UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

DIPARTIMENTO DI CHIMICA

Doctoral School in Industrial Chemistry - XXXII Cycle

A COMPUTER AIDED EXCURSUS INTO MOLECULAR BIOSYSTEMS

Francesco OLIVA

Tutor: Dr. Stefano PIERACCINI

Coordinator: Prof. Dominique ROBERTO

— Academic year 2018/2019 —

Contents

Acronyms

AMA	Aspergillomarasmine A.
aMD	Accelerated Molecular Dynamics.
AR	Antimicrobial Resistance.
CAS	Computational Alanine Scanning.
CVs	Collective Variables.
DNA	Deoxyribonucleic Acid.
EcoGGT	E. Coli GGT.
GDP	Guanosine Diphosphate.
GGT	γ-glutamyltransferase.
GTP	Guanosine Triphosphate.
MBLs	Metallo-β-lactamases.
MD	Molecular Dynamics.
MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area.
MTD	Metadynamics.
PBC	Periodic Boundary Conditions.
PPIs	Protein Protein Interactions.
PSA	Path Similarity Analysis.
QM/MM	Quantum Mechanics/Molecular Mechanics.
RAMD	Random Acceleration Molecular Dynamics.
RMSD	Root Mean Square Deviation.
SASA	Solvent Accessible Surface Analysis.

tRNA Transfer Ribonucleic Acid.

UAAs Unnatural Ammino Acids.

WTMD Well Tempered Metadynamics.

Thesis Overview

The application of computational methods to study biological and biochemical processes is a growing field in both industrial and academical environment. They can be used to explore condition that would be extremely expensive or flat out impossible to reproduce in a laboratory; to filter out molecules that would be unlikely to react with the designed target, thus reducing the price needed to produce a new drug; to understand experimental data by looking at the behaviour of molecules and proteins at an atomistic level. In this PhD thesis, molecular modeling, molecular dynamics and different enhanced sampling methods are employed to address different aspects of relevant biosystems.

- Chapter 1 is a theoretical introduction to the principles of molecular dynamics and the different approaches employed in the following chapters. It also explains the strengths and weaknesses of each method and how are they relevant in their respective fields of application.

- Chapter 2 focuses on the interaction between Aspergillomarasmine-A and the zinc ion of a Metallo-β-lactamases. Metallo-β-lactamases are a relevant target in fighting bacterial resistance due to their ability to inactivate β-lactam rings that are the core fragment in penicillin like antibiotics. To study this process we had to develop a method that was able to overcome the limitations of the force fields used to simulate the protein and the ligand, which were not able to correctly predict the coordination geometry. We could generate different conformations using well tempered metadynamics and filter them in order to obtain, after a quantomechanical minimization the correct structures.

- In Chapter 3 six nucleosides obtained by modifying the phosphate sidechain of guanosine triphosphate and their ability to interfere with the microtubule dynamics is analyzed. This study is carried out by calculating the binding free energy of the six nucleosides with the β- of tubuline. We also investigated which aminoacids are more important in the process by calculating the binding free energy difference upon mutation for each of the 36 aminoacids in the binding site.

- In Chapter 4 is described the conformational study of a small peptide containing a non natural aminoacid. We simulated the two possible enantiomers and both the possible rotamers of one of the two enantiomer with well tempered metadynamics to study the ability of the new aminoacid to act as a mimetic of a β-turn. To better understand the small peptides behaviour we also ran simulations where the NMR data were considered.

- In Chapter 5 the protein protein interactions between two monomer of the universal protein YeaZ which is composed by five identical monomer were investigated. We started by running a Solvent Accessible Surface Area analysis to identify which residues were located in the protein protein interaction surface. We then proceded to calculate which residues were important to the binding by performing a computational alanine scanning which allowed us to identify a small peptide that, in our simulations, proved to be able to bind in the interaction site. We then tried to improve the stability of the peptide by stiffening the helix.

- The final chapter describes the work carried on in collaboration with Prof. Birgit Schiøtt during my staying at Aarhus University. Here we employed accelerated molecular dynamics to simulate the binding of glutamine to γ-glutamyltransferase. We ran 90 simulations which resulted in three binding events. By analyzing these three simulations we shed light on the residues that are more important during the binding process, recognized four general configurations that are common in the successful simulations and elucidated the importance of a loop known as lid-loop in the recognition of the substrate.

Chapter 1

Theoretical Introduction

1.1 Molecular Dynamics

Molecular Dynamics (MD) [1] is a method firstly developed by Fermi, Pasta, Ulam and Tsingou [2] in the mid 1950s. In its classical form it consists in the resolution of Newton's laws of motion to calculate the force (F) acting on a system, starting in a specific position (R), with specific velocities (V). F is calculated as the negative gradient of the potential energy function $U(R)$.

$$F(R) = -\nabla U(R) \tag{1.1}$$

Equation 1.1 is iterated recursively on each particle of the system, at each step to calculate the new set of positions ($\vec{r}_i$) and velocities ($\vec{v}_i$).

$$\begin{aligned} \frac{d\vec{v}_i}{dt} &= \frac{1}{m_i}\vec{f}_i \\ \frac{d\vec{r}_i}{dt} &= \vec{v}_i \end{aligned} \tag{1.2}$$

Where $\vec{f}_i$ is the force applied to the i-th atom and m_i is its mass. In a real case the starting positions of every atom $\vec{r}_i(0) = \vec{r}_{0i}$ are reported in a file containing the crystal structure of the protein, usually obtained by X-Ray studies and uploaded on the Protein Data Bank. The starting velocities $\vec{v}_i(0) = \vec{v}_{0i}$ are assigned according to a Boltzmann distribution. A standard MD simulation will consists in many iterative steps, presented in Figure 1.1.

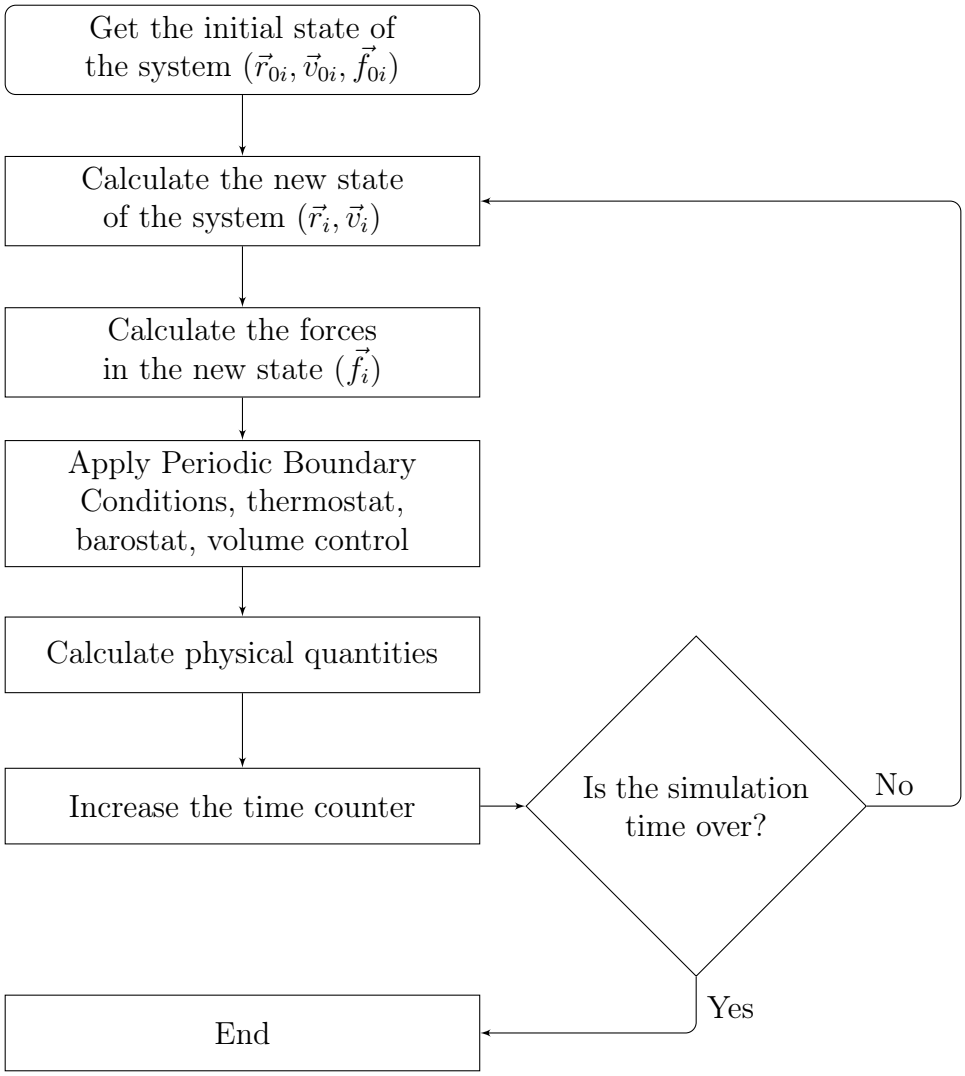


Figure 1.1: Generic algorithm of a MD simulation.

1.1.1 Integrators

To calculate the new state of the system (second step of Figure 1.1), is a Cauchy problem which is impossible to solve analytically but easy to solve using numerical methods. There are many methods that have been developed over the years to solve the problem and they are called integrators [3, 4]. The integrator used in this PhD thesis is called "leapfrog" [5] which calculates the positions and velocities as follows:

$$\begin{aligned} \vec{r}_i(t + \Delta t) &= \vec{r}_i(t) + \vec{v}_i \left(t + \frac{\Delta t}{2} \right) \Delta t \\ \vec{v}_i \left(t + \frac{\Delta t}{2} \right) &= \vec{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{\vec{f}_i(\vec{r}_i(t))}{m_i} \Delta t \end{aligned} \tag{1.3}$$

First the new positions are calculated, then the new velocities are calculated based on the new positions. It is important to note that the leapfrog integrator is not a self starting algorithm, which means that for the first step it is necessary to use another

integrator, such as Eulero's [3]. The choice of the right Δt is critical for the correct execution of the MD simulation and it is related to the chosen integrator. Since most of the time we are not interested in studying the motion of bonded atoms, which would require a time step of 1 fs we constrain all bond lengths to their equilibrium distances using the LINCS algorithm [6], which allows to use a time step equal to 2 fs.

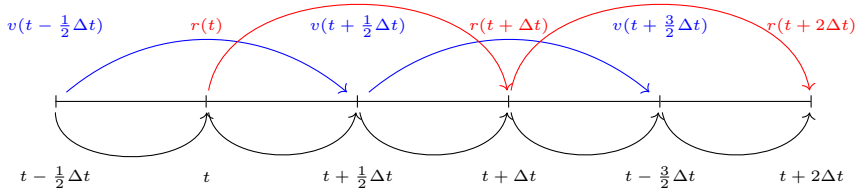


Figure 1.2: The graphical representation of the leapfrog integrator.

1.1.2 Simulating the real world

A MD simulation generates a sequence of points in the phase space. These points are connected in time and from this trajectory, ensemble averages and other properties can be calculated. An ensemble average is an average taken over a large number of replicas considered simultaneously and is defined as

$$\langle A \rangle = \frac{\int \mathrm{d}\vec{\mathrm{p}}^N \, \mathrm{d}\vec{\mathrm{r}}^N \;\; \exp(-\beta H(\vec{p}^N \vec{r}^N)) \, A(\vec{p}^N \vec{r}^N)}{\int \mathrm{d}\vec{\mathrm{p}}^N \, \mathrm{d}\vec{\mathrm{r}}^N \;\; \exp(-\beta H(\vec{p}^N \vec{r}^N))} \tag{1.4}$$

Where H is the Hamiltonian of the classical system, $\langle A \rangle$ can be also calculated as a time average thanks to Birkhoff's ergodic theorem [7] "The long time average is equal to the ensemble average in the limit of an infinite number of members in the ensemble".

$$\langle A \rangle = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} A(\vec{r}(t_i)), \tag{1.5}$$

Where n is the number of snapshots in the trajectory. The motion of the system through phase space is governed by the Hamiltonian which is defined in two components, where the first one depends on the momenta while the second one depends on particles positions.

$$\begin{aligned} H(\vec{r}, \vec{p}) &= K(\vec{p}) + U(\vec{r}) \\ K(\vec{p}) &= \sum_{i}^{N} \frac{\vec{p}_i^{\,2}}{2m_i} \end{aligned} \tag{1.6}$$

The potential energy function $U(\vec{r})$ can be written as the sum of the interactions between pairs, triplets and so on of particles.

$$U(\vec{r}) = \sum_{ij} U_{ij}(\vec{r}_i, \vec{r}_j) + \sum_{ijk} U_{ijk}(\vec{r}_i, \vec{r}_j, \vec{r}_k) + ... \tag{1.7}$$

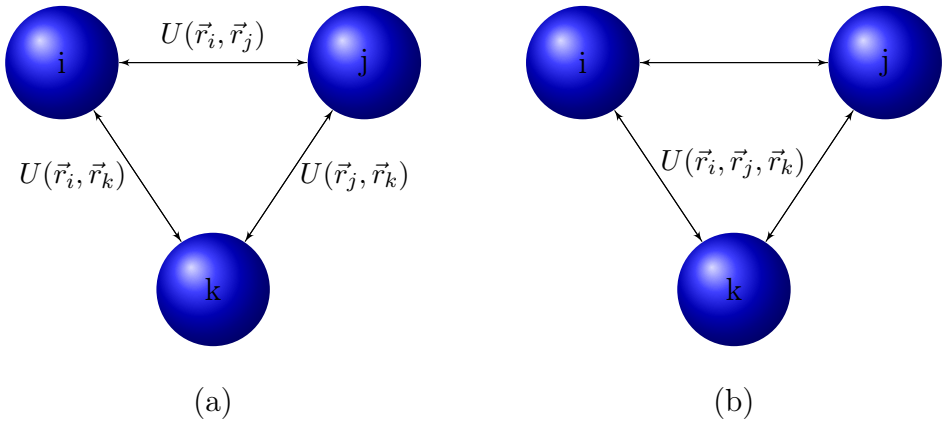


Figure 1.3: Pair interaction (a) and triplet interaction (b).

In simulations only pair terms are considered, since they dominate the interaction, so Equation 1.7 is truncated at the first term and the potential energy function is replaced with an "effective potential" which incorporate the many-body effect as an average into the first term.

$$U(\vec{r}) \approx \sum_{ij} U_{ij}^{eff}(\vec{r}_i, \vec{r}_j) \tag{1.8}$$

The potential energy function ($U(\vec{r})$) can be implemented in many ways and it depends on the force field used. The term "force field" refers to the specific functional form and set of parameters used to calculate $U(\vec{r})$. The typical $U(\vec{r})$ expression is

$$U(\vec{r}) = \underbrace{U_{str} + U_{bend} + U_{tors} + U_{cross}}_{\text{bonded terms}} + \underbrace{U_{vdW} + U_{elec}}_{\text{non-bonded terms}} \tag{1.9}$$

Where *str*, *bend*, *tors*, *cross*, *vdW* and *elec* means *stretching*, *bending*, *torsional*, *cross correlation*, *van der Waals* and *electrostatic* respectively. U_{tors} and U_{cross} are not paired terms since they involve the interaction between more than two atoms but they are usually added to correctly simulate tridimensional structure of molecules. A specific implementation is the AMBER force field [8] and the potential energy

function is calculated as follows

$$U(\vec{r}) = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[cos(n\phi+\delta)+1] + \sum_{atom}\sum_{i}\sum_{i>j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i}\sum_{i>j}\frac{q_i q_j}{4\pi\epsilon r_{ij}} \tag{1.10}$$

The summations are used to calculate the single terms of Equation 1.9 and are in the same order. The fourth term, used to calculate U_{vdW} describes the van der Waals interaction between two non bonded atoms using a Lennard-Jones potential. The parameters k_b, k_θ and k_ϕ are optimized to reproduce experimental data and quantum mechanical calculations.

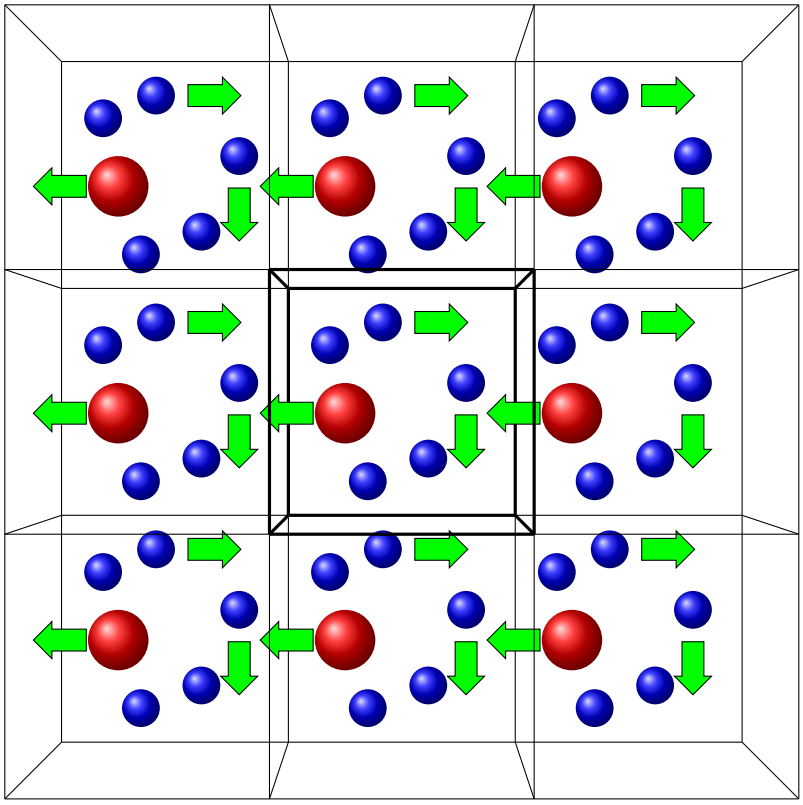


Figure 1.4: Graphical representation of the concept of Periodic Boundary Conditions.

To replicate reality, the system should contain a number of atoms of the same order of magnitude of Avogadro's number, which is currently impossible. On the other hand, a MD simulation with even hundreds of thousands of atoms would not replicate correctly the properties of any system. To overcome this problem, the concept of Periodic Boundary Conditions (PBC) is applied (Figure 1.4). When PBC are implemented, the system is replicated identical in three dimension and when a particle leaves the central box from a side, it is replaced by an replica

coming from the opposite side of the box. We should, in theory, calculate the interaction between all the infinite particles. To avoid this problem, the *minimum image convention* is employed so that every particle can interact only with a single image of every other particle. This is realized by centering a box of the same shape and dimension of the original one over the particle of which we want to calculate the interactions. To further reduce the computational cost of the simulation, usually a cutoff is implemented as well and the interactions above the threshold are not calculated. The cutoff also prevents the particles to interact with their own periodic images since the upper limit for the cutoff distance is half the length of the box's shortest side.

1.2 Protein Protein Interaction

Proteins are responsible for a large array of functions within organisms, from transporting molecules to responding to external and internal stimuli, catalysing reactions, help to provide structure to cells and many more. However, proteins rarely act alone. They tend to interact with each others and create molecular machines able to perform complicated tasks. For example, microtubules are an aggregation of thousands of dimers of tubulin and are essential in cell division, which makes them important targets for anti-cancer drugs [9]. Protein Protein Interactions (PPIs) play a crucial role in the well being of organisms and to correctly accomplish their biological task the contacts between proteins need to be formed with specific patterns, at the right time and locations [10]. Any kind of deviation from the correct PPIs network can lead to diseases, including cancer [11]

1.2.1 Importance of protein protein interaction

Being involved in so many processes inside the organism, 130000 binary interactions at any give time [12], PPIs are a natural candidate to become a focal target in drug design. For example, tubulin aggregation has been shown to be a relevant target against different diseases, cancer and Alzheimer's included [13, 14]. However PPIs mechanisms are quite different from those involved in enzymes inhibition so therapeutic molecules can not be designed in the same way. For example, every PPIs must be studied as a single case since there is not a conserved binding site like in the case enzymes of the same family [15], even between similar interactions. Furthermore there are usually many residues involved in the interaction. Even if the interaction surface is usually large, it has been shown that a ligand does not need to cover the whole surface but only a limited number of key aminoacids called "hotspots" [16, 17] this shows that not all the residues situated on the interaction surface are relevant. The hotspots are a subsection of the interaction surface that

can be used to strengthen or weaken the complex formation. To find out which residues are relevant there are two steps involved: firs, the Solvent Accessible Surface Analysis (SASA) is calculated, which gives information on all the residues that are actually on the interaction surface by calculating the the change of area accessible to the solvent in the bound state and in the monomers. Second is the alanine scanning, that can be performed both computationally or experimentally. In this step the residues on the interaction surface are mutated to alanine and the difference in the binding free energy is measured. Molecules able to interfere with the PPIs network, in both stabilizing the bound complex or destroying the key interaction that keep the network togheter, are very interesting for therapeutic approaches. The main classes of PPIs modulators are humanized monoclonal antibodies, peptides and peptidomimetics. Humanized monoclonal antibodies are the most commons due to their very high target specificity even if they have difficulties to go through the hematoencephalic and cell barriers [18]. This process is usually the first step to develop a peptide able to act as a drug, and one example of this will be discussed in this thesis (Chapter 5), the sequence is usually selected from the contact surface between protein after an analysis on the hotspots. Since peptides suffer of low bioavailabilty and poor metabolic stability the interest in peptidomimetics is another topic of rising interest that will be discussed in Chapter 4.

1.3 Free energy calculations

Gibbs free energy (G) gives important informations about structure and stability of biological systems. Over the years, many different methods have been developed to compute G [19], in this thesis the Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) [20] has been employed since it is a good compromise between the methods with low accuracy that are computationally inexpensive, such as the score functions methods [21, 22] and the ones with high accuracy but that requires a lot of computational time like thermodynamic integration [23]. To calculate the free energy variation (ΔG) between two states, like a bound and unbound states of a protein-ligand complex, one should simulate multiple times the whole process and then average all the ΔG values. This is not affordable from a computational point of view, so another way has to be used. MM-PBSA calculates the binding free energy variation ($\Delta G_{binding}$) as the difference between the ensemble average of the free energy of the complex and the summation of the free energy of the monomers. In the case of ligand-protein interaction

$$\Delta G_{binding} = \langle G_{complex} \rangle - (\langle G_{protein} \rangle + \langle G_{ligand} \rangle) \tag{1.11}$$

To obtain $\langle G_{compelx} \rangle$, $\langle G_{protein} \rangle$ and $\langle G_{ligand} \rangle$, strictly speaking, three different simulation should be necessary: one for the complex, one for the protein and on for the ligand. It is much more common to only simulate the complex and create the ensemble average of the free receptor and ligand by simply removing the appropriate atoms so Equation 1.11 becomes

$$\Delta G_{binding} = \langle G_{compelx} - (G_{protein} + G_{ligand}) \rangle_{complex} \tag{1.12}$$

G is estimated as summation of different terms

$$G = E_{bound} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS \tag{1.13}$$

The first three terms are standard MM energy terms: bonded, electrostatic and van der Waals, respectively. G_{pol} and G_{np} are, respectively, the polar and non polar contributions to the free energy which are usually obtained by using a Generalized Born model. T and S are the temperature and entropy contribution. By using only one trajectory the E_{bound} cancels out, which reduces the error. On the other end, binding induced conformational changes are ignored [24].

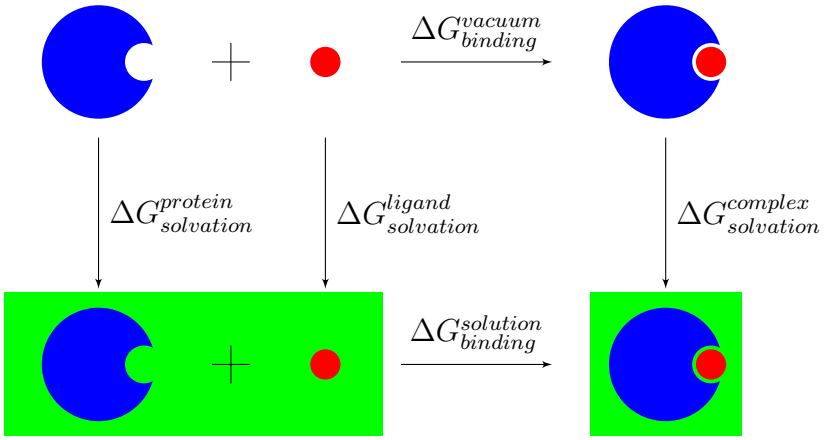


Figure 1.5: Thermodynamic cycle used to calculate ΔG.

Figure 1.5 is a graphical representation of the thermodynamic cycle used to calculate $\Delta G^{solution}_{binding}$ which is the binding free energy of the system in solution. The single terms are considered additives and $\Delta G^{solution}_{binding}$ is calculated as shown in Equation 1.14

$$\Delta G^{solution}_{binding} = \Delta G^{vacuum}_{binding} + \Delta G^{complex}_{solvation} - (\Delta G^{protein}_{solvation} + \Delta G^{ligand}_{solvation}) \tag{1.14}$$

1.3.1 Computational alanine scanning

The Computational Alanine Scanning (CAS) is a method designed to substitute the experimental one, which is a long procedure that requires DNA engineering, protein expression and purification. Luckily there is the *in silico* alternative, which allows to evaluate the binding free energy between subunits of the wild type complex (ΔG_{WT}), the mutated ones (ΔG_{mutant}) and then calculate the variation for each mutant ($\Delta\Delta G_{mutant}$).

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{WT} \tag{1.15}$$

There are three main approaches to perform CAS and the difference lies in how

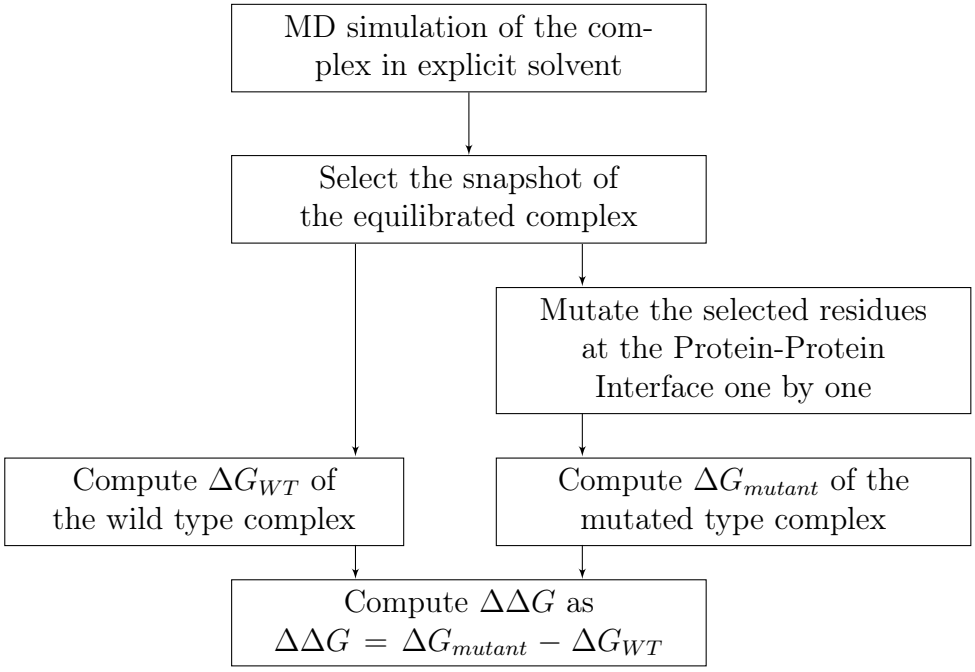


Figure 1.6: Flowchart showing the steps involved in performing the CAS.

many trajectories they uses: one, two or four. When only one trajectory is used, the wild type complex is simulated and then monomers are obtained by discarding the counterpart. The mutant proteins are obtained by post processing the original trajectory. When two trajectories are used, one is the wild type and the other one is the mutated complex. The monomers are still obtained by discarding the counterpart. In the four trajectories case, two trajectories are used to simulate the mutated and wild type and the other two simulates the monomers. In this thesis the single trajectory approach has been followed since it has been shown that it gives better results thanks to an error cancellation and the possibility to neglet the entropic terms when calculating the $\Delta\Delta G$, if four assumption are valid. The first assumption is that a point mutation in the protein does not affect significantly the

tertiary structure as shown in literature [25, 26]. Only the proteins are considered for the calculation, any metal ion, water molecule or cofactors are not taken into account when calculating the binding free energy. The terms used to calculate the free energy (ΔG) are additive. Proline residues are not mutated since this would significantly alter the backbone structure, usually also glycine are ignored.

1.3.2 Solvation terms

The solvation contribution is divided in two terms: polar and apolar. They can be both calculated using the APBS software [27].

$$\Delta G_{solvation} = \Delta G_{polar} + \Delta G_{apolar} \tag{1.16}$$

The calculation of the term is split in different steps as shown in Figure 1.7. Step one

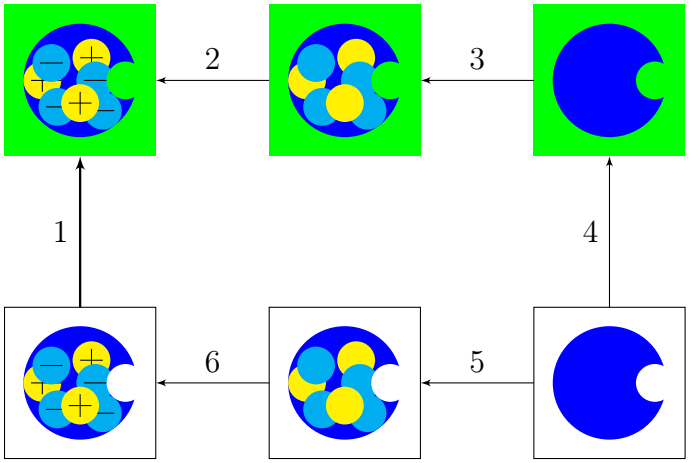


Figure 1.7: Thermodynamic cycle used to calculate $\Delta G_{solvation}$.

is the solvation energy that we want to calculate. Step two represent the charging of the solute in solution: in case there are ions present or there is an inhomogeneous dielectric they are accounted in this step. Step three represents the introduction of attractive solute-solvent dispersive interaction interactions. The fourth step introduces the repulsive solute-solvent interactions (e.g. cavity formation). Step five is used to offset unwanted energies added in Steps three and four. Finally, in step six charges are added to the solute in a vacuum or homogeneous dielectric environment in the absence of mobile ions.

Polar Solvation

The full free energy cycle is usually decomposed into polar and nonpolar parts. The polar portion is usually represented by the charging energies in steps two and six

(Figure 1.7)

$$\Delta G_{polar} = \Delta G_2 - \Delta G_6 \tag{1.17}$$

Energies returned from APBS electrostatics calculations are charging free energies. Therefore, to calculate the polar contribution to the solvation free energy, we simply need to setup two calculations corresponding to steps two and six in the free energy cycle. Note that the electrostatic charging free energies returned by APBS include self-interaction terms. These are the energies of a charge distribution interacting with itself. Such self-interaction energies are typically very large and extremely sensitive to the problem of how the charges are discretized (grid spacing, location, etc.). Therefore, it is very important that the two calculations in steps two and six are performed with identical grid spacings, lengths, and centers, in order to ensure appropriate matching (or "cancellation") of self-energy terms [28].

Apolar Solvation

The nonpolar solvation free energy is usually represented by the energy changes in steps three through five

$$\Delta G_{apolar} = \Delta G_4 - (\Delta G_3 + \Delta G_5) \tag{1.18}$$

where step four represents the energy of creating a cavity in solution and steps three to five is the energy associated with dispersive interactions between the solute and solvent. The apolar contribution, ΔG_4 is calculated as

$$\Delta G_4 = pV + \gamma A \tag{1.19}$$

where p is the solvent pressure, V is the solute volume, γ is the solvent surface tension, and A is the solute surface area. The dispersion terms (steps three and five) follow a Weeks-Chandler-Anderson framework [29]

$$\Delta G_3 - \Delta G_5 = \overline{\rho} \int_{\Omega} \mathrm{dy}\ u^{(att)}(y)\theta(y) \tag{1.20}$$

where $\overline{\rho}$ is the bulk solvent density, Ω is the problem domain, $u^{(att)}(y)$ is the attractive dispersion interaction between the solute and the solvent at point y, and $\theta(y)$ describes the solvent accessibility of point y.

1.4 Enhanced Sampling Techniques

A MD simulation aims to sample the phase space of a system. Sometimes to obtain an acceptable sampling, where all the relevant conformational states are sampled

multiple times, in an affordable amount of time may not be possible. For example biological systems are known to have rough energy landscapes with many local minima separated by high energy barriers [30], limits of the force field or due to hardware limitations may all contribute to an unsuccessful sampling. Furthermore, large conformational changes or rare events may require a timescale that ranges between tenth of microseconds to seconds to occur when the standard time for a MD tops at few microseconds. For these reasons, different approaches to increase the sampling and decrease the amount of time required to sample a statistically meaningful part pf the phase space have been developed.

1.4.1 Metadynamics

Metadynamics (MTD) is a method proposed for the first time in 2002 by Alessandro Laio and Michele Parrinello [31]. It is the first algorithm that implemented memory in the sampling by locally increasing the value of the potential energy function in the conformation already sampled by introducing a Gaussian, meaning that a previously sampled conformation will be less likely to be sampled again since it has an artificially increased energy. It looks like slowly filling the free energy surface to allow the system to leave the local minimum as shown in Figure 1.8. This approach leads to a broader exploration of the free energy surface by pushing the system out of energy minima by adding a bias potential to the Hamiltonian that acts on a selected number of degrees of freedom, often referred as Collective Variables (CVs). At any given time t the biased potential U_G can be written as in Equation 1.21

$$U_G(S,t) = \int_0^t \mathrm{dt}'\ \omega \exp\left(-\sum_{i=1}^{d} \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2}\right) \tag{1.21}$$

Where S is a set of d functions of the microscopic coordinates R of the system, σ_i is the width of the Gaussian for the ith CV and ω is an energy rate, which is constant and expressed in terms of a Gaussian height W and a deposition stride τ_G [32]

$$\omega = \frac{W}{\tau_G} \tag{1.22}$$

and the free energy surface is defined as

$$F(S) = -\frac{1}{\beta} \ln\left(\int \mathrm{dR}\ \delta(S - S(R)) e^{-\beta U(R)}\right) \tag{1.23}$$

There are two major drawbacks that comes with MTD. The first is the need to pick a set of CV that correctly describes the system which may be very difficult and the second is that U_G does not converge to the free energy of the system but oscillates around the correct value, which leads to an error proportional to the square root of

the bias potential deposition rate. Another problem is the fact that there is a chance to push the system to explore configurations which are not physically meaningful. A good set of CVs need to satisfy three conditions:

- All the slow modes need to be included.
- It needs to be able to distinguish between the initial and final state of the system and every other relevant configuration in between.
- The set needs to be composed by few CVs.

Slow modes are those motions that can not be sampled adequately in the time of the simulation, while the fast ones are those motions expected to adapt rapidly when the slow modes change. If one of the slow degrees of freedom is not included, the system will not be able to pass the energy barrier on that direction of the free energy landscape. The method has received many improvements since it was initially

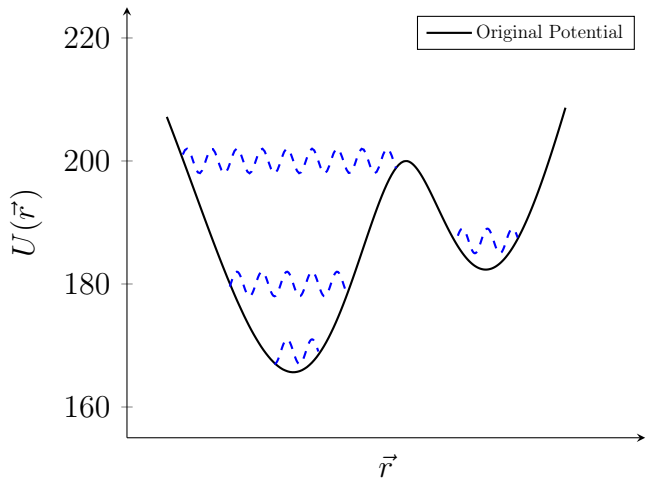


Figure 1.8: Graphical representation of MTD.

published, a very important one is the introduction of Well Tempered Metadynamics (WTMD) [33] which solves the problem of MTD not converging to the correct free energy value and the fact that in standard MTD, is not obvious when to stop the simulation. These improvements of WTMD over MTD lies are achieved reshaping the height of the deposed Gaussians during the course of the simulation. In WTMD the biased potential is expressed as follows

$$U(S,t) = k_B \Delta T \ln\left(1 + \frac{\omega N(S,t)}{k_B \Delta T}\right) \tag{1.24}$$

Where k_B is the Boltzmann's constant, ΔT is an input parameter with the dimension of a temperature and $N(S,t)$ is the histogram of the S variables collected during

the simulation and equal to

$$N(S,t) = \int_0^t \mathrm{dt}' \ \delta_{S,S(t)} \tag{1.25}$$

To relate Equation 1.24 to MTD it is sufficient to replace $\delta_{S,S(t)}$ in Equation 1.25 with a Gaussian. In practice the height of the Gaussian is reshaped modifying 1.22 with

$$W = \omega\tau_G e^{-\left(\frac{U_G(S,t)}{k_B\Delta T}\right)} \tag{1.26}$$

In this way, the bias deposition rate decreases as t^{-1} so that the dynamics of all the microscopic variables gets closer to thermodynamic equilibrium. The bias potential now converges to

$$U_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \tag{1.27}$$

Where C is an immaterial constant. In the limit of $\Delta T \to 0$, standard MD is recovered while if $\Delta T \to \infty$ the simulation is ran as MTD. Tuning ΔT regulates the exploration of the free energy surface and can avoid overfilling the minima and save computational time.
An example of CVs that will be used in Chapter 2 is the one called COORDINATION. This CVs can be used to calculate the number of contacts between two groups of atoms and is defined as

$$\sum_{i \in A} \sum_{i \in B} s_{ij} \tag{1.28}$$

Where A and B are the lists of atoms to calculate the distances, while s_{ij} is defined as

$$s_{ij} = \frac{1 - \left(\frac{\vec{r}_{ij} - d_0}{\vec{r}_{ij}}\right)^n}{1 - \left(\frac{\vec{r}_{ij} - d_0}{\vec{r}_{ij}}\right)^m} \tag{1.29}$$

d_0 is a user defined paramiter, if $\vec{r}_{ij} < d_0$ then $s(r) = 1$ while if $\vec{r}_{ij} > d_0$ the function decays smoothly to 0, n and m are two parameter, their standard values are 6 and $2n$ respectively.

