

# PPP - Potestio Prions Project

Students: *T. Dalle Nogare, A. Guadagnin Pattaro*

---

Course: *Multi-scale methods in soft matter physics [145889]* – Teacher: *Prof. R. Potestio*  
Date: *October 2, 2022*

## 1 Abstract

The structural characterization of prions represents a challenging task, both from an experimental and computational point of view. The aim of this project is to investigate the change in secondary structure of two "hand-made" chimeras, called CHI1 and CHI2, exploiting molecular dynamics simulations.

Firstly, an all-atom simulation of 1 ns was performed for both chimeras. Subsequently, two different coarse-grained methods, namely Gō model and Martini 3 were used to coarse grain the initial structures and perform longer MD simulations. An analysis exploiting structural parameters such as the radius of gyration, RMSD and RMSF was performed in order to understand the differences in behaviour between the two structures, but also the effect of temperature on the process of unfolding.

Overall, no particularly significant results were found concerning the folding process itself. However, some considerations were made on the role of coarse graining in the study of the protein folding process.

## 2 Introduction

### 2.1 Prions

In recent decades a series of diseases named transmissible spongiform encephalopathies (TSEs) have been discovered and studied. TSEs of humans include Creutzfeldt–Jakob disease and the fatal familial insomnia. Not only humans, but also non-human mammals suffer of TSEs, such as scrapie in sheep and bovine spongiform encephalopathy (BSE) in cattle. All of these neurodegenerative disorders are characterized by a spongiform changes in brain tissue associated with neuronal loss, as shown in Fig. 1, and a failure to induce inflammatory response. Moreover, since no effective treatment has been found yet, they are always fatal. By now, from the several studies conducted with the aim to understand the nature behind these diseases, it was found that the infectious agent in TSEs is the so called prion protein.

Prions are self-replicative proteins which have the ability to transfer their misfolded state onto the surrounding prions which are in their native conformation. The word *prion* was first coined by Stanley B. Prusiner in 1982 who was the first person to study its conformational changes. His studies have suggested that several among the fatal neurodegenerative diseases stem from the modification of the prion protein from its native

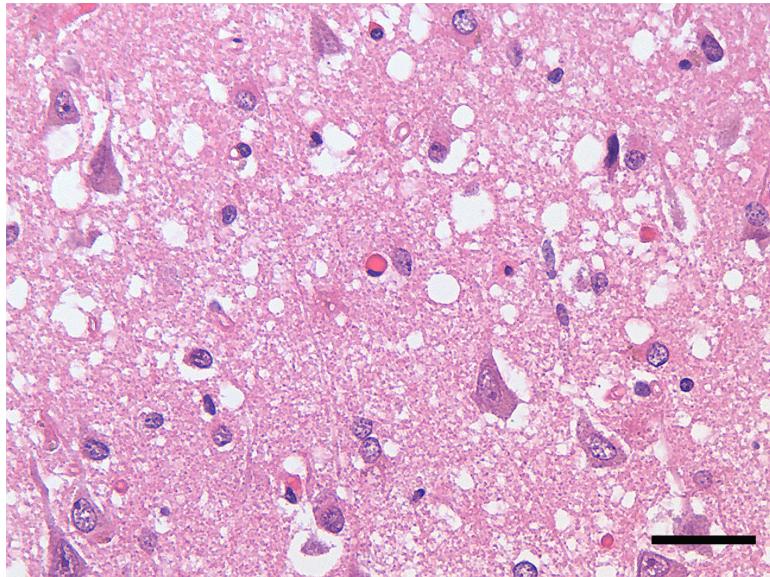


Figure 1: Photo taken with a light microscope showing spongiform degeneration in the cerebral cortex of a patient who had died of Creutzfeldt-Jakob disease (CJD). Image from Prion - Wikipedia

