Euskal Herriko Unibertsitatea (EHU/UPV)

Kimika Fakultatea/Facultad de Química

Official degree in Chemistry

Bachelor thesis on Theoretical Chemistry and Computational Modelling

The interaction landscape of amyloid β-Zn^{2+} from molecular dynamics simulations

Author: Julen Aduriz Arrizabalaga

Supervisor: Dr. David de Sancho

Co-Supervisor: Dr. Jose María Mercero

San Sebastián, June 2020

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Contents

Acknowledgement

I would like to express my sincere gratitude to:

- Dr. David de Sancho and Dr. Jose María Mercero for supporting me throughout the whole process and for bringing the best out of me.
- My family for their constant support, patience and love.
- My girlfriend Ane for all the times she was there to help me let off steam.
- My friends at the University. Martin, Leyre, Ainhoa, Jon... We made it!
- My friends in *Meltdown* for their unconditional love and support.
- Everybody in the Theoretical Chemistry group of the University for taking me like their own.
- Everybody in the BioKT group for their patience and kindness.

Abstract

Amyloid fibrils are stable forms of misfolded proteins associated with numerous neurodegenerative diseases. Among these, Alzheimer's disease may be the most prevalent, with over 50 million dementia cases reported by the World Health Organization in 2019. The molecular origin of Alzheimer's is linked to amyloid fibril formation by misfolded Aβ-peptide (Aβ). These fibrils can form aggregates that are stabilized by the presence of Zn^{2+} cations. Although many possible structures have been reported in the last few years for the Zn^{2+}-Aβ complexes, details about the molecular interactions involved are still lacking. In this work, I present a detailed computational study of the different possible structures that the system could show with their respective weights. I have employed equilibrium classical molecular dynamics simulations and the Hamiltonian replica exchange method in order to characterize these bound states of Zn^{2+}.

Laburpena

Amiloide zuntzak hainbat gaixotasun neurodegeneratiboren eragile diren proteina desegituratuen egoera egonkorrak dira. Gaixotasun hauen artean, Alzheimerra da nabarmenena. 2019an Osasunaren Munduarteko Erakundeak 50 milioi kasu baina gehiago erregistratu zituen. Alzheimerraren eragilea maila molekularrean Aβ-peptido (Aβ) amiloidearen desegituraren ondorioz sortutako zuntzak dira. Peptido pilaketa hauek Zn^{2+} katioiaren presentzian egonkortu egiten dira. Nahiz eta Zn^{2+}-Aβ sistemaren egitura ugari ezagunak diren, haien arteko interakzio molekularrei buruzko informazio eza oraindik ere handia da. Ikerketa proiektu honetan, egitura ezberdin posibleei buruzko azterketa konputazional detailatua eta bakoitzaren garrantzia aurkezten dut. Oreka klasikoko dinamika molekularrak eta *Hamiltonian replica exchange* metodoetaz baliatu naiz Zn^{2+}ak erakusten dituen koordinazio egoera ezberdinak karakterizatzeko.

Chapter 1

Introduction

Proteins are the workhorses of living organisms, undertaking most of the biological functions. About a third of the proteins encoded in eukaryotic genomes are intrinsically disordered, meaning that they do not acquire a well defined three dimensional structure in their active form. Some of these intrinsically disordered proteins (IDPs) are tightly related to diseases [1], including Alzheimer's, Parkinson's and Type II diabetes (connected, respectively, to amyloid β-peptide, α-synuclein and the islet amyloid polypeptide). In particular, Alzheimer's disease is known as the leading cause of senile dementia, of which the World Health Organization (WHO) reported over 50 million cases worldwide on 2019 [2]. These numbers are likely to increase rapidly unless effective therapeutics are developed. Even though the cause of Alzheimer's disease is not completely known, a relation with aggregation and deposition of Aβ in neural tissue is widely accepted as part of the cause of the disease.

Transition metal ions and oxidative metabolism have been proposed to play fundamental roles in the processes of aggregation and deposition of Aβ in Alzheimer's disease [3]. Binding of divalent metal ions, such as copper (Cu^{2+}), iron (Fe^{2+}) and Zinc (Zn^{2+}), with disordered fibrillogenic proteins, such as Aβ, influences the aggregation process of the protein, contributing directly to the severity of the neurodegenerative disease [4]. It has been reported that both monomeric and oligomeric forms of Aβ are toxic to neurons [5] and that said cations influence toxicity [6]. Interestingly, the ion concentration of

Zn^{2+} in the brain, which varies from 150 to 200 μM in the neocortex, is one order of magnitude higher than the ion concentration in blood. Furthermore, even though the Zn^{2+} concentration stays relatively constant throughout adult life, a significantly elevated concentrations have been found on the brains affected by Alzheimer's disease [7]. Hence, the role that Zn^{2+} ions play in Alzheimer's disease has become of great interest.

In order to understand the role plaid by metal-protein interactions in the disease, and to be able to design effective drugs targeting Aβ-Zn^{2+} complexes, characterizing the specific modes of interaction between Zn^{2+} and Aβ proteins is essential. Multiple experimental structures have been reported, such as the one for Aβ(1-16) by Zirah et al. [8], which was meant to be used as a therapeutic target. In Figure 1 we show a cartoon representation of a molecular model built combining the Zirah structure and a full sequence structure of Aβ [9], which provides a model for the protein-metal interactions [10]. Alternative bound states of the cation have also been proposed, which may have a role in amyloid formation. Theoretical and computational chemistry methods provide an ideal shortcut for identifying the different bound states of the metal to the protein and understanding the origin of the molecular interactions involved.

Here we use molecular dynamics (MD) to study the Aβ-Zn^{2+} system. Benefiting from the variety of possibilities that the technique offers, we propose a research study on different conformational states of Aβ-Zn^{2+} in an effort to find previously unknown states that could be used as therapeutic targets. As a starting point, we selected the structure that Zirah et al. proposed (PDB id 1ZE9) [8], as comparison between theoretical and experimental

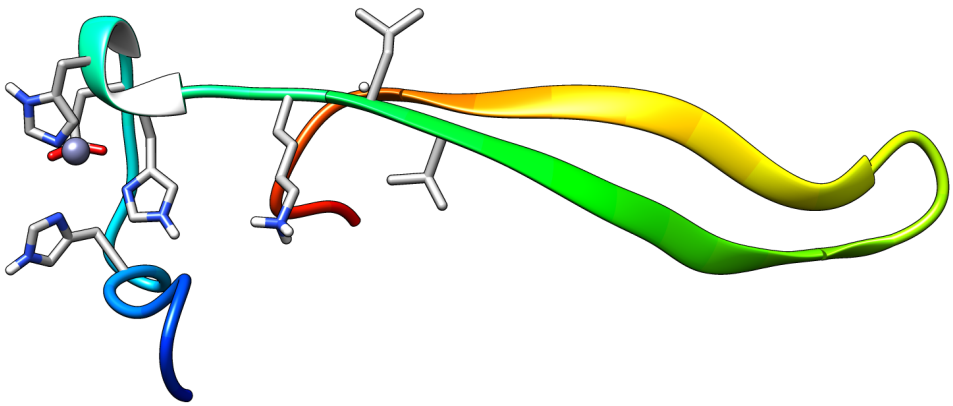

Figure 1: *Molecular model for the full-length Aβ-Zn^{2+} complex.*

results is imperative. The 16 residue system we focus on, is a minimal model for the full-length Aβ_{40} protein, but comprises all the amino acid residues that have been proposed to interact with the Zn^{2+} ion [8]. In the experiments, the Zn^{2+} was found on a tetrahedral state, coordinated to four amino acids: His6, Glu11, His13 and His14. Histidines coordinated to the Zn^{2+} ion from N-δ1, N-ε and N-δ1 respectively, whereas Glu11 was found coordinated through its carboxylate.

Although this structure has been found experimentally, it does not necessarily mean it is the only structure the system might be stable in and possibilities for different coordination states have been researched [10]. In this project, we are aspiring to find different conformations for the system, on an effort to shade some light on the matter. We find different coordination states, other than the one proposed on the reference structure, where Zn^{2+} is found coordinated to a variety of new residues previously unreported.

Chapter 2

Methods

In this work we use classical molecular dynamics (MD) with explicit solvent to study different conformational states of the system of interest. Using this methodology we can understand the interactions of the Aβ peptide with Zn^{2+}. In this section we present details of the molecular models used for the description of the protein and its interaction with the metal cation and solvent. We also describe the specific types of simulation methods that we used (i.e. equilibrium and Hamiltonian replica exchange) and the techniques for the analysis of the resulting trajectories.

