UNIVERISTÁ DEGLI STUDI DI MILANO
Doctoral School in Pharmaceutical Sciences - XXXIII Cycle
DIPARTIMENTO DI SCIENZE FARMACEUTICHE

Computational approaches to study metalloproteins and zinc-containing systems

CHIM/08

Silvia GERVASONI

Tutor: Prof. Alessandro PEDRETTI
Coordinator: Prof. Giancarlo ALDINI

Academic year 2019/2020

Abstract

Computational representation of metal-containing systems is a considerable challenge that derives from the complex nature of the metal coordination bond. Herein we tested computational methods with increasing level of accuracy to handle zinc-containing systems, starting from simplistic approaches and moving to more complex descriptions. Fragment-based molecular docking was applied for drug design of novel inhibitors of HDAC, a metallo-enzyme containing a catalytic zinc ion within the binding pocket. Although metal ion interactions were treated with classic approaches, some promising compounds have been identified. Then docking, MD simulations and MM-GBSA free energy calculations were applied to study the interaction between AChE and a new metallo-organic reactivator. The MD simulation was carried out using the 12-6-4 non-bonded model, which was able to maintain the integrity of the ligand and its correct octahedral geometry through all the trajectories. A similar workflow was applied also for drug discovery of new inhibitors of CN1, a di-peptidase containing two catalytic zinc ions within the binding pocket. Small molecules able to coordinate the metal ions were found by docking of different libraries, then the hit compounds underwent to MD simulations using the cationic dummy atoms approach, where we parametrized *de novo* the five-coordinated zinc ion. As a last step, the free binding energy was computed using the MM-GBSA method. Finally, a multiscale approach for the binding mode prediction of new metallo β-lactamase IMP-1 inhibitors was developed. Docking and MD simulation using a restrained non-bonded model were applied, followed by a QM/MM refinement step, which was able to recover the correct coordination geometry and to fix the errors introduced by the MD simulation. Different modelling techniques applied to representative case studies are reported, considering increasing level of accuracy and illustrating the advantages and weakness of each approach.

Contents

Chapter 1

General introduction

1.1 Importance of metalloproteins

The ubiquitous occurrence of metalloproteins in all organisms testify the essential role they play in biochemical processes [1]. Proteins require metal ions generally for either structural or catalytic functions. The former is necessary for protein folding where the major example is represented by the zinc-finger motif, which allows the protein to bind nucleic acids [2]. In the latter, metal ions are generally placed within the active site. Several classes of enzymes belong to this group such as oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases, where the metal ion participates in substrate recognition and binding as well as in catalytic processes (i.e. redox, electrostatic stabilizers and activators by increasing acidity, nucleophilicity and electrophilicity) [3]. Evidence of the importance of metalloproteins can be found in Protein Data Dank (PDB) [4] where almost one-third of protein crystal structures contains metal ions (Figure 1.1A), with Zn, Mg, Ca and Fe being the most abundant elements, and together they constitute almost the 85% of all the metal ions contained in proteins (Figure 1.1B).

Looking at these numbers, it is easy to imagine that metalloproteins are involved in several cellular vital functions such as DNA replication and repair [5], metabolism [6], gene regulation [7], transport [8], protein degradation [9], etc. They are also used

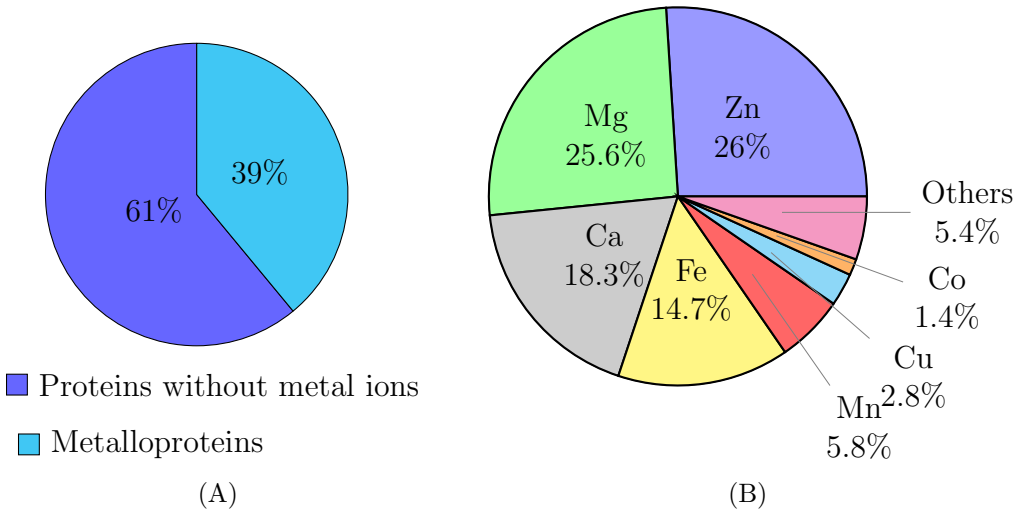


Figure 1.1: PDB statistics on metalloproteins (data retrieved on 26/02/2020). (A) Percentage of metal-containing (39 %) and not-containing (61 %) proteins. (B) Percentage of metal elements found in metalloproteins.

by microorganisms as survival mechanism like in bacteria, in the case of metallo-β-lactamases [10], or in viruses in the case of HIV integrase [11]. Their versatility has also made them an emerging tool for biocatalysis, with an increasing number of papers regarding artificial metalloenzymes [12].

For these reasons, metalloproteins have become a common target, especially in the pharmaceutical field [13]. Many new drugs approved by FDA target metalloproteins (Table 1.1), and the majority of these molecules are inhibitors that coordinate the metal ion within the active site preventing the substrate binding. For instance, histone deacetylases (HDACs) are well-known for their implication in cancer [14] and four inhibitors sharing the same mechanism of action (i.e. chelation of the catalytic zinc ion) were approved (Vorinostat, Belinostat, Panobinostat and Romidepsin) [15]. Tyrosinases are Cu-dependent enzymes that catalyse the conversion of L-tyrosine to L-DOPA followed by the oxidation of L-DOPA to L-DOPAquinone, which is further transformed into melanin. Tyrosinases are therefore a promising target for pharmaceutical regulation of melanin production [16]. As seen for HDACs, also for tyrosinases the inhibition strategy is metal-driven, where the only approved drug hydroquinone coordinates the copper ions through its hydroxyl group, preventing the entrance of the substrate. Several other drugs with this mechanism of action were approved for metalloenzymes by FDA, for a complete review see the work of Chen *et al.* [17].

This rising interest has brought to an increase of computational studies trying

Year	Number of FDA-approved drugs			NMEs for metalloenzyme targets		
	NME[a]	BLA[b]	Total	Not Metal-binding	Metal-binding	Total
2013	25	2	27	1	2	3
2014	30	11	41	2	3	5
2015	33	12	45	0	3	3
2016	15	7	22	0	1	1
2017	34	12	46	1	0	1
2013-2017	137	44	181	4	9	13

Table 1.1: FDA-approved drugs by year [17]. [a]NME: new molecular entities. [b]BLA: biologics license applications.

to interpret chemical phenomena and simulate biological mechanisms, but also to rationally design new drugs *via* ligand- and structure-based approaches [18]. There are many examples of successful computational pipelines that led to the identification of new drugs like Zolmitriptan for the treatment of migraine, Losartan for the hypertension [19], Saquinavir as HIV protease inhibitor, Dorzolamide for the glaucoma and Norfloxacin which targets bacterial DNA gyrase, which is a metelloenzyme [20]. Despite the results obtained so far, computational treatment of metalloproteins remains a critical issue that many researchers have tried to address.

1.2 Challenges of modelling metal ion-containing systems

The difficulties of the computational treatment of systems containing metal ions derive from the nature of the coordination bond, which is different from both ionic or covalent bond between s and p orbitals. Transition metals (TMs), unlike other elements of the periodic table, have d and f orbitals that give them unique properties. Given the orbitals degeneracy, TMs can have multiple spin and oxidation states that can lead to highly charged systems with pronounced long-range effects. The coordination bond is more labile than the classical covalent bond and this results in multiple coordination numbers associated with different coordination geometries, which depend also on the type of ligand that interacts with the TM. Furthermore, TMs are affected by the Jahn-Teller effect that distorts the structures of metal complexes [21]. Finally, even if TMs can form σ and π bonds, the d orbitals can

be implied also in back-bonding phenomena, further complicating the behaviour of these elements [22]. For all these reasons, the computational representation of this kind of interaction is quite complex.

In order to overcome this problem, several techniques have been proposed, which can be divided into two main categories: classical and quantum methods. In classical mechanics methods (molecular mechanics, MM), metal ions are described with force fields. The most popular ones are the Amber *ff*99 [23] and *ff*14SB [24], CHARMM 22 [25] and 36 [26], OPLS 2005 [27] and 3 [28], GROMOS 96 [29] and 11 [30]. The interaction with the metal ions can be represented through four models: non-bonded model, ligand field, bonded model and the Cationic Dummy Atoms (CDA). In the former the evaluation of the interactions between atoms is based only on the Coulombian and the Lennard-Jones potential [24], which confers large flexibility to the bond, but it remains quite simplistic and usually this model tends to underestimate the strength of the coordination bond [31]. The ligand field molecular mechanics model incorporates the ligand field stabilization energy directly into the potential energy equation of the force field [32]. In the bonded model metal ions are covalently bond with the coordinating atoms [33], resulting in greater stability compared with the non-bonded model, but this description does not allow neither the ligand exchange nor the change of coordination number that are important features of the coordination bond. In the CDA method the metal ion is considered without charge and dummy atoms with partial charges and no van der Waals radius are placed around the metal center. Therefore ligands will interact with the dummy atoms and not with the central metal ion [34]. One advantage of this model is the delocalization of the metal ion charge, and the flexibility of the bond. However, as the bonded model, also in CDA it is not possible to change the coordination number.

To overcome the limitations of the classical description, polarizable force fields have been developed (e.g. AMOEBA [35] and DRUDE [36]). These force fields have, as a great advantage, the possibility of changing the charge distribution in response to the environment, but their main drawback so far is the high computational cost [37].

Quantum mechanics (QM) methods can be classified into: *ab-initio*, semi-empirical and Density Functional Theory (DFT) [38].
The *ab-initio* methods solve the Schrödinger equation with different approximations and do not need external parameters. They can be further divided into Hartree-Fock (HF) and post-HF (i.e. MP2 [39], CC [40]). The former is quite fast and

return correct geometries, but overestimate the energy of the system, while the latter gives more accurate energies, but is computationally expensive [41]. Semi-empirical methods solve the Schrödinger equation with approximations but also with external parameters that could be obtained experimentally or by high-level *ab-initio* calculations [42]. Depending on the atoms involved in the system, several parametrizations are available (i.e. MNDO [43], AM1 [44], PM3 [45]). Compared to HF or post-HF, the semi-empirical methods are much faster, but also much less accurate. DFT methods, differently from the *ab-initio*, do not derive the properties of the molecules from the wave function, but from electron density, which is a physical observable. The energy of the molecule is a functional of the electron density and, in turn, the electron density is a function of the electron position (i.e. Cartesian coordinates). The DFT orbitals derived from the Kohn-Sham equations and their mathematical generation involve external parameters that are included into the exchange-correlation energy term [46]. DFT methods are faster than the *ab-initio* methods and generally give better results compared to the HF [47].

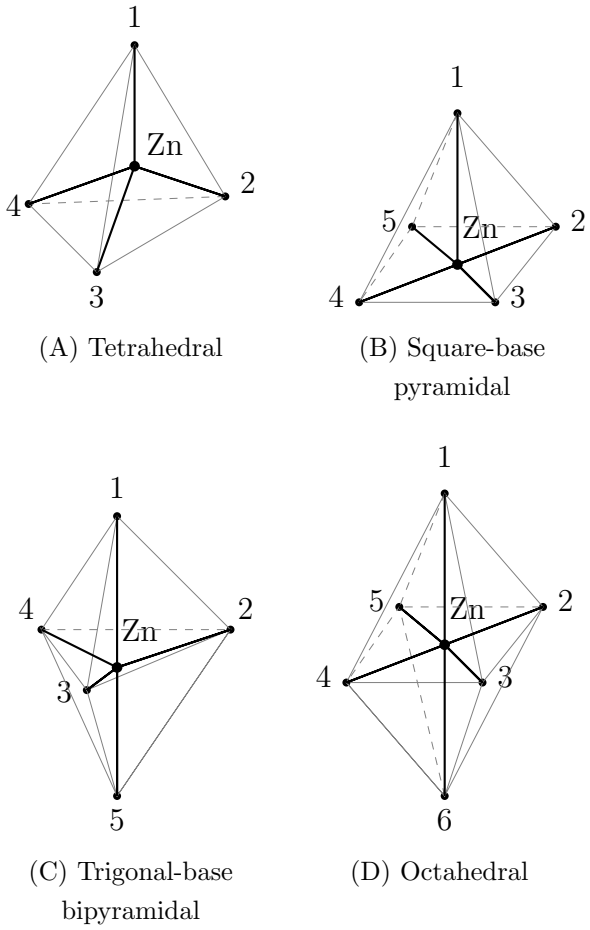
The main issue regarding QM methods is the computational cost, which limits the calculation to few hundreds of atoms. Thus, it is not possible to simulate an entire protein system with such methods. However, in 1976 Warshel and Levitt proposed a hybrid approach that combined QM with MM in one system [48], for which they won the Nobel Prize in 2013 with Martin Karplus [49]. Using this approach it is possible to model a small portion of the system with QM that, in the case of metalloproteins, would be the metal center and the nearest surroundings, while the other part of the system is treated with MM. QM/MM methods allow an accurate description of the active site of the protein, with a reasonable computational cost.

A further and more accurate description of some of these methods will be provided in **Chapter 2**.

1.3 The zinc ion

As shown in Figure 1.1B, zinc is the most abundant metal ion found in metalloproteins. Although it belongs to the *d* block of the periodic table, the zinc atom is not a proper TM since its *d* orbitals are filled. Nonetheless, it retains many peculiar properties of TMs. The zinc ion has oxidation number +2, and it can be coordinated

by four, five or six ligand atoms. Generally, given the d^{10} electronic configuration, it assumes preferentially tetrahedral geometry with four coordinating atoms [50]. However, it can be found also with coordination number of five, in which case it assumes either the trigonal bipyramidal or the square-base pyramidal geometries; and finally it can also organize into octahedral geometry with six-coordinating atoms (Figure 1.2).



(A) Tetrahedral

(B) Square-base pyramidal

(C) Trigonal-base bipyramidal

(D) Octahedral

Figure 1.2: Graphical representation of zinc coordination geometries. (A) In tetrahedral geometry the zinc ion is coordinated by four atoms, (B) in square-base pyramidal geometry by five atoms, (C) as well as in trigonal-base bipyramidal. (D) In the octahedral geometry the zinc ion is coordinated by six atoms.

In order to get insight into zinc geometries in metalloproteins, all zinc-containing proteins in PDB with a resolution smaller than 2.7 Å were collected[1] and the resulting subset of 3218 crystal structures were analysed. For each zinc ion in every structure, the coordination number, distances, angles and residues/atoms involved in the interactions were extracted (Table 1.2 and Table 1.3).

[1]Data retrieved in October 2018

					S	O	Odep	N	Npair
CN4	80.7%			min	1.32	1.59	1.42	1.75	1.64
				max	2.97	2.73	2.76	2.80	2.69
				mean	2.30	2.21	2.14	2.17	2.15
				stnd dev	0.44	0.37	0.37	0.34	0.35
CN5	13%	Square-base pyramidal	2.6%	min	2.19	1.55	1.70	1.69	1.80
				max	2.44	2.77	2.77	2.52	2.47
				mean	2.32	2.22	2.27	2.14	2.15
				stnd dev	0.10	0.37	0.37	0.37	0.34
		Trigonal bipyramidal	10.4%	max	2.23	1.59	1.70	1.94	1.72
				max	2.69	2.76	2.77	2.55	2.58
				mean	2.41	2.22	2.26	2.23	2.14
				stnd dev	0.16	0.37	0.37	0.40	0.35
CN6	6.3%			min	–	1.77	1.89	1.85	1.98
				max	–	2.77	2.76	2.72	2.64
				mean	–	2.28	2.31	2.20	2.22
				stnd dev	–	0.36	0.36	0.39	0.36

Table 1.2: Zinc coordination distances (Å) found in metallo-proteins. The data are divided according the coordination number of the zinc ion, and the percentage of each coordination geometry is reported. The coordinating atoms considered in the analysis were: sulfur (S), oxygen (O), deprotonated oxygen (Odep), nitrogen (N) and nitrogen with an available electron pair (Npair). For each atom type the minimum (min), maximum (max), average (mean) values of the coordination distance are reported, together with the corresponding standard deviation (stnd dev). These data were collected in October 2018.

Zinc ions were first filtered according to their coordination number, discarding the incomplete geometries with a number of ligands up to three, whereas, the four-, five- and six-coordinated were the actual subjects of the data analysis. The four-coordinated (CN4) zinc ions represented the highest percentage (80.7%, obs. 28431), the five-coordinated (CN5) ions were found in the 13% of the cases (obs. 457) and the six-coordinated (CN6) in the 6.3% (obs. 222), in agreement with other studies [51]. As a second step, the zinc ions were classified according to their geometry using as reference the ideal angles of the corresponding solids (Figure 1.2). The cases in which the sum of all the angles showed a difference of more than 100° from the ideal values were discarded to avoid the most distorted geometries. Ligands were also grouped according to the atom type, defining five classes: sulfur (S), oxygen (O), deprotonated oxygen (Odep), nitrogen (N) and nitrogen with an available electron pair (Npair). As expected, the increase of the coordination number led to a corre-

sponding increase of the coordination distances (Table 1.2). The sulfur coordination distances are the longest, however generally there is no significant difference between the other atom types.

	CN4	CN5				CN6		
		Square		Bipyramidal				
	109.5°	90°	180°	90°	120°	180°	90°	180°
min	55.5	15.0	135.1	18.3	105.0	150.2	42.9	135.5
max	178.1	135.0	179.7	104.9	149.7	179.5	134.7	179.8
mean	115.1	94.7	156.1	82.7	123.1	163.6	90.4	161.1
stnd dev	25.2	23.1	16.2	16.7	14.5	18.3	20.5	14.9

Table 1.3: Zinc coordination angles (°) found in metallo proteins. Data are divided according to the four zinc geometries: tetrahedral (CN4), square-base pyramidal (CN5 square), trigonal-base bipyramidal (CN5 bipyramidal) and octahedral (CN6). Angles were classified according to the ideal angles of the corresponding solids. For each coordination angle type the minimum (min), maximum (max), average (mean) values are reported, together with the corresponding standard deviation (stnd dev). Data collected in October 2018.

Table 1.3 shows that the coordination angles are often distorted, especially those which are close to 180°. Looking at the frequency of the angles, we found indeed a Gaussian-like distribution for CN4 (Figure 1.3A), while for CN6 we found two main picks corresponding to the 90° and 180° angles, but also a smaller pick at 55°, which corresponds to the square-base distortion (Figure 1.3B).

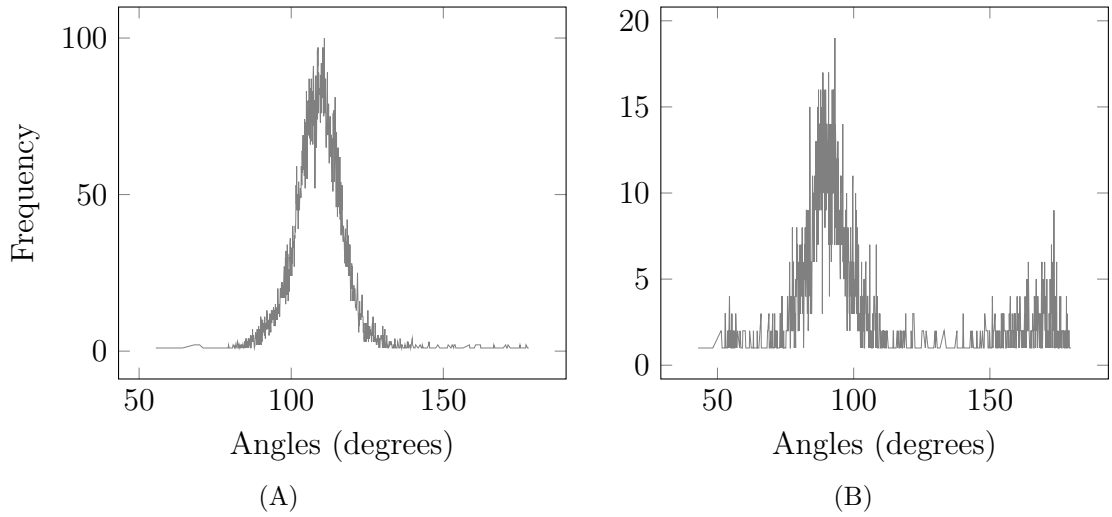


Figure 1.3: Coordination angle (°) frequency distribution, according to zinc ion geometry. (A) Tetrahedral zinc ions present a Gaussian-like distribution with a pick in correspondence to almost 110°. (B) Octahedral zinc ions show two main picks at 90° and 180°, corresponding to the ideal angle of the octahedron, and a minor at 55°, corresponding to the distorted square-bases.

Finally, we analysed which types of proteins contain the zinc ion. The majority

of zinc-containing proteins are hydrolases, followed by proteins involved in gene regulation (Figure1.4). As mentioned before, the catalytic role of zinc ion derives from its capability to stabilise and activate the ligand increasing its acidity, nucleophilicity or electrophilicity; these characteristics make the zinc ion a good candidate for hydrolases. Whereas, the gene regulation as well as the DNA/RNA binding proteins exploit mainly the zinc-finger structures.

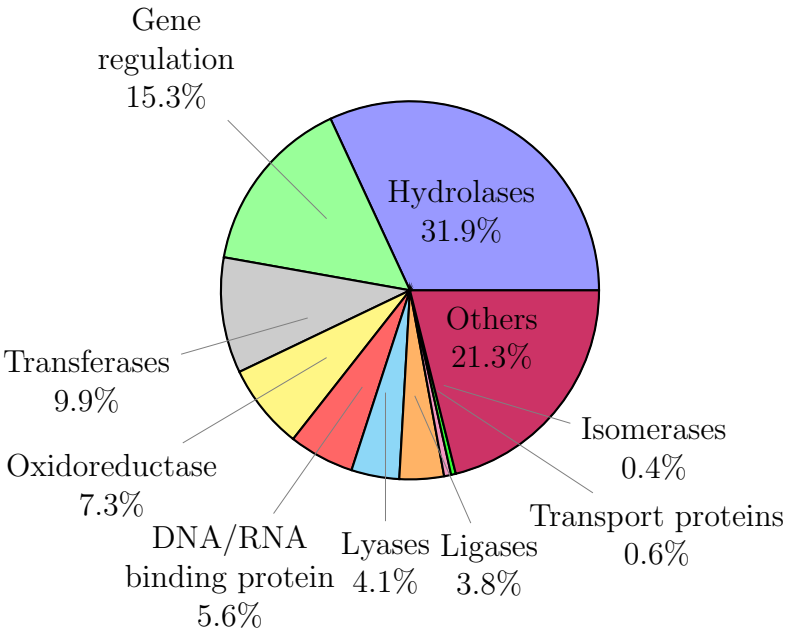


Figure 1.4: Percentage distribution of different zinc proteins. The major group is represented by hydrolases (31.9%), followed by gene regulation proteins (15.3 %).

1.4 Aim of the Thesis

The research project reported in this Thesis takes into account a representative set of case studies of zinc-containing systems, considering both metallo-enzymes and metallo-organic compounds. The computational treatments of the metal ion interactions follow a path of increasing complexity through the chapters.
When the analysis involves screening of high number of molecules, the metal ion can be treated by applying canonical methods based on molecular mechanics and introducing some empirical corrective factors to account for the specific interactions stabilized by the metals. This is the typical situation experienced when performing molecular docking analyses, and it is suitable for preliminary studies.
Whereas, when performing MD simulations of metal-containing systems, canonical MM methods cannot properly handle the complex behaviour of metal ions. As

mentioned before, classical force fields evaluate metal interaction using the classic non-bonded model, which is a summation of the electrostatic and van der Waals potential contributions. While monovalent metal ions (e.g. Na^+, K^+) are correctly handle with this model, problems arise when divalent (or more) ions are involved. A common risk of using the classic non-bonded model to treat this type of ions is to obtain an altered coordination geometry and unrealistic ligand exchanges, to the extent that a coordinated water molecule can take the place of a chelating ring. Therefore, corrections to the classic representation need to be applied to strengthen the metal coordination bond and to define a proper metal coordination geometry that would be lost using classic MM approaches. These modifications can be introduced for simulations of metallo-organic small molecules, but also for more complex environment such as metallo-proteins. This last case is of particularly difficult parametrization since the coordination geometry is not always well defined. To date, effective MM methods that describe metal interactions employ parametrizations that do not allow ligand exchanges (bonded model) or coordination number and geometry changes.
When a more exhaustive description is required, the metal ion role cannot be simplified by any MM approach. Therefore, specific simulation methodologies, such as QM/MM techniques, have to be implemented to accurately treat the metal ion and its interaction capability. In this case, neither the geometry nor the coordinating ligands need to be define *a priori* and, at the same time, the simulation errors described above are overcome. Challenges derive from the choice of the structure to use as starting point for the calculation and of the theory level to use. These two variable can indeed change the result of the calculation. Nonetheless, so far QM/MM methods are the most accurate technique for describing complex environments, like enzymes.
For each approach advantages and weaknesses are highlighted, with the goal to find the best compromise between computational cost and accuracy.
On this ground, the research followed an ideal pathway starting from simplest situations to arrive to most complex representations.

1.5 Thesis overview

The aim of the present research Thesis is the identification and evaluation of different modelling techniques, with increasing accuracy levels, to simulate various zinc-containing systems. Eventually the most successful workflow for each system will be identified, with the goal to be applied in similar cases for rational drug design.

Chapter 2 is an overview on the main computational techniques used in this Thesis. Particular emphasis is put on molecular docking, metal parametrization in molecular dynamics and QM/MM hybrid systems. For each method a brief theoretical background is given, together with a short comment on the corresponding advantages and disadvantages.

1.5.1 Chapter 3: Identification of novel HDAC inhibitors by fragment-based drug design

Histone deacetylases (HDACs) are zinc-enzymes involved in cancer development, therefore inhibitors have been proposed to block their activity. Generally HDAC inhibitors are composed of three moieties: CAP, linker and zinc binding group (ZBG). This last is often a hydroxamic acid, which has strong chelating ability, but it also brings different side effects. The aim of this work was to design new putative HDAC inhibitors without the hydroxamic acid as ZBG. To this end a fragment-based approach was applied. Three datasets corresponding to the three moieties of the HDAC inhibitors were collected. The ZBG dataset contained general metal chelator fragments, while the linker and CAP datasets included fragments derived from clinically approved HDAC inhibitors. The datasets were docked into two hybrid pockets obtained by computing the average 3D coordinates of the atoms of HDAC class I and IIB (i.e. HDAC1, 2, 3,6 and 8), and HDAC class IIA (i.e. HDAC4 and 7). The best fragments were combined with each others, obtaining the whole ligands. Furthermore, a QSAR model was developed through linear regression, using experimental data retrieved from literature, considering descriptors of both the ligands and the force field. In detail, interaction energies were computed using the pair module of NAMD2 and the CHARMM22 force field, with the classic non-bonded model. The model was applied for the activity prediction of the new molecules towards the isoform HDAC1. Promising compounds able to coordinate the zinc ion, without the hydroxamic acid, were obtained.

1.5.2 Chapter 4: Insight into the role of the zinc ion in *h*AChE reactivation: computational studies of new metallo-organic derivatives

Acetylcholinesterase (AChE) is a serine hydrolase that terminates the nervous signal through hydrolyzation of the neurotransmitter acetylcholine. Organophosphorus (OP) compounds contained in several warfare agents and pesticides are able to covalently bind AChE on the serine of the catalytic triad leading to its irreversible inhibition and, as a result, to severe effects and eventually to death. To date, the oxime-based compounds represent the most effective class of AChE reactivators. Thanks to its nucleophilicity, the oximate can perform a nucleophilic attack towards the phosphorus atom of the OP, releasing the catalytic serine. However, the efficacy of such reactivators is quite low and depends on the type of OP bound to AChE. In this chapter, a new set of zinc-containing reactivators were tested against AchE and compared to a known reactivator (i.e. HI-6). In the first part of the work, the influence of the metal ion on the nucleophilicity was assessed by semi-empirical and DFT calculations. As expected, the insertion of the zinc ion increased the negative charge of the oxime oxygen, thus increasing the nucleophilicity. In the second part of the work, docking, MD simulation and MM-GBSA free energy calculations were applied to predict the binding mode of one of these metal-containing molecules (i.e. K2), and to rationally explain the differences in efficacy with respect to HI-6. A docking protocol similar to induced-fit docking was used: K2 was first docked using flexible docking, then the complex underwent to MD simulation and finally K2 was re-docked within the AChE structure derived from the MD. A 600 ns long MD followed the docking, where K2 was parametrized using the 12-6-4 non-bonded model, which maintained the correct octahedral geometry. The OP-K2 distance was monitored along the trajectory and the binding free energy was evaluated. The results suggest that Trp86 plays a crucial role in stabilising K2 at the active site of the catalytic pocket. However, the bulky shape of the ligand hinders the descent along the narrow gorge of AChE, decreasing the reactivation efficacy.

1.5.3 Chapter 5: Validation of CN1 pharmacophore and discovery of new putative inhibitors through a combined approach

Carnosine (β-alanyl-L-histidine) is a natural dipeptide that have been described as a potential pharmacological agent owing to some positive outcomes from several pharmacological tests in animal models of human diseases. However, carnosine has limited activity in humans since the peptide upon absorption is rapidly hydrolyzed in the serum by the metallo-enzyme carnosinase (i.e. CN1; E.C. 3.4.13.20). Recently, the hypothesis of co-administration of carnosine and selective inhibitors of CN1 has been proposed. CN1 contains two catalytic zinc ions within its binding pocket, therefore one of the most promising strategy is to design metal-chelator binders. The work reported in this chapter is divided into two parts: (1) the molecular interpretation of the serum hydrolytic rate of histidine dipeptides and (2) the screening of five different datasets to identify new putative inhibitors. (1) So far a resolved complex of CN1 bound to a ligand is not available, therefore docking and DFT single point energy calculations were performed on four histidine dipeptides (i.e. carnosine, anserine, balenine and homocarnosine), and the results were validated using the serum hydrolytic rate as measure of their binding affinity. (2) Once defined a CN1 molecular interaction model, five databases including known drugs (repurposing) and natural compounds were collected and screened on CN1. The compounds with the best docking score underwent to a 100 ns molecular dynamics simulation to asses the stability of the docking pose. Particular attention was paid on the treatment of zinc ion coordination, where the cationic dummy atom approach was chosen and a new parametrization with trigonal bipyramidal geometry was designed. Finally, the binding energy of these molecules were evaluated through MM-GBSA free energy calculation. Five promising compounds were selected, which elicited a stable interaction with Arg350, as seen for carnosine and its analogues, and they appeared to have comparable binding energy.
The first part of the work described in this chapter has already been published [52].

1.5.4 Chapter 6: A multiscale approach to predict the binding mode of metallo β-lactamase inhibitors

Antibiotic resistance is a major threat to global public health. β-lactamases, which catalyse breakdown of β-lactam antibiotics, are the principal cause. Metallo β-lactamases (MBLs) represent a particular challenge because they hydrolyse almost all β-lactams, and to date no MBL inhibitor has been approved for clinical use. Molecular simulations can aid drug discovery programs, but empirical molecular mechanics (MM) methods often perform poorly for metalloproteins. In this chapter a multiscale approach is presented to model the inhibitor binding to IMP-1, a clinically important MBL containing two catalytic zinc ions, and to predict the binding mode of a novel inhibitor. Inhibitors were first docked into the IMP-1 active site, testing different docking programs and scoring functions on multiple crystal structures. Complexes were then subjected to molecular dynamics (MD) simulations and subsequently refined through QM/MM optimization with the B3LYP/6-31G(d) DFT method, thus increasing the accuracy of the method with successive steps. This workflow was tested on two IMP-1 thiolate complexes, for which it successfully reproduced the crystallographically observed binding. Therefore, this workflow was applied to predict the binding mode of an inhibitor for which a complex structure was crystallographically intractable. A 12-6-4 model was also tested in MD simulations, as well as an optimization with a SCC-DFTB QM/MM approach. The results show limitations of empirical models for treating these systems and indicated the need for higher level calculations (e.g. with DFT/MM) for reliable structural predictions. This work demonstrates a reliable computational pipeline that can be applied to design further MBL inhibitors and to other Zn-metalloenzyme systems.

Chapter 2

Theoretical background

2.1 Molecular docking

Molecular docking is a structure-based method that predicts how two molecules (usually a ligand and a protein) interact with each other's. This approach is quite fast and for this reason it is widely used in virtual screening campaigns [53]. The binding pose is achieved through two main processes: the conformational sampling and the ranking. The former samples the torsions of the ligand and, in the case of flexible docking, of the side chains of the protein, and tries to fit the ligand inside a defined binding pocket to optimize the steric complementarity and the non-bonded interactions. The latter evaluates the strength of the interactions and it gives a score, which is used to rank the poses from the best to the worse [54].

Several docking methods have been developed (i.e. protein-protein docking [55], induced-fit docking [56], ensemble docking [57], reverse docking [58], covalent docking [59], etc.), but here we will focus on protein-ligand docking considering both rigid protein with flexible ligand and flexible ligand and protein.

2.1.1 Sampling algorithms

Considering the conformational degree of freedom of the ligand (and in flexible docking, also of the protein side chains) as well as the freedom of rotation and trans-

lation, a huge number of binding modes can be computed for each protein-ligand complex. The generation of every possible complex would be too computationally expensive, so various sampling algorithms have been developed.

LUDI [60] is based on interaction sites, which are points in the space available for hydrophobic contacts or hydrogen bonds. The ligand is fitted into the interaction sites optimizing the distance between every point of interaction.

An important class of sampling algorithms are the stochastic methods, including Monte Carlo (MC) [61] and Genetic Algorithm (GA) [62] as the most popular. MC methods generate multiple docking poses through random rotation and translation of the ligand within the binding pocket. The conformations obtained with this procedure are evaluated by an energy-based threshold. If the threshold is achieved, the complex will be kept and further refined. Otherwise, the value of $f = e(-\Delta E/RT)$ is computed together with a generation of a random number from 0 to 1, if this number is minor than f, than the pose is retained for further refinement as well (i.e. Metropolis criterion) [63]. This process is iterated until a certain number of poses is collected. GA methods take a cue from Darwin's theory of evolution: the ligand is encoded by binary strings, called genes, that together make a chromosome, which represents the docking pose. The algorithm mutates randomly the chromosome and exchanges genes between chromosomes as a genetic cross-over. This process is iterated carrying on the recombinations that give the best fitness until a threshold is reached (Figure 2.1). AutoDock [64] and GOLD [65] programs use the GA method.

The Ant Colony Optimization (ACO) [66], used by the PLANTS software [67], is a method that follows the behaviour of ants in the search of the shortest path to food. In nature, when ants are looking for food, they deposit a trail of pheromone that will be followed by other ants. When the path to the food is short, the pheromone trail will be strong and many ants will follow that path. On the contrary, if the path is too long the pheromone trail will fade and few ants will follow that path. As a consequence, the intensity of the pheromone trail indicates of the quality of the path. ACO is a graph-based algorithm G = (N,A), in which every node (N) is connected to each other through an arc (i, j). The arc is associated to a variable $\tau_{i,j}$ called pheromone trail. The value of $\tau_{i,j}$ is an index of the utility of the arc to obtain the best solution. The path through the nodes followed by ants is defined by stochastic decisions. All arcs start with the same amount of pheromone, the probability that the ant k moves from the node i to the node j using the trail $\tau_{i,j}$ is given by Equation 1, where N_i^k represents all the connected nodes beside the

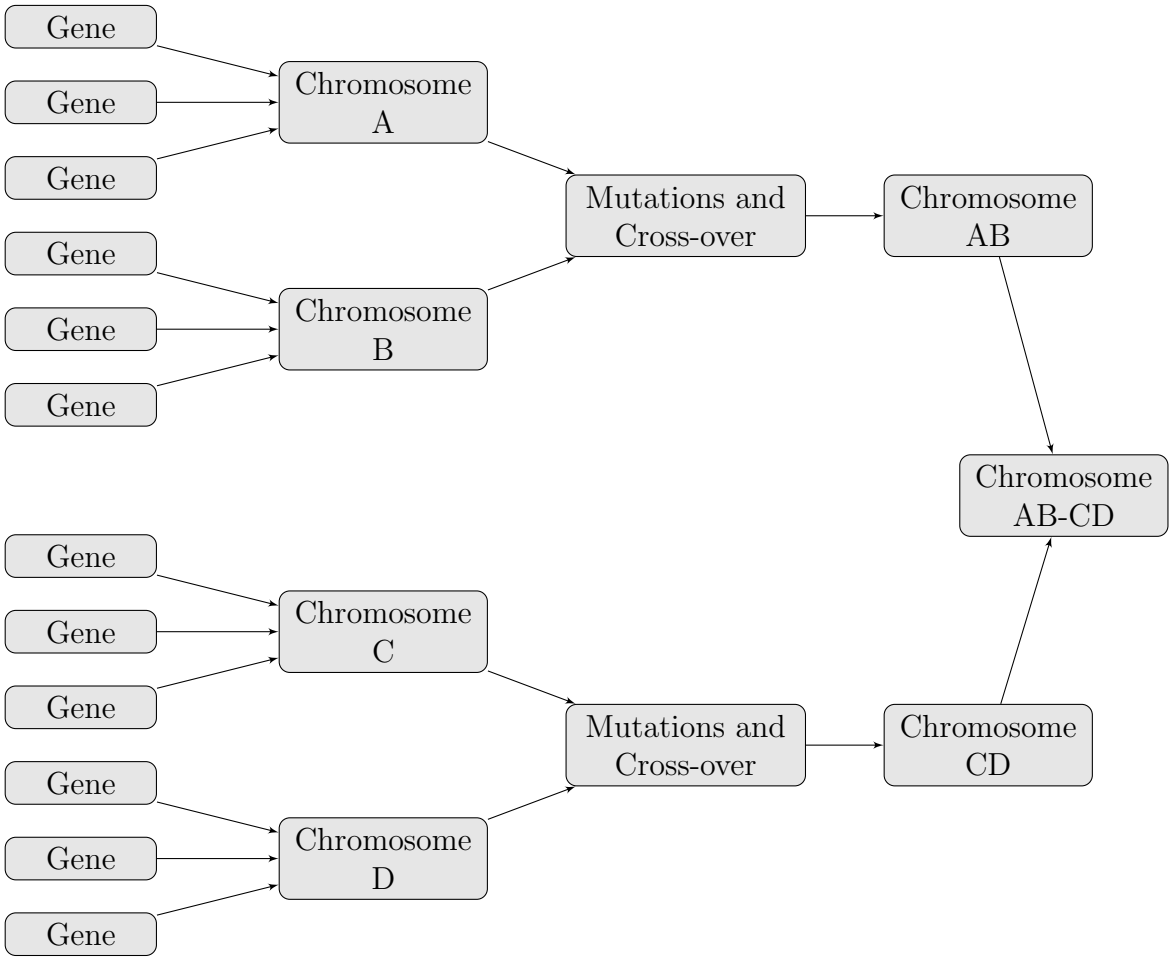


Figure 2.1: Schematic representation of genetic algorithm. The genes are the properties of the ligands (e.g. angle, torsion, hydrophobic and hydrogen contacts). The chromosomes are the poses, which undergo to recombination (which includes mutations and cross-over events) to obtain new chromosomes with an improved fitness. The process is iterated generation after generation, until a certain number of poses is reached.

predecessor node (i.e. ants can only move forward). At each iteration the pheromone level is updated using Equation 2.1-2.2 (Figure 2.2).

$$p_{ij}(k) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}}{\sum_{l \in N_l^k} \tau_{ij}^{\alpha}} & \text{if } j \in N_i^k \\ 0 & \text{if } j \notin N_i^k \end{cases} \tag{2.1}$$

$$\tau_{ij}(k+1) = \rho \tau_{ij}(k) + \Delta\tau_{ij}(k) \tag{2.2}$$

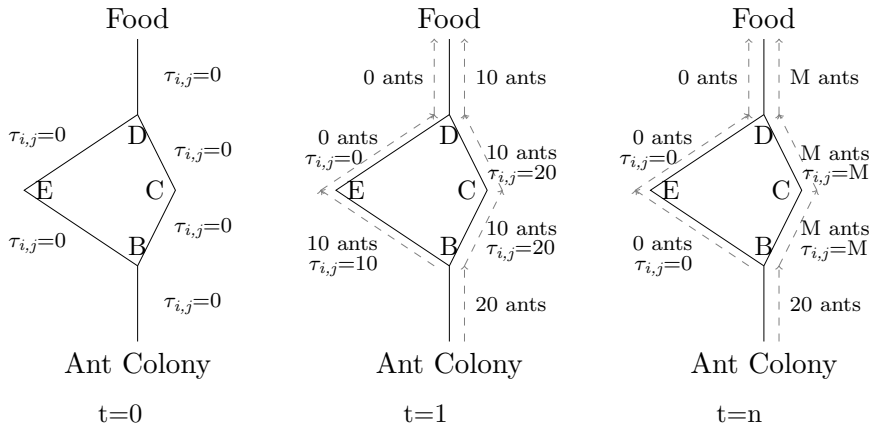


Figure 2.2: Schematic representation of ACO algorithm. The starting point is located at the Ant Colony, while the end point at the Food. The nodes are indicated with capital letters. At time zero, the pheromone trail $\tau_{i,j}$ is the same for each node, so at t=1 the same amount of ants reach nodes C and E. However, the B-E-D path is longer than the B-C-D path. As a result, at time n, all the ants will follow the B-C-D path, and none the B-E-D.

2.1.2 Scoring functions

As mentioned before, scoring functions evaluate the strength of the interaction between the ligand and the protein. The aim of scoring functions is not to compute the physics in protein-ligand binding process, but to provide a compromise between accuracy and speed [68]. Several scoring functions have been developed, and they can be classified into four groups: force field-based [69], empirical [70], knowledge-based [71] and machine learning-based [72]. Force field-based scoring functions use bonded and non-bonded terms of force field to compute the score, thus employing their parameterization. DOCK is one popular example of docking software which implements this type of scoring function, that is derived from the Amber force field [73]. The major drawback of this class is the inability to account for the desolvation effect, leading to bias towards Coulombic electrostatic interactions and, consequently, to select highly charged ligands [69]. To fix this issue, force field scoring functions were augmented by solvation energy terms, derived from either Poisson−Boltzmann (PB) or Generalized Born (GB) models [74]. The aim of empirical scoring functions is to reproduce the experimental affinity data [75] through correlation of the binding free energy to a set of variables (i.e. van der Waals, electrostatics, hydrogen bonds, etc.). These variables are associated with coefficients, obtained by regression of experimentally binding affinity data [70]. Some examples are LUDI [60], ChemScore [76] or GlideScore [77]. Knowledge-based scoring functions (e.g. DrugScore [78] and PMF-Score [79]) sum pairwise statistical potentials between protein and ligand, retrieved from experimentally determined protein-ligand complexes (Protein

Data Bank or Cambridge Structural Database). Generally they derive a distance-dependent potential using the Boltzmann law [71]. Finally, machine learning-based scoring functions are the newest. Starting from a dataset of protein-ligand complexes with annotated binding affinities (e.g. DUD-E [80]), one popular choice is to use atom pairs or structural interaction fingerprints between protein and ligand as descriptors [72]. These descriptors account for specific interactions (electrostatics, hydrogen bond, etc.), geometrical descriptors (surface or shape) and conventional ligand-based descriptors [81]. RF-Score [82] and NN-Score [83] are the most used scoring functions of this class.

Concerning how scoring functions treat metal coordination bond, two main approaches can be found. On one hand, scoring functions can implicitly evaluate the coordination bond by the classic non-bonded interaction (i.e. van der Waals and electrostatic potential), as in the case of AutoDock and AutoDock Vina (Equation 2.3)[84].

$$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hb} + \Delta G_{elect} + \Delta G_{solv} + \Delta G_{tor} \tag{2.3}$$

On the other hand, the evaluation of metal interactions can be treated explicitly, like in the case of the ChemPLP scoring function (Equation 2.4) with the terms f_{met} [85].

$$\begin{aligned} f_{PLANTS_{ChemPLP}} = f_{PLP} + f_{hb} + f_{hb-ch} + f_{hb-CHO} \\ + f_{met} + f_{met-coord} - f_{met-ch} + f_{met-coord-ch} + f_{clash} + c_{site} \end{aligned} \tag{2.4}$$

In this case, a common function is shared through all the energy terms (Equation 2.5). Equation 2.6 shows the example of the metal coordination bond function, which considers both the bond length ($f(p_r, 2.6, 3.0)$) and the angles ($f(|q_\beta - 180|, 80, 90)$).

$$f(x, x_1, x_2) := \begin{cases} 1 & if x \leq x_1 \\ \frac{x_2 - x}{x_2 - x1} & if x_1 < x \leq x_2 \\ 0 & if x > x_2 \end{cases} \tag{2.5}$$

$$f_{met} = w_{met} \cdot \sum_{p \in P_{acc-met}} f(p_r, 2.6, 3.0) \cdot \prod_{q \in p_{acc-nb}} f(|q_\beta - 180|, 80, 90) \tag{2.6}$$

Although this last type of scoring function seems to have a more solid evaluation of the coordination bond, it relays on parameterization and experimental data fitting, therefore various coordination geometries are neglected.

Docking is designed to be fast, and it is indeed used in virtual screening campaigns when datasets of thousands of compounds need to be handled. Although it could give pretty accurate results [86], when it comes to peculiar ligands or proteins, docking performance can drop [87]. The presence of metal ions represents indeed a challenge in docking, in particular due to the inability of scoring functions to correctly treat the metal ion interactions [88]. This is true for metal ions into either the ligand [89] or into the protein [90][91]. For this reason different strategies have been developed, like scoring function optimization for metal ions [92], restrained docking [88] or multi-step approaches [93][94][95].

2.2 MD Simulation

Molecular dynamics (MD) simulates atom behaviour during a certain period of time, and subject to a force [96]. An energy function (force field) computes the force experienced by every atoms in the system, while Newton's second law defines how those forces affect the atom motions. The simulation is divided into time steps, at each time step forces acting on every atom are computed and atom positions are updated. The force on a certain position in the Cartesian space r is given by the negative gradient of the potential energy U (Equation 2.7).

$$F(x) = -\nabla U(r) \tag{2.7}$$

Considering that velocity is the derivative of position and acceleration of the velocity, the equation of motion can be written as Equations 2.8, where for n atoms there are $3n$ position and $3n$ velocity coordinates.

$$\begin{aligned} \frac{dr}{dt} &= v \\ \frac{dv}{dt} &= \frac{F(r)}{m} \end{aligned} \tag{2.8}$$

An algebraic solution to this problem is impossible, therefore numerical methods are required, such as integrators. The leapfrog method [97] is the one used in this Thesis. The starting velocities of atoms are assigned with random directions and a fixed magnitude depending on temperature. Considering the time step size as $h = \Delta t$, the calculation is divided into two parts: first, the velocities are updated using the starting acceleration values (v_{ix}), while the coordinates (r_{ix}) are updated by a full time step using those preliminary velocities (Equations 2.9).

$$\begin{aligned} r_{ix}(t+h) &= r_{ix}(t) + h v_{ix}(t + \frac{h}{2}) \\ v_{ix}(t+\frac{h}{2}) &= v_{ix}(t) + \frac{h}{2} a_{ix}(t) \end{aligned} \tag{2.9}$$

Finally, these new coordinates are used to compute the final acceleration values and to update the velocities, for the second time step (Equation 2.10).

$$v_{ix}(t+h) = v_{ix}(t+\frac{h}{2}) + \frac{h}{2} a_{ix}(t+h) \tag{2.10}$$

It is clear that the choice of the correct time step is essential. Generally, the time step size should be at least 8-10 times minor than the fastest movement of the system [98]. In proteins such a movement is the C-H bond stretch, which occurs on average every 10 fs, therefore the time step should be 1 or 0.5 fs long. However, algorithms have been developed to constraint this bond (e.g. SHAKE [99], LINCS [100]), allowing the use of a larger time step (usually 2 fs) and, as a results, speeding

the simulation.

