Molecular Simulations of Metal-based Proteins: from Catalysis to Ion Transport

Thesis submitted for the degree of
Doctor Philosophiæ

Candidate:
Rolando Hong Enriquez

Supervisors:
Prof. Paolo Carloni
Dr. Alessandra Magistrato

November, 2009

Contents

Acknowledgements

It's late at night and I am about to submit my PhD thesis. This thesis owes very much to my supervisors, my colleagues, my friends, and my family. It's my great pleasure to acknowledge the people who have been contributing in one way or another to the accomplishment of my thesis.

To SISSA-related people,

First of all, I would like to thank my supervisor, Prof. Paolo Carloni. I am still wondering why he decided to give me the opportunity to start a new career in molecular simulations in spite of my complete lack of experience in this subject at the time. But today, I am really greatful for this. He has also been very hard on every mistake I have made. At the beginning, I used to get pretty angry about that; but later, I learned to be more careful and to question my own preconceived ideas. I thank Paolo for this.

I am greatful to my second supervisor, Dr. Alessandra Magistrato, for all the time she has spent following me during these 4 years at SISSA.

I would like to thank Marilisa Neri for our first helpful discussions about CG/MM method which is now used in this thesis.

To Vincenzo Carnevale and Attilio Vargiu, I am greatful for their support in the methodology of studying the polarization in the active site of proteases, which is also used in my thesis.

The valuable inputs for order parameter calculations in my thesis was the help from Giacomo Fiorin. I am greatful to him.

I want to express my gratitude to the former members of the SBP group who shared with me experiences and ideas: Sergio Pantano, Manuela Minozzi, Francesco Pontiggia, Fabio Simona, Andrea Miani, Cristina Fenolar, Claudio Anselmi, and Arturo Robertazzi.

I am indebted to Fernando Herrera for being an excellent friend and office-mate, and for guiding me through my very first days of making simulation tests.

I would love to show my gratitude also to Vanessa Leone, Fabrizio Marinelli, Nicola Soranzo, and Kamil Khafizov for their exceptional friendship and for telling me the truth, always, no matter what.

Additionally, I wish to thank all the rest of the people in the group for creating a very good working environment that favors learning from each other.

To non-SISSA people,

To Prof. Jose Luis Hernandez Caceres (Biophysics) and Prof. Julio Alvarez Gonzalez (Cardiac electrophisiology), I am inspired by their dedication to science even under very harsh conditions. I am also thankful for the privilege of their friendship.

The gratitude is also to my friends from college for all of the passionate discussions about biochemistry and science we had.

In the very tough years of college, I am indebted to have Tania, Hilda, Merceditas, and Ferrer as my second family.

I wish to thank Dulcima, Ricabal, and Julio, very dear friends and colleagues from the time I worked at Havana Medical School.

To the members of my last research group in Cuba: Luis Garcia, Carlos Martinez, Miguel Sautie, Ademar Gonzalez, Kiria Valdez, and Eduardo Tejera, I am greatful for their enthusiasm, creativity, and for sharing with me ideas, dreams, and valuable scientific computing skills during my master studies.

My thank is also to Ramon Guevara for his friendship and the invaluable help he gave me when I was learning to be an immigrant in Italy.

I owe my deepest gratitude to my mother, one of the most incredible woman I have ever known, for her unconditional love and for turning into gold everything she touches.

This thesis is dedicated to my father. With his strong character, he inculcated me to always do the right thing in spite of the problems that often comes with this choice.

I am very happy and proud to have Jaque, Rafa, Luisi, Fernand, Ling, Lazaro, Chago, Fernando, Xiomara, ... as my beloved family. I wish they were all here...

The last thank is to Trang, she has come from very far, shinning and bringing a new light to my life.

Chapter 1

INTRODUCTION

About 40% of all known proteins contain metal ions as a part of their structures in a stable or transitory way.[1] Na^+, K^+, Mg^{2+}, Ca^{2+}, Zn^{2+}, $Cu^{1+/2+}$, and $Fe^{2+/3+}$ are the metal ions most frequently found in life.[2] They have been selected over a long evolutionary time (3×10^9 years) on the basis of their availability and physicochemical properties.[3] Elucidating the molecular basis of metal handling and use in proteins helps understand an enormous variety of their functions, *e.g.,* protein folding, stabilization of enzyme catalysis, signal transduction, photosynthesis, and respiration.

Broadly speaking, metal ions may be steadily bound to proteins to perform structural or enzymatic functions,[4] or may bind rather transiently, being transferred from one protein to another.[5] It is now clear that the chemical environment around metal ions differs dramatically in the two cases. For instance, in the case of Zn^{2+} and $Cu^{+/2+}$ ions, most of the structurally characterized proteins with steadily bounded metal ions have four His/Asp residues,[4,6] while in trafficking proteins they are bound to two/three Cys residues.[5]

Computational biology approaches, including bioinformatics and molecular simulations, can shed light on structural and functional aspects of these two rather different situations. In this thesis, we have applied a large number of such approaches to Zn- and Cu-based representative proteins, which have been well characterized experimentally and are relevant from the perspective of molecular medicine. These are (i) a Zn-based enzyme, secreted by *B. anthracis*, in which a Zn^{2+} ion is steadily bound to the active site and performs an enzymatic reaction and (ii) the Atx1-Ccc2 copper transport system, a protein system that is used by Nature to deliver copper ions to extracellular copper transport proteins and copper-enzymes.

1.1 A Zn based protein: Anthrax Lethal Factor

The well known[7–11] lethal toxicity of the bacterium *B. anthracis* is caused by the release of three proteins: the **P**rotective **A**ntigen (**PA**), the **E**dema **F**actor (**EF**), and the **L**ethal **F**actor (**LF**). These proteins are involved in a complex cellular pathway (Fig. 1.1).

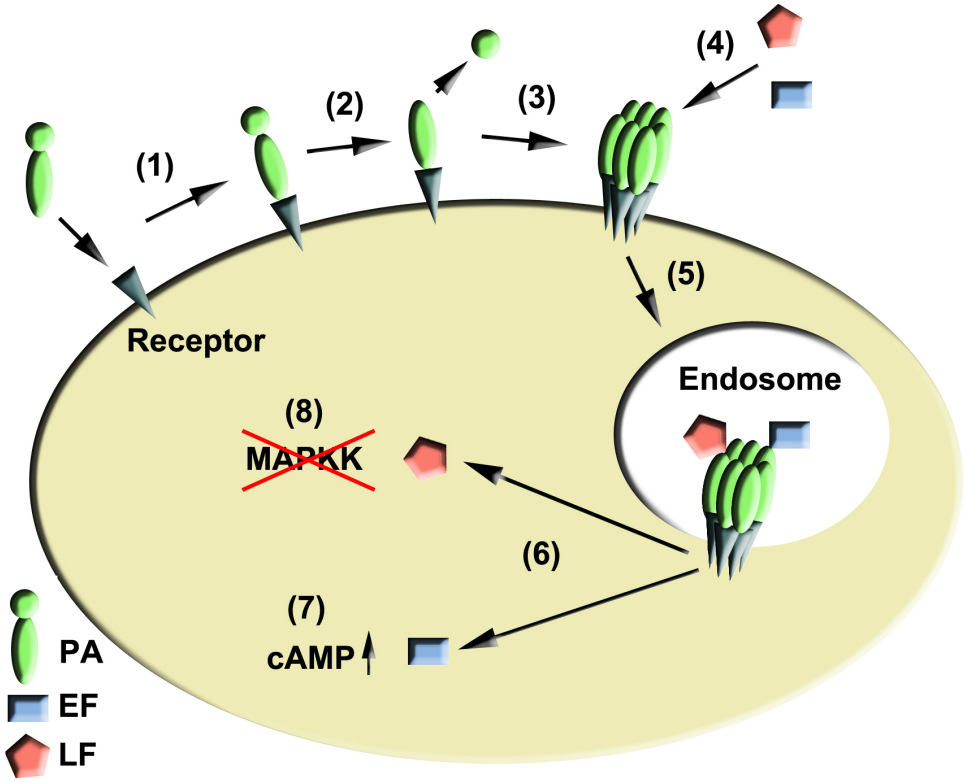


Figure 1.1: Scheme of the cellular mechanism of action responsible for the well known cytotoxicity of *B. anthracis* secreted proteins: **(1)** The Protective Antigen (PA) binds to the von Willebrand factor A domain of a membrane receptor.[12] **(2)** A furin-type enzyme cleaves PA,[13] releasing a 20 kDa fragment of PA and leaving a 63 kDa fragment bonded to the receptor. **(3)** The cleaved 63 kDa PAs self-associate and form a ring shape heptamer.[14] **(4)** Up to three molecules of Edema Factor (EF) or Lethal Factor (LF) bind with high affinity to the heptamer.[15] **(5)** The complex of the heptamer and LF and/or EF is pulled into the cell by receptor mediated endocytosis.[16] **(6)** The low pH in the endosome provokes the translocation of EF/LF to the cytoplasm.[17] **(7)** EF is a Ca^{2+} and calmodulin dependent adenylate cyclase that greatly increases the levels of cAMP perturbing the homeostasis of the cell.[18] **(8)** LF is a Zn^{2+} dependent endoprotease that cleaves the N-terminal of the Mitogen-Activated Protein Kinase Kinase (MAPKK) altering signaling pathways and finally leading to apoptosis.[19]

One of the three proteins (LF) is a large metalloprotease (750 amino acids) formed by four domains.[20] In its catalytic domain, LF holds a Zn^{2+} ion coordinated by two histidines, a glutamate, and a water molecule. The latter is probably the nucleophilic agent for the enzymatic catalysis (Fig.1.2), namely the cleavage of a peptidic bond near the N-terminal segment of members of the MAPKK family of proteins. Prompted by the high relevance of this protein for *B. anthracis* cytotoxic activity, we have performed a detailed computational study based on the available X-ray structures of LF.[20]

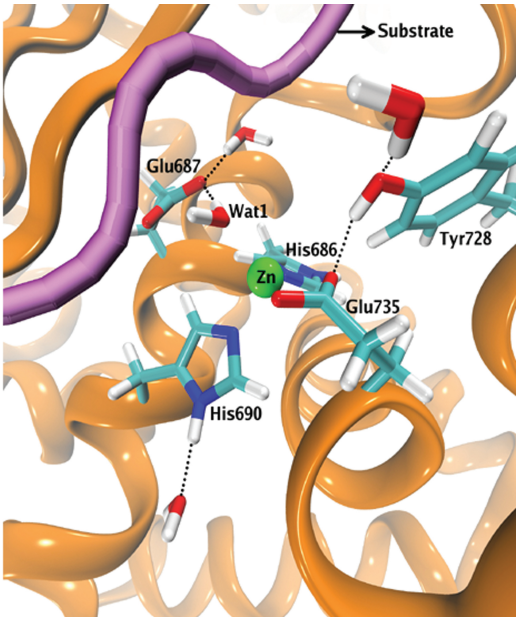


Figure 1.2: Anthrax LF active site. The residues important for the catalytic activity[20–22] are shown in licorice representation. The backbone of the substrate peptide (/VYPYPMEPT/)[21] is shown in purple.

Our calculations included Density Functional Theory (DFT), all-atom Molecular Dynamics (MD) simulations, calculations based on Coarse-Grained (CG) models and bioinformatics approaches. Based on these calculations, we predicted the H-bond pattern of the Zn active site of the protein in aqueous solution. This is absolutely crucial to model the enzymatic reaction, which in turn might help develop novel peptidomimetic inhibitors with therapeutical properties. In addition, we predicted that some parts of LF's structure may adopt a different conformation in solution from that reported in X-ray crystallography[20](*i.e.,* in the solid state). Finally, our calculations suggest that structure and energetics of the metal binding site may be tuned by specific interaction between the metal ligands and the protein frame. This feature was already found by our group several years ago for other Zn-based enzymes and could be a rather common feature of Zn-based enzymes.[23]

1.2 Cu: Atx1-Ccc2 transport system

The maintenance of adequate copper ion levels and the delivery of these ions to their target proteins are crucial aspects of cell functioning. Fig. 1.3 shows a schematic representation of these mechanisms in *S. cerevisiae*; here, these mechanisms have been well characterized experimentally.[24]. Basically identical pathways exist for humans.

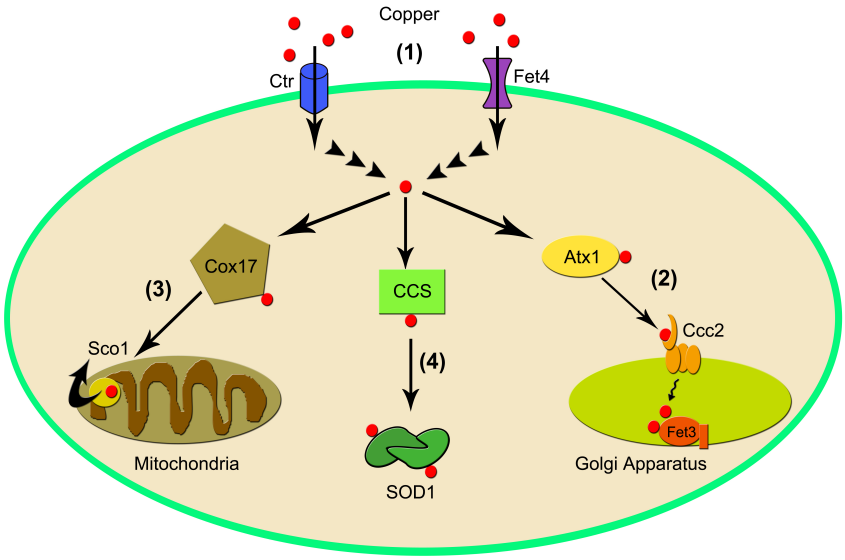


Figure 1.3: Main routes of copper metabolism in *S. cerevisiae*: (1) The Cu^+ ions enter the cell using either low affinity (Fet4)[25] or high affinity (Ctr)[26–28] copper transporters. Once inside the cell, the ions can follow three main routes. In (2), the chaperone Atx1 transports the Cu^+ ions to the membrane associated Ccc2,[29] that pumps the metal to the lumen of the Golgi apparatus; here Cu^+ is ultimately associated with Fet3, an iron uptake protein;[30,31] this route is called the copper secretory pathway, it has been studied in this thesis (Chapter 4). In (3), the chaperone Cox17 delivers the Cu^+ ions to Sco1 (or its homologue Sco2) in the internal membrane of the mitochondria; in this route, Cu^+ is finally incorporated to the respiratory enzyme cytochrome c oxidase. In (4), the chaperone CCS inserts the Cu^+ ions into the superoxide dismutase SOD1 that detoxifies superoxide radicals.[32] The correspondent human routes are rather similar.

In chapter 4, we study one of the most characterized events in the cascade, the copper secretory pathway (Pathway (2) in Fig.1.3). In this pathway, Cu^+ ions are transported by the chaperone protein Atx1 to a *trans*-Golgi associated Ccc2; then copper is inserted into copper enzymes (*e.g.,* Fet3, ceruloplasmin).[30,31,33] *S. cerevisiae* Atx1-Ccc2 copper transport system might serve as a model to understand the basic characteristics of Cu^+ transport in this kind of pathway in humans, which involve, instead of Ccc2, the multi domain proteins ATP7A and ATP7B.[34] Several mutations of the latter are the causes of Menkes and Wilson diseases.
In reality, Ccc2 is a large intra-membrane protein that consist of a Cu-pump and a soluble

domain. Atx1 interacts with Ccc2 soluble domain.[1] Ccc2 and Atx1 have the same fold that is called **H**eavy **M**etal **A**ssociated (**HMA**) domain. HMA is relatively small($\sim$ 70 amino acids) consisting of a $\beta\alpha\beta\beta\alpha\beta$ ferrodoxin-like fold[35] (Fig. 1.4). The copper binding site features the sequence MTCXXC with two conserved cysteines[36] involved in transient metal binding (Fig.1.4).

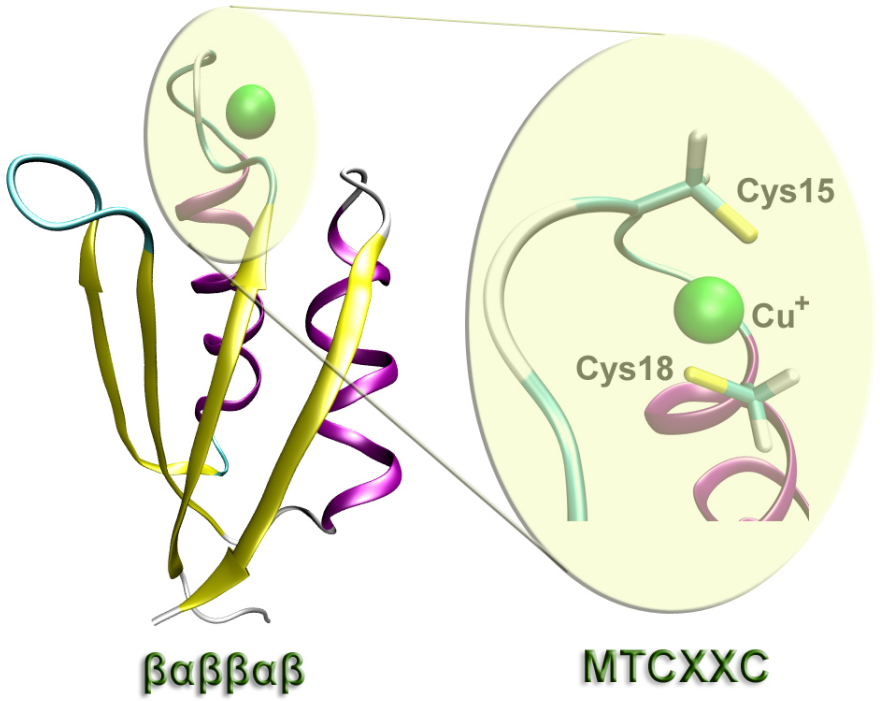


Figure 1.4: Heavy Metal Associated (HMA) domain.[37] It features the $\beta\alpha\beta\beta\alpha\beta$ ferrodoxin-like fold.[35] Two cysteines, conserved across the family,[36] bind to the copper ion in the metal binding site. The figure illustrates the case of Atx1 from *S. cerevisiae*, in which the two cysteine residues are Cys15 and Cys18.

Here we have provided qualitative insights on Cu^+ transport between Atx1 and Ccc2 (Fig.1.5). The direction of this reaction is only slightly favorable to the formation of Ccc2-Cu ($\Delta G = -0.2$ kcal/mol).[38]
Combined alanine mutation and NMR experiments have shown that the transfer occurs through several intermediates.[39] Based on a large number of biochemical,[40–42] structural,[38,43–45] and thermodynamical data,[38,46–50] here we have used Standard MD, Brownian Dynamics, Random Expulsion MD, and electrostatic modeling to provide some insights into the copper transfer process.

[1]For simplicity, in this thesis we use the term Ccc2 as an equivalent of the Ccc2 soluble domain.

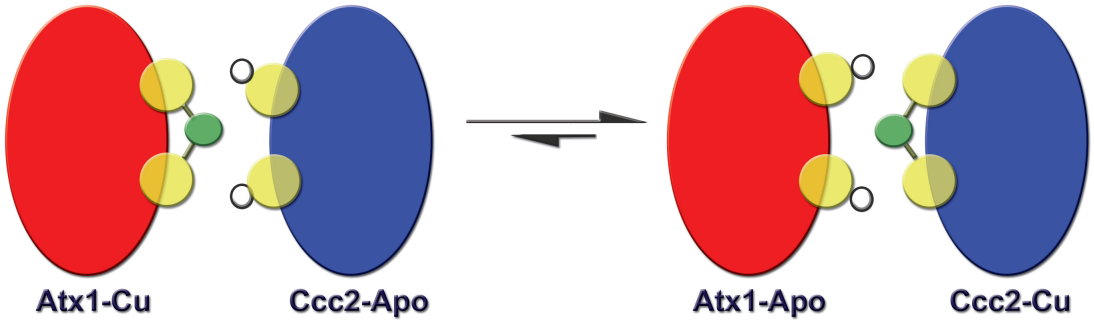


Figure 1.5: Scheme of the copper transfer reactions between Atx1 (red) and Ccc2 (blue). In this figure, are also represented the conserved cysteines directly involved in Cu^+ binding (yellow), Cu^+ (green) and hydrogens on the cysteines (white).

We focus mostly on the non-covalent association or dissociation between Atx1 and Ccc2. This is the formation of the complex *before* formation or breaking of covalent bonds (see Fig.1.5). We find that the association is faster when Cu^+ is bonded to Atx1, whilst the dissociation is faster when Cu^+ is bonded to Ccc2. In addition, we identify the residues playing a key role for protein-protein association during copper transfer from Atx1 to Ccc2 in Fig.1.5.

Chapter 2

METHODS

We describe here the basic methodological concepts of the computational techniques used in this thesis: all-atom Molecular Dynamics simulations, multiscale methods, Poisson-Boltzmann electrostatic calculations, Brownian dynamics, accelerated dynamics, standard bioinformatics algorithms and techniques for the validation and analysis of Molecular Dynamics simulations.

2.1 All-atom Molecular Dynamics

All-atom **M**olecular **D**ynamics (**MD**) simulations are the result of the application of modern computer technology to the Newton's laws of motion.
In MD simulations, abstract models of atoms and molecules are allowed to interact for a period of time following a predefined set of approximations of physics laws that are iterated numerically. With MD simulations, it is possible to study properties of statistical ensembles of atoms neglecting their internal structures and treating them at a classical (Newtonian) level. The theoretical support for such simplification comes from the Born-Oppenheimer approximation.

2.1.1 Born-Oppenheimer approximation

To describe the electronic details of a molecular system, quantum mechanics is unavoidable. For a system of N nuclei and M electrons, this would involve solving the Schrödinger equation for $3(N+M)$ variables. The Born-Oppenheimer approximation[51] allows us to reduce this mathematical problem to $3M$ variables with $3N$ parameters. This is achieved through the separation of the electronic and nuclear components of the Schrödinger equation:

$$\Psi_{\text{total}} = \Psi_{\text{electronic}} \times \Psi_{\text{nuclear}}. \tag{2.1}$$

The justification for this approach lies in the large difference between the masses of nuclei and electrons; the motions of the heavier nuclei are neglected and the Schrödinger equation is solved for the electrons using a fixed nuclear configuration. In any other method beyond the Born-Oppenheimer approximation, the concept of nuclear positions and molecular geometry becomes blurred. In the Born-Oppenheimer approximation, the nuclei move on a potential energy surface (PES); the gradient field of a PES ultimately justifies the creation of a force field based on classical

mechanics rules where the total force acting on the atoms is a function of the configuration of the atoms in the system.

2.1.2 Force Field. General concepts

A force field is a parametrization of system's electronic energy as a function of the nuclear coordinates. The use of single coordinates to represent atoms is justified by the Born-Oppenheimer approximation.[51] The parameters of a force field are normally fitted to experimental data or to high level (*e.g.*, quantum) computational data. The main idea underlying a force field is that any molecule is composed of roughly invariant "units" (*e.g.*, a C-H bond has approximately the same length and strength in any molecule). In the simpler cases, that will be referred here as "classical" force fields; the functional form is purely additive, quantifying separately the contributions for bonds, angles, dihedrals, improper angles, and non-bonded van der Waals and electrostatic interactions:

$$\begin{aligned} E_{FF} = &\sum_{\text{bonds}} K_r(r-r_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1+\cos(n\chi-\delta)) \\ &+ \sum_{\text{impropers}} K_\phi(\phi-\phi_0)^2 + \sum_{\text{non-bonded}} \left\{\epsilon_{ij}\left[\left(\frac{R_0^{ij}}{r_{ij}}\right)^{12} - \left(\frac{R_0^{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon' r_{ij}}\right\}, \end{aligned} \tag{2.2}$$

where K_r, K_θ, K_χ, and K_ϕ are force constants; r, θ, χ, ϕ, and R are the values for the current configuration; and the zero subscripts represent reference or equilibrium values. In spite of their inaccuracy,[52] the Lennard-Jones parameters for each pair of atoms are still obtained by the simple Lorentz-Berthelodt combination rules; here, for atoms i and j :

$$R_0^{ij} = (R_0^{ii} + R_0^{jj})/2, \tag{2.3}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}. \tag{2.4}$$

Other schemes for the estimation of Lennard-Jones parameters[53,54] provide better estimates, in the case of Waldman and Hagler[54] we have:

$$R_0^{ij} = \sqrt[6]{\left(\frac{(R_0^{ii})^6 + (R_0^{jj})^6}{2}\right)}, \tag{2.5}$$

$$\epsilon_{ij} = \left(\frac{(R_0^{ii})^3 (R_0^{jj})^3}{(R_0^{ii})^6 + (R_0^{jj})^6}\right) \sqrt{2(\epsilon_{ii}\epsilon_{jj})^2}. \tag{2.6}$$

As for the treatment of the electrostatic interactions in classical force fields, an usual practice is to assign fixed partial charges for the atoms during the complete MD simulation run. Partial charges can be assigned using a variety of methods;[55,56] some of these methods are based on empirical rules (*e.g.*, assigning of Gasteiger charges[57]) but a more diffuse practice is to fit the electrostatic potential calculated by electronic structure methods. Beyond this basic scheme, the rest of the available force fields use a few more characteristics with varied degrees of sophistication. The resulting force fields are sometimes called generically "second generation" (or class II)

force fields, although no strict classification exists in this respect. Among the additional features we find the use of special potential functions (*e.g.*, Morse potential for bonds[58]), explicit treatment of hydrogen bonds,[59] cross terms, and higher order terms[60,61] implemented to capture the vibrational modes of the molecules. After calculating the forces according to the described force field, the equations of motion for the particles are integrated according to Newton's laws using more or less sophisticated algorithms (see section 2.1.6).

2.1.3 Force Field. Parametrization with a metal center in the protein

Force field parametrization for metal centers in proteins faces additional problems because the metals are difficult to model in a classical force field. Particularly: (i) the bonding around metals is more varied than for organic molecules, (ii) in general, for an given number of ligands several geometrical arrangements are possible, (iii) the coordination bonds are normally ill-defined and the energy cost of a geometrical distortion is usually smaller and more variable than for a covalently bonded structure. These features give rise to an energy landscape with multiple minima and low energy barrier for interconversion between configurations. From a practical point of view, this makes unreliable the assigning of force constants in the force field; alternative non-bonded parametrizations for metal interactions are possible, but often are unstable and difficult to validate. Using more flexible functional forms (*e.g.*, Morse potentials for bonds,[58] Fourier expansion of the energy term for angle bending[62]) is an alternative to deal with these problems.

2.1.4 Periodic Boundary Conditions (PBC)

Ideally, we would like to simulate molecular systems with a number of particles in the order of magnitude of the Avogadro number ($\sim 10^{23}$ atoms). In this way, a small fraction of the atoms would be close to the boundary and we could be sure that the measured properties correspond to the interior (*i.e.,* "bulk") atoms of our simulation cell. The largest MD systems that we can simulate today ($\sim 10^6$ atoms) are still far from this objective, but if the system does not have important electrostatic interactions (*e.g.*, a Lennard-Jones system), good results can be obtained even using a small simulation cell. When electric charges are involved, it would be suitable to have a simulation cell at least as big as the Bjerrum length, λ_B (*i.e.*, the distance at which the electrostatic interactions between two elementary charges is comparable to the thermal energy); for water at 298K, $\lambda_B \simeq 0.7\ nm$. However, even if this condition is fulfilled, we cannot always discard the influence of highly charged groups at greater distances. Additionally, at large distances, direct electrostatic calculations of every single pair of interaction become unpractical; the implementation of **P**eriodic **B**oundary **C**onditions (**PBC**) alleviates this difficulty.
The introduction of PBC is equivalent to consider an infinite array of identical copies of the simulation box. However, the use of this technique comes with a price: (i) fluctuations in the system with a wavelength greater than the length of the periodic box (L) will not be properly estimated and (ii) errors in the anisotropy of the radial distribution function $g(r)$ are possible.[63]

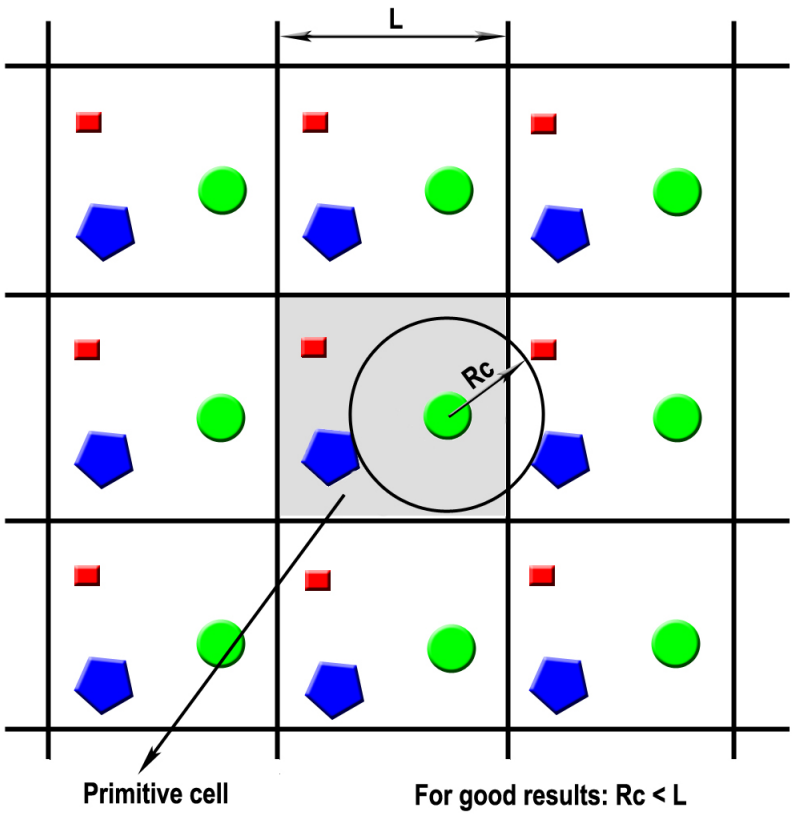


Figure 2.1: Schematic representation of the Periodic Boundary Conditions (PBC)

The error of calculating short range non-bonded interactions for a particle within a cutoff R_C can be made infinitely small by increasing this cutoff to infinity. However, for practical reasons, this is never done, a normal assumption for a good cutoff is that for any r such as $r > R_C$, the intermolecular interaction should be zero. If this condition is not fulfilled, then a systematic error is introduced when calculating the potential energy. A variety of shifting or switching functions that modify the Coulomb potential by smoothly truncation have been proposed; however, even when some of these approaches lead to stable dynamics,[64] the relationship of these modified potentials to basic electrostatics is not clear. A better treatment of long range electrostatics taking into account PBC is therefore needed.

2.1.5 Long range interactions. Particle Mesh Ewald (PME)

In systems with long range interactions (*e.g.,* Coulomb and dipolar potentials) with PBC, the introduction of a truncation technique schemes is computationally less expensive, but serious artifacts can be introduced.[64] Therefore, it is desirable to improve MD simulations schemes with long range algorithms up to the performance levels of short range cutoff schemes. A popular algorithm to do this is the **P**article **M**esh **E**wald (**PME**), based on the Ewald summation method.[65] A brief and necessarily incomplete description of the Ewald method is given below.

