.020	0.095 ± 0.017	0.149 ± 0.062	0.115 ± 0.033
284	Ala	0.095 ± 0.033	0.081 ± 0.024	0.098 ± 0.015	0.142 ± 0.048	0.119 ± 0.027
285	Ser	0.092 ± 0.037	0.076 ± 0.017	0.088 ± 0.013	0.119 ± 0.029	0.108 ± 0.027
286	Leu	0.095 ± 0.048	0.079 ± 0.026	0.097 ± 0.032	0.129 ± 0.037	0.114 ± 0.038
287	Ile	0.113 ± 0.052	0.096 ± 0.029	0.130 ± 0.055	0.169 ± 0.054	0.131 ± 0.037
288	Lys	0.122 ± 0.048	0.104 ± 0.022	0.139 ± 0.046	0.184 ± 0.058	0.135 ± 0.030
289	His	0.127 ± 0.027	0.134 ± 0.028	0.145 ± 0.031	0.178 ± 0.045	0.126 ± 0.017
290	Trp	0.097 ± 0.011	0.189 ± 0.042	0.122 ± 0.022	0.121 ± 0.027	0.105 ± 0.006
291			0.223 ± 0.017			
292			0.334 ± 0.016			
Average		0.074 ± 0.028	0.080 ± 0.037	0.090 ± 0.039	0.090 ± 0.046	0.096 ± 0.043

Table A 5-7 Differences in measured R_2 (1/ τcp) at fast (τcp = 0.625 ms) and slow (τcp = 10 ms) pulsing rates (ΔR_2 values ≥ 7s^{-1}).

Derived from TROSY-based ^{15}N-CPMG relaxation dispersion experiments, and unassigned residues (NA) for the chimeras cTEM-2m and cTEM-19m.

cTEM-2m									
Residue	**ΔR_2 (800 MHz)**			**k_{ex}**			**Population (pA)**	**α**	**Timescale of exchange**
Asp38	33.1	±	2.7	466.5	±	214.3	0.50		
Arg43	13.3	±	0.4	888.1	±	114.4	0.71	0.00	slow
Glu48	13.7	±	0.7	441.8	±	173.2	0.90	0.44	slow
Thr69	8.8	±	1.0	936.6	±	254.4	0.50		
Ser70		NA							
Tyr105	7.8	±	0.3						
Asp115		NA							
Ser130	10.5	±	1.0	587.6	±	188.4	0.50	0.00	slow
Asp131	9.0	±	0.8	295.7	±	113.5	0.50		
Glu212	19.5	±	2.1	1,403	±	582.6	0.50	0.99	intermediate
Ala217	11.2	±	1.0	1,283	±	534.8	0.50	1.30	fast
Leu221	11.2	±	1.2	1,406	±	618.8	0.98	0.22	slow
Leu225	12.3	±	0.6						
Ile231	9.5	±	0.4						
Lys234	8.9	±	1.0						
Ala237		NA							
Gly238	12.9	±	0.8	755.3	±	179.7	0.50	1.29	fast
Arg241	8.2	±	0.8	596.7	±	202.4	0.50	0.68	slow
Gly242	35.9	±	1.3						
Ser243	7.5	±	1.0	819.5	±	211.8	0.93		
Arg244	12.6	±	0.8						
Gly245	9.3	±	0.7	2,051	±	577.5	0.98		
Ile246	8.5	±	0.5	1,320	±	544.4	0.99		
Arg259	11.6	±	1.0	687.3	±	251.8	0.50		
Gly267	12.7	±	0.9	2,223	±	716.8	0.98	0.51	slow
Asn276	7.7	±	0.3	438.6	±	110.8	0.88	0.24	slow
Arg277	8.3	±	0.4	1,427	±	337.6	0.93		
Ala284	7.3	±	0.5						
Ile287	10.1	±	0.6						
cTEM-19m									
Residue	**ΔR_2 (800 MHz)**			**k_{ex}**			**Population (pA)**	**α**	**Timescale of exchange**
Glu28	11.5	±	4.3	443.0	±	125.5	0.50	1.45	fast
Thr29	12.1	±	0.6	485.6	±	130.0	0.50		
Lys32	18.6	±	0.5	323.4	±	138.7	0.93		
Asp38	9.9	±	0.6	291.1	±	138.8	0.87		
Gly41	29.5	±	1.2	1,191	±	182.7	0.93	0.87	inter
Arg43	29.3	±	1.2	619.9	±	121.6	0.93	0.23	slow
Val44	27.4	±	6.4	700.9	±	326.3	0.92		
Glu48	18.2	±	0.5						
Leu49	12.2	±	0.5						
Asp50	18.9	±	0.5						
Ser59	11.5	±	0.4	257.5	±	80.13	0.87	1.39	fast
Phe66	21.0	±	0.9	629.3	±	144.8	0.95	0.00	slow
Thr69	18.9	±	2.7						
Ser70		NA							
Phe72	8.9	±	0.9	593.1	±	166.6	0.50		
Lys73	7.5	±	0.5	1,243	±	609.5	0.50		

Gln88	18.0	±	1.1						
Tyr105	14.9	±	0.3	1,219	±	225.4	0.96	0.71	slow/inter
Ala125	23.4	±	0.8	1,091	±	225.4	0.93	1.05	inter
Ala126	8.5	±	0.4	573.6	±	194.3	0.50		
Thr128	25.9	±	2.5	1,561	±	382.1	0.94	2.02	fast
Met129	10.1	±	0.6	1,499	±	552.6	0.96	1.63	fast
Ser130	18.7	±	1.7	1,189	±	449.6	0.97	0.37	slow
Asp131	20.3	±	1.0	838.1	±	147.6	0.50	1.58	inter/fast
Asn132	30.9	±	1.3						
Glu166	7.1	±	0.4	307.2	±	188.2	0.97	0.00	slow
Asn170	12.3	±	0.7	672.8	±	176.3	0.95	0.96	inter
Glu171	10.8	±	0.9						
Gly172	30.9	±	1.5	2,105	±	842.4	0.94	0.93	inter
Lys173		NA							
Ile186	7.7	±	0.4	888.3	±	195.6	0.50	2.01	fast
Ile208	7.7	±	1.6						
Glu212	26.3	±	4.2						
Ala213	16.8	±	1.1	1,644	±	363.8	0.94		
Asp214		NA							
Val216		NA							
Ala217	27.1	±	2.7	1,761	±	484.7	0.96	0.60	slow/inter
Gly218		NA							
220 - 222		NA							
Leu225	12.9	±	0.6	577.7	±	150.4	0.50	1.64	fast
Ala227	10.9	±	0.2	769.2	±	260.7	0.50		
Trp229	11.9	±	0.4	642.2	±	187.6	0.50		
Phe230	9.2	±	0.4	1,086	±	462.8	0.50		
Ile231	19.6	±	0.7	638.9	±	196.9	0.50	1.93	fast
Ala232	26.5	±	1.0						
Asp233	18.1	±	0.8	1,105	±	215	0.96	0.79	inter
235-243		NA							
Arg244	34.7	±	4.7						
245-246		NA							
Ile247	13.2	±	0.8						
Ser258	15.9	±	1.0						
Arg259	11.3	±	0.7	760.9	±	271.4	0.95	0.90	inter
Ile260	11.3	±	0.5	441.7	±	161.8	0.50	2.03	fast
Tyr264	34.4	±	1.4	704.3	±	191.1	0.93	0.86	inter
Gly267	28.6	±	3.4	1,457	±	469.1	0.96	0.75	slow/inter
Ser268	7.8	±	1.5	1,099	±	407.2	0.50	0.58	slow
Gln269	21.1	±	1.3	1,171	±	215.3	0.91	1.87	fast
Ala270	16.4	±	1.2	385.7	±	173.5	0.50	2.04	fast
Met272		NA							
Arg275	12.1	±	0.7	706.8	±	202	0.92		
Asn276	9.8	±	0.7	293.7	±	85.38	0.90	1.57	fast
Arg277	19.5	±	0.6	539.8	±	215.8	0.95		
Gln278	16.7	±	0.6	942.2	±	124.1	0.95	0.95	inter
Ile279	18.5	±	0.9	819.8	±	121.3	0.50		
Ala280	16.4	±	1.2	504.3	±	203.1	0.50	2.03	fast
Ile282	10.7	±	0.4						
Ala284	11.8	±	0.6	805.6	±	169.8	0.82	1.35	inter
Leu286	9.6	±	0.3						
Ile287	25.6	±	1.3	336.9	±	115.5	0.90		

Table A 5-8 Comparison of the dynamism of the active-site walls using the S^2 (ps-ns), Cα-RMSF (ns-µs) and k_{ex} (µs-ms) on the different timescales.

	TEM-1			PSE-4			cTEM-2m			cTEM-17m			cTEM-19m		
	S^2	RMSF (nm)	k_{ex} (s^{-1})	S^2	RMSF (nm)	k_{ex} (s^{-1})	S^2	RMSF (nm)	k_{ex} (s^{-1})	S^2	RMSF (nm)	k_{ex} (s^{-1})	S^2	RMSF (nm)	k_{ex} (s^{-1})
Protein average	0.87 ± 0.05	0.07 ± 0.02	-	0.86 ± 0.05	0.08 ± 0.04	-	0.87 ± 0.05	0.09 ± 0.04	-	0.87 ± 0.05	0.09 ± 0.05	-	0.86 ± 0.05	0.10 ± 0.04	-
S70 (69-73)	0.91 ± 0.01	0.05 ± 0.01	-	0.90 ± 0.01	0.06 ± 0.01	-	0.89 ± 0.01	0.06 ± 0.01	937 ± 254 (69)	0.90 ± 0.01	0.06 ± 0.02	480 ± 110 (69, 72)	0.89 ± 0.01	0.06 ± 0.01	667 ± 141 (66, 72, 73)
Y105 (103-106)	0.86 ± 0.01	0.08 ± 0.02	-	0.85 ± 0.01	0.10 ± 0.03	-	0.85 ± 0.01	**0.14 ± 0.05**	-	0.86 ± 0.01	0.08 ± 0.01	300 ± 90 (105, 106)	0.85 ± 0.01	0.11 ± 0.04	1,219 ± 225 (105)
SDN (129-132)	0.89 ± 0.01	0.06 ± 0.02	-	0.87 ± 0.01	0.06 ± 0.01	-	0.88 ± 0.01	0.09 ± 0.02	480 ± 205 (130, 131)	0.88 ± 0.01	0.07 ± 0.01	775 ± 270 (128, 129)	0.87 ± 0.02	0.07 ± 0.02	979 ± 203 (128, 129, 130, 131)
Ω-loop (162-179)	0.85 ± 0.01	0.07 ± 0.01	-	0.85 ± 0.02	0.09 ± 0.02	2,019 ± 813 (173)	0.85 ± 0.02	0.10 ± 0.03	-	0.85 ± 0.03	**0.14 ± 0.09**	586 ± 134 (166, 170)	0.82 ± 0.03	**0.17 ± 0.09**	333 ± 100 (166, 170) ; Gly172: 2105 ± 842
Hinge leading in Ω-loop (155-160)	**0.79 ± 0.02**	**0.13 ± 0.01**	-	**0.80 ± 0.03**	**0.14 ± 0.07**	-	**0.81 ± 0.03**	**0.12 ± 0.02**	-	**0.80 ± 0.03**	**0.12 ± 0.04**	-	**0.79 ± 0.03**	**0.15 ± 0.02**	-
Tip of the Ω-loop (171-178)	**0.79 ± 0.01**	**0.10 ± 0.01**	-	**0.80 ± 0.03**	**0.13 ± 0.02**	2,019 ± 813 (173)	**0.78 ± 0.04**	**0.13 ± 0.03**	-	**0.80 ± 0.04**	**0.21 ± 0.15**	-	**0.78 ± 0.03**	**0.23 ± 0.12**	-
214-218	**0.80 ± 0.04**	**0.15 ± 0.05**	-	0.85 ± 0.01	0.07 ± 0.02	-	**0.76 ± 0.04**	**0.25 ± 0.10**	1,316 ± 868 (212, 217)	**0.77 ± 0.02**	**0.24 ± 0.08**	-	**0.78 ± 0.03**	**0.17 ± 0.04**	1,131 ± 423 (213, 217)
connector (214-225)	**0.81 ± 0.04**	**0.14 ± 0.05**	1,387 ± 368 (225)	0.85 ± 0.01	0.07 ± 0.02	-	**0.78 ± 0.04**	**0.21 ± 0.06**	1,339 ± 784 (212, 217, 221)	**0.78 ± 0.02**	**0.21 ± 0.07**	1,098 ± 443 (225)	**0.78 ± 0.03**	**0.17 ± 0.04**	816 ± 208 (213, 217, 225)
234-244	0.88 ± 0.01	0.06 ± 0.01	646 ± 29 (241)	0.89 ± 0.01	0.06 ± 0.01	-	0.88 ± 0.01	0.07 ± 0.01	611 ± 221 (238, 241, 243, 245, 246)	0.88 ± 0.01	0.07 ± 0.02	-	0.86 ± 0.02	0.09 ± 0.01	-
β-strand 3 (230-237)	0.88 ± 0.01	0.06 ± 0.01	-	0.90 ± 0.01	0.05 ± 0.01	-	0.89 ± 0.01	0.07 ± 0.01	-	0.89 ± 0.01	0.07 ± 0.02	910 ± 205 (231, 232, 233)	0.88 ± 0.01	0.06 ± 0.01	796 ± 250 (230, 231, 233)
Turn β-strand 3/β-strand 4 (238-242)	0.88 ± 0.01	0.07 ± 0.01	646 ± 29 (241)	0.88 ± 0.01	0.06 ± 0.01	-	0.88 ± 0.01	0.08 ± 0.01	600 ± 271 (238, 241, 243)	0.87 ± 0.01	0.08 ± 0.03	-	0.84 ± 0.02	0.12 ± 0.02	-
H1 (26-41)	0.88 ± 0.01	0.08 ± 0.01	-	0.85 ± 0.01	0.10 ± 0.01	-	0.88 ± 0.01	0.09 ± 0.01	467 ± 214 (38)	0.87 ± 0.01	0.09 ± 0.01	450 ± 60 (29, 32, 37, 38, 39, 41)	0.87 ± 0.01	0.10 ± 0.02	289 ± 53 (28, 29, 32, 38, 41)
H11 (272-288)	0.88 ± 0.01	0.10 ± 0.04	1,186 ± 335 (287)	0.86 ± 0.02	0.09 ± 0.02	-	0.87 ± 0.02	0.11 ± 0.02	800 ± 140 (276-277)	0.87 ± 0.01	0.13 ± 0.04	440 ± 50 (277, 278, 279, 281, 284, 287)	0.87 ± 0.01	0.12 ± 0.03	843 ± 61 (275, 276, 277, 278, 279, 280, 284, 287)

