



**Universitat Autònoma
de Barcelona**

**Cationic Dummy Atom Software
(CaDAS)**

by

Daniel Soler Viladrich

Supervised by Jean-Didier Maréchal and Carles Navau Ros

A thesis submitted in partial fulfilment for the
Bachelor of Physics
Bachelor of Chemistry

in the
Science Faculty
Physics Department
Chemistry Department

July 13, 2017

"Protons give an atom its identity, electrons its personality."

Bill Bryson

AUTONOMOUS UNIVERSITY OF BARCELONA

Abstract

Science Faculty
Physics Department
Chemistry Department

Physics and Chemistry Student

by [Daniel Soler Viladrich](#)

A long-standing challenge in modelling metal-containing systems via molecular mechanics tools is to rapidly define the properties of metallic centres. This project optimises the functionality of dummy-atom metallic models, by providing a time-saving software to determine the stability of organometallic systems. The software is tested on a metallopeptide model of a human carbonic anhydrase and an artificial siderophore. Results on the first system show the stability of the structure obtained with the dummy atom approach and the agreement with experimental data. No binding affinity is found for the second system suggesting the proposed geometry is implausible in accordance with the bibliography. The results show potential applicability of the software in new drug design.

Acknowledgements

I would like to use this section to thank all people who have been involved in this project.

My warmest thanks to my project tutor and supervisor Prof. Jean-Didier Maréchal to give me the opportunity and guidance along the way.

Also, my special thanks of gratitude to Jaime Rodriguez, who gave me the golden opportunity to learn from him, coming to know about so many new things.

I am also grateful to my tutor Dr. Carles Navau for his feedback, cooperation and support.

I would like to express my deepest appreciation to all Insilichem members, specially Lur, Josepe, Sergi and David Teixé whose contributions and encouragement, helped me to finalise my project.

Last but not least, I would like to thank my girlfriend, parents and friends for accepting nothing less than excellence from me.

Contents

Abstract	ii
Acknowledgements	iii
List of Figures	vi
List of Tables	viii
Abbreviations	ix
1 Introduction	1
1.1 Molecular Modelling	3
1.1.1 Quantum Mechanics	4
1.1.2 Molecular Mechanics	6
1.2 Molecular Mechanics. Modelling of Metal Systems	7
1.2.1 Covalent Bond Approach	8
1.2.2 Non Bonded Approach	8
1.2.3 Cationic Dummy Atom	8
1.3 Cationic Dummy Atoms Software	9
1.3.1 UCSF Chimera	11
1.3.2 tLEaP	11
1.3.3 OpenMM	11
2 Methodology	13
2.1 CaDAS	13
2.1.1 System Algorithms	14
2.1.1.1 Search Algorithm	15
2.1.1.2 Orientation Algorithm	17
2.1.1.3 Geometry Algorithm	17
2.1.2 System Properties	18
2.1.3 System Production	20
3 Results	22
3.1 Benchmarking CaDAS	22
3.1.1 Binding site model of a modified carbonic anhydrase.	22
3.1.2 GPG siderophore system	26
3.2 Future Improvements	32

4 Conclusion	34
A Appendix	35
A.1 Carbonic anhydrase system	35
A.1.1 Computed metal-ligand bond distances over time:	35
A.1.2 Computed metal-dummy-ligand angles over time:	37
A.2 GPG system	39
A.2.1 Computed metal-ligand bond distances over time:	39
A.2.2 Computed dummy-ligand bond distances over time:	43
Bibliography	47

List of Figures

1.1	Applications of transition metals	2
1.2	Classification of the main molecular modelling methods according to their accuracy and length scale.	4
1.3	Cationic Dummy Atoms Software Interface.	9
1.4	Zn(II) using Cationic Dummy Atoms Approach Software(CaDAS).	10
1.5	Software overall structure.	12
2.1	CaDAS flowchart.	14
2.2	System algorithms flowchart.	15
2.3	Demonstration of the geometry algorithm for an octahedron.	18
2.4	System properties flowchart.	20
2.5	System production flowchart.	21
3.1	Modified carbonic anhydrase active site.	23
3.2	Modified carbonic anhydrase active site with dummy atoms.	23
3.3	RMSD values vs frames.	24
3.4	Zn-S-C _{1Cis} -C _{2Cis} dihedral angle value over time.	26
3.5	GPG system.	27
3.6	Dummy atoms in GPG.	27
3.7	RMSD values vs frames.	28
3.8	Catechol displacement from the metal centre alignment.	29
3.9	Momentary geometry of the ligands around the metal centre with coordination number of 4.	30
3.10	Labelling of atoms in table 3.3.	31
3.11	Cage system.	32
A.1	Zn-S bond length	35
A.2	Zn-N _{His1} bond length	36
A.3	Zn-N _{His2} bond length	36
A.4	Zn-OH bond length	37
A.5	Zn-DM-S angle	37
A.6	Zn-DM-N _{His1} angle	38
A.7	Zn-DM-N _{His2} angle	38
A.8	Zn-DM-OH angle	39
A.9	Fe-OH ₁ bond length	39
A.10	Fe-OH ₂ bond length	40
A.11	Fe-OH ₃ bond length	40
A.12	Fe-OH ₄ bond length	41

A.13 Fe-OH ₅ bond length	41
A.14 Fe-OH ₆ bond length	42
A.15 Fe-O ₁ bond length	42
A.16 DM-OH ₁ bond length	43
A.17 DM-OH ₂ bond length	43
A.18 DM-OH ₃ bond length	44
A.19 DM-OH ₄ bond length	44
A.20 DM-OH ₅ bond length	45
A.21 DM-OH ₆ bond length	45
A.22 DM-O ₁ bond length	46

List of Tables

3.1	Comparison between computed metal centre-ligand distances and experimental data.	25
3.2	Computed angles for the indicated atoms, which are labelled in figure 3.2.	25
3.3	Computed angles for the indicated atoms, which are labelled in figure 3.2.	25
3.4	Dummy-Ligand bond distance and Metal-Ligand bond distance.	30

Abbreviations

CaDAS	Cationic Dummy Atom Software
CaDA	Cationic Dummy Atom method
QM	Quantum Mechanics
MM	Molecular Mechanics
QM/MM	Quantum Mechanics Molecular Mechanics
MD	Molecular Dynamics
DFT	Density Funcional Theory
M	Metal
DM	Dummy
H	Hydrogen
N	Nitrogen
Zn	Zinc
Fe	Iron
S	Sulfur
RMSD	Root Mean Square Deviation

Chapter 1

Introduction

Metal cations are of relevance in nature as they take a key role on a great number of fields such as chemistry of living systems, material construction, catalysis and many others as seen in figure 1.1. In fact, metalloproteins constitute almost a third of all proteins in biology [1]. Therefore, understanding the reasons behind structural conformation and coordination geometry of these metal centres would be of great value since it would elucidate many processes such as catalytic reactions [2], signal transduction [3] and metalloprotein and material design. In order to study the properties of metal centres in the previous processes one could rely on molecular modelling, which acts as a bridge between the microscopic world and macroscopic properties to obtain bulk property predictions. Molecular modelling is divided in two major branches depending on the type and desired atomic description of the system to study [4].

For quantitative calculations and accuracy, quantum mechanics (QM) is the best method up to date. It models electronic rearrangements during simulations which are essential for those parts of the system involved in the process. For a less accurate but faster result Molecular Mechanics (MM) is enough. The system is Newtonianly treated and no electron distribution, in other words, bond formation, is allowed. The major advantage of MM methods is that the interaction between atoms is usually derived from simple potential energy functions, overcoming the QM calculations time expenditure. On the downside, these techniques poorly reproduce the electronic interactions between metal centres and their surroundings; yet new successful molecular mechanics approaches for metallic systems have arisen in the last few years [5]. However, one still needs to undergo

a tedious parametrization process to achieve a representative model of the system to study. This is due to the fact that metal centre parametrization is much more challenging than its non metallic counterpart since metal nuclei are much larger and have variable geometries [6].

Therefore, the primary goal of this dissertation is to present a molecular modelling software to quickly parametrize metal centres in biological systems by using the so-called cationic dummy atom approach. One of the less applied but most accurate techniques when aiming to define metals in molecular systems [7]. The software automates the parametrization of metallic centres, accelerating the representation of the desired system while determining the stability and geometrical properties at a reasonable accuracy. The improvement of this software could finally lead to faster but still accurate predictions for the comprehension of these systems.

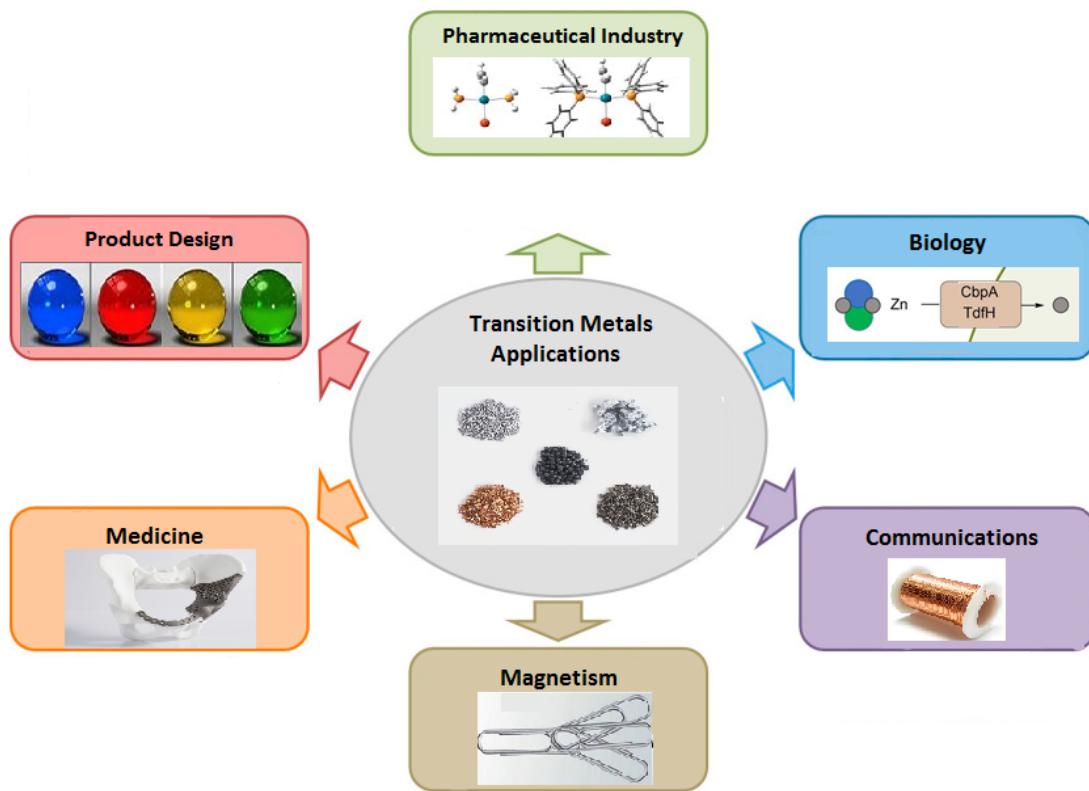


FIGURE 1.1: Applications of transition metals.

1.1 Molecular Modelling

As said in the introduction, a big number of processes in biological systems rely on the properties of transition metal ions. However, the theoretical modelling of transition metal centres is a notoriously difficult task. The metal-ligand interactions and the combination of suitable analytic interaction potentials is highly complex. This is due to the fact that at finite temperature, the system might switch dynamically between different bonding situations which are characterised by multiple coordination numbers and geometries [8].

To represent all these behaviours current modellers prefer quantum mechanics methods, which shed light to this field in the late 1980s as new and better computational approaches were created. Nevertheless, due to their high computational cost they are not feasible for large scale system where a wide conformational exploration is required as seen in figure 1.2. In order to pose a solution to this problem, Molecular Mechanics could be used as a more efficient alternative. It focuses on finding the relationship between the structure and the energy of a system. To do so, it uses the laws of classical mechanics to write the electronic energy in terms of a parametric function which mimics the energetic variation of the natural system. Unfortunately, most of these functions do not take into account the electronic degrees of freedom, experiencing large difficulties in describing the metal environment accurately. In addition, they do not account for bond breaking and forming events. To address these issues the quantum mechanical/molecular mechanical (QM/MM) method can be used to obtain the parameter-free quantum mechanical approach on certain areas of the system and keep the rest under the MM laws [9].

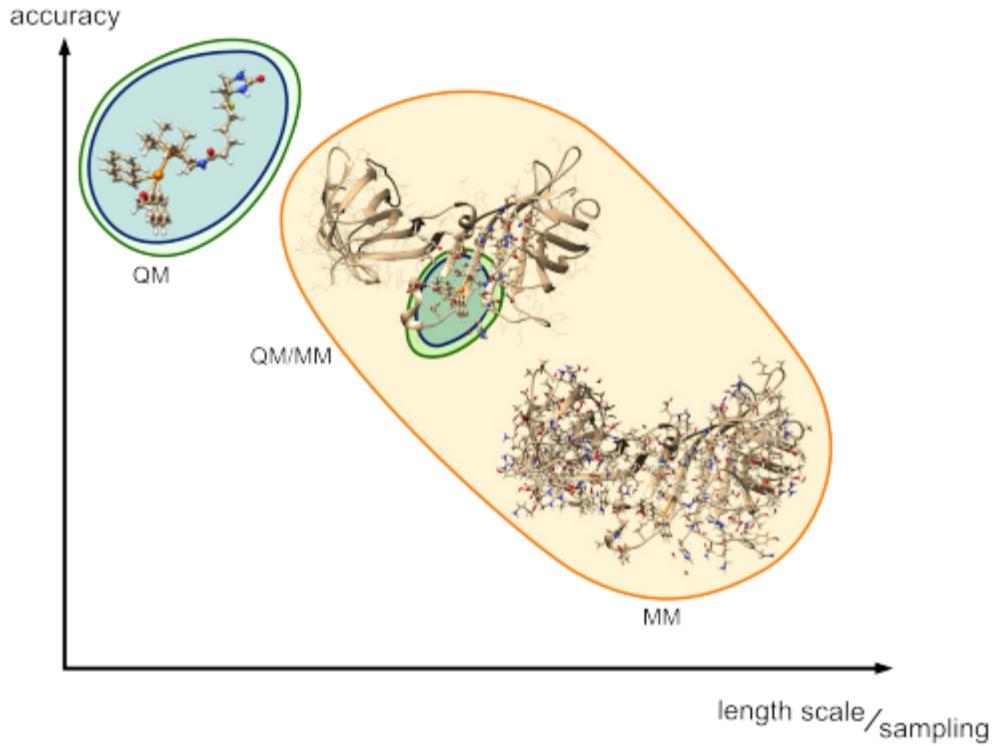


FIGURE 1.2: Classification of the main molecular modelling methods according to their accuracy and length scale.

