# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis
## „Comparative MD simulation studies of three metallo-β-lactamases from antibiotic-resistant bacteria"

verfasst von / submitted by
### Klára Přibáňová

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
### Master of Science (MSc)

Wien, 2023 / Vienna, 2023

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | UA 066863 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Masterstudium Biologische Chemie |
| Betreut von / Supervisor: | Univ.-Prof. Mag. Dr. Stefan Boresch |

# Acknowledgements

While working on my master's thesis, I met several people whom I would like to mention here.

In the first place, I would like to thank my supervisor, Univ.-Prof. Mag. Dr. Stefan Boresch, for giving me the opportunity to join the Department for Computational Biological Chemistry and for having confidence in me for the entire time I spent here. I appreciate Stefan's professionalism, his ability to spot suspicious numbers in gigantic tables within seconds, and his advice and guidance I have taken advantage of.

I would also like to express my thanks to Dr. Åsmund Kaupang from the University of Oslo for mentoring me. Thank you for introducing me to metallo-β-lactamases, for explaining everything from Linux to Python, for answering my questions, for always being ready for a meeting, and for being able to spread enthusiasm over more than a thousand kilometers.

I also want to express my gratitude to everyone whom I met at the Department of Computational Biological Chemistry. Thank you for your support, for your help, and for the cakes!

# Abstract

Antimicrobial resistance is considered a serious crisis of our time, posing a threat to human health worldwide. A prominent class of enzymes that provide bacteria with antimicrobial resistance are metallo-β-lactamases. These enzymes utilize $Zn^{2+}$ ions in their active site to catalyze the cleavage of the β-lactam ring present in all β-lactam antibiotics. They are considered a major danger in the context of antimicrobial resistance, mostly due to the current lack of effective inhibitors of these enzymes for clinical use.

This thesis concentrates on three relevant representatives of this enzyme class: NDM-1 from *Klebsiella pneumoniae*, VIM-2 from *Pseudomonas aeruginosa*, and IMP-1 from *Serratia marcescens*. Using molecular dynamics (MD) simulations, the dynamic behavior of the biologically active structures of the enzymes NDM-1, VIM-2, and IMP-1 in aqueous solution with no ligands bound in their active site was examined.

First, custom patches that describe the coordination of the two $Zn^{2+}$ ions present in the active site of the enzymes of interest were developed. The parameters of these patches were refined so that the average active site geometries of these proteins observed in the MD simulations were maintained as accurately as possible, compared to the X-ray crystallographic structures used as initial coordinates.

Second, after carrying out MD simulations with the final version of these custom patches, the stability of the enzyme structures during the MD simulations was monitored by measuring the root-mean-square deviation as a function of simulation time. The resulting stable trajectories were analyzed from several perspectives.

Selected interatomic distances and angles were measured to identify similarities and differences in the dynamics of the active sites of the enzymes. We found several indications suggesting a narrower shape of the IMP-1 active site in comparison to the other two enzymes. By contrast, we observed no significant differences regarding the number of water molecules detected within a spherical approximation of the active site. The coordination geometries of both $Zn^{2+}$ ions present in the active site of the proteins of interest were studied, and differences were discovered in the behavior of the active sites in the MD simulations in comparison to the experimental structures described in the literature. Although some observations concerning the $Zn^{2+}$ coordination might be artifacts of the chosen methodology, in particular of the classical mechanical force field used, the IMP-1 active site again behaves slightly differently from the other two enzymes. We also compared the mobility of individual residues within the full protein structures during the MD simulations. All secondary structure elements conserved across the three enzymes were well maintained during all simulations.

# Kurzfassung

Antimikrobielle Resistenz ist eine ernste, aktuelle Krise, die weltweit eine Bedrohung für die menschliche Gesundheit darstellt. Die sogenannten Metallo-β-Lactamasen sind eine Klasse von Enzymen, die Bakterien zu antimikrobieller Resistenz verhelfen. Diese Enzyme nutzen $Zn^{2+}$-Ionen in ihrem aktiven Zentrum, um die Spaltung des β-Lactamrings zu katalysieren, der in allen β-Lactam-Antibiotika vorkommt. Sie gelten als eine große Gefahr im Zusammenhang mit der antimikrobiellen Resistenz, vor allem weil es derzeit keine wirksamen Inhibitoren, die für den klinischen Einsatz zugelassen sind, gibt.

Die vorliegende Arbeit konzentriert sich auf drei relevante Vertreter dieser Enzymklasse: NDM-1 aus *Klebsiella pneumoniae*, VIM-2 aus *Pseudomonas aeruginosa*, und IMP-1 aus *Serratia marcescens*. Mithilfe von Molekulardynamiksimulationen (MD) wurde das dynamische Verhalten der biologisch aktiven Strukturen der Enzyme NDM-1, VIM-2 und IMP-1 ohne Liganden im aktiven Zentrum in wässriger Lösung untersucht.

Zunächst wurden angepasste "Patches" entwickelt, die die Koordination der beiden $Zn^{2+}$-Ionen im aktiven Zentrum der interessierenden Proteine beschreiben. Die Parameter dieser Patches wurden so verfeinert, dass die in den MD-Simulationen beobachtete durchschnittliche Geometrie des aktiven Zentrums der Proteine im Vergleich zu den als Ausgangskoordinaten verwendeten röntgenkristallographischen Strukturen so genau wie möglich erhalten blieb.

Zweitens wurde nach der Durchführung von MD-Simulationen mit der endgültigen Version dieser Patches die Stabilität der Proteinstrukturen während der MD-Simulationen durch Messung der mittleren quadratischen Abweichung als Funktion der Simulationszeit überwacht. Die resultierenden stabilen Trajektorien wurden unter verschiedenen Gesichtspunkten weiter analysiert.

Ausgewählte interatomare Abstände und Winkel wurden gemessen, um Ähnlichkeiten und Unterschiede in der Dynamik der aktiven Zentren der untersuchten Proteine zu identifizieren. Wir fanden mehrere Hinweise, die auf eine verengte Geometrie des aktiven Zentrums von IMP-1 im Vergleich zu den beiden anderen Proteinen hindeuten. Im Gegensatz dazu konnten wir keine signifikanten Unterschiede in Bezug auf die Anzahl der Wassermoleküle feststellen, die sich im Mittel innerhalb einer kugelförmigen Näherung des aktiven Zentrums befinden. Die Koordinationsgeometrien der beiden $Zn^{2+}$-Ionen, die im aktiven Zentrum der Proteine vorhanden sind, wurden untersucht, und es wurden Unterschiede im Verhalten der aktiven Zentren in den MD-Simulationen im Vergleich zu experimentellen röntgenkristallographischen Strukturen in der Literatur gefunden. Obwohl einige Beobachtungen bezüglich der $Zn^{2+}$-Koordination Artefakte der gewählten Methodik sein könnten, insbesondere des verwendeten klassischen Kraftfelds, verhält sich das aktive Zentrum von IMP-1 wiederum etwas anders als die anderen beiden Proteine. Wir verglichen auch die Mobilität einzelner Aminosäuren innerhalb der Protein-

strukturen während der MD-Simulationen. Alle in den drei Proteinen konservierten Sekundärstrukturelemente blieben während der Simulationen stabil.

# Contents

*Contents*

# List of Tables

# List of Figures

# 1. Introduction

As a result of the widespread application of antibiotics since their discovery almost a hundred years ago, the selection pressure applied to bacteria has increased. This has accelerated the manifestation of several ways in which bacteria can escape the antimicrobial effects of these drugs. Decades ago, the spread of resistant bacteria started being noticed among the scientific community, but the impact on the human population has become apparent only recently. Together with the lower rate of development and production of new antibiotics compared to their golden age, the antibiotic-resistant bacteria have started posing a significant danger to humans. This situation, called the antimicrobial resistance crisis, its causes, consequences and possible solutions are described in Chapter 2.

This thesis concentrates on three proteins involved in one of many possible mechanisms through which bacteria can gain resistance to antibiotics. These proteins—NDM-1, VIM-2, and IMP-1—belong to a class of bacterial enzymes called metallo-β-lactamases, that are capable of destroying antibiotic drugs.

Although structurally and mechanistically extremely conserved, their susceptibility to certain inhibitors has been identified as non-uniform. In order to detect potential differences among the three proteins, molecular dynamics (MD) simulations are carried out and their trajectories are thoroughly analyzed. The rationale behind the choice of these proteins of interest, the simulation methodology, and analysis procedures are described in detail in Chapter 3.

The primary goal of this thesis is to set up MD simulations of the chosen proteins in a chemically correct and meaningful manner. Several aspects of the chemical properties of the systems must be considered to obtain reliable trajectories. In this context, special attention needs to be paid to the active site. Not only the geometrical quantities, such as distances or angles, are monitored, but also other aspects of the characteristic active site of metallo-β-lactamases are studied, such as the coordination of the $Zn^{2+}$ ions. The trajectories obtained from these MD simulations can be used for further investigations of the proteins, e.g., as the starting points of quantum mechanics/molecular mechanics (QM/MM) simulations. These are, however, outside the scope of this master's thesis.

The second aim is to establish differences and similarities among the chosen proteins. Here, attention is again paid to the active site, its dynamics, and solvent accessibility. Moreover, the overall structural properties, such as the mobility and flexibility of certain regions of the proteins of interest with respect to their sequence alignment or secondary structure elements, are studied. These results are presented in Chapter 4.

# 2. Antimicrobial resistance

Bacteria that gain antimicrobial resistance (AMR) become able to bypass the effects of antimicrobial drugs, making the medical treatment ineffective and the spread of an infection easier [1]. Due to the high dissemination of resistant bacteria and the limited potential of antibiotic treatment of infections caused by these species, AMR has become a widely discussed topic in both the scientific sphere and the public media [2].

Antimicrobial resistance indeed poses a significant threat to global health. In 2019, 4.95 million deaths were associated with AMR, including 1.27 million deaths that could have been prevented if the infection had been susceptible to currently available antimicrobial drugs [3]. Deaths associated with AMR have been reported worldwide, with the most affected regions being sub-Saharan Africa and South Asia. Dominant pathogens responsible for deaths associated with AMR are *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, or *Streptococcus pneumoniae*. It has been estimated that the annual mortality rate of AMR could reach 10 million people in 2050, the most affected continents being Asia and Africa. This number is comparable with the expected mortality rate of cancer in the same year (8.2 millions) [4].

The World Health Organization (WHO) placed AMR among the top 10 global public health threats [5] and, besides that, describes the crisis as "socio-economic" [6], pointing out its impact on international trade and health care costs such as the treatment itself, extended hospital stays, intensive care units or isolated beds. The additional health care costs attributable to AMR could reach approximately 1 trillion US\$ per year by 2050 worldwide [4]. In 2015, the World Health Assembly (a forum consisting of the health ministers of all WHO member states) approved the *Global action plan on antimicrobial resistance* (GAP, [7])—a document defining the main objectives in both national and international actions against AMR. GAP emphasizes the need for awareness, research, reduction of AMR incidence, optimization of antibiotics use in humans and animals, and investment into new drugs or diagnostics. It serves as a framework for national action plans (NAPs) of the individual WHO member states. An analysis of NAPs [8] revealed that only 86 countries out of 194 WHO members established their NAPs by 2021. As the review pointed out, at least 15 of these countries reported that antibiotics were available without prescription or consumed by self-medication. These data illustrate that AMR is a high-priority problem that requires a global reaction [7].

## 2.1. AMR origin and mechanisms

AMR can be intrinsic to a certain species or acquired. Intrinsic resistance means that all members of that species exhibit the resistance and that the resistance phenotype is always

present. Alternatively, bacteria can acquire resistance by mutation of their chromosomal DNA or by horizontal gene transfer (typically by receiving a plasmid-encoded gene of resistance) [9].

Bacteria use several mechanisms to prevent the action of antimicrobial drugs: they can reduce the uptake of the drug, modify a drug target, dispose of the drug using efflux pumps or inactivate the drug [9]. The resistance development can be supported by stress factors like pH, high temperature, osmotic shock or oxidative stress [10].

To exert their antimicrobial function, most antibiotics need to get inside the bacterial cell first. Several resistance mechanisms therefore rely on preventing the drug entry [11]. Limiting the drug uptake is a common mechanism in Gram-negative bacteria. They possess a lipopolysaccharide layer in their outer membrane, which provides them with intrinsic resistance against certain antibiotic molecules. Hydrophilic antibiotics that cannot cross the outer membrane use porins to enter the cell. In order to prevent the uptake of such drugs, bacteria manipulate the number, size, and selectivity of these porins [10]. An example of resistance displayed by a bacterial community is the production of a biofilm [9]. A special case is bacteria without a cell wall, e.g., the *Mycoplasma* genus—these are naturally resistant against all antimicrobial agents that target the cell wall, i.e., β-lactams or glycopeptides.

Bacteria can alter the structure or number of molecules that are targeted by antibiotics. For example, Gram-positive bacteria can decrease the number of penicillin-binding proteins or, alternatively, increase the number of the low-affinity ones. This leads to insensitivity towards β-lactams. Mutations to enzymes against which antibiotics are targeted usually happen in the active site or in its vicinity, preserving, though, the ability of the natural substrate to bind to them. Another example of this resistance mechanism includes alterations in the DNA gyrase or topoisomerase IV structure (protection against fluoroquinolones) [9]. Alternatively, the target structures can be protected by adding a chemical group (e.g., methylation of ribosomal RNA [rRNA]) [11].

A certain concentration of an antimicrobial drug inside the cell is required to achieve the desired effect. Efflux pumps are transmembrane proteins that aim to remove toxic substances (including antibiotics) promptly from the bacterial cell before they manage to execute their effect [10]. This is done by active transport; the necessary energy is provided by ATP hydrolysis, $Na^+$ gradient or proton-motive force [9]. Mutations of transcription regulators responsible for the efflux pumps' production control were often detected in clinical isolates of resistant bacteria—these mutations lead to overexpression of the efflux pumps [11]. The pumps can be specific (e.g., to macrolides, tetracyclines) or target a wider spectrum of antibiotics [10].

Inactivation of an antimicrobial drug may lead either to its complete destruction (or to destruction of the active moiety of the molecule) or to the addition of a chemical group that disables the antibiotic action [9]. This includes acetylation, phosphorylation or adenylation of antibiotic molecules, typically aminoglycosides or chloramphenicol [10]. Drug inactivation is often a favorable choice of resistance for bacteria, as it is associated with less fitness costs than the other mechanisms described above [11]. A classical example of degradation of the antibiotic's active moiety is the hydrolysis of β-lactam antibiotics

by β-lactamases. Their mechanism of action is explained in detail in Section 2.3, after a description of β-lactam antibiotics in Section 2.2.

The development of resistant bacterial strains and their dissemination that we observe nowadays is not a natural process, but a consequence of human activity [12]. Since the beginnings of the production of antibiotics by the pharmaceutical industry in the 1940s, antibiotics have been released into the environment and have applied a selection pressure on bacteria worldwide—a condition favoring the survival of resistant microbes and thus leading to the natural selection of resistant strains.

The main driver of AMR spread is the misuse and overuse of antibiotics in human and veterinary medicine and in agriculture [13]. A positive correlation between increased antibiotic consumption and AMR was reported in many studies, both in humans [14] and animals [15]. In human medicine, antibiotics are often prescribed inappropriately to a patient when the pathogen is not identified [16]. For example, up to 50% of prescribed antibiotics in Canada were shown to be unnecessary [8]. It is, however, assumed that the largest impact on the spread of resistant bacteria was carried out by antibiotic growth promoters that used to be applied to food animals on a large scale [13]. The European Union banned all classes of growth promotion antibiotics in food-producing animals in 2006, restricting antibiotics exclusively for veterinary purposes [17].

Another important contributor to the AMR crisis is the lack of development of new antibiotics. This is caused mainly by economic reasons and regulatory obstacles, like different rules demanded in different countries or changes in regulations during the drug development campaign [16].

The prevalence of antibiotic-resistant bacteria in livestock was observed to decrease by approx. 15% when the use of antibiotics was reduced; the impact of restricting antibiotic use in livestock on the human population, however, remains unclear [18]. Nevertheless, it will not be sufficient to only reduce the consumption of antibiotics in order to stop the AMR prevalence [19]. In humans, strategies like infection prevention, vaccination, early infection detection or appropriate antibiotic use have been shown to be successful in reducing the consequences of AMR infections [20].

## 2.2. β-lactams

β-lactam antibiotics (BLAs) are the most widely used class of antibiotics [21]. All classes of BLAs share a common structural motif: the β-lactam ring, which also serves as the main pharmacophoric group of all BLAs.

BLAs are classified into several groups that differ in the ring fused to the β-lactam nucleus. The main groups are penicillin-like BLAs (in which the partner ring is a thiazolidine), cephalosporins (dihydrothiazine), carbapenems (pyrroline) and monobactams which are monocyclic. The general structures of these BLA groups are shown in Fig. 2.1. Cephalosporins account for the most prescribed BLAs in the US (47.5%). Carbapenems are active against a broad spectrum of bacterial strains, and therefore often used in cases of infections by multidrug-resistant bacteria against which most other antibiotics are not active anymore. Monobactams are only functional against aerobic Gram-negative
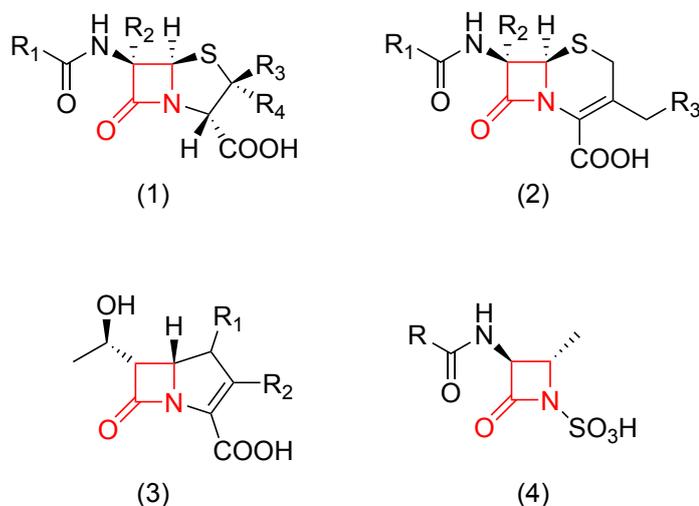
Figure 2.1.: Structures of the main BLA groups: penicillin-like BLAs (1), cephalosporins (2), carbapenems (3), and monobactams (4). The β-lactam ring is shown in red. Adapted from [23] using ChemDraw.

bacterial strains [22].

Several BLAs are produced naturally by molds (penicillins, cephalosporins) and bacteria (carbapenems, monobactams). The very first antibiotic agent, Penicillin G (benzylpenicillin), was discovered by Sir Alexander Fleming in 1928 and first put into use in the 1940s [24]. Since then, many other BLAs have been discovered and developed; their derivatives are prepared in order to broaden the spectrum of action or to modify their pharmacokinetic properties [23]. Novel derivatives include two or more β-lactam rings (either connected with a linker or fused directly together) or covalently linked BLAs with other antibiotics like quinolones, fluoroquinolones or glycopeptides.

The bactericidal effect of BLAs relies on inhibition of the cell wall synthesis. In general, BLAs act against building of the main component of the prokaryotic cell wall— peptidoglycan [24]. Its main function is to provide the cell with mechanical stability, maintain the cell shape and protect the cell from cell lysis. It is composed of linear glycan strands made of *N*-acetylglucosamine and *N*-acetylmuramic acid (linked via β-1,4-glycosidic bonds) that are cross-linked by short peptides. In the nascent peptidoglycan, these are pentapeptides that contain a D-Ala-D-Ala motif; in a mature peptidoglycan, the last D-Ala residue is removed. The carboxyl group of the remaining D-Ala residue reacts with an amino group of residue 3 (diaminopimelic acid or lysine) of another peptide, resulting in cross-linking of the glycan strands [25]. The cross-linking event is catalyzed by transpeptidases that recognize the D-Ala-D-Ala motif.

This is the point where BLAs interfere. Through the structural similarity of the β-lactam ring with the D-Ala-D-Ala motif (see Fig. 2.2), BLAs can bind irreversibly to the transpeptidase active sites. For this reason, these transpeptidases are often called penicillin-binding proteins. Consequently, no transpeptidation reaction can happen, and
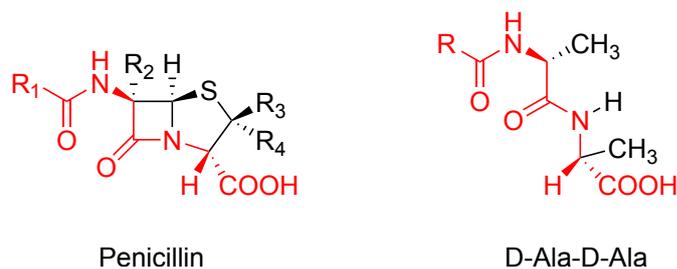
Figure 2.2.: Comparison of a penicillin-like BLA structure and the D-Ala-D-Ala motif. The common portion is shown in red. Adapted from [23] and [24] using ChemDraw.

the glycan strands are only weakly cross-linked, which is finally followed by cell lysis. BLAs are safe for humans and animals, as the animal cells do not contain peptidoglycan nor the penicillin-binding proteins [24].

## 2.3. β-lactamases

β-lactamases are hydrolytic enzymes with the ability to cleave the β-lactam ring, the common structural motif of all BLAs, which is fundamental for their functionality. Production of β-lactamases is the most common mechanism Gram-negative bacteria use against BLAs [22].

β-lactamases are usually classified based on their primary sequence similarity (the Ambler classification system [26]). This system divides β-lactamases into four classes. Classes A, C and D comprise the serine β-lactamases (SBLs), while class B includes all metallo-β-lactamases (MBLs). MBLs are additionally divided into subclasses B1, B2 and B3 which differ in structure, active-site residue composition, zinc stoichiometry or substrates [27]. Alternatively, β-lactamases are classified based on their substrate profile into narrow-spectrum β-lactamases (active against penicillins and early generation cephalosporins), extended-spectrum β-lactamases (active against most BLAs) and carbapenemases [28], or using the Bush-Jacoby-Medeiros system which is based on their substrate and inhibitor profiles [29].

β-lactamases are translated in the cytoplasm as pre-β-lactamases, i.e., nascent proteins that include an N-terminal signal peptide sequence. The signal peptide is responsible for the right cellular localization of the enzyme, and after the enzyme's translocation to the periplasm or anchoring into the outer membrane, the signal peptide is cleaved off by signal peptidase type I or type II, respectively [30]. Most β-lactamases are translocated via the Sec system [28]. This is generally the case of the plasmid-encoded enzymes; enzymes encoded on the bacterial chromosome often use the Tat pathway [28].