The total Coulomb energy of a system of N particles in a cubic box of size L and their infinite replicas introduced by PBC, can be calculated according to:

$$U = \frac{1}{2}\sum_{\mathbf{n}}^{1}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{ij,\mathbf{n}}}, \tag{2.7}$$

where $\mathbf{n}$ is the cell coordinate vector, $\mathbf{n} = (n_1, n_2, n_3) = n_1 Lx + n_2 Ly + n_3 Lz$, with x, y, z being the Cartesian coordinate unit vectors. The sum in equation (2.7) is conditionally convergent (*i.e.*, the results depend on the order of summation) and in any case the convergence is slow. The Ewald summation method substitutes the poorly convergent sum in equation (2.7) by a sum of two rapidly convergent series plus a constant-correcting term:

$$U_{\text{Ewald}} = U_r + U_i + U_0, \tag{2.8}$$

where:

$$U_r = \frac{1}{2}\sum_{i,j}^{N'}\sum_{\mathbf{n}} q_i q_j \frac{\text{erfc}(\alpha r_{ij,\mathbf{n}})}{r_{ij,\mathbf{n}}}, \tag{2.9}$$

$$U_{im} = \frac{1}{2\pi V}\sum_{i,j}^{N} q_i q_j \sum_{\mathbf{m}=0}\frac{\exp[-(\pi\mathbf{m}/\alpha)^2 + 2\pi i\mathbf{m}(r_i - r_j)]}{m^2}, \tag{2.10}$$

$$U_0 = \frac{-\alpha}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2. \tag{2.11}$$

Therefore, the Ewald sum is composed of: (i) a sum of energy terms in the real (direct) space U_r, (ii) a sum in reciprocal Fourier space U_{im} and, (iii) a constant self term, U_0. In the equations, V is the volume of the simulation box, $\mathbf{m} = (l, j, k)$ is a reciprocal space vector and $\text{erfc}(x) = 1 - \text{erf}(x) = 1 - (2/\sqrt{\pi})\int_0^x e^{-u^2} du$ is the complementary error function. The forces can be obtained by integration of equations (2.9) and (2.10); the self term is a constant and therefore does not have any contribution to the force during the simulation. Physically, the decomposition made in the Ewald method (Fig. 2.2) can be interpreted as the interactions resulting from the assignment to every point charge q_i, a Gaussian charge distribution of equal magnitude and opposite sign, ρ_i'. This charge distribution screens the interactions between charges effectively rendering these interactions short-ranged; the result is a fast convergence of the sum in real space. To balance the effect caused by the introduction of ρ_i', a second Gaussian charge distribution ρ_i'' with the same sign and magnitude of the original charge is added to each q_i. The sum of this second component is performed in reciprocal space; as an additional feature, to speed up the calculations in the reciprocal space, the **F**ast **F**ourier **T**ransform (**FFT**) is used following a general procedure: (i) using ρ_i'' and an assignment function, charge is assign to a 3D grid (mesh) extended over the simulation box; (ii) using FFT, the potentials and forces are calculated at the grid points; (iii) forces are interpolated back to the particles positions and the coordinates of the particles can then be updated in the next simulation step.

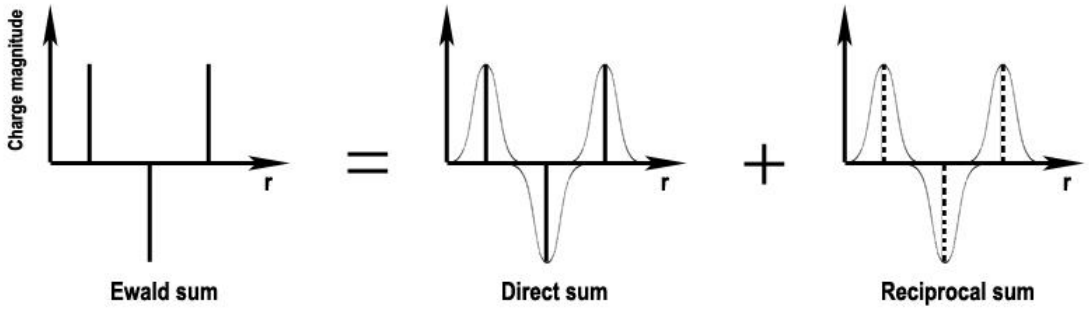


Figure 2.2: Ewald sum components for a one-dimensional point-charge system. Charges and Gaussian functions are normalized.

2.1.6 Integration methods

Newton's equations of motions for a system of N particles can only be solved numerically; therefore, in MD simulations, efficient numerical integrations methods are required. Several numerical methods to solve sets of differential equations can be found on general textbooks;[66] however, most of these methods are unpractical for MD simulations for several reasons: (i) some methods (*e.g.,* Runge-Kupta) involve more than one computationally demanding force evaluation on each MD step; (ii) other methods (*e.g.,* Euler's) are simple, fast, but inaccurate. Two types of integration methods have been successfully used for MD simulations: predictor-corrector methods the leapfrog methods; the latter are by far the most used ones; a brief description of them follows.

2.1.6.1 Leapfrog methods

Inspired by the first realistic MD simulation performed by Rahman[67] and using the same system of particles, Verlet[68] introduced a technique 10 times faster than that of Rahman. Actually, the method was already present in the Newton's Principia;[69] it was also known in astronomy as the Störmer method[70] and in the field of partial differential equations as the leapfrog method. The "Newton- Störmer-Verlet-leapfrog" scheme is simply:

$$x_{n+1} = 2x_n + x_{n-1}f(x^n), \tag{2.12}$$

where n is the current step on MD and h is the size of the time step; the current time is therefore $t = nh$. This algorithm contains no explicit velocities; if needed, velocities can be approximated by:

$$v_n = (x_{n+1} - x_{n-1})/2h. \tag{2.13}$$

Beeman[71] in an effort to combine the advantages of the method used by Rahman[67] and Verlet,[68] arrived to the following formulas:

$$x_{n+1} = x_n + hv_n + \frac{2}{3}h^2 f_n - \frac{1}{6}h^2 f_{n-1}, \tag{2.14}$$

$$v_{n+1} = v_n + \frac{1}{3}hf_{n+1} + \frac{5}{6}hf_n - \frac{1}{6}hf_{n-1}, \tag{2.15}$$

where $f_n = f(x_n)$.
Similarly to what we saw in the leap-frog algorithm, Verlet formulas can be obtained from the Beeman's using simple substitutions; therefore, the accuracy of the calculated trajectory for both methods is the same. Notably, Beeman's scheme is less sensible to noise because it estimates the velocity in a more precise way. However, another transformation of the Verlet (or leap-frog) algorithm can decrease the error in the estimation of velocity while keeping the simplicity of the formulations and occupying less memory in the computer. For these reasons, the Verlet scheme is generally preferred for MD simulations.

2.2 Thermodynamical ensembles

An isolated system is characterized by a time independent Hamiltonian. The integration of the classical equations of motion for such a system leads, in the limit of infinite sampling, to a trajectory mapping a microcanonical ensemble (NVE) of microstates. Assuming an infinite numerical precision, the result of a MD simulation should provide us with this theoretical result. However, numerical algorithms in MD simulation often create an energy drift (total energy is not kept constant) caused mainly by artifacts in: (i) the numerical integration schemes, with problems arising from the use of an unphysical finite time step Δt and (ii) the evaluation of the energy function, where problems arise from the use of simulation parameters that sacrifice accuracy for computational speed (*e.g.,* cutoff schemes for electrostatics calculations). The use of Particle Mesh Ewald method (PME, section 2.1.5) partially solves this deficiency, but also introduces other types of artifacts. It's also unfortunate that the microcanonical ensemble coming out from a standard MD simulation does not correspond to the conditions under which most experiments are carried out. For comparison with experiments, the following ensembles are more useful:

Canonical ensemble (NVT): here the temperature has an average (macroscopic) value, while the total energy of the system (*i.e.,* the hamiltonian) is allowed to fluctuate.

Isothermal-isobaric (Gibbs) ensemble (NPT): here the pressure has an average (macroscopic) value while the instantaneous volume of the system is allowed to fluctuate.

Grand-canonical ensemble (μVT): which has a constant volume and temperature (like the canonical ensemble) but is open to exchange particles with a surrounding bath. In this case the chemical potential of the different species has a specific average, but the instantaneous value (N) of the number of particles is allowed to fluctuate.

To perform MD simulations in any of the above ensembles we need to kept constant at least one intensive quantity (*e.g.,* temperature) during the simulation. This is where algorithms emulating thermostats and barostats come into play.

2.2.1 Thermostat algorithms

An algorithm that modifies the Newtonian MD scheme to generate a thermodynamical ensemble at constant temperature is called thermostat algorithm. A similar definition can be given for a barostat algorithm. Although several valid motivations can be found to implement a thermostat algorithm, an additional advantage is the possibility to fix the problem of the energy drift caused by the numerical inaccuracies of the integration and force evaluation algorithms.
In principle, the temperature of an MD system can be maintained constant simply by rescaling, *i.e.,* multiplying the velocities of the particles every time step by $\lambda = \sqrt{\frac{T_0}{T(t)}}$ where T_0 is the target temperature and $T(t)$ is the temperature at time t. However, this crude approximation can inhibit proper equilibration of the system. More sophisticated strategies involve the coupling of the system to a heat bath (*e.g.,* section 2.2.4), changing some or all velocities of the particles according to a Maxwell-Boltzmann distribution of T_0 (*e.g.,* section 2.2.2), or redefining the equations of motion so that $T(t)$ does not change (*e.g.,* section 2.2.5).

2.2.2 Stochastic thermostats: (i) Langevin thermostat; (ii) Andersen thermostat

Temperature relaxation in MD simulations can be achieved by using **S**tochastic **D**ynamics (**SD**) which relies on the integration of the Langevin equation of motion; for particle i this can be written as:

$$\mathbf{a}_i(t) = m_i^{-1}\mathbf{F}_i(t) - \gamma_i(t)\mathbf{v}_i(t) + m_i^{-1}\mathbf{R}_i(t), \tag{2.16}$$

where $\mathbf{R}_i$ is a stochastic force and γ_i is a friction coefficient.
When choosing friction coefficients for SD used as thermostat in implicit solvent simulations, one has to consider that: (i) Small values of γ_i lead to poor temperature control and recovering of the canonical ensemble at a very long simulation time with the consequent accumulation of numerical errors and energy drift. In the limit, when $\gamma_i = 0\ \forall i$, we recover MD in the microcanonical ensemble (NVE) loosing the control of the temperature. (ii) Large values of γ_i lead to a perturbation of the dynamics due to large stochastic and frictional forces. In the limit of large γ_i, when $\mathbf{a}_i(t) \approx 0$, we recover **B**rownian **D**ynamics (**BD**) with equation of motion:

$$\mathbf{v}_i(t) = \gamma_i^{-1} m_i^{-1}[\mathbf{F}_i(t) + \mathbf{R}_i(t)]. \tag{2.17}$$

The Andersen thermostat (also known as stochastic coupling scheme) relies on the integration of the Newton equations of motion; however, in each integration time step a group of atoms is selected with probability $p(\tau)\Delta t = \alpha e^{-\alpha\tau}$ where α is the reassignment frequency and τ is the time interval between two consecutive reassignments. The velocities of the selected atoms are chosen according to a Maxwell-Boltzmann probability distribution. The equation of motion of the Andersen thermostat can be written as:

$$\mathbf{a}_i(t) = m_i^{-1}\mathbf{F}_i(t) + \sum_{n=1}^{\infty} \delta(t - \sum_{m=1}^{n} \tau_{i,m})[\mathbf{v}^*_{i,n}(t) - \mathbf{v}_i(t)], \tag{2.18}$$

where $\tau_{i,n}$ with $n = 1, 2, ...$ is the series of intervals without reassignment for particle i and $\mathbf{v}^*_{i,n}$ is the randomly-reassigned velocity after the n^{th} interval.
Similarly to SD, care must be taken when choosing the collision frequency parameter α: (i) Small values of α lead to poor temperature control recovering of the canonical ensemble at a very long simulation time with the consequent accumulation of numerical errors and energy drift. In the limit, for $\alpha = 0$, we also recover MD in the microcanonical ensemble (NVE). (ii) Large values of α cause perturbation of the dynamics due to excessive collisions and velocity reassignments.

2.2.3 Temperature constraining (strong coupling)

Temperature constraining algorithms (or strong coupling methods) impose a modification of the Langevin dynamics equations by fixing the instantaneous temperature to a reference temperature without permitting further fluctuations. The reference temperature can be stated explicitly in the equations as in the Woodcock algorithm:[72]

$$\mathbf{a}_i(t) = m_i^{-1}\mathbf{F}_i(t) - (gk_BT_0)^{-1}\left[\sum_{i=1}^{N} \mathbf{v}_i(t) \cdot \mathbf{F}_i(t)\right]\mathbf{v}_i(t), \tag{2.19}$$

or imposing the condition $\frac{dT}{dt} = 0$ as in the equivalent formulation of Hoover and Evans:[73,74]

$$\mathbf{a}_i(t) = m_i^{-1}\mathbf{F}_i(t) - \left[N_{df}k_BT(t)\right]^{-1}\left[\sum_{i=1}^{n}\mathbf{v}_i(t)\cdot\mathbf{F}_i(t)\right]\mathbf{v}_i(t). \tag{2.20}$$

2.2.4 Berendsen thermostat (weak coupling)

In the Berendsen method,[75] the Langevin equations of motion are modified to remove local temperature coupling through stochastic collisions (random noise) and the temperature of the system is corrected by taking into account the way in which the deviation from the target temperature T_0 decays exponentially with some constant τ_B:

$$\frac{dT}{dt} = \tau_B^{-1}[T_0 - T(t)]. \tag{2.21}$$

The resulting equation of motion for the Berendsen thermostat is:

$$\mathbf{a}_i(t) = m_i^{-1}\mathbf{F}_i(t) - \frac{1}{2}\tau_B^{-1}\left[\frac{g}{N_{df}}\frac{T_9}{T(t)} - 1\right]\mathbf{v}_i(t). \tag{2.22}$$

In available MD algorithms, τ_B is tuned to modify the strength of the coupling between the system and the thermostat: (i) When $\tau_B \to 0$ (strong coupling), we recover the Woodcock/Hoover-Evans equations of motion. (ii) For large values of τ_B, we inactivate the thermostat and recover MD in microcanonical ensemble (NVE).

2.2.5 Nosé-Hoover thermostat (extended system)

The Nosé-Hoover thermostat is one of the most commonly used algorithms to perform isothermal MD. The method was initially introduced by Nosé [76] and subsequently reformulated into simpler equations by Hoover.[77] The idea behind the Nosé-Hoover thermostat is to redefine (extend) the equations of motion of the system for the purpose of including a new artificial dynamical variable s that acts as a time-scale parameter while keeping the temperature constant. We briefly describe bellow how this is done. First, let's consider the Newton equations for a system of N particles in the canonical form:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i}, \tag{2.23}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i = \frac{-\partial\Phi}{\partial r_i}. \tag{2.24}$$

We also know that the average kinetic energy is related to the temperature of the system:

$$\langle E_k\rangle = \left\langle\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}\right\rangle = \frac{3}{2}NkT. \tag{2.25}$$

The modified equations of motion in the Nosé-Hoover thermostat can be written as:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i}, \tag{2.26}$$

$$\frac{d\mathbf{p}_i}{dt} = F_i = \frac{-\partial\Phi}{\partial r_i} - \zeta\mathbf{p}_i, \tag{2.27}$$

where ζ is a new friction term related to the heat bath. The change in ζ is implemented as a feedback mechanism that keeps the kinetic energy at a constant value:

$$\frac{Q}{2}\frac{d\zeta}{dt} = \sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} - \frac{3}{2}NkT. \tag{2.28}$$

If the kinetic energy is larger than the average ($\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} > \frac{3}{2}NkT$), then $d\zeta > 0$; in the equations of motions this will have the effect of increasing frictional forces and consequently decreasing the velocities of the particles. The inverse analysis can be done for the case when the kinetic energy is smaller than the average. In this way, the kinetic energy fluctuates around an average value and the temperature of the simulation will be controlled by T.
Introducing a new derived variable s defined as:

$$\frac{d(\ln(s))}{dt} = \zeta, \tag{2.29}$$

we get:

$$\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + \Phi(r) + \frac{Q}{2}\zeta^2 + 3NkT\ln(s). \tag{2.30}$$

This formula clearly shows the parts of the extended system in which the two first terms represent the physical subsystem under consideration and the two last terms represent the external subsystem that mimics a heat bath. These two subsystems are coupled by the friction term ζ in the equations of motion of the thermostat (equation 2.28).

2.2.6 Barostat algorithms

A barostat algorithm modifies the Newtonian MD scheme to generate a thermodynamical ensemble at constant pressure. This is normally achieved by considering the volume of the simulation box as a variable during the MD simulation. In an homogeneous system, only the dimensions of the simulation box are varied while for inhomogeneous systems, it might be necessary to vary also the shape of the simulation box.[78,79]
Most of the available barostat algorithms have debts with the Anderson barostat[80] which is analogous to the extended system thermostats described in the previous section.
For comparison, we consider the Lagrangian (kinetic minus potential energy) of a purely Newtonian MD system of N particles:

$$\mathcal{L} = \sum_{i=1}^{N}\frac{1}{2}m_i v_i^2 - U(r_1, ..., r_N), \tag{2.31}$$

with the Lagrangian proposed by Andersen:[80]

$$\mathcal{L} = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 - U(r_1, ..., r_N) + \frac{1}{2} W \dot{V}^2(t) - P_{ext} V(t). \tag{2.32}$$

We notice two new terms: the first ($\frac{1}{2} W \dot{V}^2$) is related to the inertial properties of the system, W could be associated with the "mass" of the piston, determining the decay time of volume fluctuations, V is the volume of the simulation cell; the second term ($pV(t)$) is the "piston" represented as the familiar quantity describing "work" in the thermodynamical sense, P_{ext} is the pressure that we want to impose on the system.
From this Lagrangian, we can write the equations of motion of the particles:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3V} \frac{dV}{dt} \mathbf{r}_i, \tag{2.33}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \frac{1}{3V} \frac{dV}{dt} \mathbf{p}_i, \tag{2.34}$$

$$\frac{d^2V}{dt} = \frac{1}{W} [P(t) - P_{ext}], \tag{2.35}$$

where $P(t)$ is the instantaneous pressure. Similar equations can be written for the Berendsen barostat:[75]

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3V} \frac{dV}{dt} \mathbf{r}_i, \tag{2.36}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i, \tag{2.37}$$

$$\frac{dV}{dt} = \frac{\chi}{\tau_p} [P(t) - P_{ext}] V, \tag{2.38}$$

where χ is the isothermal compressibility and τ_p is the pressure coupling time, determining the strength of the coupling to the external pressure.
For the Langevin piston algorithm[81] the equations are:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3V} \frac{dV}{dt} \mathbf{r}_i, \tag{2.39}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \frac{1}{3V} \frac{dV}{dt} \mathbf{p}_i, \tag{2.40}$$

$$\frac{d^2V}{dt} = \frac{1}{W} [P(t) - P_{ext}] - \gamma \frac{dV}{dt} + R(t), \tag{2.41}$$

where γ is the collision frequency and $R(t)$ is a Gaussian random force with mean zero and variance $2\gamma k_B T \delta(t)/W$ in which k_B is the Boltzmann constant. Notice that when $\lambda = 0$, the Langevin piston reduces to the Andersen thermostat.

2.3 Multiscale simulations

In molecular dynamics simulations, multiscale methodologies are used to make a partition of a system in regions that can be considered with different levels of resolution or accuracy. In this sense, multiscale methods provide an advantage when we want to study phenomena occurring at large time scales, or otherwise we need to save computer resources by treating a small region of the system in a computational demanding level of calculation and a larger region of the system at a computationally simpler level. In this thesis, we used two multiscale methodologies: (i) Coarse-Grained/Molecular Mechanics (CG/MM) simulations and (ii) Quantum Mechanics/Molecular Mechanics (QM/MM) simulations. In the next sections we will briefly describe both approaches.

2.3.1 Hybrid Coarse-Grained/Molecular Mechanics (CG/MM) simulations

In the **C**oarse-**G**rained/**M**olecular **M**echanics (**CG/MM**) methodologies one part of the system is simulated in a full atomistic detail (the MM part, which is treated with typical atomistic force fields), while the rest of the system (the CG part) is simulated at a lower level of molecular detail. Some CG/MM methodologies introduce an interface region between the CG and MM parts bridging the large discontinuity between the full-atom and CG descriptions (Fig.2.3).

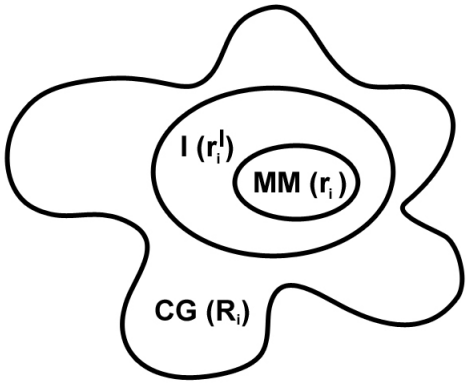


Figure 2.3: Schematic representation of the model for CG/MM simulations.

In this thesis, we used the CG/MM method developed by Neri *et al.,*[82] in which the CG part of a protein is represented by $C\alpha$ centroids instead of amino acids. The interface and MM part were treated with atomistic force fields (*e.g.,* AMBER parm98[83] and Gromos96 43a1[84]), while the CG region was treated with a Go simplified potential.[85] The total potential energy for the system can be represented by:

$$U = E_{MM} + E_{CG} + E_I + E_{MM/I} + E_{CG/I} + E_{SD}, \tag{2.42}$$

including the terms for the MM, CG, and I regions and also the cross-terms representing the interactions of the MM and CG regions with the interface; the final term represents the effect

of the solvent considered as the sum of stochastic and frictional forces proportional to the mass and velocities of the particles in the system. In the next equations, the coordinates of the atoms will be represented by r_i if the atom is in the MM region, R_i for a $C\alpha$ centroid in the CG region, and r_i^I for an atom in the interface region; the equation for the interactions in the MM region then reads:

$$E_{MM} = E_{bond} + E_{vdW} + \sum_{i>j} \frac{q_i q_j}{\varepsilon |\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.43}$$

The same functional form is used for E_{MM}, E_I, and $E_{MM/I}$ while for E_{CG} we have:

$$E_{CG} = \frac{1}{4} \sum_i K_b (|\mathbf{R}_i - \mathbf{R}_{i+1}|^2 - b_{ii+1}^2)^2 + \sum_{i>j} V_0 (1 - e^{-B_{ij}(|\mathbf{R}_i - \mathbf{R}_j| - b_{ij})})^2. \tag{2.44}$$

The first sum represents the bonded interactions between consecutive C_α centroids and the second sum represents the non-bonded Morse-like potential between C_α centroids. A similar expression was used for the interactions between atoms in the interface and the CG region:

$$E_{CG} = \frac{1}{4} \sum_i K_b (|\mathbf{R}_i - \mathbf{R}_{i+1}|^2 - b_{ii+1}^2)^2 + \frac{1}{2} \sum_{i \in [C\alpha, C\beta], j} V_0 (1 - e^{-B_{ij}(|\mathbf{r}_i^I - \mathbf{R}_j| - b_{ij})})^2. \tag{2.45}$$

The multiscale approach presented here is static in the sense that atoms belonging to a region do not change their character (CG or MM) during the simulation. Sorting a series of technical difficulties, Ensing *et al.*,[86] presented a time and space fully dynamical CG/MM methodology that includes switch functions acting on the interface region and modifying the character of the atoms. However, this method was not available at the time we were doing our simulations and so we used the static approach developed in our group[82] that has been recently proven to be useful to study a large membrane protease at the microsecond time scale.[87]

2.3.2 Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) simulations

Chemical reactions in enzymes, electron transfer in respiratory and photosynthetic pathways, and electronic excitation in rhodopsin-dependent visual systems are good biological examples of processes that occur in small spatial regions and can be properly described only through quantum mechanics calculations. However, these processes are embedded in larger macromolecular frameworks that can affect the purely quantum phenomena. As quantum methods can presently be performed on $\sim 10^2$ atoms and molecular mechanics methods to treat the large protein frameworks can be performed on $\sim 10^5$ atoms, the logical approach is to combine the two techniques: QM methods for the chemically interesting region and MM methods for the outer regions. The result of this combination has been called QM/MM methods and were first introduced by Warshel and Levitt in 1976.[88]

The partition of a system into QM and MM regions is not straightforward; a common strategy is to include hydrogen atoms (hydrogen capping) saturating the valences of the atoms that are at the interface of the partition, these hydrogen atoms will be considered for the QM calculations. In principle, a great number of probabilities are possible by combining QM and MM methods.

A requirement for a QM method is its possibility to perform self-consistent-field treatment in the presence of an external point charge field (the MM region); a popular choice is the use of Density Functional Theory (DFT) as implemented in the program CPMD[89] that provides an MM interface with the GROMOS force field.
A critical aspect in the implementation of QM/MM approaches is the treatment of the electrostatic coupling between the QM and the MM regions. Various approaches are possible depending on the extent of polarization implemented between the QM and MM regions. Table 2.1 resumes the main characteristics of different types of electrostatic couplings. In this thesis, we used the QM/MM algorithm as implemented in the Car-Parrinello molecular dynamics package (CPMD), which makes a extensive use of density functional theory.[89]

Electrostatic coupling	Computational resources	Fixed point charges in MM?	Fixed point charges in QM?	MM polarizes QM?	QM polarizes MM?
Mechanical embedding	★	✓	✓	-	-
Electrostatic embedding	★★	✓	-	✓	-
Polarized I	★★★	-	-	-	✓
Polarized II	★★★★	-	-	✓	✓

Table 2.1: Electrostatic coupling between QM and MM regions in QM/MM simulations.

2.3.3 Density Functional Theory (DFT)

Density **F**unctional **T**heory (**DFT**) is based on the Hohenberg-Khon theorem[90] which states that in the ground state, the total energy of a system can be described as a functional of the system electronic density $\rho(r)$; any other electronic density $\rho'(r)$ different from the true density will necessarily lead to a higher energy. Therefore, while in conventional *ab initio* calculations we need to work with $3N$-dimensional wavefunction ψ to solve the Schrödinger equation:

$$H\psi = E\psi, \tag{2.46}$$

in DFT we need only to work with a simple 3-dimensional electronic density function and minimize the energy functional, $E[\rho(r)]$. The advantage of DFT cannot be underestimated: while the complexity of the wavefunction increases exponentially with the number of electron, the electron density has the same number of variables, independently of the system size. In spite of the great simplification, this approach has one fundamental limitation: the exact nature of the energy functional is unknown.

2.3.4 DFT with the Kohn-Sham (KS) method

In 1965, Kohn and Sham offered a practical approach to perform DFT calculations.[91] In their approach, they made a partition of the unknown total energy functional $E[\rho(r)]$ into electrostatic energy $U[\rho(r)]$, kinetic energy $T[\rho(r)]$, and an also unknown "exchange" energy functional $E_{xc}[\rho(r)]$ for which, however, we can find fairly good approximations and is, in any case, a small fraction of the total energy functional. Resuming the **K**ohn-**S**ham (**KS**) partition can be written as:

$$E[\rho(\mathbf{r})] = U[\rho(\mathbf{r})] + T[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})]. \tag{2.47}$$

The electrostatic energy is separated in the sum of the electron-nucleus attractions and electron-electron repulsions:

$$U[\rho(\mathbf{r})] = \left(\sum_A \int \frac{-Z_A\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_A|}d\mathbf{r}\right) + \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'. \tag{2.48}$$

In KS approach, the electronic density $\rho(r)$ of an N-electron system (with N^α spin up and N^β spin down electrons) is expressed as the sum of the square moduli of singly occupied KS molecular orbitals:

$$\rho(\mathbf{r}) = \rho^\alpha(\mathbf{r}) + \rho^\beta(\mathbf{r}) = \sum_{i=1}^{N^\alpha}\left|\psi_i^\alpha(\mathbf{r})\right|^2 + \sum_{i=1}^{N^\beta}\left|\psi_i^\beta(\mathbf{r})\right|^2. \tag{2.49}$$

Then we define the kinetic energy $T[\rho(\mathbf{r})]$ as:

$$T[\rho(\mathbf{r})] = \sum_{\sigma=\alpha,\beta}\sum_i^{N^\sigma}\int \psi_i^\sigma(\mathbf{r})\frac{-\nabla^2}{2}\psi_i^\sigma(\mathbf{r})d\mathbf{r}. \tag{2.50}$$

The exchange energy functional $E_{xc}[\rho(\mathbf{r})]$ is built in a way to account for the corrections in the kinetic energy and non-classical part of the particle-particle interactions. Since the quality of DFT is limited by the quality of the approximation used to calculate $E_{xc}[\rho(\mathbf{r})]$, a great deal of research has been done in the development of sophisticated exchange energy functional forms. The parameters in the mathematical forms of exchange functional can be made to fulfill a group of properties[92] and then they are sometimes called non-empirical methods (*e.g.,* **L**ocal **D**ensity **A**pproximation (**LDA**),[91] **G**eneralized **G**radient **A**pproximations(**GGA**)[93]) or the parameters can be fitted to experimental data, and these methods are called empirical or semi-empirical although in practice a combination of these approaches is often used in the so called hybrid methods (*e.g.,* B3LYP).[94] There is no recognized "best" functional for all the systems, the "best choice" for a functional will depend on the system and properties that need to be simulated.
As we know that the total energy functional $E[\rho(\mathbf{r})]$ is minimized in the density of the ground state $\rho(\mathbf{r})$ then $E[\rho(\mathbf{r})]$ must be stationary with respect to any arbitrary variation in either of the spin densities:

$$\frac{\delta E[\rho(\mathbf{r})]}{\delta\rho^\alpha(\mathbf{r})} = \frac{\delta E[\rho(\mathbf{r})]}{\delta\rho^\beta(\mathbf{r})} = 0. \tag{2.51}$$

This condition leads to the one-electron KS equations:

$$\left(\frac{-\nabla^2}{2} - \left(\sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|}\right) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|d\mathbf{r}'} + \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta\rho^\sigma(\mathbf{R})}\right)\psi_i^\sigma(\mathbf{r}) = \epsilon_i\psi_i^\sigma(\mathbf{r}), \tag{2.52}$$

where $\sigma = \alpha, \beta$.

Once the approximated energy exchange functional has been selected, we still need to solve the KS equations; this can be done through an iterative process: we generate an initial guess for the density ($\rho^\sigma(\mathbf{r}), \sigma = \alpha, \beta$ in KS-equations) and use it to evaluate the KS-equations; the resulting set of KS-spin orbitals $\{\psi_i^\sigma(\mathbf{r})\}$, is used to generate new guesses of the density. The process is repeated until self-consistency is achieved.