1.1.1 Quantum Mechanics

Quantum Mechanics stem from the Schrodinger equation 1.1 first brought to light in the 1920's. This method treats the system as a collection of nuclei and electrons without any input concerning chemical bonds. Calculations are made by solving the system's Schrodinger equation. However, this equation cannot actually be solved for any but a one-electron system, and approximations need to be made. Depending on these approximations a wide scope of chemical models and methods can be found in terms of their capability and their "cost" [10].

$$i\hbar \frac{\partial}{\partial t} \psi(r, t) = \left[\frac{-\hbar^2}{2\mu} \nabla^2 + V(r, t) \right] \psi(r, t) \quad (1.1)$$

The most common methods are directly derived from theoretical principles, as Hartree Fock and density functional theory (DFT), typically used to solve the time-independent Schrödinger equation for a multi-electron atom or molecule. Hartree Fock [11] is a

variational method based on the Eckart theorem, which uses a linear combination of functions as in figure 1.2 to describe orbital's density and calculate an upper bound to the true ground state energy of a given molecule. Instead, DFT [12] uses the Kohn–Sham equation, 1.3, defined by a fictitious external potential denoted as $V_s(r)$ and a N-particle density function to determine the energy of a system 1.4. Even though Kohn-Sham equation is exact, the exchange-correlation energy functional is not known explicitly. That is the main reason the method is not absolutely accurate and it is also the reason for the derivation of various exchange correlation functionals.

$$\psi_{HP}(x_1, x_2, \dots, x_N) = \chi^1(x_1)\chi^2(x_2)\dots\chi^N(x_N). \quad (1.2)$$

$$[\frac{-\hbar^2}{2m} \nabla^2 + V_s(r)]\theta(r) = E_i\theta(r) \quad (1.3)$$

$$E_{tot(\rho)} = E_{kin(\rho)} + E_{ext(\rho)} + E_{H(\rho)} + E_{xc(\rho)} \quad (1.4)$$

However, these two methods agree on three major simplifications.

1. The Born–Oppenheimer approximation where the motion of atomic nuclei and electrons in a molecule can be separated as a result of their big difference in mass.
2. The momentum operator is assumed to be non-relativistic.
3. The mean field approximation is applied, instead of evaluating the electronic density of each electron over a nuclei, an average density function for all electrons is calculated.

With all the approximations above, the equation 1.5 is finally resolved obtaining the potential energy surface (from which one can get the equilibrium geometry and vibrational frequencies) and the electronic wave function $\psi(r; R)$, which contains lots of useful information about molecular properties such as dipole moments and polarizability.

$$[T_e(r) + V_{eN}(r; R) + V_{NN}(R) + V_{ee}(r)]\varphi(r; R) = E_{el}\varphi(r; R) \quad (1.5)$$

1.1.2 Molecular Mechanics

Systems treated within molecular mechanics are made up of atoms and bonds (as opposed to nuclei and electrons), and studied through forcefields, or functional forms used to determine the potential energy of the system [13]. These are split into two general terms, bonded and non-bonded interactions as seen in equation 1.6. The latter is traditionally studied through one, two or even three body interaction terms. The former, is calculated as a sum of individual terms, as seen in equation 1.7. The final aim is to calculate the molecular mechanics energy equation 1.9.

$$E_T = E_{bonded} + E_{non-bonded} \quad (1.6)$$

$$E_{bonded} = E_{stretching} + E_{bending} + E_{torsions} \quad (1.7)$$

$$E_{non-bonded} = E_{VdW} + E_{electrostatic} + E_{crossterms} \quad (1.8)$$

$$E_T = E_{stretching} + E_{bending} + E_{torsions} + E_{VdW} + E_{electrostatic} + E_{crossterms} \quad (1.9)$$

Where $E_{stretching}$ is the energy involved in the deformation of a bond, either by stretching or compression and $E_{bending}$ is the energy involved in angle bending. To determine them, bonds are treated as springs and described by equation 1.10 based on the Hook's law.

$$E_{stretching} = \sum_{bonds} k_b(r - r_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 \quad (1.10)$$

$E_{torsions}$ is the torsional angle energy, which represents the amount of energy that must be added or subtracted to the other terms to make the total energy agree with the experimental value. This one is described through the periodical function 1.11.

$$E_{torsions} = A[1 + \cos(n\tau - \phi)] \quad (1.11)$$

E_{VdW} is the energy related with the distance dependent interactions between atoms. On the one hand, Van der Waals attraction occurs at close distance, and rapidly dies off as the interacting atoms separate a few angstroms. On the other hand, repulsion occurs when the distance between interacting atoms becomes slightly less than the sum

of their contact radius. That is summed up on the equation 1.12, where the repulsive term diverges at short ranges and the A and B parameters control the depth and the position of the potential energy well. However, this is not the only equation that can describe this type of behaviour.

$$E_{vanderWaals} = -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \quad (1.12)$$

$E_{electrostatic}$ is the energy involved in interactions between atoms that are not directly bonded. This can be modeled using a Coulombic potential. Where, electrostatic energy depends on the charge of the non-bonded atoms, their interatomic distance, and a molecular dielectric expression that accounts for the attenuation of electrostatic interactions by the environment.

One application were MM is found is molecular dynamics simulations (MD) based on the ergodic hypothesis, instead of calculating the partition function to obtain ensemble averages, such as energy, values are replaced by time averages over the simulation. All of this is obviously much simpler than solving the Schrödinger equation for electron motions, but requires an explicit description of “chemical bonding”, as well as a large amount of information about the structures of molecules. Pharmaceutical companies, research groups and many other members of the scientific community have been using MD to perform ligand dockings on the protein active site, for instance in drug design, the mechanism of inhibiton of protein kinases [14] (targets for treatment of a number of diseases) has been elucidated by this technique.

1.2 Molecular Mechanics. Modelling of Metal Systems

Forcefield parameters are available for the 20 natural aminoacids and the five base pairs of nucleic acids [15]. Nonetheless, in the case of metal centres these parameters are related to the nature of the metal, its geometry and coordination. Therefore, they must be calculated and optimised for each system. In spite of the intrinsic hardships of metal system parametrization, there have been some attempts of MD simulations [16] applied to organometallic systems with the purpose of understanding the dynamic properties and analysis of the internal motions of the metal centre. Nowadays, there are multiple

MM tools to model metal ions with their associated pros and cons. The process of decision-making will depend on how we want to define our system and the answers one is wishing to extract.

1.2.1 Covalent Bond Approach

This method considers bonding between the metal centre and the surrounding ligands, treating them as springs and using the harmonic oscillator potential to describe their behaviour. As a result, a large number of parameters need to be defined and double counting of the electrostatic and Lennard-Jones interactions occurs. In addition, ligand exchange cannot be described. However, in some cases the exchange occurs at larger timescales than the duration of the MD simulation. Hence, the bonded model is an effective approach to study these systems because ligand exchange does not occur over the timescales employed.

1.2.2 Non Bonded Approach

In contrast to the covalent bond method, there is the non bond approach where metal–ligand interactions are described through electrostatic and Van der Waals potentials. Thus, allowing for ligand exchange. On the downside, it appears to be inadequate when it comes to more complex situations such as multinuclear metal centre systems, closely located metal ions and transition metals. In the latter case, the challenge is to obtain a parameter set that can simultaneously reproduce both solvation free energies and metal–water distances. Withal, it is still a very functional method to study simple systems without need for detailed parametrization.

1.2.3 Cationic Dummy Atom

Finally, one realises that there are two sides to every coin. In this case either a good description of the electrostatic interaction or a strong bonded model. Even though these approaches clash, one can still wonder if it would be possible to have an in between of both. Currently, the solution is the cationic dummy atom model. A dummy atom is a cationic atom that does not sterically interact with other atoms, but it represents the metal’s vacant orbitals, thus imposing the orientational requirement for the metal

coordinates and simulating the metal propensity to a specific coordination geometry [17]. To achieve the previous, the metal nucleus is assigned with the Van der Waals radius and some partial charge while the dummy atoms are assigned only with charge. Then, the model replaces the metal ion with a specific number of dummy atoms bonded to the same metal centre and deprotonates all the metal coordinates.

All in all, the cationic dummy atoms method tries to mimic covalent bonds while offering a more sophisticated electrostatic model [18]. However, it still requires the tedious parametrization of, not just the metal centre, but also the mass and charge on each dummy.

1.3 Cationic Dummy Atoms Software

Cationic Dummy Atoms Software (CaDAS) implements all advantages cited before whilst removing the time-wise handicap, and, as a result, composes a Molecular Dynamics input. All this is done through a light and simple interface show in figure 1.3, adapted to users of any level, for the most frequently found metal centre geometries (tetrahedral, square planar, octahedral and square pyramidal).

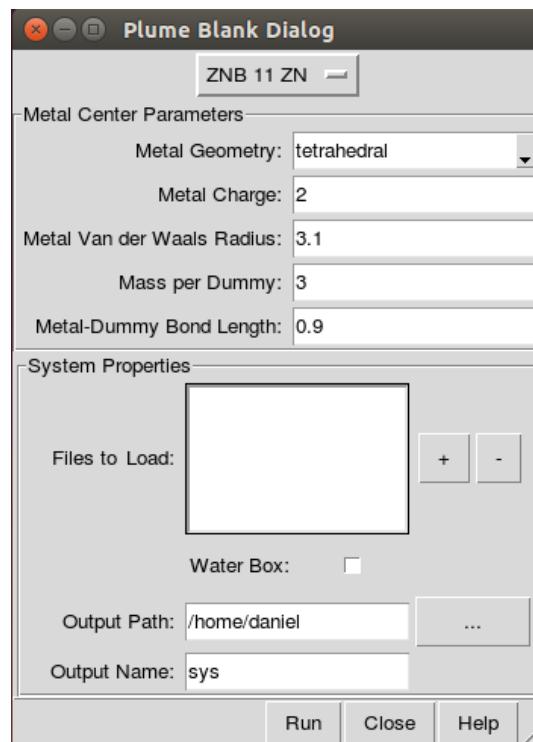


FIGURE 1.3: Cationic Dummy Atoms Software Interface.

CaDAS includes several dummy atoms that bind to the metal centre with the metal charge evenly transferred between them. In order not to lose the dummy geometry during simulation, the dummies are attached to one another creating constraints that keeps all of them packed as in figure 1.4. The spreading of the charges between the surrounding dummies tries to copy the charge distribution density inside the metal atom orbitals, making the non covalent character of the system more realistic. As for the ligand-dummy interaction, the attachment is purely electrostatic and yet, it affixes the ligand to a static position with respect to the metal, emulating a covalent bond behaviour. The largest part of the software is smoothly ran on the background creating all mass, bond, angle, dihedral and charge parameters the model needs. Hence, it removes the most laborious part of the approach simplifying the parametrization process for users. How this is done is explained in more depth in the methodology section.

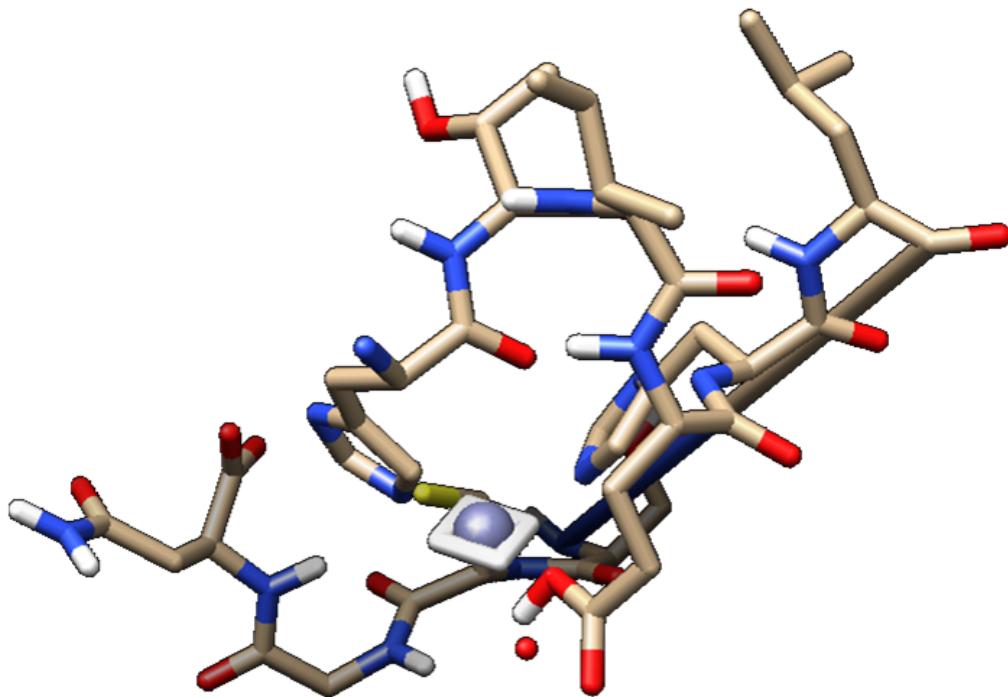


FIGURE 1.4: Zn(II) using Cationic Dummy Atoms Approach Software(CaDAS).

To implement CaDAS, Chimera UCSF and tLEaP have been used and wrapped around python language. Both of them having a set of unique characteristics that make them essential to this software. However, the launch of the MD simulation will be performed by an external high performance software called OpenMM.

1.3.1 UCSF Chimera

UCSF Chimera [19] is a highly extensible program for interactive visualisation and analysis of molecular structures and related data, including density maps, supra-molecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. In addition, UCSF Chimera gives users the opportunity to work through a python shell. The UAB research group Insilichem seized this characteristic to create PyChimera, which provides access to Chimera's modules from any Python 2.x interpreter keeping the interface out of the equation. This last tool will be one of the two major external softwares used in CaDAS, which will extract any necessary information of the system in order to create the search and orientation algorithms that will be seen later on in the methodology.

1.3.2 tLEaP

tLEaP [20] is one of the primary programs used to parametrize new systems, or to modify existing ones. It works without interface, only using a simple command line. tLEaP will be in charge of building up the system coordinates and topology, in other words, the MD input.

1.3.3 OpenMM

OpenMM [21] is a high performance toolkit for molecular simulation that can be used as a library, or as an application. Its extreme flexibility through custom forces and integrators, as well as its extreme performance through GPU Acceleration, with optimizations for AMD, NVIDIA, and Intel Integrated GPUs make the software one of the most powerful MD tools around the scientific community. However, OpenMM is a complicated library to use and sometimes not user friendly for beginners. Hence, a new software called ommprotocol was created by Jaime Rodriguez and Jean Didier Maréchal to make OpenMM MD launches simpler. Consequently, this last program will be the one handling the MD launch.

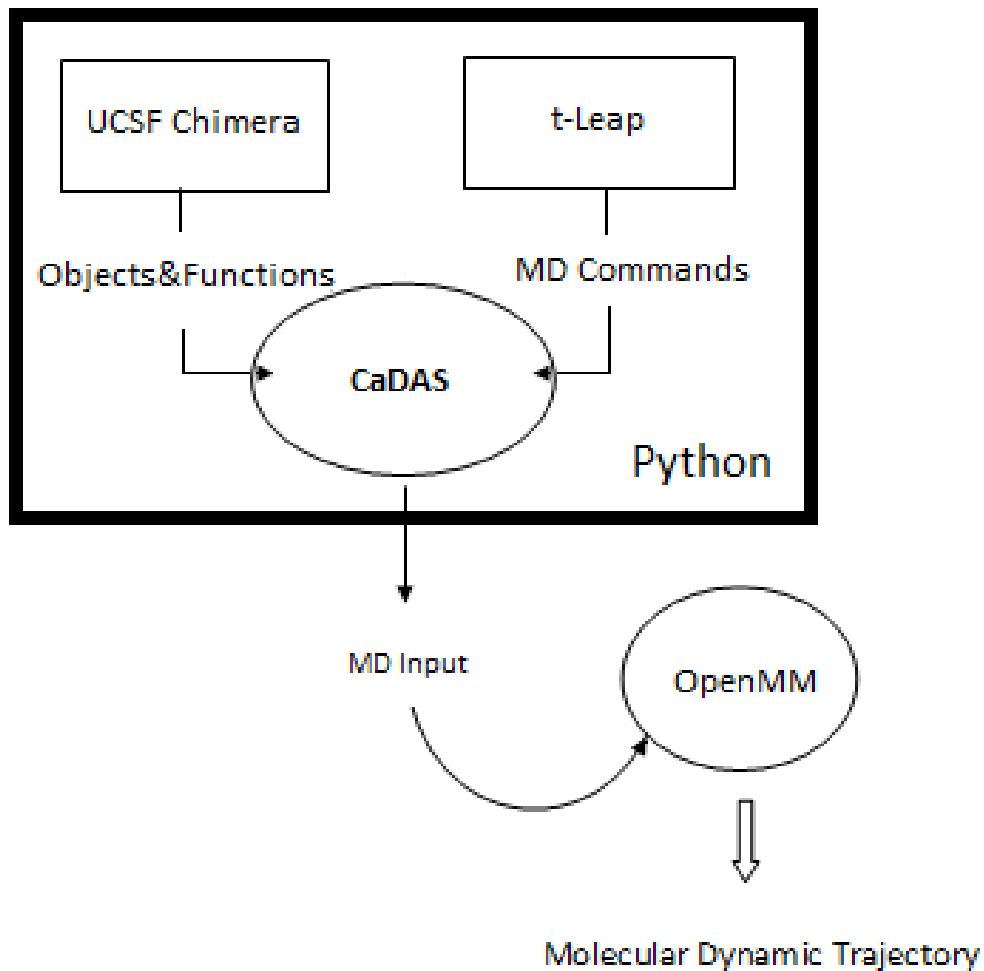


FIGURE 1.5: Software overall structure.