1.4.2 Accelerated molecular dynamics

Accelerated Molecular Dynamics (aMD) was introduced for the first time by McCammon *et al.* in 2004 [34] based on earlier works on hyperdynamics [35, 36] and Steiner *et al.* work on bias potential [37]. This method applies a bias potential every time the calculated potential is below a certain threshold to help the system to leave

the minimum and explore other part of the potential energy landscape.

$$U^*(\vec{r}) = \begin{cases} U(\vec{r}), & U(\vec{r}) \geq E. \\ U(\vec{r}) + \Delta U(\vec{r}), & U(\vec{r}) < E. \end{cases} \tag{1.30}$$

Where $U^*(\vec{r})$ is the modified potential, $\Delta U(\vec{r})$ is the boost applied to the potential and E is the energy threshold.

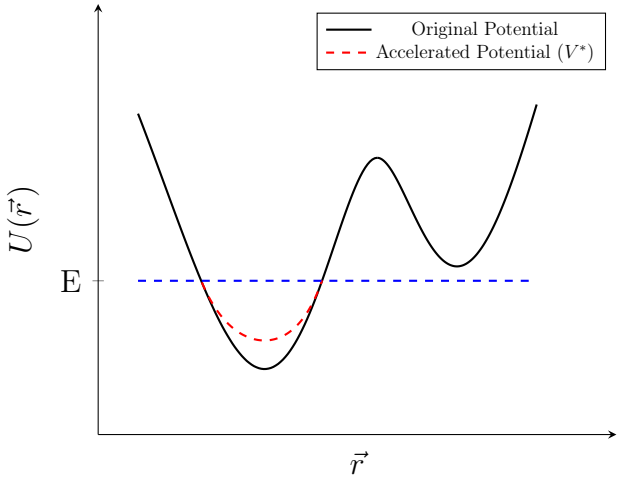


Figure 1.9: Graphical representation of how aMD modifies the potential to increase the escape rate of the system out of a minimum.

It can be proved that the method still yields to a correct canonical average of a quantity calculated with a normal potential. First Equation 1.4 is written in a more compact form:

$$\langle A \rangle = \frac{\int \text{dr } A(\vec{r}) \exp(-\beta U(\vec{r}))}{\int \text{dr } \exp(-\beta U(\vec{r}))} \tag{1.31}$$

Then the ensemble average calculated with then modified potential can be written as

$$\langle A^* \rangle = \frac{\int \text{dr } A(\vec{r}) \exp(-\beta U^*(\vec{r}))}{\int \text{dr } \exp(-\beta U^*(\vec{r}))} \tag{1.32}$$

Substituting for $U^*(\vec{r})$ from Equation 1.30

$$\langle A^* \rangle = \frac{\int \text{dr } A(\vec{r}) \exp(-\beta U(\vec{r}) - \beta \Delta U(\vec{r}))}{\int \text{dr } \exp(-\beta U(\vec{r}) - \beta \Delta U(\vec{r}))} \tag{1.33}$$

Reweighting the phase space of the modified potential by multiplying each configuration by the strength of the bias at each position $\exp(\beta \Delta U(\vec{r}))$ we obtain the

corrected ensemble average which is equivalent to Equation 1.31

$$\langle A^C \rangle = \frac{\int \text{dr } A(\vec{r}) \exp(-\beta U(\vec{r}) - \beta \Delta U(\vec{r})) \exp(\beta \Delta U(\vec{r}))}{\int \text{dr } \exp(-\beta U(\vec{r}) - \beta \Delta U(\vec{r})) \exp(\beta \Delta U(\vec{r}))} \tag{1.34}$$

The definition of the bias potential is a crucial point of the method since it has to smoothly reproduce the shape of the potential energy surface when the actual potential is below the threshold. $\Delta U(\vec{r})$ is chosen in such a way that the derivative of $U^*(\vec{r})$ has no discontinuity

$$\Delta U(\vec{r}) = \frac{(E - U(\vec{r}))^2}{\alpha + (E - U(\vec{r}))} \tag{1.35}$$

Where α is a tuning parameters that allows to calibrate how deep the modified

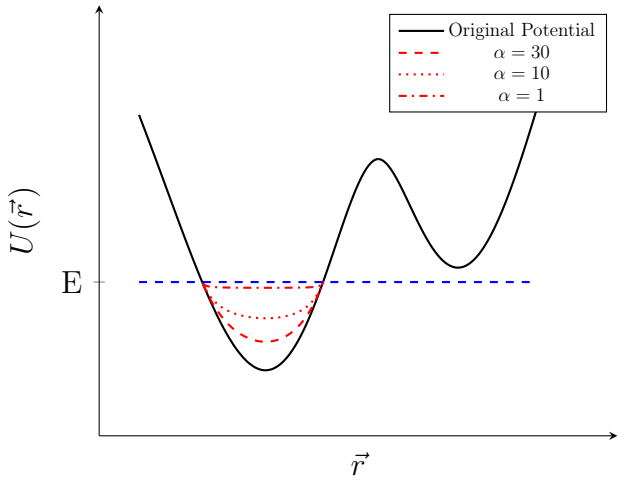


Figure 1.10: Impact of α on the modified potential.

potential energy basins needs to be and its effects are shown in Figure 1.10. If α is zero, then $U^*(\vec{r})$ is always equal to E and there is a flat surface instead of a minimum. Both α and E are important parameters and define how aggressive the acceleration will be. If E is high and α is low, the results will be a flat surface which will allow the system to freely explore all the points of the conformational space without restrictions. If E is too low, below the deepest minimum, the simulation will always be a normal MD. A further improvement of the method is the implementation of the "dual boost" approach where the potential $U(\vec{r})$ is calculated as the sum of two separate potentials

$$U(\vec{r}) = U_0(\vec{r}) + U_t(\vec{r}) \tag{1.36}$$

Where $U_t(\vec{r})$ refers to the torsional terms, while $U_0(\vec{r})$ is the potential that takes into account everything else. The modified potential $(U^*(\vec{r}))$ is given by the boost applied selectively to $U_t(\vec{r})$, which boosts the exploration of the possible conforma-

tions, plus a boost applied to the whole system, $U_T(\vec{r})$.

$$U_t^*(\vec{r}) = U_t(\vec{r}) + \Delta U_t(\vec{r}) \tag{1.37}$$

$$U^*(\vec{r}) = (U_0(\vec{r}) + U_t^*(\vec{r})) + \Delta U_T(\vec{r}) \tag{1.38}$$

This approach allows to overcome the limitations imposed by diffusive motion, such as a ligand diffusing through water molecules. Using the dual boost approach increases the diffusion of water molecules and solutes while preserving the overall water structure. The boost on the torsional terms allows to accelerate the sampling of different protein structure [38]. This approach is particularly suitable to simulate binding events of small molecules to protein, since it reduces significantly the time spent by the ligand in water, increases the chances of it leaving a non relevant pocket in case it falls in a local minimum. An application of this method will be presented in Chapter 6

Chapter 2

Metadynamics Study of Aspergillomarasmine-A as Metallo-β-Lactamases Inhibitor

2.1 Introduction

The World Health Organization defines Antimicrobial Resistance (AR) as "*The ability of a microorganism (like bacteria, viruses, and some parasites) to stop an antimicrobial (such as antibiotics, antivirals and antimalarials) from working against it. As a result, standard treatments become ineffective, infections persist and may spread to others*". The widespread use of antibiotics and their misuse [39] had triggered an evolutionary answer from microorganism that rapidly evolved the ability to resist to the most common drugs [40] making the treatment of patients difficult, costly, or even impossible. Patients with infections caused by bacteria resistant to a specific antibacterial drug generally have an increased risk of worse clinical outcomes and death, and consume more healthcare resources, than patients infected with the same bacteria not demonstrating the resistance pattern in question [41]. AR is a complex global public health challenge, and no single or simple strategy will sufficed to fully contain the emergence and spread of infectious organisms that become resistant to the available antimicrobial drugs. The development of AR is a natural phenomenon in microorganisms, and is accelerated by the selective pressure exerted by use and misuse of antimicrobial agents in humans and animals. The current lack of new antimicrobials on the horizon to replace those that become ineffective brings added urgency to the need to protect the efficacy of existing drugs [41]. A specific case of resistance is the carbapenem-resistant Gram-negative pathogens which acquired Metallo-β-lactamases (MBLs) enzymes that threatens the use of penicillin, cephalosporin and carbapenem antibiotics to treat infections [42].

2.1.1 Metallo-β-lactamases

β-lactamases production, relentlessly fueled by natural selection, is generally considered the primary route of resistance to β-lactam antibiotics. β-lactamases can be divided in two categories, based on their mechanism of β-lactam ring hydrolysis [43]. On one hand serine β-lactamases use a serine as active site residue that covalently attacks the β-lactam ring [44]. On the other hand MBLs (Figure 2.1) that use Zn^{2+} ion to activate a nucleophilic water molecule that opens the ring [45]. Several inhibitors of serine β-lactamase are known and available as co-drugs of β-lactam antibiotics [46] but no MBLs inhibitors are being tested in clinical trial due to the concerns about cross reactivity with human metal-based enzymes and the technical difficulty to develop a single inhibitors able to neutralise key clinically important MBLs [42]. A reason why these enzymes have proven difficult to study, from a MD point of view, is that the zinc ions are difficult to model with the common force fields since they can change their coordination geometry, oxidation state and connectivity [47]. A number of different force fields both polarizable [48, 49] and unpolarized [50–52] have been introduced to correctly simulate the metallo-organic interactions but since the properties of the metal ions are so deeply connected with the surrounding, the transferability of the parameters is a major limitation [47]. A class of methods used to approach these kind of system is the Quantum Mechanics/Molecular Mechanics (QM/MM) approach. This method allows to study most of the protein at a molecular mechanics level, while the active site, with the zinc ion and its surrounding is described with quantum mechanics.

2.1.2 Aspergillomarasmine A

Aspergillomarasmine A (AMA) (Figure 2.2) is a fungus-derived molecule that was discovered and reported in the early 1960s for its wilting and necrotic activity on plant leaves [53]. This molecule was re-evaluated in the 1980s as an inhibitor of angiotensin-converting enzyme [54] and in the early 1990s as a pre-clinical candidate for the inhibition of activation of human endothelin [55], a peptide that modulates blood vessel muscle contraction. The activation of this vasoconstricting peptide requires endothelin-converting enzyme, which, like angiotensin-converting enzyme, is a metalloproteinase that shares some mechanistic similarities with MBLs [42]. To better understand the inhibition mechanism that could be a direct interaction with the enzyme or a chelation of the Zn^{2+} ion by AMA we used MTD [31], WTMD [33] and QM/MM [56]. To investigate the coordination mechanism we developed our method and tested it on three other know chelating agents: Ethylenediaminetetraacetic acid (EDTA); Cyclohexanediaminetetraacetic acid (CDTA); Ethylenediamine-N, N'-disuccinic acid (EDDS) (Figure 2.3).

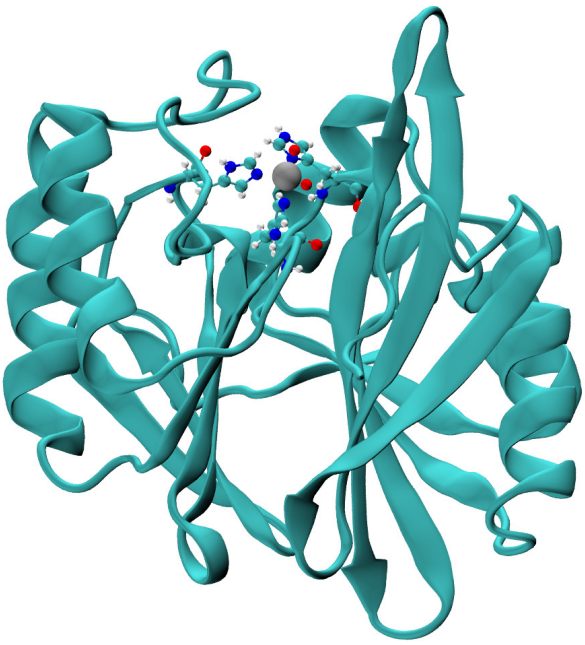


Figure 2.1: MBL from *Pseudomonas aeruginosa* (PDB code: 4D1U). The protein is shown using the New Cartoon representation, the zinc ion in silver and using the VDW representation (bigger sphere) and the three histidines and the two water molecules are shown using CPK representation.

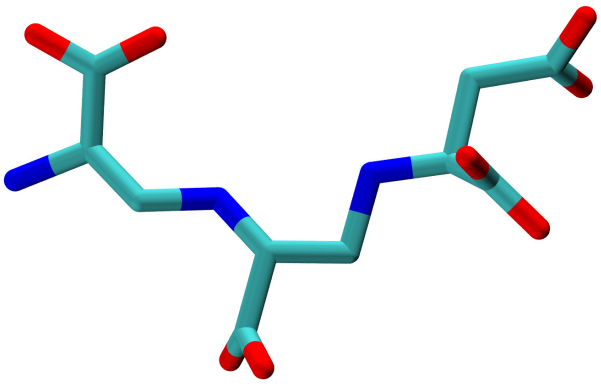
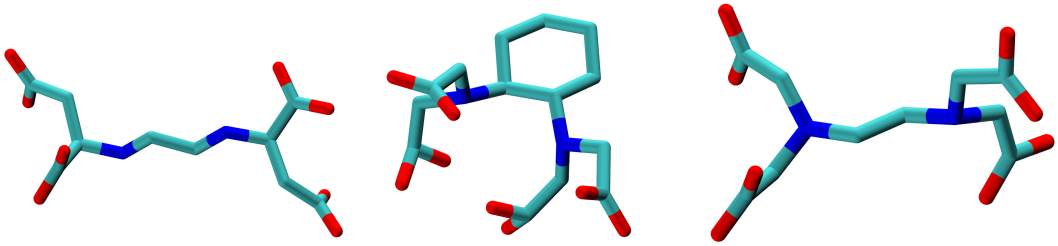


Figure 2.2: Aspergillomarasmine A. Hydrogen not shown



(a) (b) (c)

Figure 2.3: (a) EDTA, (b) CDTA, (c) EDTA. Hydrogen not shown

2.2 Computational Details

All our systems (EDTA, CDTA, EDDS, AMA and MBLs with AMA) were solvated with with TIP3P [57] water and then submitted to geometry optimization with the steepest descent algorithm with a convergence of 100 kJ·mol^{-1}·nm^{-1}. Then we performed a 1 ns NVT equilibration at 300 K, using velocity rescaling as temperature coupling algorithm and τ_T=0.01 ps. A 1 ns NPT equilibration followed, using the Berendsen pressure coupling algorithm with isotropic coupling, a reference pressure of 1 bar, τ_P=2 ps, and the reference isothermal compressibility of water: 4.5×10^{-5} bar^{-1}. After the equilibration phase, we ran 20 ns MTD on the ligands, during the development of the method and 20 ns WTMD simulations with a 8.0 kJ ·mol^{-1} bias-factor, at a reference temperature of 300 K. The force field GAFF [58] was used to model the ligands. The ligands geometry was optimized with Gaussian 09 (Revision D.01) [59] using Restricted Hartree-Fock with basis set 6-31G(d). Each optimized structure has been solvated in a cubic box with a side of 3 nm and solvated. The total number of atoms was $\approx$ 2600. Two CVs where used:

- The COORDINATION between the Zn^{2+} ion and the ligands polar atoms.
- The COORDINATION between the Zn^{2+} ion and the water molecules oxygen.

And the RATIONAL switching function, with $R_0 = 0.3$ and $D_{MAX} = 0.45$ for the ligands and $R_0 = 0.25$ and $D_{MAX} = 0.35$ for water molecules. Amber 99SB-ILDN [8] force field was used to model the MBLs. The system MBLs with AMA contained 51231 atoms, of which 3385 were protein atoms, 35 where AMA atoms and 9 Na ions were used to balance the overall charge rest where water atoms. The system was solvated in a cubic box with a side of 8 nm and then treated equilibrated like the others. We ran a 20 ns WTMD using as collective variable

- The COORDINATION between the Zn^{2+} ion and AMA polar atoms.
- The COORDINATION between the Zn^{2+} ion and the nitrogen of the three histidines (116, 118 and 196) which coordinate the zinc ion in the crystallographic structure.

And the RATIONAL switching function, with $R_0 = 0.3$ and $D_{MAX} = 0.45$ for the ligands and $R_0 = 0.25$ and $D_{MAX} = 0.35$ for histidines nitrogen atoms. Using the bash script reported in Appendix 2.A.3 we selected the middle structures of the cluster using the coordination number as filtering criteria. The selected structures where then optimized with Gaussian 09 using Restricted Hartree-Fock with basis set 6-311++G(d,p) and the default water in the cases where the implicit solvent was used.

2.3 Method Development and Results

As previously stated, when studying this system we encountered many obstacles. We needed to understand how AMA coordinates the zinc ion, which would require a polarizable force field, but we also had to correctly simulate the protein and the ligands. We decided not to use a polarizable solvent since GAFF and Amber 99SB-ILDN have been developed to work together, while a polarizable force field would not. The price of this decision was that we needed a method to consistently produce the correct coordination structure.

2.3.1 Metadynamics simulations

Firstly we simulated a simple system with only water, a zinc ion and one of our reference ligands (EDTA, EDDS or CDTA). We ran a test with EDTA usign the COORDINATION CVs applied to the zinc ion with both water and EDTA (See Section 1.4.1 for an explenation on how this CVs is implemented). The free energy surface did not show promising results since we expected to obtain as deeper minimum the one corresponding to an interaction between the zinc ion and EDTA. Due to this limitation of MTD we decided to use it as a way to generate conformations and then select the best conformations by running a cluster analysis to see if we could find some structure to use as starting point to obtain an octahedral structure. We used an inhouse written bash script reported in Appendix 2.A.3 to filter out all the structure that had a coordination number below a user defined threshold (usually 4) from the PLUMED output, with the same script we also removed all the clusters with a population below than another user defined threshold. By comparing these two arrays of structures we managed to greatly reduce the amount of structures to process with quantum mechanical optimization. In this first test we used a threshold of 3 to filter the coordination and found 22 structures to optimize with QM calculations which did not produced any octahedral structure since at least one of the carboxylic groups was facing in the opposite direction with respect to the zinc ions. So we decided to try a drastic approach. We adressed the fact that the charges assigned by the force field to the oxygen atoms in a carboxylic group were identical, but this charge distribution was not a realistic representation of the ligand-zinc complex in which one oxygen had its charge modified by the electron donation to the metal center. To better model this, we decided to move half of the original charge from one oxygen of the carboxylic group to the other. We ran again our simulations with the modified charges and we found that in this way the free energy surface looked more promising, with more stable minima correlated with a coordination number between 3 to 6. More in detail, with the charges assigned by the force field we observed the most stable minimum to be the zinc ion with 3 water

molecules and 3 EDTA atoms, while with the modified charges, the most sampled conformation became the one with 4 EDTA atoms around the zinc ion and 2 water molecules and the minimum with 6 water molecules disappeared. The results are shown in Figure 2.4.

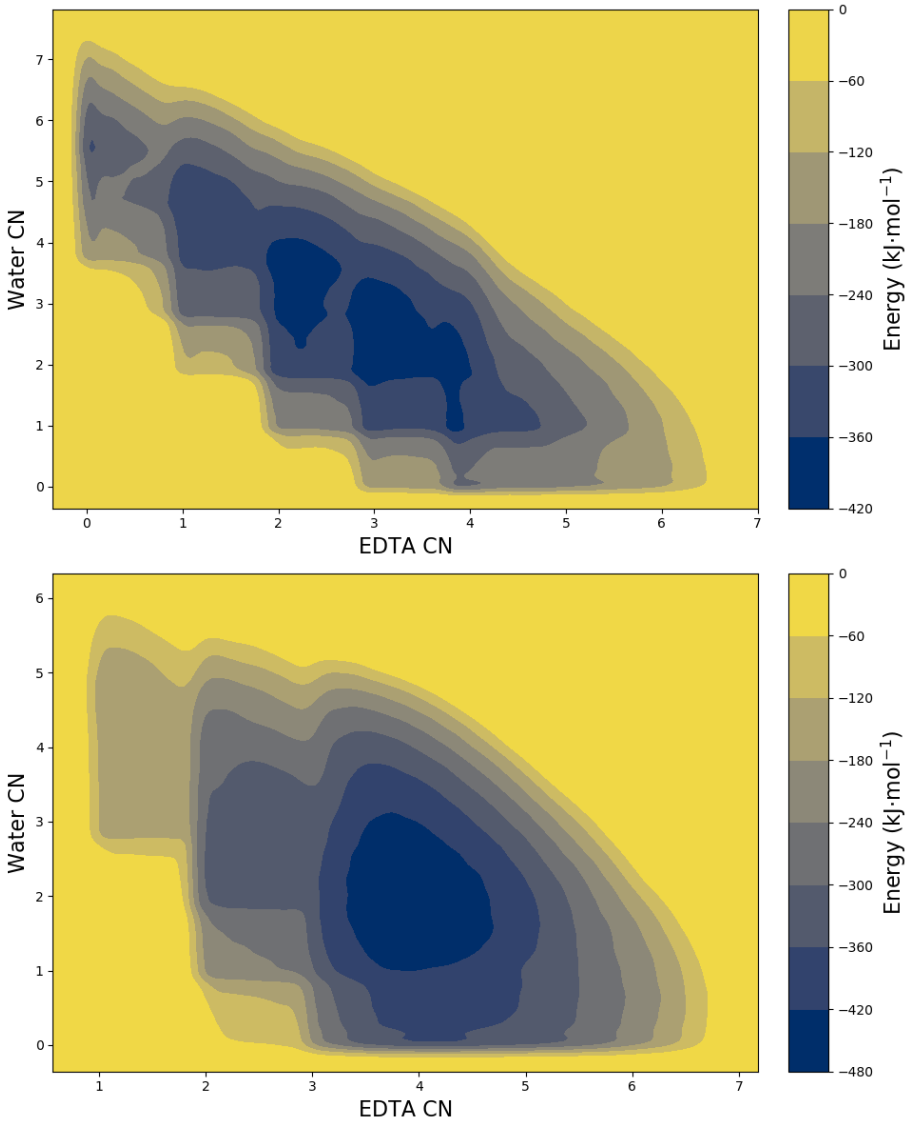


Figure 2.4: Comparison between the free energy surface calculated by analyzing the MTD simulation of the system containing EDTA, Zn^{2+} and water when the charges are modified (bottom), the minima corresponding to higher EDTA coordination number (CN) are stabilized, when the charges are not modified (top) the minimum corresponding to the zinc ion completly coordinated by water is more stable. On the axis, the CN of water (Y axes) and EDTA (X axes) with the zinc ion.

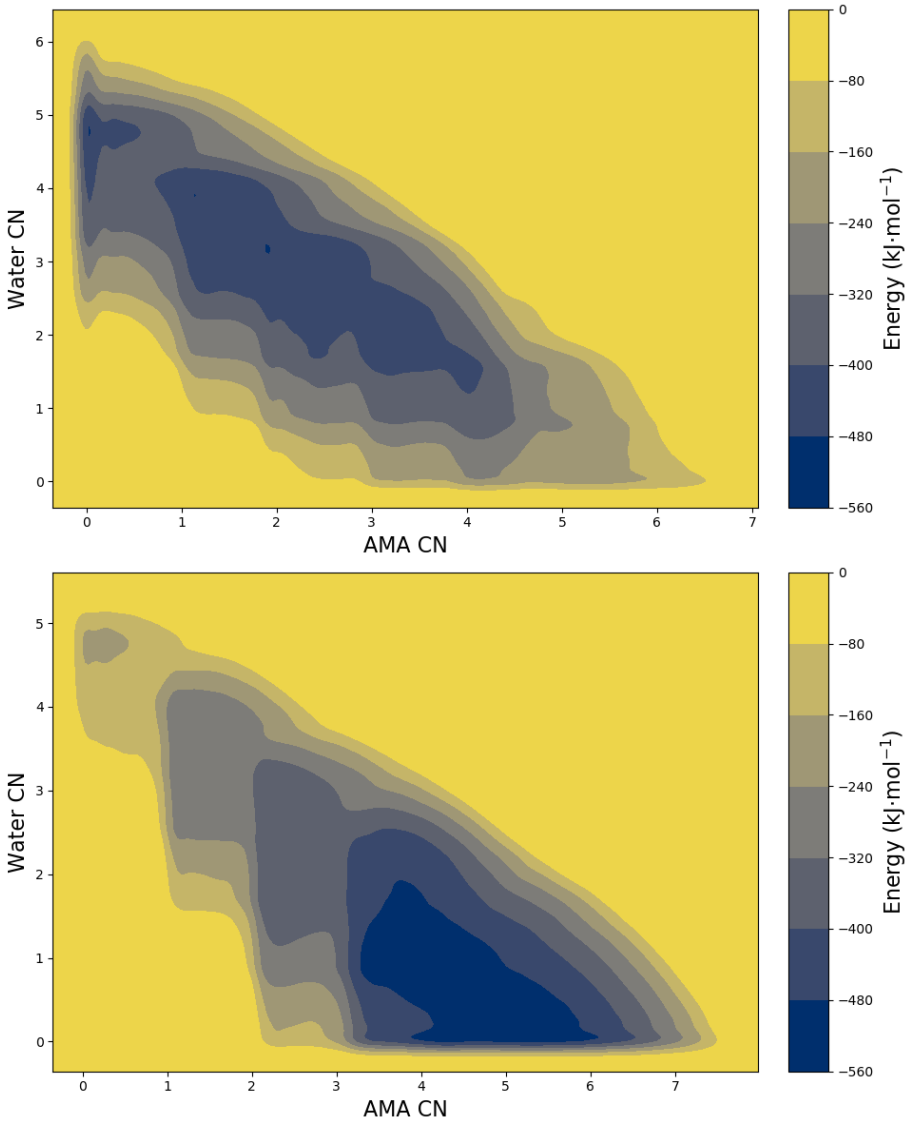


Figure 2.5: Free energy surface calculated by analyzing the MTD simulation of the system containing AMA, Zn^{2+} and water with the charges assigned by the force field (top) and modified (bottom). On the axis, the Coordination Number (CN) of water (Y axes) and AMA (X axes) with the zinc ion.

When we optimized the middle structures of the clusters, filtered by our script, in gas phase we noticed that even if most of the configurations converged to the correct coordination geometry with the same energy, the most stable structure was not one of them. To address this problem we optimized our structures using the Polarizable Continuum Model [60], which is the default one when using water as implicit solvent. In this way, we obtain the correct coordination geometry as the structure with the lowest energy. The results are shown in Table 2.1, the optimization of structure 17300 did not converged correctly during the optimization in gas phase. Notice that some structure are not recalculated with the implicit solvent since they already converged at the same structure.

Gas Phase	
Structure	Energy (kcal·mol^{-1})
10375	-1806878.638
5480	-1806877.762
19520	-1806877.762
3775	-1806877.762
4875	-1806877.762
4995	-1806877.762
5090	-1806877.762
15205	-1806877.762
7550	-1806872.645
8120	-1806872.645

Implicit Solvent	
Structure	Energy (kcal·mol^{-1})
17300	-1807042.074
19520	-1807042.074
5480	-1807042.072
8120	-1807035.571
7550	-1807035.571
10375	-1807021.075

Table 2.1: Energy of the structures extracted from a MTD simulation of EDTA with the modified charges. The optimization is ran in gas phase (left) and implicit solvent (right).

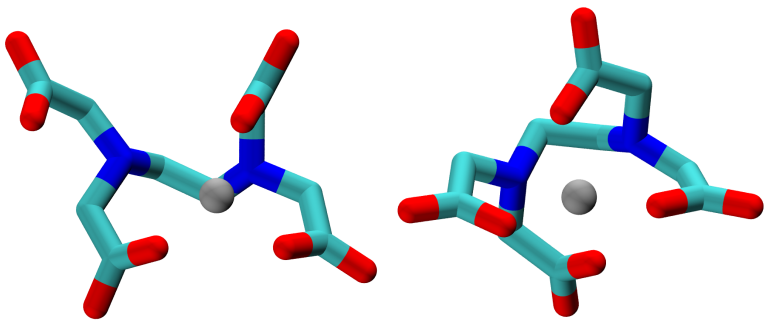


Figure 2.6: Structure 19520 optimized in gas phase (left) and with implicit solvent (right).

We used the same approach to generate octahedral structures for CDTA and EDDS, skipping the simulations with the "standard charges". MTD simulations of CDTA showed three energy minima: one with 2 water molecules and between 2 and 3 CDTA atoms coordinated to the zinc ion, one with 1 water molecule and between 4 and 5 CDTA atoms and finally one, sharp minimum, with 0 water molecules and

again between 4 and 5 CDTA atoms coordinated to the zinc ion. EDDS presents two minima, the first with between 2 and 3 water molecule and 2-3 EDDS atoms and the second one, very sharp with 0 water molecules and 5 EDDS atoms coordinated to the zinc ion. When simulating AMA we decided to test the system with both the unmodified charges and the modified ones. Once again we observed better results with the modified charges since AMA goes from not coordinating the zinc ion at all to present a broad minimum that spaces from an AMA CN of 3 to 6 (Figure 2.5). Once again, minimizing the structures with or without an implicit solvent was a key difference between obtaining or not an octahedral geometry as the lowest energy as shown in Table 2.2 and Figure 2.6.

Gas Phase	
Structure	Energy (kcal·mol^{-1})
18660	-1842013.118
10325	-1842007.219
18890	-1842006.796
8225	-1842005.660
11720	-1842004.305
19220	-1842002.884

Implicit Solvent	
Structure	Energy (kcal·mol^{-1})
10325	-1842081.740
18660	-1842079.389
19220	-1842075.514
11720	-1842058.503
8225	-1842057.717
18890	-1842054.697

Table 2.2: Energy of the structures extracted from a MTD simulation of AMA with the modified charges. The optimization is ran in gas phase (left) and implicit solvent (right).

2.3.2 Well tempered metadynamics simulations

When simulating the system with WTMD we found out that for EDTA, CDTA and EDDS there was no need to use the modified charges to obtain octahedral structures after minimizing the cluster middle structures. When simulating AMA with the force field assigned charges we observed structures similar to the one we obtained with MTD so we decided to once again modifying AMA carboxylic oxygens, this again produced pseudo octahedral structures which had a lower energy of the ones obtained previously. We now felt confident to simulate AMA with the MBLs. In this case we were interested in the coordination of the zinc ion with the protein and AMA, so we used as CVs the coordination between the three nitrogen atoms of the histidines (116, 118 and 196) that were bound to the zinc ion and the AMA atoms that can coordinate. During the simulation we observed the zinc ion to spontaneously leave its pocket and only later gets in contact with AMA. This is expected and correct due to the nature of the CVs we used but the force field can not correctly reproduce the interaction between the protein and the zinc ion and that is confirmed by the fact that the minimum corresponding to the coordination of the three histidines is shallow while the ones corresponding to zinc in water and coordinated by AMA are

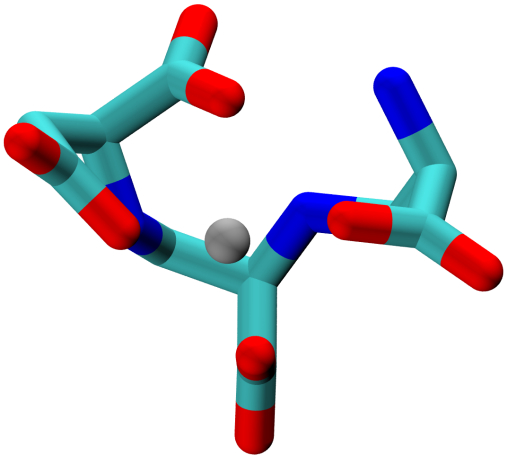


Figure 2.7: AMA lowest energy structure extracted from the WTMD trajectory with modified charges.

very deep, as shown in Figure 2.8.

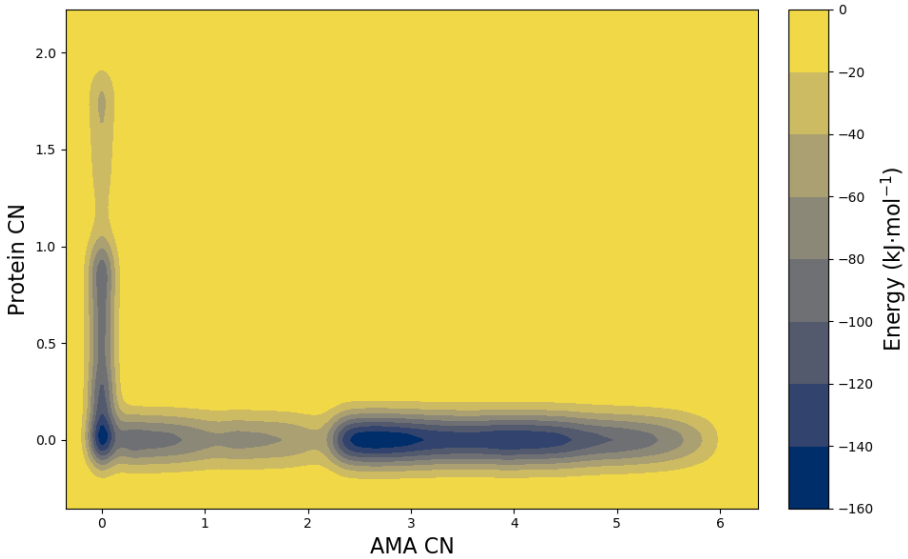


Figure 2.8: Free energy surface calculated by analyzing the WTMD simulation of the system containing AMA, Zn^{2+}, the MBL and water. On the axis, the Coordination Number (CN) of the three histidines (Y axes) and AMA (X axes) with the zinc ion.

2.3.3 Quantum mechanic/molecular mechanic

Since the force field performed poorly in part of the simulation we decided to study the system with a QM/MM approach. We simulated the four different stages of the process: the protein with the zinc ion coordinated, starting from the crystallographic structure; the protein without the zinc ion; AMA coordinated with the zinc ion, starting from the structure we obtained from WTMD simulations and AMA without zinc. The results we obtained when minimizing these structure in gas phase showed that removing the zinc ion was a big perturbation of the systems. The MBL had its energy almost doubled when the zinc ion was removed while the AMA energy was almost three time higher. These fluctuations are of course too high to have physical sense or to give any insight on the actual inhibition mechanism underlying the MBLs-AMA interaction. We tried to optimize our structures adding the implicit solvent but we could not get the calculations to converge.

System	Energy (kcal·mol^{-1})
MBL with Zn^{2+}	-1214863.330
MBL without Zn^{2+}	-667869.173
AMA with Zn^{2+}	-900133.129
AMA without Zn^{2+}	-353400.409

Table 2.3: QM/MM calculation results

2.4 Conclusions

The aim of our work was to investigate the interaction mechanism that lies behind the ability of AMA to restore the activity of penicillin like drugs. Even if we did not accomplish our primary objective we were able to correctly model the coordination process using MTD and WTMD as a method to generate different conformation. We produced a script to filter out the uninteresting conformations and select only the ones that, once optimized with QM calculations, are more likely to produce the correct coordination geometry.

2.A Appendixes

2.A.1 Free energy surfaces

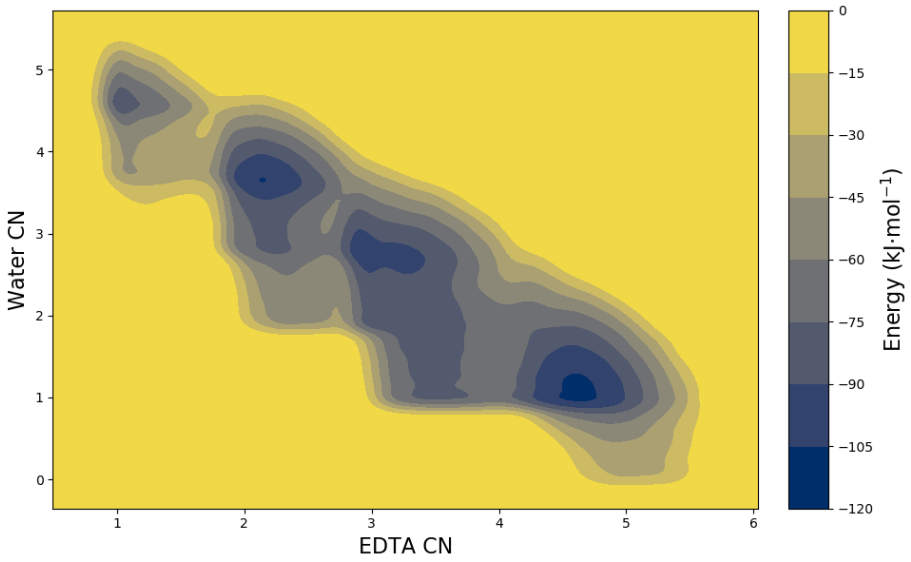


Figure 2.A.1: Free energy surface calculated by analyzing the WTMD simulation of the system containing EDTA, Zn^{2+} and water.