2.1 Molecular dynamics simulations

Classical MD simulations consist in propagating Newton's equations of motion of atoms and molecules in a system of interest in order to reproduce their physical movement. Atoms and molecules are allowed to interact for a defined period of time (typically in the ns to μs range) showing the dynamic evolution of the system. This evolution is dictated by inter-atomic forces determined by energy functions or "force fields". Using these forces, one can propagate the Cartesian coordinates of the system after small time-steps (usually, 1 to 2 fs in length). After an often very large number of time-steps, one can sample the conformational space and estimate equilibrium and dynamic properties of complex systems, serving as interface between theory and experiment. MD simulations explain their dynamic motions which are used to deduce structural and dynamic properties. In

this way, large volumes of information are obtained from these simulations [11].

2.2 Force field and water models

As mentioned above, in order to propagate Newton's equations of motion, we need to calculate the pairwise forces acting on each particle, $\mathbf{f}_i$, which we obtain as the derivative of the potential V with respect to the particle positions $\mathbf{r}_i$

$$\mathbf{f}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{2.1}$$

The potential energy (V) is determined by the force field, which is defined as a sum of energy terms

$$V = V_{\text{bonded}} + V_{\text{nonbonded}} \tag{2.2}$$

In this expression, the bonded contribution is also given as a summation

$$V_{\text{bonded}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsions}} \tag{2.3}$$

where each of the terms above have a particular functional form, which are shown below

$$V_{\text{bonds}} = k_r \left(r - r_0\right)^2 \tag{2.4}$$

$$V_{\text{angles}} = k_\theta \left(\theta - \theta_0\right)^2 \tag{2.5}$$

$$V_{\text{torsions}} = \sum_n k_n \left(\cos n\varphi\right) \tag{2.6}$$

Equations 2.4 and 2.5 above correspond to terms for bonds and angles and are defined as harmonic potentials, where k_r and k_θ are the force constants, r is the bond length for an atom pair, r_0 is corresponding equilibrium bond length, θ is the value of the angle and θ_0 is the value of the equilibrium angle. The torsional term, defined in Equation 2.6, corresponding to torsional rotations of 4 atoms about a central bond, also known as dihedrals, is defined as a sum of n terms. Nonbonded interactions are described as a sum

of a Van der Waals and an electrostatic term,

$$V_{\text{nonbonded}} = V_{\text{electrostatic}} + V_{\text{van der Waals}} \tag{2.7}$$

Van der Waals interactions are described as

$$V_{\text{van der Waals}} = \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.8}$$

where ε is the term that describes the Lennard-Jones well-depth, σ defines Lennard-Jones radius and r_{ij} is the distance between non-bonded atoms. Lastly, the electrostatic contribution is simply computed using a Coulomb expression

$$V_{\text{Electrostatics}} = \frac{q_i q_j}{4\pi D r_{ij}} \tag{2.9}$$

q refers to each atom charge and D stands for the dielectric constant. Note that hydrogen bonds are treated as simple electrostatics.

Considering the large number of atom types in amino acid residues, force fields include a vast number of parameters. These constants and variables are defined either by a geometry file or by the parameter file of the selected force field. Geometry files, such as PDB files, define distances (r and r_{ij}) and angles (θ and φ), whereas force fields parameter files define constants (k_r, k_θ and k_n), equilibrium terms (r_0 and θ_0) and LJ terms (ε and σ).

As seen in Figure 2, during a MD simulation, trajectory files are being overwritten continuously from the equations of motion and used to calculate future trajectories, but, force field parameters are defined by selecting a force field and do not change throughout the whole simulation. Parameter files of force fields are obtained from parametrization processes and are defined as constant values, this is why they are defined once. Many parametrization processes are being used to define these parameters, among some of the usual practices quantum mechanical approximations highlight, often in the gas phase, with the expectation of some correlation with condensed phase properties and empirical modifications of potentials to match experimental observables [12, 13]. The accuracy and precision of these parametrization processes determines the accuracy of each force field

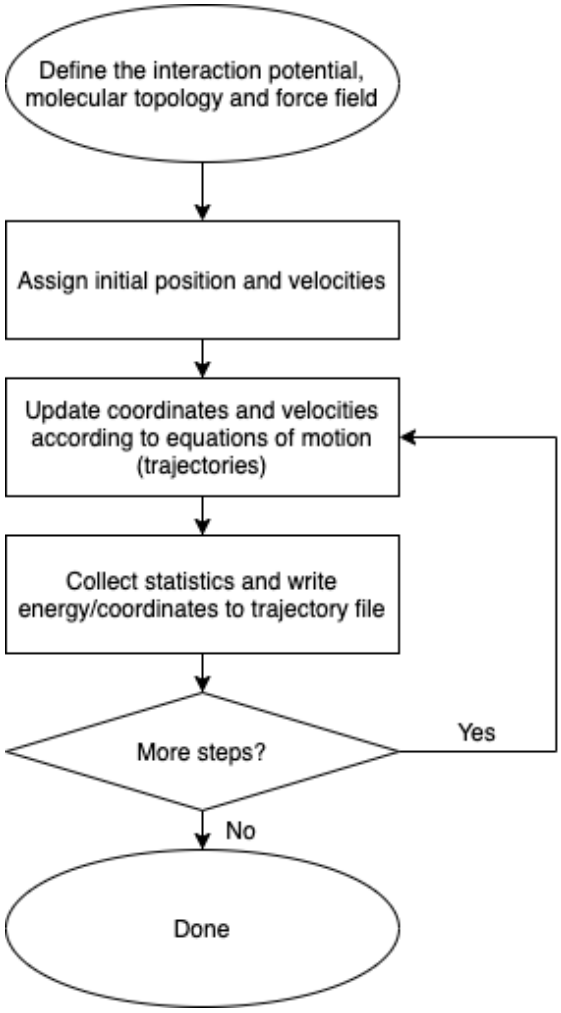

Figure 2: *Flowchart of MD simulation workflow.*

and this is why a big weight is given to a correct selection of the force field.

Intrinsically disordered proteins (IDPs) are challenging to study both experimentally and computationally. These proteins have native structures where tertiary structure is not well-defined and they must be instead described as an ensemble of interconverting conformations. It is necessary to determine the heterogeneous ensembles of conformations that they adopt to structurally characterize them. As mentioned, MD simulations are strongly dependent on the accuracy of the selected force field to replicate the system of interest and IDPs are challenging structures to characterize. For this reason, force fields have been specifically re-parameterized to recapitulate the properties of IDPs [14, 15].

Huang et al. reported very satisfactory results in 2016 with the CHARMM36m (C36m) force field after they improved accuracy in generating polypeptide backbone conforma-

tional ensembles for intrinsically disordered peptides and proteins [16]. C36m being an all-atom additive force field, it defines all variables for each atom specifically. Although force fields will define the gradient of the function of the system, the way the solvent, water in this case, is defined directly affects on the atomistic simulations, or MD simulations [17]. Relying on the results Huang et al. obtained [16], we decided to opt for the CHARMM36m force field and TIP3P water, which is a three-site model [18].

2.3 The molecular dynamic simulation workflow

In this work, MD simulations were carried out using the GROMACS software package, created by the Biophysical Chemistry Department of University of Groningen and that is designed for simulations of proteins, lipids, and nucleic acids [19]. As shown in Figure 3, preparing and running a MD simulation is a quite elaborate process. Briefly, we can split the process in three different parts: set-up, production run and analysis. In the set-up, we first convert a PDB file into GROMACS' own structure format, `gro` file. When we converted the files we had to specify the protonation state of each histidine as GROMACS sets them as HSD histidines (histidines protonated in their δ N). In our case, at least one histidine (His13) had to be defined as HSE (histidine protonated in its ε N) as it was coordinated to the Zn^{2+} from its δ N. Next, we must define the size of the simulation box. Defining a simulation box is a process that needs to be approached carefully. All of the simulations will be carried out inside the simulation box, hence, defining the dimensions of the box is very important. Taking into account that we are applying boundary conditions, if the box is too small, the periodic images of the system will interact with each other. On the contrary, if the box is too big, the calculations become bigger and slower [20]. In this case, we opted for a cubic 4 nm×4 nm×4 nm box. Lastly, we add water and ions to the system. We solvated the system with the water model selected beforehand and added Cl^- and Na^+ ions at 0.1 M concentration.