Chapter 2

Methodology

The aim of this section is to display how this project was conducted by following the guidelines shown in the previous section. In order to do that, the structure of the software will be broken down into three main parts. Prior to delving into the description of the software's structure, it would be helpful for one to understand first the general flowchart of the program.

2.1 CaDAS

The flowchart is subdivided into three main sections, as we can see in figure 2.1. In the first place, the given input is a .pdb file usually obtained from the protein data bank [22] or from another program's output, which contains the topology and coordinates of the system to be studied. Next, two different processes are applied to the system. The aim of the first process, shown in orange in figure 2.1, is the addition of the dummies while conserving the chemical properties of the system. The second, in blue, is the acquisition of the system's physical properties such as masses, coordinates, bond length, Van der Waals radius and number and position of its dummies. As it was mentioned before, the anatomy of the system is fundamental in MM, in order to get the right description of the system which is being modelled. Last but not least, shown in green, is the final phase of the software, where the program recollects the system's anatomy and its properties from the steps before, producing the simulation input.

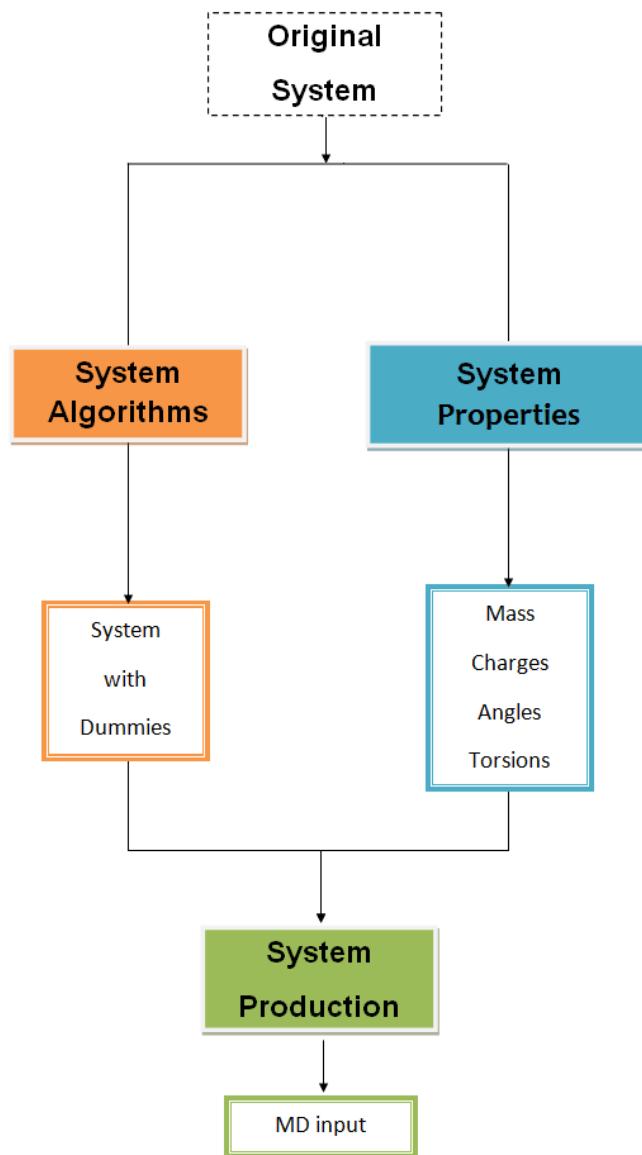


FIGURE 2.1: CaDAS flowchart.

2.1.1 System Algorithms

Prior to the dummies being included, their optimum position must be known in order to avoid problems at the very first stages of simulation. Finding out these coordinates is the main goal of this subsection. In the initial step, the metal centre ligands will be identified out of all the other atoms in the system. Having done that, the most optimum

orientation of the dummies in regards to RMSD against the initial ligand position, will be determined taking into account the user's given geometry and the ligand positions obtained in the step before. Ultimately, every dummy will be carefully placed in the system while atom types are assigned to them. This is explained more in depth later on in this section.

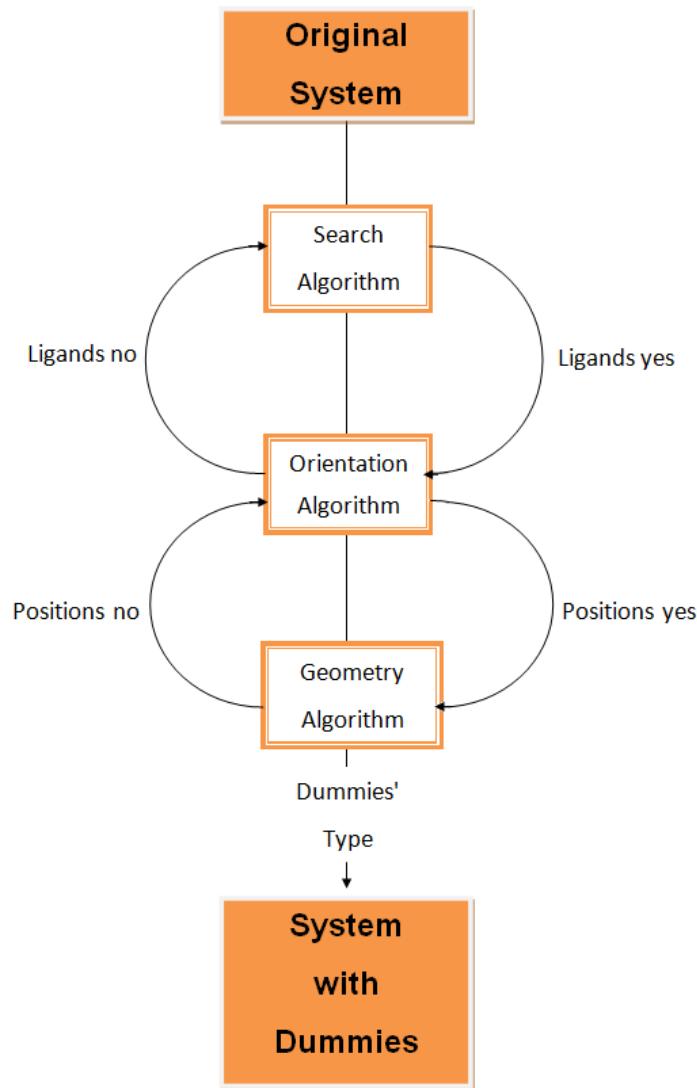


FIGURE 2.2: System algorithms flowchart.

2.1.1.1 Search Algorithm

The heart of the program is a ligand-search algorithm around the metal using UCSF Chimera objects and wrapping them with python. The algorithm starts by looking for the closest atoms around the metal. Once this search is completed, for the list of atoms

selected, the software does the following. First, for every atom in the list, the conditions stated below are checked in the order in which they are written. Once one of these conditions is not fulfilled the atom is discarded as a possible ligand and the function moves on to the next atom. Finally, the algorithm retrieves the chosen ligands as seen in the pseudo-code 1.

Conditions:

- The atom cannot be a metal. Therefore, CaDAS excludes metal-metal bridges.
- The atom valence must be greater than 5 after withdrawing the electron pair which will coordinate to the metal centre. Otherwise, this pair of electrons will be placed in an inner shell being unable to participate in the formation of a chemical bond with the metal centre.

```
valence = (atomic number - 2)%8 < 5
```

- The atom must have a lone pair of electrons or it will not be able to interact with the metal centre.
- The atom cannot be a hydrogen since H atoms do not act as ligands.
- The coordinating distance between metal and ligand must be smaller than 4Å or the interaction between both will not be strong enough to create a chemical bond [23].
- The atom must still have a free coordinating place to bond with the chosen metal.
- The angle of the atom must be close to the requested geometry. Otherwise it will not be possible to create the desired model.

```
for neighbour in neighbours_candidates:  
    if angle(neighbour, candidate_to_ligand, metal) < angle_Cutoff:  
        break
```

Algorithm 1 Search Algorithm Pseudo-Code.

```
ligands_chosen = []
for all ligands:
    if(conditions):
        ligands_chosen.append(current_molecule)
    else:
        current_molecule = next_possible_ligand
return ligands_chosen
```

2.1.1.2 Orientation Algorithm

The atoms that fulfil the conditions from above will be treated as ligands from the studied metal centre. They will undergo an orientation process to find which is the best position to place the dummies, taking into account the desired geometry and the positions of the ligands. This part of the code is mostly geometrical, as half of the dummies are fixed to the right position (on the line connecting the metal centre and the ligand) and the others are rotated around the metal while calculating the RMSD between the ideal position and the one found for each conformation. Once all of them are explored, only the best one is reported. Lastly, the algorithm returns the exact position of all dummies for this geometry as seen in the pseudo code 2.

Algorithm 2 Orientation Algorithm Pseudo-Code.

```
optimum_directions = search_orientation(system)
for direction in optimum_directions:
    direction.length = dummy_metal_bondlength #user input
    dummyposition = metal_position + direction
    dummies_xyz.append(dummyposition)
return dummies_xyz
```

2.1.1.3 Geometry Algorithm

The final step is to obtain the right dummies' position with the right atom types. An atom type is a label that defines how an atom will interact with its neighbours. In order to do that every atom type must bring angles, torsions and hybridisation associated. A mistake on these labels could lead to an undesired system topology.

To achieve the correct correlation between dummy position and type, CaDAS labels a dummy with a given type. The next step is to find the dummies that are equivalent in

type, which means they are found at a specified geometry with respect to the labelled atom. The atoms that fulfilled this condition are labelled with the same type as the reference atom, as seen in figure 2.3. This is achieved by iterating through all the other dummies coordinates searching for an approximated 180 degree angle. In case there is a unique atom (only atom within its type) as in the square pyramid, a unique label is assigned to it. This works for all the geometries of the software except the tetrahedral one, which does not need atom type assignation, due to the fact that all atoms are equivalent and consequently the same type. Once all dummies are set into place, the final topology of the system is extracted as a pdb file. Having done that, the system is now ready to be parameterized in the next section of the software.

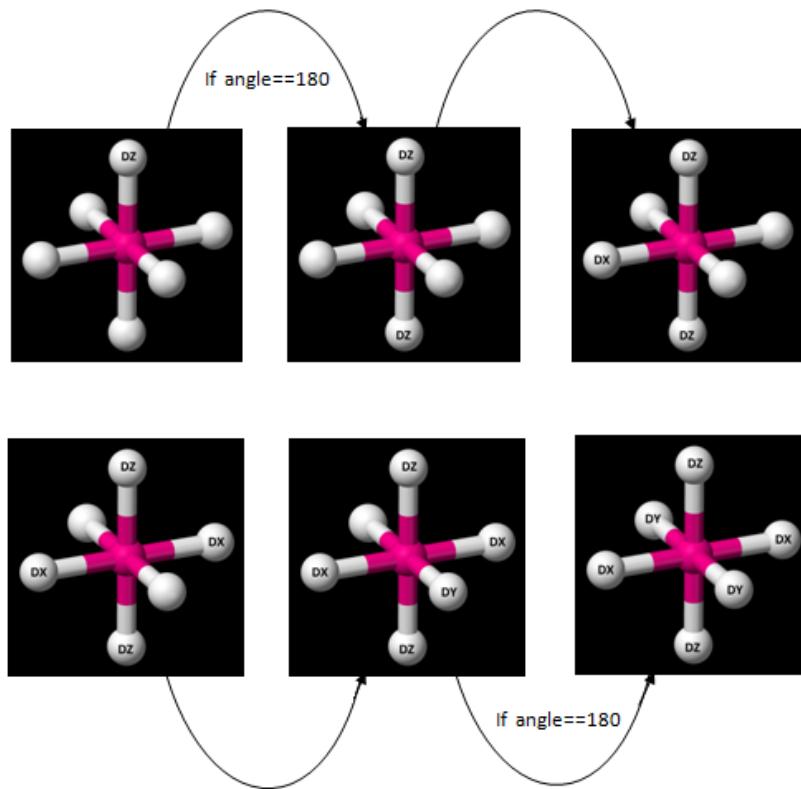


FIGURE 2.3: Demonstration of the geometry algorithm for an octahedron.

2.1.2 System Properties

This section intends to yield the metal centre properties and not the whole system properties. To do that, two processes are applied to the original system. First, a new virtual system including the metal centre and the dummy atoms is created on the computer's RAM memory. In the following step, the virtual system is modelled to the

users preference by including charge on the dummies, emulating orbital density, the metal-dummy connectivity as well as bonds between dummies, to keep them packed together. Finally, a tLEaP command will record all the preceding information in a .lib file. Second, a template file of format .frcmod, which contains all the metal system parameters such as mass, angles and torsions is filled in with the user's inputs [24]. Initially setting little portions of mass to the dummies, making them susceptible to be affected by Newtonian forces (which otherwise would be null and no movement would occur) and then, describing all the bonds, angles and torsions for each dummy type.

It is important to remember the fact that every time some portion of mass or charge is added to a dummy, it is, at the same time, deducted from the metal centre in order not to change the system's properties.

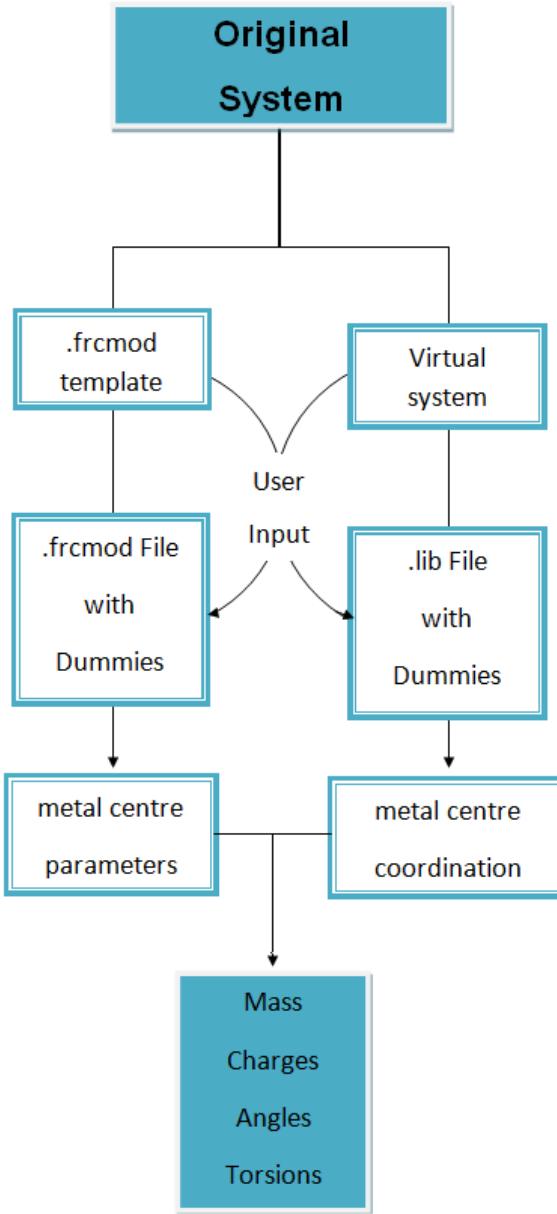


FIGURE 2.4: System properties flowchart.