The site of action of most β-lactamases is the periplasmic space. Here, the proteins are folded with the help of the periplasmic folding machinery, and MBLs also bind the $Zn^{2+}$ ions that get to the periplasmic space via diffusion through non-selective porins
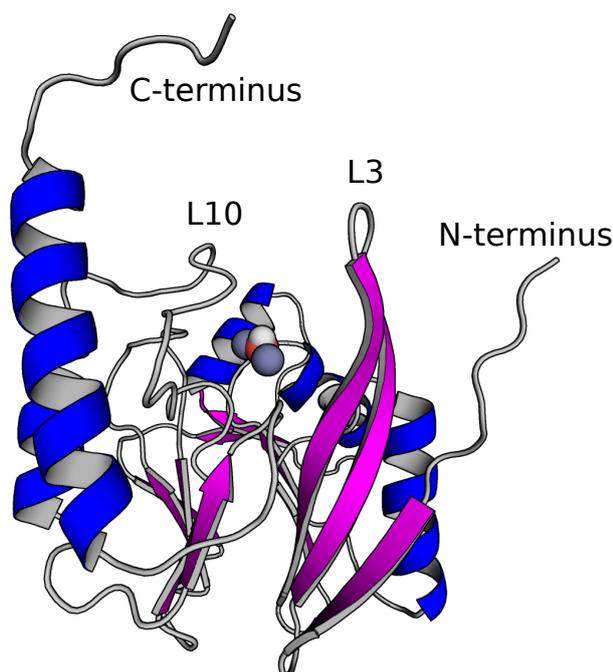
Figure 2.3.: Structure of the IMP-1 protein. Secondary structure elements are shown in color to highlight the αβ/βα fold. $Zn^{2+}$ ions and hydroxide ion, located in the active site of MBLs, are shown in the van der Waals representation.

[30]. Gram-positive bacteria lack the periplasmic space and can therefore either release the enzymes to the extracellular space or produce membrane-anchored β-lactamases [28]. Membrane-anchored enzymes (e.g., NDM-1) are also found in Gram-negative species.

Although the actions of SBLs and MBLs lead to identical results, i.e., hydrolysis of the β-lactam ring, their active site structures and correspondingly their mechanism of action are fundamentally different. These differences suggest distinct evolutionary origins of SBLs and MBLs [31]. SBLs contain a serine residue in their active site, making them structurally similar to the active site of penicillin-binding proteins. The serine residue covalently binds to the carbonyl group of the β-lactam ring, forming an ester intermediate after rupture of the β-lactam ring [22]. The catalytic activity of MBLs depends on the $Zn^{2+}$ ions present in their active site; their mechanism of action is described in more detail in Section 2.3.2.

### 2.3.1. MBL structure

MBLs display an αβ/βα sandwich fold, common to the zinc metallo-hydrolase family [32]. Although the sequence homology of MBLs is not particularly high (only 10% between distant enzymes), the overall structure is highly conserved [33]. As an example, the tertiary structure of IMP-1, a B1 class MBL, is shown in Figure 2.3.

The active site is located between the two β-sheet motifs [34]. The residues constituting the active site (which we will also refer to as first-shell residues) and the number of $Zn^{2+}$ ions differ between the MBL subclasses B1, B2 and B3; this section only focuses on the B1 MBLs.

Enzymes of the B1 subclass contain two binding sites for $Zn^{2+}$ ions—these are called site 1 and site 2 and the corresponding $Zn^{2+}$ ions are referred to as Zn1 and Zn2. Zn1 is coordinated by three histidine residues (site 1 is therefore also called the histidine site). These are generally referred to as His1, His2 and His3 in the following text. Site 2 (also referred to as the DCH[1] site) is constituted by an aspartate (Asp), a cysteine (Cys), and another histidine (His4). The residues of the B1 MBL active site belong to the most prevalent residues to coordinate $Zn^{2+}$ ions [35]. Both Zn1 and Zn2 are involved in the coordination of a water molecule located between them. As the $pK_a$ of this water molecule is lowered by the metal-coordination, it loses a proton and becomes a hydroxide ion [34]. According to QM studies, the bridging water molecule directly transmits its proton to the nearby Asp residue, resulting in a hydroxide ion and an aspartic acid [36]. The cysteine residue has been found in thiolate form [36]. The coordination sphere of Zn2 is typically envisioned to be completed by another water molecule. The resulting coordination geometries are tetrahedral for Zn1 and trigonal bipyramidal for Zn2 [33]. The geometry of the B1 MBL active site is shown in Figure 2.4.

The histidine residues of site 1 are arranged in a pattern commonly observed in zinc metal sites: two histidines (the so-called proximal histidines) are separated by a single residue; the third histidine is more distant in terms of primary structure [35]. The histidine residues coordinate the respective $Zn^{2+}$ ions with the lone pair of electrons of one of their nitrogens: either the δ- or ε-nitrogen. In the terminology of the CHARMM force field, which we will use from now on, these are referred to as ND1 and NE2, and the corresponding tautomeric forms as HSD and HSE. The other nitrogen atom of the imidazole ring is protonated (resulting in a net charge of zero on the histidine residues). Although HSE is the more prevalent tautomeric conformation for the proximal histidines in other zinc-coordinating proteins (the backbone of an HSE is further from the metal site, so there are less steric restrictions), one of the proximal histidine in MBLs (His2) coordinates Zn1 with its ND1 [35].

The Asp residue is conserved in site 2 among all currently known MBLs. Mutagenesis studies have confirmed the assumption that it is critical for the enzymes' function [37]. Earlier suggestions that the protonated Asp serves as a proton donor during the hydrolysis was, however, shown to be false [37]. The currently accepted hypothesis says that Asp positions Zn2 in the right place within the active site to make the catalysis of the hydrolysis of BLAs possible [37], and functions as a general base to aid the formation of the nucleophilic hydroxide [36].

The presence of two $Zn^{2+}$ ions in the active site is a fundamental prerequisite for the functionality of B1 MBLs. The absence of the $Zn^{2+}$ ions results in MBLs with distorted active sites; the overall fold of the protein is, however, not disrupted [33]. As already mentioned, the folding and metal coordination of MBLs take place in the periplasm—the

---

[1]Asp-Cys-His
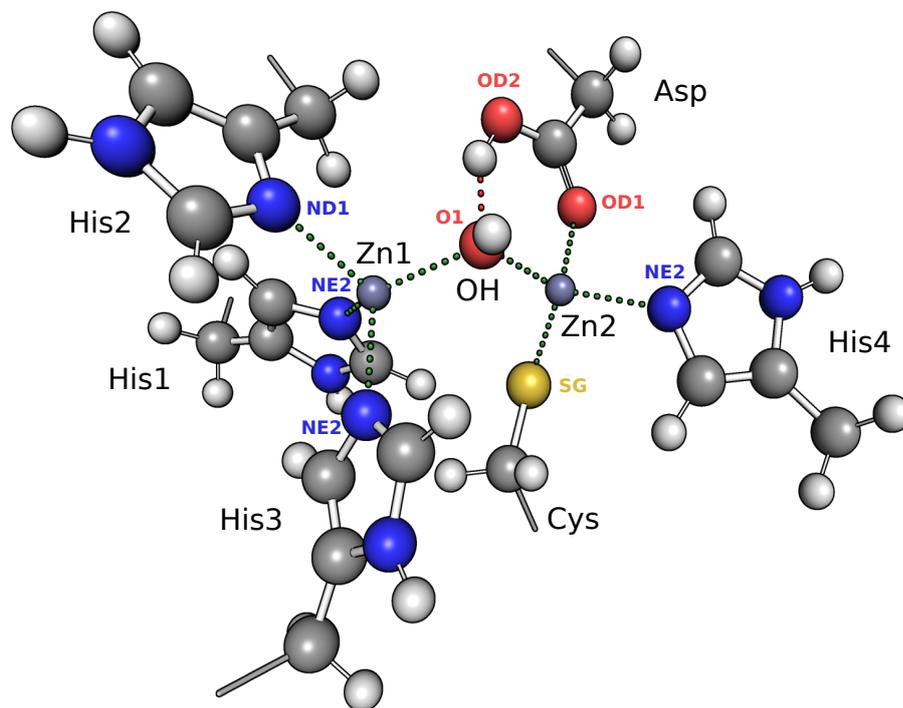
Figure 2.4.: MBLs active site. Residues are labeled in black with the residue names applicable to all B1 MBLs; these residue names are used throughout the thesis (cf. the main text). Atom names of atoms that participate in the coordination of the $Zn^{2+}$ ions are shown in color.

$Zn^{2+}$ availability in the periplasm is therefore essential. To be able to bind Zn2 even in low $Zn^{2+}$ concentration, B1 MBLs use the conserved Cys residue. This is remarkable because Cys is otherwise rarely a ligand for metal ions in oxidizing environments, such as the periplasm [38].

The B1 MBL active site is flanked by two loops: L3 and L10, that assist the catalytic activity of MBLs (see Figure 2.3). The L3 loop is highly flexible and its conformation changes after binding of a substrate or an inhibitor. The residues present in the L3 loop (mostly aromatic ones) contact the substrate substituents via hydrophobic interactions [33]. The L10 residues interact with the β-lactam carbonyl group of some BLAs, which leads to a polarization that facilitates the following nucleophilic attack (see Section 2.3.2).

Apart from the active site, there are other residues conserved among many MBLs. Some of them account for so-called second-shell residues—these are residues in a direct contact with the residues of the first coordination shell. Their contact is realized via hydrogen bonds. Second-shell residues play an important role in the orientation and polarization of the first-shell residues [33].
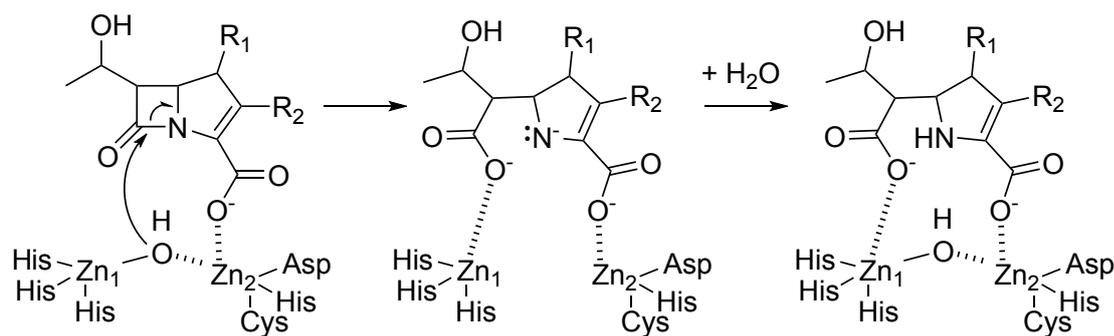
Figure 2.5.: Basic steps of the hydrolysis of a carbapenem by B1 MBLs. Adapted from [33] (Figure 3b) using ChemDraw.

## 2.3.2. MBLs mechanism of action

The hydrolysis performed by MBLs has three main steps: nucleophilic attack on the carbonyl group of the β-lactam ring, cleavage of the C-N bond and nitrogen protonation (see Figure 2.5) [33]. This results in the production of a β-amino acid, which poses no danger to the bacterium [34].

The two $Zn^{2+}$ ions of the bimetallic active site of B1 MBLs have different roles in the catalysis. The main task of Zn1 is to lower the pKa of the water molecule it coordinates. The water molecule therefore gets the character of a hydroxide ion, which serves as a better nucleophile than water [34]. It has also been proposed that Zn1 helps with polarization of the carbonyl group of the β-lactam ring. This may facilitate the nucleophilic attack by the hydroxide.

Zn2 is proposed to bind to the carboxyl group of the BLA substrate. By doing this, it detaches from the bridging hydroxide. This is a necessary event for the catalysis, as a bridging hydroxide would act as a weaker nucleophile than a singly coordinated one [39].

The nucleophilic hydroxide ion is consumed during the catalysis. This poses the need for a new water molecule to be bound by the $Zn^{2+}$ ions. This water molecule was identified as the proton donor in the last catalytic step—the nitrogen protonation. This event is the rate-determining step for all MBLs [33].

## 2.3.3. MBL representatives

The B1 subclass includes the most clinically relevant MBLs. This section introduces three of the most important B1 MBL enzymes: New Delhi MBL (NDM), Verona integron-encoded MBL (VIM) and Imipenemase (IMP), with special focus on the variants NDM-1, VIM-2 and IMP-1. A phylogenetic tree displaying these enzymes in a broader context of the B1 subclass is shown in Figure 2.6.

The first Imipenemase (IMP-1) was detected in Japan in 1991 [44]. It was isolated from a *Serratia marcescens* strain, but since then, IMP-1 has been identified in other Gram-
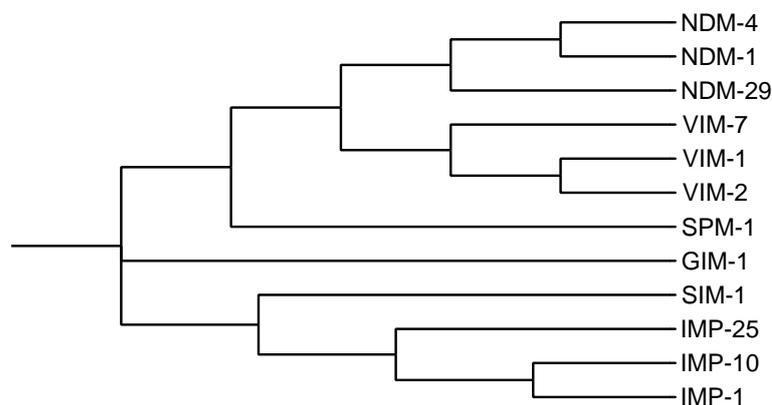
Figure 2.6.: Phylogenetic tree of selected B1 MBLs representatives from the NDM, VIM, IMP, SIM, GIM, and SPM families. The sequences used for the creation of the phylogenetic tree were (name; UniProt ID): NDM-1; C7C422, NDM-4; A0A0G2ST15, NDM-29; A0A2S1ZDP9, VIM-1; A0A0F7KYQ8, VIM-2; Q9K2N0, VIM-7; Q840P9, SPM-1; K7XF88, GIM-1; Q704V1, SIM-1; Q32T02, IMP-1; P52699, IMP-10; Q7DH52, IMP-25; D9I3U0. Sequence alignment was performed using ClustalOmega [40, 41, 42], the phylogenetic tree was created using TreeViewer [43].

negative pathogens like *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* or *Acinetobacter baumannii* [45]. The IMP-1 cluster contains 20 IMP variants, including the clinical variants IMP-6, IMP-10, IMP-25, IMP-30 and IMP-46 [45].

IMP's evolution depends on the exposure to structurally different substrates [46]. A selection pressure applied to the IMP enzymes may result, for example, in a more flexible loop that plays a role in substrate binding. This was shown in IMP-6—the loop's flexibility is increased in comparison to IMP-1 by a single substitution (S262G). As a result, IMP-6 shows a higher catalytic efficiency towards meropenem in comparison to IMP-1 and is also significantly more effective against meropenem than imipenem. In comparison, the $k_{cat}/K_m$ values of IMP-1 against these two antibiotics are similar [46].

Verona integron-encoded MBL is another large group of B1 MBLs. The first variant was detected from a *Pseudomonas aeruginosa* clinical isolate in 1999 in Verona, Italy [47]. VIM-2, even more frequent than VIM-1 [33], was reported shortly after in Marseilles, France [48]. VIM enzymes have been found as a contamination in hospital water environments, in surface water, drinking water and sewage in many countries [49]. The catalytic efficiencies of reactions with antibiotic substrates are comparable among the individual enzyme variants of the VIM group. Their differences are mostly connected to inhibitor profiles, caused by substitutions in the L10 loop [50]. VIM evolution might have also been driven towards enhancing thermal stability.

NDM-1 was isolated from a *Klebsiella pneumoniae* strain in 2009 [51]. NDMs are

most closely related to the VIM family, but generally have a low sequence similarity to other MBLs. The most common mutation among NDMs is the M154L substitution. Remarkably, an analysis of 16 NDM variants has revealed that combinations of mutations in these proteins do not show any epistatic influences [52]. Some mutations also correlate with enhanced thermostability and refolding efficiency [53].

NDMs have disseminated faster and wider than any other MBL family [52] and differ from other MBLs in several aspects. First, they show a strong binding towards penicillin, which is not a typical substrate of MBLs [51]. More importantly, NDMs are lipoproteins anchored to the inner side of the outer membrane of Gram-negative bacteria, while all other known MBLs are soluble periplasmic proteins [52].

The signal peptide of NDM enzymes contains the conserved lipobox sequence LSGC. As described by Gonzalez et al. [54], the lipobox is recognized by the Lol pathway machinery that is responsible for lipidation of the cysteine residue and transport of the protein to the outer membrane. The signal peptide is cleaved off; the lipidated cysteine residue becomes the N-terminal amino acid and constitutes the anchoring point of the whole protein. It has been shown that a substitution of this cysteine residue for an alanine (C26A) leads to production of a soluble protein [54]. There are also positively charged residues on the surface of the protein that interact with the membrane [55]. This interaction also helps define the stable protein orientation, with the active site exposed to the periplasmic space. As lipoproteins, NDM enzymes can also be incorporated into outer membrane vesicles (OMVs) which then provide protection to other bacteria in the vicinity [54].

Membrane-anchoring also plays a role in the protein's stability—the cellular localization helps with stabilization of proteins that have lost their $Zn^{2+}$ ions [52]. NDM-1 was shown to better resist zinc deprivation conditions compared to the soluble VIM-2 and also to the already mentioned C26A NDM-1 mutant [54].

First-shell residues and residues that directly interact with a substrate are conserved among all NDM variants. Interestingly, as long as the zinc concentration is sufficient, mutations observed among the NDM enzymes show no impact on the zinc affinity, substrate profiles or kinetics of the enzyme [53]. Effects of the mutations can be discerned only at too low zinc concentrations, suggesting that an increased resistance to BLAs conferred by certain variants occurs under zinc starvation first [53]. This is an important advantage that protects NDMs against the host immune responses [54]. The host's neutrophils release the metal-chelating protein calprotectin at the infection site in order to remove the $Zn^{2+}$ ions from the MBLs. That usually leads to their degradation by periplasmic proteases [30]. The stabilization of NDM enzymes in zinc deprivation conditions suggests that zinc starvation was an important driver of NDMs evolution [52].

## 2.4. Inhibitors

As described above, infections caused by antibiotic-resistant bacteria are becoming a worldwide threat, and require an appropriate reaction. A possible solution might lie in the development of inhibitors of β-lactamases that are administered to the patient together with antibiotics [22]. The inhibitor itself does not usually display any antibacterial effects,

but it inhibits the action of β-lactamases, allowing the antibiotic agent to act against the infection even when caused by a resistant pathogen.

The first attempts to search for β-lactamases inhibitors (BLIs) were made in the 1970s [21]. These attempts led to the development of the first generation of BLIs, e.g., clavulanic acid, sulbactam and tazobactam, and to their commercialization in the 1980s and 1990s [56]. Clavulanic acid is a natural compound structurally similar to penicillin; the latter two are synthetic penicillanic acid sulfones. These BLIs are used clinically in combination with several antibiotics, e.g., clavulanic acid/amoxicillin or sulbactam/ampicillin [21]. These BLIs contain the β-lactam ring and are only active against SBLs, as the mechanism of inhibition relies on processing of the inhibitor by the β-lactamase and irreversible binding to the active-site serine residue [22].

Next-generation BLIs include avibactam, relebactam or vaborbactam. In contrast to the first-generation BLIs, these compounds do not contain the β-lactam ring structure: avibactam and relebactam are diazabicyclooctanes and vaborbactam is a boronic acid derivative [22]. Their binding to β-lactamases is covalent, yet reversible. These BLIs are all clinically used; however, they are also only active against SBLs.

In other words, there are still no clinically used inhibitors of MBLs. Bahr and co-workers [33] describe several aspects that complicate the design of MBL inhibitors. Unlike the SBLs' mechanism of action, there is no covalent substrate-enzyme intermediate during the MBLs catalysis (cf. Figure 2.5). Another obstacle is the difference between the active site properties of the MBLs subclasses B1, B2 and B3. A fundamental disadvantage of zinc-dependent MBL inhibitors is the danger of off-target actions, which pose potential harm to eukaryotic metalloproteins [33].

Several strategies have been explored in the search for MBL inhibitors. One of the approaches is chelation of the $Zn^{2+}$ ions. This method works *in vitro* and allows specific detection of MBLs in microbiological assays [33]. However, many chelating agents used in such assays, e.g., EDTA, are toxic to eukaryotic cells due to their off-target effects. Therefore, attempts have been made to develop chelating agents with no toxicity against eukaryotic cells. For example, aspergillomarasmine A is a natural compound able to restore meropenem activity in mice infected with a *Klebsiella pneumoniae* strain expressing NDM-1, while showing low toxicity and no effect on blood pressure [57]. A water-soluble tris(2-pyridylmethyl)amine (TPA) derivative, ZN148, is another candidate for a zinc-chelating drug—it was shown to restore the effect of meropenem *in vitro*, while also being nontoxic in an *in vivo* mouse model [58]. Interestingly, MBL activity was restored only by approximately 30% after subsequent zinc addition—this is an indication of an irreversibility in the inhibition mechanism, and also a significant difference to chelation with EDTA, where the enzyme activity is restored to 80% after zinc addition.

Another strategy of MBL inhibition is metal ligation. In this context, thiol-based inhibitors have been widely examined, as the $Zn^{2+}$ ion is thiophilic [33]. An example of a thiol-based MBL inhibitor is captopril. L-captopril is an approved drug to treat hypertension via inhibition of angiotensin-converting enzyme (ACE) [59] and also inhibits certain MBLs. However, D-captopril was shown to inhibit some MBLs, including NDM-1, VIM-2 or IMP-1, better than the L-enantiomer [60]. In both D- and L-captopril, the thiol

group, as a thiolate, displaces the hydroxide ion from the active site by intercalating directly between the $Zn^{2+}$ ions [59]. The hydrophobic portion of the captopril molecule interacts with hydrophobic residues of the L3 loop, while the hydrophilic portion contacts residues of the L10 loop via hydrogen bonds [60].

Other zinc-ligating inhibitors are succinic acid derivatives. They act similarly to captopril—they use one of their carboxylates to replace the hydroxide ion in binding to the $Zn^{2+}$ ions [61]. The other carboxylate was found in the position of the carbonyl group of BLAs during their hydrolysis. The inhibitory effect is supported by the hydrophobic interactions between apolar residues of the MBL and the succinic acid substituents. Succinic acid itself, however, shows no inhibition of MBLs [61].

A particularly promising group of MBL inhibitors are boronates. As mentioned above, the cyclic boronate SBL inhibitor vaborbactam is a clinically approved drug, yet inactive against MBLs. Bicyclic boronates with aromatic substituents were reported as inhibitors of MBLs in 2016 [62]. X-ray crystallography revealed that oxygen atoms of the bicyclic boronate coordinate Zn1 in the same geometry that has been detected in the tetrahedral intermediate of the MBLs' catalytic mechanism (cf. Figure 2.5) [62].

At present, the most promising MBL inhibitor is taniborbactam (VNRX-5133). Taniborbactam, which also contains the bicyclic boronate moiety, can inhibit enzymes from all β-lactamase classes [63]. It has a hydrophobic (cyclohexyl) moiety that interacts with the hydrophobic residues of loops near the active site (which are present in both SBLs and MBLs) and a polar moiety (an ethylene diamine) to facilitate the entry of the inhibitor into a Gram-negative bacterial cell. The polar group does not show any interaction with SBLs; however, X-ray crystallography revealed a hydrogen bond/electrostatic interaction between the polar group and a glutamate in VIM-2 [63]. Taniborbactam is planned to be used in combination with cefepime in cases of complicated urinary tract infections. Venatorx Pharmaceuticals has recently applied for approval of this drug combination by the U.S. Food and Drug Administration (FDA) [64].

