Universidad Zaragoza
1542

Trabajo Fin de Máster

Refinando el análisis de estabilidad y cinético del desplegamiento de proteínas a partir de simulaciones de dinámica molecular

Fine Tuning Protein Stability and Unfolding Kinetics Analyses from Molecular Dynamics Simulations

Autor

Daniel Eduardo López Martínez

Director/es

Juan José Galano Frutos
Javier Sancho Sanz

Máster Universitario en Biotecnología Cuantitativa

Facultad de Ciencias

2021

ABSTRACT

In the present study, a method for detecting protein unfolding events and evaluating protein stability and kinetics from relaxation Molecular Dynamics (rMD) simulations has been fine tuned. In particular, a 2D-root mean square deviation (RMSD)-based clustering was performed on trajectories generated by rMD simulations for the protein α_3D under a broad range of temperatures and two different force fields: CHARMM27 and AMBER99SB-disp. The half-life of unfolding was measured from the clustering plots and its value was used to compute the conformational stability of the protein extrapolated at 298 K by means of an empirical model previously developed by the research group. Ladder- and ramp-based temperature scanning simulations were also explored to assess the optimum simulation approach in terms of computational resources, accuracy, and adequacy for the kind of protein stability and unfolding kinetics analyses here envisioned. Furthermore, the performance (accuracy and reliability) of the above-mentioned force fields was assessed in estimating protein stability.

KEY WORDS

Relaxation Molecular Dynamics simulations; Conformational stability; Unfolding kinetics; Half-life of unfolding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

The conformational stability of a protein refers to the free energy difference between the folded and unfolded protein conformations, in a population of folded and unfolded protein molecules that are in equilibrium (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001; Sancho, 2013). In this context, studying conformational stability can provide valuable insights about protein structure and function (Sancho, 2013; Tokuriki & Tawfik, 2009). To do so, protein stability can be determined experimentally by using spectroscopic, calorimetric techniques or Nuclear Magnetic Resonance (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001; Sancho, 2013; Tayyab, Siddiqui, & Ahmad, 1995); or it can be estimated computationally through Molecular Dynamics (MD) simulations (Galano-Frutos & Sancho, 2019; Sancho, 2013).

Molecular Dynamics is a computational technique in which biomolecules are simulated as multi-particle mechanical systems subject to the Newtonian laws of motion within a solvent box (Kumari & Akhter, 2017). As a technique performed in virtual environments, MD simulations are time and cost effective as compared with the traditional approach of carrying out experimentally the gene expression, the protein purification and activity assays (Kumari & Akhter, 2017). So far, MD simulation has been regarded as a reliable technique capable of providing precise results in studies of protein conformation and stability, thanks to the accuracy and development achieved by current force fields and the growth in computing power (Galano-Frutos, García-Cebollada, & Sancho, 2019; Hospital, Goñi, Orozco, & Gelpí, 2015).

In particular, MD simulations have been used to investigate the unfolding process of proteins through thermal and chemical denaturation (Daggett & Levitt, 1993; Day, Bennion, Ham, & Daggett, 2002; Rocco, et al., 2008). In these studies simulations normally proceed from a well-defined starting structure (the folded state) where the protein is surveyed over a period of simulation time in which conformational transitions take place until the structure is "completely" disrupted (the unfolded state) (Daggett & Levitt, 1993).

To describe the unfolding kinetics for a given protein, some parameters are used such as the unfolding rate constant and the half-life of unfolding (Tang & Pikal, 2005). In this context, an empirical model (fitted from both experimental and *in silico* data) for the calculation of conformational stability of proteins has been developed, which links the conformational stability of a protein at 298 K to the half-life of unfolding (Galano-Frutos, García-Cebollada, & Sancho, 2019):

$$\Delta G_U = 3.97 + 0.74 * \log(\tau_{1/2}) \quad (1)$$

In particular, the former empiric equation for estimating ΔG_U (in kcal/mol) through $\tau_{1/2}$ (in seconds) has been determined from the folding and unfolding rate constants of 89 two-state proteins normalized at 25.0 °C (Galano-Frutos, García-Cebollada, & Sancho, 2019). The linear fitting is shown in the Figure 1.

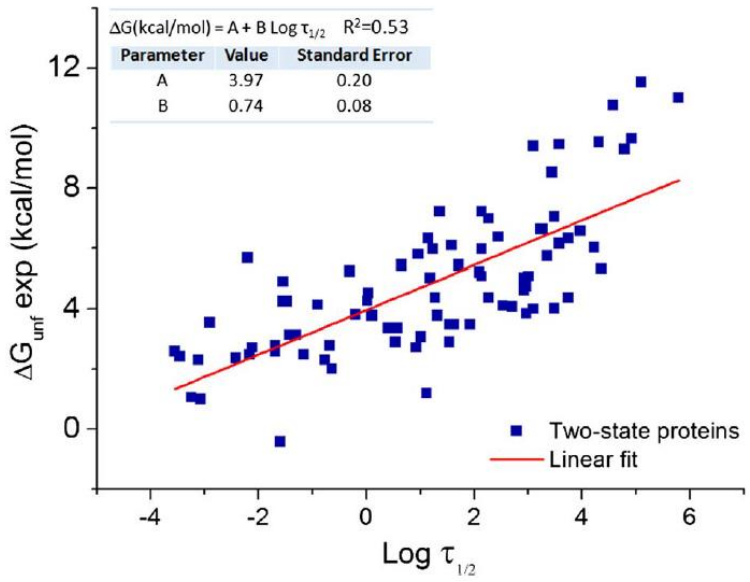

Figure 1. Correlation plot between protein conformational stability and the logarithm of unfolding half-life (Galano-Frutos, García-Cebollada, & Sancho, 2019).

It is known that the time required (τ_{x_U}) to obtain a given fraction (x_U) of unfolded molecules starting from a population of fully folded proteins can be described as:

$$\tau_{x_U} = -\tau_{1/2} * \ln(1 - x_U)/\ln(2)$$

$$\therefore \tau_{1/2} = -\frac{\tau_{x_U} * \ln(2)}{\ln(1 - x_U)} \quad (2)$$

Moreover, $\tau_{1/2}$ can also be calculated according to the rate law for a first-order reaction (Wittung-Stafshede, 2004):

$$\tau_{1/2} = \frac{\ln(2)}{k_U} \quad (3)$$

So, by combining (2) and (3):

$$-\frac{\tau_{x_U} * \ln(2)}{\ln(1 - x_U)} = \frac{\ln(2)}{k_U}$$

$$\therefore k_U = -\frac{\ln(1 - x_U)}{\tau_{x_U}} \quad (4)$$

On the other hand, from the Eyring equation we can extrapolate the unfolding kinetic constant of a protein that has been obtained at a temperature different from 298 K to this referential temperature (Galano-Frutos, García-Cebollada, & Sancho, 2019):

$$\frac{k_U^{298}}{k_U} = \frac{298\ K}{T} * e^{\left(\frac{\Delta G^{\ddagger}}{R}\right)\left(\frac{1}{T} - \frac{1}{298\ K}\right)}$$

$$\therefore k_U^{298} = k_U * \frac{298\ K}{T} * e^{\left(\frac{\Delta G^{\ddagger}}{R}\right)\left(\frac{1}{T} - \frac{1}{298\ K}\right)} \quad (5)$$

In Galano-Frutos et.al. (2019), $\Delta G^{\ddagger}$ was estimated by analyzing the experimental data of 11 proteins from 48 to 118 amino acid residues; obtaining an average activation free energy of unfolding of 19.71 ± 0.87 kcal/mol.

Then, substituting (4) in (5):

$$k_U^{298} = -\frac{\ln(1 - x_U)}{\tau_{x_U}} * \frac{298\ K}{T} * e^{\left(\frac{\Delta G^{\ddagger}}{R}\right)\left(\frac{1}{T} - \frac{1}{298\ K}\right)} \quad (6)$$

And (6) in (3):

$$\tau_{1/2} = -\frac{\ln(2)}{\frac{\ln(1 - x_U)}{\tau_{x_U}} * \frac{298\ K}{T} * e^{\left(\frac{\Delta G^{\ddagger}}{R}\right)\left(\frac{1}{T} - \frac{1}{298\ K}\right)}} \quad (7)$$

Finally, substituting (7) in (1):

$$\therefore \Delta G_U = 3.97 + 0.74 \times \log\left(-\frac{\ln(2)}{\frac{\ln(1 - x_U)}{\tau_{x_U}} * \frac{298\ K}{T} * e^{\left(\frac{\Delta G^{\ddagger}}{R}\right)\left(\frac{1}{T} - \frac{1}{298\ K}\right)}} \right) \quad (8)$$

In this way, by means of Equation (8), if we are able to measure the half-life of an unfolded protein fraction at a given temperature, we may compute the conformational stability of the protein at 298 K.

This semi-empirical model, however, may have the limitation of being dependent of the simulation temperature on the one hand, and on the other of being fitted with an empirical $\Delta G^{\ddagger}$ average which was obtained from a limited set of proteins and on which we cannot say anything about its matching with the *in silico* value. In this work, thus, we will explore the *in silico* temperature dependence in the estimation of proteins half-live, in order to optimize or reformulate, if it is the case, our original model for computing the conformational stability of proteins. At the same time, we propose a couple a

MD-based temperature scanning “experiments” that may reduce the cost of the analysis in terms of computational resources and/or even lead to more accurate results in the estimation of protein kinetics.

1.2 Objectives

1.2.1 General objective

To fine tune a MD-based method for evaluating protein stability and kinetics.

1.2.2 Specific objectives

The specific objectives are:

- To check the *in silico* temperature dependence in estimating half-life through relaxation MD simulations on a model protein.
- To conduct ladder- and ramp-based temperature scanning MD simulations to assess their accuracy and precision in estimating unfolding kinetics (half-life and *in silico* mid-denaturation temperature) of proteins.
- To explore the differences obtained in the results when a special-purpose force field like AMBER99SB-disp (designed both for structured and unstructured proteins) is used in comparison with what is obtained with CHARMM27.

1.3 Hypothesis

The most appropriate developed force field for structured and unstructured proteins (AMBER99-disp) affects the accuracy and reliability of the stability and kinetic analyzes.

2. THEORETICAL FRAMEWORK

2.1 Protein stability and unfolding kinetics

Proteins are polypeptides with a defined amino acid sequence that perform most the biological functions in living beings (Mathews, van Holde, & Ahern, 2002). In this sense, "protein molecules embody a remarkable relationship between structure and function at the molecular level" (Dill & MacCallum, 2012). As such, the ability of a protein to carry out its biological tasks depends on its tridimensional structure and the stability thereof (Sancho, 2013).

Even though the term "protein stability" may associate to phenomena such as chemical stability or kinetic stability, conformational or thermodynamic stability refers to the free energy difference between the fully folded (F) and fully unfolded (U) protein conformations in a population of folded and unfolded identical protein molecules that are in equilibrium (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001; Sancho, 2013).

The simplest scheme of transition between the F and U states is given by the two-state model (Castillo-Cano, 2012), which is often followed by small proteins (Sancho, 2013):

$$F \leftrightarrow U$$

When this process happens under ambient conditions, such as room temperature and neutral pH, the free energy differences of the chemical reaction can be estimated as follows (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001):

$$\Delta G_U = G_U - G_F$$

$$\Delta G_F = G_F - G_U$$

where ΔG_U is free-energy change of the unfolding reaction; ΔG_F, free-energy change of the folding reaction; G_U, free-energy of the unfolded state; and G_F, free-energy of the folded state.

It is considered that the major factors promoting protein conformational stability are van der Waals forces, hydrophobic interactions, hydrogen bonds, disulphide bonds and histidine ionization (Day A. , 1995; Deller, Kong, & Rupp, 2016; Huyghues-Despointes, Pace, Englander, & Scholtz, 2001). For instance, the burial of hydrophobic moieties and the formation of intramolecular H-bonds are processes that contributes significantly into the stabilization of a folded protein (Day A. , 1995; Huyghues-Despointes, Pace, Englander, & Scholtz, 2001). Whereas main destabilization factors are conformational entropy and unfavorable burial of peptide and polar groups (Day A. , 1995; Huyghues-Despointes, Pace, Englander, & Scholtz, 2001).

In this sense, the conformational stability of a protein is given by the net protein stability between the stabilizing (G_F) and destabilizing (G_U) interactions within the protein, and among the latter and the solvent (Day A. , 1995; Deller, Kong, & Rupp, 2016; Galano-Frutos & Sancho, 2019). A representation of this phenomenon can be visualized in Figure 2.

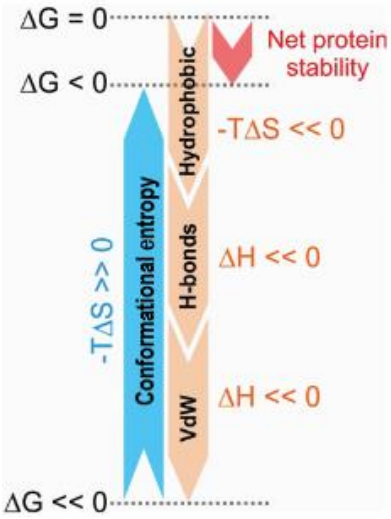

Figure 2. Cancellation of stabilizing and destabilizing interactions leading to a marginal net protein stability. Adapted from Deller, Kong, & Rupp (2016).

As shown in Figure 1, the stability of the folded conformation of a protein is a delicate balance between compensating forces. Thus, the overall free energy of a folded state is given by the small difference between two large numbers (the bulk of stabilizing and destabilizing interactions) that yields a marginal net protein stability, generally of about 5-15 kcal/mol under native conditions (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001).

This phenomenon, historically, has hindered the quantitative computational calculation of protein stability (Day A. , 1995) as highly accurate G values for both the folded and the unfolded states are required. Only in the recent years, more refined computational approaches have been developed and applied successfully for quantitatively calculating the folding energetics of model proteins within the experimental error (Galano-Frutos & Sancho, 2019).

Therefore, as a thermodynamic property, protein conformational stability can be defined as the difference in Gibbs free energy upon folding (ΔG_F) and calculated by using the Lewis equation (Sánchez-Ruiz, 2010; Sancho, 2013):

$$\Delta G_F = -R\,T\,\ln\,(K_F)$$

where R is the gas constant (0.001987 kcal/K*mol), T is the absolute temperature, and K_F is the equilibrium constant that models the fraction of molecules that are folded (x_F) relative to those unfolded (x_U) (Sancho, 2013).

Alternatively, the conformational stability is also described and reported in the literature as the free energy of unfolding (ΔG_U) (Galano-Frutos, García-Cebollada, & Sancho, 2019; Tayyab, Siddiqui, & Ahmad, 1995), and can also be evaluated using the relationship (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001; Saboury & Moosavi-Movahedi, 1995):

$$\Delta G_U = -R\,T\,\ln\,(K_U)$$

where K_U is the equilibrium constant between the fraction of molecules that are unfolded (x_U) relative to those folded (x_F) (Tayyab, Siddiqui, & Ahmad, 1995).

In this last case, ΔG_U, or also called free energy of stabilization, is the energy required to transform a population of folded proteins in solution into a population of completely unfolded proteins, and its value is positive for stable proteins (Sancho, 2013).

In this context, K_F and K_U can be calculated as a function of the fraction of molecules that are folded (x_F) and unfolded (x_U) in a given population of proteins as follows (Sancho, 2013; Saboury & Moosavi-Movahedi, 1995):

$$K_F = \frac{x_F}{x_U} = \frac{x_F}{1 - x_F}, \qquad x_U + x_F = 1$$

$$K_U = \frac{x_U}{x_F} = \frac{x_U}{1 - x_U}, \qquad x_U + x_F = 1$$

On the other hand, protein unfolding can be understood as the process in which a folded protein loses its native conformation to become an unfolded, structureless polypeptide chain (Tayyab, Siddiqui, & Ahmad, 1995). This unfolded state being characterized by many different conformations with similar free energies (Piana, Klepeis, & Shaw, 2014).

In this way, kinetic stability, while formally defined as the free-energy barrier separating the folded state from the unfolded state (Sánchez-Ruiz, 2010), it can be understood as a measure of how rapidly a protein unfolds (Day A. , 1995).

So, for a given reaction of folding/unfolding of a protein following the two-state model, if the transition is reversible, the equilibrium can be written as follows (Saboury & Moosavi-Movahedi, 1995):

$$F \underset{k_F}{\overset{k_U}{\rightleftharpoons}} U$$

where k_U and k_F are the rates of unfolding and folding, respectively.

In this case, ΔG_U is not relevant but the free energy difference between the folded and the transition state, also known as activation free energy ($\Delta G^{\ddagger}$), as the magnitude of this difference is the

one that determines the rate of unfolding (Day A. , 1995), as shown in the Eyring equation, assuming transition-state theory (Bilsel & Matthews, 2000; Sánchez-Ruiz, 2010):

$$k_U = \frac{k_B\, T}{h} * e^{-\left(\frac{\Delta G^{\ddagger}}{RT}\right)}$$

where k_u is the kinetic constant of protein unfolding; k_B, Boltzmann's constant; h, Planck's constant; T, temperature; $\Delta G^{\ddagger}$, free energy of activation; and R, ideal gas constant.

Graphically, it is possible to represent both conformational and kinetic stability as a funnel-shaped energy landscape with many high-energy unfolded structures and only a few low-energy folded proteins (Figure 3).

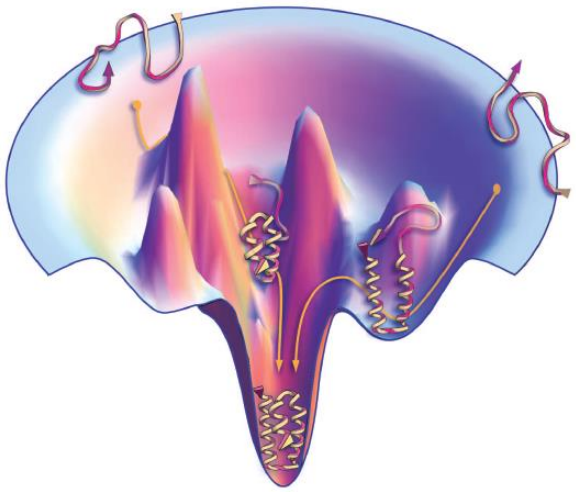

Figure 3. Energy landscape of protein molecules and conformational stability (Dill & MacCallum, 2012).

As shown in Figure 2, folded conformations are at the bottom of the energy landscape and have a narrower space in comparison with the broader space for unfolded structures (Dill & MacCallum, 2012). Thus, the conformational stability of a certain protein can be quantified by measuring ΔG_U, the free energy difference between the folded and unfolded states; whereas the kinetic stability is estimated by measuring $\Delta G^{\ddagger}$, the free energy difference between the folded and the transition states (Day A. , 1995; Sánchez-Ruiz, 2010; Sancho, 2013).

In particular, it has been written in the literature that the unfolding rate (k_U) depends on topology of the backbone, the oligomerization state, the presence of bound ligands, local sequence-structure propensities and the protein environment (Ramakrishnan, et al., 2012).