Once the initial set-up is done, one must minimize the initial configuration to avoid clashes between atoms. Next, the box is equilibrated in the NVT ensemble for a short time, with position restraints on the protein heavy atoms, which allows the water molecules to equilibrate, and large forces in the initial state to decrease. Next, another short simulation

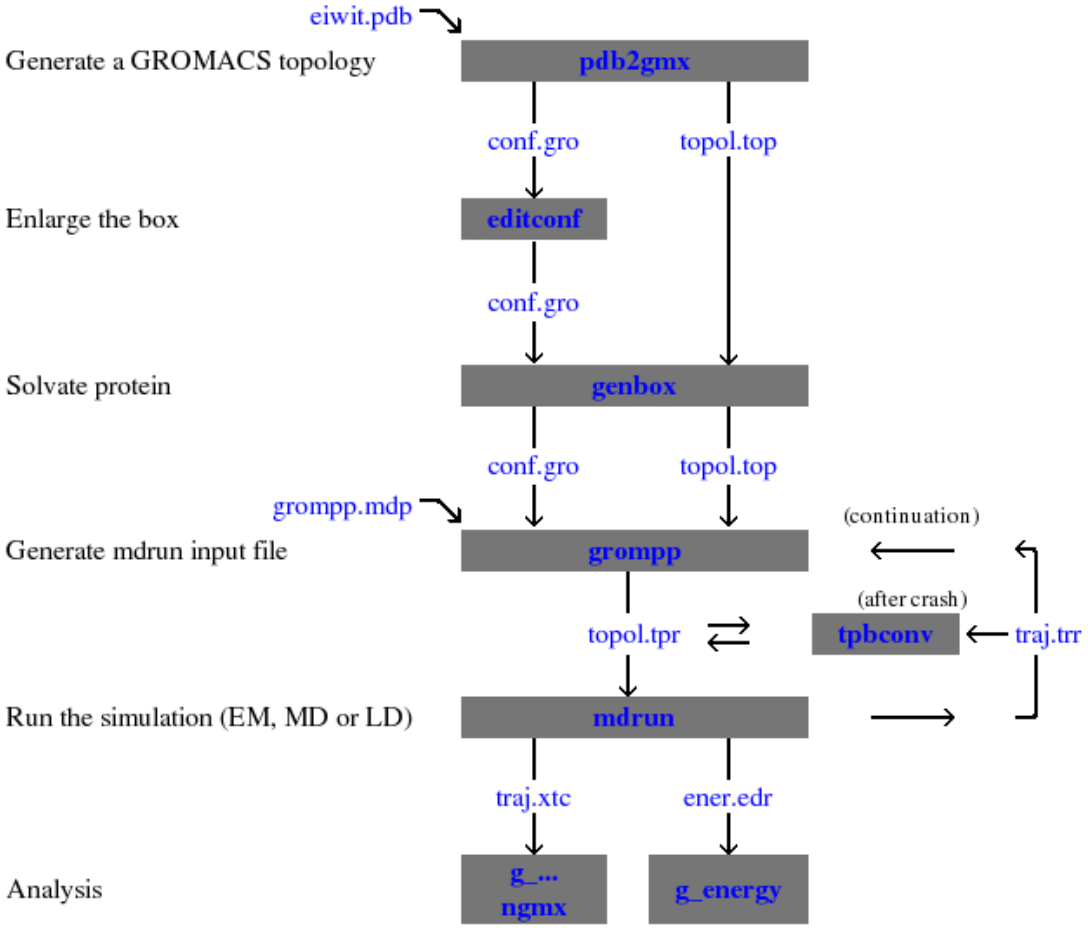

Figure 3: *Flowchart of the steps involved in running an MD simulation using GROMACS.*

is run in the NPT ensemble, now without restraints, in order to equilibrate the density at a pressure of 1 bar. From the final state of this simulation, the production simulations are run for a longer time-scale. This results in often very large datasets that must be analyzed.

2.3.1 Hamiltonian replica exchange

Due to the short integration timestep, the time-scales of MD simulations are a severe limitation of this technique. Because we are often interested in rare events like slow conformational transitions, equilibrium MD simulations are often insufficient to sample the processes of interest. As a response to this issue, many enhanced sampling techniques that allow to accelerate MD simulations have been developed [21]. There is a wide variety of techniques that come in handy, such as umbrella sampling, simulated tempering or

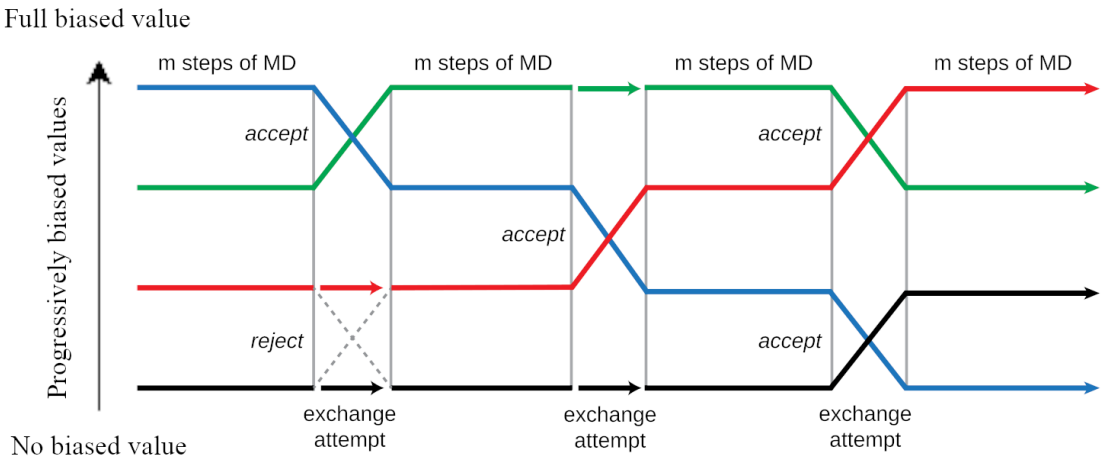

Figure 4: *Diagram showing HREMD simulation functioning.*

Hamiltonian replica exchange. Each technique has its own downsides and upsides, and therefore, depending on the approach that we are willing to give to our simulations, a different technique should be used. In our case, knowing that we are interested on recording conformational transitions, we decided to work with Hamiltonian replica exchange (HREMD) simulations.

On HREMD simulations, different replicas run in parallel exchanging coordinates every given time. Each replica is defined by a Hamiltonian with a biased variable and the biased variable is chosen beforehand to force the system to experience the events of interest. Exchanges are given with a probability and if the probability is not high enough, an exchange attempt can be rejected (as the exchange attempt between the red and black lines shown in Figure 4). This probability is directly dependent on the biased variable's values, so a correct selection of values will directly affect on the HREMD simulation's accuracy and efficiency. Probability of exchange needs to be higher than 5% to be considered a good value for sampling. The exchange probability is defined as:

$$P(1 \leftrightarrow 2) = \min\left(1, \exp\left[\left(\frac{1}{k_B T} - \frac{1}{k_B T}\right)((U_1(x_2) - U_1(x_1) + (U_2(x_1) - U_2(x_2)))\right]\right) \tag{2.10}$$

where $U_i(x_i)$ is the Hamiltonian of a replica with its biased variable value (or biased Hamiltonian) and k_B stands for the Boltzmann constant. As Figure 4 shows, exchanges occur between neighbour replicas, i.e., between i and $i \pm 1$ replicas respectively. The

reason to this is that as the difference between biased variables values grows, the probability of exchange decreases rapidly [21].

In practice, in order to change variables and give them values of interest in GROMACS, `mdp` files have to be created and edited. The separate Hamiltonians are defined by the free energy functionality of GROMACS, with swaps made between the different values of λ, which is defined, as said, in the `mdp` file. λ indicates the extent to which the Hamiltonian has been perturbed and the system has been transformed. A good selection of λ values determines a reliable probability of exchange which is crucial for a successful HREMD simulation [20].

2.4 Analysis methods

MD simulations generate a big volume of data and they require an extensive analysis to be carried out. The GROMACS software package comes with many programs other than those required to carry out MD simulations. Occasionally, the MDTraj package was used as it is particularly user-friendly for the possibility of scripting analysis using the Python programming language [22].

2.4.1 Distance metrics

For many different purposes, such as the calculation of minimal distances with periodic images or estimating the convergence calculation we had to calculate Cartesian distances between atoms from the simulation trajectories. We used of MDTraj to calculate distances between residues and Zn^{2+}, to calculate pairs of distance between different residues and Zn^{2+} or to carry some exploratory data analysis. In addition to raw distances, we also use the root mean square deviation ($RMSD$). This metric is defined as the difference of any data-set compared to a reference value,

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - d_i(0))^2} \tag{2.11}$$

where n is the number of pairwise distances, d_i is the distance for the i-th pair at an instantaneous configuration and $d_i(0)$ is the distance for the same pair in the reference

values.

2.4.2 Free energy landscapes

To analyze the data obtained from a MD simulation, a large variety of analysis methods can be used, but statistical mechanics come in specially handy. Statistical mechanics offer tools to describe the behaviour of a system in a statistical and simple way. When studying any system, we usually are interested on equilibrium state behaviour. Boltzmann law describes the average distribution of non-interacting material particles over various energy states in thermal equilibrium, which is specially useful in our case.

$$p_i = \frac{e^{-\beta E_i}}{\int_j e^{-\beta E_j}} \tag{2.12}$$

where $e^{-\beta E_i}$ tells the relative probability of a particular arrangement on a given temperature.