When performing a MD simulation, the system can be described by thermodynamic ensembles, which provide a set of configurations distributed according to: microcanonical NVE (constant number of particles N, volume V, energy E), canonical NVT (constant number of particles N, volume V, temperature T) or isothermal-isobaric NTP (constant number of particles N, temperature T, pressure P) ensembles [101]. In order to keep a constant temperature and pressure of the system, thermostat and barostat are used along the simulation (Lavengin and Berendsen are the two used in this Thesis) [102].

The calculation of non-bonded interactions is the most expensive part of an MD simulation. The short-range interaction are usually handle with a cut-off value, beyond that the interactions are smoothly switched to zero. Whereas, the long-range interactions are treated with the Particle Mesh Ewald (PME) method, which is an improvement of the traditional Ewald summation [103]. PME method is associated with the Periodic Boundary Condition (PBC), an infinite periodic systems, as the name suggests, in which a central box containing the system is duplicated infinite times in all directions of space. As a result, when a particle leaves the central box from an edge, it is replaced by a replica coming from one of its periodic images. Duplicate interactions are avoided by the minimum image criterion, which states that one particle can only interact with a single image of every other particle.

2.2.1 Force field and metal ion models

As mentioned before, in order to numerically solve the Newton's equation of motion for a system, it is necessary to know the potential energy, which should be fast to compute and straightforward. To this end, force fields are used: they are based on parameters obtained either from high-level quantum mechanics calculations or experimental measurements. The general functional form of force field is reported in Equation 2.11 in which bonded terms (bonds, angles and dihedrals, Equation 2.12-2.14) and non-bonded terms (van der Waals and Coulomb potentials, Equation 2.15-2.16) appear.

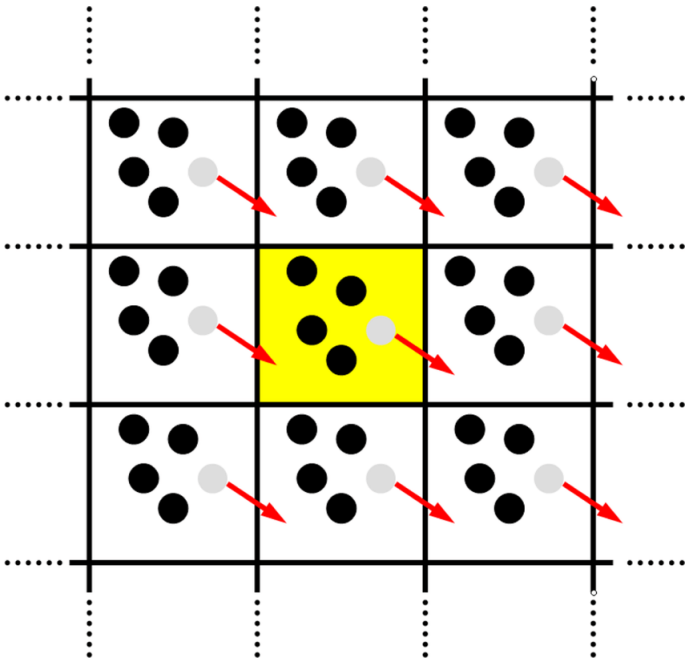


Figure 2.3: Graphic representation of PBC. The central box is coloured in yellow, while the replica are white. The red arrows point the movement of the particles from one box to another. (Figure from reference [104])

.

$$E_{tot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdw} + E_{coulomb} \tag{2.11}$$

$$E_{bonds}(r_{ij}) = \frac{1}{2}k^r_{ij}(r_{ij} - r^0_{ij})^2 \tag{2.12}$$

$$E_{angles}(\theta_{ijk}) = \frac{1}{2}k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2 \tag{2.13}$$

$$E_{dihedrals}(\phi_{ijkl} = \frac{1}{2}k^n_\phi(1 + cos(n\phi_{ijkl} - \gamma_n)) \tag{2.14}$$

$$E_{vdw}(r_{ij}) = \frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^6_{ij}} \tag{2.15}$$

$$E_{coulomb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r ij} \tag{2.16}$$

Force fields for proteins are massively used and well parametrized [105], while more difficulties are found in nucleic acids, small molecules and metal ions [106]. However, nucleic acids force fields have been recently optimized [107][108] and for small molecules general force field (e.g. GAFF [109] and CGenFF [110]) or quantum mechanics parametrization are often used [111]. Metal ion force fields lack an unique description instead and often relay on simplistic non-bonded model [112]. In this Thesis, particular attention is paid in the treatment of metal interactions, and

different models are tested. As mentioned in **Chapter 1.2**, non-bonded, bonded, ligand field and Cationic Dummy Atoms (CDA) models can be used to describe interactions involving metal ions. In this Thesis non-bonded and CDA are used.
The classic non-bonded model sums the contributions of van der Waals and electrostatics, which are computed by the Lennard-Jones (LJ) and Coulomb potential respectively (Equation 15-16). Generally this model can work on monovalent metal ions, whereas it is not able to correctly describe the behaviour of divalent or highly charged metal ions [113][114]. Therefore, it can be used in combination with distance restraints (restrained non-bonded model) to avoid any change in coordination. Moreover, the classic LJ 12-6 potential was improved for divalent and highly charged metal ion by the LJ 12-6-4, in which an r^{-4} term was added to describe the ion-induced dipole interaction (Figure 2.4) [115][116].

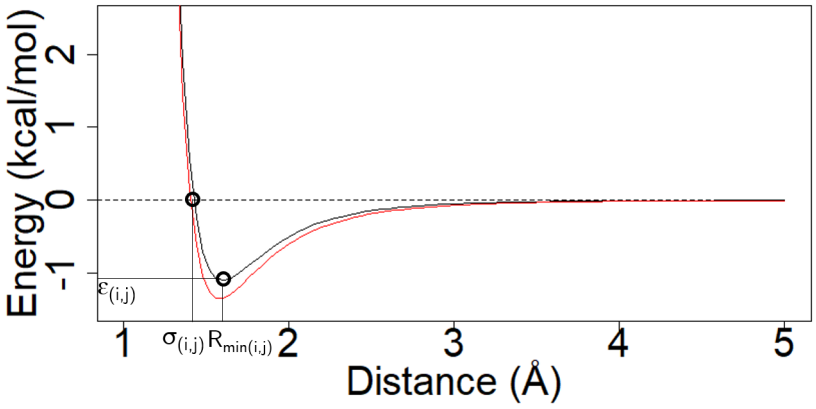


Figure 2.4: Van der Waals potential between two particles i, j. The classic LJ 12-6 model is represented as a black line, while the LJ 12-6-4 model is reported as a red line. $R_{min,i,j}$ is the distance (Å) between i and j at which the potential has its minimum value. $\sigma_{i.j}$ is the distance (Å) between i and j at which the potential is equal to zero. ϵ_{ij} is the well depth of the LJ curve.

In the CDA model dummy atoms are added in correspondence of coordination sites. These atoms have neither mass nor van der Waals radius, but they have partial charges instead, while the central metal ion is a van der Waals sphere with zero charge [117]. The dummy atoms are bonded to the central metal ion, and they are free to move around the metal center. As a result, the system is free to exchange ligand, simulating the charge distribution around the metal ion and, differently from the classic LJ 12-6 model, elicits stronger interactions with the coordinating atoms, while maintaining the correct geometry [118]. The main drawback of this

method is the need to know the coordination number *a priori*. So far, tetrahedral [119], octahedral [120] and pentagonal bipyramidal [121] geometries have been used. However, to the best of our knowledge, no CDA model have been publish for the trigonal bipyramidal geometry, which occurs often in metal coordination [122][123]. In **Chapter 5** a new parameterization of this geometry using CDA is illustrated.

2.2.2 MM-GBSA free energy calculation

Several methods have been proposed to compute the binding free energy of a complex, starting with thermodynamic integration and free energy perturbation, coming to the newest machine learning methods [124]. One of the most used and straightforward free energy method is the Molecular Mechanics-Poisson/Generalized Boltzmann Surface Area (MM-P/GBSA), an end-point method that samples the final states of a system (for instance a MD simulation trajectory) and therefore is much less expensive than pathways methods [125]. MM-P/GBSA computes the binding free energy of a ligand (L) to a protein (P) as shown in Equation 2.17, where G_{PL} is the protein-ligand complex (Figure 2.5).

$$\Delta G_{bind} = G_{PL} - G_P - G_L \tag{2.17}$$

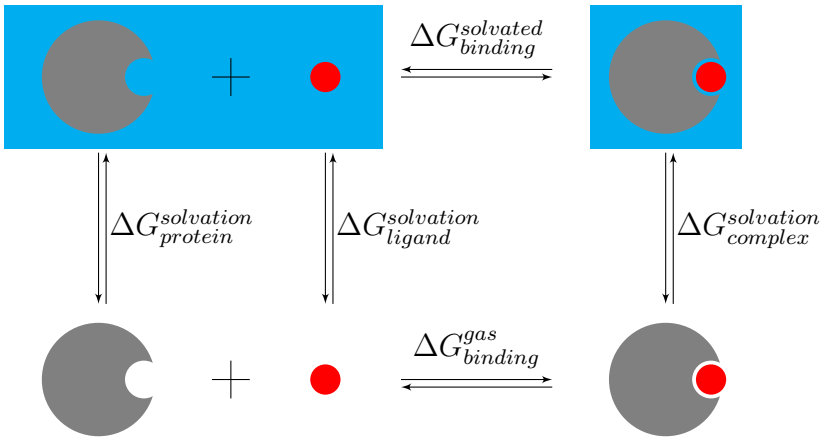


Figure 2.5: Thermodynamic cycle used for binding free energy calculations of a protein-ligand complex. Cyan boxes represent the solvated systems, while white background indicates the gas phase. $\Delta G^{solvated}_{binding}$ is the result of the MM-P/GBSA calculation.

Equation 2.17 can be decomposed into Equation 18, in which ΔE_{MM} is the difference in the gas phase molecular mechanics energy, ΔG_{sol} the solvation free

energy and $-T\Delta S$ the conformational entropy upon ligand binding [126].

$$\Delta G_{bind} = \Delta H - T\Delta S = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{2.18}$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdw} \tag{2.19}$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA} \tag{2.20}$$

Equation 2.18 can be further decomposed into Equations 2.19-2.20. ΔE_{MM} includes the bonded (ΔE_{int}) and non-bonded (ΔE_{ele}, ΔE_{vdw}) MM energy terms (Equation 2.19). ΔG_{sol} includes the polar (electrostatic solvation energy $\Delta G_{PB/GB}$) and apolar (ΔG_{SA}) contribution between the solute and the continuum solvent (Equation 2.20). ΔG_{SA} is calculated using the Solvent-Accessible Surface Area (SASA), while $\Delta G_{PB/GB}$ is computed using either the PB or GB model. The former uses the Poisson-Boltzmann equation to compute the electrostatic contribution to the free-energy, while the latter uses the Generalized Born approximation, which is faster compared to the first. Although the solvation energy calculated by PB is supposed to be more accurate than that solved by GB, different studies reported just small differences between the two methods [127][128][129]. For this reason, in this Thesis the MM-GBSA method was applied.

2.3 Hybrid QM/MM method

Molecular mechanics (MM) force fields relay on empirical parameters that are unable to describe changes in the electronic structure. These changes can involve bond breaking and forming, but also charge transfer, polarization and oxidation state changes, which are key features in the interaction with metal ions [47]. The analysis of these phenomena requires Quantum Mechanics (QM) methods. However, these methods are highly expensive and they can handle just few hundreds of atoms [130]. Therefore, a hybrid QM/MM approach was proposed to overcome the limitations of both the methods (see **Chapter 1.2**).
In this Thesis DFT and semi-empirical QM methods were used, therefore below is reported a very brief introduction on these theories.

2.3.1 Brief introduction on QM methods

Among all QM methods, Density Functional Theory (DFT) is routinely used thanks to favourable scaling with the number of atoms and its accuracy, in particular regarding metal-containing complexes [46][47]. The electron density distribution can be used to compute the ground-state energy of a molecule, which is much simpler (three variables) compared to *ab initio* methods (3N variables, where N is the number of electrons). DFT theory states that the energy of a system (in which the nuclei are fixed and only electrons move) can be expressed as a functional of the electronic density function $E[\rho]$ (Equation 2.21) [131]. However, the exact form of this functional is unknown.

$$E[\rho] = \int v(r)\rho(r)dr + J(\rho) + Ts(\rho) + E_{XC}(\rho) \tag{2.21}$$

The first term represents the electrostatic interaction between nuclei and electrons, the second the repulsion between electron density and itself and the third accounts for electronic kinetic energy. The last term $E_{XC}(\rho)$, the exchange–correlation functional, contains corrections of the previous terms, and it is the main responsible for the level of accuracy, as it is the source of errors in a DFT calculation. Furthermore, the exchange–correlation functional neglects the dispersion interactions, leading to the need of inclusion of dispersion effects by empirical corrections like the Grimme's [132]. Starting from Equation 2.21, a set of functionals can be derived to describe the orbitals by the Kohn–Sham equations [133]. The exchange–correlation functionals can be the Local Density Approximation (LDA), which is an earlier version of DFT and exhibited large errors in the calculation of bond energies [134]; the Generalised Gradient Approximation (GGA) and the most popular hybrid density functionals (e.g. B3LYP, PBE0) [135].

Semi-empirical methods derive from either Hartree−Fock (HF) or DFT methods, by the inclusion of approximation with external parameters to simplify the calculation. Semi-empirical methods based on the HF theory are NDDO, AM1, the PM series and MNDO; while based on DFT is the Density Functional Tight Binding (DFTB) [42]. Semi-empirical methods are cheaper in terms of computational cost if compared to *ab initio* or DFT thanks to the use of minimal basis sets and approximations to electron integrals. Recently Kříř *et al.* published a benchmark of different semi-empirical methods, where they showed that PM6 and DFTB obtained

the best results [136].

2.3.2 Setting up a QM/MM system

The first thing to consider setting up a QM/MM system, is to decide which part of the system will be treated by MM and QM. The QM calculation is the most demanding, thus it should contain the minimum number of atoms to study the event of interest. When working with proteins, it is often necessary to cut covalent bonds in order to divide the two subsystems (e.g. protein side chains) [137]. In this case, two main approaches are available: the frozen bond orbitals [138] and the link atoms [139]. In the former, a frozen bond orbital is added to satisfy the valence shell of the QM atom at the QM/MM interface, that could be placed using either the Local Self-Consistent Field (LSCF) or the Generalized Hybrid Orbital (GHO) method. In LSCF localized orbitals are place on the QM atom at the interface, and the one pointing at the MM atom is double occupied and blocked, while the orbitals pointing towards QM atoms are optimized and single occupied (Figure 2.6A). In GHO the orbitals are placed on the MM atom at the interface, and the one pointing at the QM atom is single occupied and frozen, while the other pointing to MM atoms are double occupied and optimized (Figure 2.6B).

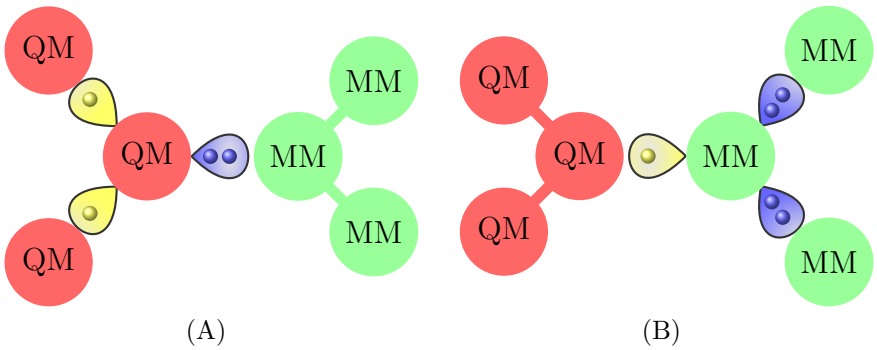


Figure 2.6: Graphical representation of the frozen bond orbital partitioning method. (A) The LSCF and (B) GHO methods are shown. QM atoms are coloured in red, MM atoms in green. The single occupied and optimized orbitals are coloured in yellow, while the double occupied and frozen orbitals are blue.

The second approach consists in the addition of a link atom, often a hydrogen atom, at an appropriate distance along the QM/MM bond vector, and it will be included only in the QM calculation [140]. In both the methods, polar bond break (as the peptide bond) should be avoid to prevent convergence issues.

The evaluation of energies within the two regions is straightforward: QM methods (i.e. HF, post-HF, DFT, semi-empirical) in the QM region and force field in the MM region. Differently, the interactions between the two portions need to be defined by a coupling scheme: either subtractive or additive [137][141]. In the subtractive one, the energy of the whole system (E_{sub}^{total}) is obtained following Equation 2.22, where E_{total}^{MM} is the MM energy computed for the entire system, $E_{QMregion}^{QM}$ is the QM energy of the QM region and $E_{QMregion}^{MM}$ is the MM energy of the QM region (Figure 2.7) [142].

$$E_{sub}^{total} = E_{total}^{MM} + E_{QMregion}^{QM} - E_{QMregion}^{MM} \tag{2.22}$$

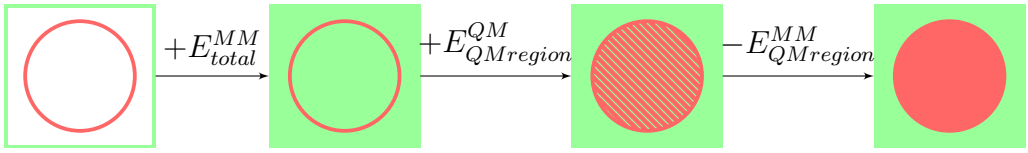


Figure 2.7: Schematic representation of the subtractive coupling scheme. First the energy of the whole system is evaluated by MM (E_{total}^{MM}, in green), then the QM energy of the QM subsystem is added ($E_{QMregion}^{QM}$, in red). Since the QM energy is computed two times, finally the MM energy of the QM subsystem is subtracted ($E_{QMregion}^{MM}$).

In the additive approach (Equation 2.23), the potential energy for the whole system is a sum of QM energy of the QM region ($E_{QMregion}^{QM}$), MM energy of the MM region ($E_{MMregion}^{MM}$) and a QM/MM coupling term ($E^{QM/MM}$) (Figure 2.8) [142].

$$E_{add}^{total} = E_{QMregion}^{QM} + E_{MMregion}^{MM} + E^{QM/MM} \tag{2.23}$$

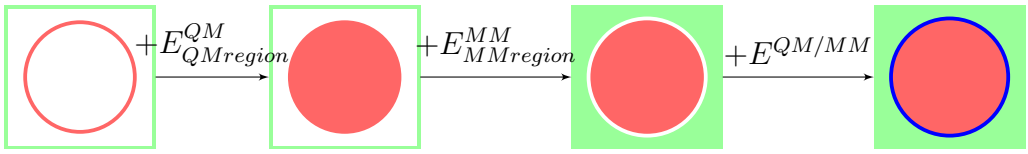


Figure 2.8: Schematic representation of the additive coupling scheme. First the energy of QM region is evaluated by QM ($E_{QMregion}^{QM}$, in red), then the MM energy of the MM subsystem is added ($E_{MMregion}^{MM}$, in green). Finally the QM/MM interface energy is also added ($E^{QM/MM}$, in blue).

In this last case, the interactions between the two subsystems are treated explicitly and they can be described with different levels of accuracy. In the mechanical embedding scheme these interactions are treated at the force field level, and it is the

most simplistic embedding scheme. By using the electrostatic embedding scheme, the electrostatic interactions between the QM and the MM regions are handled in the calculation of the electronic wave function, and the MM atoms will enter the QM Hamiltonian as one-electron operators. As a result, the atoms in the QM region can be polarized by the MM atoms [143]. Although this is a much more accurate description of the system, compared the mechanical embedding, over-polarization issues can occur, and so different charge smearing methods have been proposed [144]. In the most recent polarization embedding scheme, both the regions can polarized each other thanks to polarizable force fields. However these scheme is not fully developed [145].

Chapter 3

Identification of novel HDAC inhibitors by fragment-based drug design

3.1 HDACs as pharmaceutical targets

Transcription in eukaryotes is regulated by several mechanism on either DNA (i.e. DNA methylation [146]) or histones. Histones are strongly basic proteins able to package the DNA, which brings a strong negative charge derived from the phosphate groups of the nucleotides. Epigenetic modifications on histones include methylation [147], phosphorylation, ubiquitination [148], acylation, hydroxylation, glycation, serotonylation, glycosylation, sumoylation ADP-ribosylation and acetylation [149]. This last mechanism consists in the addition of an acetyl group on lysine residues, located upon histone tails, by histone acetylases (HATs). Such a post-tradutional modification leads to a decrease of the positive charge on histones and therefore to a loose packaging of DNA, which results in the activation of transcription [150]. HATs enzymatic activity is combined with histone deacetylases (HDACs), which on the contrary deacetylate the histones leading to a suppression of transcription. The balance between these two modifications is essential for a correct genetic transcription [151]. Changes in HAT/HDAC activity balance can lead

to altered gene expression and signaling pathways [152] that often bring to cancer [15].

HDACs can be divided into two families (zinc-dependent and NAD-dependent) and four classes. The NAD-dependent family correspond to the class III, which includes sirtuins 1-7 [153]. While the zinc-dependent family can be grouped into three classes based on sequence identity: class I which comprises the isoforms 1, 2, 3 and 8. Class II further divides into IIA (HDAC4, 5, 7 and 9) and IIB (HDAC6 and 10), finally class IV includes only HDAC11 [154].
The zinc-dependent HDACs share the same catalytic mechanism, reported in Figure 3.1, as their binding pocket is highly conserved [155].

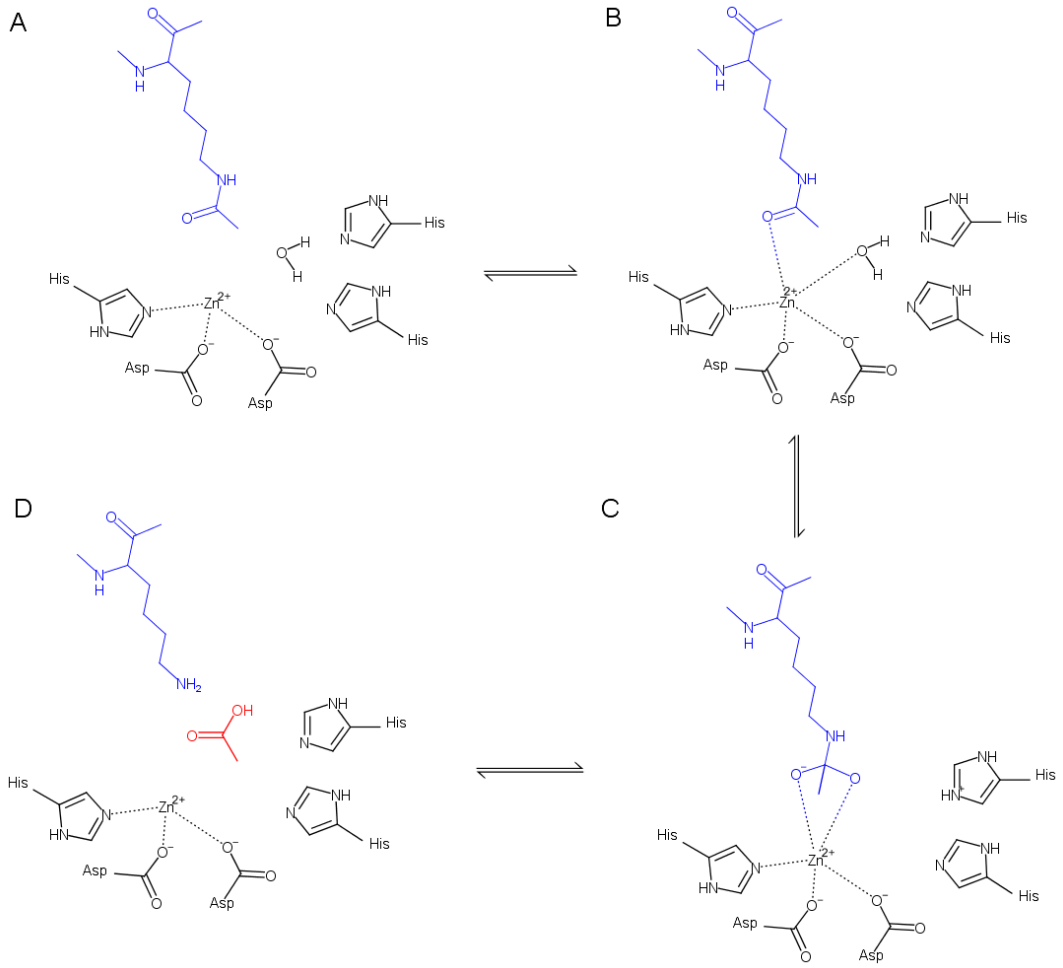


Figure 3.1: HDAC catalytic mechanism. (A) The zinc ion is coordinated by two aspartate and one histidine residues. When the acetylated lysine (in blue) enters HDAC binding site, (B) the carbonyl oxygen coordinates the zinc ion and a histidine facilitates the nucleophilic attack of a water molecule, coordinated with the zinc ion as well, to the carbonyl carbon of the lysine. (C) A tetrahedral intermediate is produced followed by a proton transfer from the histidine to the lysine, (D) which releases an acetate group (in red) and the lysine [155]. Coordination bonds are shown as dotted lines.

HDAC biding pocket is tubular and divided into three portions: at the bottom of the pocket is located the catalytic site, which includes the zinc ions and its coordinated residues, two histidine residues that participate to the catalytic reaction and usually a tyrosine that stabilizes the tetrahedral intermediate. Two aromatic phenylalanine residues are placed at the middle of the pocket, whereas the apical and external portion is surrounded by loops that change across isoform and confer specificity [156].

As mentioned before, an alteration of HAT/HDAC balance can bring to cancer. Over expression of HDACs is found indeed in prostate, gastric, lung, colon and breast tumors [15]. Therefore a promising pharmaceutical approach is targeting HDACs using small molecules as inhibitors, which bind the enzyme preventing the entrance of the substrate. So far, different HDAC inhibitors have been approved for clinical use by FDA and others are in clinical trials [157]. Generally, all HDAC inhibitors are composed by three moieties that interact with the three portions of the binding site described above: a zinc binding group (ZBG), a linker and a CAP (Figure 3.2).

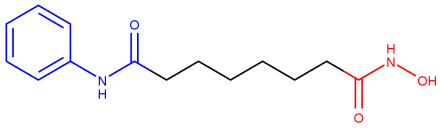


Figure 3.2: Structure of SAHA as an example of HDAC inhibitor. The CAP is coloured in blue and interacts with the apical portion of the binding pocket. The linker is coloured in black and interacts with the middle region of the pocket. The ZBG (hydroxamic acid) is coloured in red and interacts with the zinc ion at the inner region.

The CAP moiety interacts with the surface of the binding pocket, which is variable through the different HDAC isoforms and, for this reason, it confers selectivity. Nonetheless, several non-specific inhibitors (i.e. pan inhibitors) able to bind different isoforms are available. The linker portion is usually composed by an alkyl chain or a phenyl ring, that interacts with the aromatic residues at the middle of the tubular binding site. The ZBG region contains groups able to chelate the zinc ion. The most popular ZBG is the hydroxamic acid, but thiols, benzamides and trifluoromethyl ketones have been used as well [158].
Hydroxamic acid is often chosen for its strong chelating ability. However, such an ability can cause the inhibitors to bind other metal ions (e.g. iron and copper) that are essential for biological processes [159]. Furthermore, hydroxamic acid-based compounds are unstable *in vivo* as they showed low bioavailability and a short half-life ($\sim$ 2 hours) [160]. Finally, compounds containing hydroxamic acid have been

shown to be mutagenic by sulphation or acylation of their ester oxygen, that can undergo to the Lossen rearrangement (Figure 3.3) resulting in isocyanates, which serve as electrophiles in DNA carbamoylation [161][162].

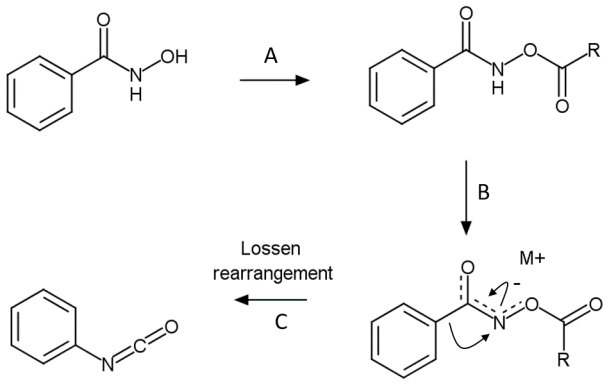


Figure 3.3: Schematic representation of the Lossen rearrangement. (A) The hydroxamic acid group undergo to either acylation or sulphation in its ester oxygen atom (label R). This is followed by (B) deprotonation of the amide nitrogen of the activated hydroxamates that eventually (C) rearrange to isocyanate by Lossen rearrangement [162].

These findings rise the need to design new HDAC inhibitors without hydroxamic acid as ZBG. In order to explore different ZBG, in this chapter a fragment-based approach is proposed. Fragment-based drug design (FBDD) has become a widely used approach thanks to the low degree of complexity of the fragments and to the high amount of molecules that derives from their combination [163]. Moreover, FBDD allows the discovery of low molecular weight ligands with an affinity range from micromolar to millimolar [164].
Here fragments corresponding to the three HDAC inhibitor portions (i.e. CAP, linker and ZBG) are combined together and tested through docking on different HDAC isoforms, with the aim to design a novel HDAC pan-inhibitor. Molecular descriptors are used, as well as descriptors derived from docking and force fields, and these descriptors were applied to a QSAR analysis, for the prediction of the binding affinity of the new molecules.

3.2 Methods

3.2.1 Protein preparation

Since the aim was to identify new pan-inhibitors, different HDAC isoforms were used in the docking study. In detail, resolved HDAC structures were chosen according to resolution and to the type of co-crystallized ligand (Table 3.1). All class I HDACs are resolved, as well as some of class II (i.e. HDAC4, 6 and 7).

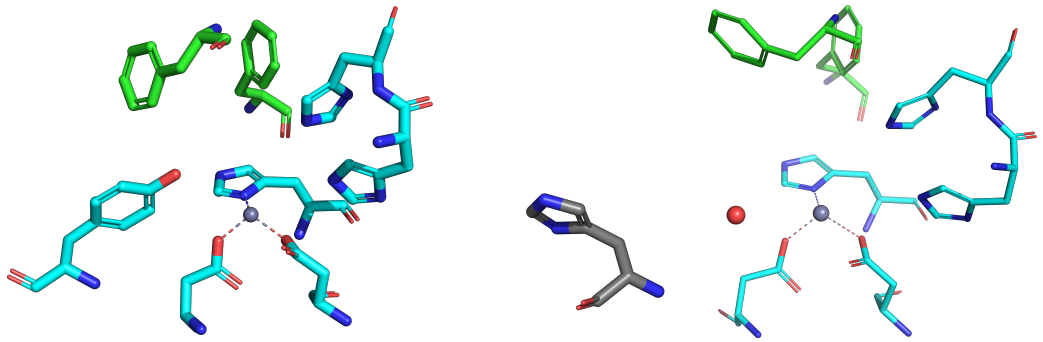
PDB Id	HDAC isoform	Resolution (Å)	Co-crystallized ligand ZBG
5ICN [165]	1	3.30	hydroxamic acid
4LXZ [166]	2	1.85	hydroxamic acid
4A69 [167]	3	2.06	–
2VQJ [168]	4	2.10	trifluoromethyl ketone
5EDU [169]	6	2.79	hydroxamic acid
3C10 [170]	7	2.00	hydroxamic acid
5THS [171]	8	1.90	hydroxamic acid

Table 3.1: HDAC crystal structures used in the docking studies. One resolved structure for each class was chosen based on the resolution and the ligand type. In the table the PDB Id, HDAC isoform and resolution (Å) are reported, as well as the ZBG contained in the co-crystallized ligand. HDAC3 crystal structure is available only in the apo form.

Hydrogen atoms were added to all the crystal structures according to physiological pH and they underwent to an energy minimization using NAMD2 [172], CHARMM22 as force field and constraints on the backbone atoms. In order to prevent the binding site collapse, the pan-inhibitor SAHA was inserted into the HDAC3 structure (crystallized in the apo form) through docking (Figure S3.1).

As mentioned in the previous section, the inner and middle portions of the binding pocket (that are the ZBG and liker binding regions) are highly conserved through HDAC isoforms (Figure S3.2). In particular, superimposition of class I and IIB proteins reveals an almost identical disposition in space of these portions, which comprise in the inner region: two aspartate and one histidine residues that coordinate the zinc ion, two histidine residues that participate in the catalytic reaction and a tyrosine that stabilise the tetrahedral intermediate. While the middle region comprises two phenylalanine orientated in a π-π stacking interaction (Figure 3.4A). The average RMSD of these residues between the five crystallized isoform is 0.98 Å. Little differences are present in class IIA, where the tyrosine is replaced by a histidine residue. However, a water molecule was frequently found near the histidine (PDB

Ids: 2VQM, 2VQO, 2VQQ [168], 4CBY [173], 3C0Z [170]). Such a disposition seems to mimic the hydroxide group of the tyrosine, as reported in Figure 3.4B. Therefore, this water molecule was retained during docking, as considered important for the ligand interaction.



(A) (B)

Figure 3.4: Comparison between inner and middle region of (A) class I-IIB and (B) class IIA. The residues of the inner region are reported in cyan and the the residues of the middle region in green. In panel (B), the water molecule replacing the hydroxyl group of tyrosine is represented as a red sphere, while the histidine taking the tyrosine place is coloured in gray. The zinc is shown as a silver sphere.

Docking studies of the ZBG and linker libraries were performed in these highly conserved regions (see below). Given the high similarity between the isoform, two hybrid pockets (one for class I and IIB, and one for class IIA) were created to save computational time. In detail, the average position of the 3D coordinates of the residues mentioned above was computed and used to generate the hybrid pockets.

3.2.2 Fragments molecular docking

Docking studies were performed using the software PLANTS [67] with speed1 accuracy, ChemPLP scoring function and generating 10 poses.
Three docking runs were performed using first the ZBG fragment library and considering the inner region as a binding pocket. Second, the ZBG fragments resulted from docking that coordinated the zinc ion were combined with the linker library

and we considered both the inner and middle region as a binding pocket. Finally, the resulting fragments that coordinated the zinc ion were combined with the CAP library and docked on the entire binding site (i.e. considering a radius of 10 Å around the zinc ion), to include the external portion of the pocket, where is located the CAP interaction site. In all of the steps, fragments that did not interact with the zinc ion were excluded.
The binding pockets were chosen with small size to avoid misplacement of the fragments, that can easily interact with different portions of the protein.

3.2.3 QSAR model creation

In order to predict the activity of the new inhibitors obtained by molecular docking, a QSAR model was developed. A training set that comprised 40 molecules with known activity (i.e. IC_{50}), ranging from pM to μM, was collected from literature (Figure S3.3). For time reasons, only HDAC1 inhibitors were considered.
Physicochemical descriptors were computed by both VEGA ZZ [174] (molecular mechanics descriptors) and Mopac2016 [175] using the PM7 Hamiltonian (semi-empirical descriptors). Furthermore, protein-ligand interaction descriptors were computed: the training set was docked into HDAC1, following the procedure described above for the whole ligands. Then, Rescore+ script, implemented in VEGA ZZ, was used to compute different interaction scores, together with the pair interaction calculation of NAMD2. The latter computes ligand-protein interaction energy using force field.

These descriptors were combined through linear regression obtaining Equation 3.1, where *Lipole* and *Dipole* are physicochemical descriptors of ligands, while E_{ff} is the force field interaction energy. The r^2 value was 0.74 and the q^2 was 0.54.

$$pIC_{50} = 6.964 - 0.469Lipole - 0.017E_{ff} - 0.047Dipole \tag{3.1}$$

However, removing one molecule (i.e. valproic acid) from the training set, the r^2 and q^2 values increased up to 0.78 and 0.72 respectively (Figure 3.5), with $\pm$ 0.34 as mean error on the prediction, and a new equation was computed (Equation 3.2). The valproic acid was considered an outlier, as its structure differs from the others of the training set (Figure S3.3), and generally from the ZBG-linker-CAP structure.

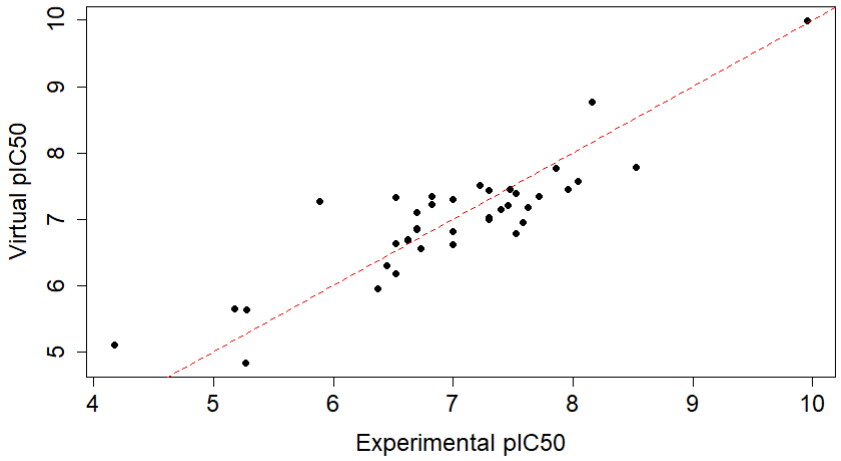


Figure 3.5: Linear regression used for the creation of the QSAR model. The regression line is coloured in red, while the experimental and calculated pIC_50 are reported on the x- and y-axis respectively. The black dots indicate the molecules of the training set.

$$pIC_{50} = 7.106 - 0.469Lipole - 0.019Elect_{ff} + 0.115VirtualLogP \tag{3.2}$$

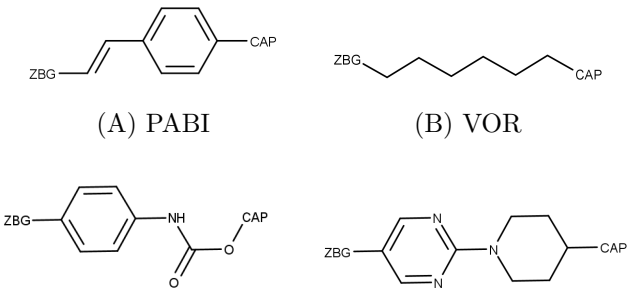
3.3 Results

3.3.1 Generation of the fragment libraries

Three fragment libraries were collected, corresponding to the three portion of HDAC inhibitors (i.e. CAP - linker - ZBG).

The ZBG library included 50 general metal chelator fragments retrieved from literature (Figure S3.4) [155][176][177][178][179]. Hydroxamic acid was inserted into the ZBG library as positive control.

The linker library included four fragments derived from known HDAC pan-inhibitors already approved for clinical use (Figure 3.6). The linker length was 6 or 7 carbon atoms, which is reported to be the best length for HDAC ligands [180].

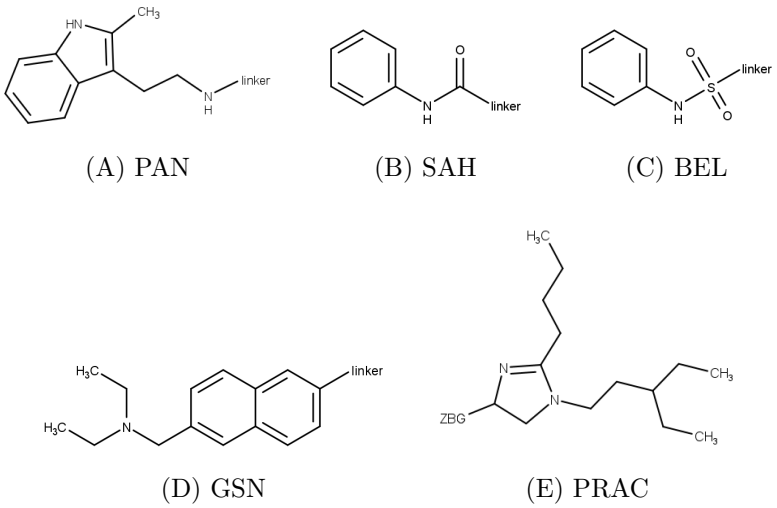


(A) PABI (B) VOR

(C) GIV (D) QSN

Figure 3.6: Linker library. The four fragments derive from approved HDAC pan-inhibitors: (A) Panobinostat, (B) Vorinostat, (C) Givinostat and (D) Quisinostat. An identifier is assigned to each fragment. ZBG and CAP binding positions are reported.

The CAP library included five fragments derived from the CAP moiety of Belinostat, Panobinostat, Vorinostat, Givinostat and Pracinostat (Figure 3.7).



(A) PAN (B) SAH (C) BEL

(D) GSN (E) PRAC

Figure 3.7: CAP library. The five fragments derived from approved HDAC pan-inhibitors: (A) Belinostat, (B) Panobinostat, (C) Vorinostat, (D) Givinostat and (E) Pracinostat. An identifier is assigned to each fragment. Linker binding position is reported.

3.3.2 Fragment docking and combination

As mentioned in the **Methods** section, docking of the ZBG library was performed using the hybrid pockets, considering just the residues of the inner portion as a binding pocket. Fragments that did not interact with the zinc ion were excluded, while the remaining fragments were selected and combined with the linker library. Table 3.2 reports the selected fragments, ranked by docking score (ChemPLP normalized by the number of heavy atoms) and divided according to the hybrid pocket

used in the docking calculation. It is worth noticing that the majority of the selected fragments are shared by the two hybrid pockets and that the hydroxamic acid fragment (HA), inserted as positive control, was ranked in the second and first positions, suggesting that the scoring function is able to identify strong-binders.

Class I & IIB		Class IIA	
Fragment identifier	ChemPLP$_{HA}$	Fragment identifier	ChemPLP$_{HA}$
PA	-31.12	HA	-30.09
HA	-30.80	PA	-29.64
BA	-30.34	BOA	-28.31
PICA	-30.14	HZ	-28.23
QUI	-29.65	QUI	-27.70
bLATT	-29.53	PICA	-27.29
HZ	-29.52	PHA	-26.87
PHA	-29.28	SA	-26.80
DIPHE	-28.62	bLATT	-26.28
DKA	-28.40	HU	-26.27
BOA	-28.27	SHYPYD	-25.53
SHYPYD	-27.69	DIPHE	-25.44
TMK	-22.30	TMK	-25.53
		BA	-19.47

Table 3.2: ZBG fragments derived from docking and selected for the next step of the fragment-based approach. Fragments are ranked according to the docking score ChemPLP normalized by the number of heavy atoms (ChemPLP$_{HA}$). Docking calculations were performed on the two hybrid pockets. The 2D structures of the fragments linked with their identifier can be found in Figure S3.4. The HA fragment refers to the hydroxamic acid included as positive control.

These fragments were combined with the linker library. For some fragments, more than one binding point for the linker was considered. In Figure 3.8 the linker binding points are indicated, with the new nomenclature.

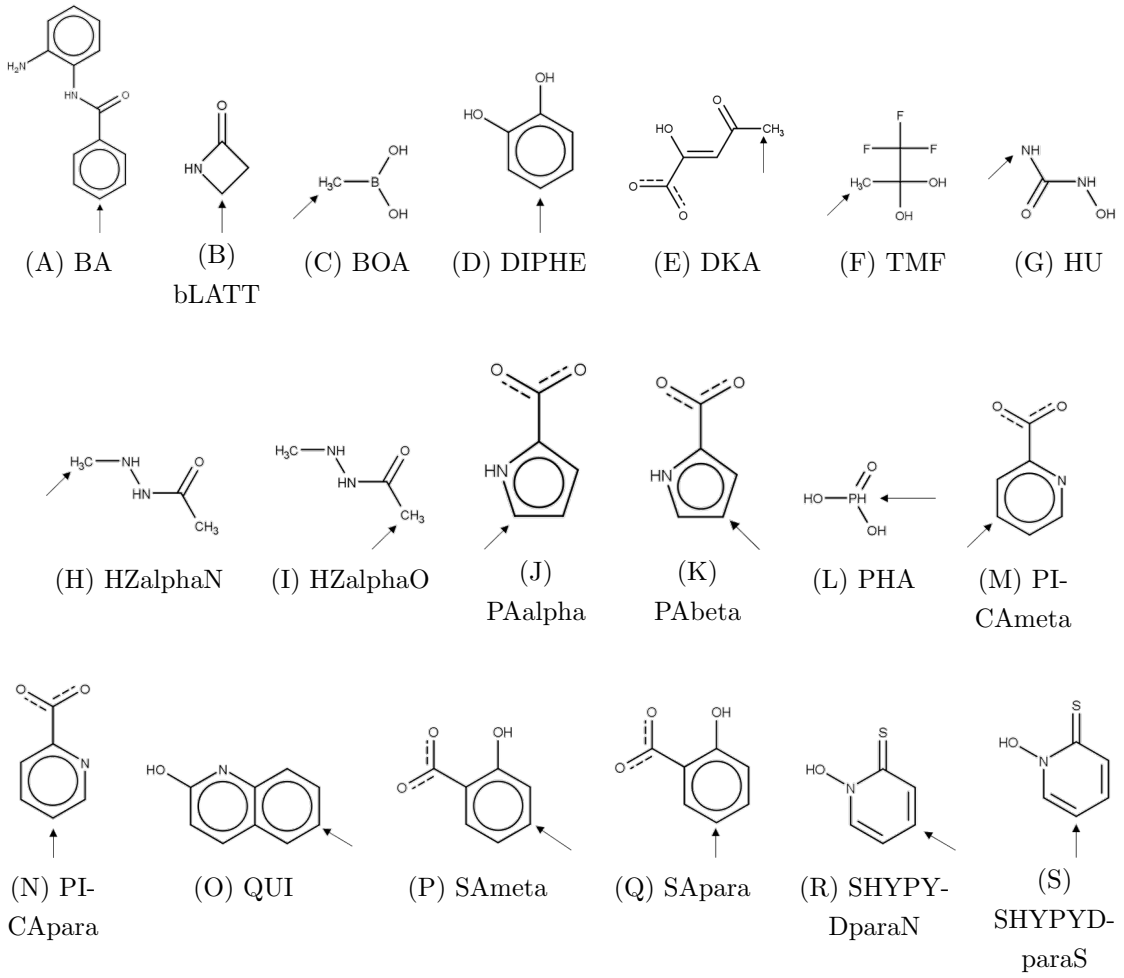


Figure 3.8: ZBG fragments selected for the combination with the linker library. The arrows indicate the points at which the linkers are attached. For some fragments, more than one binding point is reported. In these cases a new nomenclature is introduced.

The ZBG fragments BA and QUI were excluded from the combination with the linker library for their steric hindrance, and they were directly combined with the CAP fragments. The ZBG-linker fragments that interacted with the zinc ion were selected (Table 3.3). As seen for the docking results of the ZBG, most of the fragments are shared between the two pockets and the fragment HA-VOR, which derives from the inhibitor SAHA, is still included in the selection.

Finally, these fragments were combined with the CAP library, obtaining the whole ligands. These ligands were docked into each HDAC isoform (Table 3.1), considering as a binding pocket for the docking, a radius of 10 Å from the zinc ion. The resulting docking poses underwent to an energy minimization using NAMD2.

Class I & IIB		Class IIA	
Fragment identifier	ChemPLP$_{HA}$	Fragment identifier	ChemPLP$_{HA}$
TMK-PABI	-33.45	BOA-PABI	-30.04
BOA-PABI	-32.99	PHA-PABI	-28.69
BOA-PABI	-30.45	TMK-PABI	-27.87
PHA-PABI	-30.37	BOA-VOR	-26.98
HZalphaO-PABI	-29.66	HU-PABI	-25.63
TMK-VOR	-28.64	HA-VOR	-25.49
HU-PABI	-28.42	HA-PABI	-24.83
bLATT-PABI	-27.68	BOA-GIV	-24.43
PHA-VOR	-27.41	PHA-VOR	-24.13
DKA-VOR	-27.33	bLATT-VOR	-23.90
bLATT-VOR	-26.55	DKA-VOR	-23.51
HA-VOR	-26.15	HZalphaO-PABI	-23.44
PAbeta-VOR	-25.85	HU-VOR	-22.89
PICAmeta-VOR	-23.81	PICAmeta-VOR	-23.36

Table 3.3: ZBG-linker fragments derived from docking and selected for the next step of the fragment-based approach. Fragments are ranked according to the docking score ChemPLP normalized on heavy atoms (ChemPLP$_{HA}$). Docking calculations were performed on the two hybrid pockets. The 2D structures of the fragments linked with their identifier can be found in Figure S3.4 and 3.6. The HA fragment refers to the hydroxamic acid, included as positive control.