On the other hand, k_U can be used to estimate the half-life of unfolding ($\tau_{1/2}$) of a given protein, which is defined as the time it takes for half the molecules to unfold according to the rate law for a first-order reaction (Wittung-Stafshede, 2004):

$$\tau_{1/2} = \frac{\ln(2)}{k_U}$$

In this order of ideas, it is possible to relate the equilibrium constants K_F and K_U to the rates of unfolding and folding, the concentration of the states and the population fractions as follows (Castillo-Cano, 2012):

$$K_F = \frac{[F]}{[U]} = \frac{x_F}{x_U} = \frac{k_F}{k_U}$$

$$K_U = \frac{[U]}{[F]} = \frac{x_U}{x_F} = \frac{k_U}{k_F}$$

Thus, the protein conformational stabilities could be calculated through the derived expressions (Galano-Frutos, García-Cebollada, & Sancho, 2019; Wittung-Stafshede, 2004):

$$\Delta G_F = -R\,T \ln\left(\frac{k_F}{k_U}\right)$$

$$\Delta G_U = -R\,T \ln\left(\frac{k_U}{k_F}\right)$$

So, as the conformational stability of a protein is the net effect of the folding and unfolding rates, an increased protein stability will be reflected in the kinetics as a slower unfolding (k_U), a faster folding (k_F) or a combination of both (Wittung-Stafshede, 2004).

From a dynamical point of view, equations for the rate that govern the folding and unfolding processes can be written as follows (Zwanzig, 1997):

$$\frac{dx_F}{dt} = k_F\,x_U - k_U\,x_F$$

$$\frac{dx_U}{dt} = k_U\,x_F - k_F\,x_U$$

The solution of this set of equations reads as follows:

$$x_F(t) = \frac{k_F}{k_F + k_U} + \left(x_F(0) - \frac{k_F}{k_F + k_U}\right) * e^{-(k_F + k_U)\,t}$$

$$x_U(t) = \frac{k_U}{k_F + k_U} + \left(x_U(0) - \frac{k_U}{k_F + k_U}\right) * e^{-(k_F + k_U)\,t}$$

The case of three-state and more-state proteins is discussed in the literature (Sancho, 2013).

Another key concept in the quantification of the stability of proteins is the mid-denaturation temperature (T_m) that is defined as the temperature at which the concentration of the protein in its folded state equals the concentration of the protein in its unfolded protein, i.e., the temperature at which 50% of a protein population is unfolded (Ku, et al., 2009; Miotto, et al., 2019).

Finally, conformational and kinetic stability can be determined experimentally by using spectroscopic, calorimetric techniques or NMR (Huyghues-Despointes, Pace, Englander, & Scholtz, 2001; Sancho, 2013; Tayyab, Siddiqui, & Ahmad, 1995), or it could be estimated *in silico* through Molecular Dynamics simulations (Galano-Frutos & Sancho, 2019; Sancho, 2013).

2.2 Molecular Dynamics simulations

Molecular Dynamics (MD) simulations, first developed at the end of the 70s (McCammon, Gelin, & Karplus, 1977), are a computational technique that allows the simulation of biomolecules as multi-particle mechanical systems (Kumari & Akhter, 2017) by using numerical methods (Vlachakis, Bencurova, Papangelopoulos, & Kossida, 2014). This kind of simulations has evolved into a mature technique that can be used effectively to simulate ensembles of 50,000–500,000 atoms (Hospital, Goñi, Orozco, & Gelpí, 2015). MD simulations can be applied to protein systems, lipid bilayer membranes, carbohydrates and nucleic acids (Kumari & Akhter, 2017); and, nowadays, MD is also employed to describe protein properties and behavior, drug-receptor interactions, the solvation of molecules, and the conformational changes that a protein or molecule may undergo under various conditions (Vlachakis, Bencurova, Papangelopoulos, & Kossida, 2014).

As such, MD is based on the movement of biomolecules in a solvent box, wherein the movement is modeled through the Newtonian laws of motion of classical Physics. Starting structures in MD simulations requires either experimentally solved structural data or comparative modeling data (Hospital, Goñi, Orozco, & Gelpí, 2015). The details of the simulated systems can be defined at different extents; nonetheless, very often, an all-atom representation along with an explicit representation of the solvent molecules will yield more accurate results (Galano-Frutos & Sancho, 2019), at the expense of using more computational resources (Hospital, Goñi, Orozco, & Gelpí, 2015). Notwithstanding the above, the rougher coarse-grained representations are also useful for very large systems or long simulations (Hospital, Goñi, Orozco, & Gelpí, 2015).

Forces acting on every atom in MD simulations are obtained by applying the so-called force fields, which are a set of equations (also called potential or energy functions) that derive the potential energy of the system based on the molecular structure thereof (Hospital, Goñi, Orozco, & Gelpí, 2015) and parametrized interactions (Lindahl, Abraham, Hess, & van der Spoel, 2020).

As such, a force field represents two groups of molecular properties: (1) bonded interactions, which characterize stretching of bonds, bending of valence angles, and rotation of dihedrals; and (2) non-bonded interactions that evaluate electrostatic interactions, Pauli exclusion and dispersion, whose addition yields the total potential energy of the system (Vlachakis, Bencurova, Papangelopoulos, & Kossida, 2014):

$$E_{potential} = E_{bonded} + E_{nonbonded}$$

For instance, the energy function for the CHARMM force field has the following form (MacKerell, et al., 1998):

$$E_{potential} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{UB} K_{UB}(S - S_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \gamma)] + \sum_{impropers} K_{imp}(\varphi - \varphi_0)^2 + \sum_{nonbond} \epsilon\left[\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon_1 r_{ij}}$$

where $E_{potential}$ is the total potential energy of the system; K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, χ, and φ are the bond length, Urey-Bradley 1,3-distance, bond planar angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms; ϵ, Lennard-Jones well depth; $R_{min_{ij}}$, distance at the Lennard-Jones minimum; q_i, partial atomic charge on atom i; ε_1, effective dielectric constant; and r_{ij}, the distance between atoms i and j.

On the other hand, the energy function for the AMBER force field has the following form (Cornell, et al., 1995):

$$E_{potential} = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{K_\chi}{2}[1 + \cos(n\chi - \gamma)] + \sum_{i<j} \epsilon\left[\frac{A_{ij}}{{r_{ij}}^{12}} - \frac{B_{ij}}{{r_{ij}}^{6}} + \frac{q_i q_j}{\epsilon_1 r_{ij}}\right]$$

As seen above, force fields are complex equations that encompass each intramolecular contribution (bond stretching, angle bending and dihedral and improper torsions of bond rotations), repulsive interactions (van der Waals), as well as electrostatic interactions modeled through Coulomb's law (Pinak, 2006; Hospital, Goñi, Orozco, & Gelpí, 2015; Kumari & Akhter, 2017). They are approximations of the reality and can differ in accuracy and realism.

Once the forces acting on each atom are estimated, classical Newton's law of motion is applied to the system in order to compute accelerations and velocities and to update the atom coordinates/positions (Hospital, Goñi, Orozco, & Gelpí, 2015). This way, the spatial distribution of biomolecules' atoms defined by these force fields allows to generate a new set of coordinates thereof in terms of energy of the system in an iterative process (Kumari & Akhter, 2017).

The integration of the atoms' movement is performed through numerical methods. So, a time step shorter than the fastest movements in the biomolecule should be used to avoid numerical instability, which is about 1 and 2 fs for full atomistic simulations (Hospital, Goñi, Orozco, & Gelpí, 2015).

In summary, the basic algorithm for a MD simulation is shown in Figure 4.

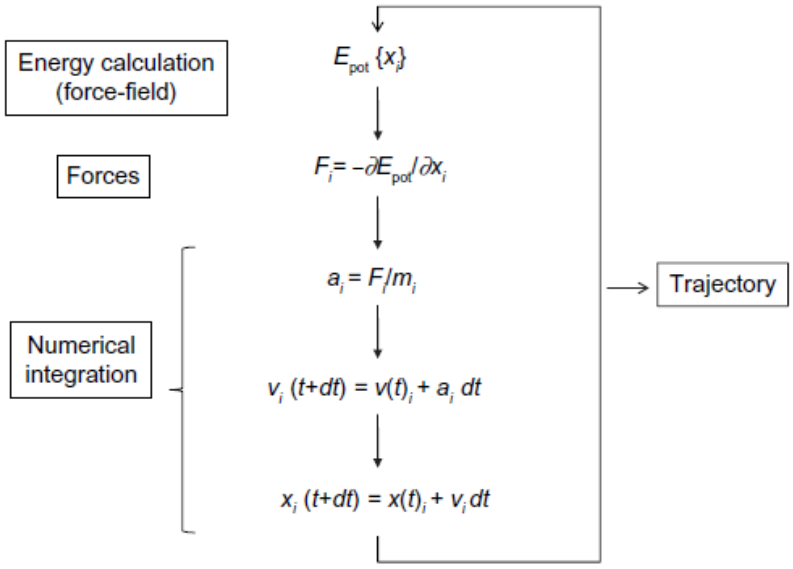

Figure 4. Basic algorithm for a MD simulation, where E_{pot} is potential energy; *t*, simulation time; *dt*, iteration time; *x*, atom coordinate; *F*, forces component; *a*, acceleration; *m*, atom mass; and *v*, velocity (Hospital, Goñi, Orozco, & Gelpí, 2015).

Nowadays, some of the most used codes for performing MD simulations are AMBER, CHARMM, GROMACS, and NAMD (Hospital, Goñi, Orozco, & Gelpí, 2015); while the most popular force fields for the simulation of biomolecules are CHARMM, AMBER, and GROMOS (Vlachakis, Bencurova, Papangelopoulos, & Kossida, 2014).

In MD simulations, it is noteworthy that statistical mechanics ensembles such as microcanonical (constant NVE), canonical (constant NVT), grand canonical (constant μVT) and isobaric-isothermal (constant NPT) are employed to produce the points of the simulation, so that they exhibit the same conditions of temperature, pressure or number of particles, depending on the ensemble used. To maintain constant properties like temperature and/or pressure in a simulation, barostats and/or thermostats should be setup (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984; Parrinello & Rahman, 1981).

In addition, to properly simulate a biomolecular system, it is often necessary to mimic the environment. The usage of Periodic Boundary Conditions (PBC) is most often recommendable, which implies using a solvation box replicated in all directions so that continue solvation is mimicked and edge effect are minimized (Katiyar & Jha, 2018). With PBC applied, a molecule exiting the box by one side will enter simultaneously by the opposite one, keeping the number of molecules inside the box constant during the whole simulation. A visual representation of the Periodic Boundary Conditions is shown in Figure 5.

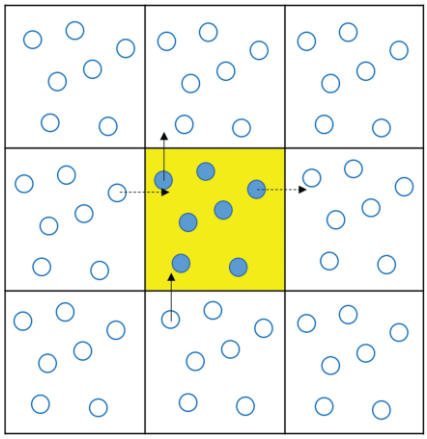

Figure 5. Visual representation of the Periodic Boundary Conditions (Katiyar & Jha, 2018).

The scheme of a MD simulation involves a more or less specific sequence of steps. First, the geometry of the biomolecule is retrieved from databases such as PDB or built in a molecule builder program (a good homology model if suitable). If necessary, the biomolecule geometry must be corrected and fixed. Then, a specific force field is selected, which will define the interatomic bonded and non-bonded parameters. Next, a simulation box is created (defined shape and size), PBC are usually setup and the box filled with solvent (water, ions and other relevant chemical species). An energy minimization step should be performed to relax the biomolecule's structure and to better accommodate water molecules in its surface (remove overlaps between molecules that results in high interatomic forces and avoid numerical instabilities). Subsequently, equilibration steps are performed from the minimized configuration to achieve a state of "thermodynamic equilibrium", in which the property of interest appears to converge to an average value. Once this thermodynamic equilibrium is attained, the productive step is carried out through a longer simulation that allows to sample many equilibrium configurations. Finally, the analysis is done to determine the properties of interest from the trajectories generated (Katiyar & Jha, 2018). The representation of a general MD simulation scheme is shown below (Figure 6).

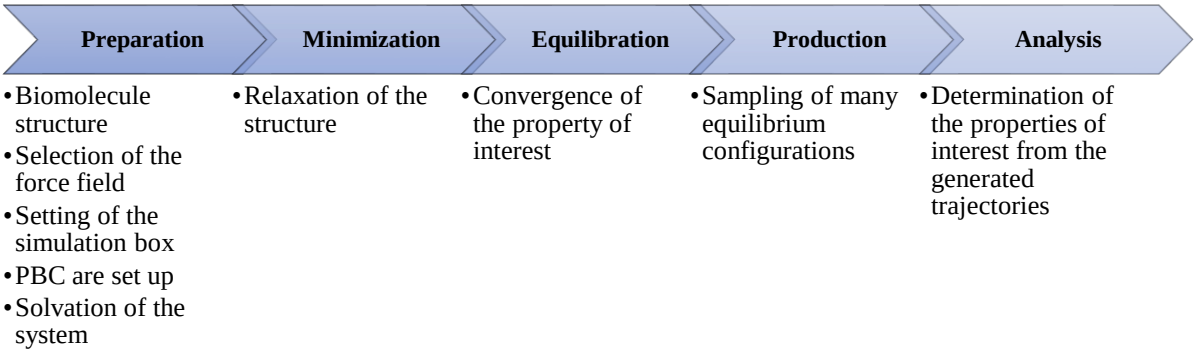

Figure 6. General MD simulation scheme. Own elaboration.

Advantageously, MD simulations are time and cost effective as compared with the experimental approach of performing gene expression, protein purification and activity assays (Kumari & Akhter, 2017).

Nevertheless, the unfolding process of a medium-size protein takes place in the order to milliseconds, or faster, under denaturing conditions (Daggett & Levitt, 1993). This timescale is accessible in experimental settings but not always achievable with present computer power due, in part, to the necessary extremely short iteration time (Hospital, Goñi, Orozco, & Gelpí, 2015) and the lack of new algorithms that can speed up the calculations (Galano-Frutos & Sancho, 2019). Therefore, to investigate the protein unfolding within the timescale available to MD simulations with explicit solvent, elevated temperatures have been used to speed up the process (Daggett & Levitt, 1993; Day, Bennion, Ham, & Daggett, 2002; Rocco, et al., 2008). This procedure has been deemed suitable for elucidating the details of protein unfolding without affecting the unfolding pathway, with a reduced computational cost (Day, Bennion, Ham, & Daggett, 2002).

On the other hand, the unfolded state of a protein is usually characterized by a huge number of different conformations with comparable free energies (Piana, Klepeis, & Shaw, 2014). So, small force field inaccuracies can significantly alter the structural and dynamical properties of the unfolded state (Piana, Klepeis, & Shaw, 2014). Nonetheless, overall, MD simulations are considered a reliable computational technique capable of providing accurate results in studies of protein conformation and stability, especially when using highly realistic settings such as all-atom representations, explicit solvent and specific solution conditions (Galano-Frutos, García-Cebollada, & Sancho, 2019; Hospital, Goñi, Orozco, & Gelpí, 2015).

2.4 Protein stability and unfolding analysis from MD simulations

There are several approaches in which MD simulations could be used to explore and sample the conformational space of a given protein (Galano-Frutos, García-Cebollada, & Sancho, 2019; Hospital, Goñi, Orozco, & Gelpí, 2015; Kirkpatrick, Gelatt, & Vecchi, 1983). In this sense, one the most used approaches involve starting from a native conformation for a given protein to then evaluate its evolution in time by MD simulations. This approach has been described as relaxation Molecular Dynamics (rMD) simulations as they deal with the relaxation of the protein native structure over time at a fixed temperature (Galano-Frutos, García-Cebollada, & Sancho, 2019).

On the other hand, ladder- and ramp-based temperature scanning MD simulations are also other types of MD experiments that involves the simulation of a given protein but in evolving temperature conditions. Having chosen a temperature range $[T_{min}, T_{max}]$, ladder-based temperature scanning MD simulations require the simulation of a temperature ladder ($T_1 = T_{min},\ T_2, \dots, T_N = T_{max}$) (Denschlag, Lingenheil, & Tavan, 2009); while ramp-based temperature scanning MD simulations involve the simulation of a slow temperature ramp from T_{min} to T_{max} at a given rate (Reyes, et al., 2018).

The folded state's stability and the unfolding process of a given protein are commonly analyzed through the description of the structural fluctuations of the macromolecule (Martínez L. , 2015). Some of the most common measures of structural fluctuations are the Root-Mean-Square-Deviation (RMSD), the Root-Mean-Square-Fluctuations (RMSF), the Radius of gyration (Rg), the Hydrogen bonds (H-bonds), the Native contacts (Nc), the Solvent Accessible Surface Area (SASA) and the Template Modelling score (TM-score) (Arnittali, Rissanou, & Harmandaris, 2019; Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009; Martínez L. , 2015; Zhang & Lazim, 2017; Zhang & Skolnick, 2004).

The RMSD is the average displacement of the atoms at an instant of the simulation relative to a reference structure, which is usually the first frame of the simulation or the crystallographic structure (Martínez L. , 2015), and it is commonly based only on the C_α atoms (Arnittali, Rissanou, & Harmandaris, 2019). So, in simpler terms, the RMSD is measure of the difference between two overlapping structures; and the smaller the deviation, the more spatially equivalent the two compared structures are (Arnittali, Rissanou, & Harmandaris, 2019). It is given by the equation (Arnittali, Rissanou, & Harmandaris, 2019):

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} m_i \, (\mathbf{r}_i - \mathbf{r}_{ref})^2}{\sum_{i=1}^{N} m_i}}$$

where m_i is the mass of atom i, $\mathbf{r}_i = (r_{i,x}, r_{i,y}, r_{i,z})$ are the coordinates of atom i at a certain instance, $\mathbf{r}_{ref} = (r_{ref,x},\ r_{ref,x}, r_{ref,z})$ are the coordinates of atom i at its reference state.

The RMSD is useful for the analysis of time-dependent motions of the structure, and it can be used to assess whether a structure is stable in the timescale of the simulation or if it is fluctuating from its initial configuration and position (Martínez L. , 2015). However, even though the RMSD is commonly used to compare protein structures, because all atoms in the structures are equally weighted in the calculation, one of the major drawbacks of this metric is that it becomes more sensitive to the local structure deviation than to global deviation when the RMSD value is big (Xu & Zhang, 2010). So, for instance, the RMSD of two protein structures could be high if the tails or some loops have a different orientation despite the global topology of the core part being the same (Xu & Zhang, 2010). Thus, based on the RMSD value alone, it is difficult to distinguish if two structures have completely different conformations.

To address traditional RMSD deficiencies, a two-dimensional RMSD (2D-RMSD) analysis involves calculating the RMSD of each frame in a trajectory to all other frames in the same or in another trajectory, thus providing more information (Wang, et al., 2020). In this way, a 2D-RMSD of a trajectory to itself can be used to gain insight into the conformational convergence of the simulation; whereas, when a 2D-RMSD is calculated to another trajectory, this analysis can be used to compare the similarity of two conformational ensembles to a more accurate extent (Wang, et al., 2020).

On the other hand, the RMSF is a measure of the displacement of a particular atom, or group of atoms, relative to the reference structure, averaged over the number of atoms (Martínez L. , 2015). This measurement allows to assess local fluctuations on the protein, and it is useful for proteins in which the global structure may not vary significantly but do it in a specific region thereof (Blasco-Puyuelo, 2019).