2.3.5 Car-Parrinello Molecular Dynamics (CPMD)

Car-**P**arrinello **M**olecular **D**ynamics (**CPMD**)[89] has come to light as a method capable of performing MD simulations in which interatomic forces are calculated. The vast majority of CPMD applications have been within DFT, specifically plane wave DFT.[89] CPMD involves the simulation and instantaneous evolution of both the electronic and nuclear degrees of freedom by means of a fictitious Lagrangian:

$$\mathcal{L} = \underbrace{\sum_I \frac{1}{2}M_I\dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2}\mu_i\langle\dot{\psi}_i|\dot{\psi}_i\rangle}_{\text{Kinetic energy}} - \underbrace{\langle\Psi_0|H_0|\Psi_0\rangle}_{\text{Potential energy}} + \underbrace{\text{constraints}}_{\text{Orthonormality}}. \tag{2.53}$$

The associated Lagrange-Euler equations are:

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{R}}_I} = \frac{\partial\mathcal{L}}{\partial\mathbf{R}_I}, \tag{2.54}$$

$$\frac{d}{dt}\frac{\delta\mathcal{L}}{\delta\dot{\psi}_I^*} = \frac{\delta\mathcal{L}}{\delta\psi_I^*}, \tag{2.55}$$

from here the Car-Parrinello equations of motion are:

$$M_I\ddot{\mathbf{R}}_I(t) = -\frac{\partial}{\partial\mathbf{R}_I}\langle\Psi_0|H_e|\Psi_0\rangle + \frac{\partial}{\partial\mathbf{R}_I}\text{ (constraints)}, \tag{2.56}$$

$$M_i\ddot{\psi}_i(t) = -\frac{\partial}{\partial\psi_i^*}\langle\Psi_0|H_e|\Psi_0\rangle + \frac{\partial}{\partial\psi_i^*}\text{ (constraints)}. \tag{2.57}$$

Note that in these equations the two components of the system, formed by quantum (electrons) and classical (nuclei) are mapped onto a purely classical system of two components with separated energy scales. In the equations, μ_i are "fictitious masses" for electrons (or more correctly for the orbitals). In this scheme we have two temperatures, an instantaneous physical temperature for the nuclei (T_N) and a fictitious temperature for the electrons (T_e) in a way that:

$$T_N \propto \sum_I M_I\dot{\mathbf{R}}_i^2, \tag{2.58}$$

$$T_e \propto \sum_i \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle. \tag{2.59}$$

By "cooling" the electrons (minimizing the electronic subsystem to be close to the Born-Oppenheimer surface), we obtain a ground-state wave function for the initial configuration of nuclei that will remain close to the ground state during the simulation as long as the temperature is kept low enough. In practice, the nuclear and electronic motions have to be separated in a way that the fast electronic system remains cold during the simulation and is at the same time able to "follow" instantaneously the slower nuclear motions. This decoupling of the nuclear and electronic subsystems is possible if their motions do not have a significant overlap in the frequency domain so that the energy transferring from "hot nuclei" to "cold electrons" is negligible at relevant time-scales.

2.3.6 Wannier centers

Localized **M**olecular **O**rbitals (**LMO**), as readily deduce from the name, define spatially confined molecular orbitals, and therefore display in a clear way which atoms are bonded or have similar structural properties in different molecules; one approach to do this is the calculations of Wannier centers,[95] that can be estimated from a QM/MM calculation in CPMD.[89] With this information, the polarization of chemical bonds can be estimated by the use of the so called **B**ond **I**onicity (**BI**) indexes[96] BI_{AB} of a bond between two atoms A and B is defined as $BI_{AB} = \frac{d_A}{d_{AB}}$ where d_A is the distance between atom A and the Boys Orbitals along the AB bond, and d_{AB} is the length of the bond between A and B. A value of $BI = 0.5$ (the Boys orbilal is in the middle of the bond) indicates absence of polarization; while values close to 0 or 1 indicate polarization.

2.4 Extracting information from MD simulations

2.4.1 Validation of MD simulations

2.4.1.1 RMSF and Debye-Waller factors

Crystallographic Debye-Waller factors describe the fluctuations of atoms about their mean positions in a crystal structure. Since crystallographic studies are time-averaged, fluctuation amplitudes include terms due to time dependent thermal motions and terms due to deviations from atomic mean positions because of disorder in the crystal:

$$\langle u_i^2 \rangle_{\text{total}} = \frac{3B_i}{8\pi^2} = \langle u_i^2 \rangle_{\text{thermal}} + \langle u_i^2 \rangle_{\text{disorder}}, \tag{2.60}$$

where B_i is the crystallographic B-factor. The disorder term is independent of temperature, while the square of the thermal amplitude for harmonic vibrations varies linearly with the absolute temperature.

2.4.1.2 Validation against NMR data: Order parameters (S^2)

The order parameters (S^2) are a measure of the degree of spatial restriction of the motion and can be calculated from NMR experimental data[97–99] and from MD simulations.[100,101] Here we used two approaches to calculate the order parameters; for the rest of the thesis, these approaches will be called: (i) contact model and (ii) model free.

(i) Contact model
In this approach, we used an empirical formula that calculates the order parameter (S_i^2) of the N-H vector from residue i by considering the close contacts experienced by the H atom of this vector and the carbonyl oxygen of the preceding residue $i-1$ with the heavy atoms k:

$$S_i^2 = \tanh(0.8 \sum_k (e^{-\mathbf{r}^O_{i-1,k}}) + 0.8(e^{-\mathbf{r}^H_{i,k}})) + b, \tag{2.61}$$

Where:

(i) $\mathbf{r}^O_{i-1,k}$ is the distance between the carbonyl oxygen of residue $i-1$ to heavy atom k.

(ii) $\mathbf{r}^H_{i,k}$ is the distance between amide proton H (from vector N-H) and heavy atom k.

(iii) k is the heavy atoms that do not belong to residues i or $i-1$.

(iv) $b = 0.1$

All distances are expressed in angstrom (Å).
We used this formula to calculate order parameters from experimental X-ray or NMR structures and also from frames of MD trajectories.

(ii) Model free
A basic assumption of the "model free" formalism[97,98] is that the internal motions and the global tumbling of a molecule can be separated so that the correlation function can be factored as:

$$C(t) = C_0(t)C_I(t), \tag{2.62}$$

with the correlation function for global tumbling being equal to:

$$C_0(t) = e^{-t/t_c}. \tag{2.63}$$

where t_c is the correlation time for global tumbling. The internal correlation is approximately equal to:

$$C_I(t) \cong S^2 + (1 + S^2)e^{-t/t_e}. \tag{2.64}$$

$C_I(t)$ is normalized $C_I(0) = 1$ and decays with an effective correlation time t_e to a plateau value S^2.
Under the "model free" formalism, in a MD trajectory, global tumbling can be eliminated by fitting the protein to a reference initial structure and the resulting N-H correlation function extracted directly from the trajectory can be fitted to equation (2.64) to estimate the order parameters.

2.4.1.3 Validation against NMR data: NMR relaxation data

In NMR experiments, the bulk magnetization of the sample is perturbed from it's equilibrium state, the emitted signal is measured as the sample returns to equilibrium; this transit from non-equilibrium to equilibrium magnetization is called relaxation. For proteins, the relaxation processes more frequently measured are those of the backbone NH bonds; in this context, it is typical to obtain the ^{15}N longitudinal relaxation rate ($R_1 = 1/T_1$) the ^{15}N transverse relaxation rate ($R_1 = 1/T_1$) and the heteronuclear **N**uclear **O**verhauser **E**ffect ($\{^1H\} - \{^{15}N\}$ **NOE**) . To extract this type of information from MD simulations, we follow a procedure described in detail in[102] that is briefly summarized bellow. We started from the fact that relaxation parameters can be expressed in terms of the spectral density:[103]

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)dt. \tag{2.65}$$

Here we can use again the factorization of the model free approach (equation (2.62)). However, the correlation time for global tumbling, t_c, cannot be accurately estimated from MD trajectories. It is necessary to take its experimental value or fit t_c as an additional parameter.
To consider richer variations of the internal autocorrelation function, $C_I(t)$ data extracted from the MD trajectories can be fitted to a normalized multi-exponential function:[104]

$$C_I(t) = A_0 + \sum_{k=1}^{5} A_k e^{-t/t_k}. \tag{2.66}$$

With $\sum_k A_k = 1 (k = 0, ..., 5), A_k \geq 0$ and $0 \leq t_k$ for all k. A_0 is equivalent to S^2 in equation (2.64).
The best fit parameters and the experimental t_c are used to calculate the spectral density function:

$$J(\omega) = \frac{A_0 2t_c}{1+(\omega t)^2} + \sum_{k=1}^{5} \frac{A_k 2t'}{1+(\omega t)^2}, \tag{2.67}$$

with $t' = t_c t_k/(t_c + t_k)$.
From the spectral density functions, we can estimate the NMR observables R_1, R_2, and NOE in a variety of ways. [105] In this work, we used the following equations[102]:

$$\frac{1}{T_1} = R_1 = d_{00}[3J(\omega_N) + J(\omega_{H-N}) + 6J(\omega_{N+H})] + c_{00}\omega_N^2 J(\omega_N), \tag{2.68}$$

$$\begin{aligned}\frac{1}{T_2} = R_2 = \frac{1}{2}d_{00}[4J(0) + 3J(\omega_N) + J(\omega_{H-N}) + 6j(\omega_N) + 6J(\omega_N)] + ... \\ ... + \frac{1}{6}c_{00}\omega_N^2[4J(0) + 3J(\omega_N)],\end{aligned} \tag{2.69}$$

$$NOE = 1 + \frac{\gamma_H}{\gamma_N} d_{00} T_1 [6J(\omega_{H+N} - J(\omega_{H-N}))], \tag{2.70}$$

where $d_{00} = (1/20)(\mu_0/4\pi)^2(h/2\pi)^2\gamma_N^2\gamma_H^2\langle r_{NH}^{-3}\rangle^2$, $c_{00} = (1/15)\Delta\sigma^2$, μ_0 is the vacuum permeability, h is the Planck's constant, γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^1H respectively, $r_{NH} = 1.02$Å is the $N-H$ bond length, and $\Delta\sigma = -160$ *ppm* is the chemical shift anisotropy of ^{15}N in an amide group. ω_N, ω_H, ω_{N+H} and ω_{H-N} are the Larmor frequencies as well as their sum and difference in radians per second of ^{15}N and ^1H spins, respectively.

2.4.2 Analysis of MD simulations

2.4.2.1 Principal Component Analysis (PCA)

Principal **C**omponent **A**nalysis (**PCA**) is one of the oldest techniques for multivariate statistical analysis. Although the method was first introduced by Pearson in 1901,[106] Hotelling[107] independently derived a method that is similar to those we use today. The aim of PCA is to reduce the dimensionality of a data set in which there is a large number of interrelated variables. The reduction is achieved by transforming the original N variables to a new set of N uncorrelated (*i.e.,* orthogonal, see Fig.2.4) variables which are ordered in a way that the first new M variables (*i.e.,* the principal components) retain most of the information (*i.e.,* variability) present in the original set of variables. For a determined principal component $PC_{(1)}$, it is mathematically equivalent to find a linear combination of the observed variables X_j, $j = 1, 2, ..., p$

$$PC_{(1)} = w_{(1)1}X_1 + w_{(1)2}X_2 + ... + w_{(1)p}X_p, \tag{2.71}$$

where the weights $w_{(1)1}, w_{(1)2}, ..., w_{(1)p}$ have been chosen to maximize the ratio of the variance of $PC_{(1)}$ to the total variation, subject to the constraint $\sum_{j=1}^{p} w_{(1)j}^2 = 1$. The second principal component, $PC_{(2)}$, is the weighted linear combination of the observed variables which is uncorrelated with the first linear combination and which accounts for the maximum amount of the remaining total variation not already accounted for by $PC_{(1)}$. In general, the principal component is then the weighted linear combination of the X's,

$$PC_{(m)} = w_{(m)1}X_1 + w_{(m)2}X_2 + ... + w_{(m)p}X_p, \tag{2.72}$$

that has the largest variance of all linear combinations uncorrelated with all of the previously extracted principal components.

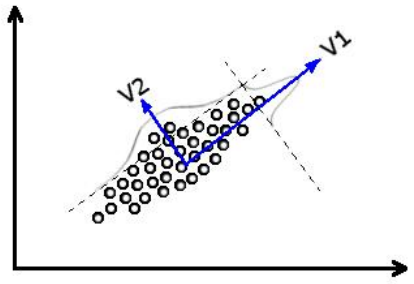


Figure 2.4: Schematic representation of the projections of variables in orthogonal principal components.

The last step in PCA is determining the number of "meaningful" components to retain. This can be achieved by several methods. A common approach is to build a Scree plot[108] which is a plot of the eigenvalues associated with each component (Fig. 2.5).

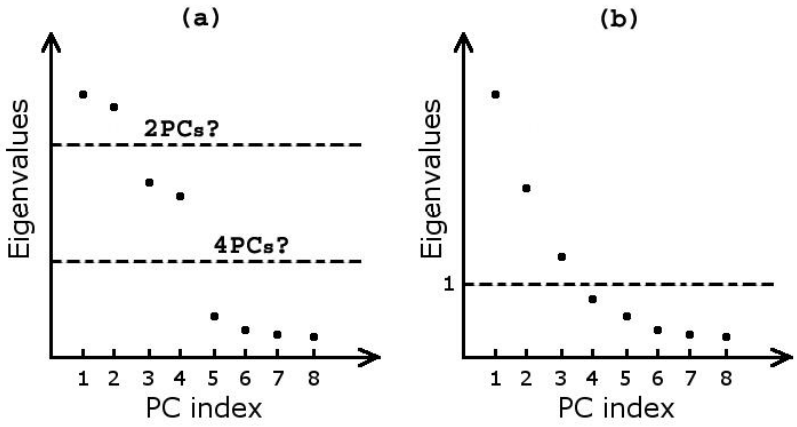


Figure 2.5: **(a)** Scree plot with two breaks which renders arbitrary the selection of the principal components. **(b)** Scree plot with no obvious break underlying the use of the Guttman-Kaiser criterion.[109–111]

In a Scree plot, it is desirable to find a "break" between components with relatively large eigenvalues and components with relatively low eigenvalues. The components that appear before the break are assumed to be meaningful and are retained; those appeared after the break are assumed to be unimportant and are not retained. This method however is corrupted by subjective decisions, besides, in some cases there can be several breaks in the data (Fig. 2.5**(a)**) and in other

cases no obvious breaks can be found (Fig. 2.5**(b)**). A simple way to deal with this drawback is to use the Guttman-Kaiser criterion.[109–111] In this approach, any component with an eigenvalue greater than 1 is considered to be meaningful and is therefore retained for further analysis (Fig. 2.5**(b)**). In some cases (*e.g.,* data has to many variables with low communalities between them), the use of the Guttman-Kaiser criterion can lead to retain a wrong (or difficult to analyze) number of components In these cases, it is sometimes useful to add some knowledge about the system from which the data is extracted when deciding how many components to retain; this is called the "interpretability" criterion. In this way, the researcher finds a compromise between a proper number of principal components that retain the maximum variability of the original data; in this way, for a "meaningful" component we also estimate the proportion of the variance that it can explain as:

$$\%\ of\ the\ PC_{(m)} = \frac{eigenvalue\ of\ the\ PC_{(m)}}{total\ eigenvalues\ of\ the\ correlation\ matrix}. \tag{2.73}$$

In practice, we perform PCA by calculating the eigenvectors of the covariance matrix of C_α fluctuations all-atom MD simulations. The PCA calculations were performed with the Dynatraj program.[112] For the selected $PC_{(m)}$ we made an assignment of residues to each rigid domain. Rigid domains and the hinges formed between them were identified using the scheme developed by Wriggers and Schulten.[113]

2.4.2.2 Normal Mode Analysis (NMA)

Normal **M**ode **A**nalysis (**NMA**) is a well known technique for dimensionality reduction. It was first used to study large-scale internal dynamics of proteins in the early 1980's.[114–116] In these analysis, it is often assumed that the normal modes with lower frequencies and larger amplitudes of motion (soft modes) are the ones that are functionally relevant. Although some experimental justification has been found for this statement,[117] this assumption cannot be considered to be true in all cases. Another assumption of NMA is that, over the range of thermal fluctuations, the conformational energy surface can be characterize by a harmonic approximation to a single energy minimum, this assumption has also been challenged both by experiments[118] and computational studies.[119] In spite of its limitations, NMA has proven to be useful and accurate in a variety of situations.[120]

To effectively calculate normal modes we need, as input, a set of initial coordinates for the atoms and a force field describing their interactions. Then three computational steps are required: (i) energy minimization of the conformational potential energy as a function of the coordinates; (ii) calculation of the Hessian matrix; (iii) diagonalization of the Hessian matrix.

At the minimum configuration, the potential energy can be expanded as a Taylor series in terms of mass-weighted coordinates, $q_i = \sqrt{m_j}\Delta X_j$, $q_{i+1} = \sqrt{m_j}\Delta Y_j$, $q_{i+2} = \sqrt{m_j}\Delta Z_j$, where j represents one of the N atoms and i represents one of the $3N$ Cartesian coordinates. The resulting Taylor expansion would have the following form:

$$U = \frac{1}{2}\sum_{i,j=1}^{3N}\frac{\partial^2 U}{\partial q_i \partial q_j} + \dots. \tag{2.74}$$

In the expansion, the first term (energy at the minimum) as well as the linear terms and higher order terms are neglected; the second derivatives calculated at the minimum are assumed to characterize the energy surface even at fluctuations far from equilibrium. The symmetric matrix F formed by these second derivatives is called the Hessian. The Lagrangian (kinetic energy minus potential energy) can then be written as:

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{3N} \dot{q_i}^2 - \frac{1}{2}\sum_{i,j=1}^{3N} \left.\frac{\partial^2 U}{\partial q_i \partial q_j}\right|_0 q_i q_j, \tag{2.75}$$

or in matrix form:

$$\mathcal{L} = \frac{1}{2}\dot{q}^t \dot{q} - \frac{1}{2} q^t F q, \tag{2.76}$$

in which, the superscript t denotes the transpose.
As F is a symmetric matrix, eq. 2.76 can be diagonalized and considered as an eigenvalue equation. Normal modes and their frequencies of oscillation are determined by the eigenvectors and eigenvalues of F. Tirion[121] made a significant simplification to this technique by eliminating the need for minimization and replacing the Hessian by a matrix whose elements are zero for any pair of atoms separated by a distance greater than a cutoff equal to the sum of their van der Waals radii plus a distance parameter R_c. If the pair of atoms are within the cutoff then the corresponding value in the matrix is calculated according to a Hookean pairwise potential with the same force constant for all pair of atoms. The B-factors predicted by this method are in good agreement with experiments.[122]

2.4.2.3 Correlated Motions

A broadly employed procedure to determine the internal correlated movements of a protein in MD simulations is the calculation of the normalized covariance matrix of the atomic fluctuations,[123] also known, for the one-dimensional case, as the Pearson coefficients.
A modification of this method that eliminates the inconsistencies of the Pearson coefficients was recently proposed;[124] the resulting generalized correlation coefficients can be defined as:

$$r_{MI}[X_i, X_j] = \{1 - e^{(-2I[X_i, X_j]/d)}\}^{-1/2}, \tag{2.77}$$

where X_i is the positional fluctuation vector for atom i and I is the mutual information.
The value of r_{MI} is in the interval [0,1] and like I, it vanishes for uncorrelated distributions.

2.4.2.4 Poisson-Boltzmann calculations

The electrostatic interaction between molecules in ionic solutions can be described by the use of Poisson-Boltzmann calculations. A starting point for this calculation is the *Poisson* equation that relates in a simple form the electrostatic potential ϕ, the charge distribution ρ, and the dielectric constant ε:

$$\nabla[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}). \tag{2.78}$$

The *Boltzmann* distribution of ions in the solvent can be expressed as:

$$q_+ = qce^{-q\phi/kT}, \tag{2.79}$$

$$q_- = -qce^{-q\phi/kT}. \tag{2.80}$$

These factors imply that following the pattern dictated by the thermal fluctuations, the negative ions will be located close to the areas where the potential is positive and vice versa. Including these terms in equation (2.78), we obtain the **P**oisson-**B**oltzmann **E**quation (**PBE**):

$$\nabla[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \kappa^2\left(\frac{kT}{q}\right)\sinh\left(\frac{q\phi(\mathbf{r})}{kT}\right) = -4\pi\rho(\mathbf{r}), \tag{2.81}$$

$$\kappa^2 = \frac{8\pi q^2 I}{kT}, \tag{2.82}$$

where I is the ionic strength in the solution.
In the general case, there is no exact solution for these equations; therefore they must be solved numerically. To do this, we build a grid on the system and solve the PBE on every node of the grid; the obtained values of the electrostatic potential per node can be mapped on the surface of the solute, which might suggest ideas about how the solute can interact with other polar molecules.

2.5 Other simulation methods

2.5.1 Brownian dynamics (BD)

In this section, we focus on the methodological aspects of **B**rownian **D**ynamics (**BD**) applied to the study of soft matter issues like the protein-protein interaction; especial interest will be devoted to measure the association rate between proteins. The simulation method used in this thesis to determine the rate constants had its origins in the group of McCammon in 1984,[125] since then several improvements has been incorporated into a program called MacroDox.[126] The idea behind the method is deceptively simple: the bimolecular rate association constant is inferred from the fraction of trajectories in which the reactant species achieve a favorable geometry for the association. In the next sections, we give the theoretical basis for this method.

2.5.1.1 Brownian dynamics algorithm

The original algorithm used in these studies was first developed by Ermack and McCammon,[127] but has important precedents with the development of Fokker-Planck equations for two[128] and n-brownian particle systems.[129]
In the present algorithm, the equation of motion of molecule 1 with respect to a reference frame of molecule 2 is given by:

$$\mathbf{r}_1 = \mathbf{r}_1^0 + \frac{2\Delta t}{k_B T}(D_a^0 \cdot \mathbf{F}_1^0 - D_{12}^0 \cdot \mathbf{F}_1^0) + \mathbf{R}_1 - \mathbf{R}_2, \tag{2.83}$$

with $D_a^0 = (D_{11}^0 + D_{22}^0)/2$; $\mathbf{r}_1^0$ and $\mathbf{r}_1$ denote the position of molecule 1 before and after the Brownian dynamics step; Δt is the time step, D_{ij} is the relative diffusion tensor between molecules i and j and $\mathbf{F}_1$ is the force on molecule 1 due to molecule 2. $\mathbf{R}_i$ are vector Gaussian numbers with mean, $\langle \mathbf{R}_i \rangle = 0$ and variance-covariance, $\langle \mathbf{R}_i \mathbf{R}_j \rangle = 2D_{ij}^0 \Delta t$ representing stochastic displacements due to solvent collisions. In this approach, hydrodynamic interactions are incorporated through the diffusion tensors, which can be approximated by:

$$D_{ij} = k_B T \left[\frac{\delta_{ij}}{4\pi\eta a_i} + (1 - \delta_{ij}) T_{ij} \right], \tag{2.84}$$

where T_{ij} is the Oseen tensor[130]:

$$T_{ij} = \frac{1}{8\pi\eta R}\left(1 + \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{\mathbf{r}_{ij}^2}\right), \tag{2.85}$$

$$R = \begin{cases} a_i + a_j, & \mathbf{r}_{ij} < a_i + a_j \\ \mathbf{r}_{ij}, & \mathbf{r}_{ij} \geq a_i + a_j. \end{cases} \tag{2.86}$$

2.5.1.2 Rate constants from Brownian dynamics

An initial and simplified approach to calculate the diffusional bimolecular rate constant is to measure the ensemble flux (*e.g.,* from BD trajectories) through a "reaction surface" of Brownian particles whose motions are governed by the steady state Smoluchoski equation:

$$\nabla \cdot D(\mathbf{r}) \cdot [\nabla - \mathbf{F}(\mathbf{r})/k_B T]\rho(\mathbf{r}) = 0, \tag{2.87}$$

or

$$\nabla \cdot [-J(\mathbf{r})] = 0, \tag{2.88}$$

where $\rho(\mathbf{r})$ is the pair probability density at configuration $\mathbf{r}$, $\mathbf{F}(\mathbf{r})$ is minus the gradient of the intermolecular potential mean force $U(\mathbf{r})$, and $D(\mathbf{r})$ is the relative diffusion tensor. The total integrated flux through the "reactive surface" is a scalar quantity J which is related to the bimolecular rate constant k as:

$$J(b) = k\rho_0. \tag{2.89}$$

For calculation purposes b is chosen in a way that for $r > b$ the effect of charge asymmetry in the interacting particles can be considered negligible (the interactions are therefore isotropic at large distances, see Fig. 2.6). If a Brownian particle reaches a point $r = q \gg b$ then the BD trajectory is not monitored anymore and the particle is considered to be at an infinite distance and therefore unlikely to interact again with its partner.

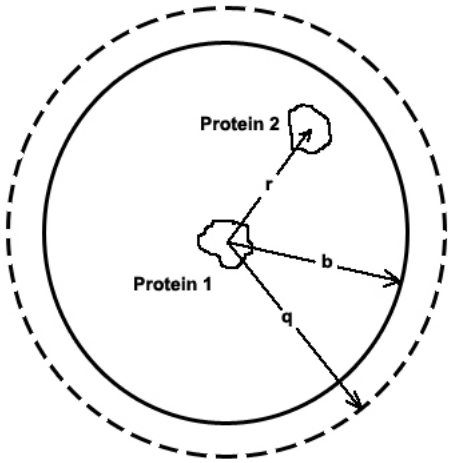


Figure 2.6: Schema for Brownian Dynamics diffusional approach to collision. The distance between proteins is r; the anisotropic inner region is labeled b and the outer isotropic region is the space $b < r < q$ where q is the truncation distance (*i.e.,* "infinite") for the simulation.

The flux $J(b)$ can be further decomposed into the conditional flux containing any particles achieving separation b for the first time multiplied by the probability p that a particle pair achieving separation b will ultimately "react" rather than escape to infinite separation. The term "react" in this section will be intended as the binding of the interacting pair of Brownian particles fulfilling certain geometrical distance criteria in the protein-protein interaction surface.
For the flux decomposition we can then write:

$$k = k_D(b)p. \tag{2.90}$$

A generalized result for $k_D(b)$ is[131]:

$$k_D(b) = \left(\int_b^\infty \left[\frac{e^{[U(\mathbf{r})/k_B T]}}{4\pi \mathbf{r}^2 D(\mathbf{r})} \right] d\mathbf{r} \right)^{-1}. \tag{2.91}$$

After this step, the complicated anisotropic inner part of the calculation is isolated in the probability term p. In order to separate the part of p that is purely diffusive in nature, a new probability β_∞ representing the likelihood that particles achieving a separation $r = b$ will have at least one collision with the reactive site rather than escape to infinite separation. Another new probability α will define the likelihood of effective "reaction" upon a given collision at the reactive site. Therefore in the simplest case for $\alpha = 1$:

$$k = k_D(b)\beta_\infty. \tag{2.92}$$

For $\alpha < 1$ and several possible events of re-collision (each with probability Δ_∞), after unsuccessful reaction attempts, we found that:

$$p = \frac{\beta_\infty \alpha}{1 - (1 - \alpha)\Delta_\infty}; \tag{2.93}$$

therefore,

$$k = k_D(b)\beta_\infty \alpha / [1 - (1 - \alpha)\Delta_\infty]. \tag{2.94}$$

The monitoring of particles until separation distance q instead of to infinity implies that the fraction of successful trajectories subject to the truncation is no longer equal to β_∞ but to a slightly different quantity β, the relation between these quantities is given by:

$$\beta_\infty = \frac{\beta}{1 - (1 - \beta)\Omega}, \tag{2.95}$$

in which Ω is the diffusion probability that a particle at $r = q$ will eventually return to $r = b$. As before, since Ω involves diffusion only in the isotropic domain $r > b$ we can write:

$$\Omega = k_D(b) \int_q^\infty \left[\frac{e^{\left[U(\mathbf{r})/k_B T\right]}}{4\pi \mathbf{r}^2 D(\mathbf{r})} \right] d\mathbf{r}. \tag{2.96}$$

The truncation correction for Δ_∞ is:

$$\Delta_\infty = \Delta + \beta_\infty (1 - \Delta). \tag{2.97}$$

Combining equations (2.94), (2.95), and (2.97), we finally obtain the rate constant:

$$k = \frac{k_D(b)\left[\frac{\beta}{1-(1-\beta)\Omega}\right]\alpha}{1 - (1 - \alpha)\left(\Delta + \left[\frac{\beta}{1-(1-\beta)\Omega}\right](1 - \Delta)\right)}. \tag{2.98}$$

2.5.2 Random Expulsion Molecular Dynamics (REMD)

There are several methods that increase the probability of observing rare events in molecular dynamics simulations; a complete mention of these strategies is not an objective of the present thesis, the interested reader can refer to excellent reviews in the literature. [132]
The specific case of **R**andom **E**xpulsion **M**olecular **D**ynamics (**REMD**) was developed[133] to investigate the escape pathways of substrates from the active site. In REMD, the probability of spontaneous substrate exit is enhanced by imposing an artificial random force on every atom of the substrate in addition to the influences of the force field. The imposed force is defined as:

$$\mathbf{F} = k\mathbf{r}_0, \tag{2.99}$$

where k is the force constant (an adjustable parameter that remains constant during the entire simulation); $\mathbf{r}_0$ is a unit vector in a random direction. A REMD simulation proceeds in a following way: (i) Starting from a complex enzyme-substrate, a random vector is chosen, (ii) a force calculated with equation (2.99) is applied to the substrate, (iii) if after n MD steps the substrate has moved to a distance minor to $r_{\min}$, then a new vector $\mathbf{r}_0$ is chosen and a force in the new direction is applied; otherwise the direction of the force is maintained, (iv) if the separation between the enzyme and substrate reaches the value $r_{\max}$ the REMD simulation is stopped.

2.6 Bioinformatics methods

2.6.1 Protein structure prediction

The prediction of the 3D structure of a protein (tertiary structure) using as input its amino acid sequence (primary structure) is called protein structure prediction and has been for many year one of the fundamental goals of bioinformatics and theoretical chemistry. More sophisticated predictions based on 3D comparative procedures (*e.g.,* homology modeling[134]) or *ab initio* predictions[135] (based purely on physico-chemical principles without considering previously solved structures) are not included in the restricted definition we used here. These approaches were not used in this thesis and therefore they will not be discussed further.
A preliminary step in many protein structure prediction is secondary structure prediction, which attempts to predict protein local features (*e.g.,* helix, sheet) from the aminoacid sequence. As secondary structure is in great extension defined by the pattern of formed hydrogen bonds, some methods, such as the classic **D**ictionary of **P**rotein **S**econdary **S**tructure (**DSSP**) method[136] pay special attention to it. Other recent methods[137] borrow some characteristics from DSSP and add additional features.