2.1.3 System Production

The final phase of the software gathers the .lib, .frcmod and .pdb files in one single input ready to be launched by the MD simulation program OpenMM. This is performed with tLEaP, which combines topology, coordinates, charges and all the remaining parameters of the system to create a unique simulation input divided into two files, the final topology

and the system coordinates. In addition, a .log error file is output in case the software lacks information on the system. All the previous is summarised in the figure 2.5.

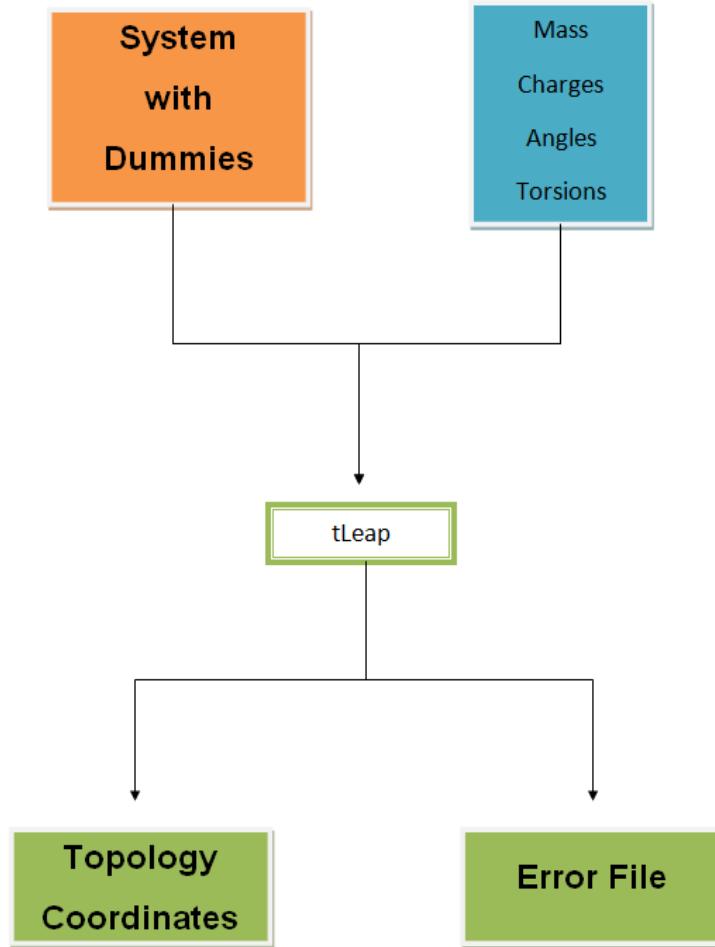


FIGURE 2.5: System production flowchart.

Chapter 3

Results

This chapter sets out the results obtained during this project. It initially assesses the test outcomes obtained through MD simulations on two different systems. The last section reports the current state of the project and its possible future improvements.

3.1 Benchmarking CaDAS

3.1.1 Binding site model of a modified carbonic anhydrase.

Carbonic anhydrases are a family of metallo-enzymes that catalyse the rapid interconversion of carbon dioxide and water to bicarbonate and protons to maintain acid-base balance in animals and humans [25]. This reversible reaction occurs at a relatively slow pace in the absence of a catalyst. The active site of these proteins is formed by a zinc ion, which interacts with the nitrogens of three histidine residues and the oxygen of a water or hydroxide molecule [26].

In 1999, the group of Dr. Pang Y.P. published the structure of the binding site of a carbonic anhydrase [27] where one of the histidines was substituted by a cysteine residue, that interacts with the metal centre through a sulfur atom as seen in the figure 3.1. This model represents a small peptide that we used for software testing.

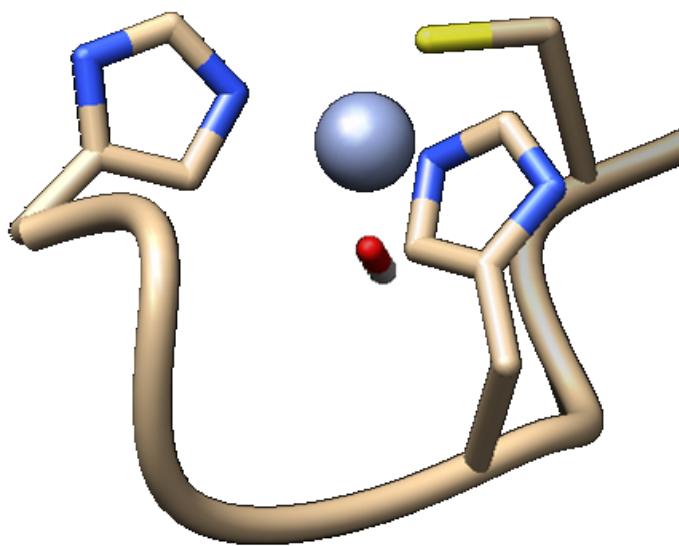


FIGURE 3.1: Modified carbonic anhydrase active site.

For this calculation, the metallopeptide is parametrized with CaDAS. It uses a tetrahedral geometry of the zinc so that it interacts through the dummy atoms with the two nitrogen atoms of the histidines, an hydroxyl and the sulfur atom of a cysteine. In this parametrization the dummy atoms coming from the metal are oriented at the beginning of the simulation to maximize geometric overlap of the metal-dummy vector with the dummy-ligand atom vector, producing an MD input for the system shown in figure 3.2.

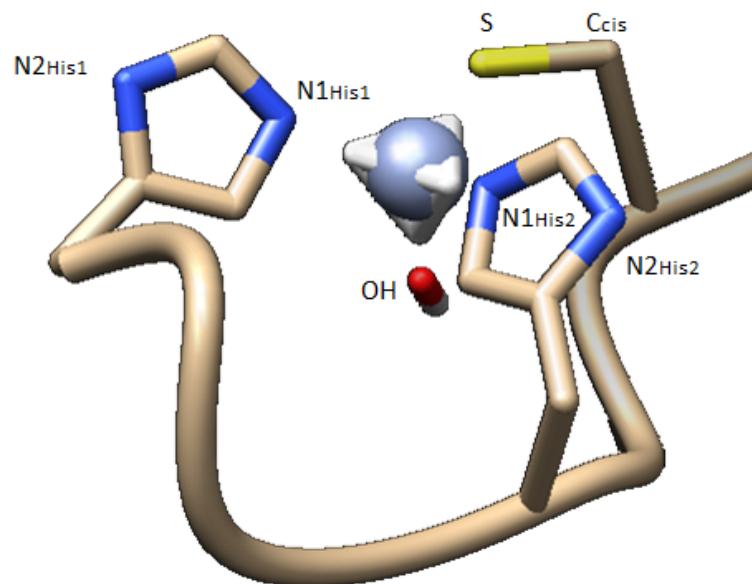


FIGURE 3.2: Modified carbonic anhydrase active site with dummy atoms.

This system was submitted to a 100 nanoseconds MD. In this case, the root mean square deviation (RMSD) is the measure of the average squared error between the current and the initial position of the system. Therefore, the stability of this parameter reflects structural stability over time. Three RMSDs were computed, the first including all the atoms of the system (blue lines in figure 3.3), a second with excluded water molecules (yellow in figure 3.3) and a last with atoms in a range of 5 Å from the active site (in green in figure 3.3).

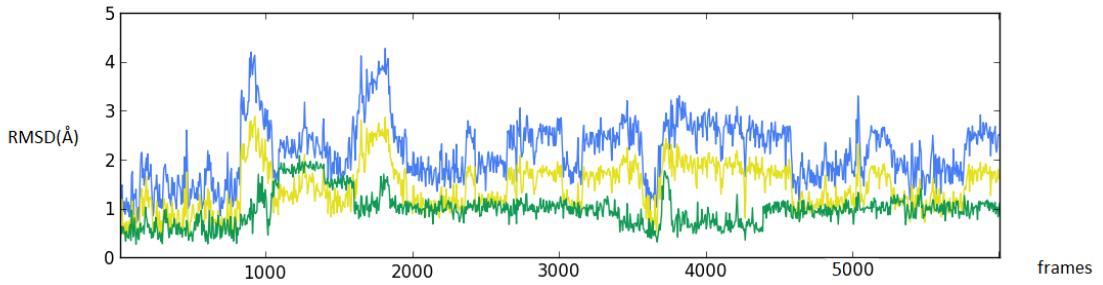


FIGURE 3.3: RMSD values vs frames.

The full RMSD fluctuates considerably towards the end of the trajectory because of the movement of molecules of water. Due to this fact, the water excluded RMSD curve is slightly steadier. Finally, the active site RMSD is stable in time as it does not vary from the frame 2000 on. These facts indicate our system is stable in time even though, one can not affirm the structure is valid until a more specific coordination geometry analysis has been made.

Next, the distances between the metal centre and its four ligands were carefully studied. They are reported in the table 3.1, accompanied by the experimental crystallographic bond lengths extracted from the crystal structure of human carbonic anhydrase II (2cba) [28]. The Zn-S bond length is obtained from a survey of zinc protein crystal structures [18]. The fact that all experimental data falls inside the confidence interval of the computed values, supports the hypothesis that this system is feasible.

Bond	Computed Bond Length (Å)	Experimental Data (Å)
Zn - S	2.2±0.3	2.3±0.1
Zn - N1 _{His1}	2.0±0.2	2.10±0.25
Zn - N1 _{His2}	2.0±0.2	2.11±0.25
Zn - O	1.9±0.2	2.05±0.25

TABLE 3.1: Comparison between computed metal centre-ligand distances and experimental data.

Next, the results extracted from the study of the angles between the metal centre and its ligands are presented in the tables 3.2 and 3.3. All the values contained in these tables are in the neighbourhood of 180°, which means the position of the chain of the ligand residue with respect to the metal centre is almost vertical. These angles create a good orientation for an effective interaction between metal and ligands.

Angle	Value
Zn - DM - S	(172±20)°
Zn - DM - N _{His1}	(171±20)°
Zn - DM - N _{His2}	(164±30)°
Zn - DM- O	(176±21)°

TABLE 3.2: Computed angles for the indicated atoms, which are labelled in figure 3.2.

Angle	Value
Zn - S - C _{Cis}	(157±10)°
Zn - N1 _{His1} - N2 _{His2}	(170±20)°
Zn - N1 _{His2} - N2 _{His1}	(169±30)°
Zn - O - H	(159±21)°

TABLE 3.3: Computed angles for the indicated atoms, which are labelled in figure 3.2.

Finally, the simulation trajectory over time is checked. When first studying the system, one could observe that the dihedral angle of the sulfur atom with respect to the metal centre was slightly smaller than awaited. Therefore, a more exhaustive study on the dihedral angle Zn-S-C1_{Cis}-C2_{Cis} was performed and presented in figure 3.4. Again, the

computed value (150°) proves how the metal and the ligand are in a good orientation for interaction. However, from the frame 1000 to the frame 1600 the cisteine residue shows a small but abrupt angle variation, which explains the increase on the RMSD showed on the figure 3.3. This phenomenon is not very relevant as this variation is found on the early stage of the simulation, where some ligands might not have been in the most efficient orientation with respect to the metal centre. In addition, the fact that no other abnormal variation is found after this event, supports this theory.

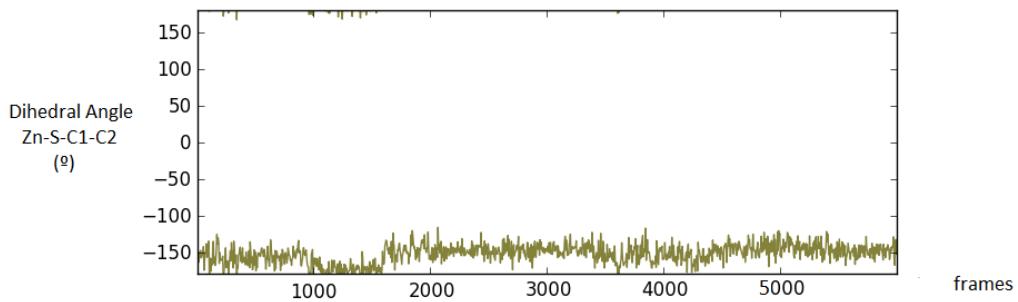


FIGURE 3.4: $\text{Zn-S-C1}_{Cis}\text{-C2}_{Cis}$ dihedral angle value over time.

All in all, the results show a stable tetrahedral Zn^{2+} structure with four defined ligands, where the computed bond length between ligands and metal yielded results close to the experimental data.

3.1.2 GPG siderophore system

Siderophores are amongst the strongest soluble Fe^{3+} iron-chelating compounds secreted by microorganisms. However, they tend to release the iron when the metal centre is reduced to Fe^{2+} [29][30]. In this section, GPG, an artificial siderophore compounded by three catechols inter-connected through two amino skeletons containing a pyrrole ring (figure 3.5), tries to chelate a Fe^{2+} ion in an octahedral geometry and serves as initial system for software testing.

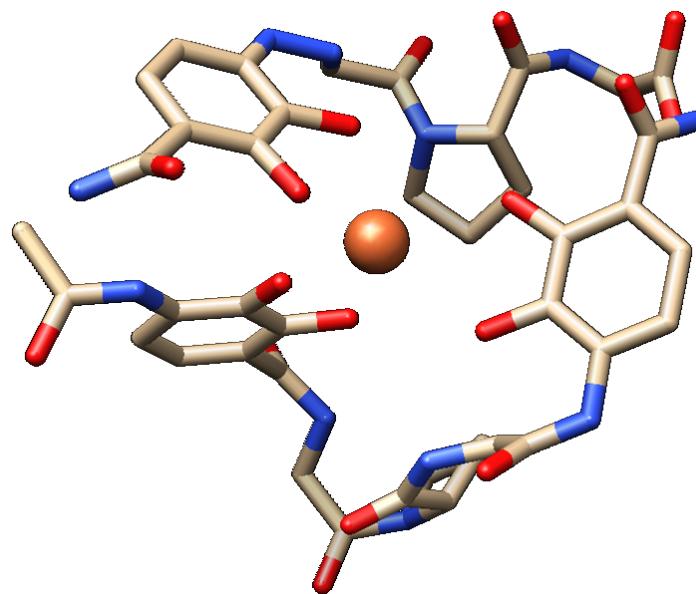


FIGURE 3.5: GPG system.

For this calculation, the siderophore is parametrized with CaDAS. It uses an octahedral geometry of the iron centre so each dummy interacts with a single oxygen atom coming from one of the catechols. As the section before, the dummy atoms were oriented before the 100ns simulation of the system shown in figure 3.6. The RMSDs obtained are presented on the figure 3.7. Where the dark green curve is the RMSD calculated including all the atoms of the system, the light green excludes water molecules and the blue one is determined with the atoms in a range of 6Å from the metal centre.