The Rg is an indicator of the size and the compactness of a protein, and it is calculated using the formula (Arnittali, Rissanou, & Harmandaris, 2019):

$$Rg = \sqrt{\frac{\sum_{i=1}^{N} m_i \, r_i^2}{\sum_{i=1}^{N} m_i}}$$

where m_i is the mass of the atom i, r_i is the distance of atom i from the center of mass of the protein.

The H-bonds refer to the number of formed hydrogen bonds in the structure. They are commonly defined based on a geometric criterion: the donor-acceptor distance must be less than or equal to 0.35 nm and the hydrogen-donor-acceptor angle must be less than or equal to 30°. The simultaneous compliance of these two conditions implies the existence of a hydrogen bond (Arnittali, Rissanou, & Harmandaris, 2019). As this kind of interactions play an important role in the maintenance of the secondary structure of a protein, they could be used a good approximation to assess any structural fluctuation of the macromolecule (Arnittali, Rissanou, & Harmandaris, 2019).

Native contacts are a measure of the number of interactions between spatially closed amino acids, which are not sequentially next to each other in the primary sequence of the protein (Zhang & Lazim, 2017). Thus, proteins that are unfolding exhibit a significant loss in native contacts over the simulation.

The SASA is a geometric measure of the degree to which the protein core is exposed to the solvent environment (Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009). So, when a protein is experiencing an unfolding process, the SASA increases, making the hydrophobic core of the protein to be in contact with the solvent. It can be estimated through several algorithms such as the maximal speed molecular surfaces or statistical approximations (Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009).

The TM-score is a scoring function given by the equation (Zhang & Skolnick, 2004):

$$TM - score = Max\left[\frac{1}{L_N}\sum_{i=1}^{L_T}\frac{1}{1+\left(\frac{d_i}{d_0}\right)^2}\right]$$

where L_N is the length of the native structure, L_T is the length of the aligned residues to the template structure, d_i is the distance between the ith pair of aligned residues and d_0 is a scale to normalize the match difference. Max denotes the maximum value after optimal spatial superposition. The value of the TM-score always lies between [0, 1], with better templates having higher TM-scores (Zhang & Skolnick, 2004). In particular, values above 0.5 can be deemed as the same conformation.

Finally, a clustering analysis is a technique in which a set of data points is partitioned into a disjoint collection of data sets called clusters where, ideally, the points in one cluster are more similar to each other than to the points from other clusters (Shao, Tanner, Thompson, & Cheatham, 2007). This tool has been successfully implemented for grouping together similar conformations visited during a MD simulation, by using several algorithms such as hierarchical, centroid-based or density-based clustering (Shao, Tanner, Thompson, & Cheatham, 2007).

3. MATERIALS AND METHODS

3.1 Computational resources and programs

MD simulations were performed on the computer cluster called Cierzo. Cierzo is a distributed memory supercomputer designed for high performance computing (Peak performance of 85.87 TeraFLOPs and 2.38 GigaFLOPS per watt consumed), from the Aragon Supercomputing Center (Cesar) hosted at the Institute for Biocomputation and Physics of Complex Systems (BIFI) of the University of Zaragoza. In particular, Intel Xeon E5-2680v3 2.5GHz cores were used, in a scheme of 96 processors per job in the production step.

Regarding the software, GROMACS 5.1.4 (Abraham, et al., 2015) and its whole set of analysis programs (version 5.0 or later) was used. For trajectory analyses a battery of Python 3 (Van Rossum & Drake, 2009) and BASH scripts were implemented. For visualizing MD trajectories and properties plots VMD 1.9.3 (Humphrey, Dalke, & Schulten, 1996), UCSF Chimera 1.14 (Pettersen, et al., 2004) and GRACE 5.1.22 (Stambulchik, 2008) were employed.

3.2 Protein

To carry out the objectives posed in this work the protein α_3D was selected, whose features and reported experimental unfolding conditions are shown in the Table 1:

Table 1. Features and reported experimental unfolding conditions of the protein α_3D.

Protein	PDB code	Class	No. of Residues	$\ln(k_U)$ at 298 K	k_U (s^{-1})	Temperature (K)	pH	Reference
De novo bundle α_3D	2A3D	α	73	7.7	2114	298	2.2	(Zhu, et al., 2003)

Protein α_3D is a single chain three-helix bundle designed to be stabilized by the packing of hydrophobic side chains. It lacks buried polar residues or structured loops that might introduce significant kinetic barriers to folding, so this protein adopts a well-defined tertiary structure (Zhu, et al., 2003). It was selected due to its small size, suited for all-atom simulations in explicit solvent, the considerable magnitude of its rate of unfolding and because it has been described as a two-state folding state protein (Zhang & Luo, 2011; Zhu, et al., 2003).

3.3 Molecular Dynamics simulations

3.3.1 Simulations conditions

The protein was simulated using the force fields CHARMM27 (Brooks, et al., 2009; MacKerell, et al., 1998) and AMBER99SB-disp (Robustelli, Piana, & Shaw, 2018), under the temperature and during the time described in Table 2.

Table 2. Production simulation conditions for rMD simulations.

Force field	Temperature* (K)	Time (μs)	pH
CHARMM27	360	2.0	2.2
	380	0.5	2.2
	400	0.5	2.2
	420	0.5	2.2
	450	0.5	2.2
	500	0.5	2.2
AMBER99SB-disp	360	2.0	2.2
	380	0.5	2.2
	450	0.5	2.2
	500	0.5	2.2

*Ten replicas each.

The generation of the topologies was carried out trying to reproduce the conditions (protonation states according to the reported pH) under which the experimental unfolding parameters were obtained for the protein (Zhu, et al., 2003). Thus, the residues of histidine, aspartic acid and glutamic acid were protonated, yielding a total charge of +12.0 e.

Furthermore, ladder- and ramp-based temperature scanning MD simulations were carried out for both CHARMM27 and AMBER99SB-disp, five replicas each. The ladder-based temperature scanning MD simulations increased their temperature in 10 K from 298 K to 488 K each 50 ns, yielding a total of 1 μs of simulation. On the other hand, in the ramp-based temperature scanning MD simulations their temperature was increased from 298 K to 498 K at a rate of 1 K/ns during a time span of 200 ns (by using simulated annealing) (Kirkpatrick, Gelatt, & Vecchi, 1983).

3.3.2 Preparation phase

For each structure, the topology file was created with the corresponding force field and water model: TIP3P for CHARMM27 (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) or TIP4Pdisp for AMBER99SB-disp (Robustelli, Piana, & Shaw, 2018).

Each protein was first minimized in vacuum by using the steepest descent minimization method during 5000 steps with the Verlet cut-off scheme for neighbor searching, with 1 nm for both the van der Waals cut-off and the Coulomb cut-off.

Next, each structure was centered in a truncated dodecahedron simulation box with a minimal distance of 1 nm from the box edges to the protein, solvated with the selected water model, and neutralized with Na^+ and Cl^- counterions.

3.3.3 Minimization step

The solvated systems were taken to its minimal energy with the steepest descent minimization algorithm, by using a tolerance buffer of 1 kJ/mol·nm and a maximum of 20000 steps of 0.001 ps (a total minimization timespan of 20 ps under completion).

Moreover, regarding the non-bonded interactions, Particle-Mesh Ewald (PME) electrostatics with 1 nm cut-off for coulombic short-range interactions and a van der Waals cut-off scheme (1 nm) with the Potential-shift-Verlet modifier for short-range interactions were used. Periodic boundary conditions were setup. These parameters were identically settled in all next simulation steps.

3.3.4 Heating step

The heating process was performed using a canonical ensemble (constant NVT) with position restraints. In particular, each system was heated gradually from an initial temperature to its final temperature by increasing the temperature with a 50/60 K-ladder every 50 ps (25000 steps of 0.002 fs each tier).

For integrating Newton's equations of motion in this step and in the subsequent ones, the leap-frog algorithm was used. The temperature was controlled using the modified Berendsen thermostat with velocities rescaling (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984; Bussi, Donadio, & Parrinello, 2007), using a time constant for coupling of 0.1 ps. The atom velocities were generated according to the Maxwell distribution.

3.3.5 Equilibration steps

Once the systems were taken to the desired temperature, they were stabilized through two sequential equilibration phases performed with the isobaric-isothermal ensemble (constant NPT).

The first equilibration phase was carried out with position restraints on the heavy atoms during 250 ps (125000 steps of 0.002 fs each), whereas the second one was done without any restraint during 500 ps (250000 steps of 0.002 fs each). Both equilibration phases utilized the modified Berendsen thermostat with velocities rescaling (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984; Bussi, Donadio, & Parrinello, 2007) for temperature coupling (using a time constant for coupling of 0.1 ps), and the Parrinello-Rahman barostat (Parrinello & Rahman, 1981) for isotropic pressure coupling (employing a time constant for coupling of 5 ps and reference pressure of 1 bar). No velocities were generated but were taken from the system resulted from the previous phase.

3.3.6 Production phase

The production phase was carried out in an isobaric-isothermal ensemble (constant NPT, same parameters as in the previous equilibration phase), during the established time in Table 2 (time steps of 0.002 fs). The coordinates, velocities and forces in each simulation were stored every 100 ps.

3.4 Analysis of trajectories

An integrated analysis of the trajectories was performed by using several tools from GROMACS (Abraham, et al., 2015) through a script built in Python3 (see Annex II). There, RMSD, Rg, RMSF, H-bonds, SASA, native contacts, TM-scores, and secondary structure plots were obtained and analyzed. As a reference structure for computing the RMSD, the one obtained from NMR was used (Walsh, Cheng, Bryson, Roder, & DeGrado, 1999).

Moreover, a 2D-RMSD-based clustering was performed to compute the RMSD of the structure in each frame when aligned and compared to the remaining frames of the trajectory. A cutoff of 0.6 nm (Reva, Finkelstein, & Skolnick, 1998) was set to carry out the clustering. , which was implemented in another specific Python3 script (see Annex III). To group the clusters, an agglomerative hierarchical clustering with average linkage was used..

Trajectory analyses were complemented with the visualization thereof in VMD (Humphrey, Dalke, & Schulten, 1996) and UCSF Chimera 1.14 (Pettersen, et al., 2004).

3.5 Estimation of the half-life of unfolding and conformational stability from relaxation MD simulations

The half-lives of unfolding at the temperature of simulation ($\tau_{1/2}$) were estimated from the 2D-RMSD-based clustering plots with ImageJ (Schneider, Rasband, & Eliceiri, 2012), by estimating the point in the time axis in which new clusters deviated significantly from the reference cluster (see, for instance, Figure 8).

Then, the rate of unfolding at the simulation temperature (k_U), the rate of unfolding at 298 K (k_U^{298}), the half-life of unfolding at 298 K ($\tau_{1/2}^{298}$), and the conformational stability at 298 K (ΔG_U) were computed with the Equations (3), (6), (7) and (8), respectively, according to the empirical model previously developed in the research group (Galano-Frutos, García-Cebollada, & Sancho, 2019).

3.6 Estimation of the temperature of unfolding from ladder- and ramp-based temperature scanning MD simulations

Finally, in the ladder- and ramp-based temperature scanning MD simulations, the temperature of unfolding was also estimated from the 2D-RMSD-based clustering plots with ImageJ, by measuring the point in time the clusters grouped from the conformations of the protein deviated significantly from the initial conformation, i.e., the time of unfolding (t_U); and then evaluating said value into the function of temperature $T(t_U)$ of the corresponding simulation:

$$T_R(t_U) = t + 298$$

for the ramp-based temperature scanning MD simulations; and

$$T_L(t_U) = \begin{cases} 298\,K \quad if \quad 0 < t \leq 50\,ns \\ 308\,K \quad if \quad 50 < t \leq 100\,ns \\ 318\,K \quad if \quad 100 < t \leq 150\,ns \\ \vdots \\ 478\,K \quad if \quad 900 < t \leq 950\,ns \\ 488\,K \quad if \quad 950 < t \leq 1000\,ns \end{cases}$$

for the ladder-based temperature scanning MD simulations.

4. RESULTS AND DISCUSSION

4.1 Determination of the adequate conditions in rMD simulations

It is considered that the best conditions for the analysis of the protein unfolding simulations are those in which it is possible to watch the whole process, from the native conformation of the protein to the one in which it is considered unfolded.

In our simulations, rMD simulations at 360K and 380K (with both force fields: CHARMM27 and AMBER99-disp) did not exhibit relevant unfolding for any of the simulated replicas. While simulations at 400K and 420 K (only performed with the CHARMM27 force field) showed in some replicas relevant unfolding events. Unfortunately, due a lack of time, the corresponding simulations at 400K and 420 K with AMBER99-disp could not be completed. On the other hand, at 450K it was possible to see the process of unfolding for all the simulated replicas with both force fields here selected. Instead, in simulations at 500 K the protein experienced a fast process of unfolding from the beginning of the production phase, making difficult to measure the half-life of unfolding.

4.2 Protein unfolding analysis

To detect the unfolding events from the native conformations for each rMD simulation, the 2D-RMSD-based clustering was used. The clustering plots of all the replicas are shown in the Annex I.

4.2.1 2D-RMSD-based clustering with CHARMM27

In the case of the rMD simulations using the CHARMM27 force field, the trajectory of the replica r1 at 380K was used as the reference for performing the clustering analysis at the temperatures of 380 K, 400 K, 420 K, 450 K, and 500 K due to its stability profile. The simulations performed at the temperature of 360 K were excluded from this analysis since they were run on a different time length (2 μs) and did not show any relevant sign of unfolding.

So, in order to save space and as an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 1 at 380 K using the CHARMM27 force field is shown in Figure 7, (see the clustering plots for all the replicas in the Annex I).

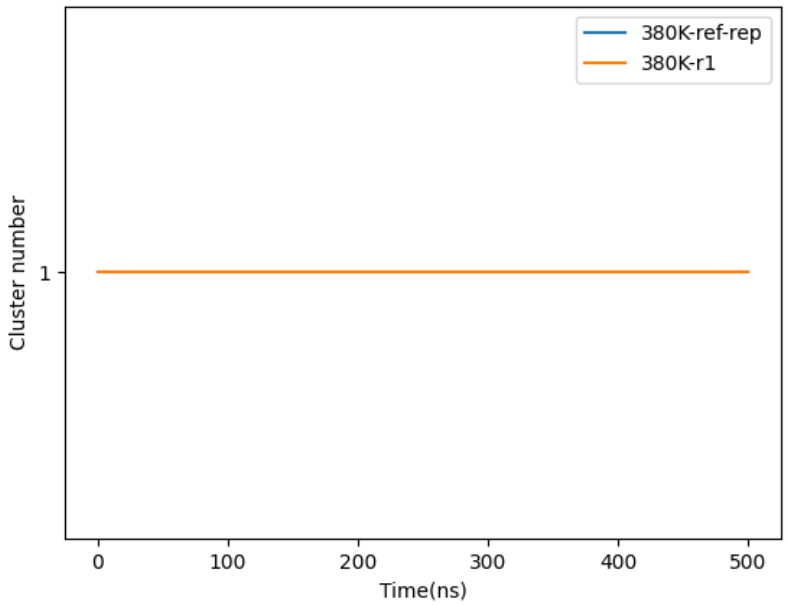

Figure 7. 2D-RMSD-based clustering plot for the replica 1 at 380 K using the CHARMM27 force field.

From the Figure 7, it is possible to observe that the protein did not exhibit any relevant structural difference in comparison with the native conformation along the time span of the simulation. This behavior was consistent in all simulated replicas (see Annex I).

As an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 2 at 400 K using the CHARMM27 force field is shown in the Figure 8 (see the clustering plots for all the replicas in the Annex I).

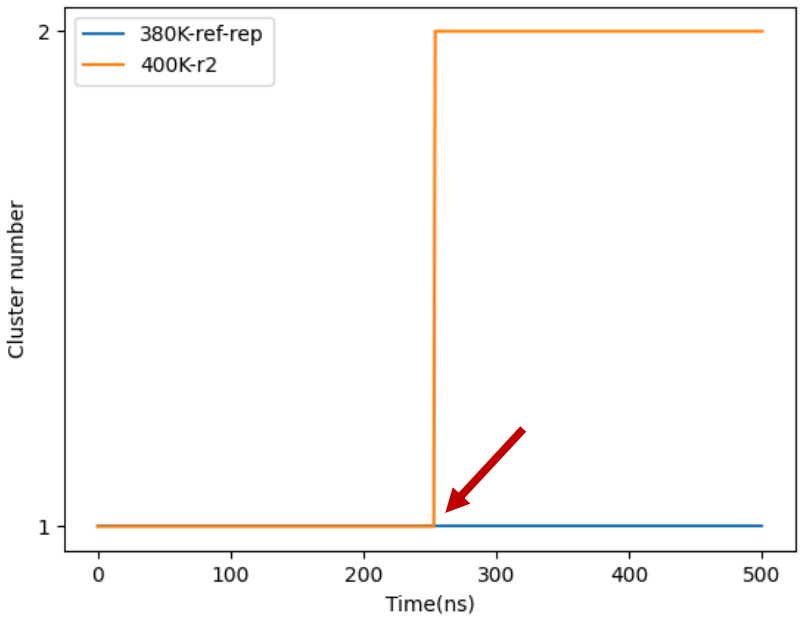

Figure 8. 2D-RMSD-based clustering plot for the replica 2 at 400 K using the CHARMM27 force field.

In this case, the replica 2 with the CHARMM27 force field at 400 K exhibited a structural variation from the native conformation at the 253.33 ns of the simulation. Nevertheless, at this temperature, only 2 replicas (r2 and r8) out of 10 showed some noticeable degree of unfolding within the 500 ns of the simulation (see Annex I).

As an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 8 at 420 K using the CHARMM27 force field is shown in the Figure 9 (see the clustering plots for all the replicas in the Annex I).

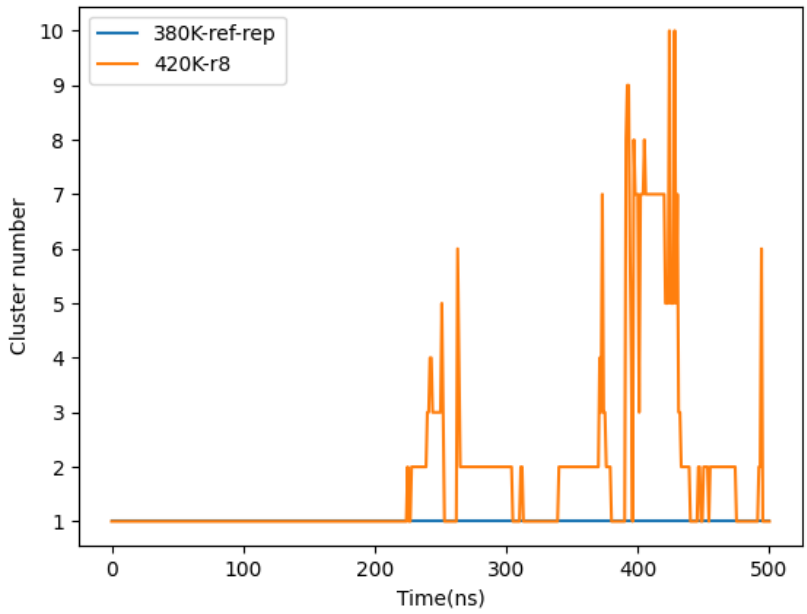

Figure 9. 2D-RMSD-based clustering plot for the simulations at 420 K using the CHARMM27 force field, replica 8.