# 3. Theory and Methods

## 3.1. Choice of structures

This thesis focuses on NDM-1, VIM-2, and IMP-1 which belong to the most concerning variants of the respective MBL subgroups [65]. As described in Chapter 2, these enzymes were first isolated from bacterial species with frequent AMR: *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Serratia marcescens*, respectively. Notably, the first two mentioned species are members of the ESKAPE list, defined by the WHO in 2017 as a list of the most critical pathogens [66]. An understanding of the MBLs structure and chemical behavior may help to develop novel β-lactamase inhibitors and to unravel the interactions between the MBL enzymes and their inhibitors.

Several studies have shown that despite the highly conserved active site and similar mechanism used by B1 MBLs to hydrolyze β-lactam drugs, individual B1 MBL representatives do not react to inhibitors with the same sensitivity. In fact, IMP-1 was reported to be less susceptible to several inhibitors compared to NDM-1 or VIM-2. For example, the bicyclic boronate inhibitor taniborbactam (cf. Section 2.4) inhibits VIM-2 strongly ($IC_{50} = 0.5$ nM), but IMP-1 to a significantly lower extent ($IC_{50} = 2.5$ μM) [67]. QPX7728, another cyclic boronate inhibitor of MBLs, has also shown a higher $IC_{50}$ in IMP-1 (610±70 nM) compared to NDM-1 and VIM-1 (55±25 nM and 14±4 nM, resp.) [68]. However, such differences have not only been observed with boronates. IMP-1 was also revealed to preserve its hydrolytic activity better than NDM-1 or VIM-2 under treatment with the metal-chelator dipicolinic acid (DPA) [54]. Another zinc–chelating inhibitor, aspergillomarasmine A, was shown less effective against IMP-7 compared to NDM-1 and VIM-2, suggesting its lower potency against IMP enzymes in general [57].

Given that the general structure of the active site of all B1 MBLs is conserved, the lack of response of IMP-1 to some inhibitors compared to other B1 MBLs suggests possible differences in the active site dynamics. In this thesis, we therefore perform two levels of comparative analyses of the B1 MBLs NDM-1, VIM-2, and IMP-1. First, the active site of these enzymes is studied in detail. Apart from searching for differences in the active site properties among the three proteins, a thorough analysis of the active site's structure is necessary to reproduce the active site's character (geometry, coordination of $Zn^{2+}$ ions) during MD simulations compared to the experimental structural data. This is a prerequisite for studying the dynamics of the active site residues. Second, the full proteins are considered, focusing on their mobility and secondary structure stability. In this thesis, only MM calculations were performed; these may serve as the starting point for future QM/MM calculations.

### 3.1.1. Sequence alignment

Sequence alignment allows the comparison of related proteins and serves as a background for further analyses. It reveals the conservation of residues in selected proteins and detects deletions or insertions that have appeared during the evolution of related proteins. Generally, this method can discover protein homology in newly identified protein sequences, predict a newly discovered protein's function, or reveal the common evolutionary origin of known protein sequences (although a lack of sequence similarity does not necessarily eliminate the possibility of a common origin) [69].

In this thesis, sequence alignment of full sequences of NDM-1, VIM-2 and IMP-1 is carried out to illustrate their common evolutionary origin and to get a more in-depth understanding of residue conservation in important regions (primarily the active site) and conservation of secondary structure elements. Most importantly, aligned sequences can reveal the presence or absence of certain structural regions in an individual protein compared to other proteins, which leads to questions about their functional role in the particular protein.

The second purpose of the sequence alignment is to establish a consensus numbering scheme for the employed structures of NDM-1, VIM-2, and IMP-1. This allows a simple, concise way of referring to conserved residues that otherwise possess a different protein-specific residue numbers.

The sequence alignment presented here was performed using ClustalOmega, an algorithm optimized for multiple sequence alignment. The web interface of the UniProt database [70] was used. The full sequence of all three proteins was used for the sequence alignment, including their signal peptides. These are not a part of the biologically active enzymes and are therefore not present in the X-ray structures used as initial coordinates, nor are they reconstructed for the MD simulations.

## 3.2. Molecular dynamics

### 3.2.1. Theoretical background

Molecular dynamics (MD) is a simulation method, permitting, among many other applications, the study of protein dynamics, revealing atomic-level information about the mobility of certain domains, conformational changes, protein folding, or ligand binding [71]. It can also be used to validate experimental results and help interpret them at atomic resolution [72]. Since the first MD simulations of simple gases in the 1950s and of proteins in the 1970s, the application of MD simulations to biomolecules has experienced a rapid growth. This was caused particularly by the increase in computational power, especially the use of general purpose graphics processing units (GPUs) [73]. These days, MD simulations take advantage of developments in X-ray crystallography and cryogenic electron microscopy that can deliver structural data of protein classes that were previously difficult to acquire (e.g., membrane proteins) [71]. Recent developments allow employing Artificial Intelligence (AI) for protein structure prediction—this method is used, e.g., by AlphaFold [74]. The AlphaFold Protein Structure Database currently contains structures

of almost all proteins in the UniProt Database [75].

MD simulations rely on experimental structures of biological macromolecules. Any missing coordinates must be provided. The calculations are started by assigning random initial velocities to all atoms. The coordinates and velocities of the atom positions are propagated as a function of time based on the interatomic forces, and their velocities. The coupled Newtonian equations of motion are solved by numerical integration, typically a variant of the Verlet algorithm [76]. With current computers, simulation lengths of hundreds of nanoseconds to a few microseconds are possible [71]. During these calculations, coordinates for all atom positions are saved at regular time intervals; these series of coordinates as a function of simulation time are referred to as trajectories. A single coordinate set in a trajectory will also be referred to as a frame.

The interactions between atoms and molecules are described by so-called classical mechanical force fields, typically consisting of terms for bonds, angles, dihedral angles and improper dihedral angles (bonded terms), and electrostatic and Lennard-Jones potentials (non-bonded terms).

$$\begin{aligned}
U &= U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} + U_{LJ} + U_{elec} = \\
&= \sum_{bonds} K_b \left(b - b_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \sum_j K_{\phi,j} \left(1 + cos\left(n_j\phi - \delta_j\right)\right) + \\
&+ \sum_{impropers} K_\omega \left(\omega - \omega_0\right)^2 + \frac{1}{2}\sum_{i,j\neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right] + \frac{1}{2}f\sum_{i,j\neq i}\frac{q_iq_j}{r_{ij}}
\end{aligned}$$

$$(3.1)$$

Bond stretching, angle bending, and improper angle bending are described by a force constant ($K_b$, $K_\theta$, $K_{omega}$) and an equilibrium value ($b_0$, $\theta_0$, $\omega_0$) around which the actual bond or angle ($b$, $\theta$, $\omega$) oscillates. Dihedral angles energies as a function of $\phi$ are expressed as a sum of cosine functions, where $k$ is the amplitude, $n$ the frequency, and $\delta$ the phase shift. The electrostatic forces are described by Coulomb's law—the potential depends on the charges of the particles participating in the interaction ($q_i$ and $q_j$) and indirectly on their distance ($r_{ij}$). The Lennard-Jones potential describes the repulsive and attractive forces as a function of distance ($r_{ij}$) between two interacting particles, $i$ and $j$; $\epsilon$ is the minimum of the function and $\sigma$ the interparticle distance at which $U_{LJ} = 0$.

To be able to evaluate MD trajectories in a statistically relevant manner, more than one simulation with the same parameters and time length are usually run. These sets of simulations will be referred to as replicas in the following text.

An important aspect to be considered is that within an MD simulation, no covalent bonds are broken, and no new bonds are formed. In the context of this thesis, this means that the protonation states of amino acids must be decided when preparing the input files (see Section 3.2.2), and do not change throughout the simulation.

## 3.2.2. Preparation of the MD simulations

The X-ray crystallographic structures used as the starting coordinates in the MD simulations were obtained from the Protein Data Bank (PDB), namely PDB IDs 5ZH1 (NDM-1), 4BZ3 (VIM-2), and 5Y5B (IMP-1). These entries were chosen based on their high resolution (1.05 Å, 1.29 Å and 1.70 Å, respectively), satisfying R-free values (0.146, 0.138 and 0.199, respectively) and the absence of organic ligands in the active site.

There are several modifications that need to be applied to the raw PDB files before they can be used in an MD simulation. Some residues may be missing in the PDB file, typically the highly mobile residues at the N- or C-termini of the expressed protein. Furthermore, protein X-ray structures do not usually have enough resolution to reveal the positions of hydrogen atoms. This means that the protonation states of several amino acids—which may affect their charge or non-bonded interactions with other residues—must be evaluated and set up before running an MD simulation.

One of the amino acids with several possible protonation states is histidine: the tautomeric form HSD is protonated at ND1, HSE at NE2. To assign the most likely protonation states to histidine residues in their specific local environment, Protoss[1] [77, 78], a tool for the determination of hydrogen coordinates and amino acid protonation states, was used. The Protoss results were visually inspected and revised in some cases. The protonation states of the active-site histidines were assessed correctly by Protoss (i.e., as described in the literature, see Section 2.3.1). In the case of histidine residues that do not belong to the first-shell residues, the Protoss results were modified if a crystal water molecule was present in the original PDB structure near a particular ND1 or NE2 atom so that it might lead to a hydrogen bond formation (His133 and His228 of NDM-1 and His34 of IMP-1). A summary of the protonation states of all histidine residues in NDM-1, VIM-2, and IMP-1 are provided in Table 3.1.

The original PDB files were modified using CHARMM-GUI Solution Builder[2] [79], a web-based tool allowing preparation of the complete system to be simulated (solvation, periodic boundary conditions) and generation of the input files for the simulations. Only the chain A, the $Zn^{2+}$ ions and crystal water molecules were selected for further processing. The protonation states of the histidine residues were set up as suggested by Protoss and the manual inspection (see Table 3.1). The zinc-coordinating cysteine residue in the active site was set to the deprotonated state (CYM in the CHARMM force field terminology).

There are multiple residues missing in the original PDB files of all three proteins that, however, are part of the biologically active enzymes and therefore must be added to the initial structures used for MD simulations. The coordinates of these residues can be either obtained from tools specialized on modelling secondary and tertiary structures based on the knowledge of the primary sequence or from relevant structural data.

The missing residues of 5Y5B_A (IMP-1, residues 1-2 and 224-228) were modelled by CHARMM-GUI's built-in GalaxyFill program[3]. The PDB file of 4BZ3_A (VIM-2) had to be modified manually, as structural elements from different sources were added to

---

[1]https://proteins.plus; accessed on 13/10/2023
[2]https://charmm-gui.org; accessed on 13/10/2023
[3]https://seoklab.org/GalaxyFill; accessed on 13/10/2023

| 5ZH1_A | 4BZ3_A | 5Y5B_A |
|---|---|---|
| HSE 61 | HSE 56 | HSE 19 |
| | | HSE 34 |
| **HSD 120** | **HSD 114** | **HSD 77** |
| **HSE 122** | **HSE 116** | **HSE 79** |
| HSD 133 | | |
| HSD 159 | HSD 153 | |
| **HSD 189** | **HSD 179** | **HSD 139** |
| HSD 228 | | |
| | HSD 229 | |
| **HSD 250** | **HSD 240** | **HSD 197** |
| HSE 261 | HSE 251 | |
| | HSE 259 | |

Table 3.1.: Protonation states in all histidine residues present in NDM-1 (5ZH1_A), VIM-2 (4BZ3_A) and IMP-1 (5Y5B_A). Sequence alignment is indicated by placing the corresponding histidines on the same line. First-shell histidine residues are shown in bold.

it. First, coordinates for the N-terminal residues 27-31, located after the signal peptide sequence [80], were obtained from a structure predicted by AlphaFold [81, 74], available via the UniProt database. Second, the helical structure of the C-terminal residues 263-266 was adopted from the PDB ID 6V1P_B.

In the case of 5ZH1_A (NDM-1), residues 29 and 270 were modelled by GalaxyFill, accessed via the CHARMM-GUI website. This assumed that the signal peptide of NDM-1 is 28 amino acids long, as stated in the publication reporting the 5ZH1 X-ray crystallographic structure [82] or in the UniProt database (entry C7C422 BLAN1_KLEPN). However, a literature search later revealed that—as described in Section 2.3.3—after lipidation of Cys26, the signal peptide is cleaved off and the lipidated cysteine becomes the first residue of the mature protein [54]. Though it would not make sense to include the lipidated Cys26 in our NDM-1 structure (as we simulate the proteins in aqueous solution), the simulated structure of NDM-1 is missing residues 27 and 28.

All remaining CHARMM-GUI settings were kept at their default values (rectangular water box with a distance of 10.0 Å to the protein edges, neutralization of the protein net charge with potassium and chloride ions to a final concentration of 0.15 M). In the last step, input files for the OpenMM MD engine were generated [83, 84].

There are two more residues that require special treatment. An essential component of the MBL active site is a hydroxide ion that is, however, often missing in the original PDB structures or labeled as a water molecule during the X-ray data post-processing. The coordinates of its atoms (O1 and H1) were derived from QM studies of the MBL active site performed by Diaz et al. [36] (complex 5BH). Selected atoms of the 5BH complex ($Zn^{2+}$ and the N, O and S atoms participating in zinc coordination) were aligned to the corresponding atoms in the structures 5ZH1_A, 4BZ3_A and 5Y5B_A to derive

| coordinated Zn$^{2+}$ ion | residue | NDM-1 | VIM-2 | IMP-1 |
|---|---|---|---|---|
| Zn1 | His1 | HSD 120 | HSD 114 | HSD 77 |
| | His2 | HSE 122 | HSE 116 | HSE 79 |
| | His3 | HSD 189 | HSD 179 | HSD 139 |
| Zn2 | Asp | ASPP 124 | ASPP 118 | ASPP 81 |
| | Cys | CYM 208 | CYM 198 | CYM 158 |
| | His4 | HSD 250 | HSD 240 | HSD 197 |

Table 3.2.: First-shell residues coordinating the Zn1 and Zn2 ions. The "residue" column states the first-shell residue names, as they are used throughout the text. The remaining columns contain the residue names as defined in CHARMM and the protein-specific residue numbers. Out of the active-site components, the hydroxide ion (coordinated by both Zn$^{2+}$ ions) is not shown; this is also referred to as "OH" in the text and has no residue number.

structure-specific OH coordinates.

The other residue is an active-site-located Asp, which has been reported to be in the protonated form by Diaz et al. [36]. The protonation modification of an Asp residue produces a protonated Asp (ASPP in the CHARMM force field terminology) and is achieved by the CHARMM built-in ASPP patch. This patch, however, assigns a proton onto the OD2 oxygen of an ASP residue. In order to reproduce the active site as described by QM and as depicted in Figure 2.4, the identities of the two oxygen atoms (OD1/OD2) were exchanged manually.

The structure of B1 MBL active site was explained in detail in Section 2.3.1. The exact protonation states of first-shell residues set up as described above, including the protein-specific residue numbers, are summed up in Table 3.2.

### 3.2.3. Running the MD simulations

Three sets of MD simulations (called Set 0, Set 1, Set 2) were prepared. The simulations were carried out using OpenMM [85]. The employed force field was CHARMM [86]. The purpose of Set 0 and Set 1 was to examine the behavior of the systems 5ZH1_A, 4BZ3_A and 5Y5B_A during MD simulations. The respective trajectories were analyzed in order to refine the simulation parameters and obtain the final set of trajectories (which used the Set 2 parameters). These are the data which are analyzed as described in Section 3.4. The results of these analyses are presented in Chapter 4. Due to the different rationales behind these three sets, a different number of replicas was run in each case. For each protein, there is only one replica of Set 0 and Set 1, and 5 replicas of Set 2. The exact definitions of the set parameters are described in Section 3.3.

In all sets, the minimization was performed using the steepest descent (50 steps) and subsequently the adopted basis Newton-Raphson (50 steps) algorithms. The system is restrained during equilibration (125,000 steps with a 1 fs time-step; every 5000th step was written out to a DCD file). The simulation length of the production phase was

100 ns for Set 0 and 200 ns for Set 1 and Set 2 (the time-step in all cases was 2 fs). All simulations ran at the constant temperature of 303.15 K with a friction coefficient of 1/ps (both equilibration and production phase) and under a constant pressure, using a Monte Carlo barostat [87, 88]. The Particle Mesh Ewald (PME) method was used to treat electrostatic interactions [89], Lennard-Jones interactions were truncated using the CHARMM force-switching function [90]. The protein atoms, $Zn^{2+}$ ions, and the hydroxide were centered and wrapped in the central water box using MDAnalysis [91, 92] in order to prevent the moving of the protein's center of mass to the "neighboring" water box, leaving the hydroxide atoms in the previous central water box which would cause a systematic mistake in the analyses.

## 3.3. Custom patches

Section 3.2.2 describes basic steps required during system preparation before running an MD simulation. In our case, this means modelling of missing terminal residues, addition of the hydroxide ion to the active site and setting the protonation state of the first-shell residues. This set of modifications will be referred to as Set 0. However, when only the Set 0 modifications are used to run a simulation, the resulting geometry of the active site does not correspond to the X-ray crystallographic data. For illustration, Table 3.3 presents a comparison of selected distances and angles within the active site in the crystal structure 5ZH1_A and the mean values of an MD simulation initiated from the same crystal structure with the Set 0 modifications applied. Only representative examples of distances and angles were selected.

As one can see in Table 3.3, the system Set 0 does not reproduce the coordination of the $Zn^{2+}$ ions by the histidines correctly. Specifically, the mean value of the distance $His1_{NE2}$–Zn1 in the MD simulation is more than 2 Å longer than in the corresponding crystal structure. Moreover, the standard deviation of this distance is high, which reflects the relatively high mobility of this residue. Similarly, there is a significant difference in the resulting values of the $His1_{NE2}$–Zn1–$His2_{ND1}$ angle, which emphasizes the wrong placement of the histidine residues within the active site. His1 manifested the most striking deviation from the X-ray structure; His2, His3 and His4 displayed better results, with the difference of the respective His–Zn distance from the X-ray structure being less than 0.2 Å (data not shown). The mean distances $Asp_{OD1}$–Zn2, $Cys_{SG}$–Zn1 and $Cys_{SG}$–Zn2 are also longer in the MD simulation than in the 5ZH1_A structure. More importantly, the angles representing the orientation of the Asp to Zn2 and the hydroxide show inappropriate mean values and high standard deviations. In the case of the OH–Zn distances, the simulation was unable to reproduce the asymmetry in the hydroxide's position observed in 5ZH1_A in which the hydroxide ion is positioned closer to Zn2.

The excessive distances between the zinc-coordinating residues and the $Zn^{2+}$ ions can be explained by non-bonded forces. In the CHARMM force field, the Lennard-Jones potential values calculated for the relevant distances in the 5ZH1_A structure are 9.96 kcal/mol for $His1_{NE2}$–Zn1 (2.09 Å), 7.36 kcal/mol for $Asp_{OD1}$–Zn2 (1.99 Å) and 21.75 kcal/mol for $Cys_{SG}$–Zn2 (2.28 Å).

*3. Theory and Methods*

|  |  | **5ZH1_A** | **Set 0** | **Set 1** |
|---|---|---|---|---|
| **Distances [Å]** | $His1_{NE2}$–$Zn1$ | 2.09 | 4.32 ± 0.43 | 2.30 ± 0.11 |
|  | $Asp_{OD1}$–$Zn2$ | 1.99 | 2.17 ± 0.09 | 2.24 ± 0.10 |
|  | $Cys_{SG}$–$Zn1$ | 3.91 | 5.00 ± 0.14 | 2.60 ± 0.08 |
|  | $Cys_{SG}$–$Zn2$ | 2.28 | 2.54 ± 0.07 | 3.67 ± 0.10 |
|  | $Asp_{OD2}$–$OH_{O1}$ | 2.80 | 3.11 ± 0.41 | 2.68 ± 0.09 |
|  | $OH_{O1}$–$Zn1$ | 2.03 | 1.89 ± 0.03 | 1.88 ± 0.03 |
|  | $OH_{O1}$–$Zn2$ | 1.95 | 1.89 ± 0.03 | 1.87 ± 0.03 |
| **Angles [°]** | $His1_{NE2}$–$Zn1$–$His2_{ND1}$ | 102.3 | 55.7 ± 5.3 | 95.7 ± 2.8 |
|  | $Cys_{CB}$–$Cys_{SG}$–$Zn2$ | 113.0 | 119.6 ± 10.4 | 117.8 ± 4.9 |
|  | $Asp_{OD1}$–$Zn2$–$Cys_{SG}$ | 110.8 | 91.7 ± 4.9 | 108.0 ± 5.1 |
|  | $Asp_{CG}$–$Asp_{OD2}$–$OH_{O1}$ | 102.8 | 85.9 ± 10.8 | 98.3 ± 4.9 |

Table 3.3.: Comparison of selected distances and angles in NDM-1 X-ray structure 5ZH1_A and MD simulations with the Set 0 and Set 1 modifications, initiated from the same X-ray structure. The simulation length with Set 0 was 100 ns, the length of the Set 1 simulation was 200 ns. For the parameters of Set 0, see Section 3.2.2 and for the parameters of Set 1, Section 3.3. In both MD simulations, 1 replica was run. The presented values were calculated as means and standard deviation of the production phases.

The results of Set 0 imply the need for custom patches that better describe the coordination of $Zn^{2+}$ ions by the first-shell residues, i.e., His–Zn, Cys–Zn and Asp–Zn. Simultaneously, this also provides a means to overcome the repulsive Lennard-Jones forces at the desired distances observed in X-ray crystallography. Application of such patches introduces pseudo-bonds between the $Zn^{2+}$ ions and the coordinating residues and allows control of the corresponding distances and angles by defining the respective force constants and equilibrium values.

The CHARMM force field includes patches that introduce covalent bonds between Cys or His and $Zn^{2+}$ available in the toppar_all36_prot_modify_res.str file. The equilibrium values for these pseudo-bonds are set as 2.32 Å for Cys–Zn bond and 2.07 Å for His–Zn bond [86]. The force constant is 20 kcal/mol.$Å^2$ in both cases. However, there is no patch in the CHARMM force field that defines the coordination of a $Zn^{2+}$ ion by the carboxylic group of an Asp residue. Moreover, our system requires a patch that specifically describes the coordination of a $Zn^{2+}$ ion by the OD1 atom of an ASPP residue. Such a patch was written manually by Åsmund Kaupang. The equilibrium bond length of the pseudo-bond $Asp_{OD1}$–$Zn2$ and equilibrium values of the relevant angles ($Asp_{OD1}$–$Zn2$–$Cys_{SG}$, $Asp_{OD1}$–$Zn2$–$His4_{NE2}$) were defined based on their values in the 5ZH1_A structure. The equilibrium value of the angle $Asp_{CG}$–$Asp_{OD1}$–$Zn2$ was set to 120° with the force constant 20 kcal/mol rad$^2$. The equilibrium value was estimated based on the expected chemical behavior of such an angle. The combination of applying these patches and parameters will be referred to as Set 1 modifications.