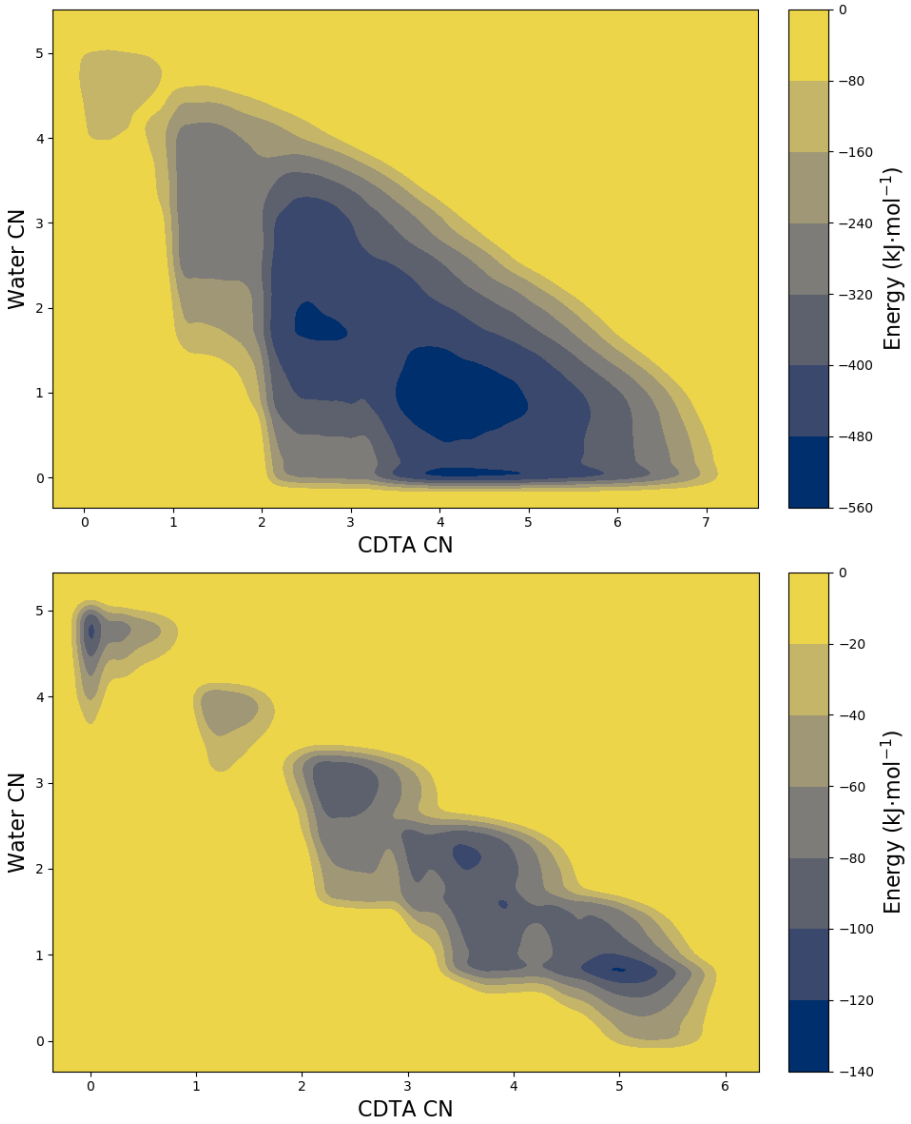


Figure 2.A.2: Free energy surface calculated by analyzing the MTD (top) and WTMD (bot) simulation of the system containing CDTA, Zn^{2+} and water.

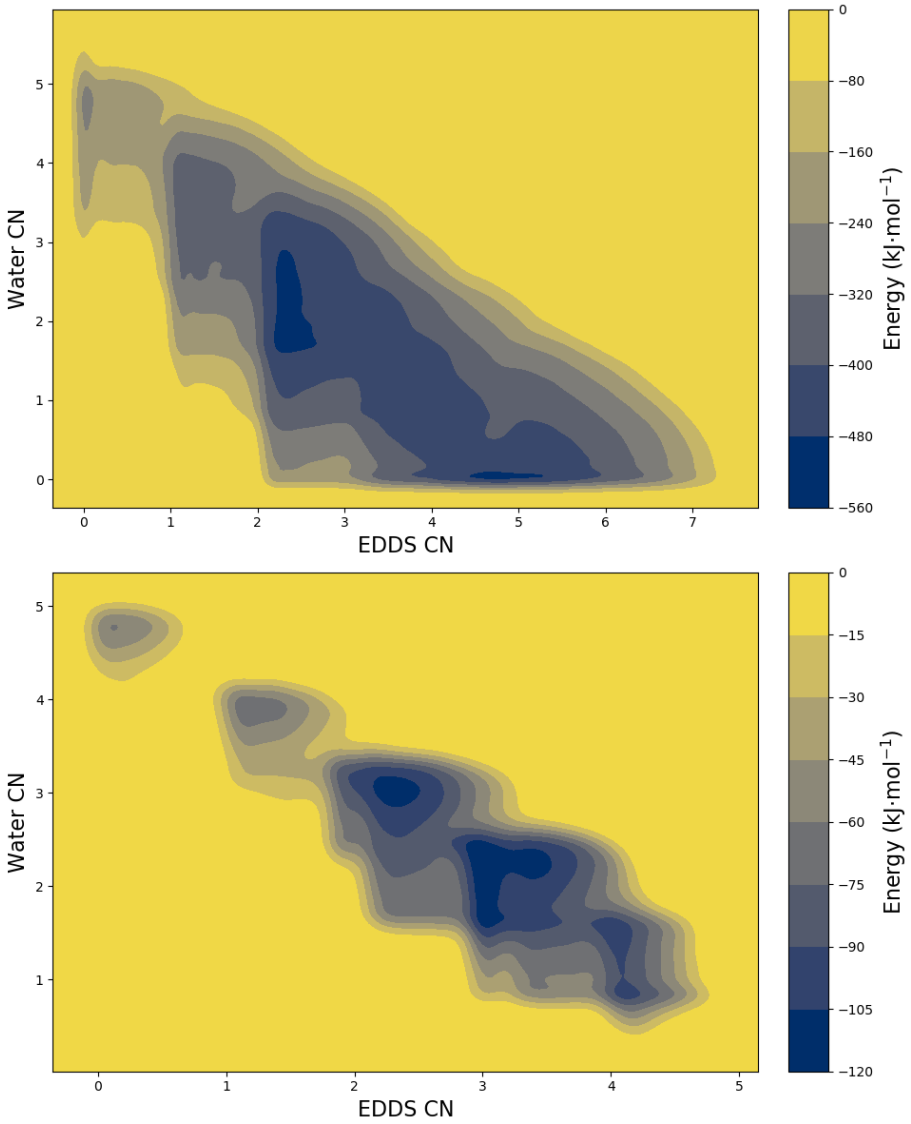


Figure 2.A.3: Free energy surface calculated by analyzing the MTD (top) and WTMD (bottom) simulation of the system containing EDDS, Zn^{2+} and water.

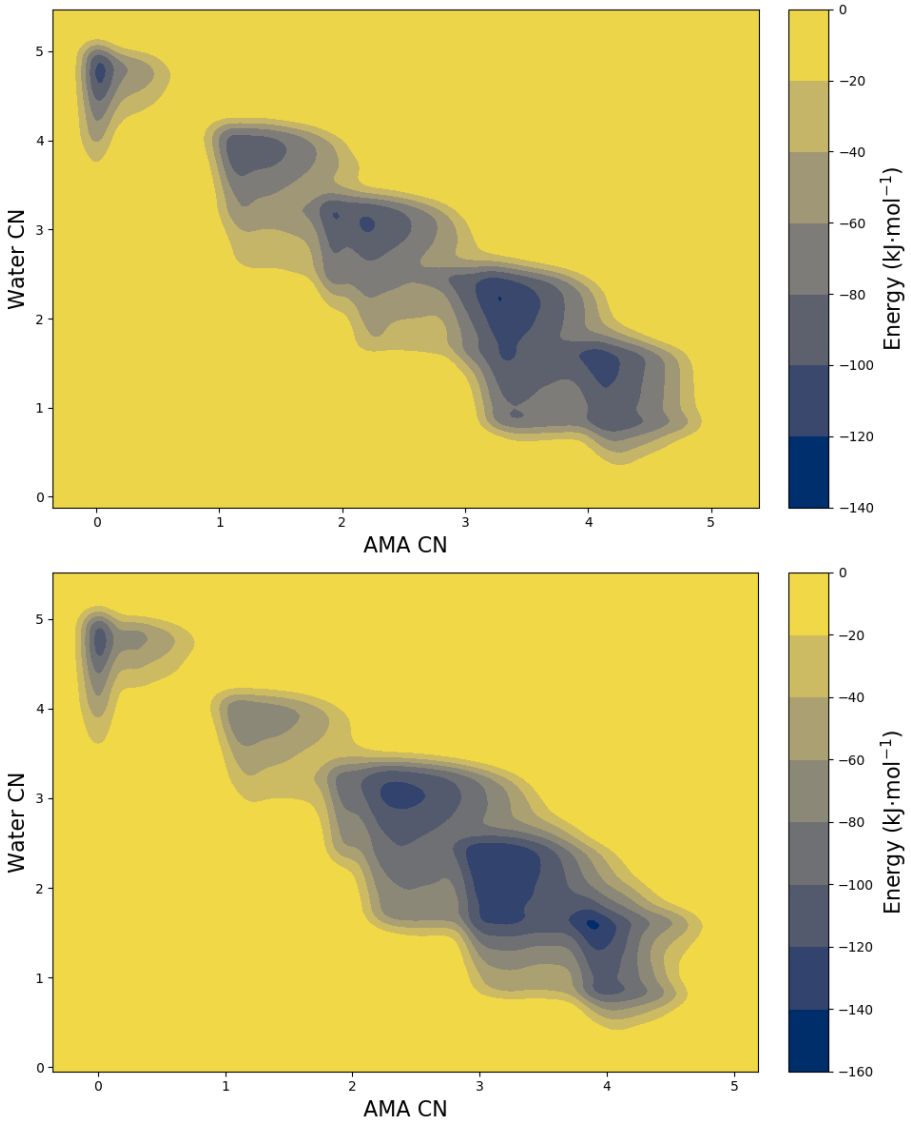


Figure 2.A.4: Free energy surface calculated by analyzing the WTMD simulation with the charges assigned by the force field (top) and modified (bottom) of the system containing AMA, Zn^{2+} and water.

2.A.2 Tables

MTD	
Structure	Energy (kcal·mol^{-1})
4355	-1904984.321
10900	-1904984.318
15240	-1904984.057
4190	-1904984.054
13445	-1904980.526
13190	-1904980.525
13210	-1904980.525
14625	-1904980.524
19260	-1904980.198
8640	-1904980.197
9545	-1904980.196
1100	-1904979.268
2500	-1904979.267
2685	-1904979.267
3835	-1904979.267

WTMD	
Structure	Energy (kcal·mol^{-1})
10705	-1904984.321
14730	-1904984.320
17165	-1904984.289
19215	-1904980.528
7670	-1904980.525
7690	-1904980.525
12990	-1904980.524
8055	-1904980.524
8575	-1904977.041
8575	-1904977.041
2235	-1904929.384
3765	-1904928.702

Table 2.A.1: Energy of the structures extracted from MTD (left) and WTMD (right) simulations of CDTA. The optimization is performed in implicit solvent. The MTD simulation is performed with the modified charges.

MTD	
Structure	Energy (kcal·mol^{-1})
3865	-1807050.517
4385	-1807049.811
4525	-1807049.810
19510	-1807043.875
2865	-1807036.197
2330	-1807035.301
19010	-1807035.299
2750	-1807035.245
2550	-1807035.244
19915	-1807034.900
4800	-1807034.900
5170	-1807034.900
5990	-1807034.900
8525	-1807034.900
7715	-1807022.596
9200	-1807022.305
7325	-1807016.725
1335	-1807006.887

WTMD	
Structure	Energy (kcal·mol^{-1})
8815	-1807023.872
9555	-1807023.872
11165	-1807023.871
14310	-1807023.871
15600	-1807023.871
19955	-1807023.871
9615	-1807023.871
9885	-1807023.871
14320	-1807023.870
14695	-1807023.032
15350	-1807017.428
11400	-1807017.427
8730	-1807017.427
19855	-1807017.327
17135	-1807017.324
18165	-1807017.324
15310	-1807017.120
18145	-1807016.809
1325	-1807004.834

Table 2.A.2: Energy of the structures extracted from MTD (left) and WTMD (right) simulations of EDDS. The optimization is performed in implicit solvent. The MTD simulation is performed with the modified charges.

WTMD	
Structure	Energy (kcal·mol^{-1})
19015	-1807042.075
19480	-1807042.075
18250	-1807042.074
17285	-1807042.073
10240	-1807042.073
17070	-1807042.073

Table 2.A.3: Energy of the structures extracted from WTMD simulation of EDTA. The optimization is performed in implicit solvent.

MTD	
Structure	Energy (kcal·mol^{-1})
11935	-1806879.127
13315	-1806878.638
7755	-1806878.638
9510	-1806878.638
9555	-1806878.638
5635	-1806876.266
7110	-1806876.266
5690	-1806874.647
6180	-1806874.647
12275	-1806866.171
3820	-1806858.511

Table 2.A.4: Energy of the structures extracted from MTD simulation of EDTA performed with the force field assigned charges. The optimization is performed in implicit solvent.

MTD	
Structure	Energy (kcal·mol^{-1})
2530	-1841987.620
3455	-1841970.301
3755	-1841970.301
4685	-1841970.301
7270	-1841965.985
7370	-1841965.985
12255	-1841948.496
11590	-1841948.495
3430	-1841886.689

Table 2.A.5: Energy of the structures extracted from MTD simulation of AMA performed with the force field assigned charges. The optimization is performed in implicit solvent.

Standard Charges	
Structure	Energy (kcal·mol^{-1})
31575	-1842042.569
27665	-1842042.568
21865	-1842042.079
4980	-1842041.736
12740	-1842038.313

Modified Charges	
Structure	Energy (kcal·mol^{-1})
12225	-1842059.609
12100	-1842059.598

Table 2.A.6: Energy of the structures extracted from a WTMD simulation of AMA. The optimization is performed in implicit solvent. The WTMD simulation is performed with the force field assigne charges (left) and the modified charges (right).

2.A.3 Result filtering script

Here is reported the bash code used to filter the coordination number and the cluster

```
#!/bin/bash

module purge
module load gromacs/5.0.7-cpu

echo -e "Never use the file extension \c"

echo -e "Have you already ran the cluster analysis? (Y/N) \c"
read o

ls *.gro -lhtr

echo -e "File .gro to use\c"
read gro

ls *.tpr -lhtr

echo -e "file tpr to use \c"
read tpr
```

The first line is used to check if it is necessary to run the cluster analysis or not. The other lines ask for inputs that are always necessary.

```
if [ $o == N ]
then

ls *.xtc -lhtr

echo -e "file xtc to use \c"
read trj

echo -e "index? (Y/N) \c"
read v

if [ $v == Y ]
then
```

```
make_ndx_mpi -f $gro.gro
wait

fi
```

In case the cluster analysis has to be run, this part of the code ask for the necessary files and if the user wants to create an index file, if the users answer positively, the code create the index file.

```
if [ $v == N ]
then
z="0"
fi

echo -e "cluster? (Y/N) \c"
read v


#index
echo -e "index group: \c"
read z


echo -e "time step to use \c"
read dt


if [ $v == Y ]
then

if [ $z != 0 ]
then
echo $z|gmx_mpi cluster -f $trj.xtc -dt $dt -n -cutoff 0.1
            -method gromos -s $tpr.tpr
wait

else

echo -e "time step to use \c"
```

```
read dt

#index
echo -e "index group: \c"
read z


echo $z|gmx_mpi cluster -f $trj.xtc -dt $dt -n -cutoff 0.1
            -method gromos -s $tpr.tpr
wait

fi
fi
fi
```

These two parts deal with the index usage and run the cluster analysis. If the users does not want to run it, it is all skipped.

```
#Ask if the user wants to extract all
#the middle structures or single snapshots

echo -e "All the middle structures of a cluster (press 1)
or single snapshots (press 2)?"

read a

#If the user wants to extract single structures

if [ $a == 2 ]

then

#Numer of snapshot to extract
echo -e "how many snapshots: \c"
read y

#Index to use
echo -e "index group: \c"
read z
```

```
#Trajectory to use
ls *.xtc
echo -e "xtc file to use\c"
read trj

i="0"

#Cicle over the number of structures to extract

while [ "$i" -lhtr  "$y" ]

do

#Ask the structure to extract
echo -e "which snapshot: \c"
read x

echo $z|trjconv_mpi -f $trj.xtc -s $tpr.tpr -n -b $x
            -e $x  -o $x.pdb -pbc whole

i=$[$i+1]

wait

done

fi

if [ $a == 1 ]

then

#In case of multiple runs, the previously
#created folders/files are eliminated

rm file_extracted -r
rm middle_structures.log

mkdir file_extracted
```

```
mkdir gro_files
mv *.gro gro_files

if [ $o == Y ]
then

#Index file
echo -e "index group: \c"
read z

ls *.xtc -lhtr

#Trajectory file
echo -e "xtc file to use: \c"
read trj

fi

#Extract the middle structures but only if they have at least 2

#Since the cluster.log file is not neatly written,
#is not possible to extract directly the middle structures

awk 'NR > 10{ print }' cluster.log > file_extracted/file1.log

#The following line removes the first columns,
#leaving only the lines that contain a
#pipe character "|"

awk '{for(z=1;z<=2;z++)$z="";$0=$0;$1=$1}1'
      file_extracted/file1.log>file_extracted/file2.log

#Now we eliminate all the lines that start with a pipe.
#this will produce a nice formatting that will also cause
#the elimination of the cluster with only one structure

awk -F '|'  ' /^\|/ {next} {print}'
      file_extracted/file2.log>file_extracted/file3.log
```

```
#Remove the clusters with less than the users defined
#number of structure

echo -e "Minimum nuber of structures: \c"
read s

awk -v s="$s" '!($1<s)'
      file_extracted/file3.log > file_extracted/file4.log

#Print the middle structures of the remaining clusters

awk '{print $4}'file_extracted/file4.log>file_extracted/file5.log

#Clean the file from possible remaining pipes

awk -F '|'  ' /^\|/ {next} {print }'
      file_extracted/file5.log > middle_structures.log

#Read the middle structures and assign them to "myarray"

declare -a myarray
readarray -t myarray < middle_structures.log

##############################################
###     Filter with coordination        ####
##############################################

echo -e "Minimum coordination number: \c"
read c

#removes the first two lines from file HILLS

awk 'NR > 2{ print }' HILLS > file_extracted/filter1.log


#IMPORTANT! The number used after NR% MUST be equal
#to the time step used in the cluster analysis

#Select only the frames that are contained in thei
```

```
#refined trajectory (if the trajectory has been refined)

awk 'NR%5==0'file_extracted/filter1.log>file_extracted/filter2.log

#Print only the column with the coordination number

awk '{print int($2)}'
          file_extracted/filter2.log>file_extracted/filter3.log

#Load the coordination number in the array

declare -a coordarray
readarray -t coordarray < file_extracted/filter3.log
i=0

 touch appo.log

touch list_gjf.log

#This loop pairs the coordination number with a specific structure
#using the "row" of the coordination number.

while (( ${#myarray[@]} > i )); do

        x=${myarray[i]}

        k=${coordarray[$x-1]}

i=$(($i+1))

printf "%s      %s\n" "$x       $k" > appo$i.log

cat appo$i.log>>appo.log

rm appo$i.log

#If the coordination is equal or greater than the threshold
#selected, the structure is noted and extracted
```

```
if [ $k -ge $c ]

then

echo $z $z|trjconv_mpi -f $trj.xtc -s $tpr.tpr -n -b $x
                -e $x -o $x.gro -pbc cluster

        printf "%s\n" "$x" > list_gjf$i.log

        cat list_gjf$i.log >> list_gjf.log

        rm list_gjf$i.log

wait

        cat $x.gro>>traiettoria

        editconf_mpi -f $x.gro -pbc -o $x.pdb

fi

done

#Clean the folder

mv *.gro file_extracted
mv traiettoria traiettoria.gro
mkdir file_extracted/pdb
mv *.pdb file_extracted/pdb
cd file_extracted/
mkdir gjf
cd ..
mv gro_files/* .
awk '{print}' list_gjf.log > list_due.log
mv list_due.log file_extracted/gjf/list_gjf.log
rm gro_files -r
fi

exit 0
```

Chapter 3

Tubulin Targeting Guanosine Derivatives

3.1 Introduction

Tubulin is a heterodimeric protein fromed by α and β-tubulin, although many different α and β-sequences have been identified, the two subunits never exist alone; they are folded together during formation in the ribosome. The tubulin dimer bounds two Guanosine Triphosphate (GTP) molecules, one on the α subunit and the other on the β subunit. While the GTP molecule bound to the α subunit plays a structural role in the dimer, the GTP molecule bound to the β subunit can be hydrolyzed and a Guanosine Diphosphate (GDP) capped dimer is more likely to fall off [61] but only if the dimer is at one end of the microtubule, a GDP-bound tubulin in the middle of a microtubule cannot spontaneously pop out of the polymer. α/β-tubulin heterodimers assemble in a head-to-tail fashion to form linear protofilaments. These protofilaments further polymerize to form a hollow tubular structure consisting of 13 protofilaments joined together. The structure has a polar organization, in which the α-tubulin subunit is located at the negative end and the β-tubulin subunit is exposed at the positive end [62]. While growth can occur at both ends of the microtubule, it is significantly faster at the plus end [63]. The equilibrium between microtubule growing and shrinking is called dynamic instability and it is regulated by the hydrolysis of the GTP molecule of the β-subunit as mentioned earlier. Tubulin, due to its ability to self assemble in a controllable manner to form microtubules plays a key role in many cellular processes, such as transportation, where microtubules act as "highways" for motor proteins or in cell division where some microtubules assists in the chromosomes separation process by attaching the kinetochore of the chromosome and then, when the kinetochore microtubules depolymerise they pull the two half of the chromosome away from each others. To alter the aggregation of tubulin dimer or the disintegration of microtubules has dramatic effects on the cells, to the

point of initiating programmed cell death [64]. Microtubule dynamic is extremely sensitive to therapeutic inhibition [65] which explains the amount of research focused on developing molecules that are able to disrupt or stabilize the formation of microtubules [66]. A lot of work has been focused to produce non hydrolysable GTP analogues [67, 68] to stabilize the microtubule and prevent its disaggregation. Here

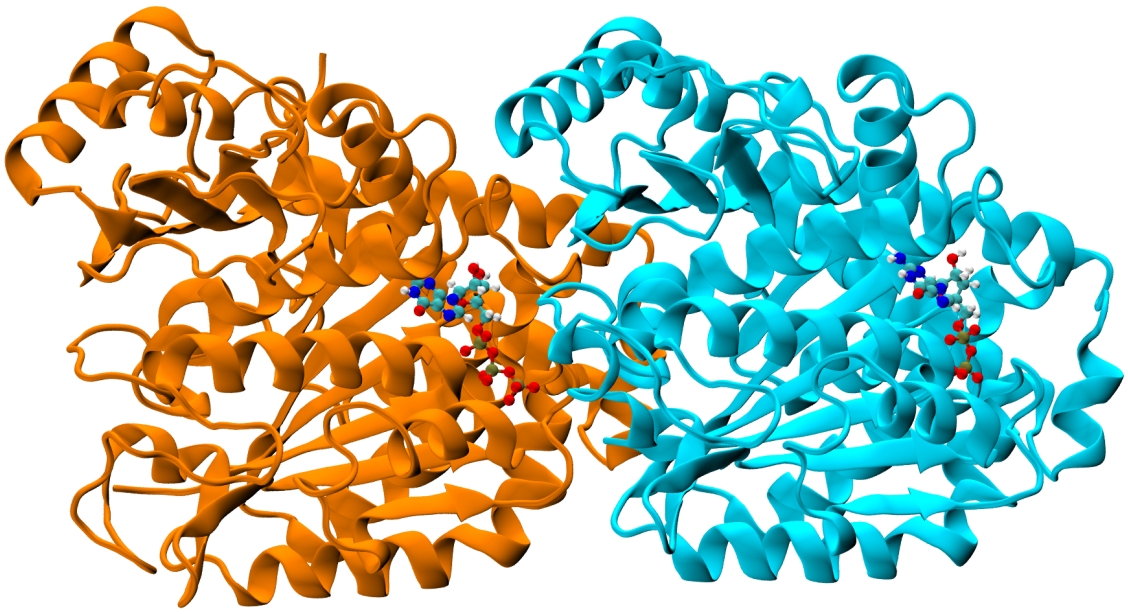


Figure 3.1: A tubulin dimer. The α subunit is on the left and it is colored in orange, the β subunit is on the right and is colored in cyan. Between the two subutin, represented in CPK a GTP molecule. Finally, on the β subunit a GDP molecule is represented in CPK.

we designed six nucleosides, sterically similar to GDP but with a positive charge instead of the three negative ones. This charge modification is aimed to disrupt the electrostatic interaction between the Mg^{2+} ion which is structural in the GTP binding site and the phosphate negative charges. We removed the phosphate chain and replaced it with an aliphatic chain, with a protonated amino group at the end, as shown in Figure 3.2. To determine which of our six compounds would be more likely to bind in the GTP binding site we calculated the binding free energy of our compounds. The nucleosides were directly built in the binding pocket by replacing the phosphate chain with the aliphatic ones.

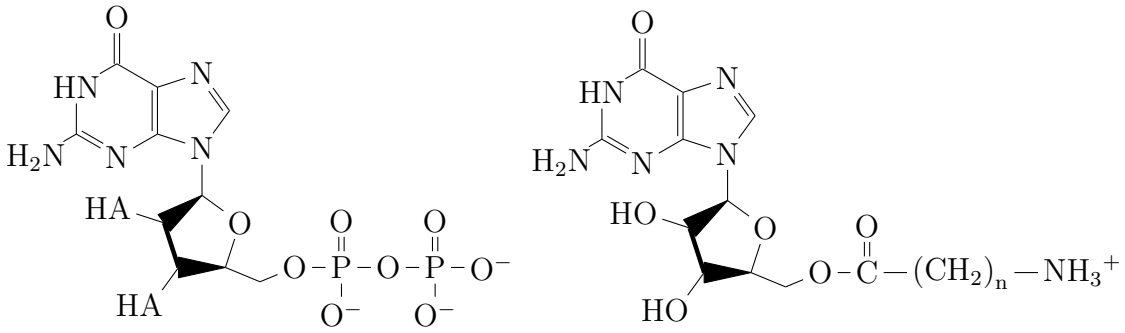


Figure 3.2: GDP (left) and our nucleosides (right). The number of CH_2 groups that determines the length of the aliphatic chain (n) goes from 1 to 6, the compounds are named after this number **Compound 1** to **Compound 6**.

3.2 Computational Details

The structure of tubuline was obtained from the protein data bank (PDB ID code 4I55 [69]). We extracted from the crystal structure two tubulin monomers labeled A and B, the structural magnesium ions contained in the A and B subunits, the GTP bound to the A subunit and the GDP bound to the B subunit, finally 27 Na^+ ions were added to balance the total charge. The system was simulated in explicit water solvent employing periodic boundary conditions with initial box dimensions 93 Å x 120 Å x 119 Å. Tubuline was described with the amber99SB [70] force field, GTP, GDP and the modified GDPs were described with GAFF [58], TIP3P [57] model was adopted for water. The total number of atoms was $\approx$ 33000 of which 13410 were part of the protein, 43 were GTP atoms, 27 were Na^+ ions to balance the overall charge, 2 Mg^{2+} ions and $\approx$ 19000 water atoms, the nucleosides were formed by a number of atoms that went from 40 to 55. The modified nucleosides were built using the nucleoside part of the crystallographic structure, to replace the phosphate chain we used tLEaP a module of AmberTools16 [71]. We optimized the initial system geometry with a 50000 steps of energy minimization with the steepest descent method. Hydrogen bond lengths were constrained to their equilibrium values with the LINCS algorithm [6], allowing time step of 2 fs. We then carried out a 200 ps equilibration in the NVT ensemble. The NVT was followed by a second equilibration of 200 ps in the NPT ensemble. Temperature and pressure were kept constant to their reference values (1 bar, 300 K respectively) through the velocity rescale algorithm [72] and the Berendsen barostat [73]. We then performed a 100 ns MD. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [74] was employed to calculate long range electrostatic interactions.

CAS To perform computational alanine scanning, 100 snapshots were extracted from each dynamics (one snapshot every 10 ps). The binding free energy variation (ΔG_{bind}) was calculated using the MM-PBSA approach. The Poisson–Boltzmann equation was solved with APBS [27] using a relative dielectric constant of 80 for the region around the protein and of 2 for the protein interior. The entropic contribution to ΔG is supposed constant in the mutated and wild type structure considering their similarity, so it has not been calculated, as discussed by Kollman *et al.* [75]. The Mg^{2+} ion was not included in the calculations since APBS is unreliable when multivalent ions are involved, especially if they are bound to the protein.

3.3 Results

The binding free energy calculations show two distinct trends for the six compounds. At first the binding free energy rapidly decreases when then aliphatic chain grows. When we calculate the binding free energy for our fourth nucleoside (Figure 3.2), the binding free energy increases suddenly and then keep slowly increasing with each carbon atom added, as shown in Figure 3.3. We then decided to investigate which

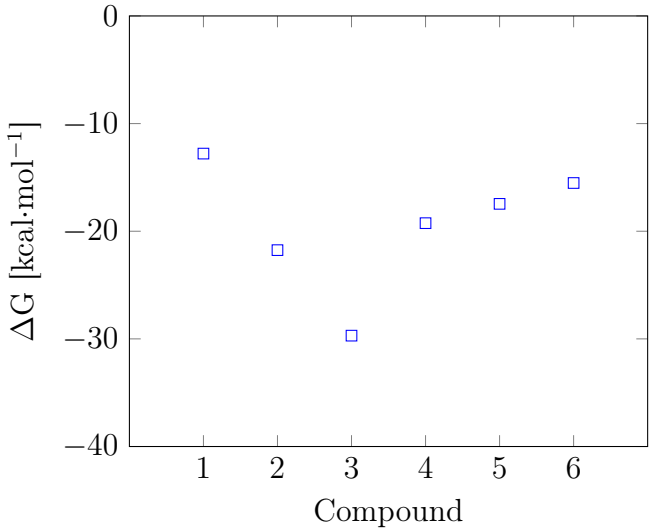


Figure 3.3: ΔG_{bind} of our six compounds with the GTP binding pocket.

aminoacids were in contact with our compounds by running a Solvent Accessible Surface Area (SASA) analysis. We considered significative a ΔSASA above 0.1 $Å^2$, with this threshold we found 36 aminoacids in contact with our compounds. We used CAS to mutate the 36 aminoacids and calculate the variation in the binding free energy ($\Delta\Delta G_{bind}$). Our result show that all 36 aminoacids are hotspots when interacting with **Compound 3** as shown in Table 3.1. From Table 3.1 we can see which nucleoside has the highest number of hotspots and warmspots, more concisely

showed in Figure 3.4. **Compound 3** shows that all the residues considered are hotspots. Residue Glu 624 is involved in an hydrogen bond with the ribose ring while Tyr 665 and Asn 669 are involved in interactions with guanine. These three hotspots are common for all the six Compounds but **Compound 1** and **Compound 6** where the interaction with Tyr 665 is missing. The results suggest that there are many factors at play in balancing the ability of our ligands to interact with the binding pocket. We noticed that while the increasing stability of the complex between β-

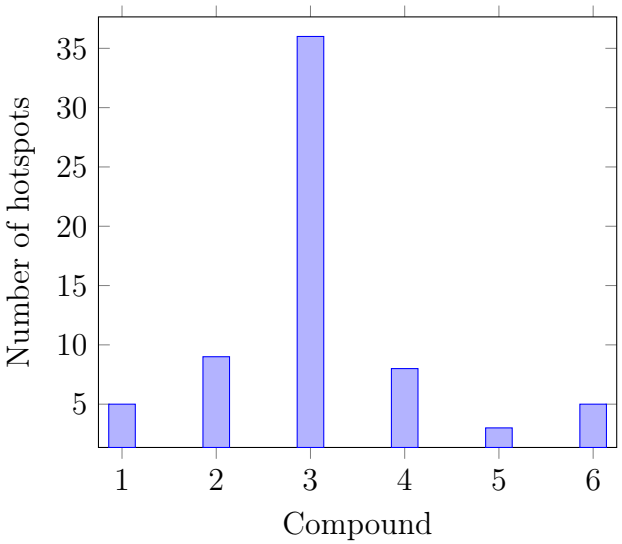


Figure 3.4: Number of hotspots for each compound tested.

tubulin and **Compound 1**, **Compound 2** and **Compound 3**, that we observe is probably due to the increasing amount of hot spots, the steep increase of the ΔG_{bind} value associated with **Compound 4** is probably related to the fact that the aliphatic side chain becomes too long to accomodate the positive charge carried by the amino group inside a side pocket and so the alyphatic chain is forced to bend towards the end and the amino group outside the side pocket as shown in Figure 3.6, while the first three compounds have a stabilizing interaction with the side binding pocket as shown in Figure 3.5. All these result are strong indicators that **Compound 3**, being the compound with the lowest binding free energy, could compete with GDP in the relative binding site and so alter the microtubule dynamic. An experimental input is now necessary to further understand if and how **Compound 3** or another compound derived from it can actually interact with the tubulin dimer and the microtubule to disrupt the dynamic instability which is an integral part of the cell life cycle.

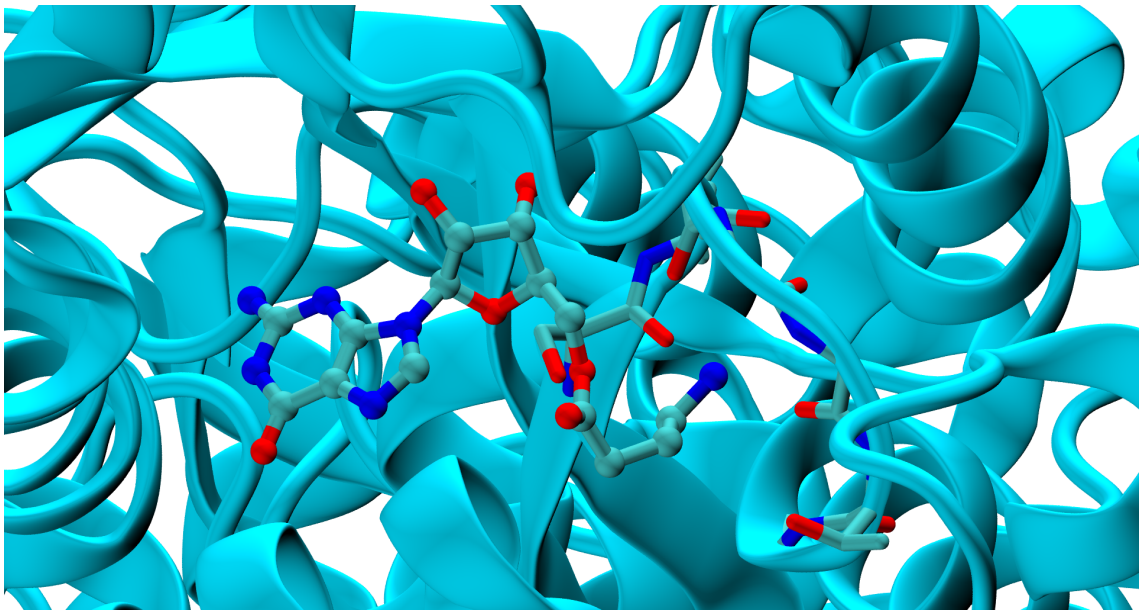


Figure 3.5: **Compound 3**, represented in CPK. The amino group is oriented towards the negative side pocket, the relevant residues are shown with the licorice representation.

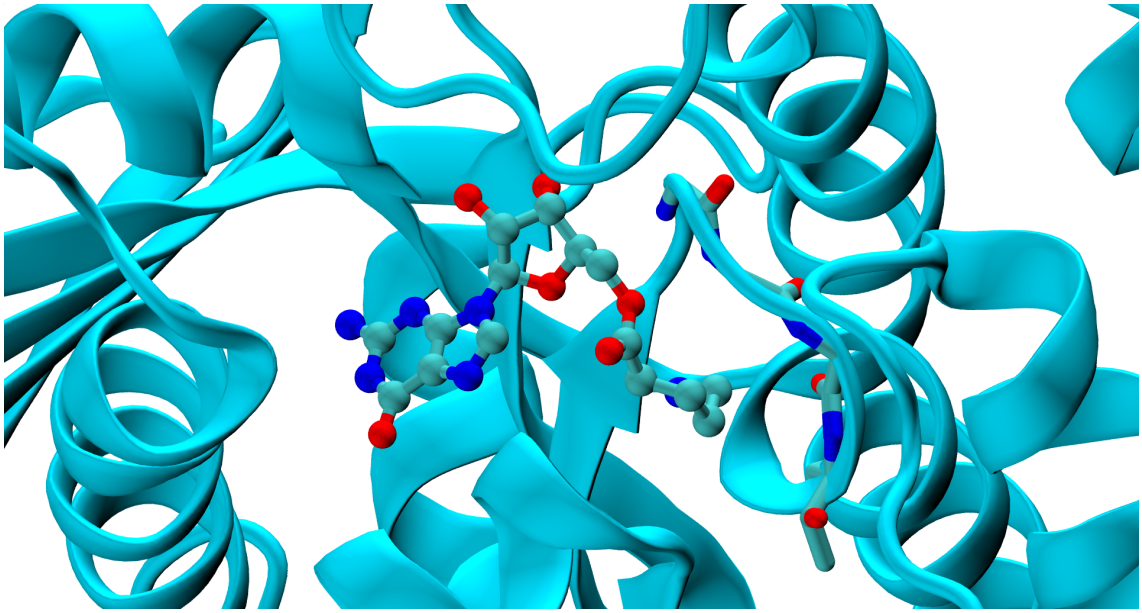


Figure 3.6: **Compound 4**, represented in CPK. The amino group faces away from the negative side pocket, the relevant residues are shown with the licorice representation.