3.3.3 Activity prediction

The QSAR model reported in Equation 3.2 was applied for the activity prediction of the new ligands, obtained from docking. Table 3.4 indicates the prediction referred to the HDAC1 isoform.

Ligand identifier	Predicted IC_{50} (nM)
bLATT-PABI-GSN	4.56
bLATT-PABI-PAN	9.99
bLATT-VOR-PAN	229.06
DKA-VOR-BELI	215.50
DKA-VOR-GSN	3.31
DKA-VOR-PAN	0.31
DKA-VOR-SAH	1869.20
HU-PABI-BELI	227.59
HU-PABI-GSN	26.66
HU-PABI-PAN	1.11
HU-PABI-SAH	465.91
HU-VOR-BELI	678.11
HU-VOR-GSN	12.95
HU-VOR-PAN	1.70
HU-VOR-SAH	1070.92
HZalphaO-PABI-BELI	6.80
HZalphaO-PABI-PAN	1.20
HZ-PABI-GSN	42.57
HZ-PABI-SAH	194.30
PAbeta-VOR-BELI	0.45
PAbeta-VOR-PAN	0.08
PAbeta-VOR-SAH	0.75
PICAmeta-VOR-SAH	1.77
QUI-BELI	6.61

Table 3.4: Predicted IC_{50} (nM) for the whole ligands, using the QSAR model of Equation 3.2. The ligands were tested on the HDAC1 isoform. Ligands containing the hydroxamic acid moiety were excluded.

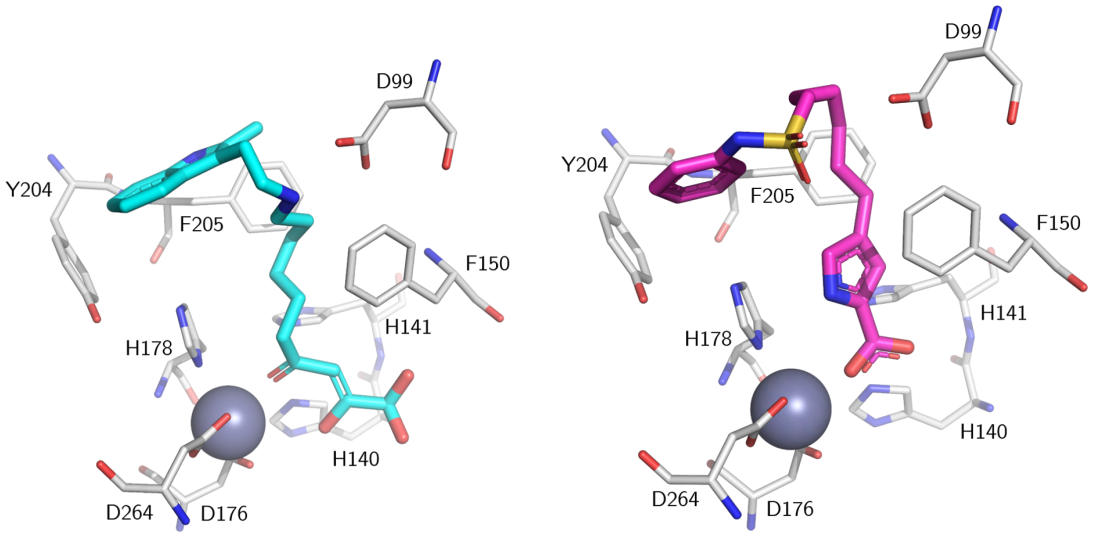
Tanimoto index was computed between the molecules that were predicted as the most active and Belinostat, Panobinostat, Givinostat, Pracinostat, Quisinostat and Vorinostat (Table 3.5). A threshold of 0.6 was chosen to exclude molecules too similar to the known HDAC inhibitors.

Ligand identifier	Bel	Pan	Giv	Pra	Qui	Vor
DKA-VOR-GSN	0.260	0.645	0.386	0.250	0.260	0.286
DKA-VOR-PAN	0.288	0.339	0.322	0.269	0.288	0.290
HU-PABI-GSN	0.344	0.448	0.398	0.340	0.344	0.329
HU-PABI-PAN	0.288	0.779	0.349	0.412	0.288	0.277
HU-VOR-GSN	0.258	0.398	0.430	0.272	0.258	0.406
HU-VOR-PAN	0.220	0.686	0.337	0.354	0.220	0.318
HZalphaO-PABI-PAN	0.310	0.843	0.324	0.450	0.310	0.263
HZ-PABI-GSN	0.262	0.385	0.333	0.301	0.263	0.208
PAbeta-VOR-PAN	0.282	0.301	0.282	0.244	0.244	0.245

Table 3.5: Tanimoto index between the best molecules, obtained from docking and QSAR analysis, and approved HDAC inhibitors: Belinostat, Panobinostat, Givinostat, Pracinostat, Quisinostat and Vorinostat. DKA-VOR-GSN, HU-PABI-PAN, HU-VOR-PAN and HZalphaO-PABI-PAN have one index higher than 0.6, therefore were excluded.

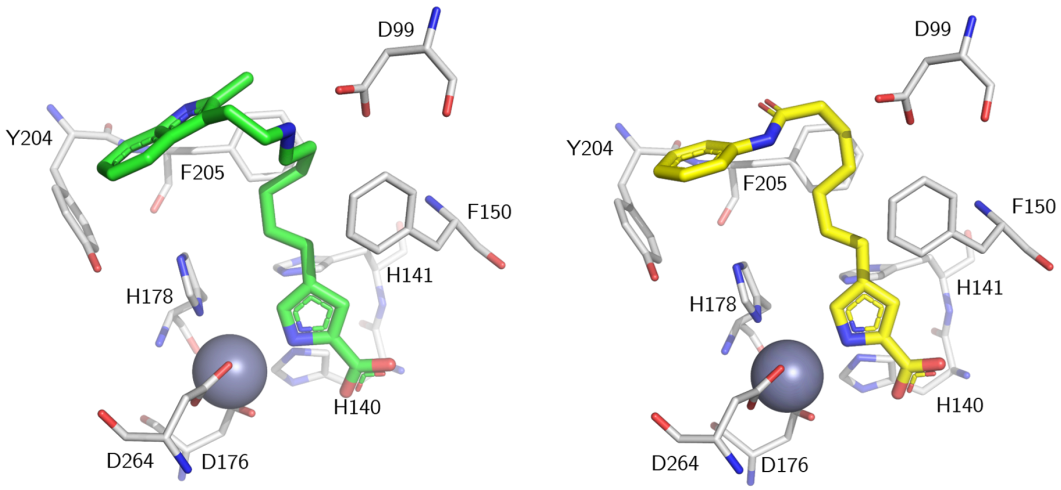
DKA-VOR-GSN, HU-PABI-PAN, HU-VOR-PAN and HZalphaO-PABI-PAN obtained a index higher than 0.6, therefore were excluded.

The four molecules predicted with a pM activity contained as ZBG the DKA and PAbeta fragments. Figure 3.9 shows their binding mode inside HDAC1.



(A) DKA-VOR-PAN

(B) PAbeta-VOR-BELI



(C) PAbeta-VOR-PAN

(D) PAbeta-VOR-SAH

Figure 3.9: HDAC1 in complex with the four molecules predicted with a pM activity. (A) DKA-VOR-PAN is coloured in cyan, (B) PAbeta-VOR-BELI in purple, (C) PAbeta-VOR-PAN in green and (D) PAbeta-VOR-SAH in yellow. The complexes were obtained through docking with PLANTS. The silver sphere is the zinc ion.

When the CAP fragment PAN (derived from Panobinostat) is included, the amine group of the ligands elicits an electrostatic interaction with Asp99 and the indole moiety is involved in hydrophobic interaction with Tyr204. This hydrophobic interaction is present also between the benzyl moiety of PAbeta-VOR-BELI and PAbeta-VOR-SAH (Figure 3.9B and D).

The six-carbon chain was chosen as linker in all of the four inhibitors, confirming its versatility [180].

The PAbeta was chosen as ZBG in three of the best ligands, and it coordinates the zinc ion with either the carboxyl group (Figure 3.9B) or the pyrrole moiety (Figure 3.9C-D). However this ZBG was expected to chelate the zinc ion with both the carboxyl and pyrrole group [176]. A similar result was obtained with the DKA ZBG, that coordinates the zinc ion, without chelation, which however was expected as well [178]. These results can be due to the scoring function used during docking (i.e. ChemPLP) that includes a specific term for the treatment of metal ion chelation. However, it was parametrized only for tetrahedral and octahedral geometries, therefore the five-coordinated zinc ion is not considered, and this could explain why the docking poses do not favour chelation over coordination.

3.4 Conclusions

HDAC inhibitors have risen a lot of interest as anticancer agents, where hydroxamic acid-based compounds represent the majority of the approved molecules. However, hydroxamic acid can bring severe side effects like mutagenesis [161][162]. Therefore, the aim of the work was to design new HDAC inhibitors without using the hydroxamic acid as ZBG. We applied a fragment-based approach to screen general metal chelator fragments, then these fragments were combined with linker and CAP fragments retrieved from approved HDAC inhibitors.
A QSAR model was created to quantify the binding affinity of these new ligands towards the HDAC1 isoform. Finally, different promising compounds were identified, with a predicted binding affinity in the range of nM/pM.

In this work we used a simplistic approach to treat the zinc ion. Such an approach comprises molecular docking, where the scoring function determinates how the ligand interacts with the metal ion, and interaction energies derived from the force field (i.e. CHARMM22), which uses the classic non-bonded model to describe metal interactions. Despite these simple descriptions of coordination bond, we were able to identify some promising compounds that will be tested in experimental assays.
Nonetheless this approach has some limitations: first in the various steps of the docking procedure, we excluded those fragments (or molecules) that did not coordinate the zinc ion. The fragments were frequently docked with the ZBG pointing towards the apical portion of the binding pocket, whereas the linker or the CAP pointed towards the zinc ion. This could results in the exclusion of potential good molecules

that were discarded due to the inability of the docking program to correctly predict their binding. Second, the force field used to compute the interaction energy tends to favour the electrostatic interactions above the metal coordination bond [112]. Third, the metal coordination term of the scoring function includes only the tetrahedral and octahedral coordination geometries, and not the five-coordinated which is frequently found in zinc-metalloproteins [85].

This approach allows a fast screening of new putative HDAC inhibitors and the identification of hit compounds. However, given these inaccuracy, further analyses should follow to better validate the results.

3.5 Supplementary Information

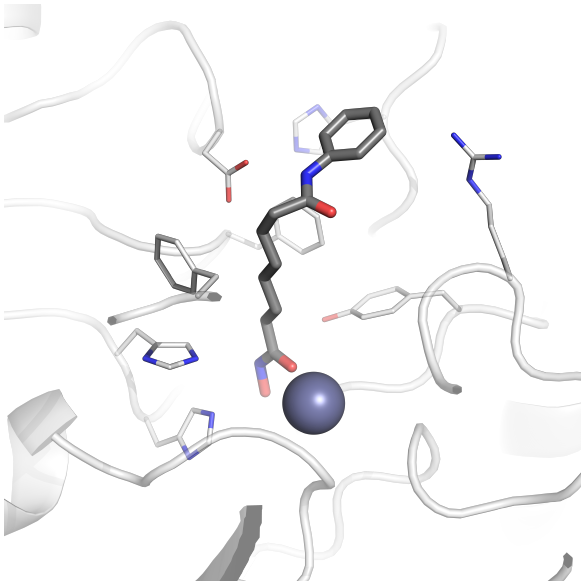
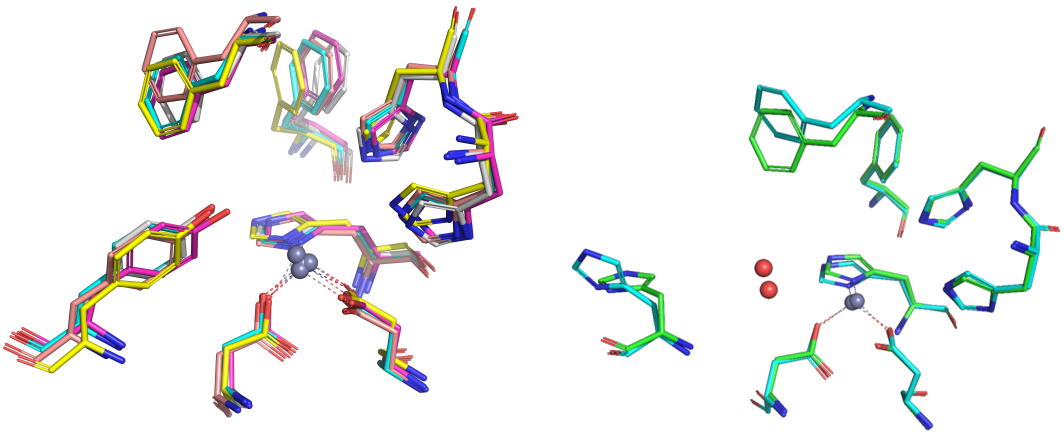
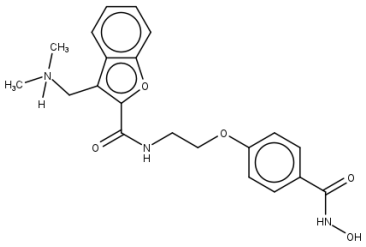


Figure S3.1: HDAC3-SAHA complex obtained through docking. HDAC3 crystal structure (PDB Id: 4A69 [167]) is resolved in the apo form. A pan inhibitor (i.e. SAHA) was inserted into the binding site to prevent its collapse during the minimization step.

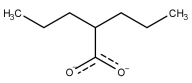


(A) (B)

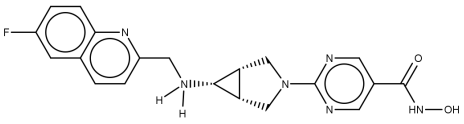
Figure S3.2: The ZBG and linker binding region are highly conserved in (A) class I and IIB and (B) class IIA. (A) Superimposition of class I and IIB crystal structures. The average RMSD of the represented residues is 0.98 Å. HDAC1 is coloured in light gray, HDAC2 in cyan, HDAC3 in purple, HDAC6 in yellow and HDAC8 in pink. (B) Superimposition of class IIA crystal structures. HDAC4 is coloured in green while HDAC7 in cyan. Zinc ions are represented as silver spheres and dotted lines represent the coordination bonds. PDB Ids for each isoform can be found in Table 3.1.



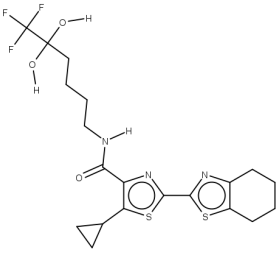
(A) Abexinostat, 7 nM [181]



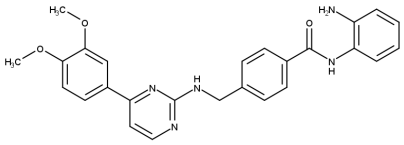
(B) Valproic acid, 400 nM [182]



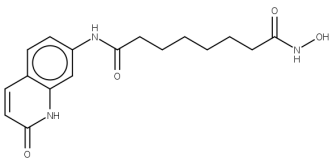
(C) CHR-3996, 3 nM [183]



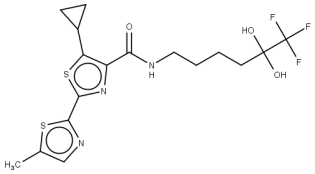
(D) 19.38 nM [184]



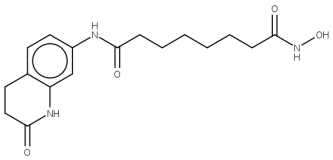
(E) 50 nM [185]



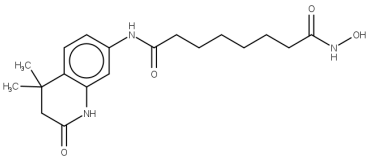
(F) 35 nM [186]



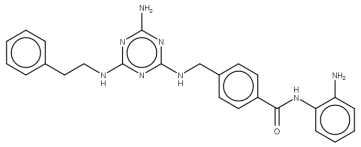
(G) 33.6 nM [184]



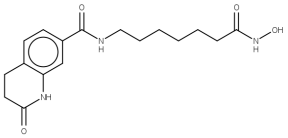
(H) 14 nM [186]



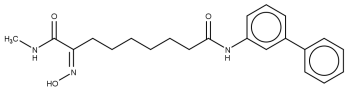
(I) 9.1 nM [186]



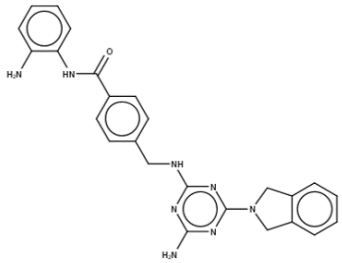
(J) 100 nM [185]



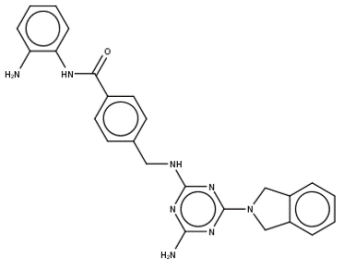
(K) 24 nM [186]



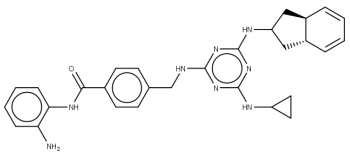
(L) 6710 nM [187]



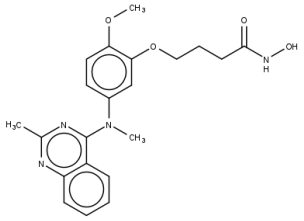
(M) 30 nM [187]



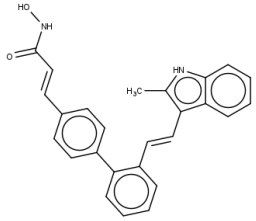
(N) 40 nM [185]



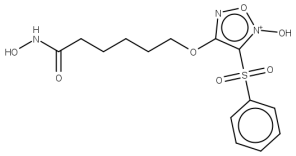
(O) 30 nM [185]



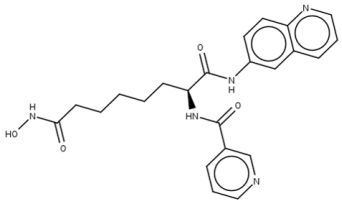
(P) 30 nM [188]



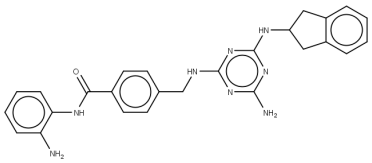
(Q) 5400 nM [189]



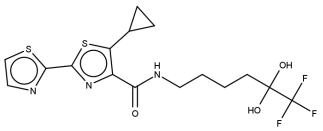
(R) 241 nM [190]



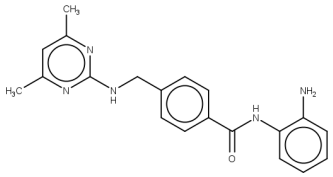
(S) 1300 nM [191]



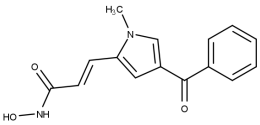
(T) 200 nM [185]



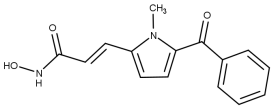
(U) 26.28 nM [184]



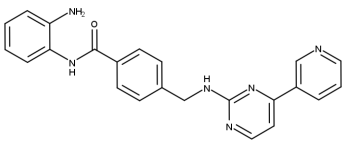
(V) 300 nM [185]



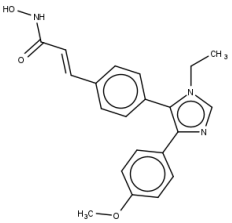
(W) 5310 nM [192]



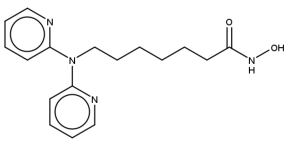
(X) 6640 nM [192]



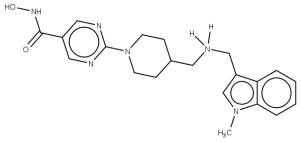
(Y) 150 nM [185]



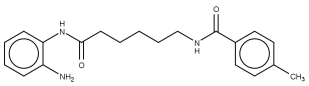
(Z) Etacrox, 187 nM [193]



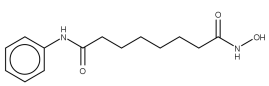
(AA) KA1010, 356 nM [194]



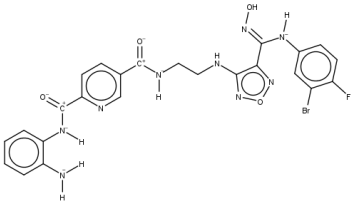
(AB) Quisinostat, 0.11 nM [195]



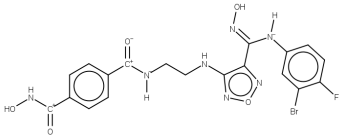
(AC) RGFP109, 60 nM [196]



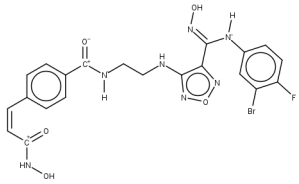
(AD) SAHA, 11 nM [197]



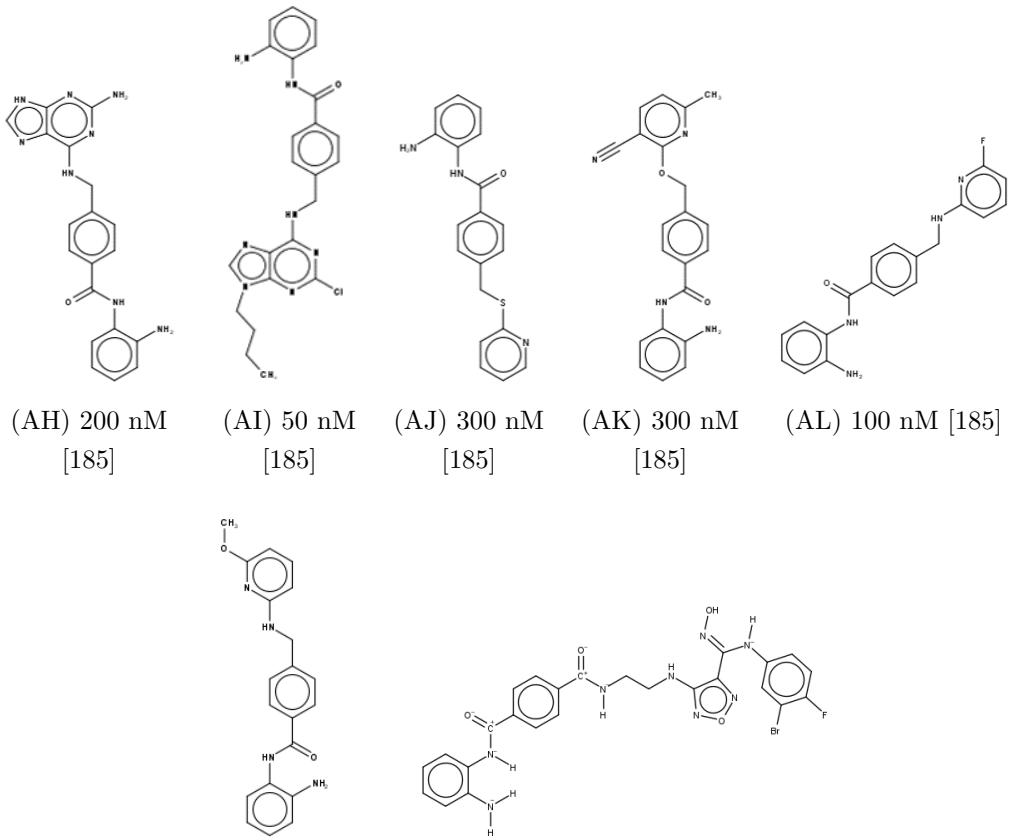
(AE) 604.4 nM [198]



(AF) 262.4 nM [198]



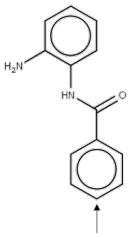
(AG) 46.2 nM [198]



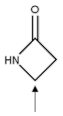
(AH) 200 nM [185] (AI) 50 nM [185] (AJ) 300 nM [185] (AK) 300 nM [185] (AL) 100 nM [185]

(AM) 100 nM [185] (AN) 66.5 nM [198]

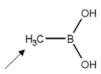
Figure S3.3: 2D structures of the 40 molecules of the training set used to create the QSAR model. For each molecule the activity (pIC_{50}) and reference are reported. Only proper names are reported.



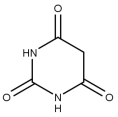
(A) BA



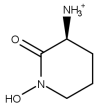
(B) bLATT



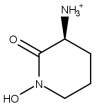
(C) BOA



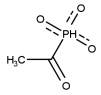
(D) BTA



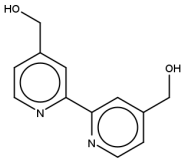
(E) CHA



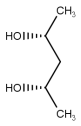
(F) CHA2



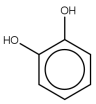
(G) CMP



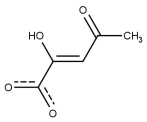
(H) DEPYR



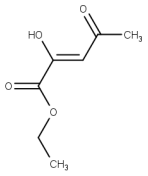
(I) DIOLO



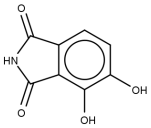
(J) DIPHE



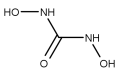
(K) DKA



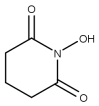
(L) DKAet



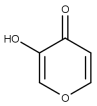
(M) DOIND



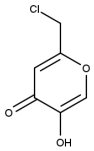
(N) HU



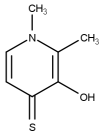
(O) HYISOQ



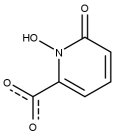
(P) HYPY



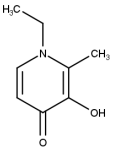
(Q) HYPYC1



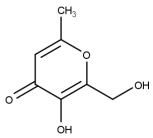
(R) HYPYD



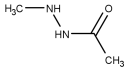
(S) HYPYDA



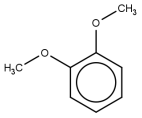
(T) HYPYDet



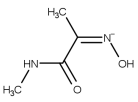
(U) HYPYmetoh



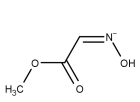
(V) HZ



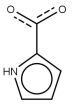
(W) NDGA



(X) OXAM



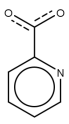
(Y) OXES



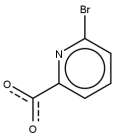
(Z) PA

(AA) PHA

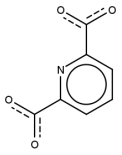
(AB) PHA2



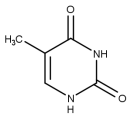
(AC) PICA



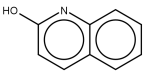
(AD) PICABr



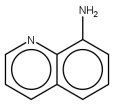
(AE) PICDA



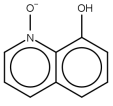
(AF) PYR



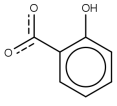
(AG) QUI



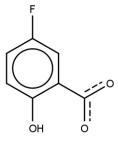
(AH) QUIAM



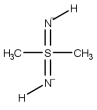
(AI) QUIOX



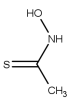
(AJ) SA



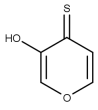
(AK) SAF



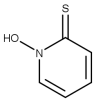
(AL) SDI



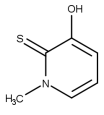
(AM) SHA



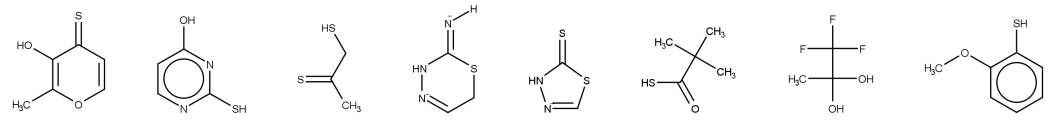
(AN) SHYPY



(AO) SHYPYD



(AP) SHYPYD2



(AQ)	(AR)	(AS)	(AT)	(AU)	(AV)	(AW)	(AX)
SHYPYmet	SPYR	SSH	TDN	TDZ	THK	TMK	TYetoh

Figure S3.4: ZBG library. The fragments are general metal chelators retrieved from literature [155][176][177][178][179]. An identifier was assigned for each fragment.

Chapter 4

Insight into the role of the zinc ion in *h*AChE reactivation: computational studies of new metallo-organic derivatives

4.1 AChE and its reactivators

Acetylcholinesterase (AChE) is a serine hydrolase that plays an essential role in terminating the nervous signal, by hydrolysing the neurotransmitter acetylcoline in the synaptic cleft [199]. The catalytic active site (CAS) contains the catalytic triad (Ser203, Glu334, His447 in the human isoform), which is located at the bottom of a 20 Å deep gorge [200]. Substrate trafficking is controlled by aromatic residues (Tyr72, Tyr124 and Trp286) at the peripheral anionic site (PAS) of the gorge [201]. Organophosphorus (OP) nerve agents irreversibly inhibit AChE through covalent binding to the catalytic serine. The recent attack against the Russian journalist Skripal in Salisbury [202], along with other attacks [203][204] shows the employment of these chemicals as warfare weapons, although the use of OPs is internationally banned [205]. However, given their widespread use in agriculture, people are more frequently poisoned by OP pesticides leading to an alarming increase of pesticides-related deaths [206]. OP severe effects are due to the covalent phosphoester conjugate that they form with the catalytic Ser203, resulting in the enzyme inactivation

and, as a consequence, in death through respiratory paralysis [207]. Generally the structure of OPs is a phosphate ester bound to two alkyl chains and a leaving group, that is removed after the conjugation (Figure 4.1A). Furthermore, over time, the inhibited enzyme can undergo to dealkylation of the conjugate and become completely resistant towards reactivation: this process is called aging [208][209].

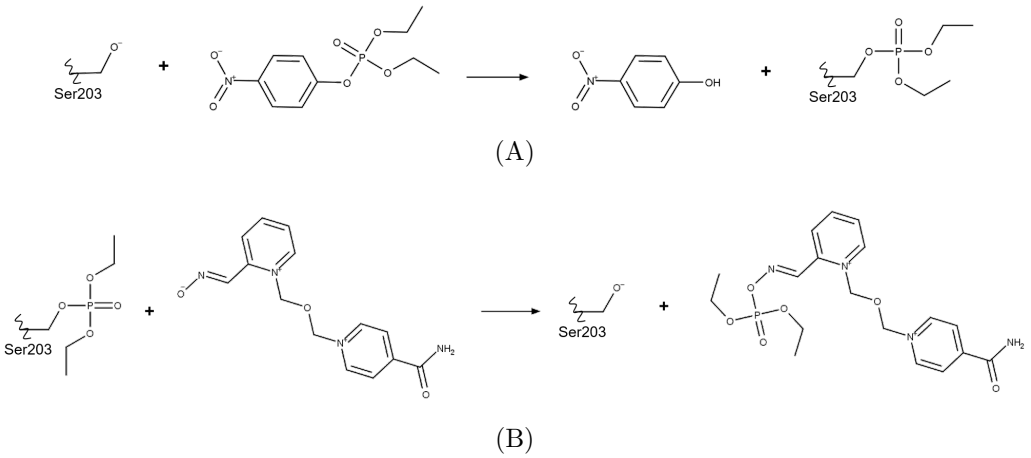


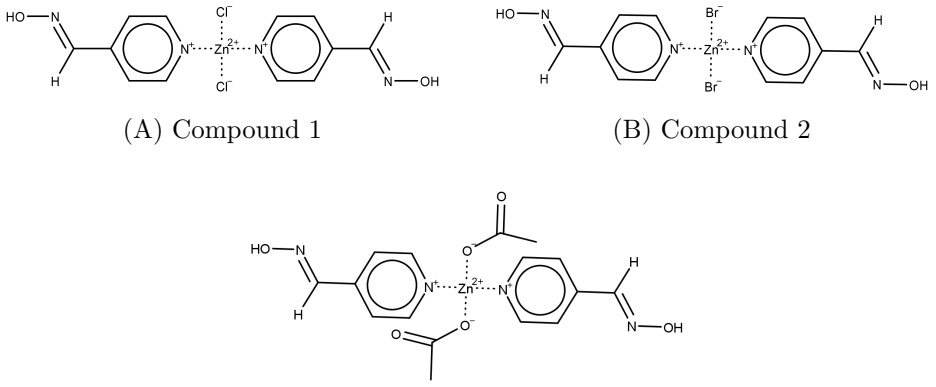
Figure 4.1: AChE inhibition and reactivation mechanisms. (A) The phosphorus atom of the OP binds the oxygen atom of the catalytic serine with loss of the leaving group and formation of the O-P bond between the enzyme and the inhibitor. (B) The oxime-based reactivator cleaves the covalent bond between OP and AChE through nucleophilic attack, releasing the catalytic serine. In the example paraoxon is the OP and HI-6 the reactivator.

The enzyme activity can be restored through administration of strong nucleophilic agents such as oximes, called reactivators, that break the strong phosphorus-oxygen bond between the OP and the Ser203 (Figure 4.1B) [210]. In the last decades, several AChE reactivators have been synthesized by many groups worldwide [211][212]. Generally, the quaternization of the pyridine nitrogen atom is essential for the reactivator efficacy, together with the position and amount of oxime groups [213]. However, none of the known oxime reactivator is effective against all OPs, due to the broad structural variability of these compounds [207]. Therefore, the development of new AChE reactivators able to remove any OP type remains an open challenge.

4.2 Aim

A promising approach is to increase the nucleophilicity of the oxime group through lowering the pK_a of the molecule. To this end, the inclusion of a metal

ion, in particular of the zinc ion, can increase the nucleophilicity of the oxime as well as the Lewis acidity of the zinc ion, and efficiently break the covalent bond between the enzyme and the inhibitor. Such an approach was proposed for the first time by Bolton and Beckett in 1964 [214] and later again by Breslow [215][216], where their findings suggested that the oximate ion and the zinc Lewis acid actions can be synergistic.



(A) Compound 1

(B) Compound 2

(C) Compound 3

Figure 4.2: First serie of mononuclear zinc-containing reactivators proposed by Konidaris *et al.* [217]. The zinc ion is in tetrahedral geometry, coordinated by two nitrogen atoms of the pyridyloxime ligands in all compounds, and (A) by two Cl^- in compound 1, (B) two Br^- in compound 2 and (C) two oxygen atoms of the acetate ions in compound 3. Coordination bond is represented with a dotted line.

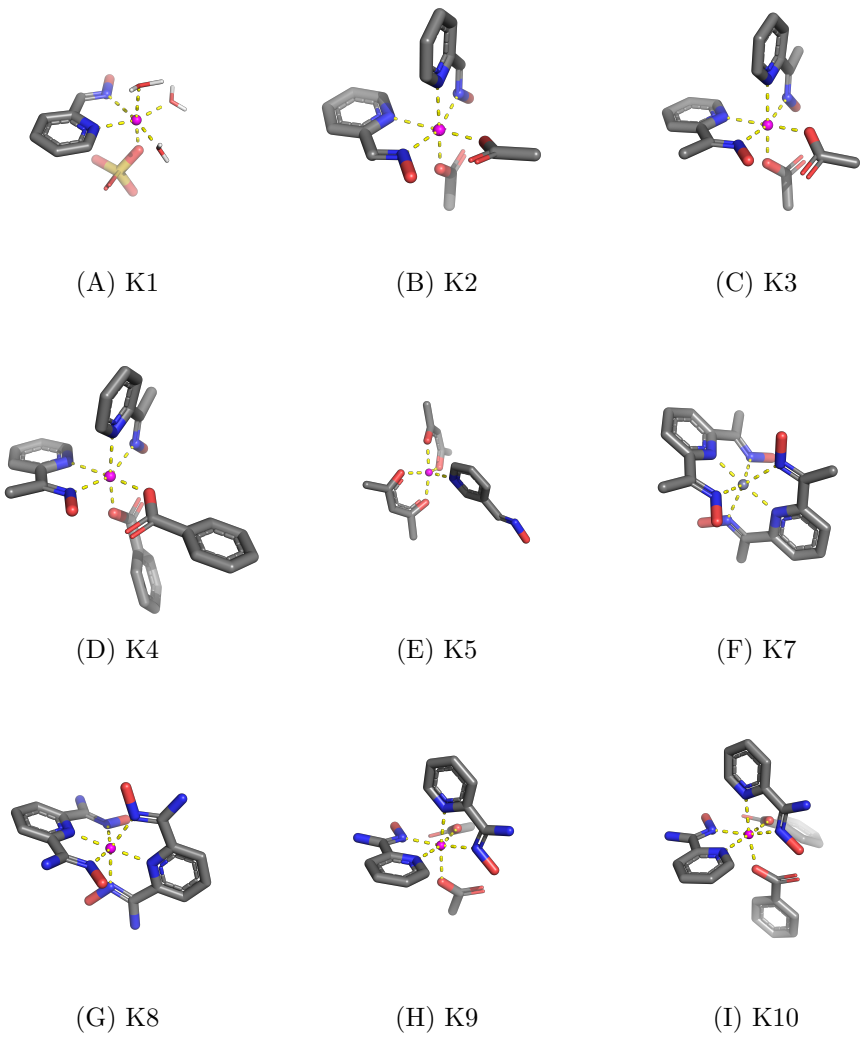
Following these findings, Konidaris *et al.* proposed zinc-pyridyloxime complexes [217] with either one (Figure 4.2) or two metal centers. These compounds were tested on paraoxon inhibited *Tc*AChE and they showed an enzyme reactivation rate up to 28.3%, which was still lower than the obidoxime's (i.e. 72.4%) (Table 4.1).

	Reactivation %		
Reactivator	1 nM	0.5 nM	0.1 nM
Compound 1	28.3	15.4	5.4
Compound 2	12.0	6.1	1.5
Compound 3	8.8	4.7	0.2
Compound 4	0.2	0.3	0.0
Obidoxime	72.4	70.6	57.3

Table 4.1: Reactivation potency (%) for paraoxon-inhibited AChE of mononuclear compounds (Compounds 1, 2, 3) and binuclear compound (Compound 4), compared with obidoxime, at different concentrations [217].

According to the authors, the poor reactivation potency might be due a to a distant orientation of the oximate towards the P–O bond. Although the modest results,

the metal-containing compounds retained their integrity in solution and the effectiveness of some ligands was found incosequential [217]. For this reason, Konidaris and colleagues synthesised a new series of zinc-containing reactivators (Figure 4.3), here called K-series, for which activity tests are currently ongoing. Nonetheless, some preliminary results showed a very similar reactivation rate, around 28-30% (data not shown).



(A) K1 (B) K2 (C) K3

(D) K4 (E) K5 (F) K7

(G) K8 (H) K9 (I) K10

Figure 4.3: K-series of AChE zinc-reactivators. The 3D structures were solved by X-ray. The zinc ion (purple sphere) is in octahedral geometry, with the exception of K5 where it is in square-base pyramidal geometry. The coordination bond is represented as a yellow dotted line.

With the aim of better exploring the role of the zinc ion in AChE reactivators and to rationalise the modest activities of these compounds, we investigated both the electronic features of the reactivators and the structural insights of the ternary complex (i.e. AChE, inhibitor, reactivator). First, a quantum mechanical

(QM) study was performed to quantify the increasing of nucleophilicity of the oxime group. Then, the dynamical behaviour of the enzyme was explored through molecular dynamics (MD) simulations. This last point is of particular interest given the well-known AChE "breathing" that influences the substrate binding [218][219][220]. For time's sake, only one compound was tested with MD simulation, that is K2, as an exemplification of all the others. To avoid any bias during the modelling procedure and to verify the reliability of the method, the entire study was performed also on the co-crystallized reactivator HI-6 (see **Supplementary Information** section).

4.3 Methods

4.3.1 QM calculations

The key step for the reactivation mechanism is the nucleophilic attack of the oxyme oxygen to the phosphorus atom of the OP [210]. Nucleophilicity is the ability of a species to react at an electron-deficient center and it is influenced by charge, electronegativity, solvent and steric hindrance. The addition of a zinc ion into oxyme-containing molecule is supposed to increase the nucleophilicity and, as a consequence, the potency of the reactivators. To verify this hypothesis, partial charge calculations of the oxyme oxygen atom were performed using both semi-empirical and density functional theory (DFT) calculation.
The semi-empirical program MOPAC2016 (Version: 20.015W) [175] was used with two different Hamiltonians: PM6-DH2 [221] and PM7 [222]. PM6-DH2 has improved and transferable H-bonding correction, which are present in some of K-series ligands, while PM7 has an improved parametrization for metal ions. Dielectric constant of 80 was set using the COSMO method for implicit solvent [223]. The DFT calculations were performed using Gaussian16 (Revison: A.03) [224], with the hybrid functional B3LYP [225], which is frequently used in presence of zinc ion [226][227][228][229], and two basis sets: a double ζ split valance 6-31G(d) and a triple ζ split valance 6-311G(d), both with polarization on heavy atoms. Grimme's GD3 correction of dispersion [230] was included, as well as Polarizable Continuum Model for the implicit solvent [231]. Electrostatic potential (ESP) charges [232] for DFT and net charges for semi-empirical calculation are reported. The K-series molecules were obtained through X-ray diffraction by Konidaris and colleagues, while the structures of the known reactivators used for comparison (i.e. 2-PAM, HI-6, obidoxime and trime-

doxime) were first optimized with molecular mechanics (MM) using CGenFF as force field [110].

4.3.2 Docking studies

Molecular docking was used to generate the ternary complex. Human AChE inhibited by paraoxon (which is the same inhibitor used in the K-series activity assays) and complexed with the reactivator HI-6 was chosen (PDB Id: 5HF9 [233]). The choice of such a structure was driven by the type of inhibitor and reactivator in complex with the enzyme, and by the quality of the resolution (i.e. 2.20 Å). However, the complex was obtained through soaking, which consists in the immersion of protein crystals into a solution containing the ligand, which eventually diffuses into protein cavities [234]. During soaking the crystal lattice could hinder conformational changes of the protein upon ligand binding [235][236], preventing ligand-induced conformational changes. Perhaps, it is no coincidence that in the majority of AChE crystal structures the oxime moiety of the reactivator is often placed relatively far from the catalytic serine (around 5-7 Å) [237][238][239][240]. Furthermore, the steric hindrance of the K-series compounds represents a challenge during classical rigid protein-flexible ligand docking. For these two reasons, an alternative docking protocol, combined with MD simulation, was designed.
First, missing residues of the crystal structure were manually added. The protonation state of the aminoacidic residues, histidine tautomers and asparagine/glutamine flipping were checked by H++ server [241], considering physiological pH of 7.4. Furthermore, Driant *et al.* [242] highlighted the importance of the protonation state of the active site residues to lower the energy barriers during the reactivation process. According to these findings, His447, belonging to the catalytic triad, was considered as protonated, Glu202, located near His447 (3-4 Å), was represented in its neutral form and the reactivator with the oximate group instead of oxime.
The reactivator was docked into AChE by GOLD [243] using the ChemPLP scoring function [244] and considering, as a binding site, a sphere centered on the paraoxon phosphorus atom with a radius of 18 Å, thus including all the residues of the gorge. Flexible docking was used, considering the side chains of Tyr72, Tyr124, Trp286, Phe295, Phe297, Phe338 and Tyr341 as flexible (Figure 4.4). Genetic algorithm (GA) parameters were set to 100 GA runs and a minimum of 100000 operations. Furthermore, the ligand QM charges computed with DFT B3LYP/6-311G(d) where used for docking. Although the combination of such an accurate method with dock-

ing can seem odd, the partial charges of ligands affect the performance of the scoring function [245], in particular considering also the presence of a metal ion. The best pose were chosen according to the best docking score.

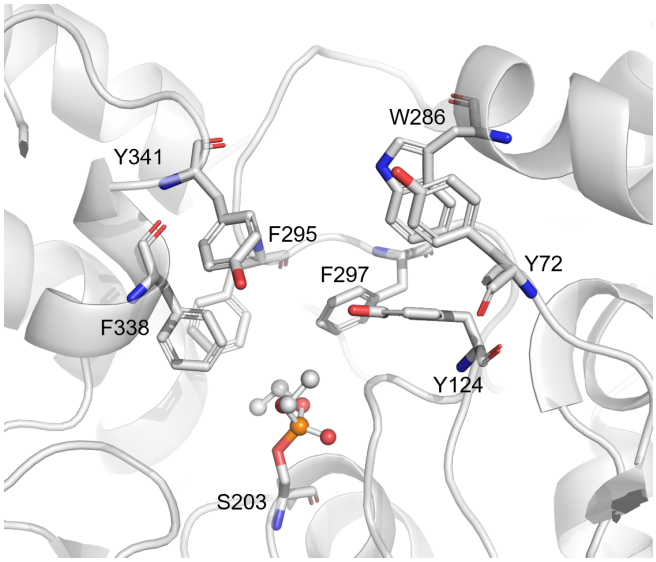


Figure 4.4: Binding pocket with residues considered as flexible during docking (represented in stick). The OP bound to Ser203 is also highlighted (ball and stick).

After the MD simulation (see below), the frame with the lowest potential energy was extracted and used for a new docking calculation. This second docking was performed following the settings described above, but considering the protein as rigid. This procedure resemble the induced-fit docking [56] and allows overcoming the limitations of steric hindrance and soaking crystallization.

4.3.3 MD simulations

The resulting complex was then solvated into a box of TIP3P water molecules, counter ions (Na^+ and Cl^-) were added to neutralize the system and to reach the ionic concentration of 0.15 M and PBC with PME were used. The protein, glycosidic adducts (i.e. N-acetyl-D-glucosamine and alpha-L-fucose) and paraoxon were parametrized with *ff14SB*, GLYCAM_06j and GAFF force fields respectively. The organic portion of the K-series ligands was treated with GAFF, while the interactions with the zinc ion were handled by the 12-6-4 model from Li-Merz *et al.* [115]. As a matter of fact, the classic non-bonded model struggled to maintain the integrity of the ligand during the simulations, whereas the 12-6-4 non-bonded model,

that is parametrized considering the octahedral geometry of the zinc ion, was able to preserve the ligand original structure. The system underwent to a three steps energy minimization using Amber18 [246]: first only hydrogen atoms were optimized, then water molecules and finally side chains. A heating phase of 10 ps followed the minimization, increasing the temperature from 0 to 300 K (using the Langevin thermostat) and applying position restraints on the α carbon atoms. SHAKE algorithm [99] was applied with a time step of 2 fs. The equilibration was conducted in the NPT ensemble (using the Berendsen barostat and maintaining the pressure at 1 atm) with position restraints on Cα for 50 ps, and then for 70 ps decreasing the weight of the restraints. The last step of equilibration was performed using the NVT ensemble and letting the protein free to move.
The production runs were conducted using the NVT ensemble, and forcing the ligand inside the binding site though distance restraints of 20 kcal/mol·Å between the oximate oxygen and the paraoxon phosphorus atoms, with an equilibrium distance of 2.00 Å, for 50 ns. Then, the simulation was run for 200 ns without restraints in order relax the system adapting it to the new position of the ligand.
The new docking pose (see above) was then used as a starting structure for further MD simulations, using the same procedure illustrated before (i.e. minimization, heating, equilibration and production). This last production run was 600 ns long, and it was performed twice, with an overall simulation time of 1.2 μs.

4.3.4 MM-GBSA free energy calculations

As a final step, binding free energy was computed using the MM-GBSA methods [127]. The Python program MM-PBSA.py [247] implemented in Amber18 was used, considering an external dielectric constant of 78.3, and an ionic concentration of 0.15 M. Generalized Born (GB) mbondi2 model was applied [248] with a surface tension value of 0.0072 kcal/mol/A^2 and a maximum distance between atom pairs of 25.0 Å. The calculation was performed for all the MD frame (30000 frame in 600 ns), for both the replica and the average value (kcal/mol) was returned.