In total, the following pseudo-bonds were applied to maintain zinc coordination:

His1$_{NE2}$–Zn1, His2$_{ND1}$–Zn1, His3$_{NE2}$–Zn1, Asp$_{OD1}$–Zn2, Cys$_{SG}$–Zn2, and His4$_{NE2}$–Zn2. The custom patches were incorporated into the input files generated by CHARMM-GUI in an automated way using macha (github.com/akaupang/macha). No patches were implemented to control the OH–Zn distances, as this central property of the active site geometry was reproduced adequately by the non-bonded interactions (cf. distances OH$_{O1}$–Zn1 and OH$_{O1}$–Zn2 in Table 3.3). All parameters of the custom patches with the Set 1 parameters are listed in Appendix A.2.

Introduction of the custom patches with the Set 1 parameters improved some deviating mean values and high standard deviations, such as the distances His1$_{NE2}$–Zn1 or Asp$_{OD2}$–OH$_{O1}$. A similar trend was observed in the case of the presented angles. However, some measurements revealed the need for improvement of the Set 1 parameters—for example, in the treatment of the Cys$_{SG}$–Zn distances. The shortcomings of the Set 1 trajectories were detected after analyzing selected geometrical properties of the system—the selection criteria are described in detail in Section 3.4.2. For a thorough discussion of the improvements made to Set 1 parameters, see Section 3.3.1.

### 3.3.1. Adjustment of the custom patches

Application of the Set 1 modifications creates pseudo-bonds between the coordination partners in the active site, except for the Zn–OH pairs. Based on the results of the MD simulations with the Set1 modifications applied, the parameters of the custom patches can be changed to make the active site geometry even more accurate with respect to the corresponding X-ray structures. Inspection of the trajectories resulting from the Set 1 simulations revealed that such refinement was indeed necessary. The reasons for the introduced changes and the exact modifications—these are referred to as Set 2—are described in the following paragraphs. The custom patches with the Set 2 parameters are listed in Appendix A.3.

**Improper dihedral angles**

The first example of chemically inappropriate behavior observed in the simulations with the Set 1 parameters was the orientation of the zinc-coordinating histidines, or more precisely, their imidazole ring planes. The most striking deviation was detected in the values of the improper dihedral angles that represent the vertical orientation of the imidazole ring plane to the respective $Zn^{2+}$ ion. The second example is the horizontal orientation of the imidazole ring plane to the $Zn^{2+}$ ions, discussed in the next section. For illustration of the terms horizontal and vertical orientation of the imidazole ring plane to the $Zn^{2+}$ ions, see Figure 3.1. The improper angles are generally defined as the angle between planes His$_{NE2}$–His$_{CE1}$–His$_{CD2}$ and His$_{CE1}$–His$_{CD2}$–Zn in the case of HSD (or the angle between planes His$_{ND1}$–His$_{CG}$–His$_{CE1}$ and His$_{CG}$–His$_{CE1}$–Zn in the case of HSE). This definition of improper angles is also expressed as His$_{NE2}$–His$_{CE1}$–His$_{CD2}$–Zn for HSD and as His$_{ND1}$–His$_{CG}$–His$_{CE1}$–Zn for HSE, where the first atom is always the central atom of the improper angle, or shortly as His–Zn improper (e.g. His1–Zn1 for the His1$_{NE2}$–His1$_{CE1}$–His1$_{CD2}$–Zn1 improper angle).
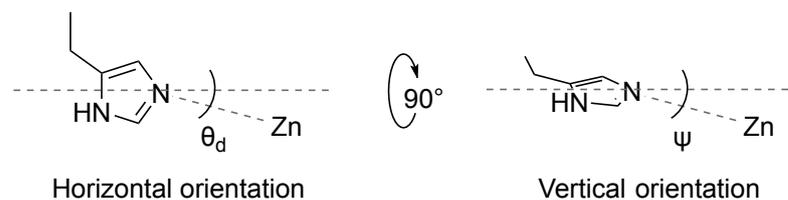
Figure 3.1.: Illustration of the horizontal orientation and vertical orientation of the imidazole ring plane to the $Zn^{2+}$ ions. Horizontal orientation is expressed as the deviation from $Zn^{2+}$ ion position from the axial symmetry of the coordinating imidazole ring ($\vartheta_d$). Vertical orientation is expressed as an improper angle (cf. main text). Created in ChemDraw.
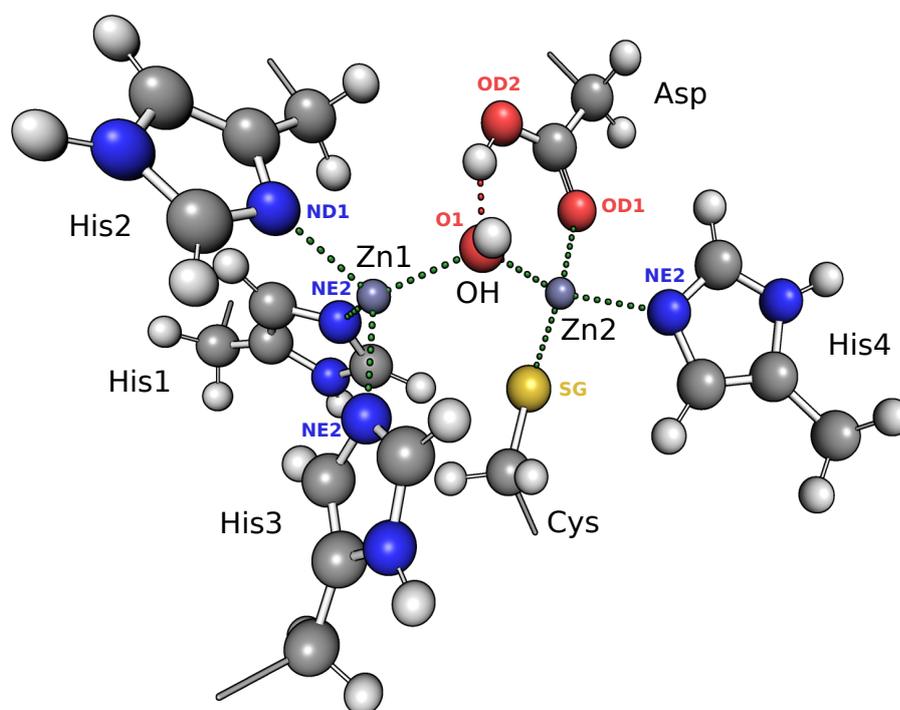


Figure 3.2.: For the reader's better orientation in the Tables 3.3, 3.4 and 3.5, Figure 2.4 is duplicated here. Residues are labeled in black with the residue names applicable to all class B1 MBLs (cf. 2.3.1). Atom names of atoms that participate in the coordination of the $Zn^{2+}$ ions are shown in color.

The improper angles observed in an MD simulation with the Set 1 parameters are listed in Table 3.4. As one sees there, the improper angles acquire diverse values. However, the observed values not only differ strongly from the values measured in the corresponding initial structures, but also do not correspond to the chemical properties of the system. The coordination of the $Zn^{2+}$ ion is maintained via the free electron pair of the NE2/ND1 atom. The $sp^2$ hybrid orbital containing this lone pair is oriented in the plane of the imidazole ring—therefore it is expected that the orientation of the imidazole ring will tend towards planar in respect to the coordinated $Zn^{2+}$ ions. In other words, the improper angles as defined above should have a value around 0°. Values of the relevant improper angles in the X-ray structures indeed do not deviate from the plane by more than 13.5° (cf. Table 3.4).

In order to be able to control the improper angles, improper terms were added to the His–Zn patch. The equilibrium value of the improper term was set to 0° as rationalized above. The force constant was set to 20 kcal/mol radian$^2$. This value was chosen based on a set of shorter simulations and is a compromise between bringing the imidazole rings into the plane of the $Zn^{2+}$ ions while not restricting the flexibility of the system too much. These improper terms are part of the Set 2 modifications.

As the Table 3.4 shows, the values of the improper angles in the MD simulations with the Set 2 parameters improved significantly. The maximal deviation from plane (mean value) is 13.8° in the case of $His3_{NE2}$–$His3_{CE1}$–$His3_{CD2}$–Zn1 in NDM-1, which is comparable to the maximal deviation from plane in the respective X-ray structure. It is, though, important to note that the crystal phase only displays static structures, and that a certain oscillation around the value measured in the crystallographic structures is expected. The Set 2 results were unable to reproduce the pattern of improper angle size in the individual X-ray structures—for example, the most planar improper angle of 5ZH1_A, His3–Zn1, has the highest value in Set 2, the same applies for His4–Zn2 of IMP-1. No significant improvement was observed in the standard deviation when comparing Set 1 and Set 2 results. To be able to compare the standard deviation values of Set 1 (1 replica) and Set 2 (5 replicas), the presented standard deviation for Set 2 is always an average value of the standard deviations of the 5 trajectories.

## Horizontal orientation of the imidazole ring plane to the $Zn^{2+}$ ions

Apart from the improper angles that represent the vertical orientation of the imidazole ring with respect to the $Zn^{2+}$ ion, the horizontal orientation of the ring plane was measured as well. It was determined as the deviation of the $Zn^{2+}$ ion position from the imaginary axis of symmetry of the imidazole ring (see Figure 3.1). The symmetry is determined as the average of angles $His_{CE1}$–$His_{NE2}$–Zn and $His_{CD2}$–$His_{NE2}$–Zn in the case of HSD (or $His_{CE1}$–$His_{ND1}$–Zn and $His_{CG}$–$His_{ND1}$–Zn in the case of HSE) for each His–Zn pair. The deviation from symmetry ($\vartheta_d$) is then calculated as the difference of the first angle ($His_{CE1}$–$His_{NE2/ND1}$–Zn) from this average value.

The data from the X-ray structures show that the $Zn^{2+}$ ions are positioned near to the imaginary axis of symmetry, with the maximal deviation being 3.9°. Although the mean values of the deviation from symmetry in the Set 1 simulations were not as

| | NDM-1 | | | VIM-2 | | | IMP-1 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5ZH1_ A | Set 1 | Set 2 | 4BZ3_ A | Set 1 | Set 2 | 5Y5B_ A | Set 1 | Set 2 |
| **Improper dihedral angle [°]** | | | | | | | | | |
| His1–Zn1 | -10.1 | -17.2 ± 7.6 | -7.5 ± 5.6 | -10.9 | -24.3 ± 6.8 | -9.9 ± 5.9 | 3.0 | -16.3 ± 9.2 | -1.1 ± 6.0 |
| His2–Zn1 | -8.8 | -44.8 ± 7.3 | -0.2 ± 8.1 | -7.5 | -45.3 ± 6.8 | -3.0 ± 8.7 | -7.0 | -50.5 ± 6.2 | 1.9 ± 8.1 |
| His3–Zn1 | 4.0 | 28.8 ± 5.6 | 13.8 ± 6.8 | 7.2 | 32.8 ± 5.7 | 12.3 ± 6.7 | -2.8 | 39.5 ± 5.7 | 6.1 ± 5.7 |
| His4–Zn2 | 13.5 | -39.7 ± 5.8 | -1.5 ± 6.6 | 6.7 | -39.6 ± 5.3 | -6.4 ± 8.1 | -0.5 | -42.2 ± 6.2 | -9.1 ± 6.3 |
| **Deviation from symmetry [°]** | | | | | | | | | |
| His1–Zn1 | 1.4 | -3.3 | 0.7 | -0.7 | -7.2 | -0.1 | 3.9 | -7.0 | 3.1 |
| His2–Zn1 | 1.1 | -6.4 | -3.6 | -0.9 | -5.6 | -0.7 | -0.1 | -4.6 | -0.2 |
| His3–Zn1 | 3.6 | -2.7 | -0.1 | 2.1 | -2.8 | -1.4 | 1.4 | -0.6 | -2.4 |
| His4–Zn2 | -2.8 | 6.0 | 7.9 | 5.0 | 2.6 | 8.3 | 3.7 | -1.8 | 8.2 |

Table 3.4.: Deviation from symmetry (horizontal orientation) and improper dihedral angles (vertical orientation) of the His1–4 imidazole ring planes to the coordinated $Zn^{2+}$ ions in the original X-ray structures and MD simulations. The average values (and standard deviations in the case of the improper dihedral angles) of Set 1 (1 replica) and Set 2 (5 replicas) are compared with the corresponding values of the X-ray structures 5ZH1_A, 4BZ3_A, 5Y5B_A. The horizontal orientation of each His–Zn pair was calculated as the difference between $His_{CE1}$–$His_{NE2/ND1}$–Zn and the mean value of the $His_{CE1}$–$His_{NE2/ND1}$–Zn and $His_{CD2/CG}$–$His_{NE2/ND1}$–Zn angles' size. The vertical orientation is represented as the improper dihedral angles $His_{NE2}$–$His_{CE1}$–$His_{CD2}$–Zn (for HSD) or as $His_{ND1}$–$His_{CE1}$–$His_{CG}$–Zn (for HSE) with the central atom being NE2/ND1.

inappropriate as the values of the improper angles, they still improved in most cases in Set 2, compared to Set 1. This was achieved by introducing changes in the relevant angle terms which affected the equilibrium values of the $His_{CE1}$–$His_{NE1/ND1}$–Zn and $His_{CD2/CG}$–$His_{NE2/ND1}$–Zn angles that were modified from their original values of 120.65° and 125.98°, to 126° and 123°, resp. The new values are average values of the two sets of angles in the X-ray structures of NDM-1, VIM-2 and IMP-1 enzymes without any ligands bound in the active site (see the list of these structures in Appendix A.1). The force constant (20 kcal/mol.radian$^2$) was the same in the Set 1 and Set 2 simulations. Interestingly, the deviation from symmetry in the His4–Zn2 coordination was increased with Set 2, compared to Set 1.

### Non-bonded fix for Cys–Zn1

Another serious problem with the Set 1 trajectories was observed for the distances $Cys_{SG}$–Zn1 and $Cys_{SG}$–Zn2 (see Table 3.5). A significant worsening of these distances, resulting in distortion of the whole active site, was observed in all 3 enzymes because of the introduction of the custom patches, specifically the Cys–Zn patch (cf. Table 3.3). The formation of the $Cys_{SG}$–Zn2 pseudo-bond resulted in disregarding of the non-bonded parameters of the $Cys_{SG}$–Zn2 pair. Specifically, the loss of electrostatic forces between the thiolate group and the Zn2 ion led to a stronger attraction of the thiolate to the other $Zn^{2+}$ ion, Zn1. This caused the $Cys_{SG}$–Zn1 distance to be more than 1 Å shorter and $Cys_{SG}$–Zn2 more than 1 Å longer in the MD simulation than in the NDM-1 and VIM-2 X-ray structures. Interestingly, the IMP-1 simulation with the Set 1 parameters equilibrated to 2.82 Å, but relaxed to the mean value of 4.56 Å within a few frames, resulting in both $Cys_{SG}$–Zn1 and $Cys_{SG}$–Zn2 distances being about 0.8 Å too long.

These observations imply the necessity of a non-bonded fix[4] for the $Cys_{SG}$–Zn pair not bound by the pseudo-bond, i.e. $Cys_{SG}$–Zn1. The default $R_{min}$ for a SG–Zn pair (3.29 Å) used in Set 1 simulations was changed to 4.50 Å in Set 2. This value was chosen based on results of several short MD simulations with different $R_{min}$ estimates. Additionally, the force constant of the $Cys_{SG}$–Zn2 pseudo-bond, defined within the Cys–Zn patch, was increased from 20 kcal/mol Å$^2$ to 200 kcal/mol Å$^2$. These modifications led to improved values of the distances $Cys_{SG}$–Zn1 and $Cys_{SG}$–Zn2 in all 3 enzymes—the mean values did not deviate from the respective X-ray structures by more than 0.26 Å.

### Equilibrium angles

The Set 2 parameters also include new equilibrium angles values, which are based on the mean angle values of X-ray structures of all 3 enzymes with no ligands bound in the active site. The main improvement was observed in the case of the $Asp_{CG}$–$Asp_{OD1}$–Zn2 angle (see Table 3.5). Its equilibrium value was modified from 120° to 138°, which led to more accurate, yet still lower values with respect to the X-ray data. Other changes were made in order to improve the correctness of the custom patches, and did not cause considerable

---

[4]In the CHARMM family of force fields, a so-called non-bonded fix overrides the mixing rules and defines Lennard-Jones parameters for a specific atom pair.

| | | NDM-1 | | | VIM-2 | | | IMP-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 5ZH1 A | Set 1 | Set 2 | 4BZ3 A | Set 1 | Set 2 | 5Y5B A | Set 1 | Set 2 |
| **Distances [Å]** | Zn1–Zn2 | 3.41 | 3.51 ± 0.06 | 3.57 ± 0.07 | 3.51 | 3.52 ± 0.06 | 3.60 ± 0.08 | 3.44 | 3.63 ± 0.06 | 3.53 ± 0.07 |
| | AspOD1–Zn2 | 1.99 | 2.24 ± 0.10 | 1.82 ± 0.10 | 2.20 | 2.25 ± 0.09 | 1.84 ± 0.10 | 2.08 | 2.03 ± 0.11 | 1.85 ± 0.10 |
| | CysSG–Zn1 | 3.91 | 2.60 ± 0.08 | 4.17 ± 0.19 | 4.25 | 2.63 ± 0.09 | 4.17 ± 0.25 | 3.75 | 4.56 ± 0.19 | 4.00 ± 0.18 |
| | CysSG–Zn2 | 2.28 | 3.67 ± 0.10 | 2.42 ± 0.04 | 2.30 | 3.67 ± 0.11 | 2.42 ± 0.04 | 2.37 | 3.24 ± 0.11 | 2.43 ± 0.04 |
| | AspHD2–OHO1 | n/a | 3.50 ± 0.12 | 1.68 ± 0.13 | n/a | 3.52 ± 0.11 | 1.65 ± 0.11 | n/a | 3.52 ± 0.15 | 1.73 ± 0.22 |
| **Angles [°]** | AspCG–AspOD1–Zn2 | 128.8 | 83.9 ± 3.5 | 120.7 ± 4.0 | 136.2 | 83.9 ± 3.4 | 121.4 ± 4.0 | 139.0 | 120.9 ± 3.9 | 120.9 ± 3.9 |
| | AspOD1–Zn2–CysSG | 110.8 | 108.0 ± 5.1 | 80.9 ± 5.1 | 100.3 | 102.4 ± 5.0 | 78.2 ± 4.8 | 98.9 | 75.3 ± 6.9 | 83.5 ± 4.8 |
| | AspOD1–Zn2–His4NE2 | 97.8 | 54.3 ± 3.8 | 66.8 ± 4.4 | 83.6 | 52.5 ± 3.7 | 65.3 ± 4.7 | 88.8 | 58.0 ± 5.6 | 69.5 ± 4.9 |
| | CysSG–Zn2–OHO1 | 115.52 | 54.9 ± 3.1 | 95.0 ± 4.1 | 115.7 | 55.1 ± 3.1 | 92.4 ± 4.3 | 92.5 | 73.3 ± 5.7 | 92.6 ± 3.6 |
| | AspOD2–AspHD2–OHO1 | n/a | 26.0 ± 5.2 | 150.7 ± 10.4 | n/a | 26.3 ± 4.8 | 154.4 ± 9.6 | n/a | 34.3 ± 5.4 | 147.3 ± 15.7 |

Table 3.5.: Comparison of selected distances and angles resulting from the Set1 and Set2 simulations. All entries have the format average ± standard deviation. For Set1, the standard deviation was derived from a single replica. For Set2, standard deviations were calculated for each trajectory and then averaged. If a standard deviation was lower than 0.05, then its value is listed as 0.0. The Set1 and Set2 results of each protein are compared with the corresponding values of the original X-ray structures (taken from chain A of PDB codes 5ZH1, 4BZ3 and 5Y5B, respectively) after step 1 of the CHARMM-GUI Solution Builder procedure and the addition of OH residue's coordinates.

differences between the Set 1 and Set 2 results in most cases. A significant difference is also observed in the $Cys_{SG}$–Zn2–$OH_{O1}$ angle, which is, however, not directly included in the custom patches. Interestingly, the IMP-1's value of this angle (92.5°) is significantly lower than in NDM-1 and VIM-2 (115.5° and 115.7°, resp.), and was also reached by the Set 2 simulations of IMP-1 in contrast to what was observed in the simulations of the other two proteins.

Another important difference between the simulations with the Set 1 and Set 2 parameters is the significant shortening of the $Asp_{HD2}$–$OH_{O1}$ distance and a significant change of the $Asp_{OD2}$–$Asp_{HD2}$–$OH_{O1}$ angle. In this case, the values measured in the MD simulations cannot be compared to X-ray crystallographic data, as the X-ray structures do not contain information about the position of the HD2 atom. Despite this, we perceive the results of Set 2 as an improvement from Set 1, as the Set 2 values correspond better to the assumed hydrogen bond between $Asp_{HD2}$ and $OH_{O1}$ (cf. Section 2.3.1 and Figure 2.4).

## 3.4. Theoretical background on analysis methods

This section provides the theoretical background and rationale behind the analysis methods used in this master's thesis. The analyses described below were only applied to trajectories obtained from MD simulations run with the Set 2 parameters (cf. Section 3.3.1). However, the measurement of a selection of geometrical properties described in Section 3.4.2 was employed to examine the trajectories of the Set 0 and Set 1 simulations and assisted in the search for parameters which would deliver more accurate results than the parameters of Set 0 and Set 1.

### 3.4.1. RMSD

Root-mean-square deviation (RMSD) is a function of distance differences between two coordinate sets over time. Here, the two coordinate sets are two protein structures that are first superimposed and then the distances between the corresponding atoms are calculated and averaged.

$$RMSD\ (t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \mathbf{x}_i(t) - \mathbf{x}_i^{\mathrm{ref}} \right)^2} \tag{3.2}$$

N is the number of atoms. The protein structure of frame 0 serves as the reference (its positions are the $x_i^{ref}$) against which the deviation of the structures in the other frames ($x_i$) is determined.

RMSD is a common measurement applied to MD trajectories in order to observe if a structure remains stable throughout the simulation. The RMSD function should reach a stable plateau. A continuous increase of the RMSD as the simulation progresses means that the protein fold deviates more and more from the (experimental) starting structure and might be an indication of a need for adjustment of the simulation parameters. Sudden

changes in the RMSD curve may imply that the protein structure is unfolding or switches to a different conformation.

The built-in RMSD function of MDAnalysis (MDAnalysis.analysis.rms module) with the default keyword arguments settings (no particle weighting) was used to calculate RMSD values for each frame of 5 replicas of each protein.

## 3.4.2. Geometrical properties

Distances, angles and improper dihedral angles within the active site were measured in order to monitor the reproduction of the X-ray crystallographic data from the PDB. These measurements also allow the comparison of the influence of the same parameters applied to NDM-1, VIM-2, and IMP-1, searching for possible differences in the active site dynamics. The MDAnalysis library was used to calculate the selected geometrical properties.