When we collect data on a given order parameter, a distribution function can be obtained from histograms. Afterwords, by deriving the distribution function validating of equation 2.12, we can obtain the free energy of the system, which will be specially useful to characterize stable conformations.

2.4.3 Principal component analysis

Lastly, we used dimensionality reduction techniques, in particular the principal component analysis (PCA). In a few words, PCA is a dimension reduction method. Its goal is to extract the important information from a data-set, to represent it as a set of new orthogonal variables called principal components (PCs). Given a collection of data as point in various dimensions, a correlation line can be defined as the one that minimizes the average squared distance from a point to the line [23]. This way, we get a description of the system where a big percentage of the variance is explained by a lower number of dimensions. Even if a data loss needs to be assumed to gain the principal components, the method guarantees that the first few coordinates will contain most of the information from the full dataset.

Chapter 3

Results and Discussion

The system remains stable in short equilibrium runs

The models we are using to sample the Aβ peptide bound to Zn^{2+} are based on approximations and their aim is to replicate natural states and interactions. For this reason, we started validating the models to ensure their integrity. In particular, we were concerned about the stability of protein-metal ion interactions, which are described by a non-bonded interaction term (see Methods). After the equilibration process, we compared the obtained results to the initial structure. We calculate the root mean square deviation, $RMSD$ for the pairwise distances for the four amino acid residues coordinated to the Zn^{2+} with respect to the reference (i.e. 1ZE9) state, which we show in Figure 5A. In this case, we were interested in knowing whether the force field parameters would maintain the coordination of the Zn^{2+} cation. A low $RMSD$ value indicates that the reference distances and the distances throughout the simulations have similar values, whereas high $RMSD$ values indicate large displacements. This way, we were able to determine whether the coordination site was correctly simulated or not. In Figure 5C we show the $RMSD$, which clearly indicates that the distances remain close to the experimental range, as we obtained extraordinarily low $RMSD$ values.

Despite the overall consistency of the simulation result with the reference, we can inspect more closely the pairwise distances for all the residues coordinated to the metal (see

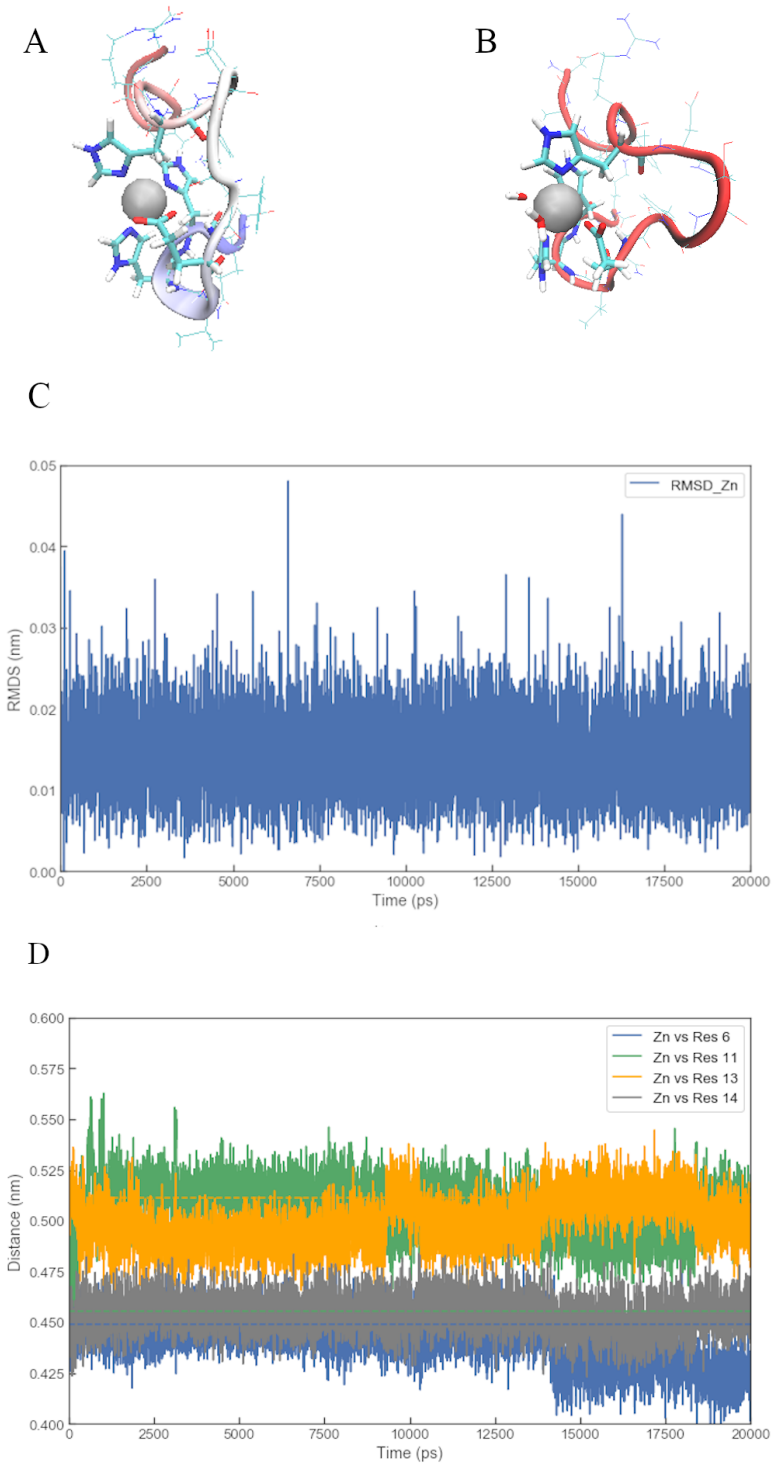

Figure 5: *Cartoon representations of the initial (A) and post-equilibration (B) states of the Aβ peptide coordinated to the Zn^{2+} cation. (C) RMSD values obtained throughout the equilibration simulation. (D) Time series data for the distances between residues of interest (His6, Glu11, His13 and His14 respectively) and Zn^{2+}. The reference distances are represented as dashed horizontal lines.*

Figure 5D). The distances are maintained close to the reference state, with the exception of Glu11. This is due to a change in the coordination sphere of the cation, which after the equilibration involves water molecules coordinating the Zn^{2+} (see Figure 5B). Hence, the metal goes from a tetrahedral to an octahedral coordination. This suggests that the Zn^{2+} model that C36m uses is not optimal. The model with which C36m describes the Zn^{2+} is a nonbonded Lennard-Jones potential. Even though it is widely used, it fails at describing some properties of metals such as solvent-ion radial distributions [24]. This is not surprising considering Zn^{2+} electron configuration is $[Ar]3d^{10}$ and d orbitals do not have a spherical shape. Despite the failure of the model in capturing the details of the coordination of Zn^{2+}, we find that the approximation is acceptable considering the agreement in the $RMSD$ and the simplicity in the approximations involved.

Perturbed Hamiltonians allow sampling bound states of Zn^{2+}

As mentioned in the Methods section, HREMD simulations are used to enhance the sampling. This requires the definition of biased Hamiltonians. In this case, we are aiming to sample different bound states of the metal, so our system should be able to visit the free state and let the Zn^{2+} diffuse around the protein to find new coordination sites. Taking into account that Zn^{2+}-protein coordinations are described as nonbonded interactions, we considered which interactions had to be switched off in the perturbed Hamiltonian. We tested three different possibilities, including both electrostatic and Van der Waals terms in the original force field, only the Van der Waals term, and none of the interactions (i.e. switching off both the electrostatics and the Van der Waals term).