k_{ex} : residues used for the fitting are indicated between bracket

Table A 5-9 Kinetic parameters for hydrolysis of the cephalosporins.

Kinetic parameters for hydrolysis of the cephalosporins cephalothin (CF), cefazolin (CZ) and cefotaxime (CTX) and penicillins benzylpenicillin (BZ) and carbenicillin (CB) by the parental β-lactamases TEM-1 and PSE-4, the chimeras cTEM-2m, cTEM-17m and cTEM-19m and the deconvoluted mutants in the cTEM-17m context (cTEM-18m(M68L) and cTEM-18m(M69T)) and the TEM-1 context (TEM-1(M68L) and TEM-1(M69T)).

	Variant	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
	TEM-1[a]	84 ± 12	180 ± 28	4.3×10^5
	cTEM-17m[a]	120 ± 8	470 ± 80	2.4×10^5
	TEM-1(M68L)	31 ± 7	100 ± 30	3.0×10^5
	cTEM-18m(M68L)	34 ± 5	80 ± 38	5.0×10^5
CF	TEM-1(M69T)	7 ± 4	130 ± 11	5.2×10^4
	cTEM-18m(M69T)	1.4 ± 0.3	12 ± 3	1.2×10^5
	cTEM-2m	2.8 ± 0.8	46 ± 19	6.1×10^4
	cTEM-19m	6 ± 0.5	68 ± 14	8.8×10^5
	PSE-4[a]	0.80 ± 0.01	64 ± 34	1.3×10^4

	TEM-1[a]	55 ± 8	130 ± 9	4.2×10^5
	cTEM-17m[a]	75 ± 21	600 ± 90	1.2×10^5
	TEM-1(M68L)	59 ± 21	500 ± 178	1.2×10^5
	cTEM-18m(M68L)	23 ± 5	170 ± 43	1.4×10^5
CZ	TEM-1(M69T)	9 ± 2	300 ± 121	3.0×10^4
	cTEM-18m(M69T)	4 ± 0.6	83 ± 18	5.0×10^4
	cTEM-2m	3.3 ± 0.1	75 ± 1	4.4×10^4
	cTEM-19m	6 ± 0.3	120 ± 5	5.0×10^4
	PSE-4[a]	1.9 ± 0.3	140 ± 39	1.3×10^4
	TEM-1[a]	0.74 ± 0.1	840 ± 160	8.8×10^2
	cTEM-17m[a]	0.14 ± 0.04	260 ± 100	5.6×10^2
	TEM-1(M68L)	0.06 ± 0.02	170 ± 72	3.6×10^2
	cTEM-18m(M68L)	0.07 ± 0.03	260 ± 73	2.5×10^2
CTX	TEM-1(M69T)	0.03 ± 0.02	280 ± 151	1.1×10^2
	cTEM-18m(M69T)	0.06 ± 0.02	430 ± 78	1.4×10^2
	cTEM-2m	0.02 ± 0.01	240 ± 40	8.2×10^1
	cTEM-19m	0.008 ± 0.002	110 ± 37	7.0×10^1
	PSE-4[a]	0.03 ± 0.01	200 ± 30	1.5×10^2
	TEM-1[a]	450 ± 100	20 ± 3	2.3×10^7
	cTEM-17m[a]	480 ± 46	28 ± 8	1.7×10^7
	TEM-1(M68L)	310 ± 19	11 ± 3	2.8×10^7
	cTEM-18m(M68L)	400 ± 50	31 ± 17	1.3×10^7
BZ	TEM-1(M69T)	120 ± 3	7 ± 1	1.7×10^7
	cTEM-18m(M69T)	20 ± 5	6 ± 4	3.3×10^6
	cTEM-2m	83 ± 2	7 ± 1	1.2×10^7
	cTEM-19m*	230 ± 19*	< 1*	-
	PSE-4[a]	630 ± 110	16 ± 2	3.9×10^7
	TEM-1[a]	92 ± 4	49 ± 12	1.9×10^6
	cTEM-17m[a]	37 ± 2	22 ± 4	1.7×10^6
	TEM-1(M68L)	51 ± 3	28 ± 5	1.8×10^6
	cTEM-18m(M68L)	65 ± 13	37 ± 14	1.8×10^6
CB	TEM-1(M69T)	31 ± 3	12 ± 5	2.6×10^6
	cTEM-18m(M69T)	10 ± 5	26 ± 21	3.9×10^6
	cTEM-2m	24 ± 1	16 ± 3	1.5×10^6
	cTEM-19m	25 ± 6	27 ± 10	9.3×10^5
	PSE-4[a]	500 ± 92	64 ± 5	7.8×10^6

[a] Data from Clouthier *et al.* [2]

*k_{cat} (apparent) and estimated K_M due to saturation at a substrate concentration ≥ 5 μM.

5.2. Figures

Residue	68	69	150	153	154	155	158	159	165	168	172	173	174	175	177	182	184	186	188
TEM-1 cTEM-2m	M L	M T	A	H	N	M	H	V	W	E	A	I	P	N	E	M	A	M	T
cTEM-17m cTEM-19m PSE-4	M L L	M T T	D	R	Q	I	K	E	I	D	G	K	L	G	L	T	K	I	S

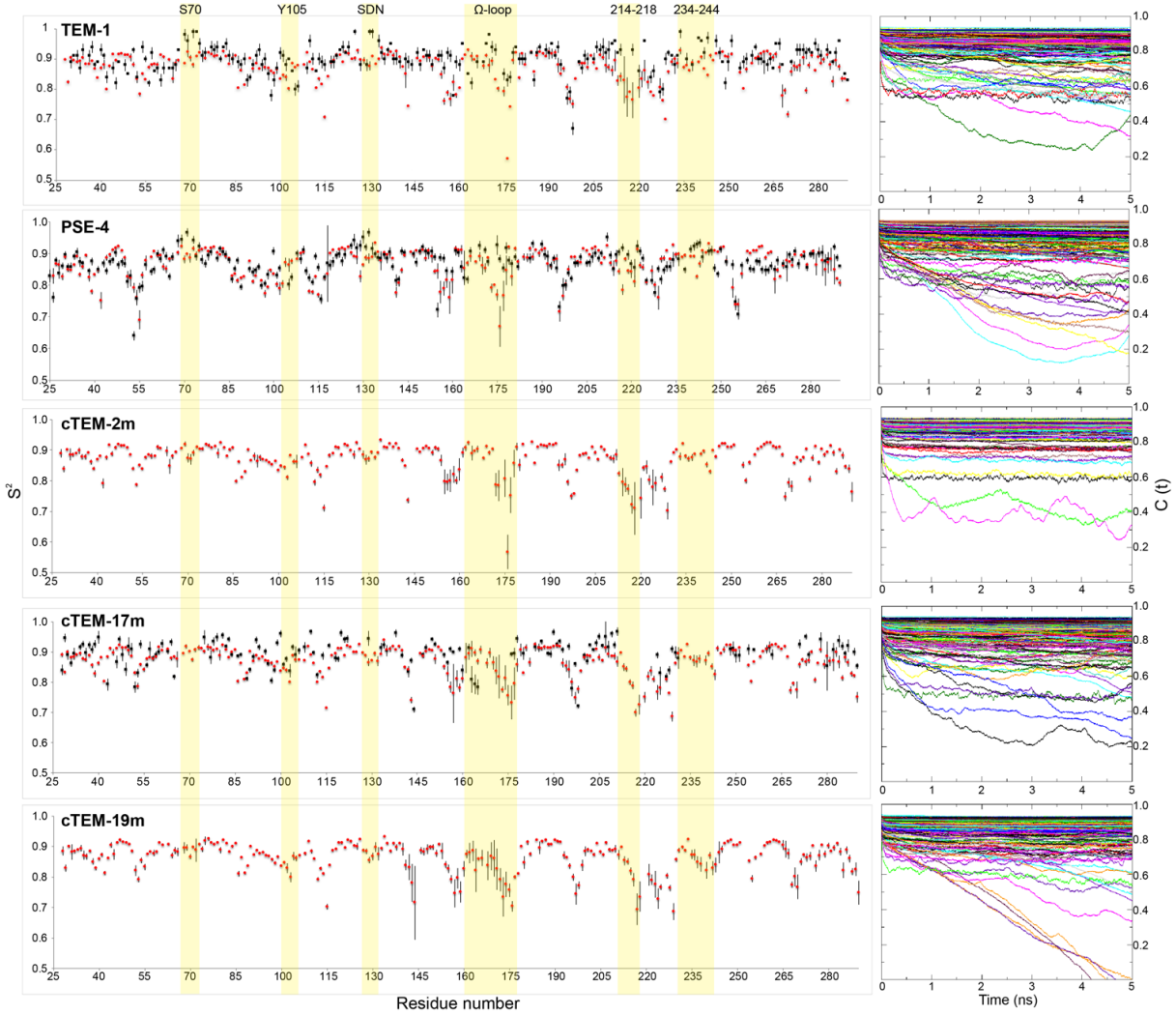


Figure A 5-1 Sequence relation in the exchanged regions of the chimeras (residues 66-73 and 150-190) to the parental TEM-1 and PSE-4 class A β-lactamases and generalized squared order parameters (S^2) along the protein sequence for the amide NH bonds.
MD-derived (red) and NMR relaxation (black) S^2 are compared for TEM-1 and PSE-4 and the chimera cTEM-17m. MD-derived S^2 (red) are shown for cTEM-2m and cTEM-19m. The mean MD-NMR S^2 difference (or Δ) was the highest for Asp176 (Δ = 0.22 in TEM-1, 0.21 in PSE-4

and 0.16 in cTEM-17m), Asp115 (Δ = 0.17 in TEM-1; no value in cTEM-17m), Gly143 (Δ = 0.16 in TEM-1 and 0.13 in cTEM-17m) and Trp229 (Δ = 0.21 in TEM-1 and 0.22 in cTEM-17m). Examples of the autocorrelation functions for all residues for one among the 100 analyzed 10 ns segments is shown on the right.