The measured distances were selected based on several criteria. First, Zn1–Zn2 was measured as an illustration of the active site width. The lengths of the pseudo-bonds defined by custom patches of zinc coordination ($His1_{NE2}$–Zn1, $His2_{ND1}$–Zn1, $His3_{NE2}$–Zn1, $His4_{NE2}$–Zn2, $Asp_{OD1}$–Zn2, $Cys_{SG}$–Zn2, cf. Section 3.3) were monitored as well. The $Cys_{SG}$–Zn1 distance was measured to monitor the effect of the introduced non-bonded fix. $Asp_{OD2}$–$OH_{O1}$ was selected as a distance demonstrating the behavior of the central part of the active site—this distance was chosen since it is not possible to compare the $Asp_{HD2}$–$OH_{O1}$ hydrogen bond with X-ray data, as there is not enough resolution for localization of the hydrogen. It is also not possible to compare OH–Zn distances with the X-ray data, as the hydroxide's coordinates were added to the initial structure by us based on the results of QM calculations [36].

Several angles were measured in order to describe the horizontal orientation of the imidazole ring planes of the first-shell histidine residues to their $Zn^{2+}$ ions. These are listed in Section 3.3.1. Additional angles were measured to monitor, e.g., the orientation of the active-site Asp and Cys residues to Zn2 ($Asp_{CG}$–$Asp_{OD1}$–Zn2, $Cys_{CB}$–$Cys_{SG}$–Zn2), the orientation of the Asp to the hydroxide ($Asp_{CG}$–$Asp_{OD2}$–$OH_{O1}$), and the respective positioning of the $Zn^{2+}$ ions and the hydroxide ion (Zn1–$OH_{O1}$–Zn2).

In order to inspect the deviation of our results from the initial structures, mean values of the selected geometrical properties found in X-ray structures of NDM-1 (22 structures), VIM-2 (9 structures), and IMP-1 (13 structures) were compared to the results of our MD simulations. Data from these X-ray structures were collected by Åsmund Kaupang. The standard deviation of the X-ray data reflects the variation between different X-ray structures of the same protein. Only PDB entries without any organic ligands bound to the active site were used; their list is presented in Appendix A.1. Contrary to the data presented in Tables 3.3, 3.4 and 3.5, the standard deviation of these angles represents the deviation of the individual replicas (i.e., the standard deviation of the mean values of the 5 replicas of Set 2 trajectories).

### 3.4.3. Zn²⁺ ion coordination

AFICS (Analysis of the First Ion Coordination Sphere) is a Python tool for analyzing the geometry of ion coordination over the course of an MD trajectory. It provides, besides other things, information about the deviation from an ideal polyhedron of a selected coordination number, quantified as an RMSD value [93].

In our case, AFICS was used to determine the number of atoms coordinating the $Zn^{2+}$ ions (O and N for Zn1 and S, O and N for Zn2) within 2.70 Å. The atoms were selected based on possible coordination partners of the respective $Zn^{2+}$ ion, i.e., the first-shell residues (cf. Section 2.3.1) and water molecules. The threshold value of 2.70 Å was chosen based on the radial distribution function that measures average counts of selected atom types within all distances up to 6 Å. The atom selection was updated every frame.

The positions of all atoms of interest present within the threshold distance build an imaginary polyhedron. For each trajectory frame, this polyhedron is superpositioned to an ideal polyhedron of the same number of vertices—or, in the AFICS terminology, the coordination number (CN). The deviation (RMSD) of the actual coordination sphere from the possible ideal polyhedra of the given coordination number is calculated according to Equation 3.3.

$$RMSD_{AFICS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a_i - b_i)^2} \tag{3.3}$$

$N$ is the number of vertices and $a$ and $b$ are the spatial positions of the ideal and measured polyhedra. In our application, CNs 4, 5 and 6 are considered. These CNs correspond to square planar or tetrahedral polyhedra (CN = 4), square pyramidal or trigonal bipyramidal polyhedra (CN = 5) and octahedral or trigonal prismatic polyhedra (CN = 6).

### 3.4.4. RMSF and secondary structure

The root-mean-square fluctuation (RMSF) provides information about the flexibility of individual residues within a protein. Thereby, it reveals which regions of the protein are, for example, more mobile than their surroundings. This can lead to the establishment of hypotheses about the function of these mobile regions. In our case, the Cα atoms were considered as sufficient representatives of the flexibility of each residue within the structure, and so only these atoms were used in the RMSF calculation.

$$RMSF_{C\alpha} = \sqrt{\langle (x_{C\alpha} - \langle x_{C\alpha} \rangle)^2 \rangle} \tag{3.4}$$

In order to determine the RMSF of the Cα atoms within a trajectory, it is necessary to perform a structural alignment of the individual frames to remove the translational and rotational movement of the whole protein from the RMSF calculation. Two methods of trajectory alignment were compared: the least-squares superposition used by the MDAnalysis.analysis.align module [91] and maximum likelihood superposition used by

Theseus[5] [94]. Use of the latter alignment method led to lower RMSF values for residues with low mobility—the RMSF values of highly mobile residues were comparable—and this alignment method was therefore chosen for the analysis. After the alignment, the MDanalysis.analysis.rms.RMSF module was used for the RMSF calculation itself. As the RMSF function is time-independent, the trajectories of the five replicas of each protein were concatenated and the RMSFs of the concatenated trajectories were determined. This was done primarily to perform the alignment on the whole set of data, which minimizes the possible errors in aligning each trajectory separately.

The aim of the RMSF measurement was not to get the RMSF values of the residues of each protein as such, but to compare the mobility of the corresponding residues in the three studied proteins, NDM-1, VIM-2, and IMP-1. The obtained RMSF results were therefore aligned based on the consensus sequence alignment of the three proteins. Additionally, the secondary structures of the respective X-ray structures were compared to examine both their conservation (using the sequence alignment) and the mobility of their secondary structure elements (using the RMSF calculation). The maintenance of the secondary structure elements during the MD simulations were monitored using the VMD plugin Timeline [95] which employs the STRIDE algorithm [96] for the secondary structure prediction.

### 3.4.5. Water molecules in the active site

The number of water molecules present in the active site of each protein was counted over the trajectory. This measurement allows to determine the openness and solvent accessibility of the active site of each protein. The active site was approximated as a sphere, with the center in the midpoint between the $Zn^{2+}$ ions. Several radii of the active site approximation were examined: from 3 Å (approximate distance from the center of the sphere to the zinc-coordinating atoms) to 9 Å (approximate distance to Cα atoms of the zinc-coordinating residues).

---

[5]https://theobald.brandeis.edu/theseus; accessed on 13/10/2023

# 4. Results

All analyses described in this chapter were performed on trajectories obtained from the MD simulations with the Set 2 parameters applied. For a description of these parameters and how they were derived, see Section 3.3.1. These simulations were always carried out in 5 replicas; the Set 2 results always refer to mean values of these replicas. Only the trajectories of the production phase were analyzed.

## 4.1. Sequence alignment

The full sequence alignment is shown in Figure 4.1. Signal peptide residues were included in the sequence alignment, although they are not present in the X-ray crystallographic structures of the biologically active proteins used as starting structures for the MD simulations, as discussed in section 2.3.1. Nevertheless, they also evidence a certain degree of homology.

Residues are numbered differently in the three PDB structures used. In the case of VIM-2, the signal peptide is included in the protein-specific numbering: 4BZ3_A starts with residue number 27. The structure of NDM-1 used as a basis for the simulation (5ZH1_A) starts with residue number 29, although it would be more correct if the residues 27 and 28 were included in the initial structure as well, as discussed in Section 3.2.2. On the other hand, in the PDB structure 5Y5B_A (IMP-1), residue number 1 is the first residue after signal peptide cleavage.

To be able to use the sequence alignment in further analyses (especially the RMSF analysis of the sequence-aligned structures), it was necessary to establish a consensus numbering scheme for these three proteins. In the consensus numbering scheme, residue 1 is the first residue of the structure of NDM-1 used in the simulations, residue 255 is the last residue of the IMP-1 sequence. The consensus numbering marks each residue that appeared in the sequence alignment in a continuous way, i.e., deletions in regions where at least one of the other proteins has a defined sequence are numbered in the consensus numbering scheme (although, of course, these deletions are not given a number in the protein-specific numbering schemes).

The sequence alignment analysis revealed a significant similarity between the studied proteins. Sequence similarity scores of protein pairs were established by ClustalOmega as follows: 34% of the residues for the NDM-1–VIM-2 pair, 34% for NDM-1–IMP-1, and 32% for VIM-2–IMP-1 are identical (including the signal peptides). When only the sequences of the biologically active proteins are considered, comparable results are obtained: 36% for the NDM-1–VIM-2 pair, 35% for NDM-1–IMP-1 and 32% for VIM-2–IMP-1. These numbers are in line with the fact that these proteins are homologous [69]. The longest

```
                                              10          20          30
          ─────────────────────────────────────────────────────────────────
                8        18        28        38        48        58
NDM-1     MELPNIMH  PVAKLSTALA  AALMLSGCMP  GEIRPTIGQQ  METGDQRFGD  LVFRQLAPNV
                3        13        23        33        43        53
VIM-2          MFK  LLSKLLVYLT  ASIMAIASPL  AFSVDSSGEY  PTVSEIPVGE  VRLYQIADGV
                                                          6          16
IMP-1      .MSKLSVFFI  F....LFCSI  ..........  .ATAAESLPD  LKIEKLDEGV


                                40        50        60        70        80        90
          ─────────────────────────────────────────────────────────────────
               68        78        88        98       108       118
NDM-1     WQHTSYLDMP  GFGAVASNGL  IVRDGGRVLV  VDTAWTDDQT  AQILNWIKQE  INLPVALAVV
               63        72        82        92       102       112
VIM-2     WSHIATQSFD  G.AVYPSNGL  IVRDGDELLL  IDTAWGAKNT  AALLAEIEKQ  IGLPVTRAVS
               26        36        46        56                   75
IMP-1     YVHTSFEEVN  GWGVVPKHGL  VVLVNAEAYL  IDTPFTAKDT  EKLVTWFVE.  RGYKIKGSIS


          * * *  100          110       120       130       140       150
          ─────────────────────────────────────────────────────────────────
              128       138       148       158       168       178
NDM-1     THAHQDKMGG  MDALHAAGIA  TYANALSNQL  APQEGMVAAQ  HSLTFAANGW  VEPATAPNFG
              122       132       142       152       158       168
VIM-2     THFHDDRVGG  VDVLRAAGVA  TYASPSTRRL  AEVEGNEIPT  HSL....EGL  SSSGDAVRFG
               85        95       105       115                 126
IMP-1     SHFHSDSTGG  IEWLNSRSIP  TYASELTNEL  LKKDGKVQAT  NSFSGV....  .....NYWLV


                       160    *    170       180    *    190       200       210
          ─────────────────────────────────────────────────────────────────
              186       196       206       216       226       236
NDM-1     ..PLKVFYPG  PGHTSDNITV  GIDGTDIAFG  GCLIKDSKAK  SLGNLGDADT  EHYAASARAF
              176       186       196       206       216       226
VIM-2     ..PVELFYPG  AAHSTDNLVV  YVPSASVLYG  GCAIYELSRT  SAGNVADADL  AEWPTSIERI
              136       146       156       163       173       183
IMP-1     KNKIEVFYPG  PGHTPDNVVV  WLPERKILFG  GCFIKP...Y  GLGNLGDANI  EAWPKSAKLL


                       220    *    230       240       250
          ─────────────────────────────────────────────────────────────────
              246       256       266
NDM-1     GAAFPKASMI  VMSHSAPDSR  AAITHTARMA  DKLR......  .....
              236       246       256       266
VIM-2     QQHYPEAQFV  IPGHGLPGGL  DLLKHTTNVV  KAHTNRSVVE  .....
              193       203       213       223
IMP-1     KSKYGKAKLV  VPSHSEVGDA  SLLKLTLEQA  VKGLNESKKP  SKPSN
```

Figure 4.1.: Sequence alignment of NDM-1, VIM-2, and IMP-1. The consensus numbering starts with the first residue of the NDM-1 protein after signal peptide cleavage. Protein-specific residue numbers (shown above each sequence) are identical to the original PDB structures 5ZH1 (NDM-1), 4BZ3 (VIM-2) and 5Y5B (IMP-1). The consensus numbering of the sequence-aligned structures is shown above all sequences, separated by a line. Identical residues in at least two of the structures are shown in bold. Asterisks mark the active site residues.

stretches of identical residues appeared between NDM-1 and the other proteins (a stretch of 9 identical residues with VIM-2, residues 47-55 using the consensus numbering, and also 9 identical residues with IMP-1, residues 156-164).

All active site residues (marked with an asterisk in Figure 4.1) are conserved. Overall, there are 41 residues conserved across all three sequences of the mature proteins out of 255 sequence positions (the length of the sequence alignment)—these are shown in bold in the Fig. 4.1. Some conserved residues appear in structural patches that seem to be important for maintaining the protein fold. For example, residues 30, 33, 49, 50 and 52 (consensus numbering) belong to a β-sheet system. A complete β-sheet is conserved in residues 111-113. Residues 157-160 build an aromatic patch with a phenylalanine, a tyrosine, followed by a kink made of a proline and a glycine. Frequently, a pattern occurs in the alignment in which two proteins share an identical residue, followed by a residue shared by a different pair of proteins, e.g., residues 43-45, 59-61 or 73-74.

## 4.2. RMSD

An RMSD calculation is a standard measure to monitor the stability of simulated protein structures over the trajectory. Here, the starting structure (frame 0) is used as the reference on which each following frame is superimposed. The more a region of a protein deviates in a specific frame from the reference structure (frame 0), the higher the RMSD of this frame. Only the backbone atoms (Cα, C, O, N) are included in the RMSD calculation, as the movements of the amino acid side chains are not relevant for the stability of the overall protein fold during the simulation.

The results of the RMSD calculations for each replica of the Set 2 simulations are presented in Figure 4.2. After 200 ns, the RMSD curve is below 2 Å in all replicas of the three proteins. Values larger than 2 Å are rare, e.g., NDM-1, rep.2 and IMP-1, rep.2. As the purpose of the RMSD measurement is to reflect the overall conformational stability of the protein, it is not necessary to include all residues—in our case, the most relevant regions of the proteins are their cores, not their highly mobile termini. Therefore, the backbone atoms of residues with an RMSF > 1 in at least one of the protein's average RMSF value (5 replicas) were excluded from the RMSD calculation (cf. Figure 4.4). This applies to residues 1–20 and 241–255 of the consensus numbering scheme of all replicas of the three proteins.

The RMSD value of the last frame averaged over the five replicas for each protein is 1.28 Å for NDM-1, 1.27 Å for VIM-2, and 1.11 Å for IMP-1. These numbers indicate that our MD simulations remain stable for the total simulation time of 200 ns per individual simulation, with no sudden changes of RMSD observed. The RMSD results pose confidence that the conditions used in the Set2 MD simulations produce relevant trajectories that can be used for further analyses.
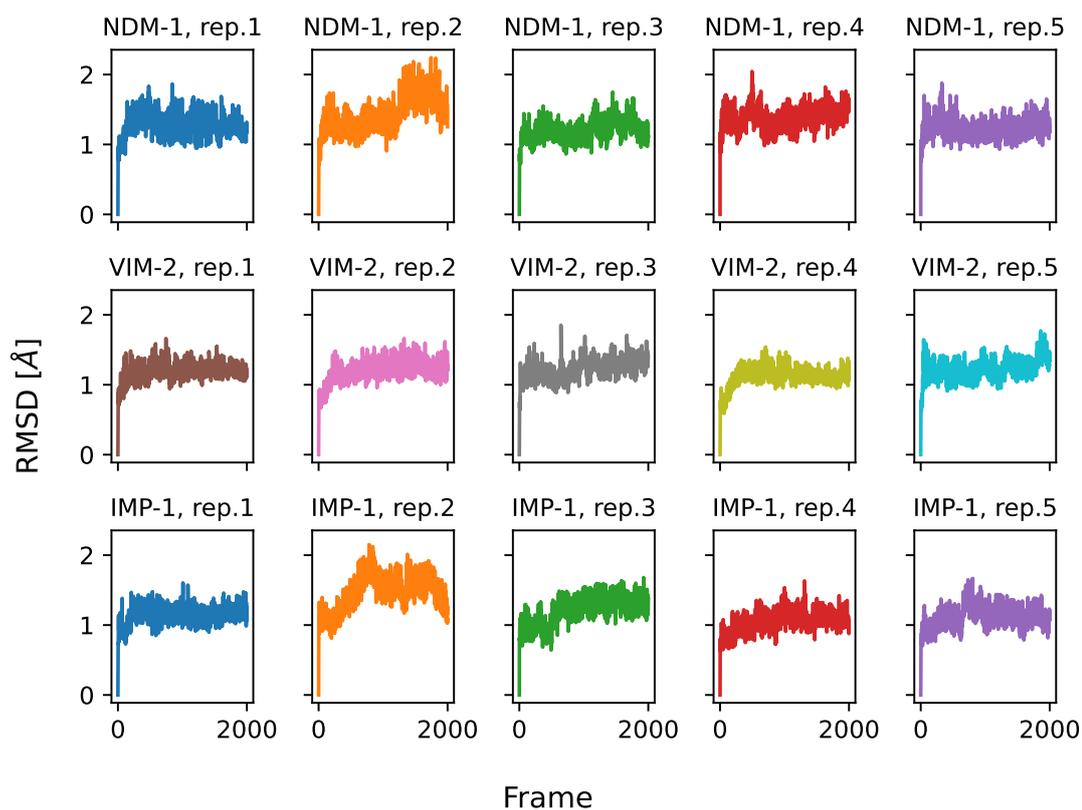
Figure 4.2.: Root-mean-square deviation (RMSD) of the backbone atoms for all MD simulation replicas (each 200 ns long) for NDM-1, VIM-2, and IMP-1. Highly mobile N- and C-terminal residues (average RMSF $> 1$ in at least one of the proteins) were excluded from the RMSD calculation—this applies to residues 1-20 and 241-255 of the consensus numbering scheme (see Figure 4.1).

## 4.3. Geometrical properties of the active site

Several distances and angles among the active-site components were measured. For the rationale behind the choice of these distances and angles, see Section 3.4.2. The selected geometrical properties are discussed from two perspectives. First, the MD simulation results of the three proteins of interest—NDM-1, VIM-2, and IMP-1—are compared. Second, the same geometrical properties were measured for, and averaged over, a selection of X-ray crystallographic structures without ligands bound in the active site (cf. Appendix A.1); these are then compared with the simulation results.

### Distances

Selected interatomic distances within the active site are listed in Table 4.1. For each distance, the MD results are shown in the first line and the average value from the X-ray structures in the second line. The MD results were obtained as follows: for each individual trajectory (i.e., each replica), the mean distance was calculated. The MD value reported in Table 4.1 is the mean value over the five replicas for each of the proteins. The mean distances of each replica were also used to calculate the standard deviations presented in the table. In other words, the standard deviations reported for the MD results allow a comparison of the replicas of each protein simulation and do not reflect the time evolution of the individual trajectories.

The Zn1–Zn2 mean distance obtained in the MD simulations of IMP-1 differs significantly from those of NDM-1 and VIM-2. Since the standard deviation for the IMP-1 result is comparable to that of the other two proteins, it is unlikely that the distinct mean value is caused by an instability of the IMP-1 system during the simulation. This is especially interesting in the context of the X-ray structure measurements, as the mean value of the IMP-1 X-ray structures is also considerably shorter, and has an even lower standard deviation—which is, by the way, relatively high in NDM-1 and VIM-2, suggesting a flexibility of the active site in its central region. The shorter mean distance and lower standard deviation of the Zn1–Zn2 distance in the IMP-1 X-ray structures suggests that the active site of the IMP-1 protein may be somewhat narrower in comparison to the NDM-1 and VIM-2 active sites.

It is also important to note that the X-ray data on the Zn1–Zn2 distances, shown in Table 4.1, are generally higher than the same distances in the X-ray structures used as initial coordinates for the MD simulations (3.41 Å in 5ZH1_A [NDM-1], 3.51 Å in 4BZ3_A [VIM-2] and 3.44 Å in 5Y5B_A [IMP-1]). However, the Zn1–Zn2 distance has been previously reported as a flexible property in dizinc MBLs, as it probably adapts to the presence of a substrate [97]. Zhang et al. [98] observed the longest Zn1–Zn2 distance ever reported in MBLs: 4.6 Å in the NDM-1 enzyme in a complex with hydrolyzed ampicillin. Their subsequent inspection of the system revealed that the Zn1–Zn2 distance indeed differs in complexes with other antibiotics (e.g., 3.96 Å in complex with hydrolyzed meropenem or 3.81 Å in complex with cefuroxime) [82].

The mean values of all His–Zn distances in the MD simulations are generally higher than in the X-ray structures. Contrary to the Zn1–Zn2 distance, the His–Zn distances

| Distances [Å] | NDM-1 | VIM-2 | IMP-1 |
|---|---|---|---|
| | MD | MD | MD |
| | X-ray | X-ray | X-ray |
| Zn1–Zn2 | 3.57 ± 0.01 | 3.60 ± 0.01 | 3.53 ± 0.01 |
| | 3.61 ± 0.35 | 3.67 ± 0.24 | 3.40 ± 0.13 |
| His1$_{NE2}$–Zn1 | 2.23 ± 0.03 | 2.25 ± 0.01 | 2.18 ± 0.02 |
| | 2.07 ± 0.04 | 2.11 ± 0.09 | 2.25 ± 0.46 |
| His2$_{ND1}$–Zn1 | 2.11 ± 0.05 | 2.08 ± 0.04 | 2.07 ± 0.00 |
| | 2.04 ± 0.13 | 1.99 ± 0.10 | 2.04 ± 0.03 |
| His3$_{NE2}$–Zn1 | 2.22 ± 0.09 | 2.14 ± 0.06 | 2.13 ± 0.01 |
| | 2.08 ± 0.08 | 2.05 ± 0.08 | 2.04 ± 0.08 |
| Asp$_{OD1}$–Zn2 | 1.82 ± 0.01 | 1.84 ± 0.01 | 1.85 ± 0.01 |
| | 2.06 ± 0.09 | 2.31 ± 0.16 | 2.02 ± 0.12 |
| Cys$_{SG}$–Zn1 | 4.17 ± 0.09 | 4.17 ± 0.01 | 4.00 ± 0.04 |
| | 3.99 ± 0.18 | 4.29 ± 0.09 | 3.82 ± 0.14 |
| Cys$_{SG}$–Zn2 | 2.42 ± 0.00 | 2.42 ± 0.00 | 2.43 ± 0.00 |
| | 2.33 ± 0.11 | 2.19 ± 0.13 | 2.30 ± 0.16 |
| His4$_{NE2}$–Zn2 | 2.27 ± 0.03 | 2.25 ± 0.03 | 2.24 ± 0.01 |
| | 2.13 ± 0.19 | 2.19 ± 0.10 | 2.17 ± 0.15 |
| Asp$_{OD2}$–OH$_{O1}$ | 2.55 ± 0.02 | 2.54 ± 0.02 | 2.57 ± 0.00 |
| | 2.96 ± 0.66 | 2.68 ± 0.07 | 3.42 ± 0.39 |

Table 4.1.: Selected interatomic distances in the active site. For each distance, the MD simulation results for NDM-1, VIM-2 and IMP-1 are shown in the first line; values from the 5 replicas were averaged and the standard deviation of each trajectory's average distance was calculated. The second line of each distance entry gives the average distance and standard deviation in the X-ray crystallographic structures without ligands bound in the active site. A standard deviation of 0.00 means that the value was less than 0.005 Å.

from the X-ray structures (e.g., for His2$_{ND1}$–Zn1 2.04±0.13 Å in NDM-1, 1.99±0.10 Å in VIM-2, 2.04±0.03 Å in IMP-1) are comparable to those observed in the original PDB structures used as the starting coordinates in the MD simulations (2.08 Å, 2.02 Å, 2.05 Å, respectively). The discrepancies between the average distances observed in the MD simulations and the X-ray structures suggest that further refinements of the His–Zn patch might be needed. As the equilibrium distance defined within this patch is set to 2.07 Å, which corresponds well to the observations in X-ray structures, it might be necessary to increase the force constant from the current value of 20 kcal/mol Å$^2$. On the other hand, this force constant allows the histidines to move relatively freely, as reflected by the differences in the mean values of the various His–Zn pairs, and their higher standard deviations. The small differences of the corresponding His–Zn distances among the three proteins indicate that the three protein systems behave similarly.