With these three different models we ran short simulations, to see whether each of these options would allow reaching an unbound state of the system. After editing the parameters that define nonbonded interactions, which can be easily done on the simulation set-up in GROMACS, we ran short simulations following the procedures described in the Methods section. After a 10 ns NVT run, we obtained the results shown in Figure 6. With the original force field parameters, the Zn^{2+} remains in the original state, resulting in very low Zn^{2+}-backbone distances. We find that switching off only the electrostatics seems to be sufficient to reach unbound states. Hence, we decided to carry out HREMD simulations

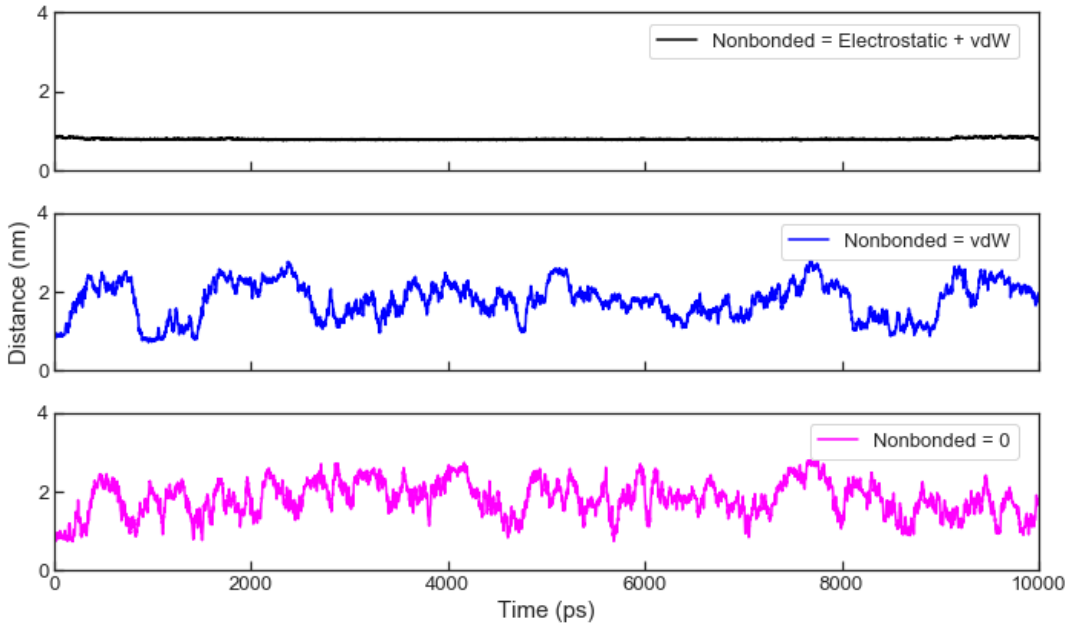

Figure 6: *Time series data for distances obtained between Zn^{2+} and protein backbone in the short simulations carried out with different hamiltonians for the Zn^{2+} cation. Top: original force field; center: van der Waals term only; bottom: no interactions.*

with a perturbed Hamiltonian including Van der Waals terms. The reason for choosing this model, instead of going all the way to switching off all the interactions, is that the smaller the perturbation the lower the number of replicas needed for a successful HREMD simulation.

Efficient HREMD simulations require many independent replicas

Once the model for the HREMD was selected, we needed to define the number of replicas in order to guarantee good sampling, which is determined by the probability of exchange (see Methods). GROMACS allowed us to create λ values from 0 to 1 (corresponding, respectively to the simulation run with the original force field and the perturbed Hamiltonian). In order to obtain a reasonably high probability of exchange with the minimum computational cost, we first run a test simulation using 11 replicas. In Figure 7A we show the histograms of the potential energy and their overlapping for each of the replicas. When running simulations with these conditions, the exchange probability for all replicas was below 5%, which would result in insufficient sampling. For this reason, we decided to run HREMD simulations using 16 replicas, resulting in a higher probability of exchange, as

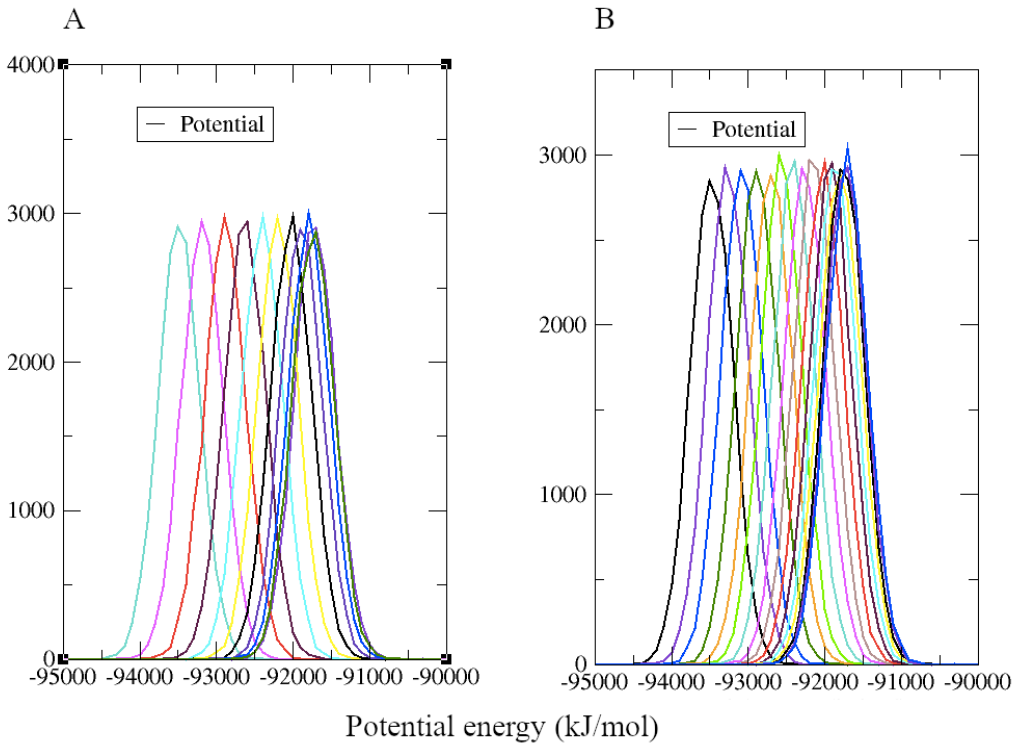

Figure 7: *Histograms of the potential energies for different numbers of replicas in HREMD runs: (A) 11 replicas and (B) 16 replicas.*

manifested in the greater overlap between energy histograms (see Figure 7B).

The binding energy landscape of Zn^{2+}-Aβ_{1-16}

Using the GPU implementation of GROMACS [25], we obtained a performance of over 45 ns/day, which is about three times higher than the obtained with CPUs, 16 ns/day. In this way, we could run 500 ns HREMD simulations for each of the 16 replicas. The first thing we noticed in the resulting simulation trajectories was that the selected box was ocassionally not big enough. Throughout the simulation the protein was seen interacting with its periodic image. This is a problem that often emerges in MD simulations of IDPs, as they sample extended conformations instead of adopting well defined, compact forms. To determine how much of the simulation data was affected by this unphysical artifact, we calculated the minimum distances between periodic images and the "original" protein chain (see Figure 8). We find that the distribution has a maximum at ~1.5 nm, but there is a small peak below 0.5 nm, corresponding to frames where periodic images are interacting

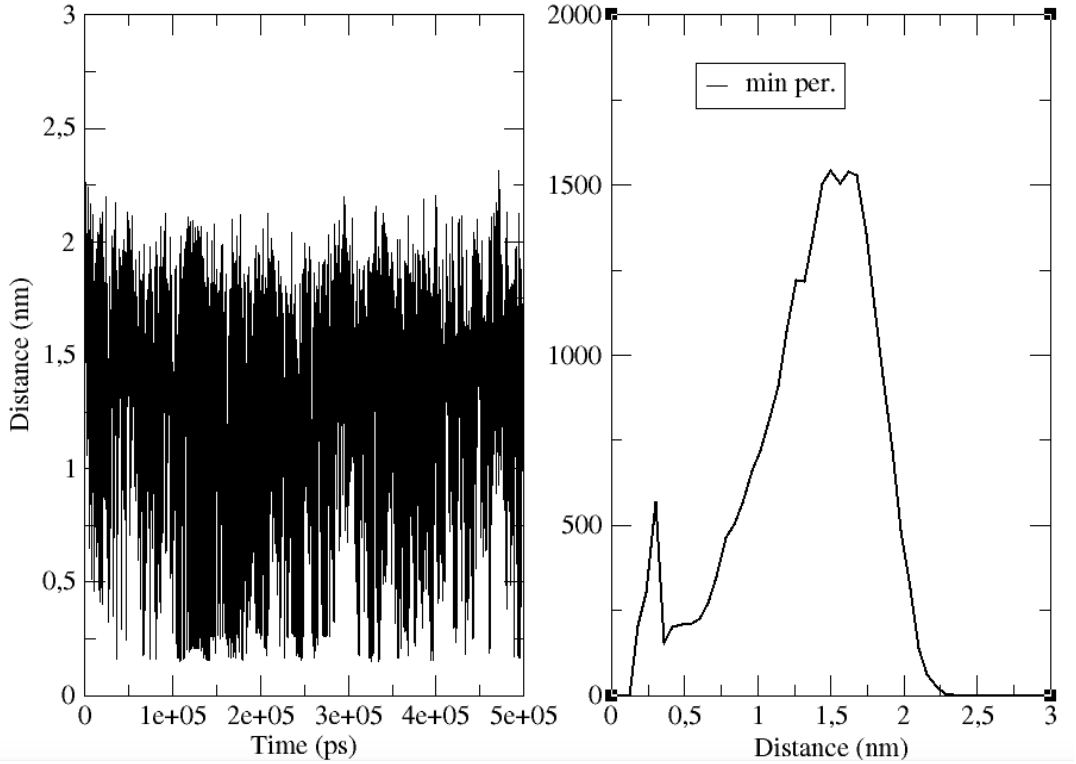

Figure 8: *Time series data (left) and histogram (right) for the minimum distances between Aβ and its periodic image throughout the HREMD simulation.*

with each other. Even though it is clear that we wanted to avoid periodic images from interacting with each other, the amount of data affected by that problem is low enough for them to be removed and still have a more than acceptable data set, avoiding a new HREMD simulation run. We deleted all frames that recorded a distance between periodic images lower than 1 nm losing a 20% of data in the process.