4.4 Results

4.4.1 Nucleophilicity analysis

To quantify the influence of the zinc ion on the partial charge of the oxygen atom of the oxime of reactivators, that is responsible for the nucleophilic attack, QM methods were applied. In detail, both semi-empirical and DFT calculations were performed. Two Hamiltonians were used for the semi-empirical calculation: PM6 with a correction for the hydrogen bonding (which are present in K2, K3, K4, K9 and K10) and PM7, that is the newest version of the PM Hamiltonians. In the DFT calculation the hybrid functional B3LYP and two different basis-set were used: a double ζ split valance 6-31G(d), and a triple ζ split valance 6-311G(d).
Although some successful examples are available [249], generally semi-empirical method struggle to generate accurate results when metal ions are involved in the calculation [250]. A recent study reported optimized molecular geometries and electronic properties calculated by the PM6 method of 91.2 million molecules catalogued in PubChem [251]. Geometries of zinc ion-containing compounds were correctly predicted only in the 65.2% of the instances, while other elements like phosphorous or sulfur were predicted correctly in the 90.9% and 97.8% of the instances respectively. Therefore DFT methods were also applied and differences were found between the double and the triple ζ split valance basis set.
In Table 4.2 the partial charges of oxime oxygen atoms are reported. Confronting some of the most popular reactivators and the K-series, with both semi-empirical and DFT methods, the mean partial charge of the oxygen atom increase in presence of the zinc ion. It's worth noticing that, in K-series ligands, the increase of accuracy of the QM method leads to an increase of the negative charge on the oxygen atom as well; while considering the organic known reactivators smaller differences are reported. This is a further proof the of need of high level methods when metal ions are involved in the calculation.

	Semi-empirical		DFT (B3LYP)	
Reactivator	PM6-DH2	PM7	6-31G(d)	6-311G(d)
2-PAM	-0.355	-0.315	-0.382	-0.415
HI-6	-0.338	-0.301	-0.361	-0.392
Obidoxime	-0.349	-0.316	-0.358	-0.388
Trimedoxime	-0.356	-0.321	-0.365	-0.395
K1	-0.380	-0.381	-0.547	-0.629
K2	-0.362	-0.303	-0.513	-0.594
K3	-0.358	-0.290	-0.549	-0.629
K4	-0.391	-0.359	-0.515	-0.578
K5	-0.406	-0.403	-0.479	-0.539
K7	-0.342	-0.307	-0.526	-0.588
K8	-0.387	-0.324	-0.504	-0.548
K9	-0.395	-0.344	-0.565	-0.628
K10	-0.423	-0.374	-0.569	-0.628
Known reactivators mean	-0.350	-0.313	-0.367	-0.398
K-series reactivators mean	-0.383	-0.343	-0.530	-0.596

Table 4.2: Oxime oxygen partial charges of the K-series and four known reactivators. Partial charges were computed using both semi-empirical and DFT methods. PM6-DH2 and PM7 Hamiltonians were used in the semi-empirical calculation, while the hybrid functional B3LYP was chosen for the DFT calculation using both a double and a triple ζ split valance basis set. The mean value of oxime oxygen atoms is reported when more than one oxime group is present. In the lower panel the mean values of the known reactivators and the K-series are reported for each method.

As expected, the farther the oxime moiety from the zinc ion, the less negative the oxygen atom (i.e. K5 with -0.539 using DFT B3LYP/6-311G(d)). This evidences suggest that the inclusion of the zinc ion can facilitate the nucleophilic attack of the reactivators through increasing of the electron density.

4.4.2 Generation of the ternary complex

To generate the ternary complex, K2 was docked into the binding site of AChE, as described in the **Methods** section. Given the K2 large size, flexible docking procedure was chosen. The side chains of seven bulky residues located at the mid

gorge (Phe295, Phe297, Tyr338, Tyr341), at the PAS (Tyr72, Tyr286) and at the CAS (Tyr124) were free to rotate [252]. The docking procedure was previously tested by redocking on nine covalently inhibited AChE structures in complex with reactivators, obtaining an average RMSD of 2.45 Å (Table S4.1). HI-6 was docked in a similar conformation, compared to the crystal structure (Figure S4.1). However, such a docking procedure was not able to place K2 near the bottom of the binding pocket, but on the contrary the ligand remained at the entrance of the gorge, with a distance between the oximate group and the OP of 10.6 Å (Figure 4.5).

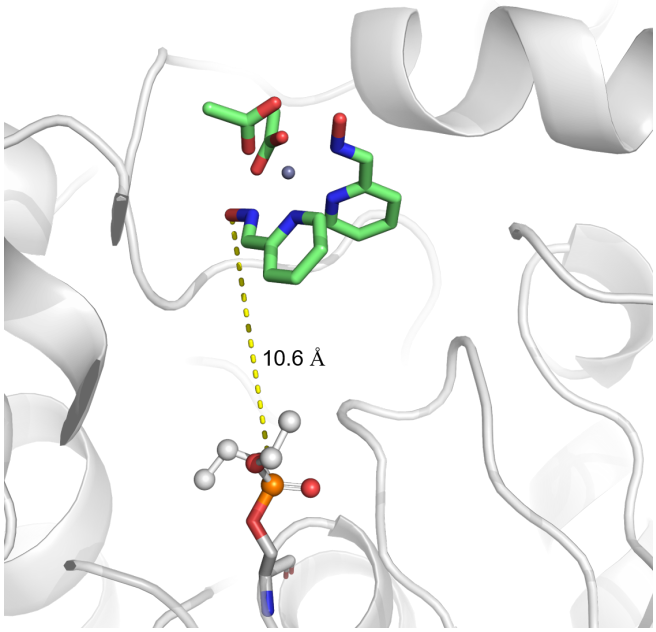


Figure 4.5: Docking pose of K2 (in green) obtained by flexible docking. K2 steric hindrance and the bulky residues at the mid gorge prevented the approach of the reactivator to the OP. The distance of 10.6 Å between K2 and paraoxon (in ball and stick) is highlighted as a yellow dotted line. The zinc ion is represented as a silver sphere.

Therefore, an alternative docking protocol was designed: the docking complex underwent an MD simulation, the frame with the lowest potential energy was extracted, and that protein structure was used for a new docking calculation (see below). This approach, which resembles induced-fit docking, allows considering the flexibility of the protein and the influence that the ligand exerts on protein conformation. Protein flexibility is of particular importance in AChE given the gorge "breathing" that facilitates the pass-through of the substrate and inhibitors [220].

The complex obtained from docking (Figure 4.5) underwent to an MD simulation in which the reactivator was moved closer to Ser203 with distance restraints for 50 ns, then the system was left free to move for 200 ns to let the enzyme adapt to the new conformation.

Figure 4.6A shows the RMSD of the protein backbone along the trajectory which, after a first phase of adjustment, reached convergence fluctuating no more than 1 Å. Thus, forcing of K2 along the gorge did not impair the protein folding. Figure 4.6B refers to K2 RMSD along the trajectory, as seen for the backbone, also the reactivator reached stability. The distance between the center of mass of K2 and paraoxon was monitored (Figure 4.6C) highlighting that after 50 ns, K2 moved away from the OP, assessing at a distance between 8 and 10 Å. The frame with the lowest potential energy was extracted for the next step of docking (Figure 4.6D). A comparable result was obtained with HI-6, with major fluctuation of HI-6 compared to K2, probably due to its higher flexibility (Figure S4.2).

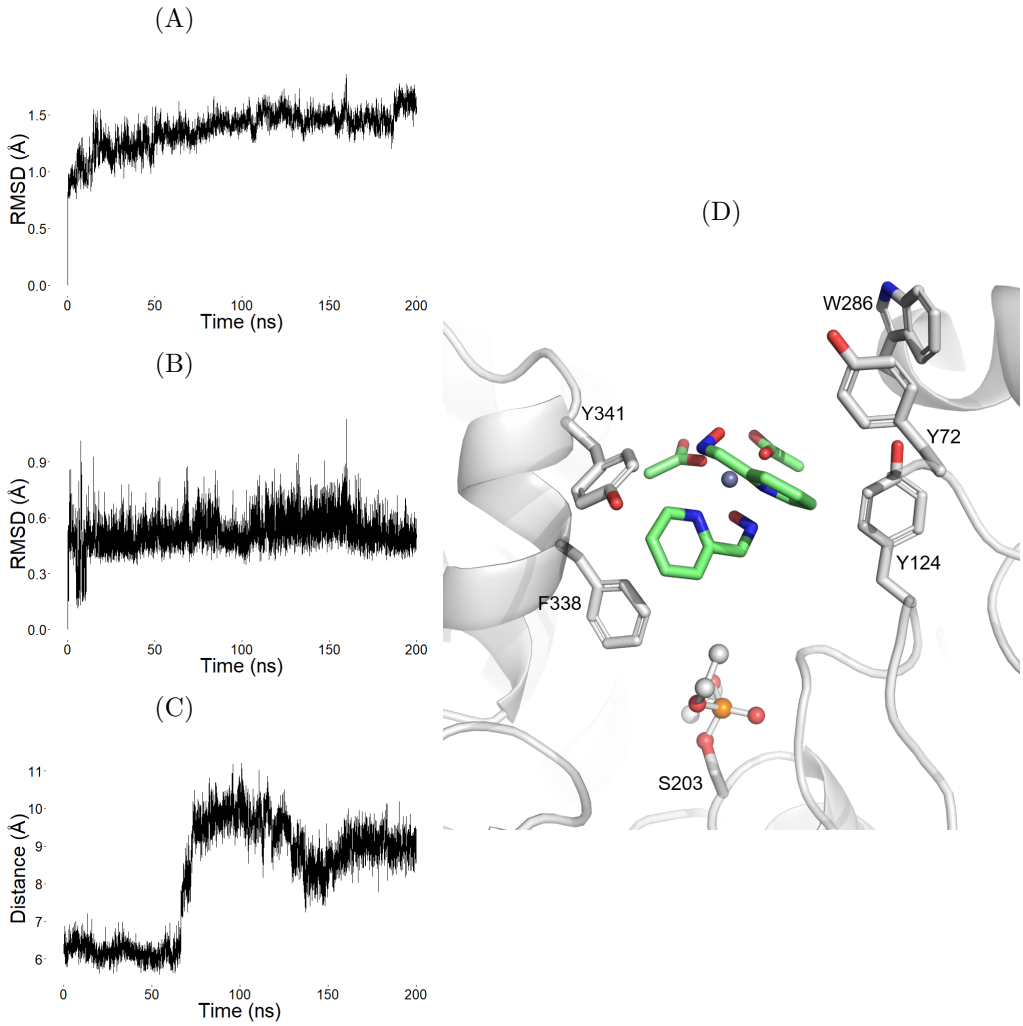


Figure 4.6: Results from K2 MD simulation. Distance restraints were used for 50 ns to force K2 entrance into the binding site, then the system was left free to move for 200 ns. (A) Protein backbone and (B) K2 RMSD fluctuation (Å) along the trajectory are reported, as well as (C) the distance (Å) between K2 and the OP. Panel (D) shows the frame chosen for the new docking run. K2 is coloured in green, the OP is represented in ball and stick, the zinc ion is represented as a silver sphere, and the bulky residues of the gorge are shown.

Water molecules, counter ions and the reactivator were removed from the system,

and a second docking calculation was performed using the new protein structure. The second docking was performed with rigid protein, as the structure was supposed to enlarge during the MD. This time, K2 was placed at the CAS (Figure 4.7), as well as HI-6 (Figure S4.3).

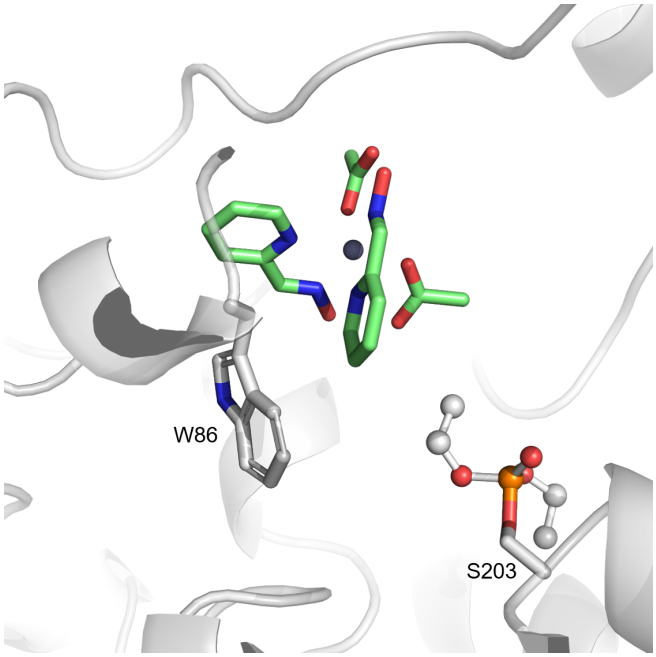


Figure 4.7: Docking pose of K2 (in green) obtained using the MD-derived structure of AChE. The reactivator interacts with Trp86 via π-π stacking interaction, at the bottom of the gorge, near the OP. Paraoxon is represented in ball and stick, while the zinc ion as a silver sphere.

The volume of the gorge cavity expanded during the trajectory leading K2 to settle at the bottom of the binding pocket, and eliciting both π-π stacking and π-cation interactions with Trp86. These interactions were also present in HI-6 docking pose, differently from the crystal structure in which this interaction is missing and HI-6 binds above the CAS. Actually, in all the available crystal structures, none of the reactivators seems to bind in an active conformation, remaining quite far from the OP (5-7 Å) [237][238][239][240]. On the contrary, HI-6 docking pose seems much more favourable for the reaction with the inhibitor. This difference could be explained by the soaking crystallization procedure, which could hinder the ligand-induced rearrangement of the protein, preventing the descent of the ligand into the gorge. Furthermore, evidences of Trp86 importance are widely reported in literature [253][254][255][256][257][258][259].

4.4.3 MD simulation analysis

To evaluate the stability of the ternary complex, the system underwent to a further 600 ns long MD simulation (with two replica). Backbone and K2 RMSD along the trajectory suggest that the convergence were reached in both the replica (Figure 4.8A-B). Nonetheless, K2-OP distance fluctuated through the trajectory: in the first replica, after almost 150 ns, K2 moved towards the upper portion of the gorge; while in the second replica it remained stable at the CAS (Figure 4.8C). A similar RMSD profile was obtained for HI-6 (Figure S4.4A-B), while the fluctuation of HI-6-OP distance (Figure S4.4C) is slightly different with a higher number of HI-6 approaches to paraoxon, compared to K2. This difference can be related to the difficulty of K2 to move along the gorge because of its steric hindrance whereas HI-6, which has an elongated shape, can easily diffuse along the binding site.

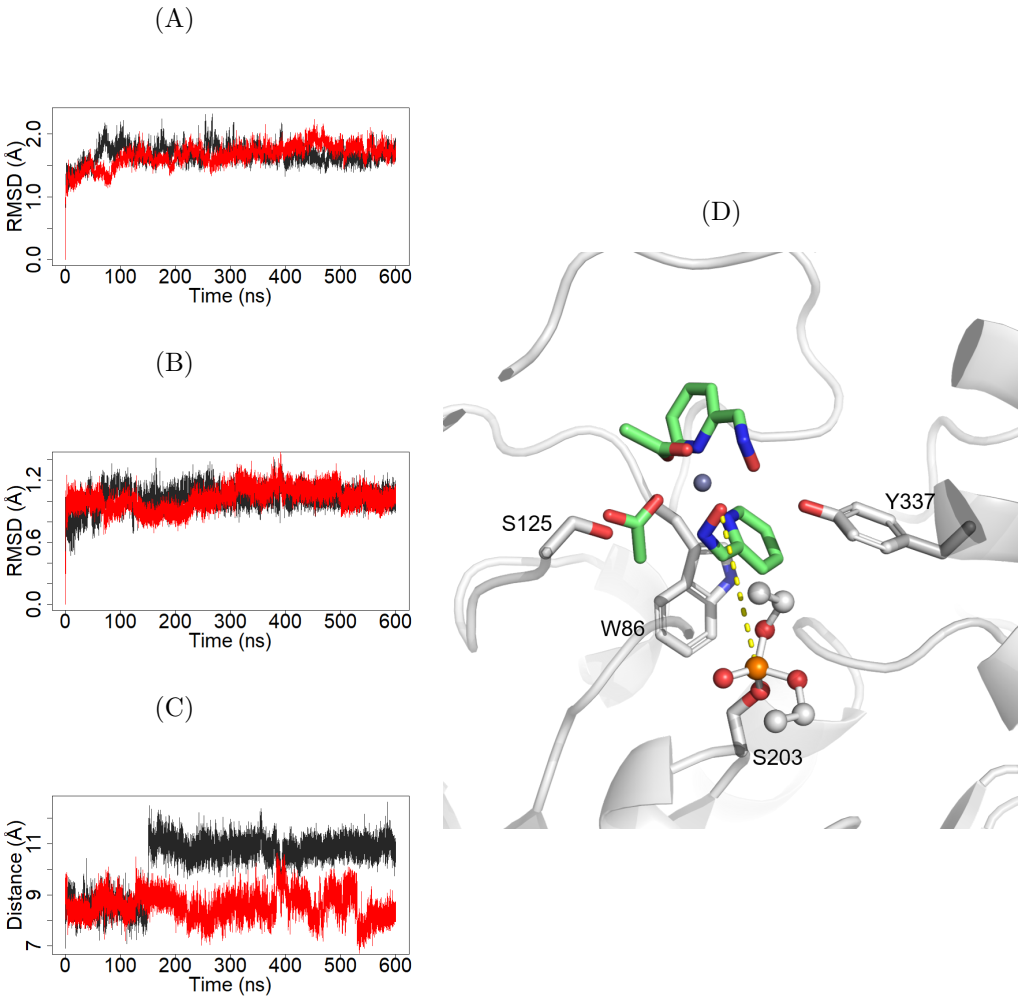


Figure 4.8: Results from K2 MD simulation. Black lines refer to the first run, while red lines refer to the second run (overall time: 1.2 μs). (A) Protein backbone and (B) K2 RMSD fluctuation (Å) along the trajectory are reported, as well as (C) the distance (Å) between K2 and the OP. Panel (D) shows a representative AChE-K2 complex extracted from the 600 ns MD: K2 is coloured in green, the OP is represented with ball and stick, the zinc ion is reported as a silver sphere. The distance between OP and the oxime is highlighted as a yellow dotted line (6.1 Å).

The trajectories were clustered based on the reactivator RMSD value, with a cut-off of 1 Å. Figure 4.8D shows the representative K2 conformation. As seen before, the interaction with Trp86 is still present, as well as H-bond with Tyr337 and Ser125.

HI-6 trajectories were clustered as well, and the representative conformation is shown in Figure S4.4D. The binding mode differs from the one of the crystal structure: π-π stacking and π-cation interactions were elicited with Trp86 and Tyr72, while Tyr337 established a H-bond interaction with the oximate group. Moreover, the HI-6-OP distance is 6.0 Å, which is shorter than the one of the crystal structure (9.7 Å).

4.4.4 Binding free energy

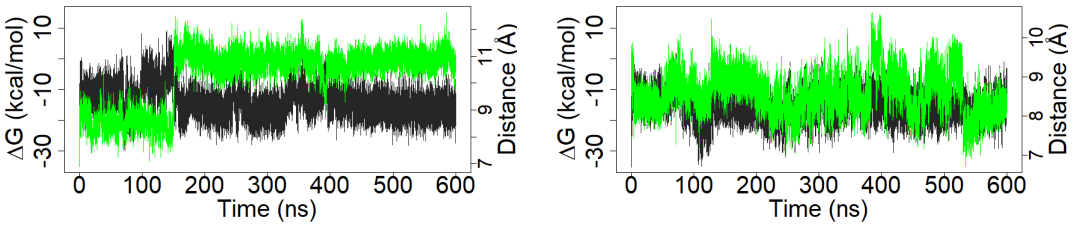
The K2-OP distance profile of the first replica and the interaction that K2 elicited with the binding site, suggest that the poor activity of K2 is due to its steric hindrance. Therefore, MM-GBSA free energy calculation was performed to asses the binding free energy of K2, compared to HI-6. The calculation was performed on all the frames of both the trajectories, and the average ΔG values of each trajectory are reported in Table 4.3.

	ΔG replica 1 (kcal/mol)	ΔG replica 2 (kcal/mol)
HI-6	-29.26 $\pm$ 4.22	-29.10 $\pm$ 3.40
K2	-14.03 $\pm$ 4.86	-15.83 $\pm$ 4.41

Table 4.3: MM-GBSA free energy calculation results. Energy values are reported in kcal/mol. The ΔG refers to the mean energy value of all the frames of the trajectories, which is composed by hydrophobic and electrostatic contribution calculated with both MM and GB methods. Values for both the replica (replica 1 and replica 2) are reported.

HI-6 and K2 binding energies differ by almost 14 kcal/mol. MD simulation results suggest a propensity of K2 to bind at the CAS, where it can elicit several interactions (see above). However the free energy calculation results highlight a weaker binding of K2 compared to HI-6. This is in according to preliminary activity data and it could explain the displacement of K2 from the CAS after almost 150 ns of the replica 1 production.

In Figure 4.9 the binding free energy trend for both the replica is reported (in black), together with the distance between K2 and paraoxon (in green). In replica 1 (Figure 4.9A) the increase of K2-OP distance after almost 150 ns was associated with a reduction of the binding energy that remained stable until the end of the trajectory. While in replica 2 (Figure 4.9B) decreasing of the K2-OP distance corresponded to a decrease of the binding energy as well.



(A) (B)

Figure 4.9: Comparison between K2 binding free energy and K2-OP distance along the MD trajectories. Results for (A) replica 1 and (B) replica 2 are reported. Black lines indicate the binding free energy and its values refer to the left y-axis (ΔG (kcal/mol)), while green lines indicate the K2-OP distance and its values refer to the right y-axis (Distance (Å)).

Although in replica 1 K2 stabilized at 11 Å away from the OP, the lowest values of energy were reached in replica 2, where K2 was placed closer to paraoxon. This suggest that K2 binding to AChE is energetically favourite in proximity of paraoxon (i.e. at the CAS), however it can also stabilised away from the OP (i.e. at the PAS).

4.5 Discussion

The key role of AChE reactivators is to unbind the OP from the catalytic Ser203, by nucleophilic attack on the phosphorus atom of the inhibitor [213]. Therefore, nucleophilicity is an essential factor for new reactivators design. The most popular functional group found in AChE reactivators is the oxime [211], nonetheless different strategies have been proposed to further increase the nucleophilicity of such a group, like including metal ions [214][215]. A new series of zinc-containing ligands (K-series) designed by Konidaris and colleagues are presented, with a moderate reactivation capacity (i.e. $\sim$30%).

Partial charge on the oxime oxygen atom has been computed as representative of nucleophilicity. Compared with other known reactivators, the K-series ligands seem to have increased nucleophilicity, thanks to the zinc ion. The fact that the lower partial charge found in K-series ligands belonged to K5, which is the compound where the oxime group is farthest from the metal ion (Figure 4.3), further confirms the role that the zinc ion plays in nucleophilicity. Partial charges were computed

by both semi-empirical and DFT methods. The former was unable to find differences between know reactivator and metal-containing K-series ligands (the average difference between the partial charges of the two types of reactivators was 0.03). On the contrary, DFT has proven to be more efficient in finding the difference derived from the insertion of zinc ion (0.163 and 0.198 with double and triple ζ split valance basis set respectively, Table 4.2), and these results highlighted the need for high level methods when metal ions are involved.

A combination of molecular docking and MD simulation was used to generate the AChE-OP-reactivator complex. Difficulties arising from the crystallization procedure (i.e. soaking), the bulky shape of the ligand and the conformational change of the enzyme (i.e. "breathing") required the development of an alternative docking protocol. K2 was first docked into the binding pocket, then it was forced to enter the bottom of the gorge by MD simulation, and finally it was re-docked into the protein structure derived from the MD. This approach allowed AChE to adapt to the ligand and to modify the binding site conformation. Although forcing the ligand inside the binding pocket could seem a harsh choice, the RMSD of protein backbone during the 200 ns production testifies that no denaturation occurred (Figure 4.6). This procedure has the same rational of the induced-fit docking, but it allows a more extensive rearrangement of the binding site [260].
A long MD simulation was performed (1.2 μs overall) on both HI-6 and K2. Interestingly, HI-6 arranged differently compared to the crystal structure [233] (Figure S4.4D), in a way that seems much more incline to react with the OP (HI-6-OP distance of 6.0 Å). Both K2 and HI-6 located at the bottom of the gorge, interacting with different aromatic residues (i.e. Tyr72, Trp86 and Tyr337), suggesting the important role of π-π and cation-π interactions. The choice of the correct parameters to treat the zinc ion interactions in K2 was essential to maintain the integrity of the ligand. Li-Merz 12-6-4 non-bonded model for divalent metal ions [115] was chosen to model K2, for its parameters tuned for the octahedral geometry.

Binding free energy between AChE and the reactivators was computed using the MM-GBSA method, showing that K2 binds to the enzyme much weaker than HI-6 (Table 4.3). However, it is worth noticing that the binding was more favourite in replica 2, where K2 was located at the CAS, compared to replica 1; so the question arises as to what is the reason of the relocation of K2 to the top of the gorge. One possible explanation is that such a movement was caused by the steric hindrance of the K2, which overcame the energy of the other interactions.

4.6 Conclusions

This study investigated the reasons for the modest reactivation activity of the K-series ligand. The results suggest that this is not due to a lack of neither nucleophilicity nor interactions, but to the steric hindrance of K2 which prevents the reactivator from remaining at the bottom of the gorge long enough to exploit its function. The insertion of the zinc ion and the presence of aromatic group carrying quaternary nitrogen seems to be functional, although a more elongated shape should be considered. In addition to these findings, this work highlights the importance of combining various methods, with different accuracy levels as well, into one workflow. Such an approach allows to perform a complete analysis of the problem, revealing multiple aspects of the same system. Finally, the method proposed here can be applied for the design of new AChE reactivators and, in general, of zinc-containing compounds.

4.7 Supplementary Information

PDB Id	Resolution (Å)	Species	RMSD (Å)
2WG1 [238]	2.21	*Torpedo Californica*	3.61
2WHP [237]	2.21	*Mus musculus*	1.75
2WHQ [237]	2.15	*Mus musculus*	3.98
5DTJ [261]	2.71	*Mus musculus*	1.55
5HF9 [233]	2.21	Human	2.52
5HFA [233]	2.22	Human	4.06
6CQW [239]	2.27	Human	1.98
6O5S [240]	2.81	Human	3.36
6O66 [240]	2.45	Human	0.89

Table S4.1: RMSD values (Å) of different AChE species obtained by reactivators redocking. All the AChE structures are covalently inhibited by OPs. Resolution values (Å) are also reported.

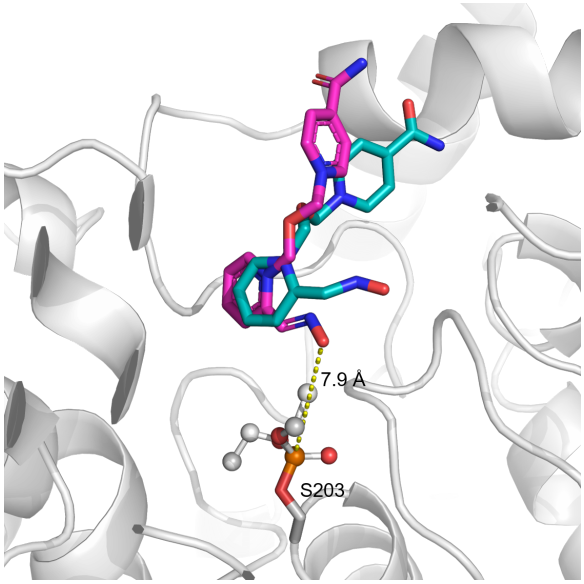


Figure S4.1: Superimposition between HI-6 in the crystal structure (in light blue) and in the docking pose obtained by flexible docking (in purple). The distance between docked HI-6 oximate and paraoxon (in ball and stick) is highlighted as a yellow dotted line. In the crystal structure (PDB Id: 5HF9 [233]) such a distance is 9.7 Å, while in the docking pose is 7.9 Å.

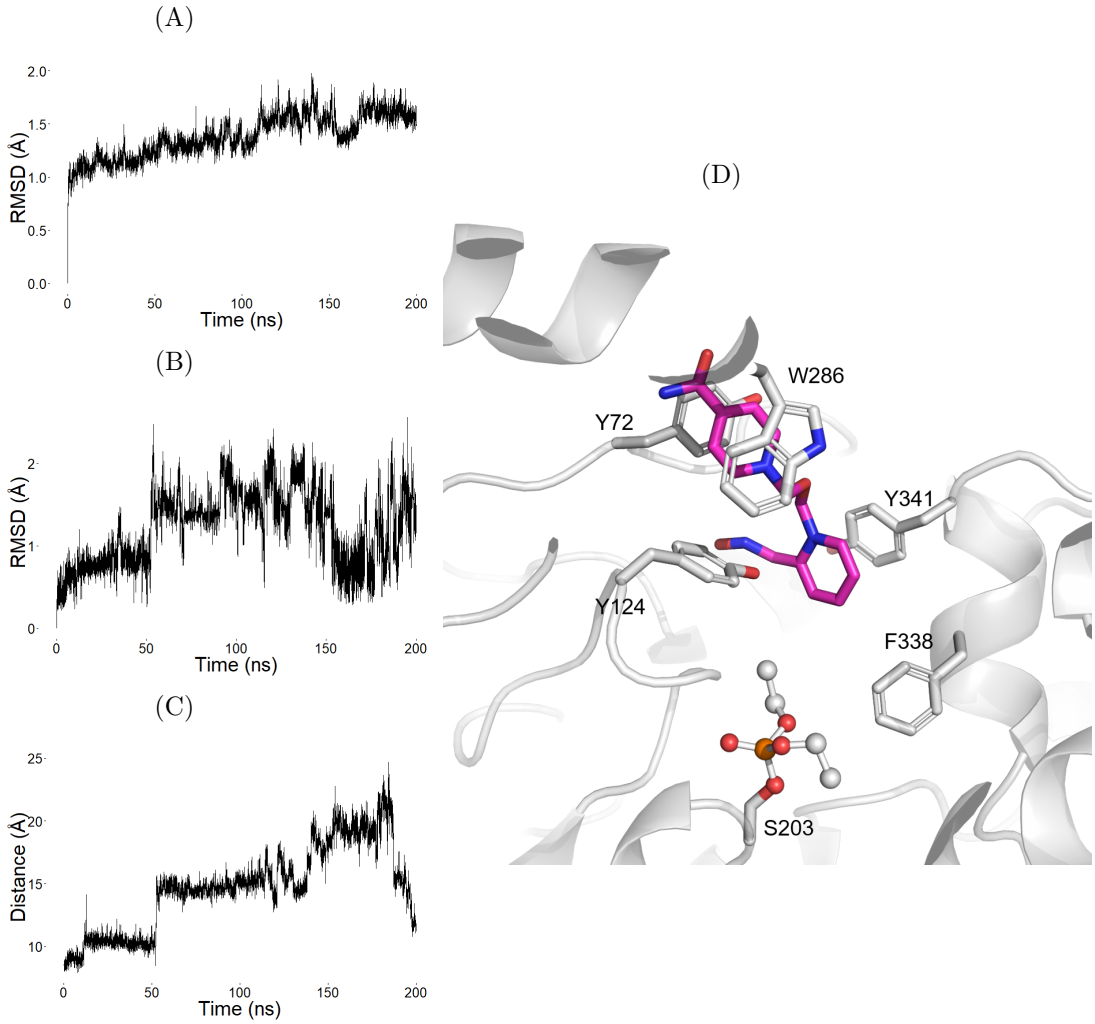


Figure S4.2: Results from HI-6 MD simulation. Distance restraints were used for 50 ns to force HI-6 entrance into the binding site, then the system was left free to move for 200 ns. (A) Protein backbone and (B) HI-6 RMSD fluctuation (Å) along the trajectory are reported, (C) as well as the distance (Å) between HI-6 and the OP. Panel (D) shows the frame chosen for the new docking run. HI-6 is coloured in purple, the OP is represented in ball and stick and the bulky residues of the gorge are shown. The main interactions between HI-6 and AChE are hydrophobic, with a clear π-π sandwich with Tyr72 and Trp286.

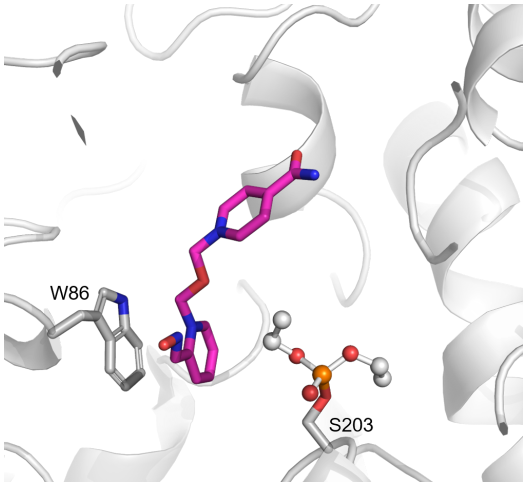


Figure S4.3: HI-6 docking pose (in purple) obtained by rigid docking, starting with the MD-derived structure of AChE. Trp86 (in stick) elicits both π - π and π - cation interaction with HI-6 pyridinium ring. Paraoxon is represented in ball and stick.

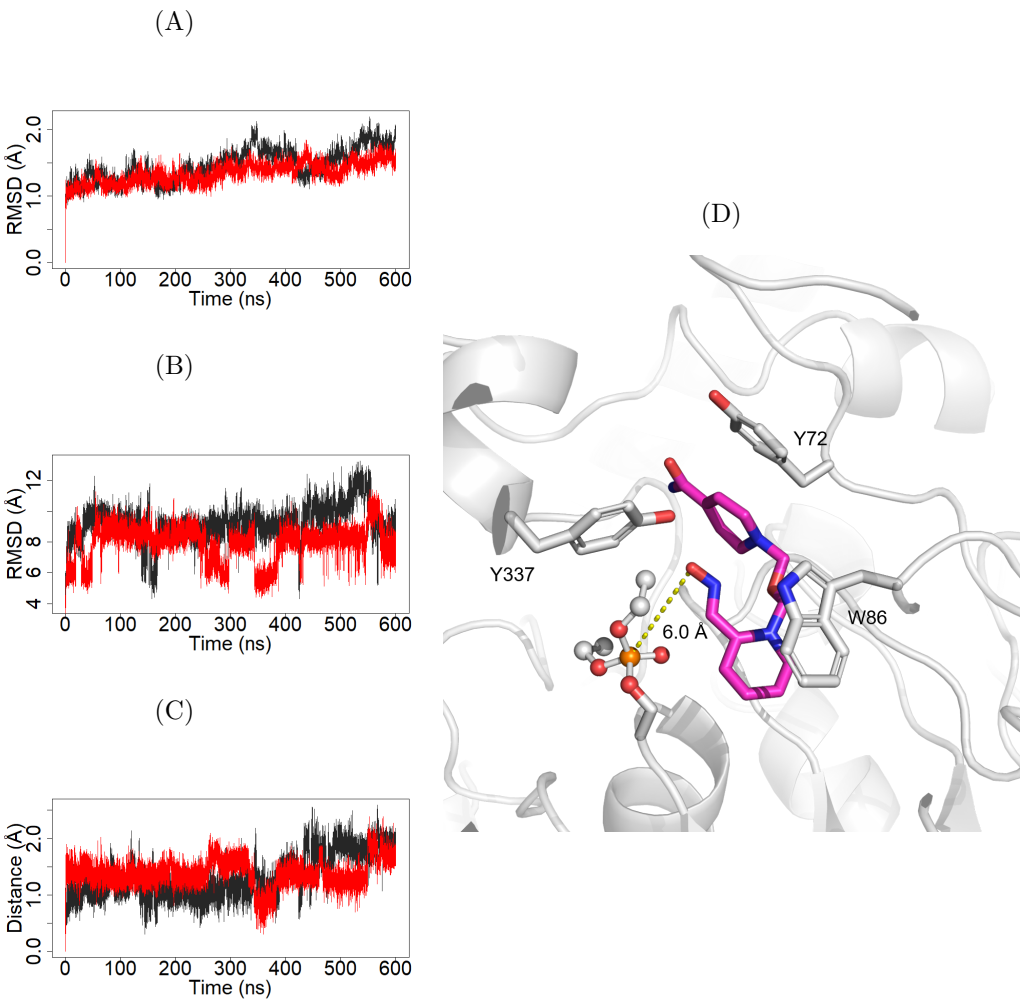


Figure S4.4: Results from HI-6 MD simulation. Black lines refer to the first run, while red lines refer to the second run (overall time: 1.2 μs). (A) Protein backbone and (B) HI-6 RMSD fluctuation (Å) along the trajectory are reported, as well as (C) the distance (Å) between the HI-6 and the OP. Panel (D) shows a representative AChE-HI-6 complex extracted from the 600 ns MD: HI-6 is coloured in purple, the OP is represented with ball and stick. The O-P distance is highlighted as a yellow dotted line.

Chapter 5

Validation of CN1 pharmacophore and discovery of new putative inhibitors through a combined approach

5.1 Carnosine and CN1

Carnosine (β-alanyl-L-histidine) is an endogenous peptide mostly present in skeletal muscle of humans, which has risen lot of interests in the scientific community owing to its properties such as metal chelation [262], buffering [263] and detoxification of reactive oxygen [264] and carbonyl species [265]. In the past decade, several studies have been carried out to evaluate the pharmacological properties of carnosine in animal models of human diseases [266]. Atherosclerosis [267][268], diabetes [269], metabolic syndrome [270][271] and neurological disorders [272] are few examples of animal models where carnosine have been tested resulting in an overall consensus around the benefits of carnosine intake for the prevention or mitigation of oxidative stress-based diseases. Nevertheless, in humans, carnosine has limited activity since the peptide is rapidly hydrolyzed upon absorption [265]. Two enzymes, namely human carnosinase (i.e. CN1; E.C. 3.4.13.20) and cytosolic nonspecific dipeptidase (i.e. CN2; EC. 3.4.13.18) are the main responsible for such a metabolism [273].
CN1, which is a homodimeric zinc-dependent metalloprotease belonging to the M20

family, is mostly produced in the brain and through the cerebrospinal fluid reaches the blood, where is the main responsible for overall hydrolysis of carnosine absorbed by diet [274]. Individual carnosinase activity can be quite different in humans due to polymorphisms occurring on CN1 gene (CNDP1) [275]. Interestingly, it has been reported that subjects with a low carnosinase activity due to specific CN1 polymorphisms have also a lower onset of diabetic nephropathy. To explain this phenomenon, it has been speculated that the less carnosine is hydrolyzed the more it is excreted by the kidney, where it can accumulate and protect renal cells from the adverse effect of high glucose levels [276].
Over the years, several pharmaceutical strategies were explored with the aim of enhancing carnosine benefits. The increase of stability towards CN1 degradation is considered crucial since carnosine seems to exert a different activity in plasma, compared to other body compartments [277]. One strategy was to design CN1-stable analogs of carnosine by chemical modification of the structure (e.g. carnosinol) [278] or by the simple inversion of the peptide chiral center (i.e. D-carnosine) [279]. Unfortunately, the inversion of the chiral center underwent a negative impact on peptide absorption, requiring the design of suitable prodrugs (i.e. D-carnosine octylester) [268]. Alternative strategy of developing CN1 inhibitors have been reported with the successful discovery of carnostatine [280] and salicyl-carnosine [281].

5.2 Aim

The lack of CN1 inhibitors and of experimental binding data makes difficult to design new molecules able to block CN1 activity. However, several studies have been made on hydrolytic rate of carnosine and its derivatives [282][283][276][280][284][285][286]. Rationalization of these experimental results can help to propose an interaction model and to generate a pharmacophore hypothesis that could be applied for the design of novel inhibitors. This work is divided into two main parts: (1) the molecular interpretation of the serum hydrolytic rate of histidine dipeptides and (2) the screening of different datasets to identify putative inhibitors. (1) In the first part, serum hydrolytic rates reported in Table 5.1 were used as reference.

	Serum hydrolytic rate (nmol/h · μL)
Carnosine	1.328 ± 0.081
Anserine	0.495 ± 0.036
Balenine	0.050 ± 0.001
Homocarnosine	0.017 ± 0.001

Table 5.1: Serum hydrolytic rate of natural histidine dipeptides. Three measures were performed at 50 μM concentration [52].

The four dipeptides reported in Figure 5.1 were first docked into CN1, then DFT single point energy calculations were performed on a sub-system of the complex, to quantify the binding affinity.

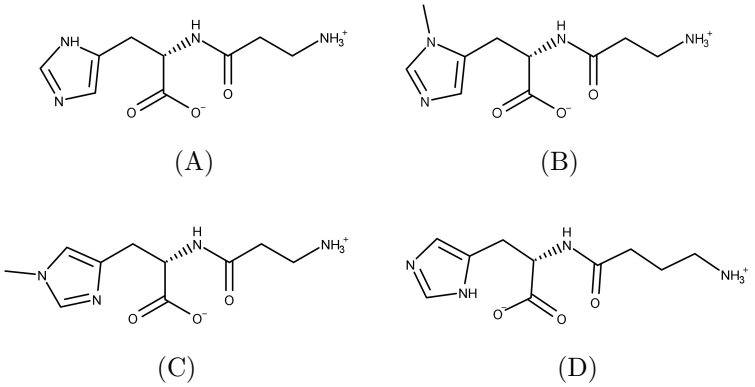


Figure 5.1: Structure of histidine dipetides. (A) carnosine, (B) anserine, (C) balenine, (D) homocarnosine. Carnosine and homocarnosine are represented in their δ tautomer.

(2) In the second part of the work, five different datasets containing overall 11214 molecules where screened by docking on CN1. The best poses underwent to 100 ns MD simulation, to assess the stability of the docking poses. Finally, binding free energies were computed by MM-GBSA, to identify new hit compounds. Carnosine and its derivatives were used as reference.

5.3 Methods

5.3.1 Protein preparation

The crystal structure of human carnosinase was retrieved from Protein Data Bank (PDB Id: 3DLJ), and it was co-crystallized with unknown atoms (probably SO_4^{2-}, according to the electron density). Simulations were based on the chain A which is resolved with better quality compared to chain B. Three missing portions of the protein were added manually (i.e. residues 77-78, 208-209 and 437-438). The protonation state of protein residues at physiologic pH, the tautomer of histidine residues and the Ans/Gln flipping were checked using the H++ server [241]. From this analysis and the previous work of Pavlin *et al.* [93], we considered Glu173 as neutral. In order to avoid the binding pocket closure, we added SO_4^{2-} within the binding pocket according to the electron density and following the geometrical orientation found in other crystal structures of bi-zinc hydrolases (e.g. carbapenamase PDB Id: 3F9O [287], β-lactamase PDB Id: 1JT1 [288]). The structure was then solvated in a cubic box of TIP3P water molecules, and counter ions (Na^+ and Cl^-) were added to neutralize the system, reaching an ionic concentration of 0.15 M. The resulting system was then optimized by a three steps minimization, using Amber18 [246], involving first the hydrogen atoms, then the water molecules and finally the side chains, keeping harmonic restraints in the Cα of the protein. The ligand (i.e. SO_4^{2-}) was parametrized using antechamber with AM1-bcc charges and GAFF, the protein using *ff14SB* and the zinc ions using the Li/Merz ion parameters of divalent to tetravalent ions [24]. Furthermore, in order to stabilize the interaction with the metal ions and their coordination geometry, distance restraints (30 Kcal/mol·Å) between the ions and the coordinating atoms of the protein were applied (Figure S7.1).

5.3.2 Molecular docking

The minimized structure of the protein was used for the docking study. Water molecules and counter ions were removed, while the SO_4^{2-} ligand was replaced by a hydroxyl ion, which is known to be implicated in the catalytic mechanism and it is found in other M20 di-zinc peptidases (e.g. PDB Id: 5UEJ, 1CG2 [289], 2PFE, 2V8D

[290]). The docking calculations of both carnosine and homocarnosine involved all the tautomers (ϵ and δ). The binding pocket used for docking was chosen by selecting a radius of 10 Å around the hydroxyl ion. Docking calculations were performed using GOLD [65] by generating 100 poses per molecule, using the ChemPLP scoring function and genetic algorithm (GA) parameters were set to 100 GA runs with a minimum of 100000 operations. Docking poses were then optimized by following the procedure described in **Protein preparation** section.

5.3.3 Cationic Dummy Atom MD simulation

After the minimization, the systems underwent to 100 ns MD simulation to asses the stability of the ligand inside the binding pocket. The heating phase of 10 ps followed the minimization, bringing the temperature up to 300 K (using the Langevin thermostat) and keeping restraints on alpha carbon atoms. SHAKE algorithm [99] was applied with a time step of 2 fs. The equilibration was conducted in the NPT ensemble (using Berendsen barostat, keeping the pressure at 1 atm) with restraints on the Cα for 50 ps, and then for 70 ps decreasing the weight of the restraints. A last step of equilibration was performed using the NVT ensemble and letting the protein free to move. PBC with PME were used. Production runs of 100 ns were conducted using the NVT ensemble.
Special attention was paid in the treatment of zinc ion interactions. In the crystal structure, both the zinc ions are in the tetrahedral geometry, each coordinated by three residues and the unknown ligand (probably SO_4^{2-}, as mentioned before). It is reasonable to assume that, similarly to other di-zinc hydrolases [291][292][123][122], when the substrate (i.e. carnosine) binds to the catalytic pocket, one zinc ion temporally changes its coordination number into five (or six), until the enzymatic reaction is done. The classic non-bonded model is unable to describe such a complex behaviour [112], therefore the cationic dummy atom model was chosen. The zinc ion coordinated by His106, Asp139, Asp202 and the unknown ligand (here called ZN1) was considered as tetrahedral, while the zinc ion coordinated by Asp139, Glu174, His452 and the unknown ligand (here called ZN2) was considered with a trigonal bipyramidal geometry in order to allow the binding of the ligand. Dummy atoms on ZN1 were added using the parameters for tetrahedral zinc ion developed by Pang [34], whereas dummy atoms on ZN2 were parametrized following Liao *et al.* model [118] and adapting it to the trigonal bipyramidal geometry (Table S7.1). As suggested in the original model, the histidine residues that coordinate the ions were

considered deprotonated [293]. The final system is shown in Figure 5.2.

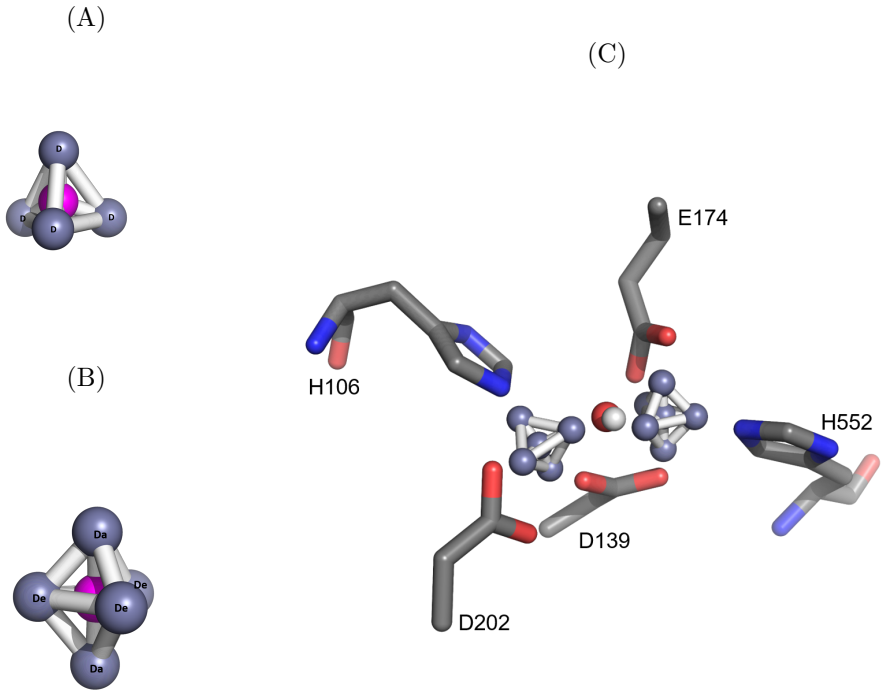


Figure 5.2: Cationic dummy atom model in the carnosinase system. (A) ZN1 and (B) ZN2 are shown. The purple sphere is the zinc ion, while the gray spheres are the dummy atoms. In ZN2, Da and De nomenclature is used to point at the axial and equatorial position respectively. (C) All the coordination sites of ZN1 are occupied, while ZN2 has one free coordination site for the ligand. A hydroxide ion is placed between the two zinc ions. Protein coordinating residues are shown.

5.3.4 Single-point energy calculation

Given the presence of the metal ions, the binding affinity was evaluated by performing a single point DFT calculation using Gaussian16 (Revison: A.03) [224] and focusing on a small system representing the binding site and the ligand (Figure S7.2). Polarizable Continuum Model was used to treat the implicit solvent [231]. B3LYP hybrid functional [225] with 6-31G(d) basis set was used, as commonly applied for calculations that involves zinc-enzymes [294][295][296].