At the temperature of 420 K, the rMD simulation in the replica 8 was able to generate a trajectory for which numerous conformations could be grouped in clusters, suggesting the occurrence of an unfolding process for the protein. It is noteworthy that this force field yielded a trajectory in which the structure returned to the native conformation in several moments during the simulation; even though its conformation changed repeatedly across the simulation. Furthermore, at this temperature and within the time span of 500 ns, 4 replicas (r3, r6, r8 and r10) out of 10 showed some noticeable degree of unfolding (see Annex I).

It is noteworthy that the 2D-RMSD-based clustering allowed to reduce the variability that it is characteristic of the 1D-RMSD plots (see Annex I), thus allowing to more easily identify the clusters of similar conformations visited during the timespan of the simulation.

As an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 4 at 450 K using the CHARMM27 force field is shown in the Figure 10 (see the clustering plots for all the replicas in the Annex I).

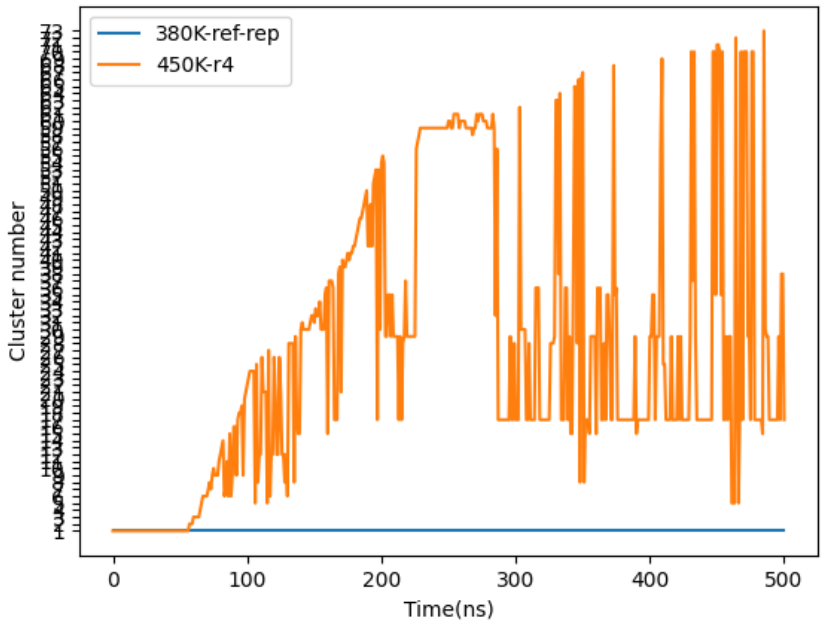

Figure 10. 2D-RMSD-based clustering plot for the replica 4 at 450 K using the CHARMM27 force field.

From Figure 10, it is possible to see that the protein experienced relevant changes in its structure due to the formation of numerous clusters far from the reference cluster. In particular, once the structure experienced a process of unfolding, it was not able to return to its native conformation later in the simulation, yielding a half-life of 55.92 ns. Nonetheless, the rest of replicas returned to a native-like conformation briefly, after crossing the border of their half-life (see Annex I). In this sense, from Figure 11 it is possible to observe the supposed unfolded state of the protein with CHARMM, in which the α-helix structures remain intact even though they are unpacked. So, it may be possible that this behavior is related with the well-known backwards that have been described for CHARMM27 in relation with the α-helix overstabilization they present (Gao, et al., 2015; Lindorff-Larsen, et al., 2012).

Figure 10 also suggests the formation of some recurring conformations that are often adopted by the unfolded structure, for instance, the cluster 18. This behavior is also consistent with the one observed for other replicas (see Annex I).

So, in order to better understand the results from this MD simulation, some frames of the trajectory of the replica 4 at 450 K using the CHARMM27 force field is shown in the Figure 11.

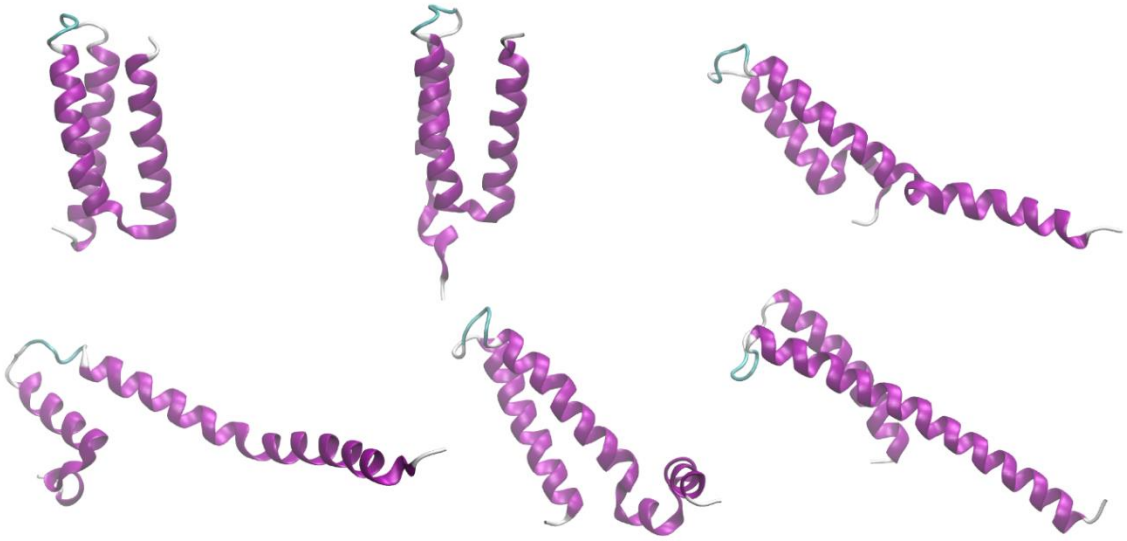

Figure 11. Protein conformations at 10 ns (top left), 50 ns (top center), 100 ns (top right), 200 ns (bottom left), 300 ns (bottom center) and 450 ns (bottom right) for the replica 4 at 450 K using the CHARMM27 force field.

Figure 11 shows the progression over time of the conformation of α_3D at 450 K of the replica 10 as an example of the ten simulated replicas (see Annex I for the unfolding analysis plots of the other replicas), ranging from the native conformation at the first 50 ns to an unfolded structure after the half-life of unfolding of 55.92 ns. In this case, it is possible to see that the disordered conformation mostly consists in the rupture of the packing of the α-helix structures, to give place to a more extended conformation but without losing the wholeness of the native α-helixes. At the temperature of 450 K and within the time span of 500 ns, 10 replicas out of 10 that were run showed a noticeable degree of unfolding (see Annex I), yielding an average half-life of 46.26 ± 13.34 ns.

Finally, as an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 4 at 500 K using the CHARMM27 force field is shown in the Figure 12 (see the clustering plots for all the replicas in the Annex I).

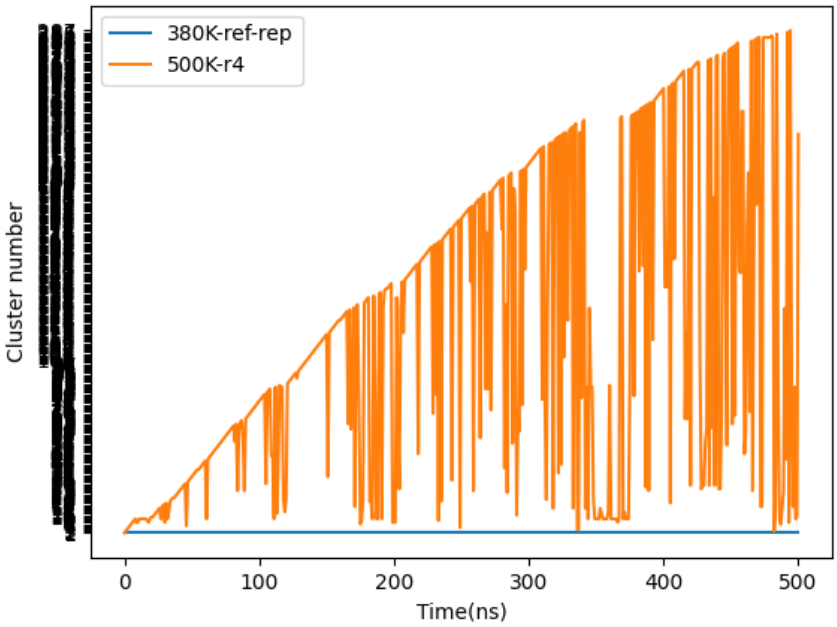

Figure 12. 2D-RMSD-based clustering plot for the replica 4 at 500 K using the CHARMM27 force field.

As with the results of 2D-RMSD-based clustering plot at 450 K, Figure 12 showed that, at 500 K, a similar behavior was observed, except that the time in which the unfolding was detected was smaller.

On the other hand, the half-life of unfolding for the replica 4 (2.27 ns) was smaller than that obtained at 450 K (55.92 ns). In this way, this 2D-RMSD-based analysis method allowed to clearly identify the point in which rMD trajectories of proteins deviate from the native conformation.

To assess whether at higher temperatures the α-helix structures observed at 450 K remained intact or unfolded, some frames of the trajectory are shown in the Figure 13.

Figure 13. Protein conformations at 1 ns (top left), 25 ns (top center), 50 ns (top right), 100 ns (bottom left), 300 ns (bottom center) and 450 ns (bottom right) of the replica 4 at 500 K using the CHARMM27 force field.

Figure 13 shows the progression over time of the conformation of α_3D at 500 K, going from the native conformation at the first few frames of the simulation to an ensemble of unfolded structures after the half-life of unfolding of 2.27 ns.

In comparison to the trajectory obtained at 450 K, it is possible to observe that, besides the disruption of the hydrogen bonds of the bundle of α-helix, some sections of the α-helix structures were also unfolded giving rise to a more disordered structure than the one obtained at 450 K. However, it is noteworthy that, for the most part of the simulation, significant portions of the α-helix structures were conserved by the force field CHARMM27.

At the temperature of 500 K and within the time span of 500 ns, 10 replicas out of 10 showed the same noticeable degree of unfolding (see Table 3 below), yielding an average half-life of 4.52 ± 0.72 ns.

4.2.2 2D-RMSD-based clustering with AMBER99SB-disp

On the other hand, regarding the rMD simulations using the AMBER99SB-disp force field, the trajectory of the replica r6 at 380K was used as the reference for performing the clustering analysis at temperatures 380 K, 450 K and 500 K. Again, the trajectories at the temperature of 360 K were excluded due to identical reasons to those stated for the simulations with CHARMM27.

So, as an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 5 at 380 K using the AMBER99SB-disp force field are shown in the Figure 14 (see the clustering plots for all the replicas in the Annex I).

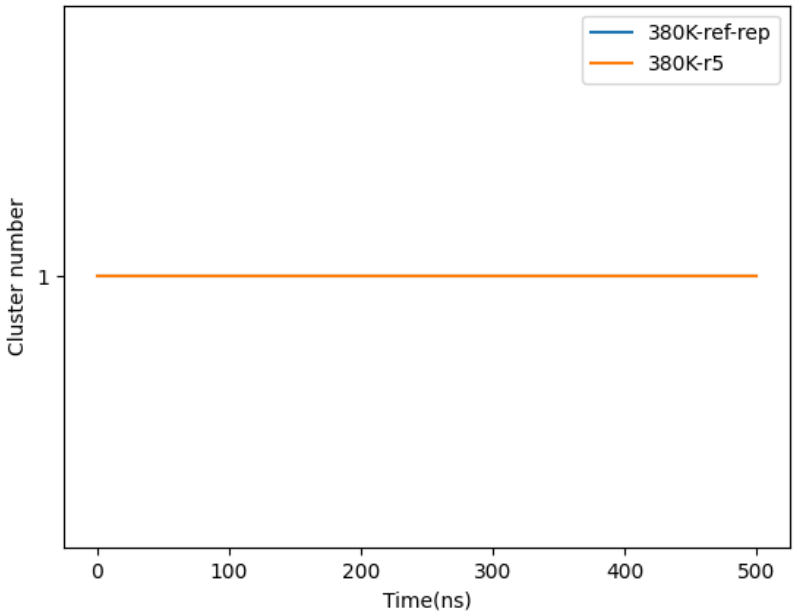

Figure 14. 2D-RMSD-based clustering plot for the replica 5 at 380 K using the AMBER99SB-disp force field.

From Figure 14, it is possible to observe that the protein did not exhibit any relevant sign of unfolding. This outcome was consistent in all simulated replicas (see Annex I).

On the other hand, as an example of the results obtained, the plot from the 2D-RMSD-based clustering at 450 K using the AMBER99SB-disp force field are shown in the Figure 12 (see the clustering plots for all the replicas in the Annex I).

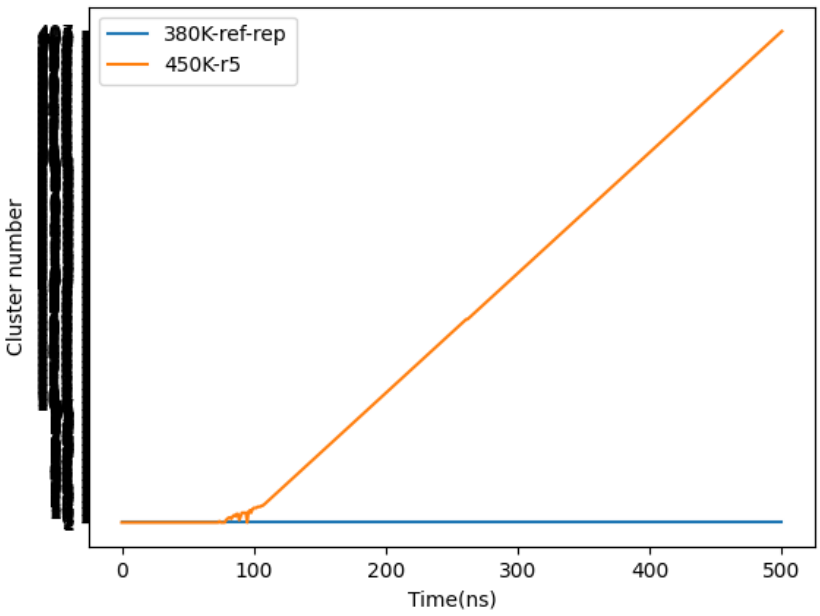

Figure 15. 2D-RMSD-based clustering plot for the replica 5 at 450 K using the AMBER99SB-disp force field.

Figure 15 shows the unfolding process of the protein at 450 K, exhibiting a half-life of unfolding of 80.28 ns. It is noteworthy that the profile of clustering build based on the trajectory generated by AMBER99SB-disp force field is slope-shaped, unlike the pattern generated by using the CHARMM27, which tend to have considerably more fluctuations.

This outcome is common in all the simulated replicas (see Annex I), in which, after a certain point after the starting of the process of unfolding. The simulations using the AMBER99SB-disp force field at 450 K yielded an average half-life of 58.05 ± 11.91 ns.

In order to assess whether a complete process of unfolding took place as suggested by the clustering method, more analyses were carried out. So, the RMSD, RMSF and TM-score for the replica 5 with AMBER99SB-disp at 450 K are shown in the Figure 16.

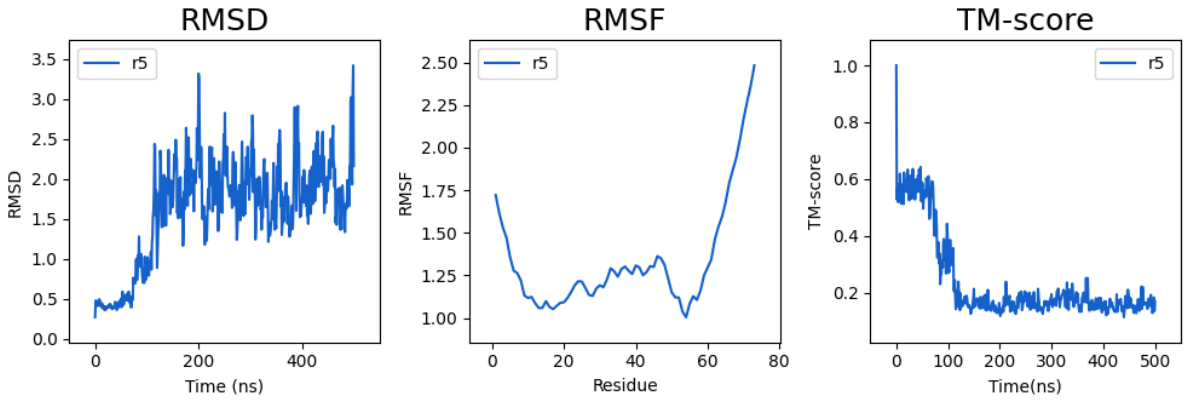

Figure 16. RMSD, RMSF and TM-score for the replica 5 at 450 K using the AMBER99SB-disp force field.

Indeed, from Figure 16, it is possible to see that the values of the RMSD increased from about 0.5 to 2.5 after approximately 100 ns of simulation, which implies that the average displacement of the atoms is significative in comparison with the NMR structure. On the other hand, the TM-score reduced its value from 1 in the first moments of the simulation to about 0.2 after approximately 100 ns, which means that the conformation of the protein after the half-life of unfolding of 80.28 ns is significatively different from the original conformation. Finally, the RMSF results suggested that, relative to the native conformation, the displacements or fluctuations of the first and last residues of the structure were important.

On the other hand, the H-bonds, native contacts and SASA estimations for the replica 5 with AMBER99SB-disp at 450 K are shown in the Figure 17.

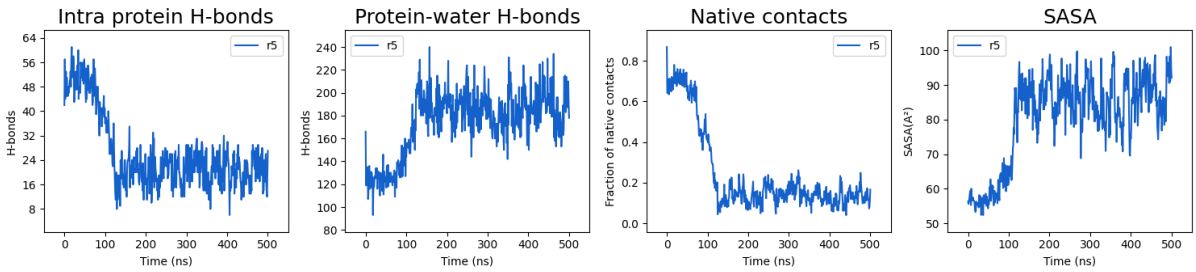

Figure 17. H-bonds, Native contacts and SASA for the replica 5 at 450 K using the AMBER99SB-disp force field.

In agreement with the above results, from Figure 17 it is possible to observe that, after the identified half-life of unfolding of 80.28 ns, intra protein H-bonds decreased significantly while the protein-water H-bonds showed a contrary behavior. As the H-bonds play an important role in the maintenance of the secondary structure of a protein (Arnittali, Rissanou, & Harmandaris, 2019), the drastic change in their numbers suggested that the protein lost important part of its secondary structure. In this sense, the diminution of the native contacts between spatially closed amino acids that are not sequentially next to each other in the primary sequence of the protein (Zhang & Lazim, 2017), and the increase in the SASA measure due to the exposition of the hydrophobic protein core to the solvent environment (Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009), also supported the idea that the protein unfolded.

Moreover, the Rg and the secondary structure for the replica 5 using the AMBER99SB-disp force field at 450 K are shown in the Figure 18.