ΔΔG (kcal·mol^{-1}) calculated with CAS						
Residue	1	2	3	4	5	6
Gln454	-1.58	**8.64**	**2.88**	-1.35	-0.16	-2.23
Cys455	0.80	1.54	**4.11**	**2.85**	1.05	**2.87**
Gln458	-0.27	0.14	**2.80**	-0.23	-0.29	0.95
Ile459	-0.04	0.98	**3.32**	0.84	0.33	0.78
Asp510	**2.92**	**4.37**	**5.51**	**5.30**	1.50	1.53
Leu511	-0.04	-0.03	**2.56**	-0.05	0.00	-0.41
Glu512	-0.27	1.61	**3.16**	-0.85	0.63	0.55
Asn542	0.07	**3.40**	**2.63**	1.72	-0.21	-1.24
Asn543	0.05	0.30	**2.64**	0.19	0.05	0.07
Tyr549	0.00	0.00	**2.55**	0.02	0.02	0.02
Leu578	-0.01	0.00	**2.67**	0.03	0.01	0.31
Hie580	0.23	0.40	**2.85**	0.22	0.09	0.33
Ser581	0.54	**2.34**	**3.37**	1.00	-0.63	0.34
Leu582	0.11	0.14	**2.95**	0.15	0.10	0.33
Thr586	0.09	-0.21	**2.56**	**2.63**	-0.01	0.02
Ser588	-0.06	0.81	**2.01**	0.48	-0.12	**2.43**
Met590	0.00	0.00	**2.86**	0.04	-0.01	0.16
Thr592	0.07	-0.17	**2.78**	-0.09	-0.03	-0.01
Val612	1.45	0.74	**2.87**	1.24	0.22	0.98
Met613	0.04	0.14	**2.66**	0.07	0.04	0.04
Pro614	0.13	0.59	**3.34**	0.37	0.14	0.11
Ser615	0.02	-0.02	**2.64**	0.00	0.05	-0.12
Val618	1.49	1.25	**2.98**	1.23	1.24	1.82
Ser619	-0.09	0.32	**2.63**	-0.38	-0.38	0.29
Asp620	-2.18	**2.65**	**3.56**	-2.23	0.31	-1.41
Thr621	-0.48	0.33	**2.54**	0.04	0.20	-0.35
Glu624	**7.59**	**26.65**	**25.63**	**16.27**	**12.04**	**12.10**
Asn627	-0.11	-0.44	**2.34**	-0.12	0.07	-0.05
Ser631	0.07	0.28	**3.07**	0.17	0.05	0.47
Ile645	0.07	0.07	**2.59**	-0.10	0.05	0.08
Asn647	**2.02**	**2.71**	**5.93**	**4.04**	1.99	1.66
Leu650	0.09	0.00	**2.49**	-0.20	-0.03	0.02
Tyr665	0.28	**4.17**	**7.18**	**4.39**	**4.47**	0.94
Leu668	0.34	0.67	**3.41**	1.13	0.75	0.36
Asn669	**2.44**	**5.27**	**7.50**	**3.83**	**3.81**	**4.15**
Val672	0.25	0.01	**2.55**	0.06	-0.09	0.03

Table 3.1: The calculated ΔΔG for every mutated residue. In bold are highlighted the ones with a ΔΔG higher than 2 kcal·mol^{-1}.

3.4 Conclusions

Our study was focused on the GTP binding pocket since the equilibrium between GTP and GDP regulates the structural stability of the microtubule positive end. In our work we tested 6 guanosine derivatives and their ability to bind the relative pocket while disrupting the network of electrostatic interactions between the phosphate groups, the residues in the pocket and the Mg^{2+} ion. To do this, we removed the phosphates and replaced them with an aliphatic chain which carries a positive charge at the end via an amino group. Our results show that **Compound 3** is very likely to bind strongly to the GTP binding pocket since it is able to interact with all the relevant residues and has the correct length to accommodate the positive charge in a side pocket where the backbone oxygens are oriented in a favorable manner.

Chapter 4

Conformational Study of Bicyclic Pyrrolidine-Isoxazoline γ Amino Acid

4.1 Introduction

Amino acids are the key building blocks of peptides and proteins which are widely exploited in different applications, from pharmaceutical chemistry and biomedicine to material science, optoelectronics and catalysis [76–79]. The insertion of Unnatural Ammino Acids (UAAs) in peptides and peptide mimics could add specific features to these molecules, such as proteolytic stability, active functional groups and new reactivity [80–83]. α,α disubstituted and β-homologs of natural amino acids have been particularly studied during the years for their ability to introduce conformational constrains in peptides and to stabilize specific secondary structures [84–87]. On the other hand, γ UAAs provide a further opportunity to engineer the available backbone through the incorporation of an additional methylene group [88, 89]. Recent studies report that γ-peptides are able to form helices, sheets and turns, whose conformational stability is increased by introducing substituents on the backbone chain [90]. Moreover, cyclic γ UAAs, as gabapentin [91, 92] and γ-cyclohexyl amino acid [93] are able to stabilize both turn and helix conformation in oligomers and in mixed α-γ and β-γ sequences. On the other hand, the use of bicyclic γ amino acids for the stabilization of the peptide structure is much less investigated [94]. We investigated the conformational behavior of both enantiomers of the bicyclic unnatural γ amino acid 4,5,6,6a-tetrahydro-3aH-pyrrolo[3,4-d]isoxazole-3-carboxylic acid (**1**). Unnatural aminoacid **1** is a conformationally constrained dipeptide analog and, in principle, it could substitute two amino acids in a peptide chain (Figure 4.1). The presence of the constrained bicyclic system could induce specific secondary structure allowing a proper orientation of the peptide arms at C- and N- termini. Furthermore, the presence of the isoxazoline ring, a core often found in biologically active compounds, could be particularly useful for future applications in the pharmaceutical

field. In fact, isoxazoline derivatives are important scaffolds found in many naturally occurring and biologically active compounds possessing a wide range of bioactivities, such as antibacterial, antifungal, antiparasitic [95–97], anticancer [98, 99]. Recently, peptidomimetics containing the isoxazoline ring have been reported as β-turn mimics [100, 101]. Starting from the (-)-(3aS6aR)-1 and (+)-(3aR6aS)-1 enantiomers,

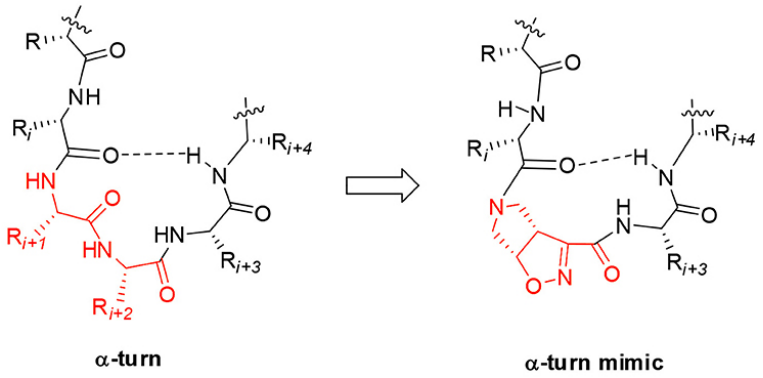


Figure 4.1: The base structure of the mimetic used in this work.

model peptides containing Leu-Val dipeptide at C-terminus and variable sequences at N-terminus were prepared (Figure 4.2). Their conformational behavior was investigated by NMR spectroscopy, FT-IR, circular dichroism, and molecular modeling. NMR and FT-IR results indicated that (+)-(3aR6aS)-1 enantiomer, in combination with glycine, is effective in stabilizing the α-turn conformation in peptides. This structural motif occurs quite often in many key sites of proteins, such as enzyme active site, and metal binding domains [102], although few molecules are known to mimic or stabilize it on isolated peptides [103–106].

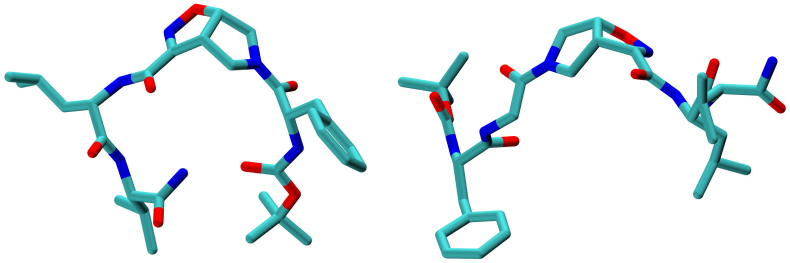


Figure 4.2: The two enantiomers studied. **Peptide 1** on the left and **Peptide 2** on the right.

4.2 Computational Details

Two peptides were modeled in explicit solvent with periodic boundary conditions using WTMD and restrained molecular dynamics simulations. Each molecule was solvated in a cubic box of 4 nm with chloroform (**Peptide 1**) and acetonitrile (**Peptide**

2) to reproduce the NMR experiment conditions. Every system has been submitted to geometry optimization with the steepest descent algorithm with a convergence of 100 kJ·mol^{-1}·nm^{-1}. Then we performed a 1 ns equilibration in NVT conditions at 300 K, followed by a 1 ns NPT equilibration at the 1 bar and at the same temperature. After the equilibration phase, we ran a 50 ns WTMD with a 8.0 kJ·mol^{-1} bias-factor, at a reference temperature of 300 K. To study the open and closed state of both peptides, we selected as CVs the distance between C_β VAL and the N-tert-butoxy carbonyl (N-Boc) quaternary carbon. A Gaussian hill with $\sigma = 0.1$ Å and an initial height of 1 kJ·mol^{-1} was added once every 100 steps. The simulation was sped up by saving the Gaussian hills on a grid that ranged from 0 to 16 Å and spaced 0.02 Å. **Peptide 2** was also simulated using a 50 ns restrained molecular dynamics. Special potentials are used for imposing restraints on the motion of the system, to include knowledge from experimental data such as the NMR derived interatomic distances. Distance restraints add a penalty to the potential when the distance between specified pairs of atoms exceeds a threshold value [107]. The threshold values used in our simulation are centered on the NOESY ones and are reported in Table 4.1. The restraints are applied in a progressive way to give time to the peptide to relax, using a τ equal to 500 ps. Since we are using multiple distance restraints, they are not necessary satisfied at each simulation step but as a time average. The two peptides have been described using the Amber99SB-ILDN Force Field [8]. 4,5,6,6a-Tetrahydro-3aH-pyrrolo[3,4-d]isoxazole-3-carboxylic acid residues were parameterized according to standard procedures as explained in previously reported work [108]. During MD and WTMD simulations, temperature was held constant with the v-rescale algorithm [72], while pressure was kept constant through the Berendsen barostat [73]. MD simulations were performed using the leap-frog algorithm with 2 fs time-step, with holonomic constraints on every bond enforced using the LINCS algorithm. Simulations and subsequent analysis were performed with the GROMACS 5.0.4 [109] program suite, while WTMD was run along with PLUMED version 2.2.2 [110].

Atom	Atom	NOESY Value (Å)
C_αGly	C_βLeu	2.73
C_αGly	H_{NH}Val	3.03
H_{NH}Leu	H_{NH}Val	2.68
H_{NH}Phe	H_{NH}Gly	2.96

Table 4.1: Values used as reference for the simulations where distance restrains were applied.

4.3 Results

From NMR studies, it was found that both model peptides **1** and **2** are present as a mixture of conformers in 2:1 ratio in $CDCl_3$ solution. The presence of these two conformers is probably due to the low-energy barrier associated to cis/trans isomerization of the tertiary amide on the pyrrolidine as it is frequently observed on the acylated proline and in the case of tertiary cyclic amides [111, 112]. Regarding **Peptide 1** this rotation led to a splitting of the (L)-Phe proton signals only. Furthermore, no significative Noesy effects were detected, and, in variable temperature experiments, the obtained $\Delta\delta/\Delta T$ coefficient is high for all the amide protons. All these data together, point towards a specific conformation where Phe and Leu-Val dipeptide are oriented in opposite directions and that the 3aS6aR stereochemistry of the scaffold is not effective in inducing specific secondary structures when used in combination with (L)-α-amino acids. A different scenario was observed for **Peptide 2**. Its two rotamers are indeed characterized by different chemical shifts, suggesting that the isomerization of the tertiary amide leads to two different structures, shown in Figure 4.3. Unfortunately, due to several overlapping signals, it was not possible to assess significative Noesy proximities. Variable temperature experiments were thus done, in order to investigate if the NH were involved in hydrogen bonds and, as a consequence, if the two isomers of **Peptide 2** were characterized by a stable conformation in solution. In the case of the major isomer, the obtained $\Delta\delta/\Delta T$ is around 4 ppb/K for NH-Phe, indicating that this proton could be involved in a weak/medium hydrogen bond. All the other amide protons had higher $\Delta\delta/\Delta T$. On the other hand, in the minor conformer the obtained $\Delta\delta/\Delta T$ is of around 2 ppb/K for NH-Val, indicating that this proton is involved in a strong hydrogen bond. Metadynamic studies were then performed on the cis/trans tertiary amide conformers of **Peptide 2**. In particular, we performed two 50 ns long WTMD simulations, using the distance between Boc quaternary carbon and C_βVal as collective variable (CV). This geometric parameter was selected as a suitable CV because it discriminates well between closed and extended states of the peptide. In this way, it could be possible to evaluate if the two conformers had an intrinsic tendency toward turn conformation. The cis isomer showed a broad free energy minimum corresponding to CV values between 5 and 8 Å, while the trans isomer showed much higher free energy values in this region, exhibiting a shallow minimum around 10–15 Å (Figure 4.4). In order to elucidate the conformation of **Peptide 2**, a restrained molecular dynamic was performed. The restraints were based on the experimental NOESY values. When the hydrogen was not uniquely defined, like in H_αGly and H_βLeu, we used the closest carbon, e.g., C_αGly and C_βLeu to implement distance restrains (Figure 4.6). The analysis showed that C_αGly-C_βLeu and C_αGly-H_{NH}Val, (a) and (b) respectively in Figure 4.6 are mutually exclusive in the trans isomer but not in

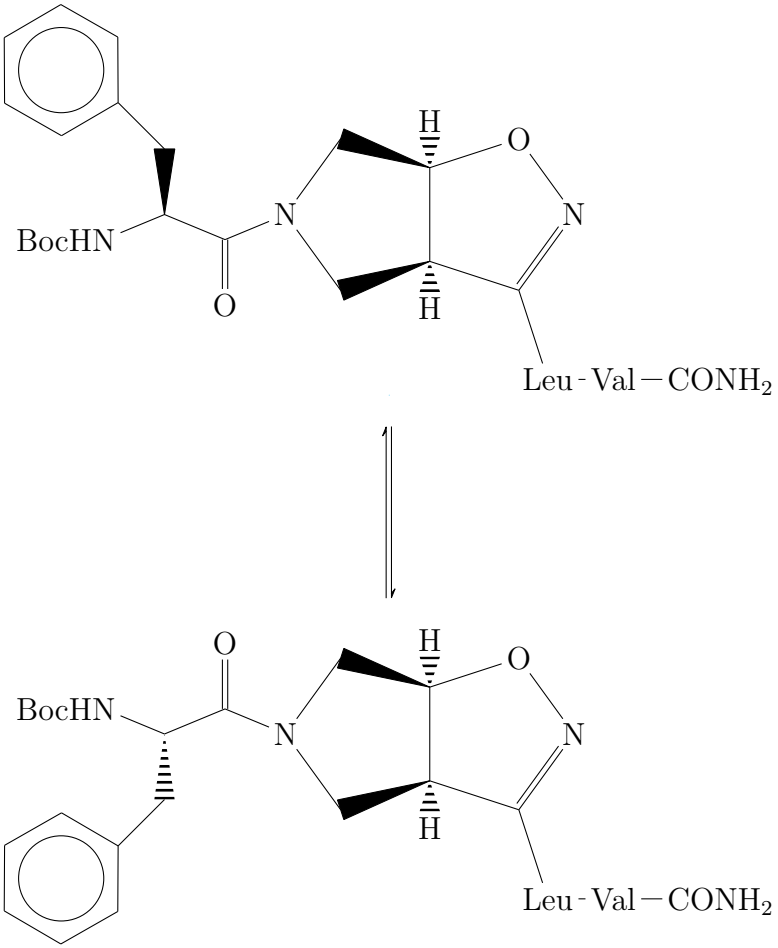


Figure 4.3: The possible rotamers of **Peptide 2**. The trans peptide is reported at the top of the picture while the cis is below.

the cis one as shown in Figure 4.5. Furthermore, H-bond analysis resulted in the observation that a H-bond is between CO-Gly and NH-Val. In Figure 4.7 the structure extracted from the MD simulation in agreement with the experimental data is reported.

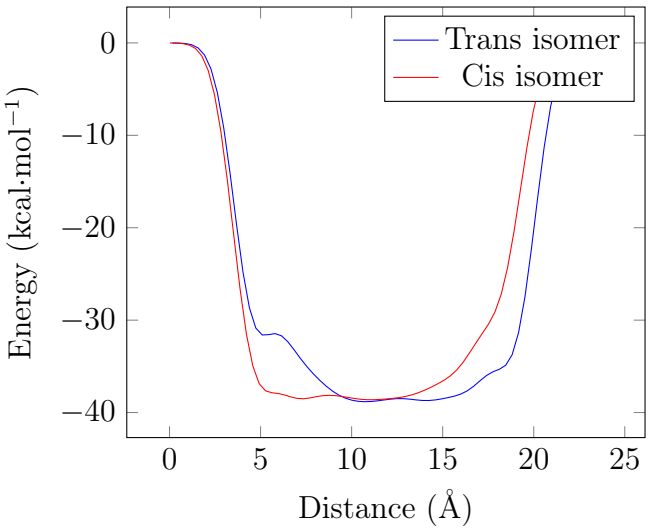


Figure 4.4: Potential energy as a function of distance between Boc quaternary carbon and C_βVal obtained from Metadynamics for the two isomers of **Peptide 2** in chloroform.

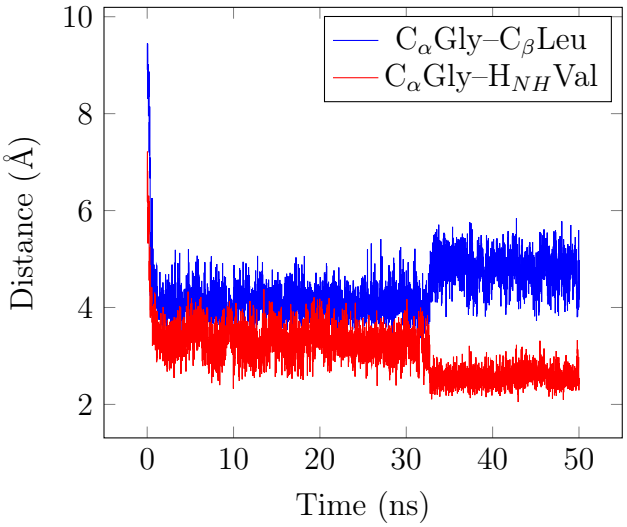


Figure 4.5: Distances between C_αGly–C_βLeu and C_αGly–H_{NH}Val.

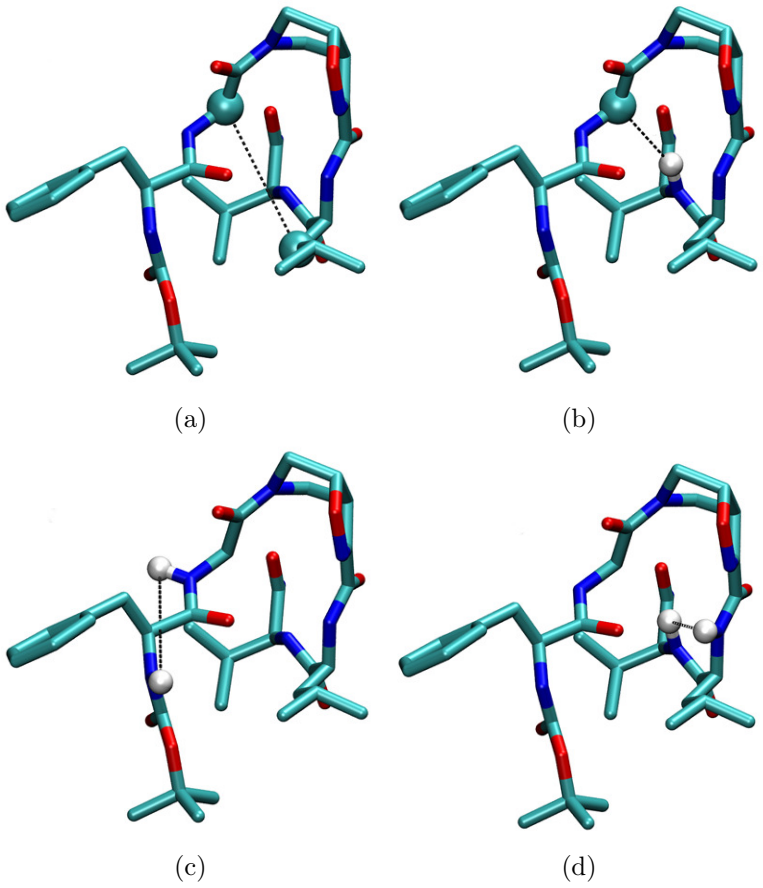


Figure 4.6: Interatomic distances restrained during the simulation on **Peptide 2**. C_αGly-C_βLeu (a). C_αGly-H_{NH}Val (b), H_{NH}Phe-H_{NH}Gly (c). H_{NH}Leu-H_{NH}Val (d). Hydrogens are omitted for clarity.

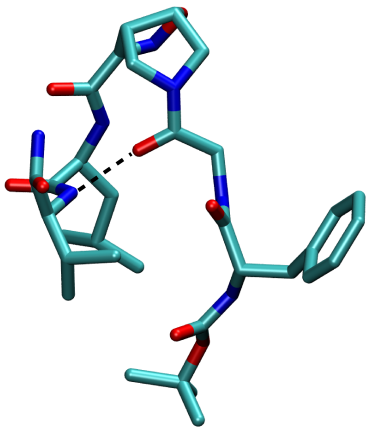


Figure 4.7: **Peptide 2** trans during the restained MD. The hidrogen bond between CO-Gly and NH-Val is highlighted.

4.4 Conclusions

Taking together both experimental and molecular modeling data, we can assume that, in solution, **Peptide 2** effectively adopts a α-turn conformation. This motif is normally formed by 5 α-amino acids and is characterized by a H-bond between i and i+4 residues. In our case, the isoxazoline scaffold replaces two of the three core amino acids 4.1 and the H-bond is formed by the CO-Gly and NH-Val, leading to the formation of a 12 member pseudo-cycle. Furthermore, the overall structure of the peptidomimetic is reinforced by medium-weak H-bonds involving NH-Leu and one C-terminus-NH2, as evidenced from NMR data. The presence of a α-turn conformation is also confirmed by the computational analysis of the dihedral angles of the residue i+3 and of the distances between residues i and i+4 as shown in Table 4.2, as their values rely within the α-turn parameters [113].

Φ_{i+3} (°)	Ψ_{i+3} (°)	$C_{\alpha_i} - C_{\alpha_{i+3}}$ (Å)	$C_{\alpha_i} - C_{\alpha_{i+4}}$ (Å)	$O_i - N_{i+4}$ (Å)
-39.2	-44.4	4.96	5.44	3.37

Table 4.2: Dihedral angles and distances.

In conclusion, as α-turns are often found on biologically active sites and few molecules have been reported to mimic or stabilize them, the ability of compound (+)-1 to stabilize α-turn conformation on isolated peptide is particularly important in view of future biological applications.

Chapter 5

In Silico Design of YeaZ Self Assembly Inhibitor Peptides

5.1 Introduction

Nucleoside modification is an essential part of Transfer Ribonucleic Acid (tRNA) maturation [114]. Over 100 modifications have been observed [115] but only few of them are universally conserved, among them threonylcarbamoyl adenosine (t^6A), found at position 37 in the anticodon loop of a subset of tRNA [116], which is extremely importat to keep tranlation fidelity [117]. *In vivo* studies on t^6A devoid yeast exhibit translation related phenotypes such as initiation and frame maintenance defects and increased nonsense suppression [118–120]. The biosynthetic pathway responsible for t^6A modifications is a challenging problem and part of it remains undiscovered [121]. Two family of universally conserved proteins have been linked to t^6A biosynthesis: YrdC/Sua5 [122] and YgjD/Kae1/Qri7 [120, 123] however *in vitro* studies suggested that other proteins should be involved in the modification process and specifically that YjeE and YeaZ form an essential interaction network [124]. Between the ones nominated here, YeaZ is the only one that is not conserved in eukaryotes and expressed only by bacteria, such as *Pseudomonas Aeruginosa* and *Escherichia Coli*, this makes it an interesting system to target since antimicrobial resistance is becoming a more and more pressing problem of our society. As shown in Figure 5.1, YeaZ crystallographic structure is a pentamer. We decided to focus our attention on a dimer to develop a small peptide that would be able to inhibit the PPIs. For this reason we focused on a dimer, shown in Figure 5.2

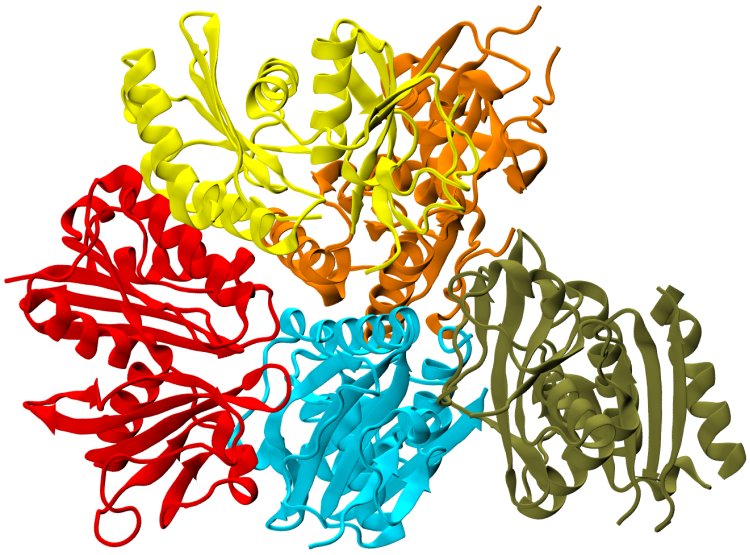


Figure 5.1: Crystallographic structure of YeaZ from *P. Aeruginosa* [125].

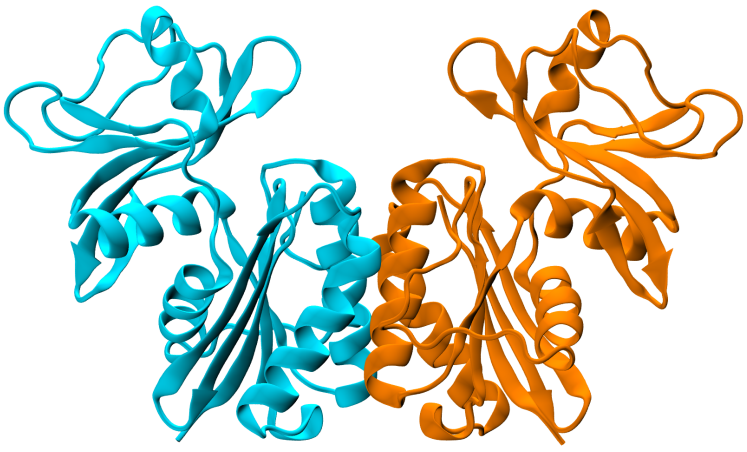


Figure 5.2: The homo-dimer we considered. On the left, in cyan, what will be reffered as "second subunit" and on the right, in orange, the "first subunit".

5.2 Computational Details

The structure of YeaZ was obtained from the protein data bank (PDB ID code 4Y0W [125]). We extracted from the crystal structure two monomers. All systems were solvated with explicit water solvent, employing periodic boundary conditions with initial box dimensions of ≈ 93 Å x 120 Å x 83 Å for the dimer simulation, ≈ 110 Å x 110 Å x 110 Å when simulating the enzime-peptide system and ≈ 62 Å x 62 Å x 62 Å when simulating the peptides by themself. The systems were described using the amber99SB-ILDN [8] force field, TIP3P [57] model was adopted for water. The number of atoms for each system are reported in Table 5.1

System	Protein	Peptide	Water	Ions	Total
Dimer	6462	0	86139	20	92621
Protein and **Peptide 1**	3231	202	26864	10	30307
Protein and **Peptide 2**	3231	194	38094	10	41461
Peptide 1	0	202	24399	0	24601
Peptide 2	0	194	24501	0	24695

Table 5.1: Number of atoms in each system.

We optimized the initial systems geometry with a 50000 steps of energy minimization with the steepest descent method. We then carried out a 200 ps equilibration in the NVT ensemble. The NVT was followed by a second equilibration of 200 ps in the NPT ensemble. Temperature and pressure were kept constant to their reference values (1 bar, 300 K respectively) through the velocity rescale algorithm [72] and the Berendsen barostat [73]. A 14 Å cutoff was applied for non-bonded interactions and the Particles Mesh Ewald algorithm [74] was employed to calculate long range electrostatic interactions. During the MD simulations all bonds lengths were constrained to their equilibrium values with the LINCS algorithm [6], allowing time step of 2 fs. Simulations and subsequent analysis were performed with the GROMACS 5.0.7 [109] program suite. To analyze which aminoacids were directly involved in the interaction between two monomers we used the SASA analysis implemented in GROMACS [126] and we considered meaningful a SASA variation above 0.1 nm^2. The CAS was run using a tool also implemented in GROMACS, GMXPBSA [127].

5.3 Results

The composition of the protein-protein interface was assessed using SASA analysis with which we calculated which aminoacids were actually in contact with the other monomer by measuring the SASA in the complex and in each monomer by itself. The

results of the analysis are shown in Table 5.2 We ran the CAS on all the aminoacids

First Subunit		Second Subunit	
Residue	ΔSASA (nm^2)	Residue	ΔSASA (nm^2)
Ala 68	0.000	Ala 68	0.000
Phe 69	0.006	Phe 69	0.079
Thr 70	**0.200**	Thr 70	**0.335**
Gly 71	0.059	Gly 71	0.000
Val 72	0.003	Val 72	0.000
Arg 73	**1.000**	Arg 73	**0.793**
Ile 74	**0.932**	Ile 74	**0.627**
Ala 75	0.000	Ala 75	0.000
Ile 76	0.030	Ile 76	**0.100**
Gly 77	**0.379**	Gly 77	**0.428**
Val 78	**0.299**	Val 78	**0.295**
Val 79	0.000	Val 79	0.000
Gln 80	**0.537**	Gln 80	**0.588**
Gly 81	**0.425**	Gly 81	**0.432**
Leu 82	**0.276**	Leu 82	**0.278**
Ala 83	0.000	Ala 83	0.000
Phe 84	**1.175**	Phe 84	**1.251**
Ala 85	**0.671**	Ala 85	**0.654**
Leu 86	**0.111**	Leu 86	**0.111**
Gln 87	0.097	Gln 87	**0.219**
Arg 88	0.000	Arg 88	0.000

Table 5.2: SASA analysis results. The aminoacids with a ΔSASA greater or equal to 0.1 nm^2 are bolded.

with a ΔSASA equal or greater than 0.1 nm^2. We mutated from Thr 70 to Leu 86 the first subunit and from Thr 70 to Gln 87 the second. We found that the first subunit has 4 residues with a ΔΔG above 2 $kcal \cdot mol^{-1}$ while the second subunit only has 2, as shown in Table 5.3. This led us to use the first subunit as starting point for our peptide. We also noticed that mutating Arg 73 into an alanine stabilizes the binding so we decided to keep this mutation for our peptide. The sequence of this peptide, from now on referred as **Peptide 1** is: AIAIGVVQGLAFAL. We then removed the rest of the first subunit, as shown in Figure 5.3 (a). Of course, by exposing the peptide directly to the solvent it is possible that it loses its ability to retain the secondary structure and to bind the second subunit. To address this two problems we ran again the CAS on the system made by the second subunit and **Peptide 1** to confirm that the hotspots are retained. As presented in Table 5.4 where the ΔΔG of the aminoacids that form **Peptide 1** are compared with the same region in the subunit and is shown that there is not a considerable difference between the two. A 411 ns simulation was run to test the stability of the peptide's secondary structure. After 100 ns we see that **Peptide 1** is still in its secondary

First Subunit		Second Subunit	
Mutation	$\Delta\Delta G$ (kcal·mol^{-1})	Mutation	$\Delta\Delta G$ (kcal·mol^{-1})
Thr70Ala	-0.5497	Thr70Ala	0.3585
Val72Ala	0.0956	Val72Ala	0.0717
Arg73Ala	-13.0972	Arg73Ala	-8.6279
Ile74Ala	**3.5133**	Ile74Ala	0.6214
Ile76Ala	0.2629	Ile76Ala	0.8843
Val78Ala	**2.1032**	Val78Ala	1.5774
Gln80Ala	-0.1673	Gln80Ala	-1.4340
Leu82Ala	**3.2982**	Leu82Ala	**2.2705**
Phe84Ala	**7.5524**	Phe84Ala	**10.6594**
Leu86Ala	1.6491	Leu86Ala	1.3862
-	-	Gln87Ala	-0.9560

Table 5.3: CAS results. The aminoacids with a $\Delta\Delta G$ greater or equal to 2 kcal·mol^{-1} are bolded and considered hotspots.

structure and well attached to the peptide as shown in Figure 5.3 (b) finally the peptide starts to lose the secondary structure towards the end of the simulation and to unbind from the second subunit the final structure is shown in Figure 5.3 (c). The peptide, when simulated by itself in water, is not stable and rapidly switches from an helix to a turn.

(a) (b) (c)

Figure 5.3: Snapshots of the second subunit and **Peptide 1**, taken at different simulation time. At the start of the simulation (a), after 100 ns (b) and the final conformation after 411 ns. The peptide stays bound to the secondary subunit for most of the simulation and starts to lose the secondary structure and to leave the binding site only towards the end.

First Subunit		**Peptide 1**	
Mutation	$\Delta\Delta$G (kcal·mol^{-1})	Mutation	$\Delta\Delta$G (kcal·mol^{-1})
Ile74Ala	**3.5133**	**Ile74Ala**	**3.1309**
Ile76Ala	0.2629	Ile76Ala	1.195
Val78Ala	**2.1032**	**Val78Ala**	**2.2227**
Gln80Ala	-0.1673	Gln80Ala	0.7887
Leu82Ala	**3.2982**	**Leu82Ala**	**2.868**
Phe84Ala	**7.5524**	**Phe84Ala**	**6.8354**

Table 5.4: Second CAS results. The aminoacids with a $\Delta\Delta$G greater or equal to 2 kcal·mol^{-1} are bolded and considered hotspots.

5.3.1 Structural stability

To increase the overall helicity of the peptide we tried to mutate the Ile 76 and Gln 80 into an Orn 76, a non standard aminoacid of which the bidimensional structure is shown in Figure 5.4, and an Asp 80 with their sidechains condensed (**Peptide 2**) as shown in Figure 5.5, to constrain the helix increasing its rigidity. This approach was already used in previous work [128]. This did not prove to be a successful approach since after a very short simulation time **Peptide 2** started to unfold and rapidly detached from the second subunit.

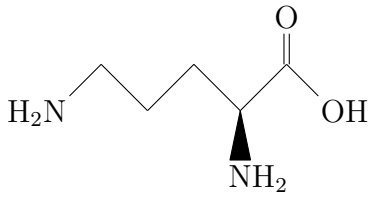


Figure 5.4: Ornithine two dimensional structure.

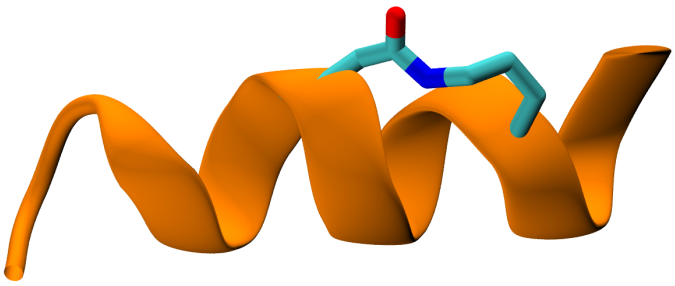
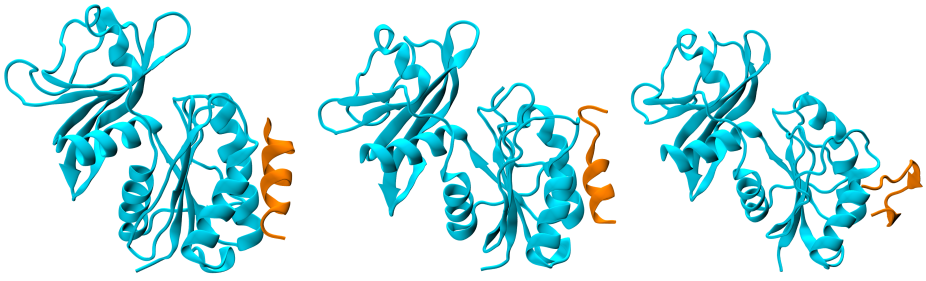


Figure 5.5: **Peptide 2**. The sidechains of Orn and Asp are drawn in licorice, the hydrogen are removed to ease the visualization.



(a) (b) (c)

Figure 5.6: Snapshots of the second subunit and **Peptide 2**, taken at different simulation time. At the start of the simulation (a), after 100 ns (b) and the final conformation after 411 ns. **Peptide 2** starts to lose the secondary structure after 100 ns and it has completely left the second subunit after 411 ns.

To assess the role of the secondary subunit to retain the helicity of the two peptides we simulated **Peptide 1** and **Peptide 2** by themself in water for 100 ns each. Our simulations show that **Peptide 1**, which was very stable when bound to the secondary subunit, is not able to retain its secondary structure by itself and rapidly loses the helical conformation as shown in Figure 5.7 where the secondary structure for each residue is reported over time (top) and as a fraction of the trajectory (bottom). As expected from the simulation of **Peptide 2** with the protein, the unfolding was even faster. This is a strong indicator that **Peptide 1** is only able to stay folded when bound to the second subunit while **Peptide 2** can not retain the helical conformation.