5.3.5 MM-GBSA free energy calculation

Binding free energy was computed using the MM-GBSA methods [127]. The Python program MM-PBSA.py [247] implemented in Amber18 was used, considering an external dielectric constant of 78.3 and an ionic concentration of 0.15 M. Generalized Born (GB) mbondi2 model was applied [248], with a surface tension value of 0.0072 kcal/mol/A^2, and a maximum distance between atom pairs of 25.0 Å. The calculation was performed for all the MD frame (5000 frame in 100 ns), and the average value (kcal/mol) was returned.

5.4 Results

5.4.1 Molecular explanation of serum hydrolysis rate

Computational studies have been addressed to explain the experimental results reported in Table 5.1. Figure 5.3 compares the docking poses computed for carnosine (in its ϵ tautomer), anserine, balenine and homocarnosine (in its ϵ tautomer) which can be seen as representative of good, intermediate and poor CN1 substrates.

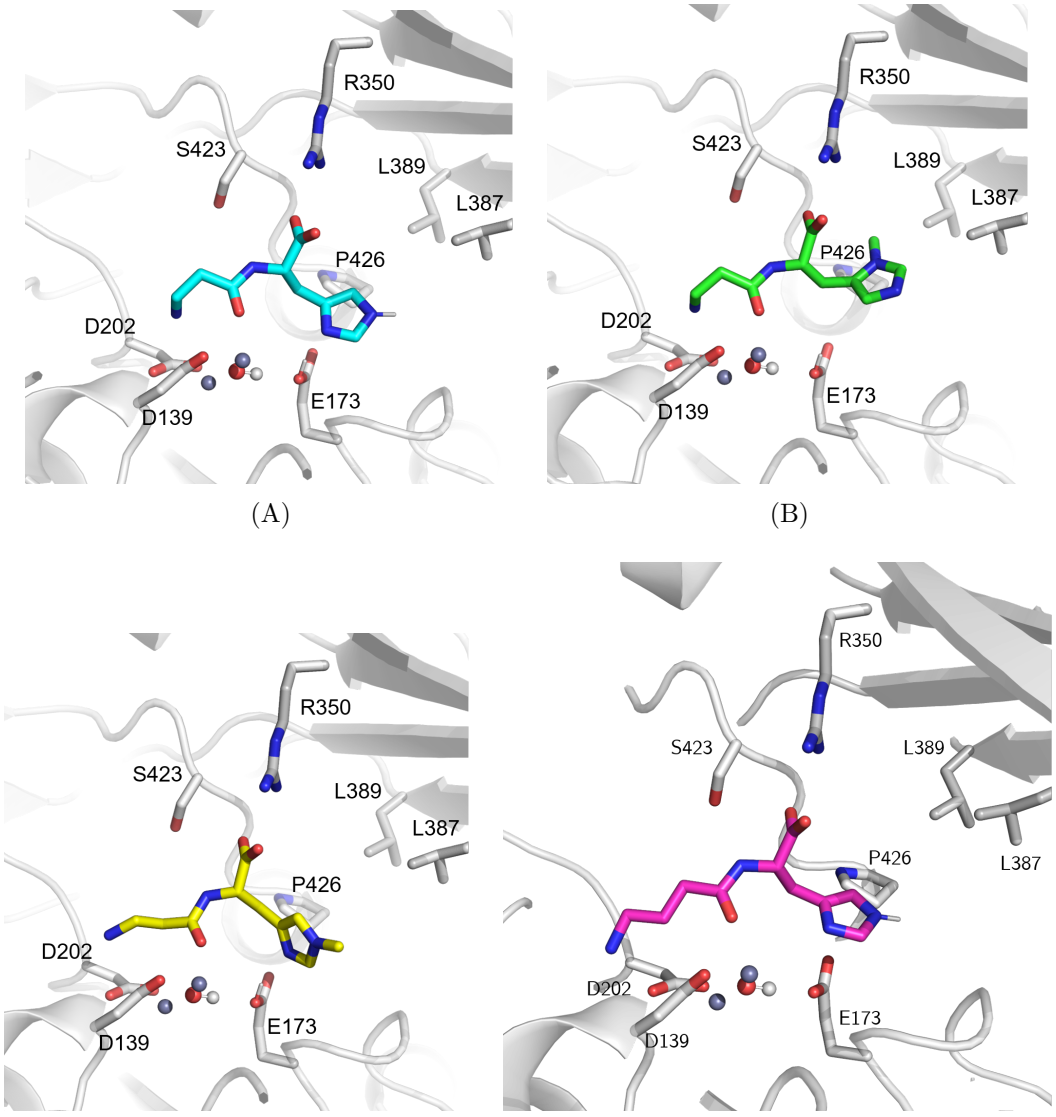


Figure 5.3: Comparison between the putative binding mode of (A) carnosine ϵ, (B) anserine, (C) balenine and (D) homocarnosine ϵ, coloured in cyan, green, yellow and purple respectively. The zinc ions are represented as silver spheres, while the hydroxide ion in ball and stick.

Carnosine assumes a pose in line with previous studies in which the carboxylate is engaged in a pivotal ion pair with Arg350 while the protonated amino group contacts Asp139 and Asp202 [93][284]. As expected, the oxygen atom of the carbonyl group is directed towards a zinc ion, while the NH group of the labile amide function approaches Ser423. Finally, the imadozole ring is inserted in a rather apolar subpocket where the Nδ atom elicits an H-bond with the neutral Glu173. This last interaction can explain the better interaction energy as computed for this tautomer compared to δ tautomer (see Table 5.2).

	E (kcal/mol)	Heavy Atoms	Enorm (kcal/mol)
Carnosine ϵ	-7281.01	16	-455.06
Carnosine δ	-7280.97	16	-455.06
Anserine	-7319.92	17	-430.58
Homocarnosine ϵ	-7206.70	17	-423.92
Homocarnosine δ	-7206.66	17	-423.92
Balenine	-7054.38	17	-414.96

Table 5.2: DFT-based interaction energies computed by single-point calculation of the minimized docking poses, using B3LYP/6-31G(d). Energy values normalized by the number of heavy atoms are also reported.

Figure 5.3B reveals that anserine assumes a pose very similar to that of carnosine apart from the imidazole ring, which exhibits a slightly shifted arrangement by which it loses the above described H-bond with Glu173 and inserts the N-methyl group into a hydrophobic niche lined by Leu387, Leu389 and Pro426. Figure 5.3C shows the putative complex of balenine which elicits the expected contacts concerning the charged termini and the labile amide, while the imidazole ring assumes a further shifted arrangement by which it retains the H-bond with Glu173 but addresses the N-methyl group towards the water molecules. Finally, the computed complexes for both homocarnosine tautomers (Figure 5.3D) confirm the detrimental role played by the modifications on the N-terminal residue since the extended γ amino butyric acid residue is unable to stabilize most of the polar contacts which characterize the carnosine complexes.

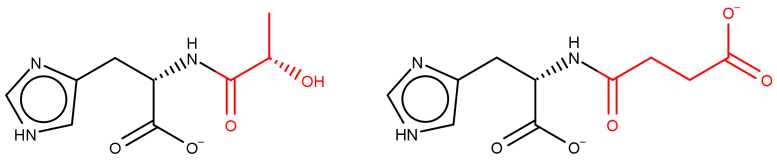
5.4.2 Identification of new hit compounds

Dataset collection. Table 5.3 reports the five databases used for the screening.

Database	Number of molecules
Known drugs[a]	5770
Hydroxyl dipeptides	11
Lactoyl amminoacids	14
Succinyl amminoacids	14
Natural compounds[b]	5404

Table 5.3: Datasets used for the virtual screening campaigns. The number of molecules contained in each dataset is reported. [a]The known drugs dataset is an in-house built library of drugs retrieved from DrugBank [297]. [b]The natural compound library was retrieved from the ChemDiv Natural-Product-Based Library (https://www.chemdiv.com/natural-product-based-library/).

Drug reporposing is a popular and effective strategy to find novel therapeutic applications for old molecules [298][299][300][301]. For this reason, an in-house made dataset of known-drugs (containing 5770 molecules extracted from DrugBank [297]) was included in the screening. The hydroxyl dipeptides dataset contains different combination of histidine- and tryptophan-based dipeptides, with a hydroxyl group in α position to the amide carbon, creating a new chiral center (Figure S7.3). This choice was made to mimic the carnostatine (Figure S7.3A) [280] and, to an extent, the intermediate of the hydrolytic reaction [93]. The datasets of lactoyl and succinyl amminoacids contained a set of amminoacids (i.e. Asp, Ans, Cys, Gln, Glu, His, Lys, Pro, Ser, Thr, Trp, Tyr, hydroxy-Lys and hydroxy-Pro) bound respectively with a lactoyl- [302] and succinyl-moiety (Figure 5.4).



(A) (B)

Figure 5.4: Example of molecules contained in the (A) lactoyl and (B) succinyl amminoacid library. Histine is used as an example. Lactoyl and succinyl moieties are highlighted in red.

As natural compounds have become more and more popular in drug discovery [303][304], a natural compound library was also collected. The molecules (5404) where retrieved from the Natural-Product-Based Library of ChemDiv (https://www.chemdiv.com/natural-product-based-library/).
For every dataset, molecules where ionized according to physiological pH and minimized using CGenFF force field [25] and the platform VEGA ZZ [174].

Dataset screening. The five datasets were docked into CN1 following the

procedure described in the **Methods** section. In particular, two steps of screening were performed: first a docking run with the generation of a single docking pose. Then the top five molecules for each dataset were chosen according to the metal interaction and the ChemPLP score normalized by the number of heavy atoms [305] and docked again producing 100 poses. Table 5.4 reports the best complexes for each dataset with the corresponding docking score.

Molecule	ChemPLP/HA (kcal/mol)	Dataset
N-hexanoyl-L-homoserine	3.35	Known drugs
N-2-Succinylarginine	2.38	
N-Succinyl Methionine	2.47	
N-2-Succinylornithine	2.92	
N2-(Carboxyethyl)-L-Arginine	2.61	
Hydroxy-carnosine-Trp	3.03	Hydroxyl dipeptides
Hydroxy-His-Thr	3.23	
Hydroxy-His-Asn	2.91	
Hydroxy-His-Tyr	2.74	
Hydroxy-His-3-hydroxy tetrahydrofuran	2.87	
Lactoyl-His	3.00	Lactoyl amminoacids
Lactoyl-Asn	3.15	
Lactoyl-Hydroxy-Lys	3.40	
Lactoyl-Pro	2.96	
Lactoyl-Tyr	2.86	
Succinyl-His	2.84	Succinyl amminoacids
Succinyl-Lys	3.28	
Succinyl-Asn	3.07	
Succinyl-Hydroxy-Lys	3.41	
Succinyl-Tyr	2.98	
Agaridoxin	3.09	Natural compounds
Anguibactin	3.06	
Bacilysin	3.05	
1-[5-(hydroxymethyl)pyrrolidin-3-yl]-3,3-dimethylurea	3.34	
N-cyclopropyl-4-(pyridin-3-yloxy)pyrrolidine-2-carboxamide	3.03	

Table 5.4: Best molecules obtained from docking. Name, score and dataset are indicated for each molecule. The docking score (ChemPLP) is normalized by the number of heavy atoms (HA).

Assessment of binding pose stability. These docking complexes underwent to a 100 ns MD simulation, to asses the stability of the docking poses. MD simulations were performed also on carnosine and its derivatives as comparison and control.

RMSD fluctuations of carnosine and its derivatives assessed around 1 Å, while the backbone's around 2-4 Å with the exception of the homocarnosine complex (Figure 5.5A-B). In this last case, the backbone RMSD reached the value of 6 Å, suggesting that the increase of the ligand dimension can affect the binding pocket morphology. As seen in the docking poses (see above), Arg350 elicited crucial interaction with the ligand. In Figure 5.5C the distance between Arg350 and the carboxyl moiety of the ligand is reported. It worth noticing that carnosine and anserine maintained a shorter distance to Arg350 compared to balenine and homocarnosine. Such an interaction could be one of the reasons for the differences in the hydrolytic rate.

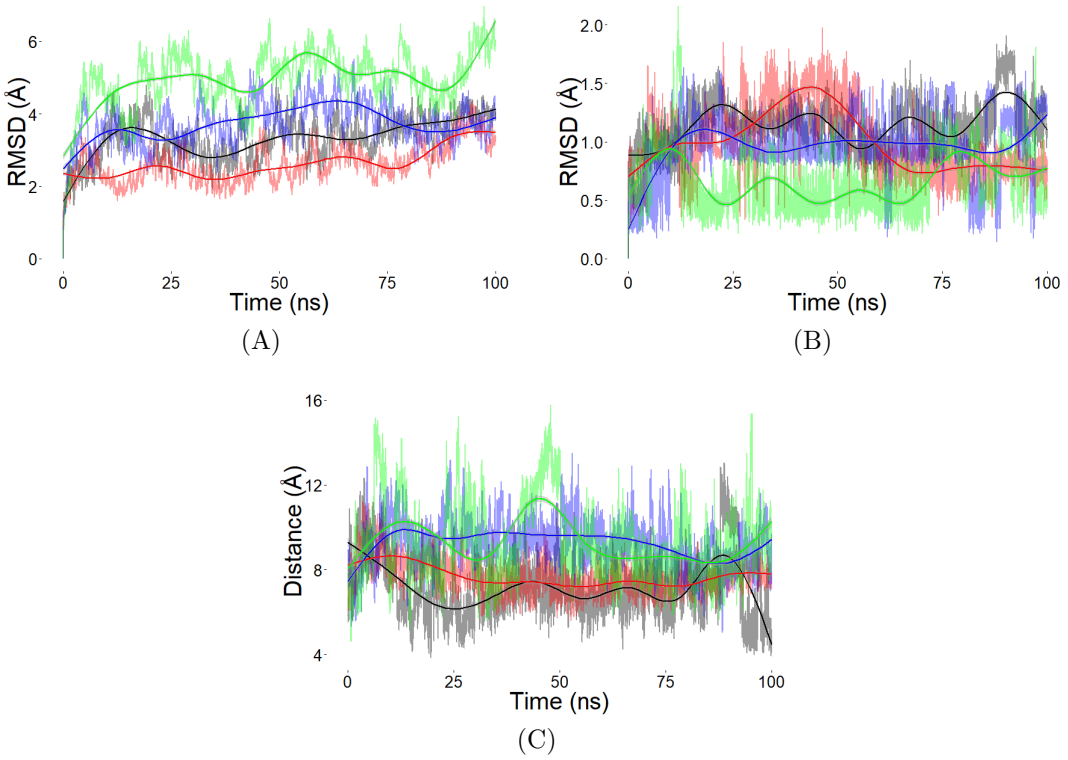


Figure 5.5: Results from the MD simulation of carnosine and its derivatives. RMSD values (Å) of the (A) protein backbone and of the (B) ligand along the trajectory. (C) Distance between Arg350 and the carboxyl group of the ligand along the trajectory. Carnosine is reported in black, anserine in red, balenine in blue and homocarnosine in green. Lines of every single value recorded during the trajectory are reported in background, and smooth lines showing the trend are reported in foreground.

After the production run, 16 molecules out of 25 maintained the coordination with ZN2, within the binding site of the enzyme. In order to further filter the best hit compounds, these 16 molecules underwent to a free energy calculation.

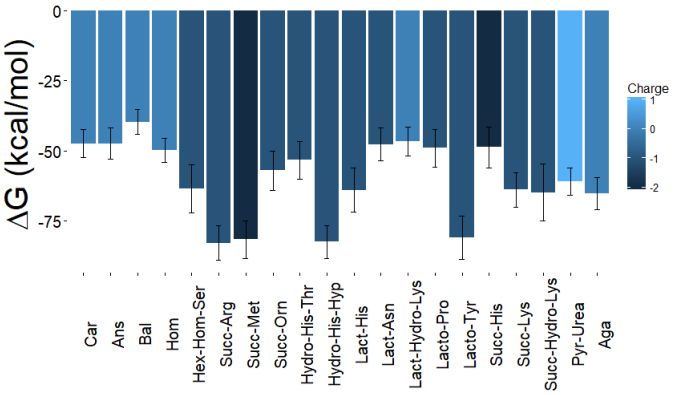
Evaluation of the free energy of binding. MM-GBSA was computed on the 100 ns trajectories stripped every 5 frames (1000 frames overall), to reduce computational time. The free energy of binding was calculated for the 16 molecules

selected from the MD simulation, and for carnosine and its analogues as comparison. Results are reported in Table 5.5. Free energy of binding of carnosine and its analogues partially agree with the experimental results: homocarnosine seems to bind the enzyme stronger than carnosine, which has also a similar binding energy to anserine. MM-GBSA is not enough accurate to perceive the slight differences between these molecules, so it is reasonable considering the mean value of the four ligands (i.e. -46.06 kcal/mol) as a reference value to identify other binders.

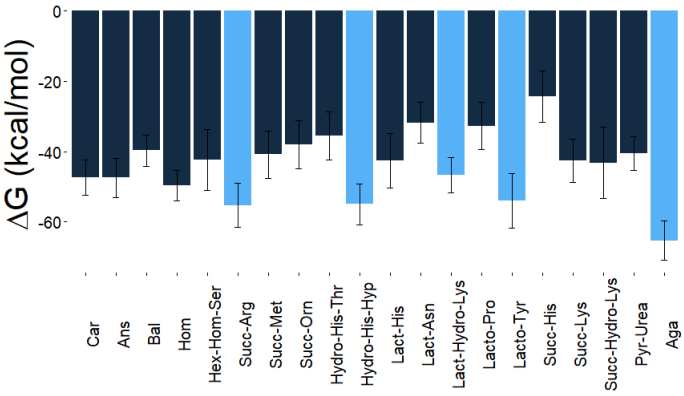
Ligand	Dataset	ΔG (kcal/mol)	Net charge	ΔG norm
Carnosine		-47.43	0	-47.43
Anserine		-47.45	0	-47.45
Balenine		-39.68	0	-39.68
Homocarnosine		-49.69	0	-49.69
N-hexanoyl-L-homoserine	Known drugs	-63.55	-1	-42.37
N-2-Succinyl arginine		-82.88	-1	-55.25
N-Succinyl methionine		-81.70	-2	-40.85
N-2-Succinyl ornithine		-57.06	-1	-38.04
Hydroxy-His-Thr	Hydroxyl dipeptides	-53.34	-1	-35.56
Hydroxy-His-3-hydroxy tetrahydrofuran		-82.46	-1	-54.97
Lactoyl-His	Lactoyl amminoacids	-63.98	-1	-42.65
Lactoyl-Asn		-47.66	-1	-31.77
Lactoyl-Hydroxy-Lys		-46.69	0	-46.69
Lactoyl-Pro		-49.05	-1	-32.70
Lactoyl-Tyr		-80.92	-1	-53.95
Succinyl-His	Succinyl amminoacids	-48.74	-2	-24.37
Succinyl-Lys		-63.92	-1	-42.61
Succinyl-Hydroxy-Lys		-64.85	-1	-43.23
1-[5-(hydroxymethyl)pyrrolidin-3-yl]-3,3-dimethylurea	Natural compounds	-60.93	1	-40.62
Agaridoxin		-65.26	0	-65.26

Table 5.5: Free energy of binding computed by MM-GBSA. Carnosine and its analogues are reported as comparison. ΔG values (kcal/mol) are normalized by the net charge of the molecule to overcome the bias derived from charged molecules [127]. In detail, ΔG values are divided by 1.5 when the net charge is 1 or -1, and by 2 when the net charge is -2. For the 16 molecules the dataset of origin is reported.

The majority of the selected molecules have a net charge different from zero. Sun *et al.* reported that MM-GBSA accuracy became worse with the increase of the ligand formal charge [127]. In order to reduce the charge-derived bias, ΔG values (kcal/mol) were normalized by dividing by 1.5 and 2 when the formal charge of the molecule was ±1 and -2 respectively. Considering the charge-normalized free energies, five molecules reached the reference value of -46.06 kcal/mol: N-2-Succinyl arginine, Hydroxy-His-3-hydroxy tetrahydrofuran, Lactoyl-Hydroxy-Lys, Lactoyl-Tyr, agaridoxin (Figure 5.6).



(A)

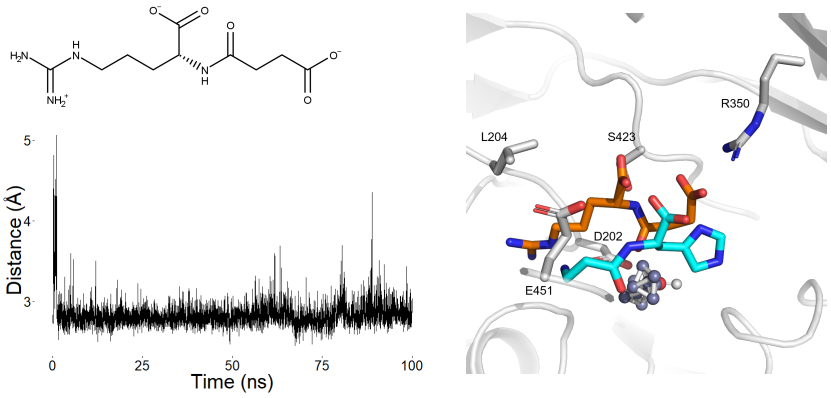


(B)

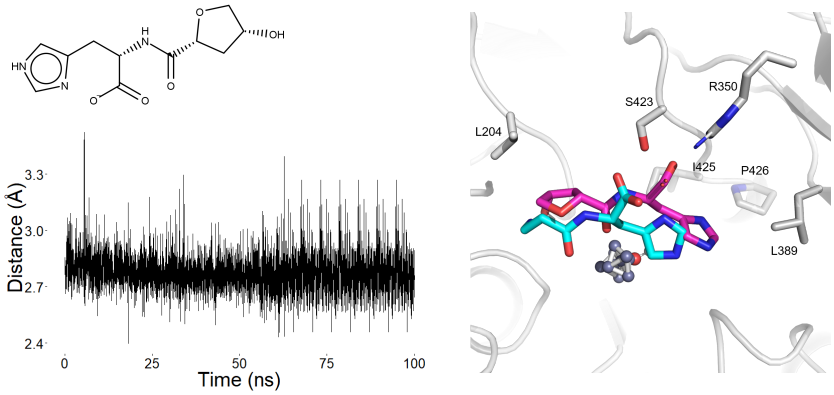
Figure 5.6: Graphical representation of the free energy calculation results. (A) ΔG values for each compound, bars are coloured based on the formal charge of the molecules. (B) ΔG values normalized on the formal charge, the best five compounds are highlighted in light blue. *X*-axis labels refer respectively to: carnosine, anserine, balenine, homocarnosine, N-hexanoyl-L-homoserine, N-2-Succinyl arginine, N-Succinyl methionine, N-2-Succinyl ornithine, Hydroxy-His-Thr, Hydroxy-His-3-hydroxy tetrahydrofuran, Lactoyl-His, Lactoyl-Ans, Lactoyl-Hydroxy-Lys, Lactoyl-Pro, Lactoyl-Tyr, Succinyl-His, Succinyl-Lys, Succinyl-Hydroxy-Lys, 1-[5-(hydroxymethyl)pyrrolidin-3-yl]-3,3-dimethylurea and agaridoxin.

The trajectory of these five compounds were clustered according to the backbone

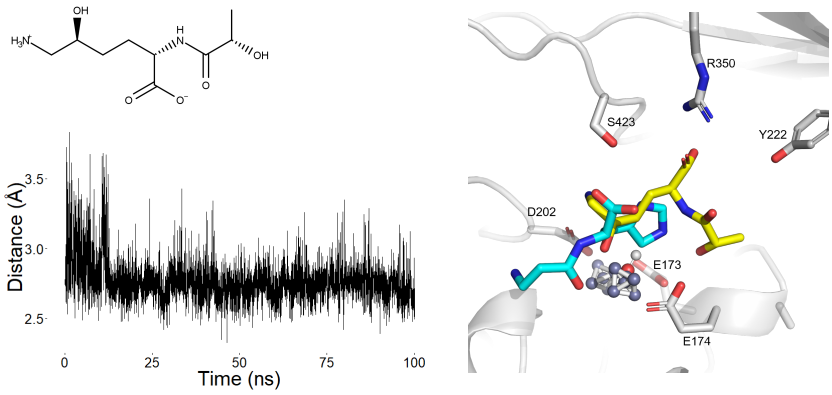
RMSD (1 Å was considered as a threshold value) and the most populated was compared with the one of carnosine (Figure 5.7).



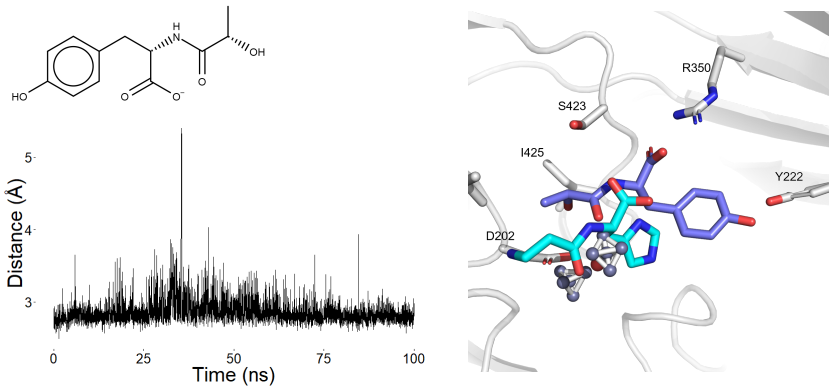
(A)



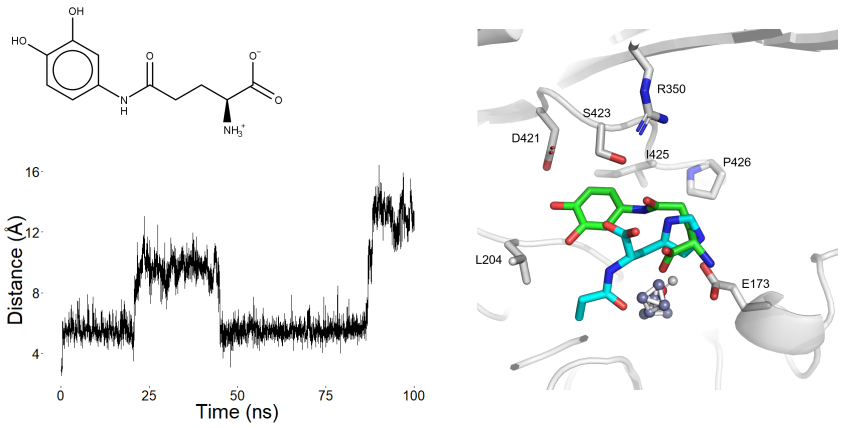
(B)



(C)



(D)



(E)

Figure 5.7: Results from the MD simulation of the five best compounds, according to the binding free energy. The 2D structure and distance fluctuation (Å) between Arg350 and the oxygen atom of the ligand along the trajectory are reported, together with the superimposition of carnosine (cyan) and the ligand within the enzyme. The structures were extracted through clusterization by backbone RMSD. (A) N-2-Succinyl arginine in orange, (B) Hydroxy-His-3-hydroxy tetrahydrofuran in purple, (C) Lactoyl-Hydroxy-Lys in yellow, (D) Lactoyl-Tyr in violet, (E) agaridoxin in green. Residues interacting with the ligand are also reported. The zinc ions are represented with their dummy atoms.

The five best compounds according to the binding free energy elicited a stable interaction with Arg350, as seen for carnosine and its analogues. Moreover, H-bonds are frequently present with Tyr222 and Ser423.

5.5 Discussion

In absence of experimental resolved structures, very little information are available on the molecular basis of the interaction between CN1 and its substrates. However different computational studies are in agreement, proposing one common inter-

action model [93][278], which was observed also in this work (Figure 5.3). Assuming that a higher hydrolytic rate corresponds to higher affinity, the monitored susceptibility of the dipeptides (i.e. carnosine and its analogues) to the CN1 activity is in clear agreement with the DFT-based interaction energies as compiled by Table 5.2. In particular, the comparison of the docking results for balenine and anserine emphasizes the crucial role of the hydrophobic interactions which only anserine is able to stabilize and which can counterbalance the missed H-bond with Glu173, the role of which appears to be rather marginal compared to that of the other detected polar interactions. These results are in agreement with those reported in previous studies [306] and emphasize that even simple modifications on N-terminal residue are not permitted, while the subpocket which harbors the histidine moiety is more tolerating especially for modifications able to reinforce the hydrophobic interactions as seen for anserine.

Using carnosine and its analogues as reference, a screening to identify novel CN1 binders was performed. Five databases were collected following different strategies: repurposing and natural compounds are popular source of hits [299][304], while hydroxyl dipeptides, as well as succinyl and lactoyl amminoacids, could resemble the intermediate of reaction, which is another promising approach in drug design [307]. From the screening of these databases, 25 molecules were selected.

Docking is able to rapidly screen a huge amount of molecules, however it is important to further validate its results [308]. Therefore 100 ns MD simulations were performed to test the stability of the docking poses and, as a matter of fact, nine molecules moved away from the binding site, even if one dummy atom was available for the interaction.

A further filter was represented by the free energy calculation using MM-GBSA. According to the results reported in Table 5.5, this method is not enough accurate to correctly rank carnosine and the other dipeptides, as opposed to the DFT single point calculation. However, a reference value of energy was identified and used as comparison for the other 16 molecules. Although MM-GBSA method allows to compute the free energy of binding along a trajectory in a short amount of time, it suffers from formal charge bias [127]. To reduce such a bias and the charge effect, a normalization based on the net charge of the molecules was performed. Following these results, five hit compounds were identified (Figure 5.6B).

The hit compounds elicited similar binding mode compared to carnosine, where the interaction with Arg350 seems to play a key role (Figure 5.5 and 5.7). In particular, N-2-succinyl arginine, hydroxy-His-3-hydroxy tetrahydrofuran and lactoyl-Tyr mimic the carnosine binding as seen in the superimposition (Figure 5.7A, B, D).

5.6 Conclusions

Carnosine exploits many beneficial effects, however its fast hydrolysis prevents the use of this dipeptide for therapeutic applications. Inhibition of CN1 can increase the availability of carnosine and, as a consequence, can lead to its use in clinics.
This work proposes five new hit compounds for the inhibition of CN1, selected from five different databases. The pipeline included docking, MD simulation and free energy calculation. DFT-based single point energy calculations show the importance of using high level methods when metal ions are included in the system, and it was able to detect binding energy differences even upon small changes in molecule structures (i.e. between anserine and balenine).
A further evidence of the need to validate docking poses arises in this study, where 9 molecules out of 25 moved away from the binding site after MD simulation. Nonetheless, the cationic dummy atom approach maintained the correct coordination geometry along the trajectories and, to the best of our knowledge, a new parametrization of 5-coordinated zinc ion was presented here for the first time.
The five selected compounds interact with the enzyme similarly to carnosine, and the binding energy equals or exceeds the one of carnosine, making them promising lead compounds.

5.7 Supplementary Information

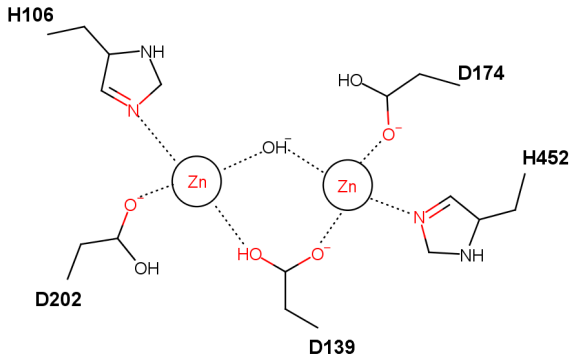


Figure S7.1: Graphical representation of the distance restraints (30 kcal/mol·Å) applied between the zinc ions and the coordinating atoms of the protein, coloured in red. The coordination bond is represented as a dot line.

Bonded parameters				
bond type[a]	K_b (kcal/mol·Å^2)	b_0 (Å)		
Da-Zn	640.0	0.92		
De-Zn	640.0	0.82		
Da-De	640.0	1.23		
De-De	640.0	1.42		
angle type[b]	K_θ (kcal/mol·rad^2)	θ (°)		
Da-Zn-Da	55.0	180.00		
De-Zn-De	55.0	120.00		
Da-Zn-De	55.0	90.00		
De-De-De	55.0	60.00		
De-Da-De	55.0	70.37		
Da-De-Da	55.0	96.58		
Da-De-Zn	55.0	48.29		
Da-De-De	55.0	54.81		
Zn-Da-De	55.0	41.71		
Zn-De-De	55.0	30.00		
diehedral type[c]	phase shift angle (°)	periodicity		
Zn-Da-De-De	322.28	2.00		
Zn-Da-De-Da	359.97	2.00		
Zn-De-De-De	0.03	2.00		
Zn-De-De-Da	65.98	2.00		
Zn-De-Da-De	322.28	2.00		
Da-Zn-Da-De	219.14	2.00		
Da-Zn-De-De	90.00	2.00		
Da-Zn-De-Da	180.00	2.00		
De-Zn-De-De	180.00	2.00		
De-Zn-De-Da	90.00	2.00		
De-Zn-Da-De	240.00	2.00		
Da-De-Da-De	37.72	2.00		
Da-De-De-De	294.02	2.00		
Da-De-De-Da	131.96	2.00		
De-Da-De-De	75.44	2.00		
De-De-Da-De	75.44	2.00		
Non-bonded parameters[d]				
atom type	mass	charge	$R_{min}/2$ (Å)	ϵ (kcal/mol)
Da	3.0	0.5	0	0
De	3.0	0.5	0	0
Zn	50.380	-0.5	1.0900	0.250

Table S7.1: Force field parameters for the trigonal bipyramidal dummy model presented in this work. [a] $U_b = K_b(b - b_0)^2$, where K_b is the force constant and b_0 is the equilibrium bond length. [b] $U_\theta = K_\theta(\theta - \theta_0)^2$, where K_θ is the force constant and θ_0 is the equilibrium angle. [c] Phase shift angle in the torsional function (in degrees) with the periodicity of the torsional barrier reported (for more information see Weiner *et al.* [309]). [d] $R_{min}/2$ is the the van der Waals radius and ϵ the 6-12 well depth potential. Zn denotes the central zinc ion, while Da and De denote axial and equatorial dummy atoms respectively (Figure 5.3).

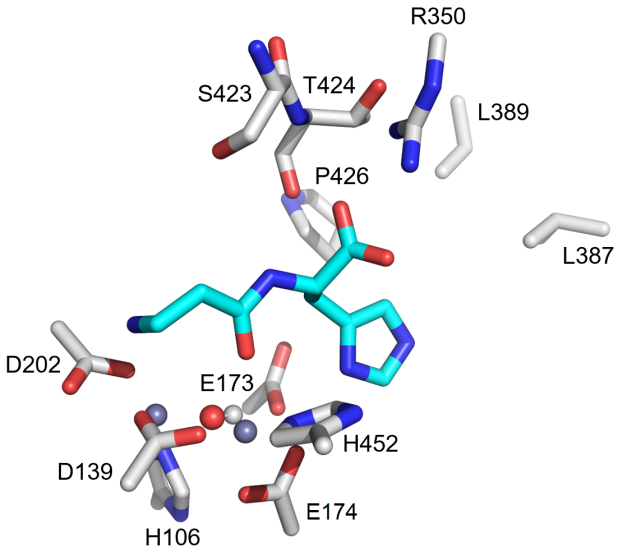


Figure S7.2: Sub-system of CN1-ligand complex used in the single point energy calculation. For the QM calculation we considered the zinc ions (silver sphere), the hydroxide ion (ball and stick), the side chain of the binding pocket residues and the ligand (cyan). Only heavy atoms are shown for clarity.

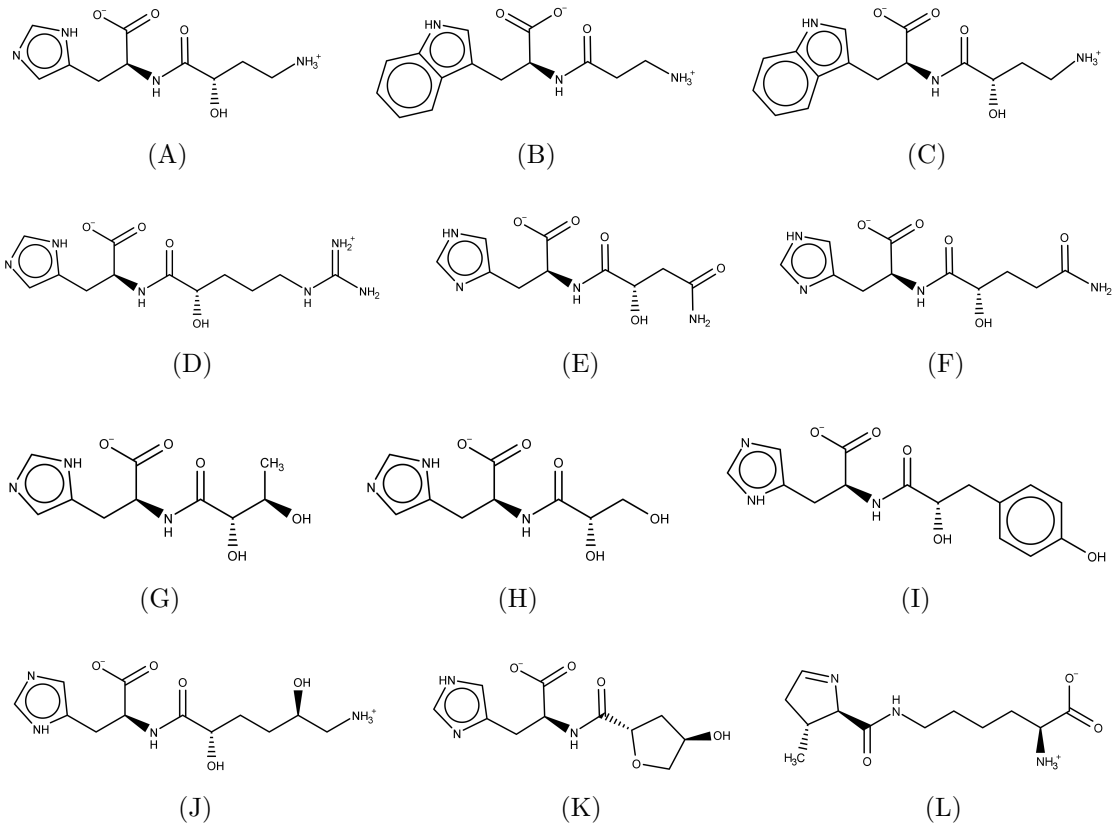


Figure S7.3: Hydroxyl dipeptides dataset. (A) Carnostatine is shown as comparison. (B) Carnosine and (C) carnostatine with substitution of the imidazole ring were also included. (D) His-Arg, (E) His-Asn, (F) His-Gln, (G) His-Thr, (H) His-Ser, (I) His-Tyr dipeptides with the addition of an hydroxyl group at the α position of the amide carbon. Histidine bound with (J) hydroxy-lysine and (K) 3-hydroxy tetrahydrofuran were included in the dataset, together with (L) pyrrolysine.

Chapter 6

A multiscale approach to predict the binding mode of metallo β-lactamase inhibitors

6.1 Antibiotic resistance and β-lactamases

The spread of antimicrobial resistance threatens our capability to treat infectious diseases [310]. Microorganisms have developed several resistance mechanisms including efflux pump production, target alteration, decreased drug uptake and drug inactivation [311]. The latter mechanism is one of the most important and is often seen in the form of the hydrolysis of β-lactam antibiotics, catalysed by serine-β-lactamases (SBLs) and metallo-β-lactamases (MBLs) [312][313]. One effective strategy to overcome this resistance mechanism is to inhibit β-lactamases [314][315][316][317]. However, while for the SBLs several compounds of different classes have already been approved for clinical use, no MBL inhibitor has yet been approved [316].

MBLs can be divided into three subclasses: B1, B2 and B3, with B1 emerging as the most clinically significant because they catalyse the hydrolysis of a broad spectrum of β-lactams [318] and several family members, notably the NDM, IMP and VIM enzymes, are disseminating among Gram-negative bacterial pathogens (e.g.

Escherichia coli, *Pseudomonas aeruginosa*) on mobile genetic elements [317][319]. MBLs contain two catalytic zinc ions within the active site which participate in the reaction mechanism by coordinating the substrate and activating a bound water molecule to undergo nucleophilic attack upon the carbonyl carbon of the scissile amide [317][320][321][322] (Figure 6.1). Thus, efforts have been made to design inhibitors that coordinate one or both the zinc ions and compete with substrate for binding to the active site [320], which could potentially be co-administered with β-lactams to restore their efficacy against MBL-producing strains.

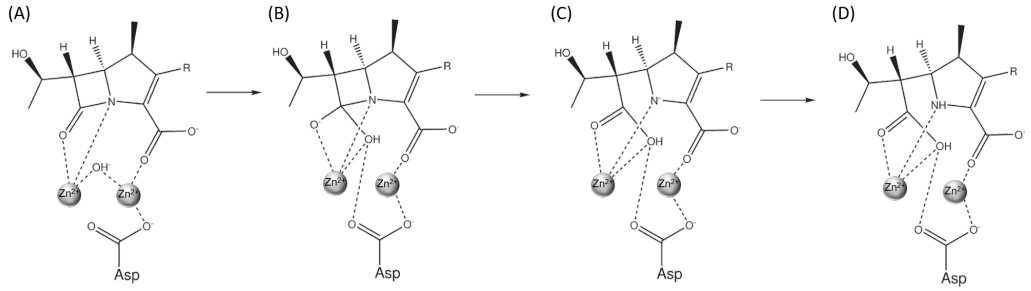


Figure 6.1: Possible mechanism of β-lactam ring hydrolysis by subclass B1 metallo-β-lactamases [323]. (A) Catalysis initiates with coordination of the zinc ions by the β-lactam (here drawn as a carbapenem), with a hydroxide ion bridging the two metal centers. (B) The carbonyl carbon of the β-lactam ring undergoes nucleophilic attack by the hydroxide ion, leading to formation of a transient tetrahedral species; (C) followed by cleavage of the C-N bond that leads to the accumulation of an anionic intermediate (D) that may be resolved by protonation of the nitrogen.

In silico drug discovery to identify new MBL inhibitors is more challenging than for SBLs because of the presence of zinc ions. Zinc ions possess d orbitals that participate in chemical bonding and, compared to s or p orbitals, have more complicated shapes. Furthermore, zinc ions can have multiple coordination numbers and different coordination geometries [51]. In molecular mechanics (MM) approaches, metal ions are often represented as point charges with van der Waals interactions [24][324]. This simplistic representation cannot reproduce all the properties of the zinc ion described above, and in consequence fails to successfully model the dynamic and versatile nature of zinc coordination in many biological systems [31][325][326]. Over the years there has been significant effort to overcome these limitations by introducing alternative MM models to treat metal ions, including bonded models, non-bonded models, ligand field molecular mechanics, cationic dummy atoms or polarizable force fields. Bonded models add bonding potentials between the metal and ligands and can also include angles, torsions, electrostatic and van der Waals terms; drawbacks include the inability to model ligand exchange and dissociation, or changes in coordination [326]. The ligand field molecular mechanics model incorporates the ligand field stabilization energy directly into the potential energy equation

of molecular mechanics [32]. The cationic dummy atom model places additional virtual atoms around the ion to mimic a covalent bond [117]. Polarizable force fields include terms to account for the polarization effect, but so far are computationally expensive and not fully developed [37]. Finally, in typical MM non-bonded approaches, interactions are evaluated by the sum of the Coulomb and Lennard-Jones terms (Equation 6.1) [112]. However, this type of model tends to underestimate the strength of the coordination, and does not represent coordination geometry preferences as well as other electronic effects [112]. For this reason, other non-bonded interaction models have been proposed, such as the restrained non-bonded model or the LJ12-6-4 model (Equation 6.2), which was tested mainly for Mg^{2+} [327][328] but also for Zn^{2+} ions [329][115] and contains an r^{-4} term that is included to describe the ion-induced dipole interaction [116].

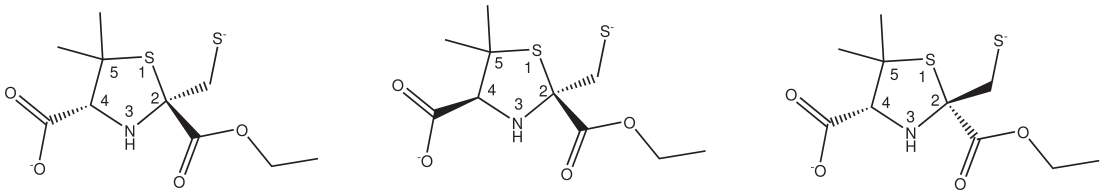
$$U_{ij}(r_{ij}) = \frac{e^2 Q_i Q_j}{r_{ij}} + \frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \tag{6.1}$$

$$U_{ij}(r_{ij}) = \frac{e^2 Q_i Q_j}{r_{ij}} - \frac{C_4^{ij}}{r_{ij}^4} + \frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \tag{6.2}$$

In this chapter, we investigate the binding of inhibitors to the imipenemase (IMP)-1, a B1 MBL that is frequently found on mobile genetic elements that facilitate its spread through bacteria [319][330][331][332]. The primary host, *Pseudomonas aeruginosa*, is a Gram-negative bacterial pathogen responsible for a variety of opportunistic infections of immunocompromised patients, and notably associated with infections of burn wounds and the lungs of individuals with cystic fibrosis [333]. In the continued absence of new antibacterial agents with anti-Gram-negative activity, infections by resistant *P. aeruginosa* strains are very difficult to treat [334]. Various compound classes have been investigated as potential MBL inhibitors [313][316][318] [320][335]; we here focus on thiolate-based compounds which exploit the high affinity of zinc for sulfur ligands and have been extensively investigated [336][337]. In consequence, relatively extensive structural data exist for thiolate complexes of IMP-1 that may be exploited for development and assessment of approaches to model inhibitor binding [336][338].

Accordingly, we have developed a multiscale approach to investigate the binding of thiolates to IMP-1, in which successive docking, molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) simulations provide a stepwise refinement of complex models. This has been applied to study a newly characterised

series of thiazolidine inhibitors (Figure 6.2), first successfully replicating two known crystal structures of IMP-1 complexes with two different thiazolidine stereoisomers (structures not yet released) and subsequently predicting the mode of binding for a third form known to inhibit IMP-1 but for which the crystal structure of its complex with the protein could not be obtained. This approach could be applied to design other potential metallo-β-lactamase inhibitors to predict their binding mode and optimise their interactions with enzyme, aiding their development as potential pharmaceutical leads. Moreover, it could be applied also to model other similar metalloprotein systems: an example is represented by di-zinc metalloenzyme human carnosinase CN1, where Pavlin *et al.* [93] adopted a similar workflow (i.e. docking, MD and DFT/MM) to simulate the binding mode of the substrate carnosine.



(A) (B) (C)

Figure 6.2: Structures of thiazolidine inhibitors (not yet published). (A) THZ (2S, 4S), (B) VCZ (2S, 4R) and (C) RSZ (2R, 4S).

6.2 Methods

6.2.1 Molecular docking

Different docking procedures were tested for crystal structures of IMP-1 (see below) co-crystallized with thiolate-based inhibitors (Figure S6.1), in order to find the docking set-up that best reproduced their interactions. The region of the protein included in docking was selected by considering whole residues within a 10 Å radius around the two zinc ions of the IMP-1 active site, which comprises the whole catalytic pocket. Docking with PLANTS [339] (see below) applied distance restraints between the sulfur atom of the ligand and the zinc ions with a weight of - 7.5 kcal/mol, and a maximal and minimal distance of 3 and 1.30 Å, respectively. A σ value of 5 was applied to increase the number of search iterations of the ACO algorithm [67] and 10000 clusters were produced, using the ChemPLP scoring function and the speed option equal to 1. Furthermore, we tested both the flexible (designated Opt1) and

rigid (Opt2, Opt3) binding pocket and we also included the water molecules present in the binding site of the crystal structures (Opt3). In Opt1 the side chains of Val31, Phe51, Lys161, Leu165, Asn167, Asp170, Trp176 and Ser198 (which face the binding site) were treated as flexible. For docking with AutoDock Vina, we considered the binding pocket as rigid, and tested the procedure with and without crystallographic water molecules (Opt5 and Opt4 respectively). The results of docking experiments were evaluated by computing the RMSD values between binding poses identified by docking and the corresponding crystal structure.