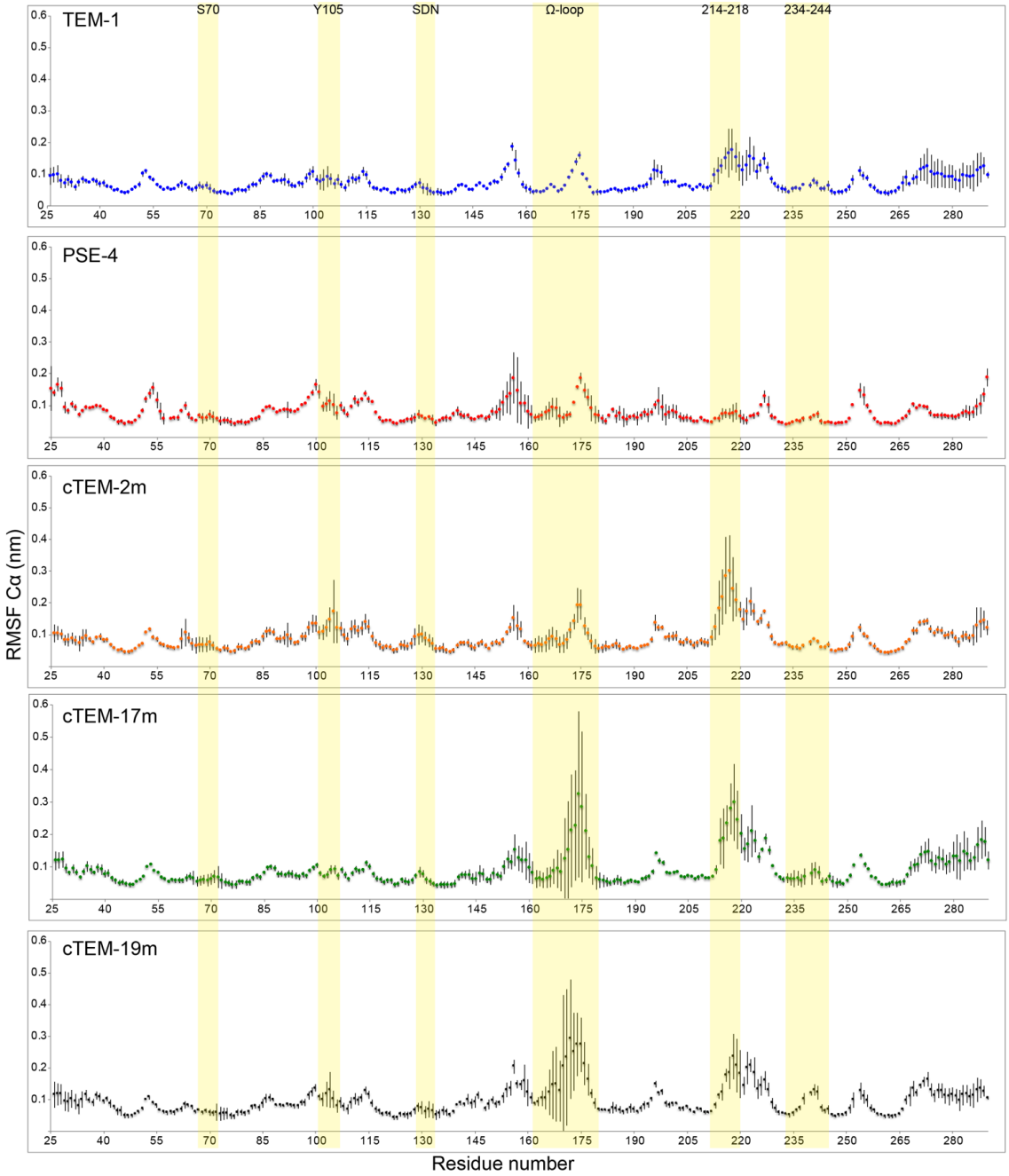


Figure A 5-2 Cα RMSF derived from triplicate 2 µs MD simulations.
For the parental class A β-lactamases TEM-1 and PSE-4 and the chimeras cTEM-2m, cTEM-17m and cTEM-19m.

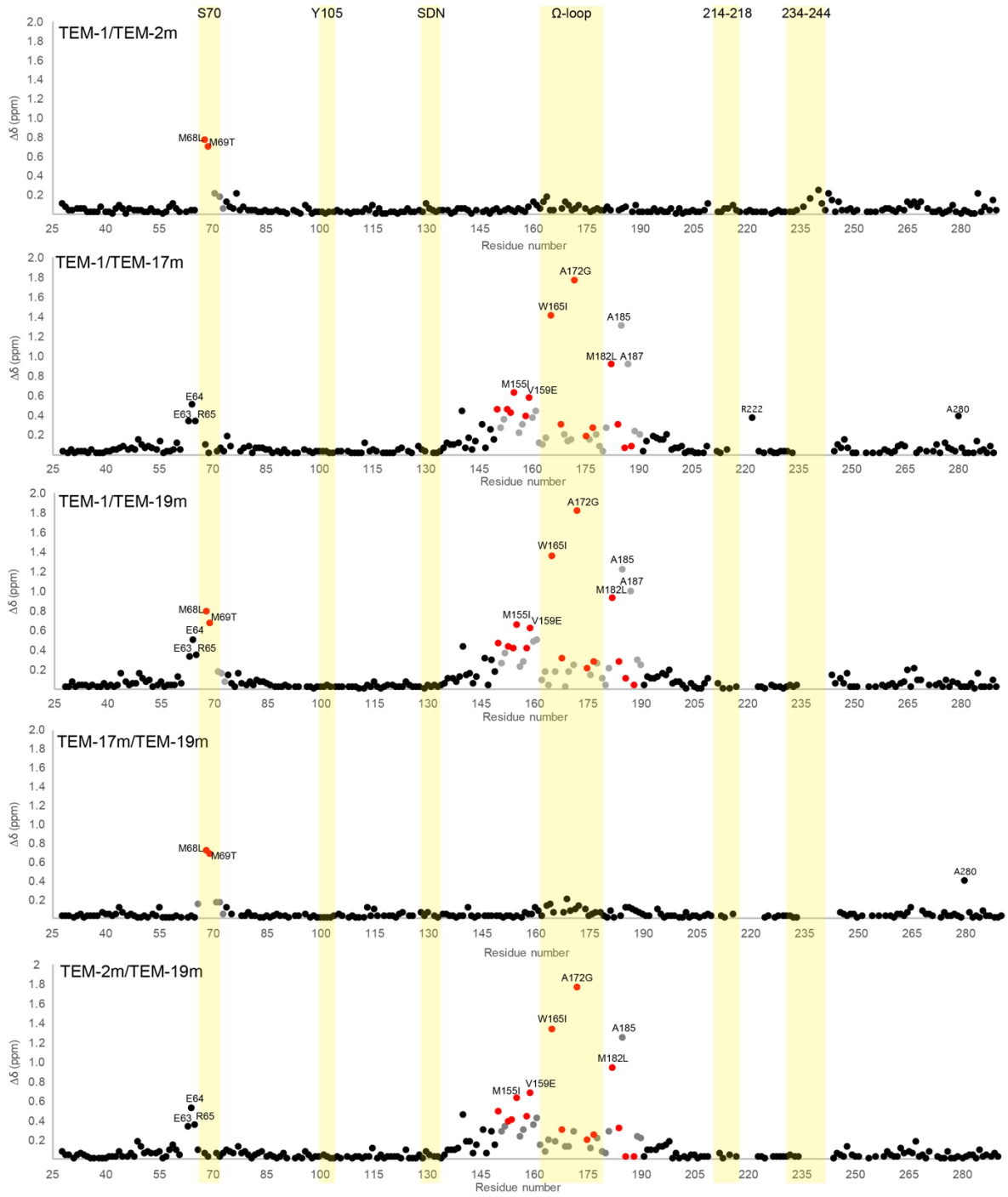
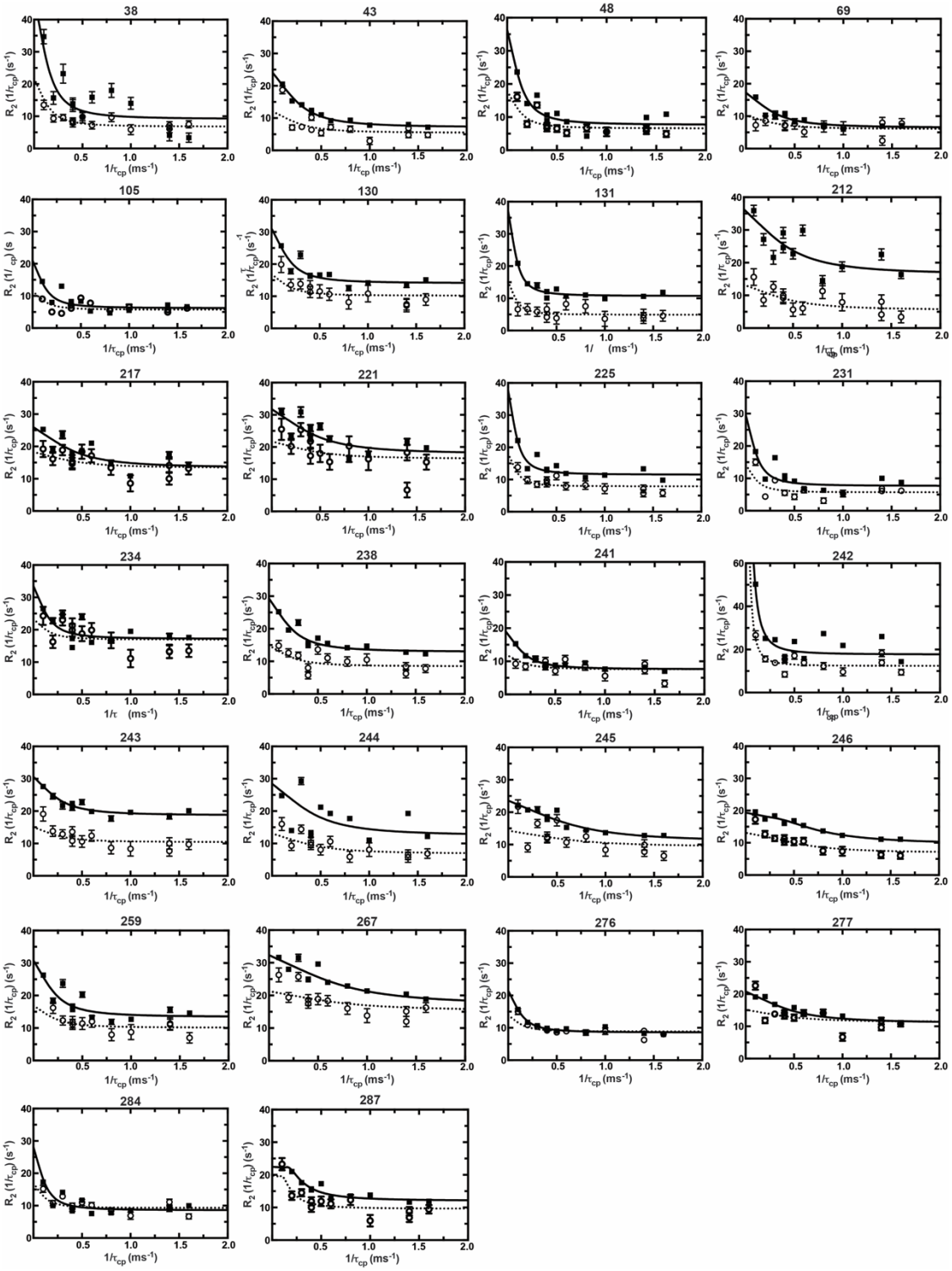