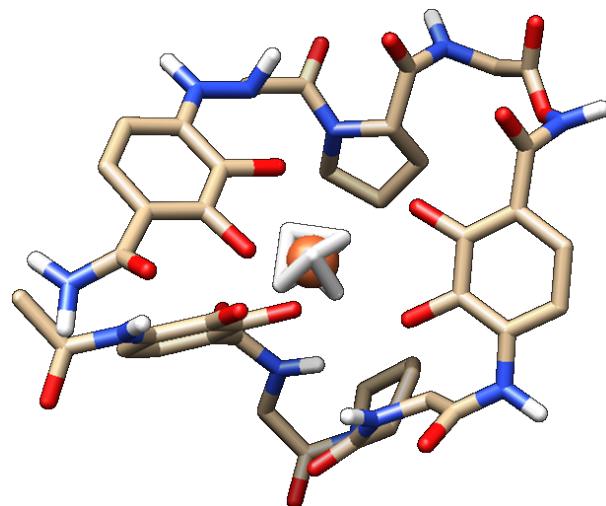


FIGURE 3.6: Dummy atoms in GPG.

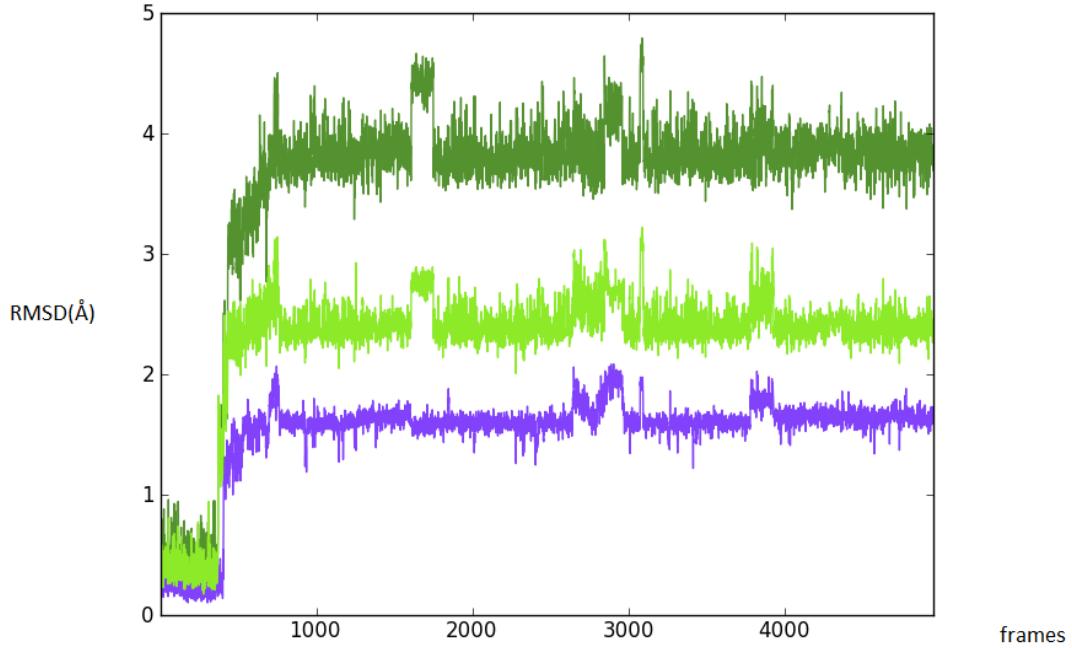


FIGURE 3.7: RMSD values vs frames.

The full RMSD has considerably high values due to the large amount of atoms included in the calculation. Nevertheless, one can appreciate a stabilisation tendency towards the end of the simulation. As expected, the RMSD without the water molecules presents smaller values than the one before. Ultimately, the radius restricted plot (in blue) clearly exposes the stabilisation process that the system undergoes. However, two brusque variations are appreciated in the later half of the simulation, which evidence the need for further studies. In order to analyse this in more depth, the frame intervals presenting these deviations will be studied separately. The first anomaly is found from the frame 2600 to the 3000 where one of the catechols stops interacting with the metal centre and moves away from it as seen in figure 3.8. The same happens from frames 3800 to 4100 with another catechol head.

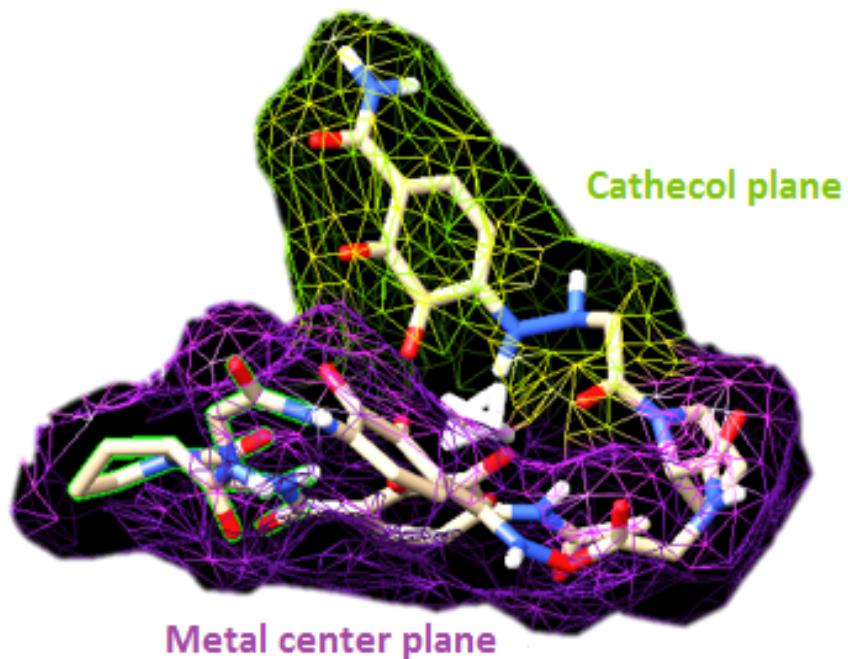


FIGURE 3.8: Catechol displacement from the metal centre alignment.

In addition, the metal centre presents a substantial geometrical distortion and looks like it only interacts with four ligands instead of six. Also, some catechols do not interact with the metal through the two hydroxide groups. Instead, they display interaction with one hydroxide and the oxygen atom coming from the closest amide group. The previous facts can be seen in figure 3.9. These lead to believe the system is not stable in a octahedral geometry and it is cyclically trying to switch to a geometry with a lower coordination number.

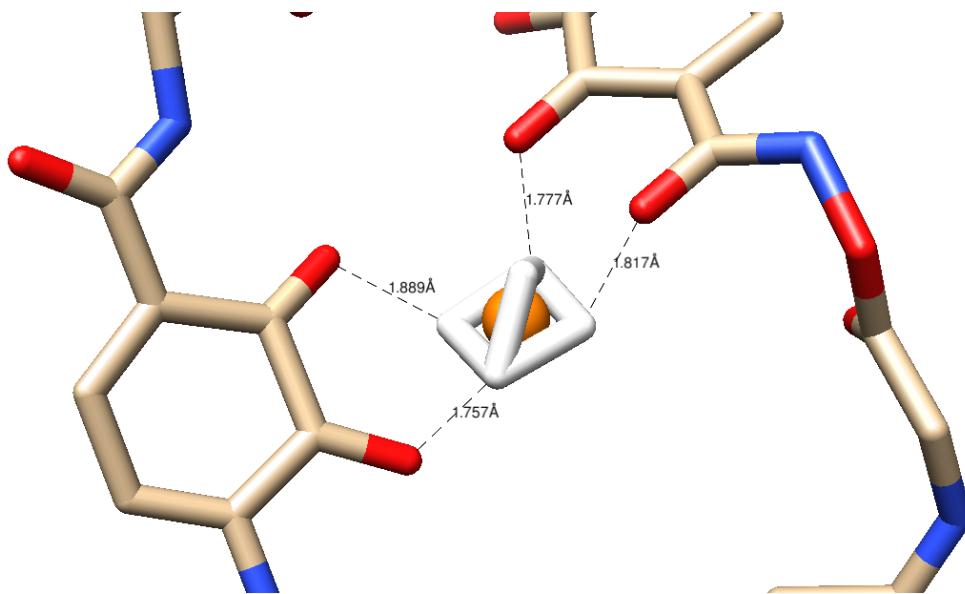


FIGURE 3.9: Momentary geometry of the ligands around the metal centre with coordination number of 4.

Even though, the RMSD plot looks stable, the two previous variations of this plot show a periodical unexpected behaviour of the system. In order to conclude whether or not the system is viable, a coordination geometry study over all the simulation is performed. The computed data obtained from this analysis is reported in table 3.10.

Atom	DZ - Atom (Å)	Fe - Atom (Å)
OH ₁	(2.4±1.3)	(2.4±0.4)
OH ₂	(5.0±2.1)	(3.9±1.3)
OH ₃	(5.1±1.5)	(5.1±0.3)
OH ₄	(2.4±2.0)	(2.3±0.3)
OH ₅	(2.5±1.2)	(2.3±0.3)
OH ₆	(2.4±1.3)	(2.5±0.4)
O ₁	(2.4±1.2)	(2.3±1.7)

TABLE 3.4: Dummy-Ligand bond distance and Metal-Ligand bond distance.

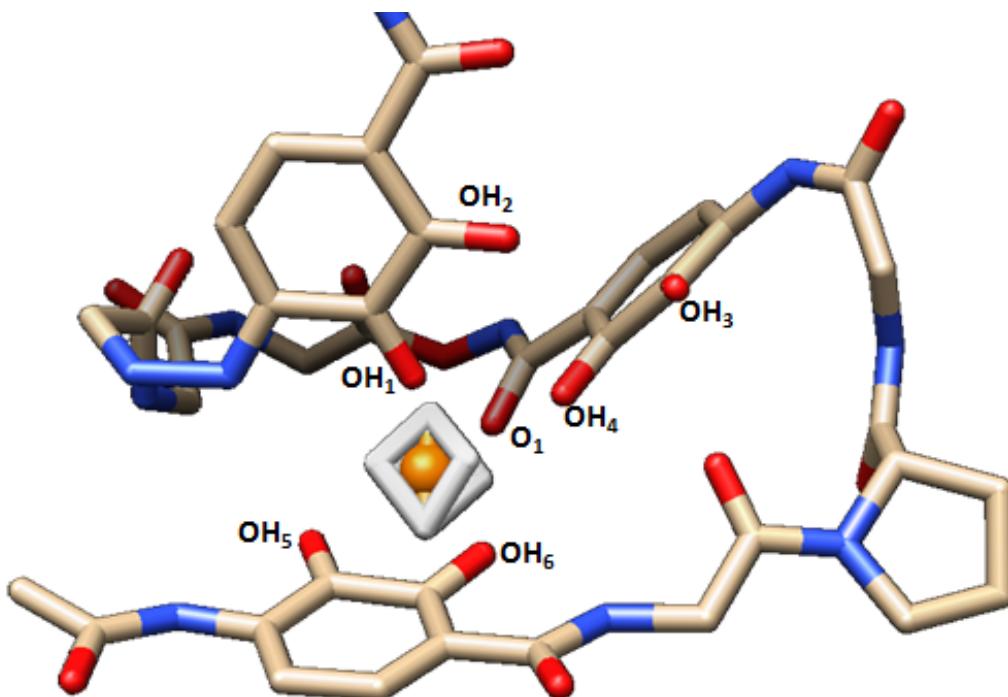


FIGURE 3.10: Labelling of atoms in table 3.3.

The values above show that the metal centre has no special binding affinity for any of the ligands. The dummy atom interacts weakly with one oxygen and consequently switches to another. Moreover, the results above support the fact that one of the catechols is excessively far from the metal centre without any possibility of bonding. All these proofs demonstrate the instability of the system making it unviable.

On the whole, the results presented in this section show an implausible structure in agreement with well known experimental and bibliographical data. Since the oxygen donor atoms that are mainly used in siderophores for iron coordination represent hard Lewis bases that allow additional strong ionic interactions between metal and ligand. Thus, siderophores display an enormous affinity towards Fe(III). However, in case of the borderline Lewis acid Fe(II), this one prefers interaction with softer donor atoms such as nitrogen or sulfur, which are only partially employed for iron coordination in natural siderophores. Furthermore, the higher electron density of Fe(II) is poorly compensated for in the oxygen-donor siderophore complexes. What is more, the low ratio of charge to ionic radius of Fe(II) compared with that of Fe(III) might be additionally unfavourable to maintain the optimal complex geometry as seen in 3.9. Thus, the capability of siderophores of forming stable complexes with Fe(II) is rather low [29].

3.2 Future Improvements

So far, the software has only been successfully applied to two systems. However, the cationic dummy atoms software promises to do many more things. For instance, one could attain to study the rapid ligand exchange mechanism in ligand substitution reactions. In order to decide whether or not it is worth to make further calculations to determine the binding energy between one coordinate or the other, in this way coming to understand the reasons behind the process, and assessing the viability of this.

On the other hand, the software has not yet been tested on large scale systems with several metal nuclei. In order to address this, a cage-shaped amino skeleton, which interacts with two zinc metal centres and eight palladiums as seen in the figure 3.11, is currently being parametrized with CaDAS.

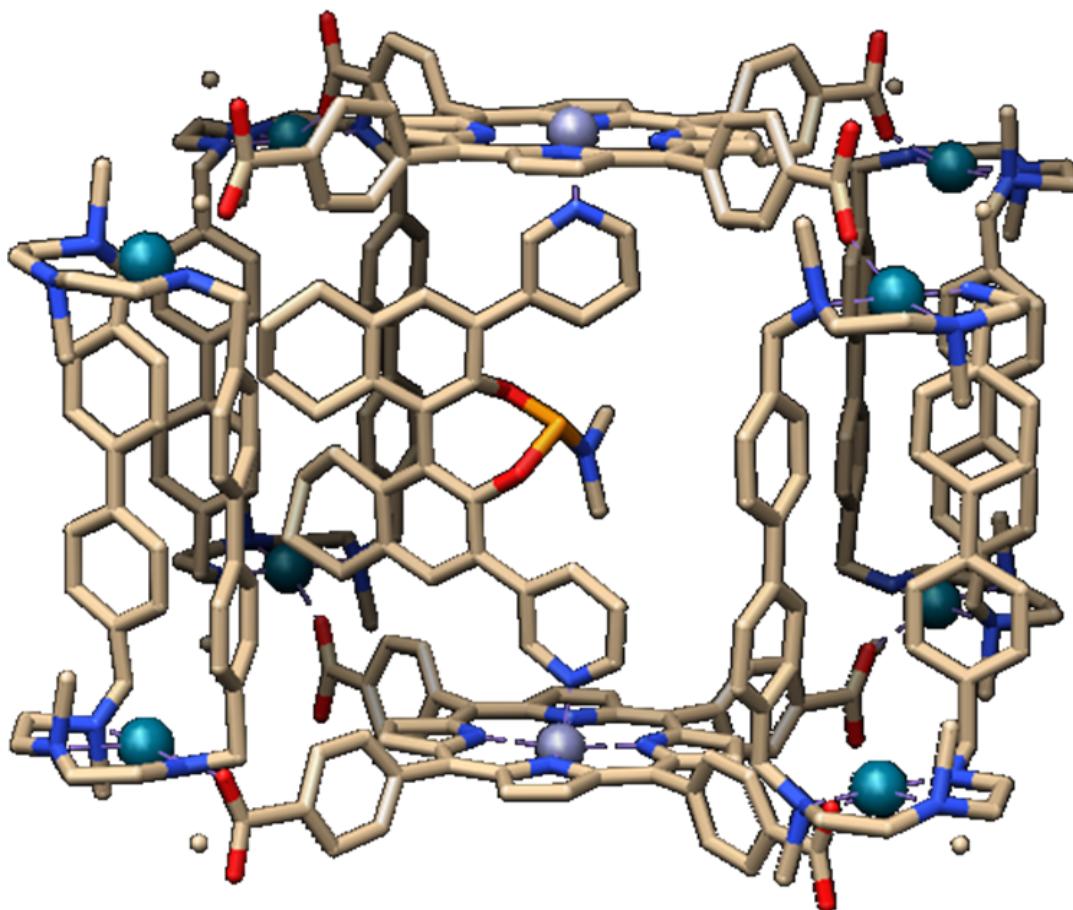


FIGURE 3.11: Cage system.