2.6.2 Hot spot residues at the protein-protein interface: FastContact algorithm

Protein-protein recognition plays an important role in a large number of biomolecular processes. Even though the structural details of a growing quantity (>28300) of protein-protein complexes has been published in the PDB database, the general rules to energetically describe these interactions are still a matter of great debate. In this thesis, we estimate the binding energies with the FastContact algorithm.[138]
FastContact (v. 2.0) screens protein binding interactions using a scoring function in which the interaction between two proteins (ΔG_{bind}) is decomposed in two terms: the standard intermolecular Coulombic electrostatic potential (ΔE_{elect}) and an empirical contact potential that captures the most important characteristics of the desolvation free energy in proteins (ΔG_{desolv}):

$$\Delta G_{bind} = \Delta E_{elect} + \Delta G_{desolv}, \tag{2.100}$$

$$\Delta G_{desolv} = g(\mathbf{r}) \sum \sum e_{ij}, \tag{2.101}$$

where e_{ij} is the atomic contact potential between atoms i in the receptor and atoms j in the ligand, $\mathbf{r}$ is the distance between atoms. The double sum is taken over all atom pairs and $g(\mathbf{r}) = 0$ if $\mathbf{r} > 7$Å, $g(\mathbf{r}) = 1$ if $\mathbf{r} < 1$Å, and between these two limits $g(\mathbf{r})$ is a smooth function.[138]

Chapter 3

ANTHRAX LF INVESTIGATED BY MD SIMULATIONS

3.1 Introduction

The anthrax infection caused by the bacterium *Bacillus anthracis* is posing significant threat in biological warfare and terrorism. If ingested or inhaled, the anthrax bacterial spores germinate, resulting in a toxaemia that is usually fatal to the host.[10,139–141] Unfortunately, the only way to intervene against anthrax intoxication is to give a generic antibiotic treatment at the early stage of the disease.[142] Thus, there is presently a tremendous effort in investigating the molecular mechanisms responsible for anthrax infection to develop new therapeutic agents. Most of anthrax toxic effects are caused by the so-called lethal toxin, a complex consisting of the **P**rotective **A**ntigen (**PA**) and **L**ethal **F**actor (**LF**) proteins.[143] **PA** is the membrane-translocating component of the complex; it binds to a host cell-surface receptor and translocates **LF** into the cytosol.[11,144–146] **LF** is a cytoplasmatic zinc metalloprotease that cleaves the N-terminal region of selected members of **M**itogen-**A**ctivated-**P**rotein-**K**inase-**K**inase (**MAPKK**) family;[147] **MAPKKs** govern the **MAPK** signaling pathway, controlling the genomic and physiological response of the cell to its environment.[148] **LF** alters different cell types, apparently, in an evolutionary conserved manner.[149–152]

X-ray crystallographic studies provided the structural determinants of **LF** in the free state (PDB: 1J7N),[20] with a segment of one of its substrates, MAPKK-2 (PDB: 1JKY),[20] with an optimized peptidic substrate (PDB: 1PWV, 1PWW)[21] and with synthetic inhibitors (PDB: 1PWP, 1ZXV).[153,154] Domain **I** of **LF** binds to **PA**, while domains **II-IV** create a long groove that holds MAPKK-2 N-term[20](Fig.3.1).

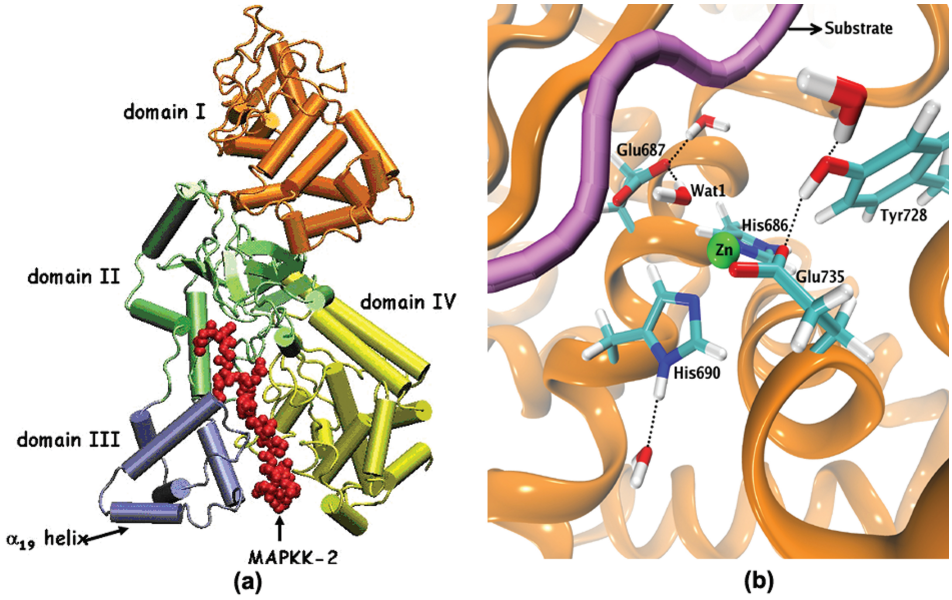


Figure 3.1: **(a)** The structure of anthrax **L**ethal **F**actor (**LF**) in complex with its MAPKK-2 substrate (red) obtained by X-ray crystallography (PDB: 1JKY),[20] includes the following domains: **I** (orange, residues 1-262) is the **P**rotective **A**ntigen (**PA**) binding domain; **II** (green, residues 263-297, 385-550) is called the **V**egetative **I**nsecticidal **P**rotein 2 (**VIP2**)-like domain because of its similarity with the ADP-ribosyltransferase from *Bacillus cereus* toxin; **III** (blue, residues 303-383), is the helix bundle domain; and **IV** (yellow, residues 552-776) is the catalytic domain. **(b)** Snapshot from the all-atom MD trajectory featuring the active site (domain **IV**) with the crucial residues for the catalytic activity[20–22] (shown in licorice representation) and the position of the optimized substrate[21] (shown in purple) used in the simulations.

Domain **IV** performs the enzymatic catalysis; it features, in its active site, a zinc ion coordinated by two histidines (His686 and His690) and a glutamate (Glu735). The tetrahedral coordination for zinc is completed with a water molecule (or an hydroxide group), which is probably the nucleophilic agent in the catalysis. Similar tetrahedral coordination spheres have also been found in related metalloproteases from the carboxypeptidase and thermolysin families.[4] Clearly, for proteolytic reactions like the one catalyzed by **LF**, the nature and protonation state of the nucleophilic agent is crucial. The zinc site may be stabilized by outer shells groups,[20,155] as found in other zinc-enzymes.[156] For **LF**, these may include: (i) Glu687, which H-bonds the catalytic water and is believed to play a key role acting as a general base during the reaction;[20] this proposal is consistent with the loss of activity of the **LF** mutant E687C;[22,157] (ii) Tyr728, which H-bonds Glu735 and whose conservative mutation to phenylalanine (Y728F) impairs the catalytic activity;[155] (iii) Glu739, although no mutagenesis data are available for this residue, it may play a role by interacting with the substrate or forming an H-bond with His686 (see PDB: 1JKY).[20] A complete structural description of the Michaelis complex may help to develop novel peptidomimetic inhibitors with therapeutical properties. In this work we provide a structural model of such complex in aqueous solution using several computational tools; *e.g.*, **D**ensity **F**unctional **T**heory (**DFT**), all-atom **M**olecular **D**ynamics (**MD**) simulations, calculations based on **C**oarse-**G**rained

(**CG**) models[82,121,158] and bioinformatics approaches.[137,159–161] Our calculations show that the nucleophilic agent for the **LF** catalyzed hydrolysis is a water molecule (not an OH group). We provide data indicating a correspondence between the substrate per residue interaction energies (with **LF**) and **LF** substrate selectivity. On the other hand, the large scale motions of the enzyme do not appear to play a fundamental role in the enzymatic reaction, as has been suggested for other proteases.[162] Finally, our results indicate that at least a part of helix α19 in domain **III**, which is present in only one solid state structure of **LF** (PDB: 1JKY),[20] assumes preferentially a coiled conformation; this is consistent with the unusually large temperature factors reported for this region in the X-ray structure.[20]

3.2 Methods

3.2.1 Construction of the LF Michaelis complex

The determination of the protonation state of **LF** active site is non-trivial. Here, using the same procedure as that of ref.[156]**(b)**, we perform **DFT** calculations on a series of models (**A**-**D**, Fig.3.2) based on the X-ray structure of the free enzyme (PDB: 1J7N).[20]

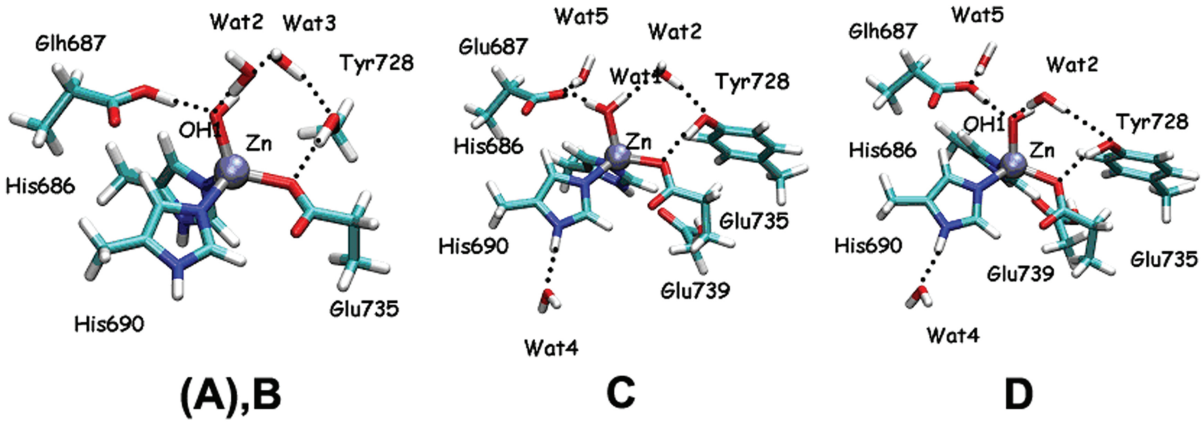


Figure 3.2: Protomers of **LF** active site considered in this work: model **A** and **B** (60 atoms); models **C** and **D** (84 atoms). Model **A** (not shown) was highly unstable; during geometry optimization it was rapidly transformed into model **B** bearing a protonated Glu687 (labeled Glh687 in the figure) and OH1 instead of Wat1. Models **C** and **D** turned out to have the same stability within the accuracy of **DFT**.

Models **A** and **B** (60 atoms) included: (i) the zinc atom and its ligands: His686 and His690, Glu735 (all cut at C_β atom, and saturated with hydrogen atoms), and either a water molecule (Wat1, model **A**) or a hydroxide group (OH1, model **B**); (ii) Glu687, either in its ionized state (model **A**) or protonated at Oε2 (model **B**); this residue, which forms an H-bond with the nucleophile, was also cut at C_β; (iii) the water molecule Wat2, detected in the X-ray structure, which H-bonds to Wat1; (iv) the water molecule Wat3, which bridges Wat2 and Tyr728 (Fig. 3.2); (v) Tyr728 (modeled as methanol), which H-bonds to Glu735. Models **C** and **D** (84 atoms) included the same groups as **A** and **B** respectively, as well as: (i) Glu739 (cut at the C_γ) which H-bonds to His686; (ii) the crystallographic water molecules Wat4 and Wat5, which H-bond to His690 and Glu687, respectively; (iii) the aromatic ring Tyr728 (cut at C_β), which H-bonds to Wat2 and Glu735. **DFT** calculations were performed using the program **CPMD**[89] with a plane waves basis set up to an energy cutoff of 70 Ry. The core/valence interactions were described using norm conserving pseudopotentials of the Martins-Troullier type.[163] Integration of the nonlocal parts of the pseudopotential was obtained via the Kleinman-Bylander scheme[164] for all of the atoms except zinc, for which a Gauss-Hermite numerical integration scheme was used. The gradient corrected Becke exchange functional and the Lee-Yang-Parr correlation functional

(BLYP) were used.[165,166] Periodic boundary conditions were applied and we used orthorhombic cells with edges $a = 16.0$Å, $b = 19.2$Å, and $c = 12.8$Å for models **A**, **B** and $a = b = 17.0$Å, and $c = 14.5$Å for models **C**, **D**. Isolated system conditions were applied.[167] The Michaelis complex was built by reproducing the protonation state resulting from **DFT** calculations, on the X-ray structure of the **LF** mutant E687C (which is unable to perform catalysis)[21] in complex with an optimized substrate featuring the **MAPKK** consensus sequence /VYPYPMEPT/ around the scissible bond (PDB: 1PWW).[21] The wild type enzyme was constructed by replacing Cys687 with a glutamic residue and by adding residues 346-367, missing in this X-ray structure, in the same conformation (α-helix) as they were found in the only X-ray structure of **LF** which provides their positions (PDB: 1JKY).[20] The histidines located outside the **LF** active site were protonated in Nπ, with exception of His35, His91, His229, His277, His309, His588, that were protonated in Nτ. The two protomers were neutralized by adding 22 potassium counterions immersed in a box of 128.47, 81.25, and 94.02Å, containing ∼26,700 water molecules. The total size of the systems was ∼92,600 atoms. In the **MAPKK** consensus sequence /VYPYPMEPT/, the conserved reactive proline (*i.e.,* the one placed between two tyrosines) was labeled P1. Residues on the left side of P1 were labeled P2, P3 to Pn, while residues on the right side were labeled P1', P2' to Pn'.[21] The structural characteristics of the Michaelis complex were studied using all-atom MD simulations, coarse-grained methods[82,121,158] and bioinformatics tools;[137,159–161] special attention was given to the stability of the reconstructed region α19 (residues 346-367). In addition, we investigated the electrostatic properties of the complex using the Poisson-Boltzmann approach and **DFT** calculations.

3.2.2 All-atom MD simulations

The AMBER parm98[83] force field was adopted for the substrate, the potassium counterions and the enzyme (**LF**) regions outside the active site. The parametrization of the Zn coordination sphere in the active site followed the procedure of ref [168] (see Appendix A.4). The electrostatic interactions were evaluated using the **P**article **M**esh **E**wald (**PME**) method.[169] A cutoff of 10Å was used for the van der Waals interactions and the real-space part of the electrostatics. The bonds involving hydrogen atoms were kept fixed using the SHAKE algorithm.[170] A time step of 2 fs was used. The initial structures were relaxed by short minimization runs of 2000 steps using the conjugate gradient energy minimization algorithm. 100 ps of MD at constant volume were then performed during which the system was gradually heated to 300 K. Constant temperature (300 K) and pressure (1 atm) production runs were performed by coupling the systems to a Berendsen thermostat and barostat.[171] The NAMD simulation software was used.[172] A trajectory of 50 ns was computed and the following properties were calculated:
(a)RMSD/RMSF. **R**oot **M**ean **S**quared **D**eviations (**RMSD**) and **R**oot **M**ean **S**quared **F**luctuations (**RMSF**) of the C_α atoms were calculated from the all-atom MD trajectory. The structural stability of the complex during the simulations was monitored by using the **RMSD**; the normalized **RMSFs** were compared with the normalized temperature factors (B-values) from the X-ray structure (PDB: 1JKY).[20]
*(b)**P**rincipal **C**omponent **A**nalysis (**PCA**).* Large scale motions were calculated as eigenvectors of the covariance matrix of C_α fluctuations, constructed from **PCA**. The **Dynatraj** program[112] was used to perform **PCA** on the last 15 ns of the all-atom MD simulation. For the first three principal components, rigid domains and hinges were identified using the scheme developed by

Wriggers and Schulten.[113] Details of these calculations are reported in Appendix A.6.

3.2.3 Hybrid Coarse-Grained/Molecular Mechanics (CG/MM) simulations

In this approach, helix α19 (residues 346-367) and protein or solvent atoms within 12.5Å from α19 were treated with an all-atom force field (**MM** region); the rest of the protein was treated with the Go simplified potential[85] (**CG** region). The effect of the solvent outside the **MM** region was considered as the sum of stochastic and frictional forces proportional to the mass and velocities of the particles in the system.[173] We used the same procedure as in Ref.[82], which is summarized in Appendix A.5. Two 60 ns CG/MM simulations using AMBER parm98[83] and Gromos96 43a1[84] force fields were performed; each of them started from a snapshot taken at 3 ns of the all-atom MD trajectory. These simulations were preceded by 1000 steps of energy minimization (using the steepest descend algorithm) followed by a gentle heating from 0K to 300 K in 500 ps. Normalized **RMSFs** for the C_α atoms were estimated and compared with the normalized temperature factors from the X-ray structure (PDB: 1JKY)[20]. Additional **CG/MM** simulations, starting from different initial structures, were computed to ensure the reproducibility of these results (see Appendix A.5).

3.2.4 Normal Mode Analysis (NMA)

NMA was performed with the NOMAD-ref server[174] on the energy-minimized structure taken from the last frame of the all-atom MD simulation. In this scheme,[121,158] the protein was represented by a network of beads connected by harmonic springs; only the interactions between the beads separated by a distance ≤ 3Å were considered.[121] Normalized C_α **RMSFs** were estimated and compared with the corresponding normalized temperature factors from the X-ray structure (PDB: 1JKY).[20] Additional **NMA** calculations, starting from different initial structures, were performed in order to ensure the reproducibility of these results (for details see Appendix A.6).

3.2.5 Bioinformatics

We investigated the propensity for disorder of helix α19 by using several prediction programs: *e.g.,* PredictProtein,[137] PSIPRED,[159] SPRITZ,[160] and HNN.[161] (see Appendix A.7 for details)

3.2.6 Electrostatics

(a) Poisson-Boltzmann calculations. Electrostatic surface potentials for both, **LF** and substrate, were calculated by solving the Poisson-Boltzmann equations with the **APBS**[175] and **PDB2PQR**[176] programs; the results were visualized using a **PYMOL** interface.[177] These calculations were made on the energy minimized structure taken from the last frame of the all-atom MD trajectory featuring model **C** in the active site (see section 3.2.1).
(b) Polarization of the active site. The polarization of selected chemical bonds in the active site was investigated using the so called **B**ond **I**onicity (**BI**) indexes[96] that can be estimated from **DFT** calculations. We considered active site models under different environments: in vacuo and with the influence of the solvent and/or **LF** electric fields. To construct the models (represented in Appendix A.2) we used 15 equally spaced frames from the last 15 ns of the all-atom MD

trajectory. BI_{AB} of a bond between two atoms A and B is defined as

$$BI_{AB} = \frac{d_A}{d_{AB}} \tag{3.1}$$

where d_A is the distance between atom A and the Boys orbitals[96] along the AB bond, and d_{AB} is the length of the bond between A and B. A value of BI=0.5 (the Boys orbital is in the middle of the bond) indicates absence of polarization; while values close to 1 or 0, indicate polarization.

3.3 Results and Discussion

The main purpose of this work is to characterize the structural and electrostatic properties of the Michaelis complex formed by **LF** and an optimized substrate that features the **MAPKK** consensus sequence /VYPYPMEPT/.[21] The first step to achieve our goal is determining the correct protonation state in **LF** active site residues using **DFT** calculations.

3.3.1 Protonation state at the active site

As mentioned in the introduction of this chapter, a critical issue in Zn-based hydrolases is the determination of the protonation states of residues in the active site (in particular that of the nucleophile).[156]

In this work, we addressed this issue by performing **DFT** calculations on increasingly complex models of the active site (Fig.3.2). The smallest models (**A**-**B**) include only the Zn site, whilst the largest models (**C**-**D**) include additional second-shell ligands of established (*e.g.,* Tyr728)[155] or putative (*e.g.,* Glu739)[20] relevance for the enzymatic reaction (see section 3.2.1 for details). The most likely protomers were defined as those associated with the lowest potential energy and with the lowest **RMSD** relative to the reference X-ray structure (**LF** in the free state, PDB code: 1J7N).[20]

In the smallest models, **A** featured Glu687 in the ionized state and Wat1 as nucleophile, whilst **B** exhibited Glu687 in its neutral state and the nucleophile was an hydroxide group. During the geometry optimization model **A** was unstable, as Wat1 transferred a proton to Oε2@Glu687, resulting in model **B** (see Fig. 3.2). The latter was instead stable; it featured a slightly distorted tetrahedral coordination geometry and establishing the H-bonds b(Hε2@Glu687, O@OH1) and b(Hη@Tyr728, Oε2@Glu735), which are also putatively present in the X-ray structure (Table 3.1). The distances Zn-X (X=coordinating atom), decreased by $\sim$0.1-0.3Å relative to the X-ray structure (Table 3.1); the **RMSD** between model **B** and the X-ray structure (which was sizeable: $\sim$0.55Å) increased; both effects were possibly caused by the limited size of the model.

In the largest models, **C** featured Glu687 in the ionized state and Wat1 as nucleophile (like **A**), whilst **D** exhibited Glu687 in its neutral state and the nucleophile was an hydroxide group (like **B**). Both models (**C** and **D**) were stable and their differences in energetics and in structural properties were not significant within the accuracy of **DFT** calculations. Some general trends in models **C** and **D** with respect to model **B** could be identified: (i) the bonds b(Zn, Nτ@His686) were shorter (Δd=-0.11 for **C** and $\Delta d = -0.10$Å for **D**), (ii) one bond b(Zn, Nτ@His690) was longer in **D** ($\Delta d = +0.03$Å) and unaltered in **C**, (iii) the bonds b(Zn, O@[OH1,Wat1]) were larger ($\Delta d = +0.08$Å for **C** and $\Delta d = +0.05$Å for **D**), (iv) the hydrogen bonds between Glu687 and the nucleophile b(Oε2@Glu687, H@[OH1,Wat1]) were longer ($\Delta d = +0.14$Å for **C** and $\Delta d = +0.02$Å for **D**). On the other hand, the hydrogen bond between Wat4 and Glu687 (not included in the smallest models) was different between models **C** and **D** (Table 3.1).

	X-ray	**B**	**C**	**D**
Bond length (b(A_i,B_j); in Å)				
b(Zn,O@[OH1,Wat1])	2.1	1.96	2.04	2.01
b(Zn,Nτ@His690)	2.1	2.04	2.04	2.07
b(Zn,Nτ@His686)	2.3	2.20	2.09	2.10
b(Zn,Oε2@Glu735)	2.3	2.00	2.01	2.03
b(Oε2@Glu687, H@[OH1,Wat1])	3.6	1.50	1.64	1.52
		(2.57)	(2.66)	(2.58)
b(Hη@Tyr728,Oε2@Glu735)	2.7	1.75	2.00	1.77
		(2.79)	(2.94)	(2.75)
b(Hπ@His690,O@Wat4)	2.9		1.96	1.95
			(2.97)	(2.97)
b(Hπ@His686,Oε1@Glu739)	2.9		1.96	1.91
			(2.93)	(2.88)
b(Oε2@Glu687,H@Wat5)	2.9		1.86	2.00
			(2.85)	(2.98)
Angles (τ(A_i, B_j, C_k), in deg)				
τ(Oε1@Glu735,Zn,O@[Wat1, OH1])	96	121	116	113
τ(Nτ@His686,Zn,O@[Wat1,OH1])	107	102	105	102
τ(Nτ@His690,Zn,O@[Wat1,OH1])	128	110	105	105
RMSD (Å)				
		0.55	0.36	0.58

Table 3.1: Comparison between calculated (**DFT**) and experimental (X-ray; PDB: 1J7N)[20] structural parameters (bond lengths (Å) and angles (deg)) for models of the free enzyme. The distances in the parenthesis are those between heavy atoms in hydrogen bonds.

3.3.2 Molecular dynamics of LF Michaelis complexes

Next, we built two Michaelis complexes using protonation states **C** and **D** for the active site (see section 3.2.1). We performed all-atom MD simulations on both Michaelis complexes.

3.3.2.1 MD of LF Michaelis complex in C protonation state

LF Michaelis complex with protonation state **C** was stable during the entire simulation (50 ns). In particular, the bond lengths in the coordination sphere had small fluctuations around their average positions during the dynamics (Table 3.2) and the substrate remained in its binding site for the entire simulation time.

Pairs of atoms	**Distance (Å)**
b(Zn,Oε2@Glu735)	2.22(0.07)
d(Zn,Oε2@Glu687)	4.8(0.2)
b(Zn,Nτ@His686)	2.05(0.05)
b(Zn,Nτ@His690)	2.15 (0.05)
b(Zn,O@Wat1)	1.98(0.06)
d(Cδ@Glu735,Cδ@Glu687)	6.9(0.2)
d(Nτ@His686,Nτ@His690)	2.9(0.1)
b(H@Wat1,Oε2@Glu687)	1.9(0.2)
b(O'@Tyr-P2,H@Wat1)	1.8(0.2)
b(Hη@Tyr728,O'@Pro-P1)	1.8(0.2)
b(Hπ@His686,Oε1@Glu739)	1.97(0.2)

Table 3.2: Distances between selected pairs of atoms in the **LF** active site, including: (i) Zn-coordination bonds (*e.g.,* b(Zn,Oε2@Glu735); (ii), hydrogen bonds (*e.g.,* b(H@Wat1,Oε2@Glu687) and key geometrical features (*e.g.,* d(Cδ@Glu735,Cδ@Glu687) of the active site. The distances were measured during the last 15 ns of the all-atom MD simulation. Standard deviations are given in parenthesis.

Besides the active site, the structure of the rest of the protein was also maintained except for helix α19 (residues 346-367), which became partially unfolded within the first 25 ns of the MD trajectory (Fig. 3.3). This unfolding was localized in the second part of α19 (residues 361-367) and caused a sudden increase in the **RMSD** of the C_α atoms; the coil-like conformation of this second part of α19 was then maintained until the end of the simulation.

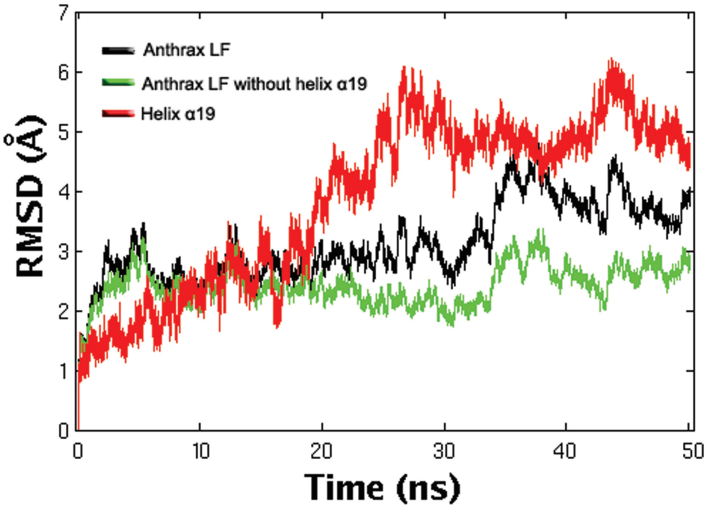


Figure 3.3: **RMSD** of **LF** backbone atoms during the 50 ns all-atom MD simulation. Note the increase in **RMSD** for helix α19 during the first 25 ns; after this time α19 gradually achieved structural stability.

To gain further insights into the instability of helix α19, we carried out calculations with two types of Coarse-Grained methods and with disorder prediction servers:
*(a)Hybrid **C**oarse-**G**rained/**M**olecular-**M**echanics (**CG/MM**) simulations.* Here, α19 and nearby atoms were treated at the **MM** level, while the rest of the system was treated at **CG** level (see section 3.2.3). Two 60 ns **CG/MM** simulations using AMBER parm98[83] and Gromos96 43a1[84] force fields for the **MM** part were performed. In both **CG/MM** simulations, the helix α19 partially unfolds within the first 25 ns (see Appendix A.5). The **RMSF** of the C_α atoms from the two 60 ns trajectories (Fig.3.4) were larger for α19 than for the rest of the protein. To ensure the reproducibility of these results, **CG/MM** simulations were performed using 15 equally spaced frames from the last 15 ns of the all-atom MD simulation (see Appendix A.5).
*(b)**N**ormal **M**ode **A**nalysis (**NMA**).* Here a **CG** elastic network of C_α atoms was built based on the energy minimized structure taken from the last frame of the all-atom MD trajectory. As obtained before for **CG/MM** MD simulations, the calculated **RMSF** of C_α atoms estimated from **NMA** for the α19 region were larger than those of the rest of the protein (Fig.3.4). To ensure the reproducibility of these results, **NMA** calculations were also performed using 15 equally spaced frames from the last 15 ns of the all-atom MD simulation (see Appendix A.6). The normalized B-values calculated from the all-atom **MD**, **NMA** and **CG/MM** simulations (Fig.3.4) agreed with those reported in the X-ray structure (PDB: 1JKY),[20] except for helix α19. Residues 361-367 from α19 assumed a coiled conformation in aqueous solution and its calculated normalized B-values were larger than those of the X-ray structure (Fig.3.4). This is consistent with the apparent difficulty to determine the solid state structure of α19. In fact, the structure of α19 has only been resolved in the X-ray structure used here as starting model for this part of the protein (PDB: 1JKY).[20]

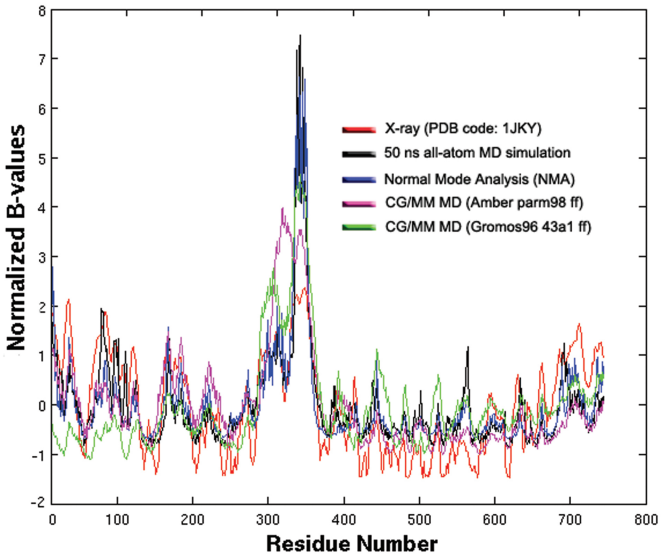


Figure 3.4: Calculated (RMSF) and experimental (X-ray; (PDB: 1JKY)[20]) normalized B-values for **LF**. For the all-atom MD and CG/MM simulations, only the last 15 ns were considered.

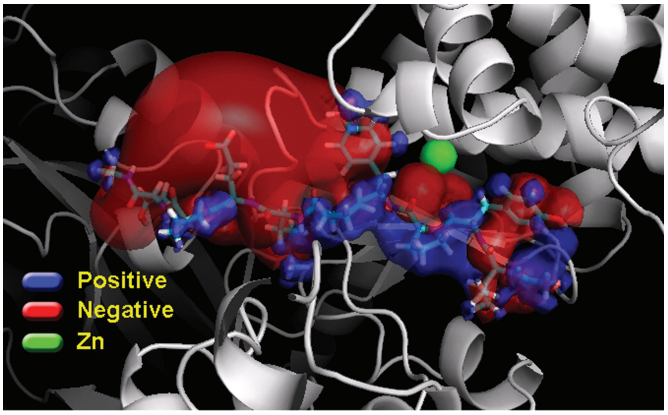
(c)Disorder prediction. The structural predictors PredictProtein,[137] PSIPRED,[159] SPRITZ,[160] and HNN,[161] indicate that at a part of α19 is a disordered region. Particularly, the last segment of this region (residues 361-367) is more likely to be a loop than an α-helix (see Appendix A.7). We describe the structural and electrostatic features of the active site as found in the complex. Only the last 15 ns of all-atom MD simulation were used for the analysis because α19 became stable only at this simulation time.

At the active site, Glu735 acts as a monodentate ligand of Zn and H-bonds to the solvent, interacting on average with ∼0.7 water molecules, as obtained by integrating the radial distribution function of Oε2@Glu687 *vs.* O@water (see Appendix A.3). The Zn is bonded with two histidine residues [b(Zn, Nτ@His686)=2.05 ± 0.05Å and b(Zn, Nτ@His690) = 2.15 ± 0.05Å].