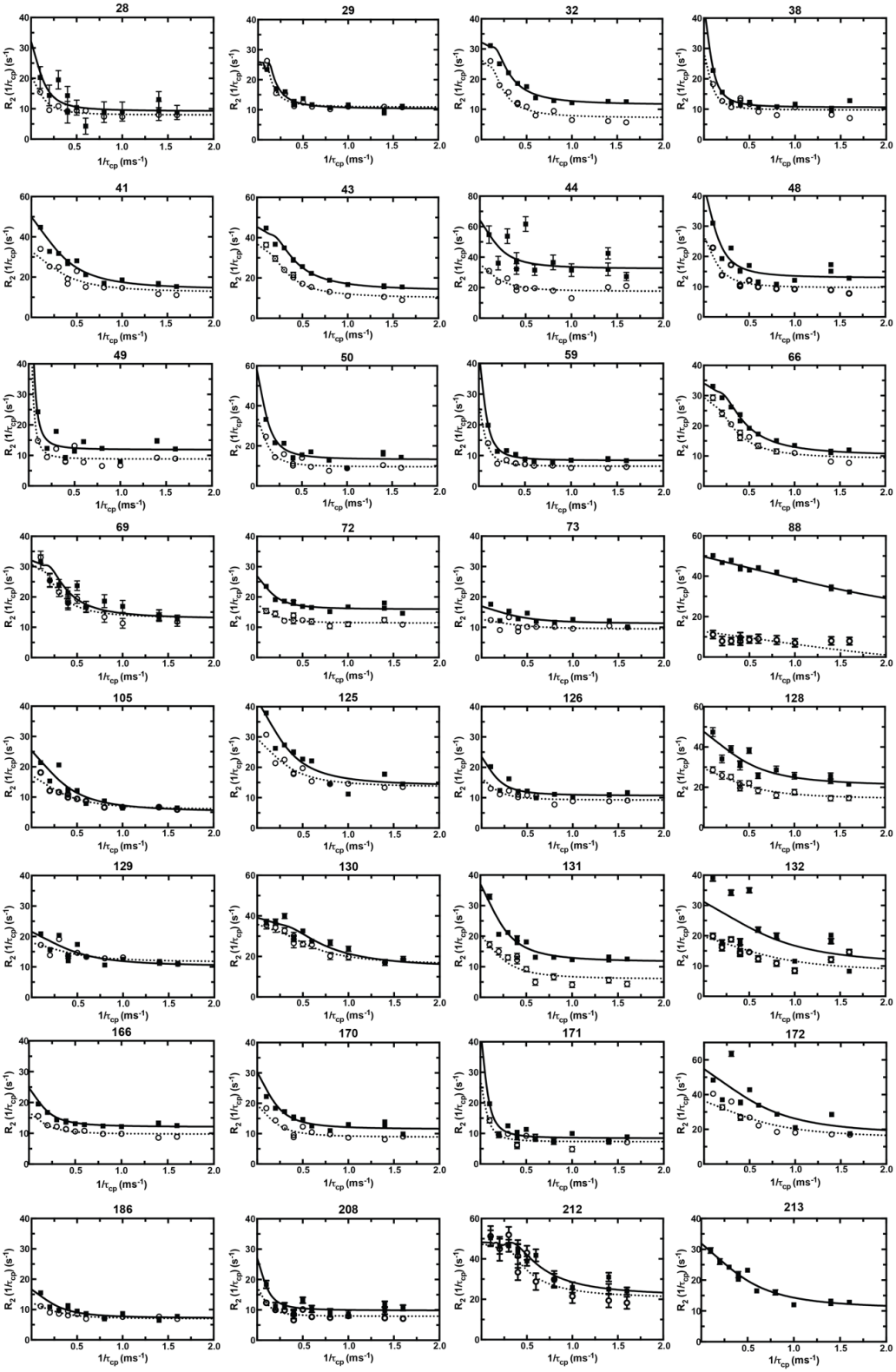
Figure A 5-3 Effect of the recombination of regions 66-73 and 150-190 on the backbone chemical shifts.

Comparison of the chimeras cTEM-2m, cTEM-17m and cTEM-19m to the parental TEM-1 class A β-lactamases and comparisons of cTEM-2m and cTEM-17m to cTEM-19m. Exchanged residues relative to TEM-1 sequence are colored red while residues in the exchanged regions (66 to 73 and 150 to 190) are colored grey.

A)


38
43
48
69
105
130
131
212
217
221
225
231
234
238
241
242
243
244
245
246
259
267
276
277
284
287
R2 (1/τcp) (s-1)
1/τcp (ms-1)

B)


28
29
32
38
41
43
44
48
49
50
59
66
69
72
73
88
105
125
126
128
129
130
131
132
166
170
171
172
186
208
212
213
$R_2(1/\tau_{cp})(s^{-1})$
$1/\tau_{cp}(ms^{-1})$

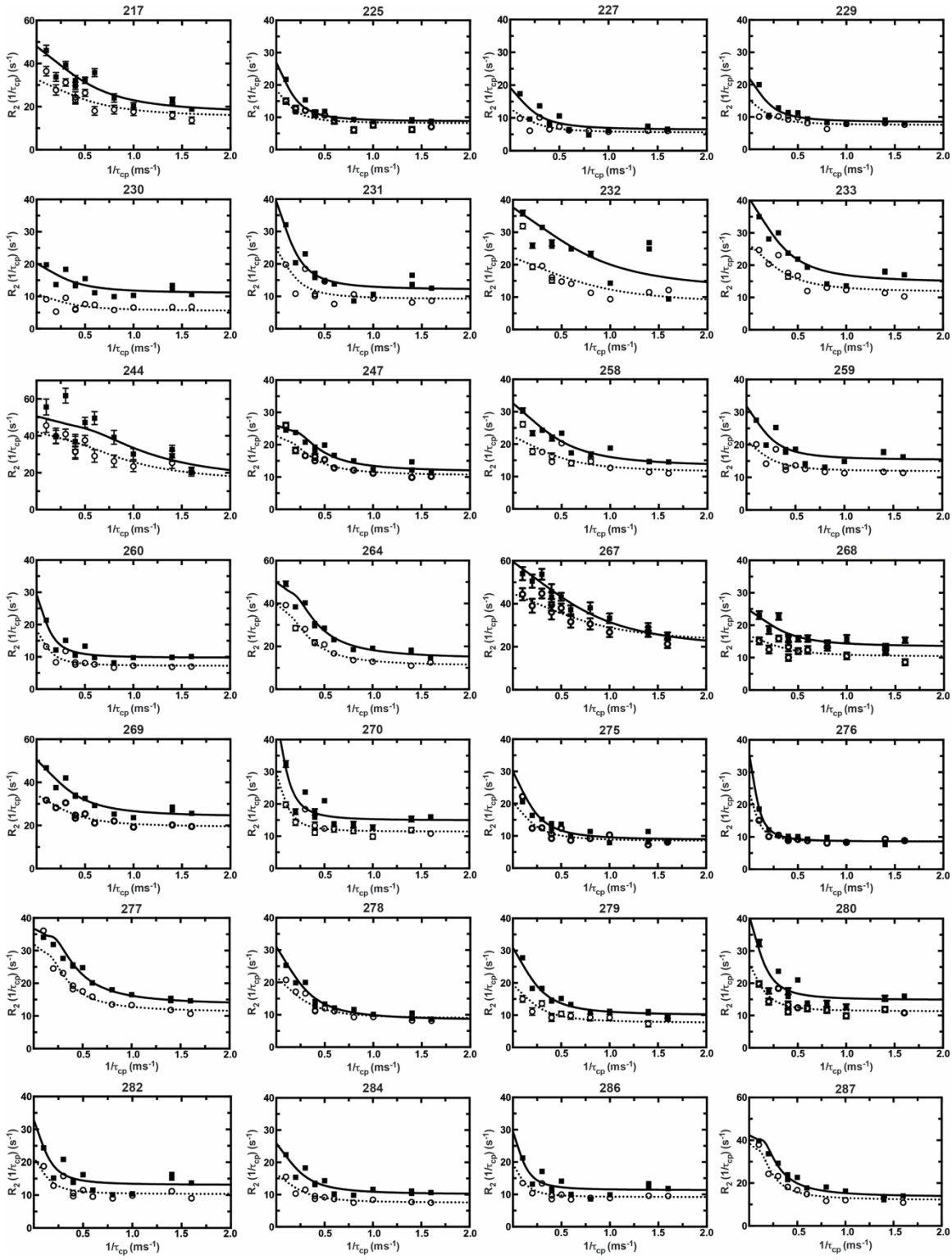


Figure A 5-4 A) and B) ^{15}N-CPMG relaxation dispersion curves for cTEM-2m at 800 MHz (full line) and 500 MHz (dashed line). The residue number appears above each plot.

5.3. Materials and methods

5.3.1. Materials

All DNA-modifying enzymes were purchased from New England Biolabs (Mississauga, ON) or Bio-Rad (Mississauga, ON). Nitrocefin was purchased from Calbiochem (Mississauga, ON). Kanamycin (Kan), isopropyl-1-thio-β-D-galactopyranoside (IPTG), components for growth media and purification materials were from Bioshop Canada (Burlington, ON). 99% Ammonium chloride ($^{15}NH_4Cl$) and 99% deuterium oxide (D_2O) were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). 99% D-glucose-^{13}C, benzylpenicillin (BZ), carbenicillin (CB), cephalotin (CF), cefazolin (CZ), cefotaxime (CTX), DEAE-Sepharose Fast Flow, Tris-HCl, PEG 4000, HEPES, ammonium sulfate and $MgCl_2$ used for protein crystallization were purchased from Sigma-Aldrich (Oakville, ON). Tazobactam was purchased from Molekula (Dorset, UK).

5.3.2. Mutagenesis and subcloning

Subcloning of the chimera cTEM-19m into pET-24 with the OmpA signal sequence was performed as previously described [2]. Oligonucleotide primers used for mutagenesis of TEM-1 and cTEM-19m were synthesized by Alpha DNA (Montréal, QC). To generate the mutants, the QuickChange site-directed mutagenesis method [3] was used with the I-Proof High Fidelity DNA polymerase (Bio-Rad, Mississauga, ON). The PCR product was *Dpn*I-digested for 1 hour at 37°C to eliminate the starting DNA template. The digested product was butanol-precipitated and transformed into *E. coli* XL1 Blue. Cells were propagated for 1h in Luria-Bertani broth (LB) then plated on LB-agar containing 50 μg/mL Kan. Colonies were picked and propagated overnight. DNA extraction was performed for confirmation of DNA sequence of the entire coding region and for retransformation into *E. coli* BL21(DE3).

5.3.3. β-Lactamase expression and purification

Protein expression for enzymatic assays and crystallisation was performed as previously reported [2, 4-6]. Briefly, cells were propagated in auto-inducing ZYP-5052 medium [7] (50 μg/mL Kan) with agitation at 37°C for two hours, then overnight at 22°C.