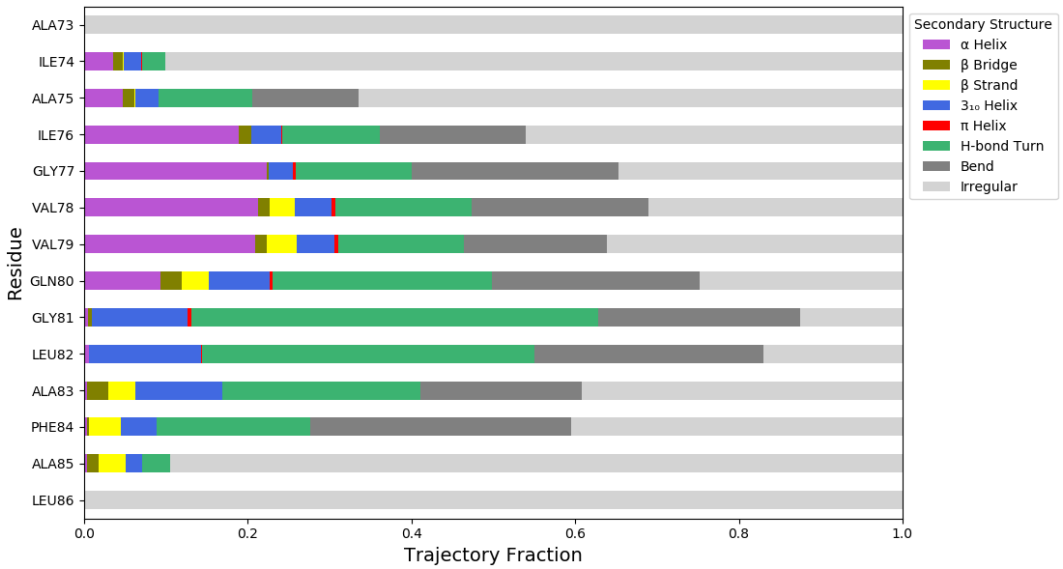


Figure 5.7: Secondary structure analysis of **Peptide 1**

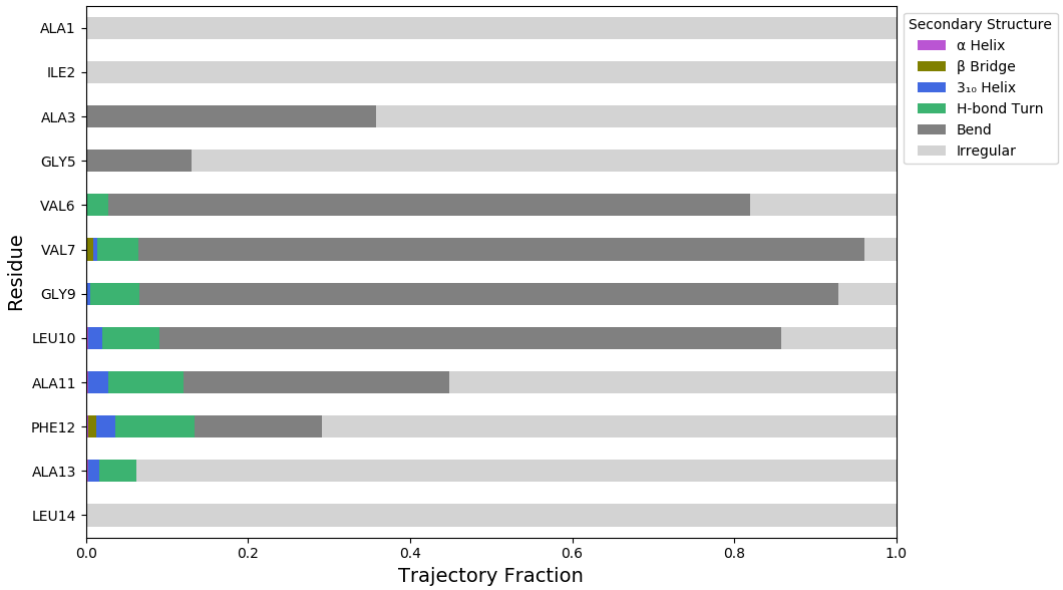


Figure 5.8: Secondary structure analysis of **Peptide 2**

5.4 Conclusions

Using SASA and CAS we were able to locate a sequence of aminoacids on the subunit one that are likely to interfere with the polymerization of YeaZ monomers by binding to a monomer and preventing another one to bind it efficiently by competitively binding the aminoacids most responsible for the polymerization. The structure stability problem resulted to be very challenging since it is not immediately clear how to improve it. We tried to stiffen the helix in order to obtain a more stable helical structure but unsuccessfully. It is possible that with a longer bridge between the two mutated residues, the helix stability would improve. Another approach could be to completely change the chemical nature of the bridge to add a chemical stabilization effect to the physical one. We also proved that neither of the two peptides is able to retain the helical structure by itself and **Peptide 1** needs the secondary subunit to be able to maintain the secondary structure for a longer period of time. This is not necessarily bad, since many peptides are known to only fold on the protein surface.

Chapter 6

Glutamine Recognition and Binding by γ-Glutamyltransferase

6.1 Introduction

The enzyme-substrate molecular recognition process is a rather complex sequence of events which takes place on the microseconds to milliseconds time scale. In particular, during the substrate recruiting stage, the ligand must extensively sample both roto-translational and conformational degrees of freedom in order to reach the bound state into the catalytic site where it can undergo chemical reactions. Such preliminary steps are nevertheless critical for enzyme activity and specificity, so an atomic level knowledge of the sequence of enzyme-substrate interactions occurring through the recruiting phase is highly desirable both to better understand the origin of such specificity and to pursue the rational design of improved enzymes for chemical synthesis [129]. In this work, we focused our attention on γ-glutamyltransferase, also referred to as γ-glutamyltransferase (GGT) (E.C. 2.3.2.2), a heterodimeric enzyme able to cleave amide bonds and widely distributed from bacteria to plants and mammals [130,131], i.e. GGT is a virulence factor of Helicobacter pylori, a bacteria, that enables its widespread infection responsible for chronic gastritis and increased risk of gastric cancer [132]. To explore the substrate recruiting and binding to GGT catalytic site, we performed aMD simulations. Using aMD we were able to simulate the binding of glutamine to EcoGGT catalytic site (Figure 6.1). All the members of the GGT superfamily are autocatalytically activated through a proteolytic cleavage, which liberates a N-terminal nucleophilic, catalytically active residue [133]. GGT cleaves γ-glutamyl compounds, e.g. glutathione, releasing the cysteinylglycine portion and forming a covalently bound γ-glutamyl-enzyme intermediate through an ester linkage between the γ-carbon atom of the glutamyl group and the oxygen atom of the catalytically active, N-terminal threonine residue of the small enzyme subunit [134] (Figure 6.1). The γ-glutamyl-enzyme intermediate is then resolved by

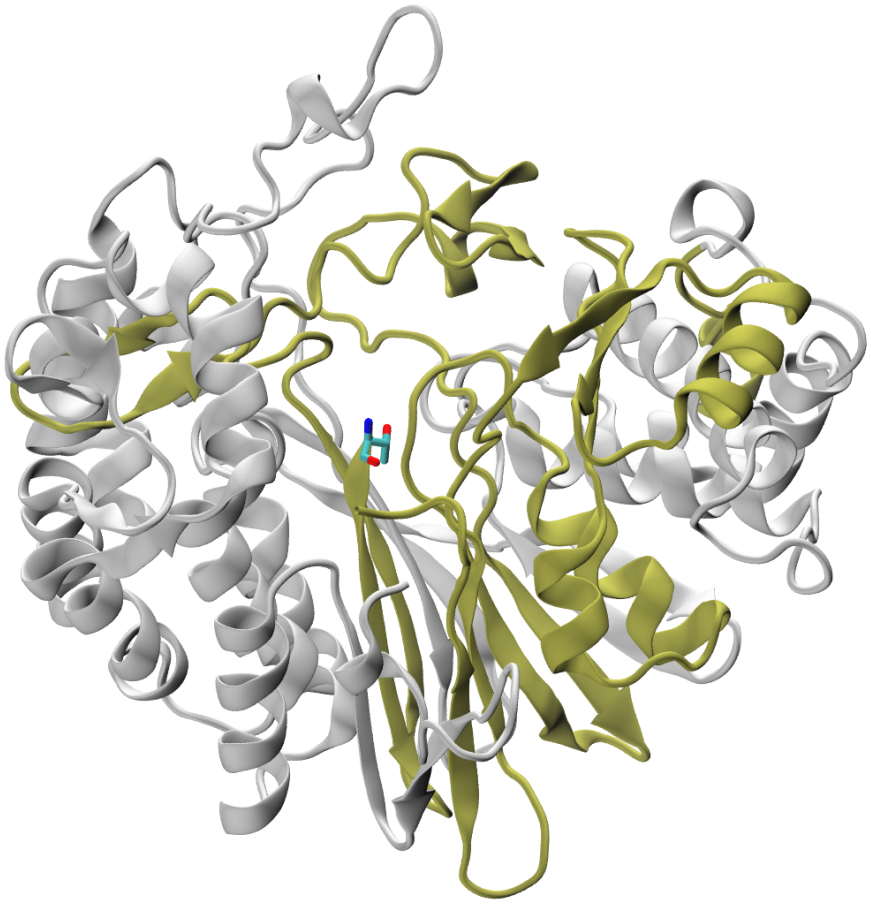


Figure 6.1: Crystal X-ray structure of E. Coli GGT (EcoGGT) (PDB ID code 2DBX). Silver and tan color ribbons represents the heavy and light sub domains. The side chain of the reactive residue (Thr 391) is represented in Licorice. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color, respectively.

a water molecule (hydrolysis) or by the transfer of the glutamyl group to an amino acid or a short peptide (transpeptidation) [130, 135–138] (Figure 6.2). EcoGGT has periplasmic localization and it was suggested that its physiological role is related to the hydrolysis of glutathione and other γ-glutamyl derivatives [139]. As a member of the NtN hydrolase superfamily, the catalytically active residue of EcoGGT was identified in the N-terminal threonine residue of the small subunit (Thr 391) [134] (Figure 6.1). The crystal structure of EcoGGT has been solved unbound, in complex with an intermediate and the corresponding hydrolysis product [140, 141] and, with the aid of structural information, light was shed on the autocatalytic process leading to enzyme maturation [142] and on some insights of the reaction mechanism [143]. Other residues essential for enzyme maturation were identified through site-directed mutagenesis [144] and/or from activity studies [145, 146]. EcoGGT found use also as a biocatalyst for the synthesis of γ-glutamyl derivatives at a preparative level [147], accepting L- but not D-amino acids and has a preference for basic and aromatic acceptor amino acids [148]. Glutamine, a high affinity substrate of EcoGGT [149],

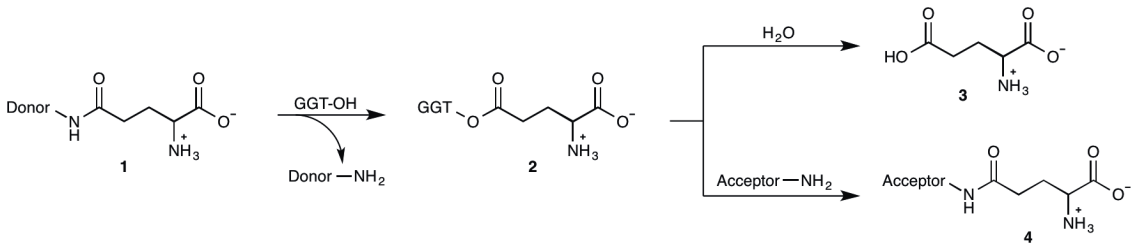


Figure 6.2: Schematic representation of GGTs mechanism. At first there is the reaction between GGT and the γ-glutamyl donor **1**, which lead to the formation of the γ-glutamyl enzyme intermediate **2**. The intermediate can react in two ways: if the γ-glutamyl moiety is transferred to a water molecule, hydrolysis occurs with liberation of glutamic acid **3**; if the acceptor substrate is a peptide, the result is a new γ-glutamyl derivative **4** through a transpeptidation reaction.

is usually employed as a cheap and readily available γ-glutamyl donor but, due to its high propensity to act also as acceptor, important amounts of γ-glutamylglutamine byproduct were obtained in these reactions. The general picture of GGT's catalytic cycle is widely described and commonly accepted and shortly described in Figure 6.2. However, a crystal structure of the EcoGGT-glutamine complex is not available, as the substrate almost immediately reacts upon binding and it is thus impossible to obtain a co-crystal. Furthermore, little is known about the molecular basis underlying the glutamine recognition by EcoGGT, and its elucidation will provide insight on the general recognition pathway of other substrates by the GGT family. Molecular dynamics (MD) simulations are increasingly used to get insight on fundamental biomolecular processes such as protein folding, allostery regulation, and protein-ligand binding [150–152]. Conventional MD simulations are usually limited to a few microseconds in conventional high-performance computing clusters, which limit the modelling of multi-microsecond time scale processes such as protein-ligand binding. To overcome these limitations, several enhanced sampling methods, such as replica exchange MD, metadynamics, umbrella sampling, transition path sampling and accelerated molecular dynamics (aMD) [31,34,153–155], have been developed to speed up phase space exploration. In particular, aMD is an enhanced sampling technique that works by reducing energy barriers of the molecular potential energy surface, adding a non-negative boost potential when the system potential is lower than a reference energy [34]. It has been employed to simulate rare events that occurs in the microseconds to milliseconds time scale, such as protein folding and ligand-protein binding, with tens to hundreds of nanoseconds long simulations [156–160]. A similar modelling approach has been used before by McCammon and coworkers [160] who applied microsecond long aMD simulations to model ligand binding to the M3 muscarinic membrane protein receptor. In this work, we employed a different computational strategy, performing an ensemble of independent aMD repeats and

focusing on a subset of trajectories where GGT-glutamine binding was observed. Our approach allowed us to model the process of substrate recruiting by the enzyme and to identify recurring clusters that could describe the approach to GGT of a glutamine molecule, from the free unbound state to its full insertion into the active site. Notably, no restraint was applied and no prior information or hypothesis concerning the enzyme-substrate interactions were used in the simulations.

6.2 Computational Details

The structure of EcoGGT was obtained from the protein data bank (PDB ID code 2DBX [141]). The crystal structure contains a dimer of the protein and the subunits are labeled A, B, C and D. The resolution was 1.7 Å, subunits C and D were removed as well as all the water molecules, two calcium ions, one molecule of glycerol and one molecule of glutamate. In the starting structure of each aMD run glutamine was placed away from EcoGGT and each run has different random seed numbers and initial velocities. The system was simulated in explicit water solvent employing periodic boundary conditions with initial box dimensions 100 Å x 101 Å x 92 Å. Glutamine and EcoGGT were described with the amber99SB [70] force field and the TIP3P [57] model was adopted for water. The total number of atoms was 89736, divided in 8093 protein atoms, 20 glutamine atoms, 14 sodium ions and 81608 water atoms (27203 water molecules). We optimized the initial system geometry with a 50000 steps of energy minimization, using OpenMM 7.2.1 CUDA kernels [161]. Hydrogen bond lengths were constrained to their equilibrium values with the SETTLE algorithm [162]. Glutamine starting position was selected by taking the last frame of the central simulation ran using Random Acceleration Molecular Dynamics (RAMD) (detailed in Appendix 6.A.1 [163]) that was applied to bring the ligand outside the binding pocket. Position restraints were imposed on heavy atoms in the protein with a spring force of 100 kcal·mol^{-1}·Å^{-2}. We then carried out a 40 ps equilibration in the NVT ensemble, slowly increasing the temperature from 50 to 300 K in 4 steps and progressively relaxing position restraints from 100 to 4 kcal·mol^{-1}·Å^{-2} in 5 steps. To control the temperature we used a Langevin integrator [164]. A time step of 2 fs was used for all the reported MD runs. The NVT was followed by a second equilibration of 50 ps in the NPT ensemble, during which restraints were fully removed in 5 steps. Pressure control was assured through an isotropic Monte Carlo barostat which attempts to change the volume every 25 steps and accepts the changes with the Metropolis Monte-Carlo criterion [165, 166]. This was followed by an equilibration run of 10 ns and a production run of 50 ns. aMD parameters were calculated using the respective total (E_T) and dihedral (E_{Dih}) averages from the 50 ns production run and were set equal to $E_T = -273982.6$

kcal·mol^{-1} and E_{Dih} = 5662.3 kcal·mol^{-1}. aMD protocol implementation was based on a dual-boosting aMD python script available at the OpenMM virtual repository [167].

6.2.1 Phrases clustering analysis

The phrases clustering analysis is a clustering method for the identification of clusters of protein residues making simultaneous contacts with a ligand through an MD trajectory [168]. It uses a distance metric based on the surprisal of a pair of residues to be in contact with a ligand as shown in

$$d(i,j) = -\log\left(\frac{n(i,j)+1}{N}\right) \tag{6.1}$$

where $n(i,j)$ is the number of simultaneous contacts for residues i and j over a trajectory containing N frames. Note that $d(i,j) \approx 0$ when N is high and the ligand is always in contact with the two residues considered and $d(i,j) = \log(N)$ when there are no simultaneous contacts throughout the trajectory. A ligand is in contact with a protein residue if their heavy-atoms are within 3 Å distance. Clustering is carried out using complete-linkage hierarchical clustering.

6.2.2 Cluster analysis

Trajectory segments from glutamine-enzyme complexes (R10, R20 and R50), were extracted and concatenated into a single data set when the (Root Mean Square Deviation (RMSD)) value between glutamine and the crystal structure was steadily below 10 Å. The binding pocket flexibility is studied by clustering analysis performed [169] using the heavy-atoms of glutamine, residues in the binding pocket (Ser 462, Asn 411, Ser 463, Gly 484, Met 464, Gly 483, Asp 433, Arg 114, Gln 430, Thr 409, Thr391, Ser 485), and residues undergoing conformational changes Pro 482 and Tyr 444. The conformational space sampled by glutamine bound in the binding pocket was analyzed by clustering on a substrate heavy-atom RMSD basis. The gromos method [170] was employed for all reported clustering analysis using a 1 Å cut-off distance. The central structure of each cluster is defined as the structure with the largest number of neighbors. Protein figures rendering was performed with VMD [171]. Two dimensional plots of the substrate-protein interactions were obtained with Ligplot+ [172].

6.3 Results

The starting simulation system consists of the EcoGGT heterodimer structure in the presence of one unbounded glutamine, while solvated in explicit solvent. We performed 90 aMD independent runs, each of them with a simulation length of c.a. 100 ns, in a high-throughput computing cluster. As the nucleophilic attack of the hydroxyl oxygen of Thr391 on the glutamine amide carbon is the first step in the formation of the γ-glutamyl-enzyme intermediate, this C-O distance has been monitored along the trajectories, in order to select those leading to potentially reactive bound conformations. In 5 out of 90 trajectories the C-O distance fell below 10 Å, showing that the substrate approached the protein active site. In 3 of these 5 trajectories, the substrate durably binds into the catalytic site until the end of the MD trajectories (trajectories named R10, R20 and R50 in Figure 6.3 left). Whereas, in the other two trajectories, namely R22 and R30, glutamine rapidly escaped from the binding pocket after less than 20 ns (Figure 6.3 left). Trajectories R20 and R50 show similar final values, while R10 stabilizes around a larger nucleophilic attack distance indicating a different substrate-pocket conformation. The small number of substrate-enzyme binding observations in an ensemble of short aMD runs mirrors the behavior of real-world enzymes, where only a small fraction of enzyme-substrate collisions has the proper steric and energetic requirements to evolve into a reactive encounter. Comparison of the five trajectories shows that a salt bridge interaction between glutamine C-terminus and Arg 114 is essential for the stable binding of glutamine into the binding pocket. The distance between the glutamine carboxyl carbon and the Arg 114 guanidinic carbon (Figure 6.3 right) steadily falls below 5 Å in successful trajectories.

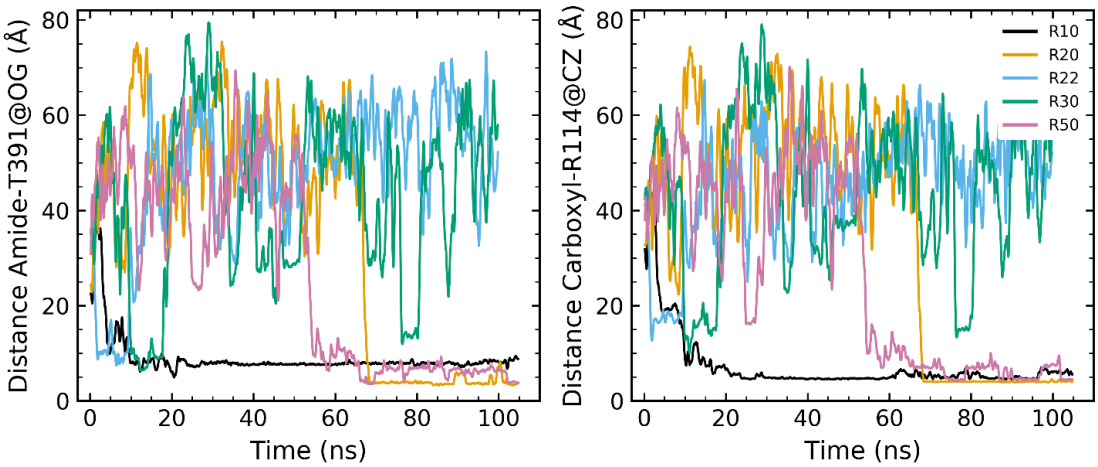


Figure 6.3: Potential energy as a function of distance between Boc quaternary carbon and C$_\beta$Val obtained from Metadynamics for the two isomers of **Peptide 2** in chloroform.

6.3.1 Comparison of simulated glutamine binding to a EcoGGT-glutamate complex

Simulated EcoGGT-substrate bound structures are compared to EcoGGT cocrystallized with glutamate, the product of the enzymatic reaction as shown in the reaction scheme in Figure 6.2. Due to the size and shape similarity between glutamine and glutamate, it is assumed that the position and orientation of the two molecules are similar before and after the reaction takes place (Figure 6.4a). This is tested by aligning the aMD generated protein structures to the crystal protein atom coordinates (PDB ID 2DBX), followed by RMSD calculations using the substrate heavy-atoms (Figure 6.4b). Figure 6.4b shows that the simulated bound structures

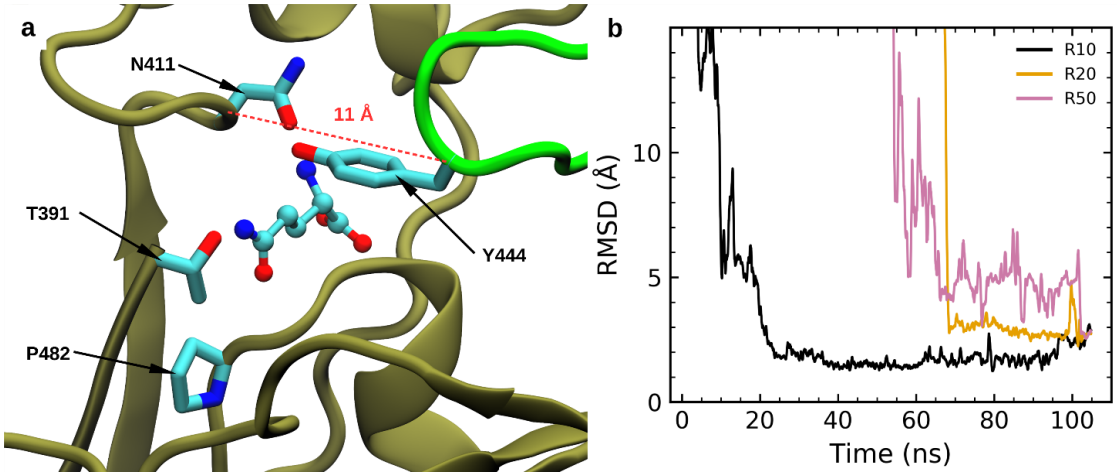


Figure 6.4: (a) Receptor-glutamine complex model built from a co-crystallized receptor-glutamate structure. Lid-loop and chain B of EcoGGT are shown in green and tan color ribbons, respectively. C-C distance between Asn 411 and Tyr 444 is highlighted. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color spheres, respectively. (b) RMSD traces of the substrate (glutamine) in R10, R20 and R50 with respect to the receptor-glutamine complex model. aMD generated protein trajectories were aligned with respect to the crystal X-ray structure and the RMSD values were calculated using only the substrate heavy atoms. Data is smoothed by moving average over 20 frames (400 ps).

agree with a model built using the product crystallographic structure and chemical intuition. In R10, glutamine spends more than 80 ns with an average RMSD c.a. 2 Å relative to the proposed substrate bound conformation, while in R20 and R50 the substrate samples conformations with RMSD values c.a. 3 Å over the last 3 ns.

6.3.2 Specific glutamine binding site interactions

The pocket structure is characterized by the phrases clustering of pair of residues simultaneously in contact with glutamine. The binding pocket contacting residues are divided into two groups denominated I and II each of them containing six residues (see groups definition in Figure 6.5). Asn 411, part of group I, forms simultaneous

contacts with residues in both groups. This is followed in less degree by Ser 462 in the same group I. The catalytic residue Thr 391 is strongly coupled with Thr 409 as part of group II. Residues location in the binding pocket can explain this analysis, where Asn 411 lies inside the pocket close to the reactive residue Thr 391 and Thr 409 (Figure 6.6). In contrast, Ser 462 is located at the near opposite end of the pocket entrance allowing it to form simultaneous contacts with most residues in both groups. Group II is characterized by its highly polar nature with residues Asp 433, Arg 114 and Gln 430 clustered together with Thr 409 and Thr 391. These five residues and Asn 411 are located on one side of the binding pocket opposite to the lid-loop and participate in frequent simultaneous contacts with the substrate providing stabilization through hydrogen bonds. This is complemented by contact analysis calculations between the glutamine and the protein, which show that Asn 411, Arg 114, Thr 409 and Asp 433 form the most frequent contacts with glutamine in the three trajectories R10, R20 and R50 (see Contact Analysis and Figures 6.A.4 - 6.A.9 in Section 6.A.5). The identification of sets of contacting residues through

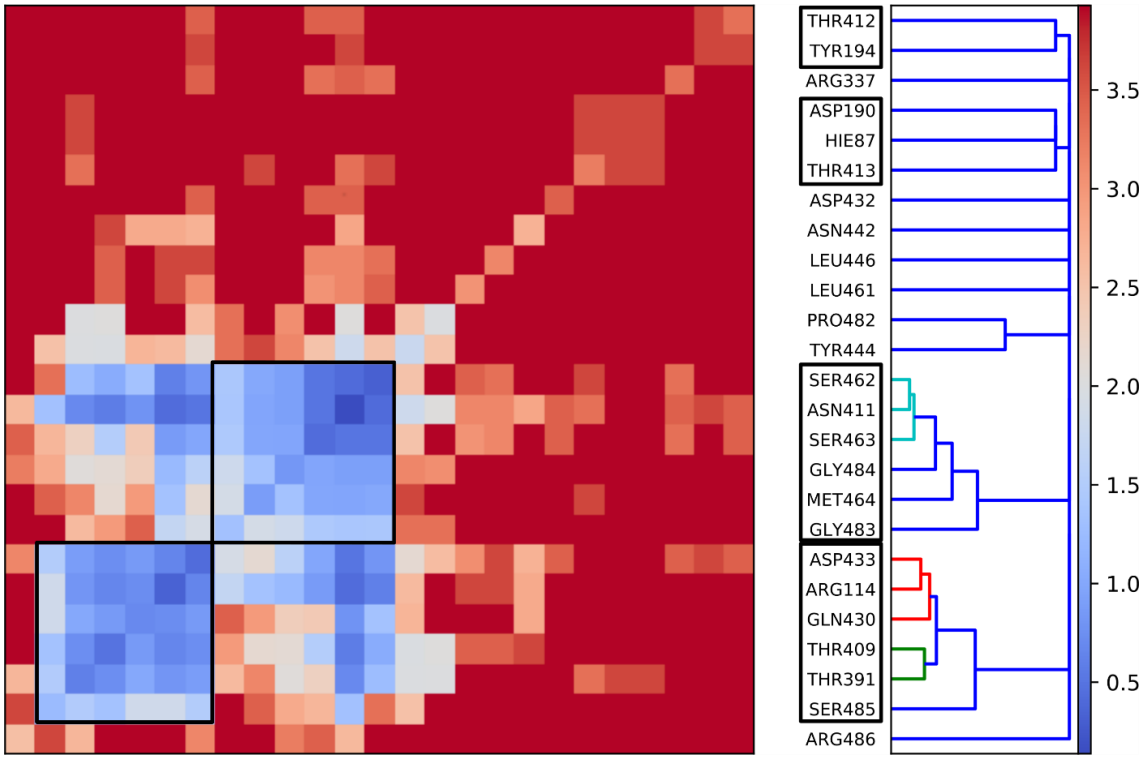


Figure 6.5: Hierarchical clustering of pair of residues forming contacts with glutamine. Group I consists of a polar subcluster (Asn 411, Ser 462 and Ser 463), and a non-polar subcluster (Met 464, Gly 483 and Gly 484). Group II contains six polar residues: Asp 433, Arg 114, Gln 430, Thr 409, Thr 391 and Ser 485. Other group residues highlighted are formed by Thr 412, Tyr 194 and Asp 190, Hie 87, Thr 413 are further described in the text. Pro 482 and Tyr 444 interactions will be further described in the text. The analysis was carried out using frames from all trajectories where the ligand is bound to the protein (described in the Methods section).

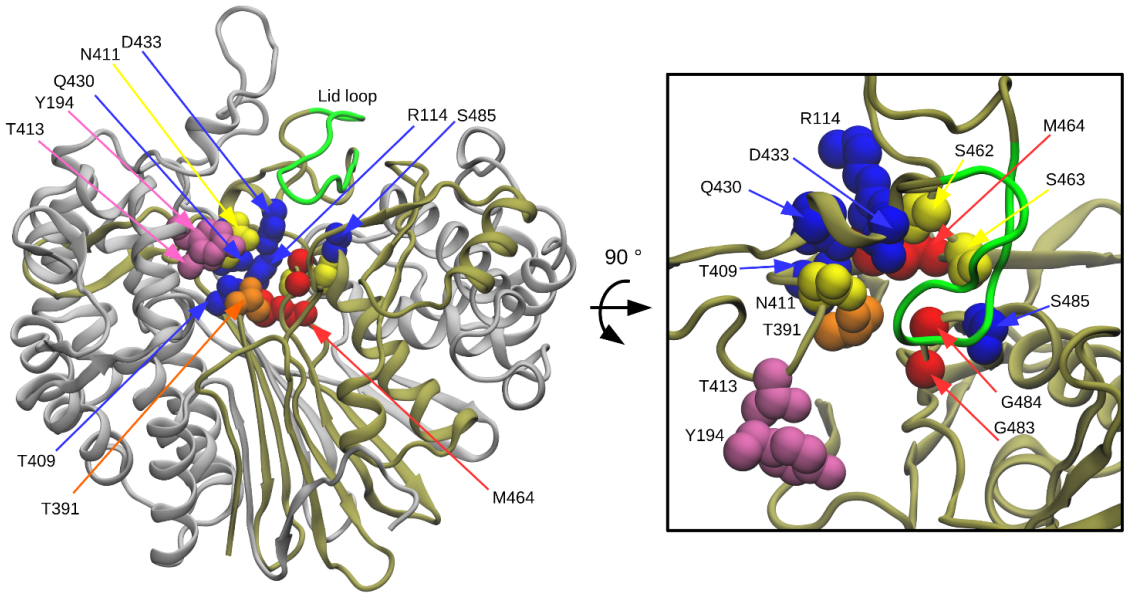


Figure 6.6: Structural representation of residues forming simultaneous contacts with glutamine. Left panel shows the EcoGGT heterodimer with subunits A and B in silver and tan ribbons representations. Lid-loop is shown in green ribbons (residues 438-449). Contacting residues are highlighted: Group I (yellow and red color sidechains in space filling representation), and group II (blue and orange color sidechains). Residues involved in the initial binding process are shown in magenta color representation, see results for further discussion. Right panel shows the top view after a 90° rotation.

phrases clustering is coupled with hydrogen bonding analysis for each glutamine bound trajectory. This allows for the determination of a common hydrogen bond pattern formation through the binding process. In all three successful trajectories, residues Thr 413 and Tyr 194 form short-lived hydrogen bonds with the substrate as it approaches the binding pocket (Figure 6.A.10). Five residues in group I (Ser 463 and Asn 411) and group II (Arg 114, Asp 433 and Thr 409) set of contacting residues are found to form hydrogen bonds when the substrate is deeply buried into the binding pocket (Figure 6.A.10). In our simulations, residues Thr 413 and Tyr 194 are involved in the initial stages of the molecular recognition process, when glutamine transiently binds in proximity of the binding pocket (see structure in Figure 6.6). Other residues such as Thr 412, Asp 190 and His 87 are found to form short lived hydrogen bond during the early binding process in some of the glutamine bound trajectories. The set of five residues, in group I and II, stabilize the substrate deep into the binding pocket through the formation of multiple hydrogen bonds, which stabilize the glutamine reactive amide carbon close to the catalytic site (Thr 391).

6.3.3 Structural role of the lid-loop

The lid-loop consists of twelve residues (438-449) of which Tyr 444 is on the lid-loop tip that allows it to be positioned on the pocket entrance side forming a hydrogen bond with Asn 411 in the crystal structure (Figure 6.6a). Phrases clustering shows that Tyr 444 forms infrequent simultaneous contacts with the substrate. Other lid-loop residues form few simultaneous contacts with the bound substrate (see also Section 6.A.5). Instead, bound glutamine is frequently separated from the lid-loop by water molecules, which is shown by calculating the number of water molecules between the substrate and Tyr 444 (Figure 6.A.11). Trajectories R20 and R50 have 5.2 and 2.1 average water molecules separating the substrate from the lid-loop over the last 15 ns, while R10 shows 0.4 water molecules. The latter displays an average decrease of 1.8 water molecules over the last 25 ns corresponding to a closure of the pocket (see the last MD frames in Figure 6.A.12). To assess lid-loop conformational changes relative to the pocket, the minimum distance between loop tip (Tyr 444), and Asn 411 is calculated (top panel in Figure 6.A.13). In all five aMD runs, the lid-loop tip fluctuates with opening and closing motions relative to a conventional MD (cMD) simulation of the apo state for over one microsecond long (bottom panel in Figure 6.A.13). A similar behavior is observed for their C-C distance, but R10 shows a closure after the substrate binding c.a. 20 ns (top panels in Figures 6.A.13 and 6.A.14). In contrast, the lid-loop adopts a stable conformation with short lived changes of a few nanoseconds through the cMD apo state (bottom panes in Figures 6.A.13 and 6.A.14). In contrast, R22 is the only aMD trajectory that shows a brief loop opening between c.a. 15 ns to 25 ns, while it shows a behavior to the apo MD simulation out of this simulation time range. During the first 10 ns, the protein has a reduced pocket entrance when the substrate is near the binding pocket (Figure 6.3, top panel in Figure 6.A.13). In R30 the substrate fails to move into the binding pocket despite the lid-loop opening (top panel in Figure 6.A.13). Our MD modelling indicates that the lid-loop acts as a steric barrier enclosing the substrate in the pocket and points to a possible function as a gating structure allowing the substrate to bind into the catalytic pocket. However, the lid opening is not the only condition necessary for the substrate binding. Favorable interactions between the pocket polar residues and the substrate can stabilize the latter in the pocket as discussed in the previous subsection.

6.3.4 Structural flexibility of the glutamine-EcoGGT binding pocket

The conformational heterogeneity of the bound substrate and the catalytic site is analyzed by structural clustering. This resulted in a set of conformations where

the top 6 clusters amount for 58% of the data set, and the top two clusters have a combined representation of 35% of the data set (Table 6.1). A main difference between the two top clusters is an opening of the lid-loop in cluster 2 measured by a 7 Å increase of the C-C distance of Tyr 444 – Asn 411 with respect to that distance in cluster 1, which results in a larger binding site entrance (Figure 6.7a and c). Another important difference, cluster 1 shows a conformational change where the N-terminus residue Thr 391 is displaced by Pro 482 (Figure 6.7a). Cluster 1 structures correspond to R10 trajectory. However, this conformational change is not observed in the second top cluster (structures from R20 and R50), in which the glutamine's amide group is in close contact with Thr 391. Common interactions between the two top cluster representatives involve hydrogen bond interactions between Arg 114 and Asn 411 with the substrate and contacts with Gly 484 (Figure 6.7b and d). Comparison between representative structures of clusters 3 and 4 shows that residues in the binding pocket undergo subtle conformational changes that allow residues e.g. Ser 462 and Ser 463 to form hydrogen bonds with the substrate (Figure 6.A.15). Meanwhile, the lid-loop retains its conformation separating the binding pocket from the solvent (Figure 6.A.15). To further dissect the origins of the conformational flexibility, clustering analysis of substrate heavy-atom coordinates showed a dominant cluster accounting for c.a. 80% of the substrate-bound data set (Table 6.A.10). Thus, the substrate relative position and orientation and the contacting residues contribute the most to the substrate-pocket conformational heterogeneity, rather than the substrate conformational flexibility. Displacement of the lid-loop suggests that the loop acts as a gating structure by opening and closing the pocket entrance. Furthermore, conformational changes of pocket residues are possibly associated with the accommodation of different substrates into the catalytic reactive site. The implications of our modeling results will be further discussed in the Discussion and Conclusions.

Cluster ID	Number of Structure	Cluster weight (%)
1	1842	21.7
2	1154	13.6
3	689	8.1
4	640	7.5
5	327	3.9
6	276	3.3

Table 6.1: Clustering analysis results of the conformational space of glutamine and binding pocket residues. The number of structures for each cluster is indicated. The cluster weight is calculated over the entire set of frames. The top six clusters represent 58.1 % of the data set. Details of clustering analysis are described in the Methods section.