After post-processing the trajectory to eliminate these unphysical conformations, the amount of simulation data left is sufficient to identify different coordination states of the Zn^{2+} cation. This is enabled by the much more efficient sampling of stable states of the system. In Figure 9 we show the calculated $RMSD$ (see Methods) for the distances between residues of the amino acids coordinated initially to the Zn^{2+} and the Zn^{2+} itself. These results must be interpreted with caution, because HREMD simulation trajectories do not contain dynamical information with physical meaning, as the coordinates from different replicas are swapped periodically, they do not describe a realistic system behaviour. When looking at the $RMSD$ values sampled in the simulation, we find that a broad range of values have been explored. This suggests that we indeed were able to create unbound

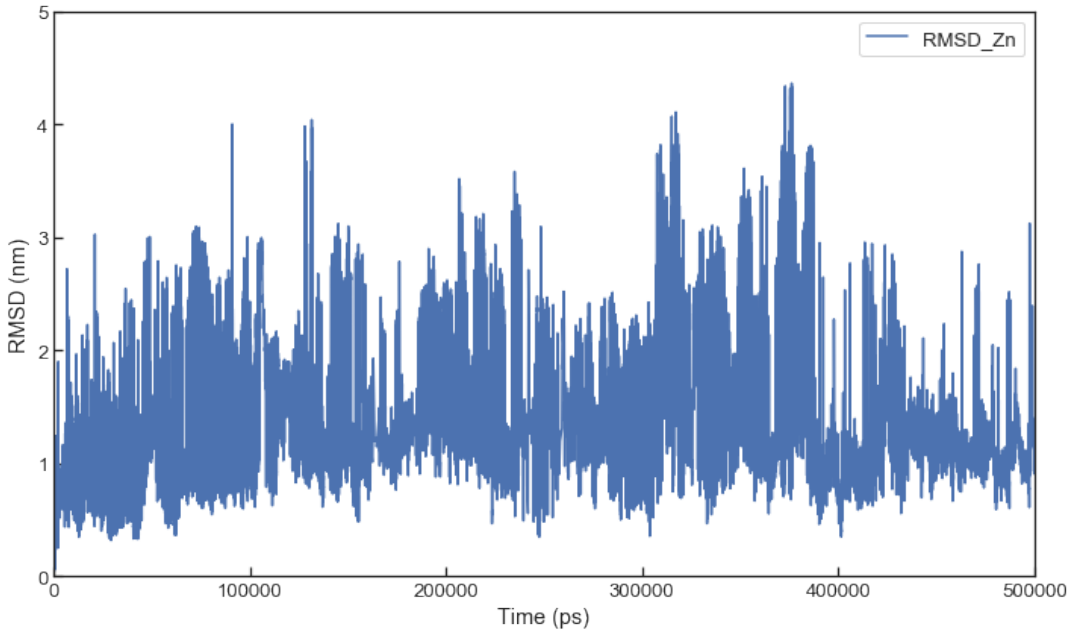

Figure 9: *RMSD of the Zn^{2+} and residues of His6, Glu11, His13 and His14 in the HREMD simulation.*

states of the Zn^{2+} throughout the HREMD simulation.

To recover a more detailed picture of the amino acids involved in interactions with the metal ion, a logical parameter to look at is the distances between Zn^{2+} and the residues that are coordinated to it, just as we did for the model validation (Figure 5D). This time though, we calculated distances of each residue, as each residue should be treated as a potential binding site for the Zn^{2+}. In Figure 10 we show the distributions of distances between each residue and the Zn^{2+}. Interestingly, Glu3, Asp7, Ser8 and Glu11 stand out significantly, with sharp narrow peaks at low distances, whereas His6, His13 and His14 do not show any low-range significant states. These results suggest that either one of the residues of reference (His6, Glu11, His13 and His14) but Glu11 show a tendency to be on a Zn^{2+} binding range. After carefully analyzing the trajectory, looking for bound states that corresponded to each residue, we found out that Ser8 never reached a bound state with Zn^{2+} and therefore Ser8 was discarded as a potential binding site for Zn^{2+} and was determined that its high-tendency to appear on a low-range of the Zn^{2+} is a consequence of the Asp7 being coordinated to the Zn^{2+}.

All of this information was represented on one dimension, so we decided to go one step

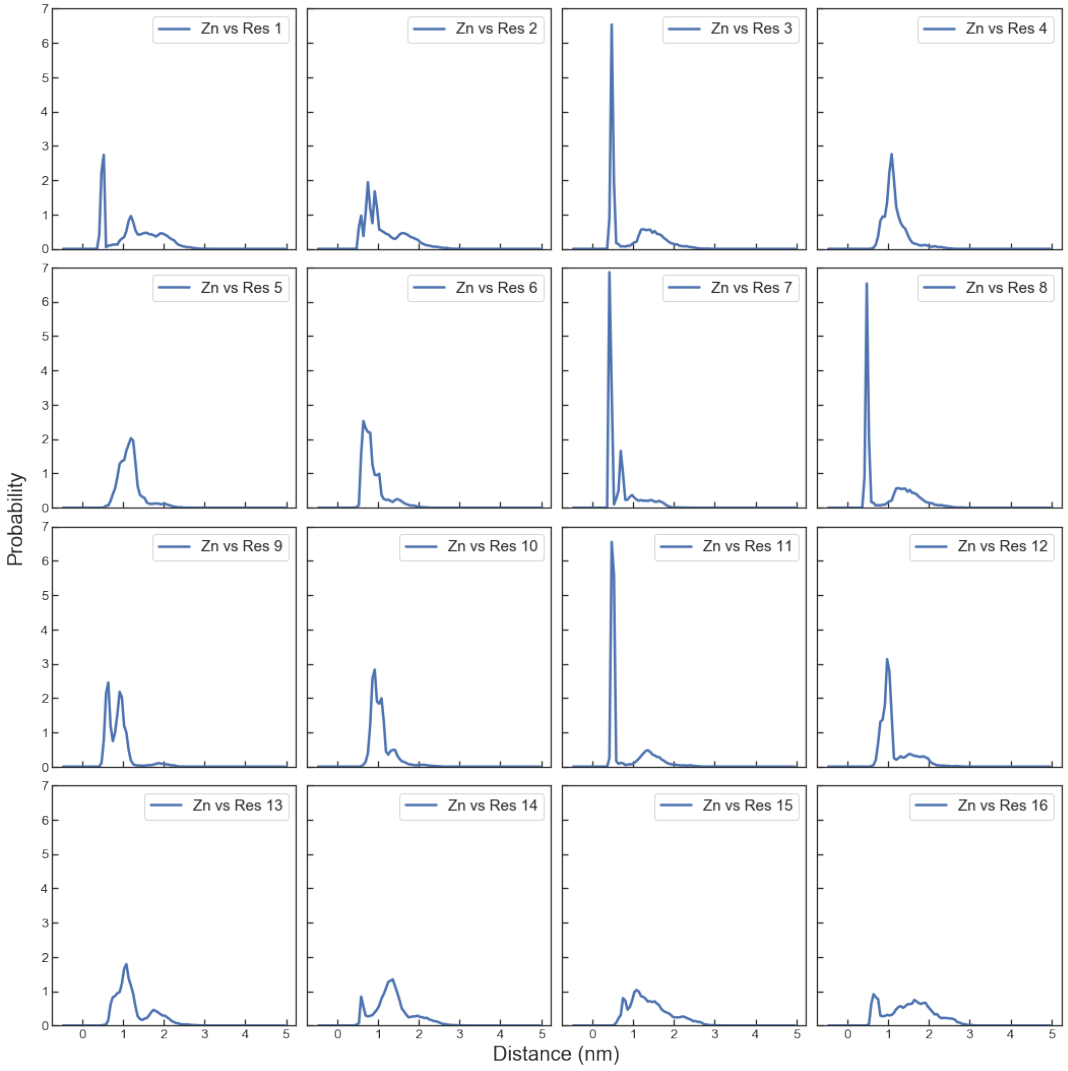

Figure 10: *Histograms of distances between each residue and* Zn^{2+}.

ahead and process the data to obtain two dimensional information, or contour plots. In these contour plots, we plotted a pair of distances between Zn^{2+} and residues of interest as two dimensional histograms. From these plots, we expected to be able to shed some light on the correlation of distances between residues and different coordination states. From the trajectory file and the results obtained in Figure 10, we selected as residues of interest the reference residues (i.e. residues seen coordinated on experimental data) and any residue that was seen coordinated to the Zn^{2+} throughout the simulation. The residues selected turned out to be: Asp1, Glu3, His6, Asp7, Glu11, His13, His14 and Lys16. We calculated contour plots for every pair of distance.