Induced cells were sedimented by centrifugation (5,000 × *g*, 30 min, 4°C) and resuspended in 10 mM Tris-HCl buffer pH 7.0. Lysis was performed using a Constant Systems cell disrupter (Northants, UK). Cellular debris were sedimented by centrifugation (20,000 × *g*, 30 min, 4°C), the supernatant was filtered through a 0.2 μm filter and the protein was purified. The chromatography was undertaken at 4°C using an ÄKTA FPLC (GE Healthcare). Following sample application, the DEAE-Sepharose Fast Flow column (1.6 cm × 30 cm) was washed with 3 column volumes (CV) of 10 mM Tris-Cl pH 7.0. A linear gradient to 200 mM Tris-Cl pH 7.0 was achieved over 4 CV, and the column was further washed over 3 CV. Fractions containing β-lactamase were identified first using a qualitative nitrocefin hydrolysis essay and then by SDS-polyacrylamide gel electrophoresis (15% (w/v) polyacrylamide gel) with Coomassie Brilliant Blue staining. Analysis of the purity was performed with Image Lab 5.0 (Bio-Rad, Hercules, CA). Enzyme concentration was determined by Bradford assay using the Bio-Rad protein assay solution (Bio-Rad, Hercules, CA), with bovine serum albumin as the standard. For protein crystallization assay, a second purification step was added. Fractions containing ≥ 75% β-lactamase following the first purification step were concentrated to a volume of 1.5 mL using a 10,000 MWCO Amicon concentrator (Millipore, Billerica, MA) and applied to a Superose 12 column (1.6 cm × 55 cm) that had been pre-equilibrated with 50 mM Tris-HCl pH 7.0. Elution was performed at a flow rate of 1 mL/min. Fractions containing β-lactamase were identified by SDS-polyacrylamide gel electrophoresis (15% (w/v) polyacrylamide gel) with zinc-imidazole staining [8]. Purity and concentration were determined as above. Fractions containing β-lactamase with a purity of ≥ 95% were concentrated to 25 mg/mL (0.8 mM). For NMR assays, uniformly [^{15}N]- and [^{2}H, ^{15}N]-labeled cTEM-19m and cTEM-2m samples were produced by overexpression in modified M9 minimal media containing $^{15}NH_4Cl$ and deuterium oxide D_2O (99%) [2, 4-6, 9]. For [^{2}H, ^{15}N]-labeled samples, cells were progressively acclimatised to deuterium oxide using M9 media with 30, 60, 75, 85 and 99% D_2O. Purification was performed as described for protein crystallization assay. Samples were dialyzed against distilled, deionized water overnight at 4°C. Purity of the final samples was verified by SDS-PAGE stained by the zinc-imidazole method. Typical yields were 25 mg/L of > 98% pure protein.

5.3.4. cTEM-19m crystallization and data collection

cTEM-19m was concentrated to 25 mg/mL in 50 mM Tris-Cl (pH 7.0) in the presence and absence of 4 mM tazobactam (5-fold molar excess relative to the protein concentration). Crystals were grown at 22°C in hanging drops prepared by mixing 1 μL of the protein solution and 1 μL of the reservoir solution (100 mM Tris-Cl (pH 8.0), 26% PEG 4000, and 0.25 M $MgCl_2$). Two single crystals, one grown in the presence of tazobactam and the other in its absence, were isolated and flash-frozen in liquid nitrogen. Diffraction data were collected from each crystal at the Canadian Macromolecular Crystallography Facility Beamline 08ID-1 (Canadian Light Source, Saskatoon, SK). Although the classical β-lactamase inhibitor tazobactam (turnover rate of 1.4 s^{-1}) [10] was included during crystallization for cTEM-19m, no corresponding electron density was observed in 4R4S.

5.3.5. cTEM-2m crystallization and data collection

cTEM-2m was concentrated to 25 mg/mL in 50 mM Tris-Cl (pH 7.0). Crystals were grown at 22°C in hanging drops prepared by mixing 1 μL of the protein solution and 1 μL of the reservoir solution, which contained 100 mM HEPES (pH 7.5) and 1.5 M ammonium sulfate. A rod-shaped crystal was removed from a crystal cluster, transferred to a cryoprotectant solution containing 10% (v/v) glycerol, 100 mM HEPES (pH 7.5) and 1.5 M ammonium sulfate, and flash-cooled in a nitrogen gas cold-stream. Diffraction data were collected at 100K as 0.5° oscillation images, with a Rigaku MicroMax 007 HF X-ray generator and a Rigaku Saturn 944 HG CCD detector.

5.3.6. Crystallization data processing and refinement

The collected diffraction images were indexed, integrated, and scaled either with HKL2000 [11] or the *xia2* package [12]. Initial phases were calculated by molecular replacement with the program PHASER [13], with a previously determined TEM-1 structure (PDB code: 1ZG4 [14]) as a search model. The structure models were further improved through iterative rounds of manual model building with COOT [15] and automated refinement with PHENIX [16] and REFMAC5 [17]. Data collection and refinement statistics are

summarized in Table A 5-2. The refined structures of cTEM-19m (PDB IDs 4R4R and 4R4S) and cTEM-2m (PDB ID: 4MEZ) have been deposited to the Protein Data Bank.

5.3.7. Active-site Volume Estimation

The active-site cavity volume of TEM-1 (PDB 1XPB), PSE-4 (PDB 1G68), cTEM-2m (PDB 4MEZ), cTEM-17m (PDB 4ID4) and cTEM-19m (PDB 4R4S) was estimated using 3V: Voss Volume Voxelator [18]. The estimation was made with a small sphere of 1.5Å radius and a large sphere of 8 Å radius.

5.3.8. Molecular dynamics simulations

Crystal structures with the best ratio of resolution and completeness of the sequence were chosen for the two parents (TEM-1: 1XPB, 1.9 Å resolution and PSE-4: 1G68, 1.95 Å resolution) and the three variants (cTEM-2m: 4MEZ, 2.05 Å resolution, cTEM-17m: 4ID4, 1.05 Å resolution and cTEM-19m: 4R4S, 1.1 Å resolution). These structures were used as a starting point for all molecular dynamics (MD) simulations. Missing side-chains in 4MEZ (LYS32, LYS55, LYS111, ASP115, LYS146, LYS1192 and ARG277), 4ID4 (LYS32, LYS55, LYS111, LYS146, GLN154, LYS158, LYS173, ASP254 and LYS256) and 4R4S (LYS 111, GLN 154, LYS 158, LYS215 and LYS256) were reconstructed using the rotamer explorer in the Structure Preparation plugin of MOE [19]. All following steps were executed using GROMACS 5.0.1. The protonation and orientation of histidines was verified using the high-resolution crystal structure of TEM-1 variant M182T (1M40), MolProbity [20] and PROPKA [21, 22]. For each system, crystallographic water molecules were conserved in a truncated dodecahedron periodic water box of SPC/E water molecules with a minimum distance of 10 Å between the box boundary and the protein. The total number of water molecules and net charge for the different systems were: 1XPB: 12,812, -7; 1G68: 12,475, -5; 4MEZ: 9,952, -7; 4ID4: 10,177, -4; 4R4R: 10,703, -4. Each system was neutralized by adding the respective number of sodium counterions. The total size of the systems, in atoms, was: 1XPB: 42,507; 1G68: 41,501; 4MEZ: 33,926, 4ID4: 34,619; 4R4R: 36,196.

For energy minimization and MD simulations, GROMACS 5.0.1 was used with the AMBER99SB-ILDN force field [23, 24]. The LINear Constraint Solver (LINCS) algorithm

was applied to all bonds containing hydrogen atoms. Each 1 ps frame was saved in a compressed XTC trajectory and each 100 ps in a full precision TRR trajectory. The Nose-Hoover thermostat and the Parrinello-Rahman barostat were used to couple the system to a constant temperature of 304.65 K and a pressure of 1 bar. The electrostatic interactions were evaluated by the particle-mesh Ewald method, and Lennard-Jones interactions were evaluated using a 1.2 nm cutoff. Each system was energy minimized using the method of the Steepest Descent to a target F_{max} of no greater than 1,000 kJ mol^{-1} nm^{-1}. The time step was chosen to be 2 fs. For equilibration, all systems were subjected to 2 ns NVT ensemble to gradually heat up the system from 10 K to 304.65 K. Next, a 20 ns NPT ensemble was generated and three conformations with the highest, average and lowest root mean-square deviation (RMSD) were chosen as starting conformations for three individual replicates. A further 2 ns NVT ensemble was performed for each repetition to heat up the system and used as starting point for the 2 μs production run. This results in a total of three 2 μs simulations with 1 ps compressed and 100 ps full precision trajectories for each of the five proteins.

5.3.9. Analysis of protein structure and dynamics

Analysis was performed using GROMACS 5.0.1 on 2 μs for each MD simulation, unless otherwise stated. All protein figures were created using PyMOL [25] or MOE (Chemical Computing Group, Montréal, QC). The stability of the system was evaluated after 100 ns and 1,000 ns by calculating the RMSD of the backbone atoms between each frame and the initial conformation. In addition, the radius of gyration was calculated for each replicate over 2 μs to ensure overall structural integrity. The root mean-square fluctuation (RMSF) was calculated based on the Cα atom for each residue and converted into b-factors after removal of the overall translational and rotational motion of the protein by superimposition onto a common reference frame.

S^2 order parameters were calculated applying the model-free approach of Lipari and Szabo [26, 27]. After removal of translational and rotational motion for each 2 μs run, the trajectory was split into 10 ns segments. For every second segment, the autocorrelation function was calculated for each N-H internuclear vector. The final S^2 value was obtained by averaging all 10 ns segments, thus representing the S^2 value over 2 μs.

5.3.10. Solution NMR experiments

All NMR experiments were performed as previously reported [4, 5]. Briefly, [^{15}N] and [^{2}H,^{15}N]-Labeled samples were concentrated to 0.6 mM using a 10,000 MWCO Amicon concentrator and buffer exchanged to 3 mM imidazole, 0.01% NaN_3, and 10% D_2O, pH 6.7. The final pH was verified using a pH meter with a microprobe. The temperature was calibrated to 31.5°C with a standard methanol sample [28] on Agilent 800 MHz (18.8T – cTEM-2m and cTEM-19m), Agilent 600 MHz (14.1T, cTEM-19m) and Agilent 500 MHz (11.7T, cTEM-2m) NMR spectrometers, each equipped with a triple-resonance probe and a pulse-field gradient. Microsecond-millisecond protein dynamics for cTEM-2m and cTEM-19m were characterized using ^{15}N TROSY relaxation-compensated Carr-Purcell-Meiboom-Gill (rcCPMG) experiments [29]. Transverse spin-relaxation data using the TROSY rcCPMG experiment were acquired in an interleaved fashion with interpulse delays, τ_{cp}, of 0.0, 0.625, 0.714 (× 2), 1.0, 1.25, 1.67, 2.0, 2.50 (× 2), 3.33, 5.0, and 10.0 ms during the nitrogen relaxation period for a constant, total relaxation time of 20 ms. All heteronuclear NMR data were processed using NMRPipe [30], in-house scripts and further analyzed with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Previously acquired and analysed NMR backbone assignments of cTEM-2m (BMRB 26586) and cTEM-19m (BMRB 26590) were used to identify the N-H resonances [5]. Conformational exchange was considered significant when the difference in measured R_2 ($\Delta R_2\,(1/\tau_{cp})$) at fast ($\tau_{cp}$ = 0.625 ms) and slow (τ_{cp} = 10 ms) pulsing rates on the 18.8T data set was greater than 7 s^{-1} in relaxation dispersion profiles. This threshold was selected based on relaxation dispersion error estimation, to eliminate false positives resulting from a poor fit and to be consistent with the previously reported dynamics of TEM-1, PSE-4 and cTEM-17m [4]. Global residue fits and model analyses were performed using the 800 MHz (cTEM-2m and cTEM-19m) in combination with the 500 MHz (cTEM-2m) or 600 MHz (cTEM-19m) relaxation dispersion data and fitting them using GraphPad Prism 5 to the full single-quantum ^{15}N-CPMG equation [31]. The residue were selected for global fitting, first according to structural proximity. Then, residues with statistically overlapping k_{ex} where fitted as a region.