Generally, the shortest His–Zn distances are again observed in IMP-1. This supports the hypothesis of the narrower active site of IMP-1 even further, as the shorter Zn1–Zn2 distance is not compensated by longer His–Zn distances. The other distances in IMP-1, between the first-shell residues and Zn2 (Asp$_{OD1}$–Zn2 and Cys$_{SG}$–Zn2), are comparable to those in NDM-1 and VIM-2.

The His1–Zn1 distance averaged over the IMP-1 X-ray structures has a surprisingly high standard deviation. This is, however, caused by a single X-ray structure (namely, PDB ID 5HH4_D) in which the His1–Zn1 distance is 3.77 Å. Omitting this outlier, the mean value of this distance becomes 2.12 Å and the standard deviation reduces to only 0.05 Å. Other distances in 5HH4_D do not deviate significantly from the rest of the X-ray structure selection.

In the case of the Asp$_{OD1}$–Zn2 distance, the X-ray structures of VIM-2 have a somewhat higher mean value in comparison to the other two proteins. Interestingly, this might be compensated by a lower value of this residue's distance to the other Zn$^{2+}$ ion: the Asp$_{OD1}$–Zn1 is slightly shorter in VIM-2 (4.56±0.16 Å) than in NDM-1 (4.62±0.19 Å) and IMP-1 (4.66±0.10 Å), suggesting a slightly different positioning of the Asp residue in the VIM-2 active site.

The position of the active-site Cys is strictly defined in the MD simulations by the application of the non-bonded fix for the Cys$_{SG}$–Zn1 distance and the high force constant (200 kcal/mol rad$^2$) of the Cys$_{SG}$–Zn2 distance defined by the Cys–Zn patch. Despite this, the Cys$_{SG}$–Zn1 is another example of a shorter distance in IMP-1, compared to NDM-1 and VIM-2 in both the MD simulations and in the X-ray structures. This is in accordance with the previous observations—e.g., the Zn1–Zn2 distance—as well as the hypothesis of the narrower active site of IMP-1.

The distance Asp$_{OD2}$–OH$_{O1}$ was measured, although there is no direct interaction between these atoms. These residues are connected via a hydrogen bond Asp$_{HD2}$–OH$_{O1}$, but the position of the hydrogen atom is not resolved in the X-ray structures. Moreover, the standard deviations for this distance obtained from the X-ray structures of NDM-1 and IMP-1 are high. By contrast, the MD results are uniform across the three proteins, and also the replicas of an individual protein differ minimally.

| Angles [°] | NDM-1 | VIM-2 | IMP-1 |
|---|---|---|---|
| | MD | MD | MD |
| | X-ray | X-ray | X-ray |
| His1$_{NE2}$–Zn1–His2$_{ND1}$ | $92.2 \pm 3.0$ | $90.5 \pm 3.2$ | $91.6 \pm 2.9$ |
| | $99.8 \pm 2.8$ | $99.0 \pm 2.7$ | $101.6 \pm 8.9$ |
| His2$_{ND1}$–Zn1–His3$_{NE2}$ | $97.4 \pm 4.0$ | $98.6 \pm 3.9$ | $98.5 \pm 3.2$ |
| | $112.2 \pm 3.3$ | $106.4 \pm 3.2$ | $104.6 \pm 4.7$ |
| Cys$_{CB}$–Cys$_{SG}$–Zn2 | $126.8 \pm 1.3$ | $124.1 \pm 0.5$ | $125.8 \pm 0.4$ |
| | $111.8 \pm 3.5$ | $108.1 \pm 2.9$ | $115.9 \pm 15.7$ |
| Asp$_{CG}$–Asp$_{OD1}$–Zn2 | $120.7 \pm 0.3$ | $121.4 \pm 0.5$ | $120.9 \pm 0.2$ |
| | $134.3 \pm 6.3$ | $138.3 \pm 1.8$ | $136.3 \pm 5.5$ |
| Asp$_{CG}$–Asp$_{OD2}$–OH$_{O1}$ | $98.0 \pm 0.3$ | $99.3 \pm 0.2$ | $96.5 \pm 0.3$ |
| | $100.5 \pm 6.8$ | $96.9 \pm 3.6$ | $86.2 \pm 5.5$ |
| Zn1–OH$_{O1}$–Zn2 | $139.4 \pm 2.2$ | $142.8 \pm 1.5$ | $137.7 \pm 1.0$ |
| | $114.9 \pm 19.6$ | $118.7 \pm 8.7$ | $108.2 \pm 8.6$ |

Table 4.2.: Selected angles in the active site. For each angle, the MD simulation results of NDM-1, VIM-2 and IMP-1 are shown in the first line (averages and standard deviations of 5 replicas for each protein). The second line of each angle entry displays the average and standard deviation over all X-ray crystallographic structures without active-site ligands.

## Angles

Several angles were measured in order to monitor the correct geometrical arrangement of the active site and its stability during the MD simulations. A selection of these angles is listed in Table 4.2. Other angles were measured in order to refine the simulation parameters—for more details see Section 3.3.1—and are therefore not presented here. Generally, the measured angles do not differ significantly among the MD simulations of the three proteins.

Even after applying the more accurate equilibrium angles of the Set 2 parameters compared to the default values of CHARMM's His–Zn and Cys–Zn patches, the average angles observed during the MD simulations generally deviate slightly from the X-ray data. For example, the equilibrium angle of the first two angles in Table 4.2, His1$_{NE2}$–Zn1–His2$_{ND1}$ and His2$_{ND1}$–Zn1–His3$_{NE2}$, was set to 104° and held by a force constant of 100 kcal/mol rad$^2$. Despite this relatively strong force constant, the average angles observed in the MD simulations are generally smaller than in the X-ray structures. It is again important to note that in the case of the angle His1$_{NE2}$–Zn1–His2$_{ND1}$ in IMP-1, the X-ray structure 5HH4_D is an outlier in terms of the orientation of the His1 residue (126.1°). If this value is removed, the resulting average and standard deviation of the remaining IMP-1 X-ray structures is 99.6±5.2°.

Similarly to the His–Zn–His angles, the angle Cys$_{CB}$–Cys$_{SG}$–Zn2 shows higher mean values than in the X-ray structures; this angle was defined in the Cys–Zn patch with

an equilibrium angle of 112° and a force constant of 20 kcal/mol rad$^2$. The angle $Asp_{CG}$–$Asp_{OD2}$–$OH_{O1}$ was well reproduced in the MD simulations with respect to the X-ray data, even though it is not defined by the $Asp_{OD1}$–Zn custom patch. On the other hand, the angle $Zn1$–$OH_{O1}$–$Zn2$ displayed higher mean values in the MD simulations compared to the X-ray structures—one should, however, also note the high standard deviation of the X-ray data.

## 4.4. $Zn^{2+}$ ions coordination analysis

The coordination geometries of the $Zn^{2+}$ ions in the MD simulations of the NDM-1, VIM-2, and IMP-1 systems were checked using the AFICS program. Relevant coordination partners were counted in each frame of the trajectories (cf. 3.4.3); the percentage of the trajectory spent by each $Zn^{2+}$ ion with the coordination number (CN) of 4, 5, and 6 is shown in Table 4.3. Based on the particular CN, the actual coordination geometry was then compared to that of an ideal polyhedron (square pyramid and trigonal bipyramid for CN = 5, octahedral and trigonal prismatic geometry for CN = 6). The deviation of the actual coordination geometry from the ideal polyhedron with the same CN is expressed as an RMSD. Figure 4.3 shows the moving average with a window of 50 frames and the corresponding standard deviation in a lighter shade.

As described in Section 2.3.1, the coordination geometry proposed for Zn1 is tetrahedral in the B1 MBLs [33]. The coordination partners of Zn1 would therefore only be His1, His2, His3 and the hydroxide ion. However, the CN of 4 was not even detected by the AFICS algorithm, as it is shown in Figure 4.3. Due to the presence of water molecules in the cut-off distance of 2.7 Å from Zn1, mostly a CN of 5 was determined. The coordination geometry corresponds approximately to a trigonal bipyramidal arrangement with the mean deviation from the regular trigonal bipyramid 0.32±0.05 Å in both NDM-1 and VIM-2, and 0.33±0.04 Å in IMP-1 (average and standard deviation over 5 replicas). This arrangement is extremely invariable in the case of IMP-1. Moreover, Zn1 of IMP-1 is coordinated by the same water molecule (TIP3409) in all replicas except for replica 3 where this particular water molecule is replaced by TIP31382 and later by TIP32201 during the last 40 ns.

Particularly in NDM-1, but also in VIM-2, a second water molecule participates in the coordination of Zn1, resulting in a CN of 6 (see the top row of Figure 4.3). The preferred arrangement of Zn1 in this case is octahedral. Here, the deviation of the actual geometry from the regular polyhedron differs among the proteins and even among the replicas of NDM-1. The average RMSD for the octahedral arrangement in replica 1 of NDM-1 is 0.39±0.04 Å, but only 0.24±0.04 Å in replica 4 and 0.26±0.06 Å in replica 5 which is even a more stable geometry compared to the trigonal bipyramid in replicas 1-3. The water molecule TIP3477 coordinating Zn1 was detected in 4 replicas of NDM-1; this molecule even stays near Zn1 for the full simulation length in replicas 2 and 5. In the case of VIM-2, the CN of 6 in a relevant part of the trajectory was only detected in replica 5 with the deviation of 0.27±0.08 Å from an ideal octahedron. There are no particular water molecules that would be present near Zn1 in more than one replica.

|  | NDM-1 | | | VIM-2 | | | IMP-1 | | |
|---|---|---|---|---|---|---|---|---|---|
| **CN** | 4 | 5 | 6 | 4 | 5 | 6 | 4 | 5 | 6 |
| Zn1 [%] | 0 | 55 | 45 | 0 | 69 | 31 | 0 | 94 | 6 |
| Zn2 [%] | 0 | 0 | 100 | 0 | 0 | 100 | 0 | 0 | 100 |

Table 4.3.: Time spent by Zn1 and Zn2 with a CN of 4, 5, and 6 expressed as a percentage of the concatenated trajectory of 5 replicas of each protein.
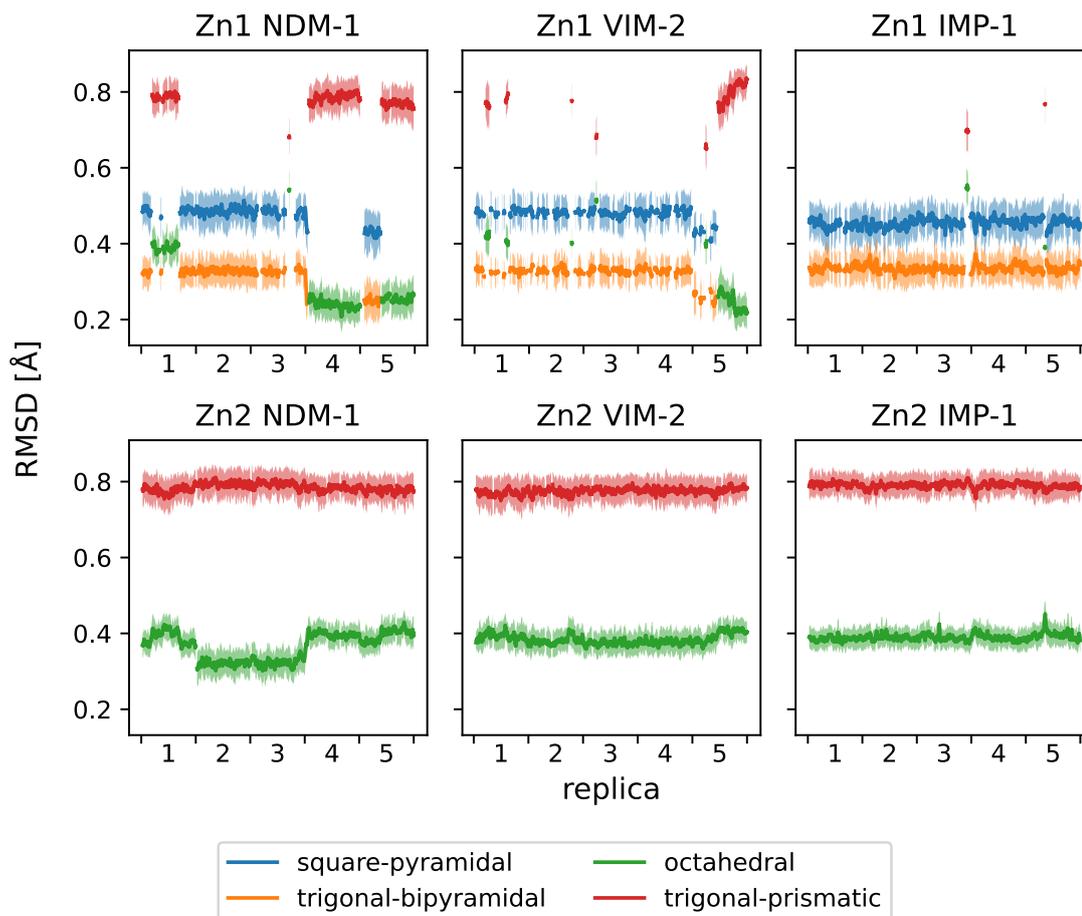


Figure 4.3.: Deviation (RMSD [Å]) of the actual coordination geometries of the $Zn^{2+}$ ions present in the active sites of NDM-1, VIM-2, and IMP-1 from the ideal polyhedra square pyramid and trigonal bipyramid (CN=5) or octahedral and trigonal prismatic structure (CN=6). The x-axis displays concatenated trajectories of 5 replicas of each protein. Data are shown as a moving average over a window of 50 frames; the corresponding standard deviation is shown in a lighter shade of the particular color.

The coordination geometry of Zn2 was reported to be a trigonal bipyramid [33], i.e., to have five coordination partners: His4, Cys, Asp, the hydroxide ion and a water molecule (cf. Section 2.3.1). In contrast to this, our MD simulations registered two water molecules coordinating Zn2 in all replicas of all three proteins, forming an octahedral arrangement. The deviation of this arrangement from a regular octahedral shape is 0.37±0.05 Å in NDM-1, 0.38±0.04 Å in VIM-2 and 0.39±0.03 Å in IMP-1 (average over 5 replicas of each protein).

The coordination of Zn2 is not only stable in terms of the low RMSD of the octahedral arrangement, but also in terms of the particular water molecules that participate in the zinc coordination. In NDM-1, Zn2 is coordinated by the water molecule TIP3611 in all replicas (all simulations of a certain protein were initiated from the same starting coordinates), in combination with TIP3519, TIP3647 or TIP3625. These water molecules remain present in the active site during the full simulation length. In the case of VIM-2, the water molecule TIP32208 participates in the Zn2-coordination in every replica, accompanied by various other water molecules. The most stable protein is again IMP-1— Zn2 is coordinated by water molecules TIP3425 and TIP3461 in all replicas, except for replica 5, where TIP3425 is accompanied first by TIP35772 and then by TIP313110 for the last approx. 60 ns.

To summarize the discussion of the zinc coordination spheres, the IMP-1 active site displayed the most stable zinc coordination within the MD simulations. This applies to the stability of the number of residues staying near Zn1 and Zn2 during the simulations and the number of water molecules exchanged by the $Zn^{2+}$ ions during the simulation time. The RMSD values recorded for IMP-1, however, do not differ significantly from the other two proteins. The exchange of the particular water molecules participating in the zinc coordination without changes in the CN are not accompanied by sudden changes in the RMSD values in any of the observed trajectory. It is also important to mention that this analysis relies solely on data from MD simulations with a classical mechanical force field. Therefore, it might be necessary to verify these results with other methods, such as QM/MM simulations.

## 4.5. RMSF and secondary structure

The RMSF results are presented in Figure 4.4. To make the results comparable between the three proteins, the RMSF is plotted against the consensus alignment of NDM-1, VIM-2 and IMP-1 (cf. sequence alignment in Figure 4.1). To put the RMSF results in a wider context, information about the secondary structure, as detected in the respective X-ray structures, is included in Figure 4.4. Only three secondary structure elements are distinguished: helix, sheet and loop. Under the term "helix", α-helices are generally understood—possible $3_{10}$ helices are explicitly mentioned in the text; "sheet" elements are β-sheets. We classify all other secondary structure elements as a "loop", including turns and coils. In order to determine the stability of the secondary structure elements observed in X-ray crystallographic data, the classification of each residue with respect to the secondary structure was monitored over the whole trajectory.
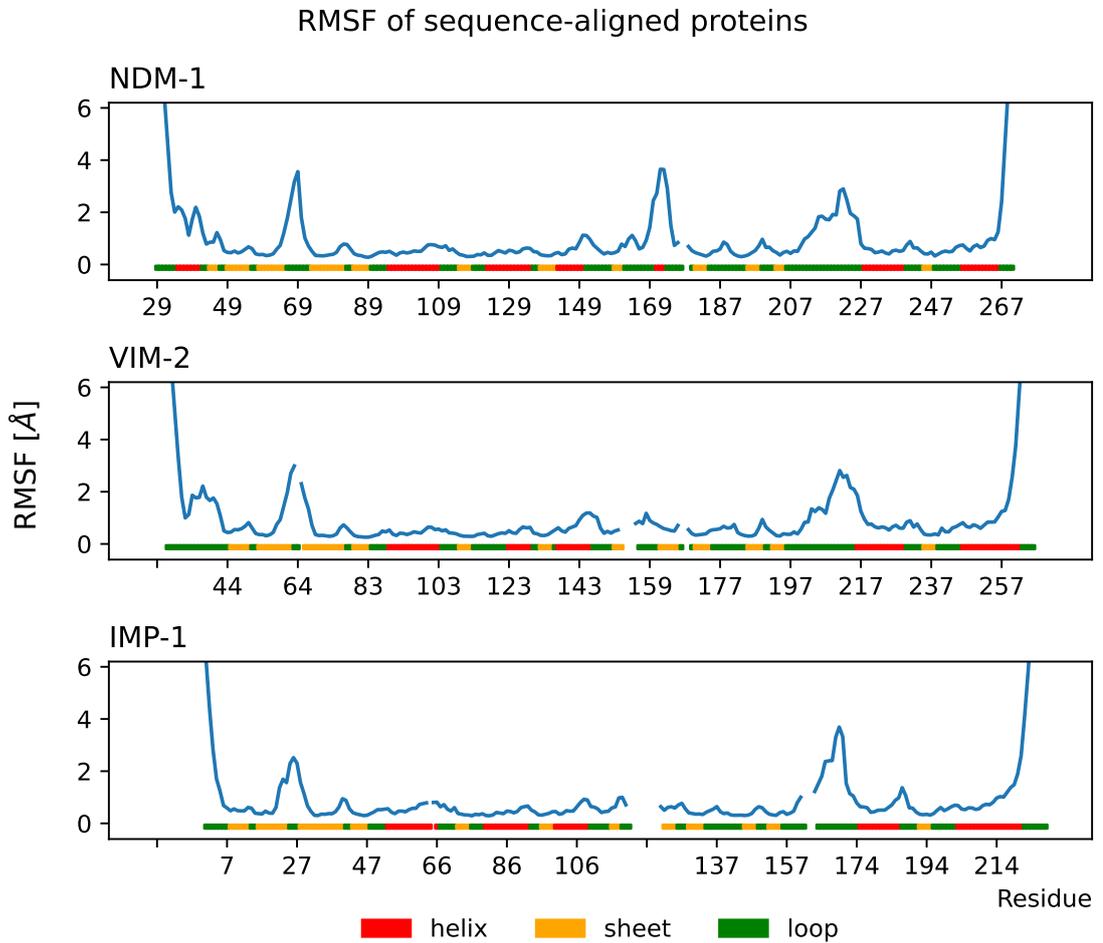
Figure 4.4.: Comparison of the C$_\alpha$ RMSFs [Å] of NDM-1, VIM-2, and IMP-1. The RMSFs of the three proteins are plotted against the protein-specific residue numbers, but respecting the consensus sequence alignment (cf. Section 4.1). For each protein, the RMSF was calculated from all 5 replicas. Color bars indicate the secondary structure of each residue as assigned in the corresponding X-ray structures: 5ZH1_A (NDM-1), 4BZ3_A (VIM-2), and 5Y5B (IMP-1).

In all three proteins, the terminal residues are highly mobile. Since, to our knowledge, no interactions of these residues with other regions of the protein have been reported in the literature, and since these residues are part of loop elements in VIM-2 and IMP-1, the high RMSF values are in accord with expectation. Interestingly, a short helix was detected at the N-terminus of NDM-1 (residues 35–41; protein-specific numbering). Its high mobility can be explained by the fact that the structure was simulated in aqueous solution, rather than anchored to a membrane. As already mentioned in Section 2.3.3, NDM-1 is a membrane-anchored protein. Apart from the lipidated Cys26, there are also several other residues that interact with the membrane. Such interactions were, for example, observed for the positively charged Arg45 [55], located near the helix 35–41. We therefore speculate that if this protein was simulated in a membrane-bound state, the RMSF of the N-terminal residues might be lower than observed here. During the simulations, the N-terminal helix is quite stable (in terms of secondary structure), except for the edge residues of the helix that are classified as loop or sheet residues in some frames.

Generally, the secondary structure elements are strongly conserved across the X-ray structures of the three proteins. There are minor qualitative differences at the N-terminus of NDM-1 (helix 35–41 already mentioned above, followed by a short β-sheet); another short helix (171–173; protein-specific numbering) is present in NDM-1 only—it will be discussed in more detail later in this section. The length of the aligned secondary structure elements conserved across the three proteins usually only differs by individual residues at the edges of the respective element. The edge residues, however, often oscillate between, e.g., a helix and a loop identity during the MD simulation. The most significant difference in length is in the region 95–107 (consensus numbering), which is only 7 residues long in VIM-2 (123–129; protein-specific numbering) compared to the full length of 13 residues in NDM-1 and IMP-1.

