

Towards in silico prediction of mutations related to antibiotic resistance

Eddy Elisée

► To cite this version:

Eddy Elisée. Towards in silico prediction of mutations related to antibiotic resistance. Cheminformatics. Université Paris Saclay (COmUE), 2019. English. NNT: 2019SACLS350. tel-02963886

HAL Id: tel-02963886 https://theses.hal.science/tel-02963886

Submitted on 12 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Towards in silico prediction of mutations related to antibiotic resistance

Vers la prédiction in silico des mutations liées à la résistance aux antibiotiques

Thèse de doctorat de l'Université Paris-Saclay préparée à l'Université Paris-Sud

Ecole doctorale n°571 Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Spécialité de doctorat : Chimie

Thèse présentée et soutenue à Gif-sur-Yvette, le 11 octobre 2019, par

Eddy Elisée

Composition du Jury :

Laurent Chaloin Directeur de Recherche, IRIM - CNRS, Montpellier	Président Rapporteur
Oliver Beckstein Associate professor, Arizona State University, Etats-Unis	Rapporteur
Guillaume Bernadat Maitre de Conférence, Université Paris-Sud, Châtenay-Malabry	Examinateur
Bogdan lorga Directeur de Recherche, ICSN - CNRS, Gif-sur-Yvette	Directeur de thèse

Acknowledgements _

"This is the end, hold your breath and count to ten..." are the first lyrics of the famous *Skyfall* movie soundtrack sung by Adèle, and as far as I am concerned, this should be a stress-relieving motto for every student defending their PhD work. Indeed, my three-year thesis comes to an end with the writing of this manuscript, and many people have contributed to make this journey an extraordinary experience.

First, I would like to sincerely thank my supervisor, Dr. Bogdan Iorga, who trusted me since the beginning of my Master 2 degree internship, and with whom I learnt much more than I expected. I also thank you for opportunities of attending national and international workshops and conferences in which I met interesting, and sometimes renowned, scientists. Lastly, thanks for being human and not a selfish tyrant; it helps to work peacefully and keep a positive mood!

Then, I would like to acknowledge Dr. Laurent Chaloin and Pr. Oliver Beckstein for having accepted to review my doctoral manuscript, and Dr. Guillaume Bernadat for having followed my work during the last three years and accepted to be a member of my defence jury.

As research can't be achieved without money, I would like to thank the doctoral school "Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)" for having funding my PhD project.

A special mention to Guillaume Loire and David Lê, the two brilliant computer scientists and guardians of the network infrastructure from Information Technology (IT) department, who saved our data several times!

Furthermore, it is important for me to thank two particular persons who convinced me, through their passion and pedagogy, to choose one path over another during my Bachelor's degree: I am referring to Pr. Anne Zaparucha who transmitted me the taste of Organic Chemistry and Pr. Véronica Beswick who taught me the basics of Computational Chemistry (force fields, molecular dynamics...) and aroused my curiosity on the notion of Chemo-Informatics. I can't imagine working alone with computers but fortunately I had the chance to meet great colleagues with whom I spent delicious moments drinking coffee or eating specialty foods from abroad. They also brought me a valuable help when I was, inter alia, fighting against scripting languages or looking for some computational tips, and it was really cool to collaborate for D3R competitions, we did a good job! This way, I would like to deeply thank Edithe, Nawel, Agustin, Ludo, Valentin, Hristo and Pascal.

Special thanks to you, Nawel, since you genuinely enlivened my eventful third year with your generosity, your warm personality and our uncontrollable fits of laughter, thank you equally for having withstand my hyperactive, sometimes crazy I concede, personality through some jukebox abilities and mewings.

I would like to thank some exceptional members of my students' committee (a.k.a CEI 2017) who contribute to make the ICSN a pleasant workplace through coffee breaks, dinners and boisterous parties; long live zouk! Therefore, Julie, Gwenaëlle, Morgane, Laure (my mate from southern corridor and my official snack provider), Simon and Mathieu, thank you very much!

I am also glad to have met amazing persons at the ICSN, namely Orianne, Pauline, Agathe, Céline the outstanding dancer, and Capucine the dazzling one-woman show.

In addition, I would like to acknowledge my dear friends who I met at different moment of my life, either in middle school, high school, university, student job, internships or elsewhere. Nevertheless, I have to write some special thanks to those who kindly took care of me during the last three years, in the good as well as the hard times: Céline, Sophie, Claire, Amélie, Aurélie, Marion, Morgane, Eva, Kalani, Emilie, Sergi, Robin, Marc and my best friends Marine and Raphaël, thank you for the genuineness of moments we have spent together.

Finally, I would like to thank my beloved family for their unwavering support, and especially, Maman for your unconditional love and for having let me free to choose my path, Earvin, my twin brother, for our great and full complicity, Mémé for being so generous toward your grandkids, and Pépé who told me in final stages of your life to stay the course since you were convinced my kind personality could take me far away; it appears that I made it, Pépé!

Thesis structure ____

This manuscript should drive you to the initial aim for which this PhD subject was developed, that is *in silico* prediction of β -lactamase mutations related to antibiotic resistance. Otherwise, in the Molecular Modeling and Structural Crystallography team at the Institut de Chimie des Substances Naturelles on the CNRS campus in Gif-sur-Yvette, we work on β -lactam antibiotic resistance. Therefore, a bibliographic chapter first outlines the antibiotics' history and reasons that led to a resistance crisis before introducing the actors of β -lactam antibiotic resistance and new strategies designed to fight against it.

Then, the main part of the manuscript includes publication-based chapters which all respect a consistent thread, namely the development and assessment of a computational workflow for calculating free energies of mutations in metal-free and metal-containing protein systems. First, modifications of OPLS-AA (Optimized Potentials for Liquid Simulations - All Atoms) force field to study zinc-containing metalloenzymes is presented. Secondly, several ligand-free proteins and protein-ligand complexes are studied with molecular dynamics simulations using the above-mentioned modified force field. Thirdly, an in-house pmx-based protocol for free energy calculations is assessed through participations in international blinded prediction challenges.

Finally, a classical "Conclusions" section closes this manuscript by summarizing achieved works and presenting perspectives regarding the prediction of β -lactamase mutations.

Besides the Bibliography section of this manuscript, to find out more about antibiotic resistance, I turn your attention to some inspiring websites, reviews and books listed below:

– L. C. Ventola, The Antibiotic Resistance Crisis. Part 1: Causes and Threats. *Pharm. Ther.*, 40(4):278–283, 2015.

– L. C. Ventola, The Antibiotic Resistance Crisis. Part 2: Management Strategies and New Agents. *Pharm. Ther.*, 40(5):344–352, 2015.

- WHO website. https://www.who.int/antimicrobial-resistance/

– K. Kon, M. Rai, Antibiotic Resistance: Mechanisms and New Antimicrobial Approaches. *Academic Press*, 2016.

Introduction

En seulement quelques années, la résistance aux antibiotiques est devenu un problème majeur de santé publique au niveau mondial du fait de l'émergence de souches bactériennes résistantes voire multi-résistantes. Certaines d'entre elles (*M. tuberculosis XDR ou PDR, A. baumannii, P. aeruginosa, Enterobacteriaceae...*) sont d'ores et déjà apparues comme résistantes à tout l'arsenal d'antibiotiques présent sur le marché incluant les carbapénèmes, principalement utilisés lors des traitements de dernier recours. De plus, bien que ce phénomène de résistance apparaisse naturellement, l'utilisation abusive d'antibiotiques dans les milieux médicaux et agricoles accélère ce processus, laissant présager l'éventualité d'un basculement dans une ère dite "post-antibiotique" dans laquelle les infections microbiennes ne pourront plus être traitées efficacement via des traitements antibiotiques.

De leur côté, les bactéries peuvent acquérir une résistance aux antibiotiques de trois manières différentes : (a) le transfert vertical, dans lequel la résistance est transmise d'une bactérie résistante à sa descendance, (b) le transfert horizontal, dans lequel la résistance est transmise à une bactérie depuis une bactérie morte (transformation), un bactériophage (transduction), ou une bactérie résistante (conjugaison), et (c) une mutation ponctuelle localisée sur le chromosome bactérien. Par ailleurs, chez les bactéries résistantes, on observe quatre mécanismes d'action pour contrer la présence d'antibiotiques dans leur environnement : la sur-expression des pompes à efflux, la diminution de la perméabilité membranaire, la modification de la cible visée par les antibiotiques et la production d'enzymes de dégradation.

Au laboratoire, on s'intéresse aux antibiotiques de type β -lactames ainsi qu'aux enzymes de dégradation, nommées β -lactamases, spécifiquement développées pour hydrolyser ces derniers.

Il existe quatre familles de β -lactames (pénicillines, céphalosporines, monobactames et carbapénèmes) ainsi que quatre grandes classes de β lactamases (A – pénicillinases, B – métalloenzymes, C – céphalosporinases et D – oxacillinases). Initialement, ces dernières étaient spécifiques d'une ou deux classes d'antibiotiques, mais de récentes enzymes présentent un spectre d'hydrolyse beaucoup plus étendu englobant la plupart voire la totalité des quatre classes de β -lactames. Les quatre classes de β -lactamases peuvent être réparties en deux familles selon la constitution de leur site actif : les sérine- β -lactamases (classes A, C et D) possèdent une sérine impliquée dans l'hydrolyse des β -lactames, alors que chez les metallo- β -lactamases (classe B), ce mécanisme est assuré par un ou deux ions zinc.

Objectifs de recherche

Cette thèse s'inscrit dans le contexte de la résistance aux antibiotiques de type β -lactames avec un double objectif : (1) étudier, par dynamique moléculaire, le comportement des β -lactamases en présence ou en absence de ligands, et (2) prédire, par calcul d'énergie libre, de nouvelles mutations ponctuelles conférant une résistance aux β -lactamases.

Développement

– I. Introduction –

Premièrement, la résistance aux antibiotiques est présentée au travers (1) de l'évolution des méthodes de soins apportés aux malades depuis la Préhistoire jusqu'à aujourd'hui, (2) de son origine via l'analyse de plusieurs exemples tirés de la nature (animaux, environnement...) ou de l'activité humaine (agriculture, hôpitaux...) et (3) de sa propagation fortement médiée par l'environnement.

Deuxièmement, les acteurs de la résistance aux β -lactames sont détaillés, en commençant par les systèmes de défense développés par les Entérobactéries, telles les porines ou les β -lactamases, pour lutter contre la pression antibiotique. Ensuite, les β -lactames sont décrits de par leur structure et leur rôle d'inhibiteurs de la protéine liant la pénicilline et, enfin, les β -lactamases sont présentées de manière non exhaustive.

Troisièmement, certaines stratégies imaginées pour lutter contre la résistance aux antibiotiques sont brièvement listées, incluant les adjuvants chimiques ou bio-mimétiques, les outils biotechnologiques (bacteriophages, anticorps, nano-matériaux...) et computationnels (algorithme de prédiction, "machine learning"...).

- II. Nouveaux paramètres OPLS-AA pour les enzymes à zinc -

Dans ce projet, de nouvelles charges CM5 ont été développées pour certains résidus en interaction avec un ion zinc présent dans le site actif de certaines métalloenzymes, dont les métallo- β -lactamases. Ces charges ont ensuite été implémentées et validées dans le champ de force OPLS-AA selon des critères géométriques (distance zinc-atome coordonnant, RMSD du site actif et de la protéine) sur un panel de 17 métalloenzymes représentatives. Ainsi, les nouveaux résidus considérés sont : HME/HMD pour l'histidine Nɛ et Nô, CYSM pour la cystéine et ASM/GLM pour l'asparagine/glutamine.

– III. Etude des β -lactamases dans le champ de force OPLS-AA –

Ce chapitre présente quatre études, par dynamique moléculaire, de sérineet métallo- β -lactamases.

Premièrement, le paramétrage de ligands variés, incluant les antibiotiques β -lactames, dans le champ de force OPLS-AA est présenté via l'utilisation du programme MOL2FF développé au laboratoire. Certains antibiotiques ont ensuite été utilisés dans une étude portant sur la caractérisation d'une nouvelle sérine- β -lactamase de la classe C, CMY-136.

Deuxièmement, deux inhibiteurs de la métallo-β-lactamase IMP-1 ont été testés sur NDM-1 au travers de simulations de dynamique moléculaire pour évaluer la robustesse, en présence d'un ligand, des paramètres développés dans le Chapitre 2.

Troisièmement, la mutation d'un résidu histidine, coordonnant un des deux ions zinc catalytiques, en glycine a été étudiée chez la métallo- β -lactamase SPM-1. Nous avions émis l'hypothèse qu'un résidu aspartate situé non loin du site actif pouvait être amené à compléter la sphère de coordination de l'ion zinc. Bien que certaines simulations aient révélé un changement spontané de conformation validant notre hypothèse, une récente étude a démontré qu'une molécule d'eau prenait la place de l'histidine mutée, structure cristallographique à l'appui.

Enfin, de nombreux calculs d'énergie libre, utilisant le protocole pmx, ont été réalisés afin d'étudier les processus de carboxylation et protonation de la lysine catalytique dans les oxacillinases.

- IV. L'aventure D3R-GC -

Le projet D3R-GC offre l'opportunité d'évaluer et de valider les protocoles développés au laboratoire au travers de compétitions internationales de prédiction à l'aveugle basées sur de la prédiction de poses et des calculs d'énergie libre. Les résultats de trois de ces compétitions sont présentés dans ce chapitre.

Conclusion

Les travaux méthodologiques présentés dans ce manuscrit de thèse ont donc conduit au développement : (1) de nouveaux paramètres pour les enzymes à zinc, implémentés dans le champ de force OPLS-AA et validés par des simulations de dynamique moléculaire sur un panel de métalloenzymes représentatives, (2) d'un protocole de paramétrisation de ligands covalents pour étudier le comportement de certains β -lactames dans CMY-136, une nouvelle β -lactamase caractérisée au laboratoire, et (3) d'un protocole de calcul d'énergie libre évalué au moyen de compétitions internationales de prédiction. Ce dernier a ensuite été utilisé pour tenter d'expliquer pourquoi la carbamylation de la sérine catalytique n'a pas lieu dans certaines oxacillinases.

Au travers de ces travaux, nous avons pu améliorer significativement notre approche computationnelle et désormais tout est en place pour une exploration exhaustive des mutations possibles dans les β -lactamases.

_____ Table of Contents _____

List of Figures xv			
List of Ta	List of Tables xvii		
Abbrevia	tions	xx	
Chap. I	Introduction	1	
1 Antibio	otics: a leap into resistance	3	
1.1 Cha	inges in humans' attitude	3	
1.2 Fro	m first antibiotics to antibiotics' era	5	
1.2.1	Louis Pasteur and Jules F. Joubert	6	
1.2.2	Paul Ehrlich	6	
1.2.3	Alexander Fleming	6	
1.2.4	Ernst B. Chain, Howard W. Florey and Edward Abraham	7	
1.2.5	Dorothy M. C. Hodgkin	8	
1.2.6	Josef Klarer, Fritz Mietzsch and Gerhard Domagk	8	
1.2.7	Selman Waksman and Harold B. Woodruff	8	
1.3 Ant	ibiotic resistance	9	
1.3.1	Definition	10	
1.3.2	Where does the antibiotic resistance come from ?	10	
1.3.3	Environmental spread of antibiotic resistance	12	
1.4 Sur	nmary	14	
2 Actors	of β-lactam resistance	15	
2.1 Ent	erobacteriaceae	15	
2.1.1	General features	15	
2.1.2	Gram-negative bacteria features	16	
2.1.3	Peptidoglycan formation	17	
2.1.4	Penicillin-binding proteins	18	

2.2 β-lactam antibiotics		
2.2.1	Structure	
2.2.2	Mode of action 20	
2.2.3	Examples of chemical modifications 21	
2.3 β-la	ctamases	
2.3.1	Classification 22	
2.3.2	Enzymatic mechanisms 23	
2.4 Sun	nmary	
3 Strateg	gies against β -lactam resistance	
3.1 Ant	imicrobial adjuvants 27	
3.1.1	β-lactamase inhibitors 28	
3.1.2	Newly discovered antimicrobials 29	
3.2 Exp	erimental strategies 32	
3.2.1	Inteins and pro-drugs 32	
3.2.2	Bacteriophages 33	
3.2.3	Monoclonal antibodies 35	
3.2.4	Environmental aspects	
3.3 The	oretical strategies	
3.3.1	Mathematical model 36	
3.3.2	Algorithm	
3.3.3	Detection	
3.3.4	Prediction	
3.4 Sun	nmary	
Chap. II	New OPLS-AA force field parameters for zinc-containing	

1 Overview	41
2 Force field analysis	43
2.1 Emergence of force fields	43

39

enzymes

2.2 Fun	damental principles 45
2.2.1	The Born-Oppenheimer approximation
2.2.2	Functional form
2.3 OP	LS-AA force field
2.3.1	History
2.3.2	OPLS-AA functional form 49
2.4 Zn-	OPLS publication
Chap. III OPLS-AA	Molecular dynamics (MD) studies of β-lactamases in force field 89
1 Overvi	ew
2 Param	eterization of covalent ligands in OPLS-AA force field 93
2.1 MO	L2FF software
2.1.1	Philosophy of OPLS-AA parameterization
2.1.2	MOL2FF algorithm
2.1.3	LigParGen service
2.2 Cov	valent parameterization
2.2.1	Implementation in MOL2FF
2.2.2	Parameterized antibiotics
2.3 Pub	lication
3 MD sin	nulations of MBL–ligand complexes143
3.1 Sys	tem setup 143
3.1.1	Protein
3.1.2	Ligands144
3.1.3	MD simulations

4 MD simulations of an intriguing SPM-1 glycine mutant		
4.1 System setup	149	
4.1.1 Protein	149	
4.1.2 MD simulations	150	
4.2 Results and Discussion	150	
5 OXA β -lactamases studies using free energy calculations		
5.1 Alchemical PMX philosophy	154	
5.2 System setup	155	
5.2.1 Protein	155	
5.2.2 Free energy calculations	156	
5.3 Results and Discussion	157	
Chap. IV Blinded prediction challenges: the D3R-GC adventure 159		
1 Overview	161	
1.1 Drug Design Data Resource (D3R)	161	
1.2 Strategy	162	
1.3 What we have learned	162	
2 D3R publications	165	
2.1 D3R–GC2 publication	166	
2.2 D3R–GC3 publication	213	
2.3 D3R–GC4 publication		
Conclusions	303	
References	308	

_____ List of Figures _____

Figure 1	From the antibiotics' discovery to the emergence of antibi-	
	otic resistance: key dates	4
Figure 2	The antibiotic "wall of fame"	5
Figure 3	Main mechanisms and acquisition pathways involved in	
	antibiotic resistance in bacteria.	11
Figure 4	Existing antibiotic classes	15
Figure 5	The Gram-staining method	17
Figure 6	Peptidic connection between two peptidoglycan layer	18
Figure 7	Peptidoglycan formation and inhibition of the penicillin-	
	binding protein (PBP).	19
Figure 8	The four β -lactam classes	20
Figure 9	Three-dimensional structures of representative β -lactamases	
	from each class.	23
Figure 10	Mechanism of carbapenem hydrolysis achieved by serine- β -	
	lactamases (A) and metallo- β -lactamases (B)	24
Figure 11	Overview of existing β -lactamase inhibitors	29
Figure 12	Original developed antibiotics.	30
Figure 13	The intein mechanism.	33
Figure 14	β -lactamase as biological scissors to release a drug inside	
	the cell	34
Figure 15	Bacteriophages' lytic and lysogenic cycles	34
Figure 16	Antibodies' mechanisms of actions	35
Figure 17	Approaches for implementing metal-coordinating residues	
	into a force field	41
Figure 18	Atoms' charge adjustement in OPLS-AA force field	93
Figure 19	SMILES description.	95
Figure 20	MOL2FF algorithm workflow	96
Figure 21	LigParGen web interface.	97
Figure 22	Covalent β -lactam parameterization using MOL2FF software.	98
Figure 23	Parameterized β -lactam antibiotics	99
Figure 24	Crystallographic structure of NDM-1 (PDB code 4hl2)	143
Figure 25	Structures of NDM–1 and IMP–1 enzymes	144
Figure 26	RMSD analysis of apo-NDM-1	145
Figure 27	RMSD analysis of NDM-1–inhibitor complexes	146
Figure 28	SPM-1 and SPM-like sequences alignment	147
Figure 29	SPM-1 crystallographic structure.	148
Figure 30	Protonation and carboxylation reactions of the lysine	153

Figure 31	Thermodynamic cycle for estimating the difference in sta-
	bility between wild type and mutant
Figure 32	Crystallographic structures of OXA enzymes
Figure 33	System setup for a mutation that modifies the charge of the
	residue
Figure 34	Single mutations in KPC, NDM-1 and OXA families 307
Figure 35	Single mutations in CMY family

_____ List of Tables _____

Table 1	Examples of atom types in OPLS-AA force field 94
Table 2	Some features of SMARTS language
Table 3	Atom equivalence for SPM-1 and SPM-like structures 149
Table 4	PDB survey on some MBLs from B2 and B3 subclasses 151
Table 5	Free energy results on carboxylation and protonation reac-
	tions in OXA enzymes of interest
Table 6	Overview of D3R datasets

Abbreviations

A. D.	Anno Domini (nostri Jesu Christi) = in the year of our Lord Jesus Christ
ADC	Acinetobacter derived cephalosporinase
AIDS	Acquired immune deficiency syndrome
AIM	Adelaide imipenemase
AMBER	Assisted model building and energy refinement
AMOEBA	Atomic multipole optimized energetics for biomolecular simulation
AMP	Antimicrobial peptide
AMR	Antimicrobial resistance
ARBs	Antibiotic-resistant bacteria
ARGs	Antibiotic resistance genes
ASCII	American standard code for information interchange
BAR	Bennett acceptance ratio
B. C.	Before Christ
BcII	From Bacillus cereus type II
BO	Born-Oppenheimer
Cas	CRISPR associated
CG	Coarse-grained
CGI	Crooks-gaussian intersection
CHARMM	Chemistry at Harvard molecular mechanics
CM1A	Charge model 1 A
CM5	Charge model 5
СМУ	Active on cephamycins
СОМ	Center of mass
СРВ	Carbapenemase-producing bacteria

CphA	Carbapenem hydrolyzing and first (A) from <i>Aeromonas hydrophila</i>
CRISPR	Clustered regularly interspaced short palindromic repeats
CTX-M	Active on cefotaxime, first isolated at Munich
D3R-GC	Drug design data resource - grand challenge
DNA	Deoxyribonucleic acid
DT50	Disappearance time 50
EDTA	Ethylenediaminetetraacetic acid
e.g.	exempli gratia = for example
ESBL	Extended spectrum beta-lactamase
GOB	From Chryseobacterium meningosepticum class B
GROMOS	Groningen molecular simulation
IC50	Inhibitory concentration 50
i.e.	id est = in other words
IONPs	Iron oxide nanoparticles
КРС	Klebsiella pneumoniae carbapenemase
L1	Labile enzyme from Stenotrophomonas (Pseudomonas, Xanthomonas) maltophilia
LBCC	Localized bond-charge corrected
LPS	Lipopolysaccharides
MAC	Membrane Attack Complex
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MBL	Metallo-β-lactamase
MD	Molecular dynamics
MDR	Multi-drug resistant
MM	Molecular mechanics
MMFF	Merck molecular force field
NAG / GlcNAc	N-acetylglucosamine

NAM / MurNAc	N-acetylmuramic acid
NDM	New Delhi metallo-β-lactamase
OMPs	Outer membrane proteins
OPLS-AA	Optimized potentials for liquid simulations - all atoms
OXA	Oxacillinases
PCR	Polymerase chain reaction
QM	Quantum mechanics
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SHV	Sulfhydryl reagent variable
SMARTS	SMILES arbitrary target specification
SMILES	Simplified molecular-input line-entry system
SPM	São Paulo metallo-β-lactamase
TEM	Temoneira - the patient who provides the first sample
USD	United States Dollar
VIM	Verona integron-encoded metallo-β-lactamase
WHO	World Health Organization

Chapter I

Introduction

99 The thoughtless person playing with penicillin treatment is morally responsible for the death of the man who succumbs to infection with the penicillin-resistant organism. I hope this evil can be averted.

— Sir Alexander Fleming (Microbiologist)

Antibiotics: a leap into resistance

1

Since the beginning of their history, human beings had to overcome critical pandemics (smallpox, Black-Death, Spanish flu, AIDS...) leading to around 500 millions deaths. However, over the years, humans succeeded in saving themselves and humanity is not as endangered as it was previously thanks to advancements in modern medicine and to the enhancement of the standard of living. This way, given the positive evolution of worldwide demography, United Nations anticipated that the world's population shall reach 9,6 billions people by 2050. Simultaneously, a new global issue has appeared in the last decades, threatening the current health and socio-economic balance: the antibiotic resistance.

In this chapter, we will first have a brief look at how humans have managed their health issues from the emergence of the human species until today. Then, starting from the discovery of penicillin, benefits of the first antibiotics will be introduced. Finally, we will discuss about a potential postantibiotics' era due to the emergence of resistances and in which humans shall not be able to survive with the current medications.

1.1 Changes in humans' attitude

Analysis of prehistoric bones, particularly dental calculus in which DNA from micro-organisms or food is trapped, revealed the Neanderthal mans' habits in terms of diet and lifestyle. Thus, some of them seemed to be self-medicating, such as ingesting mildew or poplar to get a natural antibiotic (penicillium) or a pain-killer substance (salicylic acid), respectively [202].

In 3600 B. C., the Chinese enacted the first manuscripts of pharmacy, followed by the ancient Egyptian ones in 2900 B. C., all dealing with the power of Nature, inspired from religion, mythology and mysticism. Later, in Ancient Greece and Rome, temples were dedicated to Asclepius, god of healing, and philosophers became pharmacists (e.g. Aristotle) or vice versa (e.g., Hippocrates). Thereafter, the Greek doctor Galen (131-201 A. D.) was the worthy successor of his predecessors as he is now known as the father of Pharmacy whose name is linked to the oath sworn by all new pharmacists, the "Galen oath" (similarly, "Hippocrates oath" for new doctors).

Later, the period from Middle Ages to the 17th century was marked by a return to medicinal plants and complex mixtures: this is the experimental age of apothecaries. Up to 100 compounds were mixed in potions, then used first as antidotes and cures for all ills afterwards. The Royal Declaration of 1777 and the Law of Germinal 1803 established the legal basis of French pharmacy by permanently separating it from other medical activities, and demonstrated a great desire for clarity and regulation in the preparation of remedies. Therefore, instead of only treating people, it became important to fully understand why the healing occurred.

That's why, since the end of the 18th century, pharmacists and, by extension, scientists are seeking information about compounds responsible for the observed biological activities. As reported in the Figure 1 depicting a historical timeline, many therapeutic molecules were found or synthesized; some of them will be discussed in the next subchapter.



Figure 1: From the antibiotics' discovery to the emergence of antibiotic resistance: key dates.

1.2 From first antibiotics to antibiotics' era

In 1871, François Hallopeau (1842-1919), a French dermatologist, coined the word "antibiotic" to describe all substances preventing the development of life. Later, in 1941, Selman Waksman, a Ukrainian biochemist and microbiologist living in the United States, defined the noun "antibiotic" as: "[any] chemical substance, produced by micro-organisms, which has the capacity to inhibit the growth of and even to destroy bacteria and other micro-organisms". In current usage, an antibiotic is now literally identified as an "anti" "biotic", namely any compound which kills a micro-organism or inhibit its growth, with no distinction about its origin, either natural or synthetic.

The golden age of antibiotics began with the work of brilliant scientists, rewarded by several Nobel Prizes, and whose stories will be summarized below (Figure 2).



Figure 2: The antibiotic "wall of fame". Some of these scientists received the prestigious Nobel Prize:

— Physiology or Medicine — in 1908, Paul Ehrlich and Ilya Ilyich Mechnikov "in recognition of their work on immunity"; in 1945, Sir Alexander Fleming, Ernst Boris Chain and Sir Howard Walter Florey, "for the discovery of penicillin and its curative effect in various infectious diseases"; in 1952, Selman Abraham Waksman "for his discovery of streptomycin, the first antibiotic against tuberculosis."

— Chemistry — in 1964, Dorothy Crowfoot Hodgkin, one of the three women rewarded by a Nobel Prize, "for her determinations by X-ray techniques of the structures of important biochemical substances."

1.2.1 Louis Pasteur (1822-1895) and Jules F. Joubert (1834-1910)

In 1877, these two French scientists published their work revealing that *Bacillus anthracis*, the etiologic agent of a livestock's common disease, named anthrax, can't grow significantly in animals co-infected with other pathogenic bacteria [154]. Starting with this observation, they made three assumptions: (a) the *Bacillus* could form spores and survive in a stressful environment (high temperature, pressure, pure oxygen); (b) the blood of a putrefied animal (more than 16 hours) could kill a living animal without being exposed to anthrax; and (c) the growth of *Bacillus anthracis* was delayed by other microorganisms. However, Pasteur had already described a process, known as *la lutte pour la vie* (the battle for life), in which bacteria could develop in the blood during putrefaction [172]. In addition, other bacteria such as *Pseudomonas aeruginosa* or *Streptococcus pyogenes* have been used to prevent the infection from other microorganisms [70]. This antagonism was thereafter described as circulating chemical substances released by some bacteria in order to eradicate the others [31].

1.2.2 Paul Ehrlich (1854-1915)

The German physician had an idea based on the creation of "magic bullet", fatal to microorganisms but totally safe for the host. In partnership with the German chemist Alfred Bertheim (1879-1914) and the Japanese bacteriologist Sahachiro Hata (1873-1938), Ehrlich produced organoarsenic derivatives of the drug atoxyl to treat syphilis-infected rabbits. In 1909, they finally found a compound able to successfully cure those rabbits with hopes for the treatment of patients; this is the first example of chemotherapy [63]. Nevertheless, given the severe side effects, two additional compounds with an increased solubility and a reduced toxicity were released: Salvarsan in 1909 (arsphenamine or Ehrlich 606) and Neosalvarsan in 1912. Later, in 1940s, this drug will be replaced by another promising antibiotic: penicillin.

1.2.3 Alexander Fleming (1881-1955)

In 1890s, the two scientists Ernest Duchesne and Vicenzo Tiberio were the first to discover an antibacterial action coming from mold [183] and *Penicillium* [60] but were unable to identify its precise origin. In 1928, Sir Alexander Fleming, a Scottish biologist already well-known for the discovery of the enzyme lysozyme in 1923, pinpointed a substance with abilities to inhibit the growth of microbes (bacteriostatic) and even kill other microorganisms (bactericidal).

When he returned to his laboratory, Fleming first noticed an unusual colony of mold which grew in several *Staphylococcus*-containing Petri dishes. Then, he spotted an absence of *Staphylococcus* bacteria around the mold which led to the assumption of a substance preventing the bacteria's development. Identified as *Penicillium notatum*, a microscopic fungi cultivated in the nearby laboratories, Fleming decided to name the latter substance, penicillin [68].

Further tests in animals revealed no toxicity [19, 121] but given the difficulties encountered for its isolation and purification, penicillin was restricted to a military use [6]. It was not until the work of other renowned scientists that the story of penicillin was upgraded to an industrial level.

1.2.4 Ernst B. Chain (1906-1979), Howard W. Florey (1898-1968) and Edward Abraham (1913-1999)

In 1940, Ernst Chain and Howard Florey, a German-born British biochemist and an Australian pharmacologist, respectively, decided to continue the work of Alexander Fleming on penicillin with the help of Edward Abraham, a British chemist. With a view to develop it into a useful treatment, they succeeded in isolating the molecule thanks to chromatography on an alumine column. Besides the purification process, Abraham was in charge of the elucidation of its chemical structure. In 1943, jointly with Ernst, he presented the structure as a fused two ring system [39]. In 1945, this assumption was confirmed using X-ray crystallography thanks to Dorothy Hodgkin Crowfoot [51], whose involvement in the structural characterization of several biological molecules has been very important (see below).

From 1942 to 1944, expatriated in the United States because of the World War II, Florey involved the big American companies in the large-scale production of penicillin. In 1944, one of those companies, named Pfizer, opened its first penicillin production site and provided the American soldiers which landed in Normandy during the famous D-Day. At last, from 1945, penicillin is available in all pharmacies.

1.2.5 Dorothy M. Hodgkin Crowfoot (1910-1994)

Dorothy Hodgkin Crowfoot, a British chemist, is a pioneer in the X-ray crystallography field especially in the structural characterization of biomolecules. In her track record, we find the first structure of a steroid named cholesteryl iodide (1945) [38], the penicillin (1945) [51], the final structure of vitamin B12 (1955) [24, 50], and after 35 years of work, the structure of insulin (1969) [4].

1.2.6 Josef Klarer (1898-1953), Fritz Mietzsch (1896-1958) and Gerhard Domagk (1895-1964)

Following the discovery of penicillin, in 1932, two German chemists, Josef Klarer and Fritz Mietzsch, synthesized a new drug known as sulfonamidochrysoidine (KI-730, Prontosil). Its antibacterial activity was then assessed by Gerhard Domagk, a German bacteriologist, in different type of diseases [80]. In 1935, Ernest Fourneau's group, at the Pasteur Institute (Paris) found the active ingredient to be the sulfanilamide (Figure 4), a compound already synthesized and patented in 1909 by Paul Gelmo (1879-1961) but with no consideration for its medical potential. Moreover, sulfanilamide has many advantages such as being cheap to produce, off patent and really easy to modify, that's why many companies launched large-scale production of sulfonamides derivatives [10]. To give two successful light-hearted cases, Domagk's daughter and Franklin D. Roosevelt's son had been totally cured by Prontosil, which also boosted the popularity of such chemical scaffold. Furthermore, it was the beginning of combinatorial approaches to obtain potent synergistic effects: for example, trimethoprim, a powerful antibacterial agent, combined with sulfadiazine, a sulfonamide-containing compound [190].

1.2.7 Selman Waksman (1888-1973) and Harold B. Woodruff (1917-2017)

Both experts in soils microbiology, Waksman and Woodruff discovered a novel class of antibiotics, namely aminoglycosides which include actinomycin (1940) [193], streptothricin (1942) [194], streptomycin (1943), and neomycin (1949). Streptomycin was the first antibiotic used in the treatment of tuberculosis, an infectious disease usually caused by *Mycobacterium tuberculosis*. At last, scientists had found some clues to reach the "magic bullets" dreamt by Paul Ehrlich, but as Edward Abraham noticed in 1940, bacteria seems to have developed new biological weapons in order to fight against selective pressure due to antibiotics: enzymes [2]. In the next subchapter, we will focus on the current medical situation trying to explain the origin of antibiotic resistance and why it is important to urgently adopt a suitable behavior about antibiotic use in global medicine.

1.3 Antibiotic resistance

The World Health Organization (WHO) is clear: "without urgent action, we are heading for a post-antibiotic era, in which common infections and minor injuries can once again kill". Nevertheless, the problem is more general since it will affect not only the health sector but also the worldwide economy: for example, a patient infected by one multi-drug resistant (MDR) strain will stay longer at the hospital, have more expensive health cares and could have greater risk of dying.

According to O'Neill's report https://amr-review.org, whether no reaction arise from the international community, antimicrobial resistance (AMR) might be responsible for 10 million deaths per year in 2050 and may cost over 100 trillion USD by 2050. Indeed, the main problem comes from some MDR bacteria which are now resistant to all existing antibiotics classes including carbapenems, the last-resort antimicrobial weapons.

Historically, two antibiotics, colistin and tigecyclin, were found to be powerful against pathogens. Nevertheless, after a short duration use of colistin in 1960s, it was suspended due to serious side effects, namely nephrotoxic, but its use was maintained in animals. However, to face the emergence of MDR bacteria, colistin was recently reintroduced for humans but, as feared, colistin resistance quickly appeared [149]. From the tigecyclin side, a recent study noted the emergence of plasmid-mediated high-level tigecycline resistance genes in animals, meat for consumption and humans [86].

Hence, novel antimicrobial compounds with original activities are urgently needed to ensure the survival of human species; this topic will be discussed later. In this part, we will define the antibiotic resistance before presenting the reasons which may explain this phenomenon, and we will finish on the global role of environment in the spread of resistance.

1.3.1 Definition

Antimicrobial resistance (AMR) arises when a microorganism become resistant to an antimicrobial agent for which it was first susceptible. Over the last few years, many antibiotics have been threatened by resistance and the latter may arise and spread in different ways from one microbe to another (Figure 3). First, through horizontal gene transfer, mobile integrons (i.e., genetic elements including a site-specific recombination system able to integrate, express and exchange specific DNA elements, called gene cassettes) carried on transposons (i.e., class of genetic elements that can "jump" to different locations within a genome thanks to enzymes called transposases) promote the share of resistance mechanisms. In other words, a resistance gene can be spread among bacteria with genetic vehicles, named integrons, and whose mobility is ensured by specific enzymes. Thus, a transposon carrying several antibiotic resistance cassettes might confer a multi-resistant phenotype to an organism initially susceptible or mono-resistant. Secondly, through vertical gene transfer, bacteria may transmit natural mutations on chromosomal genes to offsprings by replication. Finally, other mechanisms such as circulating DNA (called plasmids) picked up by bacteria from environment, transmission of integrons from dead bacteria or bacteriophages, or the conjugation process may lead to the spread of resistance [48, 131].

Otherwise, resistance may occur by different mechanisms, including target's modification, reduction in cell permeability, efflux systems' overexpression and production of specific enzymes (Figure 3).

1.3.2 Where does the antibiotic resistance come from ?

We previously dealt with the natural process of resistance but the main cause is still a lack of public knowledge on antibiotics, leading to a misuse even if their prescription and purchase are more and more controlled. This improper use of antibiotics involved a non-natural selective pressure on microorganisms which forced them to rapidly evolve [66, 82, 180]. Moreover, given that most antibiotics stem from natural fungi or bacteria, additional selection pressures may happen due to antibiotic-producing strains. Indeed, these strains may either transfer genes encoding resistance to their own antibiotic (generally located in the same cluster as the antibiotic biosynthesis ones) [92] or produce the antibiotic directly in the environment applying a selective pressure on neighboring organisms [7]. One may even speak about a genuine "resistome", namely a reservoir of resistance genes [156].



Figure 3: Main mechanisms and acquisition pathways involved in antibiotic bacterial resistance. Figure inspired from Figure 1 in Chellat *et al.*, 2016 [41].

According to a recent study, the antibiotic resistance might also arise from several drivers (clinical, biological, social, political, economic, environmental) and results underline, for example, a clear difference between high-income countries, in which antibiotic use remained flat over the past fifteen years, and low to middle income countries in which antibiotic use suffered a sharp increase due to economic improvements and changes in diet [191].

Another source of antibiotic resistance lies in agriculture field including feedstocks, husbandry, aquaculture [90], horticulture, animals and so on. In the next lines, we will see some examples to illustrate this.

— **Agriculture** — The prophylactic use of antibiotics in feeds (e.g., sulfonamides or other non-clinically relevant antibiotics), in other words, the process of treating to prevent disease, may have far-reaching implications for human health since it could potentially lead to the co-selection of environmental bacteria with resistance to clinically important antibiotics [28].

Furthermore, during the fattening process, some veal calves are probably fed with milk containing antimicrobial residues which is hypothesized to explain the high extended spectrum β -lactamase (ESBL) loads in animals at the entrance on farms [79]. By the same logic, antimicrobial effects on swine gastrointestinal microbiota were analyzed and outcomes revealed an emergence of antibiotic resistance genes (ARGs), in other words, another kind of "resistome" [211].

— **Companion animals** — Pets are not spared from antibiotic resistance since some veterinary studies have been carried out on cats, dogs and horses, highlighting a global lack of guidelines in this field (inappropriate doses, veterinarian-dependent treatments...) [94, 140, 179]. In addition, antibiotics used in veterinary medicine from the Mediterranean Basin include mainly tetracycline, aminoglycosides, fluoroquinolones, and polymyxins, but emergence of ESBL and carbapenemase producers in animals seems to be related to the co-selective pressure applied by the misuse of non-β-lactams rather than to the use of β-lactams [53].

— Food — In terms of food, highly virulent strains of *Arcobacter butzleri* were isolated from shellfish and characterized as multi-resistant (β-lactams, vancomycin, tetracycline), raising questions about potential risks related to their consumption [67]. Even more worrisome, carbapenem-resistant enterobacteriaceae were found in ready-to-eat vegetables from supermarkets in China, which constitutes a food safety issue [122].

1.3.3 Environmental spread of antibiotic resistance

Environment has, despite itself, a dual role in antibiotic resistance since it is simultaneously guilty about being a huge "resistome" with all existing antibiotic-producing bacteria, and victim of human or animal activities as we saw in the previous subchapter. Thus, in this part, we will itemize a few recent examples concerning the evolving environment behavior in antibiotic resistance phenomenon.

In waters and soils, microorganisms are potentially threatened by presence of antibiotics increasingly released into these environments by agricultural [127], human and animal activities. To enlighten this last point, a study has been achieved on crows used to roost in a constructed wetland near the University of Washington Bothell/Cascadia College (UWB/CC) campus. Fecal isolates from Crow Roost and water isolates had similar antibiotic resistance pattern whose approximately 40% were multi-drug resistant. However, strains did not appear to survive for long in the wetland, but it shows a possible transfer from corvids to waters. With their habits to forage on a variety of wastes (garbage dumps, hospital, animal, feedlots), crows should be considered as potential reservoirs and vectors of antibiotic resistant and pathogenic bacteria (here, *Escherichia coli*), underlining a significant environmental threat [175]. — **Soils** — In addition, main processes determining the antibiotic persistence in soils are sorption to organic particles and degradation/transformation. An in-depth study revealed a wide range of disappearance time 50 (DT50) and half-life values, depending on antibiotic physico-chemical properties, soil's features, and climatic factors (temperature, rainfall, humidity).

As a reminder, in Eco-toxicology language, half-life corresponds to the time taken for 50% transformation of a test substance when the transformation can be described by first-order kinetics (value independent of the initial concentration) and DT50 means the time needed to reduce the initial concentration of this substance by 50%.

This way, some antibiotics have a half-life higher than 100 days, such as fluoroquinolones (e.g., ciprofloxacin: 3466 days), tetracyclines (e.g., doxy-cycline: 578 days), and macrolides (e.g., azithromycin: 990 days), while easily-hydrolysable antibiotics have a half-life lower than 10 days, like β -lactams (amoxicillin: 0.6 days). Results also underlined the antibiotic impact on soil microbiome modifying both its composition (i.e., Gram-negative bacteria, Gram-positive bacteria, fungi) and its enzymatic activity. Nonetheless, authors emphasized the discovery of new genes and enzymes related to antibiotic resistance but claimed that precise estimation of the impact of antibiotics on the activity and diversity of soil microbial communities remains a great challenge [52].

— Waters — On the other hand, the emergence of antibiotic resistance in waters has been scanned through several investigations across the world. In Lake Água Preta (Brasil), used for human activities including post-treatment consumption, high dissemination of ESBL-producing bacteria was observed although authors mentioned this lake did not exhibit features of a strongly impacted environment [71].

Moreover, wastewaters have been widely analyzed in order to detect antibiotic resistance signs: in China, ESBLs-producing Escherichia coli isolated from wastewater treatment plants showcased the role of plasmids in the multiple antibiotic resistance transfer [196]. Similarly in Sweden, β -lactamase genes were identified from aquatic environments impacted by household and hospital wastewater [110], while in Israel, the disinfection of greywater to reduce the potential presence of pathogens and ARGs was advised [188]. Additionally, in the United States, carbapenemase-producing bacteria (CPB) with clinically important genotypes, including those producing *Klebsiella pneumoniae* carbapenemase (KPC) or New Delhi metallo- β -lactamase (NDM), were exposed in a global analysis of wastewater treatment plants effluent and nearby surface waters. As we saw in the subchapter 1.3.2, this is a concern for both public health and animal agriculture since introduction of CPB into intensively managed livestock populations could lead to their amplification and foodborne dissemination [132].

Nevertheless, *in silico* analysis on Global Ocean Sampling metagenomic revealed that majority of distant metallo- β -lactamases (MBLs) homologs from Atlantic, Indian, and Pacific Oceans was mainly related to a chromosomally encoded MBL, named GOB (i.e., from *Chryseobacterium meningosepticum* class B), present in *Elizabethkingia genus* and only a minority of them was related to the acquired MBL enzymes (VIM, SPM-1, and AIM-1) responsible for global resistance in hospitals. Thus, authors concluded that low antibiotic impacted marine environments, such as the ocean, are probably not the first source of the high-virulent ARGs currently threatening the public health [69].

1.4 Summary

In this global introduction about antibiotic resistance, we first developed the simultaneous evolution of humans and healing methods, from the basic use of plants in Prehistory to the elaborated design of medications by checking for compounds' chemical identity and their related biological properties, in the 19th century. Secondly, we defined what antibiotic resistance phenomenon is and where it comes from, through the analysis of several examples taken from nature (animals, environment, antibiotic-producing bacteria) and human activities (agriculture, hospital, household). Thirdly, we detailed the role of environment in the spread of antibiotic resistance, mostly supported by an anthropological help, and we noticed that even though it is not the main source of current clinical issues, environment takes part in the global dissemination of ARGs.

In our team, we work especially on β -lactam resistance, therefore next chapters will be focused on this topic.

Actors of β-lactam resistance

2

Previously, we learnt that antibiotic resistance occurs when bacteria evolve either naturally or under antibiotic pressure. Thus, to fully understand the whole process, actors of β -lactam resistance must be presented. In this subchapter, we will sequentially present Enterobacteriaceae bacteria, β lactam antibiotics and powerful biological weapons developed to fight against antibiotic pressure: β -lactamases. To further complete the topic, Figure 4 presents other existing classes of antibiotics.



Figure 4: Existing antibiotic classes.

2.1 Enterobacteriaceae

2.1.1 General features

Enterobacteriaceae family is ubiquitous in various worldwide ecological sources such as soil, water, vegetation and animals. In addition, some species are human pathogens (e.g., *Salmonella enterica*, *Shigella sp.*, *Yersinia sp.*) while others belong to the mammalian gut microbiota, so called enteric bacteria, in which they act as opportunistic pathogens (e.g., *Escherichia coli*, *Klebsiella spp.*, *Proteus spp.*).
Main transmission of intestinal infections is classically fecal–oral with various pathways: person-to-person, direct contact with animals or their environment, or consumption of contaminated food or water. Infection may also have an endogenous origin (e.g., bacterial translocation from the gut to blood), resulting in extraintestinal disease, and is more often observed in immunocompromised hosts or persons with underlying conditions such as cirrhosis or those undergoing chemotherapy.

Enterobacteriaceae are Gram-negative bacteria, thereby with specific properties further developed in the next part, facultative anaerobes, meaning they can survive either in presence or absence of molecular oxygen (O_2), and sugar fermenting. Most of them possess flagella allowing motility and type I fimbriae to adhere to their hosts cells.

2.1.2 Gram-negative bacteria features

Compared to Gram-positive bacteria, Gram-negative bacteria's cell envelope always includes (1) an outer membrane [62], containing lipopolysaccharides (LPS), such as lipid A, core polysaccharides and O-antigens in its outer side and phospholipids in its inner side, which protects (2) the thin peptidoglycan layer, leading to a non-coloration by crystal violet in Gram-staining method of bacterial identification (Figure 5), and (3) an inner membrane, also named cytoplasmic membrane, made up of phospholipids (e.g., phosphatidyl ethanolamine, phosphatidyl glycerol, phosphatidyl serine and cardiolipin in *E. coli* [162]). Unlike Gram-positive bacteria which have a dense peptidoglycan layer, Gram-negative bacteria have a thinner peptidoglycan layer which allows the alcohol penetrate the cell, thus eliminating the crystal violet during the third step.

Space delimited by inner and outer membranes, called periplasm, is filled with a concentrated gel-like substance which can sequester potentially harmful degradative enzymes such as RNAses or alkaline phosphatases. Otherwise, LPS, linked to the outer leaflet of the outer membrane thanks to a fatty acid-containing glucosamine disaccharide (lipid A), are also termed endotoxins [107] and are sensitively recognized by the immune system since their release during cell lysis can cause septiceaemia. Moreover, a classification of pathogenic strains was defined according to the O-antigens' properties after some variations between different strains of the same species were identified [163]. LPS may bind each other in presence of magnesium ions to create tight packing, enhanced by the high saturation of their acyl chains, acting as an effective barrier against hydrophobic molecules. Additionally, specific transmembrane proteins, referred to as outer membrane proteins (OMPs) or porins, contain large domains folded into β -sheets and wrapped into cylinders, whose some of them (e.g. OmpF, OmpC) allow the passive diffusion of small molecules such as monosaccharides, disaccharides and amino-acids. Porins may also limit the diffusion of hydrophilic molecules larger than around 700 Daltons.

To sum up, association between LPS and porins creates a powerful, effective and selective permeability barrier [145].



Figure 5: The Gram-staining method.

2.1.3 Peptidoglycan formation

Peptidoglycan, or murein, is one of the three layer constituting the cell wall in Gram-negative bacteria. It is built by repetition of disaccharide Nacetylglucosamine (NAG or GlcNAc) and N-acetylmuramic acid (NAM or MurNAc), cross-linked by three to five peptide side chains [192] (Figure 6), leading to a 3D mesh-like layer. Peptidoglycan monomers are synthesized inside the cell, before being attached to a membrane carrier, named bactoprenol, and transported across the inner membrane. Once in the periplasm, enzymes called transglycosidases perform a transglycosylation process, to insert the newly created monomers into the existing peptidoglycan. In this way, the C4 hydroxyl group of the GlcNAc will be joined to the C1 of MurNAc in the glycan, displacing the lipid-PP from the glycan chain [203].



Figure 6: Peptidic connection between two peptidoglycan layer. Figure modified from GlycoPedia website [https://www.glycopedia.eu/e-chapters/the-structure-of-bacterial-cell/article/peptidoglycan-molecular-structure].

2.1.4 Penicillin-binding proteins

Historically, penicillin-binding proteins (PBPs) are D,D-transpeptidases for which identification was carried out by covalent labelling with radioactive penicillin followed by gel electrophoresis [210]. PBPs catalyse the polymerization of the glycan strand (transglycosylation) and the cross-linking between glycan chains (transpeptidation), as presented in Figure 7. This enzymatic reaction follows a three-step mechanism: (1) fast and reversible formation of a non-covalent Henri–Michaelis complex between the enzyme and a peptidoglycan stem pentapeptide, called the donor strand, followed by (2) attack of the active site serine on the carbonyl carbon atom of the C-terminal D-Ala-D-Ala peptide bond, leading to the formation of an acylenzyme intermediate and the simultaneous release of the C-terminal D-Ala (acylation), and (3) deacylation in which either the shortened peptide is hydrolysed and released (carboxypeptidation), or is cross-linked with a second peptidoglycan stem peptide called the acceptor strand (transpeptidation).

As exposed in Figure 6, the donor strand is covalently linked to the glycan strands with an amide bond to the carboxyl carbon of the D-lactyl group of the MurNAc. While glycans' composition is highly conserved across species, this pentapeptide sequence is more diverse and involve D-amino-acids, namely non-natural.

In the first position from the lactyl group, we found a L-alanine (L-Ala) followed by a D-isoglutamic acid (D-iGlu), which is sometimes amidated in Gram-positive bacteria to yield a D-isoglutamine (D-iGln). Then, the γ -carbon of this residue is connected to probably the most variable amino acid within the stem peptide, usually the meso-diaminopimelic acid (m-DAP) or a D-lysine (D-Lys) in Gram-negative and Gram-positive, respectively [192]. Finally, the peptide stem ends with two D-alanines (D-Ala).

Despite a D,D-transpeptidase activity involving the formation of crossbridges, PBPs also display a D,D-endopeptidase activity to hydrolyse the cross-bridges [173].

The main interest in developing new β -lactam antibiotics lies in the structural resemblance between the D-Ala-D-Ala pattern and β -lactam ring; this key point will be explained in the next section.



Figure 7: Peptidoglycan formation and inhibition of the penicillin-binding protein (PBP). Figure inspired from Wikimedia Commons website [https://commons.wikimedia.org/wiki/File:PBP_catalysis.svg].

2.2 β-lactam antibiotics

Since the discovery of penicillin, β -lactams remain among the most commonly used antibiotics to treat microbial infections, as concluded in a recent survey about antibiotic consumption [113]. This widely use is due to the efficiency and low toxicity of β -lactams besides having a mechanism targeting a vital cellular function of bacteria, namely the cell wall synthesis, which is absent in humans cells. Otherwise, scientists have achieved numerous changes in order to improve their chemical properties such as potency, spectrum of activity, pharmacokinetics and safety profiles.

2.2.1 Structure

Nowadays, four main classes of β -lactam antibiotics are prescribed in clinics, including a monocyclic scaffold, namely monobactams [96, 182], and bicyclic scaffolds, named according to the nature of the second cycle linked to the initial four-membered β -lactam: penicillins (five-membered thiazolidine), cephalosporins (six-membered dihydrothiazine) [26, 144] and carbapenems (five-membered pyrroline) [27]. Figure 8 depicts these chemical structures together with some examples of modified antibiotics used to treat infections.





2.2.2 Mode of action

Initially, β -lactams were identified as antibacterial agents due to their structural similarity with the terminal D-Ala-D-Ala moiety [184], in other words, the β -lactam amide and adjacent carboxylate groups mimicking the peptide bond in peptidoglycan stem pentapeptide (as previously presented in section 2.1).

Expected activity results from a nucleophilic attack of the PBP active site serine on the β -lactam ring, leading to its opening (Figure 7). The newly created acyl-enzyme complex is stable and not immediately hydrolysed, thus inactivating the PBP enzyme and preventing the peptidoglycan transpeptidation process [173]. In consequence, given the subsequent bacterial cell wall weakening, osmotic pressure differences between inside and outside of the cell cause lysis and bacteria death [152].

2.2.3 Examples of chemical modifications

One golden rule: antibacterial activity has to be retained after application of selected structural changes. This way, several positions on the β -lactam scaffold should be considered: C6 in penicillins, C3 and C7 in cephalosporins, C2 in carbapenems and C3 in monobactams (Figure 8).

Some modifications led to key developments such as introduction of aminopenicillins (e.g., ampicillin) to extend penicillins' spectrum activity including Gram-negative bacteria [168], methicillin to counter penicillin-resistant *Staphylococcus aureus* strains [167], and oxyiminocephalosporins (e.g., cefotaxime, ceftazidime) to halt emergence of β -lactamase-mediated resistance in Gram-negative bacteria [130].

2.3 β-lactamases

As previously outlined, different mechanisms may cause the emergence of resistance, including, in the case of β -lactams, mutation or expression of alternative PBPs, down-regulation of porins required for β -lactam entry, efflux systems' over-expression and production of specific enzymes (Figure 3) [111]. During my thesis, I mainly focused my work on the enzymemediated resistance which arises from the activity of β -lactamases, enzymes produced by both Gram-positive and Gram-negative bacteria in order to hydrolyze the β -lactam amide [32].

Since resistance arose in *Staphylococcus aureus* via production of the PC1 enzyme [112] and methicillin was introduced to counter it [167], it's a bit of cat and mouse between scientists and bacteria. Nevertheless, introduction of new β -lactams caused the emergence of new β -lactamases either by mutation of existing families or dissemination of genes encoding new enzymes; as a reference, the β -lactamase database (www.bldb.eu) [141] now gathers information from over 4300 enzymes more or less characterized.

2.3.1 Classification

Two systems are currently used to classify β -lactamases: one based on activity, the Bush–Jacoby–Medeiros system [33,34], and another based on sequence information, the Ambler system [9]. The latter defines four distinct classes, namely class A, B, C and D, according to specific sequence patterns. Furthermore, these four classes are split into two subgroups according to their hydrolytic mechanism: classes A, C and D are defined as serine β -lactamases (SBLs), with a catalytic serine, and class B includes metallo- β -lactamases (MBLs), with one or two zinc ions as catalysts.