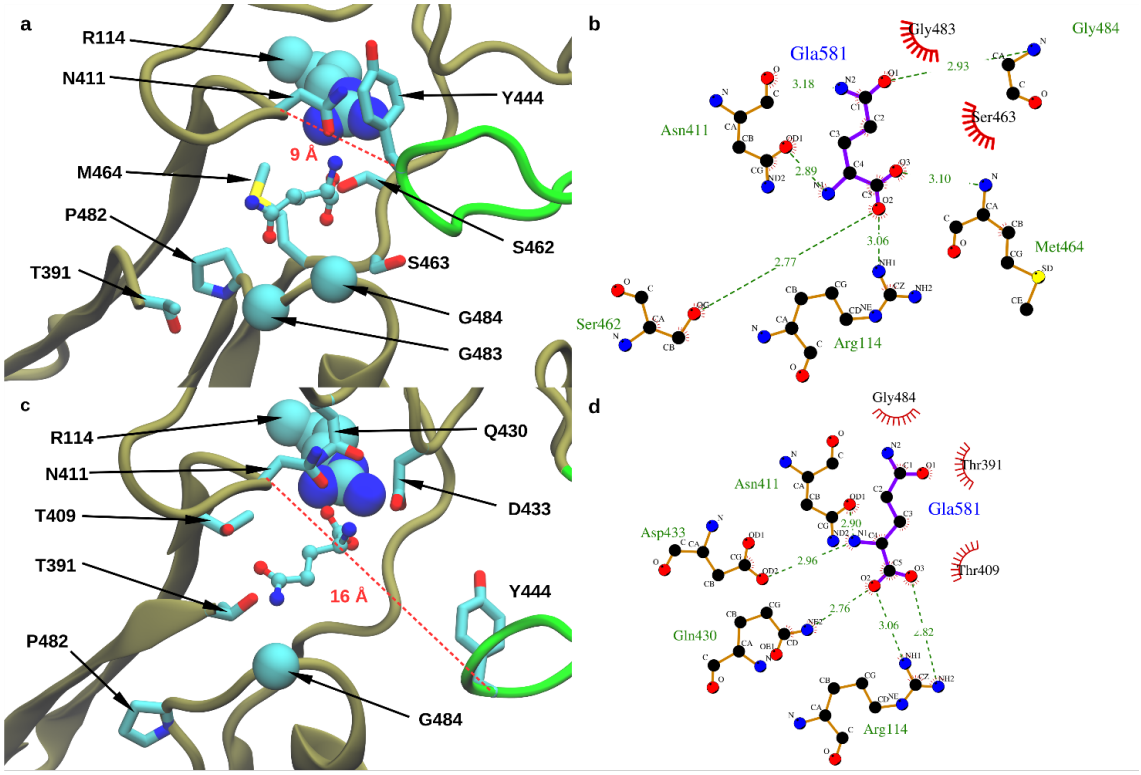


Figure 6.7: Top two cluster representative structures of glutamine bound to EcoGGT. a) Binding pocket of the representative structure of cluster 1. b) Hydrogen bond network and hydrophobic contacts of the glutamine (Gla 581), and the protein pocket of cluster 1. c) and d) Binding pocket and receptor-substrate interactions plot of representative structure of cluster 2, respectively. The substrate and sidechains involved in hydrogen bonding and hydrophobic interactions are highlighted in licorice and space-filling representation, respectively. lid-loop (residues 439-448), is shown in green ribbon. Only chain B is shown in tan ribbons. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color spheres, respectively. Two-dimensional visualization obtained with Ligplot+.

6.3.5 Recurrent configurations

Using all the results already shown, four recurring configuration are observed as shown in Figure 6.8. These configurations represent 4 of 5 clusters identified when the ligand RMSD is below 10 Å for trajectory R20. The first configuration Figure 6.8a attains a state common to all the five trajectories where glutamine approaches the binding pocket. The ligand is about 10 Å distant from the catalytic residue and does not interact directly with the residues inside the active site. It is still able to depart from the protein surface, like in run 22 and 30. This configuration is also supported by the presence of residues that make hydrogen bonds with glutamine outside the binding pocket, namely Thr 413 and Tyr 194, as shown by the Specific Binding Site Interactions subsection. In the second configuration, glutamine is oriented in a way that allows it to make a stabilizing interaction between the negative charge on the carboxyl group and the N-terminus of Thr 391. Glutamine's amino group is also oriented towards the polar residue Asn 411 and the negatively charged

Asp 433 (Figure 6.8b). The importance of these two residues is highlighted in the previous subsections. In the third configuration glutamine amino group interacts with Asn 411 and Asp 433. this interaction leads glutamine deep into the binding pocket (Figure 6.8c). Finally, in the last configuration the carboxyl group is engaged in hydrogen bonds with Arg 114, Ser 462 and Ser 463 (residues in group I and II identified by the phrases clustering). The amino group is stabilized by Asn 411 and Asp 433 previously mentioned, the amide group is now near the hydroxyl oxygen on Thr 391 ready to react (Figure 6.8d).

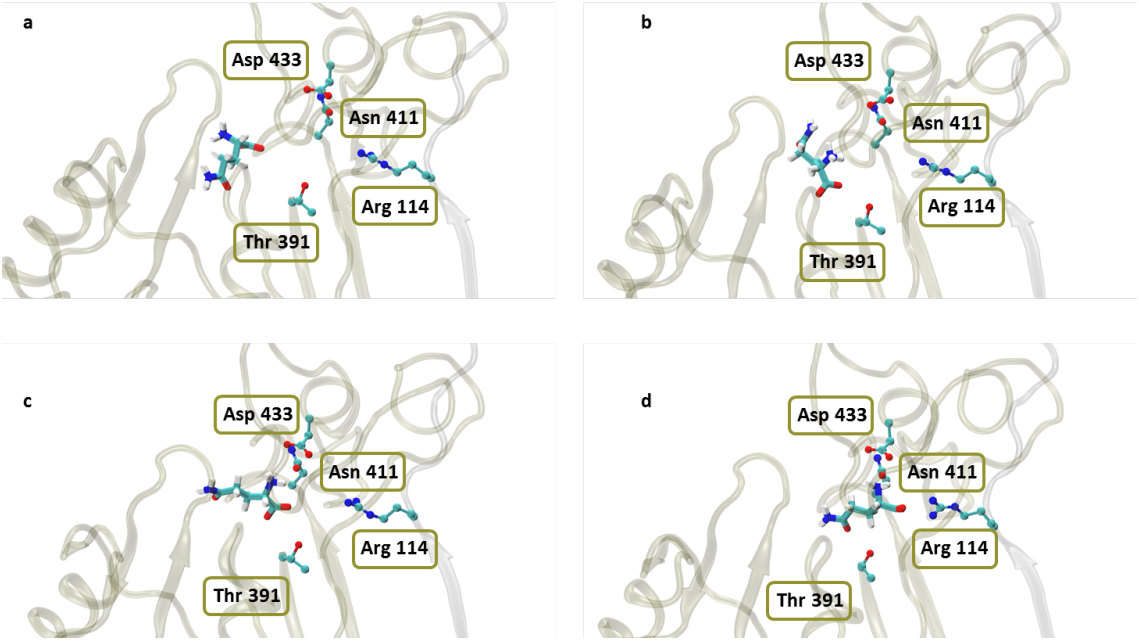


Figure 6.8: The four common configurations. These configurations represent clusters identified on the R20 trajectory segment when the ligand RMSD is below 10 Å. Clustering was performed on glutamine heavy-atoms with the gromos clustering method and cutoff radius of 3 Å. The protein structures were previously aligned relative to its initial conformation. Glutamine is represented in licorice, the relevant residues in CPK and the protein in new cartoons. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color spheres, respectively. White and magenta color ribbons represent the two subdomains of EcoGGT enzyme.

6.4 Discussion and Conclusions

The process of substrate recruiting by enzymes may encompass key features of their selectivity. We employed a computational strategy based on a swarm of aMD simulations was employed to model the approach and subsequent binding of glutamine into the EcoGGT active site. Three binding events out of 90 independent short aMD repeats were observed, mirroring the actual behavior of enzymatic reactions, where only a small fraction of enzyme-substrate encounters results in effective bind-

ing events. Our simulated bound glutamine's relative orientation and position is in good agreement with a naïvely built model of the EcoGGT bound to the substrate, based on a crystal X-ray structure of EcoGGT bound to the corresponding product obtained by Fukuyama *et al.* [141]. Based on MD postprocessing analysis, we observed four recurring configurations which could be relevant in the recognition process, imposing no restraint or knowledge based bias to the simulations. Here we defined an efficient, computationally affordable protocol for substrate recruiting modeling, which in the future could be used to study also the more complex GGT catalyzed transpeptidation reaction and applied to any enzyme-substrate system aMD simulations presented here suggest that the lid-loop functions as a gating structural element and as a steric wall confining the substrate inside the binding pocket. Opening and closing of the lid-loop is a necessary condition, but not the only one for the substrate to enter the binding site. Pocket residues interactions with glutamine can further stabilize the substrate through its binding pathway. These modelling insights are in good agreement with previous experimental work, in which the lid loop's role was probed through a variety of experimental techniques [173]. That work indicated that the loop acts as a gating mechanism allowing the selection of small substrates such as glutamine into the binding pocket of EcoGGT. aMD simulations show conformational changes of residues in the pocket, which suggests pocket flexibility over longer timescales. Note that hundred-nanoseconds long aMD simulations were shown to reproduce protein fluctuations observed at millisecond timescale for several other proteins [174–176]. The simulated EcoGGT-glutamine complex dynamics can be explained through a conformational selection model, in which the binding is controlled by the timescale of the receptor conformational changes relative to the receptor-ligand diffusion [177, 178]. A hypothesis is that the lid-loop gating mechanism is a general feature of GGT enzymes containing this conserved lid-loop and that specific site mutations on this loop may affect the enzyme selectivity. In conclusion, we not only elucidated the key configurations of GGT molecular recognition process, but we provide computational modelling that supports a lid-loop gating mechanism governing substrate binding to EcoGGT.

6.A Appendixes

6.A.1 Random acceleration molecular dynamics

We also focused on the exit pathway of glutamine and glutamate from the binding pocket using RAMD [163]. This technique applies to the center of mass of the ligand a randomly directed force calculated as shown in 6.2 where f_0 is the magnitude of the force in kcal·mol^{-1}·Å^{-1} and $\hat{r}^*$ is the random direction, which is changed every

N steps if the ligand can not move at a velocity higher than a set threshold. The minimum velocity is defined as reported in Equation 6.3

$$F_{ext} = f_0 \hat{r}^* \tag{6.2}$$

$$v_{min} = \frac{r_{min}}{N\Delta t} \tag{6.3}$$

Where r_{min} is the minimum distance in Å the ligand needs to cover between two checks. N is the number of steps in which the force is maintained, t is the time step of the simulation (2 fs). It is important to select the right set of parameters, which is characteristic for each system. A good way to decide if a particular set of parameter is good is the success rate, if the ligand gets out of the protein in 30 to 50 percent of the times it means the set can be accepted [179]. To analyze the result we used the Path Similarity Analysis (PSA) [180] approach. PSA uses the Fréchet [181, 182] and the Hausdorff [183–185] distances to calculate how similar two trajectories are. The Hausdorff distance is defined as the longest possible distance between one point of one set and the closest point of the other. Mathematically is defined as reported in Equation 6.4

$$\delta_H(P,Q) = \max\{\delta_h(P|Q), \delta_H(Q|P)\} \tag{6.4}$$

Where $\delta_H(P,Q)$is the Hausdorff distance between two paths P and Q represented, respectively, as sequences of conformations in a 3N dimensional space presented in 6.5. $\delta_h(P|Q)$ shown in 6.6 is a function that selects, out of all possible point in P, the one with the most distant nearest neighbor in Q.

$$P = \{(p_k)_{k-1}^n | p_k | \mathbb{R}^{3N}, k = 1, ..., n\} \tag{6.5}$$

$$\delta_h(P|Q) = \max_{p \in P} \min_{q \in Q} d(p,q) \tag{6.6}$$

6.A.2 Testing

To select the correct parameters we used a "trial and error" approach. Our first attempts were based on Vashisth and Abrams [179] work but at first we used less runs. It is clear from Table 6.A.1 that in our first set of simulations, where we ran only 50 simulations, we used either a force too strong (20 kcal·mol^{-1}·Å^{-1}) or too weak (5, 10 and 12 kcal·mol^{-1}·Å^{-1}). Only using 15 kcal·mol^{-1}·Å^{-1} produces an acceptable success rate, so we used this set of parameters and increased the number of runs to determine the level of reproducibility. 100 runs are not enough to have

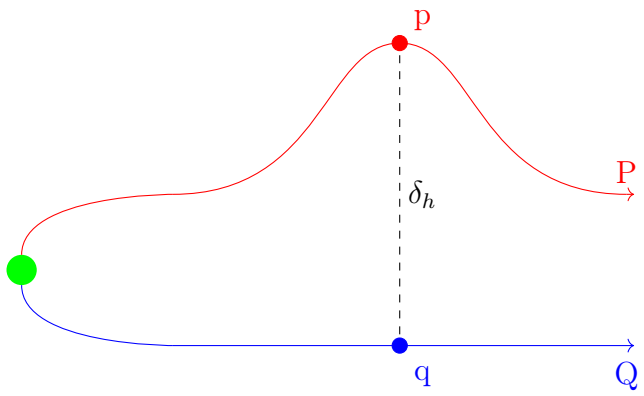


Figure 6.A.1: Hausdorff distance

F (kcal·mol^{-1}·Å^{-1})	N (steps)	r$_{min}$ (Å)	v$_{min}$ (Å·frame^{-1})	Runs	Success Rate (%)
20	10	0.006	0.0003	50	98
5	100	0.03	0.0002	50	0
10	30	0.004	0.0001	50	3
12	50	0.015	0.0002	50	10
15	10	0.005	0.0003	50	30
15	10	0.005	0.0003	100	56
15	10	0.005	0.0003	100	37
15	10	0.004	0.0002	200	50
14	10	0.005	0.0003	200	39

Table 6.A.1: Parameter tested, in chronological order, to calculate the optimal set to simulate glutamate's leaving pathway. F is the magnitude of the force applied to the center of mass of glutamate.

a reproducible success rate. A final set of simulation of 200 runs each was carried out and we found a good success rate of 39%. Since most of the testing was carried out using glutammic acid, only two set were necessary to identify a set of parameter satisfactory to simulate the glutamine system.

F (kcal·mol^{-1}·Å^{-1})	N (steps)	r$_{min}$ (Å)	v$_{min}$ (Å·frame^{-1})	Runs	Success Rate (%)
14	10	0.005	0.0005	200	15
15	10	0.005	0.0005	200	32

Table 6.A.2: Parameter tested, to calculate the optimal set to simulate glutamine's leaving pathway.

6.A.3 Clustering

To analyze our results we ran a cluster analysis on our trajectories with the PSA implementation of the Hausdorff metric. The successful trajectories were first stripped

of water and ions and then aligned, minimizing the RMSD. Then the cluster containing at least 5 trajectories separated by less than 10 Å were selected and the middle trajectory of the cluster was picked. In Figure 6.A.2 there are the 4 glutammic acid exit pathways selected with the aforementioned criteria and the representations of all the successful trajectories with the relative distances. The same was done on glutamine and presented in Figure 6.A.3.

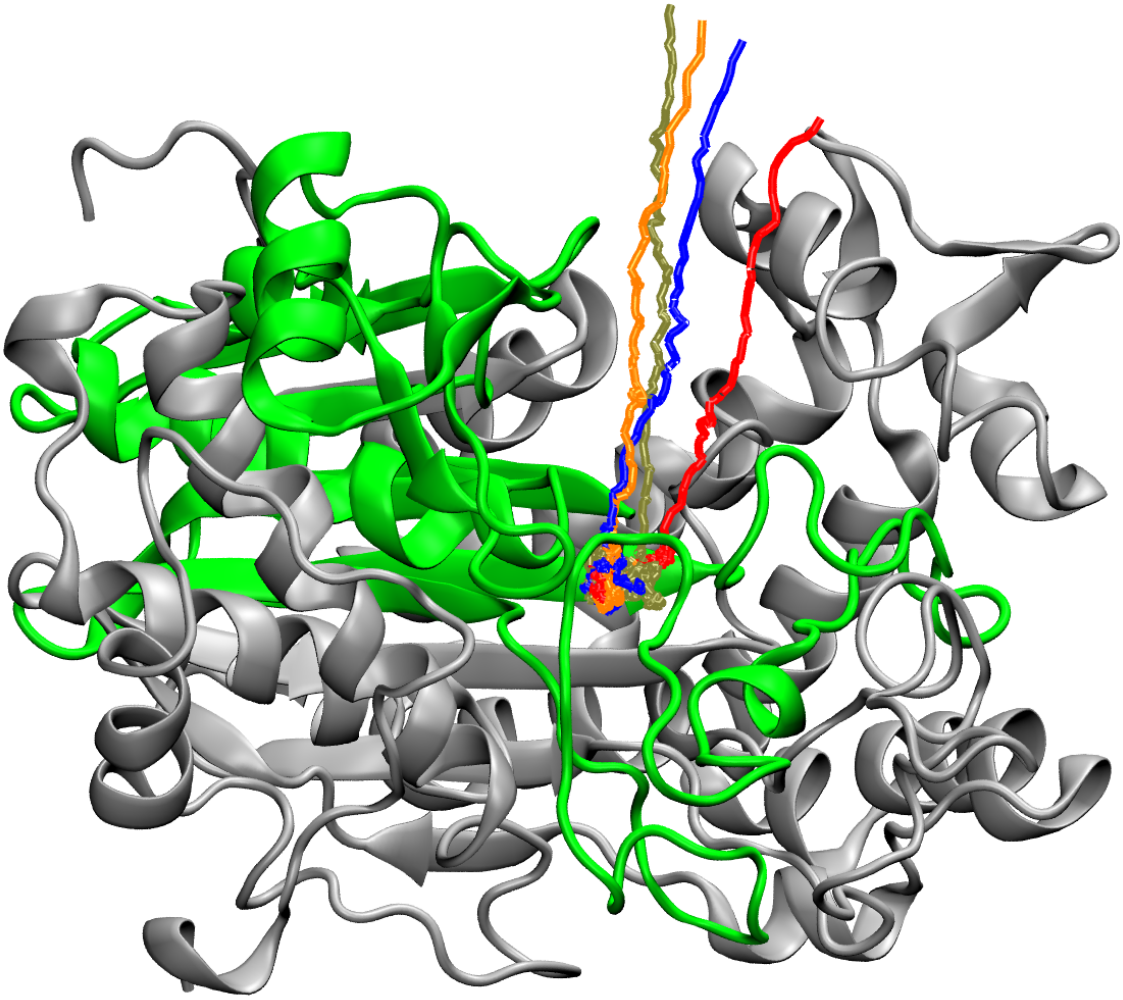


Figure 6.A.2: Glutamate exit pathways

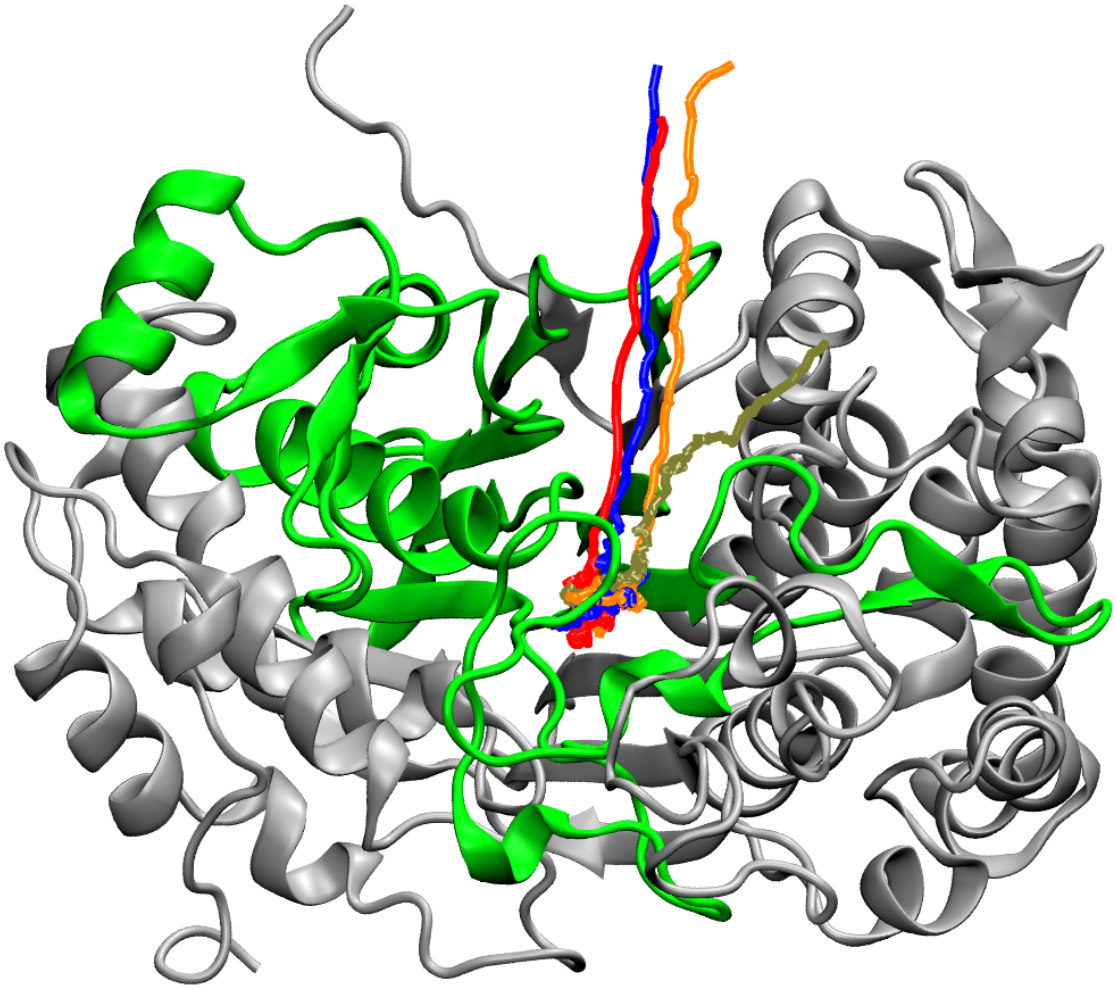


Figure 6.A.3: Glutamine exit pathways

6.A.4 Concluding remarks

Visualizing the exit paths for both glutamate and glutamine allows us to notice that, even if a few different paths are found, they only diverge after the loss of any meaningful interaction with the protein. An interesting point is that is required a stronger force to pull glutamine out of the binding pocket compared to glutamate. This make sense from a biochemical point of view, since glutamate is supposed to be expelled while glutamine is supposed to react with the active site.

6.A.5 Other figures and tables

Contact analysis

To further investigate how glutamine approaches the binding pocket, we analyzed the contacts established between glutamine and protein residues in the successful trajectories (R10, R20 and R50), identifying which ones were common to all of them. Since the time spent by glutamine in the binding pocket is different for each trajectory, we normalized the persistence time of each contact with respect to the highest one for each MD run. It is interesting to notice that the residue that always scores the highest amount of contacts is Asn 411, which is used to normalize the number of contacts. Normalized contact percentages are defined as reported in Equation 6.7

$$\text{Normalized Contacts} = \frac{C_R \cdot 100}{C_{N411}} \tag{6.7}$$

where C_R is the number of contacts between glutamine and a specific residue. C_{N411} is the number of contacts between glutamine and Asn 411. Contacts are considered between a pair of heavy-atoms with a cutoff distance of 3.0 Å. We considered in further analyses only contacts that were common to all the three successful trajectories and were observed for at least 5% of the normalized contacts population (Tables 6.A.3 - 6.A.9). Particularly important residues are Arg 114, Thr 409 and Asp 433, which interact with glutamine for more than 30% of the normalized contacts in each of the three trajectories. This is also shown in plots of contact formation as function of time plots (Figures 6.A.4 - 6.A.9). We calculated the contacts between glutamine and EcoGGT using a plugin freely available and authored by Lubos Vrbka. The results where then analyzed using a python script.

Run 10		
Residue	Number of contacts	Percentage (%)
Arg 114	3472	74.67
Lys 192	5	0.11
Tyr 194	28	0.60
Lys 218	5	0.11
Thr 391	323	6.95
Thr 409	3455	74.30
Asn 411	4650	100.00
Thr 412	371	7.98
Thr 413	73	1.57
Gln 430	7	0.15
Asp 433	1851	39.81
Tyr 444	1308	28.13
Leu 461	42	0.90
Ser 462	4289	92.24
Ser 463	4401	94.65
Met 464	3753	80.71
Gly 483	2892	62.19
Gly 484	4140	89.03
Ser 485	154	3.31
Ile 487	2	0.04
Lys 547	62	1.33
Val 573	16	0.34

Table 6.A.3: Contact analysis results for aMD trajectory number 10. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Only the residues that make contacts in all the three successful trajectories are shown.

Run 20		
Residue	Number of contacts	Percentage (%)
Arg 114	1839	100.00
Lys 192	2	0.11
Tyr 194	20	1.09
Lys 218	13	0.71
Thr 391	1758	95.6
Thr 409	1762	95.81
Asn 411	1839	100.00
Thr 412	45	2.45
Thr 413	17	0.92
Gln 430	1676	91.14
Asp 433	1571	85.43
Tyr 444	27	1.47
Leu 461	2	0.11
Ser 462	1017	55.3
Ser 463	133	7.23
Met 464	369	20.07
Gly 483	15	0.82
Gly 484	1385	75.31
Ser 485	113	6.14
Ile 487	48	2.61
Lys 547	15	0.82
Val 573	17	0.92

Table 6.A.4: Contact analysis results for aMD trajectory number 20. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Only the residues that make contacts in all the three successful trajectories are shown.

Run 50		
Residue	Number of contacts	Percentage (%)
Arg 114	841	34.95
Lys 192	17	0.71
Tyr 194	76	3.16
Lys 218	11	0.46
Thr 391	1548	64.34
Thr 409	1861	77.35
Asn 411	2406	100.00
Thr 412	473	19.66
Thr 413	146	6.07
Gln 430	876	36.41
Asp 433	1727	71.78
Tyr 444	1755	72.94
Leu 461	6	0.25
Ser 462	14	0.58
Ser 463	135	5.61
Met 464	188	7.81
Gly 483	96	3.99
Gly 484	1189	49.42
Ser 485	975	40.52
Ile 487	47	1.95
Lys 547	22	0.91
Val 573	2	0.08

Table 6.A.5: Contact analysis results for aMD trajectory number 50. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Only the residues that make contacts in all the three successful trajectories are shown.

Run 10		
Residue	Number of contacts	Percentage (%)
Arg 114	3472	74.67
Lys 192	323	6.95
Tyr 194	3455	74.30
Lys 218	4650	100.00
Thr 391	371	7.98
Thr 409	73	1.57
Asn 411	7	0.15
Thr 412	1851	39.81
Thr 413	1308	28.13
Gln 430	4289	92.24
Asp 433	4401	94.65
Tyr 444	3753	80.71
Leu 461	2892	62.19
Ser 462	4140	89.03
Ser 463	154	3.31

Table 6.A.6: Contact analysis results for aMD trajectory number 10. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Same data showed in Table 6.A.3 but only the residues that scores more than 5% in at least one of the three trajectories are shown.

Run 20		
Residue	Number of contacts	Percentage (%)
Arg 114	1839	100.00
Lys 192	1758	95.60
Tyr 194	1762	95.81
Lys 218	1839	100.00
Thr 391	45	2.45
Thr 409	17	0.92
Asn 411	1676	91.14
Thr 412	1571	85.43
Thr 413	27	1.47
Gln 430	1017	55.3
Asp 433	133	7.23
Tyr 444	369	20.07
Leu 461	15	0.82
Ser 462	1385	75.31
Ser 463	113	6.14

Table 6.A.7: Contact analysis results for aMD trajectory number 20. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Same data showed in Table 6.A.4 but only the residues that scores more than 5% in at least one of the three trajectories are shown.

Run 50		
Residue	Number of contacts	Percentage (%)
Arg 114	841	34.95
Lys 192	1548	64.34
Tyr 194	1861	77.35
Lys 218	2406	100.00
Thr 391	473	19.66
Thr 409	146	6.07
Asn 411	876	36.41
Thr 412	1727	71.78
Thr 413	1755	72.94
Gln 430	14	0.58
Asp 433	135	5.61
Tyr 444	188	7.81
Leu 461	96	3.99
Ser 462	1189	49.42
Ser 463	975	40.52

Table 6.A.8: Contact analysis results for aMD trajectory number 20. The table shows the absolute number of contacts and the percentage averaged over 411ASN's number of contacts. Same data showed in Table 6.A.5 but only the residues that scores more than 5% in at least one of the three trajectories are shown.

Residue	Average %
Arg 114	69.87
Thr 391	55.63
Thr 409	82.49
Asn 411	100
Thr 412	10.03
Thr 413	2.85
Gln 430	42.57
Asp 433	65.67
Tyr 444	34.18
Ser 462	49.37
Ser 463	35.83
Met 464	36.2
Gly 483	22.33
Gly 484	71.25
Ser 485	16.66

Table 6.A.9: Normalized Contacts average of three successful trajectories R10, R20 and R50. The averages are calculated as sum of the normalized percentages divided by three.

Cluster	Number of structures	Cluster weight (%)	RMSD (Å)
1	6775	79.9	0.91
2	1045	12.3	1.5
3	309	3.6	1.05
4	106	1.2	2.96

Table 6.A.10: Clustering of the conformational space sampled by glutamine after binding EcoGGT. The number of structures for each cluster is indicated. The cluster weight is calculated over the entire set of clusters. The RMSD is calculated relative to the proposed complex protein-glutamine structure after converting the glutamate into glutamine. Clustering analysis details are provided in the Methods section.

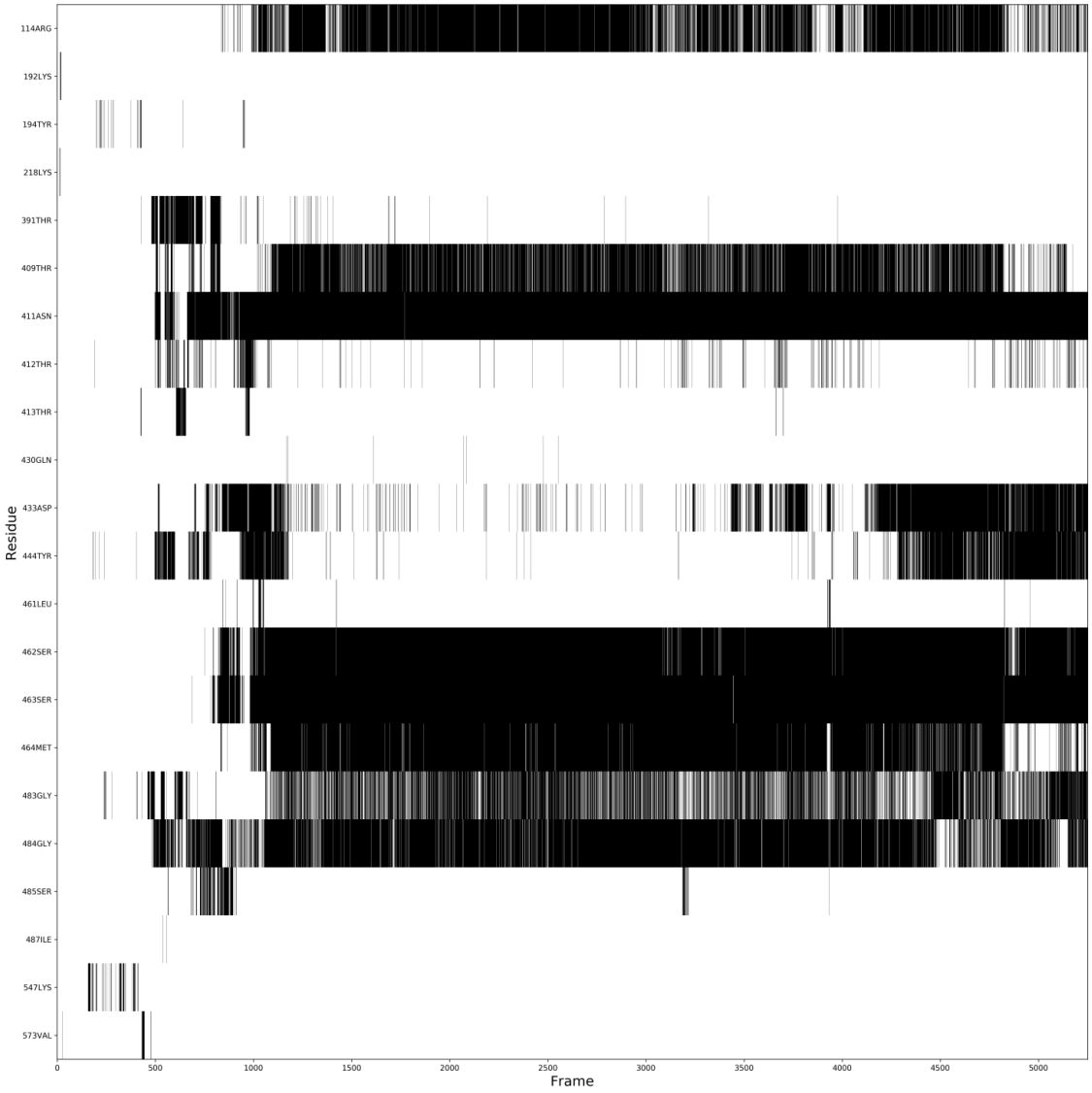


Figure 6.A.4: Contact analysis over time in aMD trajectory R10. Only the common contacts are shown

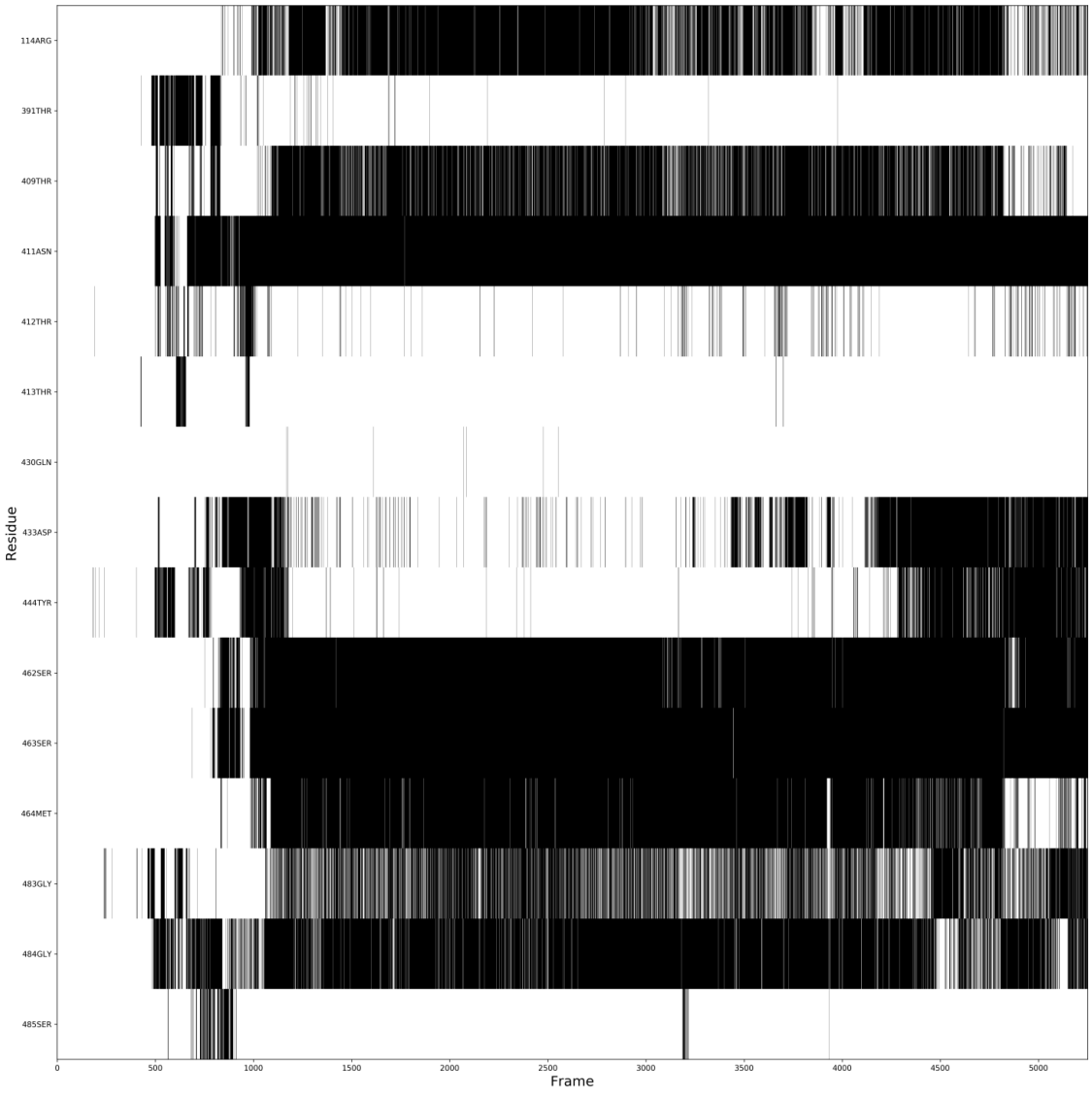


Figure 6.A.5: Contact analysis over time in aMD trajectory R10. Only the residues calculated in Table S2 are shown