conformation  $\text{PrP}^c$  (cellular) into the modified  $\text{PrP}^{sc}$  (scrapie). The normal PrP is converted into  $\text{PrP}^{sc}$  through a series of changes in its secondary structure, where a portion of  $\alpha$ -helix is refolded into  $\beta$ -sheets [1]. From Fourier-transform infrared (FTIR) spectroscopy, a content of 42% of  $\alpha$ -helices and 3% of  $\beta$ -sheets in  $\text{PrP}^c$  was measured. In contrast, when these measurements were performed on  $\text{PrP}^{sc}$ , approximately 43% of  $\alpha$ -helix and 30% of  $\beta$ -sheet were found. Moreover, the amount of  $\beta$ -sheets increased even when the N-terminally truncated protein PrP 27-30 was taken into account. [2] Unfortunately, still little is known about the structure of the prion protein in its misfolded state. As a consequence, the development of computational models together with molecular dynamics simulations are useful tools if we want to gain knowledge about the behaviour of this protein. As an example, the all atom model of the prion that was developed by Spagnolli *et al* [3] is useful to investigate mechanisms involved in protein regulation and to develop drugs. Since atomistic models require lots of computational effort, in order to investigate the secondary structure of the prion protein, coarse graining has been used.

## 2.2 Gō Model coarse graining

Coarse graining (CG) is a way of performing molecular dynamics simulations of macromolecules that is used mainly for three reasons: to reduce the number of degrees of freedom of the system, to focus on selected types of interactions, and to study the dynamics of the system for a longer time at reduced computational power; ultimately the goal is the reduction of the overall complexity of the problem. The coarse graining process usually starts from a high-resolution configuration space (the all-atom configuration) on which a mapping function that generates the new low-resolution structure is applied, after a parametrization step. The mapping function depends on the type of macromolecule we want to study and on the type of parametrization we wish to implement. An implicit assumption would be that the CG representation must be consistent with the all-atom representation. [4]

One type of model that is used for protein coarse-graining is the Gō model. This model was developed in the 1970s and it is based on the replacement of the nonbonded interactions of the protein (Van Der Waals, electrostatics) with an energy function parametrizing attractive or repulsive native-state contacts that bias, the protein towards its native structure [5], the process is summarized in Fig. 2. Another important feature of Gō-like models is the funnel-shaped energy landscape of the coarse-grained structure; this is useful since it resembles the energy landscape of a generic protein folding process [6, 7].

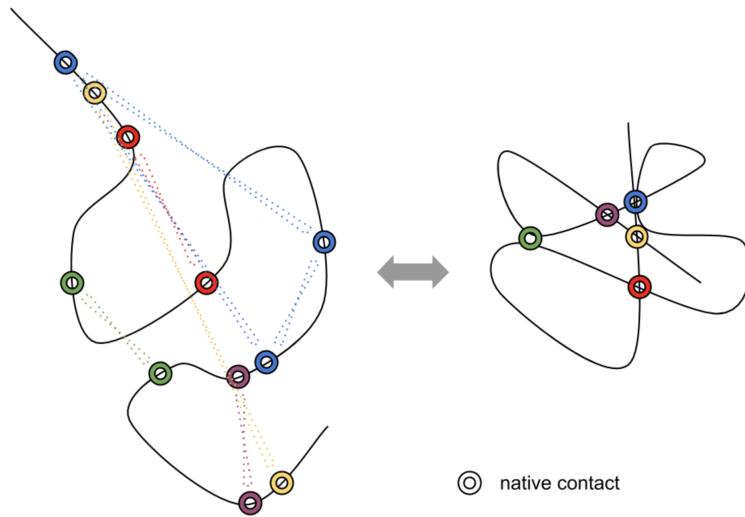


Figure 2: Graphical representation of how Gō model approximates interactions between residues to allow protein folding.

Gō model has been widely used to study properties of proteins and protein folding but the results are found to be approximate. It has been shown by performing a native-state analysis and non-native interactions analysis that Gō model generates a potential energy well that is deeper for the native structure of the transition state of a protein than the experimental values [8]. This means that Gō models are most accurate for proteins in transition states that are close to the final state. Else, the difference in potential energy generates distortions in the structure that can not help reaching the desired final state especially when the final state is a misfolded one, such as it is for prions.