— Class A represent the most studied group with representative enzymes like TEM (named after the patient Temoneira providing the first sample), SHV (Sulfhydryl reagent variable), KPC (*Klebsiella pneumoniae* carbapenemase) and CTX-M (Active on cefotaxime, first isolated at Munich). They can hydrolyse penicillins (amoxicillin, ticarcillin, piperacillin) but they are susceptible to some inhibitors such as clavulanic acid, tazobactam and sulbactam.

— Class B gathers metalloproteins whose hydrolytic activity depends on the presence of one or two zinc ions in the active site. Well-known enzymes of this class are NDM (New Delhi metallo-β-lactamase) and VIM (Verona integron-encoded metallo-β-lactamase). Otherwise, no activity on aztreonam was reported and classic inhibitors are inefficient compared to divalent ion chelators as EDTA (ethylenediaminetetraacetic acid).

— Class C β -lactamases, also termed cephalosporinases, may hydrolyse aminopenicillins and first, second and third-generation cephalosporins. They are inhibited by cloxacillin but clavulanic acid has no effect on them. CMY (Active on cephamycins) and ADC (*Acinetobacter* derived cephalosporinase) are representatives of this family.

— Class D enzymes, or oxacillinases (OXA), constitute a very heterogeneous group based on hydrolysis profile. Initially, an enzyme was identified as oxacillinase if the latter was able to hydrolyse oxacillin faster than it hydrolyses benzylpenicillins, but this definition has widely evolved regarding the variety of enzymes currently observed among this class. Moreover, certain groups are especially involved in carbapenem resistance, for example, OXA-23 and 24/40 groups in *A. baumannii*, and the OXA-48 group in Enter-obacteriaceae.



Figure 9: Three-dimensional structures of representative β-lactamases from each class. Catalytic residues are displayed in orange and zinc ions as gray spheres for serine-β-lactamases and metallo-β-lactamase, respectively. Structures have been selected with two criteria: absence of relevant ligands in the active site, namely apo-structures, and a resolution as high as possible. (A) Class A KPC-2 (PDB 5ul8 1.15 Å). (B) Class B NDM-1 (PDB 5zgy 0.95 Å). (C) Class C AmpC (PDB 1ke4 1.72 Å). (D) Class D OXA-48 (PDB 3hbr 1.90 Å). Figure inspired from Fig. 2 in Tooke *et al.*, 2019 [186].

2.3.2 Enzymatic mechanisms

First step of SBLs' mechanism is similar to PBPs one: the catalytic serine performs a nucleophilic attack on β -lactam ring resulting in its acylation (covalent complex). Then, the complex is hydrolysed via a water molecule located at a strategic position only in SBLs since there is no β -lactamase activity in PBPs. MBLs use, in turn, a zinc-activated nucleophilic water to lead the hydrolysis. Both mechanisms are reported in Figure 10.

- SBLs - Once regenerated after the nucleophilic attack of the water molecule on acyl-enzyme complex, SBLs may inactivate additional β -lactam molecules. This enzymatic reaction may be written as follows:

$$E + S \xrightarrow[]{k_1} E : S \xrightarrow[]{k_2} E - S \xrightarrow[]{k_3, H_2O} E + P$$

in which, E, S, E:S, E-S and P represent a β -lactamase, a β -lactam substrate, the Michaelis complex, the acyl-enzyme and the product with no longer antibiotic activity, respectively.



Figure 10: Mechanism of carbapenem hydrolysis achieved by serine-β-lactamases (A) and metallo-β-lactamases (B). Key residue means the residue(s) — may be the same — needed for the activation of Ser70, which is different for each SBL class: Glu166 (class A), Tyr150 (class C), carboxylated lysine KCX73 (class D), and for the activation of hydrolytic water. In MBLs mechanism, anionic nitrogen intermediate is assumed to be protonated by Asp120.

About rate constants, k_1 and k_{-1} are related to the association and the dissociation of the pre-acylation complex, k_2 to the acylation process, and k_3 to the deacylation process. Moreover, β -lactamases' kinetics may be understood thanks to various descriptors:

– K_m , the Michaelis-Menten constant [72], defined as:	$K_m = \frac{k_3 K_s}{k_2 + k_3}$
where K_s , the thermodynamic constant [45], is:	$K_s = \frac{k_{-1} + k_2}{k_1}$
– k_{cat} , the maximum turnover number, defined as:	$k_{cat} = \frac{k_2 k_3}{k_2 + k_3}$
– V_{max} , the maximal activity, defined as: where E_T is the total concentration of enzyme.	$V_{max} = k_{cat} [E_T]$

Also, K_m , expressed in terms of concentration, represents the relative affinity between E and S and the velocity at which the E-S complex is converted into E and P; the bigger K_m value is, the lower affinity is. K_m also corresponds to the substrate concentration when the velocity is equal to one half of the maximal velocity for the reaction, meaning $V_{max}/2$. — **MBLs** — Unlike SBLs, MBLs are zinc-containing enzymes with a different hydrolytic mechanism in which a water molecule, coordinated to one or two zinc ions, is activated into hydroxide anion (HO–) before opening the β-lactam ring (Figure 10). Based on the zinc reliance, class B has been subdivided into three subclasses: B1, which are enzymes fully active with either one or two ions, B2 including enzymes which use one ion and are inhibited by binding of an additional ion, and B3, with enzymes requiring two ions [18, 73, 88, 155].

Moreover, those subclasses do not have the same hydrolysis profile: B1 and B3 enzymes have a broad-spectrum substrate profile (penicillins, cephalosporins, and carbapenems) while B2 ones exhibits a narrow profile focused on carbapenems [151].

Coming back to mechanisms, MBLs, with a di-zinc active site such as *S*. *maltophilia* L1 (B3), first bind the β -lactam substrate through its carboxylate and carbonyl groups, bridged by an activated water molecule. Then, one of the zinc ions, helped by adjacent residues, polarizes the β -lactam carbonyl to facilitate the nucleophilic attack by the hydroxide ion, which is hydrogen bonded to deprotonated Asp120. A tetrahedral species is subsequently created but quickly transformed into an intermediate in which the β -lactam nitrogen is anionic. Finally, a proton, whose origin is not validated and may come from Asp120 or a close water molecule, is added to this nitrogen, whose leads to the product formation.

Similarly in *Bacillus cereus* BcII enzyme (B1), which is active in both its mononuclear and di-nuclear forms, mechanism requires the hydroxide ion to be hydrogen bonded with Asp120 and other active site residues, and to be bound to the zinc ion [49, 197]. In this case, the source of the proton needed to protonate the β -lactam nitrogen is unknown.

Finally, in the case of CphA enzyme (B2), the presence of a second zinc ion is inhibiting. The proposed mechanism includes a water molecule activated by either His118 or Asp120, rather than a zinc-mediated activation. The singular zinc ion appears to help the coordination of β -lactam nitrogen [78, 205].

2.4 Summary

In this descriptive subchapter about actors of β -lactam resistance, we first described the powerful weapons used by Enterobacteriaceae, such as the presence of an outer membrane, porins or β -lactamases, in order to fight against antibiotics pressure.

Secondly, we briefly exposed the chemical structures of developed β -lactams and we explained how they work as a peptide bond mimic of D-Ala-D-Ala pattern found in the peptidoglycan stem pentapeptide, which is a penicillin-binding proteins' substrate. This structural similarity leads to inhibition of PBPs and, consequently, prevents the cell wall formation and kills bacteria.

Thirdly, we focused on enzymes able to counter the antimicrobial property of β -lactams by hydrolysing the β -lactam ring, hence their name, β -lactamases. We also dealt with their classification, exposing four classes with their respective features (spectrum of activity, potential inhibitiors...) and two main hydrolytic mechanisms, either with a catalytic serine or with one or two zinc ions.

For further information about β -lactamases (classification, existing inhibitors, mechanisms...), I recommend the reliable review written by Tooke *et al.* [186].

Now that actors of β -lactam resistance have been reviewed, let's introduce the strategies established to limit, prevent and neutralise the spread of existing resistant strains.

Strategies against β-lactam resistance

Given the growing threat of antibiotic resistance in pathogens, novel antimicrobial strategies need to be developed. For example, targeted killing of pathogenic bacteria without harming the host microbiota is a promising strategy to cure disease and limit both antimicrobial-related dysbiosis, namely a microbial imbalance inside the body (impaired microbiota), and development of antimicrobial resistance.

An obvious strategy is the development of inhibitors active against bacteria's main functions, such as biofilm formation [153], conjugation process [23, 36], or protein translation by targeting and using the ribosomal machinery as a platform for the directed-evolution of peptide-based antibiotics [40].

Otherwise, many reviews have been published about current strategies used in the fight against antibiotic resistance [22, 42, 116, 139, 158, 174, 176] and in this subchapter, we will provide a non-exhaustive list of the most representative examples.

3.1 Antimicrobial adjuvants

This strategy relies on the synergy which may exist between compounds, either natural or synthetic but with antimicrobial properties, and existing antibiotics [81].

As a first example, iron oxide nanoparticles (IONPs), used in clinics for their biocompatibility and magnetic properties, have been conjugated with last-resort glycopeptide antibiotic teicoplanin to be magnetically directed to infection sites [13].

More recently, combination of low-concentrated polymer, termed P, and doxycycline killed over 99% of planktonic and biofilm *Pseudomonas aerug-inosa* in 20 minutes, besides reducing the rate of resistance development and restoring susceptibility to treatment in the resistant strains [143]. The polymer P is composed of several units of biocompatible oligoethylene glycol important for the low-fouling properties, hydrophobic ethylhexyl groups to induce membrane disruption, and primary amino groups to interact with the bacterial membrane (Figure 12.A).

This is also the case of β -lactamase inhibitors which have the mission to incapacitate β -lactamases while antibiotic reach its target [41].

Given the β -lactam-oriented research focus of our laboratory, this point is detailed below with a rapid overview of existing β -lactamase inhibitors.

3.1.1 β -lactamase inhibitors

Such anti- β -lactamase agents may act as (1) reversible/irreversible inhibitors with high affinity but unfavorable steric interactions (e.g., carbapenems, cephalosporins), or (2) mechanism-based "suicide inhibitors" [30]. The latter follow a mechanism of irreversible inactivation which may be written as the following equation:

$$E+I \xrightarrow[k_{k-1}]{k_{k-1}} E: I \xrightarrow{k_2} E - I^* \xrightarrow{k_3} E - I$$

in which, E, I, E:I, E-I* represent a β -lactamase, a β -lactamase inhibitor, a β -lactamase-inhibitor complex, and the permanently inactivated β -lactamase or "dead-end complex", respectively.

About rate constants, k_1 and k_{-1} are related to the association and the dissociation of the β -lactamase-inhibitor complex, and k_2 to the irreversible inactivation. Irreversible inhibition can be characterized by first-order rate constants: k_{inact} , the inactivation rate achieved with an "infinite" concentration of inactivator, and K_I , the inhibitor's concentration for which an inactivation rate equal to $k_{inact}/2$ is reached [29, 45]. However, while K_I may be approximately compared to K_m for enzyme substrates, depending on individual rate constants (k_1 , k_{-1} , k_2 and k_3), the K_I may or may not equal the equilibrium constant K_i (= k_{-1}/k_1) determined under pre-steady-state conditions.

Furthermore, the inhibitory concentration 50 (IC50) measures the amount of inhibitor required to decrease the enzyme activity to 50% of its uninhibited velocity. While an IC50 can reflect the inhibitor's affinity or k_{cat}/k_{inact} ratio, these parameters are not always concordant, namely an inhibitor can have a weak "affinity" and slowly acylate the enzyme but still yield a low IC50 because of low deacylation rates.

Otherwise, the inhibitor's activity can be evaluated by the turnover number (t_n) (also equivalent to the partition ratio [k_{cat}/k_{inact}]), defined as the number of inhibitor molecules hydrolysed per time unit before the irreversible inactivation of one enzyme [35].

Figure 11 presents some inhibitors approved or in clinical development.



Figure 11: Overview of existing β -lactamase inhibitors.

3.1.2 Newly discovered antimicrobials

Emergence of new antibacterials targeting both Gram-positive and Gramnegative bacteria is needed as soon as possible to fight against the increasing antibiotic resistance. To illustrate this point, a mini-review of developed antibiotics, in the last two years, is presented below and Figure 12 sums up their representative structures.

— **Chemical compounds** — From the early 1960s, quinolone has been improved, from a narrow to an expanded spectrum and high efficacy against both Gram-positive and Gram-negative bacteria, including mycobacteria, and anaerobes. They may act by either inhibiting bacterial nucleic acid synthesis, or disrupting the topoisomerase IV and DNA gyrase enzymes, or causing breakage of bacterial chromosomes [157].

A tetrazole compound with dual inhibitory activity on SBLs and MBLs was reported, together with crystal structures of complexes with CTX-M-14, KPC-2, both serine β -lactamases (class A) and NDM-1, a metallo- β -lactamase (class B) [187].

Tryptamine derivatives (Figure 12.B), used as antibiotic adjuvant, are able to disarm colistin resistance in Gram-negative bacteria, with no bacterial toxicity [16], as well as niclosamide and its *O*-alkylamino-tethered derivatives [206].



Figure 12: Original developed antibiotics.

6-Bromoindolglyoxylamide derivatives (Figure 12.C) play a dual role as antimicrobial agents and antibiotic enhancers in Gram-positive and Gramnegative bacteria, through a rapid membrane depolarization and permeabilization [120]. New β -lactam–tetramic acid hybrids also show promising activities against both Gram-type bacteria [44].

Cyclic boronates may inhibit all SBLs classes by mimicking tetrahedral intermediates common to SBL and MBL catalysis (Figure 11) [37, 115].

— Antimicrobial peptides (AMPs) — Antimicrobial peptides (AMPs) are highly selective, well tolerated and safe as an antimicrobial treatment [133], may rapidly kill target cells, and display a broad activity spectrum on pathogens like fungi, bacteria, virus and protozoan [54, 64, 65].

Additionally, AMPs belong to innate system of insects, amphibians and mammals [55]. Recent developments led to the discovery of novel and potent AMPs, such as helicoidal AMP with reduced hemolytic effect [119], tridecaptin-inspired antimicrobial peptides with strong activity against MDR Gram-negative bacteria [14].

In a similar vein, dimeric lysine *N*-alkylamides, developed as mimics of AMPs, revealed high specificity on bacteria over mammalian cells, through a suggested bacterial cell membrane disruption, and biofilm formation inhibition [147].

Lately, AMPs from medicinal leech *Hirudo medicinalis* have been directly identified in its genome by computational algorithms, synthesized and tested, yielding promising outcomes on bacterial growth inhibition [83].

Very recently, an antimicrobial peptide, termed thanatin, was found to be able to disrupt the outer membrane of NDM-1-producing bacteria by competitively displacing divalent cations on the outer membrane and inducing the release of lipopolysaccharides. Furthermore, an inhibition activity has been highlighted on NDM-1 enzyme through the displacement of active site zinc ions, therefore reversing carbapenem resistance in NDM-1-producing bacteria in vitro and in vivo [126].

— Promising antibiotic combinations — Novel β -lactam- β -lactamase inhibitor combinations are expected to treat carbapenem-resistant Gram-negative pathogens, like ceftazidime-avibactam and meropenem-vaborbactam. Vaborbactam (Figure 11), a boronate-based compounds known to inhibit SBLs, has been recently classified as moderate inhibitor of MBLs [117].

New drugs are in clinical development such as avibactam combined with aztreonam or ceftaroline, or the combination imipenem/cilastatin with relebactam. However, β -lactam/ β -lactamase inhibitors are ineffective against MBLs (except aztreonam-avibactam) as well as *Acinetobacter baumannii* [108].

Moreover, study of collateral sensitivity patterns in CTX-M-15 β -lactamase revealed three non-synonymous mutations with increased resistance against mecillinam or piperacillin–tazobactam while confering full susceptibility to several cephalosporin drugs. Combination of mecillinam and cefotaxime eliminated both wild-type and resistant CTX-M-15. This result underlines how the rational design of drug combinations may limit the resistance evolution of horizontally transferred genes [169].

— Natural compounds — Mechanism of action of rhodomyrtone (Figure 12.D), an antibiotic isolated from the rose myrtle *Rhodomyrtus tomentosa* used in Asian traditional medicine, has been recently elucidated and revealed that rhodomyrtone traps membrane proteins in vesicles disrupting several cellular functions, including the respiratory chain and the ATP synthase complex. Given its uncharged non-amphipathic scaffold, rhodomyrtone is not a typical membrane-inserting molecule.

Molecular dynamics simulations revealed that it transiently binds to phospholipid head groups causing distortion of lipid packing, thus providing explanations for membrane fluidization and induction of membrane curvature [171].

Besides its anti-tumor effects, embelin (Figure 12.E), a plant-based benzoquinone which is the major active constituent of *Embelia ribes* Burm fruits, can restore meropenem activity against a panel of NDM-positive pathogens by interacting directly with the zinc ion through its hydroxyl group (proved only by molecular dynamics simulation) [146].

Darwinolide (Figure 12.F), a marine natural product from the Antarctic sponge *Dendrilla membranosa*, was shown to exhibit promising activity against the biofilm phase of methicillin-resistant *Staphylococcus aureus* [178].

Numerous other scaffolds are also under investigation, such as cyclobutanone [1], functionalized benzosiloxaboroles [61], or captopril-inspired mercapto-propionamide [136]. If you are interested in reading more about β -lactamase inhibitors, I recommend the following readings: Rotondo *et al.* [170] for MBLs inhibitors, Krajnc *et al.* [115] for cyclic boronates' dual activity, and Docquier *et al.* [56] for an updated review on β -lactamase inhibitors.

3.2 Experimental strategies

3.2.1 Inteins and pro-drugs

Inteins, also named protein introns, are protein sequences embedded into a host protein, termed extein, from which they are auto-catalytically excised in a process called protein splicing. During this, the intein ligates the extein extremities, for example with a peptide bond, and allows the reconstitution of the mature protein (Figure 13).

In a recent study, authors succeeded in engineering toxin–intein antimicrobials to selectively target and kill antibiotic-resistant *Vibrio cholerae* bacteria [124]. Results showed a total extermination of resistant strains, with a rate of escape mutants around 10^{6} - 10^{8} , in zebrafish and crustacean larvae which are natural hosts for *Vibrio* species.



Figure 13: The intein mechanism. Figure inspired from Fig. 1 in Mazel *et al.*, 2019 [124].

Another strategy consists in using intrinsic enzymes of bacteria, such as β -lactamase, to trigger the release of a specific antibiotic. For example, a synthetic hybrid including a siderophore, a cephalosporin and an oxazolidinone, has proved its efficiency on Gram-negative bacteria [123]. Indeed, the cephalosporin core is hydrolysed by a β -lactamase, releasing the oxazolidinone which can inhibit its target (Figure 14).

In normal treatment, Gram-negative bacteria are not affected by oxazolidinone, due to their strong outer membrane and efflux systems. So, this strategy revealed that it is possible to bypass these barriers using a pro-drug system.

3.2.2 Bacteriophages

A bacteriophage, or informally phage, is a virus able to infect bacteria and archaea by injection of its genome, so called prophage, in the cytoplasm, with the aim to replicate itself. One strategy to counter antibiotic resistance is to use temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria (ARBs). Besides performing a typically lytic life cycle, temperate bacteriophages are able to perform a lysogenic life cycle in which introduction of prophage does not lead to bacteria death (Figure 15).



Figure 14: β -lactamase as biological scissors to release a drug inside the cell. Here, an oxazolidinone (compound 3) is released after the action of a cephalosporinase on the cephalosporin core of the initial drug (compound 1). Graphic modified from Figure 1 in Miller *et al.*, 2018 [123].



Figure 15: Bacteriophages' lytic and lysogenic cycles.

This strategy has been already used to deliver a functional clustered regularly interspaced short palindromic repeats–CRISPR-associated (CRISPR-Cas) system into the ARBs' genome [208]. Results outlined that this system destroys both resistance-conferring plasmids and genetically modified lytic phages, revealing a suitable strategy to kill only ARBs while protecting antibiotic-sensitized bacteria. According to authors, such designed phages might be used on hospital surfaces or hand sanitizers to facilitate elimination of antibiotic-resistant pathogens.

Another study about carbapenem-resistant *Acinetobacter baumannii* was carried out using lytic bacteriophage SH-Ab15519, rescuing mice from lethal *A. baumannii* lung infection without harmful side effects [93].

3.2.3 Monoclonal antibodies

A reliable state-of-the-art has been done in a recent review about the current situation of monoclonal antibodies research for MDR infections [135]. Antibodies have two main mechanisms of action which include (1) the binding of bacterial surface-exposed antigenes, inducing lysis through either the formation of the membrane attack complex (MAC) or phagocytosis, and (2) the binding to bacterial exotoxins (Figure 16).



Figure 16: Antibodies' mechanisms of actions. (A) Antibodies carry a complement to activate the Membrane Attack Complex (MAC) leading to the cell lysis. (B) Antibodies may promote the phagocystosis process by bringing target bacteria closer to phagocytic cell effector. (C) Antibodies may block exotoxins' dissemination by catching them once released.

3.2.4 Environmental aspects

This strategy aims to eliminate both ARBs and antibiotics present in environmental places. Low-energy anaerobic–aerobic treatment reactors, constructed wetlands, disinfection (chlorination), coagulation, nanomaterials and biochar are efficient processes to remove them from wastewater treatment plants [15].

To understand the last three mechanisms, coagulation is an active method to remove colloidal particles in water and treat turbidity, color, natural organic matter, and heavy metals [209], given the negative and positive charges carried by colloidal particles and coagulants, respectively.

Regarding nanomaterials, two different mechanisms are possible: (1) toxic ions are released by a functionalized nanomaterial, right after its entry within the ARB, in combination with antibiotics, and (2) synergistic effects are obtained through the same combination but, this time, with an active effect from each partners.

Finally, biochar is active charcoal, derived from the pyrolysis of carbonrich biomass, containing rich mineral elements, large specific surface area [207], and pores in which contaminants may be sorbed.

3.3 Theoretical strategies

3.3.1 Mathematical model

Most of health authorities think antibiotic use is linearly related to resistance rates. However, theoretical and mathematical models stated that relationships between antibiotic use and resistance rates is not linear, based on the existence of fitness costs associated with resistance genes. Indeed, the latter may only gain a survival advantage in environments with an antibiotic selection pressure exceeding critical thresholds.

This way, some scientists identified minimum thresholds about useresistance relationships in several resistant strains from Europe [125]. Results provided a first-step in identification of context-specific targets to streamline the antibiotic use without over-restricting them.

3.3.2 Algorithm

An original strategy has been developed based on the idea that antimicrobial peptide design could be seen as a language in which antimicrobial peptides play the role of words and grammar refers to the frequency and placement of amino acids. In addition, this language would be made up of 20 letters, one for each natural amino acid. For a sequence of x_i amino acids, the language model would try to predict the probability distribution over the 20 amino acids for the next amino acid in the sequence x_i . A probability distribution function $P(x_i|x < i)$ is learned, where x < i refers to the sequence of residues before x_i (x_1 to $x_i - 1$). According to results of LSTM network, 10 peptides were synthesized and tested against known pathogens bringing to light broad-spectrum antimicrobial activities [142].

3.3.3 Detection

A significant number of kits (~40) have been designed for accurate detection of carbapenemases (e.g., NDM, VIM, IMP, KPC, OXA-48) using different approaches such as real-time Polymerase Chain Reaction (PCR) to amplify carbapenemase genes [91, 150], or Matrix Assisted Laser Desorption Ionization – Time Of Flight (MALDI–TOF) to detect either carbapenemase activity [11] or plasmid-encoded carbapenemases [46].

3.3.4 Prediction

My thesis work follows this strategy, namely the development of an *in silico* method to calculate the free energy cost of mutations in β -lactamases and assess which one of them could give survival advantages to bacteria, in order to further develop inhibitors which may counter them. Indeed, resistance needs new mutations beneficial to bacteria while conserving functional stability brought by important and critical interactions. Thus, it severely reduces the number of possible mutational combinations. Hereafter, examples of antibiotic resistance prediction attempt are reported.

Directed evolution has been qualified as a powerful tool to predict increased antibiotic resistance [148]. In this study, TEM-1 β -lactamase has been evolved using directed evolution with cefotaxime selection. In all cases, evolution resulted in the generation of a specific combination of mutations (500-fold increased resistance). This combination was identified as equivalent to clinical isolate TEM-52, whose crystallographic structure has been solved. Authors concluded that directed evolution combined with structural analysis can be efficient to predict future mutations which may boost antibiotic resistance.

Another fine study presents molecular modeling as an invaluable tool to predict allosteric mutants' effects [118]. First, molecular dynamics simulation on β -lactamase CTX-M-9 were computed and validated through experimental assays (hydrolysis kinetics, thermostability, antibiotic susceptibility, X-ray crystallography...). Results with purified enzymes outlined an increased catalytic rate and efficiency, while mutant crystal structures did not reveal apparent changes compared to wild-type. Then, machine-learning analyses highlighted changes in the binding-pocket conformational ensemble explaining the allosteric mutations' mechanism of action.

Also, genome mining could be an efficient tool to find new Gram-negative antibiotics. Two promising cationic nonribosomal peptides, named brevicidine and laterocidine, were found via analysis of 7395 bacterial genomes and displayed bactericidal activities against *Pseudomonas aeruginosa* and colistin-resistant *Escherichia coli* [160].

Lastly, a deep-learning approach, termed Deep-ARG, has been used to predict ARGs from metagenomic data. Given the 30 antibiotic resistance categories included, Deep-ARG models has successfully predicted ARGs with both high precision (> 0.97) and recall (> 0.90). Improvements are expected for under-represented ARG categories since more data become available and all predicted results has been reported in a database, so called DeepARG-DB [12].

3.4 Summary

In this last subchapter, we first presented antibiotic adjuvants use as the main strategy to counter antibiotic resistance. Those additional compounds, chemical or bio-mimetic, synthetic or natural, help to overcome the main defensive barriers of bacteria, such as the outer membrane and β -lactamases. Then, promising experimental strategies were introduced using biological or chemical tools (bacteriophages, antibodies, pro-drugs, nano materials...). Finally, some computational strategies were exposed with a growing buzz on machine learning techniques due to explosion of digital information.

Chapter II

New OPLS-AA force field parameters for zinc-containing enzymes

1 publication:

Elisée, E., and Iorga, B.I. (2019) OPLS-AA force field parameterization of zinc-coordinating residues in metalloenzymes. – manuscript submitted to J. Chem. Inf. Model..

Overview

1

In our team, we mainly run molecular dynamics (MD) simulations in the OPLS-AA force field for which there was no residue parameterized for coordinating the zinc ion. Indeed, preliminary MD simulations of the representative metallo- β -lactamase NDM-1 using existing OPLS-AA residues led to failed coordinations from most of the coordinating residues, except aspartate.

Otherwise, in a force field, one can implement metal coordination by means of several approaches presented in Figure 17. For simplicity, we selected the non-bonded model depending on electrostatics in order to (1) develop suitable parameters for computing MD simulations of metalloenzymes using a non-laborious approach and (2) assess the robustness of such approach.



Figure 17: Approaches for implementing a metal into a force field. (A) The bondedmodel requires parameterization of bonds, angles and dihedrals involving the metal M and coordinating atoms R_{1-4} . (B) The non-bonded model simply considers electrostatic interactions, like van der Waals and Coulomb, between M and R_{1-4} . (C) The cationic dummy atoms model uses cations, mimicking valence electrons of M, to bridge the latter and R_{1-4} .

In metallo- β -lactamases, the main coordinating residues are histidine, cysteine and aspartate/glutamate, but some of them also contain a coordinating glutamine residue. That's why, my first objective was the generation of new charges for these amino acids using quantum chemistry calculations and the Charge Model 5 (CM5). Subsequent validations have been carried out through MD simulations of representative zinc-containing metalloenzymes.

This way, after a short introduction about force fields, the development and validation of the new CM5 charges are reported in a publication-based section.

Force field analysis

Scientific research is mainly based on identification and deep understanding of which biological, chemical or physical process is involved before seeking out alternatives or strategies to solve the problem of interest. This way, one powerful computational method is to perform MD simulations, namely let the biomolecule evolve and analyze its behavior in an aqueous environment.

In MD, a molecule is defined as charged points (atoms) linked by springs (bonds). The ensemble of bond lengths, bond angles and torsions, but also non-bonding van der Waals and electrostatic interactions between atoms, represent a constitutive part of a force field. Thus, the latter can be considered as a set of equations and associated constants designed to reproduce molecular geometry and selected properties of training structures, and whose robustness will strongly affect the trustworthiness of results. There are currently various generations of force fields built on a simple base — detailed in this section — and displaying their own changes.

2.1 Emergence of force fields

Hereafter, a brief history about force fields' origin and a non-exhaustive list of well-known force fields are presented.

— **1930** — D. H. Andrews suggested to broaden the spectroscopic force field concepts to carry out molecular mechanics [109].

— **1946** — F. H. Westheimer, also known for his pioneering work in physical organic chemistry especially in the exploration of dehydrogenase enzymatic mechanism, and J. E. Mayer defined a method for calculating the activation energy for the racemization of optically active biphenyl derivatives [201].

— **1947** — F. H. Westheimer succeeded in manually calculating the transition state of a tetra-substituted biphenyl [200].

— 1948 — T. L. Hill spelled the Van der Waals potential energy curve out [89].
— 1961 — J. B. Hendrickson calculated the first molecular mechanics on cycloalkanes using a computer [87].

— **1965** — K. B. Wiberg laid the groundwork for the development of a molecular mechanics type program with ability to find energy minimum on elementary cycloalkane compounds [204].

— 1976 — N. L. Allinger published a proof of concept: the first Molecular Mechanics (MM) force field to make accurate predictions of molecular structures and properties, named MM1 [8].

— **1981** — P. K. Weiner developed the Assisted Model Building and Energy Refinement (AMBER) force field using a united-atom description, in other words, hydrogen atoms are not explicitly considered but are included into the neighbouring heavy atom [198]. In 1986, S. J. Weiner published the 'all-atom' AMBER version in which hydrogen atoms are explicitly described [199].

— **1983** — Karplus' group designed the Chemistry at HARvard Molecular Mechanics (CHARMM) force field [25].

— 1984 — Groningen Molecular Simulation (GROMOS) force field development by Berendsen et al. [20].

 — 1988 — Jorgensen's group conceived the Optimized Potential for Liquid Simulations (OPLS) force field [102].

— 2001 — A reactive force field, termed ReaxFF, was elaborated by A. C.
 T. van Duin et al. to study reactive chemical systems with thousands of atoms [189].

— **2004** — J. Wang *et al.* generalize the existing AMBER force field in a general AMBER force field (GAFF) [195]. The same year, Marrink *et al.* developed a coarse grained (CG) model for lipid simulations [128].

— **2007** — Extension of the lipid CG model to biomolecular simulations, now termed MARTINI, by Marrink *et al.* [129].

— 2013 — Polarizability was successfully considered in Atomic Multipole Optimized Energetics for Biomolecular Simulation (AMOEBA) polarizable multipole force field generated by Y. Shi *et al.* [177].

However, the use of a specific force field depends on the nature of your training system. In that respect, one should employ class I force fields (AMBER, CHARMM, OPLS) for studying dynamics of proteins or nucleic acids, class II force fields (MM*n* series, Merck Molecular Force Field (MMFF)) to get some properties of organic molecules and hydrocarbons (e.g., conformational energies, vibrational spectra...), and class III force fields (polarizable and reactive force fields) to overcome the limitations of the two previous classes and reproduce physical terms derived from quantum mechanics (QM), such as total electrostatic energy and charge transfer.

Indeed, classes I and II force fields are regarded as 'additive', namely individual terms of the potential energy do not have a full physical meaning, electric field modulations and effect of the environment on the charge distribution are not included — there is no many-body effects. Additionally, limitations include the transferability principle in which the parameterization accuracy is better for compound belonging to the same class but not transferable from a force field to another and above all, it is impossible to have a break or formation of chemical bonds.

Lastly, since their release, various changes were implemented in the previously listed force fields which are still undergoing continuous improvements. To further read about them and others, I recommend the following reviews: Ponder *et al.*, 2003 [159], Monticelli *et al.*, 2013 [138], Harrison *et al.*, 2018 [85], Jing *et al.*, 2019 [97].

2.2 Fundamental principles

Force fields are basically designed under either molecular mechanics (MM) or quantum mechanics (QM) principles. In MM method, classic mechanics is used to model molecular systems whose potential energy is described as a function allowing the variation of coordinates to find the energy minimum. Furthermore, by solving the classical equations of motion, one can explore the MD — in case of proteins, water may be implemented to reproduce an aqueous environment either explicitly (real hydrogen and oxygen atoms) or implicitly with a polarizable dielectric continuum.

In QM method, one looks for accurate treatment of the system by solving the Schrödinger equation with calculations involving higher computational costs. Moreover, there are three fundamental principles: the quantization, namely the transition from classical mechanics to quantum mechanics, the wave-particle duality, in which every quantum entity is better described by using both particles and waves definitions, and the uncertainty principle, stating that there is a limit on measurements' accuracy of each physical property (e.g., given an electron's position, its speed can't be measured as accurately as its position and vice versa) due to the description of electrons as waveforms.

To clarify, entities defined by MM are present in a specific place at a specific time, whereas entities defined by QM exist in a haze of probability with a certain likelihood to be at one point or another.

In hybrid method, such as QM/MM, it is possible to accurately describe a small part of the system, such as an active site of an enzyme, by QM, and consider the surrounding parts and the solvent by MM [3]. This method requires calculations with higher computational costs but results are often more accurate. For example, metalloenzymes are suitable candidates of QM/MM approach since you may more closely reproduce the ions' properties (e.g. charge transfer, polarizability). This method is, inter alia, developed in AMOEBA force field.

2.2.1 The Born-Oppenheimer approximation

The Born-Oppenheimer (BO) approximation defines the quantum states of molecules and simplifies the Schrödinger equation. Indeed, according to Erwin Schrödinger (1887–1961), an atom is made up of a nucleus around which electrons are moving with a certain uncertainty on their exact position. Furthermore, electrons move at a certain distance from the nucleus, depending on their energy level, and thus delimiting areas in which there is a higher probability to found them, termed atomic orbitals.

This way, the BO approximation allows the separation of nuclei's motion from electrons' motion. Physically, the mass of an atomic nucleus in a molecule is more than one thousand times larger than the mass of an electron $(9.109 \times 10^{-31}$ kilograms). Thus, the nucleus is slower than electrons and its motion may be neglected. Additionally, given the elementary charge, denoted by e and equal to $1.60217634 \times 10^{-19}$ coulomb (C), the negative charge carried by one electron is -e and the global positive charge due to protons and carried by nucleus is Ze (i.e., the atomic number times the elementary charge). It results an attractive force $-Ze^2/r^2$ between the nucleus and the electron, where r is the distance from the nucleus, acting on them and causing their respective acceleration. According to Newton's second law, the velocity of an object, with a constant mass m, may be changed by an external force F such as the latter equals mass times acceleration, termed a, namely $F = m \times a$. Thus, given that the magnitude of the acceleration is inversely proportional to the mass (a = F/m), the electrons' acceleration is two thousand larger than nucleus one.

Hence, this approximation describes the electronic states of a molecule by considering the nucleus as stationary, but in different discrete positions so the electronic wavefunction may depend on nucleus' positions even if its motion is neglected.

2.2.2 Functional form

Every force field describes the potential energy of a system (U) as depending on the internal coordinates (r) of atoms, by a typical function which may look like this:

$$U(r) = U_{bonded} + U_{non-bonded}$$

where U_{bonded} is the sum of internal terms describing bonds, angles and dihedrals, $U_{non-bonded}$ is the sum of external terms accounting for interactions between non-bonded atoms separated by three or more covalent bonds, and i, j, k, l are atoms of the system, such as:

$$U_{bonded} = \sum_{bonds} U_{ij}^{stretch} + \sum_{angles} U_{ijk}^{bending} + \sum_{dihedrals} U_{ijkl}^{torsion}$$
$$U_{non-bonded} = \sum_{pairs} U_{ij}^{vdW} + \sum_{pairs} U_{ij}^{Coulomb}$$

Mathematically expressed, we obtain these equations:

$$U_{ij}^{stretch} = \frac{1}{2}k_{ij}^{stretch}(r_{ij} - r_0)^2$$

where r_{ij} is the bond between atoms *i* and *j* with an ideal length r_0 and a force constant $k_{ij}^{stretch}$.

$$U_{ijk}^{bending} = \frac{1}{2} k_{ijk}^{bending} (\theta_{ijk} - \theta_0)^2$$

where θ_{ijk} is the angle defined by atoms *i*, *j* and *k* with an ideal bond angle θ_0 and a force constant $k_{ijk}^{bending}$.

$$U_{ijkl}^{torsion} = \frac{1}{2} V_n [1 + \cos(n\phi_{ijkl} - \delta_n)]$$

where ϕ_{ijkl} is the dihedral defined by atoms *i*, *j*, *k* and *l* with an n^{th} order force constant V_n and a phase angle δ_n for which $\delta_n = 0^\circ$ and $\delta_n = 180^\circ$ refer to *n* odd and *n* even, respectively.

$$U_{ij}^{vdW} = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$

where ϵ_{ij} is the van der Waals well depth and σ_{ij} is the inter-atomic distance at which $E_{ij}^{vdw} = 0$. Mixing rules defined $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$.

$$U_{ij}^{Coulomb} = k_C \frac{q_i q_j}{r_{ij}^2}$$

where q_i and q_j are atomic charges and r_{ij} is the inter-atomic distance. $k_C = 1/4\pi\epsilon_0 = 9 \times 10^9 N.m^2.C^{-2}$ is the Coulomb's constant defined according to the vacuum electric permittivity, or electric constant, ϵ_0 .

Existing force fields are built from this general energy function with some additional terms to better fit to the desired properties. In my thesis, I used OPLS-AA force field so let's have a look at its development over the years.

2.3 OPLS-AA force field

2.3.1 History

In the early 1980s, Jorgensen and co-workers developed new potentials to simulate liquid state properties (water and about 40 organic liquids), hence the name Optimized Potentials for Liquid Simulations (OPLS). Nonbonded interactions were initially derived by comparison to liquid-state thermodynamics [99] since first applications of OPLS potentials were to perform rigid-molecule Monte Carlo simulations of liquid hydrogen fluoride [98]. Additionally, reproducible results were obtained about densities and heats of vaporization, emphasizing the reliability of developed parameters. In these methods, also termed OPLS-united atoms (OPLS-UA), aliphatic hydrogens, i.e the most numerous, are included as part of an extended atom to diminish the total number of atoms in the system, while other hydrogens are explicitly considered.

First applications on proteins [101, 102, 185] were based on potentials using a polar-hydrogen–only representation, namely only hydrogens bonded to electronegative atoms like oxygen or nitrogen were displayed, and atom types and valence (bond, angle, dihedral) parameters taken from AMBER force field (1984), hence the name AMBER/OPLS force field.

Later, an all-atom version (OPLS-AA) was developed with the same philosophy regarding the derivation of charges and van der Waals parameters from simulations on pure liquids [105, 106, 164]. In this version, torsional parameters' development consisted in fitting to HF/6-31G* energy profiles [134], while bond stretching and angle bending terms were implemented with a wide inspiration from the AMBER all-atom force field (1986).

Unlike AMBER94 force field [47] in which charges are obtained on a case-by-case basis from fitting to electrostatic potential surfaces from *ab initio* calculations, one of the main features of OPLS philosophy is the transferability principle in which OPLS parameters are easily transferred from one molecule to others if they shared similar chemical groups. Although OPLS parameters were mainly derived from condensed phase simulations, reasonable outcomes were obtained by comparing to gas-phase peptide energetics [17].

Finally, in 2000s, OPLS-AA force field underwent several changes, from the implementation of the generalized Born/surface area (GB/SA) model [104, 161] to the recent development of parameters for RNA [165], and including, inter alia, improvements of peptide and protein torsional energetics [166] and establishement of the LigParGen web server as an automatic generator of OPLS-AA parameters for organic ligands [57].

2.3.2 OPLS-AA functional form

In OPLS-AA force field, certain equations previously described in paragraph 2.2.2 were modified [100]. No additional term was implemented in E_{bond} and E_{angle} :

$$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2$$
$$E_{angle} = \sum_{angles} K_{\theta} (\theta - \theta_{eq})^2$$

However, $E_{torsion}$ and $E_{non-bonded}$ were changed as follows:

$$E_{torsion} = \sum_{dihedrals} \frac{V_1}{2} [1 + \cos(\phi_i + f_i 1) + \frac{V_2}{2} [1 - \cos(2\phi_i + f_i 2) + \frac{V_3}{2} [1 + \cos(3\phi_i + f_i 3) + \frac{V_2}{2} (1 - \cos(2\phi_i + f_i 3) + \frac{V_3}{2} (1 - \cos(2\phi_i + f_i 3) + \frac{V_$$

where ϕ_i is the dihedral angle, V_1 , V_2 , V_3 are the Fourier's, and f_1 , f_2 , f_3 are phase angles.

$$E_{non-bonded} = E_{ab} = \sum_{i}^{on \ a \ on \ b} f_{ij} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \right]$$

where E_{ab} is the interaction energy between molecules a and b. The same expression is used for intramolecular non-bonded interactions between all pairs of atoms (i < j) separated by three or more bonds. An additional "fudge factor", termed f_{ij} , equals 1 except for intramolecular 1,4-interactions for which it equals 0.5.

2.4 Zn–OPLS publication

OPLS-AA force field parameterization of zinccoordinating residues in metalloenzymes

Eddy Elisée, Bogdan I. Iorga*

Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris-Saclay, LabEx LERMIT, 91198 Gif-sur-Yvette, France

KEYWORDS

OPLS-AA force field; metalloenzymes; zinc; molecular dynamics; force field parameterization

ABSTRACT

Metalloproteins are essential players in key biological processes and their study through molecular dynamics (MD) simulations is an important step in deciphering the underlying mechanisms or in designing new drugs. In this work, we used the non-bonded model approach to develop OPLS-AA parameters for several Zn-coordinating residues, that were missing from this force field: histidine (with two coordination modes, via N ϵ and N δ) with the names HMD and HME, respectively, cysteine with the name CYSM and asparagine/glutamine with the names ASM/GLM. Their use on di-zinc systems (with metallo- β -lactamases as representative examples) required an additional constraint on the distance between the two zinc ions to maintain the correct geometry. We used

extensive MD simulations on a diverse validation dataset to show that these parameters are general and applicable to all zinc metalloenzymes.

Introduction

Acting as cofactors, metal ions play an important role in so-called metalloproteins, which are involved in key biological processes such as respiration, photosynthesis or metabolism. Their presence may affect protein folding, binding modes of substrates or mechanisms for catalysis, by producing strong electrostatic interactions.¹⁻² Therefore, the current trend in both chemistry and biochemistry is to pinpoint the structural features responsible for the high efficiency of these enzymes in order to elucidate their complex mechanism or to successfully create artificial enzymes.³

In this context, computational techniques, like molecular dynamics (MD), are suitable to analyze the behavior of metalloenzyme systems, provided that an appropriate force field including metal ions parameters is used. If these metal ions parameters are missing, there are several approaches to implement them based on the use of either bonded or non-bonded models.⁴ However, it is not obvious to correctly parameterize metal ions which are not embedded in a molecular structure, like porphyrin ring, by considering both electrostatic and dynamic aspects. For example, for about forty years, hemoglobin has been widely studied through MD simulations⁵ assuming that the coordinative saturation of the heme iron atom leads to weak electrostatic interactions between the active site and the protein. In subsequent years, force field improvements (i.e. new potential function) were introduced in order to accurately model metal-ligand interactions in metalloproteins,⁶ using bonded,⁷⁻⁹ non-bonded¹⁰⁻¹¹ or cationic dummy models.¹²⁻¹⁴

More recently, the Bonded Plus Electrostatics Model¹⁵ was used to develop the Zinc AMBER Force Field (ZAFF)⁷ and its generalized version Extended Zinc AMBER Force Field (EZAFF).¹⁶
It involves setting covalent bonds between the zinc ion and its coordinating environment besides calculating point charges. Our group also used a similar bonded model to successfully parameterize the carbonic anhydrase active site⁸ in OPLS-AA force field.¹⁷⁻²⁰ As far as we are concerned, OPLS-AA has some advantages compared to existing force fields: it was developed with the objective of reproducing condensed-phase properties of model compounds, it is currently implemented in most MD software packages, and it has been shown by others²¹⁻²² and by us during our participation to the blind prediction challenges SAMPL²³⁻²⁵ and D3R²⁶⁻²⁷ to yield accurate results in free energy calculations.

An important topic of research in our group during the recent years was to better understand the antibiotic resistance related to β -lactams, especially through computational techniques such as covalent docking, MD simulations and free energy calculations.²⁸ We are mainly focused on enzymes, named β -lactamases, that are able to hydrolyze the β -lactam antibiotics.²⁹ According to the key residues in their binding sites, these enzymes can be classified as serine- and metallo- β -lactamases (SBLs and MBLs).³⁰⁻³² Suitable OPLS-AA parameters were available for running simulations with SBLs,²⁸ but the parameters for Zn-coordinating residues were missing, which made impossible the simulations on MBLs. In these conditions, we developed appropriate force field parameters for the Zn-coordinating residues using the OPLS-AA philosophy (especially about charge transferability)¹⁷⁻²⁰ and the most recent protocols for charge computation.³³⁻³⁴ Initial attempts were made using both bonded and non-bonded models. The non-bonded model was chosen for the final implementation, as it was easier to implement, showed good performance and was more general. In this paper we describe the protocol that we followed for the development of OPLS-AA force field parameters for Zn-coordinating residues and their validation using extensive

MD simulations, showing that these parameters are general and applicable to all zinc metalloenzymes.

Methods

Structures. We considered a total of seventeen representative zinc-containing metalloproteins, featuring mono- and di-zinc coordination systems (Table 1 and Figure 1). The dataset contains metallo- β -lactamases (NDM-1, CphA, L1, GOB-18, LRA-12), human metalloenzymes (MMP-8, Carbonic Anhydrase, ACE-2) and all metalloenzymes that were used for ZAFF development.⁷ All corresponding structures were chosen according to the quality of the atomic coordinates (e.g. high resolution, no missing atoms or residues, presence of representative water molecules in the active site). The zinc coordinating residues were manually renamed with their respective new implemented residue names for MD simulations.



Figure 1. Ribbon representation of the seventeen metalloproteins used for the validation of parameters developed in this study. **a**) Angiotensin-Converting Enzyme-2 (1r42); **b**) Carbonic Anhydrase II (4q49); **c**) Carbonic Anhydrase II (1ca2); **d**) Carbonic Anhydrase II Mutant H119Q (1h9q); **e**) Matrix MetalloProteinase 8 (2oy2); **f**) *E. coli* DNA-polymerase III (1a5t); **g**) Endonuclease I PpoI (1a73); **h**) Zinc finger protein ZIF268 (1a1f); **i**) Gelatinase A (1ck7); **j**) Thermolysin (113f); **k**) 1,3-1,4-β-endoglucanase (1u0a); **l**) Fibroblast stromelysin 1 (2usn); **m**) NDM-1 metallo-β-lactamase (4hl2); **n**) CphA metallo-β-lactamase (3f9o); **o**) L1 metallo-β-lactamase (5k0w).

Protein name	PDB	Resolution (Å)	Zn environment	System ^[a]
Angiotensin-Converting Enzyme- 2 (ACE-2)	1r42	2.20	HHDO	143,132
	4q49	1.80	НННО	64,169
Carbonic Anhydrase II (CAII)	1ca2	2.00	НННО	61,975
New Delhi Metallo-β-lactamase 1 (NDM-1)	4hl2	1.05	HHHO/HCDO ^[b]	54,069
CphA metallo-β-lactamase	3f9o	2.03	HCDO	54,061
L1 metallo-β-lactamase	1sml	1.70	HHHO/HHDO ^[b]	106,318
Matrix MetalloProteinase 8 (MMP-8)	2oy2	1.50	HHHD/HHHO ^[c]	34,171
E. coli DNA-polymerase III	1a5t	2.20	CCCC	39,238
Endonuclease I PpoI	1a73	1.80	CCCH	66,914
Zinc finger protein ZIF268	1a1f	2.10	ССНН	54,583
Gelatinase A	1ck7	2.80	СННН	105,553
Thermolysin	113f	2.30	HHDO/HOOO ^[b]	84,756
1,3-1,4-β-endoglucanase	1u0a	1.64	HHDD	119,077
Fibroblast stromelysin 1	2usn	2.20	HHHD	37,498
Carbonic Anhydrase (CA) H119Q	1h9q	2.20	HHQO	62,200
GOB-18 metallo-β-lactamase	5k0w	2.61	HHDO/HHQO ^[b]	85,358
LRA-12 metallo-β-lactamase	5aeb	2.10	HHDO/HHQO ^[b]	62,713

Table 1. List of PDB structures and features of the systems used in MD simulations.

^[a] Number of atoms in the system including protein, ions and water. ^[b] Di-zinc systems. ^[c] Two different Zn-containing binding sites.

55

Zinc environment. A previous PDB survey⁷ describes the main groups of zinc coordination spheres found in metalloenzymes, revealing various ligands like histidine, aspartate, glutamate, cysteine or water molecule. Similar residues are present within the active site of MBLs, with the notable additional presence of asparagine and glutamine. Therefore, for the calculation of initial charges we considered five simplified zinc complexes as models: methanethiolate for cysteine, imidazole for histidine, monodentate and bidentate acetate for aspartate/glutamate and acetamide for asparagine/glutamine (Figure 2).



Figure 2. Zinc complexes with simplified side chains for cysteine (**A**, methanethiolate, total charge: -2), histidine (**B**, imidazole, total charge: +2), aspartate/glutamate (**C**, monodentate acetate, total charge -2 and **D**, bidentate acetate, total charge: 0) and asparagine/glutamine (**E**, acetamide, total charge: +2) residues.

Charge calculations. The initial models were created using GaussView 5.0 (Semichem Inc, Shawnee Mission, KS, USA) by paying attention to respect symmetry in order to get equivalent charges. Geometry optimizations were performed using the B3LYP method³⁵ and the 6-31G(d) basis sets. Starting from these symmetrical optimized models, Charge Model 5 (CM5) charges³⁶ were calculated using the M06-2X method³⁷ and the 6-311+G(2df,2p) basis set. One of the main advantages of CM5 approach is its applicability to any charged or uncharged molecule containing any element of the periodic table, in the gas phase or in solution. Furthermore, recent calculations on a representative dataset of compounds in condensed-phase revealed an optimal scale factor of 1.20 for CM5 charges, that provides a good correction for different approaches in treatment of aqueous solvation (e.g. pure liquid, implicit or explicit water).³⁴ Therefore, CM5 charges were computed for our model complexes (Figure 2) using Gaussian 09, revision D.01³⁸ and were scaled with the factor 1.20.

MD simulations. Molecular dynamics simulations were performed using GROMACS version 4.6³⁹ with the OPLS-AA force field.¹⁷⁻²⁰ The protein is centered in a cubic periodic box, with at least 1.0 nm on each side. The simulation box is then filled with TIP4P water molecules⁴⁰ and the system is neutralized with Na+ and Cl- ions until reaching the physiological ionic strength (150 mM). All systems contain from 34,000 (MMP-8) to 140,000 (ACE-2) atoms including protein, water and ions (see Table 1 for the complete list of values). Energy minimization was then performed until convergence using a steepest descent algorithm to remove close contacts between water molecules and the protein. Equilibration step including position restraints on protein heavy atoms was then performed for 200 ps to relax the water molecules. Production steps were carried out running five replicas of 100 ns each for every validation system. During the equilibration and

production runs, the Nosé-Hoover⁴¹⁻⁴² and Parrinello-Rahman⁴³⁻⁴⁴ methods were respectively applied for temperature and pressure coupling. Electrostatics were calculated with the particle mesh Ewald method.⁴⁵⁻⁴⁶ The LINCS algorithm⁴⁷ was chosen to constrain all bond lengths, and the time step was set at 2 fs. All simulations were performed using periodic boundary conditions and the "isothermal-isobaric" ensemble (NPT), namely when temperature T, pressure P, and number of particles N are constant. Finally, Lennard–Jones and electrostatics interactions cut-offs were both set at 1.0 nm.

Trajectory analysis. Root-mean-square-deviation (RMSD) were calculated from trajectories using the Gromacs tool g_rmsdist³⁹ (20,000 frames per replica). If a structure contains several identical metal sites in terms of Zn coordination sphere, they will be all considered in the RMSD results. Furthermore, simulations were manually inspected with VMD⁴⁸ and Zn-Zn and Zn-X distances were computed using an in-house developed script.

Results and Discussion

We started our study by checking if the existing residues from the OPLS-AA force field are able to correctly describe the structure and the dynamical behavior of zinc metalloenzymes, with a special emphasis on the zinc coordination sphere. As expected, preliminary MD simulations (50 ns) of representative MBLs showed a loss of zinc-coordination from histidine and cysteine residues, but we were satisfied to observe that the aspartate maintained the coordination to the zinc ion throughout the simulation. Similar simulations with glutamine-containing systems showed that the coordination of this residue to zinc was stable only in a limited number of cases. These preliminary results prompted us to start the development of new parameters for several residues coordinated to zinc: histidine (with two coordination modes, via Nɛ and Nô) with the names HMD and HME, respectively, cysteine with the name CYSM and asparagine/glutamine with the names ASM/GLM.

Implementation of new residues. CM5 charges were computed on the simplified Zn complexes shown in Figure 2, following the protocol described previously.³³ The partial charge of Zn ion from these complexes was modified to be +2 (the formal charge for the zinc ion in OPLS-AA force field) and the difference between the initial (CM5) and new (OPLS-AA) charges, divided by four (zinc coordination number), was added to the zinc-coordinating atom from each residue. A similar protocol was followed to connect the coordinating group with the backbone, through the C α or C β atoms, depending on the system. The charges on the connecting atoms were adjusted to keep the overall charge as an integer. In the case of histidine, according to our previous results,²⁴ the charge was changed on the exocyclic C β atom. For monodentate and bidentate models of acetate, the charges of coordinating oxygen atoms were homogenized before the implementation of a unique model.

The charge computed using the CM5 method for carboxylate oxygen atoms in aspartate/glutamate residues (-0.7) was relatively close to the OPLS-AA value (-0.8). Considering that simulations of systems containing Zn-coordinated aspartate with the original OPLS-AA parameters are stable, with stable Zn-O distances and coordination maintained throughout the simulation, we decided to keep the original OPLS-AA parameters for the Zn-coordinated aspartate/glutamate residues.

Preliminary simulations of systems containing Zn-coordinated histidine and CM5 charges also proved to be very stable, with Zn-N distances within 0.05 Å compared with the crystal structures. Therefore, these parameters were selected for further validation. In contrast, the Zn-S distances in simulations of systems containing CYSM (with the CM5 charges) had a systematic shift (0.1-0.3 Å) compared to PDB value (Table 2). In an attempt to identify more appropriate charges, we carried out seven simulations with modified charges on SG and CB atoms in CYSM (Table 3). Two Zn environments (CCCC and CHHH) were evaluated, where C and H represent cysteine and histidine residues, respectively (Table 2). Several alternative sets of charges showed an improvement in the Zn-S distance for the CCCC environment, but not for the CHHH one, so there was no combination of charges that was appropriate simultaneously for both environments. Considering these results, as well as the inherent imprecision (about 10%) on interatomic distances and angles due to refinement methods⁴⁹ or crystal structure resolution,⁵⁰ prompted us to select as the best compromise our initial CM5 charges of CYSM residue for additional validation.