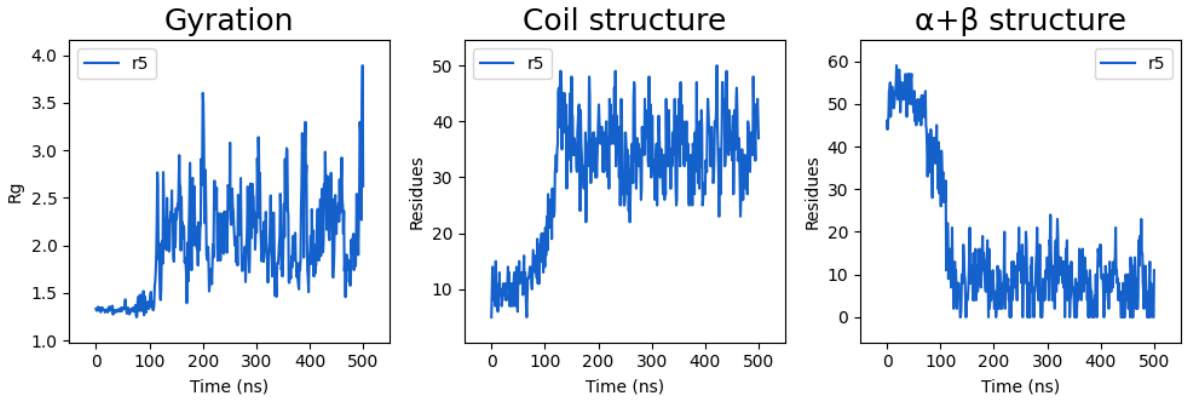

Figure 18. Rg, and secondary structure for the replica 5 at 450 K using the AMBER99SB-disp force field.

From the raise in the value of the Rg shown in the Figure 18, it is possible to state that the size and the compactness of a protein have notably diminished after approximately 100 ns of simulation. Furthermore, the gradual increase in the coil structure and the decrease in the α-helix structure after the identified half-life of unfolding, also reinforce the conclusion that the protein experienced a process of unfolding as suggested by the 2D-RMSD-based clustering analysis.

Finally, to further understand this phenomenon, some frames of the trajectory of the replica 5 using the AMBER99SB-disp force field at 450 K were assessed (Figure 19), as an example of the results obtained (see Annex I for the unfolding analysis plots of the other replicas).

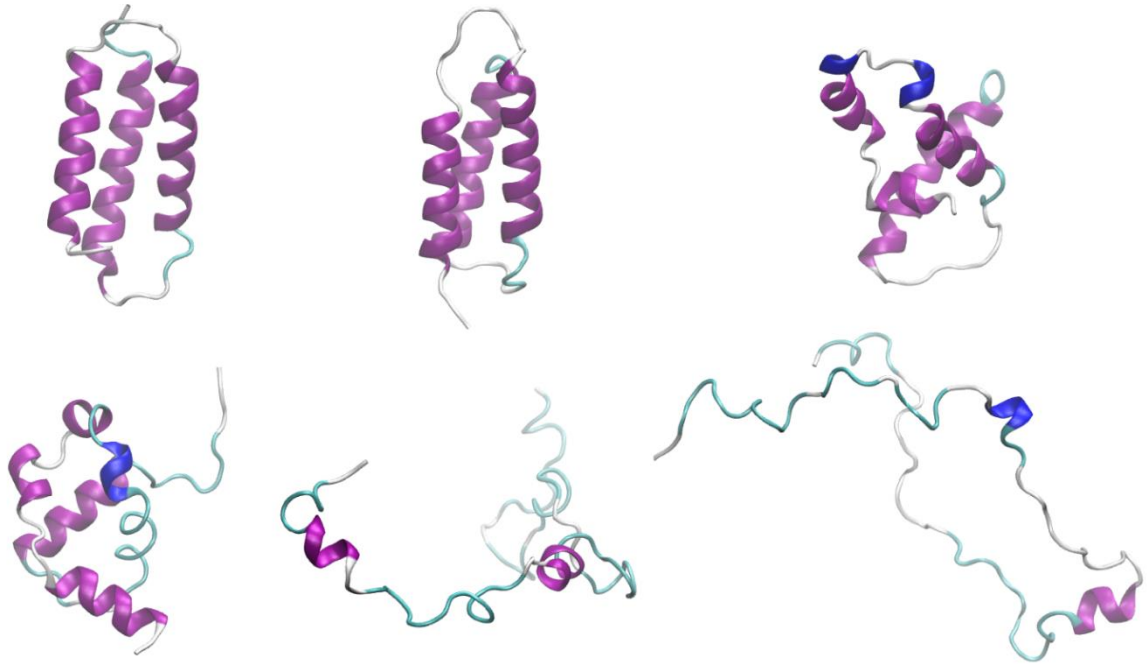

Figure 19. Protein conformations at 1 ns (top left), 60 ns (top center), 80 ns (top right), 110 ns (bottom left), 200 ns (bottom center) and 450 ns (bottom right) of the replica 5 at 450 K using the AMBER99SB-disp force field.

Figure 19 shows the progression over time of the conformation of α_3D at 450 K in the replica 5 by using the AMBER99SB-disp force field, showing a native conformation at the first 60 ns, a partially disordered formation between 80-110 ns, and a completely disrupted structure onwards.

In this case, after the half-life of unfolding has been reached (see conformation at 110 ns in Figure 19) at 80.28 ns, both the tertiary structure of the protein started to disrupt, and the α-helix structures started to unfold. Therefore, according to the results from this rMD simulation, the disruption of the tertiary and secondary structures operates simultaneously in a process of protein unfolding for α_3D.

In this exemplary trajectory, after 200 ns, the protein is completely unfolded, with only some small sections of its chains retaining some α-helix structures. This behavior is consistent in all simulated replicas (see Annex I) and indicates that the disadvantage of the usual 1D-RMSD analysis of not being able to distinguish if two structures have completely different conformations (Xu & Zhang, 2010) have been overcome by this analysis.

To assess if this behavior stands at higher temperatures, and as an example of the results obtained, the plot from the 2D-RMSD-based clustering plot for the replica 5 at 500 K using the

AMBER99SB-disp force field are shown in the Figure 20 (see the clustering plots for all the replicas in the Annex I).

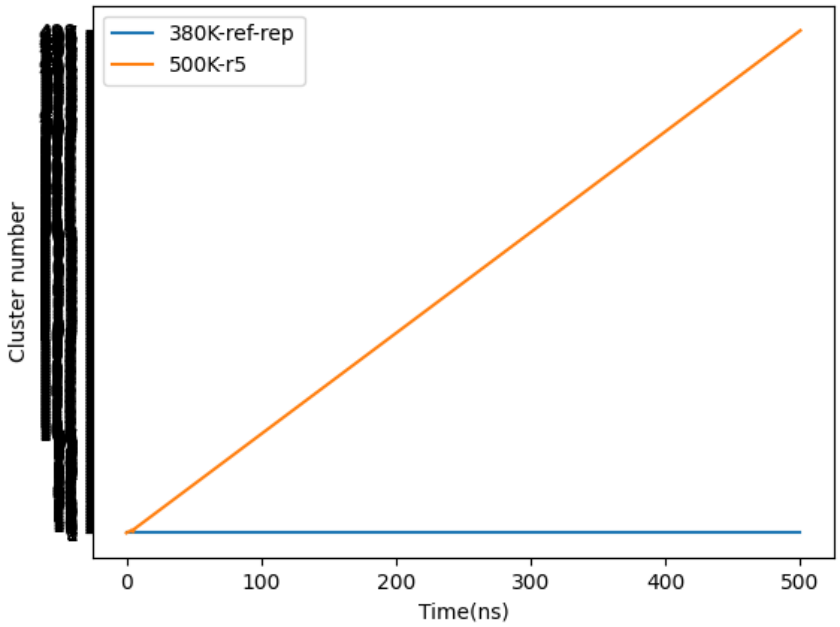

Figure 20. 2D-RMSD-based clustering plot for the replica 5 at 500 K using the AMBER99SB-disp force field.

As expectable, Figure 20 shows that the half-life of unfolding was sensibly reduced. In the case of the replica 5, to 6.57 ns at 500 K. This outcome was common in all the simulated replicas (see Annex I).

It is noteworthy that the identification of the half-life of unfolding was easier in the 2D-RMSD-based clustering plots when using the AMBER99SB-disp force field, due to the its clustering profile. The replicas using the AMBER99SB-disp force field at 500 K yielded an average half-life of 7.69 ± 1.46 ns.

Finally, some frames of the trajectory of the protein using the AMBER99SB-disp force field is shown in the Figure 21.

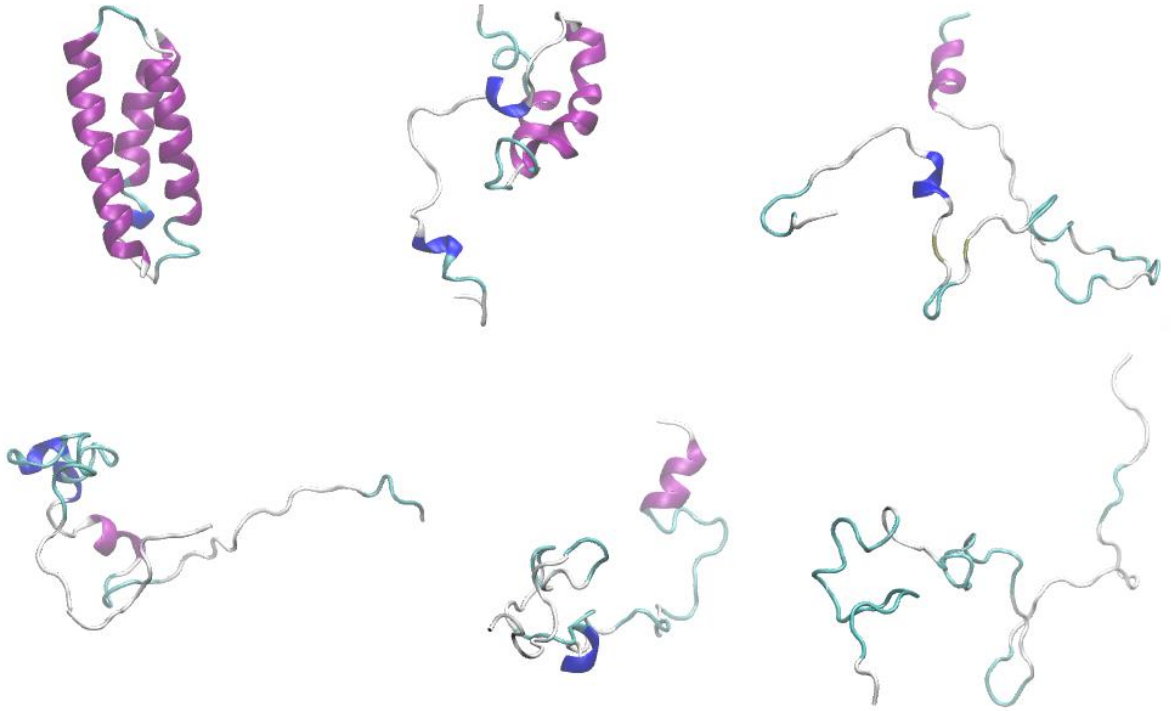

Figure 21. Protein conformations at 1 ns (top left), 5 ns (top center), 10 ns (top right), 100 ns (bottom left), 300 ns (bottom center) and 450 ns (bottom right) of the replica 5 at 500 K using the AMBER99SB-disp force field.

Figure 21 shows the progression over time of the conformation of α_3D at 500 K in the replica 5 by using the AMBER99SB-disp force field (see Annex I for the unfolding analysis plots for the other replicas), showing a native conformation at 1 ns, a partially unfolded structure at 5 ns, and a complete extended conformation onwards. It is noteworthy that the frame at 5 ns shows a protein that has lost most of its secondary structure; however, the 2D-RMSD-based clustering analysis indicates that the half-life of unfolding occurs at 6.57 ns, which suggests that the RMSD threshold of 0.6 could be higher than needed.

Again, as in the simulation at 450 K, the process of unfolding for α_3D comprised both the simultaneously disruption of the tertiary and secondary structure of the protein. So, in the Figure 21, it is possible to see that the rupture of the packing of the α-helix structures and the unfolding of the α-helixes occur at the same time at the 5 ns frame. Noteworthy, and in agreement with the results at 450 K, the simulation showed a completely structureless polypeptide chain at 450 ns, without any remaining secondary structure.

Thus, these findings obtained with the AMBER99SB-disp force field contradict the outcome from the simulations with the CHARMM27 force field, which suggested that the proteins retained the wholeness of the native α-helix structures even at high temperatures like 450 K or 500 K.

In this context, despite CHARMM27 provides a remarkable description of the folded state of proteins (Piana, Klepeis, & Shaw, 2014; Lindorff-Larsen, et al., 2012), it is known that CHARMM27

has an α-helical propensity (Gao, et al., 2015) as this force field severely over-stabilizes the formation of helical structures, generating unfolded states that are substantially more helical than those found experimentally (Lindorff-Larsen, et al., 2012). For instance, a simulation of 10 μs with CHARMM27 of human Pin1 WW domain starting from an unfolded conformation resulted in a helical structure, in completely disagreement with the native conformation of the protein which is all β-sheets (Freddolino, Liu, Gruebele, & Schulten, 2008). On the other hand, another study found that CHARMM27 predicted the formation of a helical fraction of about 0.6 in the secondary structure of ubiquitin and GB3 at 370 K, whereas the experimental data from NMR showed that there were no α-helix structures whatsoever in the proteins (Lindorff-Larsen, et al., 2012).

Therefore, we have been able to verify within this work the feature of the CHARMM27 force field of over-stabilizing α-helical structures. So, even though this forcefield is suitable for simulating folded proteins and it has been successfully used for calculating folding energetics (Galano-Frutos & Sancho, 2019), reason why this force field was selected in this study, the results from these simulations suggest that CHARMM27 is not suitable for simulating unfolded proteins.

Considering the limitations of CHARMM27 and other current force fields, AMBER99SB-disp was developed to provide high levels of accuracy in simulations of disordered protein states while maintaining the required accuracy for folded proteins (Robustelli, Piana, & Shaw, 2018). In this sense, the results from this work suggest that AMBER99SB-disp is much more suitable to simulate unfolded states, as shown in the analysis of RMSD, RMSF, TM-score, H-bonds, Native contacts, SASA, Rg, secondary structure and trajectory in the Figures 16-19.

4.3 Protein stability and unfolding kinetics analysis

4.3.1 Estimation of half-life of unfolding

Based on the results of the 2D-RMSD-based clustering, the half-life of unfolding was estimated for each replica at each temperature for the two force fields, by measuring the time in which the 2D-RMSD-based clustering plots deviates from the reference cluster (native conformation) with ImageJ. So, the half-life of unfolding and the folded and unfolded fractions for the simulations at 380 K, 400 K, 420 K, 450 K and 500 K performed with the CHARMM27 force field are shown in the Table 3.

Table 3. Half-life of unfolding and folded and unfolded fractions for the simulations performed with the CHARMM27 force field at 380 K, 400 K, 420 K, 450 K and 500 K.

	Half-life of unfolding (ns)				
	380 K	**400 K**	**420 K**	**450 K**	**500 K**
r1	>500	>500	>500	80.74	4.41
r2	>500	253.33	>500	102.50	---
r3	>500	>500	153.89	33.33	2.26
r4	>500	>500	>500	55.92	2.27
r5	>500	>500	>500	16.95	9.72
r6	>500	>500	467.78	20.28	5.51
r7	>500	>500	>500	122.96	5.51
r8	>500	489.99	224.99	5.01	3.33
r9	>500	>500	>500	14.07	3.33
r10	>500	>500	98.61	10.83	4.35
Average	>500	>371.66	>236.32	46.26	4.52
Standard Error	---	~52.92	~51.47	13.34	0.72
Folded Fraction	1	0.8	0.6	0	0
Unfolded Fraction	0	0.2	0.4	1	1

From the Table 3, the average half-life of unfolding at 450 K and 500 K yielded by the rMD simulations with CHARMM27 were 46.26 ± 13.34 ns and 4.52 ± 0.72 ns, respectively. Thus, an increase in 50 K implied a reduction in about 10 times the value of the half-life.

In the case of the temperatures 380 K, 400 K, and 420 K with CHARMM27, it was not possible to provide an accurate value for the half-life average since not all the replicas exhibited definitive unfolding events within the 500 ns of the simulations. Thus, only a lower limit is described in Table 3.

On the other hand, in the case of the replica 2 at 500 K, the simulation crashed possibly due to the high temperature; thus, it was impossible to estimate a half-life value for that replica.

Moreover, with the folded and unfolded fractions versus temperature, it was possible to build the following plot (Figure 22).

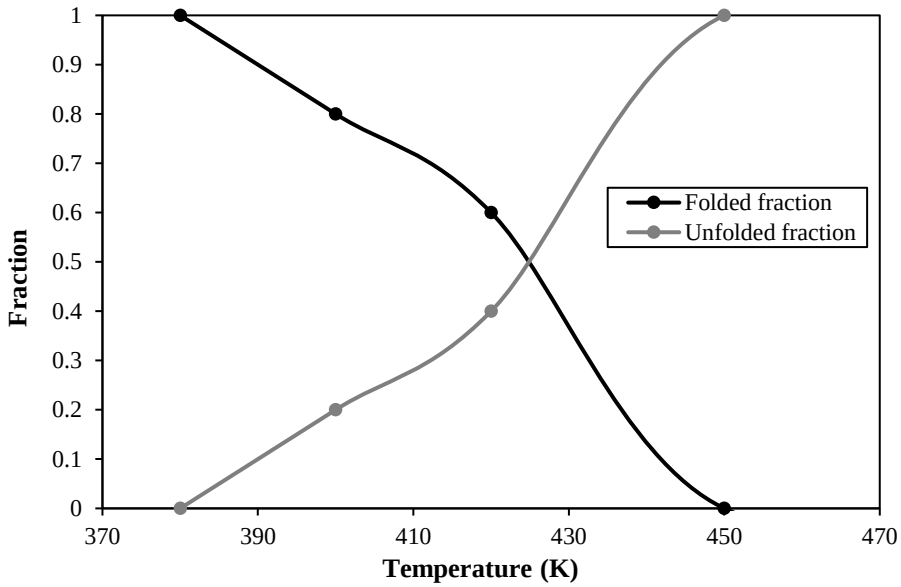

Figure 22. Plot of folded and unfolded fraction versus temperature for the results obtained through the CHARMM27 force field.

From the Figure 22, it is possible to predict that the temperature at which the half of the replicas (T_m) are exhibiting unfolding within the 500 ns of simulation with CHARMM27 is 425 K.

On the other hand, the half-life of unfolding, and the folded and unfolded fractions for the simulations at 380 K, 450 K and 500 K performed with the AMBER99SB-disp force field are shown in the Table 4.

Table 4. Half-life of unfolding and folded and unfolded fractions for the simulations performed with the AMBER99SB-disp force field at 380 K, 450 K and 500K.