One future improvement to be made is to give users the possibility to decide how to divide the charge. This can be applied, for example, in the case the system to be studied contains an exceptionally electronegative atom. Another noticeable upgrade could be to let the user choose each dummy-metal bond length, breaking the symmetry imposed by the software, making the model more realistic and more able to emulate distortions like the Jahn–Teller effect.

Last but not least, to date, CaDAS has only been applied to divalent ions. Parameters have never been tried for highly charged ions, which is for sure a compelling future research direction for this software.

Chapter 4

Conclusion

Cationic Dummy Atom Software was able to simulate the four-ligand coordination of a zinc complex inside a modified carbonic anhydrase. The distances $(2.0 \pm 0.2)\text{\AA}$ and $(2.0 \pm 0.2)\text{\AA}$ between the metal centre and the two histidines are close to the one reported from the crystallographic data $(2.10 \pm 0.25)\text{\AA}$ and $(2.11 \pm 0.25)\text{\AA}$. In addition, the bond length to the metal centre of the introduced sulfur atom $(2.2 \pm 0.3)\text{\AA}$ matched the average distance between sulfur and zinc in crystallographic structures $(2.3 \pm 0.1)\text{\AA}$.

On the other hand, CaDAS displayed no binding affinity between the metal centre and the surrounding coordinates of an octahedral Fe^{2+} siderophore, proposing that the considered geometry is implausible in agreement with the experimental data.

The results show potential applicability of the software in new drug design. However, supplementary research needs to be made in order to verify and extend the results obtained in this project.

Appendix A

Appendix

A.1 Carbonic anhydrase system

A.1.1 Computed metal-ligand bond distances over time:

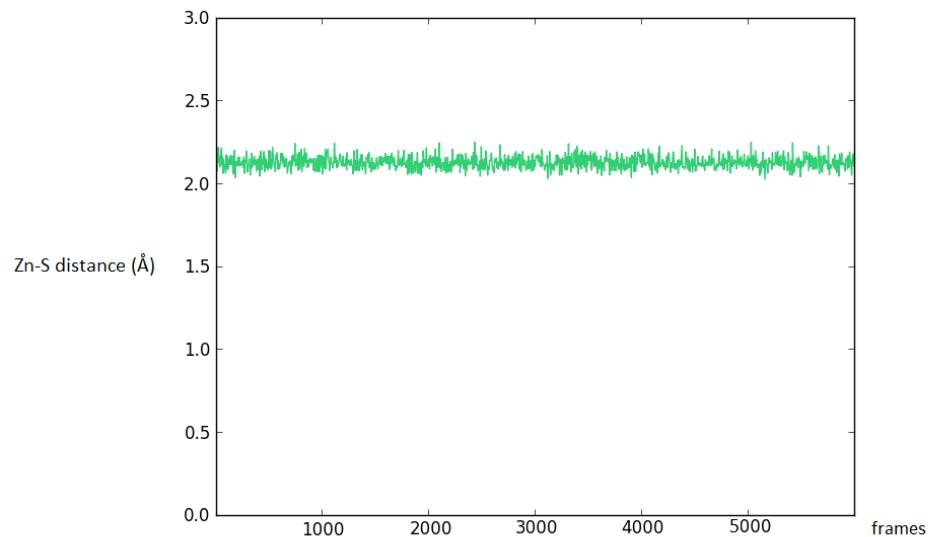
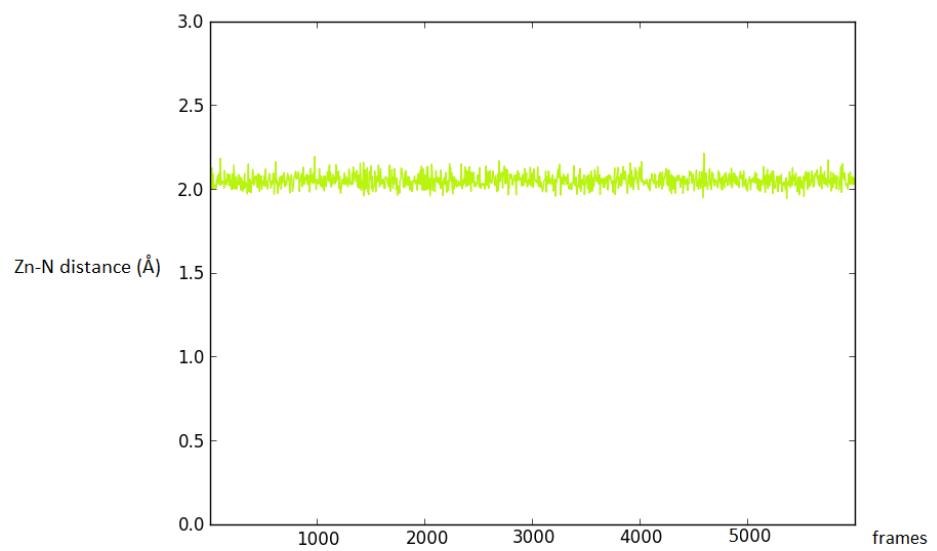
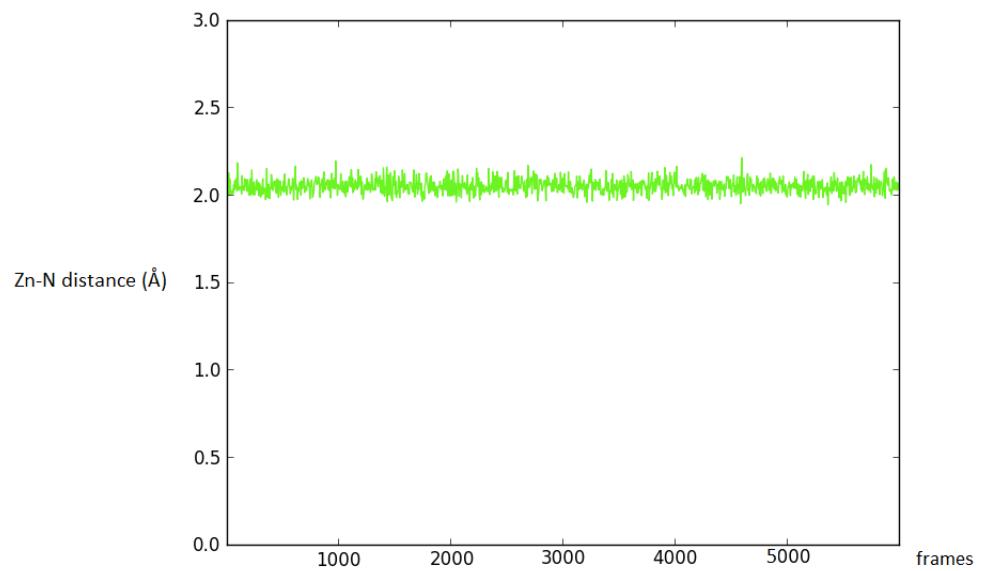


FIGURE A.1: Zn-S bond length

FIGURE A.2: Zn-N_{His1} bond lengthFIGURE A.3: Zn-N_{His2} bond length

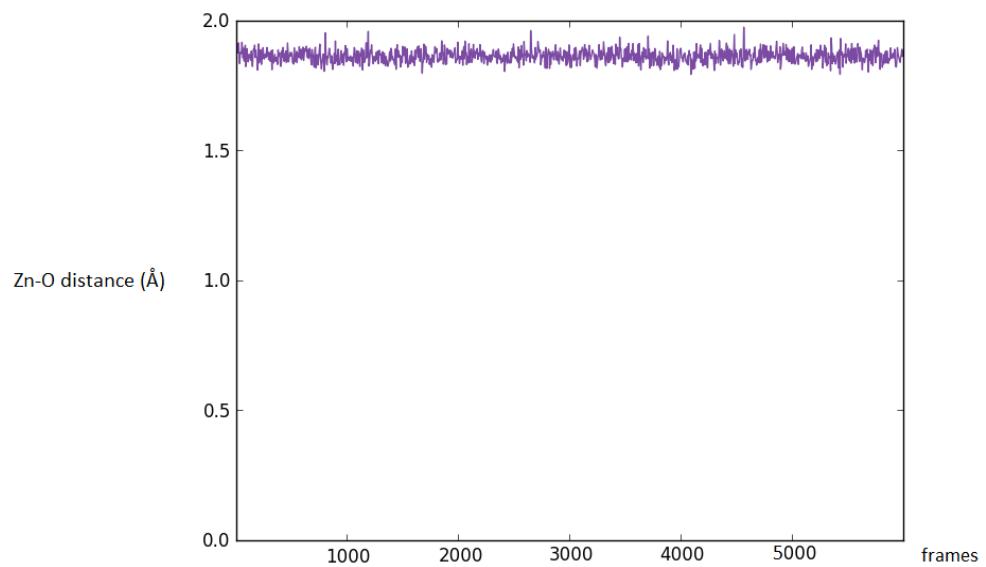


FIGURE A.4: Zn-OH bond length

A.1.2 Computed metal-dummy-ligand angles over time:

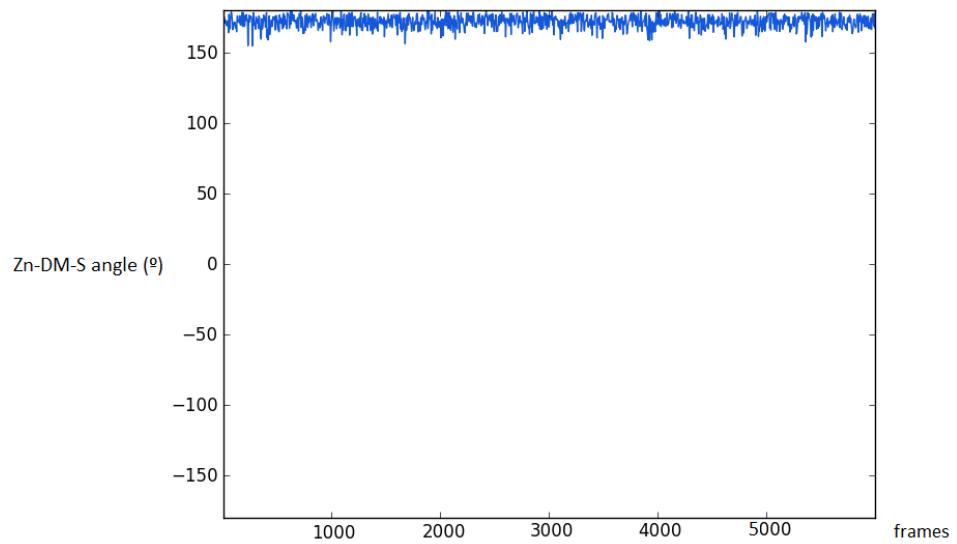
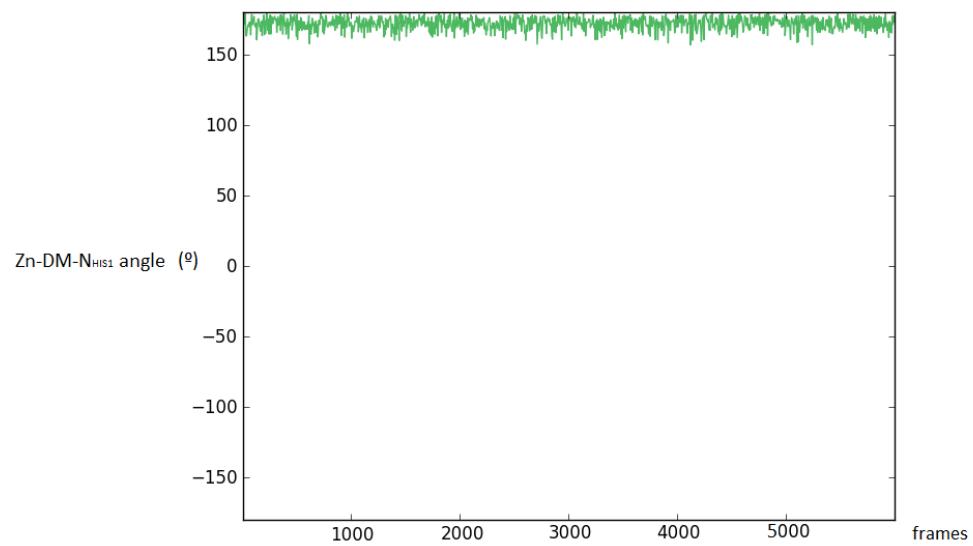
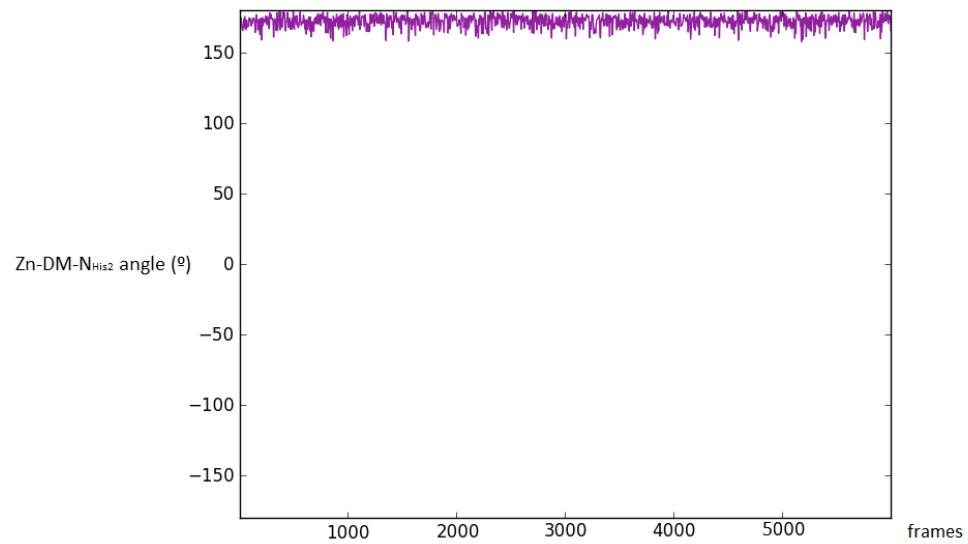