A similar protocol was applied to the new Zn-coordinating glutamine residue (GLM), with modified charges on the CD and OE1 atoms and using mono-and di-zinc proteins (Table 3). The simulations showed good Zn-O distances (less than 0.08 Å deviation compared with the crystal structure) for the mono-zinc enzyme (HHQO environment), and slightly less (0.2-0.3 Å deviation) for the di-zinc enzyme (HHQO/HHDO environment) (Table 2). Again, no significant improvement was observed from the variation of these charges, so we kept our initial CM5 charges for subsequent validation.

 Table 2. Mean errors on Zn-X distances during the optimization of CYSM and GLM residues (see

 Table 3 for the charges used in these calculations).

Residue	Zn environment	Structure (PDB code)	Mean errors (Å) ^[a]							
			CM5 ^[b]	Set 1 ^[c]	Set 2 ^[c]	Set 3 ^[c]	Set 4 ^[c]	Set 5 ^[c]	Set 6 ^[c]	Set 7 ^[c]
CYSM	CCCC (mono-zinc)	la5t	0.183	0.182	0.168	0.160	0.139	0.069	0.049	0.070
			0.112	0.100	0.091	0.079	0.053	-0.059	-0.066	0.018
			0.131	0.130	0.118	0.107	0.063	0.013	-0.017	0.012
			0.164	0.160	0.155	0.141	0.125	-0.010	-0.913	0.069
	CHHH (mono-zinc)	1ck7	0.307	0.255	0.241	0.279	0.220	0.304	0.290	0.266
GLM	HHQO (mono-zinc)	1h9q	-0.025	-0.034	-0.054	-0.057	-0.078	-0.087	0.001	
	HHQO/HHDO (di-zinc)	5k0w	0.284	0.277	0.261	0.243	0.240	0.220	0.309	

^[a] Errors were calculated as the difference between the crystallographic value and the simulation mean value. A positive value means a shorter Zn-X bond compared to crystallographic value, i.e. a stronger interaction. ^[b] Charges calculated using the CM5 method. ^[c] Exploratory sets of charges.

Table 3. Different sets of charges tested for the coordinating atoms in CYSM and GLM residues

 (see Table 2 for the results).

Residue	Atoms -	Charges							
		CM5 ^[a]	Set 1 ^[b]	Set 2 ^[b]	Set 3 ^[b]	Set 4 ^[b]	Set 5 ^[b]	Set 6 ^[b]	Set 7 ^[b]
CYSM	CB	-0.140	-0.220	-0.270	-0.320	-0.370	-0.420	-0.470	-0.520
	SG	-0.980	-0.900	-0.850	-0.800	-0.750	-0.700	-0.650	-0.600
GLM	CD	0.394	0.368	0.318	0.268	0.218	0.168	0.518	
	OE1	-0.826	-0.800	-0.750	-0.700	-0.650	-0.600	-0.950	

^[a] Charges calculated using the CM5 method. ^[b] Exploratory sets of charges.

A comparison between existing OPLS-AA residues and those developed in this study is presented in Figure 3.



Figure 3. Comparison between existing and new residues as implemented in OPLS-AA force field. Charge modifications for the side chains of following residues: glutamine GLN/GLM (A), histidine HISE/HME with protonation on N ϵ (B), cysteine CYSH/CYSM (C) and histidine HISD/HMD with protonation on N δ (D). Existing OPLS-AA charges are colored in black whereas the new ones are colored in red. Charges for the corresponding hydrogen atoms are enclosed in square brackets. A black value with no associated red one means that the charge is identical in both old and new residues.

Zn-Zn distance optimization. During simulations involving di-zinc systems, the distance between zinc ions was observed to increase to about 7 Å, instead of 3-4 Å in crystal structures,

due to electrostatic repulsion. To fix this issue, we first evaluated the Zn-Zn distances distribution of all di-zinc MBL structures present in the PDB (Figure 4), using an in-house script. The great majority of these structures show Zn-Zn bond length between 3.5 and 3.8 Å. Three of them were selected as representative structures of the minimum, mean and maximum Zn-Zn values: 1a8t (*Bacteroides fragilis* MBL, Zn-Zn distance: 2.84 Å), 5ewa (IMP-1 MBL, Zn-Zn distance: 3.71 Å) and 2nyp (*Bacillus cereus* II MBL, Zn-Zn distance: 4.73 Å).



Figure 4. Zn-Zn bond length distribution among MBL structures present in the PDB (335 chains)

We introduced an artificial bond between the two zinc ions and on the three structures selected above we tested a wide range of force constants (k_b) for the Zn-Zn bond in 5 ns MD simulations, with twelve values ranging from 700,000 to 25 kJ.mol⁻¹.nm⁻². Through these simulations we wanted to test if the corresponding crystallographic value could be reached starting from each representative structure (1a8t, 5ewa, 2nyp) with the same equilibrium bond length (b_0) set to 3.705 Å. We initially used the Gromacs default "[bonds] type 3", but simulations crashed starting for k_b values smaller than 2500 kJ.mol⁻¹.nm⁻², so we switched to a harmonic interaction, which is "[bonds] type 6" in Gromacs, that does not connect atoms by a chemical bond. The results are presented in Figure 4. In all three systems, constants higher than 25,000 kJ.mol⁻¹.nm⁻² maintain the Zn-Zn distance around the equilibrium distance with low fluctuation. Lower values increase the fluctuation and also the Zn-Zn distance, especially for the 1a8t structure, where in fact we would need to diminish it. In these conditions, we set up the following protocol that was used for all subsequent simulations with MBLs: i) b_0 is set to the Zn-Zn bond length taken from the crystallographic input structure and ii) k_b is set to a fixed value of 200,000 kJ.mol⁻¹.nm⁻², which is strong enough to maintain the Zn-Zn distance around the crystallographic value.



Figure 4. Zn-Zn distance evaluation on three representative systems using several force constants in 5 ns MD simulations: 1A8T (*Bacteroides fragilis* MBL, Zn-Zn distance in crystal structure: 2.844 Å), 5EWA (IMP-1 MBL, Zn-Zn distance in crystal structure: 3.705 Å) and 2NYP (*Bacillus cereus* II MBL, Zn-Zn distance in crystal structure: 4.731 Å). The red line indicates the Zn-Zn bond length in the PDB structure. From the top down, horizontal lines refer to the maximum, mean and minimum values of the Zn-Zn distance per training system, respectively.

Validation of new parameters. MD simulations using seventeen representative metalloenzymes (Figure 1 and Table 1) were performed to assess the validity of our new OPLS-AA parameters developed for Zn-coordinating residues. We ran five replicas of 100 ns simulation for each metalloprotein, in order to ensure reproducibility.⁵¹

RMSD values for the atoms constituting the metal sites and the protein backbone were calculated for every validation system in order to evaluate their steadiness. Globally, these plots show a good stability of the system during the 100 ns MD simulations, with values of around 1 Å (or less) for metal sites, as reported in Figure S1. Some proteins display a clear RMSD profile, such as 4q49, 1ca2 and 113f, while others underwent slight conformational changes in the metal site, described by the common RMSD evolution of the Zn-coordinating atoms and their respective residues. Moreover, in 1a5t and 3f9o, Zn-coordinating atoms are really stable and only their belonging residue move slightly. An outsider case is 1a1f, a Zinc finger protein initially crystallized with DNA double-strand (not modeled in this study) whose MD simulations revealed high RMSD values (up to 10 Å). In the first simulation (md1), the protein quickly folded in a close conformation similar to a ring, and most of the RMSD result comes from ZN201 and ZN203 contributions which are located in the metal sites at the extremities of the protein. In the second, third and fourth simulations (md2, md3 and md4, respectively), a structural rearrangement takes place, yielding an overall "helix-shaped" conformation. In the fifth simulation (md5), the closing process occurred at the beginning, but some electrostatic repulsions forced the system to go back near the starting conformation.

The distances between zinc ions and coordinating atoms were measured and mean values from all 85 simulations are reported in Table 4, showing really good results, with an absolute mean error of 0.058 Å on Zn-N distance for HME/HMD residues, 0.182 Å on Zn-S distance for CYSM residue and 0.158 on Zn-O distance for GLM residue. Furthermore, we obtained reasonable results (absolute mean errors of 0.105 Å and 0.276 Å, respectively) with the original OPLS-AA parameters for aspartate and glutamate residues. Lastly, the constraint applied to Zn-Zn distance in every di-zinc system led to an absolute mean error of only 0.051 Å. All plots are shown in Figure S2, together with a comparison with the corresponding value from the PDB structure. Finally, it is worth noting that the conformational changes that were observed to occur during the validation simulations in most cases did not disturb the Zn-X distances, thus underlining the robustness of this non-bonded approach and of our parameters.

	Mean errors (Å) ^[a]						
Residue	HMD/HME	CYSM	ASP	GLM	ZN		
Coordinating atom (X)	NE2/ND1	SG	OD1/OD2	OE1	ZN		
Non-absolute	-0.035	0.174	0.042	0.031	-0.051		
Absolute	0.058	0.182	0.105	0.158	0.051		
Ν	71	19	12	3	8		

Table 4. Mean errors on Zn-X distances from the MD validation simulations.

^[a] Errors calculated as the sum of either absolute or non-absolute differences between the crystallographic value and the simulation mean value divided by N, the number of bond instances.

We are aware of the importance of polarization in metal-containing models, which allows a better description of electrostatic and energetic properties. Polarizable force fields, like AMOEBA,⁵² or quantum mechanics/molecular mechanics (QM/MM) approaches represent interesting alternatives for the simulation of these systems, but requiring a significantly higher computational cost. In these conditions, the parameters for Zn-coordinating residues that we have developed in this work should be useful for the whole community, allowing the study of systems containing zinc metalloproteins through MD simulations using OPLS-AA force field.

Conclusion

In this work, we developed new OPLS-AA force field parameters for three new residues (HMD/HME, CYSM and ASM/GLM) using quantum chemistry calculations. These parameters were validated by molecular dynamics simulations on seventeen representative metalloenzymes with different zinc coordination spheres. These simulations showed that the non-bonded model is suitable for simulating zinc-containing metalloproteins using the OPLS-AA force field, including those containing di-zinc systems like metallo-β-lactamases, and may represent an alternative to the bonded model, which requires a very complex parameterization process. Lastly, this method could be either extended to metalloproteins containing other divalent metals (magnesium, calcium, iron) or used to implement others amino acids which may coordinate metal ions.

ASSOCIATED CONTENT

Supporting Information. Supporting Information file contains the RMSD plots for protein backbone, Zn ions, and Zn-coordinating atoms, as well as the Zn-X distances during the validation MD trajectories. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

PhD fellowship from the Université Paris-Saclay to E.E. is gratefully acknowledged. B.I.I.'s laboratory is member of the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT), supported by a grant from French National Research Agency (ANR-10-LABX-33).

REFERENCES

1. Lehninger, A. L., Role of metal ions in enzyme systems. *Physiol Rev* 1950, 30, 393-429.

2. Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A., Biomolecular electrostatics and solvation: a computational perspective. *Q Rev Biophys* **2012**, *45*, 427-491. Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny,
 S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L., Protein design: toward functional metalloenzymes. *Chem Rev* 2014, *114*, 3495-3578.

4. Li, P.; Merz, K. M., Jr., Metal Ion Modeling Using Classical Mechanics. *Chem Rev* 2017, *117*, 1564-1686.

5. Case, D. A.; Karplus, M., Dynamics of ligand binding to heme proteins. *J Mol Biol* **1979**, *132*, 343-368.

6. Vedani, A.; Huhta, D. W., A new force field for modeling metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759-4767.

7. Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M., Jr., Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J Chem Theory Comput* **2010**, *6*, 2935-2947.

8. Bernadat, G.; Supuran, C. T.; Iorga, B. I., Carbonic anhydrase binding site parameterization in OPLS-AA force field. *Bioorg. Med. Chem.* **2013**, *21*, 1427-1430.

9. Roberts, B. P.; Miller, B. R., 3rd; Roitberg, A. E.; Merz, K. M., Jr., Wide-open flaps are key to urease activity. *J Am Chem Soc* **2012**, *134*, 9934-9937.

10. Stote, R. H.; Karplus, M., Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins* **1995**, *23*, 12-31.

11. Li, P.; Merz, K. M., Jr., Taking into Account the Ion-induced Dipole Interaction in the Nonbonded Model of Ions. *J Chem Theory Comput* **2014**, *10*, 289-297.

22

12. Pang, Y. P., Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **2001**, *45*, 183-189.

13. Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C., Force field independent metal parameters using a nonbonded dummy model. *J Phys Chem B* **2014**, *118*, 4351-4362.

14. Saxena, A.; Sept, D., Multisite Ion Models That Improve Coordination and Free Energy Calculations in Molecular Dynamics Simulations. *J Chem Theory Comput* **2013**, *9*, 3538-3542.

15. Hoops, S. C.; Anderson, K. W.; Merz, K. M., Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262-8270.

16. Yu, Z.; Li, P.; Merz, K. M., Jr., Extended Zinc AMBER Force Field (EZAFF). *J Chem Theory Comput* **2018**, *14*, 242-254.

17. Jorgensen, W. L.; Tirado-Rives, J., The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* **1988**, *110*, 1657-1666.

18. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

19. Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L., Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. *J Chem Theory Comput* **2015**, *11*, 3499-3509.

20. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.

21. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S., Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740-5761.

22. Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W., Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput* **2010**, *6*, 1509-1519.

23. Beckstein, O.; Iorga, B. I., Prediction of hydration free energies for aliphatic and aromatic chloro derivatives using molecular dynamics simulations with the OPLS-AA force field. *J. Comput. Aided Mol. Des.* **2012**, *26*, 635-645.

24. Beckstein, O.; Fourrier, A.; Iorga, B. I., Prediction of hydration free energies for the SAMPL4 diverse set of compounds using molecular dynamics simulations with the OPLS-AA force field. *J. Comput. Aided Mol. Des.* **2014**, *28*, 265-276.

25. Kenney, I. M.; Beckstein, O.; Iorga, B. I., Prediction of cyclohexane-water distribution coefficients for the SAMPL5 data set using molecular dynamics simulations with the OPLS-AA force field. *J. Comput. Aided Mol. Des.* **2016**, *30*, 1045-1058.

26. Selwa, E.; Elisée, E.; Zavala, A.; Iorga, B. I., Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations. *J. Comput. Aided Mol. Des.* **2018**, *32*, 273-286.

27. Chaput, L.; Selwa, E.; Elisée, E.; Iorga, B. I., Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations. *J. Comput. Aided Mol. Des.* **2019**, *33*, 93-103.

28. Zavala, A.; Retailleau, P.; Elisée, E.; Iorga, B. I.; Naas, T., Genetic, biochemical and structural characterization of CMY-136 beta-lactamase, a peculiar CMY-2 variant. ACS Infect. Dis. 2019, 5, 528-538.

29. Naas, T.; Oueslati, S.; Bonnin, R. A.; Dabos, M. L.; Zavala, A.; Dortet, L.; Retailleau, P.; Iorga, B. I., Beta-Lactamase DataBase (BLDB) – Structure and function. *J. Enzyme Inhib. Med. Chem.* **2017**, *32*, 917-919.

30. Bonomo, R. A., beta-Lactamases: A Focus on Current Challenges. *Cold Spring Harb Perspect Med* **2017**, *7*, 1-15.

31. Bush, K., Past and Present Perspectives on beta-Lactamases. *Antimicrob Agents Chemother* **2018**, 62, e01076-18.

32. Bush, K.; Bradford, P. A., Interplay between beta-lactamases and new beta-lactamase inhibitors. *Nat Rev Microbiol* **2019**, *17*, 295-306.

33. Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation of CM5 Charges for Condensed-Phase Modeling. *J Chem Theory Comput* **2014**, *10*, 2802-2812.

34. Dodda, L. S.; Vilseck, J. Z.; Cutrona, K. J.; Jorgensen, W. L., Evaluation of CM5 Charges for Nonaqueous Condensed-Phase Modeling. *J Chem Theory Comput* **2015**, *11*, 4273-4282.

35. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 1993, 98, 5648-5652.

36. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G., Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J Chem Theory Comput* **2012**, *8*, 527-541.

37. Zhao, Y.; Truhlar, D. G., Density functionals with broad applicability in chemistry. *Acc Chem Res* **2008**, *41*, 157-167.

38. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision D.1, Gaussian Inc., Wallingford CT, USA (http://www.gaussian.com).

39. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-

26

throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845-854.

40. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.

41. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods.*J. Chem. Phys.* **1984**, *81*, 511-519.

42. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.

43. Parrinello, M.; Rahman, A., Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196-1199.

44. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

45. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

46. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.

47. Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 116-122. 48. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph.***1996**, *14*, 33-38.

49. Laskowski, R. A.; Moss, D. S.; Thornton, J. M., Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **1993**, *231*, 1049-1067.

50. Laitaoja, M.; Valjakka, J.; Janis, J., Zinc coordination spheres in protein structures. *Inorg Chem* **2013**, *52*, 10983-10991.

51. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J Chem Theory Comput* **2018**, *14*, 6127-6138.

52. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J Chem Theory Comput* **2013**, *9*, 4046-4063.

Supporting Information

OPLS-AA force field parameterization of zinccoordinating residues in metalloenzymes

Eddy Elisée, Bogdan I. Iorga*

Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris-Saclay, LabEx LERMIT, 91198 Gif-sur-Yvette, France

Corresponding Author

[•] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Table of contents

Figure S1. RMSD plots for protein backbone, Zn ions, and Zn-coordinating atoms	S2
Figure S2. Zn-X distances during the validation MD trajectories	S 7



Figure S1. RMSD plots for protein backbone, Zn ions, and Zn-coordinating atoms (black, red and blue, respectively) during the 100 ns validation trajectories for the seventeen metalloproteins studied.



Figure S1. (continued)



Figure S1. (continued)



Figure S1. (continued)

S5



Figure S1. (continued)



Figure S2. Evolution of Zn-X distances during the 100 ns validation trajectories for the seventeen metalloproteins studied (one color per replica). Corresponding values extracted from PDB survey or ZAFF benchmark are represented by plain and dashed lines, respectively.



Figure S2. (continued)



Figure S2. (continued)



Figure S2. (continued)


Figure S2. (continued)

Chapter III

Molecular dynamics (MD) studies of β -lactamases in OPLS-AA force field

1 publication:

Zavala, A., Retailleau, P., **Elisée, E.**, Iorga, B. I., and Naas, T. (2019) Genetic, biochemical and structural characterization of CMY-136 beta-lactamase, a peculiar CMY-2 variant, *ACS Infect. Dis.* 5, 528-538 DOI: 10.1021/acsinfecdis.8b00240.

Overview

1

Several metalloenzymes have been already studied during the validation work on zinc ion parameters, yielding reliable results (see Chapter II). Thus, the aim of this chapter is to present some published or ongoing side projects based on MD simulations, including free energy calculations, of serine- and metallo- β -lactamases.

First, the in-house software MOL2FF, for automatically assigning OPLS-AA parameters to covalent and free ligands, is detailed. During my PhD work, I had parameterized numerous ligands, for example, β -lactam drugs (this chapter), peptide-like ligands (confidential data) or D3R inhibitors (Chapter IV). Furthermore, some β -lactam antibiotics, presented hereafter in a publication, were used in covalent docking studies and MD simulations of CMY-136 (SBL, class C).

Secondly, two easy-to-parameterize inhibitors of IMP-1 (MBL, class B1) have been tested on NDM-1 (MBL, class B1), through MD simulations, to assess the robustness of our new metal-containing active site parameters (see Chapter II) in presence of ligands.

Thirdly, MD analysis of an interesting SPM-1 histidine-to-glycine mutant (MBL, class B1) is presented [21]. Indeed, while this mutation takes place on one of the active site zinc-coordinating histidine, β -lactamase activity is maintained. Our assumption was based on a new zinc coordination by an aspartate residue located on a loop in the proximity of the mutation site. Nevertheless, a recent crystallographic structure of this mutant tells quite another story [43].

Lastly, free energy calculations were performed, using a pmx-based protocol, in order to pursue the questions about protonation and carboxylation processes in OXA enzymes (SBL, class D).

Parameterization of covalent ligands in OPLS-AA force field

2

2.1 MOL2FF software

2.1.1 Philosophy of OPLS-AA parameterization

In Chapter II, I have described the OPLS-AA force field by mentioning, inter alia, the concept of transferability as a core feature for parameterizing ligands. Indeed, one may summarize the philosophy of OPLS-AA parameterization as follows: the charge of an atom is defined by its neighbouring atoms, for example an oxygen atom will not have the same charge in a carboxylic acid as in an alcohol. Then, atoms are gathered into elementary building blocks, i.e. chemical groups, that most of the time have integer charges and, finally, these blocks can be fused to create bigger molecules but taking care to adjust the charge of atoms at the blocks' junction – that is considering the influence of the vicinity (Figure 18).



Figure 18: Atoms' charge adjustement in OPLS-AA force field. Charge on the red carbon atom is the sum of the carbon charge in a CH_2 group (-0.120) plus the influence of secondary amine (+0.200) and thioether (+0.168) groups.

Otherwise, OPLS-AA force field displays more than 900 atom types, which is not enough for covering all diversity of chemical scaffolds. Table 1 gives some examples of OPLS atom types in which *ptype* is the particule type (A stands for Atoms), and *epsilon*, *sigma* are van der Waals parameters (see Chapter II Section 2.2.2).

This way, if the queried chemical group is not parameterized, it will be necessary to implement it either by (1) combining two existing chemical groups and adjusting the total charge, or (2) computing new charges using quantum chemistry calculations before implementing new atom types. The same principle is applicable to other parameters such as bonds, angles, dihedrals and impropers – even though most of them are already described. However, it is important to be aware of the complexity of parameterizing bonded parameters through quantum chemistry calculations, and sometimes it would be preferable to adapt existing atom types described in the force field rather than try to develop from scratch new parameters.

Table 1: Examples of atom types in OPLS-AA force field. OPLS-AA philosophy: given the charge +0.060 for the aliphatic hydrogen H, in order to build neutral groups, the carbon charge in CH group must be –0.060, in CH2 group –0.120, etc.

1	2	3	4	5	6	7	8	9		10
opls_135	СТ	6	12.01100	-0.180	А	3.50000e-01	2.76144e-01	12.01100	;	alkane CH3
opls_136	СТ	6	12.01100	-0.120	А	3.50000e-01	2.76144e-01	12.01100	;	alkane CH2
opls_137	СТ	6	12.01100	-0.060	А	3.50000e-01	2.76144e-01	12.01100	;	alkane CH
opls_138	СТ	6	12.01100	-0.240	А	3.50000e-01	2.76144e-01	12.01100	;	alkane CH4
opls_139	СТ	6	12.01100	0.000	А	3.50000e-01	2.76144e-01	12.01100	;	alkane C
opls_140	HC	1	1.00800	0.060	А	2.50000e-01	1.25520e-01	1.00800	;	alkane H
opls_145	CA	6	12.01100	-0.115	А	3.55000e-01	2.92880e-01	12.01100	;	Benzene C
opls_146	HA	1	1.00800	0.115	А	2.42000e-01	1.25520e-01	1.00800	;	Benzene H

1: opls type; 2: atom type; 3: atomic number; 4, 9: atomic mass; 5: charge; 6: ptype; 7: epsilon; 8: sigma; 10: comments.

This way, an algorithm termed MOL2FF was developed in our team during the last years to handle the automatic parameterization of ligands in OPLS-AA force field. MOL2FF uses the CACTVS Chemoinformatics Toolkit [95] for processing the chemical structures.

2.1.2 MOL2FF algorithm

The aforesaid algorithm MOL2FF was initially designed as a fast and powerful tool to automatically parameterize ligands in OPLS-AA force field.

First, the ligand must be provided by user in MOL2 format since this format keeps structural information like bond multiplicity (simple, double or triple) or atom types. Note however that three-dimensional coordinates are not modified, therefore the user needs to supply the desired pose for MD simulation – often obtained after a docking step.

Secondly, the ligand structure is analyzed and converted into a Simplified Molecular-Input Line-Entry System (SMILES) format, namely the ligand is now described by a line notation using short American Standard Code for Information Interchange (ASCII) characters (Figure 19). Indeed, the ligand is split into one main backbone and sidechains – if some rings are present, they are opened and new terminal atoms are equally numbered. Then, the SMILES format is linearly written under the longer fragment, so-called backbone, and sidechains are bracketed. With that kind of exported format, it is straightforward for some software to translate it into a two-dimensional drawing or a three-dimensional model with random coordinates.



Figure 19: SMILES description. The backbone (in green) and sidechains are identified in the SMILES pattern, located in the lower-right corner, according to their respective color. Image modified from Wikimedia Commons website [https://commons.wikimedia.org/wiki/File:SMILES.png].

Thirdly, the ligand is divided into substructural patterns, called SMILES Arbitrary Target Specification (SMARTS), easily recognised and handled by the Tool Command Language (Tcl) used by MOL2FF. Furthermore, SMARTS language is extremely precise on substructural specification and atom typing (Table 2). This way, an in-house database, gathering a number of OPLS-AA substructures in SMARTS format, has been built and in which MOL2FF algorithm looks for each chemical fragment obtained from the input structure. If there is a match, all related parameters are added to the final OPLS-AA ligand topology. If not, user has to implement a new substructure by following the instructions presented in the Section 2.1.1.

Finally, a complete ligand topology in OPLS-AA force field is built and ready to be validated in a MD simulation step carried out in explicit solvent. Once validated, the ligand topology may be implemented in Ligandbook, an online repository compiling force field parameters for small and drug-like molecules [59]. In my project, I used to employ MOL2FF for parameterizing ligands including covalent ones whose process is presented in the next section.

Furthermore, I had to fix some bugs concealed in the script: atoms involved in a covalent bond were not correctly deleted (see section 2.2.1) and charge modification was not possible on certain atoms shared by two chemical groups – this is what we call a conflict. However, further improvements should be undertaken in order to restore some basic functionalities, ease the implementation process of new parameters and better manage conflicts. Figure 20 below sums up the previously described MOL2FF workflow.

Table 2: Some features of SMARTS language.For complete description,tion,please consult the Daylight's SMARTS theory manual.[https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html].

Symbol name	Examples
aliphatic	С
aromatic	с
atomic number	[#6] carbon
anticlockwise chirality	C@H
clockwise chirality	C@@H
single bond (aliphatic)	C-C
double bond	C=O
triple bond	$C \equiv N$
aromatic bond	c:c
	Symbol name aliphatic aromatic atomic number anticlockwise chirality clockwise chirality single bond (aliphatic) double bond triple bond aromatic bond



Figure 20: MOL2FF algorithm workflow.

2.1.3 LigParGen service

Since 2017, Jorgensen's group has released the LigParGen website [57] as an automatic generator of OPLS-AA parameters for organic molecules or ligands (Figure 21).

OPLS/CM1A Parameter Ge	arGen nerator for Organic Ligands
LigParGen is a web-based service that provides force field (FF) parameters for organic molecules or ligands, offered by the Jorgensen group. LigParGen provides bond, angle, dihedral, and Lennard-Jones OPLS-AA	Step 1: Input structure
parameters with 1.14*CM1A or 1.14*CM1A-LBCC partial atomic charges.	Enter SMILES Code
Server provides parameter and topology files for commonly used molecular dynamics and Monte Carlo packages OpenMM, Gromacs, NAMD, CHARMM, LAMMPS, TINKER, CNS/X-PLOR, Q, DESMOND, BOSS and MCPRO. Also, the PQR file is generated.	OR upload MOL / PDB file (Structures MUST include all hydrogens)
Supported input formats: SMILES, MOL and PDB.	
Maximum ligand size allowed is 200 atoms.	Step 2: Options
Check this link to use LigParGen software from command-line in your local computer.	Molecule Optimization Iterations 0 ;
Please, report any issue clicking in the following link: LigParGen issues	Select charge model:
	C 1.14*CM1A-LBCC (Neutral molecules)
2,298 Pageviews	© 1.14*CM1A ¹ (Neutral or Charged molecules)
ישור. _ב עור מוד .	Molecule charge 0 \$
	Submit Molecule Sample Benzene
	¹ For charged molecules, CM1A charges are NOT scaled by a factor 1.14

Figure 21: LigParGenintuitivewebinterface[http://zarbi.chem.yale.edu/ligpargen/].

By providing a SMILES or MOL2 input, OPLS-AA bonded (bonds, angles and dihedrals) and non-bonded (Lennard-Jones) are computed by the BOSS software [103], also developed by Jorgensen's group, by analogy to the existing OPLS-AA atom types. Regarding the partial atomic charges, corrected Charge Model 1 A (CM1A) method is applied with a scale factor of 1.14 for both neutral and charged molecules. In addition, a Localized Bond-Charge Corrected (LBCC) method may be chosen for neutral molecules; this method proved to significantly reduce errors on hydration free energies [58].

Nevertheless, given the optimized, but inevitable, charge calculation step, LigParGen no longer follows the OPLS-AA philosophy about the transferability concept, thus coming back to the AMBER approach. In addition to this, new parameters, developed for the OPLS3 force field in collaboration with Schrödinger [84], are not available for open access unless one upload a ligand structure on LigParGen website and get the outcome.

However, it is interesting to parameterize a new ligand, for which one or more chemical groups can not be defined with existing OPLS-AA parameters, using LigParGen software, to possibly avoid calculating new parameters via quantum chemistry. However, LigParGen may sometimes yield incorrect parameters which need to be carefully reviewed.

2.2 Covalent parameterization

As previously mentioned in the Chapter I Section 2.3, β -lactamase enzymes include metallo- β -lactamases (MBLs) and serine- β -lactamases (SBLs). The latter perform antibiotic hydrolysis through a serine-dependent mechanism which leads to the formation of a covalent acyl-enzyme intermediate. In order to study such a complex by molecular dynamics (MD), preliminary docking has to be conducted to obtain suitable starting poses (the covalent docking technique will not be described in detail in this manuscript, see [181] for a recent review). Once obtained, the pose has to be parameterized before running MD simulations; this step is based on the parameterization of covalent ligands whose method is detailed hereafter.

2.2.1 Implementation in MOL2FF

Figure 22 shows the global idea of how covalent ligand parameterization works using MOL2FF software.



Figure 22: Covalent β -lactam parameterization using MOL2FF software. SEC stands for covalent serine which fused to the ligand constitute an acyl-complex with an integer charge, here zero. Charge unit is the elementary charge (e) and equals $1.602176565 \times 10^{-19}$ Coulomb (C). First, the ligand is prepared by removing atoms which will not belong to the future covalent bond, for example an hydroxyl group in the case of β -lactam antibiotics, resulting in an increase of the initial ligand charge by +0.080 e. Simultaneously, a covalent serine residue, termed SEC in OPLS-AA force field, was parameterized by deleting the hydrogen from the sidechain hydroxyl group resulting in a residue charge of –0.080 e. This way, when the protein-ligand complex is prepared for MD simulation, the total charge will be an offset integer.

2.2.2 Parameterized antibiotics

Figure 23 reports all β -lactam antibiotics that I have parameterized during my PhD work. Some of them, like cefuroxime, cefotaxime and ceftazidime, were used in MD simulations of CMY-136 (see next section 2.3).



Figure 23: Parameterized β-lactam antibiotics. Closed β-lactam rings are presented here but hydrolysed and covalent forms have been equally parameterized. (a): ampicillin, (b): temocillin, (c): imipenem, (d): meropenem, (e): oxacillin, (f): cloxacillin, (g): aztreonam, (h): cefotaxime, (i): cephalotin, (j): ceftazidime, (k): cefuroxime.

2.3 Publication

Page 1 of 34

ACS Infectious Diseases

1	
2	Constinuing land structural characterization of CMV 1269 lastomass
4	Genetic, biochemical and structural characterization of CN11-156 p-lactamase, a
5	peculiar CMY-2 variant
6 7	
8	
9	Agustin Zavala ^{1,2} Pascal Retailleau ¹ Eddy Elisée ¹ Bogdan I. Jorga ^{1*} Thierry Naas ^{2,3,4,5*}
10 11	Agustin Zuvalu -, Fuscul Retainedu , Eddy Ensee , Bogdan I. Jorga -, Thierry Padas
12	
13	
14 15	¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saciay, LabEx
16	LERMIT, 1 avenue de la Terrasse, Bât. 27, 91198 Gif-sur-Yvette, France
17	
18 19	² EA7361 "Structure, dynamic, function and expression of broad spectrum β -lactamases",
20	Université Paris Sud-Université Paris Saclay, LabEy LERMIT, Faculty of Medicine, 78 rue du
21	Oniversite Fairs Sud, Oniversite Fairs Sacidy, Lablex Election 1, Faculty of Medicine, 70 fac da
22	Général Leclerc, 94275 Le Kremlin-Bicêtre, France
24	3 Desterislary Hygiana unit Aggistance Dublique/Hânitaur de Daris Disêtre Hegnitel 78 me du
25	³ Bacteriology-Hygiene unit, Assistance Publique/Hopitaux de Paris, Bicette Hospitai, 78 fue du
26 27	Général Leclerc, 94275 Le Kremlin-Bicêtre, France
28	
29	⁴ Associated French National Reference Center for Antibiotic Resistance: Carbapenemase-
30 31	producing Enterobacteriaceae, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France
32	
33	⁵ Evolution and Ecology of Resistance to Antibiotics Unit, Institut Pasteur – APHP -Université
34 35	Paris Sud. 25-28 Rue du Dr Roux. 75015 Paris, France
36	Turis Suu, 25 20 Ruo uu Di Roun, 70010 Turis, Turio
37	
39	*To whom correspondence should be addressed:
40	10 whom correspondence should be addressed.
41	Bogdan I. Iorga: Institut de Chimie des Substances Naturelles, CNRS UPR 2301, 91198 Gif-sur-
43	V
44	Y vette, France. E-mail: bogdan.lorga@cnrs.lr; 1ei. +33 1 69 82 30 94; Fax. +33 1 69 07 72 47
45 46	Thierry Naas : Service de Bactériologie-Hygiène, Hôpital de Bicêtre, 78 rue du Général Leclerc,
47	
48	94275 Le Kremlin-Bicetre, France. E-mail: thierry.naas@aphp.tr; Tel: +33 1 45 21 20 19; Fax: +33
49 50	1 45 21 63 40
51	
52	
55 54	Keywords: beta-lactamase, crystal structure, docking, molecular dynamics, antibiotic resistance.
55	
56 57	CMY-2, cephalosporinase, ESAC, CAZ, TIC, Ω-loop
58	1
59	ACC Develop Dive Environment
60	ACS Paragon Plus Environment

With the widespread use and abuse of antibiotics for the past decades, antimicrobial resistance poses a serious threat to public health nowadays. β -Lactams are the most used antibiotics, and β -lactamases the most widespread resistance mechanism. Class C β -lactamases, also known as cephalosporinases, usually do not hydrolyse the latest and most potent β -lactams, expanded spectrum cephalosporinases, their resistance to inhibition by classic β -lactamase inhibitors, and the fact that they can contribute to carbapenem resistance when paired with impermeability mechanisms, means that these enzymes may still prove worrisome in the future. Here we report and characterize the CMY-136 β -lactamase, a Y221H point mutant derivative of CMY-2. CMY-136 confers an increased level of resistance to ticarcillin, cefuroxime, cefotaxime and ceftolozane/tazobactam. It is also capable of hydrolysing ticarcillin and cloxacillin, which act as inhibitors of CMY-2. X-ray crystallography and modelling experiments suggest that the hydrolytic profile alterations seem to be the result of an increased flexibility and altered conformation of the Ω -loop, caused by the Y221H mutation.

ACS Infectious Diseases

CMY-2-like enzymes are plasmidic class C β-lactamases that originated from the chromosomallyencoded AmpC of Citrobacter freundii¹. They can be found in humans, feedstock animals, and pets². CMY-2 is the most widespread plasmid-mediated AmpC (pAmpC) β -lactamase, which can be illustrated by the CMY-2-like enzymes reported to date, which amount to over 1403. They are also known as cephamycinases, because of their potential to hydrolyse these substrates, but they are also active on penicillins and first generation cephalosporins, and fairly resistant to inhibition by classical β -lactamase inhibitors (clavulanic acid, tazobactam and sulbactam)¹. They are less active against third or fourth generation cephalosporins, and almost completely inactive against carbapenems. Lately however, several class C enzymes with extended-spectrum activity, termed extended-spectrum AmpC cephalosporinases (ESACs)⁴, have been reported, among them members of CMY-2-like family. Many of these enzymes show increased activity against third generation cephalosporins such as ceftazidime or cefotaxime, and even cefepime, a fourth-generation cephalosporin. When paired with impermeability mechanisms, CMY-2 may even mediate carbapenem resistance⁵. These enzymes are of clinical importance, and their study may prove important not only in understanding how their hydrolysis profile might evolve, but also to gain insights into their structure-function relationship, which in turn may aid in the development of new inhibitors, not only for class C but also for other classes of β -lactamases. We report here a novel CMY-2-like β -lactamase, CMY-136, we characterize it both biochemically and structurally, and provide the possible structural basis behind its hydrolysis profile differences when compared to CMY-2.

RESULTS

Isolation of *bla*_{CMY-136} gene

E. coli EC13 clinical isolate, identified from a urinary tract infection, displayed an unusual cephalosporinase phenotype⁶. This isolate was resistant to ticarcillin and expanded spectrum cephalosporins, showed reduced susceptibility to cefoxitin, and showed increased susceptibility to

inhibition by clavulanic acid. Performing susceptibility testing on cloxacillin containing plates restored susceptibility to ticarcillin and expanded spectrum cephalosporins, suggesting a class C enzyme. PCR was positive for CMY-2, and sequencing revealed a novel CMY-2-like variant, which was assigned the name CMY-136, differing by a single amino acid substitution, Y221H, in a highly conserved position in class C β -lactamases³. The sequence alignment between CMY-2 and CMY-136 is shown in Fig. S1.

Phenotypical characterization of CMY-136

The antimicrobial susceptibility profile conferred by CMY-136 was compared to that of CMY-2 by transforming *E. coli* TOP10 cells with a pTOPO plasmid harbouring either blaCMY-2 or blaCMY-136 genes and performing disk diffusion susceptibility tests and MICs assays. CMY-2 and CMY-136 show similar resistance profiles, they both confer resistance to penicillins and cephalosporins, and show susceptibility to cefepime and carbapenems (see Fig. S2 for the chemical structures of these β -lactams). Based on the disk diffusion test, when compared to CMY-2, CMY-136 confers increased resistance to ticarcillin and cefotaxime, low level resistance to cefoxitin, and shows increased susceptibility to inhibition by clavulanic acid. The MICs show that when compared to CMY-2, CMY-136 confers to *E. coli* decreased resistance to piperacillin and cefoxitin, and an increased level of resistance to ticarcillin, cefuroxime and cefotaxime (Table 1). According to the breakpoints of EUCAST, *E. coli* TOP10 expressing CMY-136 is also resistance to it.

ß laatam –	E. coli TOP	E. col	E. coli TOP10 pTOPO-CMY-136				
p-lactalli	MIC (g/L)	CLSI	EUCAST	MIC	(g/L)	CLSI	EUCAST
Amoxicillin	>256	_*	R	>2:	56	_*	R
Amoxicillin + CLA ^a	32	R	R	24	4	Ι	R
Ticarcillin	32	_*	R	>2:	56	_*	R
Temocillin	24	_*	R	24	4	_*	R
Piperacillin	24	Ι	R	12	2	S	Ι
Piperacillin + TAZ^b	4	S	S	2	*	S	S
		Λ					

Table 1 MICs for E. coli TOP10 pTOPO-CMY-2 and E. coli TOP10 pTOPO-CMY-136.

ACS Paragon Plus Environment

Cefoxitin	>256	R	R	16	Ι	Ι
Cefuroxime	24	Ι	R	>256	R	R
Cefotaxime	8	R	R	>32	R	R
Ceftazidime	24	R	R	24	R	R
Ceftazidime + AVI ^c	0.38	S	S	0.75	S	S
Cefepime	0.125	S	S	0.094	S	S
Ceftolozane + TAZ ^{b}	0.5	S	S	2.5	Ι	R
Imipenem	0.25	S	S	0.19	S	S
Meropenem	0.032	S	S	0.023	S	S
Ertapenem	0.012	S	S	0.012	S	S
Aztreonam	2	S	Ι	3	S	Ι

S: susceptible; I: intermediate; R: resistant

* values not available

^aCLA: clavulanic acid

^{*b*}TAZ: tazobactam

^cAVI: avibactam

Genetic support of bla_{CMY-136} gene

Kieser⁷ extracted DNA from *E. coli* EC13 was electroporated into electrocompetent *E. coli* TOP10 cells and plated on ampicillin-containing plates (100 mg/L). Several colonies that grew on ampicillin were confirmed by PCR and sequencing to be blaCMY-136 positive. Kieser DNA of *E. coli* EC13 and of electrotransformed *E. coli* TOP10 cells run on a 0.7% agarose gel revealed a single plasmid, pCMY-136 of ca. 90 Kbp in both isolates. Whole genome sequencing using Illumina technology of *E. coli* TOP10 (pCMY-136) revealed a 90844 bp plasmid in size, that belonged to the IncI1 group, and carried a single antibiotic resistance gene, $bla_{CMY-136}$. BLAST search showed the closest deposited plasmid sequence to be the bla_{CMY-2} carrying plasmid pCVM22462 (accession number CP009566.1)⁸, isolated from a *Salmonella enterica* strain. It is 99.98% identical to pCMY-136, excluding a 3,888-bp fragment that is absent in pCMY-136. This deletion is located 1,965-bp upstream of $bla_{CMY-136}/bla_{CMY-2}$ genes. This 3,888-bp fragment contains 3 hypothetical protein ORFs and occurred contiguously downstream of an IS*1294*-like insertion sequence present on both plasmids.

Purification and biochemical characterization of CMY-136

CMY-136 and CMY-2 were overexpressed and purified by Immobilized Metal Affinity Chromatography (IMAC), and purity of both enzymes was confirmed by SDS-PAGE, where a single band of ca. 41 kDa was observed. Steady-state kinetic parameters were determined to compare the catalytic activity of CMY-136 to that of CMY-2 against several β -lactams (Table 2). CMY-136 showed around a 10-fold increase in Km for most β -lactams assayed, with however, a few exceptions: (i) cephalothin (CEF), which showed a slightly higher affinity for CMY-136 than CMY-2, although other authors have reported a lower Km for CMY-2 against CEF^{9,10,11}; (ii) cefoxitin (FOX), which showed a practically identical Km for both enzymes; (iii) β -lactams with a very high affinity for CMY-2 and a bulky, non-linear, R1 substituent, namely cefuroxime (CXM), cefotaxime (CTX) and ceftazidime (CAZ), which show a 10³-10⁴-fold increase in Km for CMY-136; (iv) other β -lactams, namely ticarcillin (TIC) and cloxacillin (CLX), for which CMY-2 shows high affinity and no hydrolysis unlike CMY-136 that does hydrolyze them (see Fig. S2 for the meaning of the R1 and R2 substituents).

CMY-136 showed a decreased k_{cat} for most of the preferred substrates of CMY-2, namely amino-penicillins and first generation cephalosporins, and also cefoxitin. However, it also shows an increased k_{cat} for cefepime and ceftolozane, and a remarkably increased k_{cat} (10²-10³-fold) for those β -lactams for which the most dramatic Km increase was also seen, CAZ, CTX, CXM. CMY-136 is also capable of hydrolysing cloxacillin and ticarcillin, which are not hydrolysed by CMY-2. In terms of their catalytic efficiency, all these translate into CMY-136 showing a lower efficiency than CMY-2 for most β -lactams tested, except for those not hydrolysed by CMY-2, TIC and CLX, and also ceftolozane.

Table 2 Steady-state kinetic parameters of β -lactamases CMY-136 and CMY-2.

	$K_{\rm m}(\mu{ m M})$		$k_{\rm cat}$ (s ⁻¹)		$k_{\rm cat}/K_{\rm m}~({\rm mM}^{-1}~{\rm s}^{-1})$	
Substrate	CMY-2	CMY-136	CMY-2	CMY-136	CMY-2	CMY-136
Benzylpenicillin	0.6^{b}	10.1*	24 ^b	7.83	40000 ^b	775
Cloxacillin	0.0002*	543	N.H.	0.47	N.H.	0.9
Ampicillin	0.16 ^c	3.5*	0.55 ^c	0.43	3437 ^c	122

ACS Paragon Plus Environment

T:	0.02*	1 7*		0.15		0.6
Ticarcillin	0.03*	1./*	N.H.	0.15	N.H.	86
Cephaloridine	130 ^a	1532	421 ^a	169	3240 ^a	110
Cephalothin	90 a	35	217 ^a	21.65	2400 a	618
Cefoxitin	0.15 ^b	0.2*	0.35 ^b	0.02	2333 ^b	111
Cefuroxime	0.005 ^b	21.6	0.014 ^b	9.42	2800 ^b	436
Cefotaxime	0.001 a	20	0.007 ^a	4.71	7000 ^a	236
Ceftazidime	0.15 <i>a</i>	2360	0.005 a	6.26	33 a	2.7
Cefepime	412	3588	0.37	1.79	0.9	0.5
Ceftolozane	200	1587	0.69	13.27	3.5	8.4
Imipenem	0.8 <i>a</i>	10.9	0.03 <i>a</i>	0.01	38 a	1.2
Aztreonam	N.H. ^b	N.H.	N.H. ^{<i>b</i>}	N.H.	N.H. ^b	N.H.

Errors for determined kinetic values are below 10%

N.H.: no hydrolysis detected with concentrations of substrate and enzymes up to 1,000 mM and 400 nM, respectively.

^{*a*} taken from ^{12,5}

^b taken from ⁹

^c taken from ¹¹

*Km values determined by inhibition assays as explained in experimental procedures

Table 3 shows IC_{50} values for aztreonam, tazobactam, and clavulanic acid. CMY-136 is similarly susceptible to inhibition by tazobactam as CMY-2^{13,14}.

Table 3 IC₅₀ values (μ M) for inhibitors of CMY-136 and CMY-2.

Inhibitor	CMY-2	CMY-136
Tazobactam	3.2ª	5.5
Clavulanic acid	N.D.	137
Aztreonam	N.D.	0.11
ND: not determined		

N.D.: not determined *a* taken from¹³

CMY-136 crystallization and X-ray crystallography

CMY-136 crystallized in 0.1 M HEPES pH 7.5; 20% w/v PEG4000; 10% isopropanol. The structure was determined by molecular replacement using the deposited CMY-2 structure (PDB: 1ZC2) as a template, and the model was refined to 1.60 Å (Table 4). R/Rfree values seem to be relatively high for the resolution obtained, which may be a consequence of merging two datasets (see experimental procedures). The asymmetric unit contains two protein chains, A and B, showing

no significant conformational differences between them. They have been modelled with 359 residues each, with clear electron density observed for all regions of the protein. Several ordered water molecules have been modelled in the structure as well as phosphate anions and isopropyl alcohol molecules, from the crystallization condition. Both chains present the typical class C fold with an α -helical region and a mixed α -helix/ β -sheet region. 99% of all residues are inside the favoured regions of the Ramachandran plot, and 1% in the allowed regions.

Table 4 Crystallography data collection and refinement statistics.

Data collection			
Space group	P 1 21 1		
Cell dimensions			
a, b, c (Å)	60.58, 58.09, 100.08		
α, β, γ (°)	90.00, 89.97, 90.00		
Resolution (Å)	16.1-1.60 (1.62-1.60)		
Rmerge	0.092 (0.67)		
I/σ(I)	9.9 (3.9)		
Completeness (%)	95.0 (96.3)		
Redundancy	4.2 (4.4)		
CC (1/2)	0.991 (0.775)		
Refinement			
Resolution range (Å)	16.10-1.60		
No. unique reflections	87,503		
Rwork/Rfree	23.4%/26.4%		
No. non-hydrogen atoms			
Protein	5,642		
Water	396		
Ligand/Ions	67		
Total	6105		
Average B, all atoms (Å2)	19.8		
Protein	19.4		
Water	25.2		
Ligand/Ions	29.6		
Root mean squared deviations			
Bond lengths (Å)	0.01		
Bond angles (°)	1.04		

Values in parenthesis correspond to the higher resolution bin.

ACS Infectious Diseases

CMY-136 shows the same overall conformation as CMY-2, and 3D alignment reveals a Ca RMSD of 0.465-0.486 Å between them. Several structural differences can be found when comparing CMY-136 (Fig. 1) to CMY-2: (i) A phosphate molecule is positioned inside the active site of both CMY-136 chains, with one of its oxygens buried inside the oxyanion hole. CMY-2 was crystallized with a citrate molecule inside the active site; (ii) The conformation adopted by the TYR221 in CMY-2, which interacts with the backbone of the Ω -loop at SER212 O and GLN215 O via two bridging waters, is different from the conformation found in CMY-136 for HIS221, whose sidechain invades the active site cavity, hydrogen bonding the aforementioned phosphate oxygen atom. The TYR221 residue and its conformation are highly conserved in all class C enzymes reported to date (with the exception of PDC-85) and in all class C crystallographic structures deposited in the PDB that have a consensus length Ω -loop; (iii) In turn, the sidechain of ASP217 turns towards the interior of the Ω loop to partially occupy the space left by the conformational change of the Y221H mutation; (iv) The conformation adopted by the backbone of the Ω -loop residues ARG204 to GLN215 in CMY-136 is different from the one found in CMY-2, that is also highly conserved in deposited class C structures with a consensus length loop, which translates into a backbone atoms RMSD between said residues in both crystal structures of 2.176 - 2.219 Å, a 4.56 - 4.68 Å displacement in the position of the Ca of VAL211, and the loss of the hydrogen bonds between GLU61 OE and VAL211 N (Fig. 2) and between VAL209 O and GLY202 N (Fig. 2B). Two water molecules, w502 and w504, occupy the usual space of residues V209-V211 backbone, and partially replace those lost hydrogen bonds (Fig. 2B); (v)The configuration of HIS210 sidechain is shifted from being positioned between, and interacting via pi-pi stacking with, TYR199 and TRP201 in CMY-2 another highly conserved feature of class C structures, that fill this space with either HIS210 or ARG210 sidechains - to not directly interacting with any other residue, and hydrogen bonding w502 (Fig. 2B). The space between TYR199 and TRP201 in the CMY-136 structure is filled by LYS290 from another CMY-136 chain in the neighbouring asymmetric unit in the crystal structure; (vi) The conformation of the active site cavity residues and the hydrogen bond network between them is

highly conserved, with the exception of TYR150, which is slightly shifted towards the phosphate anchored in the active site, with a 3.1 Å displacement of its hydroxyl oxygen atom compared to TYR150 of the CMY-2 structure (Fig. 1 and Fig. S3). A water molecule can be found occupying this position, and TYR150 is still hydrogen bonding LYS67, and now also hydrogen bonds with the phosphate molecule and possibly ASN152, but no longer interacts with LYS315.



Fig. 1 Crystal structure of CMY-136. Active site cavity of CMY-136, with a phosphate molecule interacting with the oxyanion hole and other important residues. Note the conformation of the H221

ACS Infectious Diseases

sidechain, the position shift of Y150, and the water molecule occupying its hydroxyl usual position. Hydrogen bonds are depicted as dashed yellow lines. 2Fo-Fc electron density (contoured at 2σ) is shown.



Fig. 2 Comparison of the Ω-loop of CMY-136 and CMY-2. CMY-2 and CMY-136 are shown in green and magenta, respectively. a) Front view, the conformational change in the Ω-loop for CMY-136 widens the active site cavity, displacing V211 C α about 4.6 Å, and breaking the hydrogen bond with E61. Hydrogen bonds are depicted as dashed yellow lines. Displacement depicted as dashed orange lines. b) Side view, the conformational change also displaces H210 sidechain from its usual position interacting with residues 199-201. Hydrogen bonds are displayed as colored dashed lines according to model colors.

Modelling of covalent complexes of CMY-2 and CMY-136 with β -lactams

As no crystal structure of CMY-2 co-crystallized with β -lactams is available, covalent complexes of CMY-2 and CMY-136 with ceftazidime, cefotaxime and cefuroxime were created by superposing the structures of CMY-2 and CMY-136 with the available structures of class C enzymes with

ceftazidime (PDB code 1IEL) and cefotaxime (PDB code 4KG2), and in the case of cefuroxime, the covalently bound substrate was modelled on the conformation of the cefotaxime complex, given no covalent complex of class C enzymes with cefuroxime is available and both molecules show minor differences. Results show that in the case of CMY-2, acyl-enzyme complexes have minor clashes with L293 (Fig. 3A). In the case of CMY-136 numerous overlaps with residues H221 and Y150 exist (Fig. 3C), which may explain the approximately 1000-fold increase in Km values for expanded spectrum cephalosporins for CMY-136 compared to CMY-2. Covalent complexes were then subjected to energy minimization steps, and results showed the conformational changes needed to resolve overlaps in the case of CMY-2 were smaller than those for CMY-136 (Fig. 3B and D, respectively). Both H221 and the R1 substituent of cephalosporins are displaced in opposite directions in the CMY-136 complexes, pushing the substrate outwards from the active site cavity and H221 towards the Ω -loop. In the case of CMY-2, displacements were smaller in magnitude, and they tended to approach the R1 substituent towards Y221 (see Fig. S2 for the meaning of the R1 substituent).

Page 13 of 34





Fig. 3. Modelling of covalent complexes of CMY-2 and CMY-136 with ceftazidime. Covalent complexes were made by superposing the structures of CMY-2 or CMY-136 on the covalent complex of ceftazidime with AmpC (PDB code 11EL). These were subjected to energy minimization. Covalent complex of CMY-2 with ceftazidime before (a) and after (b) minimization, as well as covalent complex of CMY-136 with ceftazidime before (c) and after (d) minimization are shown. Only relevant residues from CMY enzymes and ceftazidime are shown as sticks for simplicity. Overlaps between the ligand and the protein are depicted as red lines. In figures **b** and **d**,

both the minimized (cyan) and initial (green for CMY-2, orange for CMY-136) structures are shown for comparison.

Molecular dynamics simulations of CMY-2 and CMY-136

To investigate the role of the Y221 mutation on the dynamics of the protein structure, 10 ns molecular dynamics (MD) simulations were performed in triplicate with four different systems (Fig. S4): CMY-2, CMY-136, CMY-136 with the Ω -loop modelled on the CMY-2 structure, and CMY-2 with the Ω -loop modelled on the CMY-136 structure. For these MD simulations, citrate and phosphate ions were removed from the active sites of CMY-2 and CMY-136 structures, respectively. Root-mean-square deviation (RMSD) analysis of the trajectories reveals that the simulations reached stability within 2 ns and remained stable after that. Root-mean-square fluctuation (RMSF) analysis of C α atoms in CMY-2 and CMY-136, expressed as B-factors, reveals some differences between them, regarding the flexibility of several parts of the structure (Fig. 4A). CMY-136 displays increased flexibility in the Ω -loop and H7 helix, from R204 up to H221, except for a short segment in the middle. An increased flexibility can also be found for part of the Q120 loop, part of H10 and H11 helixes, and the C terminus of B3 strand, S318 and T319. Other loops around CMY-136 also have increased flexibility compared to CMY-2, to a lesser extent. There are also segments of decreased flexibility towards the exterior of the structure, contiguous to the previously described regions, the N-terminus of H10 helix and its preceding loop, C-terminus of H11 helix, H4 helix, G320, part of helixes H5 and H6, directly behind H4 helix in the tertiary structure, and also S343, next to the C-terminus of strand B3. It should be noted that R349, in the middle of H11, also shows decreased flexibility. Regions with increased flexibility can be found surrounding and forming the walls of the active site cavity, with regions of decreased flexibility further away (Fig. 4B).



Fig. 4. Altered flexibility of CMY-136 Structure. a) CMY-2 and CMY-136 RMSF analysis, expressed as Cα B-factors. Compared to CMY-2, several regions displayed increased or decreased flexibility in the CMY-136 structure during MD simulation: part of the Ω-loop (R204-Y221), Q120 loop (P118-R126), C-terminus of B3 strand (S318-G320), helixes H4 (D127-L131), H5 (A160-G167), H6 (S169-M174), H10 (P277-A295), H11 (S343-Q361). b) Surface representation of CMY-136 in the acyl-enzyme complex with CAZ, which is represented as green sticks. The surface of S64 is colored in purple. The surface of regions with increased flexibility, determined by Cα B-factors, are colored in yellow, and those with decreased flexibility in red.