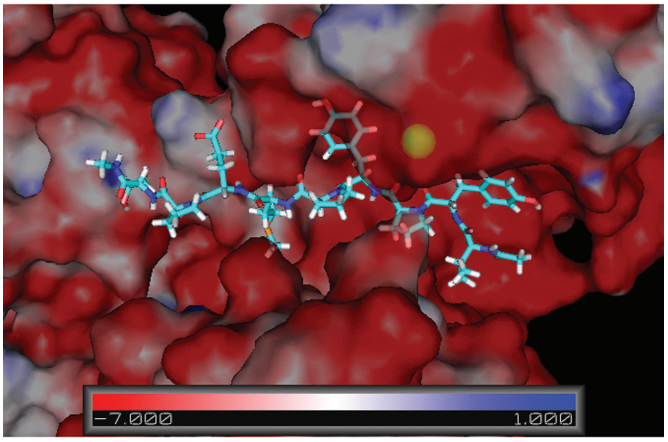
The catalytic water molecule, Wat1, H-bonds to Glu687, which is deprotonated; Glu687 is believed to act as a general base during catalysis (*i.e.,* accepting an hydrogen ion from Wat1).[20,178] Besides the H-bond to Wat1, Glu687 also interacts, on average, with ∼1 water molecule from the solvent. On the other hand Wat1 also H-bonds to Tyr-P2 (Table 3.2); this interaction could help to orient Wat1 in a proper position to perform the hydrolysis of the substrate.

Tyr728 forms an H-bond with the reactive carbonyl group of a proline residue in the substrate (O'@Pro-P1, Table 3.2); this interaction was intermittent and was established for ∼35% of the total simulation time. By forming this H-bond the substrate finds a proper orientation for the nucleophilic attack, therefore this interaction might play a role for the catalysis, providing a plausible, yet speculative explanation on why the Y728F mutant is not catalytically active.[155] Also Tyr728 is exposed towards the solvent and it H-bonds, on average, to ∼2.0 water molecules. O'@Pro-P1 is also water exposed, interacting, on average, with ∼1 water molecule.

The motions of Tyr728 in the enzyme and Tyr-P1' in the substrate are correlated (see Appendix A.1). The aromatic rings of these two residues lay at a distance smaller than 5Å for approximately 53% of the simulation time, being Tyr-P1' accommodated in the hydrophobic pocket (S1') of the enzyme which is partially formed by Tyr728 (Fig. 3.5, panel **(b)**). Tyr-P1' also forms an hydrogen bond with the carbonyl group in the Val675 backbone (Hη@Tyr-P1'O'@Val675=1.8 ± 0.1Å); this interaction was maintained for 80% of the simulation time.



a



b

Figure 3.5: Electrostatic isosurfaces on the Michaelis complex. **(a)** Optimized MAPKK2-like substrate. Note the complementarity of the bi-lobular negative patch (created by O'@Pro-P1 and O'@Tyr-P1') around the Zn ion, while a larger negative patch (formed by Glu-P4') is outside the negative groove of **LF**. **(b) LF** active site groove. Note the tyrosine on the left side of the Zn ion (Tyr-P1') in the hydrophobic pocket (S1') of the enzyme.

Poisson-Boltzmann calculations show that the substrate fits in the groove of the enzyme forming complementary electrostatic interactions (Fig.3.5). Note in particular, the interaction between a bi-lobular negative patch on the substrate (formed by O'@Pro-P1 and O'@Tyr-P1') and the positive Zn ion. During the **RMSD**-stabilized dynamics (last 15 ns of the all-atom MD simulation), the interactions between **LF** and the optimized substrate were similarly partitioned between electrostatic (-68±25 kcal/mol) and van der Waals (-82±6 kcal/mol) interactions. The substrate per residue interaction energies with **LF** were in good agreement with the experimental selectivity

results obtained by Turk et al.[21] (see Appendix A.8). Tyr-P1' was the energetically dominant residue for **LF** interaction, being electrostatic contributions (-23±4 kcal/mol) dominant over van der Waals interactions (-18±2 kcal/mol), which agrees with the strong H-bond formed between Tyr-P1' and the **LF** backbone described in the previous paragraph.
The electrostatic polarization of the active site may play a role in the enzymatic catalysis.[23] Here we compare the polarization of the bonds involved in substrate cleavage of **LF** with homologous reactive bonds of other two proteases: the aspartyl protease from **H**uman **I**mmunodeficiency **V**irus of type **1** (**HIV-1** PR),[23,179] which is believed to use a water molecule for the hydrolysis;[179] and the serine protease **Furin**,[23] which uses the hydroxylic group from a serine residue as nucleophile. As a measure of polarization we used the Bond Ionicity indexes[96] (see section 3.2.6).

	Furin	**HIV-1 PR**	**Anthrax LF**
$d(N_{pep}\text{-}BO^1)/d(N_{pep}\text{-}C_{pep})$	0.33(0.02)	0.31(0.03)	0.37(0.02)
$d(N_{pep}\text{-}BO^2)/d(N_{pep}\text{-}C_{pep})$	0.38(0.01)	0.38(0.02)	0.35(0.02)
$d(O_{pep}\text{-}BO^1{}_{lone})$	0.33(0.01)	0.34(0.01)	0.32(0.01)
$d(O_{pep}\text{-}BO^2{}_{lone})$	0.31(0.01)	0.32(0.01)	0.31(0.01)
$d(O_{pep}\text{-}BO^1{}_{C=O})/d(O_{pep}\text{-}C_{pep})$	0.37(0.01)	0.38(0.01)	0.39(0.01)
$d(O_{pep}\text{-}BO^2{}_{C=O})/d(O_{pep}\text{-}C_{pep})$	0.38(0.01)	0.39(0.01)	0.39(0.01)
$d(O_{Hyd}\text{-}BO^1{}_{lone})$	0.31(0.01)		
$d(O_{Hyd}\text{-}BO^2{}_{lone})$	0.31(0.01)		
$d(O_{Hyd}\text{-}BO_{O-H})/d(O_{Hyd}\text{-}H_{Hyd})$	0.50(0.01)		
$d(O_{Hyd}\text{-}BO_{C-O})/d(O_{Hyd}\text{-}C)$	0.39(0.01)		
$d(O_{Wat}\text{-}BO^1{}_{lone})$		0.32(0.01)	0.28(0.04)
$d(O_{Wat}\text{-}BO^2{}_{lone})$		0.33(0.01)	0.43(0.06)
$d(O_{Wat}\text{-}BO_{O-H1})/d(O_{Wat}\text{-}H_{Wat})$		0.47(0.02)	0.50(0.01)
$d(O_{Wat}\text{-}BO_{O-H2})/d(O_{Wat}\text{-}H_{Wat})$		0.53(0.02)	0.51(0.01)

Table 3.3: Comparison between **B**ond **I**onicity indexes (**BIs**)[96] in the reactive bonds of three prototypical proteases: **Furin**, **HIV**-1 PR and **LF**. The location of the **B**oys **O**rbitals (**BOs**) for **LF** is given in Appendix A.2. In the table, for an atom type "X", we refer as X_{pep} the atom from the peptidic bond that undergoes nucleophilic attack, X_{Wat} refers to an atom of the nucleophilic water molecule and X_{Hyd} refers to an atom from the nucleophilic hydroxylic group (*i.e.,* in SER side chain). The calculations were performed on 15 equally spaced frames taken from the last (equilibrated) 15 ns of the all-atom MD trajectory. Standard deviations are given in parenthesis.

Based on the values of the BI's of C_{pep}=O_{pep} and N_{pep}-C_{pep} bonds in the substrate (Table 3.3, see Appendix A.2 for details), we conclude that substrate's reactive bonds in **LF** are less polarized than those of **HIV-1** PR and **Furin**. In addition, the water (Wat1) O-H bonds were also less polarized in **LF** than in **HIV-1** PR: (i) in **LF**, BI_{O-H1}@Wat1(LF) = 0.50±0.01, BI_{O-H2}@Wat1(LF) = 0.51±0.01, (ii) in **HIV-1** PR , BI_{O-H1}@Wat1(HIV-1 PR) = 0.47±0.02 and BI_{O-H2}@Wat1(HIV-1 PR) = 0.53±0.02. However, the lone pairs on the oxygen atom of the

catalytic water (represented as d(O_{Wat}-$BO^{1[2]}_{lone}$) in Table 3.3) were more asymmetric in **LF** than in **HIV-1** PR simply due to the coordination of water to Zn. Our results also suggest that small polarization effects are induced by the **LF** scaffold, while the presence of zinc has a more critical role for the catalytic efficiency of **LF** (Tables 3.3 and 3.4).

	Charged	**No Charge in protein**	**No Charge in solvent**	**No Charge**
d(N_{pep}-BO^1)/d(N_{pep}-C_{pep})	0.37(0.02)	0.38(0.02)	0.37(0.02)	0.37(0.02)
d(N_{pep}-BO^2)/d(N_{pep}-C_{pep})	0.35(0.02)	0.36(0.02)	0.35(0.02)	0.36(0.02)
d(O_{pep}-$BO^1{}_{lone}$)	0.32(0.01)	0.31(0.01)	0.32(0.01)	0.32(0.01)
d(O_{pep}-$BO^2{}_{lone}$)	0.31(0.01)	0.31(0.01)	0.30(0.01)	0.30(0.01)
d(O_{pep}-$BO^1{}_{C=O}$)/d(O_{pep}-C_{pep})	0.39(0.01)	0.40(0.01)	0.40(0.01)	0.40(0.01)
d(O_{pep}-$BO^2{}_{C=O}$)/d(O_{pep}-C_{pep})	0.39(0.01)	0.39(0.03)	0.40(0.01)	0.40(0.01)
d(O_{Wat}-$BO^1{}_{lone}$)	0.28(0.04)	0.28(0.05)	0.28(0.02)	0.27(0.02)
d(O_{Wat}-$BO^2{}_{lone}$)	0.43(0.06)	0.47(0.05)	0.43(0.04)	0.45(0.04)
d(O_{Wat}-BO_{O-H1})/d(O_{Wat}-H_{Wat})	0.50(0.01)	0.49(0.01)	0.50(0.01)	0.49(0.01)
d(O_{Wat}-BO_{O-H2})/d(O_{Wat}-H_{Wat})	0.51(0.01)	0.51(0.01)	0.51(0.01)	0.51(0.01)

Table 3.4: Comparison between **B**ond **I**onicity indexes (**BIs**)[96] in the reactive bonds of **LF** under different electrostatic conditions showing the effect of the protein and solvent electrostatic properties on the polarization of the active site. Four conditions were evaluated: (i)"**Charged**", calculation assigning charges to all atoms in the system; (ii)"**No charge in protein**", calculation assigning charges equal to zero in all atoms of the protein; (iii)"**No charge in solvent**", calculation assigning charges equal to zero to the atoms of all the water molecules; (iv)"**No charge**", calculation assigning charges equal to zero to all atoms in the system. Atomic charges were assigned using the AMBER parm98 force field.[83] The calculations were performed on 15 equally spaced frames taken from the last (equilibrated) 15 ns of the all-atom MD trajectory. Standard deviations are given in parenthesis.

The large scale motions may play a role for substrate recognition and/or for enzymatic catalysis.[23,162,179,180] We explore this issue by performing **PCA**[112] on the last (equilibrated) 15 ns of the all-atom MD trajectory. For the sake of simplicity, we included only the first three **P**rincipal **C**omponents (**PCs**) that accounts for ∼54% of the overall motion (Fig.3.6).

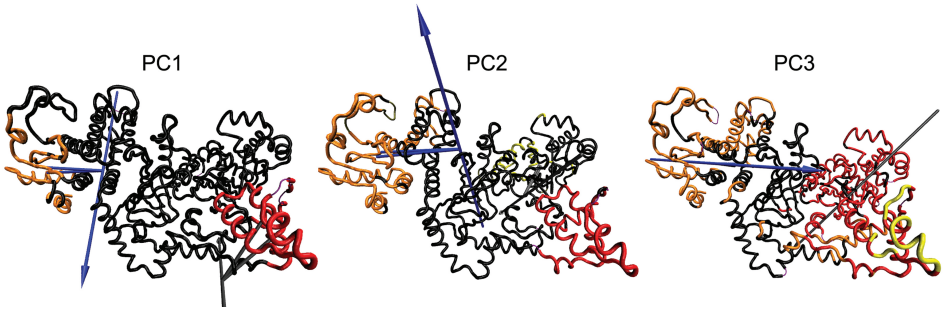


Figure 3.6: The first three **P**rincipal **C**omponents (PCs) as calculated with Dynatraj[112] from the last 15 ns of the all-atom MD simulation. Rigid domains, as calculated with the procedure of Wriggers and Schulten,[113] are depicted in different colors; details of these calculations can be found in Appendix A.6. The arrows indicate the effective rotation axis between the two adjacent rigid domains. The direction of the arrow represents the sense (*e.g.,* clockwise) used to find the rotations axis between two rigid domains and it is ultimately dependent of the choice of the "reference" rigid domain (the one kept steady during the calculation of the rotational angle). Note that PC3 was able to capture a movement in the α19 region (PC3, yellow), pointing to the higher flexibility of this region (see Fig.3.4).

In the first three **PCs**, we identified the large motions involving domain **III** (*i.e.,* the domain that includes helix α19, see Fig.3.1). In particular, **PC3** was able to capture a relatively independent movement of α19 with respect to domain **III** (Fig.3.6). The observed large scale motions of **LF** did not involve the active site (see Table3.2), similarly to what was found in another protease studied with a similar computational setup, the serine protease **Furin**.[23,179]

3.3.2.2 MD of LF Michaelis complex in D protonation state

This protonation state turned out to be already unstable in the first 0.5 ns of the simulation: the H-bond network was disrupted because of a rotation of Glu687 about the C_γ-C_δ bond, allowing the entrance of additional water and the departure of the substrate from the active site. This complex was therefore discarded.

3.4 Conclusions

We characterize the structural properties of the Michaelis complex formed by anthrax **LF** and an optimized substrate using several computational tools. Our findings can be summarized as follows:

(i) The second shell ligands affect the energetics of Zn active site as has been observed in other Zn-based enzymes.[23,156,157] Our calculations confirm that second shell ligands have an influence on the protonation state of the active site; stabilizing (by ∼3kcal/mol) the hydrogen bond network around it.

(ii) The nucleophilic agent is a Zn-bound water molecule (not an OH group) which forms a hydrogen bond with Glu687.

(iii) Per residue substrate selectivity of **LF** is in good agreement with interaction energies calculated from the all-atom MD simulation (model **C**). Specifically, Tyr-P1' had the greatest interaction energies with **LF**, which were dominated by electrostatic contributions; featured mainly by a strong H-bond established inside the hydrophobic pocket of **LF** between Hη@Tyr-P1' and the backbone of the enzyme (O'@Val675). These results help to explain the selectivity of **LF** for substrates with tyrosines in the vicinity of the reactive proline.[21]

(iv) The **LF** scaffold induces small polarization effects on the active site. A larger polarization was observed for the lone pairs of the nucleophilic agent Wat1 (see (ii)); the obvious cause for this effect is the Wat1 coordination to Zn.

(v) Large-scale motions do not affect the structure of the **LF** active site; it is therefore unlikely that these motions could play a mechanical role during the first step of the catalysis.

(vi) Part of helix α19 (residues 361-367) assumes a coil-like conformation in aqueous solution.

Chapter 4

ATX1-CCC2 COPPER TRANSPORT SYSTEM

4.1 Introduction

Intracellular copper ions ($Cu^{+/2+}$) not bound to proteins are toxic.[181] They may react with oxygen reactive species to yield hydroxyl radicals which can damage proteins, lipids and nucleic acids.[182] Fortunately, intracellular copper ion concentrations are kept close to zero due to the presence of a highly Cu–chelating environment in the cytosol.[183] Copper transport inside the cell is assisted by several proteins.[40] One of the best characterized pathways is the copper secretory pathway (see Chapter 1, Fig. 1.3). In this pathway, Cu^{+} ions are transferred from chaperon Atx1 to the membrane associated Ccc2 following the scheme in Fig. 4.1. The direction of the reaction is only slightly favorable to the formation of Ccc2-Cu ($\Delta G = -0.2$ kcal/mol).[38]
In this chapter, we characterize some aspects of the copper transport between Atx1 and Ccc2 summarized in Fig. 4.1, by molecular simulations. We focus on the non-covalent association of these proteins schematized in Fig. 4.2.

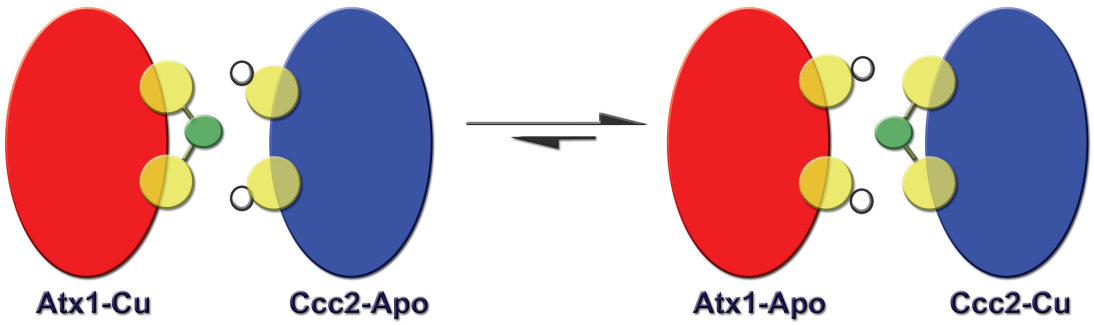


Figure 4.1: Scheme of the copper transfer reactions between Atx1 (red) and Ccc2 (blue). In this figure, are also represented the conserved cysteines directly involved in Cu^{+} binding (yellow), Cu^{+} (green), and hydrogens on the cysteines (white).

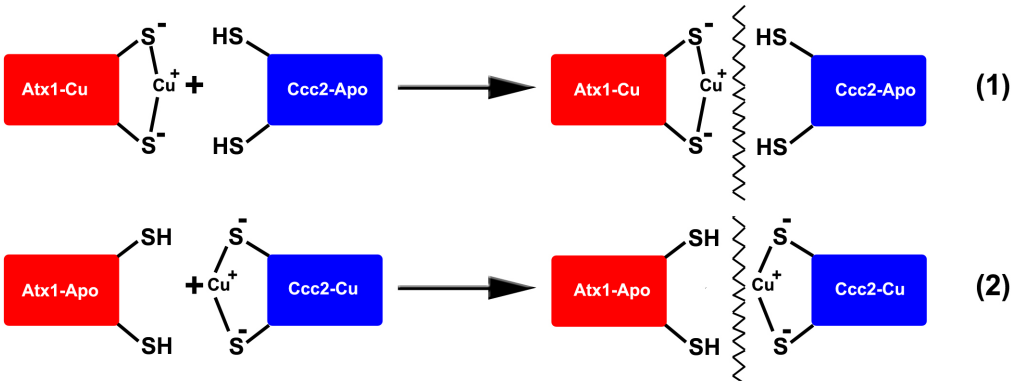


Figure 4.2: Possible types of encounters between Atx1 and Ccc2 species that lead to Cu^{+} transport.[38,39] The break line indicates complex formation only through non-covalent interactions.

We provide qualitative insights into such reactions by investigating: (i) protein-protein non-covalent association through Brownian Dynamics simulations (section 4.2.1), (ii) protein-protein dissociation through Random Expulsion Molecular Dynamics (REMD)(section 4.2.2), (iii) the correlation of motions of Atx1-[Apo,Cu], Ccc2-[Apo,Cu], and Atx1-Cu-Ccc2 (section 4.2.3), and (iv) Poisson-Boltzmann electrostatic calculations of Atx1-[Apo,Cu], Ccc2-[Apo,Cu], and Atx1-Cu-Ccc2 (section 4.2.4).

4.2 Results

4.2.1 Kinetics of protein-protein non-covalent association: Brownian Dynamics (BD) simulations

We estimate here the kinetics associated with the encounters of Atx1 and Ccc2 in their Apo and Cu-bonded states by **B**rownian **D**ynamics (**BD**). To do this, we make the reasonable assumption that the binding processes leading to Cu^+ transport (Fig. 4.2) occurs first by the encounter Atx1 and Ccc2 forming a non-covalent complex (*e.g.,* the final product of reaction (1)) in which the cysteines in one partner (Ccc2) are protonated and the cysteines in the other partner (Atx1) are deprotonated and establish bonds with Cu^+. Subsequently, copper and proton transfers occur until Cu^+ passes completely from Atx1 to Ccc2. In a similar way, Cu^+ transfer from Ccc2 to Atx1 should start with the binding process represented in reaction (2) (Fig.4.2). Comparison is made with events which cannot lead to Cu^+ transport, namely, the encounters between the Apo forms and between the Cu-bonded forms (reactions (3) and (4) in Fig.4.3).

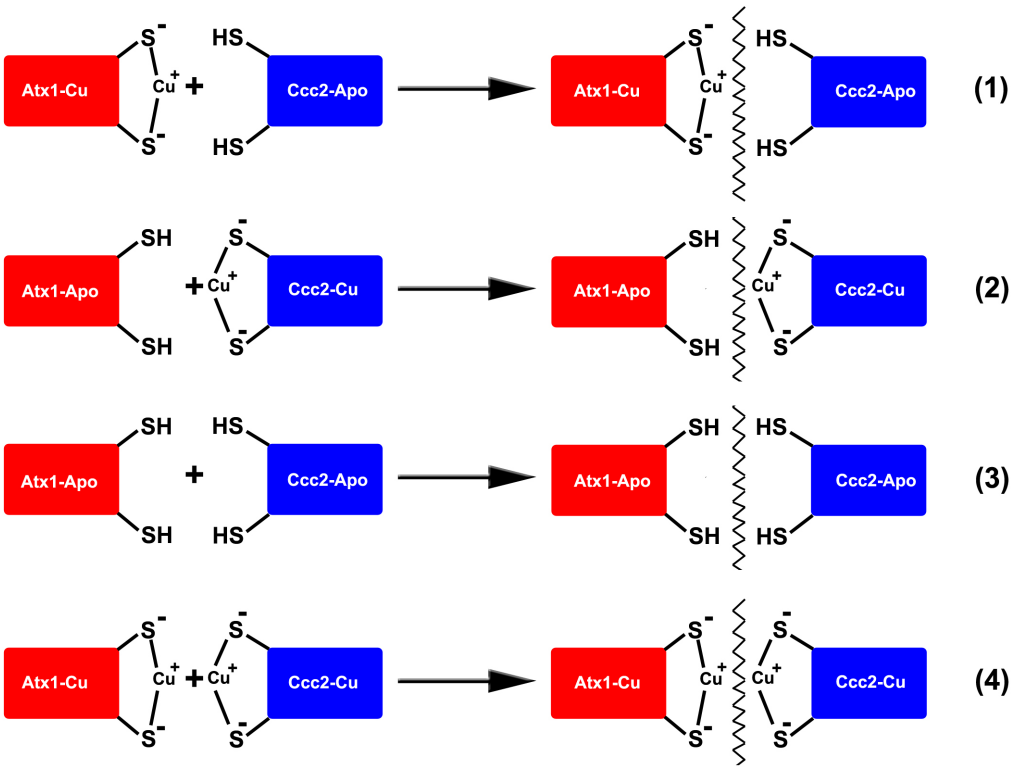


Figure 4.3: Possible types of encounters between Atx1 and Ccc2. Besides the types of encounters represented in Fig.4.2, in this section, we also investigated the associations between the Apo forms (3) and the copper bonded forms (4). The break line indicates complex formation only through non-covalent interactions.

The geometrical criterion used here to define the bound state for the complexes in Fig. 4.3 is

a cutoff distance between atoms belonging to surface residues of Atx1 and Ccc2. The selected atoms are: (i) the sulfur atoms of Cys15@Atx1 and Cys13@Ccc2, (ii) the ammonium nitrogen of Lys59@Atx1, and the carboxylic carbon of Asp61@Ccc2 (Fig. 4.4). The first pair was selected because these conserved Cysteines[36] on each protein, participate directly in the copper transfer reactions.[39] The second pair consists of two key residues for protein-protein association, as identified by the FastContact algorithm[138] (see Appendix B.2 for details). BD simulations were performed until both distances were lower than 3.5, 4.0, and 4.5 Å (Fig. 4.5).

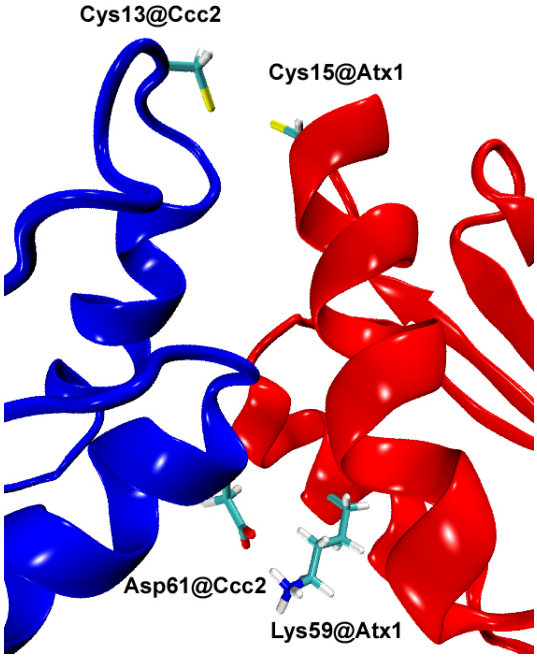


Figure 4.4: Schematic representation of the amino acids at the Ccc2-Atx1 interface used to define the geometrical distance criteria for Brownian dynamics simulations of the encounters between Atx1 and Ccc2 each in the Apo and Cu-bonded states. The bound states were defined in base of cutoff distances (see Text) between the sulfur atoms of Cys15@Atx1 and Cys13@Ccc2 as well as that between the ammonium nitrogen at Lys59@Atx1 and the carboxylic carbon at Asp61@Ccc2.

Our calculated kinetic constants provide information about the relative association kinetics of the processes represented in Fig. 4.3. For all three choices of the cutoff used here, it turns out that the slowest process is related to the Atx1-Cu/Ccc2-Cu association (Fig. 4.5). This may be caused, at least in part, by the Coulomb repulsion between the two sites containing copper. The productive processes have intermediate rates, being slightly faster the association of Atx1-Cu with Ccc2-Apo; this is in agreement with a favored transfer of Cu^+ from Atx1 to Ccc2. The fastest process was the association between Apo forms. Once the non-covalent complexes portrayed in Fig.4.2 are formed, several chemical reactions are needed for a complete Cu^+ transfer between these proteins (Fig.4.1) Further investigations with methods like QM/MM[88] are needed to address this issue. It should also be noted that experimental data for these reactions is unfortunately lacking so far.

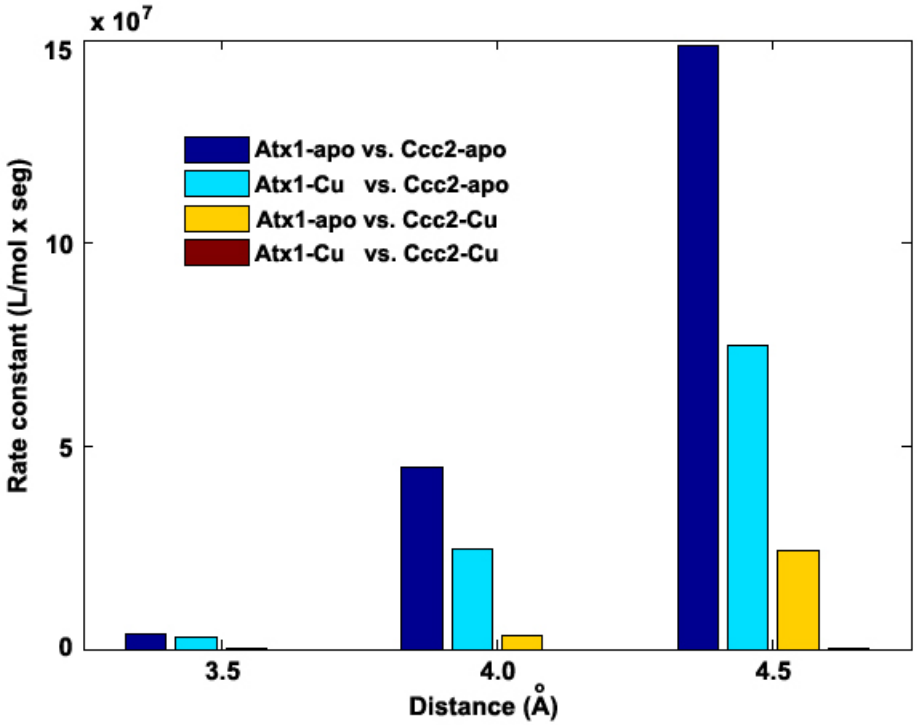


Figure 4.5: Association rate constants for the encounters between Atx1 and Ccc2 in their Apo and Cu-bonded states. Rate constants were calculated for three different distances between the pairs of amino acids chosen to define the bound state (Fig.4.4) of the possible non-covalent complexes (Fig. 4.3). The calculated rate constants are comparable to those reported in other BD studies.[184]

4.2.2 Protein-protein dissociation: Random Expulsion Molecular Dynamics (REMD)

We estimate here quantitatively the dissociation kinetics of the non-covalent complexes Atx1-Cu/Ccc2-Apo and Atx1-Apo/Ccc2-Cu by **R**andom **E**xpulsion **M**olecular **D**ynamics (**REMD**). A similar approach has been used to study qualitatively the kinetics of other biomolecular unbinding processes.[185] The complexes investigated in our study were obtained from the only available NMR structure of a reaction intermediate.[39]
Kinetics is here estimated in terms of a probability density of calculated expulsion times (ET's) for the separation of the complexes, as done in ref.[185] The ET is defined as the time employed by REMD to separate the centers of mass of Atx1 and Ccc2 up to a certain cutoff (See sections 2.5.2 and 4.4.3 for details).