	Half-life of unfolding (ns)		
	380 K	**450 K**	**500 K**
r1	>500	35.93	3.40
r2	>500	105.28	4.31
r3	>500	115.55	---
r4	>500	15.01	7.69
r5	>500	80.28	6.57
r6	>500	32.08	15.37
r7	>500	15.42	---
r8	>500	30.37	14.31
r9	>500	58.70	4.38
r10	>500	91.85	5.51
Average	>500	58.05	7.69
Standard Error	---	11.91	1.46
Folded Fraction	1	0	0
Unfolded Fraction	0	1	1

From Table 4, the average half-life of unfolding at 450 K and 500 K yielded by the simulations with AMBER99SB-disp were 58.05 ± 11.91 ns and 7.69 ± 1.46 ns, respectively, which are in the same range as the ones obtained with CHARMM27. Nonetheless, the percentage differences in the half-life of unfolding estimated with CHARMM27 and AMBER99SB-disp were 22.6% and 51.9% at 450 K and 500 K, respectively. So, overall, the simulations with AMBER99SB-disp seems to lead to slightly larger values for the half-life of unfolding. Furthermore, unlike with CHARMM27, the increase from 450 K to 500 K only decreased the half-life of unfolding in 7.5 times for the simulations with AMBER99SB-disp.

On the other hand, in the case of the replicas 3 and 7 at 500 K, the simulations crashed due to the high temperature; thus, it was impossible to estimate a half-life value for these replicas as well.

4.3.2 Estimation of the conformational stability and kinetic unfolding parameters

Once the values of half-life of unfolding were estimated based on the 2D-RMSD-based clustering analysis, equations (3), (6), (7) and (8) from the empirical model (Galano-Frutos, García-Cebollada, & Sancho, 2019) were used to calculate the kinetic unfolding parameters for the protein α_3D. So, the half-life of unfolding at the temperature of simulation ($\tau_{1/2}$), the rate of unfolding at temperature of simulation (k_U), the rate of unfolding at 298 K (k_U^{298}), the half-life of unfolding at 298 K ($\tau_{1/2}^{298}$), and the conformational stability at 298 K (ΔG_U) estimated with the CHARMM27 force field are shown in the Table 5.

Table 5. Kinetic unfolding parameters for the simulations performed with the CHARMM27 force field at 380 K, 400 K, 420 K, 450 K and 500 K.

Temperature	**380 K**	**400 K**	**420 K**	**450 K**	**500 K**
$\tau_{1/2}$ **(ns)**	>500.00	>371.66 ± 52.92	>236.32 ± 51.47	46.25 ± 13.34	4.52 ± 0.72
k_U [a] **(s^{-1})**	<1.39x10^6	<1.87x10^6	<2.93x10^6	1.50x10^7	1.53x10^8
ln (k_U)	<14.14	<14.44	<14.89	16.52	18.85
k_U^{298} [b] **(s^{-1})**	<11.99 ± 120.60	<68.61 ± 46.26	<45.94 ± 24.70	865.63 ± 1944.40	879.93 ± 1986.09
ln (k_U^{298})	<2.48	<4.23	<3.83	6.76	6.78
$\tau_{1/2}^{298}$ [c] **(ms)**	>57.79 ± 581.13	>10.10 ± 6.81	>15.09 ± 8.11	0.80 ± 1.80	0.79 ± 1.78
ΔG_U [d] **(kcal/mol)**	>3.05 ± 3.43	>2.49 ± 0.47	>2.62 ± 0.43	1.68 ±0.96	1.67 ± 0.97

Calculations done with: [a] Equation (3); [b] Equation (6); [c] Equation (7); and [d] Equation (8).

Even though the estimations for ΔG_U are inaccurate for the temperatures 380 K, 400 K and 420 K due to the impossibility of observe unfolding events for all the replicas within the timespan of the

simulation, it is noteworthy that the estimations for ΔG_U coincide for the simulations at 450 K and 500 K, in a value of about 1.68 ± 0.97 kcal/mol. Moreover, it is also noticeable that the estimations for $ln\ (k_U^{298})$ and $\tau_{1/2}^{298}$ computed from the simulations at 450 K and 500 K also coincide in a value of about 6.77 and about 0.80 ± 1.79 ms, respectively.

Finally, the kinetic unfolding parameters for the simulations performed with the AMBER99SB-disp force field at 380 K, 450 K and 500 K are shown in the Table 6.

Table 6. Kinetic unfolding parameters for the simulations performed with the AMBER99SB-disp force field at 380 K, 450 K and 500 K.

Temperature	**380 K**	**450 K**	**500 K**
$\boldsymbol{\tau_{1/2}}$ **(ns)**	>500.00	58.05 ± 11.91	7.69 ± 1.46
$\boldsymbol{k_U}$ **[a] (s^{-1})**	<1.39x10^6	1.19x10^7	9.01x10^7
ln $(\boldsymbol{k_U})$	<14.14	16.30	18.32
$\boldsymbol{k_U^{298}}$ **[b] (s^{-1})**	<11.99 ± 120.60	689.84 ± 1543.22	517.16 ± 1168.47
ln $(\boldsymbol{k_U^{298}})$	<2.48	6.54	6.25
$\boldsymbol{\tau_{1/2}^{298}}$ **[c] (ms)**	>57.79 ± 581.13	1.00 ± 2.22	1.34 ± 3.03
$\boldsymbol{\Delta G_U}$ **[d] (kcal/mol)**	>3.05 ± 3.43	1.75 ± 0.96	1.84 ± 0.96

Calculations done with: [a] Equation (3); [b] Equation (6); [c] Equation (7); and [d] Equation (8).

Likewise, even though the estimations for the kinetic unfolding parameters are inaccurate for the simulations at 380 K, it is also noteworthy that the estimations for $ln\ (k_U^{298})$, $\tau_{1/2}^{298}$ and ΔG_U mostly coincide for the simulations at 450 K and 500 K performed with AMBER99SB-disp, in a range of 6.25-6.54, 1.00-1.34 ms, and 1.75-1.84 kcal/mol, respectively. Remarkably, it is considered that the estimation for ΔG_U from both simulations at 450 K and 500 K is consistent as both values of 1.75 ± 0.96 kcal/mol and 1.84 ± 0.96 kcal/mol are within their respective estimated standard errors.

So, in general, the simulations with AMBER99SB-disp yielded higher values of $\tau_{1/2}^{298}$ and ΔG_U and smaller values of ln (k_U^{298}) than those obtained with CHARMM27. In particular, for AMBER99SB-disp, the values of $\tau_{1/2}^{298}$ were 25.5% and 70.1% higher at 450 K and 500 K, respectively; the values of ln (k_U^{298}) were 3.4% and 7.8% smaller at 450 K and 500 K, respectively; and the values of ΔG_U were 4.3% and 10.2% higher at 450 K and 500 K, respectively. So, AMBER99SB-disp predicts a higher conformational stability for α_3D than CHARMM27.

Nonetheless, the results obtained with CHARMM27 at 450 K and 500 K converged in a more precise range of stability, 1.68-1.67 kcal/mol; whereas the simulations with AMBER99SB-disp at the same temperatures generated more disperse estimations from 1.75 to 1.84 kcal/mol, within the range of standard error though.

On the other hand, from both Table 5 and Table 6, it is noteworthy that both the estimated half-life of unfolding at 298K and the calculated conformational stability (ΔG_U) exhibited a high dependence on the temperature of simulation. This behavior could be explained due to the greater uncertainties found at the rMD simulations performed at lowest temperatures, in which fewer replicas showed unfolding events with defined half-lives. In this sense, we have to rely more on the values of the half-life and conformational stabilities observed at the highest temperatures, in which it was possible to observe unfolding events for all the simulated replicas. Furthermore, it is remarkable that the stability data obtained at these highest temperatures is very similar to the experimental one (Zhu, et al., 2003) at 298K shown in Table 7.

Table 7. Experimental values and percent error of the predicted unfolding kinetic parameters and conformational stability with CHARMM27 and AMBER99SB-disp force fields at 450 K and 500 K.

	Experimental values[a]	**Percent Error[b]**			
		CHARMM27		**AMBER99SB-disp**	
		450 K	**500 K**	**450 K**	**500 K**
$\boldsymbol{k_U^{298}}$ **(s^{-1})**	2114	-59.05%	-58.37%	-67.37%	-75.54%
ln ($\boldsymbol{k_U^{298}}$)	7.7	-11.66%	-11.45%	-14.63%	-18.39%
$\boldsymbol{\tau_{1/2}^{298}}$ **(ms)**	0.33	144.21%	140.24%	206.44%	308.76%
$\boldsymbol{\Delta G_U}$ **(kcal/mol)**	1.39 ± 0.20	20.62%	20.24%	25.86%	32.51%

[a] Data measured by time-resolved IR spectroscopy (Zhu, et al., 2003).

[b] Percent Error (%E) calculated as: $\%E = \frac{Estimated\ value - Experimental\ value}{Experimental\ value}$

In this context, Table 7 shows that both force fields have overestimated ΔG_U. This is due, in turn, to an overestimation of the half-life of unfolding, which may suggest that the RMSD threshold of 0.6 was higher than needed as delayed the identification of the unfolding events. Nonetheless, it is noteworthy that the force field AMBER99SB-disp yielded more inaccurate ΔG_U values than CHARMM27, even though AMBER99SB-disp is more suitable for simulating unfolded states of proteins. However, it is difficult to draw final conclusions about the accuracy of both force fields with only 10 replicas (or less in some cases).

Thus, according to these results, it is advisable for future work to perform a 2D-RMSD-based clustering with different RMSD thresholds in order to assess whether it is possible to fine tune the measurement of $\tau_{1/2}$ in way that the predicted ΔG_U is more accurate and closer to the reported experimental data. Moreover, it is also suggestable to perform the 2D-RMSD-based clustering analysis with other clustering algorithms like divisive or refinement algorithms (Shao, Tanner, Thompson, & Cheatham, 2007), and/or other linkage methods such as single-linkage, centroid-linkage or complete-linkage clustering, in order to determine which one yields the more accurate results.

4.4 Temperature of unfolding in ladder- and ramp-based simulations

From the rMD simulations whose results are shown above, it was possible to estimate the half-life of unfolding as well as the conformational stability of the protein α_3D. Nonetheless, in that approach, the identification of the right temperature to observe unfolding events required to carry out several simulations at different temperatures above 298 K with ten replicas each. Thus, finding the adequate temperature of simulation of 450 K demanded the investment of considerable time and computer resources.

In this context, trying new MD-based approaches that allow us rMD simulation to optimize the time and the computer resources required to carry out accurate estimations like those here discussed for protein stability and kinetics studies may be actually advantageous.

To explore alternatives to rMD simulations, to assess their accuracy and precision in estimating unfolding kinetics (half-life and *in silico* mid-denaturation temperature) of proteins and also to evaluate the effect of the CHARMM27 and AMBER99SB-disp force fields in the *in silico* temperature of unfolding, ladder- and ramp-based temperature scanning MD simulations were carried out. Again, protein conformations were classified into different groups according to its variability through a 2D-RMSD-based clustering. The clustering plots of all the replicas are shown in the Annex I.

4.4.1 CHARMM27

As an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 5 of the ladder-based temperature scanning MD simulations using the CHARMM27 force field is shown in the Figure 23 (see the clustering plots for all the replicas in the Annex I).

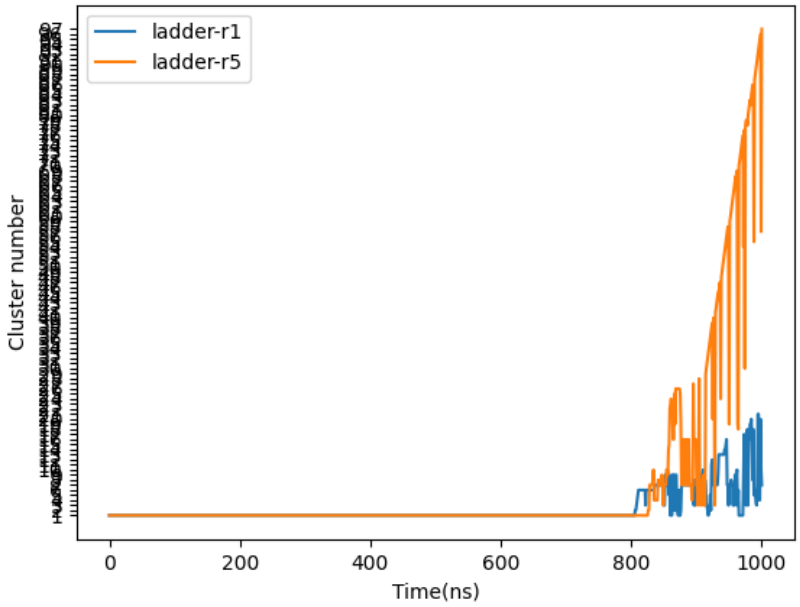

Figure 23. 2D-RMSD-based clustering plot for the replica 5 of the ladder-based temperature scanning MD simulations using CHARMM27.

From Figure 23, it is possible to observe that the protein α_3D unfolded at the 826.67 ns of simulation, which implied a temperature of unfolding of 458 K. This outcome was consistent in most of the simulated ladder-based temperature scanning MD simulations, in which four replicas out of five the protein unfolded at 458 K.

Besides, in order to assess whether CHARMM27 still exhibited the identified α-helical propensity (Gao, et al., 2015), some frames of the trajectory of the replica 5 of the ladder-based temperature scanning MD simulations are shown in the Figure 24.

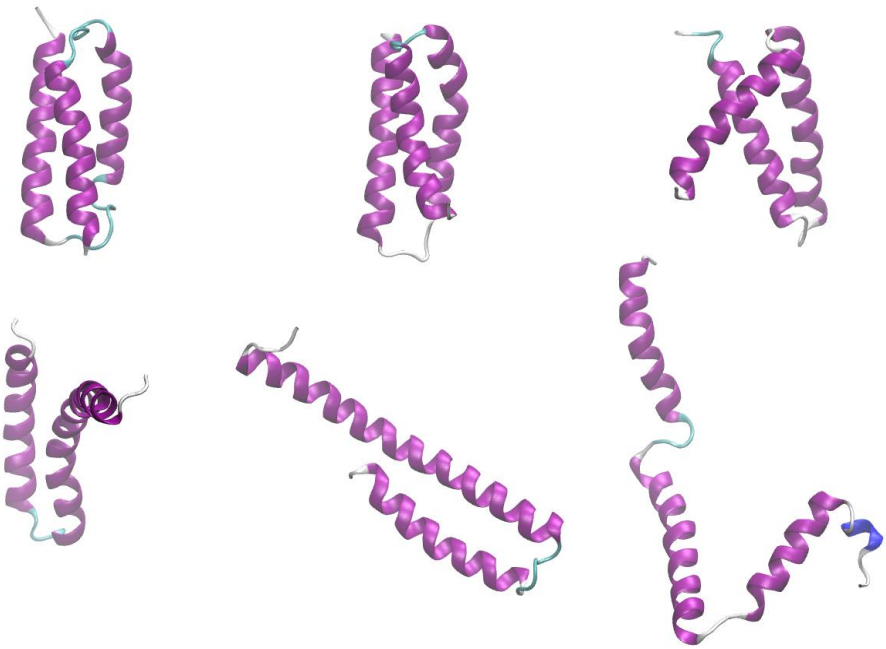

Figure 24. Protein conformations at 1 ns and 298 K (top left), 500 ns and 398 K (top center), 827 ns and 458 K (top right), 833 ns and 458 K (bottom left), 891 ns and 468 K (bottom center) and 1000 ns and 488 K (bottom right) for the replica 5 of the ladder-based temperature scanning MD simulations with CHARMM27.

Thus, from Figure 24 it is possible to observe that CHARMM27 still exhibited a tendency in over stabilizing α-helixes, as discussed above.

On the other hand, as an example of the obtained results, the plot from the 2D-RMSD-based clustering for the replica 4 of the ramp-based temperature scanning MD simulations using the CHARMM27 force field are shown in the Figure 25 (see the clustering plots for all the replicas in the Annex I).

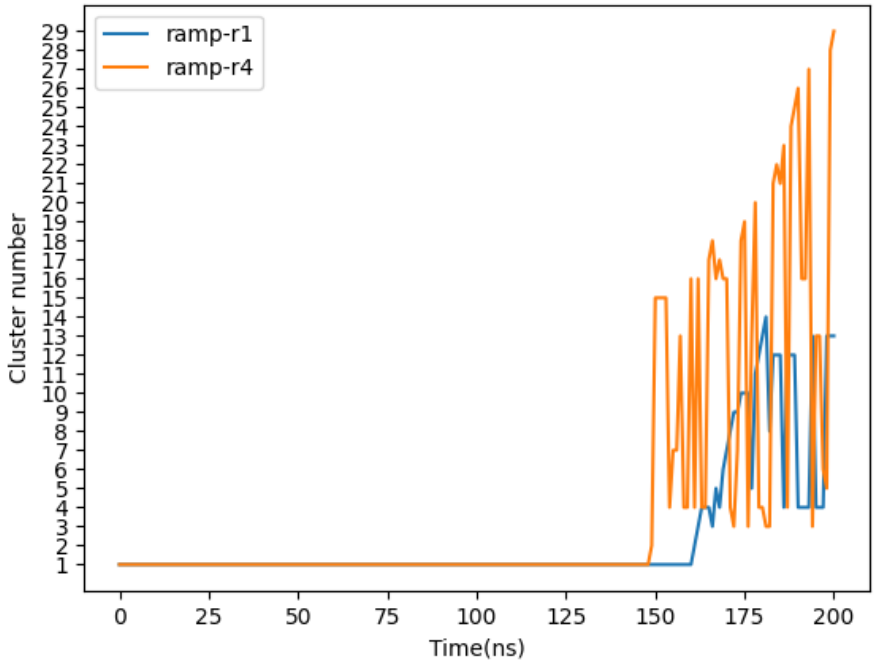

Figure 25. 2D-RMSD-based clustering plot for the replica 4 of the ramp-based temperature scanning MD simulations using CHARMM27.

On the other hand, from Figure 25 it is possible to see that the protein α_3D unfolded at the 191.36 ns of simulation at a temperature of unfolding of 489 K, a temperature superior than that obtained from the ladder-based temperature scanning MD simulations.

Likewise, in order to assess the protein conformation evolution in the ramp-based temperature scanning MD simulations, some frames of the trajectory of the replica 4 are shown in the Figure 26.