As one can see in Figure 4.4, there are several peaks in the RMSF plots of all three proteins. First, the peak around the residue 69 of NDM-1, 64 of VIM-2 or 27 of IMP-1 (protein-specific numbering) corresponds to the L3 loop. In the consensus numbering scheme, this loop encompasses residues 37–45. The flexibility of the L3 loop allows MBLs to process various classes of BLAs [99]. The most mobile residues of the L3 loop are Phe70 in NDM-1 with a mean RMSF value of 3.34 Å, Gly64 in VIM-2 (2.27 Å) and Gly27 in IMP-1 (2.36 Å).

Another RMSF peak common to all three proteins corresponds to the L10 loop. In NDM-1, this loop covers the residues 206–227 (protein-specific numbering); in VIM-2, the loop is two residues shorter (196–215) and in IMP-1, there is a deletion in the central part of the loop (156–174) in comparison to the two other proteins. IMP-1 shows the highest flexibility of the L10 loop in the MD simulations. Its most mobile residue in the L10 loop is Gly169, with RMSF = 2.84 Å. In VIM-2, it is Ala212 (2.37 Å), and in NDM-1, Gly222 (2.36 Å)—these are corresponding residues in the sequence alignment (residue number 196 in the consensus numbering scheme).

There is one significant peak in the RMSF plot of NDM-1 that is not present in the other two proteins, i.e., the region 169–176 (protein-specific numbering). Moreover, this

region includes the above-mentioned helix 171–173, which is also absent from VIM-2 and IMP-1. The most mobile residues of the region are Thr173 (3.28 Å) and Ala174 (3.24 Å). This region represents a considerable structural difference from the other two proteins. However, it is relatively distant from the active site, so a direct influence on the first-shell residues can be excluded. Moreover, the helical structure of the region is not very stable in the MD simulations. Initially, it is mostly classified as a $3_{10}$ helix, but later on, it is classified as a turn.

## 4.6. Water molecules in the active site

The number of water molecules within selected distances from the active site center (3–9 Å) is listed in Table 4.4. For each protein, the mean values of the five replicas were averaged. The standard deviation was determined for each individual trajectory of a certain protein, and averaged.

As one can see, there are no significant differences among the three proteins in the average number of water molecules within the spherical approximation of the active site. Moreover, the number of water molecules varies very little during the simulation time. The number of water molecules in the 5 replicas of each protein do not show any dissimilarities as well. This also indicates that the accessibility and interactions with water molecules of the active sites of NDM-1, VIM-2, and IMP-1 remain stable throughout the simulations. A closer look at water molecules coordinating the $Zn^{2+}$ ions was provided in Section 4.4.

The similarity of the content of water molecules within the active sites of the three proteins can be explained by the imperfection of the spherical approximation of the active site volume. Further analyses of the number of water molecules in the active site would require a more appropriate approximation, using sophisticated tools to determine the exact shape of the active site. Such an approach would allow us either to detect potential differences, or to securely declare that there are no differences between the proteins of interest in this context.

| Distance [Å] | NDM-1 | VIM-2 | IMP-1 |
|:---:|:---:|:---:|:---:|
| 3 | $3 \pm 1$ | $3 \pm 1$ | $3 \pm 1$ |
| 4 | $4 \pm 1$ | $4 \pm 1$ | $4 \pm 1$ |
| 5 | $7 \pm 1$ | $7 \pm 1$ | $8 \pm 1$ |
| 6 | $13 \pm 2$ | $13 \pm 2$ | $14 \pm 2$ |
| 7 | $21 \pm 2$ | $21 \pm 2$ | $22 \pm 2$ |
| 8 | $32 \pm 3$ | $30 \pm 3$ | $31 \pm 3$ |
| 9 | $44 \pm 4$ | $41 \pm 4$ | $43 \pm 4$ |

Table 4.4.: The number of water molecules inside a spherical approximation of the active site with selected radii (i.e., distances from the center of the active sites of NDM-1, VIM-2, and IMP-1). The center of the active site is defined as the midpoint between the two $Zn^{2+}$ ions. Mean values represent the average over the trajectories of the 5 replicas of each protein; standard deviation is the average of 5 standard deviation values of each individual replica.

# 5. Conclusion

MD simulations of three MBLs—NDM-1, VIM-2, and IMP-1—were performed with the aim to provide a more profound insight into the dynamics of these enzymes. Special attention was paid to the active site. To obtain reliable trajectories from the MD simulations, i.e., trajectories in which the active site is reproduced correctly with respect to the initial X-ray crystallographic structures, several stereochemical characteristics of the system were monitored. Based on the first round of results, simulation parameters were refined to achieve a more accurate representation of the active site geometries. The coordination of the active site $Zn^{2+}$ ions by the first-shell residues had to be described by covalent pseudo-bonds; these were introduced by three custom patches—the CHARMM-provided His–Zn patch and Cys–Zn patch, as well as the $ASPP_{OD1}$–Zn patch written by Åsmund Kaupang.

A significant part of this master's thesis focused on the refinement of these custom patches. To better reproduce the orientation of the imidazole ring planes with respect to the $Zn^{2+}$ ions, improper dihedral angle terms were added to the His–Zn patch. To prevent distortions of the active site observed in early trajectories, the parameters of the Cys–Zn patch were modified and a non-bonded fix term was added. Furthermore, the force constants and angle equilibrium values of all three patches were customized.

The above-described refinement of the simulation parameters allowed us to run MD simulations that produced trajectories reliable enough to be used for studying the dynamics of the active site in more detail. Several distances and angles within the active site were selected in order to discover differences in the behavior of the active site during the simulation. While no major differences among the three proteins were found, we, nevertheless, hypothesize that the active site of IMP-1 may be narrower compared to the active sites of NDM-1 and VIM-2 based on several observed average distances. The coordination geometries of the $Zn^{2+}$ ions present in the active site were also monitored during the simulations, with special attention paid to the water molecules participating in the $Zn^{2+}$-coordination. The results suggest that the conditions of the IMP-1 active site, as reproduced in the MD simulations, allow a more stable trigonal bipyramidal arrangement of Zn1-coordination and octahedral coordination geometry of Zn2. Apart from this, the number of water molecules present in a spherical approximation of the active site was determined, but no significant differences between the proteins were observed.

The stability of the simulated proteins was controlled by monitoring the RMSD function over the course of the simulations. The flexibility of individual residues was determined via RMSF calculations. To compare the RMSFs of the three proteins, a sequence alignment was carried out. The resulting consensus alignment also revealed that the secondary structure elements are strongly conserved across the three proteins, and that the secondary structure elements remain stable during the MD simulations.

*5. Conclusion*

The employed methods and analysis procedures have certain limitations. First, as reported in Section 3.2.2, the NDM-1 structure used as the starting coordinates of the MD simulations lacks two N-terminal residues that follow the lipidated cysteine residue anchoring NDM-1 to a membrane. However, we do not expect this limitation to have serious effects on the results presented, as the focus of this thesis was on the active site, which is located distantly from the N-terminus. Furthermore, it had no effect on the sequence alignment, as the full sequences of the proteins were used. On the other hand, simulations of the NDM-1 protein anchored to a membrane might produce different results, especially in analyses in which the full structures were considered. This applies in particular to the RMSF analysis, as there are several residues on the surface of the NDM-1 structure that were reported in the literature to interact with the membrane. The presence of a membrane might therefore result in a lower flexibility of certain regions of NDM-1.

Several analyses revealed minor differences of the IMP-1 protein in comparison to NDM-1 and VIM-2. However, these observations are not sufficient to allow us to hypothesize about why the IMP-1 protein responds less well to treatment with certain inhibitors, as described in Section 3.1. Further investigations are therefore required to answer these questions. The MD simulations performed here could, for example, be used as starting points of QM/MM simulations that might describe the behavior of the active site more accurately.

Another phenomenon that was out of the scope of this thesis is the interaction of the residues of the first shell with the second-shell residues. As the active site of B1 MBLs is conserved not only in its composition, but also in its catalytic mechanism, an analysis of the second-shell residues and their possible impact on the active site might reveal differences that could bring more insight into the observed dissimilarities in the enzymes' response to inhibitors.

# Bibliography

[1] World Health Organization. Antimicrobial resistance, November 17 2021. URL `https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance`. Accessed on 13/10/2023.

[2] Allison Way, Maria Bond, Bradley Nanna, and Erik S. Wright. Evaluating the long-term portrayal of antibiotic resistance in major U.S. newspapers. *BMC Public Health*, 23:1343, 2023. doi: 10.1186/s12889-023-16203-8.

[3] Christopher J Murray, Kevin Shunji Ikuta, Fablina Sharara, Lucien Swetschinski, Gisela Robles Aguilar, Authia Gray, Chieh Han, Catherine Bisignano, Puja Rao, and Eve et al. Wool. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, 399(10325):629–655, 2022. doi: 10.1016/s0140-6736(21)02724-0.

[4] Porooshat Dadgostar. Antimicrobial resistance: Implications and costs. *Infection and Drug Resistance*, 12:3903–3910, 2019. doi: 10.2147/idr.s234610.

[5] World Health Organization. Ten threats to global health in 2019. URL `https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019`. Accessed on 13/10/2023.

[6] World Health Organization. Incentivising the development of new antibacterial treatments 2023: Progress report by the Global AMR R&D Hub & WHO. 2023.

[7] World Health Organization. Global action plan on antimicrobial resistance. 2015.

[8] Angela Willemsen, Simon Reid, and Yibeltal Assefa. A review of national action plans on antimicrobial resistance: strengths and weaknesses. *Antimicrobial Resistance and Infection Control*, 11:1–13, 2022. doi: 10.1186/s13756-022-01130-x.

[9] Wanda C Reygaert. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiology*, 4:482–501, 2018. doi: 10.3934/microbiol.2018.3.482.

[10] Shilpa Ray, Susmita Das, and Mrutyunjay Suar. *Molecular mechanism of drug resistance*, pages 47–110. Springer International Publishing, 2017. ISBN 9783319486833. doi: 10.1007/978-3-319-48683-3_3.

[11] Elizabeth M. Darby, Eleftheria Trampari, Pauline Siasat, Maria Solsona Gaya, Ilyas Alav, Mark A. Webber, and Jessica M.A. Blair. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology*, 21:280–295, 2022. doi: 10.1038/s41579-022-00820-y.

*Bibliography*

[12] Julian Davies and Dorothy Davies. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 74:417–433, 2010. doi: 10.1128/mmbr .00016-10.

[13] Nikola Vidovic and Sinisa Vidovic. Antimicrobial resistance and food animals: Influence of livestock environment on the emergence and dissemination of antimicrobial resistance. *Antibiotics*, 9:52, 2020. doi: 10.3390/antibiotics9020052.

[14] Brian G. Bell, Francois Schellevis, Ellen Stobberingh, Herman Goossens, and Mike Pringle. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infectious Diseases*, 14:1–25, 2014. doi: 10.1186/1471 -2334-14-13.

[15] Sakib Rahman and Aidan Hollis. The effect of antibiotic usage on resistance in humans and food-producing animals: a longitudinal, One Health analysis using European data. *Frontiers in Public Health*, 11:1170426, 2023. doi: 10.3389/fpubh.2023.1170426.

[16] C. Lee Ventola. The antibiotic resistance crisis, Part 1: Causes and threats. *Pharmacy and Therapeutics*, 40:277–283, 2015.

[17] European Union. Ban on antibiotics as growth promoters in animal feed enters into effect, December 22 2005. URL `https://ec.europa.eu/commission/presscorne r/detail/en/IP_05_1687`. Accessed on 13/10/2023.

[18] Karen L. Tang, Niamh P. Caffrey, Diego B. Nóbrega, Susan C. Cork, Paul E. Ronksley, Herman W. Barkema, Alicia J. Polachek, Heather Ganshorn, Nishan Sharma, James D. Kellner, and William A. Ghali. Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis. *The Lancet Planetary Health*, 1:e316–e327, 2017. doi: 10.1016/s2542-5196(17)30141-9.

[19] Kasim Allel, Lucy Day, Alisa Hamilton, Leesa Lin, Luis Furuya-Kanamori, Catrin E. Moore, Thomas Van Boeckel, Ramanan Laxminarayan, and Laith Yakob. Global antimicrobial-resistance drivers: an ecological country-level study at the human–animal interface. *The Lancet Planetary Health*, 7:e291–e303, 2023. doi: 10.1016/s2542-5196(23)00026-8.

[20] Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2019. 2019. doi: 10.15620/cdc:82532.

[21] Karen Bush and Patricia A. Bradford. β-lactams and β-lactamase inhibitors: An overview. *Cold Spring Harbor Perspectives in Medicine*, 6, 2016. doi: 10.1101/cshperspect.a025247.

[22] Silvana Alfei and Anna Maria Schito. β-lactam antibiotics and β-lactamase enzymes inhibitors, Part 2: Our limited resources. *Pharmaceuticals*, 15:476, 2022. doi: 10.3390/ph15040476.

[23] Margherita De Rosa, Anna Verdino, Annunziata Soriente, and Anna Marabotti. The odd couple(s): An overview of beta-lactam antibiotics bearing more than one pharmacophoric group. *International Journal of Molecular Sciences*, 22:1–21, 2021. doi: 10.3390/ijms22020617.

[24] Lidia Moreira Lima, Bianca Nascimento Monteiro Da Silva, Gisele Barbosa, and J Barreiro. β-lactam antibiotics: An overview from a medicinal chemistry perspective. *European Journal of Medicinal Chemistry*, 2020. doi: 10.1016/j.ejmech.2020.112829.

[25] Waldemar Vollmer, Didier Blanot, and Miguel A. De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32:149–167, 2008. doi: 10.1111/j.1574-6976.2007.00094.x.

[26] R. P. Ambler. The structure of β-lactamases. *Philosophical Transactions of the Royal Society B*, 289:321–331, 1980. doi: 10.1098/rstb.1980.0049.

[27] Sara E. Boyd, David M. Livermore, David C. Hooper, and William W. Hope. Metallo-β-lactamases: Structure, function, epidemiology, treatment options, and the development pipeline. *Antimicrobial Agents and Chemotherapy*, 64, 2020. doi: 10.1128/aac.00397-20.

[28] Nikol Kaderabkova, Manasa Bharathwaj, R. Christopher D. Furniss, Diego Gonzalez, Tracy Palmer, and Despoina A.I. Mavridou. The biogenesis of β-lactamase enzymes. *Microbiology*, 168:1217, 2022. doi: 10.1099/mic.0.001217.

[29] K. Bush, G. A. Jacoby, and A. A. Medeiros. A functional classification scheme for β-lactamases and its correlation with molecular structure. *Antimicrobial Agents and Chemotherapy*, 39:1211–1233, 1995. doi: 10.1128/aac.39.6.1211.

[30] Carolina López, Juliana Delmonti, Robert A. Bonomo, and Alejandro J. Vila. Deciphering the evolution of metallo-β-lactamases: A journey from the test tube to the bacterial periplasm. *Journal of Biological Chemistry*, 298:101665, 2022. doi: 10.1016/j.jbc.2022.101665.

[31] Timothy Palzkill. Metallo-β-lactamase structure and function. *Annals of the New York Academy of Sciences*, 1277:91–104, 2013. doi: 10.1111/j.1749-6632.2012.06796.x.

[32] Hiromi Daiyasu, Kazuya Osaka, Yoshizumi Ishino, and Hiroyuki Toh. Expansion of the zinc metallo-hydrolase family of the β-lactamase fold. *FEBS Letters*, 503:1–6, 2001. doi: 10.1016/s0014-5793(01)02686-2.

[33] Guillermo Bahr, Lisandro J. González, and Alejandro J. Vila. Metallo-β-lactamases in the age of multidrug resistance: From structure and mechanism to evolution, dissemination, and inhibitor design. *Chemical Reviews*, 121:7957–8094, 2021. doi: 10.1021/acs.chemrev.1c00138.

*Bibliography*

[34] Michael I. Page and Adriana Badarau. The mechanisms of catalysis by metallo β-lactamases. *Bioinorganic Chemistry and Applications*, 2008, 2008. doi: 10.1155/2008/576297.

[35] Samuel Karlin and Zhan Yang Zhu. Classification of mononuclear zinc metal sites in protein structures. *Proceedings of the National Academy of Sciences of the United States of America*, 94:14231–14236, 1997. doi: 10.1073/pnas.94.26.14231.

[36] Natalia Díaz, Dimas Suárez, and Kenneth M. Merz. Zinc metallo-β-lactamase from *Bacteroides fragilis*: A quantum chemical study on model systems of the active site. *Journal of the American Chemical Society*, 122:4197–4208, 2000. doi: 10.1021/ja994462s.

[37] Leticia I. Llarrull, Stella M. Fabiane, Jason M. Kowalski, Brian Bennett, Brian J. Sutton, and Alejandro J. Vila. Asp-120 locates Zn2 for optimal metallo-β-lactamase activity. *Journal of Biological Chemistry*, 282:18276–18285, 2007. doi: 10.1074/jbc.M700742200.

[38] Isabelle S. Arts, Alexandra Gennaris, and Jean François Collet. Reducing systems protecting the bacterial cell envelope from oxidative damage. *FEBS Letters*, 589: 1559–1568, 2015. doi: 10.1016/j.febslet.2015.04.057.

[39] Mariá Natalia Lisa, Antonela R. Palacios, Mahesh Aitha, Mariano M. González, DIego M. Moreno, Michael W. Crowder, Robert A. Bonomo, James Spencer, David L. Tierney, Leticia I. Llarrull, and Alejandro J. Vila. A general reaction mechanism for carbapenem hydrolysis by mononuclear and binuclear metallo-β-lactamases. *Nature Communications*, 8:1–11, 2017. doi: 10.1038/s41467-017-00601-9.

[40] Fabian Sievers, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D. Thompson, and Desmond G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539, 2011. doi: 10.1038/msb.2011.75.

[41] Fabian Sievers and Desmond G. Higgins. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, 27:135–145, 2018. doi: 10.1002/pro.3290.

[42] Andreas D. Baxevanis, Gary D. Bader, and David Scott Wishart. *Bioinformatics*. Wiley, 2020. ISBN 978-1-119-33558-0.

[43] Giorgio Bianchini and Patricia Sánchez-Baracaldo. Treeviewer version 2.1.0, 2023. URL https://doi.org/10.5281/zenodo.7768344.

[44] E. Osano, Y. Arakawa, R. Wacharotayankun, M. Ohta, T. Horii, H. Ito, F. Yoshimura, and N. Kato. Molecular characterization of an enterobacterial metallo beta-lactamase found in a clinical isolate of *Serratia marcescens* that shows imipenem resistance. *Antimicrobial Agents and Chemotherapy*, 38:71–78, 1994. doi: 10.1128/aac.38.1.71.

56

[45] Zishuo Cheng, Christopher R. Bethel, Pei W. Thomas, Ben A. Shurina, J. P. Alao, Caitlyn A. Thomas, Kundi Yang, Steven H. Marshall, Huan Zhang, and Aidan M. Sturgill et al. Carbapenem use is driving the evolution of imipenemase 1 variants. *Antimicrobial Agents and Chemotherapy*, 65, 2021. doi: 10.1128/aac.01714-20.

[46] Keizo Yamamoto, Hideaki Tanaka, Genji Kurisu, Ryuichi Nakano, Hisakazu Yano, and Hiromi Sakai. Structural insights into the substrate specificity of IMP-6 and IMP-1 metallo-β-lactamases. *Journal of Biochemistry*, 173:21–30, 2023. doi: 10.109 3/jb/mvac080.

[47] Laura Lauretti, Maria Letizia Riccio, Annarita Mazzariol, Giuseppe Cornaglia, Gianfranco Amicosante, Roberta Fontana, and Gian Maria Rossolini. Cloning and characterization of *bla*VIM, a new integron-borne metallo-β-lactamase gene from a Pseudomonas aeruginosa clinical isolate. *Antimicrobial Agents and Chemotherapy*, 43:1584–1590, 1999. doi: 10.1128/aac.43.7.1584.

[48] Laurent Poirel, Thierry Naas, Delphine Nicolas, Louis Collet, Samuel Bellais, Jean Didier Cavallo, and Patrice Nordmann. Characterization of VIM-2, a carbapenem-hydrolyzing metallo-β-lactamase and its plasmid- and integron-borne gene from a *Pseudomonas aeruginosa* clinical isolate in france. *Antimicrobial Agents and Chemotherapy*, 44:891–897, 2000. doi: 10.1128/aac.44.4.891-897.2000.

[49] Philipp Kohler, Nathalie Tijet, Hyunjin C. Kim, Jennie Johnstone, Tom Edge, Samir N. Patel, Christine Seah, Barbara Willey, Brenda Coleman, and Karen Green et al. Dissemination of Verona Integron-encoded Metallo-β-lactamase among clinical and environmental Enterobacteriaceae isolates in Ontario, Canada. *Scientific Reports*, 10:1–10, 2020. doi: 10.1038/s41598-020-75247-7.

[50] Anne Makena, Azer Düzgün, Jürgen Brem, Michael A. McDonough, Anna M. Rydzik, Martine I. Abboud, Ayşegül Saral, Ayşegül Çiçek, Cemal Sandalli, and Christopher J. Schofield. Comparison of verona integron-borne metallo-β-lactamase (VIM) variants reveals differences in stability and inhibition profiles. *Antimicrobial Agents and Chemotherapy*, 60:1377–1384, 2016. doi: 10.1128/aac.01768-15.

[51] Dongeun Yong, Mark A. Toleman, Christian G. Giske, Hyun S. Cho, Kristina Sundman, Kyungwon Lee, and Timothy R. Walsh. Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrobial Agents and Chemotherapy*, 53:5046–5054, 2009. doi: 10.1128/aac.00774-09.

[52] Guillermo Bahr, Luisina Vitor-Horen, Christopher R. Bethel, Robert A. Bonomo, Lisandro J. Gonzalez, and Alejandro J. Vila. Clinical evolution of New Delhi Metallo-β-lactamase (NDM) optimizes resistance under Zn(II) deprivation. *Antimicrobial Agents and Chemotherapy*, 62, 2018. doi: 10.1128/aac.01849-17.

[53] Zishuo Cheng, Pei W. Thomas, Lincheng Ju, Alexander Bergstrom, Kelly Mason, Delaney Clayton, Callie Miller, Christopher R. Bethel, Jamie VanPelt, and David L. Tierney et al. Evolution of New Delhi metallo-β-lactamase (NDM) in the clinic: Effects of NDM mutations on stability, zinc affinity, and mono-zinc activity. *Journal of Biological Chemistry*, 293:12606–12618, 2018. doi: 10.1074/jbc.ra118.003835.

[54] Lisandro J. González, Guillermo Bahr, Toshiki G. Nakashige, Elizabeth M. Nolan, Robert A. Bonomo, and Alejandro J. Vila. Membrane anchoring stabilizes and favors secretion of New Delhi metallo-β-lactamase. *Nature Chemical Biology*, 12:516–522, 2016. doi: 10.1038/nchembio.2083.