FIGURE A.5: Zn-DM-S angle

FIGURE A.6: Zn-DM-N_{His1} angleFIGURE A.7: Zn-DM-N_{His2} angle

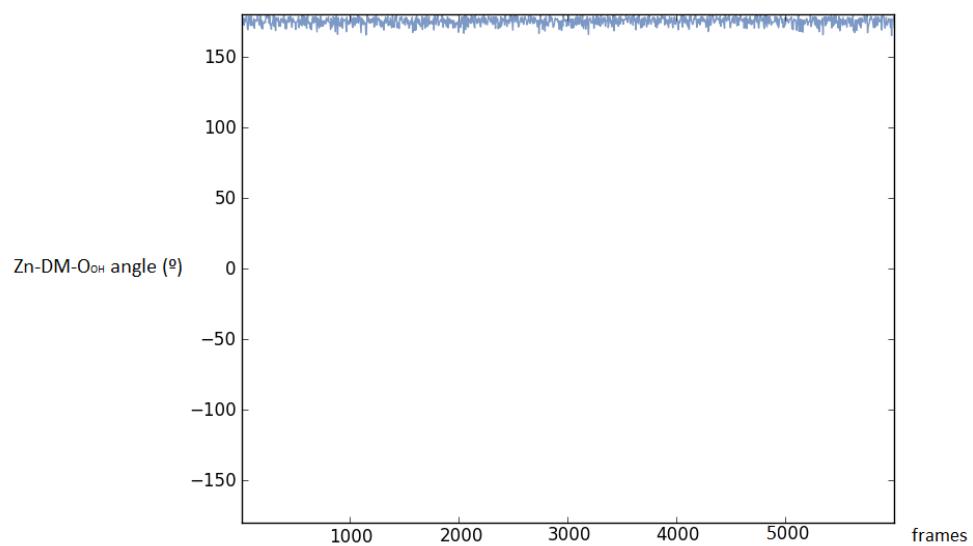


FIGURE A.8: Zn-DM-OH angle

A.2 GPG system

A.2.1 Computed metal-ligand bond distances over time:

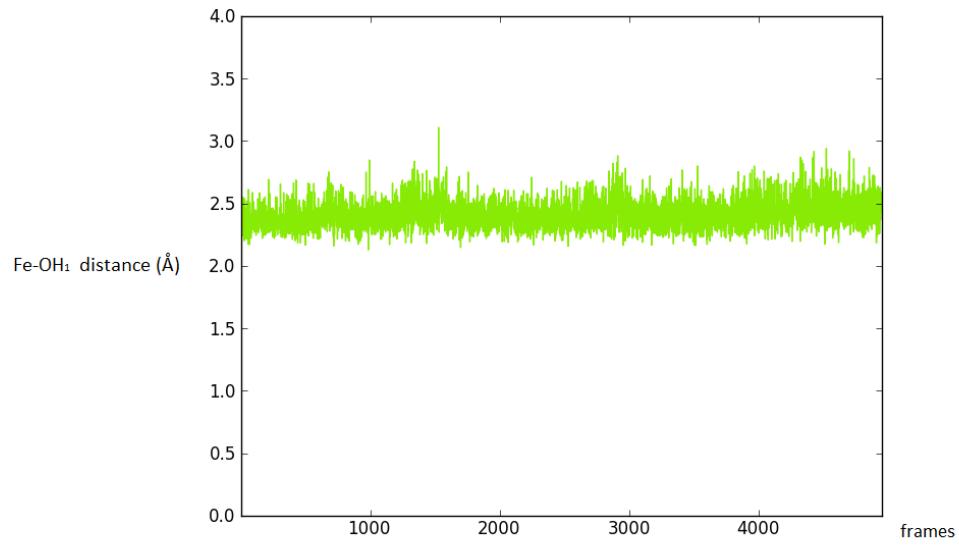
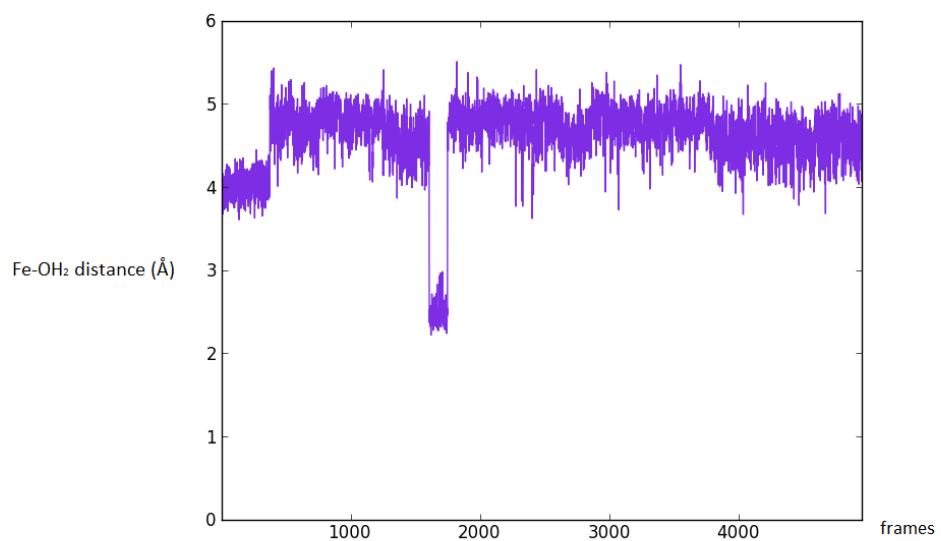
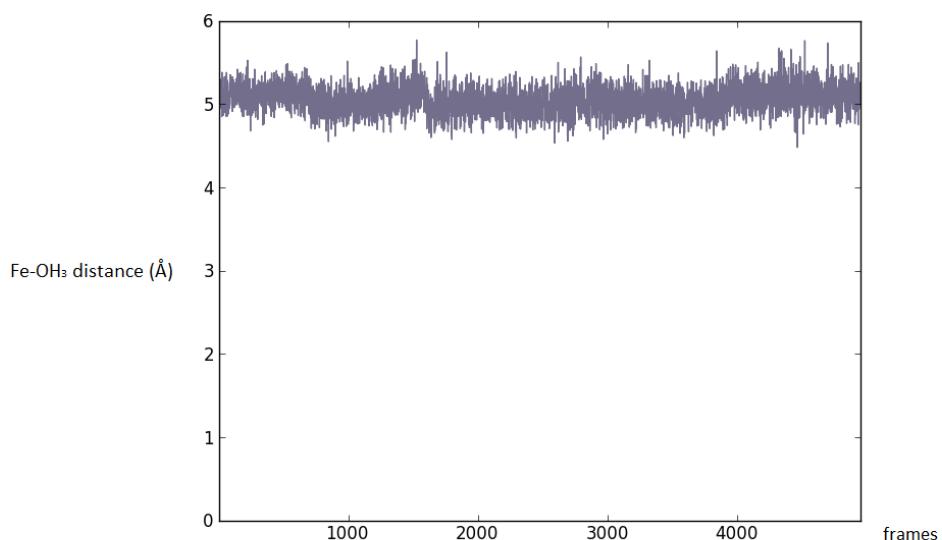
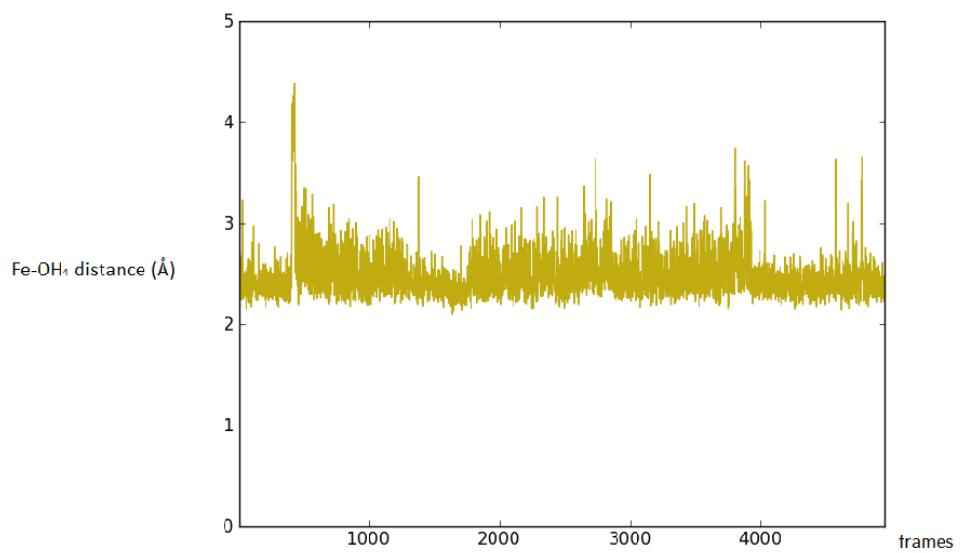
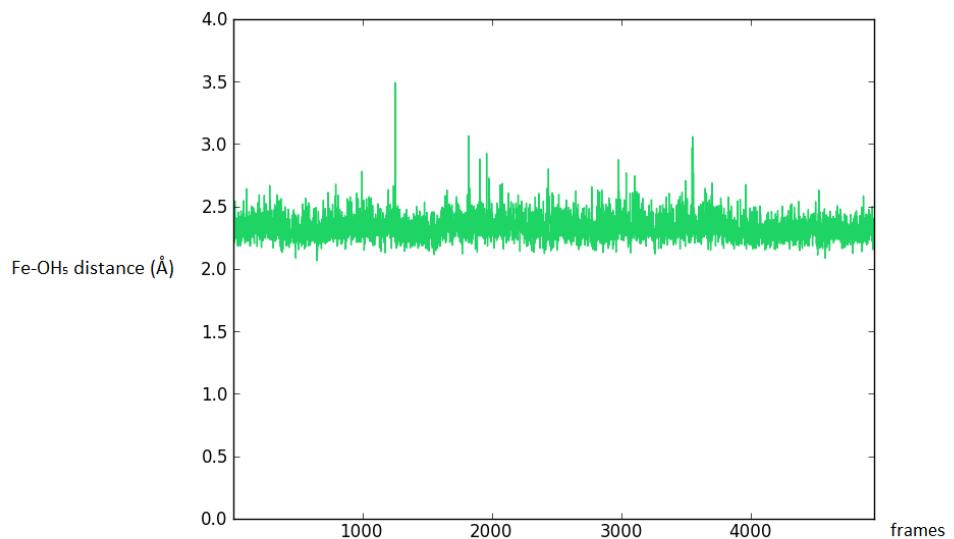
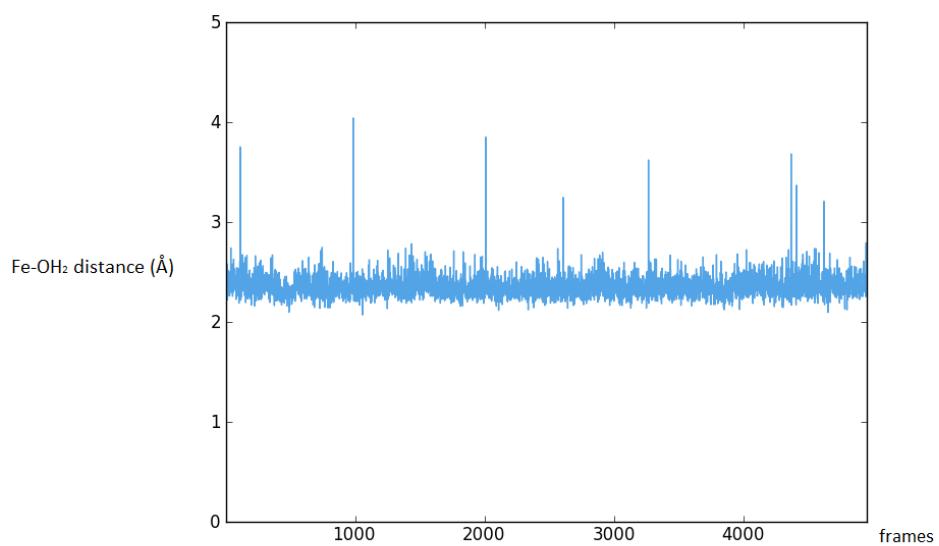
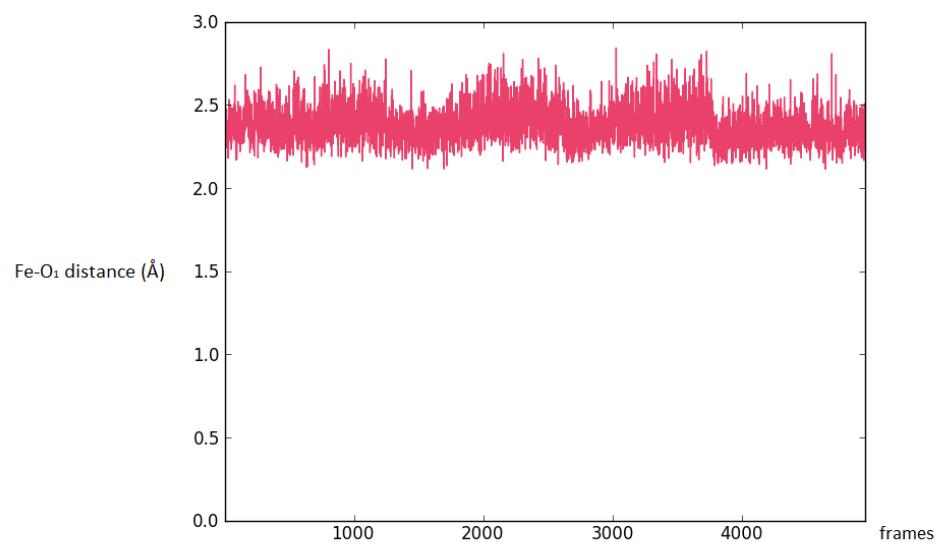


FIGURE A.9: Fe-OH₁ bond length

FIGURE A.10: Fe-OH₂ bond lengthFIGURE A.11: Fe-OH₃ bond length

FIGURE A.12: Fe-OH₄ bond lengthFIGURE A.13: Fe-OH₅ bond length

FIGURE A.14: Fe-OH₆ bond lengthFIGURE A.15: Fe-O₁ bond length

A.2.2 Computed dummy-ligand bond distances over time:

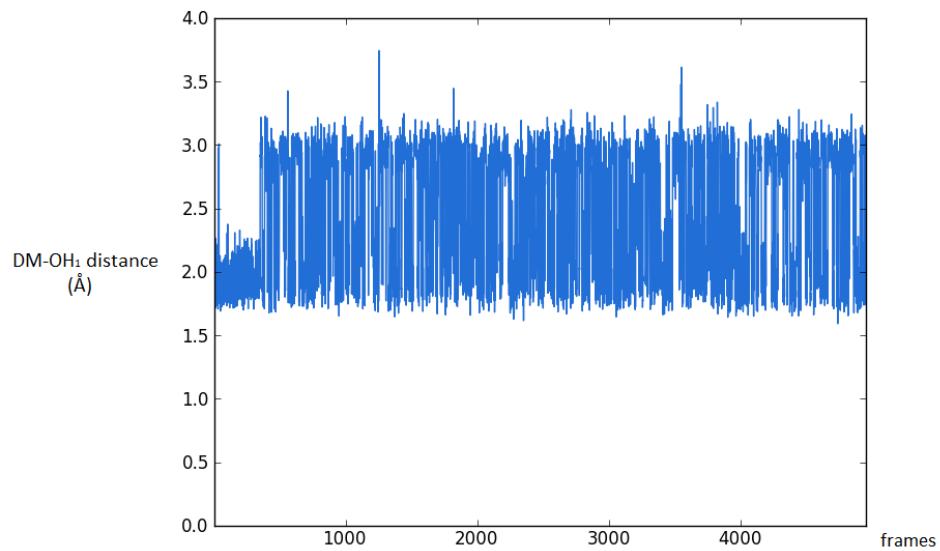


FIGURE A.16: DM-OH₁ bond length

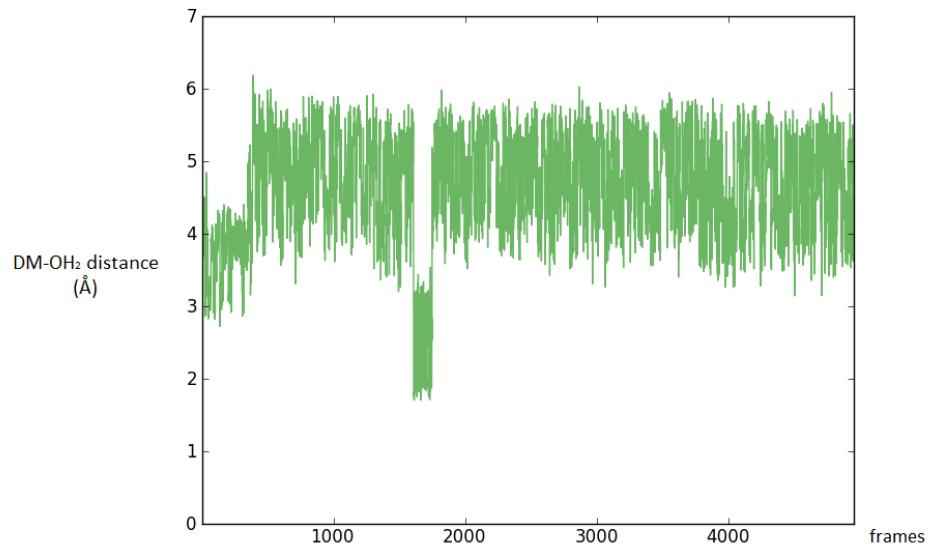


FIGURE A.17: DM-OH₂ bond length

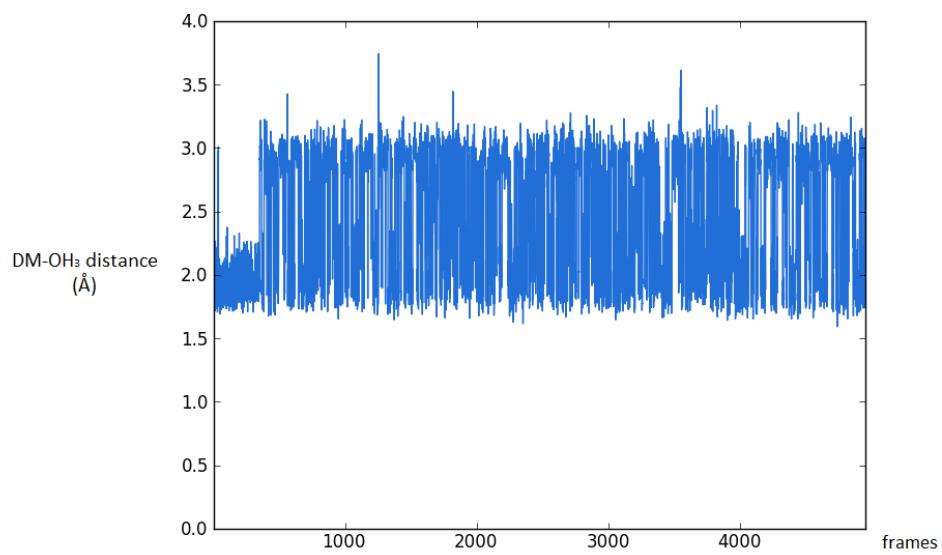


FIGURE A.18: DM-OH₃ bond length

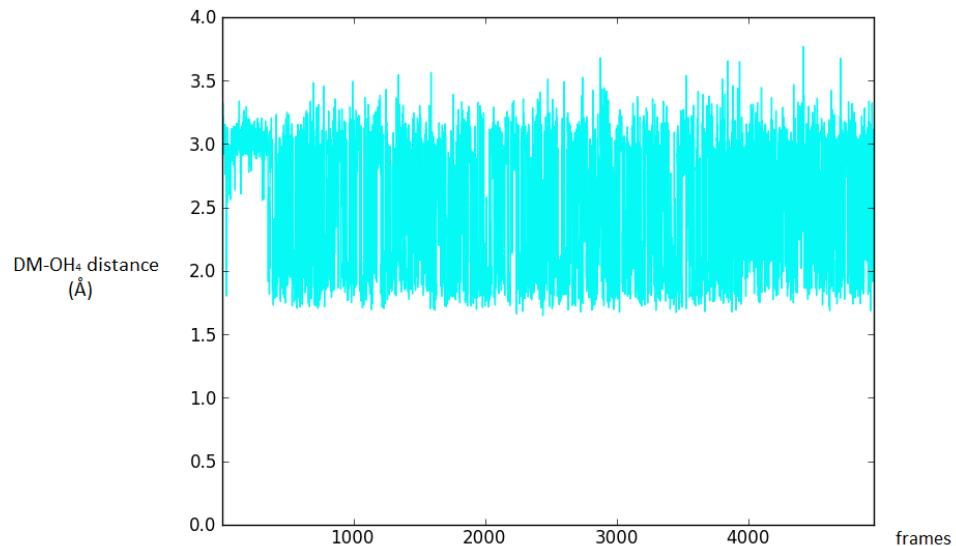


FIGURE A.19: DM-OH₄ bond length

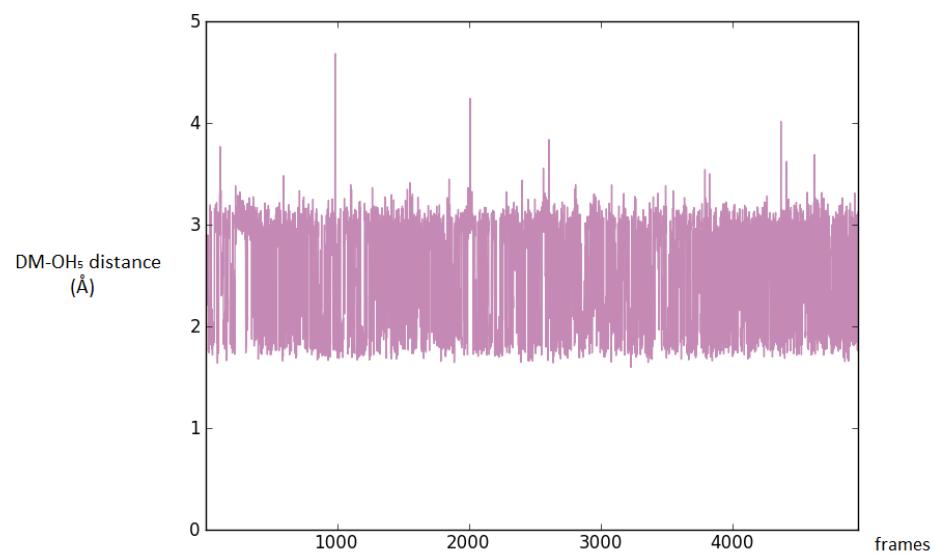


FIGURE A.20: DM-OH₅ bond length

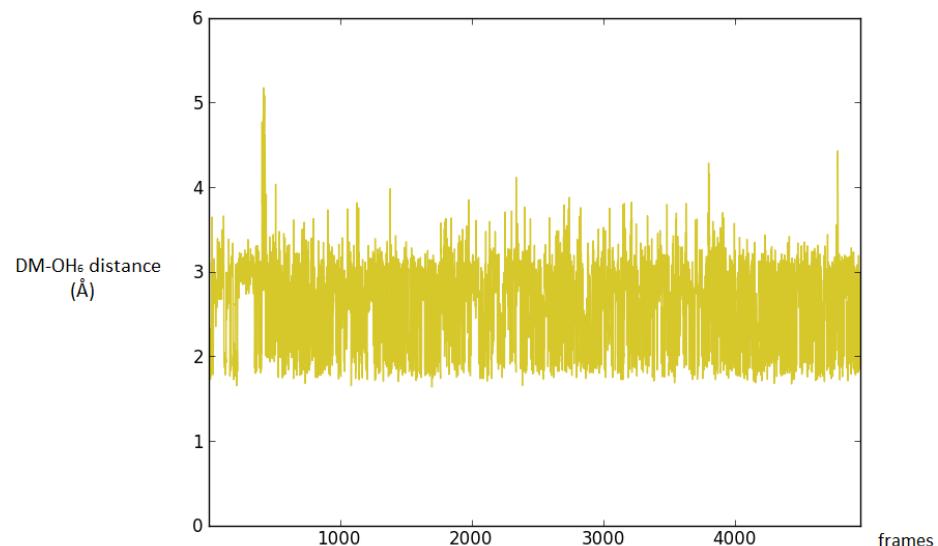


FIGURE A.21: DM-OH₆ bond length

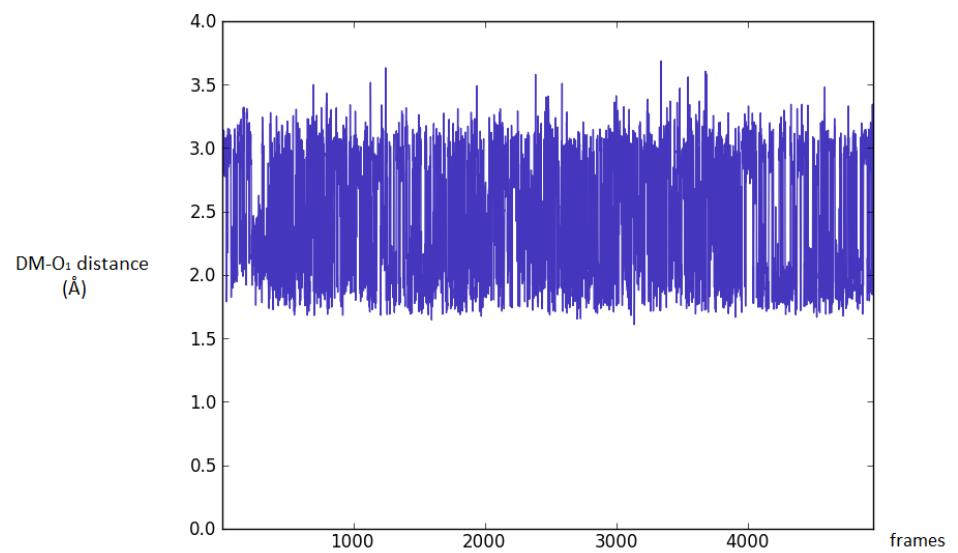


FIGURE A.22: DM-O₁ bond length

Bibliography

- [1] Denis Wirtz. *Cell biophysics*. Comprehensive biophysics / editor-in-chief Edward H. Egelman, vol. 7. Elsevier, Academic Press, 1 edition, 2012.
- [2] David W Christianson and J David Cox. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. *Annual review of biochemistry*, 68(1):33–57, 1999.
- [3] TV OHalloran. Roles of transition metals in signal transduction and gene expression. In *FASEB journal*, volume 10, pages 2125–2125. Federation AMER SOC EXP BIOL 9650 rockville pike, bethesda, MD 20814-3998, 1996.
- [4] Chérif F Matta. How dependent are molecular and atomic properties on the electronic structure method? comparison of hartree-fock, dft, and mp2 on a biologically relevant set of molecules. *Journal of computational chemistry*, 31(6):1297–1311, 2010.
- [5] Yuan-Ping 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 Bioinformatics*, 45(3):183–189, 2001.
- [6] Pengfei Li and Kenneth M Merz Jr. Metal ion modeling using classical mechanics. *Chemical reviews*, 117(3):1564–1686, 2017.
- [7] Fernanda Duarte, Paul Bauer, Alexandre Barrozo, Beat Anton Amrein, Miha Purg, Johan Åqvist, and Shina Caroline Lynn Kamerlin. Force field independent metal parameters using a nonbonded dummy model. *The Journal of Physical Chemistry B*, 118(16):4351–4362, 2014.

- [8] Yang Jiang, Haiyang Zhang, Ziheng Cui, and Tianwei Tan. Modeling coordination-directed self-assembly of m2l4 nanocapsule featuring competitive guest encapsulation. *The Journal of Physical Chemistry Letters*, 8(9):2082–2086, 2017.
- [9] Maria Carola Colombo, Leonardo Guidoni, Alessandro Laio, Alessandra Magistrato, Patrick Maurer, Stefano Piana, Ute Röhrig, Katrin Spiegel, Marialore Sulpizi, Joost VandeVondele, et al. Hybrid qm/mm car-parrinello simulations of catalytic and enzymatic reactions. *CHIMIA International Journal for Chemistry*, 56(1-2):13–19, 2002.
- [10] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.
- [11] C David Sherrill. An introduction to hartree-fock molecular orbital theory. *Georgia inst. of technology*, 2000.
- [12] Thomas R Cundari. Computational studies of transition metal- main group multiple bonding. *Chemical reviews*, 100(2):807–818, 2000.
- [13] Clark R Landis, Daniel M Root, and Thomas Cleveland. Molecular mechanics force fields for modeling inorganic and organometallic compounds. *Reviews in Computational Chemistry, Volume 6*, pages 73–148, 1995.
- [14] Martin EM Noble, Jane A Endicott, and Louise N Johnson. Protein kinase inhibitors: insights into drug design from structure. *Science*, 303(5665):1800–1805, 2004.
- [15] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [16] Lucia Banci. Molecular dynamics simulations of metalloproteins. *Current opinion in chemical biology*, 7(1):143–149, 2003.
- [17] Johan Aqvist and Arieh Warshel. Free energy relationships in metalloenzyme-catalyzed reactions. calculations of the effects of metal ion substitutions in staphylococcal nuclease. *Journal of the American Chemical Society*, 112(8):2860–2868, 1990.

- [18] Yuan-Ping Pang, KUN Xu, Jamal El Yazal, and Franklyn G Prendergast. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Science*, 9(10):1857–1865, 2000.
- [19] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.
- [20] David A Pearlman, David A Case, James W Caldwell, Wilson S Ross, Thomas E Cheatham, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.
- [21] Peter Eastman and Vijay Pande. Openmm: a hardware-independent framework for molecular simulations. *Computing in science & engineering*, 12(4):34–39, 2010.
- [22] Frances C Bernstein, Thomas F Koetzle, Graheme JB Williams, Edgar F Meyer, Michael D Brice, John R Rodgers, Olga Kennard, Takehiko Shimanouchi, and Mit-suo Tasumi. The protein data bank. *The FEBS Journal*, 80(2):319–324, 1977.
- [23] Anders Nimmermark, Lars Öhrström, and Jan Reedijk. Metal-ligand bond lengths and strengths: are they correlated? a detailed csd analysis. *Zeitschrift für Kristallographie-Crystalline Materials*, 228(7):311–317, 2013.
- [24] Markus A Lill and Matthew L Danielson. Computer-aided drug design platform using pymol. *Journal of computer-aided molecular design*, 25(1):13–19, 2011.
- [25] W Richard Chegwidden and Nicholas D Carter. Introduction to the carbonic anhydrases. In *The Carbonic Anhydrases*, pages 13–28. Springer, 2000.
- [26] Laura L Kiefer, Steven A Paterno, and Carol A Fierke. Hydrogen bond network in the metal binding site of carbonic anhydrase enhances zinc affinity and catalytic efficiency. *Journal of the american chemical society*, 117(26):6831–6837, 1995.

- [27] Yuan-Ping Pang. Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. *Journal of Molecular Modeling*, 5(10):196–202, 1999.
- [28] Kjell Håkansson, Maurits Carlsson, L Anders Svensson, and Anders Liljas. Structure of native and apo carbonic anhydrase ii and structure of some of its anion-ligand complexes. *Journal of molecular biology*, 227(4):1192–1204, 1992.
- [29] Marcus Miethke and Mohamed A Marahiel. Siderophore-based iron acquisition and pathogen control. *Microbiology and molecular biology reviews*, 71(3):413–451, 2007.
- [30] JB Neilands. Siderophores: structure and function of microbial iron transport compounds. *Journal of Biological Chemistry*, 270(45):26723–26726, 1995.