The MD simulations also demonstrate how the sidechain of Y221 in CMY-2 remains constantly in the same conformation, pointing outwards of S64, with its hydroxyl hydrogen bonding with the backbone of the Ω -loop directly or through bridging waters (Fig. 5A). In contrast, H221 in CMY-136 displays much more flexibility, adopting several conformations in the active site throughout the MD simulation (Fig. 5B and 6). Other parts of the Ω -loop also show increased sidechain flexibility (Fig. 6), namely H210-S212 and Y199-W201, which normally interact closely with each other and with E61.



Fig. 5. Flexibility of TYR 221 in CMY-2 and HIS 221 in CMY-136. A) Snapshot from the CMY-2 MD simulation depicting the only conformation of Y221 observed throughout the 10 ns simulation. B) H221 conformations observed in four representative snapshots of the CMY-136 MD simulation. One of the conformations adopted is similar to that of Y221 for CMY-2 seen in panel A.



Fig. 6. Flexibility of Ω -loop in CMY-136 vs CMY-2. Individual atom B-factors of CMY-136 and CMY-2 Ω -loops. Note the increased flexibility of residue 221 in CMY-136, and also of regions H210-S212 and Y199-W201, which interact in CMY-2 as shown in Fig. 2.

This same H221 flexibility can be observed for the simulation of the CMY-136 model with the Ω loop, H7 and H221 sidechain starting with the same conformation as the CMY-2 crystal structure. In contrast, for the CMY-2 model with the Ω -loop, H7 and Y221 sidechain starting with the same conformation as the CMY-136 crystal structure, the native conformation of Y221 is not reached. Simulations for both enzymes starting either with the Ω -loop in the conformation observed in their own crystal structure, or in that of the other enzyme, seem to show that the conformation acquired by the Ω -loop during the simulations is independent of starting conformation. Snapshots of these four simulations were extracted every 1 ns and fitted on the CMY-2 or CMY-136 crystal structure

as references. Ω -loop C α RMSD between the snapshot and the reference was determined, to assess conformational differences (Fig. S5). The analysis shows that both CMY-136 simulations, starting with a Ω -loop, H7 and H221 conformation as that observed in either the CMY-136 crystal structure or the CMY-2 crystal structure, present a lower RMSD compared to CMY-136 than to CMY-2 crystal structures towards the end of the simulation. Similarly, for the CMY-2 simulation starting from the CMY-2 crystal structure conformation, some fluctuations can be observed during the simulation but finally the RMSD is also lower for its own native crystal structure than for that of the other enzyme, and in the same magnitudes as observed for CMY-136. In contrast, in the simulation of CMY-2 starting with the Ω -loop, H7 and Y221 conformation of the CMY-136 loop does not converge to the CMY-2 crystal structure conformation and remains closer to the CMY-136 crystal structure conformation. As mentioned earlier, for this simulation, Y221 does not acquire the native conformation observed in all class C structures.

DISCUSSION

We report here a novel β -lactamase belonging to the CMY-2 family, with an unusual mutation, Y221H, in a highly conserved position in class C β -lactamases. Analysis of the plasmid harbouring $bla_{CMY-136}$ suggests it may have originated by mutation deriving from the bla_{CMY-2} harbouring plasmid pCVM22462. It is noteworthy that the 3888bp fragment absent from pCMY-136 but present in pCMY-2 occurs directly adjacent to IS1294, which may suggest its involvement in the deletion of this fragment¹⁵.

The biochemical characterization of CMY-136 suggests it to be less efficient than CMY-2 in hydrolysing most β -lactams, the exception being ceftolozane, and also those not hydrolysed by the latter, cloxacillin and ticarcillin. The results for these three β -lactams corroborate what can be observed in the MIC assays. However, other significant differences in MICs for CMY-136 when compared to CMY-2 (increased MIC for cefotaxime and cefuroxime, and decreased MIC for cefoxitin) are not reflected in the catalytic efficiencies determined. In this regard, it has been

ACS Infectious Diseases

previously suggested that for class C enzymes, k_{cat} may better represent the degree of resistance towards β -lactams^{16,17}. Therefore, the comparison of k_{cat} may explain better what is observed for the MICs. A 10-fold increase in k_{cat} , or bigger, is observed for CMY-136 for cloxacillin, ticarcillin, cefuroxime, cefotaxime, ceftazidime, and ceftolozane, and a 10-fold decrease for cephalothin and cefoxitin. This corresponds well with observed MIC results, except for cloxacillin and cephalothin, for which MICs were not determined, and ceftazidime, which showed no difference in MIC between CMY-136 and CMY-2.

These results for CMY-136 are consistent with previously published results for similar cases. PDC-85, a variant of the chromosomally-encoded AmpC of *Pseudomonas aeruginosa*, is the only other class C enzyme presenting a Y221H mutation¹⁸. Similarly to CMY-136, the PDC-85 variant showed an increase in MIC for ticarcillin and ceftolozane/tazobactam, and a decreased susceptibility to inhibition by cloxacillin, probably related to CMY-136 being capable of hydrolysing cloxacillin. In a different study, it was observed that the *E. coli* AmpC Y221G laboratory mutant showed an increased K_m and k_{cat} for cefotaxime, while at the same time significantly losing activity on cephalothin¹⁷. The authors evidenced that this mutation decreases the enzyme thermal stability, increase flexibility, and expand the active site cavity, and attribute the increase in cefotaxime hydrolysis to these effects, allowing for a better accommodation of the acylenzyme complex into a catalytically competent conformation for de-acylation. For cephalothin, they propose the now missing stabilizing interactions between the substrate and the Y221G mutation to be responsible for the loss of hydrolysis.

The crystal structure of CMY-136 shows the same conformation as that of CMY-2, except for those differences described in the Results section. Just as for the Y221G AmpC mutant study, in CMY-136 the space no longer occupied by Y221 is partially occupied by D217, which turns towards the active site cavity. Another possible common feature is the expansion of the active site cavity, as the CMY-136 Ω -loop conformation displaces the V211 C α from its usual position in class C structures by about 4.6Å. It is unclear whether the changes observed in the CMY-136

conformation are the result of the crystal configuration and interactions between neighbouring chains in it, the crystallization with a phosphate molecule in the active site, the consequences of the Y221H mutation, or a combination of them.

In order to investigate the high differences in affinity for expanded spectrum cephalosporins between CMY-2 and CMY-136, covalent complexes were created by superimposing the structures with available complexes and subjected to energy minimization, to evidence the conformations adopted by the acyl-enzyme complexes. As shown in Fig. 3, minimization does not lead to important conformational changes in the case of CMY-2. Superposition of the CMY-136 crystal structure on the complexes structures shows that there would be significant overlapping between the R1 sidechain of β -lactams and the H221 sidechain. This serves to explain why CMY-136 may show an increased Km for most β -lactams: in order for the β -lactam and the enzyme to interact, either the R1 sidechain or the H221 sidechain, or both, must adopt a different conformation from the one seen in most structures of class C enzymes in complex with substrates or the CMY-136 structure, respectively. Such conformational change most likely implies an energetic cost, that in turn represents an increase in Km. Furthermore, this higher energetic cost – and increase in Km – for CMY-136 interacting with most β -lactams seems to correlate with how big, bulky, ramified or rigid is the R1 substituent. There are much larger increases in Km for cloxacillin, ceftazidime, cefuroxime and cefotaxime than for those with a simpler R1 sidechain like benzylpenicillin or cephalothin. One possible explanation as for why such Km increase is less intense or doesn't exist for CEF or FOX compared to CAZ or CXM, is that the R1 sidechain of the former is not as bulky, ramified, or rigid as for the others, and thus they may prove easier to accommodate inside the active site of CMY-136 with its Y221H mutation, implying less movement restrain for H221, or making more favourable contacts and hydrogen bonds, and thus with a lesser cost in terms of binding energy and Km.

The analysis of MD simulations suggests that the CMY-136 structure may show increased flexibility around the R1 and R2 pockets in the active site. The R2 pocket is usually described as

ACS Infectious Diseases

surrounded by the R2 loop (residues 289-307) and the proximal residues of helixes H10 and H11^{1,19}, and the R1 pocket is defined ad surrounded by the Ω loop (residues 178-226)^{1,19} (Fig. S6). The R1 substituent of ceftazidime, and other β -lactams, also interacts with other residues surrounding it, namely Q120, N152, residues 218-220 and S343 (Fig. S6). As previously proposed by other authors^{9,20}, an increased k_{cat} may be a consequence of an increased protein flexibility, that may allow the acyl-enzyme complexes to more often adopt a conformation competent for being attacked by the hydrolytic water and be released. In the case of CMY-136, a change in the Ω -loop conformation, widening the active site cavity by displacing V211, may also participate in the k_{cat} increase observed for certain substrates. As observed by the superposition and RMSD analysis of the MD simulations, the CMY-136 Ω -loop tends to adopt a conformation more similar to the one observed in the CMY-136 structure than to that observed in the CMY-2 structure, regardless of the starting conformation. For CMY-2 starting with the Ω -loop in its native conformation, some fluctuation is observed but finally a conformation more similar to this native one than to the one in CMY-136 crystal structure is adopted. As mentioned earlier, this is not observed for the CMY-2 simulation starting with a CMY-136 Ω -loop conformation. In this simulation the loop remains more flexible and the Y221 sidechain adopts different conformations, as observed in the case of CMY-136. The stability of the native Y221 conformation (conserved in class C structures) in the CMY-2 MD simulation, and the fact that this conformation is not reached in a 10 ns simulation starting from a different conformation, suggests that the native conformation of the Ω -loop of class C enzymes may be quite stable, and a relatively high energetic barrier might have to be overcome to reach, or come out of it. In contrast, H221 in CMY-136 easily adopts several conformations throughout the simulations, starting from either crystal structure conformations (CMY-2 and CMY-136). This in turn may help to explain the experimental data observed, i.e. a flexible H221 would allow CMY-136 to hydrolyse β -lactams in spite of its conformation in the crystal structure. Were H221 to remain in this conformation, CMY-136 would most likely be incapable of interacting with its substrates, especially those with a more rigid, bulkier, ramified R1 sidechain. The fact that H221

can adopt different conformations during the MD simulation, one of them similar to the Y221 conformation in CMY-2, helps to imagine how it may still accommodate the substrates in its active site cavity. On the other hand, this flexibility may also explain the increase in Km observed for most β -lactams. Although H221 seems to be much more flexible than Y221 and may adopt several conformations, once CMY-136 is interacting with the substrate, H221 would probably lose much of its movement freedom. This would imply an energetic cost, and an increase in Km for the reaction. Such increase in Km would clearly not be observed in the case of CMY-2 or other class C enzymes with Y221 in its native conformation, as it is probably static enough that no movement restrain is imposed on it upon binding of the substrate.

CONCLUSION

CMY-136 is a novel β -lactamase, which belongs to the CMY-2-like family. It possesses a single unusual mutation, Y221H, in a position highly conserved in all class C β -lactamases. We compared here its genetic environment to that of the closest plasmid harbouring CMY-2 and characterized CMY-136 both biochemically and structurally. Finally, *in silico* modelling techniques suggested a possible structural explanation for the differences observed compared to CMY-2. These results allowed us to better assess the potential of class C β -lactamases for their spectrum extension and for their potential threat to public health. Finally, the identification of the Y221H mutation in a natural variant of CMY-2 and which confers resistance to ceftolozane/tazobactam prior to the use of this latter antibiotic in clinics is worrisome, since it may indicate a more rapid appearance of resistance, as this type of enzyme may be selected.

EXPERIMENTAL PROCEDURES

Bacterial strains

ACS Infectious Diseases

*bla*_{CMY-136} was recovered from a urinary tract *E. coli* isolate EC13⁶. *bla*_{CMY-2} was recovered from the *E. cloacae* ec-204 strain (Bicêtre strain collection). *E. coli* TOP10 (Invitrogen, Saint-Aubin, France) and *E. coli* BL21 (DE3) (Novagen, VWR International, Fontenay-sous-Bois, France) were used for cloning experiments and protein overproduction, respectively.

Plasmid extraction and cloning

Plasmid DNA from *E. coli* EC13 was extracted using the Kieser method⁷. The Kieser extracted DNA was used to transform *E. coli* TOP10 strain by electroporation. The electroporants were plated on TSA plate containing ampicillin (100 μ g/ml). Transformants were analyzed by PCR using the primers CMY-2A and CMY-2B (5'-aaaaacatatgatgaaaaaatcgttatgctgc-3' and 5'-aaaaggatccttattgcagcttttcaagaatgc-3', respectively). From the transformants harbouring *bla*_{CMY-2-like} genes, plasmid DNA was extracted using Kieser's method and subsequently analysed on a 0.7% agarose gel stained with ethidium bromide. Plasmids of ca. 154, 66, 48, and 7 kb of *E. coli* NCTC 50192 were used as plasmid size markers²¹.

Plasmid sequencing (WGS)

Plasmid DNA was extracted from an *E. coli* TOP10 clone harbouring the wild type pCMY-136 plasmid using the QIAGEN Plasmid Maxi Kit following the manufacturer's instructions. The DNA concentration and purity were controlled by a Qubit[®] 2.0 Fluorometer using the dsDNA HS and/or BR assay kit (Life technologies, Carlsbad, CA, US). The DNA library was prepared using the Nextera XT-v3 kit (Illumina, San Diego, CA, US) according to the manufacturer's instructions and then run on Miseq (Illumina) for generating paired-end 300-bp reads. De novo assembly was performed by CLC Genomics Workbench v9.5.3 (Qiagen, Hilden, Germany) after quality trimming (Qs \geq 20) with word size 34.

The acquired antimicrobial resistance genes were identified by uploading the assembled plasmid to the Resfinder server v2.1 (<u>http://cge.cbs.dtu.dk/services/ResFinder-2.1</u>)²². Plasmid incompatibility

group was obtained by uploading the plasmid sequence to PlasmidFinder server 1.3 (<u>https://cge.cbs.dtu.dk/services/PlasmidFinder/</u>)²³. Plasmid sequence was blasted (<u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u>)²⁴ and compared to the best hit using Artemis Comparison Tool²⁵.

Cloning of *bla*_{CMY-136} and *bla*_{CMY-2} genes

PCR amplification of $bla_{CMY-136}$ and bla_{CMY-2} genes was performed using total DNA extraction of E. coli EC13 and Enterobacter cloacae ec-204, and primers CMY-2A (5'aaaaacatatgatgaaaaaatcgttatgctgc-3') and CMY-2B (5'-aaaaggatccttattgcagcttttcaagaatgc-3'). Amplicons were cloned into pCR®-Blunt II-TOPO® cloning plasmid (Invitrogen, Illkirch, France) under regulation by the pLac promoter. The recombinant pTOPO-cmy-136 and pTOPO-cmy-2 plasmids were electroporated into the E. coli TOP10 strain and then selected using TSA-plate containing kanamycin (50 mg/L). bla_{CMY-136} and bla_{CMY-2} were amplified using primers INFcmy-2A (5'-aaggagatatacatatgatgaaaaaatcgttatgct-3') and INFcmy-2B (5'-VWR International, Fontenay-sous-Bois, France) using the NEBuilder® HiFiDNA Assembly Cloning Kit (New England BioLabs®Inc, United Kingdom), following the manufacturer's instructions. Recombinant plasmids pET41b-cmy-136 and pET41b-cmy-2 were electroporated into electrocompetent E. coli BL21 (DE3) and selected using TSA-plates containing kanamycin (50 mg/L).

Recombinant plasmids were extracted using the Qiagen miniprep kit and both strands of the inserts were sequenced using M13 F and M13 R primers, for the pCR®-Blunt II-TOPO® plasmid (Invitrogen, Illkirch, France), and T7 promoter and T7 terminator primers, for pET-41b(+) (Novagen, VWR International, Fontenay-sous-Bois, France), with an automated sequencer (ABI Prism 3100; Applied Biosystems, Les Ulis, France). The nucleotide sequences were analysed using
software available at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov).

Antimicrobial agents, susceptibility testing and microbiological techniques

Antimicrobial susceptibilities were determined by disk diffusion technique on Mueller-Hinton agar (Bio-Rad, Marnes-La-Coquette, France) and interpreted according to the EUCAST breakpoints, updated in 2015 (http://www.eucast.org). Minimal inhibitory concentration (MIC) values were determined using the Etest technique (BioMérieux, Paris, France).

β-Lactamase purification

Overnight cultures of *E. coli* BL21 (DE3) harbouring either pET41b-cmy-2 or pET41b-cmy-136 recombinant plasmids were used to inoculate 2 L of BHI broth containing 50 mg/L kanamycin. Bacteria were cultured at 37°C until reaching an OD of 0.6 at 600 nm, and protein expression was induced overnight at 25°C with 0.2 mM IPTG. Cultures were then centrifuged at 6000 g for 15 min and the pellets resuspended with 10 mL of Buffer A (20 mM PBS, 175 mM K₂SO₄, 40 mM imidazol, pH 7.40). Bacterial cells were disrupted by sonication and protein solution was clarified by centrifugation at 10,000 g for 1 h at 4°C. The supernatant was then centrifuged at 48,000 g for 1 h at 4°C. CMY-136 or CMY-2 was purified using a NTA-Nickel pseudo-affinity chromatography column (GE Healthcare, Freiburg, Germany). Elution was performed in a gradient of 0 to 100% Buffer B (20 mM PBS, 175 mM K₂SO₄, 500 mM imidazol, pH 7.40). Purity was assessed by SDS–PAGE, and pure fractions were pooled and dialyzed against 100 mM sodium phosphate buffer (pH 7.4) 50 mM potassium sulphate and concentrated up to 6.4 mg/ml using Vivaspin® columns (GE Healthcare, Freiburg, Germany). Protein concentration was determined using Bradford Protein assay (Bio-Rad) ²⁶.

Steady-state kinetic parameters

Kinetic parameters of purified CMY-136 and CMY-2 were determined at 100mM sodium phosphate buffer (pH 7.0), with extinction coefficients as detailed in Table S1. When available, published values from previous studies for CMY-2 were used for comparison with those determined for CMY-136 (Table 2). The k_{cat} and Km values were determined by analysing hydrolysis of β lactams under initial-rate conditions with an ULTROSPEC 2000 model UV spectrophotometer (Amersham Pharmacia Biotech) using the Eadie–Hoffstee linearization of the Michaelis–Menten equation. The different β -lactams were purchased from Sigma-Aldrich (Saint-Quentin-Fallavier, France). Low Km values were determined as competitive inhibition constants, Ki, in the presence of a good reporter substrate (cephalothin)¹¹.

Protein crystallization and crystallography

Initial crystallization screenings of CMY-136 were set up using the Mosquito® HTS (TTP LabTech) in four crystallization screening suites: Classics, AmSO4, PEGs and PEGs II (Qiagen/NeXtal). Plates were incubated at 293K in the ROCK IMAGER 1000 (Formulatrix, Inc). The crystal was transferred to a cryo-protectant solution consisting of the mother liquor supplemented with 25% glycerol and flash-frozen in liquid nitrogen.

Diffraction data was collected at 100 K in a nitrogen cryostream on the PROXIMA2 beamline at the SOLEIL synchrotron (Saint-Aubin, France). The data were indexed and integrated with XDS²⁷ in the autoPROC toolbox²⁸. Two incomplete datasets were merged to try to make as complete a dataset as possible, using autoPROC. Data scaling was performed using AIMLESS²⁹ from the CCP4 suite³⁰. Data collection and refinement statistics are given in Table 4. The structure of CMY-136 was solved by the molecular replacement method with Phaser³¹ using the structure of CMY-2 (PDB code 1ZC2) as a search model. The model was rebuilt manually in Coot³² and then refined using BUSTER-TNT³³ with local noncrystallographic symmetry (NCS) restraints and a translation–libration–screw (TLS) description of B factors³⁴. The quality of the final refined model was assessed using MolProbity³⁵. Crystal structure images were generated using PyMOL³⁶.

Structure analysis, covalent complexes and molecular dynamics simulations

CMY-136 structure comparison to the previously published CMY-2 structure (PDB code 1ZC2) and analysis of docking results were performed with UCSF Chimera package³⁷.

Covalent complexes of CMY-2 and CMY-136 with ceftazidime, cefotaxime and cefuroxime were created by superposing the structures of CMY-2 and CMY-136 with the available structures of class C enzymes with ceftazidime (PDB code 1IEL) and cefotaxime (PDB code 4KG2), and modelling cefuroxime on the conformation of cefotaxime. Ligand structures were generated with 3D Structure Generator CORINA Classic (Molecular Networks GmbH, Nuremberg, Germany).

Molecular dynamics simulations of CMY-2 and CMY-136 were performed with Gromacs v4.6³⁷ using the OPLS-AA force field.³⁸ The covalent ligands were parametrized using a modified version of our in house MOL2FF tool. Each system was energy-minimized until convergence using a steepest descents algorithm. Molecular dynamics with position restraints for 200 ps was then performed followed by the production run of 10 ns, which was repeated three times to ensure the reproducibility. During the position restraints and production runs, the Parinello-Rahman method was used for pressure coupling, and the temperature was coupled using the Nosé-Hoover method at 300 K. Electrostatics were calculated with the particle mesh Ewald method. The P-LINCS algorithm was used to constrain bond lengths, and a time step of 2 fs was used throughout.

Nucleotide sequences accession numbers and PDB deposition.

The amino acid sequence of $bla_{CMY-136}$ gene and the nucleotide sequence of its natural plasmid pCMY-136 have been submitted to the EMBL/Genbank nucleotide sequence database under the accession numbers AVR61040.1 and MG844436.1, respectively. The crystallographic structure of CMY-136 has been deposited to the PDB, accession code 6G9T. Authors will release the atomic coordinates and experimental data upon article publication.

AKNOWLEDGEMENTS

We acknowledge SOLEIL for provision of synchrotron radiation facilities (proposal ID BAG20150780) in using PROXIMA beamlines. This work was supported by the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT) [grant number ANR-10-LABX-33], by the JPIAMR transnational project DesInMBL [grant number ANR-14-JAMR-0002] and by the Région Ile-de-France (DIM Malinf).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interests with the content of this article.

Supporting Information. Sequence alignment of CMY-2 and CMY-136 (Fig. S1), chemical structures of β -lactams and β -lactamase inhibitors used in this study (Fig. S2), different conformations of Y150 in the CMY-2 and CMY-136 structures (Fig. S3), RMSD of the MD simulations (Fig. S4), RMSD analysis of Ω -loop from MD trajectories (Fig. S5), R1 and R2 pockets in CMY-2 (Fig. S6) and extinction coefficients and wavelengths for β -lactams (Table S1).

REFERENCES

- Jacoby, G. A. (2009) AmpC Beta-Lactamases. *Clin. Microbiol. Rev.*, 22 (1), 161–182 DOI 10.1128/CMR.00036-08.
- Fang, L.-X.; Sun, J.; Li, L.; Deng, H.; Huang, T.; Yang, Q.-E.; Li, X.; Chen, M.-Y.; Liao, X.-P.; Liu, Y.-H. (2015) Dissemination of the Chromosomally Encoded CMY-2
 Cephalosporinase Gene in Escherichia Coli Isolated from Animals. *Int. J. Antimicrob. Agents*, *46* (2), 209–213 DOI 10.1016/j.ijantimicag.2015.04.003.
- Naas, T.; Oueslati, S.; Bonnin, R. A.; Dabos, M. L.; Zavala, A.; Dortet, L.; Retailleau, P.;
 Iorga, B. I. (2017) Beta-Lactamase Database (BLDB) Structure and Function. *J. Enzyme Inhib. Med. Chem.*, *32* (1), 917–919 DOI 10.1080/14756366.2017.1344235.

1 2		
2 3 4	(4)	Powers, R. A. (2016) Structural and Functional Aspects of Extended-Spectrum AmpC
5 6		Cephalosporinases. Curr. Drug Targets, 17 (9), 1051-1060 DOI
7		10.2174/1573399811666150615144707.
9 10	(5)	Dahyot, S.; Broutin, I.; de Champs, C.; Guillon, H.; Mammeri, H. (2013) Contribution of
11 12		Asparagine 346 Residue to the Carbapenemase Activity of CMY-2 β -Lactamase. <i>FEMS</i>
13 14		Microbiol. Lett., 345 (2), 147–153 DOI 10.1111/1574-6968.12199.
15 16	(6)	Mammeri, H.; Poirel, L.; Fortineau, N.; Nordmann, P. (2006) Naturally Occurring Extended-
17 18		Spectrum Cephalosporinases in Escherichia Coli. Antimicrob. Agents Chemother., 50 (7),
19 20		2573–2576 DOI 10.1128/AAC.01633-05.
21 22 22	(7)	Kieser, T. (1984) Factors Affecting the Isolation of CCC DNA from Streptomyces Lividans
23 24 25		and Escherichia Coli. Plasmid, 12 (1), 19-36 DOI 10.1016/0147-619X(84)90063-5.
25 26 27	(8)	Cao, G.; Allard, M. W.; Hoffmann, M.; Monday, S. R.; Muruvanda, T.; Luo, Y.; Payne, J.;
28 29		Rump, L.; Meng, K.; Zhao, S.; et al. (2015) Complete Sequences of Six IncA/C Plasmids of
30 31		Multidrug-Resistant Salmonella Enterica Subsp. Enterica Serotype Newport. Genome
32 33		Announc., 3 (1), e00027-15 DOI 10.1128/genomeA.00027-15.
34 35	(9)	Kotsakis, S. D.; Miriagou, V.; Vetouli, E. E.; Bozavoutoglou, E.; Lebessi, E.; Tzelepi, E.;
36 37		Tzouvelekis, L. S. (2015) Increased Hydrolysis of Oximino-β-Lactams by CMY-107, a
38 39		Tyr199Cys Mutant Form of CMY-2 Produced by Escherichia Coli. Antimicrob. Agents
40 41 42		Chemother., 59 (12), 7894–7898 DOI 10.1128/AAC.01793-15.
42 43	(10)	Endimiani, A.; Doi, Y.; Bethel, C. R.; Taracila, M.; Adams-Haduch, J. M.; O'Keefe, A.;
44 45 46		Hujer, A. M.; Paterson, D. L.; Skalweit, M. J.; Page, M. G. P.; et al. (2010) Enhancing
47 48		Resistance to Cephalosporins in Class C Beta-Lactamases: Impact of Gly214Glu in CMY-2.
49 50		Biochemistry, 49 (5), 1014–1023 DOI 10.1021/bi9015549.
51 52	(11)	Bauvois, C.; Ibuka, A. S.; Celso, A.; Alba, J.; Ishii, Y.; Frère, JM.; Galleni, M. (2005)
53 54		Kinetic Properties of Four Plasmid-Mediated AmpC Beta-Lactamases. Antimicrob. Agents
55 56		Chemother., 49 (10), 4240–4246 DOI 10.1128/AAC.49.10.4240-4246.2005.
57 58		29
59 60		ACS Paragon Plus Environment

- (12) Dahyot, S.; Mammeri, H. (2012) Hydrolysis Spectrum Extension of CMY-2-like β Lactamases Resulting from Structural Alteration in the Y-X-N Loop. *Antimicrob. Agents Chemother.*, 56 (3), 1151–1156 DOI 10.1128/AAC.05630-11.
- (13) Skalweit, M. J.; Li, M.; Conklin, B. C.; Taracila, M. A.; Hutton, R. A. (2013) N152G, -S, and -T Substitutions in CMY-2 β-Lactamase Increase Catalytic Efficiency for Cefoxitin and Inactivation Rates for Tazobactam. *Antimicrob. Agents Chemother.*, *57* (4), 1596–1602 DOI 10.1128/AAC.01334-12.
- (14) Kotsakis, S. D.; Papagiannitsis, C. C.; Tzelepi, E.; Tzouvelekis, L. S.; Miriagou, V. (2009) Extended-Spectrum Properties of CMY-30, a Val211Gly Mutant of CMY-2 Cephalosporinase. *Antimicrob. Agents Chemother.*, *53* (8), 3520–3523 DOI 10.1128/AAC.00219-09.
- Kusumoto, M.; Ooka, T.; Nishiya, Y.; Ogura, Y.; Saito, T.; Sekine, Y.; Iwata, T.; Akiba, M.;
 Hayashi, T. (2011) Insertion Sequence-Excision Enhancer Removes Transposable Elements
 from Bacterial Genomes and Induces Various Genomic Deletions. *Nat. Commun.*, 2 (1), 152
 DOI 10.1038/ncomms1152.
- Lefurgy, S. T.; Malashkevich, V. N.; Aguilan, J. T.; Nieves, E.; Mundorff, E. C.; Biju, B.;
 Noel, M. A.; Toro, R.; Baiwir, D.; Papp-Wallace, K. M.; et al. (2016) Analysis of the
 Structure and Function of FOX-4 Cephamycinase. *Antimicrob. Agents Chemother.*, 60 (2),
 717–728 DOI 10.1128/AAC.01887-15.
- (17) Thomas, V. L.; McReynolds, A. C.; Shoichet, B. K. (2010) Structural Bases for Stability-Function Tradeoffs in Antibiotic Resistance. *J. Mol. Biol.*, *396* (1), 47–59 DOI 10.1016/j.jmb.2009.11.005.
- Berrazeg, M.; Jeannot, K.; Ntsogo Enguéné, V. Y.; Broutin, I.; Loeffert, S.; Fournier, D.;
 Plésiat, P. (2015) Mutations in β-Lactamase AmpC Increase Resistance of Pseudomonas
 Aeruginosa Isolates to Antipseudomonal Cephalosporins. *Antimicrob. Agents Chemother.*,
 59 (10), 6248–6255 DOI 10.1128/AAC.00825-15.

ACS Paragon Plus Environment

ACS Infectious Diseases

- A. Powers, R. (2016) Structural and Functional Aspects of Extended-Spectrum AmpC (19) Cephalosporinases. Curr. Drug Targets, 17 (9), 1051-1060 DOI 10.2174/1573399811666150615144707. (20)Kotsakis, S. D.; Tzouvelekis, L. S.; Petinaki, E.; Tzelepi, E.; Miriagou, V. (2011) Effects of the Val211Gly Substitution on Molecular Dynamics of the CMY-2 Cephalosporinase: Implications on Hydrolysis of Expanded-Spectrum Cephalosporins. Proteins Struct. Funct. Bioinforma., 79 (11), 3180-3192 DOI 10.1002/prot.23150. (21)Dortet, L.; Oueslati, S.; Jeannot, K.; Tandé, D.; Naas, T.; Nordmann, P. (2015) Genetic and Biochemical Characterization of OXA-405, an OXA-48-Type Extended-Spectrum β -Lactamase without Significant Carbapenemase Activity. Antimicrob. Agents Chemother., 59 (7), 3823-3828 DOI 10.1128/AAC.05058-14. (22)Zankari, E.; Hasman, H.; Cosentino, S.; Vestergaard, M.; Rasmussen, S.; Lund, O.; Aarestrup, F. M.; Larsen, M. V. (2012) Identification of Acquired Antimicrobial Resistance Genes. J. Antimicrob. Chemother., 67 (11), 2640–2644 DOI 10.1093/jac/dks261. Carattoli, A.; Zankari, E.; García-Fernández, A.; Voldby Larsen, M.; Lund, O.; Villa, L.; (23)Møller Aarestrup, F.; Hasman, H. (2014) In Silico Detection and Typing of Plasmids Using PlasmidFinder and Plasmid Multilocus Sequence Typing. Antimicrob. Agents Chemother., 58 (7), 3895-3903 DOI 10.1128/AAC.02412-14. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. (1990) Basic Local (24)Alignment Search Tool. J. Mol. Biol., 215 (3), 403-410 DOI 10.1016/S0022-2836(05)80360-2.
 - (25) Carver, T. J.; Rutherford, K. M.; Berriman, M.; Rajandream, M.-A.; Barrell, B. G.; Parkhill, J. (2005) ACT: The Artemis Comparison Tool. *Bioinformatics*, *21* (16), 3422–3423 DOI 10.1093/bioinformatics/bti553.
 - (26) Bradford, M. M. (1976) A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.*, 72,

248-254.

- (27) Kabsch, W. (2010) Integration, Scaling, Space-Group Assignment and Post-Refinement.
 Acta Crystallogr. D. Biol. Crystallogr., 66 (Pt 2), 133–144 DOI
 10.1107/S0907444909047374.
- (28) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.;
 Bricogne, G. (2011) Data Processing and Analysis with the *AutoPROC* Toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 67 (4), 293–302 DOI 10.1107/S0907444911007773.
- (29) Evans, P. R.; Murshudov, G. N. (2013) How Good Are My Data and What Is the Resolution? *Acta Crystallogr. D. Biol. Crystallogr.*, 69 (Pt 7), 1204–1214 DOI 10.1107/S0907444913000061.
- Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.;
 Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. (2011) Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D. Biol. Crystallogr.*, 67 (Pt 4), 235–242 DOI 10.1107/S0907444910045749.
- McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read,
 R. J. (2007) *Phaser* Crystallographic Software. *J. Appl. Crystallogr.*, 40 (4), 658–674 DOI 10.1107/S0021889807021206.
- (32) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. (2010) Features and Development of *Coot. Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66 (4), 486–501 DOI 10.1107/S0907444910007493.
- Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.;
 Sharff, A.; Smart, O. S.; Vonrhein, C.; et al. (2011) AutoBUSTER, Version 1.6. 0. *Glob. Phasing Ltd, Cambridge, UK.*,.
- (34) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.;
 Winn, M. D.; Long, F.; Vagin, A. A. (2011) *REFMAC* 5 for the Refinement of
 Macromolecular Crystal Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 67 (4),

355–367 DOI 10.1107/S0907444911001314.

- (35) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.;
 Murray, L. W.; Richardson, J. S.; Richardson, D. C. (2010) *MolProbity* : All-Atom Structure
 Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66 (1), 12–21 DOI 10.1107/S0907444909042073.
- (36) Schrödinger, L. *The PyMOL Molecular Graphics System, Version~1.8.*
- (37) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. (2004) UCSF Chimera A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.*, 25 (13), 1605–1612 DOI 10.1002/jcc.20084.

For Table of Contents Use Only

CMY-136 β-lactamase



Supplementary Information

Genetic, biochemical and structural characterization of CMY-136 β-lactamase, a peculiar CMY-2 variant

Agustin Zavala^{1,2}, Pascal Retailleau¹, Eddy Elisée¹, Bogdan I. Iorga^{1*}, Thierry Naas^{2,3,4,5*}

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, LabEx LERMIT, 1 avenue de la Terrasse, Bât. 27, 91198 Gif-sur-Yvette, France

² EA7361 "Structure, dynamic, function and expression of broad spectrum β-lactamases", Université Paris Sud, Université Paris Saclay, LabEx LERMIT, Faculty of Medicine, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France

³ Bacteriology-Hygiene unit, Assistance Publique/Hôpitaux de Paris, Bicêtre Hospital, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France

⁴ Associated French National Reference Center for Antibiotic Resistance: Carbapenemaseproducing Enterobacteriaceae, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France

⁵ Evolution and Ecology of Resistance to Antibiotics Unit, Institut Pasteur – APHP -Université Paris Sud, 25-28 Rue du Dr Roux, 75015 Paris, France

*To whom correspondence should be addressed:

Bogdan I. Iorga: Institut de Chimie des Substances Naturelles, CNRS UPR 2301, 91198 Gif-sur-Yvette, France. E-mail: bogdan.iorga@cnrs.fr; Tel. +33 1 69 82 30 94; Fax. +33 1 69 07 72 47

Thierry Naas : Service de Bactériologie-Hygiène, Hôpital de Bicêtre, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France. E-mail: thierry.naas@aphp.fr; Tel: +33 1 45 21 20 19; Fax: +33 1 45 21 63 40

Content Page Fig. S1: Sequence alignment of CMY-2 and CMY-136 S2 Fig. S2: Chemical structures of β -lactams and β -lactamase inhibitors used in this study S3 Fig. S3: Different conformations of Y150 in the CMY-2 and CMY-136 structures S4 Fig. S4: RMSD of the MD simulations S5 Fig. S5: RMSD analysis of Ω -loop from MD trajectories **S6** Fig. S6: R1 and R2 pockets in CMY-2 **S**7 Table S1: Extinction coefficients and wavelengths for β-lactams **S**8



Fig. S1 Sequence alignment of CMY-2 and CMY-136. Sequence alignment of CMY-2 and CMY-136 showing the position of the mutated residue and secondary structure elements as defined for CMY-2 (PDB code 1ZC2).

Characterization of CMY-136 beta-lactamase



Fig. S2 Chemical structures of β -lactams and β -lactamase inhibitors used in this study. The chemical structures of the β -lactam antibiotics or β -lactamase inhibitors mentioned in this study (MICs, enzyme kinetics or molecular modelling experiments) are shown here. The R1 and R2 substituents are colored in red and blue, respectively.



Fig. S3 Different conformations of Y150 in the CMY-2 and CMY-136 structures. Superposition of CMY-2 (green) and CMY-136 (cyan) structures shows that the conformation adopted by Y150, a mechanistically important residue, in the CMY-136 structure differs from that observed in CMY-2, with a shift of 3.1 Å in its hydroxyl position. Other important residues are shown for reference.



Fig. S4 RMSD of the MD simulations. RMSD of the 10 ns simulations of CMY-2 (A) and CMY-136 (B) carried out in triplicate.



Fig. S5. RMSD analysis of Ω-loop from MD trajectories. Ω-loop Cα RMSD analysis of the MD trajectories compared to CMY-2 and CMY-136 crystal structures. RMSD was calculated between crystal structure Ω-loops, and MD simulation snapshots taken every 1 ns. Regardless of starting conformation, CMY-136 structures tend to converge to a conformation closer to that of CMY-136 crystal structure, and far from that of CMY-2 crystal structure. For CMY-2 simulations, when starting from the native structure conformation, Ω-loop conformation is maintained closer to it than to that of CMY-136 Ω-loop crystal structure. When starting from a CMY-136 Ω-loop conformation, however, CMY-2 Ω-loop conformation does not converge towards its own Ω-loop crystal structure conformation. Ω-loop RMSD between both crystal structures is represented by the purple dashed line. "CMY-2 MD": simulations for the CMY-2 structure starting from its own crystal structure conformation; "CMY-136 MD": simulations for the CMY-136 structure starting from its own crystal structure conformation; "CMY-136 MD": simulations for the CMY-136 structure starting from its own crystal structure conformation; "CMY-2(conf. 136) MD": simulations for the CMY-2 structure Ω -loop conformation; "CMY-136 crystal structure Ω -loop modelled on the CMY-136 crystal structure Ω -loop conformation; "CMY-136 crystal structure Ω -loop conforma



Fig. S6 R1 and R2 pockets in CMY-2. Surface representation of CMY-2 showing the different parts of the protein defining the R1 and R2 pockets. The R2 pocket (colored in red) is usually defined as surrounded by the R2 loop (residues 289-307) and the proximal parts of helixes H10 and H11. The R1 pocket (colored in yellow) is usually defined as surrounded by the Ω loop (residues 178-226). Apart from this, the R1 substituent of ceftazidime can interact with other surrounding residues: Q120, N152, S343 and residues 318-320. For reference, ceftazidime is shown as green sticks.

Table S1 Extinction coefficients and wavelengths for β -lactams. Extinction coefficients and

Substrate	Wavelength (nm)	$\Delta \epsilon (M^{-1} cm^{-1})$
Benzylpenicillin	232	1,100
Cloxacillin	220	700
Ampicillin	235	1,050
Ticarcillin	235	1,050
Cefaloridine	255	9,360
Cefalotin	262	7,960
Cefoxitin	265	7,380
Cefuroxime	262	7,800
Cefotaxime	265	6,260
Ceftazidime	260	8,660
Cefepime	264	8,240
Ceftolozane	259	1,900
Imipenem	297	9,210
Aztreonam	318	640

wavelengths used for the different substrates in the kinetic parameter determinations.

MD simulations of MBL–ligand complexes

After validation of parameters for the new OPLS-AA residues on ligandfree metalloproteins, I wanted to test their robustness on MBL-ligand complexes. This work was carried out at the beginning of my thesis, therefore, given that at that time I hadn't parameterized any β -lactam antibiotics, I chose the representative NDM-1 enzyme and two easy-to-parameterize succinic acid inhibitors of IMP-1.

3.1 System setup

3.1.1 Protein

The selected NDM-1 structure 4hl2 (Figure 24) was prepared as follows: (1) zinc-coordinating residues were renamed according to the new developed residues (see Chapter II), namely HMD for HIS[120, 189, 250], HME for HIS122 and CYSM for CYS208, (2) ampicillin ligand, water and all non-protein molecules – except the two active site zinc ions – were deleted, and (3) checking of the structure revealed no missing atoms and residues.



Figure 24: Crystallographic structure of NDM-1 (PDB code 4hl2). Non-protein residues (water, ligands, non-zinc ions) were removed for the MD simulation.

3.1.2 Ligands

Considering the structural similarity between IMP-1 and NDM-1 structures (metallo- β -lactamase B1 subclass), two IMP-1 structures each containing a succinic acid inhibitor have been aligned on NDM-1 (Figure 25). This alignment revealed no important steric hindrance (Figure 25.A), therefore initial poses for MD simulations were not obtained by docking calculations but by transferring the BYS (PDB code 1jje) and BDS (PDB code 1jjt) inhibitors' coordinates into the NDM-1 structure (Figure 25.B-C). Then, the two compounds were successfully parameterized using MOL2FF software.



Figure 25: Structures of NDM–1 and IMP–1 enzymes. (A) Structural alignment of NDM–1 (PDB code 4hl2) and IMP–1 (PDB code 1jjt) colored in medium purple and dark cyan, respectively. Surface representation of NDM–1 enzyme (PDB code 4hl2) complexed with (B) BYS inhibitor (PDB code 1jje) and (C) BDS (PDB code 1jjt) inhibitor.

3.1.3 MD simulations

Simulations were carried out using OPLS-AA force field implemented in GROMACS version 4.6.0. The protocol is the same as reported in Chapter II. Both NDM-1–inhibitor complexes contain around 54,000 atoms including protein, water and ions. Production steps included one 100 ns simulation for each complex and for the apo NDM-1 structure, namely with no ligand.

3.2 Results and Discussion

First, a 100 ns MD simulation was performed using a prepared apostructure of NDM-1 and its dynamic evolution has been analysed by computing Root-Mean-Square Deviation (RMSD) of the backbone (Figure 26).





Outcomes revealed an overall stability of the protein even if deeper investigations on the active site Zn-Zn distance pinpointed a sligth increase, from 3.5 Å to 5 Å, due to repulsion effects. This issue has been fixed later on, by setting a non-covalent bond between the two zinc ions (see Chapter II).

Then, NDM-1 enzyme complexed with inhibitors have been studied through 100 ns MD simulations and its dynamic evolution has been also described under RMSD calculations reported in Figure 27. From these two RMSD plots, we may draw some common conclusions: (1) absence of drastic changes in the RMSD curves reveals a probable equilibrium in protein stability, even if a simple RMSD analysis may be not enough to fully characterize an equilibrium state [114], (2) RMSD fluctuations through intermediate states underline a certain flexibility of the protein active site, especially in cases (a) and (d) where phenyl and 1,3-benzodioxole groups evolved from horizontal to vertical conformations, and (3) phenyl group seems to be less involved in interactions with the protein than the 1,3-benzodioxole one.

Otherwise, zinc-coordination, achieved by protein residues and inhibitors, is not disturbed along both simulations as well as the distance between zinc ions, even without setting a non-covalent bond.



Figure 27: RMSD analysis of NDM-1–inhibitor complexes along 100 ns simulations. In both plots, black color and red color refer to NDM-1 backbone and ligand atoms, respectively. (A) BDS inhibitor. (B) BYS inhibitor. For each screenshot (a-e), initial/intermediate poses are colored in (a) beige/green, (b) beige/blue, (c) blue/pink, (d) beige/pink, and (e) beige/blue.

In this section, we learnt that MD simulations of MBLs, with or without ligand, may be carried out in a stable way using new developed OPLS-AA residues and a correct ligand parameterization. Indeed, electrostatic charges seem to be sufficient to ensure a stable system in the presence of a ligand – assumption supported by other simulations on MBLs complexed with peptide-like ligands (confidential data). In addition, suitable MD initial poses may be obtained in some cases by aligning related crystallographic protein structures instead of running docking calculations.

MD simulations of an intriguing SPM-1 glycine mutant

A powerful search tool has been developed in our team, named ResMiner, to retrieve β -lactamase sequences in genomic and metagenomic data. Coming from one of those searches, six interesting MBLs sequences were identified. Their closest known analogs were São Paulo metallo- β -lactamase 1 (SPM-1) sequences with 35% of sequence identity; therefore we decided to term them SPM-like.

However, SPM-1 belongs to the MBL B1 subclass in which one of the two coordination pattern is normally HxHxD, with x being a "random" amino acid. But in these new sequences, the first histidine (H) is replaced by a glycine (G), which yields a GxHxD pattern (Figure 28).

	1	11	21	31	41	51
4BP0-A OGU548	1					M E R G
WP_013 WP_050 WP_053	1	T P R M T R A V L A	I L W L L P S L A L	V S V A C G G S E H	V T E V A S V S S T	T S A P G V A S R S
WP_061 WP_083	1	MDYICKH	DAFIANSIEK	EDHIIPFIN	LIKEEWRMKN	IAIFVLVGVF
4BP0-A OGU548	61 1	71 F L L V N V H H A Q	81 G P K S S N N N F K T V K V N	91 D H V D L P Y N L T D Y L T	101 A T K I D S D V F V I N E I E K D V Y R	111 V T D R D F Y S S N V S H R F P WQ A N
WP_013 WP_050 WP_053	12 R N V M L I I F F V 61 A P A A V V Q V K G 1 M S K K V K F	L T D L S L I A Q T G E A A G A L P A S C L L L L O T L T	Q P E Y P V I R L N A E G L P E T K L S E T S C L T N G T N	D E L E E D L V P K Y K L S E E L O	V R E I L P N A F V I R Q L A P D V H L I S K I K K D T V I	I T H K F P W G G N I T H T F P W P A N V T H S Y P W P A N
WP_061 WP_083	1	RLNVLSQEKQ	VGLPANERVF	PIGEIGDKLY	AEEIQESVYV	Q A N I N H E F P W P S N
4BP0-A OGU548 WP_013 WP_050 WP_053 WP_061	121 36 V L V A K M L D G T 59 S L L V R V G D K D 66 S L V V I G E K Y 115 A L L V E M A N G D 58 S L L V L M N S K Y 4 S M V V D I S P K D	131 V V I V S S P F E N F V L A D T P I E N A V F V D T P Y T P L V L C D P T Y T V I L W I D T P Y T P L L I I D A P W T S	141 L G T Q T L M D W V T G S E A L V E W M E A T E N V L D W I D A M R L V L A W M E A T A Q V L K W I A S T R D L L K W A	151 A K T M K P . K K V R K E F G D . I K L N K Q Y G N . R Q F D E R Y G K . R R I D E E I G S G Y S L K D T F G D . R K I	161 VAINTHFHLD TVINCGFHID IEINTGYHVD VALNTGFHID VEINTGFHID TAINTGSHMD	171 G T G G N E I Y K K C L G G N H Y M I N N L G G N D A L L H R V G G N A A L L E N L G G N Q E L I K R V A G N D V F L A
WP_083 4BP0-A	108 S M V V E M E T G D 181 95 M G A E T W S S D L	IVLVDTPYTP 191 TKQLRLEENK	AATKQLLDWI 201 KDRI.KA.A	HKRFGE.REI 211 EFYKNEDLKR	I A I N T <mark>G F H F D</mark> 221 R I L S S H P V P A	NLGGNQVLVQ 231 DNVFDLKQG.
WP_013 WP_050 WP_053 WP_061 WP_083	125 R N I P I I G S D K 174 R G I P V Y G S D L 118 R N I P I Y G S D L 63 E G A D V Y S S D L 167 K D I P V Y G S S L	T V S L L R E R G E T A R L V T T R S E T R Q L L D S N S A T I K Y V K K T E L L I Y E Y G E	A T R Q L T M G W L A H R A W V R A S V S T M A D M H I W L S K P T E L K S L A A S R E L M L S W L	E G P G N E K F L K D D G A I A A Q D E K Y T K Y R D S R V S D P A I K A N K P E T K E L Y N	STKNLQPTAP RHETIPYVGP VFETQPMVPP VYSQFIFYKP EFAKMKIREA VYQQIPYVPP	DHIFKAEEG. SQIFQLTEG. DRVFPIAEG. THTFDINIE. NHSFPLKEG. TEIFNLDKVD
4BP0-A OGU548 WP_013 WP_050 WP_053 WP_061 WP_083	241 151 . K V F S F S N E L 174 . L T L H I G K E D 184 . Y H F T V G D E P 230 . L S L T F G E E K 177 . Q S I K L G K N E 119 . K T L S F G E K K 227 L Q N F R F G N E E	251 V E V S F P G P A H I D I Y Y P G P A H I E V F F P G E T H V V V H H P G P G H A I I Y Y P G P T H A E I F Y P G H G H V E I Y Y P G Q S H	261 S P D N V V V Y F P T K D N I V V F F P A P D N I V V Y F P A P D N V V V F P T Y D N L V V Y L P T K D N V V V Y L P S P D N L V V Y F P 321	271 KKKLLFGGCM EKGILYGGCM ERKILFGGCM AHRLLFGGCM DYKILFGGCM DYKILFGGCM 231	281 I K P K E L G V K A L E Q S N L G L R V G N G T G V A A G E R I G I F N A N A K K M G I V G L P K L G I K S L D S K D L G 241	291 Y L G D A N V K AW N Q S D A D F E AW N R AD A N M D T W N T T D G D L A R W Y V K D G N I E E W Y I K E A D L K EW N T A D A N L Q E W
4BP0-A OGU548 WP_013 WP_050 WP_053 WP_061 WP_083	207 P D S A R R L K 233 P A S A Q E L L D K 241 K S S V E R L R 287 G D A I R D L Q 236 S K S L E Y L K E R 176 P K A L D K M S 287 P K S V K K V . F R	K F D A K I V I P G Y P D A K I V V P G D F D C V A V I P G R F K A F S I V P G F P D I K T V I P G R F D A K W V I P G	HGEWGGPEMV HGMWGDLIIV HGIRFDPGVI HGSRLDPGLL HGSPGDSTLI HGTSYSPDLI	NKTIKVAEKA RHTLELLKKR ENTISVLP DHTLALLASA AHTKEIIDAS EHTKKLVK KYTLOLCEFK	V G E M R L S P K	

Figure 28: SPM-1 and SPM-like sequences alignment. Red and black frames refer to H/GxHxD zinc-coordination pattern and a highly conserved apartate residue, respectively. 4BP0-A: SPM-1 crystallographic sequence (PDB code 4bp0, chain A). OGU548 and the WP series: the six SPM-like sequences. Alignment done with UCSF Chimera software which uses the Clustal X Colour Scheme [http://www.jalview.org/help/html/colourSchemes/clustal.html]. Alternatively, other MBLs subclasses display slightly different patterns such as NxHxD for B2 subclass, with an asparagine (N) in place of histidine, and HxHxDH for B3 subclass.

Looking at these new sequences, we asked ourselves a priori if this glycine mutation could be concerted with other mutations not involved in zinc coordination but allowing another residue to replace the coordination initially operated by histidine. In other words, could another residue appropriately positioned replace histidine in zinc coordination?

After investigations, an aspartate residue was found to be highly conserved among the seven sequences (Figure 28) and positioned on a flexible loop. The sidechain of this aspartate is located at 7.4 Å from the zinc ion coordinated by the mutated histidine (Figure 29).



Figure 29: SPM-1 crystallographic structure (PDB code 4bp0). Non-protein residues (water, ligands, non-zinc ions) were removed. On the left, an overview of SPM-1 secondary structures. On the right, a zoom on the active site, with the mutated histidine and the highly conserved aspartate residues colored in green and dark cyan, respectively.

This way, my objective was to run quite long MD simulations of SPM-1 and SPM-like histidine-to-glycine mutants in order to validate or not our assumption. It was also another opportunity to assess the parameters for our new OPLS-AA residues presented in Chapter II.

4.1 System setup

4.1.1 Protein

Only three crystallographic structures were available for SPM-1 at the time of experiments: 2fhx (only one zinc ion), 5ls3 (Y58C mutant and missing residues) and the selected one, 4bp0 (complete). Nowadays, there are two additional available structures: 5ndb (complexed with cyclobutanone inhibitor) and 5nde (apo-form). After a structural alignment, no difference was observed compared to 4bp0, except Phe151 residue which is slightly buried in the active site to interact with the inhibitor's dichloro-substituent.

The selected SPM-1 structure 4bp0 (Figure 29) was prepared as follows: (1) zinc-coordinating residues were renamed under the new developed residues (see Chapter II), namely HMD for HIS[108, 197, 258], HME for HIS110 and CYSM for CYS216, (2) water and all non-protein molecules – except the two active site zinc ions – were deleted, and (3) checking of the structure revealed no missing atoms and residues.

For each SPM-like sequence, fifty models were built by homology modelling on 4bp0 structure, using UCSF Chimera Interface to MODELLER, and one was finally selected according to the best DOPE score. The six selected models were then prepared following the same protocol as for 4bp0 and for which atom equivalence is reported in the Table 3.

	His/Gly	His	His	Asp	Cys	His	Asp
4BP0-A	108	110	197	112	216	258	200
WP_061	80	82	169	84	188	226	172
WP_050	79	81	168	83	187	225	171
OGU548	79	81	168	83	187	229	171
WP_053	80	82	172	84	191	233	175
WP_083	80	82	174	84	193	235	177
WP_013	79	81	171	83	190	228	174

Table 3: Atom equivalence for SPM-1 (4BP0-A) and SPM-like (OGU548 and WP series) structures. Colors refers to the active site residues depicted in Figure 29: green for the mutated histidine and dark cyan for the highly conserved aspartate.

4.1.2 MD simulations

Simulations were carried out using OPLS-AA force field implemented in GROMACS version 4.6.0. The protocol is the same as reported in Chapter II. All systems contain from 49,000 atoms (WP_013) to 67,000 atoms (SPM-1), including protein, water and ions. Production steps included one 500 ns simulation per system, which represents about a one-week calculation time on 40-cores clustered computers.

4.2 Results and Discussion

Analysis of 500 ns trajectories required to keep only 1 frame over 100. Then, overall analysis of each system have been visually undertaken. However, atom numbering may have changed during the different MD preparation steps, therefore new numberings for histidine-to-glycine mutation (G) and highly conserved aspartate (D) are reported below just after the corresponding sequence name.

SPM-1 (D200/G108). At the beginning, D200 is located at 4.5 Å from zinc ion and coordinated to it (\sim 2 Å) starting from frame 25.

OGU548 (D166/G74), WP_061 (D140/G48) and WP_083 (D177/G80). D166, D140 and D177 can not access to the active site since they are involved in several hydrogen bonds all along the simulation.

WP_013 (D169/G74). Since the first frame, D169 interact with zinc ion and this interaction is conserved until the end.

WP_050 (D166/G74). D166 is blocked by F75 residue and the active site is practically unreachable all along the simulation.

WP_053 (D176/G81). D176 is buried in the protein, and moves away from the active site until reaching the solvent-exposed surface.