6.2.2 Molecular dynamics simulations

Amber 18 [246] was used for MD simulations. Complexes derived from docking were solvated in a box of TIP3P water molecules, with 10 Å between the protein and the edges of the box. The systems were neutralized with Na^+ and Cl^- ions at an ionic concentration of 0.15 M, and periodic boundary conditions were applied. The ligands were prepared in Antechamber with AM1-BCC charges and GAFF as force field [109], while the protein was parameterized using the *ff14SB* force field [105]. Given the challenges associated with evaluating the interactions between the zinc ions and their coordinating atoms, we tested two different non-bonded models. First, the restrained non-bonded model (LJ12-6-R), which combine a non-bonded model (Equation 6.1) with distance restraints of 30 kcal/mol·Å between the two zinc ions and their coordinating protein atoms (Figure S6.2), and equilibrium distances as in the respective crystal structures (Table 6.1). Second, a 12-6-4 non-bonded model (LJ12-6-4) (Equation 6.2) in which an extra term is added to the Lennard-Jones equation [115] to represent the charge-induced dipole interaction.

Residue number	Residue name	Atom name	Zinc ion	Distance (Å)
139	His	NE2	ZN1	2.1
77	His	NE2	ZN1	2.1
79	His	ND1	ZN1	2.0
81	Asp	OD2	ZN2	2.0
197	His	NE2	ZN2	2.1
158	Cys	SG	ZN2	2.3

Table 6.1: Distance restraints of the LJ12-6-R model. Equilibrium distances of the restrained non-bonded model between the zinc ions and the protein coordinating atoms are reported in Å.

We applied the SHAKE algorithm [99] and a time step of 2 fs. The simulation

protocol included energy minimization of the hydrogen atoms followed by minimization of the water molecules, and finally of the whole complexes, using 100 steps of steepest descent and 2900 steps of the conjugate gradient algorithm. The system was then heated to 300 K using the Langevin thermostat for 20 ps. Equilibration was performed at a pressure of 1 atm with the Berendsen barostat for 240 ps. Finally, we computed the production run using the NVT ensemble over 100 ns.

6.2.3 QM/MM optimization

For QM/MM calculations we tested an approximate DFT-based method, SCC-DFTB (self-consistent charge-density functional tight binding, abbreviated here as DFTB, in particular DFTB3 that is a third order expansion of DFTB [340]) and a DFT method [140]. The latter originates from the Kohn–Sham formulation. SCC-DFTB is in turn derived from DFT by approximation and parameterization of multi-center electron integrals, resulting in a significant increase in speed of calculation at the expense of accuracy [341]. SCC-DFTB/MM geometry optimization was performed with DFTB3 for the QM region and the *ff14SB* force field for the MM region using Amber. DFT QM/MM optimization was carried out using ChemShell (www.chemshell.org), which combines MM and QM software: in detail, the MM portion of the system was handled by DL-Poly [342] using the Amber *ff14SB* force field, while Gaussian09 handled the QM portion, with electrostatic embedding. In both DFT/MM and DFTB3/MM optimizations, the QM region was defined as the zinc ions, the ligand, any water molecule coordinated to the zinc ions, the side chains of the zinc coordinating residues (His77, His79, Asp81, His139, Cys158, Lys161 and His197), and the side chain of the conserved residue Lys161 that interacts with the ligands (Figure S6.3). The QM region included the ligand and the residues in the nearest neighbourhood, and its net charge was +1. A charge state close to neutrality is generally desirable in the QM region to avoid unbalance charge interactions at the boundary of the two regions [140]. In the DFT optimization, a sphere of 8 Å around the QM region was considered as the active portion, while the remaining residues were fixed to decrease the computational cost. For the QM portion, we tested both the B3LYP and PBE0 hybrid functionals with the 6-31G(d) basis set and the DFT-D2 Grimme dispersion correction [343], which is important for predicting accurate structures [344]. We considered the zinc ions as closed shells. Covalent C-C bonds between the MM and the QM regions were capped with hydrogen atoms.

6.3 Results

6.3.1 Method validation

We envisaged a workflow for developing models of IMP complexes with thiolate inhibitors that involved successive steps (docking, MD, QM/MM) of higher levels of theory, to improve the accuracy of the final model. As described in Methods, for each step we tested multiple settings, in order to find the best combination to describe the structure and interactions with the metal ions.

We first tested docking options and scoring functions for thiolate-based inhibitors (Table 6.1). Molecular docking predicts the binding mode of a ligand to a target molecule, usually a receptor or an enzyme. The resulting poses are generated by the combination of the conformational sampling algorithm, which explores the conformers of the ligand, and the scoring function that ranks those poses. When metal ions are included in the docking region, the scoring function needs to address a further level of complexity that can lead to the failure of the prediction of the correct binding mode [345]. We tested two docking tools that use two different search algorithms: the ant colony optimization (ACO) and the genetic algorithm (GA), namely PLANTS [67] and AutoDock Vina [346], respectively. Moreover, these docking tools apply different scoring functions: ChemPLP from PLANTS derives from the combination of ChemScore and Piecewise Linear Potential (PLP) and includes a specific term to evaluate the coordination interactions [85], while the AutoDock Vina Energy was calculated from the AutoDock4 scoring function and does not include specific terms for metal coordination [346]. Both these tools were tested with different settings in order to determine the setup that most reliably reproduced the crystal structures of protein-ligand complexes.

Literature and database searches identified crystal structures of IMP complexes with seven different thiolates (Figure S6.1), to which we added a recently determined structures of IMP complexes with two different thiazolidine stereoisomers (structures not yet released). These nine thiolate-based inhibitors were redocked into the corresponding (ligand-free) IMP-1 crystal structures using either PLANTS or AutoDock Vina and alternative treatments for the protein binding pocket, with the aim of obtaining an average ligand RMSD to the experimental crystal structure of less than 2 Å. Table 6.2 shows that the use of flexible side chains in the binding pocket (Opt1) resulted in relatively high RMSD values. Indeed, for both PLANTS

and AutoDock Vina, lower RMSD values were generally obtained using a rigid biding pocket (Opt2, Opt4), and these further improved (for 6 out of 9 thiolate ligands) when PLANTS was used with crystallographic water molecules added to the binding pocket (Opt3). This finding highlights the important role of interactions involving the water network in the binding sites of ligands [347][348][349]. However, with one exception (PDB id: 2DOO [350]), improvement was generally not observed when AutoDock Vina was used with water molecules included. Comparing the average RMSD values we identified the Opt3 approach as providing the best results, with 6 out of 9 ligands reaching the target RMSD of less than 2 Å with respect to the crystal structure. Redocking of mercaptopropionamide to 2DOO was significantly the worse, probably due to the large size and flexibility of the ligand compared to the others tested here, but also to the low resolution of the crystal structure, moreover it performed poorly in all the docking setting indeed.

IMP-1 PDB Id	Resolution (Å)	Inhibitor	PLANTS Opt1[a]	PLANTS Opt2[b]	PLANTS Opt3[c]	AutoDock Vina Opt4[d]	AutoDock Vina Opt5[e]
VCZ	1.87	Thiazolidine (VCZ)	5.56	4.27	1.91	4.01	4.25
THZ	1.87	Thiazolidine (THZ)	4.74	3.49	4.57	4.25	4.32
1DD6 [351]	2.00	Mercaptocarboxylate	9.46	2.34	1.70	4.17	4.58
1VGN [352]	2.63	Ethanethiolate	8.41	2.94	2.94	5.24	5.18
2DOO [350]	2.43	Mercaptopropionamide	4.39	4.30	7.68	11.08	5.1
4C1F [338]	2.01	L-captopril	2.13	5.18	1.18	3.91	3.91
4C1G [338]	1.71	D-captopril	2.01	2.07	1.73	4.46	4.46
5EV8 [353]	2.30	Bisthiazolidine	2.13	1.09	1.63	3.52	3.52
5EWA [353]	2.30	Bisthiazolidine	1.37	1.23	1.29	2.94	2.95

Table 6.2: RMSD values of the ligands between the crystal structure and redocked pose. All the RMSD values are reported in Å. [a]Opt1: S-Zn distance restraints, flexible binding pocket. [b]Opt2: S-Zn distance restraints, rigid binding pocket. [c]Opt3: S-Zn distance restraints, rigid binding pocket, crystallographic water molecules. [d]Opt4: rigid binding pocket. [e]Opt5: rigid binding pocket, crystallographic water molecules.

The results of the best docking runs (Opt3) for the two thiazolidine compounds, for which the complex structures were available, were then subjected to molecular dynamics (MD) simulations to improve the quality of the models obtained. These were run using either the restrained non-bonded (LJ12-6-R) model or the LJ12-6-4 model; averaged RMSD values for the production run frames are reported in Table 6.3.

		MD model	THZ	VCZ
MD simulation	Binding site[a]	LJ12-6-R	1.93	1.78
		LJ12-6-4	1.20	1.62
	Ligand	LJ12-6-R	1.06	1.19
		LJ12-6-4	1.21	1.23
QM/MM optimization	Binding site[a]	LJ12-6-R	1.33	1.22
		LJ12-6-4	1.17	1.44
	Ligand	LJ12-6-R	0.65	1.36
		LJ12-6-4	1.09	1.34

Table 6.3: Average RMSD (Å) between the crystal structures of VCZ and THZ complexes. RMSD values of the MD simulated (upper panel) and B3LYP 6-31G(d)/MM optimized complexes (lower panel) are reported. [a]Binding site residues: Glu23, Val25, Trp28, Val31, Phe51, Ala53, His77, His79, Ser80, Asp81, His139, Thr140, Cys158, Lys161, Gly166, Asn167, Ser196, His197, HO^-, ZN1, ZN2.

In the two MD simulations, the average RMSD values (compared to crystal structures) of the binding site residues for the LJ12-6-4 model are lower than those for the LJ12-6-R, although in the latter the inhibitor RMSD is slightly smaller. In general, for both the treatments, the difference in coordination bond lengths compared to the crystal structure is small (Figure 6.3).

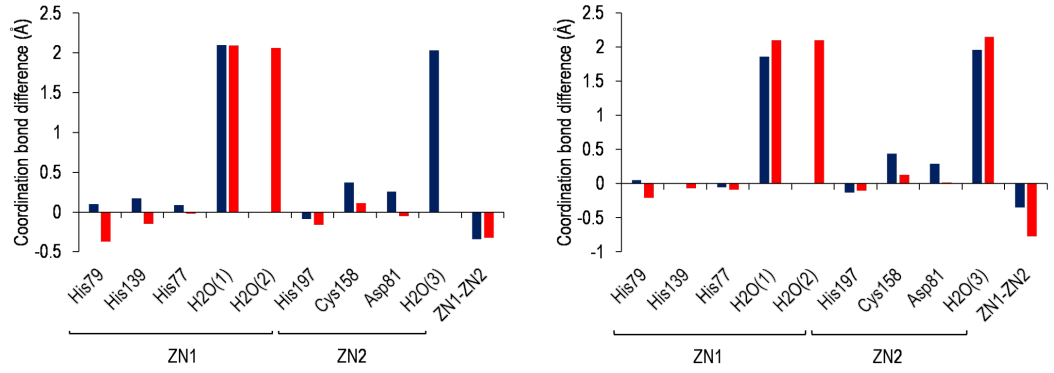


Figure 6.3: Comparison of coordination distances for crystal and MD-simulated structures (last frames) using the LJ12-6-R (blue) and LJ12-6-4 (red) models. Panel (A) refers to the THZ complex, panel (B) to the VCZ complex. Water molecules that coordinate the zinc ions are also reported.

In the crystal structures of both thiazolidine complexes, the two zinc ions are each in a tetrahedral geometry, with the first (ZN1) coordinated by three histidine residues (His77, His79, His139) and by the thiolate of the inhibitor, whilst the second (ZN2) is coordinated by Cys158, Asp81, His197 and by the inhibitor thiolate,

which, as in other complexes of IMP-1 with thiolate-based inhibitors, replaces the Zn-bridging hydroxide present in structures without ligands [354]. However, during the MD simulations using either models, the coordination number of the zinc ions tends to increase, through addition of extra water molecules, up to a limit of six. This tendency is enhanced when the LJ12-6-4 model is used. It worth noticing that generally the distance between the two zinc ions did not change significantly (i.e. the major difference was 0.77 Å, found in the VCZ LJ12-6-4 simulation).

Finally, we refined the complexes by QM/MM geometry optimization. In the refinement step we used two different methods: DFTB3 and DFT, and in the latter we tested two functionals (B3LYP and PBE0). The DFTB3/MM method was tested to establish whether it could improve structures from MM MD (e.g. whether it would decrease the RMSD and recover the tetrahedral geometry of the zinc ions that was observed in the crystal structures). Thus, the complexes underwent to a DFTB3/MM geometry optimization. We found none or slight improvement in RMSD values (a decrease of 0.43 Å RMSD of the binding site in THZ) (Table S6.1, Figure S6.4), but the geometry of the zinc ions was not recovered: they remained octahedrally coordinated (Figure S6.5), even if DFTB3 was parameterized and tested also considering tetrahedral Zn-protein [355]. Accordingly, DFTB3/MM was applied for other Zn tetrahedral complexes, but with further refinement with DFT [356].
Therefore, DFT methods were then employed to provide a more accurate description of the metal ions, with application of the hybrid functionals B3LYP and PBE0 which are commonly used in QM/MM studies of zinc-containing proteins [47][357][358]. Both the functionals gave similar results: there are only very small differences between the two calculations (Tables 6.3, Table S6.1). The results obtained with the PBE0/MM treatment are reported in the **Supplementary Information** section. Final RMSD values, with respect to the experimental crystal structures, are reported in Table 6.3. Compared to MD treatments, by using the QM/MM optimization the RMSD values decreased, in particular when starting from the last frame of the LJ12-6-R MD simulation. Moreover, in this case, the zinc coordination distances after DFT/MM optimization are almost identical to those measured in the crystal structure (Figure 6.4) and the additional Zn-coordinating water molecules that were introduced during MD simulations moved away from the zinc ion, with the result that the starting (crystallographic) coordination geometry was restored. In contrast, in DFT/MM calculations starting from the last frame of the MD simulation using the LJ12-6-4 model, the zinc ions remain in the octahedral geometry. The distance between the two zinc ions also changed significantly using the LJ12-6-4 model, and the QM/MM optimization was unable to relocate the two ions at

the correct separation distance. This demonstrates the need to apply restraints in MM MD simulations to maintain the original active site structure and metal ion coordination.

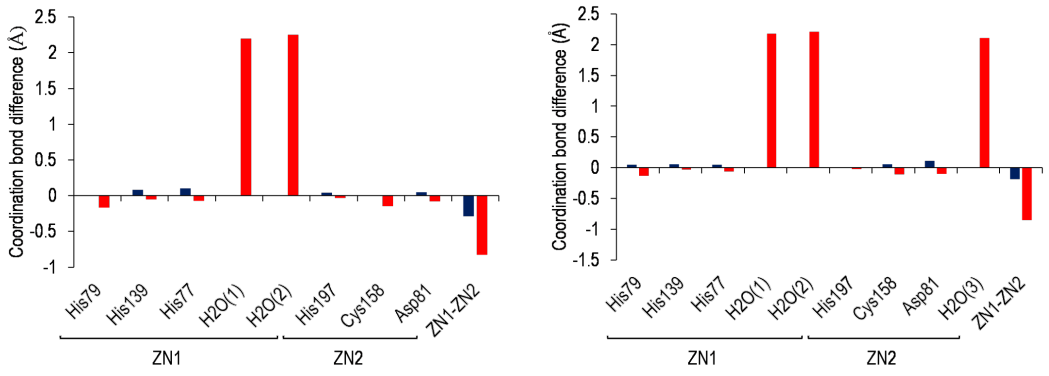


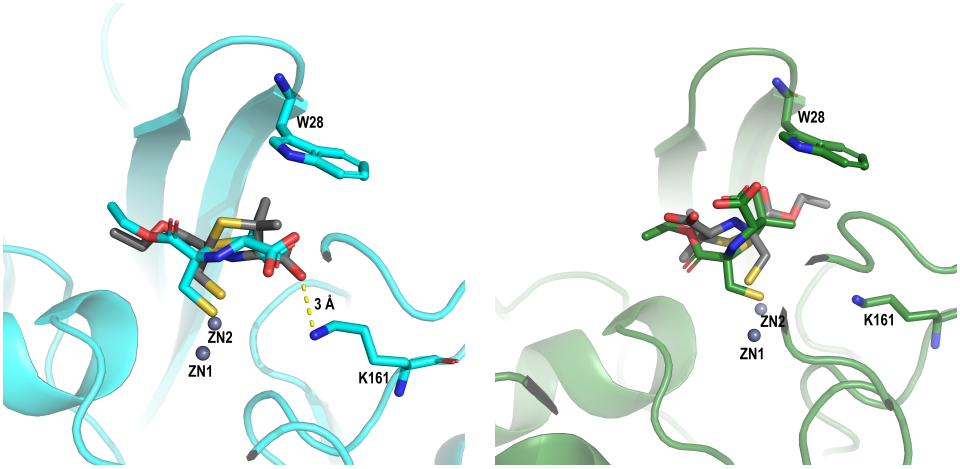
Figure 6.4: Differences of the coordination bond length between the crystal structure and the B3LYP/6-31G(d) QM/MM optimized complexes, starting from the last frame of the MD simulation of the LJ12-6-R (blue) and the LJ12-6-4 (red) models. Panel (A) refers to the THZ complex, while panel (B) to the VCZ complex. Water molecules that coordinate the zinc ions are also reported.

The results indicate that QM/MM optimization of MD simulations, carried out using the LJ12-6-R model, leads to restoration of the inhibitor binding mode observed in the crystal structure, with correct coordination geometry of the zinc ions. This approach thus successfully replicates crystallographically observed THZ and VCZ binding modes.

6.3.2 Prediction of RSZ binding mode to IMP-1

RSZ, as well as THZ and VCZ, acts as micromolar inhibitor of IMP-1 (data not shown). However its crystal structure in complexes with IMP-1 could not be obtained, even though biochemical data indicated that the potency was comparable to that of compounds whose IMP-1 complexes could be structurally characterised. Accordingly, we sought to apply the approach developed here to model interactions of RSZ with IMP-1. RSZ is respectively a diastereoisomer and enantiomer of THZ and VCZ. As it was unclear which of these complexes was likely to better accommodate the binding of RSZ, docking experiments were carried out using IMP-1 conformations derived from both the THZ and VCZ crystal structures, and subsequent simulations were undertaken in parallel. Based upon the results above, molecular docking of RSZ into IMP-1 in both conformations (i.e. from the THZ and

VCZ complexes) used the Opt3 approach described above. Figure 6.5 shows the binding modes obtained using the two crystal structures. The orientations of RSZ differ significantly between the two models, with the inhibitor carboxylate group oriented towards the Lys161 side chain in the model based upon the THZ structure (with a distance of 3 Å between the carboxylate group and the lysine amino group), while in that based upon the VCZ structure the inhibitor thiazolidine core is rotated by almost 180° and the interaction with Lys161 is lost. Thus, further simulations tested which of the two conformations generated stable binding of the inhibitor to the IMP-1 active site.



(A) (B)

Figure 6.5: Superimposition of the docked poses of the RSZ (gray) with the crystallized inhibitors using: (A) the THZ protein structure (coloured in cyan) and (B) the VCZ protein structure (coloured in green). The ligands are represented in stick. The interaction between the carboxylate group of the ligand and Lys161 is highlighted with a dotted line.

To this end, we carried out MD simulations on both the structures using both the LJ12-6-R and LJ12-6-4 models. The averaged RMSD values of the binding site of all the MD frames are reported in Table 6.4. We obtained higher RMSD values (of both the ligand and the binding site) using the VCZ structure suggesting that, in this structure, the protein was not in the best conformation to accommodate and establish interaction with RSZ. In contrast, when using the THZ structure lower RMSD values of the binding site were obtained and the ligand remained more stable with an RMSD along the trajectory that fluctuated around 0.5 Å (Figure S6.6). These values were little changed after QM/MM optimization using the B3LYP hybrid functional with the 6-31G(d) basis set (Table 6.4).

	MD model	THZ crystal structure	VCZ crystal structure
MD simulation	LJ12-6-R	1.20	2.21
	LJ12-6-4	1.37	1.53
QM/MM optimization	From LJ12-6-R	1.31	2.58
	From LJ12-6-4	1.49	1.73

Table 6.4: Averaged RMSD of the binding site (Å) versus crystal structures of VCZ and THZ and all the frames of the MD simulation (upper panel) and the B3LYP/6-31G(d)/MM optimized complexes (lower panel), in respect with the binding, using the two models.

As observed for the simulations described above, the use of the LJ12-6-4 model resulted in increased coordination numbers for the active site zinc ions, and B3LYP/6-31G(d)/MM optimization did not restore the tetrahedral coordination geometry present in the crystal structures of either the THZ or VCZ complexes. Figure 6.6 illustrates the differences in the coordination bond lengths with respect to the starting crystal structures in simulations using the LJ12-6-R model. During MD simulations based upon both starting structures, an additional water molecule (H2O(1) in Figure 6.6) coordinates the first zinc ion, changing the original tetrahedral coordination into five-coordinated geometry. Moreover, in simulations based upon the VCZ structure, the residue Asp81 moved from the starting position and coordinated both the zinc ions, which contrasts with the architecture reported in published crystal structures of IMP-1 and its complexes [338][350][351][352][353]. Notably, after the QM/MM optimization of the structure from THZ simulated with the LJ12-6-R model, the additional water molecule (H2O(1)) that coordinated ZN1 moved away (i.e. from 2 to 2.6 Å, where we did not consider it as coordinated), restoring the tetrahedral geometry of ZN1. In contrast, in simulations starting from the VCZ structure, both the additional water molecule (H2O(1)) and Asp81 maintained coordination with ZN1 even after the QM/MM optimization. These results suggest that the complex obtained from simulations based upon the VCZ structure is not reliable, given the major changes in the coordination sphere of the metal center. On the contrary, the final complex derived from simulations based upon the THZ structure is consistent with crystallographically determined coordination distances and geometries between the inhibitor thiolate and the zinc ions reported in the literature for other thiolate inhibitors [51] (Table S6.2).

Given these data, we considered as the final reliable model of the binding mode, the complex obtained using the THZ structure, in which we docked the inhibitor

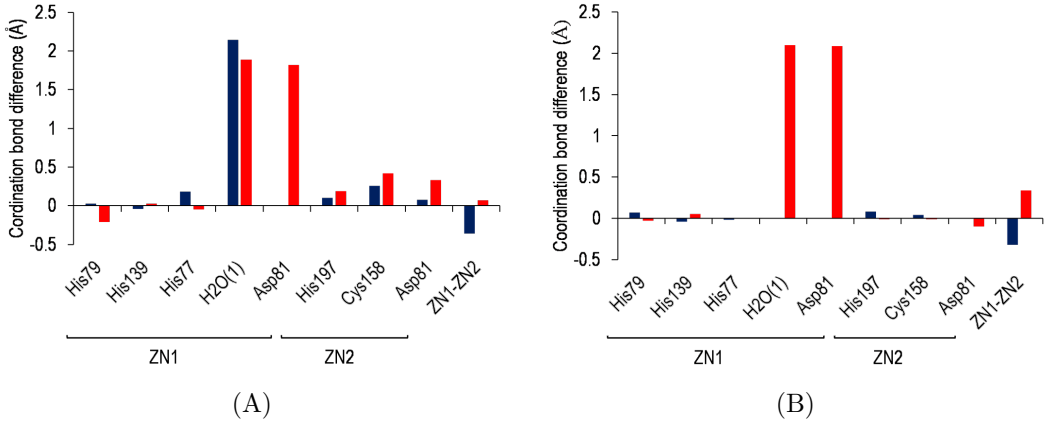


Figure 6.6: Differences of the coordination bond length between the crystal structure and (A) the last frame of the LJ12-6-R MD simulation and (B) the B3LYP/6-31G(d) QM/MM optimization, for the THZ (blue) and VCZ (red) structures. The water molecule that coordinate the zinc ions is also reported.

using the Opt3 treatment, and that subsequently underwent MD simulation using the LJ12-6-R model and finally B3LYP QM/MM optimization.

The final predicted binding mode for RSZ is illustrated in Figure 6.7. The thiolate coordinates both zinc ions, and inhibitor binding is stabilized by the interaction between the carboxylate group of the ligand and the amino group of the Lys161 side chain. Notably, the conformation of Trp28 changed, compared to its position in the crystal structure of the THZ complex. Conformational flexibility of this residue and the extended loop (sometimes termed the L3 loop) upon which it sits, has been previously described in multiple studies [353][359][360][361]. The RMSD of the IMP-1 binding site (including Trp28, Glu23, Val25 and Val31, which are located in the L3 loop and face towards the binding site) compared with the starting THZ complex crystal structure is 1.31 Å, indicating that the complexes remained stable during the simulation.

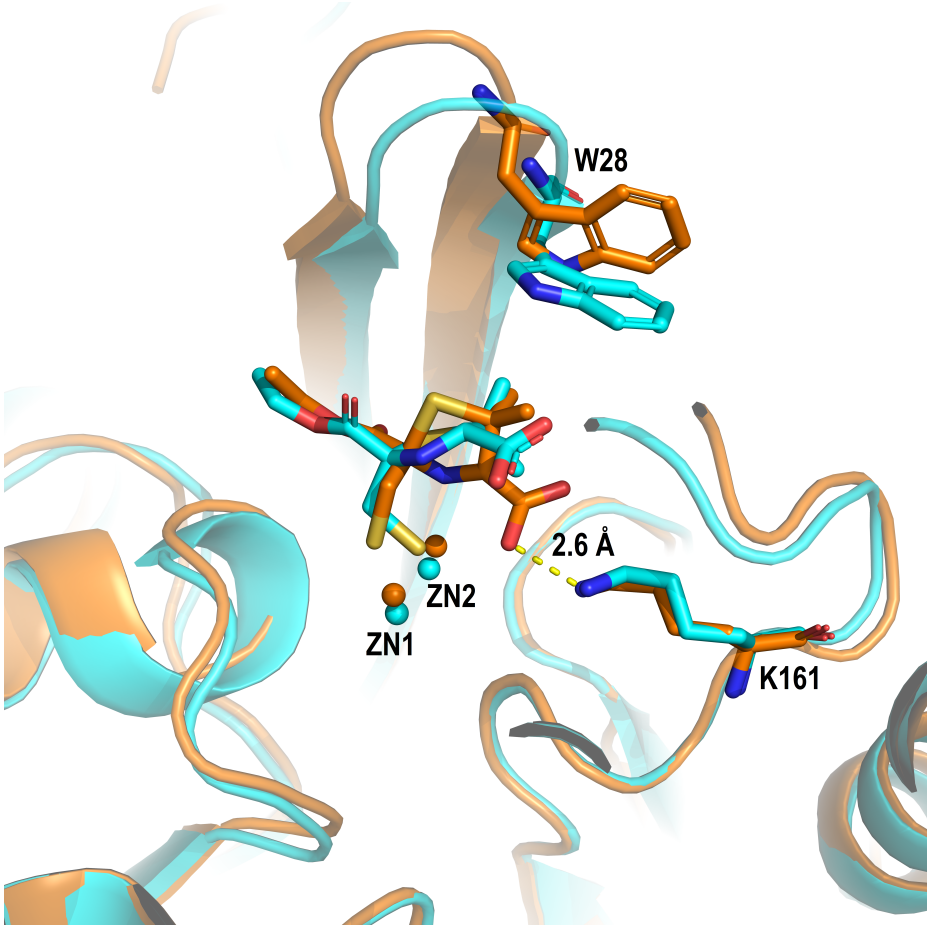


Figure 6.7: Superimposition of the crystal structure of IMP-1 THZ (cyan) and the predicted binding mode of RSZ (orange), obtained after QM/MM optimization. The ligands are represented in stick. The interaction between the carboxylate group of the ligand and Lys161 is highlighted with a dotted line

6.4 Discussion

The drug discovery process often relies upon molecular docking and MD simulations, which can usefully be combined [260][57][362]. However, when considering metalloproteins, docking approaches suffer from important limitations due to the functional forms used to describe interactions of metal ions and related difficulties in their interactions (e.g. polarization) [363]. It is well known that scoring functions struggle to correctly rank docking poses in metalloproteins [364][365]. Similarly, despite several efforts to develop non-bonded parameters for metal ions [114][113], MD simulations with typical empirical MM force fields also often fail to maintain protein-ligand complexes with correct metal ion coordination numbers and geometries [112][115][366]. We found that the LJ12-6-4 non-bonded model led to changes

from the original tetrahedral coordination into an octahedral geometry through addition of water molecules on both zinc ions. This incorrect coordination could not be reversed, even with QM/MM calculations. This variability of coordination geometry, which is an important contributor to the versatility and biological activity of metalloprotein zinc centres, compromises the stability of models of the binding site and the reliability of predictions of ligand binding. In this work, we tested workflows to predict the binding mode of thiolate inhibitors to metallo β-lactamases, in which the accuracy of the method increases with the application of MD simulations followed by QM/MM modelling, and we tested the workflows ability to reproduce the binding modes seen in crystal structures. The overall approach has potential applications in drug discovery/design for zinc-containing proteins.

The best docking results were achieved by including the crystallographic water molecules, considering the protein as rigid and increasing the contribution of the interaction of the thiolate with the zinc ions, through addition of distance restraints to the ChemPLP scoring function [90]. Subsequently, MD simulations refined the docking poses, with the lowest RMSD values observed using a restrained non-bonded model. As in docking experiments, restraints were necessary in MD simulations to maintain the active site structure, demonstrating the limitations of empirical force fields for modelling zinc sites. In MD, restraints were applied between each zinc ion and the coordinating protein atoms. These MD simulations produced protein-ligand complexes in which the coordination number of the zinc ions was still increased compared to crystal structures, though to a lesser extent in simulations using the LJ12-6-R as opposed to the LJ12-6-4 model. Further QM/MM calculations showed that the approximate SCC-DFTB3 method was unable to restore the experimentally observed coordination geometries of the metal ions, while the use of better DFT methods (i.e. B3LYP/6-31G(d)) enabled recovery of the tetrahedral geometries of the zinc ions present in the crystal structures as well as a significant reduction in RMSD. Our results thus demonstrate that this last step is of particular importance in obtaining correct coordination geometries that cannot be maintained using just MM MD simulations. QM/MM optimization, with reasonable levels of DFT QM treatment, is required to generate reasonable geometries for these complexes.

6.5 Conclusions

In this work we tested several techniques and their ability to reproduce the inhibitor binding modes observed in crystal structures. However, we also identified

methods that failed handling the metal coordination sphere. In detail, docking without the inclusion of crystallographic water molecules was not able to produce poses with an RMSD lower than 2 Å compared with the corresponding original crystal structure. The results also demonstrate limitations of standard MM models for MD simulations of these systems. MD simulations with a LJ12-6-4 model increased the zinc ions coordination number up to six; therefore this model may be well suited for zinc sites with octahedral coordination geometry [327][367], but less so for other architectures. Application of coordination restraints with a standard non-bonded model gave better results. Finally, we could not obtain the correct coordination geometry of the zinc ions using geometry optimization with the semi-empirical SCC-DFTB3/MM method. In contrast, QM/M geometry optimization with DFT methods restored the experimentally observed coordination. The results indicate the need for a high level/quantum chemical treatment for accurate modelling for the zinc metalloenzyme active sites and their interactions. Finally, we applied the workflow to predict the binding mode for a thiazolidine MBL inhibitor for which the crystal structure of the complex with IMP-1 could not be obtained. Our workflow produced a solid model for an inhibitor complex that is biochemically realistic both with respect to zinc coordination and interactions with active site residues (Lys161) known to be important to ligand binding by IMP-1 [353][293][292].

This work presents a validated, multiscale approach to computational prediction of IMP-1 interactions with thiolate inhibitors that is suitable for application to other MBL inhibitors and other zinc-containing proteins.

6.6 Supplementary Information

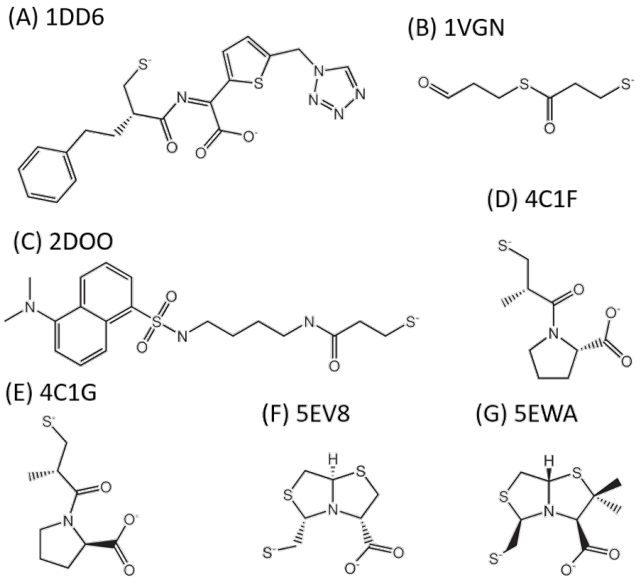


Figure S6.1: Structures of IMP-1 inhibitors for the redocking validation. (A) (2-mercaptomethyl-4-phenyl-butyrylimino)-(5-tetrazol-1-ylmethyl-thiophen-2-yl)-acetic acid, (B) 3-oxo-3-[(3-oxopropyl)sulfanyl]propane-1-thiolate, (C) dansylC4SH; D, L-captopril, (E) D-captopril, (F) D-CS319, (G) L-VC26. The PDB code of every compound is reported above the structure.

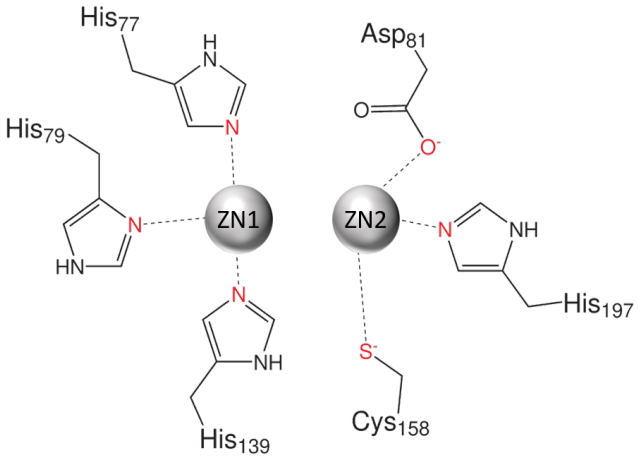


Figure S6.2: Schematic representation of the IMP-1 active site. The atoms restrained in the LJ12-6-R model are highlighted in red.

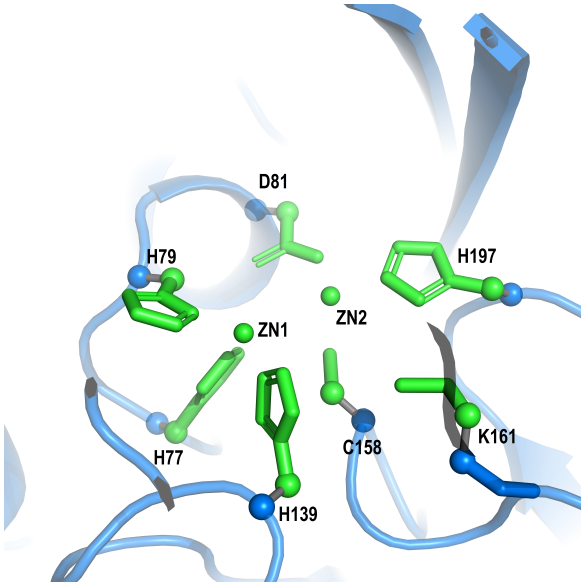


Figure S6.3: QM and MM regions of IMP-1 active site. The QM atoms are represented as green sticks, while the MM region is coloured in blue. Bonds between the two regions are shown in gray and atoms at the interface are represented as balls.

	QM/MM theory	THZ	VCZ
Binding site[a]	PBE0	1.43	1.87
	DFTB	1.50	1.88
Ligand	PBE0	0.66	1.37
	DFTB	0.80	1.36

Table S6.1: RMSD expressed in Å between the crystal structures of VCZ and THZ and the QM/MM optimized complexes, in respect with the binding site and the ligand, using DFT (PBE0) and DFTB3. [a]Binding site residues: Glu23, Val25, Trp28, Val31, Phe51, Ala53, His77, His79, Ser80, Asp81, His139, Thr140, Cys158, Lys161, Gly166, Asn167, Ser196, His197, HO^-, ZN1, ZN2.

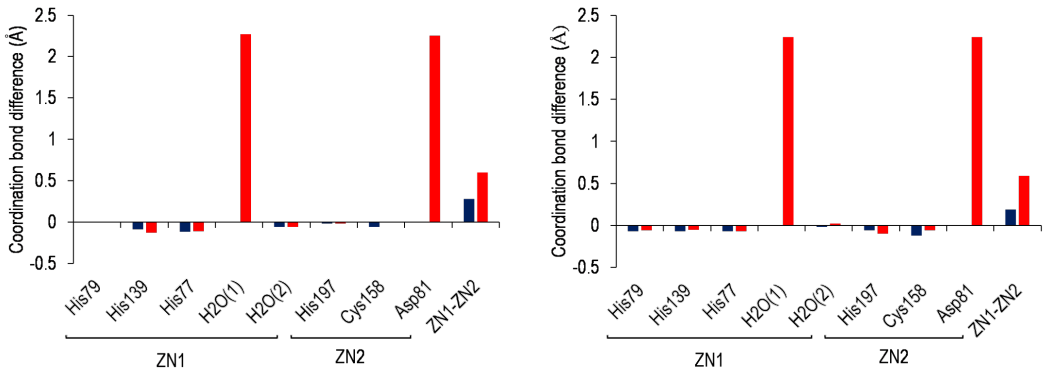
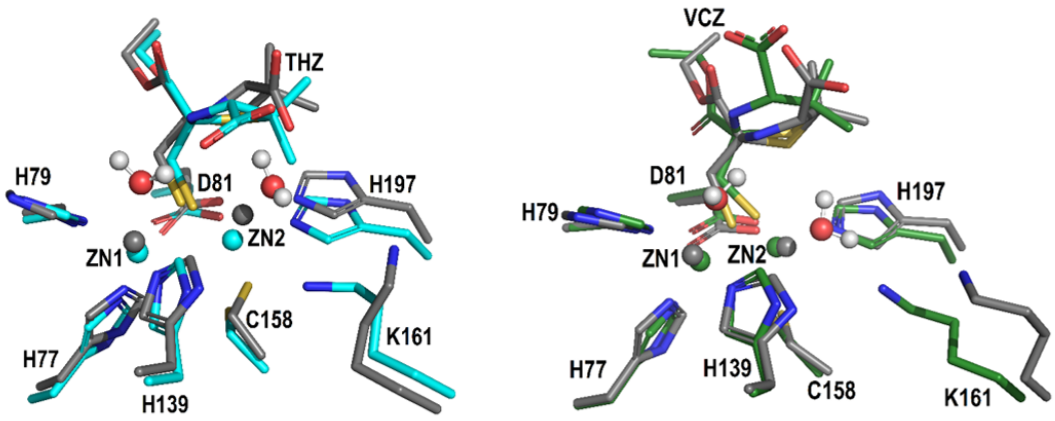
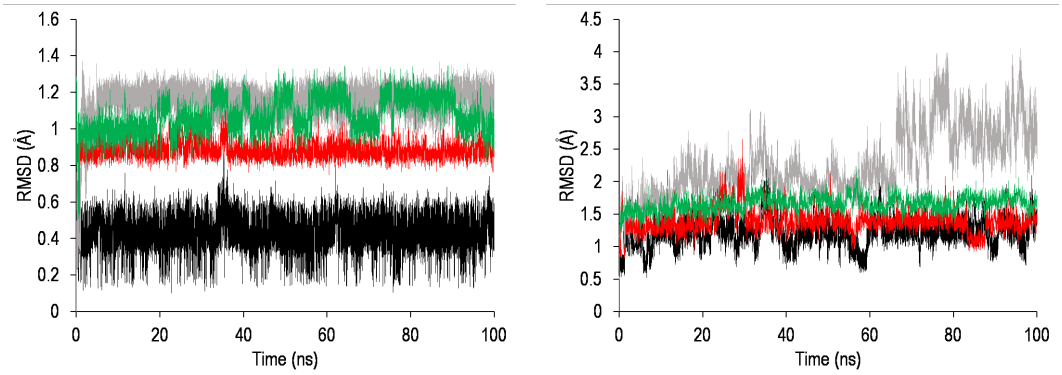


Figure S6.4: Differences in coordination bond length between the original crystal structure and its final QM/MM optimized complexes using DFT (PBE0) (red) and DFTB3 (blue), starting from the last frame of the MD simulation of the LJ12-6-R model. Panel (A) refers to the THZ complex, while panel (B) to the VCZ complex. Water molecules that coordinate the zinc ions are also reported.



(A) (B)

Figure S6.5: Superimpositions of the DFTB3 optimized models with the starting crystal structures. (A) THZ crystal structure (cyan) and the corresponding optimized structure in gray. (B) VCZ crystal structure (green) and the corresponding optimized structure in gray. Water molecules are represented in ball and stick.



(A) (B)

Figure S6.6: CRMSD (Å) fluctuation during the MD of (A) RSZ and of (B) the binding site during the MD simulations. In black the results using the THZ structure with the LJ12-6-R model, in red the THZ structure with the LJ12-6-4, in gray the VCZ structure with the LJ12-6-R model and in green the VCZ structure with the LJ12-6-4 model.

	MD				QM/MM B3LYP				QM/MM PBE0	
	LJ12-6-R		LJ12-6-4		From LJ12-6-R		From LJ12-6-4		From LJ12-6-R	
	THZ	VCZ	THZ	VCZ	THZ	VCZ	THZ	VCZ	THZ	VCZ
ZN1	1.98	2.16	2.15	2.24	2.28	3.19	2.43	2.33	2.27	3.11
ZN2	2.24	1.96	2.21	2.17	2.62	2.26	2.45	2.29	2.59	2.24

Table S6.2: Coordination distances (Å) between the zinc ions and the thiolate of RSZ. THZ and VCZ refer to the starting structures of the protein used in the calculations.

Chapter 7

Conclusions and Future perspectives

The main objective of this Thesis is to investigate computational approaches to handle zinc-containing systems. In detail, we tested modelling techniques with different levels of accuracy, starting from simplistic approaches and moving toward complex description of the metal coordination bond.

A classic approach was applied in **Chapter 3** for the design of new HDAC inhibitors, where the metal ion interactions were described using scoring function and the binding affinity was evaluated using a force field, which uses the classic non-bonded model. Such an approach allows a fast screening of large library of molecules (or fragments) and does not require complex settings. However, it has some limitations: good candidates can be excluded from the selection due to missing metal interaction, which could depend on the inability of the docking software to correctly place the ligand inside the binding pocket. A second limitation derives from the difficulty of scoring function to represent all the metal coordination geometries and therefore a reliable binding mode of the ligand, that is confirmed by the preference of metal coordination over metal chelation in the resulting complexes.
The 12-6-4 non-bonded model was used in both **Chapter 4** and **6**. In **Chapter 4** this model was used for the parametrization of a metallo-organic AChE reactivator, which included a zinc ion complexed in octahedral geometry. During the MD simulations the reactivator maintained its integrity as well as the octahedral geometry, allowing a solid description of its dynamic behaviour into AChE. In cases in which the metal ion is included into a ligand (and not into a protein), a popular strategy

is to use the bonded model, with parameters derived from QM calculations [368]. Although such an approach could better reproduce the properties of the specific ligand, it could be time consuming. In the case of the reactivator K2 we obtained good results even with a more straightforward parametrization.
Different results were obtained when the 12-6-4 non-bonded model was applied to the metallo-enzyme IMP-1 in **Chapter 6**. In this case the geometries of the zinc ions resulted distorted compared to the crystal structure used as reference, suggesting the increase of difficulty in modelling metal ions included into complex environments, like proteins.
In **Chapter 5** the cationic dummy atom approach was applied for drug discovery of new CN1 inhibitors. We decided to model one catalytic zinc ion as tetrahedral (using Pang's parameters [34]) and the other zinc ion as five-coordinated. For this last parametrization, *de novo* parameters were created considering a trigonal-base bipyramidal coordination geometry. The cationic dummy atoms were able to maintain the correct coordination geometries with both the ligands and the protein residues. At the same time, exchange ligand events could occur without altering the coordination site integrity. On the contrary, changes in neither coordination number nor geometry can occur. This is the main limitation of this method that forces the choice of the coordination geometry *a priori*.
The highest level of accuracy was reached in **Chapter 6** using a QM/MM geometry optimization for the modelling of the binding mode of a new β-lactamase inhibitor. A multiscale approach that comprised docking, MD simulation and DFT/MM geometry optimization was developed. In the MD simulation a combination of the classic non-bonded model with distance restraints was applied to model zinc ion interactions. MD simulations returned overall good results (i.e. complexes with general low RMSD values with respect to the crystal structures), however distortions in the coordination geometries were found. Therefore, a DFT/MM optimization step was added, and it was able to fix the errors introduced by the MD simulation, recovering the correct coordination geometry. QM methods are able to better describe metal interactions and to overcome the approximations introduced by force fields, their main limitation is linked to the computational cost and to the complexity of the set-up.

Although the parametrization issues mentioned above, this work shows that zinc-containing biological systems can be simulated by combinations of different methods designed *ad hoc* for each case. This combination allows to compensate errors introduced by classic approaches (e.g. docking and MD simulation) and to analyse the system under multiple points of view (classic MM and QM). Moreover,

the pipelines proposed here can be applied to similar systems, which comprise both metalloproteins and organometallic ligands.

The constant increase of computational power has broadened our possibilities in terms of modelling techniques. In metal ion modelling lots of expectations have been risen by polarizable force fields [37][369][370] and *ab initio* force fields [371] derived by machine learning techniques [372][373][374][375]. To date these methods are still not fully developed and tested, however they suggest a path towards an improvement of parametrization that will include further aspects of metal interaction that are still neglected and simplified by classical force fields.

Acknowledgements

The work of **Chapter 4** was in collaboration with Dr. Kostantis Konidaris from the University of Patras.

The work of **Chapter 5** was in collaboration with Prof. Giancarlo Aldini, Dr. Luca Regazzoni and Dr. Ettore Gilardoni from the Universitá degli Studi di Milano. Computational resources provided by INDACO Platform, which is a project of High Performance Computing at the Universitá degli Studi di Milano (http://www.unimi.it).

The work of **Chapter 6** was in collaboration with Prof. Adrian Mulholland, Prof. James Spencer and Dr. Philip Hinchliffe from the University of Bristol.
This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bris.ac.uk/acrc/.

Bibliography

[1] A. Rosato, Y. Valasatava, and C. Andreini, "Minimal Functional Sites in Metalloproteins and Their Usage in Structural Bioinformatics," *International Journal of Molecular Sciences*, vol. 17, no. 5, p. 671, 2016.

[2] C. Frauer, A. Rottach, D. Meilinger, S. Bultmann, K. Fellinger, S. Hasenöder, M. Wang, W. Qin, J. Söding, F. Spada, and H. Leonhardt, "Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1," *PLoS ONE*, vol. 6, no. 2, p. e16627, 2011.

[3] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton, "Metal ions in biological catalysis: from enzyme databases to general principles," *JBIC Journal of Biological Inorganic Chemistry*, vol. 13, no. 8, pp. 1205–1218, 2008.

[4] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[5] C. Zhang, F. Zhang, P. Zhou, and C. Zhang, "Functional role of metalloproteins in genome stability," *Frontiers in Biology*, vol. 11, no. 2, pp. 119–131, 2016.

[6] P. Rydberg, F. S. Jorgensen, and L. Olsen, "Use of density functional theory in drug metabolism studies," *Expert Opinion on Drug Metabolism and Toxicology*, vol. 10, no. 2, pp. 215–227, 2014.

[7] M. Cassandri, A. Smirnov, F. Novelli, C. Pitolli, M. Agostini, M. Malewicz, G. Melino, and G. Raschellà, "Zinc-finger proteins in health and disease," *Cell Death Discovery*, vol. 3, no. 1, pp. 1–12, 2017.

[8] G. Gao, J. Li, Y. Zhang, and Y. Z. Chang, "Cellular iron metabolism and regulation," *Advances in Experimental Medicine and Biology*, vol. 1173, pp. 21–32, 2019.

[9] A. Jabłońska-Trypuć, M. Matejczyk, and S. Rosochacki, "Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 31, no. 1, pp. 177–183, 2016.

[10] M. Laws, A. Shaaban, and K. M. Rahman, "Antibiotic resistance breakers: current approaches and future directions," *FEMS Microbiology Reviews*, vol. 43, pp. 490–516, 2019.

[11] A. Engelman and P. Cherepanov, "Retroviral Integrase Structure and DNA Recombination Mechanism," *Microbiology Spectrum*, vol. 2, no. 6, pp. 1–22, 2014.