5.3.11. Enzyme kinetics

K_M and k_{cat} parameters were determined at room temperature in 50 mM sodium phosphate buffer, pH 7.0. The following extinction coefficients (and concentration ranges) were used: $\Delta\varepsilon_{232\ nm}$ = 900 $M^{-1}cm^{-1}$ for BZ [32] (5-240 μM), $\Delta\varepsilon_{232\ nm}$ = 1,190 $M^{-1}cm^{-1}$ for CB [33] (5-240 μM), $\Delta\varepsilon_{262\ nm}$ = 7,960 $M^{-1}cm^{-1}$ for CF [32] (5-250 μM) , $\Delta\varepsilon_{260\ nm}$ = 7,900 $M^{-1}cm^{-1}$ for CZ [34] (5-250 μM), and $\Delta\varepsilon_{264\ nm}$ = 7,250 $M^{-1}cm^{-1}$ for CTX [35] (10-240 μM). Substrate hydrolysis was monitored according to initial steady-state velocities by using a Cary100 Bio UV-visible spectrophotometer (Agilent Technologies Canada, Montréal, QC) and quartz cuvettes with a path length of either 1 cm (CF, CZ, and CTX) or 10 cm (BZ and CB). Six substrate concentrations were used, generally flanking the K_M values for TEM-1. The enzyme concentrations were maintained within 5 to 30 nM. The kinetic parameters were determined from the hydrolysis rates calculated from the initial linear portion of the curve. Data fitting to the Henri Michaelis-Menten equation for the substrates CB and BZ was carried out with GraphPad Prism (GraphPad Software, San Diego, CA). For CF, CZ, and CTX a Lineweaver-Burk model (1/V *vs* 1/[S]) was used with Microsoft Excel due to inability to saturate the enzyme, a common issue with these substrates.

5.4. References

1. Zeil, C., et al., *Network Analysis of Sequence-Function Relationships and Exploration of Sequence Space of TEM beta-Lactamases.* Antimicrob Agents Chemother, 2016. **60**(5): p. 2709-2717.
2. Clouthier, C.M., et al., *Chimeric beta-lactamases: global conservation of parental function and fast time-scale dynamics with increased slow motions.* PloS one, 2012. **7**(12): p. e52283.
3. Papworth C, B.J., Braman J, Wright DA., *Site-directed mutagenesis in one day with >80% efficiency.* Strategies, 1996. **9**: p. 3-4.
4. Gobeil, S.M., et al., *Maintenance of native-like protein dynamics may not be required for engineering functional proteins.* Chemistry & biology, 2014. **21**(10): p. 1330-1340.

5. Gobeil, S.M., et al., *15N, 13C and 1H backbone resonance assignments of an artificially engineered TEM-1/PSE-4 class A beta-lactamase chimera and its deconvoluted mutant.* Biomol NMR Assign, 2016. **10**(1): p. 93-99.
6. Morin, S., et al., *Backbone resonance assignments of an artificially engineered TEM-1/PSE-4 Class A beta-lactamase chimera.* Biomol NMR Assign, 2010. **4**(2): p. 127-130.
7. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-234.
8. Simpson, R.J., *Zinc/Imidazole procedure for visualization of proteins in gels by negative staining.* CSH Protoc, 2007. **2007**(8): p. pdb prot4701.
9. Morin, S., R.C. Levesque, and S.M. Gagne, *1H, 13C, and 15N backbone resonance assignments for PSE-4, a 29.5 kDa class A beta-lactamase from Pseudomonas aeruginosa.* Journal of biomolecular NMR, 2006. **36 Suppl 1**: p. 11.
10. Swaren, P., et al., *X-ray structure of the Asn276Asp variant of the Escherichia coli TEM-1 beta-lactamase: direct observation of electrostatic modulation in resistance to inactivation by clavulanic acid.* Biochemistry, 1999. **38**(30): p. 9570-9576.
11. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Method Enzymol, 1997. **276**: p. 307-326.
12. Winter, G., C.M. Lobley, and S.M. Prince, *Decision making in xia2.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1260-1273.
13. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
14. Stec, B., et al., *Structure of the wild-type TEM-1 beta-lactamase at 1.55 A and the mutant enzyme Ser70Ala at 2.1 A suggest the mode of noncovalent catalysis for the mutant enzyme.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 8): p. 1072-1079.
15. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-2132.
16. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-221.

17. Vagin, A.A., et al., *REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use.* Acta Crystallogr D, 2004. **60**: p. 2184-2195.
18. Voss, N.R. and M. Gerstein, *3V: cavity, channel and cleft volume calculator and extractor.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W555-562.
19. (MOE), M.O.E., *2016*. Chemical Computing Group Inc.
20. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
21. Sondergaard, C.R., et al., *Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values.* J Chem Theory Comput, 2011. **7**(7): p. 2284-2295.
22. Olsson, M.H., et al., *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* J Chem Theory Comput, 2011. **7**(2): p. 525-537.
23. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-1958.
24. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1-2**: p. 19-25.
25. Schrödinger, L., *The PyMOL Molecular Graphics System*.
26. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.* J Am Chem Soc, 1982. **104**(17): p. 4546-4559.
27. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results.* J Am Chem Soc, 1982. **104**(17): p. 4559-4570.
28. John Cavanagh, W.J.F., Arthur G. Palmer III, Mark Rance and Nicholas J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*. Second Edition ed. 2007, (Burlington, MA: Elsevier Academic Press).
29. Loria, J.P., M. Rance, and A.G. Palmer, 3rd, *A TROSY CPMG sequence for characterizing chemical exchange in large proteins.* Journal of biomolecular NMR, 1999. **15**(2): p. 151-155.

30. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-293.
31. Manley, G. and J.P. Loria, *NMR insights into protein allostery.* Arch Biochem Biophys, 2012. **519**(2): p. 223-231.
32. Bouthors, A.T., et al., *Role of residues 104, 164, 166, 238 and 240 in the substrate profile of PER-1 beta-lactamase hydrolysing third-generation cephalosporins.* Biochemical Journal, 1998. **330**: p. 1443-1449.
33. Lim, D., et al., *Insights into the molecular basis for the carbenicillinase activity of PSE-4 beta-lactamase from crystallographic and kinetic studies.* Biochemistry, 2001. **40**(2): p. 395-402.
34. Tribuddharat, C., et al., *Burkholderia pseudomallei class A beta-lactamase mutations that confer selective resistance against ceftazidime or clavulanic acid inhibition.* Antimicrob Agents Ch, 2003. **47**(7): p. 2082-2087.
35. Cantu, C. and T. Palzkill, *The role of residue 238 of TEM-1 beta-lactamase in the hydrolysis of extended-spectrum antibiotics.* Journal of Biological Chemistry, 1998. **273**(41): p. 26603-26609.

Annex 6 - Revealing the hidden information about energy barriers of gas in structural ensemble

6.1. Tables

Table A 6-1 $\Delta G_{s,gas}$ free energy of solvation.
Comparison of $\Delta G_{s,gas}$ free energy of solvation of O_2, CO and N_2 calculated using Henry's law and ILS.

Gas	K_h[1]	$\Delta G_{s,gas,Henry}$ [kJ/mol]	$\Delta G_{s,gas,ILS}$ [kJ/mol]
O_2	0.032	8.58	8.99±0.14
CO	0.015	10.48	11.12±0.64
N_2	0.023	9.41	9.74±0.19

[1]Values from [1]

Table A 6-2 Number of exit events per channel and egress time of O_2 gas probe in free MD simulation.
Results are rank-ordered according to egress time.

# of exits	Egress time [ns]								
	1a	2a	2b	2d	2f	3	5	W	S
1	10	6.2		5.2	4.0		25.5		1.0
2		10.7		30	26.1		36.9		4.9
3		17.8							5.8
4		19							7.1
5									17.6

Table A 6-3 S^2 values of CYP102A1.

S^2 values and standard deviation for CYP102A1 calculated using the N-H vector autocorrelation function over ten 20 ns simulations

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
1	THR			36	LEU	0.85	0.02
2	ILE	0.66	0.08	37	GLY	0.82	0.03
3	LYS	0.82	0.02	38	GLU	0.79	0.04
4	GLU	0.77	0.03	39	ILE	0.90	0.01
5	MET	0.72	0.02	40	PHE	0.90	0.01
6	PRO			41	LYS	0.91	0.00
7	GLN	0.88	0.01	42	PHE	0.90	0.01
8	PRO			43	GLU	0.88	0.01
9	LYS	0.87	0.01	44	ALA	0.89	0.01
10	THR	0.84	0.01	45	PRO		
11	PHE	0.76	0.02	46	GLY	0.76	0.03
12	GLY	0.82	0.01	47	ARG	0.82	0.04
13	GLU	0.79	0.01	48	VAL	0.83	0.04
14	LEU	0.65	0.02	49	THR	0.89	0.01
15	LYS	0.87	0.01	50	ARG	0.87	0.02
16	ASN	0.85	0.01	51	TYR	0.88	0.02
17	LEU	0.89	0.01	52	LEU	0.88	0.01
18	PRO			53	SER	0.91	0.01
19	LEU	0.87	0.01	54	SER	0.81	0.03
20	LEU	0.87	0.02	55	GLN	0.92	0.01
21	ASN	0.79	0.03	56	ARG	0.92	0.01
22	THR	0.62	0.04	57	LEU	0.90	0.01
23	ASP	0.79	0.03	58	ILE	0.91	0.01
24	LYS	0.84	0.01	59	LYS	0.93	0.00
25	PRO			60	GLU	0.87	0.01
26	VAL	0.89	0.01	61	ALA	0.88	0.01
27	GLN	0.88	0.01	62	CYS	0.89	0.00
28	ALA	0.85	0.02	63	ASP	0.71	0.04
29	LEU	0.89	0.01	64	GLU	0.81	0.02
30	MET	0.91	0.01	65	SER	0.86	0.02
31	LYS	0.87	0.01	66	ARG	0.79	0.02
32	ILE	0.87	0.01	67	PHE	0.90	0.01
33	ALA	0.91	0.00	68	ASP	0.90	0.01
34	ASP	0.83	0.02	69	LYS	0.91	0.01

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
35	GLU	0.78	0.02	70	ASN	0.89	0.01
71	LEU	0.86	0.01	111	ALA	0.80	0.03
72	SER	0.86	0.02	112	MET	0.87	0.03
73	GLN	0.88	0.01	113	LYS	0.85	0.03
74	ALA	0.89	0.01	114	GLY	0.86	0.03
75	LEU	0.89	0.01	115	TYR	0.78	0.04
76	LYS	0.91	0.00	116	HIS	0.88	0.01
77	PHE	0.86	0.01	117	ALA	0.87	0.02
78	VAL	0.87	0.01	118	MET	0.81	0.02
79	ARG	0.92	0.00	119	MET	0.79	0.07
80	ASP	0.89	0.00	120	VAL	0.81	0.06
81	PHE	0.89	0.01	121	ASP	0.84	0.03
82	ALA	0.86	0.01	122	ILE	0.86	0.02
83	GLY	0.90	0.01	123	ALA	0.89	0.01
84	ASP	0.88	0.01	124	VAL	0.90	0.01
85	GLY	0.90	0.01	125	GLN	0.89	0.01
86	LEU	0.92	0.01	126	LEU	0.91	0.01
87	PHE	0.89	0.02	127	VAL	0.92	0.00
88	THR	0.86	0.03	128	GLN	0.91	0.01
89	SER	0.84	0.02	129	LYS	0.85	0.02
90	TRP	0.92	0.01	130	TRP	0.90	0.01
91	THR	0.92	0.00	131	GLU	0.90	0.01
92	HIS	0.81	0.04	132	ARG	0.84	0.01
93	GLU	0.87	0.01	133	LEU	0.64	0.05
94	LYS	0.88	0.03	134	ASN	0.75	0.04
95	ASN	0.89	0.05	135	ALA	0.78	0.03
96	TRP	0.87	0.04	136	ASP	0.71	0.04
97	LYS	0.90	0.02	137	GLU	0.69	0.04
98	LYS	0.89	0.02	138	HIS	0.76	0.04
99	ALA	0.85	0.03	139	ILE	0.86	0.02
100	HIS	0.92	0.01	140	GLU	0.82	0.04
101	ASN	0.85	0.04	141	VAL	0.79	0.05
102	ILE	0.79	0.02	142	PRO		
103	LEU	0.88	0.01	143	GLU	0.81	0.03
104	LEU	0.88	0.01	144	ASP	0.89	0.03
105	PRO			145	MET	0.92	0.01
106	SER	0.78	0.03	146	THR	0.88	0.01