Regarding the previous results, only two simulations validated our assumption: SPM-1 and WP_013. Nevertheless, these positive results might come from simulation artefacts since further investigations on outputs stem from the energy minimization step revealed some defects. In the case of SPM-1 and WP_013, minimized structures showed an aspartate very close to the zinc ion (between 4 and 5 Å). Thus, we may assume this convergence is due to either an "environmental factor", namely with no cumbersome residues, or an issue in parameterization of zinc-zinc interaction. The latter assumption seemed to be supported by the analysis of zinc-zinc distances which underscored an important point: distances were longer than in the initial 4bp0 structure (3.4 Å), with values around 5–6 Å. Ergo, we doubted about the veracity of our positive results since the zinc coordination could come from the zinc displacement, i.e. zinc repulsion, rather than from the aspartate. One solution will be to run new simulations with a non-covalent bond set between the two zinc ions, as defined in Chapter II, and observe if the aspartate residue will move towards the zinc in these conditions.

I also ran a 500 ns MD simulation of NDM-1 in which Asp192 is located at the same place as Asp200 in SPM-1. Then, I created the H120G mutant and followed the same MD protocol. Results did not show a new zinc coordination by the Asp192 residue and distance between zinc ions was very large (\sim 8 Å).

In contrast, I carried out a survey upon the PDB website by looking for structures of representative MBLs from B2 and B3 subclasses, in order to study the presence or absence of this highly conserved aspartate residue (Table 4).

Table 4:	PDB	survey	on some	MBLs	from	B2	(CphA)	and	B3	(L1,	SMB-1,	BJP-1,
	AIM-	1, FEZ-	1, GOB-1	.8, RM3	3) sub	clas	ses.					

CphA	L1	SMB-1	BJP-1	AIM-1	FEZ-1	GOB-18	RM3
1X8G 1X8H 1X8I 2GKL 2QDS 3F9O 3FAI 3IOF 3IOG 3SW3 3T9M	1SML 2AIO 2FM6 2FU6 2FU7 2FU8 2FU9 2GFJ 2GFK 2H6A 2HB9 2QDT 2QIN 2QJS 5DPX 5EVB 5EVD 5EVK 5HH5 5HH6	3VPE 3VQZ 5AXO 5AXR 5AYA 5BI5 5BIU	2GMN 3LVZ 3M8T 5NGG 5N5W 5WCM	4AWZ 4P62 4AWY 4AX0 4AX1	1JT1 1K07 1L9Y	5KOW	5IQK

In B2 subclass, structures of only one representative MBL can be found, that is CphA, in which the equivalent residue of Asp200 in SPM-1 is far from the HHH site, so there is no point in mutating histidine to glycine.

In B3 subclass, the L1, AIM-1, RM3, SMB-1 and GOB-18 enzymes have a glycine instead of the aspartate of interest, and no adjacent aspartate has been observed.

Moreover, I could reveal another interesting residue at position 220 (an aspartate in AIM-1, RM3 and SMB-1, an asparagine in GOB-18, and a cysteine in BJP-1) which potentially could coordinate the zinc ion if His116 was mutated to glycine. This is in contrast with FEZ-1, which has a glycine at this postion 220.

Otherwise, some MBLs display a coordination pattern with less coordinating residues, and that does not affect the active site structure, for example L1 (B3 subclass). Therefore, even if such kind of mutations – histidine to glycine – may not affect the β -lactamase activity, e.g. in *Sediminispirochaeta smaragdinae* metallo- β -lactamase (SPS-1) enzyme [21], it is also possible that no residue replace the mutated histidine. Nonetheless, a recent study published a crystallographic structure of the histidine-to-glycine SPS-1 mutant revealing the presence of a water molecule as the third zinc-coordinating group [43]. Given the static property of crystallographic structures, there is no reason to suggest that this observation reflects necessarily the reality or the dynamic evolution of the enzyme in which some neighbouring residue might also play a part in zinc coordination or not.

In this section, we learnt that zinc-coordinating histidine mutation into a glycine residue might potentially lead to a structural rearrangement of the active site, especially through the movement of an highly conserved aspartate residue. Nevertheless, this observation was not fully validated due to an artifact with the zinc-zinc distance in the MD simulations that is now fixed. The recent publication of a crystallographic structure for this mutant showed the coordination of a water molecule instead of the movement of the aspartate residue. In these conditions, new MD simulations should be carried out using the noncovalent zinc-zinc bond developed in Chapter II.

In addition, MD simulations could yield more or less reliable results as long as a suitable and well-parameterized system is used; this case is one more example of how challenging it is to identify false positive outcomes in order to avoid drawing misleading conclusions.

OXA β-lactamases studies using free energy calculations

In this study, we focused on the analysis of oxacillinases (OXA), which belong to the D class in serine- β -lactamase family, through alchemical free energy calculations. Normally, these enzymes display a peculiar feature: for example in OXA-48, lysine residue (Lys73) located near the catalytic serine (Ser70) has to be carboxylated, in the presence of CO₂, in order to achieve β -lactam hydrolysis. Indeed, the carboxylated Lys73, also termed KCX residue, is involved in several hydrogen bonds with other active site residues through its new terminal carbamate group and seems to be involved in (1) the catalytic Ser70 activation during the acylation process, and (2) the activation of a water molecule to hydrolyse the acyl-enzyme complex. For further details, one can read the updated review written by Tooke *et al.* [186].

Nevertheless, some of these enzymes are not or only partially carboxylated, as shown by the crystal structures of OXA-145 [137] or OXA-427 (Zavala, A., Retailleau, P., Bogaerts, P., Glupczynski, Y., Naas, T., Iorga, B.I., unpublished), with no clear reasons. Therefore, given that the carbamylation of this residue involves two chemical transformations, namely either protonation or carboxylation (Figure 30), we investigated the influence of the active site environment on this process, using free energy calculations.



Figure 30: Protonation and carboxylation reactions of the lysine. Depending on active site conditions, neutral lysine (middle) may be protonated (right) or carboxylated (left).

In the case where lysine is not carboxylated, three reasons are plausible: (1) the lysine is protonated, so no possible carboxylation, (2) atmospheric carbon dioxide could not enter the active site, or (3) the carboxylation occurs but decarboxylation happens shortly after, in other words it is more favorable for the lysine to be neutral than being carbamylated. When it comes to assess if a process is favorable or not, a powerful computational strategy is to undertake free energy calculations.

This way, we decided to study the protonation $(NH_2 \rightleftharpoons NH_3^+)$ and the carboxylation $(NH_2 \rightleftharpoons KCX)$ reactions using an in-house modified version of the software PMX developed by de Groot's team which implies non-equilibrium free energy calculations [5, 74–77].

5.1 Alchemical PMX philosophy

The PMX protocol is an alchemical-based method to compute free energies whose detailed workflow is reported in Figure 5 of the D3R-GC2 paper (Chapter IV section 2.1), but outlines will be also described hereafter.

The philosophy of alchemical transformations is based on the possibility to simulate atoms' modification, by changing their intrinsic properties, resulting in totally different or non-physical (dummy) species. That change corresponds to a certain amount of free energy which can be computed along the simulation. In order to achieve that, a thermodynamic cycle describing folded and unfolded states may be drawn (Figure 31).



Figure 31: Thermodynamic cycle for estimating the difference in stability between wild type and mutant.

Nevertheless, vertical paths are not computationnally suitable since the folding time of a small peptide might be limiting; thus, folding free energies ΔG_2 and ΔG_3 can not be directly obtained. One solution would be to get free energies ΔG_1 and ΔG_4 from horizontal paths by running alchemical mutations in which an amino acid A is morphed into another amino acid B, and the corresponding free energy change is computed.

The key point of such alchemical method is the generation of hybrid topologies in which both aminoacids A and B contribute under a timedependent coordinate λ describing their completeness state. Then, $\lambda = 0$ and $\lambda = 1$ will refer to a complete atomic description of A and B residues, respectively; for example, at $\lambda = 0.3$, the system is defined by 30% of residue B and 70% of residue A; at $\lambda = 0.8$, it will include 80% of residue B and 20% of residue A. On this basis, if this transformation is performed in both folded and unfolded states, therefore the sum of all free energy values ($\Delta\Delta G$) must equal zero because free energy is a state variable, i.e. independent of the chemical path:

$$\Delta\Delta G = \Delta G_1 + \Delta G_2 - \Delta G_3 - \Delta G_4 = 0$$
$$\Delta\Delta G = \Delta G_3 - \Delta G_2 = \Delta G_1 - \Delta G_4$$

Given that the structural information on unfolded protein is unknown, the unfolded state is assumed to be modeled by a small peptide.

5.2 System setup

5.2.1 Protein

The selected OXA-48, OXA-145, and OXA-427 structures were downloaded from PDB website (Figure 32) and prepared as follows: (1) ligands, water and all non-protein molecules were deleted, and (2) checking of the structure to complete missing atoms and residues.



Figure 32: Crystallographic structures of OXA-48 (A, PDB code 4s2p), OXA-145 (B, PDB code 4yin) and OXA-427 (C, PDB code 6huh, not yet released). Lysine of interest and catalytic serine are colored in coral.

5.2.2 Free energy calculations

Simulations were carried out using OPLS-AA force field implemented in GROMACS version 4.6.0. Production steps included one equilibrium 10 ns simulation for every state A ($\lambda = 0$) and B ($\lambda = 1$). Snapshots from equilibrium runs were then extracted (100 from each simulation) and each of them was the starting structure of a 50 ps simulation in which alchemical morph between the two states was undertaken. Finally, work values over every non-equilibrium transition were extracted to estimate free energy differences under the Crooks Fluctuation Theorem and using Crooks-Gaussian Intersection (CGI) and Bennett Acceptance Ratio (BAR) as estimators.

Here, from the neutral lysine to the negatively charged KCX residue or to the positively charged lysine, both carboxylation and protonation reactions involved a mutation that modifies the charge of the residue, so the system was designed as one box including the protein (folded state) and a Gly-X-Gly tripeptide (unfolded state), with X stands for the initial or mutated residue according to the path direction (A \rightarrow B or B \rightarrow A), separated by 30 Å (Figure 33). For example, in A \rightleftharpoons B reaction, the protein residue is converted from state A to state B while the tripeptide residue is converted from state B to state A. This way, the system has a conserved charge along the simulation and resulting free energies may be related to its global work.



Figure 33: System setup for a mutation that modifies the charge of the residue. The system has a conserved charge during the whole simulation. The tripeptide Gly-X-Gly corresponds to the unfolded state of the protein.

In this part, I also adapted the script for system preparation, in which the system protein+tripeptide was not correctly centered in the box. Indeed, GROMACS software centers by default the system on its center of mass (COM) but given the protein size compared to the tripeptide, the overall COM was located near the protein COM; the tripeptide was therefore positioned outside the box.

5.3 Results and Discussion

Results are reported for each selected mutation in the Table 5. Accuracy is assessed as the absolute sum of forward and reverse pathways and outcomes are the mean of three independent simulations. Results in light mauve and white lines were computed using CGI and BAR estimators, respectively.

Nevertheless, Aldeghi et al. [5] recommend using BAR estimator, instead of CGI, since it gathers all available work values from forward and reverse directions without making assumptions about the work distributions' shape. There is another estimator termed Jarzynski which is one-directional, namely it considers only the work values coming from either forward or reverse directions. However, given its exponential form, this estimator is statistically flawed and therefore values at the tail of the work distribution, i.e. rare events in which there is a minor work dissipation, will mostly contribute to the average and free energy will converge slowly to the true value.

CGI and BAR estimators are sensitive to the lack of overlap between the two distributions and large ΔG uncertainties mean an insuficient overlap. There are several tips to increase the quality of the ΔG estimate by improving the overlap: (1) slower transitions will keep the system closer to equilibrium, resulting in reducing the work dissipation, and (2) additional non-equilibrium transitions will increase the probability of observing rare events. Finally, multiple replicates of equilibrium and non-equilibrium simulations are advised to assess the precision of free energy estimates.

In our case, accurate results (less than 2 kcal.mol⁻¹ and low uncertainties) were globally obtained for the protonation process revealing a non-protonation of the lysine residue in all proteins studied. Thus, the noncarboxylation of OXA-145 and OXA-427 might be the consequence of an unfavored carboxylation process. Nevertheless, our preliminary calculations for this carboxylation process are not precise enough to assess this hypothesis and further investigations will be carried out in order to pinpoint the disruptive elements. Indeed, for the carboxylation reaction, means colored in red include positive and negative free energies which are not trustworthy since we don't know which simulation over the three pointed in the right direction, either favorable or unfavorable.

Furthermore, another approach based on a strict loss of atoms, and no longer on the creation of atoms involving use of dummies, will be tested (see Chapter IV section 1.3).

Table 5: Free energy results (kcal.mol⁻¹) on carboxylation and protonation reactions in OXA enzymes of interest. Acc.: Accuracy, KCX: carbamylated lysine. Green and red colors mean validated and to-investigate values, respectively. Numbers in brackets refer to standard deviation. Results in light mauve and white lines were computed using CGI and BAR estimators, respectively.

	$\mathbf{NH}_2 \rightarrow \mathbf{NH}_3^+$	$\mathbf{NH}_3^+ { ightarrow} \mathbf{NH}_2$	Acc.	$\mathbf{NH}_2 { ightarrow} \mathbf{KCX}$	$\textbf{KCX} {\rightarrow} \textbf{NH}_2$	Acc.
OXA-48	12.17 (0.72)	-12.19 (0.62)	0.03	-1.67 (1.01)	1.29 (1,02)	0.38
0/11/10	11.44 (0.63)	-12.27 (0.50)	0.83	-1.82 (0.83)	1.92 (0.76)	0.11
OVA 14E	13.00 (0.66)	-12.12 (0.59)	0.88	-1.77 (1.01)	-1.00 (0.96)	2.77
0/11-1-15	13.46 (0.54)	-11.62 (0.43)	1.84	-1.47 (0.98)	-0.80 (0.51)	2.27
OXA-427	15.01 (1.04)	-13.40 (0.97)	1.61	-0.13 (0.95)	-0.18 (1.06)	0.31
0//-12/	15.53 (1.17)	-13.65 (0.53)	1.87	-1.92 (1.01)	-0.23 (0.70)	2.15
$OXA_{48}(N76I)$	12.54 (0.77)	-11.08 (0.62)	1.46	-1.01 (1.11)	1.95 (0.94)	0.94
OM-+0(N/0L)	12.58 (0.74)	-11.38 (0.66)	1.20	-2.01 (0.93)	2.47 (0.94)	0.46
OXA AR(N76S)	10.93 (0.71)	-12.54 (0.67)	1.61	-0.51 (1.22)	1.18 (0.97)	0.67
OM-+0(N/03)	10.98 (0.72)	-12.11 (0.51)	1.12	0.21 (0.93)	2.52 (1.23)	2.73
OYA-48(N76H)	12.82 (0.78)	-9.94 (0.69)	2.89	0.42 (1.05)	-0.04 (0.98)	0.38
0/01-40(10/011)	13.25 (0.44)	-9.01 (0.54)	4.24	-0.85 (0.72)	-0.18 (1.16)	1.03
OX4-48(N76I)	13.88 (0.78)	-13.18 (0.67)	0.69	-2.84 (1.03)	0.47 (0.95)	2.37
0///-+0(11/01)	14.86 (0.66)	-12.92 (0.83)	1.93	-2.22 (0.74)	0.86 (0.73)	1.35
OXA-48(V120T)	12.67 (0.68)	-11.35 (0.68)	1.31	-3.04 (1.02)	2.03 (1.09)	1.01
	12.38 (0.58)	-10.93 (0.44)	1.45	-3.44 (1.08)	2.08 (0.77)	1.36
	15.33 (0.82)	-14.36 (0.86)	0.98	-0.55 (1.16)	-0.62 (1.26)	1.17
0201-40(1201)	14.91 (0.62)	-13.46 (0.54)	1.45	-0.28 (0.71)	0.56 (1.27)	0.28
OXA-48(V120I)	12.91 (0.67)	-15.96 (0.81)	3.04	-0.16 (1.18)	2.22 (1.02)	2.06
0///-+0(1201)	13.56 (0.91)	-16.34 (0.82)	2.78	-0.79 (0.91)	1.98 (1.04)	1.19
OXA A (V1201)	13.80 (0.77)	-12.73 (0.84)	1.07	-1.44 (1.02)	0.29 (0.91)	1.15
0AA-40(V120L)	13.05 (0.69)	-12.15 (0.95)	0.90	-1.59 (0.69)	0.87 (0.77)	0.72
OXA-48(V120M)	14.34 (0.80)	-13.51 (0.79)	0.83	-1.27 (1.09)	3.10 (1.06)	1.83
022-40(12010)	13.99 (0.45)	-13.14 (0.55)	0.85	-1.30 (0.63)	2.03 (0.60)	0.73
OXA-48(I 158I)	12.36 (0.65)	-14.75 (0.73)	2.39	-1.39 (1.09)	1.41 (1.11)	0.03
OAA-40(L1301)	12.56 (0.42)	-14.50 (0.52)	1.94	-0.22 (0.97)	2.83 (0.51)	2.61
OXA_48(I 158V)	12.57 (0.74)	-13.04 (0.70)	0.48	-1.73 (1.02)	1.47 (1.04)	0.26
OW10(F130A)	11.45 (0.44)	-12.62 (0.61)	1.17	-1.90 (0.74)	0.85 (0.88)	1.06
OVA 49(1150D)	12.23 (0.83)	-12.26 (0.84)	0.03	-3.32 (0.98)	1.69 (1.11)	1.62
UAA-40(L130P)	11.87 (0.78)	-11.28 (0.82)	0.60	-5.02 (1.10)	1.95 (1.02)	3.06

Chapter IV

Blinded prediction challenges: the D3R-GC adventure

3 publications:

Selwa, E., Elisée, E., Zavala, A., and Iorga, B. I. (2018) Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations, *J. Comput. Aided Mol. Des.* 32, 273-286 DOI: 10.1007/s10822-017-0054-1.

Chaput, L., Selwa, E., **Elisée, E.**, and Iorga, B. I. (2019) Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations, *J. Comput. Aided Mol. Des.* 33, 93-103 DOI: 10.1007/s10822-018-0161-7.

Elisée, E., Gapsys, V., Mele, N., Chaput, L., Selwa, E., de Groot, B. L. and Iorga, B. I. (2019) Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset, *J. Comput. Aided Mol. Des.* DOI: 10.1007/s10822-019-00232-w.
Overview

1

1.1 Drug Design Data Resource (D3R)

The D3R project gives opportunities to assess and validate computational protocols through blinded prediction Grand Challenge (D3R–GC) based on pose-prediction, affinity ranking, and free energy calculations. First, a dataset of ligands in SMILES format, namely with no three-dimensional coordinates, along with one protein target are generously provided by well-known pharmaceutical companies to D3R organizers (Table 6).

 Table 6: Overview of D3R datasets, each including a main protein target, a series of ligands and some protein-ligand structures solved by X-ray crystallography.

	D3R–GC2	D3R–GC3	D3R–GC4
Providers	Roche	Janssen	Novartis
Targets	Farnesoid X	Cathepsin S	BACE-1
Ligands	102	136	154
X-ray structures	36	26	20

Then, the challenge timeline is divided in two main stages. In Stage 1, participants have to blindly predict: (1) the crystallographic poses of a given number of ligands (Table 1, "X-ray structures"), (2) the affinity ranking for all D3R ligands (Table 1, "Ligands") and, optionally, (3) the absolute or relative binding affinities for some D3R ligands listed in a free energy (FE) subset.

In other words, participants must provide docking poses and scores, which will be assessed by D3R organizers through RMSD calculations, between submitted poses and crystallographic ones, and comparison between the submitted ranking (scoring) and a ranking done according to experimental values for the protein-ligand interaction (IC50).

In Stage 2, participants have to perform the same submission as previously but now considering structural information taken from the crystallographic structures released at the end of Stage 1.

1.2 Strategy

Since our first participation in international computational challenges, we have continuously improved our entire prediction workflow involving two steps: (1) a preliminary analysis of publicly available structural and biochemical data to identify the most appropriate docking software and scoring function for the system of interest, and (2) prediction of binding modes (pose prediction) and of relative affinities of ligands (scoring), using the selected parameters.

Indeed, given that the first step always requires performing docking calculations in order to rank the ligand dataset and get suitable poses for the second step of free energy calculations, our golden rule is to deeply search for both existing structural and affinity data about the training target.

This way, we are able to (1) assume whether a ligand might be active or inactive on the target, based on similarities between D3R ligands and existing tested compound, and (2) appreciate the target's flexibility in presence or absence of interesting ligands in order to cluster different conformations and select one or more representative structures for docking.

Beyond the care taken in the preliminary step, we were particularly concerned with the free energy calculations' part. As mentioned in the previous chapter, we have designed a pmx-based protocol, detailed in the D3R-GC2 paper (Chapter IV Section 2.1), to compute the relative binding free energy of ligands, in a selected target, through alchemical transformations. Our main objective was obviously to evaluate it with the D3R training datasets.

1.3 What we have learned

In hindsight and given our experience, we can draw conclusions on our strategy. First, the correct prediction of poses does not only rely on the selected docking algorithm or software but also on the careful preparation of ligand (protonation states...) and protein (conformations, crystallographic waters...) structures. Furthermore, known ligand poses may drive new ligand's positioning if both share structural similarities, leading to better docking results. We especially experimented this during the last D3R-GC4 competition by building BACE ligands from similar crystallographic ones before performing docking calculations, and this proved to be a highly efficient approach as our prediction was the best in this section.

This way, we obtained excellent outcomes in pose prediction that would probably not have occured equally with tested docking softwares considering the D3R ligands' macrocycle core which brings much more degrees of freedom. However, an accurate pose does not necessarily mean an accurate score-ranking since it may strongly depends on the robustness of the core docking algorithm.

Free energy calculations, in turn, yielded unbalanced outcomes through the three D3R challenges. In D3R–GC2, there were two FE subsets built with sulfonamide-containing compounds: one termed set1, containing ligands with a 4,5,6,7-tetrahydro-1H-indole core, and another termed set2, gathering ligands with a spiro core. While great results were obtained for sulfo, with a submission being ranked #2, really not competitive ones were obtained for spiranes, with a submission being ranked #20. Different reasons were assumed but one appeared to be the most probable, namely two potential non-optimal dihedrals related to the sulfonamide group which appeared to be compensated by other dihedrals in set1 but not in set2.

In D3R–GC3, FE ligands were similar to those from set1 but with a sulfonamide group linked to a 4,5,6,7-tetrahydro-1H-indazole core. Issues about wrong dihedrals in sulfonamide group were fixed but in this challenge, better results were obtained by using only the co-crystallized ligands, highlighting a global bad positioning of peripheral groups which might affect the ranking. Nevertheless, MD simulations of docking poses in explicit solvent significantly improved the ranking. All this corroborates with the idea that no good quality scoring predictions can be obtained from inaccurate docking poses.

In D3R–GC4, BACE ligands from FE dataset were challenging since it deals with big changes in macrocycle core, including its opening. Also, we observed a limitation of our pmx-based protocol on a simple case involving a terminal *N*-dimethyl amide group. Indeed, its disappearance was not a problem but its appearance led to the failure of simulations. The solution to this problem might be the gathering of ligands around a common and simpler scaffold, thus avoiding the addition of dummy atoms. As we don't perform minimization steps before running the morphing process, some dummy atoms of the ligand could appear at the same place of a water molecule or residue atom, which may cause, in certain cases, clashes leading to a crash of the simulation.

D3R publications

2

Lastly, participants are invited to publish their results in the Journal of Computer-Aided Molecular Design, providing an excellent opportunity for sharing information about developed computational tools, encountered issues, solutions and breakthrough strategies. During my PhD, I have contributed to three D3R challenges as follows: preliminary data mining on each target (all), docking calculations (D3R-GC2) and free energy calculations (D3R-GC4). They are included in the next publication-based chapters and organised by year: D3R-GC2 (2017), D3R-GC3 (2018) and D3R-GC4 (2019).

2.1 D3R–GC2 publication

Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations

Edithe Selwa, l,‡ Eddy Elisee, l,‡ Agustin Zavala, l,‡ Bogdan I. Iorga l,*

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gifsur-Yvette, France

Corresponding Author

* Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

[‡] These authors contributed equally.

KEYWORDS: docking ; scoring function ; Gold ; Vina ; Autodock ; Farnesoid X receptor ; FXR ; D3R ; Drug Design Data Resource ; Grand Challenge 2

ABSTRACT

Our participation to the D3R Grand Challenge 2 involved a protocol in two steps, with an initial analysis of the available structural data from the PDB allowing the selection of the most appropriate combination of docking software and scoring function. Subsequent docking calculations showed that the pose prediction can be carried out with a certain precision, but this is dependent on the specific nature of the ligands. The correct ranking of docking poses is still a problem and cannot be successful in the absence of good pose predictions. Our free energy calculations on two different subsets provided contrasted results, which might have the origin in non-optimal force field parameters associated with the sulfonamide chemical moiety.

INTRODUCTION

Drug Design Data Resource (D3R, https://drugdesigndata.org/) organizes, on a regular basis, blinded prediction challenges with the aim to evaluate the performance of tools and protocols that are used in real-life computer-aided drug discovery projects. To achieve this, datasets presenting different levels of difficulty are presented to the community, which is asked to predict, in "blind" conditions, the binding modes and the relative affinities of compounds.

The D3R Grand Challenge 2, which was held in 2016, was focused on a single protein, farnesoid X receptor (FXR, Figure 1), a target with multiple potential applications that has received much attention during the recent years [1-30].



Figure 1. Mesh surface representation of a representative crystal structure (PDB code 3OLF) of the FXR target. The binding site, as defined for our docking studies, is colored in red, and the ligand is colored in green.

In Phase 1 the participants were asked to provide affinity predictions for 102 FXR ligands and pose predictions for 36 of them. In Phase 2 the participants were required to provide the same affinity predictions as in Phase 1, taking into account the additional structural data (36 new protein-ligand complexes) released at the end of Phase 1.

Figure 2 shows the chemical structures of compounds from FXR dataset for which the pose predictions were required. Most of the compounds included in this dataset can be organized in four homogeneous classes based on their chemical structures (benzimidazoles, isoxazoles, sulfonamides, spiro compounds), and the remaining ones presented inhomogeneous structures and were included in a group called miscellaneous. Biological activities data were available for some compounds from this dataset [31-33]. The exact composition of each group can be found in the Electronic Supplementary Material, as well as the structures of the entire FXR dataset, containing 102 ligands used for ranking prediction (Figure S1).



















FXR_16

FXR_17









4



Figure 2. Chemical structures of the 36 FXR ligands included in Phase 1 for pose prediction (compound **FXR_33** was ultimately retired from the pose prediction analysis).

Additionally, the participants were asked to predict the relative affinities for two homogeneous subsets of compounds that are suited for free energy calculations. The structures of compounds (count of 15 and 18, respectively) included in the two free energy subsets are presented in Figures 3 and 4.



Figure 3. Chemical structures of the 15 FXR ligands included in free energy set1 (sulfonamides).



Figure 4. Chemical structures of the 18 FXR ligands included in free energy set2 (spiro compounds).

METHODS

Protein structures. We found 27 crystal structures available in the Protein Data Bank (PDB) [34] for FXR (see the Electronic Supplementary Material for the complete list). These structures constituted our evaluation dataset. All ligands, ions and solvent molecules that were present were manually removed, then the structures were superimposed on the reference structure apo FXR provided by the D3R Grand Challenge organizers, in order to conserve the same coordinate system through the whole process. Missing residues in the structures were added using Modeller 9v12 [35]. Hydrogen atoms were added using Hermes, the graphical interface of Gold v5.2.2 [36] software, or with AutoDock Tools [37] prior to docking.

Ligands. Ligand structures from the evaluation dataset were retrieved from PDB in the SMILES format and they were converted into three-dimensional MOL2 files using CORINA v3.60 (http://www.molecular-networks.com/). This protocol was used instead of retrieving directly the three-dimensional coordinates from the PDB in order to avoid any bias in the docking process that might be related to the initial coordinates of the ligands. Ligand structures used in Phase 1 were obtained from the SMILES strings provided by organizers upon conversion into three-dimensional MOL2 files using CORINA. Three-dimensional coordinates of the ligands used in Phase 2 were built using UCSF Chimera [38], by superimposing their common backbone on the released FXR_17, FXR_10 and FXR_12 crystal structures and by manual addition of the appropriate substituents. In all cases, the protonation state for all compounds was adjusted at physiological pH using LigPrep (Schrödinger, http://www.schrodinger.com/).

Docking. In the preliminary analysis step, several docking software and scoring functions have been tested (re-docking and cross-docking) for their ability to reproduce the protein-ligand complexes from the evaluation dataset: Gold [36] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [39] and AutoDock [37]. Default parameters were used in all cases for docking, except with Gold, where a search efficiency of 200% was used in order to better explore the conformational space. The binding sites were considered with Gold as spheres with a 20 Å radius around the C α atom of Ala288 (numbering from the 1OSV structure). With Vina and AutoDock, the binding sites were defined as a 40 x 40 x 40 $Å^3$ cube centered on the same atom. The protein was considered to be either rigid, or with a few key residues (Leu291/Asn297/Met332/Arg335/Ser336/His451/Trp458 or Arg335 only) from the binding site as flexible. As a result of preliminary analysis, Gold with the ASP scoring function and the rigid protein was used in the Phase 1 predictions. In Phase 2, the rescoring of the FXR complexes was carried out using Gold with the ASP scoring function. For submission, the protein structures were converted into PDB format using UCSF Chimera [38], and the docking poses were converted into MOL format using CORINA (the MOL format corresponds to the SDF output format in CORINA). Unfortunately, the ligand conversion with CORINA was carried out initially without the option "-d no3d", which led to the generation of new coordinates and therefore invalid conformations in our "rjyhz" submission. The results reported here (named "rivhz revised") represent the correct poses, obtained with the option "-d no3d".

Free energy calculations. The protein used for free energy calculation was taken from PDB database with PDB entry corresponding to 3FLI. Three-dimensional coordinates of ligands were built using UCSF Chimera [38], by superimposing common backbone on released FXR_17, FXR_10 and FXR_12 structures. In the set2, the structure solved by X-ray crystallography for

FXR 12 had two alternative positions for the aromatic ring substituent (AA and AB). Both of them were considered in our calculations, and the one leading to the most favorable energy chosen for the submission. Alchemical free energies were calculated using Gromacs [17] and OPLS-AA force field [40,41], some scripts from the PMX software [42-44] and some in house developed scripts. The main steps of this protocol are presented in Figure 5. Hybrid structures and topologies were built using a modified version of the MOL2FF package developed in our team. Hybrid topologies represent simultaneously both ligands, the contribution of each structure being controlled by a parameter λ . For example, a λ value of 0 represents exclusively the ligand A, a λ value of 1 represents exclusively the ligand B, and a λ value of 0.3 represents a contribution of 70% of the ligand A and a contribution of 30% of the ligand B. FXR 91 was used as reference structure for set1, and FXR 10 or FXR 12 for set2. Equilibrium 10 ns MD simulations were performed for the two states (corresponding to lambda 0 and 1), using Gromacs [17] and OPLS-AA force field [40,41]. Snapshots from the equilibrium runs were extracted to spawn 100 simulations of 50 ps each to alchemically morph between the two states of the system. The work values over every non-equilibrium transition were extracted and further used to estimate the free energy differences relying on the Crooks Fluctuation Theorem and utilizing Crooks-Gaussian Intersection as estimator. When the charge was the same in the two ligands considered for the alchemical transformation, separate calculations were carried out for the transformation A into B of the protein-ligand complex and of the ligand alone, the relative free energy of binding being the difference between the corresponding work for these two transformations (see Figures S2 and S3 in the Electronic Supplementary Material file). When the charge was different between the two ligands, a single box containing the protein-ligand complex and the ligand alone, separated by 30 Å, was considered. The ligand from the complex was converted from state A into B, whereas the ligand alone was converted simultaneously from state B into A. In these conditions, the overall charge of the system was conserved during the whole simulation, and the relative free energy of binding between ligands corresponds to the global work for this system (see Figure S4 in the Electronic Supplementary Material file).

Step 1: Generation of hybrid topology



Step 2: Conformational sampling using equilibrium MD



.



Step 3: Structure morphing using non-equilibrium MD







Figure 5. The main steps of the protocol used for the calculation of alchemical free energies.

Graphics. Chemical structures were depicted using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/), images for protein structures were generated using PyMol 1.8.1 (Schrödinger, http://www.pymol.org/) and histograms were obtained using the R package (http://www.r-project.org).

Statistics. Statistics were computed using the online tools available at http://www.sthda.com/english/rsthda/correlation.php.

Chemoinformatics. Tanimoto similarities were computed using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/).

RESULTS AND DISCUSSION

In our previous participations to the SAMPL3 (2011) [45], SAMPL4 (2013) [46], CSAR (2014) [47] and D3R Grand Challenge (2015) [48] docking and virtual screening challenges we followed an approach involving two steps. The first step consists in a preliminary analysis of information available in literature (structural, and in some cases enzymatic data), which allows the identification of the best combination of docking software and scoring function that are suited for studying the system of interest. In the second step, the combination of docking software and scoring function is used to predict the binding modes (pose prediction) and the relative affinities of ligands (scoring). As our previous studies [45-48] highlighted the importance of using enhanced genetic algorithm parameters for docking (a search efficiency of 200%), in this work we used the same parameters in order to ensure an adequate conformational sampling of docking conformations.

Preliminary analysis

We found 27 crystal structures of FXR that were available in the Protein Data Bank (PDB) [34]. These structures were organized in five distinct groups, according to their conformation and the ligand present in the binding site (see the Electronic Supplementary Material for a complete list of these structures and the exact composition of each group). A representative structure was selected from each group, based on the crystal structure resolution and the lack of missing residues. The three-dimensional structure of protein in these structures is well conserved, with the exception of two fragments (residues 258-285 and 335-358) that are very flexible (Figure 6). These five structures, together with the apo FXR structure provided by the organizers, constitute our evaluation dataset, which was used in re-docking and cross-docking calculations using the FXR ligands from all the 27 structures available and several combinations of docking software and scoring functions: Gold with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina and Autodock.



Figure 6. Representative 6 FXR PDB structures superimposed: a) general view, showing a very good global conservation of structural features; b) zoom on residues 258-285 and 335-358, highlighting the conformational flexibility of these fragments. The structures are represented as

follows: 1OSV (green), 3FLI (cyan), 3OLF (magenta), 4WVD (yellow), FXR apo (wheat), 3HC5 (gray).

RMSD values compared with the native ligands from the crystallographic structures were calculated for all docking poses. In order to evaluate the accuracy of docking and scoring, we have considered the lowest RMSD value and the RMSD value of the best ranking pose for each combination protein-ligand-(docking software)-(scoring function). AutoDock provided very poor results, with most of the docking conformations positioned outside the binding site, whereas Gold/ASP, followed by Gold/GoldScore and Vina, could reproduce rather well the native protein-ligand complexes, especially in the cross-docking calculations. As expected, the redocking results outperformed the cross-docking results. It was also observed that the combination of a protein and a ligand belonging to the same group was more favorable than a combination of a protein and a ligand from different groups.

Pose prediction and scoring

The 102 FXR ligands from the D3R Grand Challenge 2 dataset containing 180 ligands were docked on the 6 representative FXR structures shown in Figure 6 using Gold with the ASP scoring function. For each ligand, the best-ranked docking conformation was selected and the overall ranking was submitted, as well as the coordinates for the ligands FXR_1 to FXR_36. For the ligands belonging to a group for which crystal structures were available (e.g. benzimidazoles, isoxazoles), the RMSD was calculated using the largest common fragment, and the conformations with the best RMSD were selected for a second submission. The RMSD calculation was realized using an in house developed script based on CACTVS Chemoinformatics Toolkit. The poses from the spiro and sulfonamides groups were visually

inspected using UCSF Chimera. Only the 3FLI and the APO structures provided docking poses with a carboxylate group (that is present in most spiro structures and in FXR_101 from the sulfonamides group) interacting with Arg335. This was considered as the correct orientation, since most of the crystalized ligands show the same kind of interaction. Overall, poses obtained with the structure 3OLF were selected for benzimidazoles, with 1OSV for steroids, with 3HC5 for isoxazoles and with 3FLI for all others.

The performance of submissions for pose prediction (best RMSD and RMSD of pose 1) is presented in Figure 7, showing a relatively good result that we obtained in this category compared with the other participants.



Figure 7. Performance of Phase 1 pose prediction submissions (Kendall Tau) for the FXR D3R Grand Challenge 2 dataset: best RMSD (**a**) and RMSD of pose 1 (**b**). Our submissions are colored in red (see text for details).

Our scoring results for the two submissions in Phase 1 were very modest, with Kendall Tau values of 0.13 and 0.072. Table S1 from the Supplementary Information file shows the rank of the best RMSD pose for compounds with existing reference structural data (53 compounds out of 102 compounds from the dataset). A mean value of 4.68 (out of 10 poses in each case) was

obtained for this rank, which is quite low. If we also consider that for the remaining 49 compounds with no reference structure available we have no information about the docking reliability, these data altogether might explain the incorrect scoring prediction.

The crystallographic structures of the 36 FXR complexes proposed for pose prediction were released at the end of Phase 1. A comparison of several representative docking poses and the corresponding crystallographic conformations is provided in Figure 8. We predicted well the conformation of most benzimidazoles, but the other three groups (isoxazoles, spiro compounds and sulfonamides) were more challenging, and we could predict correctly only the overall orientation of the ligand, but not the details of the interaction with the binding site. The compounds from the miscellaneous group were even more difficult, and in some cases our prediction was completely opposite compared to the crystal structure.



Figure 8. Comparison of our docking poses (cyan) with crystal structure conformations (magenta) for representative FXR ligands from different families: a) benzimidazoles (FXR_21/3OLF, RMSD 0.97 Å); b) isoxazoles (FXR_4/3HC5, RMSD 3.87 Å); c) spiro compounds (FXR_10/3FLI, RMSD 2.85 Å); d) sulfonamides (FXR_16/3FLI, RMSD 2.03 Å); e) miscellaneous (FXR_34/10SV, RMSD 3.76 Å); f) miscellaneous (FXR_5/3FLI, RMSD 4.70 Å).

In Phase 2, the three-dimensional coordinates of the ligands FXR_37 to FXR_102 were built using UCSF Chimera [38], by superimposing their common backbone on the released FXR_17, FXR_10 and FXR_12 crystal structures. The protein-ligand complexes of these ligands, together with the 36 ligands from the crystal structures, were rescored using Gold with the ASP scoring function and the results were slightly improved compared with Phase 1, with a Kendall Tau value of 0.17.

Free energy calculations

The free energy calculations were carried out using a protocol adapted from the methodology implemented within the PMX software [42-44]. An important advantage of our procedure is the possibility to simulate transformations involving charge modification, which is relatively difficult or even impossible using other protocols (see the Methods section and Figures S2, S3 and S4 in the Electronic Supplementary Material for more details).

We obtained very good results for the free energy prediction of the set1 (sulfonamides), our submission **nszkx** being ranked #2. However, the corresponding submission **2ytv8** for set2 (spiro compounds) was not at all competitive, being ranked #20 (Figure 9). After the end of the D3R Grand Challenge 2 we have recomputed all data after fixing a bug in the hybrid topologies, and

also using docking poses instead of crystal structures (equivalent of Phase 1 calculations carried out retrospectively) and using AMBER/GAFF force field instead of OPLS-AA (Figure 9).



Figure 9. Performance of Phase 2 free energy submissions for set1 (**a**) and set2 (**b**). The correlation coefficients are represented as follows: Kendall tau in blue, Spearman rho in light blue and Pearson r in cyan. Our submissions are colored in dark red, red and pink, respectively (nszkx and 2ytv8). The results obtained on recomputed simulations after fixing a bug in the

hybrid topologies are represented in different shades of green (nszkx_af and 2ytv8_af). The results from simulations using docking poses instead of crystal structures (equivalent of Phase 1 calculations carried out retrospectively) are represented in magenta (nszkx_d). The results from simulations using AMBER/GAFF force field instead of OPLS-AA are represented in orange (nszkx_am).

Tables 1 and 2 contain the detailed computed values for set1 and set2, respectively, together with the corresponding statistics (Kendall's rank correlation tau, Spearman's rank correlation rho and Pearson's product-moment correlation r).

For set1, similar results were obtained with OPLS-AA before and after correction, as well as with AMBER/GAFF force field. However, when docking poses were used as initial coordinates (which represents Phase 1 calculations carried out retrospectively), no correlation was obtained. This corroborates with the pose prediction results, showing that no good quality predictions can be obtained from inaccurate docking poses.

		OPLS-AA before correction (nszkx)		OPLS-AA after correction (nszkx_af)	
	Experimental IC50 (µM)	$\Delta\Delta G (kJ/mol)$	$\Delta\Delta G$ error (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	$\Delta\Delta G \text{ error } (kJ/mol)$
FXR_17	0.79	-9.32	1.16	-18.35	1.12
FXR_45	28.85	-3.14	2.33	-21.21	2.15
FXR_46	62.37	-3.81	21.82	-12.98	1.26
FXR_47	20.96	NA	NA	-18.61	2.41
FXR_48	100.00	NA	NA	5.36	1.75
FXR_49	100.00	-3.73	1.48	-9.47	1.02
FXR_91	29.63	0.00	0.00	0.00	0.00
FXR_93	46.66	-8.98	1.26	-4.04	0.39
FXR_95	32.17	-8.39	1.45	-6.63	1.21
FXR_96	58.86	-19.12	1.62	-21.36	2.15
FXR_98	13.14	-21.07	1.42	-12.95	0.97
FXR_99	100.00	-6.34	1.59	-11.07	0.56
FXR_100	19.14	-15.86	2.49	-25.81	3.05
FXR_101	27.64	-36.24	3.38	-27.94	3.69
FXR_102	29.23	-0.15	2.85	13.45	2.17
		Correlation coefficient	p-value	Correlation coefficient	p-value

Table 1. Free energies computed for set1. FXR_91 was used as reference compound.

Kendall's rank correlation tau	0.2452	0.2455	0.2512	0.1962
Spearman's rank correlation rho	0.3851	0.1937	0.4444	0.0970
Pearson's product-moment correlation r	0.2648	0.3819	0.3029	0.2726

For set2, the results were quite disappointing, with negative correlations with OPLS-AA before and after the correction. Apparent better correlations were obtained with OPLS-AA using the docking poses and with AMBER/GAFF, but they are not representative since they were computed only for 12 and 6 values, all of them belonging to the FXR_12 subset (see Table 2 for the compounds belonging to the FXR_10 and FXR_12 subsets), so we decided not to represent them in Figure 9.

Table 2. Free energies computed for set2. The FXR_10 subset contains the five compounds marked with a star, and the FXR_12 subset contains the remaining compounds. FXR_10 and FXR_12 were used as reference compounds for each subset, then all free energies from subset FXR_10 were translated relative to FXR_12.

		OPLS-AA before correction (2ytv8)		OPLS-AA after corr	OPLS-AA after correction (2ytv8_af)	
	Experimental IC50 (µM)	$\Delta\Delta G$ (kJ/mol)	$\Delta\Delta G$ error (kJ/mol)	$\Delta\Delta G (kJ/mol)$	$\Delta\Delta G$ error (kJ/mol)	
FXR_10*	5.64	-10.69	0.81	-4.85	0.48	
FXR_12	0.06	0.00	0.00	0.00	0.00	
FXR_38*	100.00	-27.49	-	-18.86	-	
FXR_41	100.00	-15.19	4.15	-3.24	4.27	
FXR_73*	11.22	10.77	-	-14.83	-	
FXR_74	0.66	-3.16	-	-2.57	0.16	
FXR_75*	100.00	-71.30	-	-32.41	-	
FXR_76	41.83	-0.01	0.62	-1.78	66.29	
FXR_77	0.25	-2.35	0.35	-2.47	0.23	
FXR_78	0.03	-15.94	1.08	-5.89	19.75	
FXR_79*	4.15	2.39	-	14.84	-	
FXR_81	2.69	-11.09	0.73	-10.97	18.40	
FXR_82	0.18	7.14	0.85	2.60	0.58	
FXR_83	0.33	-1.75	0.31	5.02	0.77	
FXR_84	4.54	-5.87	0.66	4.09	17.88	
FXR_85	0.30	-9.67	0.40	-5.10	90.68	
FXR_88	0.54	-3.73	0.40	0.81	0.33	
FXR_89	0.74	-10.31	75.22	-4.15	75.34	
		Correlation coefficient	p-value	Correlation coefficient	p-value	
Kendall's rank correlation tau		-0.2772	0.1107	-0.2772	0.1107	
Spearman's rank correlation rho		-0.3320	0.1784	-0.3630	0.1387	
Pearson's product-moment correlation r		-0.6900	0.0015	-0.6105	0.0071	

We tried to find a rational explanation for the discrepancy of the results obtained for set1 and set2, using the same protocol. Among the possible hypotheses, we can mention: i) the intrinsic greater structural diversity in set2 compared with set1; ii) incorrect force field parameters and iii) insufficient conformation sampling of ligands.

To validate the first hypothesis, we computed the Tanimoto similarity matrix for set1 and set2 (see Tables S2 and S3 in Electronic Supplementary Material). The global mean values of Tanimoto similarity for the two datasets are very close, 0.88 and 0.86, respectively, suggesting a similar degree of diversity. However, a visual inspection of the two datasets shows that set1 is quite homogeneous, with variations on the substitution pattern of a single phenyl ring. On the other hand, compounds from set2 contain variations on two fragments: one can be a diversely substituted phenyl ring, and the other can be either a thienyl ring or a diversely substituted phenyl ring. According to the presence or not of the thienyl ring, set2 can be divided into two subsets, which have FXR 10 and FXR 12 as representative compounds. We computed the statistics separately on these two subsets and the results are presented in Table 3. Compared with the whole set2, only a small improvement in the correlation with experimental data is observed for the FXR_12 subset. However, for the FXR_10 subset we observe almost a perfect anticorrelation with the experimental data. Overall, this analysis shows that the differences between the structural diversity of set1 and set2 are too small to be discriminated by descriptors such as Tanimoto similarity, but the set2 is more diverse and can be divided in two subsets. One of these subsets, containing a thienyl substituent, has a major negative impact in the prediction of free energies for set2.

Table 3. Statistics computed for the subsets FXR_10 and FXR_12 of set2. See Table 2 and text for the list of compounds included in each subset.

	OPLS-AA before correction		OPLS-AA after correction	
FXR_10 subset	Correlation	p-value	Correlation	p-value
	coefficient	_	coefficient	_
Kendall's rank correlation tau	-0.5270	0.2065	-0.9487	0.0230
Spearman's rank correlation	-0.6669	0.2189	-0.9747	0.0048
rho				
Pearson's product-moment	-0.8377	0.0766	-0.7737	0.1247
correlation r				
FXR_12 subset	Correlation	p-value	Correlation	p-value
	coefficient		coefficient	
Kendall's rank correlation tau	-0.2564	0.2519	-0.0513	0.8577
Spearman's rank correlation	-0.2692	0.3733	-0.0330	0.9206
rho				
Pearson's product-moment	-0.3261	0.2769	-0.0902	0.7696
correlation r				

To evaluate the pertinence of the second hypothesis, we analyzed the conformational distribution of the ligands FXR_17, FXR_10 and FXR_12, as representative structures for set1 and the two subsets of set2, in two force fields, OPLS-AA and AMBER/GAFF. In each case, we extracted and superimposed all the 501 conformations from the 10 ns molecular dynamics simulation of the ligand alone in water. The result is presented in Figure S5 (Electronic Supplementary Material).

For compound FXR_17, we observe 4 main differences between the distributions OPLS-AA (a) and AMBER/GAFF (b): i) the phenyl ring is mostly parallel with the bicyclic system in a, and perpendicular in b; ii) the amide group is mostly perpendicular with the bicyclic system in a, and parallel in b; iii) the distribution of the thienyl ring around the dihedral C-N-S-C is restricted to a very narrow window in a, whereas in b there are two larger windows in opposite positions, showing in the latter case an unrestricted exchange between these two positions; iv) in a the thienyl ring shows equivalent populations of both faces, whereas in b the rotation around the dihedral N-S-C-S is very much restricted. However, as the predictions of set1 using either OPLS-AA or AMBER/GAFF are very similar (see Figure 9a and Table 1), these differences should not have a major contribution or, more probably, should cancel mutually.

For compound FXR_10, we observe in **a** a restricted rotation around the C-N-S-C dihedral and, in opposition with FXR_17, an impossible rotation around the N-S-C-S dihedral, probably because of the close proximity of the amide oxygen. In the case of **b**, we observe a free rotation around the C-N-S-C dihedral and a restricted rotation around the N-S-C-S dihedral, similar with FXR_17.

Finally, we observe for FXR_12 a restricted rotation around the C-N-S-CA dihedral in **a** and a free rotation around the N-S-CA-CA dihedral (the chlorine substituent is positioned equally on both sides), whereas in **b** the rotation around the C-N-S-CA dihedral is relatively free in the conditions of simulations, but the rotation around the N-S-CA-CA dihedral is almost completely restricted.

These results suggest the possible existence of two non-optimal dihedrals associated with the sulfonamide group, similarly with a recent report regarding the incorrect conformational sampling of linezolid [49]. For set1, their influence might be compensated by two other dihedrals, which is not the case for set2. Additionally, in the FXR_12 subset there are two atropoisomers that can contribute to the overall binding energy, whereas in our calculations we have considered only one, the most favorable.

The third hypothesis, insufficient conformation sampling of ligands, is not very probable given the length of our molecular dynamics simulations and the relative rigidity of the ligands. If the conformation space is not sampled correctly, this should be more due to inadequate force field parameters than to insufficient length of simulations. Along the same lines, in a few specific cases, the standard deviation of our predictions is unusually high (see the " $\Delta\Delta G$ error" columns in Tables 1 and 2), especially for set2.

CONCLUSIONS

We used in this work a protocol in two steps, involving an initial analysis of the available structural data from the PDB, which allows the selection of the most appropriate combination of docking software and scoring function. Subsequent docking calculations showed that the pose prediction can be carried out with a certain precision, but this is dependent on the specific nature of the ligands. The correct ranking of docking poses is still a problem and cannot be successful in the absence of good pose predictions. Our free energy calculations on two different subsets provided contrasted results, which might have the origin in non-optimal force field parameters associated with the sulfonamide chemical moiety.

Electronic Supplementary Material. The Electronic Supplementary Material contains the chemical structures of the entire FXR dataset, the rank of best RMSD poses, conformational distribution of representative ligands, Tanimoto similarity matrices and a schematic description of the systems and thermodynamic cycles used for free energy calculations.

Acknowledgments. We thank Prof. Bert de Groot for helpful discussions. The comments and suggestions of the reviewers are also acknowledged, as they greatly contributed to improve the manuscript. This work was supported by the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT) [grant number ANR-10-LABX-33], by the JPIAMR transnational project DesInMBL [grant number ANR-14-JAMR-0002] and by the Région Ile-de-France (DIM Malinf).

REFERENCES

Bishop-Bailey D (2004) FXR as a novel therapeutic target for vascular disease. Drug news
 & perspectives 17 (8):499-504

2. Claudel T, Sturm E, Kuipers F, Staels B (2004) The farnesoid X receptor: a novel drug target? Expert opinion on investigational drugs 13 (9):1135-1148. doi:10.1517/13543784.13.9.1135

3. Pellicciari R, Costantino G, Fiorucci S (2005) Farnesoid X receptor: from structure to potential clinical applications. J Med Chem 48 (17):5383-5403. doi:10.1021/jm0582221

4. Westin S, Heyman RA, Martin R (2005) FXR, a therapeutic target for bile acid and lipid disorders. Mini reviews in medicinal chemistry 5 (8):719-727

5. Cai SY, Boyer JL (2006) FXR: a target for cholestatic syndromes? Expert opinion on therapeutic targets 10 (3):409-421. doi:10.1517/14728222.10.3.409

6. Lee FY, Lee H, Hubbert ML, Edwards PA, Zhang Y (2006) FXR, a multipurpose nuclear receptor. Trends in biochemical sciences 31 (10):572-580. doi:10.1016/j.tibs.2006.08.002

7. Cariou B, Staels B (2007) FXR: a promising target for the metabolic syndrome? Trends Pharmacol Sci 28 (5):236-243. doi:10.1016/j.tips.2007.03.002

8. Wang YD, Chen WD, Huang W (2008) FXR, a target for different diseases. Histology and histopathology 23 (5):621-627. doi:10.14670/hh-23.621

9. Zimber A, Gespach C (2008) Bile acids and derivatives, their nuclear receptors FXR, PXR and ligands: role in health and disease and their therapeutic potential. Anti-cancer agents in medicinal chemistry 8 (5):540-563

10. Crawley ML (2010) Farnesoid X receptor modulators: a patent review. Expert Opin Ther Pat 20 (8):1047-1057. doi:10.1517/13543776.2010.496777

 Fiorucci S, Mencarelli A, Distrutti E, Palladino G, Cipriani S (2010) Targetting farnesoid-X-receptor: from medicinal chemistry to disease treatment. Current medicinal chemistry 17 (2):139-159

12. Mencarelli A, Fiorucci S (2010) FXR an emerging therapeutic target for the treatment of atherosclerosis. Journal of cellular and molecular medicine 14 (1-2):79-92. doi:10.1111/j.1582-4934.2009.00997.x

13. Teodoro JS, Rolo AP, Palmeira CM (2011) Hepatic FXR: key regulator of whole-body energy metabolism. Trends in endocrinology and metabolism: TEM 22 (11):458-466. doi:10.1016/j.tem.2011.07.002

14. Adorini L, Pruzanski M, Shapiro D (2012) Farnesoid X receptor targeting to treat nonalcoholic steatohepatitis. Drug Discov Today 17 (17-18):988-997. doi:10.1016/j.drudis.2012.05.012

15. Fiorucci S, Mencarelli A, Distrutti E, Zampella A (2012) Farnesoid X receptor: from medicinal chemistry to clinical applications. Future medicinal chemistry 4 (7):877-891. doi:10.4155/fmc.12.41

16. Fiorucci S, Zampella A, Distrutti E (2012) Development of FXR, PXR and CAR agonists and antagonists for treatment of liver disorders. Current topics in medicinal chemistry 12 (6):605-624

17. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and

28

highly parallel open source molecular simulation toolkit. Bioinformatics (Oxford, England) 29 (7):845-854. doi:10.1093/bioinformatics/btt055

18. Carotti A, Marinozzi M, Custodi C, Cerra B, Pellicciari R, Gioiello A, Macchiarulo A (2014) Beyond bile acids: targeting Farnesoid X Receptor (FXR) with natural and synthetic ligands. Current topics in medicinal chemistry 14 (19):2129-2142

19. Fiorucci S, Distrutti E, Ricci P, Giuliano V, Donini A, Baldelli F (2014) Targeting FXR in cholestasis: hype or hope. Expert opinion on therapeutic targets 18 (12):1449-1459. doi:10.1517/14728222.2014.956087

20. Gege C, Kinzel O, Steeneck C, Schulz A, Kremoser C (2014) Knocking on FXR's door: the "hammerhead"-structure series of FXR agonists - amphiphilic isoxazoles with potent in vitro and in vivo activities. Current topics in medicinal chemistry 14 (19):2143-2158

21. Huang H, Xu Y, Zhu J, Li J (2014) Recent advances in non-steroidal FXR antagonists development for therapeutic applications. Current topics in medicinal chemistry 14 (19):2175-2187

22. Lamers C, Schubert-Zsilavecz M, Merk D (2014) Medicinal chemistry and pharmacological effects of Farnesoid X Receptor (FXR) antagonists. Current topics in medicinal chemistry 14 (19):2188-2205

23. Ali AH, Carey EJ, Lindor KD (2015) Recent advances in the development of farnesoid X receptor agonists. Annals of translational medicine 3 (1):5. doi:10.3978/j.issn.2305-5839.2014.12.06

24. Carr RM, Reid AE (2015) FXR agonists as therapeutic agents for non-alcoholic fatty liver disease. Current atherosclerosis reports 17 (4):500. doi:10.1007/s11883-015-0500-2

25. Koutsounas I, Theocharis S, Delladetsima I, Patsouris E, Giaginis C (2015) Farnesoid x receptor in human metabolism and disease: the interplay between gene polymorphisms, clinical phenotypes and disease susceptibility. Expert opinion on drug metabolism & toxicology 11 (4):523-532. doi:10.1517/17425255.2014.999664

26. Sanyal AJ (2015) Use of farnesoid X receptor agonists to treat nonalcoholic fatty liver disease. Digestive diseases (Basel, Switzerland) 33 (3):426-432. doi:10.1159/000371698

27. Sepe V, Distrutti E, Fiorucci S, Zampella A (2015) Farnesoid X receptor modulators (2011
2014): a patent review. Expert Opin Ther Pat 25 (8):885-896.
doi:10.1517/13543776.2015.1045413

28. Sepe V, Distrutti E, Limongelli V, Fiorucci S, Zampella A (2015) Steroidal scaffolds as FXR and GPBAR1 ligands: from chemistry to therapeutical application. Future medicinal chemistry 7 (9):1109-1135. doi:10.4155/fmc.15.54

29. Alawad AS, Levy C (2016) FXR Agonists: From Bench to Bedside, a Guide for Clinicians. Digestive diseases and sciences 61 (12):3395-3404. doi:10.1007/s10620-016-4334-8

30. De Magalhaes Filho CD, Downes M, Evans RM (2017) Farnesoid X Receptor an Emerging Target to Combat Obesity. Digestive diseases (Basel, Switzerland) 35 (3):185-190. doi:10.1159/000450909

31. Feng S, Yang M, Zhang Z, Wang Z, Hong D, Richter H, Benson GM, Bleicher K, Grether U, Martin RE, Plancher JM, Kuhn B, Rudolph MG, Chen L (2009) Identification of an N-oxide
pyridine GW4064 analog as a potent FXR agonist. Bioorganic & medicinal chemistry letters 19 (9):2595-2598. doi:10.1016/j.bmcl.2009.03.008

32. Richter HG, Benson GM, Bleicher KH, Blum D, Chaput E, Clemann N, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S (2011) Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. Bioorganic & medicinal chemistry letters 21 (4):1134-1140. doi:10.1016/j.bmcl.2010.12.123

33. Richter HG, Benson GM, Blum D, Chaput E, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S, Bleicher KH (2011) Discovery of novel and orally active FXR agonists for the potential treatment of dyslipidemia & diabetes. Bioorganic & medicinal chemistry letters 21 (1):191-194. doi:10.1016/j.bmcl.2010.11.039

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28 (1):235-242

35. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234 (3):779-815. doi:10.1006/jmbi.1993.1626

36. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein– ligand docking using GOLD. Proteins Struct Funct Bioinf 52 (4):609-623. doi:10.1002/prot.10465 37. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30 (16):2785-2791. doi:10.1002/jcc.21256

38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera -- a visualization system for exploratory research and analysis. J Comput Chem 25 (13):1605-1612. doi:10.1002/jcc.20084

39. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31 (2):455-461. doi:10.1002/jcc.21334

40. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105 (28):6474-6487. doi:10.1021/jp003919d

41. Robertson MJ, Tirado-Rives J, Jorgensen WL (2015) Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. J Chem Theory Comput 11 (7):3499-3509. doi:10.1021/acs.jctc.5b00356

42. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem 36 (5):348-354. doi:10.1002/jcc.23804

43. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. Methods in molecular biology (Clifton, NJ) 1215:173-209. doi:10.1007/978-1-4939-1465-4_9

44. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angewandte Chemie (International ed in English) 55 (26):7364-7368. doi:10.1002/anie.201510054

45. Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. J Comput Aided Mol Des 26 (5):595-601. doi:10.1007/s10822-011-9526-x

46. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. J Comput Aided Mol Des 28 (4):455-462. doi:10.1007/s10822-014-9707-5

47. Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. Journal of chemical information and modeling 56 (6):996-1003. doi:10.1021/acs.jcim.5b00337

48. Selwa E, Martiny VY, Iorga BI (2016) Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets. J Comput Aided Mol Des 30 (9):829-839. doi:10.1007/s10822-016-9983-3

49. Grunenberg J, Licari G (2016) Effective in silico prediction of new oxazolidinone antibiotics: Force field simulations of the antibiotic-ribosome complex supervised by experiment and electronic structure methods. Beilstein J Org Chem 12:415-428. doi:10.3762/bjoc.12.45

Electronic Supplementary Material

Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations

Edithe Selwa,^{1,‡} Eddy Elisee,^{1,‡} Agustin Zavala,^{1,‡} Bogdan I. Iorga^{1,*}

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France

Corresponding Author

* Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

[‡] These authors contributed equally.