### 2.3 Martini 3 coarse graining

An alternative method to perform coarse-graining, that allows to overcome part of the issues discussed for the Gō model, is represented by Martini. Martini coarse-graining relies on a four-to-one mapping scheme, namely four heavy atoms are mapped into a unique CG bead. Nonbonded interactions between neutral beads are described by Lennard-Jones potentials, charged beads also include Coulombic interactions. In this model, a differentiation of the beads both for their charge content and dimension is present, allowing a more specific representation of forces. However, a new version of the Martini model is required in order to overcome issues concerning the overestimation of the strength of interactions among molecules.

The reparametrization of beads as well as the introduction of new labels represent the main differences between the Martini 3 model and its previous versions. For what concerns

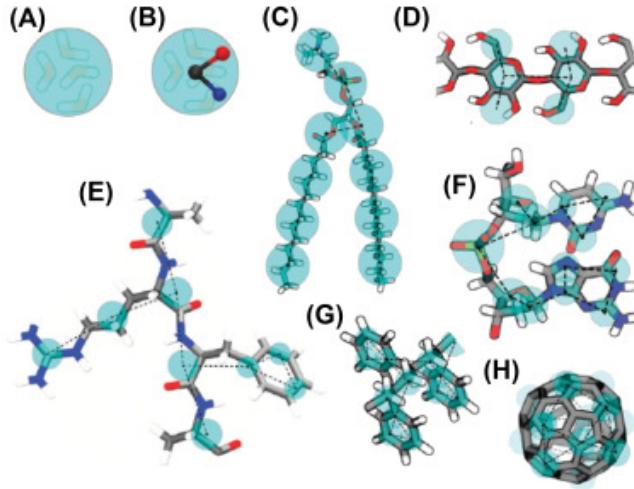


Figure 3: Examples of Martini mapping of different molecules. Image from Coarse Grained Model

the reorganization of beads, both self and cross-interactions are included. Moreover, they have been separated into groups depending both on their size and chemical type. For the latter one, three different groups have been created: organic, ions and water beads (Fig. 3).

Also a revised interaction matrix as well as intermediate interactions levels have been introduced with the aim of smoothing the transition between different chemical types. For what concerns protein-protein interactions, in Martini 3 the geometry of molecules has changed from a center of mass to a size-shape concept. This has the advantage of preserving the volume of molecules, leading to a more realistic molecular packing as well as a controlled interaction density of the model.

However, also this force field presents some drawbacks such as the limited structural detail during the modeling process due to the representation of multiple atoms with isotropic interaction sites. [9]

## 3 Methods

### 3.1 Our model structure

For the purpose of the project, a custom prion model (CHI1) is created starting from a human correctly-folded one (1FKC, residues 90-231) [10] that should correspond to PrP27-30 as cited in Section 2.1., and a mouse misfolded structure (MP) previously built by Spagnolli *et al.* [3] with a 4-rung  $\beta$ -solenoid (4R $\beta$ S) architecture. MP was based off experimental data and structural constraints, such as the presence of the  $\beta$ -solenoid loop. As reported in section 2.1, since PrP conversion occurs due to a conformational change of a PrP<sup>c</sup> when bound to a PrP<sup>Sc</sup>, the choice of building such a chimera model to try and replicate the folding mechanism was straightforward. The structure of CHI1 is reported in Fig. 4.