From Figure 11 we can highlight four distance pairs that show a high-tendency to show

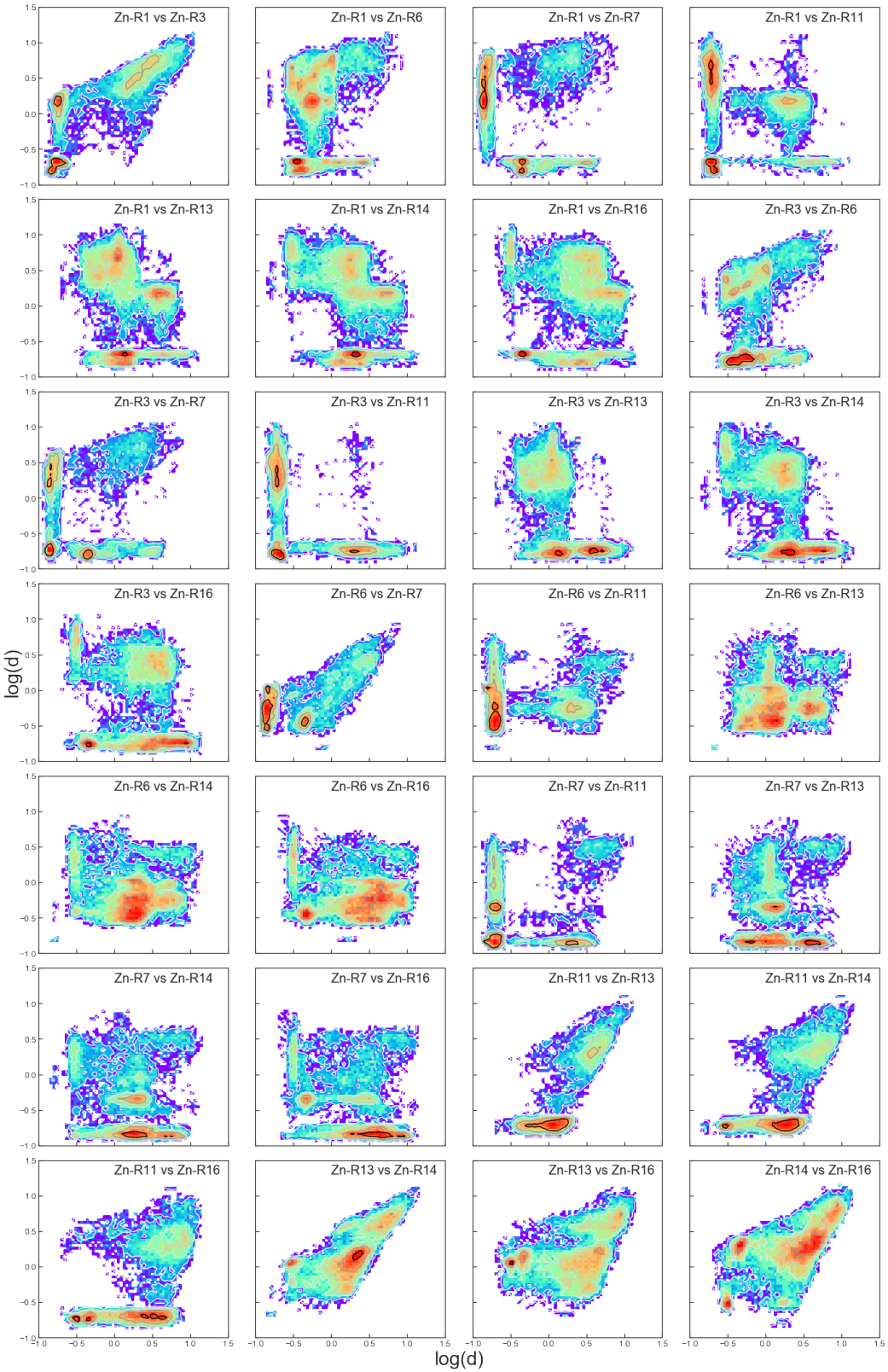

Figure 11: *Contour plots of all pair of distances of the residues of interest represented on a logarithmic scale.*

bound states at the same time through the high density of probability their contour plots show at low-range values. Those plots are the ones that show the correlation between Asp1 and Glu11; Glu3 and Asp7; Glu3 and Glu11; and Asp7 and Glu11. Among these, we could clearly identify that residue 7 and residue 11 are recurrent, but we were still unable to describe a correlation between them and the rest of residues of interest, and under no circumstances between the whole protein.

Dimensionality reduction captures correlations between residue pairs

From Figure 11 we can conclude that the information for the distance between individual residues and the Zn^{2+} cation is correlated. For this reason, the free energy for each individual residue is not particularly informative (see Figure 12A). Alternatively one can try to extract information using dimensionality reduction techniques, that will aggregate into a few variables the combined information of all the pairwise distances between amino acids and the cation. Specifically, we use the PCA method (see Methods). We find that just one principal component, PC1, can describe over the 70% of the system variance (see Figure 12B). We decided to keep this last free energy expression, as we found that the minimum values states were well represented (said minimum values are shown with vertical lines in Figure 12B). Knowing that minimal free energy values, as mentioned, represent the most stable states on a system, we found four different stable conformations.

These four conformations correspond to different coordinations: Asp7 and Glu11 (both with monodentate coordination) coordinated to Zn^{2+} octahedrally (Figure 12 C); Asp7 and Glu11 (both with monodentate coordination) coordinated to Zn^{2+} octahedrally (Figure 12 D); Glu11 and Lys16 (both with bidentate coordination) coordinated to Zn^{2+} octahedrally (Figure 12 E); and Asp1, Glu3 and Glu11 (all with monodentate coordination) coordinated to Zn^{2+} octahedrally (Figure 12 F). We also report that, as shown in Figures 12 C, D, E and F, all conformations contain water in the coordination sphere. Surprisingly, the reference state is never reached throughout the simulation.

After obtaining these coordinates, we could deduce just one thing: the force field we had chosen to carry the simulations might not be well parametrized as all of the stable states had come from acetate-Zn^{2+} coordination. We had a look at the sixteen amino

acid sequence: (1) Asp-Ala-Glu-Phe-Arg-His-Asp-Ser-Gly-Tyr-Glu-Val-His-His-Glu-Lys (16). Looking at the sequence carefully, on the one hand, we see that the only amino acids to show an acetate in their R group are aspartic acid (Asp), which correspond to Asp1 and Asp7, and glutamic acid (Glu), which correspond to Glu3 and Glu11. On the other hand, even if lysines do not have an acetate in their R groups, Lys16 was acetylated. This means that after the unbound state was reached, the Zn^{2+} exclusively coordinated to acetate groups. We presume that, as previously mentioned, there is a miscalculation on the parametrization of histidine's *N*s charges that if re-parametrized could success at describing the system more precisely. In order to confirm further this conclusion, QM calculations starting from structures shown in Figure 12 C, D, E and F conformations are going to be performed.