Table of contents

Protein Data Bank (PDB) structures used for the preliminary analysis	S 3
FXR ligands regroupment according to their chemical structure	S 3
Figure S1. Chemical structures of FXR dataset	S4-S11
Figure S2. Performance of Phase 1 pose prediction submissions	S12
Figure S3. Performance of Phase 2 free energy submissions for set2	S12
Figure S4. Thermodynamic cycle for the calculation of relative binding affinities	S 13
Figure S5. System used for the calculation of relative binding affinities (charge conserving)	S 13
Figure S6. System used for the calculation of relative binding affinities (charge modifying)	S14

Protein Data Bank (PDB) structures available

27 structures were available in the PDB for FXR at the moment when the D3R Grand Challenge 2 took place. They were organized in 5 distinct groups, according to the type of ligand and the binding site conformation. The representative structure for each group (based on the crystal structure resolution and lack of missing residues) is colored in red.

Group 1 (isoxazoles): 3dct, 3dcu, 3gd2, 3hc5, 3hc6, 3p88, 3p89, 3rut, 3ruu, 3rvf

Group 2 (benzimidazoles): 30kh, 30ki, 30lf, 30mk, 30mm, 30of, 300k

Group 3 (FXR_5-like): 311b, 3fli

Group 4 (steroid, FXR_34-like): 3bej, 1osv, 1ot7, 4qe6

Group 5 (miscellaneous): 10sh, 40iv, 4qe8, 4wvd

FXR ligands regroupment according to their chemical structure

Isoxazoles: FXR_4, FXR_23, FXR_33, FXR_65

Benzimidazoles: FXR_6, FXR_7, FXR_8, FXR_9, FXR_13, FXR_14, FXR_19, FXR_20, FXR_21, FXR_22, FXR_24, FXR_25, FXR_26, FXR_27, FXR_28, FXR_29, FXR_30, FXR_31, FXR_32, FXR_35, FXR_36, FXR_37, FXR_39, FXR_40, FXR_42, FXR_50, FXR_51, FXR_52, FXR_53, FXR_54, FXR_55, FXR_56, FXR_57, FXR_58, FXR_59, FXR_60, FXR_61, FXR_62, FXR_63, FXR_64, FXR_66, FXR_67, FXR_68, FXR_69, FXR_70, FXR_71, FXR_72

Spiro compounds: FXR_10, FXR_11, FXR_12, FXR_38, FXR_41, FXR_73, FXR_74, FXR_75, FXR_76, FXR_77, FXR_78, FXR_79, FXR_80, FXR_81, FXR_82, FXR_83, FXR_84, FXR_85, FXR_86, FXR_87, FXR_88, FXR_89

Sulfonamides: FXR_15, FXR_16, FXR_17, FXR_43, FXR_44, FXR_45, FXR_46, FXR_47, FXR_48, FXR_49, FXR_90, FXR_91, FXR_92, FXR_93, FXR_94, FXR_95, FXR_96, FXR_97, FXR_98, FXR_99, FXR_100, FXR_101, FXR_102

Miscellaneous: FXR_1, FXR_2, FXR_3, FXR_5, FXR_18, FXR_34

S2

Figure S1. Chemical structures of the entire FXR dataset, containing 102 ligands used for ranking prediction.









FXR_16



















S4







































































FXR_60

































FXR_91

FXR_92

FXR_93







FXR_94





















Figure S2. Performance of Phase 1 pose prediction submissions (Kendall Tau) for the FXR D3R Grand Challenge 2 dataset (RMSD for pose 1). Our submission is colored in red.

Figure S3. Performance of Phase 2 free energy submissions for set2 (Kendall Tau in blue, Spearman Rho in light blue and Pearson R in cyan). Our submissions are colored in different shades of red, respectively.





Figure S4. Thermodynamic cycle for the calculation of relative binding affinities of ligands L_1 and L_2 compared with the protein P.

Figure S5. Schematic representation of the system used for the calculation of relative binding affinities of ligands L_1 and L_2 for the protein P, in the case of a charge conserving structural change on the ligand.

Charge conserving structural change for ligand



 $\Delta \Delta \mathbf{G} = \Delta \mathbf{G4} - \Delta \mathbf{G2}$

Figure S6. Schematic representation of the system used for the calculation of relative binding affinities of ligands L_1 and L_2 for the protein P, in the case of a charge modifying structural change on the ligand.

Charge modifying mutation



2.2 D3R–GC3 publication

Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations

Ludovic Chaput,^{1,2,‡} Edithe Selwa,^{1,‡} Eddy Elisée,^{1,‡} Bogdan I. Iorga^{1,*}

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gifsur-Yvette, France

² Department of Nephrology and Dialysis, AP-HP, Tenon Hospital, INSERM UMR_S 1155,
75020 Paris, France

Corresponding Author

* Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

[‡] These authors contributed equally.

1

KEYWORDS: docking ; ranking ; scoring function ; free energy calculations ; Gold ; Vina ; Glide ; Gromacs ; PMX ; cathepsin S inhibitors ; D3R ; Drug Design Data Resource ; Grand Challenge 3 ; D3RGC3

ABSTRACT

During the last few years, we have developed a docking protocol involving two steps: i) the choice of the most appropriate docking software and parameters for the system of interest using structural and functional information available in public databases (PDB, ChEMBL, PubChem Assay, BindingDB, etc.); ii) the docking of ligand dataset to provide a prediction for the binding modes and ranking of ligands. We applied this protocol to the D3R Grand Challenge 3 dataset containing cathepsin S (CatS) inhibitors. Considering the size and conformational flexibility of ligands, the docking calculations afforded reasonable overall pose predictions, which are however dependent on the specific nature of each ligand. As expected, the correct ranking of docking poses is still challenging. Post-processing of docking poses with molecular dynamics simulations in explicit solvent provided a significantly better prediction, whereas free energy calculations on a subset of compounds brought no significant improvement in the ranking prediction compared with the direct ranking obtained from the scoring function.

INTRODUCTION

Drug Design Data Resource (D3R, https://drugdesigndata.org/) organizes, on a regular basis, blinded prediction challenges with the aim to evaluate the performance of tools and protocols that are used in real-life computer-aided drug discovery projects. To achieve this, datasets presenting different levels of difficulty are presented to the community, which is asked to predict, in "blind" conditions, the binding modes and the relative affinities of compounds.

The D3R Grand Challenge 3, which was held in 2017, was focused on a single protein, cathepsin S (CatS, Figure 1), a target with multiple potential applications that has received much attention during the recent years [1-5].



Figure 1. Solid surface representation of a representative crystal structure of cathepsin S (PDB code 3IEJ) with the ligand 599 in stick representation. The binding site, as defined for our docking studies, is colored in yellow, and the ligand is colored in cyan.

In Phase 1A the participants were asked to predict the crystallographic poses of 24 CatS ligands, the affinity ranking for 136 CatS ligands and the relative binding affinities for a designated free energy subset of 33 compounds. In phase 1B, the organizers released the corresponding receptor structures for the 24 CatS ligands, and asked to predict the crystallographic poses using the released receptor structures. In Phase 2 the participants were required to provide the same affinity predictions for the subset of 136 ligands and relative free

energy for the subset of 33 compounds as in Phase 1, but taking into account the additional structural data (24 new protein-ligand complexes) released at the end of Phase 1B. The CatS dataset containing the 136 ligands used for ranking prediction can be found in the Electronic Supplementary Material (Figure S3).

Figure 2 shows the chemical structures of the 24 compounds from CatS dataset for which the pose predictions were required. Additionally, the participants were asked to predict the relative affinities for one homogeneous subset of compounds that is suited for free energy calculations. The structures of the 33 CatS inhibitors included in the free energy subset are presented in Figure 3.



Figure 2. Chemical structures of the 24 CatS ligands included in Phase 1 for pose prediction.



Figure 3. Chemical structures of the 33 CatS ligands included in the free energy set.



Figure 3. Chemical structures of the 33 CatS ligands included in the free energy set (continued).

METHODS

Protein structures. We found 38 crystal structures of CatS available in the Protein Data Bank (PDB) [6] from which 31 were human CatS and 7 were mouse CatS (see Table S1 in the Electronic Supplementary Material for the complete list). These structures were superposed, showing no conformational variability for the backbone and for most sidechains. The only exceptions were the sidechains of residues Lys64, Arg141 and Asn161, which are very flexible, adopting multiple conformations, and of residues Phe146 and Phe211, which adopt two major

orientations each. The residues Lys64 and Arg141 are relatively far from the binding site and therefore their flexibility was not taken into account for the docking process. In these conditions, we generated four representative conformations of CatS (named CatS_conf1 to CatS_conf4) starting from the structure 1GLO (apo) and adjusting the orientations of Phe146 and Phe211. These conformers, with the ions and solvent molecules manually removed, were superimposed on the reference structure CatS (containing the SO4 ion) provided by the D3R Grand Challenge 3 organizers, in order to conserve the same coordinate system through the whole process. Hydrogen atoms were added using Hermes, the graphical interface of Gold v5.2.2 software [7], or with Maestro, the graphical interface of Schrödinger software (http://www.schrodinger.com), prior to docking.

Ligands. A training set of CatS ligands was constituted by gathering from ChEMBL, PubChem Assay and BindingDB databases the compounds with known activity (IC50) for CatS (1591 unique compounds). The structures of these compounds were retrieved in SMILES format and were converted into three-dimensional MOL2 files using Ligprep v37014 (Schrödinger, http://www.schrodinger.com). The same conversion from SMILES into MOL2 formats was applied to the D3RGC3 dataset of CatS inhibitors provided by organizers. In all cases, the protonation state for the compounds was adjusted at physiological pH using LigPrep (Schrödinger, http://www.schrodinger.com).

Preliminary docking. In the preliminary analysis step, several docking software and scoring functions have been tested for their ability to predict relative affinity ranking and reproduce the protein-ligand complexes: Gold [7] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [8] and Glide (Schrödinger, http://www.schrodinger.com). Default parameters were used in all cases for docking, except with Gold, where a search efficiency of 200% was

used in order to better explore the conformational space. For Glide, the standard precision (SP) mode was used. The binding sites were considered with Gold as spheres with a 20 Å radius around a virtual point with coordinates (-7,12,-8). With Vina, the binding sites were defined as a $25 \times 25 \times 25 \text{ Å}^3$ cube centered on the same virtual point. With Glide, the binding site was defined with a inner box of 13 x 12 x 13 Å³ and an outer box of 30 x 29 x 30 Å³. Glide attempts to position the ligand center within the inner box volume, while the outer box is defined to contain all the ligand atoms. The docking calculations were carried out with the four CatS conformers described above and with the two structures (containing a DMSO molecule and a SO4 ion, respectively) provided by the organizers. All structures were rigid during the docking process. Ten docking poses were generated for each ligand from the training set and from the CatS D3RGC3 dataset.

Phase 1A docking. From the preliminary analysis, Gold with the ASP scoring function were identified as the most adapted for the given system, and therefore were used for Phase 1 predictions. The conformation CatS_conf4 gave the best results in the preliminary docking step (data not shown) and was used for the docking of most compounds from the CatS D3RGC3 dataset. Notable exceptions are: i) for compounds CatS_4 and CatS_6, which contain a different scaffold, the structure 3MPF was used; ii) for compound CatS_14, the structure provided by organizers containing the DMSO molecule was used; iii) for compounds CatS_2, CatS_2, CatS_17, CatS_20, CatS_22, CatS_23 and CatS_24, the structure provided by organizers containing the SO4 ion was used. The two flip/flop forms of the amide side chain from Asn161 were considered, and then best pose (in terms of score or RMSD) selected. The same Gold docking parameters as described above were used, except that 100 docking poses were generated for each ligand. In the post-docking processing procedure, docking complexes were used as input

structures for 10 ns molecular dynamics simulations in explicit solvent using Gromacs and AMBER 99SB force field. During the simulation, constraints were applied on the protein backbone and on the ligand core fragment that was also constrained during docking. The first 1ns was discarded and the remaining snapshots clusterized (using the g_cluster tool available in Gromacs) in order to provide a single cluster. The representative structure of this cluster was selected for submission (1 structure per compound).

Phase 1B docking. After the release of the 24 crystal structures corresponding to the compounds CatS_1 to CatS_24 (but without the coordinates of the ligands), we carried out docking calculations for each of these structures with its native ligand using the same conditions as in Phase 1A (Gold with the ASP scoring function, 100 docking poses generated for each ligand).

Phase 2 docking. Our experience from Phase 1A using Gold/ASP with scaffold match constraints showed that the atoms with constraints applied were not always well aligned with the reference structure, even when very strong constraints are used (increase of the weight parameter, that determines how closely ligand atoms fit onto the scaffold, up to 2000). Therefore, in Phase 2 we used the Glide/SP protocol described above, with the exception of scaffold constraints that were applied on the common substructure of ligands with either CatS_15, CatS_20 or CatS_24. The best Glide/SP docking scores were selected for Phase 2 ranking prediction submission (1 per compound). The related docking poses for the 33 CatS compounds from the free energy dataset were used as the starting points for following free energy calculations.

11

Free energy calculations. The complexes selected in the previous step represented the input for relative alchemical free energies calculations, using the protocol described in a previous study [9]. Hybrid structures and topologies were built using a modified version of the MOL2FF package developed in our team. CatS 67 was used as reference structure for the free energy calculations. Briefly, equilibrium 10 ns MD simulations were performed for the two states (corresponding to lambda 0 and 1), using Gromacs [10] and OPLS-AA force field [11,12]. Snapshots from the equilibrium runs were extracted to spawn 100 simulations of 50 ps each to alchemically morph between the two states of the system. The work values over every nonequilibrium transition were extracted using some scripts from the PMX software [13-15] in order to estimate the free energy differences relying on the Crooks Fluctuation Theorem and utilizing Crooks-Gaussian Intersection as estimator. As described previously [9], when the two ligands considered for the alchemical transformation had the same charge, separate calculations in explicit solvent were carried out for the transformation of the protein-ligand complex and of the ligand alone. The relative free energy of binding is then calculated as the difference between the corresponding work for these two transformations. When the two ligands had different charges, a single box containing the protein-ligand complex and the ligand alone, separated by 30 Å, was used. The ligand from the complex was converted in one direction, whereas the ligand alone was converted simultaneously in the opposite direction. Thus, the overall charge of the system is conserved during the whole simulation, and the global work for this system represents the relative free energy of binding between ligands.

Submission preparation. Protein structures were converted into PDB format for submission using UCSF Chimera [16], and the docking poses were converted into MOL format using CORINA v3.60 (http://www.molecular-networks.com/), as the MOL format corresponds to the SDF output format in CORINA.

Graphics. Chemical structures were depicted using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/), images for protein structures were generated using PyMol 1.8.1 (Schrödinger, http://www.pymol.org/).

Statistics. Statistics were computed using the R-package (http://www.r-project.org).

Chemoinformatics. Tanimoto similarities were computed using Canvas (Schrödinger).

RESULTS AND DISCUSSION

During the last years we developed a protocol for docking and virtual screening that proved to be relatively successful in our participation to different blind prediction challenges: SAMPL3 (2011) [17], SAMPL4 (2013) [18], CSAR (2014) [19], D3R Grand Challenge (2015) [20] and D3R Grand Challenge 2 (2016) [9]. This protocol involves two steps, the first one representing a preliminary analysis of publicly available structural and biochemical data in order to identify the most appropriate docking software and scoring function for the system of interest. In the second step, we use these parameters for the prediction of binding modes (pose prediction) and of relative affinities of ligands (scoring). Enhanced genetic algorithm parameters for docking (a search efficiency of 200%) proved to be beneficial in our previous studies [17-20,9], especially for big and flexible ligands. Therefore, the same parameters were used in this work to explore as exhaustively as possible the docking conformations of ligands from the CatS D3RGC3 dataset.

Preliminary analysis

The superposition of the 38 crystal structures of CatS identified in the PDB (Table S1) showed very little conformational variability on backbone and on most side chains, with the exception of a few residues shown in Figure 4. Among these structures, 19 contain a ligand in the active site, and 4 of these ligands (599, 23Z, N2A and N2D from the structures 3IEJ, 3KWN, 3MPE and 3MPF, respectively) share a common scaffold with some compounds from the CatS D3RGC3 dataset. The three-dimensional structure of protein in these structures is well conserved, with the exception of two side chains of Phe211 and Phe146 that adopted alternative orientations, as found in 3 representative structures (PDB codes 2F1G, 3N4C, 3MPF). Therefore, we used an apo structure of CatS (PDB code 1GLO) to build 4 conformers, representing the four combinations of representative conformations for Phe211 and Phe146 side chains. In addition, 2 structures (containing a DMSO molecule and a SO4 ion) were provided by the D3RGC3 organizers to take into account specific ligand-DMSO or ligand-SO4 interactions. These six structures were used in the preliminary docking calculations.



Figure 4. All 38 CatS structures available in the PDB superimposed (cartoon representation). Flexible residues are shown as sticks.

A training set containing 1591 unique CatS ligands (known IC50) was created with data from ChEMBL, PubChem Assay and BindingDB databases and used to assess the performance of different docking software in terms of pose prediction and relative ranking of docking poses: Gold with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina and Glide.

The analysis of the docking results showed that no docking program was able to provide a good prediction for the relative affinity ranking. The Spearman's and Kendall's correlation coefficients between docking scores and pIC50 values of ligands from the training set are reported in Table S2.

Surprisingly, the Spearman's and Kendall's correlation coefficients (Table S2) seem to indicate in some cases an anti-correlation between docking scores and pIC50. Indeed, the scatter plot Score versus pIC50 (Figure S2) seems to indicate the existence of two clusters of compounds (two distinct clouds of points). To investigate this observation, we calculated the distance matrix between the compounds from the training set, based on the molprint 2D descriptor, the Tanimoto distance (Canvas, Schrödinger), and a hierarchical cluster analysis (R). These results highlighted the existence of a large cluster of compounds (colored in blue) that tend to have favorable docking scores despite low to moderate pIC50 (Figures S1 and S2, red/blue dendrogram and scatterplot). This blue cluster can explain in part some anti-correlations observed between the docking scores and pIC50 values (Table S2). Further calculation of Tanimoto distances between the compounds from the CatS D3RGC3 and the training set identified 292 compounds with a Tanimoto score >0.5 compared with at least one of the 141 CatS D3RGC3 compounds, and all these compounds belong to the blue cluster (Figures S1 and S2). The Spearman's and Kendall's correlation coefficients for the 292 compounds from the blue cluster (Table S2) show that Vina provided the best results, with correlation coefficients between 0.29 and 0.47 (Spearman's) and between 0.21 and 0.34 (Kendall's), depending on the CatS

conformation used for docking. Other docking programs and scoring functions did not provide good correlations for the 292 compounds from the blue cluster.

The ability of the docking programs to reproduce crystallographic poses was also evaluated. The 141 compounds from the D3RGC3 CatS dataset were docked on the four representative CatS conformations using different docking software and scoring function combinations and the RMSD were calculated between the common substructure of these compounds with the ligand 599 from the structure 31EJ (Figure 5). The mean RMSD values for the 141 compounds from the D3RGC3 CatS dataset are reported in the Table 1, showing that Gold with the ASP scoring function clearly outperformed the other docking programs and scoring functions for all the 4 CatS conformers, with mean RMSD values between 2 and 3 Å. In these conditions, Gold with the ASP scoring function and the CatS_conf4 structure were selected for the next step.



Figure 5. a) Ligand 599 in the active site of cathepsin S (crystal structure, PDB code 3IEJ); b) Chemical structure of ligand 599. The common substructure of this ligand with the compounds from CatS D3RGC3 dataset is represented in sticks (a) and colored in blue (b).

Table 1. Analysis of preliminary docking poses for the 141 compounds from the CatS dataset using four representative conformations of CatS. For all 10 docking poses of each ligand, RMSD values (Å) were calculated using the common substructure between the ligand and compound 599 (see Figure 5). The best RMSD value for each ligand was retained and the overall mean values are shown below.

	CatS_conf1	CatS_conf2	CatS_conf3	CatS_conf4
Gold/Goldscore	4.32	4.72	5.09	4.07
Gold/Chemscore	3.40	3.54	3.46	3.36
Gold/ASP	2.52	2.69	2.91	2.59
Gold/ChemPLP	3.55	4.00	3.98	3.56
Glide/SP	3.19	4.86	4.64	3.40
Vina	3.18	5.78	5.63	3.37

Pose prediction and scoring (Phase 1A)

The 141 CatS ligands were docked with Gold and the ASP scoring function, generating 100 poses for each ligand. The CatS_conf4 conformer was used for docking the vast majority of compounds, with some exceptions (see the Methods section for details).

Three versions of the protocol were tested: i) one without any constraints; ii) one with constraints on the common substructure between the ligand and one of PDB ligands 599, 23Z and N2A and iii) one with the same constraints as above, followed by a 10 ns molecular

dynamics simulation in explicit solvent. The details of all submissions are presented in Table 2. The "best RMSD" submission without constraints (entry 3) and the "best score" submission with constraints (entry 5) provided similar performances (RMSD of 4.36 Å and 4.34 Å, respectively). The post-processing of docking poses with molecular dynamics (MD) simulations (entry 7) clearly improved the quality of the prediction (RMSD of 2.96 Å). Thus, it seems that such MD simulations are beneficial for a better positioning of flexible substituents.

Table 2. Overview of all predictions submitted and their performance (see the Electronic

 Supplementary Material for the plots corresponding to these results).

Entry	Docking protocol	Pose selection	Stage	Prediction type	Submission code	Number of predictions submitted	Mean RMSD (Å)	Spearman correlation	Kendall correlation
1	Gold/ASP no constraints	Best score	1A	pose prediction	gior3	24	5.19	_	_
2	Gold/ASP no constraints	Best score	1A	ranking	dhr26	136	_	0.39±0.08	0.28±0.06
3	Gold/ASP no constraints	Best RMSD	1A	pose prediction	tjwks	24	4.36	_	_
4	Gold/ASP no constraints	Best RMSD	1A	ranking	v3c55	136	_	0.34±0.08	0.24±0.06
5	Gold/ASP scaffold match constraint	Best score	1A	pose prediction	5pmnj	24	4.34	_	_
6	Gold/ASP scaffold match constraint	Best score	1A	ranking	jg6d4	136	_	0.43±0.08	0.29±0.05
7	Gold/ASP scaffold match constraint	MD (Gromacs, Amber 99SB)	1A	pose prediction	va7rj	23	2.96	_	_
8	Vina no constraints	Best score	1A	ranking	_	_	_	0.42	0.29
----	---	--------------	----	------------------------	-------	----	------	-----------	-----------
9	Gold/ASP no constraints	Best score	1B	pose prediction	8wv4r	24	8.49	_	_
10	Gold/ASP no constraints	Best RMSD	1B	pose prediction	8byha	24	3.95	_	_
11	Glide/SP scaffold match constraint	Best score	2	ranking	87q5m	33	_	0.34±0.18	0.20±0.14
12	Glide/SP scaffold match constraint	Best score	2	ranking free energy	js3r3	33	_	0.35±0.17	0.25±0.12

The performance of pose prediction is variable within the dataset, depending on the nature of the ligand, as shown in Figure 6. When only half of the ligand structure is correctly positioned (a), very high RMSD values are obtained, but when the core of the ligand is correctly positioned (b and c), moderate to very good RMSD values can be expected.



Figure 6. Comparison of our Phase 1A best ranking poses without constraints (submission "gior3", cyan) with the corresponding crystal structure conformations (magenta) for compounds CatS_7 (a, RMSD 7.84 Å), CatS_17 (b, RMSD 4.33 Å) and CatS_20 (c, RMSD 1.03 Å).

Pose prediction (Phase 1B)

In Phase 1B, we carried out docking calculations using the 24 crystal structures corresponding to the compounds CatS_1 to CatS_24 that were released at the end of Phase 1A (but without the coordinates of the ligands) with their native ligands. The results from Table 2 (entries 9 and 10) show that the re-docking of ligands into their native structures provided only a slight improvement in the pose prediction for the "best RMSD" submission (from 4.36 Å to 3.95 Å) and a very bad pose prediction for the "best score" submission (8.49 Å instead of 5.19 Å).

Scoring and free energy calculations (Phase 2)

In phase 2, we only predicted the relative ranking for the 33 compounds of the free energy set (Table 2, entries 11 and 12). We used the Glide/SP protocol with scaffold constraints applied on the common substructure of ligands with either CatS_15, CatS_20 or CatS_24. The poses with the best docking score (1 pose per compound) were submitted for Phase 2 ranking prediction. The same CatS complexes were also used for relative free energy calculations, which were carried out using a protocol described in a previous paper [9]. The compound CatS_67 was used as reference, and the calculations were designed in order to minimize the transformations involving a change in the global charge of the ligand (Figure 7). Our previous study [9] has highlighted some potential OPLS-AA force field issues involving non-optimal dihedrals

associated with the sulfonamide group that are now fixed (manuscript in preparation), the simulations from this study using optimal parameters.



Figure 7. Diagram showing the pathways used for relative free energy calculations. The molecule (m) number corresponds to the CatS compounds numbering in the free energy dataset (see Figure 3). The compounds highlighted in green, orange and blue are positively charged, neutral and negatively charged, respectively.

The correlation coefficients obtained in Phase 2 are not as good as those obtained in Phase 1A, indicating that Glide is not performing well in ranking the poses from this free energy subset (Table 2, entry 11). Similarly, the free energy calculations do not improve the quality of the

ranking prediction compared with the direct ranking prediction from the scoring function (Table 2, entry 12). However, much better results are obtained when considering only the cocrystallized ligands (RMSEc of 0.32 ± 0.24 kcal/mol, Kendall's τ of 1.0 ± 0.47 , Spearman's ρ of 1.0 ± 0.49 , Pearson's *r* of 0.83 ± 0.49), indicating a possible bad positioning of the peripheral groups that may affect the correct ranking of the compounds.

CONCLUSIONS

In this challenge we used a protocol involving an initial assessment of the most appropriate docking software and parameters using structural and functional information available in public databases followed by the docking of a CatS inhibitors dataset to provide a prediction for the binding modes and ranking of ligands. The size and conformational flexibility of ligands were important factors to take into account, and the docking calculations afforded reasonable overall pose predictions, which are however dependent on the specific nature of each ligand. In general, the rigid core of the ligands was well positioned, but the conformation of the flexible substituents was more difficult to predict, in spite of the high value of search efficiency used in the docking poses is still a problematic issue. Post-processing of docking poses with molecular dynamics simulations in explicit solvent can significantly improve the prediction. No significant improvement in the ranking prediction, compared with the direct ranking obtained from the scoring function, was provided by free energy calculations.

Electronic Supplementary Material. The Electronic Supplementary Material contains the list of CatS crystal structures from the PDB, the chemical structures of the scoring CatS D3RGC3 dataset, and the plots showing the performance of our submissions.

Acknowledgments. We thank Prof. Bert de Groot for helpful discussions. This work was supported by the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT) [grant number ANR-10-LABX-33], by the JPIAMR transnational project DesInMBL [grant number ANR-14-JAMR-0002] and by the Région Ile-de-France (DIM Malinf).

REFERENCES

1. Thurmond RL, Sun S, Karlsson L, Edwards JP (2005) Cathepsin S inhibitors as novel immunomodulators. Current opinion in investigational drugs (London, England : 2000) 6 (5):473-482

2. Link JO, Zipfel S (2006) Advances in cathepsin S inhibitor design. Current opinion in drug discovery & development 9 (4):471-482

3. Wiener JJ, Sun S, Thurmond RL (2010) Recent advances in the design of cathepsin S inhibitors. Current topics in medicinal chemistry 10 (7):717-732

4. Lee-Dutra A, Wiener DK, Sun S (2011) Cathepsin S inhibitors: 2004-2010. Expert Opin Ther Pat 21 (3):311-337. doi:10.1517/13543776.2011.553800

5. Wilkinson RD, Williams R, Scott CJ, Burden RE (2015) Cathepsin S: therapeutic, diagnostic, and prognostic potential. Biological chemistry 396 (8):867-882. doi:10.1515/hsz-2015-0114

6. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28 (1):235-242

7. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein– ligand docking using GOLD. Proteins Struct Funct Bioinf 52 (4):609-623. doi:10.1002/prot.10465 8. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31 (2):455-461. doi:10.1002/jcc.21334

9. Selwa E, Elisée E, Zavala A, Iorga BI (2018) Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations. J Comput Aided Mol Des 32 (1):273-286. doi:10.1007/s10822-017-0054-1

10. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics (Oxford, England) 29 (7):845-854. doi:10.1093/bioinformatics/btt055

11. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105 (28):6474-6487. doi:10.1021/jp003919d

12. Robertson MJ, Tirado-Rives J, Jorgensen WL (2015) Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. J Chem Theory Comput 11 (7):3499-3509. doi:10.1021/acs.jctc.5b00356

13. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem 36 (5):348-354. doi:10.1002/jcc.23804

14. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. Methods in molecular biology (Clifton, NJ) 1215:173-209. doi:10.1007/978-1-4939-1465-4_9

25

15. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angewandte Chemie (International ed in English) 55 (26):7364-7368. doi:10.1002/anie.201510054

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera -- a visualization system for exploratory research and analysis. J Comput Chem 25 (13):1605-1612. doi:10.1002/jcc.20084

17. Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. J Comput Aided Mol Des 26 (5):595-601. doi:10.1007/s10822-011-9526-x

18. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. J Comput Aided Mol Des 28 (4):455-462. doi:10.1007/s10822-014-9707-5

19. Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. Journal of chemical information and modeling 56 (6):996-1003. doi:10.1021/acs.jcim.5b00337

20. Selwa E, Martiny VY, Iorga BI (2016) Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets. J Comput Aided Mol Des 30 (9):829-839. doi:10.1007/s10822-016-9983-3

Electronic Supplementary Material

Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations

Ludovic Chaput,128 Edithe Selwa,18 Eddy Elisée,18 Bogdan I. Iorga18

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France

² Department of Nephrology and Dialysis, AP-HP, Tenon Hospital, INSERM UMR_S 1155, 75020 Paris, France

Corresponding Author

[•] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

^a These authors contributed equally.

Table of contents

Table S1. Cathepsin S structures available in the PDB	S 3
Table S2. Kendall's τ and Spearman's ϱ correlations for the training set	S 4
Figure S1. Dendrogram for the training set	S 5
Figure S2. Correlation of docking scores with pIC50 for the training set	S 5
Figure S3. Chemical structures of the score CatS dataset	S6-S11
Figure S4. Performance of Phase 1A pose prediction submissions (complete submissions)	S12
Figure S5. Performance of Phase 1A pose prediction submissions (incomplete submissions)	S12
Figure S6. Performance of Phase 1B pose prediction submissions	S13
Figure S7. Performance of Phase 1A ranking prediction submissions	S13
Figure S8. Performance of Phase 2 ranking prediction submissions	S14
Figure S9. Performance of Phase 2 free energy prediction submissions	S14

PDB code	Resolution (Å)	Ligand	Organism
1GLO	2.2	_	Homo sapiens
1MS6	1.9	BLN	Homo sapiens
1NPZ	2.0	C1P	Homo sapiens
1NQC	1.8	C4P	Homo sapiens
2C0Y	2.1	-	Homo sapiens
2F1G	1.9	GNF	Homo sapiens
2FQ9	1.65	-	Homo sapiens
2FRA	1.9	-	Homo sapiens
2FRQ	1.6	_	Homo sapiens
2FT2	1.7	-	Homo sapiens
2FUD	1.95	-	Homo sapiens
2FYE	2.2	-	Homo sapiens
2G6D	2.5	-	Homo sapiens
2G7Y	2.0	-	Homo sapiens
2H7J	1.5	-	Homo sapiens
2HH5	1.8	GNQ	Homo sapiens
2HHN	1.55	GNQ	Homo sapiens
2HXZ	1.9	-	Homo sapiens
2OP3	1.6	TF5	Homo sapiens
2R9M	1.97	Y11	Homo sapiens
2R9N	2.0	-	Homo sapiens
2R9O	2.0	-	Homo sapiens
3IEJ	2.18	599	Homo sapiens
3KWN/5QC4	2.1	23Z	Homo sapiens
3MPE/5QBY	2.25	N2A	Homo sapiens
3MPF/5QBV	1.8	N2D	Homo sapiens
3N3G	1.6	93N	Homo sapiens
3N4C	1.9	EF3	Homo sapiens
30VX	1.49	O64	Homo sapiens
4P6E	1.8	2FC	Homo sapiens
4P6G	1.58	2FZ	Homo sapiens
4BPV	2.0	-	Mus musculus
4BQV	1.7	-	Mus musculus
4BS5	1.25	MG2	Mus musculus
4BS6	1.2	_	Mus musculus
4BSQ	1.96	QQV	Mus musculus
4MZO	1.47	_	Mus musculus
4MZS	1.85	-	Mus musculus

Table S1. Cathepsin S structures available in the PDB

Table S2. Kendall's τ and Spearman's ϱ correlations for the whole training set (black) and of a subset of
compounds having a Tanimoto similarity score of >0.5 with any compound from the CatS D3RGC3 dataset
(blue). See also Figure S1.CatS conf1CatS conf2CatS conf3CatS conf4

	CatS_conf1		CatS_conf2		CatS_conf3		CatS_conf4	
	Kendall	Spearman	Kendall	Spearman	Kendall	Spearman	Kendall	Spearman
Gold	-0.05	-0.07	-0.07	-0.10	-0.10	-0.14	-0.06	-0.09
(Goldscore)	0.04	-0.18	-0.13	-0.21	0.04	-0.23	0.07	-0.09
Gold	-0.20	-0.29	-0.20	-0.28	-0.22	-0.33	-0.18	-0.26
(Chemscore)	-0.12	0.07	-0.14	-0.20	-0.16	0.06	-0.07	0.10
C = 11 (A CD)	-0.07	-0.10	-0.09	-0.13	-0.07	-0.10	-0.08	-0.12
Gold (ASP)	0.04	0.07	-0.05	-0.08	0.00	0.00	0.07	0.10
Gold	-0.09	-0.14	-0.08	-0.12	-0.12	-0.18	-0.10	-0.14
(ChemPLP)	-0.02	-0.02	-0.11	-0.16	-0.08	-0.11	0.06	0.09
Glide (SP)	0.11	0.17	-0.03	-0.04	-0.03	-0.05	0.07	0.10
	0.12	0.17	-0.13	-0.19	-0.06	-0.10	0.02	0.03
Vina	0.04	0.06	0.00	0.00	0.03	0.04	0.04	0.06
	0.25	0.35	0.21	0.29	0.29	0.42	0.34	0.47
	0.20	0.00					0.01	



Figure S1. Dendrogram showing that the training set contains two main clusters (red and blue, left) and that the second cluster contains three distinct sub-clusters (magenta, orange and green, right). The compounds that are similar with the CatS ligands from the D3RGC3 dataset belong all to the blue cluster, and are homogeneously distributed within the three sub-clusters magenta, orange and green.



Figure S2. Correlation of Gold ASP (left) and Vina (right) docking scores (using CatS_conf4) with pIC50 for all compounds from the training set. The compounds are colored according to the two main clusters (see Figure S1).



244 Chapter 2 D3R publications











CatS_87

CatS_91

ÇatS_95





CatS_89

CatS_86







CatS_88



CatS_90







CatS_94







CatS_97

CatS_101





S9





Figure S3. Chemical structures of the score CatS dataset, containing 136 ligands used for ranking prediction.



Grand Challenge 3 - Pose Prediction Method - Cathepsin S (Stage1A)

Pose RMSDs (Å) - Pose 1 Closest Pose Average Pose

Receipt ID

Green bar indicates your p redictions (requires login) Figure S4. Performance of Phase 1A pose prediction submissions for the CatS D3R Grand Challenge 3 dataset (RMSD for pose 1, mean value over all compounds, complete submissions). Our submissions are colored in green. See Table 2 in the main text for details.

Grand Challenge 3 - Pose Prediction Method - Cathepsin S (Stage1A)

Pose RMSDs (Å) - Pose 1 Closest Pose Average Pose Compound: Mean over all 11 10 RMSD (Å) 2 0 Pi5VU Valzj 1873³ CINIVA G CCXI Paktu buexw 8paf Receipt ID Green bar indicates your predictions (requires login)

Figure S5. Performance of Phase 1A pose prediction submissions for the CatS D3R Grand Challenge 3 dataset (RMSD for pose 1, mean value over all compounds, incomplete submissions). Our submission is colored in green. See Table 2 in the main text for details.



Grand Challenge 3 - Pose Prediction Method - Cathepsin S (Stage1B)

Figure S6. Performance of Phase 1B pose prediction submissions for the CatS D3R Grand Challenge 3 dataset (RMSD for pose 1, mean value over all compounds). Our submissions are colored in green. See Table 2 in the main text for details.

Grand Challenge 3 - Active/Inactive Classification - Cathepsin S (Stage 1A)

Kendall's τ Spearman's ρ



Figure S7. Performance of Phase 1A ranking prediction submissions (Kendall Tau, mean value over all compounds) for the CatS D3R Grand Challenge 3 dataset. Our submissions are highlighted in green. See Table 2 in the main text for details.

Grand Challenge 3 - Active/Inactive Classification - Cathepsin S (Stage 2) Kendall's T Spearman's p



Figure S8. Performance of Phase 2 ranking prediction submissions (Kendall Tau, mean value over all compounds, incomplete submissions) for the CatS D3R Grand Challenge 3 dataset. Our submission is highlighted in green. See Table 2 in the main text for details.



Grand Challenge 3 - Active/Inactive Classification - Cathepsin S (Stage 2)

Figure S9. Performance of Phase 2 free energy submissions (Kendall Tau, mean value over all compounds) for the CatS D3R Grand Challenge 3 dataset. Our submission is highlighted in green. See Table 2 in the main text for details.

2.3 D3R–GC4 publication

Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset

Eddy Elisée · Vytautas Gapsys · Nawel Mele · Ludovic Chaput · Edithe Selwa · Bert L. de Groot · Bogdan I. Iorga

Received: date / Accepted: date

Abstract Using the D3R Grand Challenge 4 dataset containing Beta-secretase 1 (BACE) and Cathepsin S (CatS) inhibitors, we have evaluated the performance of

Research reported in this publication was supported by grants ANR-10-LABX-33 (LabEx LERMIT) and ANR-14-JAMR-0002 (JPIAMR) from the French National Research Agency (ANR), by the Région Ilede-France (DIM Malinf), by the Université Paris-Saclay (Globetalkers 2019) and by European Research Council grant ERC-2012-ADG_20120314 (grant agreement 322947).

E. Elisée

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France

V. Gapsys

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

N. Mele

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France

L. Chaput

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France and Sorbonne Université, UPMC Paris 06, Institut National de la Santé et de la Recherche Médicale, Unité Mixte de Recherche S 1155, Paris, France

E. Selwa

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France

B. L. de Groot Max Planck Institute for Biophysical Chemistry, Göttingen, Germany Tel.: +49 551 201 2308 Fax: +49 551 201 2302 E-mail: bgroot@gwdg.de

B. I. Iorga

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, Labex LERMIT, Gif-sur-Yvette, France Tel.: +33 1 69 82 30 94 Fax: +33 1 69 07 72 47

E-mail: bogdan.iorga@cnrs.fr

our in-house docking workflow that involves in the first step the selection of the most suitable docking software for the system of interest based on structural and functional information available in public databases, followed by the docking of the dataset to predict the binding modes and ranking of ligands. The macrocyclic nature of the BACE ligands brought additional challenges, which were dealt with by a careful preparation of the three-dimensional input structures for ligands. This provided top-performing predictions for BACE, in contrast with CatS, where the predictions in the absence of guiding constraints provided poor results. These results highlight the importance of previous structural knowledge that is needed for correct predictions on some challenging targets. After the end of the challenge, we also carried out free energy calculations (i.e. in a non-blinded manner) for CatS using the pmx software and several force fields (AMBER, Charmm). Using knowledge based starting pose construction allowed reaching remarkable accuracy for the CatS free energy estimates. Interestingly, we show that the use of a consensus result, by averaging the results from different force fields, increases the prediction accuracy.

Keywords molecular docking \cdot free energy calculations \cdot molecular dynamics \cdot pmx \cdot D3R challenge \cdot Beta secretase 1 \cdot Cathepsin S \cdot Inhibitors

1 Introduction

Over the years, the blinded prediction challenges regularly organized by the Drug Design Data Resource (D3R, https://drugdesigndata.org/) became valuable opportunities to assess the quality and performance of in-house methodologies often used in computer-aided drug discovery projects. The concept is simple: datasets are generously provided by known pharmaceutical companies and presented to the computational chemistry community, which is invited to blindly predict binding modes and relative affinities of ligands.

The D3R Grand Challenge 4 was organized in 2018 and was based on two protein targets: cathepsin S (CatS, Fig. 1a), which was already present in the previous D3R Grand Challenge 3, and beta-secretase 1 (BACE, Fig. 1b). Cathepsin S is a lysosomal cysteine protease involved, inter alia, in antigen presentation, nociception, itch, pain and whose regulation may be important in several diseases as psoriasis, rheumatoid arthritis and glioblastoma [1–9]. Beta-secretase 1 is an aspartic-acid protease involved in the formation of myelin sheaths in peripheral nerve cells and in Alzheimer's disease by cleavage of the amyloid precursor protein, leading, in the end, to amyloid- β peptide aggregation [10–13]. However, recent publications relate the possible fail of BACE inhibitors in Alzheimer's trials [14, 15].

The BACE subchallenge included three stages. In Phase 1A, the participants were asked to predict the crystallographic poses of 20 BACE ligands (Fig. 2), the affinity ranking for 154 BACE ligands (Fig. S1 in the Electronic Supplementary Material) and the relative binding affinities for a designated free energy subset of 34 BACE ligands (Fig. 3). In Phase 1B, organizers released the corresponding receptor structures (without ligands) for the 20 BACE ligands composing the pose prediction subset, and the participants could repeat the pose prediction using this additional information. In Phase 2, organizers released the complexes from the pose prediction subset and the

2



Fig. 1 Solid surface representation of representative crystal structures of BACE (**a**, PDB code 3K5C [16]) and cathepsin S (**b**, PDB code 1GLO [17]) with the ligands (**BACE_68** and **CatS_86**, respectively) in stick representation and colored in cyan. The binding sites, defined as spheres of 20 Å around the atom OG1 of THR72 in **a** and around a virtual point with coordinates [-7,8,-4] in **b**, are colored in pink and orange, respectively.

participants were required to repeat the affinity predictions for the subset of 154 ligands and the relative free energy for the subset of 34 compounds as in Phase 1A, but taking into account the structural information from the 20 released protein-ligand complexes. The CatS subchallenge included a single stage, Phase 2, with ranking prediction for two datasets of 459 ligands (structures not shown) and 39 ligands (Fig. 4), the latter being designed for relative free energy calculations.

2 Methods

2.1 Protein structures

Analysis of Protein Data Bank [18] revealed 341 available crystal structures of human BACE but only 337 of them, including four synthetic structures (1M4H [19], 4TRW [20], 4TRY [20], 4TRZ [20]), were considered in this work since two structures were not released when the panel of proteins has been built (6EQM [21] and 6DMI [22]) and two proteins contained useless extra-domains (1UJJ [23], 1UJK [23]). Furthermore, four apo structures were found but only one was complete, namely without missing residues (1SGZ [24]).

The three-dimensional alignment of all these structures revealed a perfectly conserved fold except for three regions: the first turn near the active site (VAL309-ASP318), the second turn (LYS9-GLY13) with open (e.g. structure 2VIZ [25]) and closed (e.g. structure 2VIJ [26]) conformations, and the third turn (PRO70-GLY74).

The proteins were protonated at physiological pH using the appropriate tool corresponding to each docking software: Hermes for GOLD and Maestro for Glide and Vina.



Fig. 2 Chemical structures of the D3R GC4 BACE pose prediction dataset.

The structural analysis for cathepsin S was carried out during our participation to the D3R Grand Challenge 3 in 2017 [27] and, given the similarity of the CatS ligands proposed in the previous and the present challenges, we decided to use in this study the same protocol, with the exception of the constraints during the docking process, which were not used anymore.

2.2 Ligands

Three-dimensional coordinates for all compounds were generated in MOL2 format using two protocols: i) random conformations were built from SMILES strings using LIGPREP V45011 (Schrödinger, http://www.schrodinger.com), software also used to adjust protonation states at physiological pH; and ii) bioactive conformations were built manually from the closest macrocycle-containing BACE ligands available in the PDB (see the Electronic Supplementary Material for the correspondence between these structures) or from the closest CatS ligands released during the previous D3R GC3 challenge (haan-CatS_11, wcgq-CatS_10 and mekm-CatS_13), using UCSF Chimera [28].

The ajustement of protonation states at physiological pH for all ligands makes that the CatS ligands have a total charge of +2, which is in agreement with the instructions provided by the organizers of D3R GC4 challenge stating that "all compounds in both the BACE and CatS free energy sets [are expected to have] a charge of +2 at the assay pH values of 4.5 (BACE) and 5.0 (CatS)." In contrast, our BACE ligands have a total charge of +1, which is in apparent contradiction with these instructions. We were concerned about this difference and we investigated in more detail the BACE binding



Fig. 3 Chemical structures of the D3R GC4 BACE free energy dataset.

site residues that would interact with this second charge (which in individual BACE ligands should be positioned in different regions of space). We could evidence no specific ionic interaction and therefore, we continued our study with ligands having a total charge of +1. To the best of our knowledge, there is no specific study providing an experimental proof that the BACE inhibitors from this series have a total charge of +2 when bound to the protein.

We organized the BACE ligands in four groups, according to the nature of the macrocyclic core: (A) aliphatic macrocycle, (B) one aromatic ring, (C) two aromatic rings, (D) no macrocycle. The substituents bound to the macrocycle are hereinafter referred to as "lateral chains". Table S1 in the Electronic Supplementary Material sums up the design plan of the 158 ligands.



Fig. 4 Chemical structures of the D3R GC4 CatS free energy dataset.

2.3 Preliminary docking

A training set was created with 14 BACE structures containing ligands similar with those from the D3R GC4 dataset (similarity evaluated using Tanimoto method). IC50 values could be retrieved from BindingDB database [29] for 12 of them (see Table S2 in the Electronic Supplementary Material).

The alignment of these structures showed a reasonable conserved three-dimensional structure, except for two amino acids: i) the conformation of ARG235 in the structure 3DV1 [30] clashes with the ligands from structures 2VIJ [26], 2VJ7 [31], 2VNM [32], 3K5C [16] and 4DPF [33], whereas other equivalent arginines, e.g. ARG296 in the structure 2VIJ [26], show no clashes; b) the conformation of GLN73 in the structure 3DV1 [30], like most equivalent glutamines in this position, clashes with all ligands except those from the 3K5C [16], 3K5F [34] and 3IVH [35] structures.

We finally chose the 3K5C structure for all subsequent docking studies as its active site is open enough to accept all other ligands from the training set (only 4DPF ligand has a phenyl substituent, as in **BACE_69** and **BACE_82**, which is slightly clashing with the 3K5C surface, but no amino acid in its environment seems to be an issue).

In the next step, several docking software and scoring functions have been tested for their ability to predict relative affinity ranking and reproduce the protein-ligand complexes from this training dataset: GOLD [36] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [37] and Glide (Schrödinger, http: //www.schrodinger.com). Default parameters were used in all cases for docking, except with GOLD, where a search efficiency of 200% and the "flip ring corners" option were used in order to better explore the conformational space and of the ligands macrocycles. For Glide, the standard precision (SP) mode was used. The binding sites were considered with GOLD as spheres with a 20 Å radius around the OG1 atom of THR72 (numbering from the 3K5C structure). With Vina, the binding sites were defined as a 62 x 64 x 80 Å³ cube centered on the same atom. The protein structure (3K5C) was considered to be rigid during the docking process and the ligands fully flexible. Fifty docking poses were generated for each ligand from the training set.

2.4 Phase 1A docking

From the preliminary analysis, Gold with the GoldScore scoring function were identified as the most adapted for the BACE system (see Section 3.1 and Table 1 for a more detailed discussion), and therefore were used for Phase 1 predictions with default values except search efficiency which was set to 200%. To explore ring conformations and the ligand flexibility the options flip ring corners, flip pyramidal N and flip amide bonds were activated. All the ligands were docked onto the 3K5C [16] structure, the ligands being fully flexible and the receptor rigid. The docking calculation was carried out with two hydrogen bond constraints involving the backbone oxygen atoms of residues GLY230 and GLY34 and any atoms of the ligands. Additional docking calculations were also carried out with ligands generated from SMILES, and also without any constraints. In all these cases, as expected, the poses are slightly worse compared to those obtained with constraints, and only the latter were submitted.

Analysis of the training dataset has emphasized a conserved hydrogen bond in all the ligands involving the backbone oxygen atom of residue GLY230 and also one hydrogen bond, founds in most of the ligands, involving the backbone oxygen atom of residue GLY34. Therefore, docking simulations were carried out with hydrogen bond constraints on either or both oxygen atoms, and also without constraints, for all compounds from the BACE D3RGC4 dataset, with 100 docking poses generated for each ligand.

The RMSD between the common substructure of the macrocyclic core and the side chain parts of the ligand and of one of PDB ligands AR9 (3DV1 [30]), 0BI (3K5C [16]), OLG (4DPF [33]), 0XA (4GMI), 1R8 (4KE0 [38]) and BAV (3DV5 [30]) were calculated using an in house developed script based on CACTVS Chemoinformatics Toolkit (http://www.xemistry.com/). The two RMSD values were added and the poses with the smaller sum were selected for submission (1 score per compound).

The second submission (1 score per compound) consisted of the best ranking pose for each docking.

Protein structures were converted into PDB format for submission using UCSF Chimera [28], and the docking poses were converted into MOL format using CO-RINA version 3.60 (http://www.molecular-networks.com) (the MOL format corresponds to the SDF output format in CORINA).

2.5 Phase 1B docking

After the release of the 20 crystal structures corresponding to the compounds **BACE_1** to **BACE_20** (but without the coordinates of the ligands), we carried out docking calculations with each ligand individually onto the corresponding structure released at the end of Phase 1A (Gold with the GoldScore scoring function, 100 docking poses generated for each ligand). Each ligand was individually docked onto the corresponding structure released at the end of Phase 1A. The docking calculation was carried out with two hydrogen bond constraints involving the backbone oxygen atoms of residues GLY230 and GLY34 and any atoms of the ligands.

2.6 Phase 2 docking

As no significantly new structural information was brought by the 20 protein-ligand complexes released at the end of Phase 1B, for the BACE subchallenge in Phase 2 we have submitted the same files as in Phase 1A.

For the CatS subchallenge we used our protocol from Grand Challenge 3 (2017) [27] and performed the docking with Gold using ASP scoring function and 100 poses per ligand. In contrast with our previous predictions of CatS inhibitors [27], this time no constraints were used during the docking process.

2.7 Free energy calculations

All free energy calculations were carried out after the end of the challenge (i.e. in a non-blinded manner).

The D3R GC4 cathepsin S free energy dataset was composed of 39 structurally similar CatS ligands (Fig. 4).

Free energies for Cathepsin S binders were estimated using non-equilibrium free energy calculation protocol [39]. Firstly, a mapping between ligands was established (Fig. 5): ligand **CatS_79** was used as an initial reference connecting to the other molecules (gray edges in the graph). This way double free energy differences ($\Delta \Delta G$) for every compound with respect to **CatS_79** could be evaluated. Subsequently, redundancies in the graph were introduced (red edges) to allow for formation of cycles that could be further used to correct for under-sampling [40]. In total 109 transitions were considered.