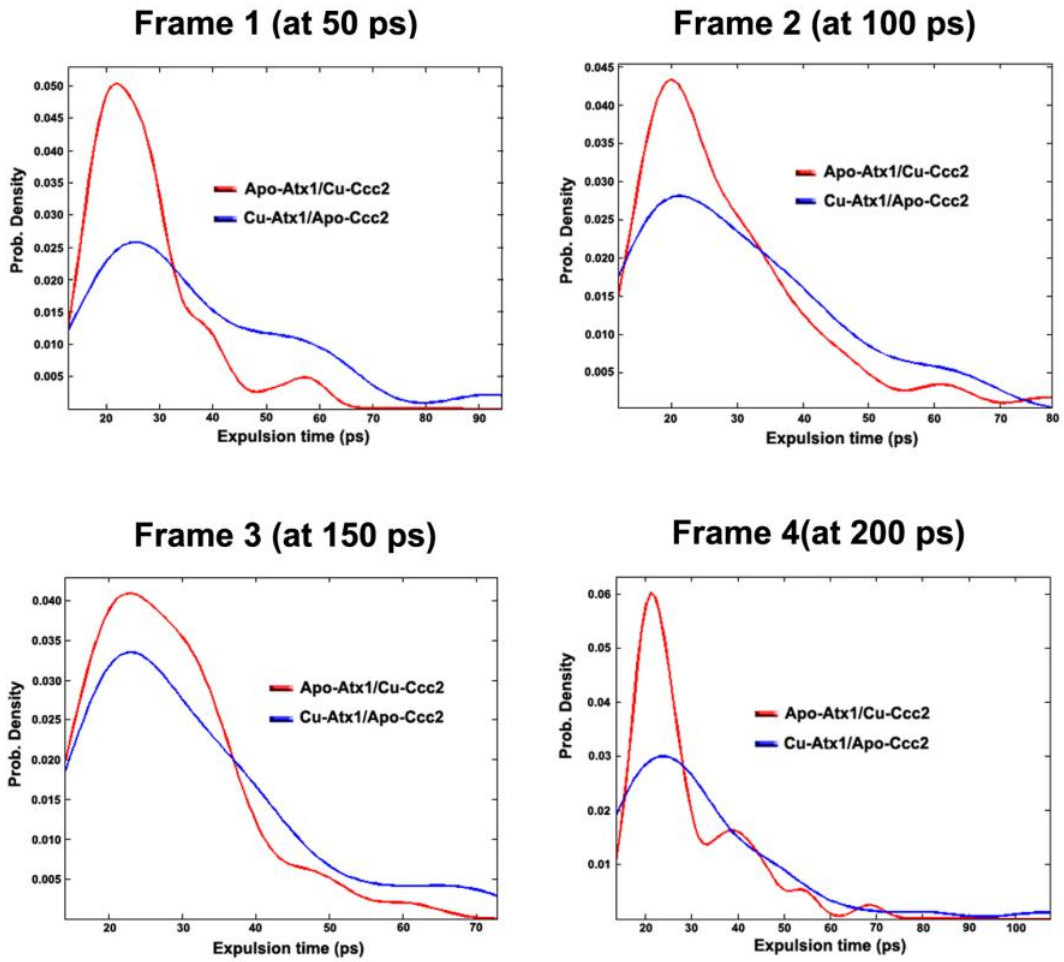


Figure 4.6: Calculated probability density distribution of the REMD expulsion times for the separation of Atx1-Cu/Ccc2-Apo and Atx1-Apo/Ccc2-Cu.

The calculated probability densities show that the complex Atx1-Cu/Ccc2-Apo separates more slowly than Atx1-Apo/Ccc2-Cu (Fig.4.6). These results are consistent with the Brownian Dynamics simulations in the previous section. Note that, also in this case, these calculations cannot account for the breaking and formation of chemical bonds which occur during Cu^+ transport.

4.2.3 Correlation of motions of Atx1-[Apo,Cu], Ccc2-[Apo,Cu] and Atx1-Cu-Ccc2

In the next two paragraphs, we show the correlations of Atx1 and Ccc2 proteins during Cu^+ transfer (see reactions in Fig. 4.1). Particularly, we investigated the conformational fluctuations of the Atx1-[Apo,Cu] and Ccc2-[Apo,Cu] species and the reaction intermediate Atx1-Cu-Ccc2. For these systems, NMR structures are available.[36,39,186]

The C_α **C**orrelation **M**atrices (**CM's**)[124] of Atx1-[Apo,Cu] and Ccc2-[Apo,Cu] as well as their respective dimeric states in Atx1-Cu-Ccc2 were calculated from their **MD** simulations in explicit water solution (Fig. 4.7, section 2.4.2.3 and 4.4.4 for details). The CM's describe the correlation motions between C_α atoms during the dynamics. A value of +1 for a pair of C_α atoms means that the motions of the two atoms are fully correlated. A value of 0 indicates no correlation.
Fig. 4.7 suggests that the conformational fluctuations of the Apo species do not change dramatically to arrive at the dimeric state, as the CM's of the single components and that of the complex are not too dissimilar. On the other hand, the conformational fluctuations of the Cu-bonded species differ from the dimeric state. At a speculative level, these differences might affect the efficiency of the association process.

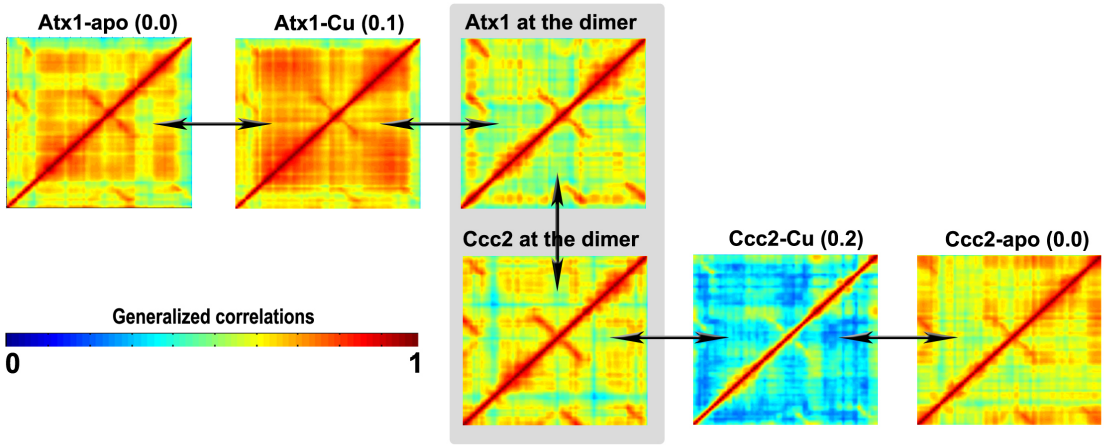


Figure 4.7: Correlation matrices (CM's) of Atx1-[Apo,Cu], Ccc2-[Apo,Cu] and Atx1-Cu-Ccc2. The overall difference between the mean values of the CM's of the single proteins and those of the complex are reported. In all circumstances, the standard deviations of the reported values were equal to 0.1.

4.2.4 Poisson-Boltzmann electrostatic calculations of Atx1-[Apo,Cu] and Ccc2-[Apo,Cu]

We calculate the electrostatic potential using the **P**oisson-**B**oltzmann **E**quations (**PBE**) for representative **MD** structures of Atx1-[Apo,Cu] and Ccc2-[Apo,Cu] (Fig. 4.8, see section 2.4.2.4 for details).
The electrostatic potential of Ccc2 changes dramatically upon Cu^+ uptake whilst that of Atx1 is much less affected. Interestingly, the interacting surfaces of Atx1-Cu and Ccc2-Apo become complementary upon Cu^+ uptake and favor the association of the two proteins; specifically, Atx1-Cu is positively charged and Ccc2-Apo is negatively charged. Instead, the interacting surfaces of Atx1-Apo and Ccc2-Cu are not electrostatically complementary.
These results suggest that electrostatic interactions stabilize the Atx1-Cu/Ccc2-Apo complex more than Atx1-Apo/Ccc2-Cu. These findings are consistent with the experimental fact that

copper transfer occurs preferentially from Atx1 to Ccc2.[38]

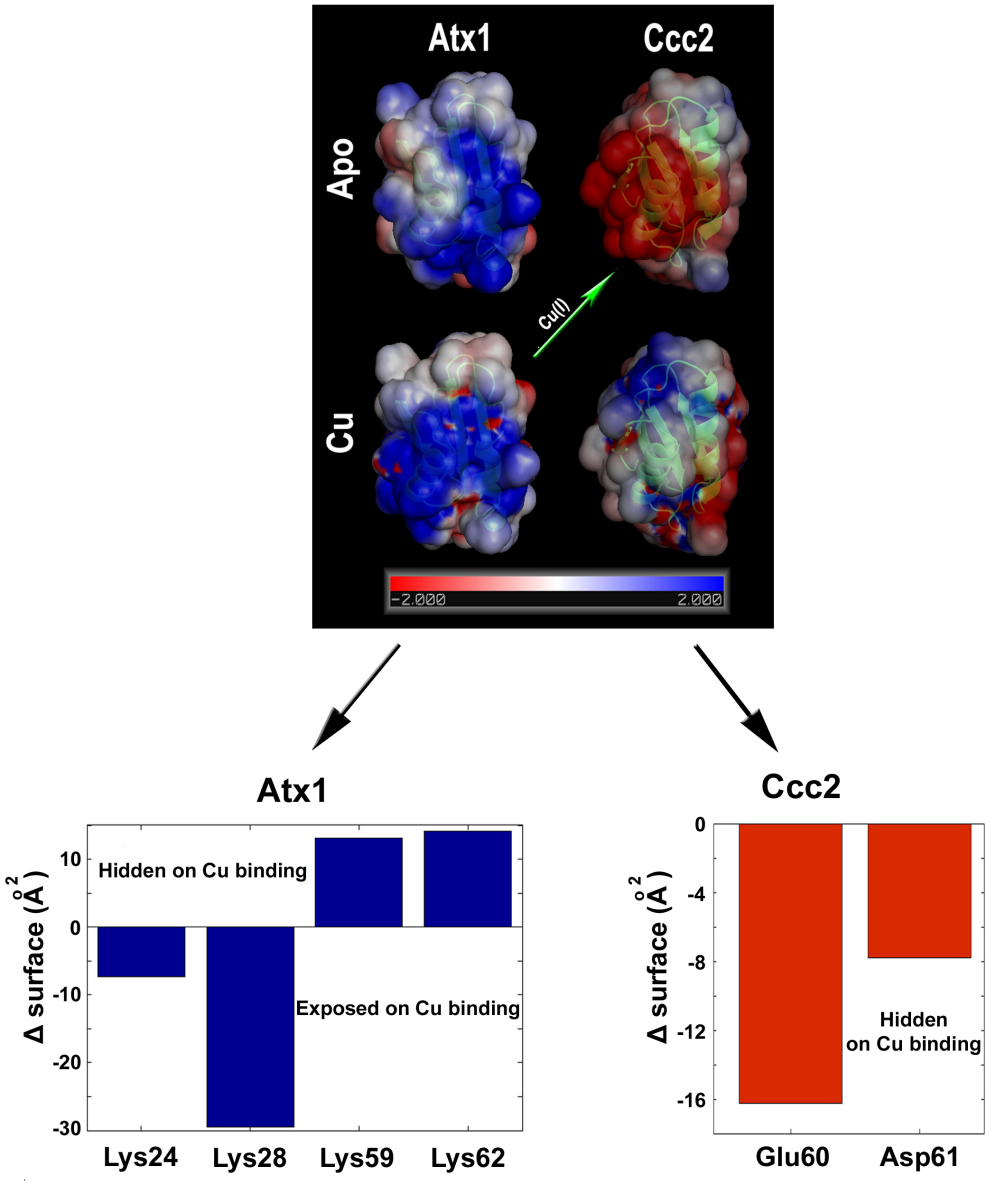


Figure 4.8: *Top:* Electrostatic potential (kT/e) of the contact surface between Atx1 and Ccc2 in their Apo and Cu-bonded states. *Bottom:* Change in solvent accessible surface area of hot spot residues in Atx1 and Ccc2 on copper binding. The fact that the electrostatic potential of Atx1 does not change largely upon copper binding might be caused, at least in part, by a compensatory mechanism: some positive residues hide and others get exposed on copper binding; while on Ccc2, the change to a more positive potential is due to the hiding of negative residues from the protein surface.

Next, we identify the residues which may play a key role for Atx1-Cu/Ccc2-Apo binding ("hot spots"). We used the FastContact algorithm[138] which has proven to have high accuracy in predicting hot spot residues for docking complexes during CAPRI experiments.[187,188] Details of our FastContact calculations can be found in Appendix B.2. The hot spots turn out to be the following, [Lys24,Lys28,Lys59,Lys62]@Atx1 and [Glu60,Asp61,Asp65]@Ccc2.
All of these residues are charged, therefore they are expected to play a role for the electrostatic potential. However, such role may be different in the Apo and Cu-bonded forms as their conformations may vary in these two states. This is the case of [Glu60,Asp61]@Ccc2, which are solvent exposed in the Apo form and fully buried upon Cu^+ binding (Fig. 4.8). Therefore, these residues are likely to be important for the observed change in electrostatic potential.[1] Instead, Asp65@Ccc2 does not change significantly its solvent exposure (less than 5% of the total change experienced by all of them, [Glu60,Asp61,Asp65]@Ccc2). These results suggest that the electrostatic surface of Ccc2 becomes more positive upon Cu^+ binding due, at least in part, to the decreased exposure of Glu60 and Asp61.
Residues [Lys24,Lys28]@Atx1 pass from an exposed to a buried conformation upon Cu^+ uptake, whilst residues [Lys59,Lys62]@Atx1 pass from a buried to an exposed conformation (Fig. 4.8). Thus, one possible reason, among others, of the much less pronounced electrostatic potential change upon Cu^+ uptake of Atx1 relative to Ccc2 might be some compensation effect of Atx1 hot spot residues.

4.3 Conclusions

Copper transport through the secretory pathway (Fig. 1.3) involves the transfer of Cu^+ from the chaperone Atx1 to Ccc2, a soluble domain of a Golgi associated Cu-pump.[189]
Based on our calculations, we suggest that:

(i) The non-covalent association of Atx1 and Ccc2 is faster when Cu^+ binds to Atx1 (Fig. 4.5).

(ii) The non-covalent dissociation of Atx1 and Ccc2 is faster when Cu^+ binds to Ccc2 (Fig. 4.6).

(iii) Changes in solvent-exposure of key residues at the proteins' surfaces play an important role for the experimentally observed larger thermodynamical stability of Ccc2-Cu relative to Atx1-Cu (Fig. 4.8).

QM/MM calculations could now be performed to complement the investigation on the kinetics of the process. They could be used to study the bond/forming and bond/breaking processes associated with copper transfer.

[1] Electrostatic potential calculations of mutants in the 60 and 61 positions could help further address this issue.

4.4 Methods

4.4.1 All-atom Molecular Dynamics

Systems: Seven systems were considered: Atx1-[Apo,Cu], Ccc2-[Apo,Cu], the reaction intermediate Atx1-Cu-Ccc2, and the non-covalent complexes Atx1-Cu/Ccc2-Apo and Atx1-Apo/Ccc2-Cu. NMR structures for the first 5 systems are available,[36,39,186] the non-covalent complexes were built from the structure of the reaction intermediate Atx1-Cu-Ccc2. Ion concentration for all systems was set to 0.1M, matching NMR experimental conditions.[36,39,186] Monomeric systems contained 15200 atoms and dimeric systems contained 37200 atoms.
Force fields: We used two force fields in our calculations:

(I) for Atx1-[Apo,Cu], Ccc2-[Apo,Cu], and the reaction intermediate Atx1-Cu-Apo, we adopted the Charmm32b1 force field with the CMAP procedure.[190] In the copper binding site, we used a previously reported bonded parametrization.[191]

(II) for non-covalent complexes Atx1-Cu/Ccc2-Apo and Atx1-Apo/Ccc2-Cu, we adopted the Amber94 force field.[192] In the copper binding site, we used a previously reported bonded parametrization.[193]

Conditions: The electrostatic interactions were evaluated using the Particle Mesh Ewald (PME) method.[169,194,195] A cutoff of 10Å was used for the van der Waals interactions and the real-space part of the electrostatics. The bonds involving hydrogen atoms were kept fixed using SHAKE algorithm.[170] A time step of 1 fs was used. Constant temperature (298K) and pressure (1 atm) simulations were achieved by coupling the systems to a Berendsen thermostat and barostat.[171]
Protocols: For each force field above mentioned, we used a different simulation protocol:

(I) *Charmm32b1 force field.* Initial configurations of the systems were relaxed by short minimization runs of 2000 steps using conjugate gradient energy minimization. 100 ps of MD at constant volume were then performed, during which the system was gradually heated to 298K. After equilibration, production runs of 30 ns at constant temperature and pressure were obtained. The NAMD software (v2.6) was used.[172] Representative structures from these simulations were used for Brownian Dynamics (BD) simulations and Poisson-Boltzmann calculations; moreover, from the 30 ns trajectories, generalized correlations were calculated. This protocol was validated by comparing calculated and experimental values of the NMR order parameters (Appendix B.1). The comparison was fairly satisfactory.

(II) *Amber94 force field.* Initial configurations of the non-covalent complexes were relaxed during 100 steps using steepest descent minimization. After that we performed: 20 ps MD simulation at constant pressure (1 atm) with restrains on the heavy atoms, 50 ps unrestrained MD simulation at constant pressure (1 atm), 200 ps MD equilibration at constant volume and 200 ps production runs. From the production runs, we took 4 equally spaced frames as starting structures for Random Expulsion Molecular Dynamics (REMD). Backbone RMSD of the initial structures for REMD were within 1.9Å of the representative structure of the reaction intermediate, Atx1-Cu-Ccc2.

4.4.2 Brownian Dynamics (BD)

These calculations were based on representative MD structures of Atx1-[Apo,Cu] and Ccc2-[Apo,Cu]. Solvent ionic strength was set to 0.1 M and temperature was set to 298K in all systems, matching experimental conditions.[36,39,186] Atomic charges were assigned using the Tanford-Kirkwood method.[196,197] Dielectric constants of the solvent and the solutes were 78 and 4 respectively. The electrostatic potential was determined by numerically solving the Poisson-Boltzmann Equations (PBE) using the Warwicker-Watson method;[198] the solutions of PBE were mapped on two kinds of grids: an inner grid of 0.5Å resolution for anisotropic interactions and an outer grid of 1.0Å resolution for isotropic calculations. For the isotropic/anisotropic region definition (see Fig. 2.6) we followed the following criteria: (i) if the proteins were closer than a distance b=100Å, then the interactions between the proteins were considered anisotropic; (ii) if the distance between the proteins arrived to q=200Å, then the BD trajectory was stopped; (iii) if the distance r between proteins fulfills the condition $b < r < q$, then the interactions between the proteins were considered isotropic. Parameters b and q were selected to be larger than those reported in the literature for similar studies.[184]
In all BD simulations, random orientation initial conditions were selected for both proteins, Atx1 and Ccc2; they were also allowed to rotate freely around their centers of mass. In the BD simulation of a diffusing pair of partners (*e.g.,* Atx1-Cu *vs.* Ccc2-Apo), a total of 10^6 BD trajectories were calculated using the Ermak-McCammon algorithm.[127] BD simulations were run until the geometrical distance criteria for the association (see Fig. 4.4) was fulfilled. For each distance criterion we calculated the corresponding rate constants of association. The algorithm for BD simulations used in this thesis is that of the software MacroDox.[126]

4.4.3 Random Expulsion Molecular Dynamics (REMD)

Initial structures for REMD of the non-covalent complexes Atx1-Cu/Ccc2-Apo and Atx1-Apo/Ccc2-Cu were taken from short MD trajectories at 50, 100, 150, and 200 ps (see section 4.4.1). From each initial structure, 50 REMD simulations were performed for a total of 400 REMD trajectories. The seed for the random number generator used to choose the direction of the force unit vector to be applied to one of the partners in the complex was varied for every simulation (see section 2.5.2); the force constant was $k = 100$ kJ mol^{-1} nm^{-1}. Forces were applied for $N = 10$ consecutive MD steps; if after this time the separation of the proteins was lower than $r_{\min}$=0.01Å, a new random vector was generated. To create a minimal deformation of the protein's 3D structures under REMD, the values of k and $r_{\min}$ were selected to be slightly lower than what is usual in this type of simulations.[133] The expulsion time for a simulation was calculated as the time employed to separate the centers of mass of the Atx1 and Ccc2 up to $r_{\max}$=40Å. The probability density distributions of the expulsion times were estimated using a kernel non-parametric function of the Gaussian type. REMD was implemented by their authors as a patch to Amber8.

4.4.4 Generalized Correlations

C_α full non-linear generalized **C**orrelation **M**atrices (**CM's**)[124] of Atx1-[Apo,Cu], Ccc2-[Apo,Cu] and Atx1-Cu-Ccc2 were calculated from MD trajectories. Prior to the calculations, overall translations and rotations of the proteins inside the box were removed by least square fitting to a reference structure. Raw CM's were imported into the Matlab software for further processing.

4.4.5 Poisson-Boltzmann calculations

Atomic charges for the representative structures of Atx1-[Apo,Cu] and Ccc2-[Apo,Cu] were calculated using the program PDB2PQR.[176] For the core calculations, a grid with a density of 20 points/Å^2 was built with the utility pzyse.py; temperature of the system was set to 298K and ion concentration was set to 0.1M, matching experimental conditions. Dielectric constants for the solvent and the solutes were 78 and 4 respectively. To define the solvent accessible surface, a solvent molecule of radius 1.4Å (water) was selected. Electrostatic surfaces were smoothed with cubic spline algorithm. Full nonlinear Poisson-Boltzmann Equations (PBE) were solved and corresponding raw data files were saved for further analysis. All core Poisson-Boltzmann calculations were performed with the APBS solver.[175] Input data with initial conditions and final visualization of the results were done with an APBS Pymol interface.[177]

Chapter 5

CONCLUDING REMARKS

Metal coordination in proteins differs dramatically depending on the biological function. In zinc and copper enzymes, Zn and Cu ions steadily bound to proteins are predominantly tetra-coordinated by His and Asp residues.[4,6] In most of these proteins, the metal has structural or catalytic functions.[4] On the other hand, Zn and Cu ions transiently bound to proteins are predominantly bi/tri-coordinated by Cys residues.[5,6] Prototype proteins in the latter group are those participating in intracellular Zn and Cu trafficking.[5]

Here, we have applied computational biology methods to two systems of interest for molecular medicine, a Zn-enzyme secreted by *B. anthracis*[10,139] and a pair of Cu-proteins of pivotal importance for copper trafficking in the cell.[29,40]

The Zn-enzyme studied in this thesis is the anthrax **L**ethal **F**actor (**LF**), a protease playing a key role in anthrax disease.[147] In LF, the Zn^{2+} ion is tetra-coordinated by two histidines, a glutamate, and a water molecule.[20] We investigated the structure of its Michaelis complex with an optimized MAPKK-like substrate[21] using several computational methods including Density Functional Theory (DFT), Molecular Dynamics (MD), and Coarse-Grained techniques. Our calculations suggest that: (i) the presence of second-shell ligands is crucial for tuning the structure, energetics, and protonation state of the metal binding site, as found in other Zn-based enzymes;[156] (ii) the nucleophilic agent is a Zn-bound water molecule; (iii) substrate binding to the active site groove is mainly stabilized by electrostatic interactions; (iv) the bonds most likely involved in the substrate hydrolysis are only mildly polarized by the protein scaffold; and (v) part of helix α19, which is present in one solid state structure of LF (PDB: 1JKY), assumes a coiled conformation.

We hope that these findings may help develop novel peptidomimetic inhibitors with therapeutical properties. Based on our calculations, one could now investigate the enzymatic reaction by QM/MM methods.[199]

The Cu-based system is involved in a key pathway of copper trafficking.[41] In this pathway, Cu^{+} ions hop from the chaperone protein Atx1 to the *trans*-Golgi associated Ccc2 in *S. cerevisiae*[29]. Both Atx1 and Ccc2 bind Cu^{+} transiently using two conserved Cys residues.[36] The Atx1-Ccc2 copper transport system might help understand the Cu^{+} transport in humans.[42] It may also shed light on aberrant processes associated with it, such as Menkes and Wilson disease-linked mutations.[200]

Using Brownian Dynamics (BD), Random Expulsion Molecular Dynamics (REMD), and Poisson-

Boltzmann electrostatics calculations, we investigated qualitative aspects of association/dissociation kinetics between the possible non-covalent species of these proteins. Our calculations point to a preferential transfer of Cu^{+} from Atx1 to Ccc2 as experimentally observed.[38] However, the picture emerging from these studies is necessarily incomplete due to the limitations of studying only non-covalent binding/unbinding processes, neglecting the chemical reactions involved. In addition, these calculations showed that the residues [Lys24,Lys28,Lys59,Lys62]@Atx1 and [Glu60,Asp61,Asp65]@Ccc2 play a key role for the formation of the Atx1-Cu-Ccc2 complex, which is present *in vitro*.[39] Also in this system, QM/MM calculations could provide further insights by describing the covalent bond forming and breaking associated with copper transport.

Appendix A

Anthrax LF: additional calculations

A.1 Correlated Motions

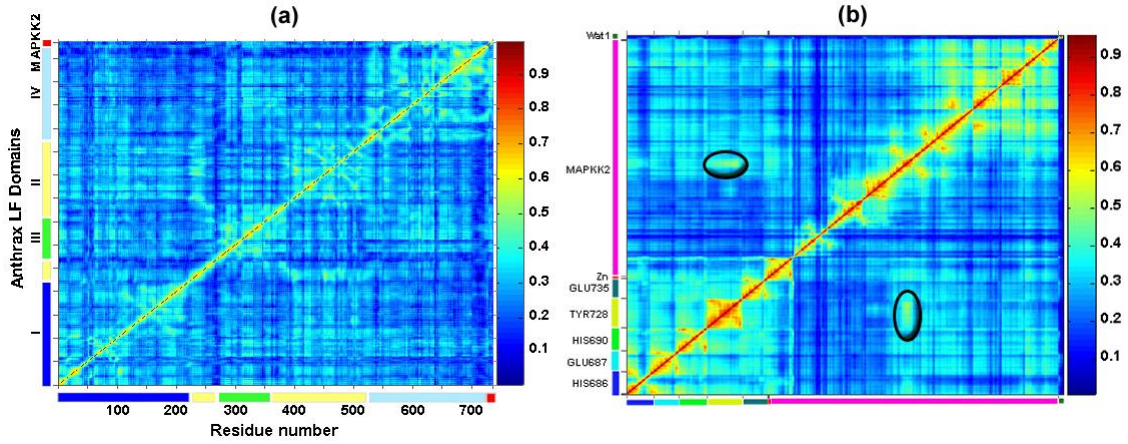


Figure A.1: Model **C**: **(a)** Generalized correlation matrix[124] for the complex formed by LF and the MAPKK2-like substrate (see section 2.4.2.3 in Methods), calculated over the last 15 ns of the all-atom MD simulation. Domains in LF were represented as in ref.[20] (see Fig.3.1); where domain **I** is the Protective Antigen (PA) binding domain, domain **II** is the Vegetative Insecticidal Protein 2 (VIP2)-like domain (VIP2 PDB: 1Q2S), domain **III** is an α-helical bundle required for LF activity and domain **IV** is the catalytic domain, MAPKK2 is an optimized substrate of the enzyme. **(b)** Generalized correlation matrix[124] for the active site residues, Wat1 water molecule (See Fig. 3.2, model **C**) and the MAPKK2-like substrate. The correlation between the motions of Tyr728 in the enzyme and Tyr-P1' in the substrate is indicated with circles.

A.2 Polarization of the active site

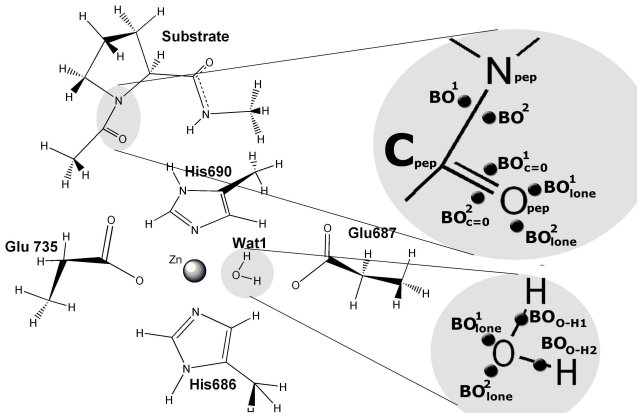


Figure A.2: Model **C** Boys Orbitals (BOs, see sections 2.3.6 and 3.2.6) from the peptidic bond of the substrate that undergoes the nucleophilic attack as well as those of the nucleophilic water.

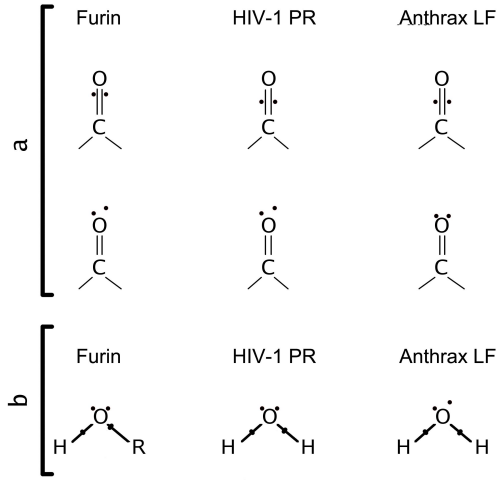


Figure A.3: Schematic representation of the BOs positions along reactive centers in Furin, HIV-1 PR and anthrax LF (see Table3.3, Chapter 3); **(a)** BOs from the carbonyl group of the substrates (see Fig.A.2), **(b)** BOs of the nucleophilic agents (a water molecule (*e.g.,* Wat1) for HIV-1 PR and anthrax LF; a serine side chain hydroxide group in Furin).

A.3 Hydration of the active site

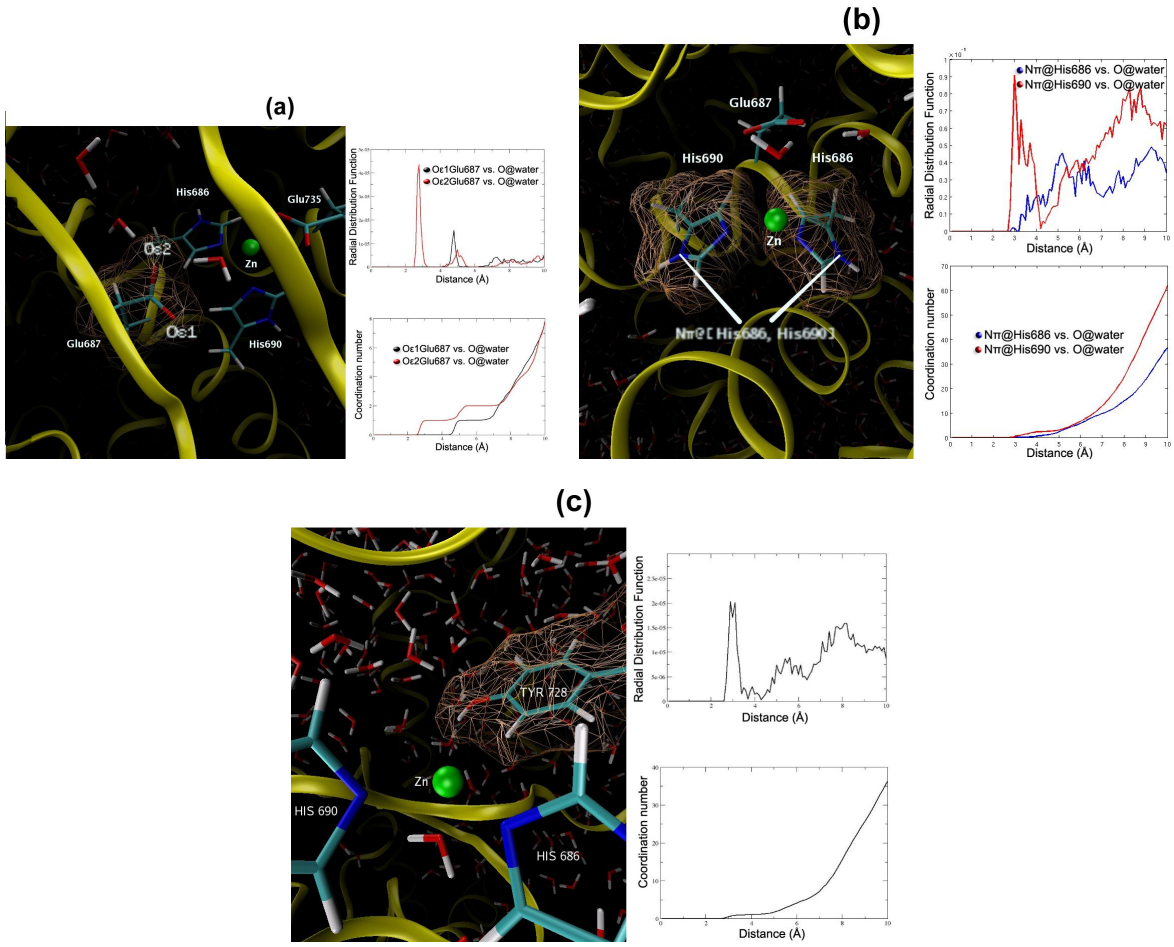


Figure A.4: Radial Distribution Functions (RDFs)of selected active site atoms that can form hydrogen bonds with water, as calculated from the last 15 ns of the all-atom MD simulation (LF Michaelis complex, model **C**). The atoms analyzed were: **(a)** Glu687 sidechain oxygens ([Oε1,Oε2]@Glu687), **(b)** the Nπ of the Zn-coordinating histidines (Nπ@[His686,His690], and **(c)** the hydroxylic oxygen of Tyr728 (Oη@Tyr728).