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
107	PHE	0.84	0.02	147	ARG	0.91	0.01
108	SER	0.80	0.05	148	LEU	0.88	0.01
109	GLN	0.86	0.05	149	THR	0.92	0.01
110	GLN	0.78	0.05	150	LEU	0.91	0.01
151	ASP	0.90	0.01	191	ALA	0.73	0.03
152	THR	0.90	0.01	192	ASN	0.72	0.05
153	ILE	0.93	0.00	193	PRO		
154	GLY	0.92	0.00	194	ASP	0.69	0.05
155	LEU	0.91	0.01	195	ASP	0.79	0.02
156	CYS	0.90	0.01	196	PRO		
157	GLY	0.91	0.00	197	ALA		0.03
158	PHE	0.92	0.00	198	TYR	0.70	0.03
159	ASN	0.89	0.01	199	ASP	0.84	0.02
160	TYR	0.79	0.02	200	GLU	0.87	0.02
161	ARG	0.84	0.02	201	ASN	0.85	0.01
162	PHE	0.84	0.02	202	LYS	0.87	0.01
163	ASN	0.90	0.01	203	ARG	0.89	0.01
164	SER	0.92	0.01	204	GLN	0.86	0.02
165	PHE	0.87	0.01	205	PHE	0.90	0.01
166	TYR	0.82	0.03	206	GLN	0.90	0.01
167	ARG	0.58	0.06	207	GLU	0.82	0.04
168	ASP	0.71	0.04	208	ASP	0.87	0.02
169	GLN	0.67	0.04	209	ILE	0.91	0.01
170	PRO			210	LYS	0.86	0.03
171	HIS	0.80	0.04	211	VAL	0.84	0.06
172	PRO			212	MET	0.88	0.02
173	PHE	0.90	0.01	213	ASN	0.80	0.07
174	ILE	0.90	0.01	214	ASP	0.84	0.04
175	THR	0.90	0.01	215	LEU	0.75	0.04
176	SER	0.88	0.01	216	VAL	0.81	0.05
177	MET	0.91	0.01	217	ASP	0.84	0.04
178	VAL	0.92	0.00	218	LYS	0.85	0.03
179	ARG	0.90	0.00	219	ILE	0.86	0.02
180	ALA	0.90	0.00	220	ILE	0.89	0.01
181	LEU	0.91	0.00	221	ALA	0.88	0.01
182	ASP	0.88	0.01	222	ASP	0.83	0.02
183	GLU	0.90	0.00	223	ARG	0.86	0.04

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
184	ALA	0.90	0.01	224	LYS	0.82	0.03
185	MET	0.88	0.01	225	ALA	0.74	0.04
186	ASN	0.86	0.01	226	SER		
187	LYS	0.89	0.02	227	GLY		
188	LEU	0.87	0.01	228	GLU		
189	GLN	0.82	0.02	229	GLN		
190	ARG	0.64	0.04	230	SER	0.71	0.17
231	ASP	0.79	0.04	271	GLY	0.92	0.00
232	ASP	0.79	0.04	272	LEU	0.92	0.00
233	LEU	0.93	0.00	273	LEU	0.93	0.00
234	LEU	0.92	0.00	274	SER	0.91	0.00
235	THR	0.92	0.00	275	PHE	0.92	0.00
236	HIS	0.91	0.00	276	ALA	0.93	0.00
237	MET	0.82	0.06	277	LEU	0.91	0.01
238	LEU	0.89	0.01	278	TYR	0.91	0.00
239	ASN	0.80	0.03	279	PHE	0.92	0.00
240	GLY			280	LEU	0.90	0.00
241	LYS	0.78	0.02	281	VAL	0.91	0.00
242	ASP	0.86	0.05	282	LYS	0.88	0.01
243	PRO			283	ASN	0.88	0.00
244	GLU	0.73	0.04	284	PRO		
245	THR	0.71	0.03	285	HIS	0.89	0.00
246	GLY	0.83	0.01	286	VAL	0.85	0.01
247	GLU	0.75	0.03	287	LEU	0.91	0.00
248	PRO			288	GLN	0.91	0.00
249	LEU	0.87	0.02	289	LYS	0.84	0.02
250	ASP	0.82	0.02	290	ALA	0.90	0.00
251	ASP	0.87	0.02	291	ALA	0.92	0.00
252	GLU	0.85	0.02	292	GLU	0.89	0.01
253	ASN	0.80	0.03	293	GLU	0.88	0.01
254	ILE	0.91	0.02	294	ALA	0.91	0.01
255	ARG	0.92	0.01	295	ALA	0.81	0.05
256	TYR	0.84	0.02	296	ARG	0.82	0.01
257	GLN	0.90	0.00	297	VAL	0.86	0.03
258	ILE	0.92	0.01	298	LEU	0.88	0.01
259	ILE	0.87	0.03	299	VAL		
260	THR	0.91	0.01	300	ASP	0.74	0.04

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
261	PHE	0.91	0.01	301	PRO		
262	LEU	0.88	0.02	302	VAL		
263	ILE	0.87	0.03	303	PRO		
264	ALA	0.88	0.01	304	SER	0.77	0.03
265	GLY	0.86	0.01	305	TYR	0.79	0.04
266	HIS	0.84	0.05	306	LYS	0.78	0.06
267	GLU	0.85	0.03	307	GLN		
268	THR	0.91	0.01	308	VAL	0.86	0.05
269	THR	0.88	0.02	309	LYS	0.89	0.03
270	SER	0.91	0.01	310	GLN	0.82	0.05
311	LEU	0.87	0.01	351	ASP	0.82	0.03
312	LYS			352	GLU	0.85	0.02
313	TYR	0.89	0.01	353	LEU	0.90	0.01
314	VAL	0.93	0.01	354	MET	0.89	0.02
315	GLY	0.87	0.02	355	VAL	0.88	0.01
316	MET	0.90	0.01	356	LEU	0.90	0.01
317	VAL	0.92	0.00	357	ILE	0.88	0.02
318	LEU	0.91	0.01	358	PRO		
319	ASN	0.91	0.01	359	GLN	0.89	0.01
320	GLU	0.92	0.01	360	LEU	0.91	0.01
321	ALA	0.93	0.00	361	HIS	0.93	0.00
322	LEU	0.92	0.00	362	ARG	0.91	0.01
323	ARG	0.89	0.01	363	ASP	0.88	0.01
324	LEU	0.90	0.01	364	LYS	0.86	0.02
325	TRP	0.92	0.00	365	THR	0.85	0.02
326	PRO			366	ILE	0.85	0.02
327	THR	0.90	0.00	367	TRP	0.85	0.03
328	ALA	0.90	0.01	368	GLY	0.79	0.02
329	PRO			369	ASP	0.73	0.04
330	ALA	0.89	0.00	370	ASP	0.74	0.03
331	PHE	0.86	0.01	371	VAL	0.87	0.01
332	SER	0.88	0.02	372	GLU	0.90	0.01
333	LEU	0.91	0.01	373	GLU	0.86	0.01
334	TYR	0.93	0.00	374	PHE	0.85	0.01
335	ALA	0.90	0.01	375	ARG	0.91	0.00
336	LYS	0.90	0.01	376	PRO		
337	GLU	0.85	0.02	377	GLU	0.91	0.00

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
338	ASP	0.86	0.01	378	ARG	0.91	0.00
339	THR	0.71	0.06	379	PHE	0.90	0.00
340	VAL	0.82	0.01	380	GLU	0.87	0.01
341	LEU	0.87	0.02	381	ASN	0.84	0.02
342	GLY	0.90	0.02	382	PRO		
343	GLY	0.87	0.01	383	SER	0.81	0.04
344	GLU	0.83	0.03	384	ALA	0.78	0.04
345	TYR	0.87	0.01	385	ILE	0.67	0.04
346	PRO			386	PRO		
347	LEU	0.86	0.01	387	GLN	0.85	0.01
348	GLU	0.83	0.03	388	HIS	0.86	0.01
349	LYS	0.88	0.02	389	ALA	0.82	0.01
350	GLY	0.88	0.02	390	PHE	0.89	0.01
391	LYS	0.89	0.00	431	LEU		
392	PRO			432	ASP	0.74	0.06
393	PHE	0.89	0.00	433	ILE	0.87	0.02
394	GLY	0.86	0.01	434	LYS	0.89	0.01
395	ASN	0.89	0.01	435	GLU	0.90	0.01
396	GLY	0.88	0.01	436	THR	0.86	0.02
397	GLN	0.88	0.01	437	LEU	0.82	0.04
398	ARG	0.88	0.02	438	THR	0.86	0.03
399	ALA	0.88	0.01	439	LEU	0.89	0.02
400	CYS	0.90	0.00	440	LYS	0.92	0.01
401	ILE	0.86	0.03	441	PRO		
402	GLY	0.89	0.02	442	GLU	0.87	0.02
403	GLN	0.89	0.02	443	GLY	0.79	0.03
404	GLN	0.87	0.04	444	PHE	0.79	0.02
405	PHE	0.86	0.04	445	VAL	0.85	0.03
406	ALA	0.91	0.01	446	VAL	0.90	0.01
407	LEU	0.87	0.03	447	LYS	0.90	0.01
408	HIS	0.74	0.04	448	ALA	0.85	0.02
409	GLU	0.87	0.04	449	LYS	0.79	0.06
410	ALA	0.91	0.01	450	SER	0.82	0.01
411	THR	0.89	0.02	451	LYS	0.86	0.02
412	LEU	0.90	0.01	452	LYS	0.86	0.01
413	VAL	0.89	0.02	453	ILE	0.84	0.02
414	LEU	0.91	0.01	454	PRO		

Residue #	aa	S^2	ΔS^2	Residue #	aa	S^2	ΔS^2
415	GLY	0.92	0.00	455	LEU		
416	MET	0.92	0.00	456	GLY		
417	MET	0.90	0.01	457	GLY		
418	LEU	0.92	0.00				
419	LYS	0.89	0.01				
420	HIS	0.88	0.01				
421	PHE	0.91	0.01				
422	ASP	0.89	0.01				
423	PHE	0.81	0.03				
424	GLU	0.89	0.01				
425	ASP	0.87	0.02				
426	HIS	0.84	0.04				
427	THR						
428	ASN						
429	TYR						
430	GLU						

6.2. Figures

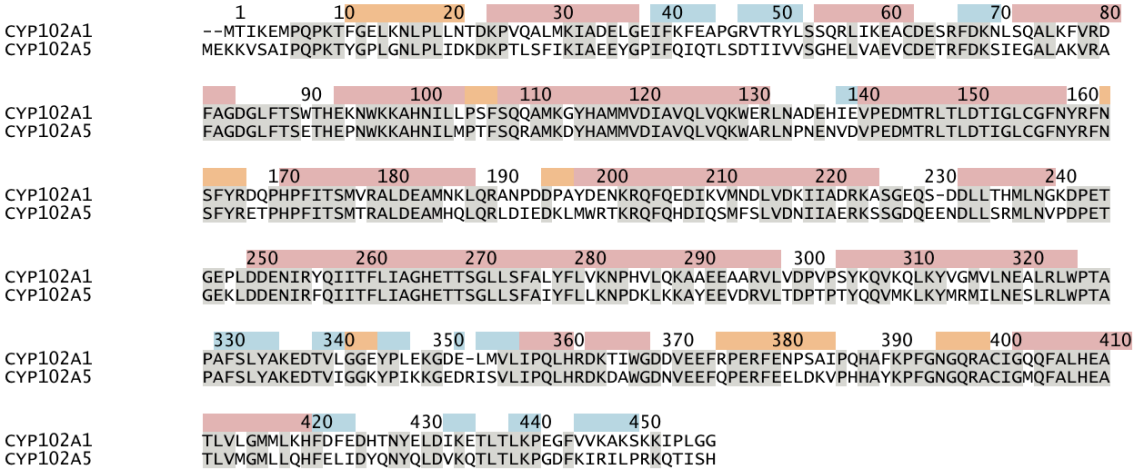


Figure A 6-1 Pairwise sequence alignment between CYPA102A1 and CYP102A5.
The residue numbering is based on the structure of CYP102A1 (PDB 2HPD). Secondary sequence elements are highlighted as follows: 3_{10}-helices in orange, α-helices in pink and β-sheets in blue. Conserved residues are shown in grey. Alignment was performed using MUSCLE[2].