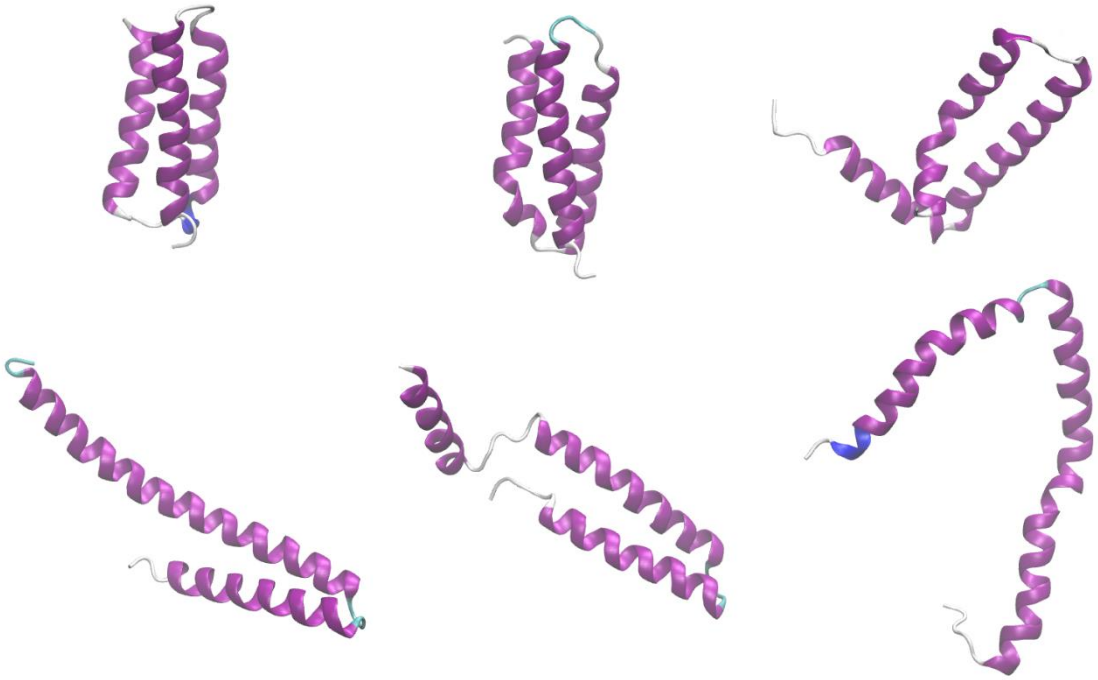

Figure 26. Protein conformations at 1 ns and 299 K (top left), 100 ns and 398 K (top center), 150 ns and 448 K (top right), 160 ns and 458 K (bottom left), 180 ns and 478 K (bottom center) and 200 ns and 198 K (bottom right) for the replica 5 of the ramp-based temperature scanning MD simulations with CHARMM27.

From the Figure 25, it is possible to watch again the tendency of CHARMM27 in over stabilizing the α-helix structures.

A summary of the times and temperatures of unfolding for both ladder- and ramp-based temperature scanning MD simulations with CHARMM27 is shown in the Table 8.

Table 8. Times and temperatures of unfolding for ladder- and ramp-based temperature scanning MD simulations with CHARMM27.

	Ladder-based temperature scanning MD simulations		**Ramp-based temperature scanning MD simulations**	
Replica	**Time of unfolding (ns)**	**Temperature of unfolding (K)**	**Time of unfolding (ns)**	**Temperature of unfolding (K)**
r1	806.66	458	163.64	462
r2	813.33	458	167.73	466
r3	804.44	458	159.55	458
r4	853.33	468	152.73	451
r5	826.67	458	191.36	489
Average	820.89	460	167.00	465
Standard Error	8.99	2.0	6.57	6.6

In view of the results from Table 8, it is possible to observe that the average temperature of unfolding with CHARMM27 was 460 ± 2.0 K for the ladder-based temperature scanning MD

simulations; while it was 465 ± 6.6 K for the ramp-based temperature scanning MD simulations, a slightly increase of 5 K, but still within the standard error.

It is remarkable that these results are consistent with those from the rMD simulations at 450 K and 500 K, in which ten replicas out of ten exhibited unfolding events within the 0.5 μs of simulation; whereas in the rMD simulations at 420 K only four replicas out of ten showed a significant degree of unfolding.

4.4.2 AMBER99SB-disp

As an example of the obtained results, the plot from the 2D-RMSD-based clustering for the replica 3 of the ladder-based temperature scanning MD simulations using the AMBER99SB-disp force field is shown in the Figure 26 (see the clustering plots for all the replicas in the Annex I).

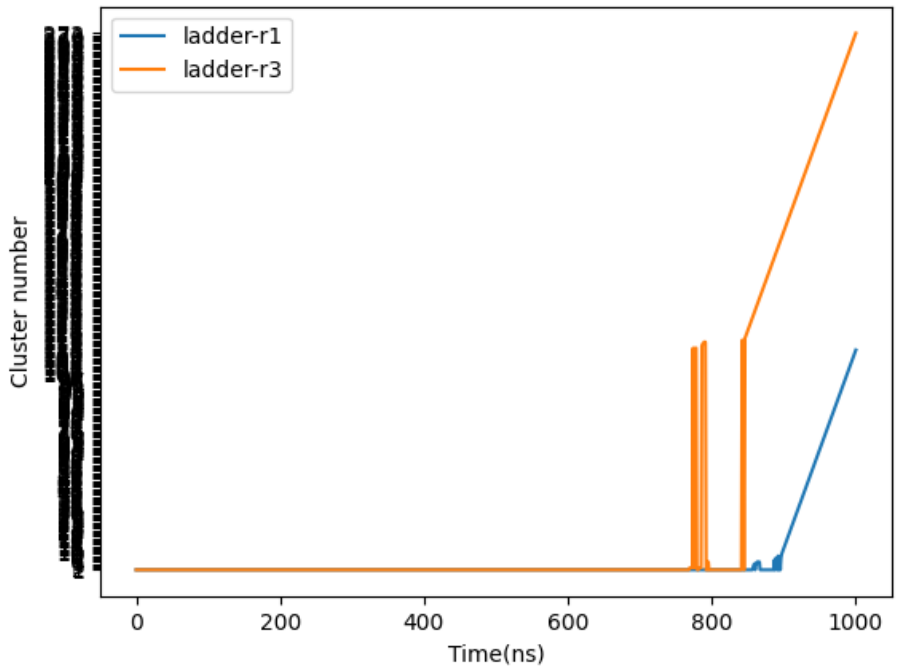

Figure 27. 2D-RMSD-based clustering plot for the replica 3 of the ladder-based temperature scanning MD simulations using AMBER99SB-disp.

From Figure 27, it is possible to observe that the protein α_3D unfolded at the 842.22 ns of simulation, which implied a temperature of unfolding of 458 K.

Likewise, in order to assess whether AMBER99SB-disp simulated unfolding states properly by using a ladder-based temperature scanning MD simulations, some frames of the trajectory of the replica 3 are show in the Figure 28.

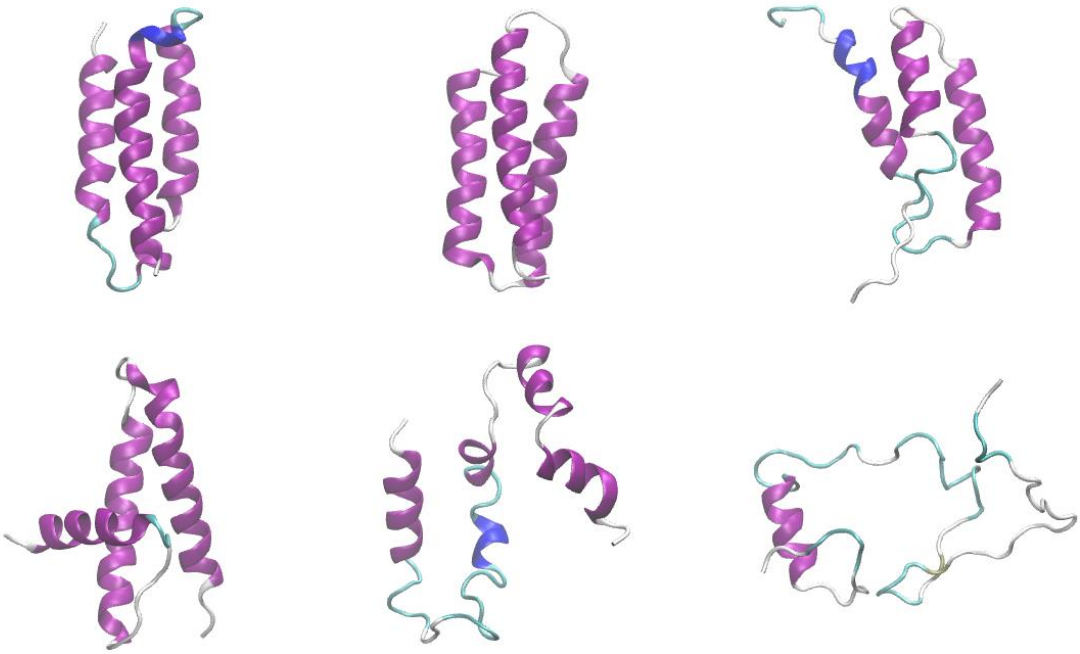

Figure 28. Protein conformations at 1 ns and 298 K (top left), 500 ns and 398 K (top center), 773 ns and 448 K (top right), 831 ns and 458 K (bottom left), 848 ns and 458 K (bottom center) and 1000 ns and 488 K (bottom right) for the replica 3 of the ladder-based temperature scanning MD simulations with AMBER99SB-disp.

Therefore, from Figure 28 it is possible to declare that AMBER99SB-disp in ladder-based temperature scanning MD simulations models properly the unfolded states. Furthermore, it is noteworthy how the protein α_3D at 773 ns and 448 K started to lose its α-helix structure but at 831 ns and 458 K recovered part of the same. Nevertheless, the protein is clearly unfolded at 848 ns and 458 K.

On the other hand, as an example of the results obtained, the plot from the 2D-RMSD-based clustering for the replica 1 of the ramp-based temperature scanning MD simulations using the AMBER99SB-disp force field is shown in the Figure 29 (see the clustering plots for all the replicas in the Annex I).

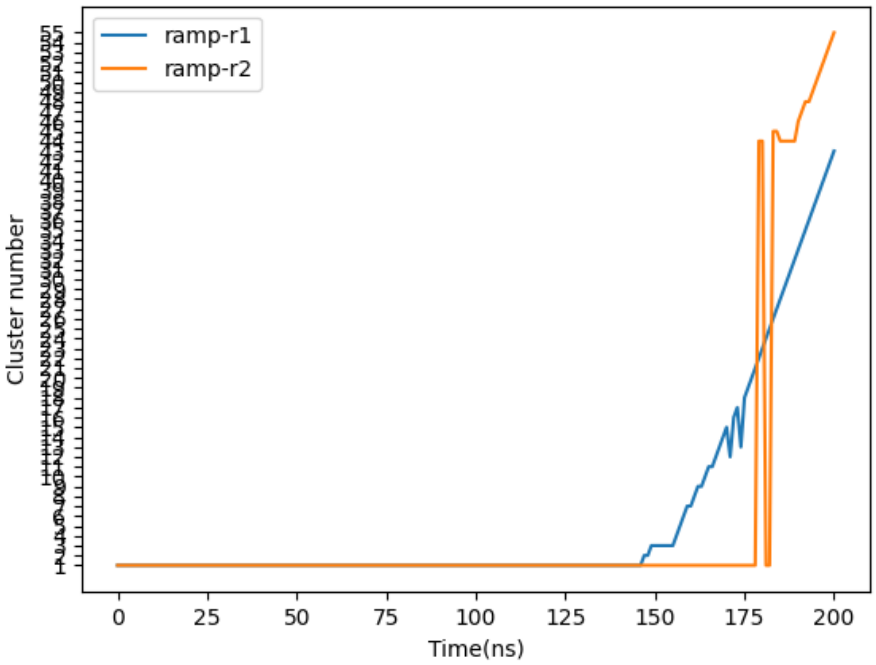

Figure 29. 2D-RMSD-based clustering plot for the replica 1 of the ramp-based temperature scanning MD simulations using AMBER99SB-disp.

From Figure 29 it is possible to watch that the protein α_3D unfolded at the 149.09 ns of simulation at a temperature of unfolding of 447 K.

Similarly, in order to assess the protein conformation evolution in the ramp-based temperature scanning MD simulations, some frames of the trajectory of the replica 4 are shown in the Figure 30.

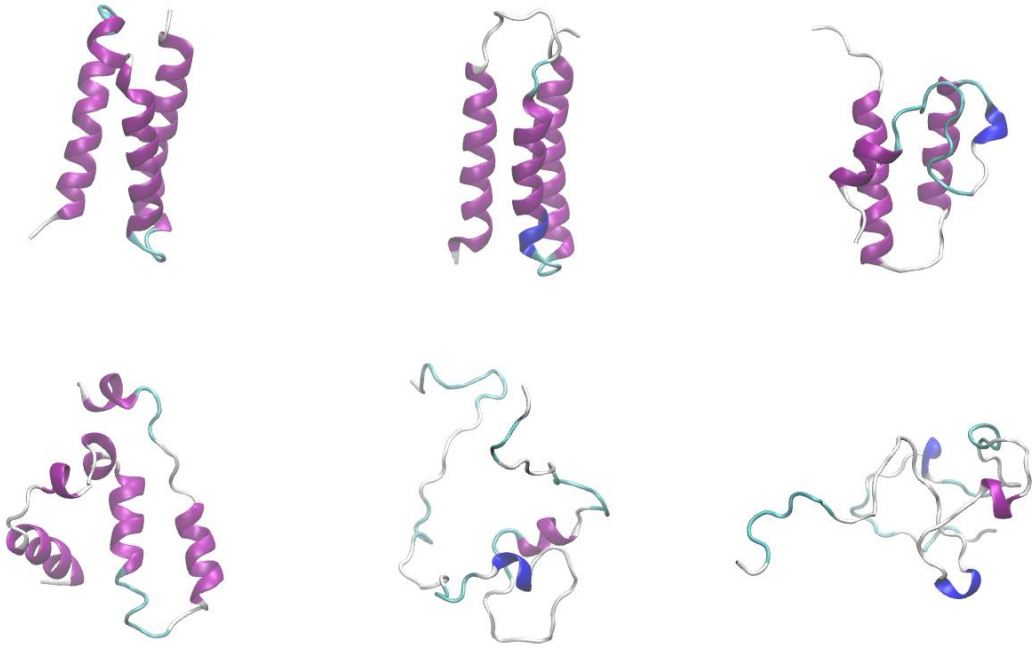

Figure 30. Protein conformations at 1 ns and 299 K (top left), 120 ns and 418 K (top center), 150 ns and 448 K (top right), 160 ns and 458 K (bottom left), 180 ns and 478 K (bottom center) and 200 ns and 498 K (bottom right) for the replica 5 of the ramp-based temperature scanning MD simulations with AMBER99SB-disp.

From Figure 30, it is possible to state that the ramp-based temperature scanning MD simulations with AMBER99SB-disp are also successful in simulating the unfolding states of protein.

A summary of the times and temperatures of unfolding for both ladder- and ramp-based temperature scanning MD simulations with AMBER99SB-disp is shown in the Table 9.

Table 9. Times and temperatures of unfolding for ladder- and ramp-based temperature scanning MD simulations with AMBER99SB-disp.

	Ladder-based temperature scanning MD simulations		**Ramp-based temperature scanning MD simulations**	
Replica	**Time of unfolding (ns)**	**Temperature of unfolding (K)**	**Time of unfolding (ns)**	**Temperature of unfolding (K)**
r1	893.34	468	149.09	447
r2	880.01	468	185.91	484
r3	842.22	458	180.10	478
r4	875.56	468	>200.00	>498
r5	851.11	468	194.55	493
Average	868.45	466	177.41*	480*
Standard Error	9.47	2.0	9.9*	8.9*

* r4 was not included in the computation of the average and the standard error for the ramp-based temperature scanning MD simulations.

According to Table 9, the average temperature of unfolding with AMBER99SB-disp was 466 ± 2.0 K for the ladder-based temperature scanning MD simulations; and 480 ± 8.9 K for the ramp-based temperature scanning MD simulations. And, in this case, unlike the simulations with CHARMM27, the results from both approaches were not within the standard error. Besides, in the ramp-based temperature scanning MD simulations, it was not possible to observe unfolding events for all the replicas as the replica 4 did not unfolded during the time span of the simulation (see Annex I).

4.5 Comparison of MD-based approaches and fine tuning of the analysis method

rMD simulations are time and computing-demanding due to the timescales required to observe unfolding events and due to the necessity of performing several replicas at different temperatures. For that reason, in order to optimize resources and fine tune the present analysis method for assessing protein stability and unfolding kinetics, temperature ladder- and ramp-based temperature scanning MD simulations were performed using the CHARMM27 and AMBER99SB-disp force fields.

A comparison of the folded and unfolded states from the rMD, ladder- and ramp-based temperature scanning MD simulations using both force fields is shown in the Figure 31.

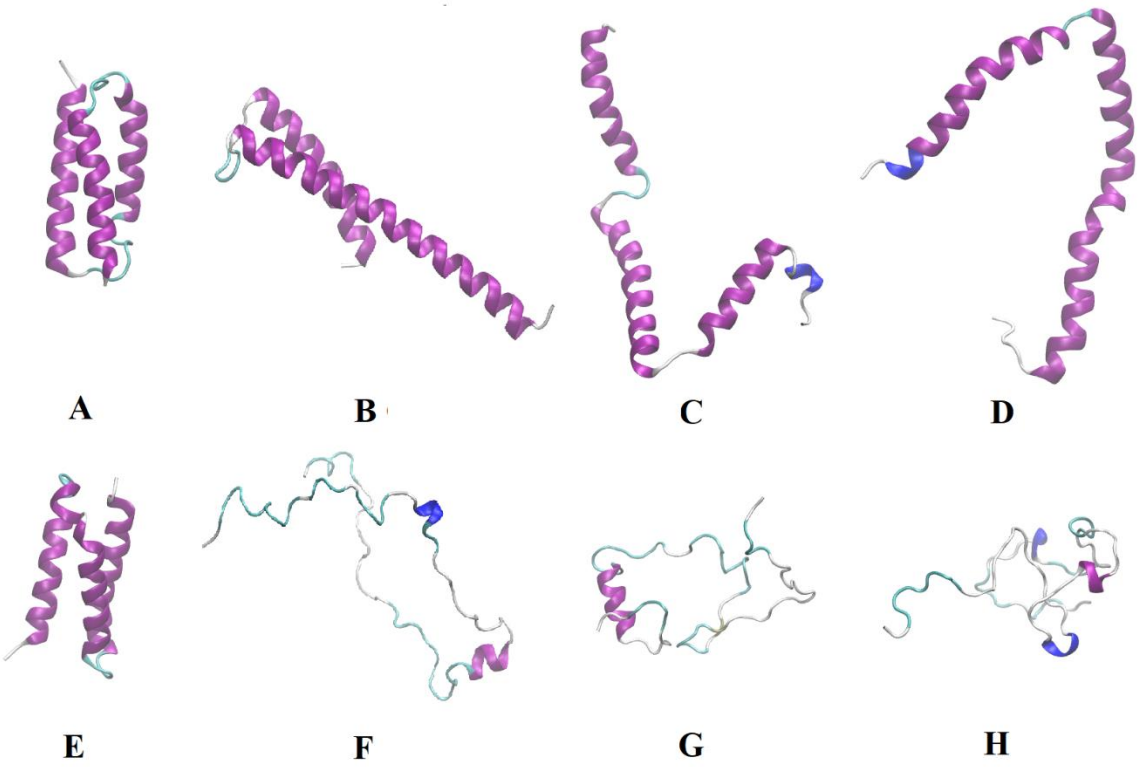

Figure 31. Protein conformations with CHARMM27: (**A**) Folded state, (**B**) Unfolded state with rMD simulation, (**C**) Unfolded state with ladder-based temperature scanning MD simulations, (**D**) Unfolded state with ramp-based temperature scanning MD simulations; and with AMBER99SB-disp: (**E**) Folded state, (**F**) Unfolded state with rMD simulation, (**G**) Unfolded state with ladder-based temperature scanning MD simulations, (**H**) Unfolded state with ramp-based temperature scanning MD simulations.

From Figure 31, it is possible to observe that no major differences exist between the unfolded states from the three types of simulations, for both force fields. In this sense, these results suggest that ladder- and ramp-based temperature scanning MD simulations are as suitable as rMD simulations for sampling the conformation of proteins and observing unfolding events.