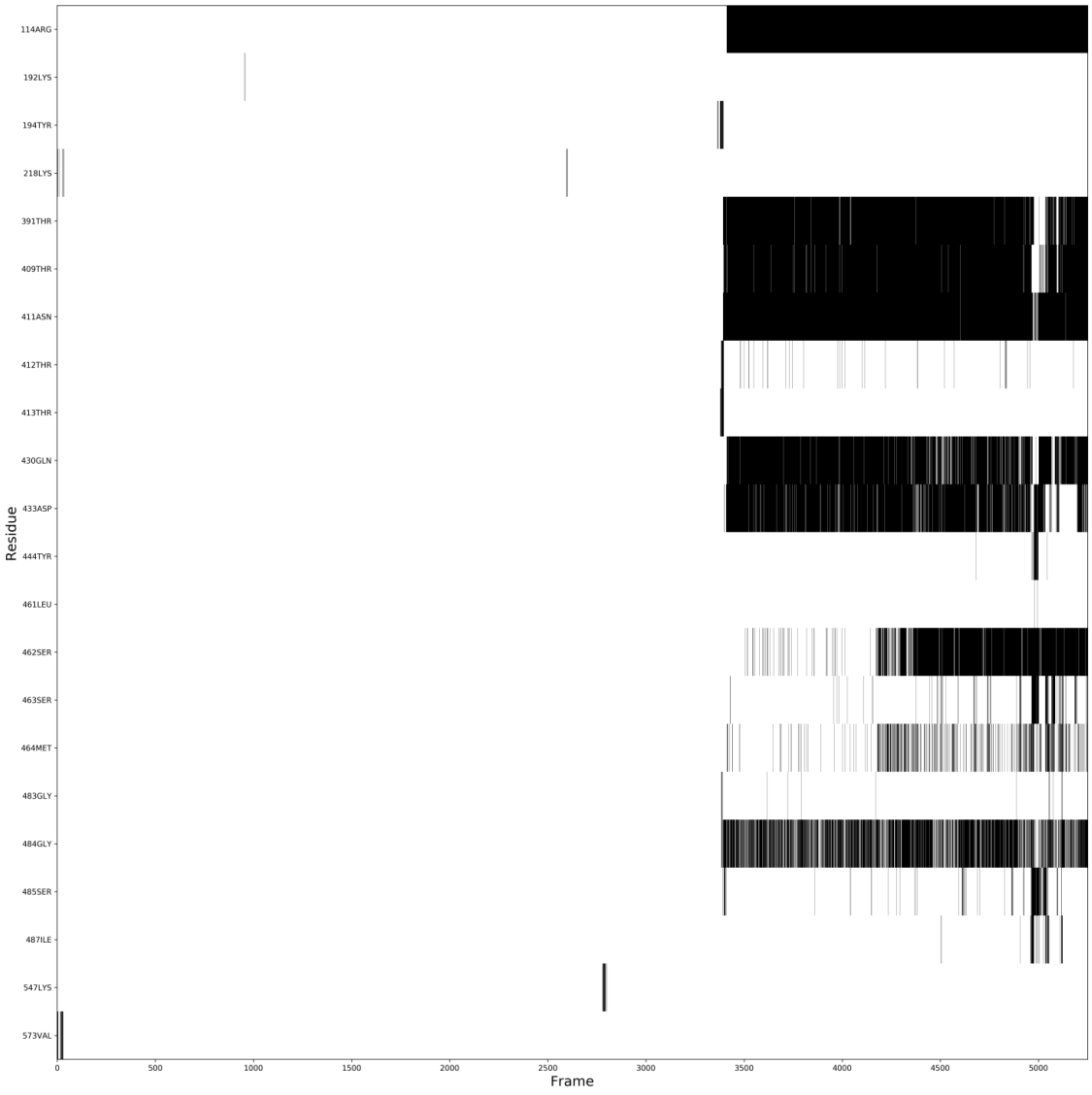


Figure 6.A.6: Contact analysis over time in aMD trajectory R20. Only the common contacts are shown

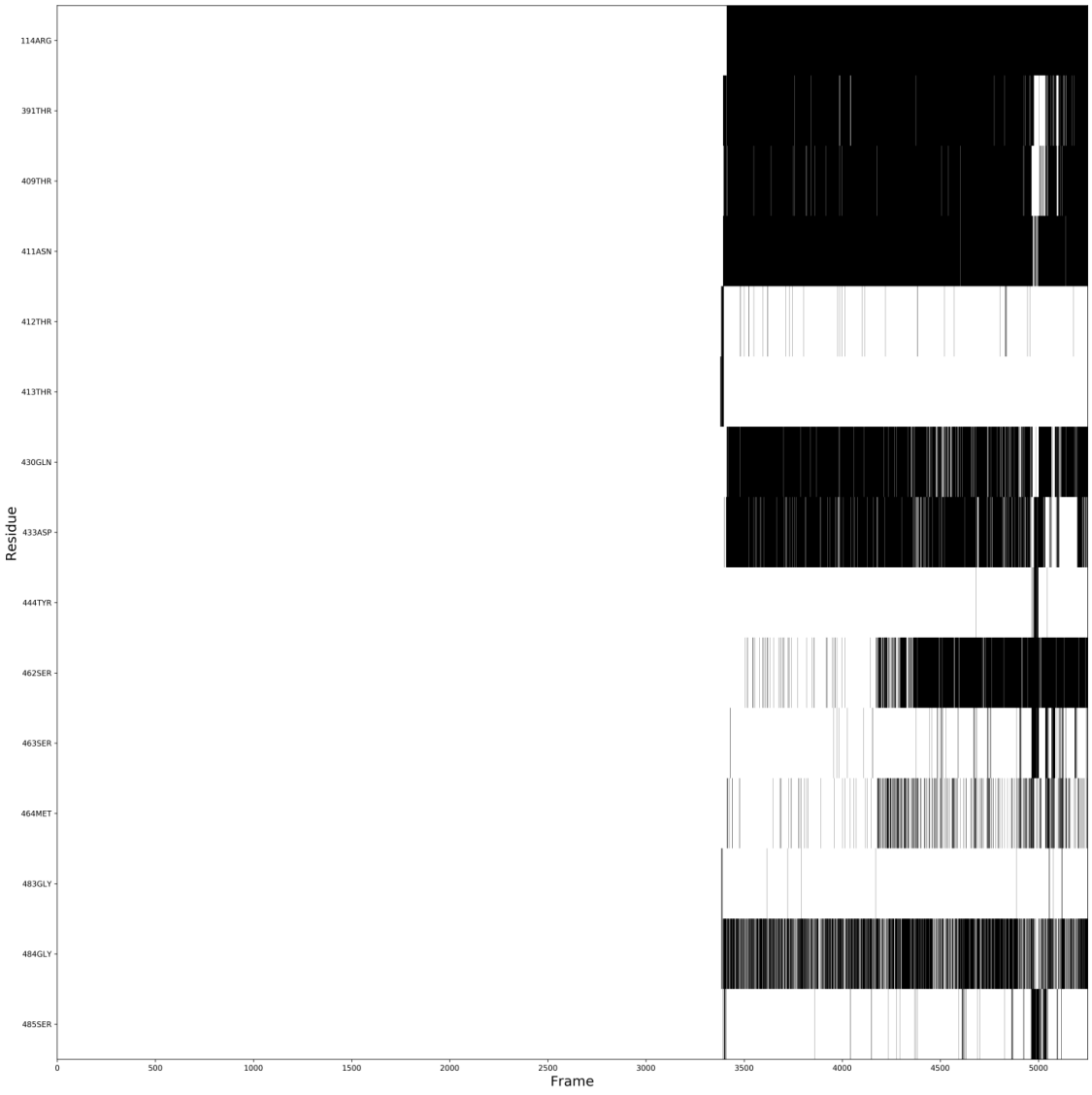


Figure 6.A.7: Contact analysis over time in aMD trajectory R20. Only the residues calculated in Table S2 are shown

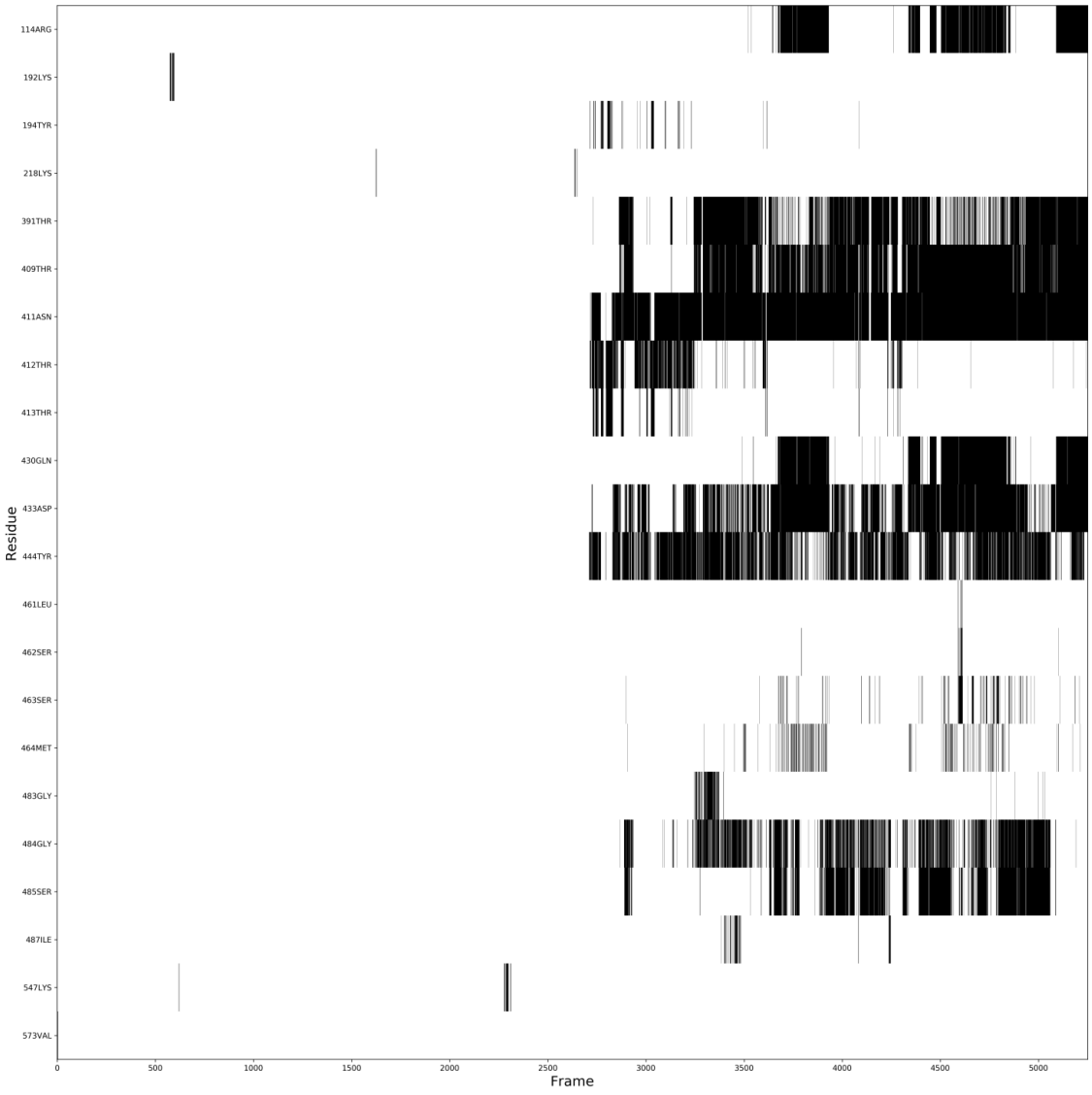


Figure 6.A.8: Contact analysis over time in aMD trajectory R50. Only the common contacts are shown

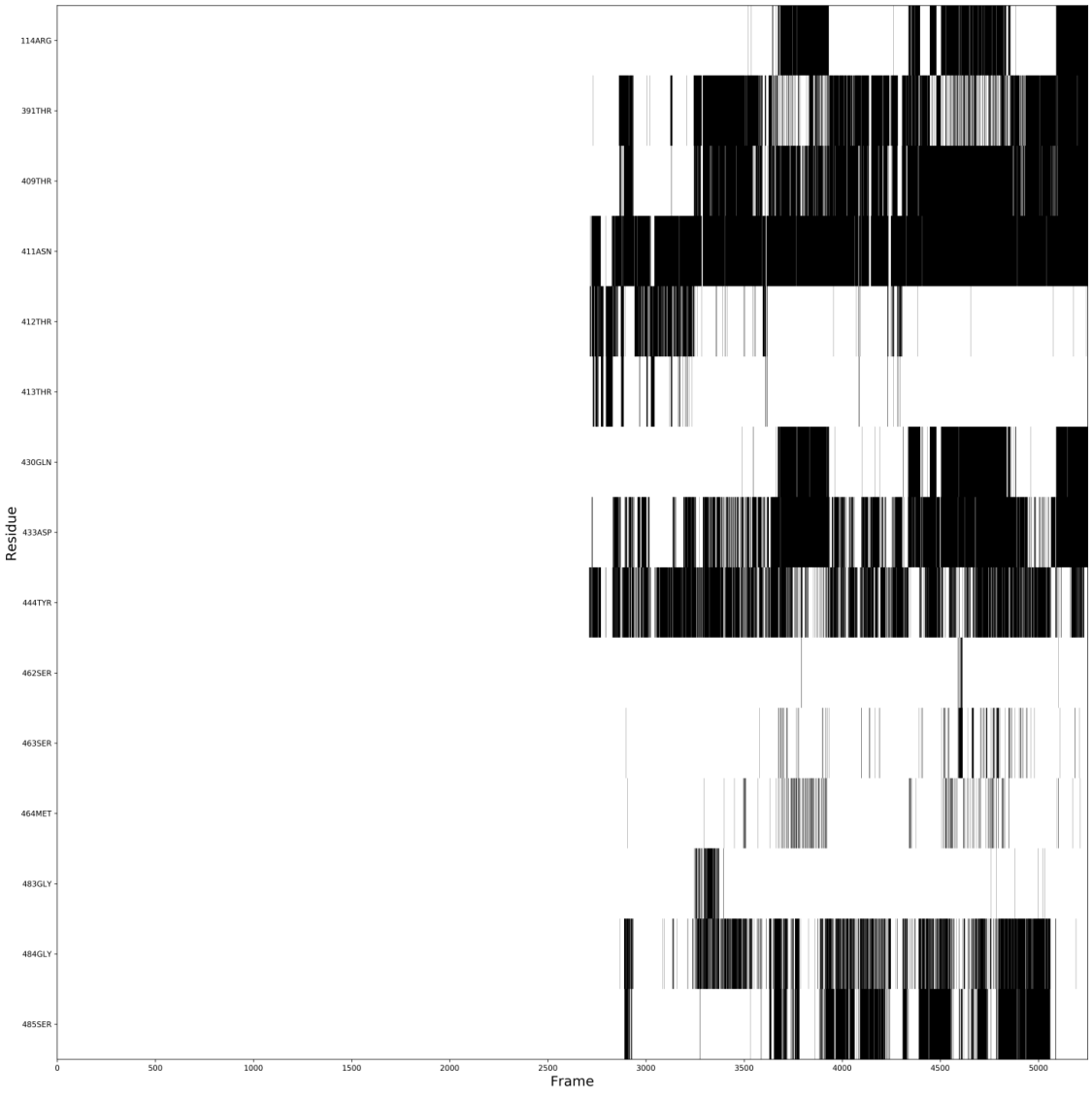


Figure 6.A.9: Contact analysis over time in aMD trajectory R50. Only the residues calculated in Table S2 are shown

Hydrogen Bonds

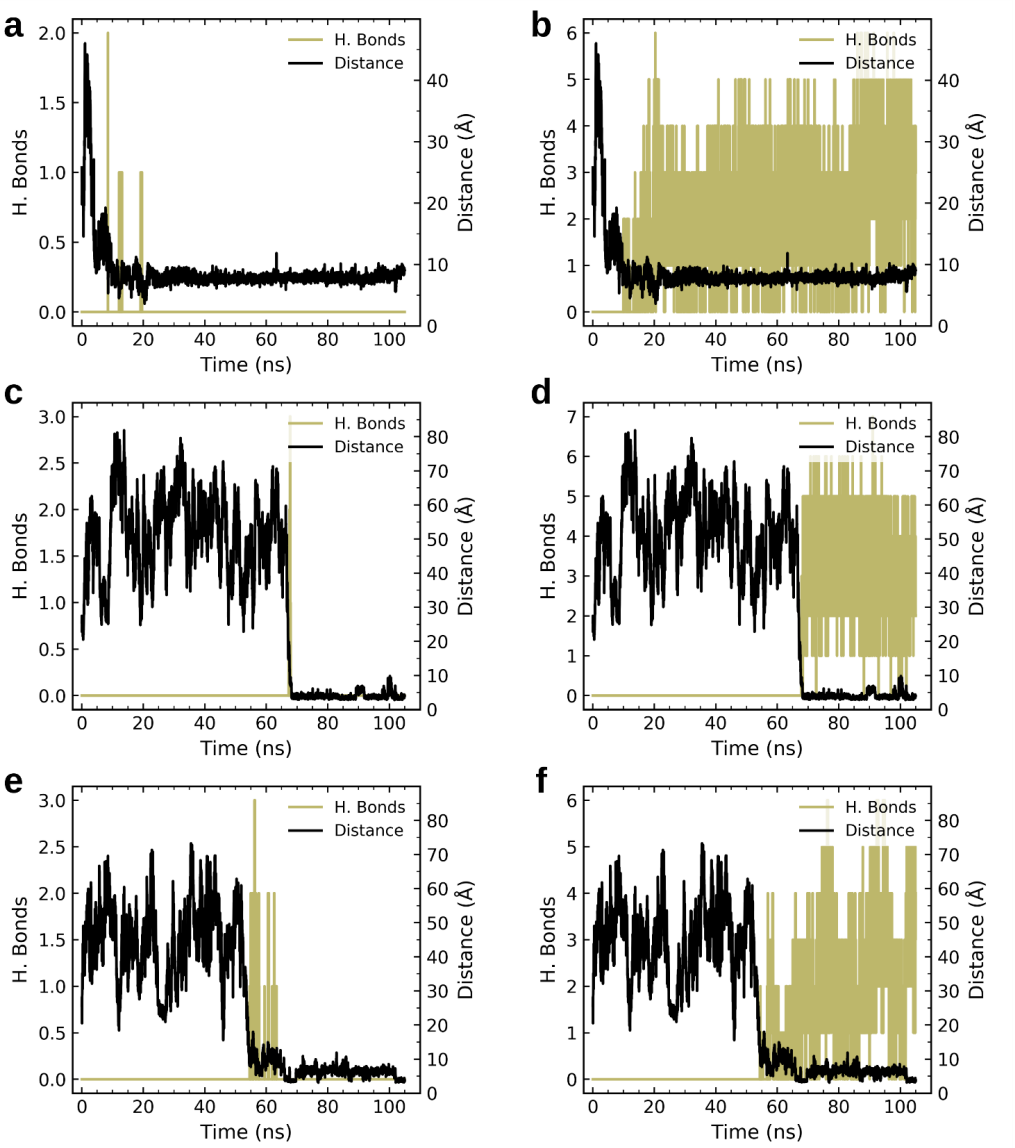


Figure 6.A.10: Traces of number of hydrogen bonds and distance between the glutamine carboxylic carbon and the guanidine carbon in Arg 114. Two sets of residues are used for the analysis. Entry pocket residues set: Thr 413 and Tyr 194. Binding pocket residues set: Ser 463, Asn 411, Arg 114, Asp 433 and Thr 409. Panels **a** and **b** shows the number of entry and pocket residue hydrogen bonds for R10, respectively. Panels **c** and **d** display the number of hydrogen bonds for R20 as explained above. Panels e and **f** show number of hydrogen bonds for R50 as described above

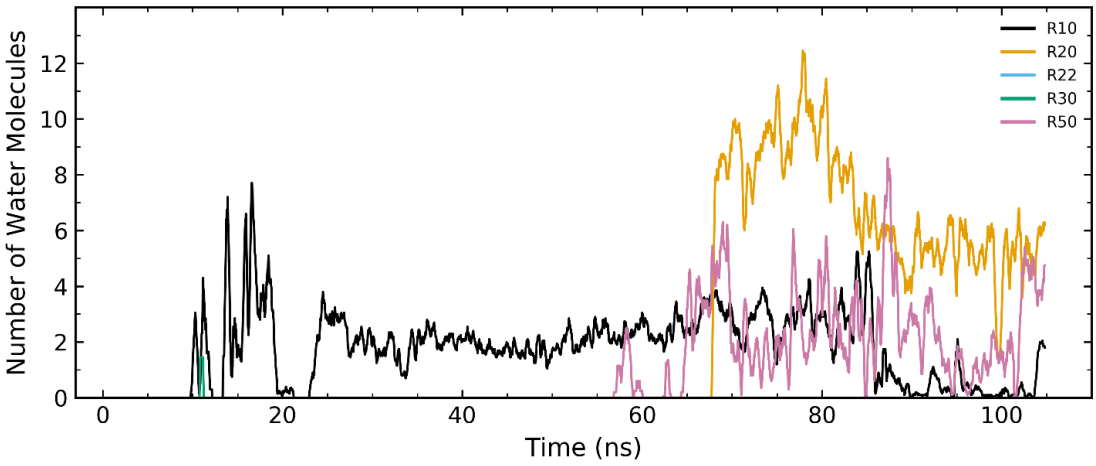


Figure 6.A.11: Number of water molecules inside a box containing glutamine and Tyr 444 for substrate-binding aMD trajectories R10, R20 and R50. R22 and R30 are trajectories where the substrate fails to bind. The box is defined by the minimum and maximum coordinates of heavy-atoms atoms in glutamine and Tyr 444 sidechain. Water molecules are counted when inside the box and within 7 of both glutamine and Tyr 444. Calculation is carried out when glutamine's carboxyl carbon is within a sphere of radius 8 centered at the guanidine carbon of Arg 114. Water selection is updated for each frame. Data is smoothed by moving average over 20 frames (400 ps).

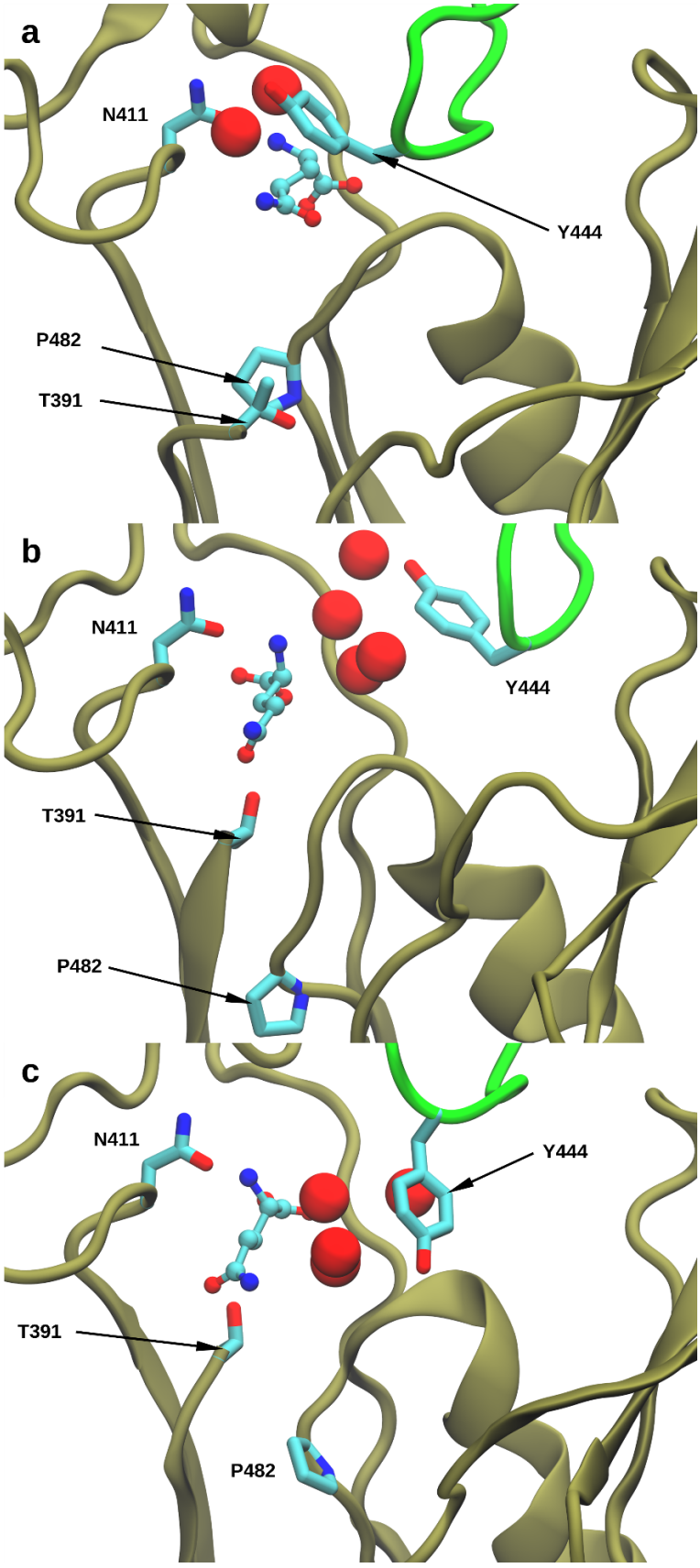


Figure 6.A.12: EcoGGT-glutamine complexes corresponding to the last frames of the following aMD trajectories: R10 (a), R20 (b), R50 (c). Glutamine represented in ball and sticks. lid-loop and chain B are shown in green and tan ribbons, respectively. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color spheres, respectively. Oxygen atoms of water molecules are shown in red space filled representation.

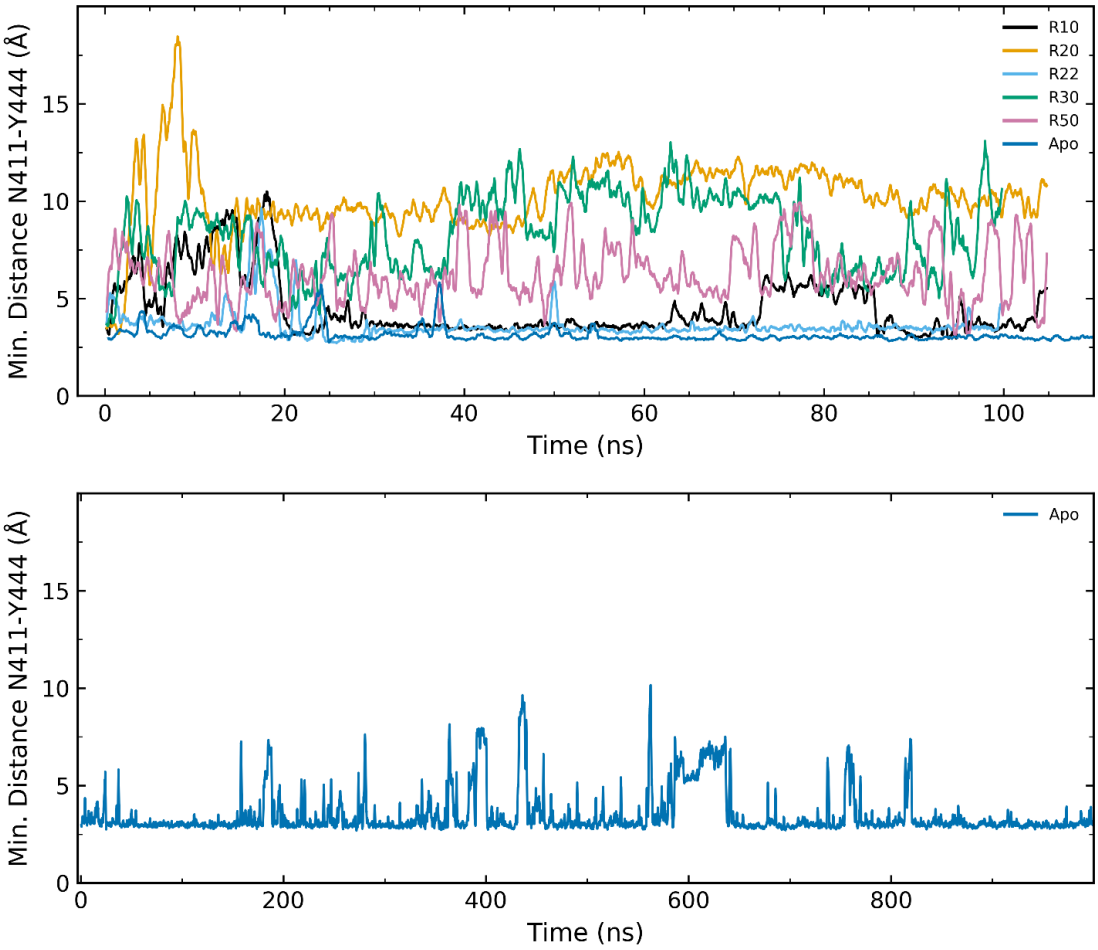


Figure 6.A.13: Minimum distance traces between heavy-atoms of Asn 411 and Tyr 444. (Top panel) aMD runs and conventional MD of apo EcoGGT. Stable receptor-substrate complexes are formed in trajectories R10, R20 and R50. R22 and R30 are trajectories where the substrate fails to bind. (Bottom panel) Extended conventional MD of apo EcoGGT. Data is smoothed by moving average over 20 frames (400 ps for aMD runs and 800 ps for conventional MD).

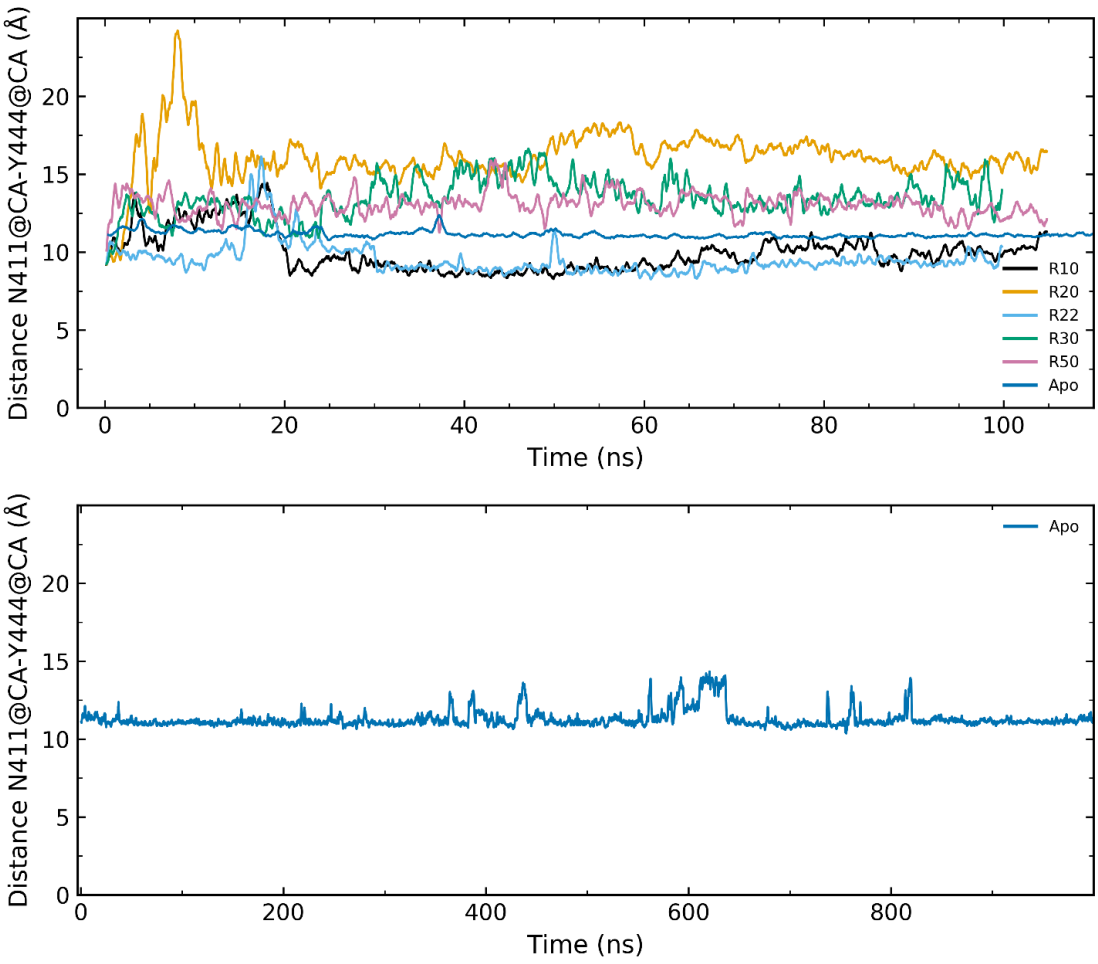


Figure 6.A.14: Distance values between C- C of Asn 411 and Tyr 444 versus simulation time. (Top panel) aMD runs and conventional MD of apo EcoGGT. Stable receptor-substrate complexes are formed in trajectories R10, R20 and R50. R22 and R30 are trajectories where the substrate fails to bind. (Bottom panel) Extended conventional MD of apo EcoGGT. Data is smoothed by moving average over 20 frames (400 ps for aMD runs and 800 ps for conventional MD).

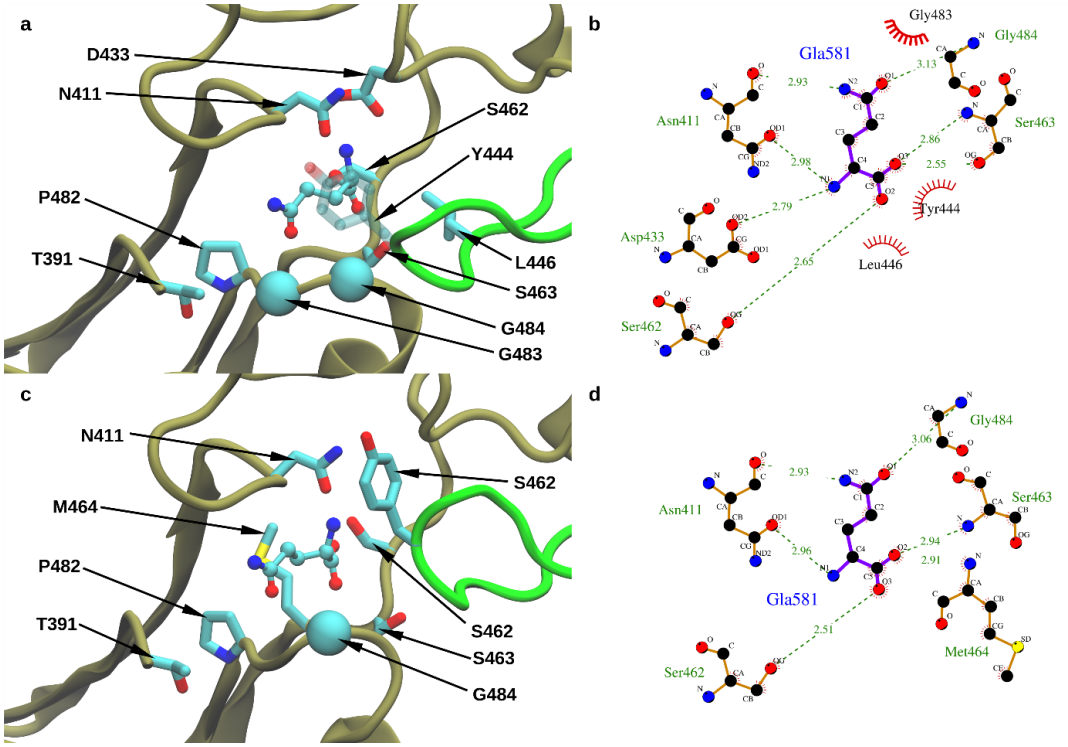


Figure 6.A.15: Structural representation of clusters 3 and 4. C-C distance between Tyr 444-Asn 411 of 10 Å and 9 Å, respectively. a) Binding pocket of the representative structure of cluster 3. Tyr 444 is shown in transparent licorice representation. b) Hydrogen bond network and hydrophobic contacts of the glutamine (Gla 581), and the protein pocket in the representative structure of cluster 3. c) and d) Binding pocket and receptor-substrate interactions plot of the representative structure of cluster 4, respectively. The substrate and protein residues involved in hydrogen bonding and hydrophobic interactions are highlighted in licorice and space-filling representation, respectively. lid-loop (residues 439-448), is shown in green ribbons. Only chain B is shown in tan ribbons. Nitrogen, carbon and oxygen atoms are represented in blue, cyan and red color spheres, respectively. Two-dimensional visualization obtained with Ligplot+.

Final Considerations

We used different computational approaches, such as CAS, free energy calculations, QM/MM, aMD and WTMD to study many relevant biosystems. the AR is a recurrent topic in this thesis and it is a pressing issues in modern medicine, since more and more microorganism are developing strategies to resist standard treatments. The topic is discussed in Chapter 2 where the interaction between MBLs and AMA was studied. The former protein is known to have a key role in disrupting the ability of penicillin like antibiotics to work, and AMA is a possible inhibitor of MBLs. In this Chapter we developed a method to study the coordination state, without using computational heavy techniques like Car–Parrinello molecular dynamics [186]. We used WTMD to generate possible conformations, a script to filter out the uninteresting frames and QM/MM to assess the energy of the remaining conformations. Chapter 5 also approaches AR since YeaZ has a key role during the translation of Deoxyribonucleic Acid (DNA) in bacteria. We studied the PPIs of two of the five monomers that form the quaternary structure and used SASA and CAS to identify a small sequence on the protein protein surface that contained different aminoacids important to the binding. Whit this information we proposed a promising peptide that in our simulations was able to stably bind the monomer in a way that should prevent the aggregation of the pentamer. This peptide is currently being synthesized. In Chapter 6 we used aMD to study the binding process of glutamine into the active site of GGT which is a key protein in *Escherichia Coli*, a well know pathogen, and other bacteria. Running 90 simulations of 105 ns each, allowed us to observe three binding event in the same time usually necessary to observe one with standard MD. We observed four recurring configurations which could be relevant in the recognition process, imposing no restraint or knowledge based bias to the simulations. We also studied the motion of the lid-loop and which residues are more likely to be important during the binding process. Another topic we focused on is the study of tubulin. In Chapter 3 we addressed the interaction between six nucleosides obtained by removing the phosphate chain of a GDP molecule and replacing it with an aliphatic chain of increasing length ending with a protonated amino group. We then calculated the binding free energy between the nucleosides and the β-tubulin subunit. **Compound 3** proved to be the best at binding in the pocket. To elu-

cidate the trend in ligand-protein interaction, we performed a CAS which led to the identification of key residues in the binding pocket. **Compound 3** is currently being synthesized. We also focused our attention on peptidomimetics. We studied a UAAs, both possible isomer and the rotamer of **Peptide 2**. Using WTMD we investigated the ability of the UAAs to emulate a β-turn, finding that **Peptide 2** was more likely to maintain a closed conformation similar to a β-turn. Using standard MD and experimental data we also investigated the possible conformations of the peptides. This work culminated in a paper that has been publised [187].

Bibliography

[1] Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[2] Enrico Fermi, John R Pasta, Stanislaw Ulam, and Mary Tsingou. Studies of the nonlinear problems. Technical report, Los Alamos Scientific Lab., N. Mex., 1955.

[3] Leonhard Euler. *Institutionum calculi integralis*, volume 1. impensis Academiae imperialis scientiarum, 1824.

[4] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[5] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. crc Press, 1988.

[6] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472, 1997.

[7] George D Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

[8] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[9] Maria Kavallaris. Microtubules and resistance to tubulin-binding agents. *Nature Reviews Cancer*, 10(3):194, 2010.

[10] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6(6):e1000807, 2010.

[11] Ratna Rajesh Thangudu, Stephen H Bryant, Anna R Panchenko, and Thomas Madej. Modulating protein–protein interactions with small molecules: the importance of binding hotspots. *Journal of molecular biology*, 415(2):443–453, 2012.