[55] Alessio Prunotto, Guillermo Bahr, Lisandro J. González, Alejandro J. Vila, and Matteo Dal Peraro. Molecular bases of the membrane association mechanism potentiating antibiotic resistance by New Delhi Metallo-β-lactamase 1. *ACS Infectious Diseases*, 6:2719–2731, 2020. doi: 10.1021/acsinfecdis.0C00341.

[56] Kamaleddin H.M.E. Tehrani and Nathaniel I. Martin. β-lactam/β-lactamase inhibitor combinations: an update. *MedChemComm*, 9:1439–1456, 2018. doi: 10.1039/C8MD 00342D.

[57] Andrew M. King, Sarah A. Reid-Yu, Wenliang Wang, Dustin T. King, Gianfranco De Pascale, Natalie C. Strynadka, Timothy R. Walsh, Brian K. Coombes, and Gerard D. Wright. Aspergillomarasmine A overcomes metallo-β-lactamase antibiotic resistance. *Nature*, 510:503–506, 2014. doi: 10.1038/nature13445.

[58] Ørjan Samuelsen, Ove Alexander Høgmoen Åstrand, Christopher Fröhlich, Adam Heikal, Susann Skagseth, Trine Josefine Olsen Carlsen, Hanna Kirsti S. Leiros, Annette Bayer, Christian Schnaars, and Geir Kildahl-Andersen et al. Zn148 is a modular synthetic metallo-β-lactamase inhibitor that reverses carbapenem resistance in Gram-negative pathogens *in vivo*. *Antimicrobial Agents and Chemotherapy*, 64, 2020. doi: 10.1128/aac.02415-19.

[59] Dustin T. King, Liam J. Worrall, Robert Gruninger, and Natalie C.J. Strynadka. New Delhi Metallo-β-lactamase: Structural insights into β-lactam recognition and inhibition. *Journal of the American Chemical Society*, 134:11362–11365, 2012. doi: 10.1021/ja303579d.

[60] Jürgen Brem, Sander S. Van Berkel, David Zollman, Sook Y. Lee, Opher Gileadi, Peter J. McHugh, Timothy R. Walsh, Michael A. McDonough, and Christopher J. Schofield. Structural basis of metallo-β-lactamase inhibition by captopril stereoisomers. *Antimicrobial Agents and Chemotherapy*, 60:142–150, 2016. doi: 10.1128/aac.01335-1 5.

[61] Jeffrey H. Toney, Gail G. Hammond, Paula M.D. Fitzgerald, Nandini Sharma, James M. Balkovec, Gregory P. Rouen, Steven H. Olson, Milton L. Hammond,

Mark L. Greenlee, and Ying Duo Gao. Succinic acids as potent inhibitors of plasmid-borne IMP-1 metallo-β-lactamase. *Journal of Biological Chemistry*, 276:31913–31918, 2001. doi: 10.1074/jbc.M104742200.

[62] Jürgen Brem, Ricky Cain, Samuel Cahill, Michael A. McDonough, Ian J. Clifton, Juan Carlos Jiménez-Castellanos, Matthew B. Avison, James Spencer, Colin W.G. Fishwick, and Christopher J. Schofield. Structural basis of metallo-β-lactamase, serine-β-lactamase and penicillin-binding protein inhibition by cyclic boronates. *Nature Communications*, 7:1–8, 2016. doi: 10.1038/ncomms12406.

[63] Bin Liu, Robert E.Lee Trout, Guo Hua Chu, Daniel Mcgarry, Randy W. Jackson, Jodie C. Hamrick, Denis M. Daigle, Susan M. Cusick, Cecilia Pozzi, and Filomena De Luca et al. Discovery of Taniborbactam (VNRX-5133): A broad-spectrum serine- and metallo-β-lactamase inhibitor for carbapenem-resistant bacterial infections. *Journal of Medicinal Chemistry*, 63:2789–2801, 2020. doi: 10.1021/acs.jmedchem.9 B01518.

[64] Press releases | Venatorx Pharmaceuticals, August 15 2023. URL https://www.ve natorx.com/press-releases/venatorx-pharmaceuticals-announces-fda-acc eptance-and-priority-review-of-new-drug-application-for-cefepime-tan iborbactam-to-treat-complicated-urinary-tract-infections-cuti-includi ng-pyelonephritis/. Accessed on 13/10/2023.

[65] Maria F. Mojica, Robert A. Bonomo, and Walter Fast. B1-metallo-beta-lactamases: Where do we stand? *Current drug targets*, 17:1029–1050, 2016. doi: 10.2174/138945 0116666151001105622.

[66] Giuseppe Mancuso, Angelina Midiri, Elisabetta Gerace, and Carmelo Biondo. Bacterial antibiotic resistance: the most critical pathogens. *Pathogens*, 10, 2021. doi: 10.3390/pathogens10101310.

[67] Alen Krajnc, Jürgen Brem, Philip Hinchliffe, Karina Calvopiña, Tharindi D. Panduwawala, Pauline A. Lang, Jos J.A.G. Kamps, Jonathan M. Tyrrell, Emma Widlake, and Benjamin G. Saward et al. Bicyclic boronate VNRX-5133 inhibits metallo- and serine-β-lactamases. *Journal of Medicinal Chemistry*, 62:8544–8556, 2019. doi: 10.1021/acs.jmedchem.9B00911.

[68] Ruslan Tsivkovski, Maxim Totrov, and Olga Lomovskaya. Biochemical characterization of QPX7728, a new ultrabroad-spectrum beta-lactamase inhibitor of serine and metallo-beta-lactamases. *Antimicrobial Agents and Chemotherapy*, 64, 2020. doi: 10.1128/aac.00130-20.

[69] William R. Pearson. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, 0 3, 2013. doi: 10.1002/0471250953.BI0301S42.

[70] Alex Bateman, Maria Jesus Martin, Claire O'Donovan, Michele Magrane, Rolf Apweiler, Emanuele Alpi, Ricardo Antunes, Joanna Arganiska, Benoit Bely, and

Mark Bingley et al. UniProt: a hub for protein information. *Nucleic Acids Research*, 43:D204–D212, 2015. doi: 10.1093/nar/gku989.

[71] Scott A. Hollingsworth and Ron O. Dror. Molecular dynamics simulation for all. *Neuron*, 99:1129–1143, 2018. doi: 10.1016/j.neuron.2018.08.011.

[72] Ron O. Dror, Robert M. Dirks, J. P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452, 2012. doi: 10.1146/annurev-biophys-042910-15524 5.

[73] M. J. Harvey, G. Giupponi, and G. De Fabritiis. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *Journal of Chemical Theory and Computation*, 5:1632–1639, 2009. doi: 10.1021/CT9000685.

[74] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, and Agata Laydon et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50:D439–D444, 2022. doi: 10.1093/nar/gkab1061.

[75] Mihaly Varadi and Sameer Velankar. The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics*, 23:2200128, 2023. doi: 10.1002/pmic.202200128.

[76] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017. ISBN 9780191841439. doi: 10.1093/OSO/9780198803195.001.0001.

[77] Tobias Lippert and Matthias Rarey. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *Journal of Cheminformatics*, 1:1–12, 2009. doi: 10.1186/1758-2946-1-13.

[78] Stefan Bietz, Sascha Urbaczek, Benjamin Schulz, and Matthias Rarey. Protoss: A holistic approach to predict tautomers and protonation states in protein-ligand complexes. *Journal of Cheminformatics*, 6:1–12, 2014. doi: 10.1186/1758-2946-6-12.

[79] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29:1859–1865, 2008. doi: 10.1002/jcc.20945.

[80] John Z. Chen, Douglas M. Fowler, and Nobuhiko Tokuriki. Comprehensive exploration of the translocation, stability and substrate recognition requirements in VIM-2 lactamase. *eLife*, 9:1–31, 2020. doi: 10.7554/elife.56707.

[81] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, and

Anna Potapenko et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021. doi: 10.1038/s41586-021-03819-2.

[82] Hongmin Zhang, Guixing Ma, Yifan Zhu, Lingxiao Zeng, Ashfaq Ahmad, Changzhi Wang, Bo Pang, Huiyan Fang, Liqing Zhao, and Quan Hao. Active-site conformational fluctuations promote the enzymatic activity of NDM-1. *Antimicrobial Agents and Chemotherapy*, 62, 2018. doi: 10.1128/aac.01579-18.

[83] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, and S. Boresch et al. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545–1614, 2009. doi: 10.1002/jcc.21287.

[84] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, and Yifei Qi et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 12:405–413, 2016. doi: 10.1021/acs.jctc.5B00935.

[85] Peter Eastman, Mark S. Friedrichs, John D. Chodera, Randall J. Radmer, Christopher M. Bruns, Joy P. Ku, Kyle A. Beauchamp, Thomas J. Lane, Lee Ping Wang, and Diwakar Shukla et al. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation*, 9:461–469, 2013. doi: 10.1021/CT300857J.

[86] Jing Huang and Alexander D. Mackerell. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34:2135–2145, 2013. doi: 10.1002/jcc.23354.

[87] Johan Åqvist, Petra Wennerström, Martin Nervall, Sinisa Bjelic, and Bjørn O. Brandsdal. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters*, 384(4-6):288–294, 2004. doi: 10.1016/j.cplett.2003.12.039.

[88] Kim Hung Chow and David M. Ferguson. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer Physics Communications*, 91(1-3):283–289, 1995. doi: 10.1016/0010-4655(95)00059-o.

[89] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103(19):8577–8593, 1995. doi: 10.1063/1.470117.

[90] Peter J. Steinbach and Bernard R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry*, 15 (7):667–683, 1994. doi: 10.1002/jcc.540150702.

*Bibliography*

[91] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32:2319–2327, 2011. doi: 10.1002/jcc.21787.

[92] Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein. MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016. doi: 10.25080/majora-629E541A-00E.

[93] Stuart J. McElhany, Thomas J. Summers, Richard C. Shiery, and David C. Cantu. Analysis of the first ion coordination sphere: A toolkit to analyze the coordination sphere of ions. *Journal of Chemical Information and Modeling*, 63:2699–2706, 2023. doi: 10.1021/acs.jcim.3c00294.

[94] Douglas L. Theobald and Deborah S. Wuttke. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, 22: 2171–2172, 2006. doi: 10.1093/bioinformatics/btl332.

[95] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[96] D. Frishman and P. Argos. Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics*, 23:566–579, 1995.

[97] Robert M. Breece, Zhenxin Hu, Brian Bennett, Michael W. Crowder, and David L. Tierney. Motion of the zinc ions in catalysis by a dizinc metallo-β-lactamase. *Journal of the American Chemical Society*, 131:11642–11643, 2009. doi: 10.1021/ja902534b.

[98] Hong Min Zhang and Quan Hao. Crystal structure of NDM-1 reveals a common β-lactam hydrolysis mechanism. *The FASEB Journal*, 25:2574–2582, 2011. doi: 10.1096/fj.11-184036.

[99] Joanna E. Raczynska, Barbara Imiolczyk, Marlena Komorowska, Joanna Sliwiak, Justyna Czyrko-Horczak, Krzysztof Brzezinski, and Mariusz Jaskolski. Flexible loops of New Delhi metallo-β-lactamase modulate its activity towards different substrates. *International Journal of Biological Macromolecules*, 158:104–115, 2020. doi: 10.1016/j.ijbiomac.2020.04.219.

# A. Appendix

## A.1. PDB codes of X-ray structures without active-site bound ligands

| NDM-1 | VIM-2 | IMP-1 |
|---|---|---|
| 3SPU_A | 1KO3_A | 4C1F_B |
| 3SPU_B | 4BZ3_A | 4C1G_B |
| 3SPU_C | 4BZ3_B | 5EV6_A |
| 3SPU_D | 4NQ2_A | 5EV6_B |
| 3SPU_E | 5ACU_A | 5EV6_C |
| 5N0H_B | 5LM6_B | 5EV6_D |
| 5NBK_A | 5YD7_A | 5EWA_D |
| 5NBK_C | 5YD7_B | 5HH4_C |
| 5ZGT_A | 6TMB_A | 5HH4_D |
| 5ZGT_B | | 5Y5B_A |
| 5ZGU_A | | 6JKA_A |
| 5ZGV_A | | 6JKA_C |
| 5ZGV_B | | 6JKA_D |
| 5ZGW_A | | |
| 5ZGX_A | | |
| 5ZGY_A | | |
| 5ZGZ_A | | |
| 5ZH1_A | | |
| 5ZH1_B | | |
| 6TWT_A | | |
| 6TWT_B | | |
| 6ZYP_B | | |

## A.2. Custom patches with Set 1 parameters

```
read rtf card append
* HSD_NE2---Zn covalent patch
* HSE_ND1---Zn covalent patch
* Topology
*
36  1

PRES ZNHD           0.00 ! patch to covalently link HSD to Zn
                         ! Patch is 1-HSD and 2-Zn
BOND 1NE2 2ZN

PRES ZNHE           0.00 ! patch to covalently link HSE to Zn
                         ! Patch is 1-Hse and 2-Zn
BOND 1ND1 2ZN

end

read para card append flex

BONDS
! HSE (HSD) - Zn
NR2  ZN     20.000     2.07   ! Neutral His-Zn bond to maintain coordination

ANGLES
! HSE (HSD) - Zn
! Equilibrium value is the average value of 40 PDB p53 structures
CPH2 NR2  ZN     20.000   120.6500
CPH1 NR2  ZN     20.000   125.9800
SS   ZN   NR2    100.000  108.1100 ! Cys-Zn-His, equilibrium value based on PDB survey data
! Equilibrium value is the average value of 98.00
NR2  ZN   NR2    100.000  107.85   ! His-Zn-His, equilibrium value based on PDB survey data
                                   ! but standard deviation is large (SD = 10.32)

DIHEDRALS
! HSE (HSD) - Zn
SS   ZN   NR2  CPH1      0.0000  3     0.00
SS   ZN   NR2  CPH2      0.0000  3     0.00
CT2  SS   ZN   NR2       0.0000  3     0.00
ZN   NR2  CPH2 NR1       0.0000  3     0.00
ZN   NR2  CPH1 CPH1      0.0000  3     0.00
ZN   NR2  CPH1 CT2       0.0000  3     0.00
ZN   NR2  CPH2 HR1       0.0000  3     0.00
CS   SS   ZN   NR2       0.0000  3     0.00
! HSE(HSD) - Zn - HSE(HSD) (2 His ligands)
NR2  ZN   NR2  CPH1      0.0000  3     0.00
NR2  ZN   NR2  CPH2      0.0000  3     0.00
HR3  CPH1 NR2  ZN        0.0000  3     0.00

end

read rtf card append
* ASPP_OD1---Zn covalent patch
* Topology
*
36  1

!default first none last none
PRES ZNAP         0.00 ! patch to covalently link ASPP_OD1 to Zn
BOND 1OD1 2ZN

end

read para card append flex
* ASPP_OD1---Zn covalent patch
* Parameters
*
                                          ! ZN+RES: ATOM-TYPES
BONDS
OB  ZN    20.000    1.986                  ! ASPP: OD1-ZN
```

```
ANGLES
! ASPP-ZN angles
CD   OB   ZN      20.000   120.0000              ! ASPP: CG-OD1-ZN

! ASPP-ZN-CYM angles
OB   ZN   SS      20.000   111.0000 !5zh1       ! ASPP+CYM: OD1-ZN-SG

! ASPP-ZN-HSD/HSE angles
OB   ZN   NR2     20.000    98.0000 !5zh1       ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1

DIHEDRALS
! ASPP-ZN intramolecular dihedrals
CT2  CD   OB   ZN        0.0000  3     0.00 ! ASPP: CB-CG-OD1-ZN
OH1  CD   OB   ZN        0.0000  3     0.00 ! ASPP: OD2-CG-OD1-ZN

! ASPP-ZN-CYM dihedrals
CD   OB   ZN   SS        0.0000  3     0.00 ! ASPP+CYM: CG-OD1-ZN-SG
OB   ZN   SS   CS        0.0000  3     0.00 ! ASPP+CYM: OD1-ZN-SG-CB

! ASPP-ZN-HSD/HSE dihedrals
CD   OB   ZN   NR2       0.0000  3     0.00 ! ASPP+HSD/HSE: CG-OD1-ZN-NE2/ND1
OB   ZN   NR2  CPH1      0.0000  3     0.00 ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1-CG
OB   ZN   NR2  CPH2      0.0000  3     0.00 ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1-CE1

end

read rtf card append
* CYM_SG---Zn covalent patch
* Topology
*
36   1

PRES ZN_C          0.00 ! patch to covalently link anionic Cys to Zn
                        ! Patch is 1-Cys and 2-Zn
BOND 1SG 2ZN

end

read para card append flex

!Do NOT add parameters with all CGenFF atom types. These should be
!added to par_all36_cgenff.prm.

BOND
SS   ZN     20.000    2.32   ! Anionic Cys-Zn bond to maintain coordination

ANGLE
SS   ZN   SS    100.000   111.7700 ! equilibrium value based on PDB survey data
CS   SS   ZN     20.000    95.0000 ! Bo from ethylmethylsulfide CT2-S-CT3, FC guess
!System with two adjacent Zn2+ atoms ligated with CYS
ZN   SS   ZN    100.000   103.8000 ! PDB 6J9J ZN1801:ZN/CYS1516:SG/ZN1802:ZN = 103.793816, FC guess

DIHE
! Cys (CSN) - Zn
X    CT1  CS   X         0.2000  3     0.00 ! From X    CT1  CT2  X

CS   SS   ZN   SS        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
HA2  CS   SS   ZN        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
CT1  CS   SS   ZN        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
!System with two adjacent Zn2+ atoms ligated with CYS
ZN   SS   ZN   SS        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond

end
```

## A.3. Custom patches with Set 2 parameters

```
read rtf card append
* HSD_NE2---Zn covalent patch
* HSE_ND1---Zn covalent patch
* Topology
*
36  1

PRES ZNHD          0.00 ! patch to covalently link HSD to Zn
                        ! Patch is 1-HSD and 2-Zn
BOND 1NE2 2ZN
IMPR 1NE2 1CE1 1CD2 2ZN

PRES ZNHE          0.00 ! patch to covalently link HSE to Zn
                        ! Patch is 1-Hse and 2-Zn
BOND 1ND1 2ZN
IMPR 1ND1 1CE1 1CG 2ZN

end

read para card append flex

BONDS
! HSE (HSD) - Zn
NR2  ZN      20.000     2.07   ! Neutral His-Zn bond to maintain coordination

ANGLES
! HSE (HSD) - Zn
! Equilibrium value is the average value of 40 PDB p53 structures
CPH2 NR2   ZN      20.000   126.0000 ! 120.6500
CPH1 NR2   ZN      20.000   123.0000 ! 125.9800
SS   ZN    NR2    100.000   105.0000 ! Cys-Zn-His, equilibrium value based on PDB survey data
! Equilibrium value is the average value of 98.00
NR2  ZN    NR2    100.000   104.0000 ! His-Zn-His, equilibrium value based on PDB survey data
                                     ! but standard deviation is large (SD = 10.32)

DIHEDRALS
! HSE (HSD) - Zn
SS    ZN    NR2   CPH1     0.0000  3     0.00
SS    ZN    NR2   CPH2     0.0000  3     0.00
CT2   SS    ZN    NR2      0.0000  3     0.00
ZN    NR2   CPH2  NR1      0.0000  3     0.00
ZN    NR2   CPH1  CPH1     0.0000  3     0.00
ZN    NR2   CPH1  CT2      0.0000  3     0.00
ZN    NR2   CPH2  HR1      0.0000  3     0.00
CS    SS    ZN    NR2      0.0000  3     0.00
! HSE(HSD) - Zn - HSE(HSD) (2 His ligands)
NR2   ZN    NR2   CPH1     0.0000  3     0.00
NR2   ZN    NR2   CPH2     0.0000  3     0.00
HR3   CPH1  NR2   ZN       0.0000  3     0.00

IMPROPERS
!atom types           Kpsi                 psi0
NR2   CPH2 CPH1 ZN     20.0000         0     0.0000 !

end

read rtf card append
* ASPP_OD1---Zn covalent patch
* Topology
*
36  1

!default first none last none
PRES ZNAP        0.00 ! patch to covalently link ASPP_OD1 to Zn
BOND 1OD1 2ZN

end

read para card append flex
```

```
* ASPP_OD1---Zn covalent patch
* Parameters
*
                                      ! ZN+RES: ATOM-TYPES
BONDS
OB  ZN       20.000     1.986       ! ASPP: OD1-ZN

ANGLES
! ASPP-ZN angles
CD  OB  ZN     50.000   138.0000            ! ASPP: CG-OD1-ZN; 120+18 based on X-ray

! ASPP-ZN-CYM angles
OB  ZN  SS     20.000   105.0000    !111.0000 !5zh1     ! ASPP+CYM: OD1-ZN-SG

! ASPP-ZN-HSD/HSE angles
OB  ZN  NR2    20.000    91.0000    !98.0000 !5zh1      ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1

DIHEDRALS
! ASPP-ZN intramolecular dihedrals
CT2 CD  OB  ZN       0.0000  3     0.00 ! ASPP: CB-CG-OD1-ZN
OH1 CD  OB  ZN       0.0000  3     0.00 ! ASPP: OD2-CG-OD1-ZN

! ASPP-ZN-CYM dihedrals
CD  OB  ZN  SS       0.0000  3     0.00 ! ASPP+CYM: CG-OD1-ZN-SG
OB  ZN  SS  CS       0.0000  3     0.00 ! ASPP+CYM: OD1-ZN-SG-CB

! ASPP-ZN-HSD/HSE dihedrals
CD  OB  ZN  NR2      0.0000  3     0.00 ! ASPP+HSD/HSE: CG-OD1-ZN-NE2/ND1
OB  ZN  NR2 CPH1     0.0000  3     0.00 ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1-CG
OB  ZN  NR2 CPH2     0.0000  3     0.00 ! ASPP+HSD/HSE: OD1-ZN-NE2/ND1-CE1

end

read rtf card append
* CYM_SG---Zn covalent patch
* Topology
*
36  1

PRES ZN_C        0.00 ! patch to covalently link anionic Cys to Zn
                      ! Patch is 1-Cys and 2-Zn
BOND 1SG 2ZN

end

read para card append flex

BOND
SS  ZN       200.00     2.32   ! Anionic Cys-Zn bond to maintain coordination

ANGLE
SS  ZN  SS   100.000   111.7700 !equilibrium value based on PDB survey data
CS  SS  ZN    20.000   112.0000 !Bo from ethylmethylsulfide CT2-S-CT3, FC guess
ZN  SS  ZN   100.000   103.8000 !PDB 6J9J ZN1801:ZN/CYS1516:SG/ZN1802:ZN = 103.793816, FC guess

DIHE
! Cys (CSN) - Zn
X    CT1 CS   X        0.2000  3     0.00 ! From X    CT1  CT2  X

CS  SS  ZN  SS        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
HA2 CS  SS  ZN        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
CT1 CS  SS  ZN        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond
ZN  SS  ZN  SS        0.0000  3     0.00 ! Dummy for Anionic Cys-Zn bond

end

read para card append flex
* Modified parameters for ZN - S nonbonded interaction
* Parameters
*

NBFIX
ZN      SS        -0.3428     4.5000
```

*A. Appendix*

end