On the other hand, in order to further compare the results obtained from rMD, temperature ladder- and ramp-based temperature scanning MD simulations with both force fields, the Table 10 shows a summary of the average *in silico* T_m and their standard errors.

Table 10. Average *in silico* T_m for rMD, ladder- and ramp-based temperature scanning MD simulations with CHARMM27 and AMBER99SB-disp.

CHARMM27			**AMBER99SB-disp**		
rMD	**ladder-based**	**ramp-based**	**rMD**	**ladder-based**	**ramp-based**
<450 ± 0.0 K	460 ± 2.0 K	465 ± 6.6 K	<450 ± 0.0 K	466 ± 2.0 K	480 ± 8.9 K

From the rMD simulations, it has been confirmed that ten replicas out of ten unfolds within the time span of the simulation (0.5 μs) at 450 K. Thus, from Table 10, it is possible to verify that both ladder- and ramp-based temperature scanning MD simulations are accurate in estimating the *in silico* mid-denaturation temperature of the model protein α_3D, with the estimations obtained with CHARMM27 being slightly more accurate than those obtained with AMBER99SB-disp.

On the other hand, it is also noticeable that the ramp-based temperature scanning MD simulations required higher temperatures to observe unfolding events in regard to the ladder- ones, for both force fields. However, the estimations of the average temperature of unfolding from ladder- and ramp-based temperature scanning MD simulations with CHARMM27 are within the range of standard error, whereas the estimations with AMBER99SB-disp are not.

It is also notably that the ramp-based temperature scanning MD simulations yielded slightly higher standard errors than those obtained from the ladder-based temperature scanning MD simulations. In this sense, the standard errors of the ramp-based temperature scanning MD simulations with CHARMM27 and AMBER99SB-disp were 6.6 K and 8.9 K, respectively; while the standard error of the ladder-based temperature scanning MD simulations was 2.0 K for both force fields. Thus, if one type of simulation had to be used to determine the temperature of unfolding before running rMD simulations, the results from this study suggest using ladder- instead of ramps-based temperature scanning MD simulations.

Furthermore, the estimations of the average temperatures of unfolding obtained with the ladder-based temperature scanning MD simulations with both force fields, 460 ± 2.0 K for CHARMM27 and 466 ± 2.0 K for AMBER99SB-disp were closer to the tested temperature of 450 K in rMD simulations, which proved adequate to be used for the estimations of the unfolding kinetic parameters and conformational stability. On the contrary, the extreme case can be seen with the ramp-based temperature scanning MD simulations with AMBER99SB-disp, with an average temperature of unfolding of 480 ± 8.9 K, much closer to the tested temperature of 500 K in rMD simulations, at which the quantification of the conformational stability was more difficult due to the extremely short half-life of unfolding. Thus, also from this point of view, the results from this study suggested that ladder- are preferable over ramps-based temperature scanning MD simulations.

On the other hand, these findings suggest that the simulations performed with AMBER99SB-disp require slightly higher temperatures to give place to unfolding events as the average temperature of unfolding in the ladder-based temperature scanning MD simulations with AMBER99SB-disp was 466 ± 2.0 K; whereas this value was 460 ± 2.0 K with CHARMM27. This behavior is consistent with the results from the ramp-based temperature scanning MD simulations as well, where even one replica with simulated with AMBER99SB-disp did not show any unfolding event whatsoever.

In this sense, this outcome from the ladder- and ramp-based temperature scanning MD simulations agrees with the results from the rMD simulations, which showed that AMBER99SB-disp seems to slightly over stabilize the whole protein whereas CHARMM27 seems to over stabilize the α-helices, as described above (see Table 7). Thus, paradoxically, even though AMBER99SB-disp is a force field more suited to simulate both folded and unfolded states, it seems that CHARMM27 could be more accurate in estimating half-life of unfolding and conformational stability through the previously described empirical model (Galano-Frutos, García-Cebollada, & Sancho, 2019).

In this context, based on the results of the present study, to perform stability and kinetics unfolding analyses on proteins ladder-based temperature scanning MD simulations could be a solution to optimize our methodology in terms of computational resources with the additional advantageous of increasing the accuracy in the estimation of the T_m. Importantly, it should be noticed that these results have been obtained with only one model protein, and that some other proteins with different sizes should also be approached in order to extract more definitive clues and conclusions on the MD-based methods, ladder and ramp, here addressed.

Even though it is true that the estimation of the half-life of unfolding should have been carried out be manually with ImageJ, the 2D-RMSD-based clustering allowed to simplify the interpretation of the fluctuations in the protein conformations observed during the simulation, easing the measurement of the half-life of unfolding by comparing the clustering profile of a given protein to that of one reference properly selected (folded stable protein).

In this context, a research line currently being developed in the research group is the creation of algorithms to automatically detect and account relevant unfolding events in MD trajectories. It should be said that this task could more easily accomplished with 2D-RMSD-based clustering performed on trajectories generated with AMBER99SB-disp, where the unfolding clustering profiles show steadily ascendant behaviours indicating a more extended unfolding process (more unstructured unfolded state) in the protein being simulated.

As another recommendation for future work, the question remains whether it would be also possible to envision a model for accurately estimating the conformational stability of proteins from a ladder- or ramp-based temperature scanning MD simulation method, as done for rMD through Equation 8. With the rMD approach, it is possible to estimate the half-life of unfolding and the ΔG_U because simulations are performed in equilibrium and at constant temperature. Nonetheless, this is not what happened in ladder- or ramp-based MD simulations, so that, new ideas about how to estimate the half-lives and ΔG_U when these methods are implemented should still come out.

5. CONCLUSIONS

In the present study, through the realization of multiple relaxation MD simulations at a broad range of temperatures with the two force fields CHARMM27 and AMBER99SB-disp, it was possible to verify the *in silico* temperature dependence for estimating half-life on the model protein α_3D. This behavior could have been due to the greater uncertainties found at the rMD simulations performed at lower temperatures, in which fewer replicas showed unfolding events with defined half-lives. In this sense, it was necessary to rely more on the values of half-life and conformational stabilities observed at higher temperatures, in which it was possible to observe unfolding events for all the simulated replicas. Furthermore, it was remarkable that the conformational stability estimations obtained at these higher temperatures were very similar to the experimental data.

Moreover, it was possible to verify that, in comparison with rMD simulations, ladder- and ramp-based temperature scanning MD simulations were accurate in estimating the *in silico* mid-denaturation temperature of the model protein α_3D, with the estimations obtained with CHARMM27 being slightly more accurate than those obtained with AMBER99SB-disp. In addition, it was also found that the estimations of ladder-based MD simulations were also slightly more accurate than those obtained with ramp-based MD simulations. Nevertheless, remains pending the issue of accurately estimating the half-life and the conformational stability of proteins from these kind of MD methods.

Finally, the performance of the force fields CHARMM27 and AMBER99SB-disp was assessed in simulating structured and unstructured states in rMD, ladder- and ramp-based MD simulations, observing that CHARMM27 had a propensity to over-stabilize α-helix structures in unfolded states; while AMBER99SB-disp was more adequate in simulating those disordered conformations. Nonetheless, even though the CHARMM27 force field exhibited an α-helical propensity, it was able to yield more accurate estimations of the conformational stability than AMBER99SB-disp in rMD, which seemed to slightly over stabilize the overall protein structure.

6. BIBLIOGRAPHY

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1-2*, 19-25. doi:10.1016/j.softx.2015.06.001

Arnittali, M., Rissanou, A. N., & Harmandaris, V. (2019). Structure Of Biomolecules Through Molecular Dynamics Simulations. *Procedia Computer Science, 156*, 69-78. doi:10.1016/j.procs.2019.08.181

Berendsen, H. J., Postma, J. P., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics, 81*(8), 3684-3690. doi:10.1063/1.448118

Bilsel, O., & Matthews, C. R. (2000). Barriers in protein folding reactions. *Advances in Protein Chemistry, 53*, 153-207. doi:10.1016/S0065-3233(00)53004-6

Blasco-Puyuelo, S. (2019). *Evaluation of the mutome associated to Rett syndrome.* Master Thesis, University of Zaragoza, Zaragoza. Retrieved from https://zaguan.unizar.es/record/86652/files/TAZ-TFM-2019-905.pdf

Brooks, B. R., Brooks III, C. L., MacKerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., . . . Karplus, M. (2009). CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry, 30*(10), 1545-1614. doi:10.1002/jcc.21287

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics, 126*(1), 1-7. doi:10.1063/1.2408420

Castillo-Cano, V. (2012). *Using Small Globular Proteins to Study Folding Stability and Aggregation.* Doctoral Thesis, Universitat Autònoma de Barcelona, Barcelona. Retrieved from https://www.tesisenred.net/bitstream/handle/10803/107824/vcc1de1.pdf?sequence=1&isAllowed=y

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. F., . . . Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society, 117*(19), 5179-5197. doi:10.1021/ja00124a002

Daggett, V., & Levitt, M. (1993). Protein Unfolding Pathways Explored Through Molecular Dynamics Simulations. *Journal of Molecular Biology, 232*(2), 600-619. doi:10.1006/jmbi.1993.1414

Day, A. (1995). *The Source of Stability in Proteins.* Retrieved from Birkbeck, University of London: http://www.cryst.bbk.ac.uk/PPS2/projects/day/TDayDiss/index.html

Day, R., Bennion, B. J., Ham, S., & Daggett, V. (2002). Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding. *Journal of Molecular Biology, 322*(1), 189-203. doi:10.1016/S0022-2836(02)00672-1

Deller, M. C., Kong, L., & Rupp, B. (2016). Protein stability: a crystallographer's perspective. *Acta Crystallographica, F72*, 72-95. doi:10.1107/S2053230X15024619

Denschlag, R., Lingenheil, M., & Tavan, P. (2009). Optimal temperature ladders in replica exchange simulations. *Chemical Physics Letters, 473*(1), 193-195. doi:10.1016/j.cplett.2009.03.053

Dill, K. A., & MacCallum, J. L. (2012). The Protein-Folding Problem, 50 Years On. *Science, 338*, 1042-1046. doi:10.1126/science.1219021

Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R., & Meiler, J. (2009). Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *Journal of Molecular Modeling, 15*(9), 1093–1108. doi:10.1007/s00894-009-0454-9

Freddolino, P. L., Liu, F., Gruebele, M., & Schulten, K. (2008). Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophysical Journal, 94*(10), 75-77. doi:10.1529/biophysj.108.131565

Galano-Frutos, J. J., & Sancho, J. (2019). Accurate Calculation of Barnase and SNase Folding Energetics Using Short Molecular Dynamics Simulations and an Atomistic Model of the Unfolded Ensemble: Evaluation of Force Fields and Water Models. *Journal of Chemical Information and Modeling, 59*, 4350-4360. doi:10.1021/acs.jcim.9b00430

Galano-Frutos, J. J., García-Cebollada, H., & Sancho, J. (2019). Molecular dynamics simulations for genetic interpretation in protein coding regions: where we are, where to go and when. *Briefings in Bioinformatics*(bbz146), 1–17. doi:10.1093/bib/bbz146

Gao, Y., Li, Y., Mou, L., Lin, B., Zhang, J. Z., & Mei, Y. (2015). Correct folding of an α-helix and a β-hairpin using a polarized 2D torsional potential. *Scientific Reports, 5*, 10359. doi:10.1038/srep10359

Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry, 8*, 37-47. doi:10.2147/AABC.S70333

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics, 14*(1), 33-38. doi:10.1016/0263-7855(96)00018-5

Huyghues-Despointes, B. M., Pace, C. N., Englander, W. S., & Scholtz, J. M. (2001). Measuring the Conformational Stability of a Protein by Hydrogen Exchange. In K. P. Murphy (Ed.), *Methods in Molecular Biology* (Vol. 168, pp. 69-92). Humana Press. doi:10.1385/1-59259-193-0:069

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics, 79*(2), 926-935. doi:10.1063/1.445869

Katiyar, R. S., & Jha, P. K. (2018). Molecular simulations in drug delivery: Opportunities and challenges. *WIREs Computational Molecular Science, e1358*, 1-18. doi:10.1002/wcms.1358

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science, 220*(4598), 671-680. doi:10.1126/science.220.4598.671

Ku, T., Lu, P., Chan, C., Wang, T., Lai, S., Lyu, P., & Hsiao, N. (2009). Predicting melting temperature directly from protein sequences. *Computational Biology and Chemistry, 33*(6), 445-450. doi:10.1016/j.compbiolchem.2009.10.002

Kumari, I., & Akhter, Y. (2017). Molecular Dynamics simulations, challenges and opportunities: a biologist's prospective. *Current Protein and Peptide Science, 18*(11), 1163-1179. doi:10.2174/1389203718666170622074741

Lindahl, E., Abraham, M. J., Hess, B., & van der Spoel, D. (2020). *GROMACS Documentation Release 2020.4.* doi:10.5281/zenodo.4054996

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Systematic Validation of Protein Force Fields against Experimental Data. *PLoSONE, 7*(2), e32131. doi:10.1371/journal.pone.0032131

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., . . . Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B, 102*(18), 3586-35616. doi:10.1021/jp973084f

Martínez, L. (2015). Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLoS One, 10*(3), e0119264. doi:10.1371/journal.pone.0119264

Mathews, C. K., van Holde, K. E., & Ahern, K. G. (2002). *Bioquímica* (Third ed.). Madrid: Pearson.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature, 267*(5612), 585-90. doi:10.1038/267585a0

Miotto, M., Olimpieri, P. P., Di Rienzo, L., Ambrosetti, F., Corsi, P., Lepore, R., . . . Milanetti, E. (2019). Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics, 35*(15), 2569–2577. doi:10.1093/bioinformatics/bty1011

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics, 52*(12), 7182-7190. doi:10.1063/1.328693

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-- a visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605-1612. doi:10.1002/jcc.20084

Piana, S., Klepeis, J. L., & Shaw, D. E. (2014). Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current Opinion in Structural Biology, 24*, 98-105. doi:10.1016/j.sbi.2013.12.006

Pinak, M. (2006). Enzymatic recognition of radiation-produced oxidative DNA lesion. Molecular dynamics approach. In E. B. Starikov, J. P. Lewis, & S. Tanaka (Eds.), *Modern Methods for Theoretical Physical Chemistry of Biopolymers* (pp. 191-210). Elsevier Science. doi:10.1016/B978-044452220-7/50074-5

Ramakrishnan, V., Srinivasan, S. P., Salem, S. M., Matthews, S. J., Colón, W., Zaki, M., & Bystroff, C. (2012). Geofold: topology-based protein unfolding pathways capture the effects of engineered disulfides on kinetic stability. *Proteins, 80*(3), 920-934. doi:10.1002/prot.23249

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 å? *Folding and Design, 3*(2), 141-147. doi:10.1016/S1359-0278(98)00019-4

Reyes, P. N., Valencia, F. J., Vega, H., Ruestes, C., Rogan, J., Valdivia, J. A., & Kiwi, M. (2018). On the stability of hollow nanoparticles and the simulation temperature ramp. *Inorganic Chemistry Frontiers, 5*, 1139-1144. doi:10.1039/C7QI00822H

Robustelli, P., Piana, S., & Shaw, D. E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America, 115*(21), E4758-E4766. doi:10.1073/pnas.1800690115

Rocco, A. G., Mollica, L., Ricchiuto, P., Baptista, A. M., Gianazza, E., & Eberini, I. (2008). Characterization of the Protein Unfolding Processes Induced by Urea and Temperature. *Biophysical Journal, 94*(6), 2241-2251. doi:10.1529/biophysj.107.115535

Saboury, A. A., & Moosavi-Movahedi, A. A. (1995). Derivation of the thermodynamic parameters involved in the elucidation of protein thermal profiles. *Biochemical Education, 23*(3), 164-167. doi:10.1016/0307-4412(95)00049-9

Sánchez-Ruiz, J. M. (2010). Protein kinetic stability. *Biophysical Chemistry, 148*(1-3), 1-15. doi:10.1016/j.bpc.2010.02.004

Sancho, J. (2013). The stability of 2-state, 3-state and more-state proteins from simple spectroscopic techniques... plus the structure of the equilibrium intermediates at the same time. *Archives of Biochemistry and Biophysics, 531*, 4-13. doi:10.1016/j.abb.2012.10.014

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods, 9*(7), 671–675. doi:doi:10.1038/nmeth.2089

Shao, J., Tanner, S. W., Thompson, N., & Cheatham, T. E. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation, 3*(6), 2312-2334. doi:10.1021/ct700119m

Stambulchik, E. (2008). *Grace*. Retrieved from WIS Plasma Laboratory: https://plasma-gate.weizmann.ac.il/Grace/

Tang, X. C., & Pikal, M. J. (2005). Measurement of the kinetics of protein unfolding in viscous systems and implications for protein stability in freeze-drying. *Pharmaceutical Research, 22*(7), 1176-1185. doi:10.1007/s11095-005-6036-3

Tayyab, S., Siddiqui, M. U., & Ahmad, N. (1995). Experimental Determination of the Free Energy of Unfolding of Proteins. *Biochemical education, 23*(3), 162-164. doi:10.1016/0307-4412(95)00005-N

Tokuriki, N., & Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology, 19*(5), 596-604. doi:10.1016/j.sbi.2009.08.003

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual.* Scotts Valley: CreateSpace.

Vlachakis, D., Bencurova, E., Papangelopoulos, N., & Kossida, S. (2014). Chapter Seven - Current State-of-the-Art Molecular Dynamics Methods and Applications. In R. Donev (Ed.), *Advances in Protein Chemistry and Structural Biology* (Vol. 94, pp. 269-313). Academic Press. doi:10.1016/B978-0-12-800168-4.00007-X

Walsh, S. T., Cheng, H., Bryson, J. W., Roder, H., & DeGrado, W. F. (1999). Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proceedings of the National Academy of Sciences of the United States of America, 96*(10), 5486-5491. doi:10.1073/pnas.96.10.5486

Wang, L., Alibay, I., Meli, R., Torchala, M., Zhuang, Y., Gowers, R. J., & Beckstein, O. (2020, June 26). *Calculating the pairwise RMSD of a trajectory*. Retrieved from MDAnalysis User Guide: https://userguide.mdanalysis.org/stable/examples/analysis/alignment_and_rms/pairwise_rmsd.html

Wittung-Stafshede, P. (2004). Slow unfolding explains high stability of thermostable ferredoxins: common mechanism governing thermostability? *Biochimica et Biophysica Acta, 1700*(1), 1-4. doi:10.1016/j.bbapap.2004.04.002

Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics, 26*(7), 889–895. doi:10.1093/bioinformatics/btq066

Zhang, D., & Lazim, R. (2017). Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports, 7*(44651). doi:10.1038/srep44651

Zhang, Y., & Luo, L. (2011). The dynamical contact order: Protein folding rate parameters based on quantum conformational transitions. *Science China Life Sciences, 54*(4), 386–392. doi:10.1007/s11427-011-4158-x

Zhang, Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins, 57*(4), 702-710. doi:10.1002/prot.20264

Zhu, Y., Alonso, D. O., Maki, K., Huang, C., Lahr, S. J., Daggett, V., . . . Gai, F. (2003). Ultrafast folding of alpha3D: a de novo designed three-helix bundle protein. *Proceedings of the National Academy of Sciences of the United States of America, 100*(26), 15486-15491. doi:10.1073/pnas.2136623100

Zwanzig, R. (1997). Two-state models of protein folding kinetics. *Proceedings of the National Academy of Sciences of the United States of America, 94*(1), 148–150. doi:10.1073/pnas.94.1.148