[12] F. Nastri, D. D'Alonzo, L. Leone, G. Zambrano, V. Pavone, and A. Lombardi, "Engineering Metalloprotein Functions in Designed and Native Scaffolds," *Trends in Biochemical Sciences*, vol. 44, no. 12, pp. 1022–1040, 2019.

[13] Y. Yang, X.-Q. Hu, Q.-S. Li, X.-X. Zhang, B.-F. Ruan, J. Xu, and C. Liao, "Metalloprotein Inhibitors for the Treatment of Human Diseases," *Current Topics in Medicinal Chemistry*, vol. 16, no. 4, pp. 384–396, 2015.

[14] A. C. West, R. W. Johnstone, C. Olsen, M. Tan, S. Spange, T. Wagner, T. Heinzel, O. Krämer, M. Fraga, H. Ozdag, W. Weichert, W. Weichert, C. Krusche, Y. Minamiya, H. Osada, Y. Tatematsu, H. Saito, Y. Yatabe, T. Mitsudomi, T. Takahashi, T. Rikimaru, W. Weichert, M. Glozak, E. Seto, G. Eot-Houllier, G. Fulcrand, L. Magnaghi-Jaulin, C. Jaulin, S. Saji, L. Marquard, L. Gjerdrum, I. Christensen, P. Jensen, M. Sehested, E. Ralfkiaer, J. Luo, F. Su, D. Chen, A. Shiloh, W. Gu, H. Duan, C. Heckman, L. Boxer, V. Gelmetti, J. Zhang, M. Fanelli, S. Minucci, P. Pelicci, M. Lazar, R. Lin, L. Nagy, S. Inoue, W. Shao, W. Miller, R. Evans, A. Insinga, A. Nebbioso, B. Taylor, S. Ropero, S. Bhaskara, M. Heideman, F. Santoro, K. Lin, Y. Wang, C. Chen, C. Ho, W. Su, Y. Jou, S. Pulukuri, B. Gorantla, J. Rao, J. Vojinovic, C. Bonfils, M. Guenther, O. Barak, M. Lazar, Y. Zhang, H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, D. Reinberg, M. Bantscheff, J. Bradner, M. Deardorff, G. Scott, A. Leder, P. Leder, M. Riggs, R. Whittaker, J. Neumann, V. Ingram, J. Bolden, M. Peart, R. Johnstone, L. Ellis, R. Lindemann, A. Newbold, R. Lindemann, L. Cluse, K. Whitecross, A. Dear, R. Johnstone, J. Vrana, Y. Zhang, M. Adachi, R. Kawamura, K. Imai, V. Fantin, S. Fotheringham, O. Khan, M. New, A. Vannini, J. Bolden, M. Duvic, E. Olsen, S. Whittaker, R. Piekarz, A. Nebbioso, V. Carafa, R. Benedetti, L. Altucci, M. New, H. Olzscha, N. L. Thangue, T. Qiu, B. Gryder, Q. Sodji, A. Oyelere, U. Banerji, M. Dong, J. Bressi, L. Zhang, M. Li, L. Zhang, W. Xu, F. Grignani, X. Chen, C. Sandi, E. Soragni, C. Xu, H. Plasterer, V. Jacques, J. Rusche, J. Gottesfeld, I. Oehme, J. Somoza, S. Balasubramanian, J. Ramos, W. Luo, M. Sirisawad, E. Verner, J. Buggy, C. Reilly, R. Tao, A. Wilson, D. Mottet, W. Fischle, M. Bottomley, A. Schuetz, M. Lobera, S. Haggarty, K. Koeller, J. Wong, C. Grozinger, S. Schreiber, Y. Kawaguchi, J. Kovacs, A. McLaurin, J. Vance, A. Ito, T. Yao, J. Kovacs, P. Bali, L. Santo, T. Hideshima, T. Hideshima, P. Richardson, K. Anderson, C. Hubbert, B. Gopalan, G. Delcuve, D. Khan, J. Davie, L. Needham, H. Deubzer, Z. Zhang, H. Adams, F. Fritzsche, S. Dirnhofer, G. Kristiansen, A. Tzankov, S. Hiebert, G. David, L. Alland, S. Hong, C. Wong, R. DePinho, A. Dejean, P. Dhordain, A. Wilson, T. Yamaguchi, F. Santoro, M. Haberland, A. Johnson, M. Mokalled, R. Montgomery, E. Olson, S. Senese, M. Glozak, N. Sengupta, X. Zhang, E. Seto, R. Wilting, J. Song, K. Jung, K. Harms, X. Chen, D. Moreno, S. Hong, G. David, C. Wong, A. Dejean, M. Privalsky, O. Kramer, A. Atsumi, A. Tomita, H. Kiyoi, T. Naoe, X. Sun, L. Wei, Q. Chen, R. Terek, T. Milde, S. Malik, S. Kotian, S. Liyanarachchi, A. Zelent, J. Parvin, Z. Zhang, J. Park, D. Rhodes, L. Gao, M. Cueto, F. Asselbergs, P. Atadja, M. Duvic, A. Younes, G. Giaccone, L. Gore, A. Hauschild, Q. Ryan, A. Younes, A. Rambaldi, A. Razak, and W. Yong, "New and emerging HDAC inhibitors for cancer treatment," *Journal of Clinical Investigation*, vol. 124, no. 1, pp. 30–39, 2014.

[15] T. Eckschlager, J. Plch, M. Stiborova, and J. Hrabeta, "Histone deacetylase inhibitors as anticancer drugs," *International Journal of Molecular Sciences*, vol. 18, no. 7, p. 1414, 2017.

[16] S. Zolghadri, A. Bahrami, M. T. Hassan Khan, J. Munoz-Munoz, F. Garcia-Molina, F. Garcia-Canovas, and A. A. Saboury, "A comprehensive review on tyrosinase inhibitors," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 34, no. 1, pp. 279–309, 2019.

[17] A. Y. Chen, R. N. Adamek, B. L. Dick, C. V. Credille, C. N. Morrison, and S. M. Cohen, "Targeting Metalloenzymes for Therapeutic Intervention," *Chem. Rev*, vol. 119, pp. 1323–1455, 2019.

[18] M. Rouffet and S. M. Cohen, "Emerging trends in metalloprotein inhibition," *Dalton Transactions*, vol. 40, no. 14, pp. 3445–3454, 2011.

[19] S. P. Leelananda and S. Lindert, "Computational methods in drug discovery," *Beilstein Journal of Organic Chemistry*, vol. 12, no. 1, pp. 2694–2718, 2016.

[20] F. D. Prieto-Martínez, E. López-López, K. Eurídice Juárez-Mercado, and J. L. Medina-Franco, "Computational Drug Design Methods—Current and Future Perspectives," in *In Silico Drug Design* (Elsevier Inc. All, ed.), no. 3, pp. 19–44, Elsevier, 2019.

[21] S. E. Howson, N. P. Chmel, G. J. Clarkson, R. J. Deeth, D. H. Simpson, and P. Scott, "Jahn-Teller effects on π-stacking and stereoselectivity in the phenylethaniminopyridine trischelates Cu(NN$'$) 32+," *Dalton Transactions*, vol. 41, no. 15, pp. 4477–4483, 2012.

[22] R. H. Crabtree, *The organometallic chemistry of the transition metals*. John Wiley & Sons, Inc., 4 ed., 2005.

[23] J. Wang, P. Cieplak, and P. A. Kollman, "How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? Keywords: additive force field; nonadditive force field; restrained electrostatic potential (RESP); torsion," *Journal of Computational Chemistry*, vol. 21, no. 12, pp. 1049–1074, 2000.

[24] P. Li, B. P. Roberts, D. K. Chakravorty, and K. M. Merz, "Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent," *Journal of Chemical Theory and Computation*, vol. 9, no. 6, pp. 2733–2748, 2013.

[25] A. D. Mackerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *J. Phys. Chem.*, vol. 102, pp. 3586–3616, 1998.

[26] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, and A. D. MacKerell, "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, no. 1, pp. 71–73, 2016.

[27] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," *Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[28] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, "OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins," *Journal of Chemical Theory and Computation*, vol. 12, no. 1, pp. 281–296, 2016.

[29] W. R. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. Van Gunsteren, "The GROMOS biomolecular simulation program package," *Journal of Physical Chemistry A*, vol. 103, no. 19, pp. 3596–3607, 1999.

[30] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, and W. F. Van Gunsteren, "Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation," *Computer Physics Communications*, vol. 183, no. 4, pp. 890–903, 2012.

[31] D. V. Sakharov and C. Lim, "Zn protein simulations including charge transfer and local polarization effects," *Journal of the American Chemical Society*, vol. 127, no. 13, pp. 4921–4929, 2005.

[32] R. J. Deeth, N. Fey, and B. Williams-Hubbard, "DommiMOE: An implementation of ligand field molecular mechanics in the molecular operating environment," *Journal of Computational Chemistry*, vol. 26, no. 2, pp. 123–130, 2005.

[33] P. Li and K. M. Merz, "MCPB.py: A Python Based Metal Center Parameter Builder," *Journal of Chemical Information and Modeling*, vol. 56, no. 4, pp. 599–604, 2016.

[34] Y.-P. Pang, "Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method," *Proteins: Structure, Function, and Genetics*, vol. 45, no. 3, pp. 183–189, 2001.

[35] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins," *J Chem Theory Comput.*, vol. 9, no. 9, pp. 4046–4063, 2013.

[36] J. Chowdhary, E. Harder, P. E. Lopes, L. Huang, A. D. MacKerell, and B. Roux, "A polarizable force field of dipalmitoylphosphatidylcholine based on the classical drude model for molecular dynamics simulations of lipids," *Journal of Physical Chemistry B*, vol. 117, no. 31, pp. 9142–9160, 2013.

[37] C. M. Baker, "Polarizable force fields for molecular dynamics simulations of biomolecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 5, no. 2, pp. 241–254, 2015.

[38] S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife, and D. R. Salahub, "Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review," *International Journal of Quantum Chemistry*, vol. 118, no. 9, p. e25558, 2018.

[39] C. Moller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Physical Review*, vol. 46, no. 7, pp. 618–622, 1934.

[40] H. G. Kummel, "A biography of the coupled cluster method," *International Journal of Modern Physics B*, vol. 17, no. 28, pp. 5311–5325, 2003.

[41] V. Magnasco, "Post-Hartree–Fock methods," in *Elementary Molecular Quantum Mechanics*, pp. 681–722, Elsevier, 2013.

[42] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, "Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications," *Chemical Reviews*, vol. 116, no. 9, pp. 5301–5337, 2016.

[43] W. Thiel and A. A. Voityuk, "Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group," *J. Phys. Chem.*, vol. 100, pp. 616–626, 1996.

[44] M. J. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. Stewart, "AM1: A New General Purpose Quantum Mechanical Molecular Model," *Journal of the American Chemical Society*, vol. 107, no. 13, pp. 3902–3909, 1985.

[45] J. J. Stewart, "Optimization of parameters for semiempirical methods II. Applications," *Journal of Computational Chemistry*, vol. 10, no. 2, pp. 221–264, 1989.

[46] J. N. Harvey, "On the accuracy of density functional theory in transition metal chemistry," *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, vol. 102, pp. 203–226, 2006.

[47] P. Vidossich and A. Magistrato, "QM/MM molecular dynamics studies of metal binding proteins," *Biomolecules*, vol. 4, no. 3, pp. 616–645, 2014.

[48] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, no. 2, pp. 227–249, 1976.

[49] Www.nobelprize.org/prizes/chemistry/, "The Nobel Prize in Chemistry."

[50] P. Barak and P. A. Helmke, "The Chemistry of Zinc," in *Zinc in Soils and Plants* (A. D. Robson, ed.), ch. 1, pp. 1–13, Kluwer Academic Publishers, 1993.

[51] M. Laitaoja, J. Valjakka, and J. Jänis, "Zinc Coordination Spheres in Protein Structures," *Inorganic Chemistry*, vol. 52, no. 19, pp. 10983–10991, 2013.

[52] E. Gilardoni, S. Gervasoni, M. Maspero, C. Dallanoce, G. Vistoli, M. Carini, G. Aldini, and L. Regazzoni, "Development of a direct LC-ESI-MS method for the measurement of human serum carnosinase activity," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 189, p. 113440, 2020.

[53] L. Ferreira, R. dos Santos, G. Oliva, and A. Andricopulo, "Molecular Docking and Structure-Based Drug Design Strategies," *Molecules*, vol. 20, no. 7, pp. 13384–13421, 2015.

[54] I. A. Guedes, C. S. de Magalhães, and L. E. Dardenne, "Receptor-ligand molecular docking," *Biophysical Reviews*, vol. 6, no. 1, pp. 75–87, 2014.

[55] I. A. Vakser, "Protein-protein docking: From interaction to interactome," *Biophysical Journal*, vol. 107, no. 8, pp. 1785–1793, 2014.

[56] W. Sherman, H. S. Beard, and R. Farid, "Use of an induced fit receptor structure in virtual screening," *Chemical Biology and Drug Design*, vol. 67, no. 1, pp. 83–84, 2006.

[57] R. E. Amaro, J. Baudry, J. Chodera, O. Demir, J. A. McCammon, Y. Miao, and J. C. Smith, "Ensemble Docking in Drug Discovery," *Biophysical Journal*, vol. 114, no. 10, pp. 2271–2278, 2018.

[58] P. S. Kharkar, S. Warrier, and R. S. Gaud, "Reverse docking: A powerful tool for drug repositioning and drug rescue," *Future Medicinal Chemistry*, vol. 6, no. 3, pp. 333–342, 2014.

[59] A. Scarpino, G. G. Ferenczy, and G. M. Keseru, "Comparative Evaluation of Covalent Docking Tools," *Journal of Chemical Information and Modeling*, vol. 58, no. 7, pp. 1441–1458, 2018.

[60] H. J. Böhm, "The computer program LUDI: A new method for the de novo design of enzyme inhibitors," *Journal of Computer-Aided Molecular Design*, vol. 6, no. 1, pp. 61–78, 1992.

[61] R. J. Read, T. N. Hart, M. D. Cummings, and S. R. Ness, "Monte Carlo algorithms for docking to proteins," *Supramolecular Chemistry*, vol. 6, no. 1-2, pp. 135–140, 1995.

[62] B. Guan, C. Zhang, and J. Ning, "Genetic algorithm with a crossover elitist preservation mechanism for protein–ligand docking," *AMB Express*, vol. 7, no. 1, pp. 714–725, 2017.

[63] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[64] G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *Journal of Computational Chemistry*, vol. 30, no. 16, pp. 2785–2791, 2009.

[65] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, "Improved Protein - Ligand Docking Using GOLD," *Proteins Structure Function And Bioinformatics*, vol. 52, pp. 609–623, 2003.

[66] A. K. Deb, "Introduction to soft computing techniques: Artificial neural networks, fuzzy logic and genetic algorithms," in *Soft Computing in Textile Engineering*, pp. 3–24, Elsevier Inc., 2011.

[67] O. Korb, T. Stützle, and T. E. Exner, "PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design," in *Lecture Notes in Computer Science*, pp. 247–258, Springer Berlin Heidelberg, 2006.

[68] J. Liu and R. Wang, "Classification of current scoring functions," *Journal of Chemical Information and Modeling*, vol. 55, no. 3, pp. 475–482, 2015.

[69] S. Y. Huang, S. Z. Grinter, and X. Zou, "Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions," *Physical Chemistry Chemical Physics*, vol. 12, no. 40, pp. 12899–12908, 2010.

[70] I. A. Guedes, F. S. Pereira, and L. E. Dardenne, "Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges," *Frontiers in Pharmacology*, vol. 9, p. 1089, 2018.

[71] Q. Shen, B. Xiong, M. Zheng, X. Luo, C. Luo, X. Liu, Y. Du, J. Li, W. Zhu, J. Shen, and H. Jiang, "Knowledge-based scoring functions in drug design: 2. can the knowledge base be enriched?," *Journal of Chemical Information and Modeling*, vol. 51, no. 2, pp. 386–397, 2011.

[72] M. Wójcikowski, P. J. Ballester, and P. Siedlecki, "Performance of machine-learning scoring functions in structure-based virtual screening," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[73] E. C. Meng, B. K. Shoichet, and I. D. Kuntz, "Automated docking with grid-based energy evaluation," *Journal of Computational Chemistry*, vol. 13, no. 4, pp. 505–524, 1992.

[74] M.-J. Hsieh and R. Luo, "Physical scoring function based on AMBER force field and Poisson-Boltzmann implicit solvent for protein structure prediction," *Proteins: Structure, Function, and Bioinformatics*, vol. 56, no. 3, pp. 475–486, 2004.

[75] L. P. Pason and C. A. Sotriffer, "Empirical Scoring Functions for Affinity Prediction of Protein-ligand Complexes," *Molecular Informatics*, vol. 35, no. 11-12, pp. 541–548, 2016.

[76] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," *Journal of Computer-Aided Molecular Design*, vol. 11, no. 5, pp. 425–445, 1997.

[77] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, and D. E. S. Research, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, pp. 1739–1749, 2004.

[78] H. F. Velec, H. Gohlke, and G. Klebe, "DrugScoreCSD-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction," *Journal of Medicinal Chemistry*, vol. 48, no. 20, pp. 6296–6303, 2005.

[79] M. Xue, M. Zheng, B. Xiong, Y. Li, H. Jiang, and J. Shen, "Knowledge-based scoring functions in drug design. 1. Developing a target-specific method for kinase-ligand interactions," *Journal of Chemical Information and Modeling*, vol. 50, no. 8, pp. 1378–1386, 2010.

[80] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking," *J. Med. Chem.*, vol. 55, p. 6582−6594, 2012.

[81] Q. U. Ain, A. Aleksandrova, F. D. Roessler, and P. J. Ballester, "Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 5, no. 6, pp. 405–424, 2015.

[82] P. J. Ballester and J. B. Mitchell, "A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking," *Bioinformatics*, vol. 26, no. 9, pp. 1169–1175, 2010.

[83] J. D. Durrant and J. A. McCammon, "NNScore 2.0: A neural-network receptor-ligand scoring function," *Journal of Chemical Information and Modeling*, vol. 51, no. 11, pp. 2897–2903, 2011.

[84] A. D. Hill and P. J. Reilly, "Scoring functions for autodock," *Methods in Molecular Biology*, vol. 1273, pp. 467–474, 2015.

[85] O. Korb, T. Stützle, and T. E. Exner, "Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS," *J. Chem. Inf. Model.*, vol. 49, pp. 84–96, 2009.

[86] Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, and T. Hou, "Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power," *Physical Chemistry Chemical Physics*, vol. 18, no. 18, pp. 12964–12975, 2016.

[87] F. D. Prieto-Martínez, M. Arciniega, and J. L. Medina-Franco, "Molecular docking: current advances and challenges," *TIP Revista Especializada en Ciencias Químico-Biológicas*, vol. 21, pp. 65–87, 2018.

[88] G. Sciortino, J. Rodríguez-Guerra Pedregal, A. Lledós, E. Garribba, and J.-D. Maréchal, "Prediction of the interaction of metallic moieties with proteins: An update for protein-ligand docking techniques," *Journal of Computational Chemistry*, vol. 39, no. 1, pp. 42–51, 2018.

[89] E. Ortega-Carrasco, A. Lledós, and J.-D. Maréchal, "Assessing protein-ligand docking for the binding of organometallic compounds to proteins," *Journal of Computational Chemistry*, vol. 35, no. 3, pp. 192–198, 2014.

[90] S. S. Çlnaroğlu and E. Timuçin, "Comparative Assessment of Seven Docking Programs on a Nonredundant Metalloprotein Subset of the PDBbind Refined," *Journal of Chemical Information and Modeling*, vol. 59, no. 9, pp. 3846–3859, 2019.

[91] X. Hu, S. Balaz, and W. H. Shelver, "A practical approach to docking of zinc metalloproteinase inhibitors," *Journal of Molecular Graphics and Modelling*, vol. 22, no. 4, pp. 293–307, 2004.

[92] D. Santos-Martins, S. Forli, M. J. Ramos, and A. J. Olson, "AutoDock4 Zn : An Improved AutoDock Force Field for Small-Molecule Docking to Zinc Metalloproteins," *Journal of Chemical Information and Modeling*, vol. 54, no. 8, pp. 2371–2379, 2014.

[93] M. Pavlin, G. Rossetti, M. De Vivo, and P. Carloni, "Carnosine and Homocarnosine Degradation Mechanisms by the Human Carnosinase Enzyme CN1: Insights from Multiscale Simulations," *Biochemistry*, vol. 55, no. 19, pp. 2772–2784, 2016.

[94] V. Muñoz Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodriguez-Guerra, A. Lledós, and J. D. Maréchal, "Toward the computational design of artificial metalloenzymes: From protein-ligand docking to multiscale approaches," *ACS Catalysis*, vol. 5, no. 4, pp. 2469–2480, 2015.

[95] R. Lonsdale, S. L. Rouse, M. S. Sansom, and A. J. Mulholland, "A Multiscale Approach to Modelling Drug Metabolism by Membrane-Bound Cytochrome P450 Enzymes," *PLoS Computational Biology*, vol. 10, no. 7, p. e1003714, 2014.

[96] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, vol. 2. Cambridge University Press, 2 ed., 2004.

[97] R. Hockney, "The potential calculation and some applications," *Methods Comput. Phys.*, vol. 9, pp. 136–211, 1970.

[98] J. I. Choe and B. Kim, "Determination of proper time step for molecular dynamics simulation," *Bulletin of the Korean Chemical Society*, vol. 21, no. 4, pp. 419–424, 2000.

[99] V. Kräutler, W. F. Van Gunsteren, and P. H. Hünenberger, "A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 22, no. 5, pp. 501–508, 2001.

[100] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: A Linear Constraint Solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[101] J. Wereszczynski and J. A. McCammon, "Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition," *Quarterly Reviews of Biophysics*, vol. 45, no. 1, pp. 1–25, 2012.

[102] J. E. Basconi and M. R. Shirts, "Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, pp. 2887–2899, 2013.

[103] C. Sagui and T. A. Darden, "Molecular dynamics simulations of biomolecules: Long-Range Electrostatic Effects," *Annu. Rev. Biophys. Biomol. Struct*, vol. 28, pp. 155–79, 1999.

[104] M. O. Steinhauser and S. Hiermaier, "A review of computational methods in materials science: Examples from shock-wave and polymer physics," *International Journal of Molecular Sciences*, vol. 10, no. 12, pp. 5135–5216, 2009.

[105] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3696–3713, 2015.

[106] J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, and B. H. Morrow, "Review of force fields and intermolecular potentials used in atomistic computational materials research," *Applied Physics Reviews*, vol. 5, no. 3, 2018.

[107] P. Kührova, V. Mlynsky, M. Zgarbova, M. Krepl, G. Bussi, R. B. Best, M. Otyepka, J. Sponer, and P. Banas, "Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions," *Journal of Chemical Theory and Computation*, vol. 15, no. 5, pp. 3288–3305, 2019.

[108] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham, "Assessing the Current State of Amber Force Field Modifications for DNA," *Journal of Chemical Theory and Computation*, vol. 12, no. 8, pp. 4114–4127, 2016.

[109] D. A. Case, J. Wang, R. M. Wolf, J. W. Caldwell, and P. A. Kollman, "Development and testing of a general amber force field.," *Journal of Computational Chemistry*, vol. 25, pp. 1157–1174, 2004.

[110] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *Journal of Computational Chemistry*, vol. 31, no. 4, pp. 671–690, 2010.

[111] F. Y. Lin and A. D. MacKerell, "Force Fields for Small Molecules," *Methods in Molecular Biology*, vol. 2022, pp. 21–54, 2019.

[112] P. Li and K. M. Merz, "Metal Ion Modeling Using Classical Mechanics," *Chemical Reviews*, vol. 117, no. 3, pp. 1564–1686, 2017.

[113] F. Fracchia, G. Del Frate, G. Mancini, W. Rocchia, and V. Barone, "Force Field Parametrization of Metal Ions from Statistical Learning Techniques," *Journal of Chemical Theory and Computation*, vol. 14, no. 1, pp. 255–273, 2018.

[114] S. Mamatkulov and N. Schwierz, "Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties," *Journal of Chemical Physics*, vol. 148, no. 7, p. 074504, 2018.

[115] P. Li, L. F. Song, and K. M. Merz, "Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water," *Journal of Physical Chemistry B*, vol. 119, no. 3, pp. 883–895, 2015.

[116] P. Li and K. M. Merz, "Taking into account the ion-induced dipole interaction in the nonbonded model of ions," *Journal of Chemical Theory and Computation*, vol. 10, no. 1, pp. 289–297, 2014.

[117] F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist, and S. C. L. Kamerlin, "Force field independent metal parameters using a nonbonded dummy model," *Journal of Physical Chemistry B*, vol. 118, pp. 4351–4362, 2014.

[118] Q. Liao, A. Pabis, B. Strodel, S. Caroline, and L. Kamerlin, "Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions," *J. Phys. Chem. Lett*, vol. 8, pp. 5408–5414, 2017.

[119] Y.-P. Pang, "Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment," *Journal of Molecular Modeling*, vol. 5, no. 10, pp. 196–202, 1999.

[120] Q. Liao, S. C. L. Kamerlin, and B. Strodel, "Development and Application of a Nonbonded Cu2+ Model That Includes the Jahn-Teller Effect," *Journal of Physical Chemistry Letters*, vol. 6, no. 13, pp. 2657–2662, 2015.

[121] A. Saxena and D. Sept, "Multisite Ion Models That Improve Coordination and Free Energy Calculations in Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, vol. 9, p. 3538−3542, 2013.

[122] G. Parkin, "Synthetic Analogues Relevant to the Structure and Function of Zinc Enzymes," *Chemical Reviews*, vol. 104, no. 2, pp. 699–767, 2004.

[123] R. C. Holz, K. P. Bzymek, and S. I. Swierczek, "Co-catalytic metallopeptidases as pharmaceutical targets," *Current Opinion in Chemical Biology*, vol. 7, no. 2, pp. 197–206, 2003.

[124] K. A. Armacost, S. Riniker, and Z. Cournia, "Novel Directions in Free Energy Methods and Applications," *Journal of Chemical Information and Modeling*, vol. 60, no. 1, pp. 1–5, 2020.

[125] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities," *Expert Opinion on Drug Discovery*, vol. 10, no. 5, pp. 449–461, 2015.

[126] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. Zhang, and T. Hou, "End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design," *Chemical Reviews*, vol. 119, no. 16, pp. 9478–9508, 2019.

[127] H. Sun, Y. Li, S. Tian, L. Xu, and T. Hou, "Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set," *Physical Chemistry Chemical Physics*, vol. 16, no. 31, pp. 16719–16729, 2014.

[128] D. S. Karlov, M. I. Lavrov, V. A. Palyulin, and N. S. Zefirov, "MM-GBSA and MM-PBSA performance in activity evaluation of AMPA receptor positive allosteric modulators," *Journal of Biomolecular Structure and Dynamics*, vol. 36, no. 10, pp. 2508–2516, 2018.

[129] S. I. Virtanen, S. P. Niinivehmas, and O. T. Pentikäinen, "Case-specific performance of MM-PBSA, MM-GBSA, and SIE in virtual screening," *Journal of Molecular Graphics and Modelling*, vol. 62, pp. 303–318, 2015.

[130] H. Lin and D. G. Truhlar, "QM/MM: What have we learned, where are we, and where do we go from here?," *Theoretical Chemistry Accounts*, vol. 117, no. 2, pp. 185–199, 2007.

[131] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review*, vol. 136, no. 3B, p. B864, 1964.

[132] E. Caldeweyher, C. Bannwarth, and S. Grimme, "Extension of the D3 dispersion coefficient model," *Journal of Chemical Physics*, vol. 147, no. 3, p. 034112, 2017.

[133] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, no. 4A, p. A1133, 1965.

[134] R. A. Friesner and V. Guallar, "Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (Qm/Mm) Methods for Studying Enzymatic Catalysis," *Annual Review of Physical Chemistry*, vol. 56, no. 1, pp. 389–427, 2005.

[135] A. V. Arbuznikov, "Hybrid exchange correlation functionals and potentials: Concept elaboration," *Journal of Structural Chemistry*, vol. 48, no. SUPPL. 1, pp. S1–S31, 2007.

[136] K. Kříž and J. Řezáč, "Benchmarking of Semiempirical Quantum-Mechanical Methods on Systems Relevant to Computer-Aided Drug Design," *Journal of Chemical Information and Modeling*, vol. 60, no. 3, pp. 1453–1460, 2020.

[137] J. Kang, Y. Hagiwara, and M. Tateno, "Biological applications of hybrid quantum mechanics/molecular mechanics calculation.," *Journal of biomedicine & biotechnology*, vol. 2012, pp. 1–11, 2012.

[138] V. Kairys and J. H. Jensen, "QM/MM boundaries across covalent bonds: A frozen localized molecular orbital-based approach for the effective fragment potential method," *Journal of Physical Chemistry A*, vol. 104, no. 28, pp. 6656–6665, 2000.

[139] H. M. Senn and W. Thiel, "QM/MM studies of enzymes," *Current Opinion in Chemical Biology*, vol. 11, no. 2, pp. 182–187, 2007.

[140] R. Lonsdale, J. N. Harvey, A. J. Mulholland, and R. Lonsdale, "A practical guide to modelling enzyme-catalysed reactions," *Chem Soc Rev*, vol. 41, pp. 3025–3038, 2012.

[141] S. Hayashi, Y. Uchida, T. Hasegawa, M. Higashi, T. Kosugi, and M. Kamiya, "QM/MM Geometry Optimization on Extensive Free-Energy Surfaces for Examination of Enzymatic Reactions and Design of Novel Functional Properties of Proteins," *Annual Review of Physical Chemistry*, vol. 68, no. 1, pp. 135–154, 2017.

[142] M. W. Van Der Kamp and A. J. Mulholland, "Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology," *Biochemistry*, vol. 52, pp. 2708–2728, 2013.

[143] A. O. Dohn, "Multiscale electrostatic embedding simulations for modeling structure and dynamics of molecules in solution: A tutorial review," *International Journal of Quantum Chemistry*, pp. 1–22, 2020.

[144] K. Schwinn, N. Ferré, and M. Huix-Rotllant, "Analytic QM/MM atomic charge derivatives avoiding the scaling of coupled perturbed equations with the MM subsystem size," *Journal of Chemical Physics*, vol. 151, no. 4, p. 041102, 2019.

[145] M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini, and B. Mennucci, "Polarizable embedding QM/MM: The future gold standard for complex (bio)systems?," *Physical Chemistry Chemical Physics*, vol. 22, no. 26, pp. 14433–14448, 2020.

[146] Z.-x. Chen and A. D. Riggs, "DNA methylation and demethylation in mammals.," *The Journal of biological chemistry*, vol. 286, no. 21, pp. 18347–53, 2011.

[147] C. Martin and Y. Zhang, "The diverse functions of histone lysine methylation," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 11, pp. 838–849, 2005.

[148] J. Cao and Q. Yan, "Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer.," *Frontiers in oncology*, vol. 2, p. 26, 2012.

[149] Z. Dai, V. Ramesh, and J. W. Locasale, "The evolving metabolic landscape of chromatin biology and epigenetics," *Nature Reviews Genetics*, pp. 1–17, 2020.

[150] F. Fiorentino, A. Mai, and D. Rotili, "Lysine Acetyltransferase Inhibitors From Natural Sources," *Frontiers in Pharmacology*, vol. 11, p. 1243, 2020.

[151] H. Rodriguez, H. Rafehi, M. Bhave, and A. El-Osta, "Metabolism and chromatin dynamics in health and disease.," *Molecular aspects of medicine*, vol. 54, pp. 1–15, 2017.

[152] E. E. Hull, M. R. Montgomery, and K. J. Leyva, "HDAC Inhibitors as Epigenetic Regulators of the Immune System: Impacts on Cancer Therapy and Inflammatory Diseases," *BioMed Research International*, vol. 2016, p. 8797206, 2016.

[153] R. B. Parmigiani, W. S. Xu, G. Venta-Perez, H. Erdjument-Bromage, M. Yaneva, P. Tempst, and P. A. Marks, "HDAC6 is a specific deacetylase of peroxiredoxins and is involved in redox regulation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 28, pp. 9633–8, 2008.

[154] M. Haberland, R. L. Montgomery, and E. N. Olson, "The many roles of histone deacetylases in development and physiology: Implications for disease and therapy," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 32–42, 2009.

[155] P. M. Lombardi, K. E. Cole, D. P. Dowling, and D. W. Christianson, "Structure, Mechanism, and Inhibition of Histone Deacetylases and Related Metalloenzymes," *Curr Opin Struct Biol.*, vol. 21, no. 6, pp. 735–743, 2011.

[156] A. b. Uba and K. Yelekçi, "Exploration of the binding pocket of histone deacetylases: The design of potent and isoform-selective inhibitors," *Turkish Journal of Biology*, vol. 41, no. 6, pp. 901–918, 2017.

[157] S. Yoon and G. H. Eom, "HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases," *Chonnam Medical Journal*, vol. 52, no. 1, pp. 1–11, 2016.

[158] K. J. Falkenberg and R. W. Johnstone, "Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders," *Nature Reviews Drug Discovery*, vol. 13, pp. 673–691, 2014.

[159] V. Patil, Q. H. Sodji, J. R. Kornacki, M. Mrksich, and A. K. Oyelere, "3-Hydroxypyridin-2-thione as Novel Zinc Binding Group for Selective Histone Deacetylase Inhibition," *Journal of Medicinal Chemistry*, vol. 56, no. 9, pp. 3492–3506, 2013.

[160] B. S. Mann, J. R. Johnson, K. He, R. Sridhara, S. Abraham, B. P. Booth, L. Verbois, D. E. Morse, J. M. Jee, S. Pope, R. S. Harapanhalli, R. Dagher, A. Farrell, R. Justice, and R. Pazdur, "Vorinostat for treatment of cutaneous manifestations of advanced primary cutaneous T-cell lymphoma.," *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 13, no. 8, pp. 2318–22, 2007.

[161] G. F. Smith, *Designing drugs to avoid toxicity*, vol. 50. Elsevier B.V., 2011.

[162] S. Shen and A. P. Kozikowski, "Why Hydroxamates May Not Be the Best Histone Deacetylase Inhibitors-What Some May Have Forgotten or Would Rather Forget?," *ChemMedChem*, vol. 11, no. 1, pp. 15–21, 2016.

[163] P. Kirsch, A. M. Hartman, A. K. Hirsch, and M. Empting, "Concepts and core principles of fragment-based drug design," *Molecules*, vol. 24, no. 23, p. 4309, 2019.

[164] M. Bissaro, M. Sturlese, and S. Moro, "The rise of molecular simulations in fragment-based drug design (FBDD): an overview," *Drug Discovery Today*, 2020.

[165] P. J. Watson, C. J. Millard, A. M. Riley, N. S. Robertson, L. C. Wright, H. Y. Godage, S. M. Cowley, A. G. Jamieson, B. V. L. Potter, and J. W. R. Schwabe, "Insights into the activation mechanism of class I HDAC complexes by inositol phosphates," *Nature Communications*, vol. 7, p. 11262, 2016.

[166] B. E. Lauffer, R. Mintzer, R. Fong, S. Mukund, C. Tam, I. Zilberleyb, B. Flicke, A. Ritscher, G. Fedorowicz, R. Vallero, D. F. Ortwine, J. Gunzner, Z. Modrusan, L. Neumann, C. M. Koth, J. S. Kaminker, C. E. Heise, and P. Steiner, "Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability," *Journal of Biological Chemistry*, vol. 288, no. 37, pp. 26926–26943, 2013.

[167] P. J. Watson, L. Fairall, G. M. Santos, and J. W. Schwabe, "Structure of HDAC3 bound to co-repressor and inositol tetraphosphate," *Nature*, vol. 481, no. 7381, pp. 335–340, 2012.

[168] M. J. Bottomley, P. Lo Surdo, P. Di Giovine, A. Cirillo, R. Scarpelli, F. Ferrigno, P. Jones, P. Neddermann, R. De Francesco, C. Steinkühler, P. Gallinari, and A. Carfí, "Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain.," *The Journal of biological chemistry*, vol. 283, no. 39, pp. 26694–704, 2008.

[169] Y. Hai and D. W. Christianson, "Histone deacetylase 6 structure and molecular basis of catalysis and inhibition.," *Nature chemical biology*, vol. 12, no. 9, pp. 741–7, 2016.

[170] A. Schuetz, J. Min, A. Allali-Hassani, M. Schapira, M. Shuen, P. Loppnau, R. Mazitschek, N. P. Kwiatkowski, T. A. Lewis, R. L. Maglathin, T. H. McLean, A. Bochkarev, A. N. Plotnikov, M. Vedadi, and C. H. Arrowsmith, "Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity," *Journal of Biological Chemistry*, vol. 283, no. 17, pp. 11355–11363, 2008.

[171] N. J. Porter, N. H. Christianson, C. Decroos, and D. W. Christianson, "Structural and Functional Influence of the Glycine-Rich Loop G302GGGY on the Catalytic Tyrosine of Histone Deacetylase 8," *Biochemistry*, vol. 55, no. 48, pp. 6718–6729, 2016.

[172] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005.

[173] R. W. Bürli, C. A. Luckhurst, O. Aziz, K. L. Matthews, D. Yates, K. A. Lyons, M. Beconi, G. McAllister, P. Breccia, A. J. Stott, S. D. Penrose, M. Wall, M. Lamers, P. Leonard, I. Müller, C. M. Richardson, R. Jarvis, L. Stones, S. Hughes, G. Wishart, A. F. Haughan, C. O'Connell, T. Mead, H. McNeil, J. Vann, J. Mangette, M. Maillard, V. Beaumont, I. Munoz-Sanjuan, and C. Dominguez, "Design, synthesis, and biological evaluation of potent and selective class IIa histone deacetylase (HDAC) inhibitors as a potential therapy for huntington's disease," *Journal of Medicinal Chemistry*, vol. 56, no. 24, pp. 9934–9954, 2013.

[174] A. Pedretti, L. Villa, and G. Vistoli, "VEGA - An open platform to develop chemo-bioinformatics applications, using plug-in architecture and script programming," *Journal of Computer-Aided Molecular Design*, vol. 18, no. 3, pp. 167–173, 2004.

[175] J. J. P. Stewart, "MOPAC2016," 2016.

[176] J. A. Jacobsen, J. L. Fullagar, M. T. Miller, and S. M. Cohen, "Identifying Chelators for Metalloprotein Inhibitors Using a Fragment-Based Approach," *Journal of Medicinal Chemistry*, vol. 54, no. 2, pp. 591–602, 2011.

[177] D. P. Martin, Z. S. Hann, and S. M. Cohen, "Metalloprotein-inhibitor binding: Human carbonic anhydrase II as a model for probing metal-ligand interactions in a metalloprotein active site," *Inorg Chem.*, vol. 52, no. 21, pp. 1–30, 2014.

[178] Z. Hajimahdi and A. Zarghi, "Progress in HIV-1 Integrase Inhibitors : A Review of their Chemical Structure Diversity," *Iranian Journal of Pharmaceutical Research*, vol. 15, no. 4, pp. 595–628, 2016.

[179] K. Chen, L. Xu, and O. Wiest, "Computational Exploration of Zinc Binding Groups for HDAC Inhibition," *The Journal of Organic Chemistry*, vol. 78, no. 10, pp. 5051–5055, 2013.

[180] B. Lecointre, R. Narozny, M. T. Borrello, J. Senger, A. Chakrabarti, M. Jung, M. Marek, C. Romier, J. Melesina, W. Sippl, L. Bischoff, and A. Ganesan, "Isoform-selective HDAC1/6/8 inhibitors with an imidazo-ketopiperazine cap containing stereochemical diversity," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1748, p. 20170364, 2018.

[181] X. Qian, W. J. LaRochelle, G. Ara, F. Wu, K. D. Petersen, A. Thougaard, M. Sehested, H. S. Lichenstein, and M. Jeffers, "Activity of PXD101, a histone deacetylase inhibitor, in preclinical ovarian cancer studies," *Molecular Cancer Therapeutics*, vol. 5, no. 8, pp. 2086–2095, 2006.

[182] C. J. Phiel, F. Zhang, E. Y. Huang, M. G. Guenther, M. A. Lazar, and P. S. Klein, "Histone Deacetylase is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen," *Journal of Biological Chemistry*, vol. 276, no. 39, pp. 36734–36741, 2001.

[183] D. Moffat, S. Patel, F. Day, A. Belfield, A. Donald, M. Rowlands, J. Wibawa, D. Brotherton, L. Stimson, V. Clark, J. Owen, L. Bawden, G. Box, E. Bone, P. Mortenson, A. Hardcastle, S. Van Meurs, S. Eccles, F. Raynaud, and W. Aherne, "Discovery of 2-(6-{[(6-Fluoroquinolin-2-yl)methyl]amino}bicyclo[3.1.0]hex- 3-yl)- N -hydroxypyrimidine-5-carboxamide (CHR-3996), a class i selective orally active histone deacetylase inhibitor," *Journal of Medicinal Chemistry*, vol. 53, no. 24, pp. 8663–8678, 2010.

[184] C.-J. Gong, A.-H. Gao, Y.-M. Zhang, M.-B. Su, F. Chen, L. Sheng, Y.-B. Zhou, J.-Y. Li, J. Li, and F.-J. Nan, "Design, synthesis and biological evaluation of bisthiazole-based trifluoromethyl ketone derivatives as potent HDAC inhibitors with improved cellular efficacy," *European Journal of Medicinal Chemistry*, vol. 112, pp. 81–90, 2016.

[185] I. Paquin, S. Raeppel, S. Leit, F. Gaudette, N. Zhou, O. Moradei, O. Saavedra, N. Bernstein, F. Raeppel, G. Bouchain, S. Fréchette, S. H. Woo, A. Vaisburg, M. Fournel, A. Kalita, M. F. Robert, A. Lu, M. C. Trachy-Bourget, P. T. Yan, J. Liu, J. Rahil, A. R. MacLeod, J. M. Besterman, Z. Li, and D. Delorme, "Design and synthesis of 4-[(s-triazin-2-ylamino)methyl]-N-(2-aminophenyl)-benzamides and their analogues as a novel class of histone deacetylase inhibitors," *Bioorganic and Medicinal Chemistry Letters*, vol. 18, no. 3, pp. 1067–1071, 2008.

[186] T. Tashima, H. Murata, and H. Kodama, "Design and synthesis of novel and highly-active pan-histone deacetylase (pan-HDAC) inhibitors," *Bioorganic and Medicinal Chemistry*, vol. 22, no. 14, pp. 3720–3731, 2014.

[187] C. B. Botta, W. Cabri, E. Cini, L. De Cesare, C. Fattorusso, G. Giannini, M. Persico, A. Petrella, F. Rondinelli, M. Rodriquez, A. Russo, and M. Taddei, "Oxime amides as a novel zinc binding group in histone deacetylase inhibitors: Synthesis, biological activity, and computational evaluation," *Journal of Medicinal Chemistry*, vol. 54, no. 7, pp. 2165–2182, 2011.

[188] C. Chen, X. Hou, G. Wang, W. Pan, X. Yang, Y. Zhang, and H. Fang, "Design, synthesis and biological evaluation of quinoline derivatives as HDAC class I inhibitors," *European Journal of Medicinal Chemistry*, vol. 133, pp. 11–23, 2017.

[189] R. Cincinelli, V. Zwick, L. Musso, V. Zuco, M. De Cesare, F. Zunino, C. Simoes-Pires, A. Nurisso, G. Giannini, M. Cuendet, and S. Dallavalle, "Biphenyl-4-yl-acrylohydroxamic acids: Identification of a novel indolyl-substituted HDAC inhibitor with antitumor activity," *European Journal of Medicinal Chemistry*, vol. 112, pp. 99–105, 2016.

[190] W. Duan, J. Li, E. S. Inks, C. J. Chou, Y. Jia, X. Chu, X. Li, W. Xu, and Y. Zhang, "Design, Synthesis, and Antitumor Evaluation of Novel Histone Deacetylase Inhibitors Equipped with a Phenylsulfonylfuroxan Module as a Nitric Oxide Donor," *Journal of Medicinal Chemistry*, vol. 58, no. 10, pp. 4325–4338, 2015.

[191] S. Belvedere, D. J. Witter, J. Yan, J. P. Secrist, V. Richon, and T. A. Miller, "Aminosuberoyl hydroxamic acids (ASHAs): A potent new class of HDAC inhibitors," *Bioorganic and Medicinal Chemistry Letters*, vol. 17, no. 14, pp. 3969–3971, 2007.

[192] S. Massa, A. Mai, G. Sbardella, M. Esposito, R. Ragno, P. Loidl, and G. Brosch, "3-(4-Aroyl-1H-pyrrol-2-yl)-N-hydroxy-2-propenamides, a new class of synthetic histone deacetylase inhibitors," *Journal of Medicinal Chemistry*, vol. 44, no. 13, pp. 2069–2072, 2001.

[193] K. Mahal, P. Kahlen, B. Biersack, and R. Schobert, "4-(1-Ethyl-4-anisyl-imidazol-5-yl)-N-hydroxycinnamide – A new pleiotropic HDAC inhibitor targeting cancer cell signalling and cytoskeletal organisation," *Experimental Cell Research*, vol. 336, no. 2, pp. 263–275, 2015.

[194] J. D. Ellis, D. A. Neil, N. G. Inston, E. Jenkinson, M. T. Drayson, P. Hampson, S. J. Shuttleworth, A. R. Ready, and M. Cobbold, "Inhibition of histone deacetylase 6 reveals a potent immunosuppressant effect in models of transplantation," *Transplantation*, vol. 100, no. 8, pp. 1667–1674, 2016.

[195] J. Arts, P. King, A. Mariën, W. Floren, A. Beliën, L. Janssen, I. Pilatte, B. Roux, L. Decrane, R. Gilissen, I. Hickson, V. Vreys, E. Cox, K. Bol, W. Talloen, I. Goris, L. Andries, M. Du Jardin, M. Janicot, M. Page, K. Van Emelen, and P. Angibaud, "JNJ-26481585, a novel "second-generation" oral histone deacetylase inhibitor, shows broad-spectrum preclinical antitumoral activity," *Clinical Cancer Research*, vol. 15, no. 22, pp. 6841–6851, 2009.

[196] M. Rai, E. Soragni, C. J. Chou, G. Barnes, S. Jones, J. R. Rusche, J. M. Gottesfeld, and M. Pandolfo, "Two new pimelic diphenylamide HDAC inhibitors induce sustained frataxin upregulation in cells from Friedreich's ataxia patients and in a mouse model," *PLoS ONE*, vol. 5, no. 1, p. e8825, 2010.

[197] Z. Yang, T. Wang, F. Wang, T. Niu, Z. Liu, X. Chen, C. Long, M. Tang, D. Cao, X. Wang, W. Xiang, Y. Yi, L. Ma, J. You, and L. Chen, "Discovery of Selective Histone Deacetylase 6 Inhibitors Using the Quinazolin as the Cap for the Treatment of Cancer," *Journal of Medicinal Chemistry*, vol. 59, p. 1455−1470, 2016.

[198] J. Jiao, H. Fang, X. Wang, P. Guan, Y. Yuan, and W. Xu, "Design, synthesis and preliminary biological evaluation of N-hydroxy-4-(3-phenylpropanamido)benzamide (HPPB) derivatives as novel histone deacetylase inhibitors," *European Journal of Medicinal Chemistry*, vol. 44, no. 11, pp. 4470–4476, 2009.

[199] D. M. Quinn, "Acetylcholinesterase: Enzyme Structure, Reaction Dynamics, and Virtual Transition States," *Chemical Reviews*, vol. 87, no. 5, pp. 955–979, 1987.

[200] Y. Xu, S. Cheng, J. L. Sussman, I. Silman, and H. Jiang, "Computational studies on acetylcholinesterases," 2017.

[201] Y. Bourne, P. Taylor, Z. Radić, and P. Marchot, "Structural insights into ligand interactions at the acetylcholinesterase peripheral anionic site," *EMBO Journal*, vol. 22, no. 1, pp. 1–12, 2003.

[202] W. Booth, "British lab confirms Novichok was used in spy attack but can't determine source," *The Washington Post*, 2018.

[203] A. B. Kingsley, P.; Barnard, "Banned Nerve Agent Sarin Used in Syria Chemical Attack, Turkey Says," *The New York Times*, 2017.

[204] N. Paddock, R. C.; Sang-Hun, C.; Wade, "In Kim Jong-nam's Death, North Korea Lets Loose a Weapon of Mass Destruction," *The New York Times*, 2017.

[205] OPCW, "Organisation for the Prohibition of Chemical Weapons."