[12] Laura Bonetta. Protein-protein interactions: Interactome under construction. *Nature*, 468(7325):851–854, 2010.

[13] Michelle Arkin. Protein–protein interactions and cancer: small molecules going in for the kill. *Current opinion in chemical biology*, 9(3):317–324, 2005.

[14] Warren J Strittmatter, Ann M Saunders, Michel Goedert, Karl H Weisgraber, Li-Ming Dong, Ross Jakes, David Y Huang, Margaret Pericak-Vance, Donald Schmechel, and Allen D Roses. Isoform-specific interactions of apolipoprotein e with microtubule-associated protein tau: implications for alzheimer disease. *Proceedings of the National Academy of Sciences*, 91(23):11183–11186, 1994.

[15] Patrick Chène. Drugs targeting protein–protein interactions. *ChemMedChem*, 1(4):400–411, 2006.

[16] Tom L Blundell, David F Burke, Dimitri Chirgadze, Venugopal Dhanaraj, Marko Hyvönen, C Axel Innis, Emilio Parisini, Luca Pellegrini, Muhammed Sayed, and B Lynn Sibanda. Protein-protein interactions in receptor activation and intracellular signalling. *Biological chemistry*, 381(9-10):955–959, 2000.

[17] Andrew A Bogan and Kurt S Thorn. Anatomy of hot spots in protein interfaces. *Journal of molecular biology*, 280(1):1–9, 1998.

[18] Peter J Hudson and Christelle Souriau. Engineered antibodies. *Nature medicine*, 9(1):129–134, 2003.

[19] Christophe Chipot and Andrew Pohorille. *Free energy calculations*. Springer, 2007.

[20] Irina Massova and Peter A Kollman. Combined molecular mechanical and continuum solvent approach (mm-pbsa/gbsa) to predict ligand binding. *Perspectives in drug discovery and design*, 18(1):113–135, 2000.

[21] Hans-Joachim Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of computer-aided molecular design*, 8(3):243–256, 1994.

[22] Ajay N Jain. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *Journal of computer-aided molecular design*, 10(5):427–440, 1996.

[23] Hiroaki Gouda, Irwin D Kuntz, David A Case, and Peter A Kollman. Free energy calculations for theophylline binding to an rna aptamer: Comparison of mm-pbsa and thermodynamic integration methods. *Biopolymers*, 68(1):16–34, 2003.

[24] Shuanghong Huo, Irina Massova, and Peter A Kollman. Computational alanine scanning of the 1: 1 human growth hormone–receptor complex. *Journal of computational chemistry*, 23(1):15–27, 2002.

[25] Shuanghong Huo, Irina Massova, and Peter A Kollman. Computational alanine scanning of the 1: 1 human growth hormone–receptor complex. *Journal of computational chemistry*, 23(1):15–27, 2002.

[26] Irina S Moreira, Pedro A Fernandes, and Maria J Ramos. Computational alanine scanning mutagenesis—an improved methodological approach. *Journal of Computational Chemistry*, 28(3):644–654, 2007.

[27] Nathan A Baker, David Sept, Simpson Joseph, Michael J Holst, and J Andrew McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.

[28] Federico Fogolari, Alessandro Brigo, and Henriette Molinari. The poisson–boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, 15(6):377–392, 2002.

[29] Emilio Gallicchio, Linda Yu Zhang, and Ronald M Levy. The sgb/np hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *Journal of computational chemistry*, 23(5):517–529, 2002.

[30] Rafael C Bernardi, Marcelo CR Melo, and Klaus Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(5):872–877, 2015.

[31] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[32] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

[33] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.

[34] Donald Hamelberg, John Mongan, and J Andrew McCammon. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120(24):11919–11929, 2004.

[35] Arthur F Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(20):3908, 1997.

[36] Arthur F Voter. A method for accelerating the molecular dynamics simulation of infrequent events. *The Journal of chemical physics*, 106(11):4665–4677, 1997.

[37] Max M Steiner, Pierre-Alain Genilloud, and John W Wilkins. Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme. *Physical Review B*, 57(17):10236, 1998.

[38] Donald Hamelberg, Cesar Augusto F de Oliveira, and J Andrew McCammon. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of chemical physics*, 127(15):10B614, 2007.

[39] Joshua Lederberg and Polly F Harrison. *Antimicrobial resistance: issues and options*. National Academies Press, 1998.

[40] Richard Wise, Tony Hart, Otto Cars, Marc Streulens, Reinen Helmuth, Pentti Huovinen, and Marc Sprenger. Antimicrobial resistance, 1998.

[41] World Health Organization. *Antimicrobial resistance: global report on surveillance*. World Health Organization, 2014.

[42] Andrew M King, Sarah A Reid-Yu, Wenliang Wang, Dustin T King, Gianfranco De Pascale, Natalie C Strynadka, Timothy R Walsh, Brian K Coombes, and Gerard D Wright. Aspergillomarasmine a overcomes metallo-β-lactamase antibiotic resistance. *Nature*, 510(7506):503, 2014.

[43] Dimas Suárez, Edward N Brothers, and Kenneth M Merz. Insights into the structure and dynamics of the dinuclear zinc β-lactamase site from bacteroides fragilis. *Biochemistry*, 41(21):6615–6630, 2002.

[44] Hesna Yigit, Anne Marie Queenan, Gregory J Anderson, Antonio Domenech-Sanchez, James W Biddle, Christine D Steward, Sebastian Alberti, Karen

Bush, and Fred C Tenover. Novel carbapenem-hydrolyzing β-lactamase, kpc-1, from a carbapenem-resistant strain of klebsiella pneumoniae. *Antimicrobial agents and chemotherapy*, 45(4):1151–1161, 2001.

[45] Karen Bush. Proliferation and significance of clinically relevant β-lactamases. *Annals of the New York Academy of Sciences*, 1277(1):84–90, 2013.

[46] Sarah M Drawz and Robert A Bonomo. Three decades of β-lactamase inhibitors. *Clinical microbiology reviews*, 23(1):160–201, 2010.

[47] Pengfei Li and Kenneth M Merz Jr. Metal ion modeling using classical mechanics. *Chemical reviews*, 117(3):1564–1686, 2017.

[48] Thomas A Halgren and Wolfgang Damm. Polarizable force fields. *Current opinion in structural biology*, 11(2):236–242, 2001.

[49] Richard A Friesner. Modeling polarization in proteins and protein–ligand complexes: Methods and preliminary results. *Advances in protein chemistry*, 72:79–104, 2005.

[50] George R Brubaker and David W Johnson. Molecular mechanics calculations in coordination chemistry. *Coordination chemistry reviews*, 53, 1984.

[51] Robert D Hancock and Arthur E Martell. Ligand design for selective complexation of metal ions in aqueous solution. *Chemical Reviews*, 89(8):1875–1914, 1989.

[52] Peter Comba. Metal ion selectivity and molecular modeling. *Coordination chemistry reviews*, 185:81–98, 1999.

[53] Anne-Lise Haenni, M Robert, W Vetter, L Roux, M Barbier, and E Lederer. Structure chimique des aspergillomarasmines a et b. *Helvetica Chimica Acta*, 48(4):729–750, 1965.

[54] Yoji Mikami and Toshiyuki Suzuki. Novel microbial inhibitors of angiotensin-converting enzyme, aspergillomarasmines a and b. *Agricultural and biological chemistry*, 47(11):2693–2695, 1983.

[55] Koshi Arai, Nami Ashikawa, Yasukazu Nakakita, Akihiro Matsuura, Naoki Ashizawa, and Masanobu Munekata. Aspergillomarasmine a and b, potent microbial inhibitors of endothelin-converting enzyme. *Bioscience, biotechnology, and biochemistry*, 57(11):1944–1945, 1993.

[56] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2):227–249, 1976.

[57] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[58] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[59] Michael J Frisch, GW Trucks, HB Schlegel, GE Scuseria, MAn Robb, JR Cheeseman, G Scalmani, V Barone, B Mennucci, GA Petersson, et al. Gaussian 09, revision d. 01, gaussian. *Inc.: Wallingford, CT*, 2009.

[60] S Miertuš, E Scrocco, and J Tomasi. Electrostatic interaction of a solute with a continuum. a direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, 55(1):117–129, 1981.

[61] Richard C Weisenberg, William J Deery, and Peter J Dickinson. Tubulin-nucleotide interactions during the polymerization and depolymerization of microtubules. *Biochemistry*, 15(19):4248–4254, 1976.

[62] Mary Ann Jordan and Leslie Wilson. Microtubules as a target for anticancer drugs. *Nature Reviews Cancer*, 4(4):253, 2004.

[63] RA Walker, ET O'brien, NK Pryer, MF Soboeiro, WA Voter, HP Erickson, and ED Salmon. Dynamic instability of individual microtubules analyzed by video light microscopy: rate constants and transition frequencies. *The Journal of cell biology*, 107(4):1437–1448, 1988.

[64] Peter K Sorger, Max Dobles, Regis Tournebize, and Anthony A Hyman. Coupling cell division and cell death to microtubule dynamics. *Current opinion in cell biology*, 9(6):807–814, 1997.

[65] Charles Dumontet and Mary Ann Jordan. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nature reviews Drug discovery*, 9(10):790, 2010.

[66] Eiman Mukhtar, Vaqar Mustafa Adhami, and Hasan Mukhtar. Targeting microtubules by natural agents for cancer therapy. *Molecular cancer therapeutics*, 13(2):275–284, 2014.

[67] Dirk-Jan Scheffers, Tanneke Den Blaauwen, and Arnold JM Driessen. Non-hydrolysable gtp-γ-s stabilizes the ftsz polymer in a gdp-bound state. *Molecular microbiology*, 35(5):1211–1219, 2000.

[68] Len Stephens, Trevor R Jackson, and Phillip T Hawkins. Activation of phosphatidylinositol 4, 5-bisphosphate supply by agonists and non-hydrolysable gtp analogues. *Biochemical Journal*, 296(2):481–488, 1993.

[69] Andrea E Prota, Katja Bargsten, Didier Zurwerra, Jessica J Field, José Fernando Díaz, Karl-Heinz Altmann, and Michel O Steinmetz. Molecular mechanism of action of microtubule-stabilizing anticancer agents. *Science*, 339(6119):587–590, 2013.

[70] John L Klepeis, Kresten Lindorff-Larsen, Ron O Dror, and David E Shaw. Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology*, 19(2):120–127, 2009.

[71] DS Cerutti, TE Cheatham III, TA Darden, RE Duke, TJ Giese, H Gohlke, AW Goetz, N Homeyer, S Izadi, P Janowski, et al. Amber 2016, 2016.

[72] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[73] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and JR Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[74] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An log n log(n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[75] Irina Massova and Peter A Kollman. Computational alanine scanning to probe protein- protein interactions: a novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, 121(36):8133–8143, 1999.

[76] Xiao-Xiang Zhang, Henry S Eden, and Xiaoyuan Chen. Peptides in cancer nanomedicine: drug carriers, targeting ligands and protease substrates. *Journal of controlled release*, 159(1):2–13, 2012.

[77] Viktoria Mikhalevich, Ioana Craciun, Myrto Kyropoulou, Cornelia G Palivan, and Wolfgang Meier. Amphiphilic peptide self-assembly: expansion to hybrid materials. *Biomacromolecules*, 18(11):3471–3480, 2017.

[78] Alicia López-Andarias, Javier López-Andarias, Carmen Atienza, Francisco J Chichón, José L Carrascosa, and Nazario Martín. Tuning optoelectronic and chiroptic properties of peptide-based materials by controlling the pathway complexity. *Chemistry–A European Journal*, 24(30):7755–7760, 2018.

[79] Danielle M Raymond and Bradley L Nilsson. Multicomponent peptide assemblies. *Chemical Society Reviews*, 47(10):3659–3720, 2018.

[80] Francesca Clerici, Emanuela Erba, Maria Luisa Gelmi, and Sara Pellegrino. Non-standard amino acids and peptides: from self-assembly to nanomaterials. *Tetrahedron letters*, 57(50):5540–5550, 2016.

[81] Sara Pellegrino, Giorgio Facchetti, Alessandro Contini, Maria Luisa Gelmi, Emanuela Erba, Raffaella Gandolfi, and Isabella Rimoldi. Ctr-1 mets7 motif inspiring new peptide ligands for cu (i)-catalyzed asymmetric henry reactions under green conditions. *RSC Advances*, 6(75):71529–71533, 2016.

[82] Sara Pellegrino, Nicolò Tonali, Emanuela Erba, Julia Kaffy, Myriam Taverna, Alessandro Contini, Mark Taylor, David Allsop, Maria Luisa Gelmi, and Sandrine Ongeri. β-hairpin mimics containing a piperidine–pyrrolidine scaffold modulate the β-amyloid aggregation process preserving the monomer species. *Chemical science*, 8(2):1295–1302, 2017.

[83] Raffaella Bucci, Priyadip Das, Filomena Iannuzzi, Marco Feligioni, Raffaella Gandolfi, Maria Luisa Gelmi, Meital Reches, and Sara Pellegrino. Self-assembly of an amphipathic $\alpha\alpha\beta$-tripeptide into cationic spherical particles for intracellular delivery. *Organic & biomolecular chemistry*, 15(32):6773–6779, 2017.

[84] Andrea Bonetti, Sara Pellegrino, Priyadip Das, Sivan Yuran, Raffaella Bucci, Nicola Ferri, Fiorella Meneghetti, Carlo Castellano, Meital Reches, and Maria Luisa Gelmi. Dipeptide nanotubes containing unnatural fluorine-substituted β2, 3-diarylamino acid and l-alanine as candidates for biomedical applications. *Organic letters*, 17(18):4468–4471, 2015.

[85] Sara Pellegrino, Andrea Bonetti, Francesca Clerici, Alessandro Contini, Alessandro Moretto, Raffaella Soave, and Maria Luisa Gelmi. 1 h-azepine-2-oxo-5-amino-5-carboxylic acid: A 310 helix inducer and an effective tool for functionalized gold nanoparticles. *The Journal of organic chemistry*, 80(11):5507–5516, 2015.

[86] Roberto Fanelli, Dorothee Berthomieu, Claude Didierjean, Abdelatif Doudouh, Aurelien Lebrun, Jean Martinez, and Florine Cavelier. Hydrophobic α, α-disubstituted disilylated tesdpg induces incipient 310-helix in short tripeptide sequence. *Organic letters*, 19(11):2937–2940, 2017.

[87] Hiroyuki Kobayashi, Takashi Misawa, Kenji Matsuno, and Yosuke Demizu. Preorganized cyclic α, α-disubstituted α-amino acids bearing functionalized

side chains that act as peptide-helix inducers. *The Journal of organic chemistry*, 82(19):10722–10726, 2017.

[88] Prema G Vasudev, Sunanda Chatterjee, Narayanaswamy Shamala, and Padmanabhan Balaram. Structural chemistry of peptides containing backbone expanded amino acid residues: conformational features of β, γ, and hybrid peptides. *Chemical reviews*, 111(2):657–687, 2010.

[89] Margareta Sohora, Mario Vazdar, Irena Sovic, Kata Mlinaric-Majerski, and Nikola Basaric. Photocyclization of tetra-and pentapeptides containing adamantylphthalimide and phenylalanines: reaction efficiency and diastereoselectivity. *The Journal of organic chemistry*, 83(24):14905–14922, 2018.

[90] Francelin Bouillere, Sophie Thétiot-Laurent, Cyrille Kouklovsky, and Valérie Alezra. Foldamers containing γ-amino acid residues or their analogues: structural features and applications. *Amino Acids*, 41(3):687–707, 2011.

[91] Sunanda Chatterjee, Prema G Vasudev, Srinivasarao Raghothama, Chandrasekharan Ramakrishnan, Narayanaswamy Shamala, and Padmanabhan Balaram. Expanding the peptide β-turn in $\alpha\gamma$ hybrid sequences: 12 atom hydrogen bonded helical and hairpin turns. *Journal of the American Chemical Society*, 131(16):5956–5965, 2009.

[92] Maruthi Konda, Rohit G Jadhav, Sayan Maiti, Shaikh M Mobin, Brice Kauffmann, and Apurba K Das. Understanding the conformational analysis of gababutin based hybrid peptides. *Organic & biomolecular chemistry*, 16(10):1728–1735, 2018.

[93] Li Guo, Weicheng Zhang, Ilia A Guzei, Lara C Spencer, and Samuel H Gellman. New preorganized γ-amino acids as foldamer building blocks. *Organic letters*, 14(10):2582–2585, 2012.

[94] Fabrizio Machetti, Alessandro Ferrali, Gloria Menchi, Ernesto G Occhiato, and Antonio Guarna. Oligomers of enantiopure bicyclic γ/δ-amino acids (btaa). 1. synthesis and conformational analysis of 3-aza-6, 8-dioxabicyclo [3.2. 1] octane-7-carboxylic acid oligomers (polybtg). *Organic letters*, 2(25):3987–3990, 2000.

[95] Paola Conti, Andrea Pinto, Pui E Wong, Louise L Major, Lucia Tamborini, Maria C Iannuzzi, Carlo De Micheli, Michael P Barrett, and Terry K Smith. Synthesis and in vitro/in vivo evaluation of the antitrypanosomal activity of 3-bromoacivicin, a potent ctp synthetase inhibitor. *ChemMedChem*, 6(2):329–333, 2011.

[96] Stefano Bruno, Andrea Pinto, Gianluca Paredi, Lucia Tamborini, Carlo De Micheli, Valeria La Pietra, Luciana Marinelli, Ettore Novellino, Paola Conti, and Andrea Mozzarelli. Discovery of covalent inhibitors of glyceraldehyde-3-phosphate dehydrogenase, a target for the treatment of malaria. *Journal of medicinal chemistry*, 57(17):7465–7471, 2014.

[97] Andrea Pinto, Lucia Tamborini, Gregorio Cullia, Paola Conti, and Carlo De Micheli. Inspired by nature: The 3-halo-4, 5-dihydroisoxazole moiety as a novel molecular warhead for the design of covalent inhibitors. *ChemMedChem*, 11(1):10–14, 2016.

[98] Sabrina Castellano, Dirk Kuck, Monica Viviano, Jakyung Yoo, Fabian López-Vallejo, Paola Conti, Lucia Tamborini, Andrea Pinto, José L Medina-Franco, and Gianluca Sbardella. Synthesis and biochemical evaluation of δ2-isoxazoline derivatives as dna methyltransferase 1 inhibitors. *Journal of medicinal chemistry*, 54(21):7663–7677, 2011.

[99] Kamalneet Kaur, Vinod Kumar, Anil Kumar Sharma, and Girish Kumar Gupta. Isoxazoline containing natural products as anticancer agents: a review. *European journal of medicinal chemistry*, 77:121–133, 2014.

[100] Raffaella Bucci, Sabrina Giofré, Francesca Clerici, Alessandro Contini, Andrea Pinto, Emanuela Erba, Raffaella Soave, Sara Pellegrino, and Maria Luisa Gelmi. Tetrahydro-4 h-(pyrrolo [3, 4-d] isoxazol-3-yl) methanamine: A bicyclic diamino scaffold stabilizing parallel turn conformations. *The Journal of organic chemistry*, 83(19):11493–11501, 2018.

[101] Misal Giuseppe Memeo, Marco Bruschi, Luca Bergonzi, Giovanni Desimoni, Giuseppe Faita, and Paolo Quadrelli. Cyclopenta [d] isoxazoline β-turn mimics: synthetic approach, turn driving force, scope, and limitations. *ACS Omega*, 3(10):13551–13558, 2018.

[102] René T Wintjens, Marianne J Rooman, and Shoshana J Wodak. Automatic classification and analysis of $\alpha\alpha$-turn motifs in proteins. *Journal of molecular biology*, 255(1):235–253, 1996.

[103] Michael J Kelso, Renée L Beyer, Huy N Hoang, Ami S Lakdawala, James P Snyder, Warren V Oliver, Tom A Robertson, Trevor G Appleton, and David P Fairlie. α-turn mimetics: short peptide α-helices composed of cyclic metallopentapeptide modules. *Journal of the American Chemical Society*, 126(15):4828–4842, 2004.

[104] Huy N Hoang, Russell W Driver, Renée L Beyer, Alpeshkumar K Malde, Giang T Le, Giovanni Abbenante, Alan E Mark, and David P Fairlie. Protein α-turns recreated in structurally stable small molecules. *Angewandte Chemie International Edition*, 50(47):11107–11111, 2011.

[105] Yarkali Krishna, Shrikant Sharma, Ravi S Ampapathi, and Dipankar Koley. Furan-based locked z-vinylogous γ-amino acid stabilizing protein α-turn in water-soluble cyclic $\alpha 3\gamma$ tetrapeptides. *Organic letters*, 16(8):2084–2087, 2014.

[106] Lei Wang, Pascale Coric, Kexin Zhu, Wang-Qing Liu, Michel Vidal, Serge Bouaziz, and Sylvain Broussy. Synthesis and characterization of water-soluble macrocyclic peptides stabilizing protein α-turn. *Organic & biomolecular chemistry*, 16(3):459–471, 2018.

[107] MJ Abraham, D Van Der Spoel, E Lindahl, and B Hess. the gromacs development team, gromacs user manual version 5.0. 4. *J. Mol. Model*, 2014.

[108] Enrico Gandini, Federico Dapiaggi, Francesco Oliva, Stefano Pieraccini, and Maurizio Sironi. Well-tempered metadynamics based method to evaluate universal peptidomimetics. *Chemical Physics Letters*, 706:729–735, 2018.

[109] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[110] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.

[111] Jonas S Laursen, Jens Engel-Andreasen, Peter Fristrup, Pernille Harris, and Christian A Olsen. Cis–trans amide bond rotamers in β-peptoids and peptoids: Evaluation of stereoelectronic effects in backbone and side chains. *Journal of the American Chemical Society*, 135(7):2835–2844, 2013.

[112] Sara Pellegrino, Alessandro Contini, Maria Luisa Gelmi, Leonardo Lo Presti, Raffaella Soave, and Emanuela Erba. Asymmetric modular synthesis of a semirigid dipeptide mimetic by cascade cycloaddition/ring rearrangement and borohydride reduction. *The Journal of organic chemistry*, 79(7):3094–3102, 2014.

[113] Vincenzo Pavone, Girolamo Gaeta, Angela Lombardi, Flavia Nastri, Ornella Maglio, Carla Isernia, and Michele Saviano. Discovering protein secondary structures: Classification and description of isolated α-turns. *Biopolymers*, 38(6):705–721, 1996.

[114] Henri Grosjean. *DNA and RNA modification enzymes: structure, mechanism, function and evolution.* CRC Press, 2009.

[115] Frank Jühling, Mario Mörl, Roland K Hartmann, Mathias Sprinzl, Peter F Stadler, and Joern Pütz. trnadb 2009: compilation of trna sequences and trna genes. *Nucleic acids research*, 37(suppl_1):D159–D162, 2008.

[116] Christopher Deutsch, Basma El Yacoubi, Valérie de Crécy-Lagard, and Dirk Iwata-Reuyl. Biosynthesis of threonylcarbamoyl adenosine (t6a), a universal trna nucleoside. *Journal of Biological Chemistry*, 287(17):13666–13673, 2012.

[117] Paul F Agris, Franck AP Vendeix, and William D Graham. trna's wobble decoding of the genome: 40 years of modification. *Journal of molecular biology*, 366(1):1–13, 2007.

[118] Marie-Claire Daugeron, Tineke L Lenstra, Martina Frizzarin, Basma El Yacoubi, Xipeng Liu, Agnes Baudin-Baillieu, Philip Lijnzaad, Laurence Decourty, Cosmin Saveanu, Alain Jacquier, Frank C P Holstege, Valérie de Crécy-Lagard, Herman van Tilbeurgh, and Domenico Libri. Gcn4 misregulation reveals a direct role for the evolutionary conserved ekc/keops in the t6a modification of trnas. *Nucleic acids research*, 39(14):6148–6160, 2011.

[119] Changyi A Lin, Steven R Ellis, and Heather L True. The sua5 protein is essential for normal translational regulation in yeast. *Molecular and cellular biology*, 30(1):354–363, 2010.

[120] Basma El Yacoubi, Isabelle Hatin, Christopher Deutsch, Tamer Kahveci, Jean-Pierre Rousset, Dirk Iwata-Reuyl, Alexey G Murzin, and Valérie de Crécy-Lagard. A role for the universal kae1/qri7/ygjd (cog0533) family in trna modification. *The EMBO journal*, 30(5):882–893, 2011.

[121] Valérie de Crécy-Lagard. Identification of genes encoding trna modification enzymes by comparative genomics. *Methods in enzymology*, 425:153–183, 2007.

[122] Basma El Yacoubi, Benjamin Lyons, Yulien Cruz, Robert Reddy, Brian Nordin, Fabio Agnelli, James R Williamson, Paul Schimmel, Manal A Swairjo, and Valerie de Crecy-Lagard. The universal yrdc/sua5 family is required for the formation of threonylcarbamoyladenosine in trna. *Nucleic acids research*, 37(9):2894–2909, 2009.

[123] Madhusudhan Srinivasan, Preeti Mehta, Yao Yu, Evelyn Prugar, Eugene V Koonin, A Wali Karzai, and Rolf Sternglanz. The highly conserved keops/ekc complex is essential for a universal trna modification, t6a. *The EMBO journal*, 30(5):873–881, 2011.

[124] Jennifer I Handford, Bérengère Ize, Grant Buchanan, Gareth P Butland, Jack Greenblatt, Andrew Emili, and Tracy Palmer. Conserved network of proteins essential for bacterial viability. *Journal of bacteriology*, 191(15):4732–4749, 2009.

[125] Davide Vecchietti, Silvia Ferrara, Ruggero Rusmini, Raffaella Macchi, Mario Milani, and Giovanni Bertoni. Crystal structure of yeaz from pseudomonas aeruginosa. *Biochemical and biophysical research communications*, 470(2):460–465, 2016.

[126] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, Chris Sander, and Michael Scharf. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1995.

[127] Cristina Paissoni, Dimitrios Spiliotopoulos, Giovanna Musco, and Andrea Spitaleri. Gmxpbsa 2.1: A gromacs tool to perform mm/pbsa and computational alanine scanning. *Computer Physics Communications*, 186:105–107, 2015.

[128] Stefano Pieraccini, Stefano Rendine, Chacko Jobichen, Prerna Domadia, J Sivaraman, Pierangelo Francescato, Giovanna Speranza, and Maurizio Sironi. Computer aided design of ftsz targeting oligopeptides. *RSC advances*, 3(6):1739–1743, 2013.

[129] Lutz Hilterhaus, Andreas Liese, Ulrich Kettling, and Garabed Antranikian. *Applied biocatalysis: from fundamental science to industrial applications*. John Wiley & Sons, 2016.

[130] Suresh S Tate and Alton Meister. γ-glutamyl transpeptidase: catalytic, structural and functional aspects. In *The Biological Effects of Glutamic Acid and Its Derivatives*, pages 357–368. Springer, 1981.

[131] Immacolata Castellano and Antonello Merlino. γ-glutamyltranspeptidases: sequence, structure, biochemical properties, and biotechnological applications. *Cellular and molecular life sciences*, 69(20):3381–3394, 2012.

[132] Vittorio Ricci, Maria Giannouli, Marco Romano, and Raffaele Zarrilli. Helicobacter pylori gamma-glutamyl transpeptidase and its pathogenic role. *World journal of gastroenterology: WJG*, 20(3):630, 2014.

[133] Carita Oinonen and Juha Rouvinen. Structural comparison of ntn-hydrolases. *Protein Science*, 9(12):2329–2337, 2000.

[134] Makoto Inoue, Jun Hiratake, Hideyuki Suzuki, Hidehiko Kumagai, and Kanzo Sakata. Identification of catalytic nucleophile of escherichia coli γ-glutamyltranspeptidase by γ-monofluorophosphono derivative of glutamic acid: N-terminal thr-391 in small subunit is the nucleophile. *Biochemistry*, 39(26):7764–7771, 2000.

[135] R Donald Allison. [51] γ-glutamyl transpeptidase: Kinetics and mechanism. In *Methods in enzymology*, volume 113, pages 419–437. Elsevier, 1985.

[136] Naoyuki Taniguchi and Yoshitaka Ikeda. c-glutamyl transpeptidase: catalytic mechanism and gene expression. *Adv Enzymol Relat Areas Mol Biol*, 72:239–278, 1998.

[137] Gregory A Thompson and Alton Meister. Hydrolysis and transfer reactions catalyzed by γ-glutamyl transpeptidase; evidence for separate substrate sites and for high affinity of l-cystine. *Biochemical and biophysical research communications*, 71(1):32–36, 1976.

[138] Suresh S Tate and Alton Meister. Interaction of γ-glutamyl transpeptidase with amino acids, dipeptides, and derivatives and analogs of glutathione. *Journal of Biological Chemistry*, 249(23):7593–7602, 1974.

[139] CS Hanes, FJR Hird, and FA Isherwood. Enzymic transpeptidation reactions involving γ-glutamyl peptides and α-amino-acyl peptides. *Biochemical Journal*, 51(1):25, 1952.

[140] Hiroaki Sakai, Noriyoshi Sakabe, Kyoyu Sasaki, Wataru Hashimoto, Hideyuki Suzuki, Hiromi Tachi, Hidehiko Kumagai, and Kiwako Sakabe. A preliminary description of the crystal structure of γ-glutamyltranspeptidase from e. coli k-12. *The Journal of Biochemistry*, 120(1):26–28, 1996.

[141] Toshihiro Okada, Hideyuki Suzuki, Kei Wada, Hidehiko Kumagai, and Keiichi Fukuyama. Crystal structures of γ-glutamyltranspeptidase from escherichia coli, a key enzyme in glutathione metabolism, and its reaction intermediate. *Proceedings of the National Academy of Sciences*, 103(17):6471–6476, 2006.

[142] Hideyuki Suzuki and Hidehiko Kumagai. Autocatalytic processing of γ-glutamyltranspeptidase. *Journal of Biological Chemistry*, 277(45):43536–43543, 2002.

[143] Kei Wada, Jun Hiratake, Machiko Irie, Toshihiro Okada, Chiaki Yamada, Hidehiko Kumagai, Hideyuki Suzuki, and Keiichi Fukuyama. Crystal structures of escherichia coli γ-glutamyltranspeptidase in complex with azaserine

and acivicin: Novel mechanistic implication for inhibition by glutamine antagonists. *Journal of molecular biology*, 380(2):361–372, 2008.

[144] Wataru Hashimoto, Hideyuki Suzuki, Kenji Yamamoto, and Hidehiko Kumagai. Effect of site-directed mutations on processing and activity of γ-glutamyltranspeptidase of escherichia coli k-12. *The Journal of Biochemistry*, 118(1):75–80, 1995.

[145] Ping-Lin Ong, Ya-Feng Yao, Yih-Ming Weng, Wen-Hwei Hsu, and Long-Liu Lin. Residues arg114 and arg337 are critical for the proper function of escherichia coli γ-glutamyltranspeptidase. *Biochemical and biophysical research communications*, 366(2):294–300, 2008.

[146] Wen-Hwei Hsu, Ping-Lin Ong, Shih-Chun Chen, and Long-Liu Lin. Contribution of ser463 residue to the enzymatic and autoprocessing activities of escherichia coli gamma-glutamyltranspeptidase. *Indian J Biochem Biophys*, 46(4):281–288, 2009.

[147] Hideyuki Suzuki, C Yamada, and K Kato. γ-glutamyl compounds and their enzymatic production using bacterial γ-glutamyltranspeptidase. *Amino acids*, 32(3):333–340, 2007.

[148] Hideyuku Suzuki, Hidehiko Kumagai, and Tatsurokuro Tochikura. gamma-glutamyltranspeptidase from escherichia coli k-12: purification and properties. *Journal of bacteriology*, 168(3):1325–1331, 1986.

[149] Keigo Shibayama, Jun-ichi Wachino, Yoshichika Arakawa, Massoud Saidijam, Nicholas G Rutherford, and Peter JF Henderson. Metabolism of glutamine and glutathione via γ-glutamyltranspeptidase and glutamate transport in helicobacter pylori: possible significance in the pathophysiology of the organism. *Molecular microbiology*, 64(2):396–406, 2007.

[150] Albert C Pan, David W Borhani, Ron O Dror, and David E Shaw. Molecular determinants of drug–receptor binding kinetics. *Drug discovery today*, 18(13-14):667–673, 2013.

[151] Robert B Best. Atomistic molecular simulations of protein folding. *Current opinion in structural biology*, 22(1):52–61, 2012.

[152] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

[153] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[154] Glenn M Torrie and John P Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28(4):578–581, 1974.

[155] Christoph Dellago, Peter G Bolhuis, Félix S Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *The Journal of Chemical Physics*, 108(5):1964–1977, 1998.

[156] Urmi Doshi and Donald Hamelberg. Achieving rigorous accelerated conformational sampling in explicit solvent. *The Journal of Physical Chemistry Letters*, 5(7):1217–1224, 2014.

[157] Jose C Flores-Canales and Maria Kurnikova. Targeting electrostatic interactions in accelerated molecular dynamics with application to protein partial unfolding. *Journal of chemical theory and computation*, 11(6):2550–2559, 2015.

[158] Yinglong Miao, Victoria A Feher, and J Andrew McCammon. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation*, 11(8):3584–3595, 2015.

[159] Yinglong Miao and J Andrew McCammon. Graded activation and free energy landscapes of a muscarinic g-protein–coupled receptor. *Proceedings of the National Academy of Sciences*, 113(43):12162–12167, 2016.

[160] Kalli Kappel, Yinglong Miao, and J Andrew McCammon. Accelerated molecular dynamics simulations of ligand binding to a muscarinic g-protein-coupled receptor. *Quarterly reviews of biophysics*, 48(4):479–487, 2015.

[161] Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659, 2017.

[162] Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry*, 13(8):952–962, 1992.

[163] Susanna K Lüdemann, Valère Lounnas, and Rebecca C Wade. How do substrates enter and products exit the buried active site of cytochrome p450cam? 1. random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *Journal of molecular biology*, 303(5):797–811, 2000.

[164] Eric Vanden-Eijnden and Giovanni Ciccotti. Second-order integrators for langevin equations with holonomic constraints. *Chemical physics letters*, 429(1-3):310–316, 2006.

[165] Kim-Hung Chow and David M Ferguson. Isothermal-isobaric molecular dynamics simulations with monte carlo volume sampling. *Computer physics communications*, 91(1-3):283–289, 1995.

[166] Johan Åqvist, Petra Wennerström, Martin Nervall, Sinisa Bjelic, and Bjørn O Brandsdal. Molecular dynamics simulations of water and biomolecules with a monte carlo constant pressure algorithm. *Chemical physics letters*, 384(4-6):288–294, 2004.

[167] Steffen Lindert, Denis Bucher, Peter Eastman, Vijay Pande, and J Andrew McCammon. Accelerated molecular dynamics simulations with the amoeba polarizable force field on graphics processing units. *Journal of chemical theory and computation*, 9(11):4684–4691, 2013.

[168] Lisbeth Ravnkilde Kjølbye, Anne Laustsen, Mikkel Vestergaard, Xavier Periole, Leonardo De Maria, Allan Svendsen, Andrea Coletta, and Birgit Schiøtt. Molecular modeling investigation of the interaction between humicola insolens cutinase and sds surfactant suggests a mechanism for enzyme inactivation. *Journal of chemical information and modeling*, 59(5):1977–1987, 2019.

[169] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[170] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999.

[171] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[172] Roman A Laskowski and Mark B Swindells. Ligplot+: multiple ligand–protein interaction diagrams for drug discovery, 2011.

[173] Cinzia Calvio, Fabio Romagnuolo, Francesca Vulcano, Giovanna Speranza, and Carlo F Morelli. Evidences on the role of the lid loop of γ-glutamyltransferases (ggt) in substrate selection. *Enzyme and microbial technology*, 114:55–62, 2018.

[174] Levi CT Pierce, Romelia Salomon-Ferrer, Cesar Augusto F. de Oliveira, J Andrew McCammon, and Ross C Walker. Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of chemical theory and computation*, 8(9):2997–3002, 2012.

[175] Phineus RL Markwick, Guillaume Bouvignies, and Martin Blackledge. Exploring multiple timescale motions in protein gb3 using accelerated molecular dynamics and nmr spectroscopy. *Journal of the American Chemical Society*, 129(15):4724–4730, 2007.

[176] Yinglong Miao, Sara E Nichols, Paul M Gasper, Vincent T Metzger, and J Andrew McCammon. Activation and dynamic network of the m2 muscarinic receptor. *Proceedings of the National Academy of Sciences*, 110(27):10982–10987, 2013.

[177] J Andrew McCammon and Scott H Northrup. Gated binding of ligands to proteins. *Nature*, 293(5830):316, 1981.

[178] Huan-Xiang Zhou. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophysical journal*, 98(6):L15–L17, 2010.

[179] Harish Vashisth and Cameron F Abrams. Ligand escape pathways and (un) binding free energy calculations for the hexameric insulin-phenol complex. *Biophysical journal*, 95(9):4193–4204, 2008.

[180] Sean L Seyler, Avishek Kumar, Michael F Thorpe, and Oliver Beckstein. Path similarity analysis: a method for quantifying macromolecular pathways. *PLoS computational biology*, 11(10):e1004568, 2015.

[181] M Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1):1–72, 1906.

[182] Helmut Alt and Michael Godau. Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02):75–91, 1995.

[183] Daniel P Huttenlocher, Gregory A Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.

[184] Helmut Alt, Bernd Behrends, and Johannes Blömer. Approximate matching of polygonal shapes. *Annals of Mathematics and Artificial Intelligence*, 13(3-4):251–265, 1995.

[185] Helmut Alt and Ludmila Scharf. Computing the hausdorff distance between curved objects. *International Journal of Computational Geometry & Applications*, 18(04):307–320, 2008.

[186] Richard Car and Mark Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical review letters*, 55(22):2471, 1985.

[187] Francesco Oliva, Raffaella Bucci, Lucia Tamborini, Stefano Pieraccini, Andrea Pinto, and Sara Pellegrino. Bicyclic pyrrolidine-isoxazoline γ amino acid: A constrained scaffold for stabilizing α-turn conformation in isolated peptides. *Frontiers in chemistry*, 7, 2019.