Fig. 5 Graph of CatS ligand mappings for relative free energy calculations. Circle size and color encode the node connectivity, i.e. the number of edges connected to a node (ligand). Gray edges mark the connections between every ligand and compound CatS_79; the additional red edges have been introduced as redundancies to allow for cycle closure correction.

For the free energy calculations of CatS inhibitors we used the same ligand structures that were used as input for docking calculations (see Section 2.2 for details).

For every transition between two ligands hybrid structures and topologies were generated using pmx package [41]. Prior to starting molecular dynamics simulations, the solvated systems were energy minimized by keeping the position restraints of 1000 kJ/mol⁻¹nm⁻² on all heavy atoms. Afterwards, a 10 ps constant temperature and volume simulations were performed still retaining the position restraints on heavy atoms. Starting from the end states of the NVT simulations, equilibrium simulations in isothermal-isobaric conditions for the ligands in solution and bound to protein were performed for 6 ns. Subsequently, 80 snapshots were extracted equidistantly in time from the last 4 ns of each trajectory and rapid 200 ps transitions were spawned driving the system from one physical state (ligand A) to another (ligand B). The nonequilibrium work values were recorded for every transition and related to the free energy difference based on the Crooks Fluctuation Theorem [42]. Maximum likelihood estimator was used to estimate free energy difference [43]. The whole procedure, including equilibrium sampling and non-equilibrium transitions, was repeated three times. The final free energy estimate was calculated as a mean of the three replicas. The errors for individual free energy differences were calculated as standard errors of the three repeats.

All the molecular dynamics simulations were performed with Gromacs [44]. Two protein force fields were used: Amber99SB*ILDN [45–47] and Charmm36m [48]. For ligands the generalized amber force field (GAFF v2.0) [49] was used in combination with Amber99sb*ILDN. Atom types and charges for ligands in Charmm force field were assigned using MATCH algorithm [50]; CGenFF 4.1 [51] was used for bonded parameters. The systems were solvated with TIP3P water [52]. Na⁺ and Cl⁻ ions were added to neutralize the simulation box and reach 150 mM salt concentration. Particle Mesh Ewald [53, 54] was used to treat long range electrostatics with the real space cutoff of 1.1 nm, Fourier spacing of 0.12 nm. Van der Waals interactions were smoothly switched off between 1.0 and 1.1 nm. All bonds were constrained by the LINCS [55] algorithm. Temperature was kept at 298 K by means of the velocity rescaling thermostat [56] with the time constant of 0.1 ps. Pressure of 1 bar was controlled with the Parrinello-Rahman barostat [57] with the time constant of 5 ps.

For reconstructing ΔG values from double free energy differences ($\Delta \Delta G$) CatS_79 was used as a reference. Free energies were calculated with GAFF (Amber) and CGenFF (Charmm) force fields. Furthermore, a consensus result was obtained by averaging the results from GAFF and CGenFF force fields [58]. The overall calculation accuracy was assessed by means of root mean squared error (RMSE), Pearson's, Kendall's and Spearman's correlation coefficients. The errors for these measures were obtained by means of bootstrapping. The bootstrap procedure followed closely the one described in previous D3R challenges [59]: namely, 10,000 samplings with replacement were performed from the calculated and experimental value sets. The experimental values were modified by adding Gaussian noise with the mean of zero and standard deviation of RTln *I*_{err}, whith *I*_{err} = 2.5.

The D3R GC4 BACE free energy dataset was composed of 34 structurally similar BACE ligands (Fig. 3). During the free energy calculations we have encountered difficulties with the stability of simulations and convergence problems. These might be due, at least in part, to the variations in the macrocyclic scaffold of BACE ligands from this dataset.

10

2.8 Graphics

The figures representing protein-ligand complexes were generated with Pymol [60], and the chemical structures were drawn using ChemDraw Professional 16.0 (PerkinElmer Informatics). The CatS ligand perturbation network was built with NetworkX [61]. The results of CatS calculations were visualized using Matplotlib [62] library.

3 Results and Discussion

From our participation to previous docking and virtual screening challenges, SAMPL3 (2011) [63], SAMPL4 (2013) [64], CSAR (2014) [65], D3R Grand Challenge (2015) [66], D3R Grand Challenge 2 (2016) [67] and D3R Grand Challenge 3 (2017) [27], we developed a protocol for docking and virtual screening that proved to be relatively successful. This protocol involves two steps, the first one representing a preliminary analysis of publicly available structural and biochemical data in order to identify the most appropriate docking software and scoring function for the system of interest. In the second step, we use these parameters for the prediction of binding modes (pose prediction) and of relative affinities of ligands (scoring). Enhanced genetic algorithm parameters for docking (a search efficiency of 200%) proved to be beneficial in our previous studies [27, 63–67], especially for big and flexible ligands. Thus, in this work we used the same parameters in order to ensure an adequate conformational sampling of docking conformations.

3.1 Preliminary analysis

In the preliminary analysis step, the ligands from the training set were docked using the structure 3K5C [16] and different combinations of docking software and scoring functions, and the results analyzed for accuracy in terms of pose prediction and ranking. In order to evaluate the accuracy of docking and scoring, we have considered the lowest RMSD value and the RMSD value of the best ranking pose for each combination protein-ligand-(docking software)-(scoring function) reported in Table 1. The RMSD values were computed for all atoms, with the crystallographic structures of these ligands as reference (see Table S2 for the list of ligands with known structures composing the training dataset.) Table 1 shows that Gold with the Gold-Score scoring function clearly outperformed the other docking programs and scoring functions in the pose prediction of the training set. Vina provided very poor results, with most of the docking conformations positioned outside the binding site, whereas Gold/GoldScore, followed by Gold/ChemScore and Glide/SP, could reproduce rather well the native protein-ligand complexes. In these conditions Gold with GoldScore with the 3K5C structure were used for the following steps.

No significant correlation could be found between the docking score and biological activity (pIC50) for the 12 compounds from the training set for which experimental biological data were available.

		0
Software/Scoring function	Method for generating ligand 3D structure	Average RMSD for best pose (A)
Glide/SP	SMILES	3.24
GOLD/GoldScore	SMILES	2.48
GOLD/ChemScore	SMILES	3.14
GOLD/ChemPLP	SMILES	2.95
GOLD/ASP	SMILES	3.50
Vina	SMILES	6.87
Glide/SP	PDB	3.14
GOLD/GoldScore	PDB	2.35
GOLD/ChemScore	PDB	2.85
GOLD/ChemPLP	PDB	3.27
GOLD/ASP	PDB	3.19
Vina	PDB	6.41

 Table 1 Preliminary evaluation of the training dataset. See Table S2 for the list of ligands with known structures composing the training dataset.

3.2 Phase 1A: Pose prediction and ranking (BACE)

The 154 BACE ligands were docked on 3K5C [16] structure using Gold and the GoldScore scoring function generated 50 poses for each ligand. Three versions of the protocol were tested (see the Methods section for details) with and without hydrogen bonds constraints. For the ligands belonging to a group for which crystal structures were available, the RMSD was calculated between the common substructure of the macrocyclic core and the side chain parts of the ligand and of one of PDB ligands AR9 (3DV1 [30]), 0BI (3K5C [16]), OLG (4DPF [33]), 0XA (4GMI), 1R8 (4KE0 [38]) and BAV (3DV5 [30]) using an in house developed script based on CACTVS Chemoinformatics Toolkit (http://www.xemistry.com/). The two RMSD values were added and the poses with the smaller sum were selected for submission in phases 1 and 2. RMSD values compared with the part of the native ligands from the crystallographic structures were calculated for all docking poses. We obtained RMSD values calculated on the common substructure with ligands with known structures ranging from 1.52 Å to 1.90 Å for the "Best RMSD" prediction and from 1.81 Å to 2.30 Å for the "Best Score" prediction. The results demonstrate that the use of constraints induced similar performances in the two cases and, as expected, the poses without constraints are slightly worse compared to those obtained with constraints, and only the latter were submitted.

The release of the Phase 1A results showed that our two BACE ranking predictions performed the best in this competition, and that the "Best RMSD" pose prediction was ranked 7th. The pose predictions for the best and the worst compounds are presented in Fig. 6.

3.3 Phase 1B: Pose prediction (BACE)

In Phase 1B, we carried out docking calculations using the 20 crystal structures corresponding to the compounds **BACE_1** to **BACE_20** that were released at the end of Phase 1A (but without the coordinates of the ligands) with their native ligands. The



Fig. 6 Comparison of our BACE Phase 1A best and worst pose predictions for compounds **BACE_3** (a, green) and **BACE_7** (b, cyan) with the corresponding BACE complexes that were released at the end of the challenge (RMSD 0.39 Å and 2.60 Å, respectively).

re-docking of ligands into their native structures provided only a slight improvement in the pose prediction. Similar to the phase 1A, adding constraints did not improve the predictions, so only simulations with one constraint were submitted.

3.4 Phase 2: Ranking (BACE and CatS)

As no significantly new structural information was brought by the 20 protein-ligand complexes released at the end of Phase 1B, for the BACE subchallenge in Phase 2 we decided to submit the same results as in Phase 1A.

Concerning CatS simulations, similar as it was done for BACE, RMSD calculations between the common substructure of the ligand and of one of PDB ligands B8J (5QBU), BCJ (5QC6), BJJ (5QCG), B9S (5QC1) and BFV (5QC8) were calculated using an in house developed script based on CACTVS Chemoinformatics Toolkit (http://www.xemistry.com/). The ranking with the best final RMSD poses were selected for submission (see Section 2.4 for details about RMSD calculation).

The pose prediction is variable within the CatS docking simulations, inducing poorer results compared with BACE. Some ligands are incorrectly positioned as illustrated in Fig. 7a, but when the core of the ligand is correctly positioned (Fig. 7b), a significant improvement in RMSD is observed.

3.5 Free energy calculations

For timing reasons, no free energy predictions were submitted during the challenge, and all free energy calculations are presented retrospectively.

Overall, calculated free energies are well within 1 kcal/mol deviation from the experimental measurements in terms of AUE and RMSE_C (Table S3,Fig. 8). In terms of absolute agreement with the experimental Δ G values, as measured by AUE and RMSE_C, our calculations using GAFF force field outperformed all the other submissions to the D3R GC4 (Fig. 8). When predicting the overall trend (Pearson correlation) and ligand ranking by their binding affinity (Kendall's and Spearman's correlation), GAFF based calculations were only superseded by another set of submissions utilizing the same ligand force field in an alchemical free energy calculation setup.



Fig. 7 Comparison of our CatS Phase 2 best ranking poses for compounds **CatS_105** (a, cyan) and **CatS_245** (b, green) with the ligand **BCJ** (magenta) from the CatS crystal structure 5QC6 (RMSD 11.20 Å and 1.88 Å, respectively, on the macrocyclic core).

The free energy estimates based on the CGenFF and Charmm36m force field combination performed slightly, but consistently worse with respect to all the considered quality measures.

Prediction accuracy further increased when applying the consensus approach. This observation is consistent with the previously reported observations for protein thermostabilities [58], DNA nucleotide mutations [68], protein-ligand interactions [69].

Of the 39 estimated free energy values, the consensus force field approach only in 3 cases shows a deviation from experiment larger than 1 kcal/mol. One of the major sources of error in the current calculation procedure comes from the starting pose definition. In the current setup, pose construction based on the structural data from the previous D3R GC3 challenge (Section 2.2) proved to yield highly accurate results. In other attempts, where we probed the accuracy of alchemical predictions when starting from a distorted set of poses the estimation accuracy markedly decresed. All in all, starting pose generation needs to be taken with care as it largely modulates accuracy achievable with alchemical free energy calculations.

Force field	AUE (kcal/mol)	RMSE _C (kcal/mol)	Pearson's r	Kendall's $ au$	Spearman's ρ
GAFF	0.40 ± 0.09	0.45 ± 0.10	0.72 ± 0.14	0.49 ± 0.12	0.67 ± 0.16
CGenFF	0.48 ± 0.10	0.50 ± 0.11	0.60 ± 0.16	0.39 ± 0.13	0.54 ± 0.17
Consensus	0.37 ± 0.10	0.43 ± 0.10	0.72 ± 0.15	0.49 ± 0.12	0.67 ± 0.16

 Table 2
 Summary of the calculated alchemical free energy accuracies in terms of average unsigned error (AUE), root mean squared error (RMSE), Pearson's, Kendall's and Spearman's correlation coefficients.

4 Conclusions

Using the D3R Grand Challenge 4 dataset containing Beta-secretase 1 (BACE) and Cathepsin S (CatS) inhibitors, we have evaluated the performance of our in-house



Fig. 8 Summary of the CatS free energy calculation results. The calculations performed in this work (GAFF - blue, CGenFF - red, consensus - yellow) are shown together with the other submissions: **a** RMSE_C, **b** Pearson correlation **c**, Kendall's τ and **d** Spearman's correlation.

docking workflow that involves in the first step the selection of the most suitable docking software for the system of interest based on structural and functional information available in public databases, followed by the docking of the dataset to predict the binding modes and ranking of ligands. The macrocyclic nature of the BACE ligands brought additional challenges, which were dealt with by a careful preparation of the three-dimensional input structures for ligands. This provided top-performing predictions for BACE, in contrast with CatS, where the predictions in the absence of guiding constraints provided poor results. These results highlight the importance of previous structural knowledge that is needed for correct predictions on some challenging targets. After the end of the challenge, we also carried out free energy calculations (i.e. in a non-blinded manner) for CatS using the pmx software and several force fields (AMBER, Charmm). Using knowledge based starting pose construction allowed reaching remarkable accuracy for the CatS free energy estimates. Interestingly, we show that the use of a consensus result, by averaging the results from different force fields, increases the prediction accuracy.

References

- 1. Reddy VB, Sun S, Azimi E, Elmariah SB, Dong X, Lerner EA (2015) Redefining the concept of protease-activated receptors: cathepsin S evokes itch via activation of Mrgprs. Nat Commun 6:7864, DOI 10.1038/ncomms8864
- 2. Ainscough JS, Macleod T, McGonagle D, Brakefield R, Baron JM, Alase A, Wittmann M, Stacey M (2017) Cathepsin S is the major activator of the
psoriasis-associated proinflammatory cytokine IL-36γ. Proc Natl Acad Sci U S A 114(13):E2748–E2757, DOI 10.1073/pnas.1620954114

- Elmariah SB, Reddy VB, Lerner EA (2014) Cathepsin S signals via PAR2 and generates a novel tethered ligand receptor agonist. PLoS One 9(6):e99,702, DOI 10.1371/journal.pone.0099702
- Xu J, Wang H, Ding K, Lu X, Li T, Wang J, Wang C, Wang J (2013) Inhibition of cathepsin S produces neuroprotective effects after traumatic brain injury in mice. Mediators Inflamm 2013:187,873, DOI 10.1155/2013/187873
- 5. Thurmond RL, Sun S, Karlsson L, Edwards JP (2005) Cathepsin S inhibitors as novel immunomodulators. Curr Opin Investig Drugs 6(5):473–82
- Link JO, Zipfel S (2006) Advances in cathepsin S inhibitor design. Curr Opin Drug Discov Devel 9(4):471–82
- Wiener JJM, Sun S, Thurmond RL (2010) Recent advances in the design of cathepsin S inhibitors. Curr Top Med Chem 10(7):717–32
- Lee-Dutra A, Wiener DK, Sun S (2011) Cathepsin S inhibitors: 2004-2010. Expert Opin Ther Pat 21(3):311–37, DOI 10.1517/13543776.2011.553800
- Wilkinson RDA, Williams R, Scott CJ, Burden RE (2015) Cathepsin S: therapeutic, diagnostic, and prognostic potential. Biol Chem 396(8):867–82, DOI 10.1515/hsz-2015-0114
- Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, Teplow DB, Ross S, Amarante P, Loeloff R, Luo Y, Fisher S, Fuller J, Edenson S, Lile J, Jarosinski MA, Biere AL, Curran E, Burgess T, Louis JC, Collins F, Treanor J, Rogers G, Citron M (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 286(5440):735–41
- Willem M, Garratt AN, Novak B, Citron M, Kaufmann S, Rittger A, DeStrooper B, Saftig P, Birchmeier C, Haass C (2006) Control of peripheral nerve myelination by the beta-secretase BACE1. Science 314(5799):664–6, DOI 10.1126/science.1132341
- 12. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhangale T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jönsson EG, Palotie A, Behrens TW, Magnusson OT, Kong A, Thorsteinsdottir U, Watts RJ, Stefansson K (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488(7409):96–9, DOI 10.1038/nature11283
- 13. Baxter EW, Conway KA, Kennis L, Bischoff F, Mercken MH, Winter HLD, Reynolds CH, Tounge BA, Luo C, Scott MK, Huang Y, Braeken M, Pieters SMA, Berthelot DJC, Masure S, Bruinzeel WD, Jordan AD, Parker MH, Boyd RE, Qu J, Alexander RS, Brenneman DE, Reitz AB (2007) 2-Amino-3,4dihydroquinazolines as inhibitors of BACE-1 (beta-site APP cleaving enzyme): Use of structure based design to convert a micromolar hit into a nanomolar lead. J Med Chem 50(18):4261–4, DOI 10.1021/jm0705408
- Panza F, Lozupone M, Solfrizzi V, Sardone R, Piccininni C, Dibello V, Stallone R, Giannelli G, Bellomo A, Greco A, Daniele A, Seripa D, Logroscino G, Imbimbo BP (2018) BACE inhibitors in clinical development for the

treatment of Alzheimer's disease. Expert Rev Neurother 18(11):847–857, DOI 10.1080/14737175.2018.1531706

- Burki T (2018) Alzheimer's disease research: the future of BACE inhibitors. Lancet 391(10139):2486, DOI 10.1016/S0140-6736(18)31425-9
- 16. Lerchner A, Machauer R, Betschart C, Veenstra S, Rueeger H, McCarthy C, Tintelnot-Blomley M, Jaton AL, Rabe S, Desrayaud S, Enz A, Staufenbiel M, Paganetti P, Rondeau JM, Neumann U (2010) Macrocyclic BACE-1 inhibitors acutely reduce Abeta in brain after po application. Bioorg Med Chem Lett 20(2):603–7, DOI 10.1016/j.bmcl.2009.11.092
- Turkenburg JP, Lamers MBAC, Brzozowski AM, Wright LM, Hubbard RE, Sturt SL, Williams DH (2002) Structure of a Cys25–>Ser mutant of human cathepsin S. Acta Crystallogr D Biol Crystallogr 58(Pt 3):451–5, DOI 10.1107/s0907444901021825
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–42, DOI 10.1093/nar/28.1.235
- Hong L, Turner RT 3rd, Koelsch G, Shin D, Ghosh AK, Tang J (2002) Crystal structure of memapsin 2 (beta-secretase) in complex with an inhibitor OM00-3. Biochemistry 41(36):10,963–7, DOI 10.1021/bi026232n
- Hattori Y, Kobayashi K, Deguchi A, Nohara Y, Akiyama T, Teruya K, Sanjoh A, Nakagawa A, Yamashita E, Akaji K (2015) Evaluation of transition-state mimics in a superior BACE1 cleavage sequence as peptidemimetic BACE1 inhibitors. Bioorg Med Chem 23(17):5626 – 5640, DOI https://doi.org/10.1016/j.bmc.2015.07.023
- 21. Neumann U, Ufer M, Jacobson LH, Rouzade-Dominguez ML, Huledal G, Kolly C, Lüönd RM, Machauer R, Veenstra SJ, Hurth K, Rueeger H, Tintelnot-Blomley M, Staufenbiel M, Shimshek DR, Perrot L, Frieauff W, Dubost V, Schiller H, Vogg B, Beltz K, Avrameas A, Kretz S, Pezous N, Rondeau JM, Beckmann N, Hartmann A, Vormfelde S, David OJ, Galli B, Ramos R, Graf A, Lopez Lopez C (2018) The BACE-1 inhibitor CNP520 for prevention trials in Alzheimer's disease. EMBO Mol Med 10(11), DOI 10.15252/emmm.201809316
- 22. van Zundert GCP, Hudson BM, de Oliveira SHP, Keedy DA, Fonseca R, Heliou A, Suresh P, Borrelli K, Day T, Fraser JS, van den Bedem H (2018) qFitligand reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps. J Med Chem 61(24):11,183–11,198, DOI 10.1021/acs.jmedchem.8b01292
- Shiba T, Kametaka S, Kawasaki M, Shibata M, Waguri S, Uchiyama Y, Wakatsuki S (2004) Insights into the phosphoregulation of beta-secretase sorting signal by the VHS domain of GGA1. Traffic 5(6):437–48, DOI 10.1111/j.1600-0854.2004.00188.x
- Hong L, Tang J (2004) Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. Biochemistry 43(16):4689–95, DOI 10.1021/bi0498252
- Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G,

Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs). Bioorg Med Chem Lett 18(3):1011–6, DOI 10.1016/j.bmcl.2007.12.017

- 26. Beswick P, Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Gleave R, Hawkins J, Hussain I, Johnson CN, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Skidmore J, Soleil V, Smith KJ, Stanway S, Stemp G, Stuart A, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 3: identification of hydroxy ethylamines (HEAs) with nanomolar potency in cells. Bioorg Med Chem Lett 18(3):1022–6, DOI 10.1016/j.bmcl.2007.12.020
- Chaput L, Selwa E, Elisée E, Iorga BI (2019) Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations. J Comput Aided Mol Des 33(1):93–103, DOI 10.1007/s10822-018-0161-7
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–12, DOI 10.1002/jcc.20084
- Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) Bindingdb: a webaccessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35(Database issue):D198–201, DOI 10.1093/nar/gkl999
- Machauer R, Laumen K, Veenstra S, Rondeau JM, Tintelnot-Blomley M, Betschart C, Jaton AL, Desrayaud S, Staufenbiel M, Rabe S, Paganetti P, Neumann U (2009) Macrocyclic peptidomimetic beta-secretase (BACE-1) inhibitors with activity in vivo. Bioorg Med Chem Lett 19(5):1366–70, DOI 10.1016/j.bmcl.2009.01.055
- 31. Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character. Bioorg Med Chem Lett 18(3):1017–21, DOI 10.1016/j.bmcl.2007.12.019
- 32. Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, East P, Hawkins J, Howes C, Hussain I, Jeffrey P, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2008) Second generation of hydroxyethy-lamine BACE-1 inhibitors: optimizing potency and oral bioavailability. J Med Chem 51(11):3313–7, DOI 10.1021/jm800138h
- Sandgren V, Agback T, Johansson PO, Lindberg J, Kvarnström I, Samuelsson B, Belda O, Dahlgren A (2012) Highly potent macrocyclic BACE-1 inhibitors incorporating a hydroxyethylamine core: design, synthesis and X-ray crystal structures of enzyme inhibitor complexes. Bioorg Med Chem 20(14):4377–89, DOI 10.1016/j.bmc.2012.05.039
- 34. Hanessian S, Shao Z, Betschart C, Rondeau JM, Neumann U, Tintelnot-Blomley M (2010) Structure-based design and synthesis of novel P2/P3 modified, non-peptidic beta-secretase (BACE-1) inhibitors. Bioorg Med Chem Lett

20(6):1924-7, DOI 10.1016/j.bmcl.2010.01.139

- 35. Sealy JM, Truong AP, Tso L, Probst GD, Aquino J, Hom RK, Jagodzinska BM, Dressen D, Wone DWG, Brogley L, John V, Tung JS, Pleiss MA, Tucker JA, Konradi AW, Dappen MS, Toth G, Pan H, Ruslim L, Miller J, Bova MP, Sinha S, Quinn KP, Sauer JM (2009) Design and synthesis of cell potent BACE-1 inhibitors: structure-activity relationship of P1' substituents. Bioorg Med Chem Lett 19(22):6386–91, DOI 10.1016/j.bmcl.2009.09.061
- Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52(4):609–23, DOI 10.1002/prot.10465
- Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–61, DOI 10.1002/jcc.21334
- Pennington LD, Whittington DA, Bartberger MD, Jordan SR, Monenschein H, Nguyen TT, Yang BH, Xue QM, Vounatsos F, Wahl RC, Chen K, Wood S, Citron M, Patel VF, Hitchcock SA, Zhong W (2013) Hydroxyethylaminebased inhibitors of BACE1: P1–P3 macrocyclization can improve potency, selectivity, and cell activity. Bioorg Med Chem Lett 23(15):4459 – 4464, DOI https://doi.org/10.1016/j.bmcl.2013.05.028
- Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. Molecular Modeling of Proteins pp 173–209
- Wang L, Deng Y, Knight JL, Wu Y, Kim B, Sherman W, Shelley JC, Lin T, Abel R (2013) Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors. J Chem Theory Comput 9(2):1282–1293
- Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem 36(5):348–354
- 42. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys Rev E 60(3):2721–2726
- Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. Phys Rev Lett 91(14):140,601
- 44. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19–25
- 45. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins Struct Funct Bioinf 65(3):712–725
- 46. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. J Phys Chem B 113(26):9004–9015
- Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinf 78(8):1950–1958
- 48. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell Jr AD (2017) CHARMM36m: An improved force field for folded and

intrinsically disordered proteins. Nat Methods 14:71-73

- 49. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
- Yesselman JD, Price DJ, Knight JL, Brooks III CL (2012) MATCH: An atomtyping toolset for molecular mechanics force fields. J Comput Chem 33(2):189– 202
- 51. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, D MJA (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671–690
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
- 53. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J Chem Phys 98(12):10,089–10,092
- 54. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19):8577–8593
- 55. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472
- 56. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014,101
- Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52(12):7182–7190
- Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. Angew Chem Int Ed 128(26)
- 59. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK (2016) D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. J Comput Aided Mol Des 30(9):651–668, DOI 10.1007/s10822-016-9946-8
- 60. Schrödinger, LLC (2015) The PyMOL molecular graphics system, version 1.8.0.3
- 61. Hagberg A, Swart P, Schult DA (2008) Exploring network structure, dynamics, and function using NetworkX. Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States)
- Hunter JD (2007) Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9(3):90–95, DOI 10.1109/MCSE.2007.55
- Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. J Comput Aided Mol Des 26(5):595–601, DOI 10.1007/s10822-011-9526-x
- 64. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. J Comput Aided Mol Des 28(4):455–62, DOI 10.1007/s10822-014-9707-5
- Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. J Chem Inf Model 56(6):996–

1003, DOI 10.1021/acs.jcim.5b00337

- 66. Selwa E, Martiny VY, Iorga BI (2016) Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets. J Comput Aided Mol Des 30(9):829–839, DOI 10.1007/s10822-016-9983-3
- 67. Selwa E, Elisée E, Zavala A, Iorga BI (2018) Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations. J Comput Aided Mol Des 32(1):273–286, DOI 10.1007/s10822-017-0054-1
- 68. Gapsys V, de Groot BL (2017) Alchemical free energy calculations for nucleotide mutations in protein–DNA complexes. J Chem Theory Comput 13(12):6275–6289
- 69. Aldeghi M, Gapsys V, de Groot BL (2018) Accurate estimation of ligand binding affinity changes upon protein mutation. ACS Central Sci 4(12):1708–1718

Electronic Supplementary Material

Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset

Eddy Elisée,¹ Vytautas Gapsys,² Nawel Mele,¹ Ludovic Chaput,^{1,3} Edithe Selwa,¹ Bert L. de Groot^{2,*} Bogdan I. Iorga^{1,*}

¹ Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Saclay, LabEx LERMIT, 91198 Gif-sur-Yvette, France

² Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

³ Sorbonne Université, UPMC Paris 06, Institut National de la Santé et de la Recherche Médicale, Unité Mixte de Recherche S 1155, Paris, France

Corresponding Authors

* Phone: +49 551 201 2308 ; Fax: +49 551 201 2302; Email: bgroot@gwdg.de (B.L.dG.).

* Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Table of contents

Table S1. Templates used for manually building the three-dimensional BACE structures	S2
Table S2. Ligands composing the BACE training dataset	S8
Table S3. Absolute alchemical free energies for the Cathepsin S ligand set	S 8
Table S4. Relative alchemical free energies for the Cathepsin S ligand set	S10
Figure S1. Chemical structures of the 158 ligands composing the BACE dataset	S14

Ligand	Template macrocycle	Template lateral chains	Type of macrocycle and number of atoms
BACE_1	0N1	AR9	B15
BACE_2	BAV	1R6/8	A16
BACE_3	BAV	BAV	A16
BACE_4	BAV	0BI	A16
BACE_5	BAV	0XA	A16
BACE_6	0N1	BAV	B15
BACE_7	0N1	BAV	B15
BACE_8	OBI	BAV	B16
BACE_9	OBI	BAV	B16
BACE_10	OBI	BAV	B16
BACE_11	BAV	BAV	B16
BACE_12	OBI	BAV	C15
BACE_13	OBI	-	C16
BACE_14	OBI	0BI	C16
BACE_15	OBI	BAV	C16
BACE_16	OBI	BAV	C16
BACE_17	OBI	-	C16
BACE_18	OBI	BAV	B16
BACE_19	1QT	AR9	B14
BACE_20	-	AR9	-
BACE_21	BAV	BAV	A16
BACE_22	BAV	BAV	A16
BACE_23	BAV	0BI	A16
BACE_24	BAV	BAV	A16
BACE_25	BAV	BAV	A16
BACE_26	BAV	BAV	A16

Table S1. Templates used for manually building the three-dimensional BACE structures

BACE_27	BAV	0XA	A16
BACE_28	BAV	BAV	A16
BACE_29	BAV	0XA	A16
BACE_30	BAV	0XA	A16
BACE_31	BAV	0XA	A16
BACE_32	BAV	BAV	A16
BACE_33	BAV	BAV	A16
BACE_34	AR9	BAV	A15
BACE_35	BAV	BAV	A16
BACE_36	BAV	BAV	A16
BACE_37	AR9	BAV	A15
BACE_38	BAV	OBI	A16
BACE_39	BAV	0XA	A16
BACE_40	0N1	BAV	B15
BACE_41	0BI	BAV	B16
BACE_42	OBI	BAV	B16
BACE_43	OBI	BAV	B16
BACE_44	0N 1	BAV	B15
BACE_45	OBI	BAV	B16
BACE_46	OBI	BAV	B16
BACE_47	OBI	BAV	B16
BACE_48	OBI	0XA	B16
BACE_49	OBI	BAV	B16
BACE_50	0BI	BAV	B16
BACE_51	0BI	BAV	B16
BACE_52	0BI	0BI	B16
BACE_53	0BI	BAV	B16
BACE_54	0N1	BAV	B15
BACE_55	0BI	BAV	B16

BACE_56	OBI	BAV	B16
BACE_57	OBI	0XA	B16
BACE_58	OBI	0XA	B16
BACE_59	OBI	0BI	B16
BACE_60	OBI	BAV	B16
BACE_61	OBI	BAV	B16
BACE_62	OBI	BAV	B16
BACE_63	OBI	1R6/8	B16
BACE_64	OBI	0XA	B16
BACE_65	OBI	BAV	B16
BACE_66	OBI	BAV	B16
BACE_67	OBI	0XA	B16
BACE_68	OBI	0BI	B16
BACE_69	OBI	BAV	B16
BACE_70	OBI	BAV	B16
BACE_71	OBI	BAV	B16
BACE_72	OBI	BAV	B16
BACE_73	OBI	BAV	B16
BACE_74	OBI	BAV	B16
BACE_75	OBI	0BI	B16
BACE_76	OBI	BAV	B16
BACE_77	OBI	0BI	B16
BACE_78	OBI	BAV	B16
BACE_79	OBI	0XA	B16
BACE_80	OBI	BAV	B16
BACE_81	OBI	0BI	B16
BACE_82	OBI	BAV	B16
BACE_83	OBI	0BI	B16
BACE_84	OBI	0XA	B16

BACE_85	BAV	BAV	A16
BACE_86	BAV	BAV	A16
BACE_87	BAV	BAV	B16
BACE_88	BAV	BAV	C15
BACE_89	OBI	BAV	B17
BACE_90	BAV	0BI	B16
BACE_91	BAV	1R6/8	B16
BACE_92	BAV	BAV	B16
BACE_93	BAV	BAV	B16
BACE_94	OBI	BAV	B15
BACE_95	0BI	0BI	B15
BACE_96	OBI	0XA	C16
BACE_97	OBI	BAV	C15
BACE_98	0BI	0BI	C15
BACE_99	0BI	BAV	C15
BACE_100	OBI	BAV	C16
BACE_101	OBI	BAV	C16
BACE_102	OBI	0BI	C16
BACE_103	OBI	BAV	C16
BACE_104	OBI	-	C16
BACE_105	OBI	BAV	C16
BACE_106	OBI	BAV	C15
BACE_107	OBI	BAV	C15
BACE_108	OBI	0XA	C15
BACE_109	OBI	0XA	C16
BACE_110	OBI	0XA	C16
BACE_111	OBI	1R6/8	C16
BACE_112	0BI	0BI	C16
BACE_113	OBI	BAV	C15

BACE_114	OBI	BAV	C16
BACE_115	OBI	BAV	C15
BACE_116	OBI	BAV	C15
BACE_117	OBI	0XA	C16
BACE_118	OBI	BAV	C15
BACE_119	OBI	0BI	C15
BACE_120	OBI	BAV	C15
BACE_121	OBI	BAV	C15
BACE_122	OBI	BAV	C16
BACE_123	OBI	BAV	C16
BACE_124	OBI	BAV	C16
BACE_125	OBI	0XA	C16
BACE_126	OBI	-	C16
BACE_127	OBI	0XA	C16
BACE_128	OBI	1R6/8	C16
BACE_129	OBI	0XA	C16
BACE_130	OBI	BAV	C16
BACE_131	OBI	0BI	C16
BACE_132	OBI	BAV	C15
BACE_133	OBI	0BI	C15
BACE_134	OBI	BAV	C15
BACE_135	OBI	BAV	C15
BACE_136	OBI	BAV	C15
BACE_137	OBI	BAV	C15
BACE_138	OBI	BAV	C16
BACE_139	OBI	BAV	C16
BACE_140	OBI	0BI	C15
BACE_141	OBI	BAV	C16
BACE_142	OBI	-	B16

BACE_143	BAV	BAV	A16
BACE_144	OBI	0BI	C17
BACE_145	-	0BI	-
BACE_146	-	BAV	-
BACE_147	AR9	AR9	A15
BACE_148	BAV	AR9	A16
BACE_149	BAV	AR9	A16
BACE_150	BAV	AR9	A16
BACE_151	BAV	AR9	A16
BACE_152	0N1	AR9	B15
BACE_153	BAV	AR9	B16
BACE_154	OBI	AR9	C16
BACE_155	OBI	AR9	C15
BACE_156	OBI	AR9	B15
BACE_157	BAV	AR9	C15
BACE_158	BAV	AR9	C15

PDB structure	Ligand	IC50 (µM)	pIC50
3K5C	OBI	17	4.77
4DPF	0LG	400	3.40
4DPI	0N1	78	4.11
4GMI	0XA	_	_
4K8S	1QT	36	4.44
4KE1	1R6	2.5	5.60
4KE0	1R8	17	4.77
2PH8	35A	1800	2.74
3DV1	AR9	590	3.23
2F3F	AXF	190	3.72
2F3E	AXQ	156	3.81
3DV5	BAV	27	4.57
2QZK	I21	27	4.57
1XS7	MMI	_	_

Table S2. Ligands composing the BACE training dataset

Table S3. Absolute alchemical free energies for the Cathepsin S ligand set after applying cycle closure correction (kcal/mol). The GAFF and CGenFF values were obtained by averaging results from 3 independent repeats.

Ligand	Experiment	GAFF	CGenFF	Consensus
CatS_4	-8.93	-9.24±0.40	-10.03±0.07	-9.65±0.21
CatS_9	-9.41	-9.01±0.63	-10.09±0.46	-9.57±0.39
CatS_29	-9.70	-10.06±0.10	-9.90±0.07	-9.98±0.06
CatS_30	-9.37	-9.57±0.27	-10.01±0.89	-9.79±0.47
CatS_52	-8.88	-10.07±0.39	-10.32±0.20	-10.21±0.22
CatS_76	-9.17	-9.03±0.28	-9.57±0.10	-9.31±0.15
CatS_79	-9.61	-9.54±0.26	-9.54±0.26	-9.54±0.19
CatS_84	-8.90	-8.73±0.26	-9.24±0.11	-9.01±0.14
CatS_122	-10.57	-11.10±0.52	-10.58±0.84	-10.85±0.49
CatS_127	-9.78	-10.19±0.26	-9.58±0.26	-9.88±0.19

-8.25	-8.83±1.13	-8.73±0.26	-8.77±0.58
-9.02	-9.78±0.36	-9.72±0.82	-9.76±0.45
-9.11	-9.53±0.61	-9.48±0.37	-9.50±0.36
-9.29	-8.81±0.35	-9.42±0.09	-9.12±0.18
-8.53	-8.83±0.44	-9.31±0.32	-9.08±0.27
-8.55	-8.90±0.18	-9.27±0.14	-9.08±0.11
-9.06	-8.70±0.37	-9.84±0.67	-9.26±0.38
-8.97	-9.37±0.36	-9.52±0.09	-9.44±0.18
-10.02	-9.67±0.62	-10.05±0.20	-9.86±0.33
-9.41	-9.20±0.37	-9.74±0.24	-9.46±0.22
-9.82	-9.64±0.25	-10.18±0.13	-9.91±0.14
-8.50	-9.07±0.14	-8.56±0.54	-8.82±0.28
-8.28	-7.82±0.54	-8.98±0.24	-8.39±0.30
-9.45	-9.10±0.35	-9.05±0.41	-9.07±0.27
-9.33	-9.26±0.40	-9.44±0.10	-9.35±0.21
-9.71	-9.19±0.27	-9.62±0.12	-9.41±0.15
-10.09	-9.60±0.19	-9.54±0.15	-9.57±0.12
-9.45	-9.19±0.51	-9.41±0.16	-9.30±0.27
-10.09	-9.87±0.10	-9.58±0.18	-9.72±0.10
-10.33	-9.73±0.34	-9.85±0.21	-9.79±0.20
-10.57	-10.71±0.16	-10.42±0.43	-10.56±0.23
-9.26	-9.36±0.10	-9.69±0.19	-9.51±0.11
-9.91	-9.68±0.71	-10.03±0.50	-9.86±0.43
-8.57	-9.16±0.26	-9.46±0.06	-9.31±0.13
-9.87	-9.94±0.21	-9.78±0.85	-9.85±0.44
-9.55	-9.27±0.22	-9.58±0.09	-9.43±0.12
-8.36	-9.22±0.20	-9.52±0.07	-9.38±0.11
-8.63	-9.60±0.06	-9.86±0.13	-9.74±0.07
-8.30	-8.84±0.12	-9.58±0.22	-9.21±0.13
	-8.25 -9.02 -9.11 -9.29 -8.53 -8.55 -9.06 -8.97 -10.02 -9.41 -9.82 -8.50 -8.28 -9.45 -9.33 -9.71 -10.09 -9.45 -10.09 -9.45 -10.09 -9.45 -10.09 -9.45 -10.09 -9.45 -10.09 -9.45 -10.57 -9.26 -9.91 -8.57 -9.87 -9.87 -9.55 -8.36 -8.63 -8.30	-8.25 -8.83 ± 1.13 -9.02 -9.78 ± 0.36 -9.11 -9.53 ± 0.61 -9.29 -8.81 ± 0.35 -8.53 -8.83 ± 0.44 -8.55 -8.90 ± 0.18 -9.06 -8.70 ± 0.37 -8.97 -9.37 ± 0.36 -10.02 -9.67 ± 0.62 -9.41 -9.20 ± 0.37 -9.82 -9.64 ± 0.25 -8.50 -9.07 ± 0.14 -8.28 -7.82 ± 0.54 -9.45 -9.10 ± 0.35 -9.33 -9.26 ± 0.40 -9.71 -9.19 ± 0.27 -10.09 -9.60 ± 0.19 -9.45 -9.19 ± 0.51 -10.09 -9.87 ± 0.10 -10.33 -9.73 ± 0.34 -10.57 -10.71 ± 0.16 -9.26 -9.36 ± 0.71 -8.57 -9.16 ± 0.26 -9.87 -9.94 ± 0.21 -9.55 -9.27 ± 0.22 -8.36 -9.22 ± 0.20 -8.63 -9.60 ± 0.06 -8.30 -8.84 ± 0.12	-8.25 -8.83 ± 1.13 -8.73 ± 0.26 -9.02 -9.78 ± 0.36 -9.72 ± 0.82 -9.11 -9.53 ± 0.61 -9.48 ± 0.37 -9.29 -8.81 ± 0.35 -9.42 ± 0.09 -8.53 -8.83 ± 0.44 -9.31 ± 0.32 -8.55 -8.90 ± 0.18 -9.27 ± 0.14 -9.06 -8.70 ± 0.37 -9.84 ± 0.67 -8.97 -9.37 ± 0.36 -9.52 ± 0.09 -10.02 -9.67 ± 0.62 -10.05 ± 0.20 -9.41 -9.20 ± 0.37 -9.74 ± 0.24 -9.82 -9.64 ± 0.25 -10.18 ± 0.13 -8.50 -9.07 ± 0.14 -8.56 ± 0.54 -8.28 -7.82 ± 0.54 -8.98 ± 0.24 -9.45 -9.10 ± 0.35 -9.05 ± 0.41 -9.33 -9.26 ± 0.40 -9.44 ± 0.10 -9.71 -9.19 ± 0.27 -9.62 ± 0.12 -10.09 -9.60 ± 0.19 -9.54 ± 0.15 -9.45 -9.19 ± 0.51 -9.41 ± 0.16 -10.33 -9.73 ± 0.34 -9.85 ± 0.21 -10.57 -10.71 ± 0.16 -10.42 ± 0.43 -9.26 -9.36 ± 0.10 -9.69 ± 0.19 -9.91 -9.68 ± 0.71 -10.03 ± 0.50 -8.57 -9.16 ± 0.26 -9.46 ± 0.06 -9.87 -9.94 ± 0.21 -9.78 ± 0.85 -9.55 -9.27 ± 0.22 -9.58 ± 0.09 -8.36 -9.22 ± 0.20 -9.52 ± 0.07 -8.63 -9.60 ± 0.06 -9.86 ± 0.13 -8.30 -8.84 ± 0.12 -9.58 ± 0.22

Edge GAFF CGenFF Consensus CatS_79-CatS_127 -0.43±0.37 0.05 ± 0.37 -0.19±0.26 CatS_79-CatS_130 0.90 ± 1.11 1.35 ± 0.21 1.12 ± 0.57 CatS_79-CatS_131 -0.08 ± 0.36 -0.35 ± 0.82 -0.22 ± 0.45 CatS_79-CatS_132 -0.25 ± 0.61 0.28 ± 0.37 0.01 ± 0.36 0.45 ± 0.16 0.44 ± 0.09 CatS_79-CatS_139 0.43 ± 0.08 0.00 ± 0.30 0.32 ± 0.23 CatS_79-CatS_141 0.64 ± 0.36 CatS_79-CatS_142 0.72 ± 0.13 0.11±0.41 0.41 ± 0.22 CatS_79-CatS_144 0.43 ± 0.37 -0.20 ± 0.67 0.12 ± 0.38 CatS_79-CatS_153 -0.15 ± 0.34 0.13 ± 0.18 -0.01±0.19 CatS_79-CatS_155 0.24 ± 0.51 -0.67±0.17 -0.21±0.27 CatS_79-CatS_4 0.22 ± 0.43 -0.48 ± 0.69 -0.13±0.41 CatS_79-CatS_157 0.14 ± 0.35 -0.31±0.14 -0.09±0.19 CatS_79-CatS_160 0.02 ± 0.34 -0.94 ± 0.27 -0.46 ± 0.22 CatS_79-CatS_162 0.47 ± 0.14 1.78 ± 0.54 1.13 ± 0.28 -0.35 ± 0.24 CatS_79-CatS_165 0.67 ± 0.30 1.70 ± 0.54 0.38±0.35 CatS_79-CatS_167 0.14 ± 0.41 0.26 ± 0.27 CatS_79-CatS_172 0.55 ± 0.43 -0.08 ± 0.08 0.24 ± 0.22 CatS_79-CatS_174 0.10 ± 0.27 0.12 ± 0.12 0.11±0.15 CatS_79-CatS_176 -0.18±0.19 0.15±0.15 -0.02 ± 0.12 CatS 79-CatS 177 0.44 ± 0.17 -0.06 ± 0.11 0.19 ± 0.10 CatS_79-CatS_178 -0.17±0.54 -0.40 ± 0.22 -0.28 ± 0.29 CatS_79-CatS_9 0.61 ± 0.30 -0.72±0.41 -0.05±0.25 CatS_79-CatS_180 -0.25 ± 0.24 -0.47 ± 0.17 -0.36±0.15 CatS_79-CatS_181 -1.29 ± 0.16 -0.89 ± 0.43 -1.09 ± 0.23

Table S4. Relative alchemical free energies for the Cathepsin S ligand set prior to applying cycle closure correction (kcal/mol). The GAFF and CGenFF values were obtained by averaging results from 3 independent repeats.

CatS_79-CatS_184	0.39±0.12	-0.25±0.18	0.07±0.11
CatS_79-CatS_185	0.52±0.62	-0.44±0.49	0.04±0.39
CatS_79-CatS_186	0.37±0.10	0.04±0.33	0.21±0.17
CatS_79-CatS_191	-0.59±0.18	0.63±0.83	0.02±0.43
CatS_79-CatS_195	0.36±0.34	0.06±0.06	0.21±0.17
CatS_79-CatS_252	0.17±0.20	0.10±0.07	0.13±0.11
CatS_79-CatS_253	-0.26±0.62	-0.27±0.25	-0.26±0.33
CatS_79-CatS_255	1.13±0.11	-0.15±0.19	0.49±0.11
CatS_79-CatS_29	-0.12±0.43	-0.75±0.69	-0.43±0.41
CatS_79-CatS_30	-0.52±0.25	-0.52±0.89	-0.52±0.46
CatS_79-CatS_52	-0.56±0.50	-0.11±0.31	-0.34±0.30
CatS_79-CatS_76	0.02±0.18	-0.17±0.04	-0.07±0.09
CatS_79-CatS_84	1.18±0.18	0.08±0.27	0.63±0.16
CatS_79-CatS_122	-1.45±0.52	-0.85±0.84	-1.15±0.49
CatS_127-CatS_142	1.60±0.29	0.36±0.22	0.98±0.18
CatS_130-CatS_142	-0.12±0.25	-0.33±0.20	-0.23±0.16
CatS_131-CatS_252	0.65±0.77	0.36±0.22	0.50 ± 0.40
CatS_132-CatS_142	0.32±0.30	0.42±0.27	0.37±0.20
CatS_139-CatS_76	-0.11±0.12	-0.28±0.08	-0.20±0.07
CatS_141-CatS_252	-0.84±0.23	-0.38±0.48	-0.61±0.27
CatS_142-CatS_139	0.16±0.34	-0.40±0.32	-0.12±0.24
CatS_144-CatS_142	-0.19±0.30	0.58±0.50	0.20±0.29
CatS_155-CatS_30	0.40±0.27	-0.19±0.69	0.11±0.37
CatS_4-CatS_9	0.45±0.57	0.13±0.10	0.29±0.29
CatS_157-CatS_153	-0.04±0.25	0.22±0.10	0.09±0.14
CatS_160-CatS_180	-0.23±0.35	0.30±0.19	0.03±0.20
CatS_162-CatS_167	-0.29±0.56	-0.29±0.28	-0.29±0.32
CatS_165-CatS_142	-1.10±0.22	-1.16±0.19	-1.13±0.15
CatS_167-CatS_252	-0.44±0.28	-0.24±0.09	-0.34±0.15

CatS_174-CatS_177	-0.26±0.26	0.16±0.20	-0.05±0.16
CatS_176-CatS_180	-0.23±0.09	-0.15±0.17	-0.19±0.09
CatS_177-CatS_172	0.13±0.57	-0.09±0.14	0.02±0.29
CatS_178-CatS_191	0.04±0.15	-0.04±0.25	0.00±0.15
CatS_9-CatS_52	-0.89±0.55	-0.68±0.29	-0.78±0.31
CatS_180-CatS_172	0.16±0.31	0.42±0.16	0.29±0.18
CatS_181-CatS_167	1.84±0.33	1.25±0.26	1.55±0.21
CatS_184-CatS_186	0.19±0.33	0.21±0.34	0.20±0.24
CatS_185-CatS_139	1.68±0.50	0.58±0.13	1.13±0.26
CatS_191-CatS_30	0.20±0.31	0.44±0.58	0.32±0.33
CatS_195-CatS_76	0.49±0.31	0.28±0.12	0.39±0.17
CatS_252-CatS_177	0.70±0.30	-0.04±0.14	0.33±0.17
CatS_253-CatS_255	0.52±0.08	0.37±0.18	0.45±0.10
CatS_255-CatS_252	-0.55±0.47	0.03±0.14	-0.26±0.24
CatS_29-CatS_30	0.63±0.15	0.22±0.09	0.43±0.09
CatS_30-CatS_178	-0.65±0.17	1.26±0.31	0.31±0.18
CatS_52-CatS_4	0.73±0.07	0.59±0.28	0.66±0.14
CatS_84-CatS_139	-0.14±0.25	-0.38±0.20	-0.26±0.16
CatS_122-CatS_139	2.49±0.57	1.27±0.67	1.88±0.44
CatS_127-CatS_139	1.28±0.13	0.21±0.47	0.74±0.24
CatS_130-CatS_139	0.24±0.77	-0.43±0.51	-0.09±0.46
CatS_131-CatS_139	1.02±0.41	-0.02±0.30	0.50±0.25
CatS_132-CatS_139	0.75±0.39	0.05±0.38	0.40±0.27
CatS_139-CatS_186	-0.23±0.63	0.03±0.17	-0.10±0.33
CatS_141-CatS_84	0.46±0.37	0.01±0.16	0.24±0.20
CatS_142-CatS_84	0.13±0.05	-0.27±0.29	-0.07±0.15
CatS_144-CatS_139	-0.53±0.07	0.50±0.26	-0.02±0.13
CatS_153-CatS_76	-0.24±0.13	-0.12±0.39	-0.18±0.21
CatS_155-CatS_153	0.36±0.50	0.53±0.13	0.45±0.26

CatS_4-CatS_255	-0.00±0.40	0.57±0.85	0.28±0.47
CatS_157-CatS_184	-0.11±0.14	0.15±0.27	0.02±0.15
CatS_160-CatS_157	0.69±0.34	0.17±0.19	0.43±0.19
CatS_162-CatS_252	-0.43±0.33	0.17±0.81	-0.13±0.44
CatS_165-CatS_130	-1.03±0.77	0.20±0.23	-0.42±0.40
CatS_167-CatS_177	-0.15±0.40	-0.37±0.15	-0.26±0.22
CatS_172-CatS_186	0.27±0.14	0.03±0.10	0.15±0.09
CatS_174-CatS_172	-0.08±0.20	0.44±0.14	0.18±0.12
CatS_176-CatS_177	0.39±0.20	0.12±0.20	0.25±0.14
CatS_177-CatS_184	-0.37±0.17	-0.17±0.10	-0.27±0.10
CatS_178-CatS_155	-0.13±0.23	-0.13±0.12	-0.13±0.13
CatS_9-CatS_157	-0.05±0.28	0.85±0.55	0.40±0.31
CatS_180-CatS_177	0.55±0.17	0.40±0.13	0.47±0.11
CatS_181-CatS_252	0.59±0.31	1.54 ± 0.73	1.06 ± 0.40
CatS_185-CatS_252	0.30±0.14	0.59±0.22	0.44±0.13
CatS_186-CatS_177	0.12±0.37	0.08 ± 0.08	0.10±0.19
CatS_191-CatS_155	0.33±0.19	0.08±0.13	0.21±0.12
CatS_195-CatS_184	-0.15±0.27	-0.22±0.23	-0.18±0.18
CatS_252-CatS_139	0.09±0.09	-0.04±0.26	0.02 ± 0.14
CatS_253-CatS_252	0.43±0.15	0.30±0.43	0.36±0.23
CatS_255-CatS_139	-0.05±0.60	0.27±0.23	0.11±0.32
CatS_29-CatS_155	0.62±0.22	-0.89±0.89	-0.13±0.46
CatS_30-CatS_76	0.28±0.52	0.18±0.55	0.23±0.38
CatS_52-CatS_167	1.26±0.95	1.21±0.48	1.24±0.53
CatS_76-CatS_84	0.17±0.13	0.35±0.22	0.26±0.13
CatS_84-CatS_252	0.07±0.23	-0.63±0.16	-0.28±0.14
CatS_122-CatS_252	1.78±0.58	1.15±0.62	1.47 ± 0.42







BACE_5





BACE_9













BACE_19

BACE_20









BACE_23

BACE_24





BACE_26





BACE_27

BACE_28







BACE_31







BACE_34





BACE_35

BACE_36



BACE_37















НО

HN

ΗN

0

CI

BACE_54



BACE_56



НΟ ΗN 0 0 HN С ΝH

















0 s=0

CI









0

ÓН

JН

°-й

ΗŃ

HO

BACE_108

CI



BACE_107













BACE_114

ΗŇ

BACE_112



HN

ΗŅ

0

Ĥ

ОН

O

▲ ОН

О

Ĥ



BACE_118

BACE_119







BACE_123



BACE_125







OH





BACE_133



HN HN O HO N BACE_137



BACE_139



HN HN O HN HN O BACE_134

,





BACE_138









BACE_145







н

ΗŃ

Ē

BACE_150

) (



Figure S1. Chemical structures of the 158 ligands composing the BACE dataset. The pose prediction subset is composed of 20 compounds, **BACE_1** to **BACE_20**. The ranking subset is composed of 154 compounds, i.e. all with the exception of **BACE_2**, **BACE_3**, **BACE_17** and **BACE_18**.

Conclusions

99 I am turned into a sort of machine for observing facts and grinding out conclusions.

— Charles Darwin (Naturalist)
— **Conclusions** — In this thesis, the main objective was to perform *in silico* predictions of β -lactamase mutations using a pmx-based free energy protocol. This way, methodological developments have been carried out in order to (1) study SBLs (considering the covalent bond in protein-ligand complexes) and MBLs (considering the zinc ions) by MD simulations, and (2) compute free energies in a reliable and robust manner. Summary of the achieved works is presented below, chapter per chapter.

In Chapter II, new zinc-coordinating residues have been parameterized to study zinc-containing metalloenzymes in OPLS-AA force field. New parameters were successfully assessed and validated by MD simulations of fourteen representative metalloenzymes. This work also revealed a misrepresentation of zinc-zinc distances in di-zinc systems, probably due to a repulsion phenomenon that may occurs when two positive charges are close to each other, here every zinc ion display a +2 charge. This issue was corrected by implementing a non-covalent bond between the two zinc ions based on the crystallographic distance.

In Chapter III, several MD studies were performed in order to keep verifying our methodological developments: (a) implementation of covalent bond was validated on SBLs-ligand systems, (b) MBLs-ligand systems were analyzed using initial ligand poses obtained by superimposition of related crystallographic protein structures, resulting in overall stable systems and pinpointing ligands as stabilizers of zinc-zinc interaction without applying a non-covalent bond, (c) zinc-coordinating histidine mutations into a glycine residue led to the expected structural rearrangement of the active site by displacement of an adjacent aspartate residue, but uncertainties on this observation arose due to the instability of the zinc-zinc distance, issue that was fixed later, and (d) encouraging results were obtained for the protonation reaction of a lysine residue in the active site of SBLs by using our pmx-based protocol – nevertheless, further investigations needs to be undertaken to improve the reliability of simulations for the carboxylation reaction.