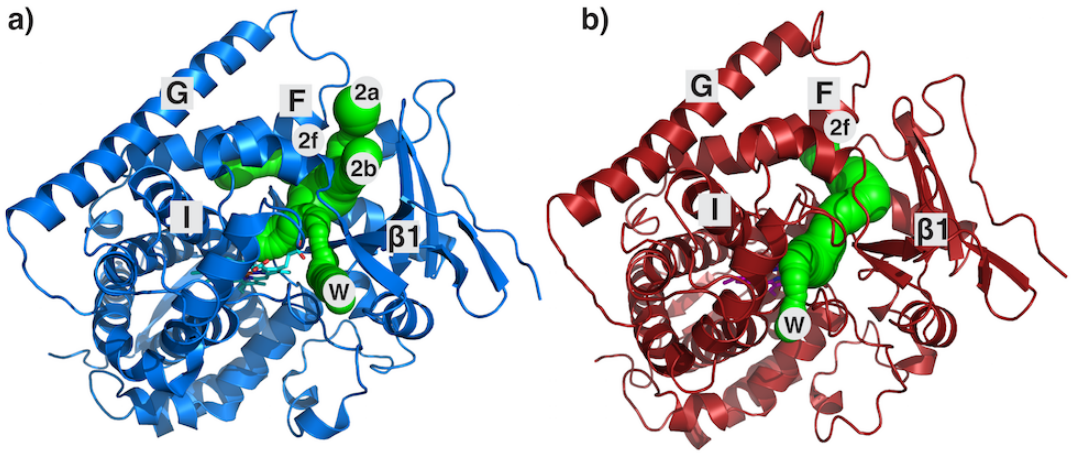


Figure A 6-2 CAVER3 channel search results for A) CYP102A1 (Protein: blue, Heme: cyan; PDB 2HPD) and B) the homology model of CYP102A5 (Protein: red, Heme: pink). *The channels are represented by a chain of spheres with a diameter correlated to the volume available at this position. For orientation, important helices and sheets are identified by square labels and the channels by round labels.*

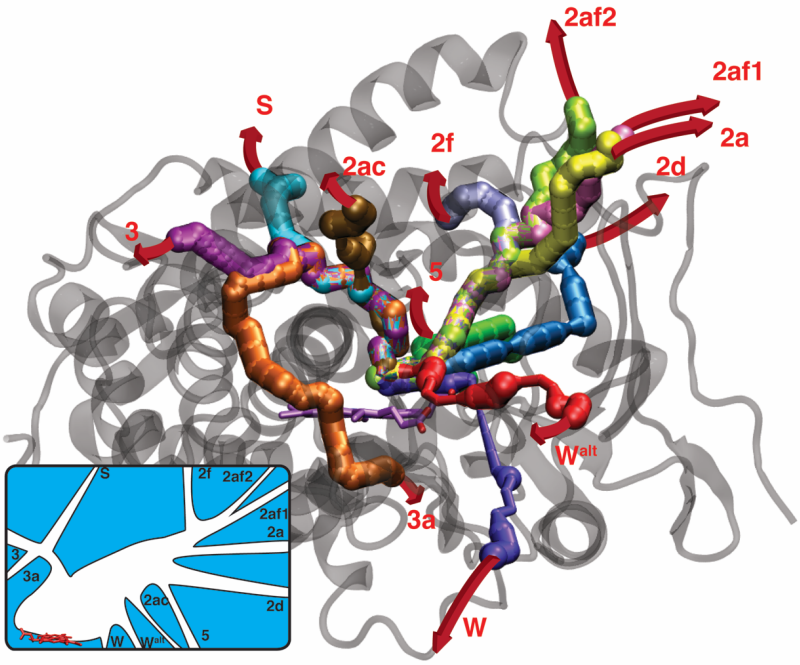


Figure A 6-3 Channels mapped by ILS in the homology model of CYP102A5. *The heme group is in sticks. The nomenclature of channels is based on Cojocaru et al.[3] with our addition of 3a, W^{alt}, 2af1 and 2af2. The diameter along each colored tube represents the free energy of O_2: the thinner the tube, the higher the energy barrier for oxygen.* ***Inset:*** *Layout of all channels identified in CYP102A5 using ILS. Drawing not to scale.*

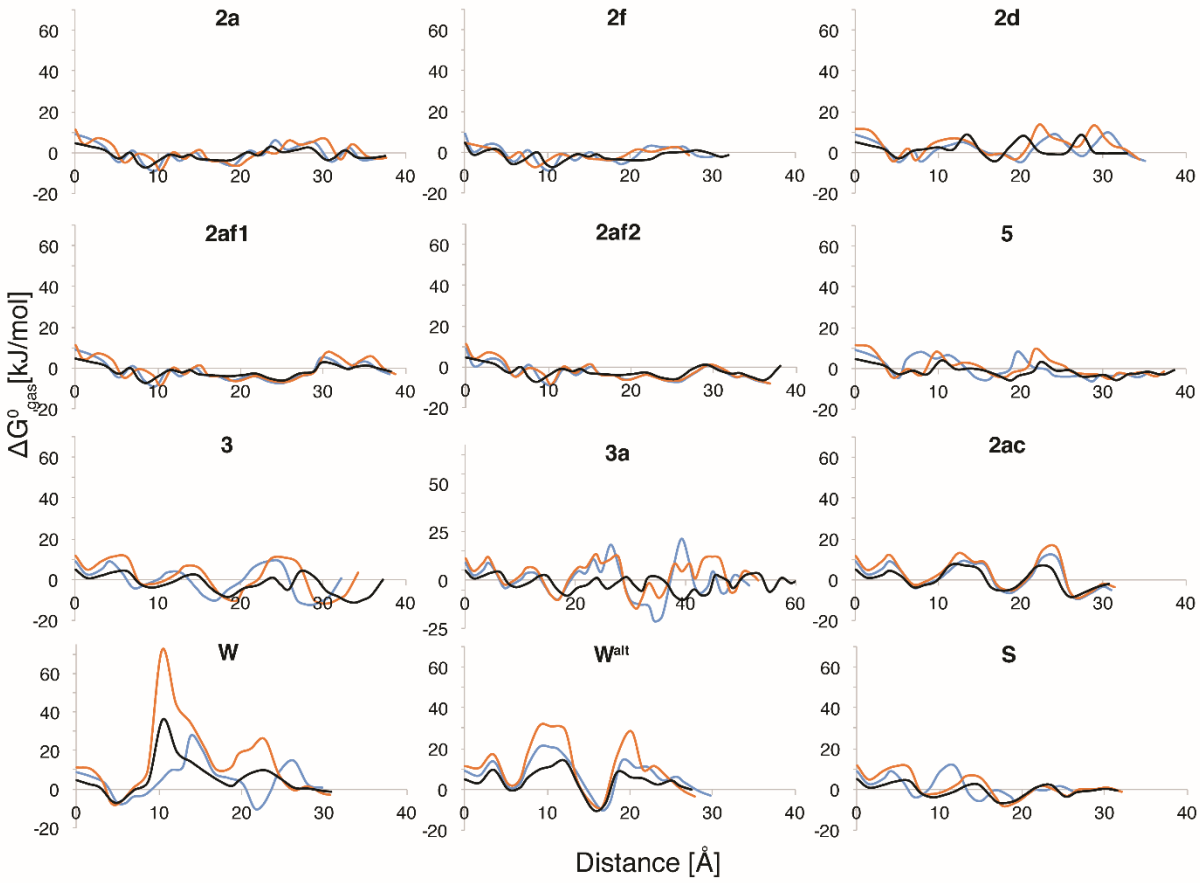


Figure A 6-4 Free energy profiles for tunnels in CYP102A5.
Free energy profiles for oxygen (black), carbon monoxide (orange) and nitrogen (blue) movement along the channels of CYP102A5. The distance is the optimal path length with the lowest sum of energies on each grid point excluding grid points outside the cut-off value.

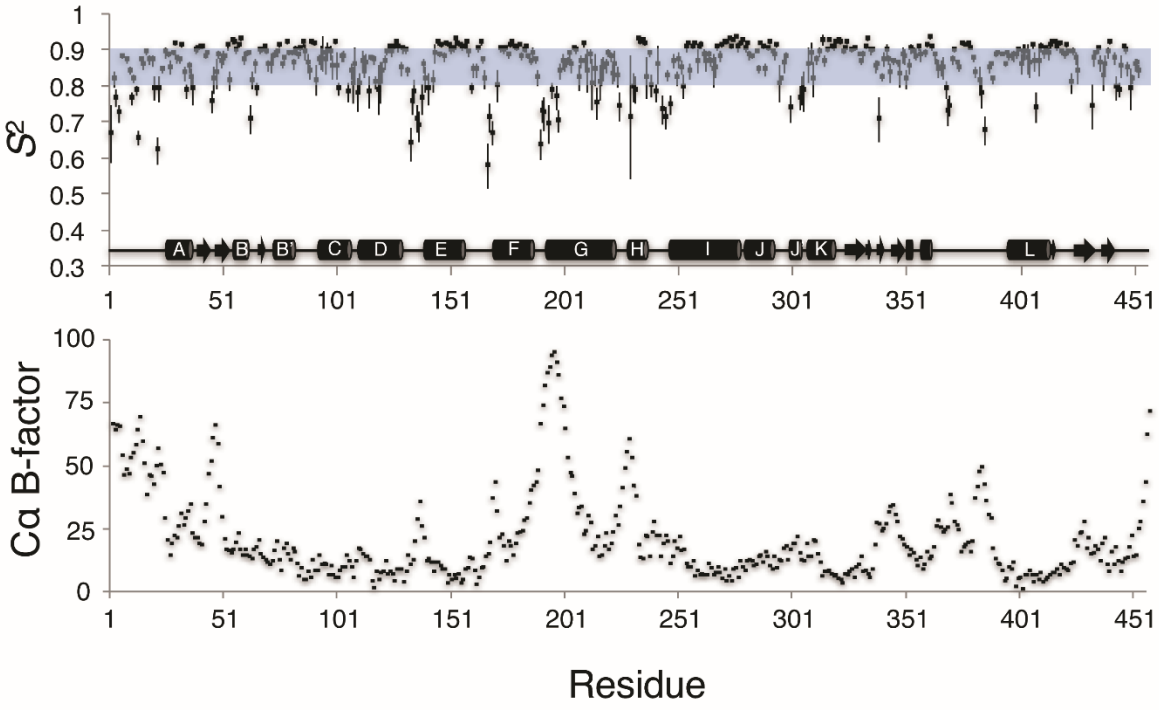


Figure A 6-5 Comparison of S^2 values and Cα b-factors for CYP102A1.
Upper panel: The calculated S^2 values of CYP102A1 according to the amino acid sequence and the secondary structure elements. The blue stripe is centered at the average S^2 and is two standard deviations wide. α-Helices are represented as cylinders and follow the nomenclature of Hasemann et al., whereas the β-sheets are depicted as arrows [4]. Lower panel: Cα b-factors for CYP102A1 based on PDB 2HPD.

6.3. Videos

Video A 6-1 O_2 diffusion in CYP102A1.
Direct O_2 egress by diffusion over the course of a 40 ns free MD simulation. The O_2 molecule migrated solely through the 2d channel.

Video A 6-2 O_2 hopping between egress channels.
O_2 hopping shown by diffusion over the course of a 40 ns free MD simulation. The O_2 molecule first entered channel 2f, then crossed over to channel 2b and finally to channel 2a. After a retention time of 5 ns in 2a, the molecule exited CYP102A1 using channel 2a. Residues Q189 and R47 are shown in sticks and are main contributors to the formation of the transient subchannels.

6.4. References

1. Sander, R., *Compilation of Henry's law constants (version 4.0) for water as solvent.* Atmos Chem Phys, 2015. **15**(8): p. 4399-4981.
2. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-1797.
3. Cojocaru, V., P.J. Winn, and R.C. Wade, *The ins and outs of cytochrome P450s.* Biochim Biophys Acta, 2007. **1770**(3): p. 390-401.
4. Hasemann, C.A., et al., *Structure and function of cytochromes P450: a comparative analysis of three crystal structures.* Structure, 1995. **3**(1): p. 41-62.