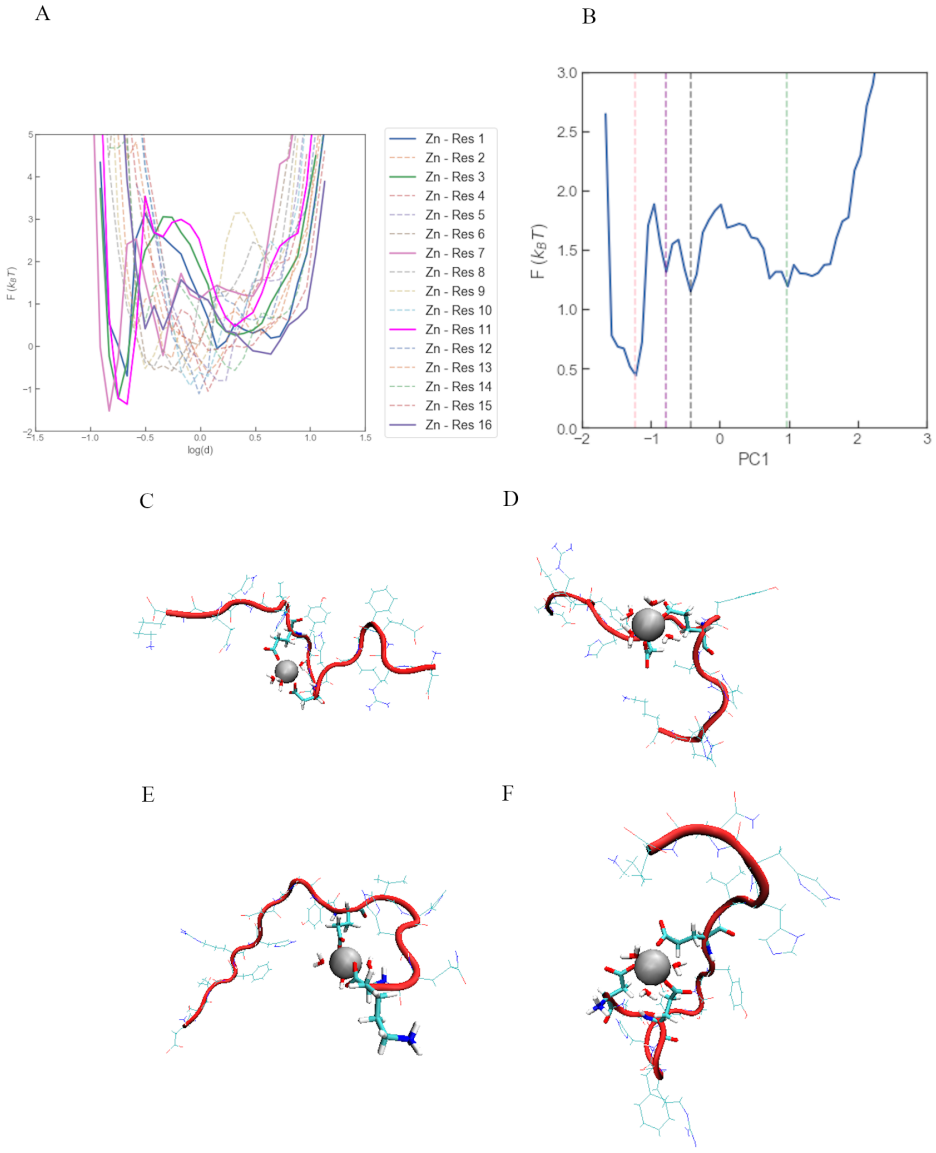

Figure 12: *Free energies for residues depending to their distance to Zn^{2+} (A) and the free energy obtained from PC1 (B). Different stable conformations obtained through the interpretation of PC1's free energy (C, D, E, F respectively).*

Chapter 4

Conclusions

- Classical molecular dynamics simulations with the Hamiltonian replica exchange technique can be used successfully to probe different conformational states of metal-protein complexes.

- Limitations in the interaction models for the Zn^{2+} cation can hamper our ability to reproduce the correct coordination. There is hence room for improving the results if we are able to better describe Zn^{2+} model in the future, for example with models such as the dummy models by Strodel et al. [26].

- The size of the simulation box must be carefully selected in order to avoid unphysical effects. In our case a bigger box would have prevented the sampling of states where the protein molecule was interacting with its periodic box.

- CHARMM36m force field's parametrization fails at representing histidine's N charges as we were unable to record a conformation where the Zn^{2+} ion was bonded to a imidazole N rather than an O from an acetyl group. We propose a further examination of the charges given to the problematic Ns and Os comparing with Quantum Mechanics calculations.

Ondorioak

- Dinamika molekular klasikoak eta *Hamiltonian replica exchange* teknika, metal-proteina konformazio ezberdinak aurkitzeko era arrakastatsuan erabili daitezke.
- Zn^{2+} katioi modeloen mugek koordinazio egoera zuzenak aurkitzeko gaitasuna oztopatzen dute. Hori dela eta, emaitzak hobetzeko aukera dago Zn^{2+} katioia hobeto deskribatzen duen modelo bat erabiliz gero, Strodel et. al-ek proposatzen dituzten *dummy model*-ak besteak beste [26].
- Simulazio kaxaren dimentsioak kontu handiz aukeratu behar dira esanahi fisikorik ez duten elkarrekintzak ekiditeko. Gure kasuan, tamaina handiagoko kaxa baten erabilerak proteina eta bere irudi periodikoaren arteko elkarrekintzak eragotziko lituzke.
- CHARMM36m *force field*-aren parametrizazioak akatsak ditu histidinetako N-en kargak egokitzerakoan. Izan ere, ezin izan dugu Zn^{2+} katioia imidazoletako N-etara koordinatuta ikusi, beti azetato taldeetako O-etara koordinatuta zegoelako. Mekanika kuantikoaz baliatuz, N eta O problematikoen kargen azterketa sakona proposatzen dugu.

Bibliography

[1] Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: Introducing the d2 concept. *Annual Review of Biophysics* **37**, 215–246 (2008).

[2] Dementia (2019). URL `https://www.who.int/health-topics/dementia`.

[3] Barnham, K. J. & Bush, A. I. Metals in alzheimer's and parkinson's diseases. *Current Opinion in Chemical Biology* **12**, 222 – 228 (2008). Biocatalysis and biotransformation/Bioinorganic chemistry.

[4] Wise-Scira, O., Xu, L., Perry, G. & Coskuner, O. Structures and free energy landscapes of aqueous zinc(ii)-bound amyloid-(1–40) and zinc(ii)-bound amyloid-(1–42) with dynamics. *JBIC Journal of Biological Inorganic Chemistry* **17**, 927–938 (2012).

[5] Hardy, J. & Selkoe, D. J. The amyloid hypothesis of alzheimer's disease: Progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).

[6] Chen, T. *et al.* Effects of cyclen and cyclam on zinc(ii)- and copper(ii)-induced amyloid -peptide aggregation and neurotoxicity. *Inorganic Chemistry* **48**, 5801–5809 (2009).

[7] Deibel, M., Ehmann, W. & Markesbery, W. Copper, iron, and zinc imbalances in severely degenerated brain regions in alzheimer's disease: possible relation to oxidative stress. *Journal of the Neurological Sciences* **143**, 137 – 142 (1996).

[8] Zirah, S. *et al.* Structural changes of region 1-16 of the alzheimer disease amyloid -peptide upon zinc binding andin vitroaging. *Journal of Biological Chemistry* **281**, 2151–2161 (2005).

[9] Hoyer, W., Grönwall, C., Jonsson, A., Ståhl, S. & Härd, T. Stabilization of a -hairpin in monomeric alzheimer's amyloid- peptide inhibits amyloid formation. *Proceedings of the National Academy of Sciences* **105**, 5099–5104 (2008).

[10] Miller, Y., Ma, B. & Nussinov, R. Zinc ions promote alzheimer aβ aggregation via population shift of polymorphic states. *Proceedings of the National Academy of Sciences* **107**, 9490–9495 (2010).

[11] Schlick, T. *Molecular Dynamics: Basics*, 425–461 (Springer New York, New York, NY, 2010). URL `https://doi.org/10.1007/978-1-4419-6351-2_13`.

[12] Aduri, R. *et al.* Amber force field parameters for the naturally occurring modified nucleosides in rna. *Journal of Chemical Theory and Computation* **3**, 1464–1475 (2007).

[13] Siu, S. W., Pluhackova, K. & Bockmann, R. A. Optimization of the opls-aa force field for long hydrocarbons. *Journal of Chemical Theory and Computation* **8**, 1459–1470 (2012).

[14] Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115**, E4758–E4766 (2018).

[15] Rauscher, S. *et al.* Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *Journal of Chemical Theory and Computation* **11**, 5513–5524 (2015).

[16] Huang, J. *et al.* Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71–73 (2017).

[17] Shabane, P. S., Izadi, S. & Onufriev, A. V. General purpose water model can improve atomistic simulations of intrinsically disordered proteins. *Journal of Chemical Theory and Computation* **15**, 2620–2634 (2019).

[18] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).

[19] Van Der Spoel, D. *et al.* Gromacs: fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701–1718 (2005).

[20] Lemkul, J. From proteins to perturbed hamiltonians: A suite of tutorials for the gromacs-2018 molecular simulation package [article v1. 0]. *Living Journal of Computational Molecular Science* **1**, 5068 (2018).

[21] Bussi, G. Hamiltonian replica exchange in gromacs: a flexible implementation. *Molecular Physics* **112**, 379–384 (2014).

[22] McGibbon, R. T. *et al.* Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal* **109**, 1528 – 1532 (2015).

[23] Abdi, H. & Williams, L. J. Principal component analysis. *WIREs Computational Statistics* **2**, 433–459 (2010).

[24] MacDermott-Opeskin, H., McDevitt, C. A. & O'Mara, M. L. Comparing nonbonded metal ion models in the divalent cation binding protein psaa. *Journal of Chemical Theory and Computation* **16**, 1913–1923 (2020).

[25] Lindahl, Abraham, Hess & van der Spoel. Gromacs 2020 source code (2020).

[26] Strodel, B. & Coskuner-Weber, O. Transition metal ion interactions with disordered amyloid-β peptides in the pathogenesis of alzheimer's disease: Insights from computational chemistry studies. *Journal of Chemical Information and Modeling* **59**, 1782–1805 (2019).