In Chapter IV, several participations in D3R-GC competitions provided us a certain overview of our computational workflow's strengths and weaknesses: we significantly improved our docking protocol by using available structural information and several docking softwares, while we faced the limitations of our free energy protocol which failed in computing trustworthy results for docking poses since they were often not correctly overlapped therefore reducing calculations' accuracy. At the end of this PhD journey, if I were to draw conclusions about my experience in Molecular Modeling field, I would retain the following nonexhaustive tips:

– Take the time to properly set up your training system by checking all parameters without blindly trust existing scripts: prepare your protein or ligand with a correct parameterization in the selected force field, think about which element could be important for the experiment (waters, cofactors, ions...), run fairly long simulations to identify potential instabilities.

– Perform replicas to assess, inter alia, the reproducibility of an observed event or the reliability of computed values such as free energies.

- Be aware of new released publications in your specific field but also in related ones to maybe catch up on new ideas.

– Don't hesitate to ask your working colleagues, boss or anyone on a suitable forum if you face a problem, they could have a solution to fix it !

— Perspectives — Given the different improvements designed within the allocated time of this thesis, everything appears to be ready for first in silico predictions of β -lactamase mutations in the presence or absence of ligand. The study would be focused on one representative enzyme of each class: KPC-2 (A), NDM-1 (B), CMY-2 (C) and OXA-48 (D). Figures 34 and 35 report all found single mutations connecting members of a same family; indeed, two simultaneous mutations could be equally carried out but first milestone is to validate our pmx-based protocol on single mutations. Nevertheless, we might encounter some inconsistencies with NDM-1 enzyme since it is difficult to correctly reproduce the genuine electrostatic and polarisable zinc ions' environment in normal MD simulations. In practice, we may compare computed free energies to experimental ones in the case of ligand-free proteins or to binding affinities in the case of protein-ligand complexes. The aim is to find some mutations which could enhance the antibiotic resistance by either promoting their hydrolysis or preventing inhibitors' action, therefore susceptible to arise in the future. This way, new inhibitors or antibiotics could be designed by taking into account this information in order to negate the effect of such mutants.

Moreover, regarding histidine-to-glycine mutants in SPM-like proteins, additional MD simulations must be run with the developed non-covalent bond between the two active site zinc ions, in order to evaluate once more our assumption about the potential replacement of mutated zinc-coordinating histidine by the neighbouring aspartate. Finally, the protocol involving a loss rather than a gain of atoms will be investigated to achieve the carboxylation process in OXA-48 and its mutants, OXA-145 and OXA-427.



Figure 34: Single mutations in KPC, NDM-1 and OXA families.



Figure 35: Single mutations in CMY family.

References

99 Redde Caesari quae sunt Caesaris, et quae sunt Dei Deo. Literally, render unto Caesar the things that are Caesar's, and unto God the things that are God's.

— Gospel According to Marc, XII, 13-17 (New Testament, the Bible)

- [1] M. I. Abboud, M. Kosmopoulou, A. P. Krismanich, J. W. Johnson, P. Hinchliffe, J. Brem, T. D. W. Claridge, J. Spencer, C. J. Schofield, and G. I. Dmitrienko. Cyclobutanone Mimics of Intermediates in Metallo-β-Lactamase Catalysis. *Chem. Eur. J.*, 24(22):5734–5737, 2018.
- [2] E. P. Abraham and E. Chain. An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, 146:837, 1940.
- [3] O. Acevedo and W. L. Jorgensen. Advances in quantum and molecular mechanical (qm/mm) simulations for organic and enzymatic reactions. *Acc. Chem. Res.*, 43(1):142–151, 2010.
- [4] M. J. Adams, T. L. Blundell, E. J. Dodson, G. G. Dodson, M. Vijayan, E. N. Baker, M. M. Harding, D. C. Hodgkin, B. Rimmer, and S. Sheat. Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature*, 224:491– 495, 1969.
- [5] M. Aldeghi, B. L. de Groot, and V. Gapsys. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation. In Sikosek T. (eds) Computational Methods in Protein Evolution. Methods in Molecular Biology, volume 1851, pages 19–47. Humana Press, New York, NY, 2019.
- [6] S. A. Alharbi, M. Wainwright, T. A. Alahmadi, H. B. Salleeh, A. A. Faden, and A. Chinnathambi. What if Fleming had not discovered penicillin? *Saudi J. Biol. Sci.*, 21(4):289–293, 2014.
- [7] H. K. Allen, J. Donato, H. H. Wang, K. A. Cloud-Hansen, J. Davies, and J. Handelsman. Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.*, 8(4):251–259, 2010.
- [8] N. L. Allinger. Calculation of molecular structure and energy by forcefield methods. *Adv. Phys. Org. Chem.*, 13:1–82, 1976.
- [9] R. P. Ambler. The structure of β -lactamases. *Phil. Trans. R. Soc. Lond. B*, 289:321–331, 1980.
- [10] R. I. Aminov. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front. Microbiol.*, 1(134):1–7, 2010.
- [11] A. Anantharajah, B. Tossens, N. Olive, B. Kabamba-Mukadi, H. Rodriguez-Villalobos, and A. Verroken. Performance evaluation of the mbt star®-carba ivd assay for the detection of carbapenemases with maldi-tof ms. *Front. Microbiol.*, 10:1413, 2019.
- [12] G. Arango-Argoty, E. Garner, A. Pruden, L. S. Heath, P. Vikesland, and L. Zhang. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6(23):1–15, 2018.

- [13] I. Armenia, G. L. Marcone, F. Berini, V. T. Orlandi, C. Pirrone, E. Martegani, R. Gornati, G. Bernardini, and F. Marinelli. Magnetic Nanoconjugated Teicoplanin: A Novel Tool for Bacterial Infection Site Targeting. *Front. Microbiol.*, 9(2270):1–17, 2018.
- [14] R. D. Ballantine, C. E. McCallion, E. Nassour, S. Tokajian, and S. A. Cochrane. Tridecaptin-inspired antimicrobial peptides with activity against multidrug-resistant Gram-negative bacteria. *Med. Chem. Comm.*, 10(3):484–487, 2019.
- [15] F. Barancheshme and M. Munir. Strategies to Combat Antibiotic Resistance in the Wastewater Treatment Plants. *Front. Microbiol.*, 8(2603):1–12, 2018.
- [16] W. T. Barker, C. E. Chandler, R. J. Melander, R. K. Ernst, and C. Melander. Tryptamine derivatives disarm colistin resistance in polymyxinresistant gram-negative bacteria. *Bioorg. Med. Chem.*, 27(9):1776– 1788, 2019.
- [17] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, and R. A. Friesner. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.*, 119(25):5908–5920, 1997.
- [18] C. Bebrone. Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem. Pharmacol.*, 74(12):1686–1701, 2007.
- [19] J. W. Bennett and K.-T. Chung. Alexander Fleming and the discovery of penicillin. *Adv. Appl. Microbiol.*, 49:163–184, 2001.
- [20] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.
- [21] F. Berglund, N. P. Marathe, T. Österlund, J. Bengtsson-Palme, S. Kotsakis, C.-F. Flach, D. G. J. Larsson, and E. Kristiansson. Identification of 76 novel B1 metallo- β -lactamases through large-scale screening of genomic and metagenomic data. *Microbiome*, 5(1):134, 2017.
- [22] R. A. Bonomo. β-Lactamases: A Focus on Current Challenges. Cold Spring Harb. Perspect. Med., 7(1):1–15, 2017.
- [23] E. Boudaher and C. L. Shaffer. Inhibiting bacterial secretion systems in the fight against antibiotic resistance. *Med. Chem. Comm.*, 10(5):682– 692, 2019.
- [24] C. Brink, D. Crowfoot, J. Lindsey, J. Pickworth, J. H. Robertson, and J. G. White. X-ray Crystallographic Evidence on the Structure of Vitamin B12. *Nature*, 174(4443):1169–1171, 1954.

- [25] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.
- [26] G. Brotzu. Ricerche su di un nuovo antibiotico. *Lavori dell'Istituto d'Igiene di Cagliari*, 1948.
- [27] A. G. Brown, D. Butterworth, M. Cole, G. Hanscomb, J. D. Hood,
 C. Reading, and G. N. Rolinson. Naturally-occurring β-lactamase inhibitors with antibacterial activity. *J. Antibiot.*, 29(6):668–669, 1976.
- [28] E. E. F. Brown, A. Cooper, C. Carrillo, and B. Blais. Selection of Multidrug-Resistant Bacteria in Medicated Animal Feeds. *Front. Microbiol.*, 10(456):1–8, 2019.
- [29] K. Bush. Evaluation of enzyme inhibition data in screening for new drugs. *Drugs Exp. Clin. Res.*, 12(6–7):565–576, 1986.
- [30] K. Bush. β-lactamase Inhibitors from Laboratory to Clinic. *Clin. Microbiol. Rev.*, 1(1):109–123, 1988.
- [31] K. Bush. The coming of age of antibiotics: discovery and therapeutic value. *Ann. N.Y. Acad. Sci.*, 1213(1):1–4, 2010.
- [32] K. Bush. Past and Present Perspectives on β-Lactamases. Antimicrob. Agents Chemother., 62(10):1–20, 2018.
- [33] K. Bush and G. A. Jacoby. Updated Functional Classification of -Lactamases. *Antimicrob. Agents Chemother.*, 54(3):969–976, 2010.
- [34] K. Bush, G. A. Jacoby, and A. A. Medeiros. A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob. Agents Chemother.*, 39(6):1211–1233, 1995.
- [35] K. Bush, C. Macalintal, B. A. Rasmussen, V. J. Lee, and Y. Yang. Kinetic interactions of tazobactam with beta-lactamases from all major structural classes. *Antimicrob. Agents Chemother.*, 37(4):851–858, 1993.
- [36] E. Cabezón, F. de la Cruz, and I. Arechaga. Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria. *Front. Microbiol.*, 8(2329):1–7, 2017.
- [37] S. T. Cahill, J. M. Tyrrell, I. H. Navratilova, K. Calvopiña, S. W. Robinson, C. T. Lohans, M. A. McDonough, R. Cain, C. W. Fishwick, M. B. Avison, T. R. Walsh, C. J. Schofield, and J. Brem. Studies on the inhibition of AmpC and other β-lactamases by cyclic boronates. *Biochim. Biophys. Acta, Gen. Subj.*, 1863(4):742–748, 2019.

- [38] C. H. Carlisle, D. Crowfoot, and J. D. Bernal. The crystal structure of cholesteryl iodide. *Proc. R. Soc. Lond. A*, 184(996):64–83, 1945.
- [39] E. Chain. The chemical structure of the penicillins. *Nobel Lectures, Physiology or Medicine 1942–1962*, pages 110–143, 1964.
- [40] J. Charon, A. Manteca, and C. A. Innis. Using the Bacterial Ribosome as a Discovery Platform for Peptide-Based Antibiotics. *Biochemistry*, 58(2):75–84, 2019.
- [41] M. F. Chellat, L. Raguž, and R. Riedl. Targeting Antibiotic Resistance. Angew. Chem. Int. Ed., 55(23):6600–6626, 2016.
- [42] A. V. Cheng and W. M. Wuest. Signed, Sealed, Delivered: Conjugate and Prodrug Strategies as Targeted Delivery Vectors for Antibiotics. *ACS Infect. Dis.*, 5(6):816–828, 2019.
- [43] Z. Cheng, J. VanPelt, A. Bergstrom, C. Bethel, A. Katko, C. Miller, K. Mason, E. Cumming, H. Zhang, R. L. Kimble, S. Fullington, S. L. Bretz, J. C. Nix, R. A. Bonomo, D. L. Tierney, R. C. Page, and M. W. Crowder. A noncanonical metal center drives the activity of the sedi-minispirochaeta smaragdinae metallo-β-lactamase sps-1. *Biochemistry*, 57(35):5218–5229, 2018. PMID: 30106565.
- [44] P. T. Cherian, M. N. Cheramie, R. K. Marreddy, D. M. Fernando, J. G. Hurdle, and R. E. Lee. New β -lactam Tetramic acid hybrids show promising antibacterial activities. *Bioorg. Med. Chem. Lett.*, 28(18):3105–3112, 2018.
- [45] R. A. Copeland. Evaluation of enzyme inhibitors in drug discovery. a guide for medicinal chemists and pharmacologists. *Methods Biochem. Anal.*, 46:1–265, 2005.
- [46] M. Cordovana, M. Kostrzewa, J. Glandorf, M. Bienia, S. Ambretti, and A. B. Pranada. A full maldi-based approach to detect plasmid-encoded kpc-producing klebsiella pneumoniae. *Front. Microbiol.*, 9:2854, 2018.
- [47] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.
- [48] G. Cox and G. D. Wright. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *Int. J. Med. Microbiol.*, 303(6-7):287– 292, 2013.
- [49] M. W. Crowder, J. Spencer, and A. J. Vila. Metallo-β-lactamases: Novel Weaponry for Antibiotic Resistance in Bacteria. Acc. Chem. Res., 39(10):721–728, 2006.

- [50] D. Crowfoot, J. Pickworth, J. H. Robertson, K. N. Trueblood, R. J. Prosen, and J. G. White. The Crystal Structure of the Hexacarboxylic Acid derived from B12 and the Molecular Structure of the Vitamin. *Nature*, 176(4477):325–328, 1955.
- [51] D. M. Crowfoot, C. W. Bunn, B. W. Rogers-Low, and A. Turner-Jones. The x-ray crystallographic investigation of the structure of penicillin. *Princeton University Press, Chapter XI*, pages 310–66, 1949.
- [52] M. Cycoń, A. Mrozik, and Z. Piotrowska-Seget. Antibiotics in the Soil Environment—Degradation and Their Impact on Microbial Activity and Diversity. *Front. Microbiol.*, 10(338):1–45, 2019.
- [53] I. Dandachi, S. Chabou, Z. Daoud, and J.-M. Rolain. Prevalence and Emergence of Extended-Spectrum Cephalosporin-, Carbapenem- and Colistin-Resistant Gram Negative Bacteria of Animal Origin in the Mediterranean Basin. *Front. Microbiol.*, 9(2299):1–26, 2018.
- [54] E. de la Vega, N. A. O'Leary, J. E. Shockey, J. Robalino, C. Payne, C. L. Browdy, G. W. Warr, and P. S. Gross. Anti-lipopolysaccharide factor in Litopenaeus vannamei (LvALF): A broad spectrum antimicrobial peptide essential for shrimp immunity against bacterial and fungal infection. *Mol. Immunol.*, 45(7):1916–1925, apr 2008.
- [55] G. Diamond, N. Beckloff, A. Weinberg, and K. O. Kisich. The roles of antimicrobial peptides in innate host defense. *Curr. Pharm. Des.*, 15(21):2377–92, 2009.
- [56] J. D. Docquier and S. Mangani. An update on β -lactamase inhibitor discovery and development. *Drug Resist. Updat.*, 36:13–29, 2018.
- [57] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.*, 45(W1):W331–W336, jul 2017.
- [58] L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen. 1.14xCM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. J. Phys. Chem. B, 121(15):3864– 3870, 2017.
- [59] J. Domański, O. Beckstein, and B. I. Iorga. Ligandbook: an online repository for small and drug-like molecule force field parameters. *Bioinformatics*, 33(11):1747–1749, 2017.
- [60] E. Duchesne. Contribution à l'étude de la concurrence vitale chez les micro-organismes : antagonisme entre les moisissures et les microbes. Alexandre Rey, 1897.

- [61] K. Durka, A. E. Laudy, Ł. Charzewski, M. Urban, K. Stępień, S. Tyski, K. A. Krzyśko, and S. Luliński. Antimicrobial and KPC/AmpC inhibitory activity of functionalized benzosiloxaboroles. *Eur. J. Med. Chem.*, 171:11–24, 2019.
- [62] A. J. F. Egan. Bacterial outer membrane constriction. *Mol. Microbiol.*, 107(6):676–687, 2018.
- [63] P. Ehrlich and S. Hata. *Die experimentelle Chemotherapie der Spirillosen*. Julius Springer, 1910.
- [64] C. M. Ernst and A. Peschel. Broad-spectrum antimicrobial peptide resistance by MprF-mediated aminoacylation and flipping of phospholipids. *Mol. Microbiol.*, 80(2):290–299, 2011.
- [65] R. L. Fahrner, T. Dieckmann, S. S. Harwig, R. I. Lehrer, D. Eisenberg, and J. Feigon. Solution structure of protegrin-1, a broad-spectrum antimicrobial peptide from porcine leukocytes. *Chem. Biol.*, 3(7):543– 550, 1996.
- [66] R. J. Fair and Y. Tor. Antibiotics and Bacterial Resistance in the 21st Century. *Perspect. Med. Chem.*, 6:25–64, 2014.
- [67] F. Fanelli, A. Di Pinto, A. Mottola, G. Mule, D. Chieffi, F. Baruzzi, G. Tantillo, and V. Fusco. Genomic Characterization of Arcobacter butzleri Isolated From Shellfish: Novel Insight Into Antibiotic Resistance and Virulence Determinants. *Front. Microbiol.*, 10(670):1–17, 2019.
- [68] A. Fleming. On the antibacterial action of cultures of a penicillium with special reference to their use in the isolation of B. influenzae. *Br. J. Exp. Pathol.*, 10(3):226–236, 1929.
- [69] E. L. Fonseca, B. G. Andrade, and A. C. Vicente. The resistome of low-impacted marine environments is composed by distant Metallo-βlactamases homologs. *Front. Microbiol.*, 9(677):1–7, 2018.
- [70] J. W. Foster and H. B. Woodruff. Antibiotic substances produced by bacteria. Ann. N.Y. Acad. Sci., 48(2):87–98, 1946.
- [71] D. Y. Freitas, S. Araújo, A. R. C. Folador, R. T. J. Ramos, J. S. N. Azevedo, M. Tacão, A. Silva, I. Henriques, and R. A. Baraúna. Extended Spectrum Beta-Lactamase-Producing Gram-Negative Bacteria Recovered From an Amazonian Lake Near the City of Belém, Brazil. *Front. Microbiol.*, 10(364):1–13, 2019.
- [72] M. Galleni and J.-M. Frère. Kinetics of β-lactamases and penicillinbinding proteins. In *Enzyme-Mediated Resistance to Antibiotics*, pages 195–213. American Society of Microbiology, 2007.

- [73] M. Galleni, J. Lamotte-brasseur, and G. M. Rossolini. Standard Numbering Scheme for Class B β-Lactamases. Antimicrob. Agents Chemother., 45(3):660–663, 2001.
- [74] V. Gapsys and B. L. De Groot. pmx Webserver: A User Friendly Interface for Alchemistry. J. Chem. Inf. Model., 57(2):109–114, 2017.
- [75] V. Gapsys, S. Michielssens, J. H. Peters, B. L. D. Groot, and H. Leonov. Molecular Modeling of Proteins, volume 1215 of Methods in Molecular Biology. Springer New York, New York, NY, 2015.
- [76] V. Gapsys, S. Michielssens, D. Seeliger, and B. L. de Groot. pmx: Automated protein structure and topology generation for alchemical perturbations. J. Comput. Chem., 36(5):348–354, feb 2015.
- [77] V. Gapsys, S. Michielssens, D. Seeliger, and B. L. de Groot. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chem. Int. Ed.*, 55(26):7364–7368, jun 2016.
- [78] G. Garau, C. Bebrone, C. Anne, M. Galleni, J.-M. Frère, and O. Dideberg. A Metallo-β-lactamase Enzyme in Action: Crystal Structures of the Monozinc Carbapenemase CphA and its Complex with Biapenem. *J. Mol. Biol.*, 345(4):785–795, 2005.
- [79] E. Gay, M. Bour, G. Cazeau, N. Jarrige, C. Martineau, J.-Y. Madec, and M. Haenni. Antimicrobial Usages and Antimicrobial Resistance in Commensal Escherichia coli From Veal Calves in France: Evolution During the Fattening Process. *Front. Microbiol.*, 10(792):1–10, 2019.
- [80] D. Gerhard. Ein Beitrag zur Chemotherapie der bakteriellen Infektionen. *Dtsch med Wochenschr*, 61(7):250–253, 1935.
- [81] C. González-Bello. Antibiotic adjuvants A strategy to unlock bacterial resistance to antibiotics. *Bioorg. Med. Chem. Lett.*, 27(18):4221–4228, 2017.
- [82] H. Goossens, M. Ferech, R. Vanderstichele, and M. Elseviers. Outpatient antibiotic use in Europe and association with resistance: a retrospective database study. *Lancet*, 365:579–587, 2005.
- [83] E. Grafskaia, K. Nadezhdin, I. Talyzina, N. Polina, O. Podgorny, E. Pavlova, P. Bashkirov, D. Kharlampieva, P. Bobrovsky, I. Latsis, V. Manuvera, V. Babenko, V. Trukhan, A. Arseniev, D. Klinov, and V. Lazarev. Medicinal leech antimicrobial peptides lacking toxicity represent a promising alternative strategy to combat antibiotic-resistant pathogens. *Eur. J. Med. Chem.*, 180:143–153, 2019.

- [84] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner. Opls3: A force field providing broad coverage of drug-like small molecules and proteins. J. Chem. Theory Comput., 12(1):281–296, 2016.
- [85] J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, and B. H. Morrow. Review of force fields and intermolecular potentials used in atomistic computational materials research. *Appl. Phys. Rev.*, 5(3):031104, 2018.
- [86] T. He, R. Wang, D. Liu, T. R. Walsh, R. Zhang, Y. Lv, Y. Ke, Q. Ji, R. Wei, Z. Liu, Y. Shen, G. Wang, L. Sun, L. Lei, Z. Lv, Y. Li, M. Pang, L. Wang, Q. Sun, Y. Fu, H. Song, Y. Hao, Z. Shen, S. Wang, G. Chen, C. Wu, J. Shen, and Y. Wang. Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. *Nat. Microbiol.*, 2019.
- [87] J. B. Hendrickson. Molecular geometry. i. machine computation of the common rings. J. Am. Chem. Soc., 83(22):4537–4547, 1961.
- [88] M. Hernandez Valladares, A. Felici, G. Weber, H. W. Adolph, M. Zeppezauer, G. M. Rossolini, G. Amicosante, J.-M. Frère, and M. Galleni. Zn(II) Dependence of the Aeromonas hydrophila AE036 Metallo- β -lactamase Activity and Stability. *Biochemistry*, 36(38):11534–11541, 1997.
- [89] T. L. Hill. Steric effects. i. van der waals potential energy curves. J. Chem. Phys., 16(4):399–404, 1948.
- [90] B. Hong, Y. Ba, L. Niu, F. Lou, Z. Zhang, H. Liu, Y. Pan, and Y. Zhao. A Comprehensive Research on Antibiotic Resistance Genes in Microbiota of Aquatic Animals. *Front. Microbiol.*, 9(1617):1–10, 2018.
- [91] J. S. Hong, D. Kim, E.-J. Yoon, H. Lee, and S. H. Jeong. Performance evaluation of the pana realtyper[™] cre kit for detecting carbapenemase genes in gram-negative bacilli. J. Glob. Antimicrob. Resist., 18:100–103, 2019.
- [92] D. A. Hopwood. How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them? *Mol. Microbiol.*, 63(4):937–940, 2007.
- [93] Y. Hua, T. Luo, Y. Yang, D. Dong, R. Wang, Y. Wang, M. Xu, X. Guo, F. Hu, and P. He. Phage Therapy as a Promising New Treatment for Lung Infection Caused by Carbapenem-Resistant Acinetobacter baumannii in Mice. *Front. Microbiol.*, 8(2659):1–11, 2018.

- [94] L. A. Hughes, G. Pinchbeck, R. Callaby, S. Dawson, P. Clegg, and N. Williams. Antimicrobial prescribing practice in UK equine veterinary practice. *Equine Vet. J.*, 45(2):141–147, 2013.
- [95] W. Ihlenfeldt, Y. Takahashi, H. Abe, and S. Sasaki. Computation and management of chemical properties in CACTVS: An extensible networked approach toward modularity and compatibility. *J. Chem. Inf. Comput. Sci.*, 34(1):109–116 (http://www.xemistry.com/), 1994.
- [96] A. Imada, K. Kitano, K. Kintaka, M. Muroi, and M. Asai. Sulfazecin and isosulfazecin, novel β -lactam antibiotics of bacterial origin. *Nature*, 289(5798):590–591, 1981.
- [97] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annu. Rev. Biophys.*, 48(1):371–394, 2019.
- [98] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids.
 10. transferable intermolecular potential functions for water, alcohols, and ethers. application to liquid water. J. Am. Chem. Soc., 103(2):335–340, 1981.
- [99] W. L. Jorgensen. Opls force fields. In *Encyclopedia of Computational Chemistry*, volume 3, pages 1986–1989. Wiley New York, 1998.
- [100] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc., 118(45):11225– 11236, 1996.
- [101] W. L. Jorgensen and C. J. Swenson. Optimized intermolecular potential functions for amides and peptides. hydration of amides. J. Am. Chem. Soc., 107(6):1489–1496, 1985.
- [102] W. L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc., 110(6):1657–1666, mar 1988.
- [103] W. L. Jorgensen and J. Tirado–Rives. Molecular modeling of organic and biomolecular systems using boss and mcpro. J. Comput. Chem., 26(16):1689–1700, 2005.
- [104] W. L. Jorgensen, J. P. Ulmschneider, and J. Tirado-Rives. Free energies of hydration from a generalized born model and an all-atom force field. J. Phys. Chem. B, 108(41):16264–16270, 2004.
- [105] G. Kaminski and W. L. Jorgensen. Performance of the amber94, mmff94, and opls-aa force fields for modeling organic liquids. J. Phys. Chem., 100(46):18010–18013, 1996.

- [106] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B, 105(28):6474–6487, jul 2001.
- [107] Y. Kamio and H. Nikaido. Outer membrane of Salmonella typhimurium: accessibility of phospholipid head groups to phospholipase C and cyanogen bromide activated dextran in the external medium. *Biochemistry*, 15(12):2561–2570, 1976.
- [108] I. Karaiskos, I. Galani, M. Souli, and H. Giamarellou. Novel β-lactamβ-lactamase inhibitor combinations: expectations for the treatment of carbapenem-resistant Gram-negative pathogens. *Expert. Opin. Drug Metab. Toxicol.*, 15(2):133–149, 2019.
- [109] C. F. Kettering, L. W. Shutts, and D. H. Andrews. A representation of the dynamic properties of molecules by mechanical models. *Phys. Rev.*, 36(3):531–543, 1930.
- [110] F. A. Khan, B. Söderquist, and J. Jass. Prevalence and Diversity of Antibiotic Resistance Genes in Swedish Aquatic Environments Impacted by Household and Hospital Wastewater. *Front. Microbiol.*, 10(688):1– 12, 2019.
- [111] D. T. King, S. Sobhanifar, and N. C. J. Strynadka. The mechanisms of resistance to β-lactam antibiotics. In *Handbook of Antimicrobial Resistance*, pages 177–201. Springer New York, 2017.
- [112] W. M. M. Kirby. Extraction of a highly potent penicillin inactivator from penicillin resistant staphylococci. *Science*, 99(2579):452–453, 1943.
- [113] E. Y. Klein, T. P. Van Boeckel, E. M. Martinez, S. Pant, S. Gandra, S. A. Levin, H. Goossens, and R. Laxminarayan. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proc. Natl. Acad. Sci.*, 115(15):E3463–E3470, 2018.
- [114] B. Knapp, S. Frantal, M. Cibena, W. Schreiner, and P. Bauer. Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *J. Comput. Biol.*, 18(8):997–1005, 2011.
- [115] A. Krajnc, P. A. Lang, T. D. Panduwawala, J. Brem, and C. J. Schofield. Will morphing boron-based inhibitors beat the β -lactamases? *Curr. Opin. Chem. Biol.*, 50:101–110, 2019.
- [116] M. Lakemeyer, W. Zhao, F. A. Mandl, P. Hammann, and S. A. Sieber. Thinking Outside the Box-Novel Antibacterials To Tackle the Resistance Crisis. *Angew. Chem. Int. Ed.*, 57(44):14440–14475, 2018.

- [117] G. W. Langley, R. Cain, J. M. Tyrrell, P. Hinchliffe, K. Calvopiña, C. L. Tooke, E. Widlake, C. G. Dowson, J. Spencer, T. R. Walsh, C. J. Schofield, and J. Brem. Profiling interactions of vaborbactam with metallo-β-lactamases. *Bioorg. Med. Chem. Lett.*, 29(15):1981–1984, 2019.
- [118] M. J. Latallo, G. A. Cortina, S. Faham, R. K. Nakamoto, and P. M. Kasson. Predicting allosteric mutants that increase activity of a major antibiotic resistance enzyme. *Chem. Sci.*, 8(9):6484–6492, 2017.
- [119] P. Lee, C. Chu, Y. Tsai, Y. Chuang, and F. Lung. Design, Synthesis, and Antimicrobial Activities of Novel Functional Peptides Against Grampositive and Gram-negative Bacteria. *Chem. Biol. Drug Des.*, in press (doi:10.1111/cbdd.13535), 2019.
- [120] S. A. Li, M. M. Cadelis, K. Sue, M. Blanchet, N. Vidal, J. M. Brunel, M.-L. Bourguet-Kondracki, and B. R. Copp. 6-Bromoindolglyoxylamido derivatives as antimicrobial agents and antibiotic enhancers. *Bioorg. Med. Chem.*, 27(10):2090–2099, 2019.
- [121] B. L. Ligon. Penicillin: its discovery and early development. *Semin. Pediatr. Infect. Dis.*, 15(1):52–57, 2004.
- [122] B. T. Liu, X. Y. Zhang, S. W. Wan, J. J. Hao, R. D. Jiang, and F. J. Song. Characteristics of carbapenem-resistant Enterobacteriaceae in ready-to-eat vegetables in China. *Front. Microbiol.*, 9(1147):1–12, 2018.
- [123] R. Liu, P. A. Miller, S. B. Vakulenko, N. K. Stewart, W. C. Boggess, and M. J. Miller. A Synthetic Dual Drug Sideromycin Induces Gram-Negative Bacteria To Commit Suicide with a Gram-Positive Antibiotic. *J. Med. Chem.*, 61(9):3845–3854, 2018.
- [124] R. López-Igual, J. Bernal-Bayard, A. Rodríguez-Patón, J.-M. Ghigo, and D. Mazel. Engineered toxin–intein antimicrobials can selectively target and kill antibiotic-resistant bacteria in mixed populations. *Nat. Biotechnol.*, 37(7):755–760, 2019.
- [125] J.-M. López-Lozano, T. Lawes, C. Nebot, A. Beyaert, X. Bertrand, D. Hocquet, M. Aldeyab, M. Scott, G. Conlon-Bingham, D. Farren, G. Kardos, A. Fésűs, J. Rodríguez-Baño, P. Retamar, N. Gonzalo-Jiménez, and I. M. Gould. A nonlinear time-series analysis approach to identify thresholds in associations between population antibiotic use and rates of resistance. *Nat. Microbiol.*, 4(7):1160–1172, 2019.

- [126] B. Ma, C. Fang, L. Lu, M. Wang, X. Xue, Y. Zhou, M. Li, Y. Hu, X. Luo, and Z. Hou. The antimicrobial peptide thanatin disrupts the bacterial outer membrane and inactivates the ndm-1 metallo-β-lactamase. *Nat. Commun.*, 10(1):3517, 2019.
- [127] C. Manyi-Loh, S. Mamphweli, E. Meyer, and A. Okoh. Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications. *Molecules*, 23(4):1–48, 2018.
- [128] S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse grained model for semiquantitative lipid simulations. J. Phys. Chem. B, 108(2):750–760, 2004.
- [129] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The martini force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007.
- [130] W. F. Marshall and J. E. Blair. The Cephalosporins. Mayo Clin. Proc., 74:187–195, 1999.
- [131] J. L. Martinez and F. Baquero. Mutation Frequencies and Antibiotic Resistance. *Antimicrob. Agents Chemother.*, 44(7):1771–1777, 2000.
- [132] D. A. Mathys, D. F. Mollenkopf, S. M. Feicht, R. J. Adams, A. L. Albers, D. M. Stuever, S. V. Grooters, G. A. Ballash, J. B. Daniels, and T. E. Wittum. Carbapenemase-producing Enterobacteriaceae and Aeromonas spp. present in wastewater treatment plant effluent and nearby surface waters in the US. *PLoS ONE*, 14(6):1–12, 2019.
- [133] K. Matsuzaki. Control of cell selectivity of antimicrobial peptides. Biochim. Biophys. Acta - Biomembr., 1788(8):1687–1692, 2009.
- [134] D. S. Maxwell, J. Tirado-Rives, and W. L. Jorgensen. A comprehensive study of the rotational energy profiles of organic systems by ab initio mo theory, forming a basis for peptide torsional parameters. *J. Comput. Chem.*, 16(8):984–1010, 1995.
- [135] M. J. McConnell. Where are we with monoclonal antibodies for multidrug-resistant infections? *Drug Discov. Today*, 24(5):1132–1138, 2019.
- [136] Z. Meng, M.-L. Tang, L. Yu, Y. Liang, J. Han, C. Zhang, F. Hu, J.-M. Yu, and X. Sun. Novel Mercapto Propionamide Derivatives with Potent New Delhi Metallo-β-Lactamase-1 Inhibitory Activity and Low Toxicity. *ACS Infect. Dis.*, 5(6):903–916, 2019.

- [137] D. Meziane-Cherif, R. Bonnet, A. Haouz, and P. Courvalin. Structural insights into the loss of penicillinase and the gain of ceftazidimase activities by OXA-145 β-lactamase in Pseudomonas aeruginosa. J. Antimicrob. Chemother., 71(2):395–402, 11 2015.
- [138] L. Monticelli and P. D. Tieleman. Force fields for classical molecular dynamics. In *Biomolecular Simulations: Methods and Protocols*, pages 197–213. Humana Press, 2013.
- [139] M. S. Mulani, E. E. Kamble, S. N. Kumkar, M. S. Tawre, and K. R. Pardesi. Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.*, 10(539):1– 24, 2019.
- [140] C. P. Murphy, R. J. Reid-Smith, P. Boerlin, J. S. Weese, J. F. Prescott, N. Janecko, and S. A. McEwen. Out-patient antimicrobial drug use in dogs and cats for new disease events from community companion animal practices in Ontario. *Can. Vet. J.*, 53(3):291–298, 2012.
- [141] T. Naas, S. Oueslati, R. A. Bonnin, M. L. Dabos, A. Zavala, L. Dortet,
 P. Retailleau, and B. I. Iorga. Beta-lactamase database (BLDB) structure and function. *J. Enzyme Inhib. Med. Chem.*, 32(1):917–919, 2017.
- [142] D. Nagarajan, T. Nagarajan, N. Roy, O. Kulkarni, S. Ravichandran, M. Mishra, D. Chakravortty, and N. Chandra. Computational antimicrobial peptide design and evaluation against multidrug-resistant clinical isolates of bacteria. J. Biol. Chem., 293(10):3492–3509, 2018.
- [143] R. Namivandi-Zangeneh, Z. Sadrearhami, D. Dutta, M. Willcox, E. H. H. Wong, and C. Boyer. Synergy between Synthetic Antimicrobial Polymer and Antibiotics: A Promising Platform To Combat Multidrug-Resistant Bacteria. ACS Infect. Dis., in press (doi:10.1021/acsinfecdis.9b00049), 2019.
- [144] G. G. F. Newton and E. P. Abraham. Degradation, structure and some derivatives of cephalosporin N. *Biochem. J.*, 58(1):103–111, 1954.
- [145] H. Nikaido. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol. Mol. Biol. Rev.*, 67(4):593–656, 1996.
- [146] N.-Z. Ning, X. Liu, F. Chen, P. Zhou, L. Hu, J. Huang, Z. Li, J. Huang, T. Li, and H. Wang. Embelin Restores Carbapenem Efficacy against NDM-1-Positive Pathogens. *Front. Microbiol.*, 9(71):1–8, 2018.
- [147] Y. Niu, M. Wang, Y. Cao, A. Nimmagadda, J. Hu, Y. Wu, J. Cai, and X.-S. Ye. Rational Design of Dimeric Lysine N-Alkylamides as Potent and Broad-Spectrum Antibacterial Agents. J. Med. Chem., 61(7):2865– 2874, 2018.

- [148] M. C. Orencia, J. S. Yoon, J. E. Ness, W. P. Stemmer, and R. C. Stevens. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.*, 8(3):238–242, 2001.
- [149] W. H. Organization. The detection and reporting of colistin resistance. World Health Organization, https://apps.who.int/iris/handle/10665/277175, 2018.
- [150] S. Oueslati, D. Girlich, L. Dortet, and T. Naas. Evaluation of the amplidiag CarbaRVRE kit for accurate detection of carbapenemaseproducing bacteria. J. Clin. Microbiol., 56(3):1–9, 2018.
- [151] T. Palzkill. Metallo-β-lactamase structure and function. Ann. N.Y. Acad. Sci., 1277(1):91–104, 2013.
- [152] J. T. Park and J. L. Strominger. Mode of Action of Penicillin. Science, 125(3238):99–101, 1957.
- [153] B. Parrino, D. Schillaci, I. Carnevale, E. Giovannetti, P. Diana, G. Cirrincione, and S. Cascioferro. Synthetic small molecules as anti-biofilm agents in the struggle against antibiotic resistance. *Eur. J. Med. Chem.*, 161:154–178, 2019.
- [154] L. Pasteur and J. F. Joubert. Charbon et septicémie. C. R. Acad. Sci., 85:101–15, 1877.
- [155] R. Paul-Soto, R. Bauer, J.-M. Frère, M. Galleni, W. Meyer-Klaucke, H. Nolting, G. M. Rossolini, D. de Seny, M. Hernandez-Valladares, M. Zeppezauer, and H.-W. Adolph. Mono- and Binuclear Zn2+ β-Lactamase. J. Biol. Chem., 274(19):13242–13249, 1999.
- [156] E. Peterson and P. Kaur. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Front. Microbiol.*, 9(2928):1–21, 2018.
- [157] T. D. M. Pham, Z. Ziora, and M. Blaskovich. Quinolone Antibiotics. Med. Chem. Comm., in press (doi:10.1039/C9MD00120D), 2019.
- [158] S. H. Podolsky. The evolving response to antibiotic resistance (1945–2018). *Palgrave Commun.*, 4(124):1–8, 2018.
- [159] J. Ponder and D. Case. Force fields for protein simulations. protein simulations. *Adv. Prot. Chem*, 66:27–85, 2003.
- [160] X. Qian, S. Zhang, S. Xue, W. Mao, M. Xu, W. Xu, and H. Xie. A carbapenem-based fluorescence assay for the screening of metallo-βlactamase inhibitors. *Bioorg. Med. Chem. Lett.*, 29(2):322–325, 2018.

- [161] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still. The gb/sa continuum model for solvation. a fast analytical method for the calculation of approximate born radii. *J. Phys. Chem. A*, 101(16):3005–3014, 1997.
- [162] C. R. H. Raetz and W. Dowhan. Biosynthesis and function of phospholipids in Escherichia coli. J. Biol. Chem., 265(3):1235–1238, 1990.
- [163] C. R. H. Raetz and C. Whitfield. Lipopolysaccharide Endotoxins. Annu. Rev. Biochem., 71(1):635–700, 2002.
- [164] R. C. Rizzo and W. L. Jorgensen. OPLS All-Atom Model for Amines: Resolution of the Amine Hydration Problem. J. Am. Chem. Soc., 121(20):4827–4836, 1999.
- [165] M. J. Robertson, Y. Qian, M. C. Robinson, J. Tirado-Rives, and W. L. Jorgensen. Development and testing of the opls-aa/m force field for rna. J. Chem. Theory Comput., 15(4):2734–2742, 2019.
- [166] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. J. Chem. Theory Comput., 11(7):3499–3509, 2015.
- [167] G. Rolinson, F. Batchelor, S. Stevens, J. Wood, and E. Chain. Bacteriological studies on a new penicillin–BRL.1241. *Lancet*, 276(7150):564– 567, 1960.
- [168] G. N. Rolinson and S. Stevens. Microbiological Studies on a New Broad-spectrum Penicillin, "Penbritin". *Br. Med. J.*, 2(5246):191–196, 1961.
- [169] C. E. H. Rosenkilde, C. Munck, A. Porse, M. Linkevicius, D. I. Andersson, and M. O. A. Sommer. Collateral sensitivity constrains resistance evolution of the CTX-M-15 β -lactamase. *Nat. Commun.*, 10(618):1–10, 2019.
- [170] C. M. Rotondo and G. D. Wright. Inhibitors of metallo- β -lactamases. *Curr. Opin. Microbiol.*, 39:96–105, 2017.
- [171] D. Saeloh, V. Tipmanee, K. K. Jim, M. P. Dekker, W. Bitter, S. P. Voravuthikunchai, M. Wenzel, and L. W. Hamoen. The novel antibiotic rhodomyrtone traps membrane proteins in vesicles with increased fluidity. *PLOS Pathog.*, 14(2):1–35, 2018.
- [172] E. R. Sams, M. Whiteley, and K. H. Turner. 'The battle for life': Pasteur, anthrax, and the first probiotics. J. Med. Microbiol., 63:1573–1574, 2014.
- [173] E. Sauvage, F. Kerff, M. Terrak, J. A. Ayala, and P. Charlier. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.*, 32(2):234–258, 2008.

- [174] D. Schillaci, V. Spanò, B. Parrino, A. Carbone, A. Montalbano, P. Barraja, P. Diana, G. Cirrincione, and S. Cascioferro. Pharmaceutical Approaches to Target Antibiotic Resistance Mechanisms. *J. Med. Chem.*, 60(20):8268–8297, 2017.
- [175] K. Sen, T. Berglund, M. A. Soares, B. Taheri, Y. Ma, L. Khalil, M. Fridge, J. Lu, and R. J. Turner. Antibiotic Resistance of E. coli Isolated From a Constructed Wetland Dominated by a Crow Roost, With Emphasis on ESBL and AmpC Containing E. coli. *Front. Microbiol.*, 10(1034):1–13, 2019.
- [176] C. Shi, J. Chen, X. Kang, X. Shen, X. Lao, and H. Zheng. Approaches for the discovery of metallo-β-lactamase inhibitors: A review. *Chem. Biol. Drug Des.*, in press (doi:10.1111/cbdd.13526), 2019.
- [177] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren. Polarizable atomic multipole-based amoeba force field for proteins. J. Chem. Theory Comput., 9(9):4046–4063, 2013.
- [178] T. Siemon, S. Steinhauer, and M. Christmann. Synthesis of (+)-Darwinolide, a Biofilm-Penetrating Anti-MRSA Agent. Angew. Chem. Int. Ed., 58(4), 2019.
- [179] D. A. Singleton, F. Sánchez-Vizcaíno, S. Dawson, P. H. Jones, P. J. Noble, G. L. Pinchbeck, N. J. Williams, and A. D. Radford. Patterns of antimicrobial agent prescription in a sentinel population of canine and feline veterinary practices in the United Kingdom. *Vet. J.*, 224:18–24, 2017.
- [180] D. W. Smith. Decreased Antimicrobial Resistance Following Changes in Antibiotic Use. Surg. Infect., 1(1):73–78, 2000.
- [181] C. Sotriffer. Docking of covalent ligands: Challenges and approaches. *Mol. Inf.*, 37(9-10):1800062, 2018.
- [182] R. B. Sykes, C. M. Cimarusti, D. P. Bonner, K. Bush, D. M. Floyd, N. H. Georgopapadakou, W. H. Koster, W. C. Liu, W. L. Parker, P. A. Principe, M. L. Rathnum, W. A. Slusarchyk, W. H. Trejo, and J. S. Wells. Monocyclic β -lactam antibiotics produced by bacteria. *Nature*, 291(5815):489–491, 1981.
- [183] V. Tiberio. Sugli estratti di alcune muffe. Annali d'Igiene Sperimentale, 5:91–103, 1895.
- [184] D. J. Tipper and J. L. Strominger. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc. Natl. Acad. Sci.*, 54(4):1133–1141, 1965.

- [185] J. Tirado-Rives and W. L. Jorgensen. Molecular dynamics of proteins with the opls potential functions. simulation of the third domain of silver pheasant ovomucoid in water. J. Am. Chem. Soc., 112(7):2773– 2781, 1990.
- [186] C. L. Tooke, P. Hinchliffe, E. C. Bragginton, C. K. Colenso, V. H. Hirvonen, Y. Takebayashi, and J. Spencer. β-Lactamases and β-Lactamase Inhibitors in the 21st Century. J. Mol. Biol., in press (doi:10.1016/j.jmb.2019.04.002), 2019.
- [187] N. J. Torelli, A. Akhtar, K. DeFrees, P. Jaishankar, O. A. Pemberton, X. Zhang, C. Johnson, A. R. Renslo, and Y. Chen. Active-Site Druggability of Carbapenemases and Broad-Spectrum Inhibitor Discovery. *ACS Infect. Dis.*, 5(6):1013–1021, 2019.
- [188] E. Troiano, L. Beneduce, A. Gross, and Z. Ronen. Antibiotic-Resistant Bacteria in Greywater and Greywater-Irrigated Soils. *Front. Microbiol.*, 9(2666):1–13, 2018.
- [189] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. Reaxff: A reactive force field for hydrocarbons. J. Phys. Chem. A, 105(41):9396– 9409, 2001.
- [190] A. S. J. P. A. M. van Miert. The sulfonamide-diaminopyrimidine story*. J. vet. Pharmacol. Therap., 17:309–316, 1994.
- [191] P. Vikesland, E. Garner, S. Gupta, S. Kang, A. Maile-Moskowitz, and N. Zhu. Differential Drivers of Antimicrobial Resistance across the World. Acc. Chem. Res., 52(4):916–924, 2019.
- [192] W. Vollmer, D. Blanot, and M. A. De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.*, 32(2):149–167, 2008.
- [193] S. A. Waksman and H. B. Woodruff. The soil as a source of microorganisms antagonistic to disease-producing bacteria. *J. Bacteriol.*, 40(4):581–600, 1940.
- [194] S. A. Waksman and H. B. Woodruff. Streptothricin, a new selective bacteriostatic and bactericidal agent, particularly active against gramnegative bacteria. *Exp. Biol. Med.*, 49(2):207–210, 1942.
- [195] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

- [196] Y.-L. Wang, S. Liu, Z.-J. Yu, Y. Lei, M.-Y. Huang, Y.-H. Yan, Q. Ma, Y. Zheng, H. Deng, Y. Sun, C. Wu, Y. Yu, Q. Chen, Z. Wang, Y. Wu, and G.-B. Li. Structure-Based Development of (1-(3'-Mercaptopropanamido)methyl)boronic Acid Derived Broad-Spectrum, Dual-Action Inhibitors of Metallo- and Serine-β-Lactamases. *J. Med. Chem.*, in press (doi:10.1021/acs.jmedchem.9b00735), 2019.
- [197] Z. Wang, W. Fast, A. M. Valentine, and S. J. Benkovic. Metallo- β -lactamase: structure and mechanism. *Curr. Opin. Chem. Biol.*, 3(5):614–622, 1999.
- [198] P. K. Weiner and P. A. Kollman. Amber: Assisted model building with energy refinement. a general program for modeling molecules and their interactions. J. Comput. Chem., 2(3):287–303, 1981.
- [199] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, 7(2):230–252, 1986.
- [200] F. H. Westheimer. A calculation of the energy of activation for the racemization of 2,2'-dibromo-4,4'-dicarboxydiphenyl. J. Chem. Phys., 15(5):252–260, 1947.
- [201] F. H. Westheimer and J. E. Mayer. The theory of the racemization of optically active derivatives of diphenyl. J. Chem. Phys., 14(12):733– 738, 1946.
- [202] L. S. Weyrich, S. Duchene, J. Soubrier, L. Arriola, B. Llamas, J. Breen, A. G. Morris, K. W. Alt, D. Caramelli, V. Dresely, M. Farrell, A. G. Farrer, M. Francken, N. Gully, W. Haak, K. Hardy, K. Harvati, P. Held, E. C. Holmes, J. Kaidonis, C. Lalueza-Fox, M. De La Rasilla, A. Rosas, P. Semal, A. Soltysiak, G. Townsend, D. Usai, J. Wahl, D. H. Huson, K. Dobney, and A. Cooper. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, 544(7650):357– 361, 2017.
- [203] D. White, J. Drummond, and C. Fuqua. *The Physiology and Biochemistry of Prokaryotes*. Oxford University Press, 2012.
- [204] K. B. Wiberg. A scheme for strain energy minimization. application to the cycloalkanes. *J. Am. Chem. Soc.*, 87(5):1070–1078, 1965.
- [205] D. Xu, D. Xie, and H. Guo. Catalytic Mechanism of Class B2 Metallo- β -lactamase. J. Biol. Chem., 281(13):8740–8747, 2006.

- [206] J. Xu, M. E. Pachón-Ibáñez, T. Cebrero-Cangueiro, H. Chen, J. Sánchez-Céspedes, and J. Zhou. Discovery of niclosamide and its Oalkylamino-tethered derivatives as potent antibacterial agents against carbapenemase-producing and/or colistin resistant Enterobacteriaceae isolates. *Bioorg. Med. Chem. Lett.*, 29(11):1399–1402, 2019.
- [207] M. Ye, M. Sun, Y. Feng, J. Wan, S. Xie, D. Tian, Y. Zhao, J. Wu, F. Hu, H. Li, and X. Jiang. Effect of biochar amendment on the control of soil sulfonamides, antibiotic-resistant bacteria, and gene enrichment in lettuce tissues. *J. Hazard. Mater.*, 309:219–227, 2016.
- [208] I. Yosef, M. Manor, R. Kiro, and U. Qimron. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci.*, 112(23):7267–7272, 2015.
- [209] M. Zainal-Abideen, A. Aris, F. Yusof, Z. Abdul-Majid, A. Selamat, and S. I. Omar. Optimizing the coagulation process in a drinking water treatment plant – comparison between traditional and statistical experimental design jar tests. *Water Sci. Technol.*, 65(3):496–503, 2012.
- [210] A. Zapun, C. Contreras-Martel, and T. Vernet. Penicillin-binding proteins and β -lactam resistance. *FEMS Microbiol. Rev.*, 32(2):361–385, 2008.
- [211] M. Zeineldin, B. Aldridge, and J. Lowe. Antimicrobial Effects on Swine Gastrointestinal Microbiota and Their Accompanying Antibiotic Resistome. *Front. Microbiol.*, 10(1035):1–14, 2019.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification. Any contributions from colleagues are explicitly referenced in the attached publications. Finally, parts of this work have been published in five publications:

Selwa, E., **Elisée, E.**, Zavala, A., and Iorga, B. I. (2018) Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations, *J. Comput. Aided Mol. Des.* 32, 273-286 DOI: 10.1007/s10822-017-0054-1.

Zavala, A., Retailleau, P., **Elisée, E.**, Iorga, B. I., and Naas, T. (2019) Genetic, biochemical and structural characterization of CMY-136 beta-lactamase, a peculiar CMY-2 variant, *ACS Infect. Dis.* 5, 528-538 DOI: 10.1021/acsinfecdis.8b00240.

Chaput, L., Selwa, E., **Elisée, E.**, and Iorga, B. I. (2019) Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations, *J. Comput. Aided Mol. Des.* 33, 93-103 DOI: 10.1007/s10822-018-0161-7.

Elisée, E., Gapsys, V., Mele, N., Chaput, L., Selwa, E., de Groot, B. L. and Iorga, B. I. (2019) Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset, *J. Comput. Aided Mol. Des.* DOI: 10.1007/s10822-019-00232-w.

Elisée, E., and Iorga, B.I. (2019) OPLS-AA force field parameterization of zinc-coordinating residues in metalloenzymes. – manuscript submitted to J. Chem. Inf. Model..

Gif-sur-Yvette, September 4, 2019

Eddy Elisée

Colophon

This thesis was typeset with $\mathbb{E}T_{E}X 2_{\varepsilon}$. It uses a modified version of *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the Clean Thesis style at http://cleanthesis.der-ric.de/.



ÉCOLE DOCTORALE

Sciences chimiques : molécules, matériaux, instrumentation et biosystèmes

Titre : Vers la prédiction in silico des mutations liées à la résistance aux antibiotiques.

Mots clés : Résistance aux antibiotiques, β -lactamase, dynamique moléculaire, calculs d'énergie libre, métalloenzyme, champ de force OPLS-AA.

Résumé : La résistance aux antibiotiques est une menace sérieuse pour la santé publique. En effet, si on ne change pas rapidement notre consommation excessive d'antibiotiques, la situation actuelle va se dégrader jusqu'à basculer dans une ère dite "post-antibiotique", dans laquelle plus aucun antibiotique ne sera efficace contre les infections microbiennes. Bien que ce phénomène de résistance apparaisse naturellement, l'utilisation abusive d'antibiotiques accélère le processus. De plus, la présence de pathogènes multi-résistants neutralise l'effet des traitements existants et dans le cas de chirurgies courantes (césariennes, transplantations d'organe...), la situation peut rapidement s'aggraver voire devenir mortelle.

C'est pourquoi des directives, émanant des autorités sanitaires, doivent être mises en place afin de contrôler l'utilisation des médicaments, et ce, à tous les niveaux de la société, des individus au secteur agricole en passant par les professionnels de santé et les industries pharmaceutiques. Le monde de la recherche scientifique, quant à elle, doit trouver des nouvelles stratégies pour enrayer la propagation de la résistance.

Dans ce contexte, cette thèse a pour objectif le développement d'une méthode de prédiction, par calculs d'énergie libre, des mutations de β -lactamases favorables à l'hydrolyse des β -lactames. Ces travaux méthodologiques ont donc conduit au développement : (1) de nouveaux paramètres pour les enzymes à zinc, implémentés dans le champ de force OPLS-AA et validés par des simulations de dynamique moléculaire sur un panel de métalloenzymes représentatives, (2) d'un protocole de paramétrisation de ligands covalents pour étudier le comportement de certains β -lactames dans CMY-136, une nouvelle β -lactamase caractérisée au laboratoire, et (3) d'un protocole de calcul d'énergie libre évalué au moyen de compétitions internationales de prédiction. Ce dernier a ensuite été utilisé pour tenter d'expliquer pourquoi la carbamylation de la sérine catalytique n'a pas lieu dans certaines oxacillinases.

Au travers de ces travaux, nous avons pu améliorer significativement notre approche computationnelle et désormais tout est en place pour une exploration exhaustive des mutations possibles dans les β -lactamases.

Title : Towards in silico prediction of mutations related to antibiotic resistance.

Keywords : Antibiotic resistance, β -lactamase, molecular dynamics, free energy calculations, metalloenzyme, OPLS-AA force field.

Abstract : Antibiotic resistance is a global concern threatening worldwide health. Indeed, if we don't change our overconsumption of antibiotics, the current situation could worsen until a "post-antibiotic" era in which existing treatment would be ineffective against microbial infections. Despite the natural occurrence of antibiotic resistance, the misuse of antibiotics is speeding up the process. Furthermore, presence of multi-resistant pathogens negates the effect of modern treatments and usual surgeries (caesarean sections, organ transplantations...) might be riskier in the future, or even lethal.

That's why, common guidelines have to be edicted by health authorities in order to control antibiotic use at every level of society, from individuals to healthcare industry including health professionals and agriculture sector. As for scientific research, new strategies have to be considered in order to limit the spread of antibiotic resistance.

In that context, the presented thesis aimed at developing a protocol to predict, by free energy calculations, β -lactamase mutations which could promote the hydolysis of β -lactams antibiotics. In order to achieve that, we developed several methodological approaches including : (1) new parameters for zinc enzymes implemented in OPLS-AA force field and thereafter validated using molecular dynamics simulations of representative zinccontaining metalloenzymes, (2) a protocol to parameterize covalent ligands in order to analyze the dynamical behavior of some β -lactams in CMY-136, a novel β -lactamase recently characterized in our laboratory, and (3) a pmx-based free energy protocol. The latter was also assessed through several international blinded prediction challenges, and finally used to find out why carbamylation of the catalytic serine is not observed in certain OXA enzymes.

Throughout this work, we made significant improvements in our protocol, and now everything is in place for an exhaustive prediction of possible mutations in β -lactamases.