[206] S. Suratman, J. W. Edwards, and K. Babina, "Organophosphate pesticides exposure among farmworkers: Pathways and risk of adverse health effects," *Reviews on Environmental Health*, vol. 30, no. 1, pp. 65–79, 2015.

[207] M. B. Colovic, D. Z. Krstic, T. D. Lazarevic-Pasti, A. M. Bondzic, and V. M. Vasic, "Acetylcholinesterase Inhibitors: Pharmacology and Toxicology," *Current Neuropharmacology*, vol. 11, no. 3, pp. 315–335, 2013.

[208] G. S. Sirin, Y. Zhou, L. Lior-Hoffmann, S. Wang, and Y. Zhang, "Aging mechanism of soman inhibited acetylcholinesterase," *Journal of Physical Chemistry B*, vol. 116, no. 40, pp. 12199–12207, 2012.

[209] Q. Zhuang, A. Young, C. S. Callam, C. A. McElroy, Ö. D. Ekici, R. J. Yoder, and C. M. Hadad, "Efforts toward treatments against aging of organophosphorus-inhibited acetylcholinesterase," *Annals of the New York Academy of Sciences*, vol. 1374, no. 1, pp. 94–104, 2016.

[210] L. Wong, Z. Radić, R. J. Brüggemann, N. Hosea, H. A. Berman, and P. Taylor, "Mechanism of oxime reactivation of acetylcholinesterase analyzed by chirality and mutagenesis," *Biochemistry*, vol. 39, no. 19, pp. 5750–5757, 2000.

[211] M. N. Faiz Norrrahim, M. A. Idayu Abdul Razak, N. A. Ahmad Shah, H. Kasim, W. Y. Wan Yusoff, N. A. Halim, S. A. Mohd Nor, S. H. Jamal, K. K. Ong, W. M. Zin Wan Yunus, V. F. Knight, and N. A. Mohd Kasim, "Recent developments on oximes to improve the blood brain barrier penetration for the treatment of organophosphorus poisoning: A review," *RSC Advances*, vol. 10, no. 8, pp. 4465–4489, 2020.

[212] L. Gorecki, J. Korabecny, K. Musilek, D. Malinak, E. Nepovimova, R. Dolezal, D. Jun, O. Soukup, and K. Kuca, "SAR study to find optimal cholinesterase reactivator against organophosphorous nerve agents and pesticides," *Archives of Toxicology*, vol. 90, no. 12, pp. 2831–2859, 2016.

[213] K. Kuca, D. Jun, and K. Musilek, "Structural Requirements of Acetylcholinesterase Reactivators," *Mini-Reviews in Medicinal Chemistry*, vol. 6, no. 3, pp. 269–277, 2006.

[214] S. Bolton and A. Beckett, "Metal chelates as potential reactivators of organic phosphate poisoned acetylchlinesterase," *Journal of Pharmaceutical Sciences*, vol. 53, no. 1, pp. 55–60, 1964.

[215] R. Breslow and D. Chipman, "Mixed Metal Complexes as Enzyme Models. I. Intracomplex Nucleophilic Catalysis by an Oxime Anion," *Journal of the American Chemical Society*, vol. 87, no. 18, pp. 4195–4196, 1965.

[216] R. Breslow and L. E. Overman, "An "Artificial Enzyme" Combining a Metal Catalytic Group and a Hydrophobic Binding Cavity," *Journal of the American Chemical Society*, vol. 92, no. 4, pp. 1075–1077, 1970.

[217] K. F. Konidaris, G. A. Dalkas, E. Katsoulakou, G. Pairas, C. P. Raptopoulou, F. N. Lamari, G. A. Spyroulias, and E. Manessi-Zoupa, "ZnII/pyridyloxime complexes as potential reactivators of OP-inhibited acetylcholinesterase: In vitro and docking simulation studies," *Journal of Inorganic Biochemistry*, vol. 134, pp. 12–19, 2014.

[218] I. Silman and J. L. Sussman, "Acetylcholinesterase: How is structure related to function?," *Chemico-Biological Interactions*, vol. 175, pp. 3–10, 2008.

[219] A. S. Kiametis, M. A. Silva, L. A. Romeiro, J. B. Martins, and R. Gargano, "Potential acetylcholinesterase inhibitors: molecular docking, molecular dynamics, and in silico prediction," *Journal of Molecular Modeling*, vol. 23, no. 2, p. 67, 2017.

[220] S. Cheng, W. Song, X. Yuan, and Y. Xu, "Gorge Motions of Acetylcholinesterase Revealed by Microsecond Molecular Dynamics Simulations," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.

[221] M. Korth, M. Pitoňák, J. Řezáč, and P. Hobza, "A transferable H-bonding correction for semiempirical quantum-chemical methods," *Journal of Chemical Theory and Computation*, vol. 6, no. 1, pp. 344–352, 2010.

[222] J. J. P. Stewart, "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters," *J Mol Model*, vol. 19, pp. 1–32, 2013.

[223] A. Klamt and G. Schüürmann, "COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *Journal of the Chemical Society, Perkin Transactions 2*, no. 5, pp. 799–805, 1993.

[224] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, M. C. X. Li, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, and H. P. Hratchian, "Gaussian 16, Revision A.03," 2016.

[225] J. Tirado-Rives and W. L. Jorgensen, "Performance of B3LYP density functional methods for a large set of organic molecules," *Journal of Chemical Theory and Computation*, vol. 4, no. 2, pp. 297–306, 2008.

[226] A. K. Singha Deb, P. Sahu, A. Boda, S. M. Ali, K. T. Shenoy, and D. Upadhyay, "DFT and MD simulation supplemented experiments for isotopic fractionation of zinc compounds using a macrocyclic crown ether appended polymeric resin," *Physical chemistry chemical physics : PCCP*, vol. 22, no. 26, p. 14682, 2020.

[227] S. M. Wagoner, M. Deeconda, K. L. Cumpian, R. Ortiz, S. Chinthala, and L. A. Angel, "The multiple conformational charge states of zinc(II) coordination by 2His-2Cys oligopeptide investigated by ion mobility-mass spectrometry, density functional theory and theoretical collision cross sections," *Journal of Mass Spectrometry*, vol. 51, no. 12, pp. 1120–1129, 2016.

[228] N. Díaz, D. Suárez, and K. M. Merz, "Hydration of zinc ions: Theoretical study of [Zn(H 2 O) 4](H 2 O) 82+ and [Zn(H 2 O) 6](H 2 O) 62+," *Chemical Physics Letters*, vol. 326, no. 3-4, pp. 288–292, 2000.

[229] A. C. Ekennia, D. C. Onwudiwe, L. O. Olasunkanmi, A. A. Osowole, and E. E. Ebenso, "Synthesis, DFT Calculation, and Antimicrobial Studies of Novel Zn(II), Co(II), Cu(II), and Mn(II) Heteroleptic Complexes Containing Benzoylacetone and Dithiocarbamate," *Bioinorg Chem Appl.*, vol. 2015, pp. 1–12, 2015.

[230] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *Journal of Chemical Physics*, vol. 132, no. 15, p. 24101, 2010.

[231] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical Reviews*, vol. 105, no. 8, pp. 2999–3093, 2005.

[232] F. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, "The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building," *Physical Chemistry Chemical Physics*, vol. 12, no. 28, pp. 7821–7839, 2010.

[233] M. C. Franklin, M. J. Rudolph, C. Ginter, M. S. Cassidy, and J. Cheung, "Structures of paraoxon-inhibited human acetylcholinesterase reveal perturbations of the acyl loop and the dimer interface," *Proteins: Structure, Function, and Bioinformatics*, vol. 84, no. 9, pp. 1246–1256, 2016.

[234] J. Newman, "A review of techniques for maximizing diffraction from a protein crystal in stilla," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 27–31, 2006.

[235] I. W. Mcnae, D. Kan, G. Kontopidis, A. Patterson, P. Taylor, L. Worrall, and M. D. Walkinshaw, "Studying protein-ligand interactions using protein crystallography," *Crystallography Reviews*, vol. 11, no. 1, pp. 61–71, 2005.

[236] I. Müller, "Guidelines for the successful generation of protein-ligand complex crystals," *Acta Crystallographica Section D: Structural Biology*, vol. 73, no. 2, pp. 79–92, 2017.

[237] F. Ekström, A. Hörnberg, E. Artursson, L. G. Hammarström, G. Schneider, and Y. P. Pang, "Structure of HI-6-sarin-acetylcholinesterase determined by x-ray crystallography and molecular dynamics simulation: Reactivator mechanism and design," *PLoS ONE*, vol. 4, no. 6, 2009.

[238] B. Sanson, F. Nachon, J. P. Colletier, M. T. Froment, L. Toker, H. M. Greenblatt, J. L. Sussman, Y. Ashani, P. Masson, I. Silman, and M. Weik, "Crystallographic snapshots of nonaged and aged conjugates of soman with acetylcholinesterase, and of a ternary complex of the aged conjugate with pralidoxime," *Journal of Medicinal Chemistry*, vol. 52, no. 23, pp. 7593–7603, 2009.

[239] S. M. Bester, M. A. Guelta, J. Cheung, M. D. Winemiller, S. Y. Bae, J. Myslinski, S. D. Pegan, and J. J. Height, "Structural Insights of Stereospecific Inhibition of Human Acetylcholinesterase by VX and Subsequent Reactivation by HI-6," *Chemical Research in Toxicology*, vol. 31, no. 12, pp. 1405–1417, 2018.

[240] O. Gerlits, X. Kong, X. Cheng, T. Wymore, D. K. Blumenthal, P. Taylor, Z. Radić, and A. Kovalevsky, "Productive reorientation of a bound oxime reactivator revealed in room temperature X-ray structures of native and VX-inhibited human acetylcholinesterase," *Journal of Biological Chemistry*, vol. 294, no. 27, pp. 10607–10618, 2019.

[241] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev, "H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations," *Nucleic Acids Research*, vol. 40, pp. 537–541, 2012.

[242] T. Driant, F. Nachon, C. Ollivier, P.-Y. Renard, and E. Derat, "On the Influence of the Protonation States of Active Site Residues on AChE Reactivation: A QM/MM Approach," *ChemBioChem*, vol. 18, no. 7, pp. 666–675, 2017.

[243] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *Journal of Molecular Biology*, vol. 267, pp. 727–748, 1997.

[244] Y. Li, L. Han, Z. Liu, and R. Wang, "Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results," *Journal of Chemical Information and Modeling*, vol. 54, no. 6, pp. 1717–1736, 2014.

[245] A. A. Adeniyi and M. E. Soliman, "Implementing QM in docking calculations: is it a waste of computational time?," *Drug Discovery Today*, vol. 22, no. 8, pp. 1216–1223, 2017.

[246] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the Amber biomolecular simulation package," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 2, pp. 198–210, 2013.

[247] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, "MMPBSA.py: An efficient program for end-state free energy calculations," *Journal of Chemical Theory and Computation*, vol. 8, no. 9, pp. 3314–3321, 2012.

[248] A. Onufriev, D. Bashford, and D. A. Case, "Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model," *Proteins: Structure, Function and Genetics*, vol. 55, no. 2, pp. 383–394, 2004.

[249] R. L. M. Gieseking, M. A. Ratner, and G. C. Schatz, "Benchmarking Semiempirical Methods To Compute Electrochemical Formal Potentials," *J. Phys. Chem. A*, vol. 122, p. 6809−6818 Article, 2018.

[250] Y. Minenkov, D. I. Sharapa, and L. Cavallo, "Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy," *J. Chem. Theory Comput.*, vol. 14, p. 3428−3439, 2018.

[251] M. Nakata, T. Shimazaki, M. Hashimoto, and T. Maeda, "PubChemQC PM6: A dataset of 221 million molecules with optimized molecular geometries and electronic properties." 2019.

[252] Y. Pourshojaei, A. Abiri, K. Eskandari, Z. Haghighijoo, N. Edraki, and A. Asadipour, "Phenoxyethyl Piperidine/Morpholine Derivatives as PAS and CAS Inhibitors of Cholinesterases: Insights for Future Drug Design," *Scientific Reports*, vol. 9, no. 1, pp. 1–19, 2019.

[253] Z. Radic, D. M. Quinn, D. C. Vellom, S. Camp, and P. Taylor, "Allosteric control of acetylcholinesterase catalysis by fasciculin," *Journal of Biological Chemistry*, vol. 270, no. 35, pp. 20391–20399, 1995.

[254] A. Ranjan, A. Chauhan, and T. Jindal, "In-silico and in-vitro evaluation of human acetylcholinesterase inhibition by organophosphates," *Environmental Toxicology and Pharmacology*, vol. 57, pp. 131–140, 2018.

[255] A. Ranjan, S. Ghosh, A. Chauhan, and T. Jindal, "Molecular Docking and Site Directed Mutagenic Approach To Investigate The Role Of Trp86 Of Human Acetylcholinesterase With Organophosphates," *International Journal of Pharmaceutical Sciences and Research*, vol. 7, no. 9, pp. 3802–3809, 2016.

[256] A. Ranjan, N. Patil, A. Chauhan, and T. Jindal, "Mechanistic of Organophosphate Mediated Inhibition of Human Acetylcholinesterase by Molecular Docking," *J Bioinform, Genomics, Proteomics*, vol. 3, no. 2, p. 1032, 2018.

[257] A. Asghar, M. Yousuf, G. Fareed, R. Nazir, A. Hassan, A. Maalik, T. Noor, N. Iqbal, and L. Rasheed, "Synthesis, acetylcholinesterase (AChE) and butyrylcholinesterase (BuChE) activities, and molecular docking studies of a novel compound based on combination of flurbiprofen and isoniazide," *RSC Advances*, vol. 10, no. 33, pp. 19346–19352, 2020.

[258] Z. Jin, C. Zhang, M. Liu, S. Jiao, J. Zhao, X. Liu, H. Lin, D. Chi-cheong Wan, and C. Hu, "Synthesis, biological activity, molecular docking studies of a novel series of 3-Aryl-7H-thiazolo[3,2-b]-1,2,4-triazin-7-one derivatives as the acetylcholinesterase inhibitors," *Journal of Biomolecular Structure and Dynamics*, vol. 0, no. 0, pp. 1–12, 2020.

[259] A. Ranjan, A. Kumar, K. Gulati, S. Thakur, and T. Jindal, "Role of Aromatic Amino Acids in Stabilizing Organophosphate and Human Acetylcholinesterase Complex," *Journal of Current Pharma Research*, vol. 5, no. 4, pp. 1632–1639, 2015.

[260] V. Salmaso and S. Moro, "Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview," *Frontiers in Pharmacology*, vol. 9, p. 923, 2018.

[261] F. S. Katz, S. Pecic, T. H. Tran, I. Trakht, L. Schneider, Z. Zhu, L. Ton-That, M. Luzac, V. Zlatanic, S. Damera, J. MacDonald, D. W. Landry, L. Tong, and M. N. Stojanovic, "Discovery of New Classes of Compounds that Reactivate Acetylcholinesterase Inhibited by Organophosphates," *ChemBioChem*, vol. 16, no. 15, pp. 2205–2215, 2015.

[262] K. Chmielewska, K. Dzierzbicka, I. Inkielewicz-Stępniak, and M. Przybyłowska, "Therapeutic Potential of Carnosine and Its Derivatives in the Treatment of Human Diseases," *Chem. Res. Toxicol.*, vol. 33, p. 1561−1578, 2020.

[263] L. Carroll, A. Karton, L. Radom, M. J. Davies, and D. I. Pattison, "Carnosine and Carcinine Derivatives Rapidly React with Hypochlorous Acid to Form Chloramines and Dichloramines," *Chemical Research in Toxicology*, vol. 32, no. 3, pp. 513–525, 2019.

[264] V. D. Prokopieva, E. G. Yarygina, N. A. Bokhan, S. A. Ivanova, and S. D. Meo, "Use of Carnosine for Oxidative Stress Reduction in Different Pathologies," *Oxidative Medicine and Cellular Longevity*, vol. 2016, pp. 1–8, 2016.

[265] A. A. Boldyrev, G. Aldini, and W. Derave, "Physiology and pathophysiology of carnosine," *Physiological Reviews*, vol. 93, no. 4, pp. 1803–1845, 2013.

[266] W. Derave, B. De Courten, and S. P. Baba, "An update on carnosine and anserine research," *Amino Acids*, vol. 51, no. 1, pp. 1–4, 2019.

[267] B. E. Brown, C. H. Kim, F. R. Torpy, C. A. Bursill, L. S. McRobb, A. K. Heather, M. J. Davies, and D. M. Van Reyk, "Supplementation with carnosine decreases plasma triglycerides andmodulates atherosclerotic plaque composition in diabetic apoE-/- mice," *Atherosclerosis*, vol. 232, no. 2, pp. 403–409, 2014.

[268] S. Menini, C. Iacobini, C. Ricci, A. Scipioni, C. B. Fantauzzi, A. Giaccari, E. Salomone, R. Canevotti, A. Lapolla, M. Orioli, G. Aldini, and G. Pugliese, "D-carnosine octylester attenuates atherosclerosis and renal disease in ApoE null mice fed a Western diet through reduction of carbonyl stress and inflammation," *British Journal of Pharmacology*, vol. 166, no. 4, pp. 1344–1356, 2012.

[269] T. Albrecht, M. Schilperoort, S. Zhang, J. D. Braun, J. Qiu, A. Rodriguez, D. O. Pastene, B. K. Krämer, H. Köppel, H. Baelde, E. De Heer, A. A. Altomare, L. Regazzoni, A. Denisi, G. Aldini, J. Van Den Born, B. A. Yard, and S. J. Hauske, "Carnosine attenuates the development of both type 2 diabetes and diabetic nephropathy in BTBR ob/ob mice," *Scientific Reports*, vol. 7, no. 1, pp. 1–16, 2017.

[270] G. Aldini, M. Orioli, G. Rossoni, F. Savi, P. Braidotti, G. Vistoli, K. J. Yeum, G. Negrisoli, and M. Carini, "The carbonyl scavenger carnosine ameliorates dyslipidaemia and renal function in Zucker obese rats," *Journal of Cellular and Molecular Medicine*, vol. 15, no. 6, pp. 1339–1354, 2011.

[271] N. A. Al-Sawalha, O. Y. Alshogran, M. S. Awawdeh, and B. A. Almomani, "The effects of L-Carnosine on development of metabolic syndrome in rats," *Life Sciences*, vol. 237, p. 116905, 2019.

[272] M. Schön, A. Mousa, M. Berk, W. L. Chia, J. Ukropec, A. Majid, B. Ukropcová, and B. De Courten, "The potential of carnosine in brain-related disorders: A comprehensive review of current evidence," *Nutrients*, vol. 11, no. 6, p. 1196, 2019.

[273] M. Teufel, V. Saudek, J. P. Ledig, A. Bernhardt, S. Boularand, A. Carreau, N. J. Cairns, C. Carter, D. J. Cowley, D. Duverger, A. J. Ganzhorn, C. Guenet, B. Heintzelmann, V. Laucher, C. Sauvage, and T. Smirnova, "Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase," *Journal of Biological Chemistry*, vol. 278, no. 8, pp. 6521–6531, 2003.

[274] P. Schoen, H. Everts, T. De Boer, and W. Van Oeveren, "Serum Carnosinase Activity in Plasma and Serum: Validation of a Method and Values in Cardiopulmonary Bypass Surgery," *Clinical Chemistry*, vol. 49, no. 11, pp. 1930–1932, 2003.

[275] K. G. Santos, "The (CTG) n repeat polymorphism in CNDP1 gene: New insights into an old molecule," *Indian Journal of Medical Research*, vol. 144, pp. 6–8, 2016.

[276] B. Janssen, D. Hohenadel, P. Brinkkoetter, V. Peters, N. Rind, C. Fischer, I. Rychlik, M. Cerna, M. Romzova, E. De Heer, H. Baelde, S. J. Bakker, M. Zirie, E. Rondeau, P. Mathieson, M. A. Saleem, J. Meyer, H. Koppel, S. Sauerhoefer, C. R. Bartram, P. Nawroth, H. P. Hammes, B. A. Yard, J. Zschocke, and F. J. Van Der Woude, "Carnosine as a protective factor in diabetic nephropathy: Association with a leucine repeat of the carnosinase gene CNDP1," *Diabetes*, vol. 54, no. 8, pp. 2320–2327, 2005.

[277] S. Stegen, B. Stegen, G. Aldini, A. Altomare, L. Cannizzaro, M. Orioli, S. Gerlo, L. Deldicque, M. Ramaekers, P. Hespel, and W. Derave, "Plasma carnosine, but not muscle carnosine, attenuates high-fat diet-induced metabolic stress," *Applied Physiology, Nutrition and Metabolism*, vol. 40, no. 9, pp. 868–876, 2015.

[278] E. J. Anderson, G. Vistoli, L. A. Katunga, K. Funai, L. Regazzoni, T. Blake Monroe, E. Gilardoni, L. Cannizzaro, M. Colzani, D. De Maddis, G. Rossoni, R. Canevotti, S. Gagliardi, M. Carini, and G. Aldini, "A carnosine analog mitigates metabolic disorders of obesity by reducing carbonyl stress," *Journal of Clinical Investigation*, vol. 128, no. 12, pp. 5280–5293, 2018.

[279] M. Orioli, G. Vistoli, L. Regazzoni, A. Pedretti, A. Lapolla, G. Rossoni, R. Canevotti, L. Gamberoni, M. Previtali, M. Carini, and G. Aldini, "Design, Synthesis, ADME Properties, and Pharmacological Activities of β-Alanyl-D-histidine (D-Carnosine) Prodrugs with Improved Bioavailability," *ChemMedChem*, vol. 6, no. 7, pp. 1269–1282, 2011.

[280] J. Qiu, S. J. Hauske, S. Zhang, A. Rodriguez-Niño, T. Albrecht, D. O. Pastene, J. van den Born, H. van Goor, S. Ruf, M. Kohlmann, M. Teufel, B. K. Krämer, H. P. Hammes, V. Peters, B. A. Yard, and A. Kannt, "Identification and characterisation of carnostatine (SAN9812), a potent and selective carnosinase (CN1) inhibitor with in vivo activity," *Amino Acids*, vol. 51, no. 1, pp. 7–16, 2019.

[281] O. I. Kulikova, S. L. Stvolinsky, V. A. Migulin, L. A. Andreeva, I. Y. Nagaev, O. M. Lopacheva, K. N. Kulichenkova, A. V. Lopachev, I. E. Trubitsina, and T. N. Fedorova, "A new derivative of acetylsalicylic acid and carnosine: synthesis, physical and chemical properties, biological activity," *DARU, Journal of Pharmaceutical Sciences*, vol. 28, no. 1, pp. 119–130, 2020.

[282] A. Pegova, H. Abe, and A. Boldyrev, "Hydrolysis of carnosine and related compounds by mammalian carnosinases," *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, vol. 127, no. 4, pp. 443–446, 2000.

[283] V. Lanza, F. Bellia, R. D'Agata, G. Grasso, E. Rizzarelli, and G. Vecchio, "New glycoside derivatives of carnosine and analogs resistant to carnosinase hydrolysis: Synthesis and characterization of their copper(II) complexes," *Journal of Inorganic Biochemistry*, vol. 105, no. 2, pp. 181–188, 2011.

[284] V. Peters, C. P. Schmitt, T. Weigand, K. Klingbeil, C. Thiel, A. van den Berg, V. Calabrese, P. Nawroth, T. Fleming, E. Forsberg, A. H. Wagner, M. Hecker, and G. Vistoli, "Allosteric inhibition of carnosinase (CN1) by inducing a conformational shift," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 32, no. 1, pp. 1102–1110, 2017.

[285] V. Peters, E. E. Jansen, C. Jakobs, E. Riedl, B. Janssen, B. A. Yard, J. Wedel, G. F. Hoffmann, J. Zschocke, D. Gotthardt, C. Fischer, and H. Köppel, "Anserine inhibits carnosine degradation but in human serum carnosinase (CN1) is not correlated with histidine dipeptide concentration," *Clinica Chimica Acta*, vol. 412, no. 3-4, pp. 263–267, 2011.

[286] V. Peters, M. Kebbewar, E. W. Jansen, C. Jakobs, E. Riedl, H. Koeppel, D. Frey, K. Adelmann, K. Klingbeil, M. MacK, G. F. Hoffmann, B. Janssen, J. Zschocke, and B. A. Yard, "Relevance of allosteric conformations and homocarnosine concentration on carnosinase activity," *Amino Acids*, vol. 38, no. 5, pp. 1607–1615, 2010.

[287] C. Bebrone, H. Delbrück, M. B. Kupper, P. Schlömer, C. Willmann, J. M. Frère, R. Fischer, M. Galleni, and K. M. Hoffmann, "The structure of the dizinc subclass B2 metallo-β-lactamase CphA reveals that the second inhibitory zinc ion binds in the histidine site," *Antimicrobial Agents and Chemotherapy*, vol. 53, no. 10, pp. 4464–4471, 2009.

[288] I. García-Sáez, P. S. Mercuri, C. Papamicael, R. Kahn, J. M. Frère, M. Galleni, G. M. Rossolini, and O. Dideberg, "Three-dimensional structure of FEZ-1, a monomeric subclass B3 metallo-β-lactamase from Fluoribacter gormanii, in native form and in complex with D-captopril," *Journal of Molecular Biology*, vol. 325, no. 4, pp. 651–660, 2003.

[289] S. Rowsell, R. A. Pauptit, A. D. Tucker, R. G. Melton, D. M. Blow, and P. Brick, "Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy," *Structure*, vol. 5, no. 3, pp. 337–347, 1997.

[290] S. Lundgren, B. Andersen, J. Piškur, and D. Dobritzsch, "Crystal structures of yeast β-alanine synthase complexes reveal the mode of substrate binding and large scale domain closure movements," *Journal of Biological Chemistry*, vol. 282, no. 49, pp. 36037–36047, 2007.

[291] R. R. Crichton, "Zinc – Lewis Acid and Gene Regulator," in *Biological Inorganic Chemistry* (Elsevier B.V. All, ed.), pp. 229–246, Elsevier, 2012.

[292] P. Hinchliffe, C. A. Tanner, A. P. Krismanich, G. Labbé, V. J. Goodfellow, L. Marrone, A. Y. Desoky, K. Calvopiña, E. E. Whittle, F. Zeng, M. B. Avison, N. C. Bols, S. Siemann, J. Spencer, and G. I. Dmitrienko, "Structural and Kinetic Studies of the Potent Inhibition of Metallo-β-lactamases by 6-Phosphonomethylpyridine-2-carboxylates," *Biochemistry*, vol. 57, no. 12, pp. 1880–1892, 2018.

[293] P. Oelschlaeger, R. D. Schmid, and J. È. Rgen Pleiss, "Insight into the mechanism of the IMP-1 metallo-β-lactamase by molecular dynamics simulations," *Protein Engineering, Design and Selection*, vol. 16, no. 5, pp. 341–350, 2003.

[294] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "The carboxylate shift in zinc enzymes: A computational study," *Journal of the American Chemical Society*, vol. 129, no. 5, pp. 1378–1385, 2007.

[295] N. Drici and M. A. Krallafa, "Effect of mutation on the stabilization energy of HIV-1 zinc fingers: a hybrid local self-consistent field/molecular mechanics investigation," *Journal of Biological Inorganic Chemistry*, vol. 22, no. 1, pp. 109–119, 2017.

[296] G. Frison and G. Ohanessian, "A comparative study of semiempirical, ab initio, and DFT methods in evaluating metal-ligand bond strength, proton affinity, and interactions between first and second shell ligands in Zn-biomimetic complexes," *Journal of Computational Chemistry*, vol. 29, pp. 416–433, 2008.

[297] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: a comprehensive resource for in silico drug discovery and exploration.," *Nucleic acids research*, vol. 34, pp. D668–72, 2006.

[298] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed, "Drug repurposing: Progress, challenges and recommendations," *Nature Reviews Drug Discovery*, vol. 18, no. 1, pp. 41–58, 2018.

[299] P. Polamreddy and N. Gattu, "The drug repurposing landscape from 2012 to 2017: evolution, challenges, and possible solutions," *Drug Discovery Today*, vol. 24, no. 3, pp. 789–795, 2019.

[300] H. S. GNS, S. GR, M. Murahari, and M. Krishnamurthy, "An update on Drug Repurposing: Re-written saga of the drug's fate," *Biomedicine and Pharmacotherapy*, vol. 110, pp. 700–716, 2019.

[301] R. Kumar, S. Harilal, S. V. Gupta, J. Jose, D. G. Thomas, M. S. Uddin, M. A. Shah, and B. Mathew, "Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work," *European Journal of Medicinal Chemistry*, vol. 182, p. 111602, 2019.

[302] R. S. Jansen, R. Addie, R. Merkx, A. Fish, S. Mahakena, O. B. Bleijerveld, M. Altelaar, L. IJlst, R. J. Wanders, P. Borst, and K. Van De Wetering, "N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 21, pp. 6601–6606, 2015.

[303] B. Shen, "A New Golden Age of Natural Products Drug Discovery," *Cell*, vol. 163, no. 6, pp. 1297–1300, 2015.

[304] N. E. Thomford, D. A. Senthebane, A. Rowe, D. Munro, P. Seele, A. Maroyi, and K. Dzobo, "Natural products for drug discovery in the 21st century: Innovations for novel drug discovery," *International Journal of Molecular Sciences*, vol. 19, no. 6, p. 1578, 2018.

[305] Q. Luo, L. Zhao, J. Hu, H. Jin, Z. Liu, and L. Zhang, "The scoring bias in reverse docking and the score normalization strategy to improve success rate of target fishing," *PLoS ONE*, vol. 12, no. 2, p. e0171433, 2017.

[306] I. Everaert, G. Baron, S. Barbaresi, E. Gilardoni, C. Coppa, M. Carini, G. Vistoli, T. Bex, J. Stautemas, L. Blancquaert, W. Derave, G. Aldini, and L. Regazzoni, "Development and validation of a sensitive LC–MS/MS assay for the quantification of anserine in human plasma and urine and its application to pharmacokinetic study," *Amino Acids*, vol. 51, no. 1, pp. 103–114, 2019.

[307] V. L. Schramm, "Transition states, analogues, and drug development," *ACS Chemical Biology*, vol. 8, no. 1, pp. 71–81, 2013.

[308] Y. C. Chen, "Beware of docking!," *Trends in Pharmacological Sciences*, vol. 36, no. 2, pp. 78–95, 2015.

[309] S. J. Weiner, P. A. Kollman, U. C. Singh, D. A. Case, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins," *Journal of the American Chemical Society*, vol. 106, no. 3, pp. 765–784, 1984.

[310] WHO, "Global Antimicrobial Resistance Surveillance System (GLASS) Report."

[311] J. M. Munita and C. A. Arias, "Mechanisms of Antibiotic Resistance," *Microbiology Spectrum*, vol. 4, no. 2, pp. 1–34, 2016.

[312] K. Bush and G. A. Jacoby, "Updated functional classification of beta-lactamases," *Antimicrobial agents and chemotherapy*, vol. 54, no. 3, pp. 969–76, 2010.

[313] C. L. Tooke, P. Hinchliffe, E. C. Bragginton, C. K. Colenso, V. H. Hirvonen, Y. Takebayashi, and J. Spencer, "β-Lactamases and β-Lactamase Inhibitors in the 21st Century," *Journal of Molecular Biology*, vol. 431, no. 18, pp. 3472–3500, 2019.

[314] R. A. Fritz, J. H. Alzate-Morales, J. Spencer, A. J. Mulholland, and M. W. Van Der Kamp, "Multiscale Simulations of Clavulanate Inhibition Identify the Reactive Complex in Class A β-Lactamases and Predict the Efficiency of Inhibition," *Biochemistry*, vol. 57, no. 26, pp. 3560–3563, 2018.

[315] R. Pereira, V. W. H. Rabelo, A. Sibajev, P. A. Abreu, and H. C. Castro, "Class A β-lactamases and inhibitors: In silico analysis of the binding mode and the relationship with resistance," *Journal of Biotechnology*, vol. 279, pp. 37–46, 2018.

[316] K. H. Tehrani and N. I. Martin, "β-lactam/β-lactamase inhibitor combinations: an update," *MedChemComm*, vol. 9, no. 9, pp. 1439–1456, 2018.

[317] C. L. Tooke, P. Hinchliffe, A. Krajnc, A. J. Mulholland, J. Brem, C. J. Schofield, and J. Spencer, "Cyclic boronates as versatile scaffolds for KPC-2 b-lactamase inhibition," *RSC Medicinal Chemistry*, vol. 11, no. 4, pp. 491–496, 2020.

[318] M. F. Mojica, R. A. Bonomo, and W. Fast, "B1-Metallo-beta-Lactamases : Where Do We Stand?," *Current Drug Targets*, vol. 17, no. 9, pp. 1029–1050, 2016.

[319] K. Bush and P. A. Bradford, "Epidemiology of β-lactamase-producing pathogens," *Clinical Microbiology Reviews*, vol. 33, no. 2, 2020.

[320] C. M. Rotondo and G. D. Wright, "Inhibitors of metallo-β-lactamases," *Current Opinion in Microbiology*, vol. 39, no. 11, pp. 96–105, 2017.

[321] M. N. Lisa, A. R. Palacios, M. Aitha, M. M. González, D. M. Moreno, M. W. Crowder, R. A. Bonomo, J. Spencer, D. L. Tierney, L. I. Llarrull, and A. J. Vila, "A general reaction mechanism for carbapenem hydrolysis by mononuclear and binuclear metallo-β-lactamases," *Nature Communications*, vol. 8, no. 1, pp. 1–11, 2017.

[322] M. R. Meini, L. I. Llarrull, and A. J. Vila, "Overcoming differences: The catalytic mechanism of metallo-β-lactamases," *FEBS Letters*, vol. 589, no. 22, pp. 3419–3432, 2015.

[323] T. Palzkill, "Metallo-β-lactamase structure and function," *Annals of the New York Academy of Sciences*, vol. 1277, no. 1, pp. 91–104, 2013.

[324] D. J. Huggins, P. C. Biggin, M. A. Dämgen, J. W. Essex, S. A. Harris, R. H. Henchman, S. Khalid, A. Kuzmanic, C. A. Laughton, J. Michel, A. J. Mulholland, E. Rosta, M. S. Sansom, and M. W. van der Kamp, "Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 9, no. 3, p. e1393, 2019.

[325] K. A. Alzahrani and R. J. Deeth, "Molecular modeling of zinc paddlewheel molecular complexes and the pores of a flexible metal organic framework," *Journal of Molecular Modeling*, vol. 22, no. 4, 2016.

[326] M. B. Peters, Y. Yang, B. Wang, L. Fusti-Molnar, M. N. Weaver, and K. M. Merz, "Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF)," *Journal of Chemical Theory and Computation*, vol. 6, no. 9, pp. 2935–2947, 2010.

[327] Z. Zuo and J. Liu, "Assessing the Performance of the Nonbonded Mg2+ Models in a Two-Metal-Dependent Ribonuclease," *Journal of Chemical Information and Modeling*, vol. 59, no. 1, pp. 399–408, 2019.

[328] P. Pokorná, M. Krepl, H. Kruse, and J. Šponer, "MD and QM/MM Study of the Quaternary HutP Homohexamer Complex with mRNA, l-Histidine Ligand, and Mg 2+," *Journal of Chemical Theory and Computation*, vol. 13, no. 11, pp. 5658–5670, 2017.

[329] Z. Yu, P. Li, and K. M. Merz, "Extended Zinc AMBER Force Field (EZAFF)," *Journal of Chemical Theory and Computation*, vol. 14, no. 1, pp. 242–254, 2018.

[330] S. Kouda, M. Ohara, M. Onodera, Y. Fujiue, M. Sasaki, T. Kohara, S. Kashiyama, S. Hayashida, T. Harino, T. Tsuji, H. Itaha, N. Gotoh, A. Matsubara, T. Usui, and M. Sugai, "Increased prevalence and clonal dissemination of multidrug-resistant Pseudomonas aeruginosa with the blaIMP-1 gene cassette in Hiroshima," *Journal of Antimicrobial Chemotherapy*, vol. 64, no. 1, pp. 46–51, 2009.

[331] W. H. Zhao and Z. Q. Hu, "IMP-type metallo-β-lactamases in Gram-negative bacilli: Distribution, phylogeny, and association with integrons," *Critical Reviews in Microbiology*, vol. 37, no. 3, pp. 214–226, 2011.

[332] G. Cornaglia, H. Giamarellou, and G. M. Rossolini, "Metallo-β-lactamases: A last frontier for β-lactams?," *The Lancet Infectious Diseases*, vol. 11, no. 5, pp. 381–393, 2011.

[333] S. L. Gellatly and R. E. Hancock, "Pseudomonas aeruginosa: New insights into pathogenesis and host defenses," *Pathogens and Disease*, vol. 67, no. 3, pp. 159–173, 2013.

[334] L. Nguyen, J. Garcia, K. Gruenberg, and C. MacDougall, "Multidrug-Resistant Pseudomonas Infections: Hard to Treat, But Hope on the Horizon?," *Current Infectious Disease Reports*, vol. 20, no. 8, p. 23, 2018.

[335] L. C. Ju, Z. Cheng, W. Fast, R. A. Bonomo, and M. W. Crowder, "The Continuing Challenge of Metallo-β-Lactamase Inhibition: Mechanism Matters," *Trends in Pharmacological Sciences*, vol. 39, no. 7, pp. 635–647, 2018.

[336] K. H. M. E. Tehrani and N. I. Martin, "Thiol-Containing Metallo-β-Lactamase Inhibitors Resensitize Resistant Gram-Negative Bacteria to Meropenem," *ACS Infectious Diseases*, vol. 3, pp. 711–717, 2017.

[337] F. M. Klingler, T. A. Wichelhaus, D. Frank, J. Cuesta-Bernal, J. El-Delik, H. F. Müller, H. Sjuts, S. Göttig, A. Koenigs, K. M. Pos, D. Pogoryelov, and E. Proschak, "Approved drugs containing thiols as inhibitors of metallo-β-lactamases: Strategy to combat multidrug-resistant bacteria," *Journal of Medicinal Chemistry*, vol. 58, no. 8, pp. 3626–3630, 2015.

[338] J. Brem, S. S. Van Berkel, D. Zollman, S. Y. Lee, O. Gileadi, P. J. McHugh, T. R. Walsh, M. A. McDonough, and C. J. Schofield, "Structural basis of metallo-β-lactamase inhibition by captopril stereoisomers," *Antimicrobial Agents and Chemotherapy*, vol. 60, no. 1, pp. 142–150, 2016.

[339] O. Korb, T. Stützle, and T. E. Exner, "An ant colony optimization approach to flexible protein-ligand docking," *Swarm Intell*, vol. 1, pp. 115–134, 2007.

[340] M. Gaus, Q. Cui, and M. Elstner, "DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB)," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 931–948, 2011.

[341] Q. Cui and M. Elstner, "Density functional tight binding: Values of semi-empirical methods in an ab initio era," *Physical Chemistry Chemical Physics*, vol. 16, no. 28, pp. 14368–14377, 2014.

[342] I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, "DL_POLY_3: New dimensions in molecular dynamics simulations via massive parallelism," *Journal of Materials Chemistry*, vol. 16, no. 20, pp. 1911–1918, 2006.

[343] S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *Journal of Computational Chemistry*, vol. 27, no. 15, pp. 1787–1799, 2006.

[344] R. Lonsdale, J. N. Harvey, and A. J. Mulholland, "Effects of dispersion in density functional based quantum mechanical/molecular mechanical calculations on cytochrome P450 catalyzed reactions," *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4637–4645, 2012.

[345] P. Chaskar, V. Zoete, and U. F. Röhrig, "Toward On-The-Fly Quantum Mechanical/Molecular Mechanical (QM/MM) Docking: Development and Benchmark of a Scoring Function," *Journal of Chemical Information and Modeling*, vol. 54, no. 11, pp. 3137–3152, 2014.

[346] O. T. Olson and A. J., "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading," *J Comput Chem.*, vol. 31, no. 2, pp. 455–461, 2010.

[347] E. G. R. de Arruda, B. A. Rocha, M. Barrionuevo, H. M. Adalsteinsson, F. E. Galdino, W. Loh, F. Lima, and C. Abbehausen, "Zn(II) coordination sphere and chemical structure influence over the reactivity of metallo-β-lactamase model compounds," *Dalton Transactions*, vol. 48, no. 9, pp. 2900–2916, 2018.

[348] D. P. Martin, P. G. Blachly, J. A. Mccammon, and S. M. Cohen, "Exploring the Influence of the Protein Environment on Metal-Binding Pharmacophores," *J. Med. Chem.*, vol. 57, pp. 7126–7135, 2014.

[349] D. P. Kapusta, D. A. Firsov, M. G. Khrenova, B. L. Grigorenko, and A. V. Nemukhin, "Effect of solvation water shells on enzyme active sites in zinc-dependent hydrolases," *Structural Chemistry*, vol. 30, no. 2, pp. 481–488, 2019.

[350] H. Kurosaki, Y. Yamaguchi, H. Yasuzawa, W. Jin, Y. Yamagata, and Y. Arakawa, "Probing, Inhibition, and Crystallographic Characterization of Metallo-β-lactamase (IMP-1) with Fluorescent Agents Containing Dansyl and Thiol Groups," *ChemMedChem*, vol. 1, no. 9, pp. 969–972, 2006.

[351] N. O. Concha, C. A. Janson, P. Rowling, S. Pearson, C. A. Cheever, B. P. Clarke, C. Lewis, M. Galleni, J. M. Frère, D. J. Payne, J. H. Bateson, and S. S. Abdel-Meguid, "Crystal structure of the IMP-1 metallo β-lactamase from Pseudomonas aeruginosa and its complex with a mercaptocarboxylate inhibitor: Binding determinants of a potent, broad-spectrum inhibitor," *Biochemistry*, vol. 39, no. 15, pp. 4288–4298, 2000.

[352] H. Kurosaki, Y. Yamaguchi, T. Higashi, K. Soga, S. Matsueda, H. Yumoto, S. Misumi, Y. Yamagata, Y. Arakawa, and M. Goto, "Irreversible Inhibition of Metallo-β-lactamase (IMP-1) by 3-(3-Mercaptopropionylsulfanyl)propionic Acid Pentafluorophenyl Ester," *Angewandte Chemie International Edition*, vol. 44, no. 25, pp. 3861–3864, 2005.

[353] P. Hinchliffe, M. M. González, M. F. Mojica, J. M. González, V. Castillo, C. Saiz, M. Kosmopoulou, C. L. Tooke, L. I. Llarrull, G. Mahler, R. A. Bonomo, A. J. Vila, and J. Spencer, "Cross-class metallo-β-lactamase inhibition by bisthiazolidines reveals multiple binding modes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 26, pp. E3745–E3754, 2016.

[354] N. O. Concha, B. A. Rasmussen, K. Bush, and O. Herzberg, "Crystal structure of the wide-spectrum binuclear zinc β- lactamase from Bacteroides fragilis," *Structure*, vol. 4, no. 7, pp. 823–836, 1996.

[355] X. Lu, M. Gaus, M. Elstner, Q. Cui, and W. L. J. Festschrift, "Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications The traditional semiempirical QM methods based on the Neglect-of-Diatomic-Differential-Overlap (NDDO) approxima-Special Issue," *J. Phys. Chem. B*, vol. 119, p. 45, 2015.

[356] D. K. Chakravorty, B. Wang, C. W. Lee, D. P. Giedroc, and K. M. Merz, "Simulations of allosteric motions in the zinc sensor CzrA," *Journal of the American Chemical Society*, vol. 134, no. 7, pp. 3367–3376, 2012.

[357] S. Wu, D. Xu, and H. Guo, "QM/MM Studies of Monozinc b-Lactamase CphA Suggest That the Crystal Structure of an Enzyme-Intermediate Complex Represents a Minor Pathway," *J. AM. CHEM. SOC.*, vol. 132, pp. 17986–17988, 2010.

[358] M. G. Khrenova and A. V. Nemukhin, "Modeling the Transient Kinetics of the L1 Metallo-β-Lactamase," *Journal of Physical Chemistry B*, vol. 122, no. 4, pp. 1378–1386, 2018.

[359] J. J. Huntley, W. Fast, S. J. Benkovic, P. E. Wright, and H. J. Dyson, "Role of a solvent-exposed tryptophan in the recognition and binding of antibiotic substrates for a metallo-β-lactamase," *Protein Science*, vol. 12, pp. 1368–1375, 2003.

[360] C. Moali, C. Anne, J. Lamotte-Brasseur, S. Groslambert, B. Devreese, J. Van Beeumen, M. Galleni, and J. M. Frère, "Analysis of the importance of the metallo-β-lactamase active site loop in substrate binding and catalysis," *Chemistry and Biology*, vol. 10, no. 4, pp. 319–329, 2003.

[361] C. Montagner, M. Nigen, O. Jacquin, N. Willet, M. Dumoulin, A. I. Karsisiotis, G. C. Roberts, C. Damblon, C. Redfield, and A. Matagne, "The role of active site flexible loops in catalysis and of zinc in conformational stability of Bacillus cereus 569/H/9 β-lactamase," *Journal of Biological Chemistry*, vol. 291, no. 31, pp. 16124–16137, 2016.

[362] Y. Ge, M. van der Kamp, M. Malaisree, D. Liu, Y. Liu, and A. J. Mulholland, "Identification of the quinolinedione inhibitor binding site in Cdc25 phosphatase B through docking and molecular dynamics simulations," *Journal of Computer-Aided Molecular Design*, vol. 31, no. 11, pp. 995–1007, 2017.

[363] J. J. Irwin, F. M. Raushel, and B. K. Shoichet, "Virtual screening against metalloenzymes for inhibitors and substrates," *Biochemistry*, vol. 44, no. 37, pp. 12316–12328, 2005.

[364] A. Pecina, S. Haldar, J. Fanfrlík, R. Meier, J. Řezáč, M. Lepšík, and P. Hobza, "SQM/COSMO Scoring Function at the DFTB3-D3H4 Level: Unique Identification of Native Protein-Ligand Poses," *Journal of Chemical Information and Modeling*, vol. 57, no. 2, pp. 127–132, 2017.

[365] A. Pecina, R. Meier, J. Fanfrlík, M. Lepšík, J. Řezáč, P. Hobza, and C. Baldauf, "The SQM/COSMO filter: Reliable native pose identification based on the quantum-mechanical description of protein-ligand interactions and implicit COSMO solvation," *Chemical Communications*, vol. 52, no. 16, pp. 3312–3315, 2016.

[366] Z. Jing, R. Qi, C. Liu, and P. Ren, "Study of interactions between metal ions and protein model compounds by energy decomposition analyses and the AMOEBA force field," *Journal of Chemical Physics*, vol. 147, no. 16, p. 161733, 2017.

[367] C. Bergonzo, K. B. Hall, and T. E. Cheatham, "Divalent ion dependent conformational changes in an RNA stem-loop observed by molecular dynamics," *Journal of Chemical Theory and Computation*, vol. 12, no. 7, pp. 3382–3389, 2016.

[368] K. Shahrokh, A. Orendt, G. S. Yost, and T. E. Cheatham, "Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle," *Journal of Computational Chemistry*, vol. 33, no. 2, pp. 119–133, 2012.

[369] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren, "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications," *Annual Review ofBiophysics*, vol. 48, pp. 371–394, 2019.

[370] V. S. Inakollu, D. P. Geerke, C. N. Rowley, and H. Yu, "Polarisable force fields: what do they add in biomolecular simulations?," *Current Opinion in Structural Biology*, vol. 61, pp. 182–190, 2020.

[371] P. Xu, E. B. Guidez, C. Bertoni, and M. S. Gordon, "Perspective: Ab initio force field methods derived from quantum mechanics," *Journal of Chemical Physics*, vol. 148, no. 9, p. 90901, 2018.

[372] Y. J. Zhang, A. Khorshidi, G. Kastlunger, and A. A. Peterson, "The potential for machine learning in hybrid QM/MM calculations," *Journal of Chemical Physics*, vol. 148, no. 24, p. 241740, 2018.

[373] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, "Machine learning force fields: Construction, validation, and outlook," *Journal of Physical Chemistry C*, vol. 121, no. 1, pp. 511–522, 2017.

[374] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, "Machine Learning of Coarse-Grained Molecular Dynamics Force Fields," *ACS Central Science*, vol. 5, no. 5, pp. 755–767, 2019.

[375] Y. Li, H. Li, F. C. Pickard, B. Narayanan, F. G. Sen, M. K. Chan, S. K. Sankaranarayanan, B. R. Brooks, and B. Roux, "Machine Learning Force Field Parameters from Ab Initio Data," *Journal of Chemical Theory and Computation*, vol. 13, no. 9, pp. 4492–4503, 2017.