A.4 Parametrization of the zinc center and its coordination sphere

The development of AMBER parm98[83] force field parameters for the Zn-coordination sphere follows the procedure described in refs.[168] These parameters are reported here. The torsional potential barriers for the Zn-ligand interactions were set to zero. The van der Waals (vdW) parameters for Zn were taken from ref.[201]

Glu735		
		RESP Charge
N	N	-0.516300
H	H	0.293600
CA	CT	0.103570
HA	H1	0.174370
CB	CT	-0.064636
HB2	HC	0.009940
HB3	HC	0.009940
CG	CT	0.091869
HG2	HC	-0.021826
HG3	HC	-0.021826
CD	C	0.597668
OE1	O2	-0.667376
OE2	O2	-0.667376
C	C	0.536600
O	O	-0.581900

Glu687		
		RESP Charge
N	N	-0.516300
H	H	0.293600
CA	CT	0.084952
HA	H1	0.155752
CB	CT	0.041258
HB2	HC	-0.027296
HB3	HC	-0.027296
CG	CT	0.045003
HG2	HC	-0.030588
HG3	HC	-0.030588
CD	C	0.613381
OE1	O2	-0.644953
OE2	O2	-0.671807
C	C	0.536600
O	O	-0.581900

His690		
		RESP Charge
N	N	-0.415700
H	H	0.271900
CA	CT	0.084984
HA	H1	0.154283
CB	CT	-0.327246
HB2	HC	0.124867
HB3	HC	0.124867
CG	CC	0.004411
ND1	NA	0.001490
HD1	H	0.244746
CE1	CR	-0.047544
HE1	H5	0.173686
NX1	NX	-0.434857
CD2	CV	0.001298
HD2	H4	0.152164
C	C	0.597300
O	O	-0.567900

His686		
		RESP Charge
N	N	-0.415700
H	H	0.271900
CA	CT	0.051962
HA	H1	0.121262
CB	CT	-0.134628
HB2	HC	0.058825
HB3	HC	0.058825
CG	CC	0.042446
ND1	NA	0.018433
HD1	H	0.232455
CE1	CR	-0.084193
HE1	H5	0.184349
NX1	NX	-0.587521
CD2	CV	0.102181
HD2	H4	-0.014074
C	C	0.597300
O	O	-0.567900

Tyr728		
		RESP Charge
N	N	-0.415700
H	H	0.271900
CA	CT	0.023363
HA	H1	0.112363
CB	CT	-0.044990
HB2	HC	0.021326
HB3	HC	0.021326
CG	CA	0.072938
CD1	CA	-0.238500
HD1	HA	0.099650
CE1	CA	-0.033378
HE1	HA	0.072652
CZ	C	0.206741
OH	OH	-0.613199
HH	HO	0.451643
CE2	CA	-0.253301
HE2	HA	0.237100
CD2	CA	-0.091927
HD2	HA	0.110036
C	C	0.597300
O	O	-0.567900

Wat1		
		RESP Charge
O	OW	-0.838585
H1	HW	0.429785
H2	HW	0.438797

Atomic weights		
Zn	65.38	active site Zn
NX	14.01	N in His690

Bonds		
NB-Zn	121.0	2.110
NX-CV	410.0	1.394
NX-Zn	121.0	2.000
CR-NX	488.0	1.335
O2-Zn	180.0	2.020
Zn-OW	80.0	2.050

Angles		
CV-NB-Zn	20.0	127.50
NB-Zn-NX	20.0	104.56
NB-Zn-O2	30.0	103.85
NB-Zn-OW	30.0	103.66
CR-NB-Zn	20.0	127.50
CV-NX-Zn	20.0	122.73
NX-CV-H4	35.0	120.10
NX-Zn-O2	30.0	124.83
NX-Zn-OW	30.0	112.08
H5-CR-NX	35.0	125.14
CR-NX-CV	70.0	117.00
CR-NX-Zn	20.0	126.92
NA-CR-NX	70.0	120.00
CC-CV-NX	70.0	120.00
C -O2-Zn	30.0	113.98
O2-Zn-OW	30.0	105.86
Zn-OW-HW	18.0	113.66

Dihedrals				
X-CR-NX-X	2	10.00	180.0	2.0
X-CV-NX-X	2	4.80	180.0	2.0
X-CR-NY-X	2	10.00	180.0	2.0
X-CV-NY-X	2	4.80	180.0	2.0
H4-CV-NB-Zn	1	0.00	0.0	1.0
CV-NB-Zn-NX	1	0.00	0.0	1.0
CV-NB-Zn-O2	1	0.00	0.0	1.0
CV-NB-Zn-OH	1	0.00	0.0	1.0
NB-Zn-NX-CV	1	0.00	0.0	1.0
NB-Zn-NX-CR	1	0.00	0.0	1.0
NB-Zn-O2-C	1	0.00	0.0	1.0
NB-Zn-OW-HW	1	0.00	0.0	1.0
H5-CR-NB-Zn	1	0.00	0.0	1.0
CR-NB-Zn-NX	1	0.00	0.0	1.0
CR-NB-Zn-O2	1	0.00	0.0	1.0
CR-NB-Zn-OW	1	0.00	0.0	1.0
NA-CR-NB-Zn	1	0.00	0.0	1.0
CC-CV-NB-Zn	1	0.00	0.0	1.0
H4-CV-NB-Zn	1	0.00	0.0	1.0
CV-NX-Zn-O2	1	0.00	0.0	1.0
CV-NX-Zn-OW	1	0.00	0.0	1.0
NX-Zn-O2-C	1	0.00	0.0	1.0
NX-Zn-OW-HW	1	0.00	0.0	1.0
H5-CR-NX-CV	1	0.00	0.0	1.0
H5-CR-NX-Zn	1	0.00	0.0	1.0
CR-NX-CV-H4	1	0.00	0.0	1.0
CR-NX-Zn-O2	1	0.00	0.0	1.0
CR-NX-Zn-OW	1	0.00	0.0	1.0
NA-CR-NX-CV	1	0.00	0.0	1.0
NA-CR-NX-Zn	1	0.00	0.0	1.0
CC-CV-NX-Zn	1	0.00	0.0	1.0
CC-CV-NX-CR	1	0.00	0.0	1.0
O2-C-O2-Zn	1	0.00	0.0	1.0
O2-Zn-OW-HW	1	0.00	0.0	1.0
C-O2-Zn-OW	1	0.00	0.0	1.0
CT-C-O2-Zn	1	0.00	0.0	1.0
CV-NB-Zn-OW	1	0.00	0.0	1.0
NB-Zn-OW-HW	1	0.00	0.0	1.0
CR-NB-Zn-OW	1	0.00	0.0	1.0
CV-NX-Zn-OW	1	0.00	0.0	1.0
NX-Zn-OW-HW	1	0.00	0.0	1.0
CR-NX-Zn-OW	1	0.00	0.0	1.0
Zn-OW-HW-HW	1	0.00	0.0	1.0
Zn-OW-HW-HW	1	0.00	0.0	1.0

A.5 Hybrid Coarse-Grained/Molecular-Mechanics (CG/MM) simulations

Helix α19 (residues 346-367) as well as the protein and solvent atoms within 12.5Å from it were treated with either the AMBER parm98[83] or the Gromos96 43a1[84] force field (MM region); the rest of the protein was treated with the Go simplified potential[85] (CG region). The effects of the solvent outside the MM region were considered to be the sum of stochastic and frictional forces proportional to the mass and velocities of the particles in the system.[173] A cutoff of 14Å was selected for the vdW interactions and for the electrostatics. The bonds involving hydrogen atoms were kept fixed with SHAKE algorithm.[170,202] A time step of 2 fs was applied. Constant temperature (300 K) and pressure (1 atm) simulations were performed by coupling the system to a Berendsen thermostat and barostat.[171] The simulations were preceded by 1000 steps of energy minimization using the steepest descend algorithm and by a gentle heating of the system from 0K to 300K in 500 ps. Helix α19 unfolded within the first 25 ns of CG/MM simulations (Fig.A.5), confirming the results of all-atom MD simulations (Figs.3.3 and 3.4, Chapter 3).

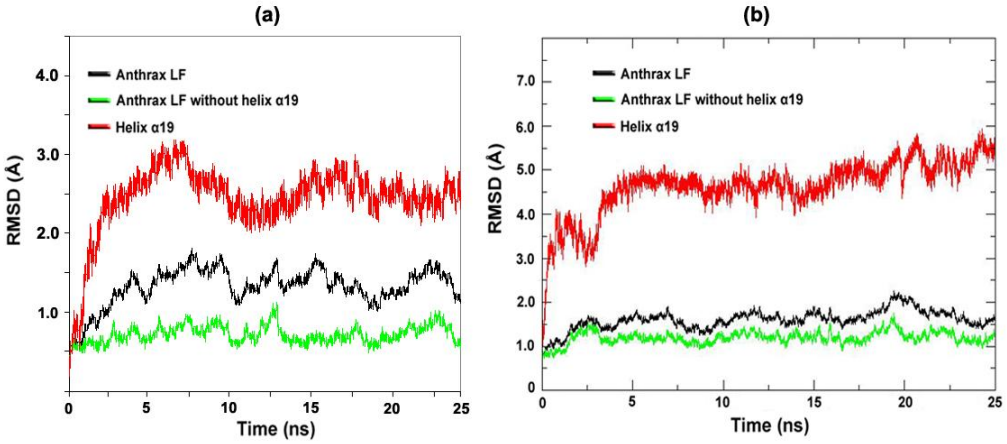


Figure A.5: Root Mean Square Displacement (RMSD) of LF C_α atoms during the first 25 ns of the CG/MM simulation using **(a)** AMBER parm98[83] force field and **(b)** Gromos96 43a1[84] force field. Note that helix α19 unfolds within the first 25 ns of the CG/MM simulations.

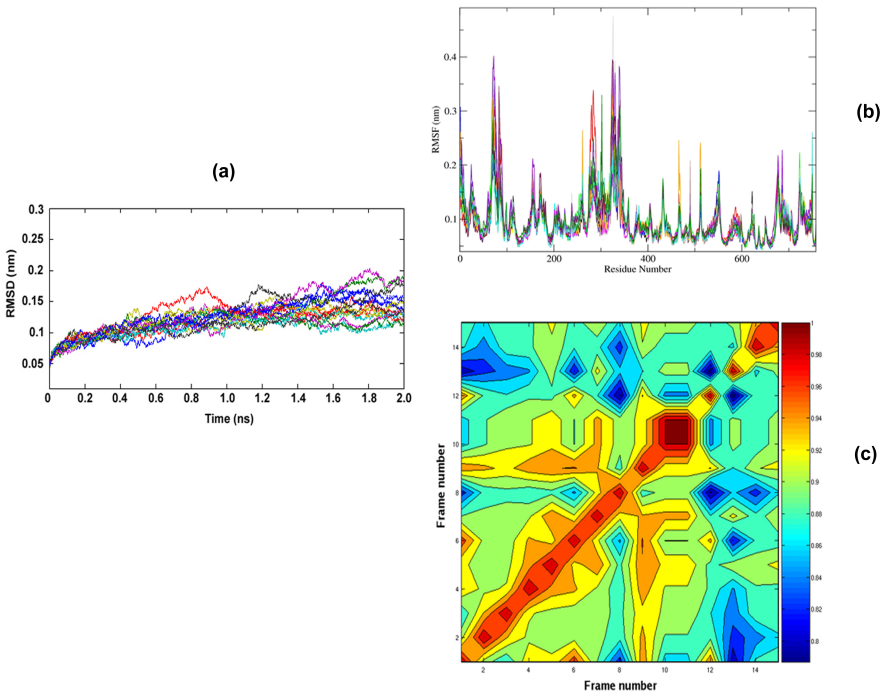


Figure A.6: **(a)** C_α RMSDs and **(b)** C_α RMSF for 15 (2ns long) CG/MM simulations performed with AMBER parm98[83] force field and using as initial structures, equally spaced snapshots (from 36 ns to 50 ns) of the all-atom MD simulation. In **(c)**, we observe a good overlap between each pair of RMSF series (the correlation coefficients were always higher than 0.8). These RMSF results are in good agreement with the calculations presented in Fig.3.4

A.6 Principal Component Analysis (PCA) and Normal Mode Analysis (NMA)

A.6.1 PCA

PCA[112] was performed on the last 15 ns of the all-atom MD simulation featuring the Michaelis complex with protonation state **C** in the active site. The first three principal components (PC1-PC3) accounted for ∼30%, 14% and 9% respectively; the others, which represent 6% or less of the total motion, were not included in the analysis. The extreme conformations of PC1-PC3 were used as input to detect the motion hinges as in ref.[113]

A.6.2 NMA

NMA large scale motions were calculated starting from the energy minimized structure taken from the last frame of the 50 ns all-atom MD simulation. These motions were represented into 20

normal modes; of these, the first three of them, which represented respectively 9%, 8% and 9% of the motions, were selected in order to compare them with their respective principal components. The extreme conformations of these three normal modes were used as input to detect the motion's hinges as in ref.[113]

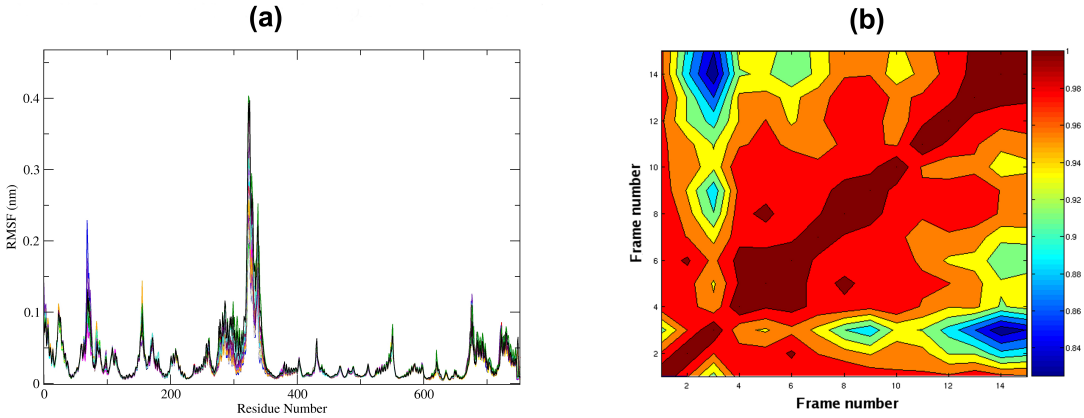


Figure A.7: **(a)** RMSF calculated from NMA on the same 15 structures used as starting points for the CG/MM calculations (see section A.5) . **(b)** The correlation coefficients between each pair of RMSF series was always higher than 0.84.

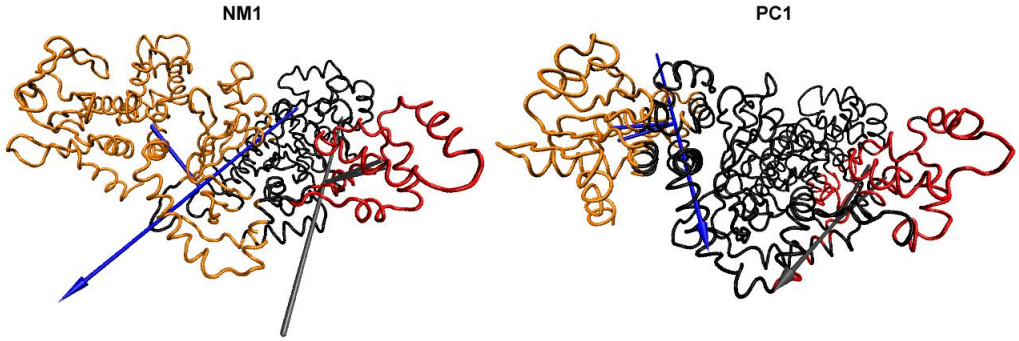


Figure A.8: Large scale motions as described by the first normal mode and the first principal component of LF: rigid domains are depicted in different colors; the arrows indicate the effective rotation axis (*i.e.,* hinges, calculated as in ref[113]) between two adjacent rigid domains.

		NM1	**NM2**	**NM3**	**PC1**	**PC2**	**PC3**
Hinge 0-1	Angle	∼23±1	∼26±1	∼19±1	∼10±11	∼5±23	∼8±11
	RMSD$_{er}$ (Å)	0.94	0.86	0.77	1.23	1.74	1.24
	Error($\Delta\varepsilon$)	0.56%	0.36%	0.23%	5.71%	20.66%	9.97%
Hinge 0-2	Angle	∼17±2	∼12±8	∼8±6	∼14±18	∼13±20	
	RMSD$_{er}$ (Å)	0.65	0.96	0.78	2.68	2.34	
	Error($\Delta\varepsilon$)	0.97%	15.84%	16.94%	14.99%	17.64%	100%

Table A.1: Hinge parameters for the first three normal modes NM1-NM3 and the first three principal components PC1-PC3 of anthrax LF. For each hinge, the effective rotation angle, the RMSD between the extreme conformations of the moving domain and the relative errors ($\Delta\varepsilon$) are reported here.

A.7 Structural predictions of α19

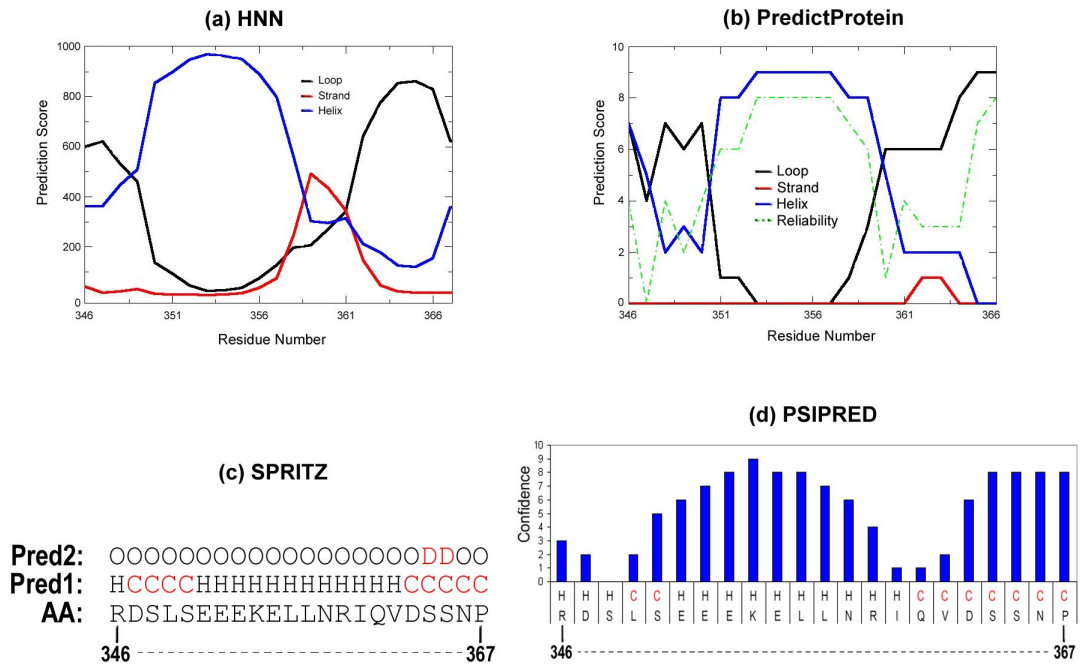


Figure A.9: Secondary structure predictions for α19 using: **(a)** Hierarchical Neural Network (HNN),[161] **(b)** PredictProtein[137] server. For both, HNN and PredictProtein servers prediction scores (for helix, strand, and loop segments) with respect to the residues number are shown; PredictProtein provides also the reliability of the predictions, **(c)** SPRITZ[160] gives a prediction for secondary structure (Pred1) for each residue indicated by a letter (H and C for helix and coil , respectively) and a disorder measure (Pred2) symbolized by "O" and "D" for order and disorder, respectively. **(d)** PSIPRED[159] provides a prediction for secondary structure of each residue as in SPRITZ and this comes along with the confidence of each prediction. All these servers provide a consistent picture in which part of α19 is likely to be a random coil.

A.8 Energetic basis for LF substrate selectivity

	Val	Tyr	Pro	Tyr	Pro	Met	Glu
	P3	P2	P1	P1'	P2'	P3'	P4'
Selectivity	1.5	3.1	–	3.0	1.9	1.3	1.6
Total energy (kcal/mol)	(-11 ± 2)	(-19 ± 4)	(-5 ± 5)	(-42 ± 4)	(-20 ± 2)	(-18 ± 3)	(0 ± 24)
Elect (kcal/mol)	(-1 ± 1)	(-3 ± 3)	(3 ± 5)	(-23 ± 4)	(-10 ± 2)	(-4 ± 2)	(7 ± 24)
vdW (kcal/mol)	(-9 ± 2)	(-15 ± 1)	(-8 ± 2)	(-18 ± 2)	(-9 ± 1)	(-14 ± 2)	(-7 ± 2)
CorrCoef (Total vs. Elect)	0.3373	0.9324	0.9683	0.8424	0.8344	0.8059	0.9973
CorrCoef (Total vs. vdW)	0.8255	0.3692	-0.1502	0.0718	0.5711	0.7734	-0.2887

Table A.2: Per residue substrate selectivity of anthrax LF. Experimental selectivity values were taken from Turk *et al.,*[21] The interaction energies (Total, electrostatics (Elect.), van der Waals (vdW)) between the substrate residues and LF were calculated during the last 15 ns of the all-atom MD simulation featuring model **C**. The correlation coefficients between the time series of the interaction energies were also calculated.

Appendix B

Atx1-Ccc2 copper transport system: additional calculations

B.1 Comparison between calculated and experimental NMR order parameters (S^2)

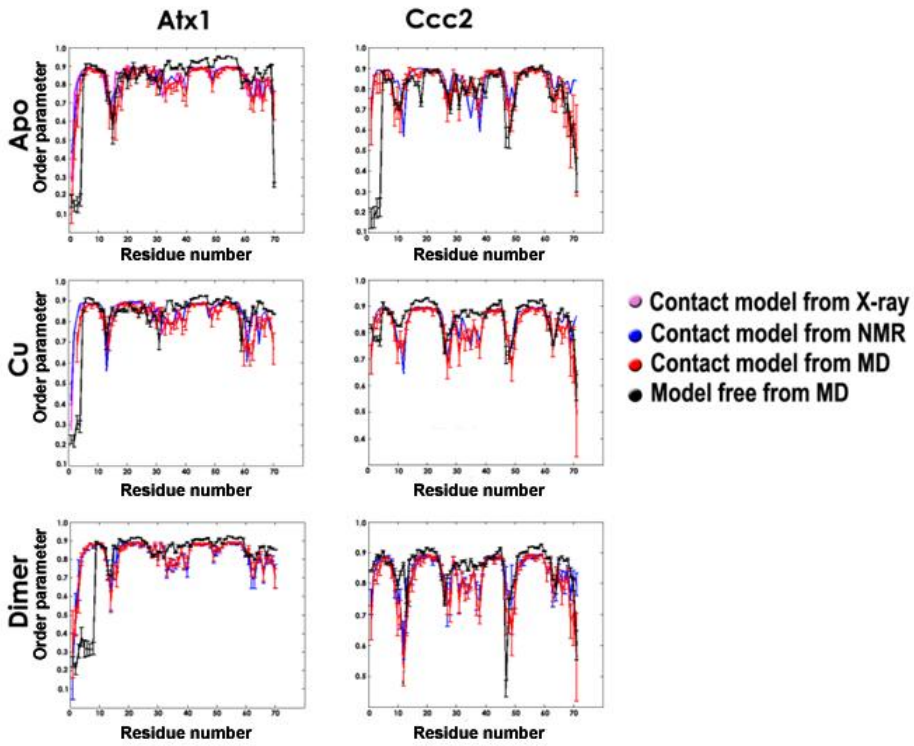


Figure B.1: Order parameters calculated from individual frames of molecular dynamics trajectory, NMR and X-ray structures using the contact model and calculated from MD trajectories using the Model free approach (see section 2.4.1.2). The systems under assessment were: Atx1-[Apo,Cu], Ccc2-[Apo,Cu] and Atx1-Cu-Ccc2 (dimer). Note the general agreement of the order parameters for the systems under study independently of the model used for the analysis.

B.2 Identification of *hot spot* residues on Atx1 and Ccc2

Atx1 and Ccc2 residues participating in protein-protein interactions during copper transfer were identified by screening the binding interactions with an empirical scoring function using the program FastContact[138]. As input for this algorithm we used 30 equally spaced frames from the last 30 ns of the Atx1-Cu-Ccc2 MD simulation. The identified residues were [Lys24,Lys28,Lys59,Lys62]@Atx1 and [Glu60,Asp61,Asp65]@Ccc2. Details can be seen in tables B.1 and B.2

Aminoacid	ΔG_{desolv}	ΔE_{elect}	ΔG_{bind}
Lys24	2.9 ± 0.7	-12 ± 2	-9 ± 2
Lys28	1.8 ± 0.4	-9 ± 2	-7 ± 2
Lys59	0.6 ± 0.4	-8 ± 4	-7 ± 4
Lys62	1.2 ± 0.9	-6 ± 4	-5 ± 3

Table B.1: FastContact[138] identified *hot spot* residues on the interacting surface of Atx1. All residues are positively charged lysines.

Aminoacid	ΔG_{desolv}	ΔE_{elect}	ΔG_{bind}
Asp61	2 ± 1	-15 ± 9	-13 ± 8
Asp65	0.9 ± 1	-7 ± 3	-6 ± 3
Glu60	1.5 ± 0.6	-6 ± 3	-5 ± 2

Table B.2: FastContact[138] identified *hot spot* residues on the interacting surface of Ccc2. All residues are charged negatively(aspartic and glutamic acid). It is therefore not surprising that the contribution of electrostatic interactions (ΔE_{elect}) with the positive charges on Atx1 have a dominant effect over the desolvation free energy (ΔG_{desolv}).

Bibliography

[1] Lovell T, Himo F, Han W-G, and Noodleman L. Density functional methods applied to metalloenzymes. *Coord Chem Rev*, 238–239:211–232, 2003.

[2] Bertini I, Siegel A, and Siegel H. *2001 Handbook on metalloproteins*. New York: Marcel Dekker, 2001.

[3] Williams RJP. The natural selection of the chemical elements. *Cell Mol Life Sci*, 53: 816–828, 1997.

[4] Auld DS. Zinc coordination sphere in biochemical zinc sites. *Biometals*, 14:271–313, 2001.

[5] Finney L and O'Halloran TV. Transition metal speciation in the cell: insights from the chemistry of metal ion receptor. *Science*, 300:931–936, 2003.

[6] Dokmanic I, Sikic M, and Tomic S. Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acod residues involved in the coordination. *Acta Cryst*, D64:257–263, 2008.

[7] Meselson M, Guillemin J, and Hug-Jones M. The Sverdlovsk anthrax outbreak of 1979. *Science*, 266:1202–1208, 1994.

[8] Swartz MN. Recognition and managment of anthrax - an update. *N Engl J Med*, 345(22): 1621–1626, 2001.

[9] Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen M, Friedlander AM, Hauer J, McDade J, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, and Tonat K. Anthrax as a biological weapon: medical and public health management. *JAMA*, 281(18): 1735–1745, 1999.

[10] Friedlander AM. Anthrax: clinical features, pathogenesis, and potential biological warfare threat. *Curr Clin Top Infect Dis*, 20:335–349, 2000.

[11] Abrami L, Reig N, and van der Goot FG. Anthrax toxin: the long and winding road that leads to the kill. *Trends Microbiol*, 13:72–78, 2005.

[12] Bradley KA, Mogridge J, Mourez M, Collier J, and Young JAT. Identification of the cellular receptor of anthrax toxin. *Nature*, 414:225–229, 2001.

[13] Klimpel KR, Molloy SS, Thomas G, and Leppla SH. Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *PNAS*, 89(21):10277–10281, 1992.

[14] Lacy DB, Wigelsworth DJ, Melnyk RA, Harrison SC, and Collier RJ. Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation. *PNAS*, 101(36):13147–13151, 2004.

[15] Mogridge J, Cunningaham K, and Collier RJ. Stoichiometry of anthrax toxin complexes. *Biochemistry*, 41:1079–1082, 2002.

[16] Abrami L, Liu S, Cosson P, Leppla SH, and van der Goot FG. Anthrax toxin triggers endocytocys of its receptor via a lipid raft-mediated clathrin-dependent process. *J Cell Biol*, 160:321–328, 2003.

[17] Collier RJ and Young JA. Anthrax toxin. *J Cell Biol*, 160:321–328, 2003.

[18] Leppla SH. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentration of eukaryotic cells. *PNAS*, 79(10):3162–3166, 1982.

[19] Vitale G, Pellizzari R, Recchi C, Napolitani G, Mock M, and Montecucco C. Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/theonine phosphorilation of MAPKs in cultured macrophages. *Biochem Biophys Res Comm*, 248:706–711, 1998.

[20] Pannifer AD, Wong TY, Schwarzenbacher R, Renatus M, Petosa C, Collier RJ, Bienkowska J, Lacy DB, Park S, Leppla SH, Hanna P, and Liddington RC. Crystal structure of the anthrax lethal factor. *Nature*, 414:229–233, 2001.

[21] Turk BJ, Wong TY, Schwarzenbacher R, Jarrell ET, Leppla SH, Collier RJ, Liddington RC, and Cantley LC. The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nat Struct Mol Biol*, 11:60–66, 2004.

[22] Hammond SE and Hanna PC. Lethal factor active-site mutations affect catalytic activity in vitro. *Infect Immun*, 66:2374–2378, 1998.

[23] Carnevale V, Raugei S, and Carloni P. Aspects of the reaction mechanism of serine and aspartic proteases investigated by molecular simulations. *Biophys J*, submitted, 2007.

[24] Huffman DL and O'Halloran TV. Function, structure and mechanism of intracellular copper trafficking proteins. *Ann Rev Biochem*, 70:677–701, 2001.

[25] Hasset R, Dix DR, Eide DJ, and Kosman DJ. The Fe(II) permease Fet4p functions as a low affinity copper transporter and supports normal copper traffiking in Saccharomyces cerevisiae. *Biochem J*, 351(Pt 2):477–484, 2000.

[26] Dancis A, Yuan DS, Haile D, Askwith C, Eide D, Mohele C, Kaplan J, and Klausner RD. Molecular characterization of a copper transport protein in S. cerevisiae: an unexpected role for copper in iron transport. *Cell*, 76:393–402, 1994.

[27] Zhou H and Thiele DJ. Identification of a novel high affinity copper transport complex in the fission yeast Schizosaccharomyces pombe. *J Biol Chem*, 276:20529–20535, 2001.

[28] Gitschier J Zhou B. hCTR1: a human gene for copper uptake identified by complementation in yeast. *PNAS*, 94:7481–7486, 1997.

[29] O'Halloran TV and Cullota VC. Metallochaperones, an intracellural shuttle service for metal ions. *J Biol Chem*, 275(33):25057–25060, 2000.

[30] De Silva DM, Askwith CC, Eide D, and Kaplan J. The Fet3 gene product required for high affinity iron transport in yeast is a cell surface ferroxidase. *J Biol Chem*, 270:1098–1101, 1995.

[31] Singh A, Severance S, Kaur N, Wiltsie W, and Kosman DJ. Assembly, Activation, and Trafficking of the Fet3pFtr1p High Affinity Iron Permease Complex in Saccharomyces cerevisiae. *J Biol Chem*, 281(19):13355–13364, 2006.

[32] McCord JM and Fridovich I. Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *J Biol Chem*, 244:6049–6055, 1969.

[33] Stoj C and Kosman DJ. Cuprous oxidase activity of yeast Fet3 and human ceruloplasmin: implications for function. *FEBS letters*, 554(3):422–426, 2003.

[34] La Fontaine S and Mercer JFB. Trafficking of the copper ATPases, ATP7A and ATP7B: role in copper homeostasis. *Archives of Biochemistry and Biophysics*, 463(2):149–167, 2007.

[35] Hubbard TJP, Murzin AG, Brenner SE, and Chlothia C. SCOP: a structural classification of proteins database. *Nucleic Acid Res*, 25:236–239, 1997.

[36] Banci L, Bertini I, Ciofi-Baffoni S, Huffman DL, and O'Halloran TV. Solution structure of the yeast copper transporter domain Ccc2a in the Apo and Cu(I)-loaded states. *J Biol Chem*, 276:8415–8426, 2001.

[37] Jordan IK, Natale DA, Koonin EV, and Galperin MY. Independent evolution of Heavy Metal-Associated domains in copper chaperones and copper-transporting ATPases. *J Mol Evol*, 53:622–633, 2001.

[38] Huffman DL and O'Halloran TV. Energetics of copper trafficking between the Atx1 metallochaperone and the intracellular copper transporter, Ccc2. *J Biol Chem*, 275(5):18611–18614, 2000.

[39] Banci L, Bertini I, Cantini F, Felli IC, Gonnelli L, Hadjiliadis N, Pierattelli R, Rosato A, and Voulgaris P. The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. *Nat Chem Biol*, 2:367–368, 2006.

[40] Harrison MD, Jones CE, Solioz M, and Dameron CT. Intracellular copper routing: the role of copper chaperones. *Trends Biochem Sci*, 25(1):29–32, 2000.

[41] Field LS, Luk E, and Culotta VC. Copper chaperones: personal escorts for metal ions. *J Bioenerg Biomembr*, 34(5):373–379, 2002.

[42] Singleton C and Le Brun NE. Atx1-like chaperones and their cognate P-type ATPases: copper-binding and transfer. *Biometals*, 20:275–289, 2007.

[43] De Feo CJ, Aller SG, and Unger VM. A structural perspective on copper uptake in eukaryotes. *Biometals*, 20:705–716, 2007.

[44] Arnesano F, Banci L, Bertini I, Ciofi-Baffoni S, Molteni E, Huffman DL, and O'Halloran TV. Metallochaperones and metal-transporting ATPases: a comparative analysis of sequences and structures. *Genome Res*, 12:255–271, 2002.

[45] Markossian KA and Kurganov BI. Copper chaperones, intracellular copper trafficking proteins. Function, structure, and mechanism of action. *Biochemistry (Moscow)*, 68(8): 1013–1025, 2003.

[46] Bounce J, Achila D, Hetrick E, Lesley L, and Huffman DL. Copper transfer studies between the N-terminal copper binding domains one and four of human Wilson protein. *Biochimica et Biophysica Acta*, 1760:907–912, 2006.

[47] Wernimont A, Yatsunyk LA, and Rosenzweig AC. Binding of copper(I) by the Wilson disease protein and its copper chaperone. *J Biol Chem*, 279(13):12269–12276, 2004.

[48] Yatsunyk LA and Rosenzweig AC. Cu(I) binding and transfer by the N terminus of the Wilson disease protein. *J Biol Chem*, 282:8622–8631, 2007.

[49] Strausak D, Howie MK, Firth SD, Schlicksupp A, Pipkorn R, Multhaup G, and Mercer JBF. Kinetic analysis of the interaction of the copper chaperone Atox1 with the metal binding sites of the Menkes protein. *J Biol Chem*, 278(23):20821–20827, 2003.

[50] Banci L, Bertini I, Ciofi-Baffoni S, Chasapis CT, Hadjiliadis N, and Rosato A. An NMR study of the interaction between the human copper(I) chaperone and the second and fifth metal-binding domains of the Menkes protein. *FEBS J*, 272:865–871, 2005.

[51] Born M and Oppenheimer JR. Zur Quantentheorie del Molekeln. *Ann Phys*, 84:457–484, 1927.

[52] Delhommelle J and Millie P. Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulations. *Mol Phys*, 99(8): 619–625, 2001.

[53] Kong J. Combining rules for intermolecular potential parameters. II. Rules for the Lennard-Jones (12-6) potential and the Morse potential. *J Chem Phys*, 59:2464, 1973.

[54] Waldman M and Hagler AT. New combining rules for rare gas van der waals parameters. *J Comput Chem*, 14(9):1077–1084, 1993.

[55] Meister J and Schwarz WHE. Principal components of ionicity. *J Phys Chem*, 98:8245–8252, 1994.

[56] Jensen F. *Introduction to computational chemistry*. John Wiley and Sons Ltd, 2 edition, 2007.

[57] Gasteiger J and Saller H. Calculation of the charge distribution in conjugate systems by a quantification of the resonance concept. *Ang Chem Int Ed.*, 24:687–689, 1985.

[58] Morse PM. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys Rev*, 34:57–64, 1929.

[59] Hagler AT, Huler E, and Lifson S. Energy functions for peptide and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J Am Chem Soc*, 96(17):5319–5327, 1974.

[60] Ewig CS, Berry R, Dinur U, Hill JR, Hwang MJ, Li H, Liang C, Maple J, Peng Z, Stockfish TP, Thacher TS, Yan L, Ni X, and Hagler AT. Derivation of class II force fields. VIII. derivation of a general quantum mechanical force field for organic compounds. *J Comput Chem*, 22(15):1782–1800, 2001.

[61] Palmo K, Mannfors B, Mirkin NG, and Krimm S. Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers*, 68: 383–394, 2003.

[62] Allured VS, Kelley CM, and Landis CR. SHAPES empirical force field: new treatment of angular potentials and its applications to square-planar transition-metal complexes. *J Am Chem Soc*, 113:1–12, 1991.

[63] Mandell MJ. On the properties of a periodic fluid. *J Stat Phys*, 15:299–305, 1976.

[64] Steinbach PJ and Brooks BR. New sherical cutoff methods for long range forces in macromolecular simulations. *J Comp Chem*, 15:667–683, 1994.

[65] Ewald P. Die berechnug optischer und elecktrostatischer gitterpotentiale. *Ann Phys*, 64: 253–289, 1921.

[66] Fausett E. *Numerical methods: algorithms and applications.* Prentice Hall, New Jersey, 2003.

[67] Rahman A. Correlation in the emotion of atoms in liquid argon. *Phys Rev*, 136(2A): A405–A411, 1964.

[68] Verlet L. Computer "experiments" on classical Fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev*, 159(1):98–103, 1967.

[69] Newton I. *The Mathematical Principles of Natural Philosophy.* The Philosophical Library, Inc. Copyright 1964, 1964. Originally published in 1687 as: De Philosophiae Naturalis Principia Mathematica.

[70] Strmer C. Sur les trajectories des corpuscules lectriss :. *Arch Sci Phys Nat, Genve*, 24: 5–18,113–158,221–247, 1907.

[71] Beeman D. Some multistep methods for use in molecular dynamics calculations. *J Comp Phys*, 20:130–139, 1976.

[72] Woodcock LV. Isothermal molecular dynamics calculations for liquid salts. *Chem Phys Lett*, 10:257–261, 1971.

[73] Hoover WG, Ladd AJC, and Moran B. High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Phys Rev Lett*, 48:1818–1820, 1982.

[74] Evans DJ. Computer "experiment" for nonlinear thermodynamics of Couette flow. *J Chem Phys*, 78(6):3297–3302, 1983.

[75] Berendsen HJC, Postma JPM, van Gusteren WF, DiNola A, and Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81:3684–3690, 1984.

[76] Nos S. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*, 52:255–268, 1984.

[77] Hoover WG. Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A*, 31: 1695–1697, 1985.

[78] Parrinello M and Rahman A. Crystal structures and pair potentials. A molecular-dynamics study. *Phys Rev Lett*, 45:1196–1199, 1980.

[79] Parrinello M and Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*, 52(12):7182–7190, 1981.

[80] Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys*, 72:2384–2393, 1980.

[81] Feller SE, Zhang Y, Pastor RW, and Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys*, 103(11):4613–4621, 1995.

[82] Neri M, Anselmi C, Cascella M, Maritan A, and Carloni P. Coarse-Grained Model of proteins incorporating atomistic detail of the active site. *Phys Rev Lett*, 95:218102, 2005.

[83] **(a)** Cheatham TE III, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat.*J Biomol Struct Dyn* 16:845-862, 1999. **(b)** Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel GL, Kollman PA. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics an1d free energy calculations to simulate the structural and energetic properties of molecules.*Comput Phys Commun.* 91:1-41, 1995.

[84] van Gunsteren WF, Daura X, and Mark AE. *Encyclopedia of Computational Chemistry*, chapter GROMOS force field, page 1211. New York: John Wiley and Sons, 1998.

[85] Noguti T and Go N. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature*, 296:776–778, 1982.

[86] Ensing B, Nielsen SO, Moore PB, Klein ML, and Parrinello M. Energy conservation in adaptive hybrid atomistic/coarse-grained molecular dynamics. *JCTC*, 3:1100–1105, 2007.

[87] Neri M, Baaden M, Carnevale V, Anselmi C, Maritan A, and Carloni P. Microseconds dynamics simulations of the outer-membrane protease T. *Biophys J*, 94:71–78, 2008.

[88] Levitt M. Warshel A. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*, 103(2): 227–249, 1976.

[89] CPMD, Copyright IBM Corp 1990-2006, Copyright MPI für Festkörperforschung Stuttgart 1997-2001, .

[90] Hohenberg P and Kohn W. Inhomogeneous electron gas. *Phys Rev*, 136(3B):B864–B871, 1964.

[91] Kohn W and Sham LJ. Self-consistent equations including exchange and correlation effects. *Phys Rev*, 140(4A):A1133–A1138, 1965.

[92] Perdew JP, Tao J, Staroverov VN, and Scuseria GE. Meta-generalized gradient approximation: explanation of a realistic nonmpirical density functional. *J Chem Phys*, 120:6898, 2004.

[93] Perdew JP. Accurate density functional for the energy: real-space cutoff of the gradient expansion for the exchange hole. *Phys Rev Lett*, 55(16):1665–1668, 1985.

[94] Stevens P, Devlin JF, Chavolowski CF, and Frisch MJ. Ab initio calculation of vibrational absortion and circular dichroism spectra using density functional force fields. *J Phys Chem*, 98(45):11623–11627, 1994.

[95] Marzari N and Vandebilt D. Maximally localized generalized Wannier functions for composite energy bands. *Phys Rev B*, 56:12847, 1997.

[96] **(a)** Alber F, Folkers G, Carloni P. Dimethyl phosphate: stereoelectronic versus environmental effects. *J. Phys. Chem. B* 103:6121-6126, 1999. **(b)** Silvestrelli PL, Marzari N, Vanderbilt D, Parrinello M. Maximally-localized Wannier functions for disordered systems: application to amorphous silicon. *Solid State Commun* 107:7-11, 1998. **(c)** Berghold G, Mundy CJ, Romero AH, Hutter J, Parrinello M. General and efficient algorithms for obtaining maximally localized Wannier functions.*Phys. Rev. B* 61:10040-10048,2000. **(d)** Magistrato A, Robertazzi A, Carloni P. Nitrogen fixation by a molybdenum catalyst mimicking the function of the nitrogenase enzyme: A critical evaluation of DFT and solvent effects.*J. Chem. Theor. Comp.* 3:1708-1720, 2007., .

[97] Lipari G and Szabo A. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*, 104:4546–4559, 1982.

[98] Lipari G and Szabo A. Model free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc*, 104:4559–4570, 1982.

[99] Kay LE, Torkia DA, and Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: applications to staphylococcal nuclease. *Biochemistry*, 28:8972–8979, 1989.

[100] Wriggers W, Mehler E, Pitici F, Weinstein H, and Schulten K. Structure and dynamics of calmodulin in solution. *Biophys J*, 74:1622–1639, 1998.

[101] Fiorin G, Biekofsky RR, Pastore A, and Carloni P. Unwinding the helical linker of calcium-loaded calmodulin: a molecular dynamics study. *Proteins*, 61:829–839, 2005.

[102] Showalter SA and Brüschweiler R. Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: applications to the AMBER99SB force field. *J Chem Theory Comput*, 3:961–975, 2007.

[103] **(a)** Wangness RK, Bloch F. The dynamical theory of nuclear induction. *Phys Rev*, 89:728-739, 1953. **(b)** Bloch F. Dynamical theory of nuclear induction 2. *Phys Rev* 102:104-135, 1956. **(c)** Redfield AG. On the theory of relaxation processes. *IBM J Res Dev* 1:19-31, 1957. **(d)** Redfield AG. The theory of relaxation processes. *Adv Magn Reson* 1:1-32, 1965.

[104] Bremi T, Brüschweiler R, and Ernst RR. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide. *J Am Chem Soc*, 119:4272–4284, 1997.

[105] Jarymowycz VA and Stone MJ. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev*, 106:1624–1671, 2006.

[106] Pearson K. On lines and planes of closest fits to systems of points in space 1901. *Phil Mag*, 2(6):559–572, 1901.

[107] Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol*, 24:417–441,498–520, 1933.

[108] Cattell RB. Scree test for the number of factors. *Multivariate Behavior Research*, 1:140–161, 1966.

[109] Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas*, 20:141–151, 1960.

[110] Kaiser HF. A note on Guttman's lower bound for the number of common factors. *Journal of Statistical Psycology*, 14:1–1, 1961.

[111] Guttman L. Some necessary conditions for common factor analysis. *Psycometria*, 19: 149–161, 1954.

[112] Barret CP, Hall BA, and Noble EM. Dynamite: a simple way to gain insight into protein motions. *Acta Cryst*, D60:2280–2287, 2004.

[113] Wriggers W and Schulten K. Protein domain movements: Detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins*, 29:1–14, 1997.

[114] Go N, Noguti T, and Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *PNAS*, 80(12):3696–3700, 1983.

[115] Brooks B and Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *PNAS*, 80(21):6571–6575, 1983.

[116] Levitt M, Sander C, and Stern PS. Normal-mode analysis of a protein: Bovine pancreatic trypsin inhibitor. *Int J Quant Chem: Quant Biol Symp*, 10:181–199, 1983.

[117] Ma J and Karplus M. Ligand-induced conformational changes in ras p21: s normal mode and energy minimization analysis. *J Mol Biol*, 274(1):114–131, 1997.

[118] Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, and Gunsalus IC. Dynamics of ligand-binding to myoglobin. *Biochemistry*, 14(24):5355–5373, 1975.

[119] Elber R and Karplus M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, 235:318–321, 1987.

[120] Cui Q and Bahar I. *Normal mode analysis: theory and applications to biological and chemical systems*. Chapman and Hall/CRC mathematical and computational biology series. CRC press, 2006.

[121] Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett*, 77:1905–1908, 1996.

[122] Bahar I, Atilgan AR, and Erman B. Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Folding Des*, 2:173–181, 1997.

[123] Ichiye T and Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular-dynamics and normal mode simulations. *Proteins*, 11:205–217, 1991.

[124] Lange OF and Grubmüller H. Generalized correlations for biomolecular dynamics. *Proteins*, 62:1053–1061, 2006.

[125] Northrup S, Allison SA, and McCammon JA. Brownian dynamics of diffusion-influenced bimolecular reactions. *J Chem Phys*, 80:1517–1524, 1984.

[126] . URL `http://iweb.tntech.edu/macrodox/macrodox.html`.

[127] Ermak D and McCammon JA. Brownian dynamics with hydrodynamic interactions. *J Chem Phys*, 69(4):1352–1360, 1978.

[128] Mazo R. On the theory of Brownian motion. III. Two-body distribution function. *J. Stat. Phys.*, 1(4):559–562, 1969.

[129] Deutch M and Oppenheim I. Molecular theory of Brownian motion for several particles. *J Chem Phys*, 54:3547–3555, 1971.

[130] Oseen CW. *Hydrodynamik*. Akademische, Leipzig, 1927.

[131] Northrup S and Hynes JT. Short range caging effects for reactions in solution. I. Reaction rate constants and short range caging picture. *J Chem Phys*, 71(2):871–883, 1979.

[132] Elber R. Long-timescale simulation methods. *Current Opinion in Structural Biology*, 15 (2):151–156, 2005.

[133] Lüdemann SK, Lounnas V, and Wade R. How do substrates enter and products exit the buried active site of citochrome P450? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanism. *J Mol Biol*, 303(5):797–811, 2000.

[134] Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, and Sali A. Comparative protein structure modelling of genes and genomes. *Annu Rev Biophys Biomol Struct*, 29: 291–235, 2000.

[135] Bonneau R and Baker D. Ab initio protein structure prediction: progress and prospects. *Annu Rev Biophys Biomol Struct*, 30:173–189, 2001.

[136] Kabsch W and Sanders C. Dictionary of protein secondary structure: patterns recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[137] Rost B, Yachdav G, and Liu J. The PredictProtein server. *Nucl Acids Res*, 32(Web Server Issue):W321–W326, 2004.

[138] Camacho CJ and Zhang C. FastContact: rapid estimate of contact and binding free energies. *Bioinformatics*, 21(10):2534–2536, 2005.

[139] Mock M and Mignot T. Anthrax toxins and the host: a story of intimacy. *Cell Microbiol*, 5:15–23, 2003.

[140] Guidi-Rotani C. The alveolar macrophage: The Trojan horse of Bacillus anthracis. *Trends Microbiol*, 10:405–409, 2002.

[141] Dixon TC, Fadl AA, Koheler TM, Swanson JA, and Hanna PC. Early Bacillus anthracis-macrophage interactions: intracellular survival and escape. *Cell Microbiol*, 2:453–463, 2000.

[142] Shoop WL, Xiong Y, Woods A, Guo J, Pivnichny JV, Felcetto T, Michael BF, Bansal A, Cummings RT, Cinnungam BR, Friedlander AM, Douglas CM, Patel SB, Wisniewski D, Scapin G, Spaolowe SP, Zaller DM, Chapman KT, Scolnick EM, Schmatz DM, Bartizal K, MacCoss M, and Hermes JD. Anthrax lethal factor inhibition. *PNAS*, 102:7958–7963, 2005.

[143] Pezard C, Berche P, and Mock M. Contribution of individual toxin components to virulence of Bacillus anthracis. *Infect Immun*, 59:3472–3477, 1991.

[144] Petosa C, Collier RJ, Klimpel KR, Leppla SH, and Liddington RC. Crystal structure of the anthrax toxin protective antigen. *Nature*, 385:833–838, 1997.

[145] Montecucco C, Tonello F, and Zanotti G. Stop the killer: how to inhibit the anthrax lethal factor metalloprotease. *Trends Biochem Sci*, 29:282–285, 2004.

[146] Hultgreen SJ Bann JC. Structural Biology: Anthrax hijacks host receptor. *Nature*, 430: 843–844, 2004.

[147] Duesbery NS, Webb CP, Leppla SH, Gordon VM, Klimpel KR, Copeland TD, Ahn NG, Oskarsonn MK, Fukasawa K, Paul KD, and Vande Woude GF. Proteolytic inactivation of MAP-Kinase-Kinase by anthrax lethal factor. *Science*, 280:734–737, 1998.

[148] Weston CR, Lambright DG, and Davis RJ. MAP Kinase signalling specificity. *Science*, 296:2345–2347, 2002.

[149] Friedlander AM. Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. *J Biol Chem*, 261(16):7123–7126, 1986.

[150] Kirby JE. Anthrax lethal toxin induces human endothelial cell apoptosis. *Infect Immun*, 72(1):430–439, 2004.

[151] Milne JC, Furlong D, Hanna PC, Well JS, and Collier RJ. Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *J Biol Chem*, 269:20607–20612, 1994.

[152] Guichard A, Park JM, Cruz-Moreno B, Karin, and M Bier E. Anthrax lethal factor and edema factor act on conserved targets in Drosophila. *PNAS*, 103:3244–3249, 2006.

[153] Panchal RG, Hermone AR, Nguyen TL, Wong TY, Schwarzenbacher R, Schmidt J, Lane D, McGrath C, Turk BE, Burnett J, Aman MJ, Little S, Sausville EA, Zaharevitz DW, Cantley LC, Liddington RC, Gussio R, and Bavari S. Identification of small molecular inhibitors of anthrax lethal factor. *Nat Struct Mol Biol*, 11:67–72, 2004.

[154] Forino M, Johnson S, Wong TY, Rozanov D, Savinov AY, Li W, Fattorusso R, Becattini B, Orry AJ, Abagyan RA, Smith JW, Alibek K, Liddington RC, Strongin AY, and Pellecchia M. Efficient synthetic inhibitors of anthrax lethal factor. *PNAS*, 102(27):9499–9504, 2005.

[155] Tonello F, Naletto L, Romanello V, Dal Molin F, and Montecucco C. Tyrosine-728 and Glutamic acid 735 are essential for the metalloproteolitic activity of the lethal factor of Bacillus anthracis. *Biochem Biophys Res Commun*, 131:496–502, 2004.

[156] **(a)** Magistrato A, DeGrado WF, Laio A, Rothlisberger U, VandeVondele J, Klein ML. Characterization of the dizinc analogue of the synthetic diiron protein DF1 using ab initio and hybrid quantum/classical molecular dynamics simulations.*J Phys Chem B* 107:4182-4188, 2003. **(b)** Dal Peraro M, Vila AJ, Carloni P. Structural determinants and hydrogen-bond network of the mononuclear zinc(II)-β-lactamase active site.*J Biol Inorg Chem* 7:704-712, 2002. **(c)** Gervasio FL, Schettino V, Mangani S, Krack M, Carloni P, Parrinello M. Influence of outer-shell metal ligands on the structural and electronic properties of horse liver alcohol dehydrogenase zinc active site. *J Phys Chem B* 107:6886-6892, 2003.

[157] Klimpel KR, Arora N, and Lepla SH. Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for lethal toxin activity. *Mol Microbiol*, 13 (6):1093–1100, 1994.

[158] Tozzini V. Coarse grained models for proteins. *Curr Opin Struct Biol*, 15:144–150, 2005.

[159] Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, and Jones DT. Protein structure prediction servers at University College London. *Nucl Acids Res*, 33:W36–W38, 2005.

[160] Vullo A, Bortolami O, Pollastri G, and Tosatto SCE. Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucl Acids Res*, 34:W164–W168, 2006.

[161] Guermeur Y, Geourjon C, Gallinari P, and Deleage G. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, 15 (5):413–421, 1999.

[162] Piana S, Carloni P, and Parrinello M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 Protease. *J Mol Biol*, 319:567–583, 2002.

[163] Trouiller N and Martins JL. Efficient pseudopotentials for plane-wave calculation. *Phys Rev B*, 43:1993–2006, 1991.

[164] Kleinman L and Bylander DM. Efficacious Form for Model Pseudopotentials. *Phys Rev Lett*, 48:1425–1428, 1982.

[165] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behaviour. *Phys Rev A*, 38:3098–3100, 1998.

[166] Lee C, Yang W, and Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B*, 37:785–789, 1988.

[167] Barnett RN and Landman U. Born-Oppenheimer molecular-dynamics simulations of finite systems: Structure and dynamics of (H2O)2. *Phys Rev B*, 48:2081–2097, 1993.

[168] **(a)** Surez D, Daz N, Merz KM Jr. Molecular dynamics simulations of the mononuclear zinc-β-lactamase from Bacillus cereus complexed with Benzylpenicillin and a quantum chemical study of the reaction mechanism.*J Am Chem Soc* 123:9867-9879, 2001. **(b)** Suarez D, Diaz N, Merz KM Jr. Molecular dynamics simulation of the dinuclear zinc-β-lactamase from Bacteroides fragilis complexed with imipenem.*J Comp Chem* 28:1587-1600, 2002. **(c)** Dal Peraro M, Villa AJ, Carloni P. Substrate binding to Mononuclear metallo-β-lactamase from Bacillus cereus. *Proteins* 54:412-423, 2004.

[169] Cheatham TE III, Miller JL, Fox T, Darden TA, and Kollman PA. Molecular dynamics simulations on solvated biomolecular systems: the Particle Mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *J Am Chem Soc*, 117:4193–4194, 1995.

[170] Ryckaert JP, Ciccotti G, and Berendsen HJC. Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comp Phys*, 23:327–341, 1977.

[171] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, and Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 81:3684–3690, 1984.

[172] **(a)** Laxmikant K, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: greater scalability for parallel molecular dynamics.*J Comp Phys* 151:283-312, 1999. **(b)** Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD.*J Comput Chem* 26(16):1781-1802, 2005.

[173] Doi M. *Introduction to Polymer Physics*. Oxford: Oxford University Press, 1996.

[174] Lindahl E, Azuara C, Koehl P, and Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom Normal Mode Analysis. *Nucleic Acids Res*, 34:W52–W56, 2006.

[175] Bank R and Holst M. A new paradigm for parallel adaptive meshing. *SIAM J. Sci. Comput.*, 22:1411–1443, 2000.

[176] Dolinsky T, Nielsen J, McCammon A, and Baker N. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatic calculations. *Nucleic Acid. Res.*, 32:W665–W667, 2004.

[177] . URL `http://www-personal.umich.edu/$\sim$mlerner/PYMOL/`.

[178] Jedrzejas M. The structure and function of novel proteins of Bacillus anthracis and other spore forming bacteria: Development of novel prophylactic and therapeutic agents. *Critical Rev Biochem Mol Biol*, 37:339–373, 2002.

[179] Piana S, Bucher D, Carloni P, and Rothlisberger U. Reaction mechanism of HIV protease by hybrid Car-Parrinello/Classical MD simulations. *J Phys Chem B*, 108:11139–11149, 2004.

[180] Perryman AL, Lin JH, and McCammon JA. HIV-1 Protease Molecular Dynamics of a Wild-Type and of the V82F/I84V Mutant: Possible Contributions to Drug Resistance and a Potential New Target Site for Drugs. *Protein Sci*, 13:1108–1123, 2004.

[181] Dameron CT and Harrison MD. Mechanism for protection against copper toxicity. *Am J Clin Nutr*, 67:1091S–1097S, 1998.

[182] Halliwell B and Gutteridge JM. Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochem J*, 219(1):1–14, 1984.

[183] Rae TD, Schmidt PJ, Pufahl RA, Culotta VC, and O'Halloran TV. Undetectable intracellular free copper: the requirement of a copper chaperone for superoxide dismutase. *Science*, 284:805–808, 1999.

[184] Ermakova E. Brownian dynamics simulation of the competitive reactions: binase dimerization and the association of binase and barstar. *Biophys Chem*, 130(1–2):26–31, 2007.

[185] Lüdemann SK, Lounnas V, and Wade R. How do substrates enter and products exit the buried active site of citochrome P450? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J Mol Biol*, 303(5):813–830, 2000.

[186] Arnesano F, Banci L, Bertini I, Huffman DL, and O'Halloran TV. Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1. *Biochemistry*, 40:1528–1539, 2001.

[187] Bueno M and Camacho CJ. Acidic groups docked to well defined wetted pockets at the core of the binding interface: a tale of scoring and missing protein interactions in CAPRI. *Proteins*, 69:786–792, 2007.

[188] Ritchie D. Recent progress and future directions in protein-protein docking. *Current protein and peptide science*, 9:1–15, 2008.

[189] Arnesano A, Banci L, Bertini I, Cantini F, Ciofi-Baffoni S, Huffman DL, and O'Halloran TV. Characterization of the binding interface between the copper chaperone Atx1 and the first cytosolic domain of Ccc2 ATPase. *J Biol Chem*, 276:41365–41376, 2001.

[190] MacKerell AD, Feig M Jr, and Brooks III CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in moleculat dynamics simulations. *J Comput Chem*, 25:1400–1415, 2004.

[191] Fuchs JF, Nedev H, Poger D, Ferrand M, Brenner V, Dognon JP, and Crouzy S. New model potentials for sulfur-copper(I) and sulfur-mercury(II) interactions in proteins: from ab initio to molecular dynamics. *J Comput Chem*, 27(7):837–856, 2006.

[192] Cornell WD, Cieplak P, Bayly CL, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*, 117:5179–5197, 1995.

[193] Op't Holt B and Merz KM Jr. Insights into Cu(I) exchange in HAH1 using quantum mechanical and molecular simulations. *Biochemistry*, 46:8816–8826, 2007.

[194] Darden T, York D, and Pedersen L. Particle Mesh Ewald: an N-log(N) method for Ewald sum in large systems. *J Chem Phys*, 98:10089–10092, 1993.

[195] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG. A smooth particle mesh Ewald potential. *J Chem Phys*, 103:8577–8592, 1995.

[196] Tanford C and Kirkwood JG. Theory of protein titration curves. I. General equations for impenetrable spheres. *J Am Chem Soc*, 79:5333–5339, 1957.

[197] Matthew JB and Gurd FRN. Calculation of electrostatic interactions in proteins. *Methods Enzymol*, 130:671–679, 1986.

[198] Warwicker J and Watson HC. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J Mol Biol*, 157:671–679, 1982.

[199] Dal Peraro M, Ruggerone P, Raugei S, Gervasio FL, and Carloni P. Investigating biological systems using first principles car-parrinello molecular dynamics simulations. *Current Opinion in Structural Biology*, 17(2):149–156, 2007.

[200] Cox DW and Moore SDP. Copper transporting P-type ATPases and human disease. *J Bioenergetics and Biomembranes*, 34(5):333–338, 2002.

[201] Hoops SC, Anderson KW, and Merz KM Jr. Force field design for metalloproteins. *J Am Chem Soc*, 113:8262–8270, 1991.

[202] Miyamoto S and Kollman PA. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J Comp Chem*, 13:952–962, 1992.