Following CHI1, a second chimera model, reported in Fig. 5, was generated (CHI2). The base PrP<sup>Sc</sup> model used is the same as CHI1, a partially-misfolded PrP<sup>c</sup> model used is a sheep prion structure generated by AlphaFold (AF) [11–13]. The main difference between

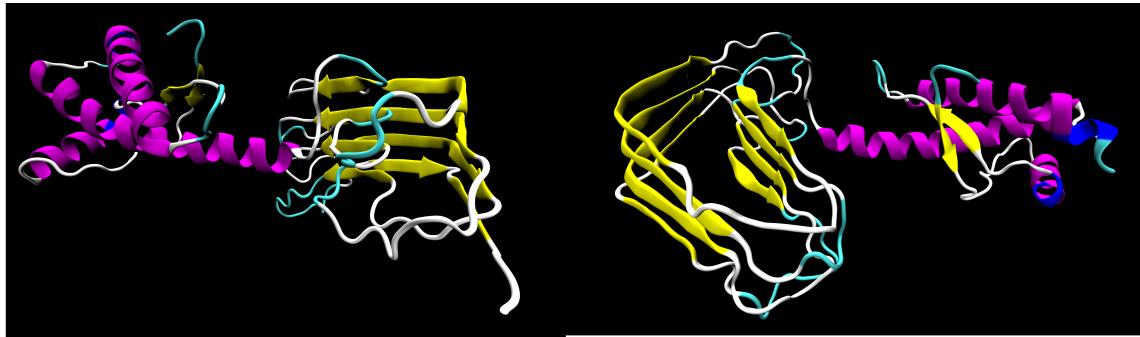


Figure 4: Rendering of CHI1 obtained with VMD, seen from two different points of view to highlight the  $\alpha$ -helix and the misfolded  $\beta$ -sheet

AF and 1FKC models is the presence of an additional chain after the SER 231 terminus in 1FKC that might be relevant for the misfolding process.

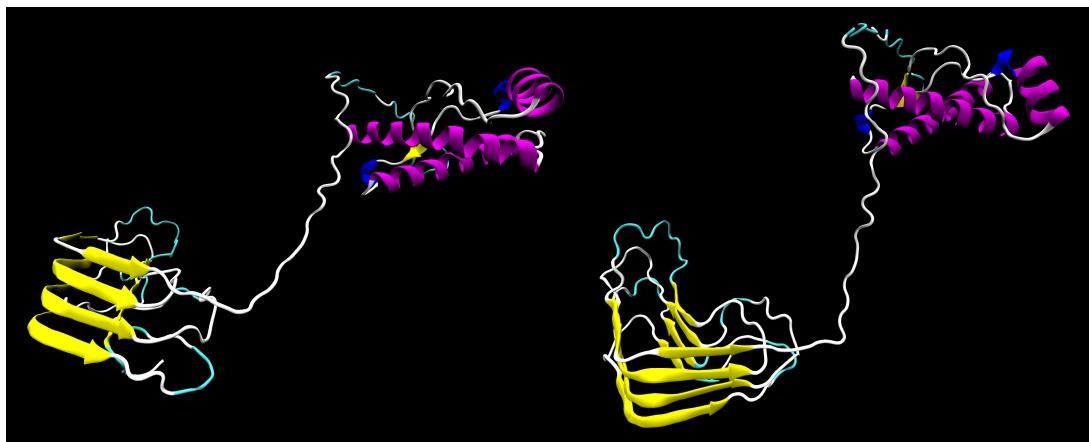


Figure 5: Rendering of CHI2 obtained with VMD, seen from two different points of view

In both models, the two structures were joint together by creating a peptide bond with the *Build Structure* function of Chimera software [14] between MP GLY 89 (N-terminus) and 1FKC SER 231 (C-terminus) for CHI1 and between between MP SER 230 (N-terminus) and AF GLY 1 (C-terminus) for CHI2.

The two newly generated secondary structures were stabilized in Chimera by using the Kabsch-Sander algorithm [15]: `ksdssp -c 0.5 -h 3 -s 3 -v`

### 3.2 Simulation Methods

Simulations were performed using GROMACS 2022.1 on three different computers, a Macbook Pro 2018 (Intel i5 quad-core CPU), a Macbook Pro 2020 (Apple M1 chip) both with MacOS Monterey 12.5.1, and a custom-built computer with Manjaro Linux 5.18.17-1 (Intel i7 octa-core CPU).

Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 [14], and on VMD (Visual Molecular Dynamics) [16, 17].

## All-atom simulations

The topology for both structures were generated using AMBER 99SB-ILDN force field [18], that is suitable for protein folding. The protein was then placed in a cubic box with width 50.0 Å for CHI1 and 100.0 Å for CHI2; the box was solvated with TIP3P water. The total charge was balanced by removing solvent molecules and adding ions so to have a 0.15 M NaCl concentration.

AMBER steepest descent algorithm was used to perform the energy minimisation, for both CHI1 and CHI2 with a maximum force allowed of 1000 kJ/mol/nm.

The second step was an equilibration run for temperature and pressure, with position of the protein restraint in both cases. For the equilibration of temperature, a modified Berendsen thermostat (V-rescale) at 300 K was used; pressure coupling was performed at 1 bar with an isotropic C-rescale algorithm. Both the processes were 100 ps long.

Trajectories were written during the production run, that was always based off the two equilibration steps. For each model, two simulations of different duration were performed: the first was 100 ps long, the second 1000 ps long. It was not possible to perform longer simulations using atomistic models because they would have been too computationally expensive for the available resources. Hydrogen bonds-containing moieties were considered as constraints and Van der Waals interactions were parametrized as Particle Mesh Ewald (PME). Each trajectory was aligned on the backbone of the protein and centered in the box using the command `gmx trjconv`. This generates the trajectory file that was used for data analysis.

## Gō model simulations

Topologies for Gō model coarse graining simulations were generated by means of SMOG server [19]. This website allows one to generate coarse grained representations of macromolecules based on a force-field that can be customized. In this work, the c-alpha model was used [6]. This model is characterized by the representation of the residues in a protein by a bead centered in the  $C_\alpha$  position, the dynamics follows a Gō-like Hamiltonian that only considers native interactions with the general form:

$$\begin{aligned}
 V = & \sum_{\text{bonds}} \varepsilon_r (r - r_0)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers/planar}} \varepsilon_\chi (\chi - \chi_0)^2 + \sum_{\text{backbone}} \varepsilon_{BB} F_D(\phi) \\
 & + \sum_{\text{sidechain}} \varepsilon_{SD} F_D(\phi) + \sum_{\text{contacts}} \varepsilon_C \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{\text{non-contacts}} \varepsilon_{NC} \left( \frac{\sigma_{ij}}{r} \right)^{12}
 \end{aligned}
 \quad (1)$$

where  $F_D(\phi) = [1 - \cos(\phi - \phi_0)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_0))]$

For both models, the contact map that was used was the shadow map with maximum contact distance of 5.5 Å and atom "shadowing" radius of 1.0 Å. The c-alpha force field was customized in the following way:

- strength of each contact: 1.0 Å
- strength of each dihedral: 1.0 Å
- excluded volume distance: 2.5 Å
- functional form for native contacts: 12-10

The box was  $30 \times 30 \times 30$  Å for CHI1 and  $50 \times 50 \times 50$  Å for CHI2; the difference is due to their different length. Settings for heterogeneous masses and unrecognition of hydrogens were selected.

After generating the input structures, the energy files `calpha.mdp` and `calphatable.mdp` were used to run the simulation on GROMACS 2022.2. In this case, only the production run was executed. It is worth to note that in c-alpha simulations there is no solvent nor ions, so the simulation was operated in *vacuum*. As for the AA, two runs per model were done, a shorter one of 100 ps and a longer one of 1000 ps, both followed by the trajectory alignment step.

### Martini 3 simulations

Input files for the simulation were generated using the web-based platform CHARMM-GUI, developed by Dr. Im's research team at Lehigh University, Bethlehem. [20] The tool *Martini Maker - Solution Builder* was used to perform the coarse graining of both CHI1 and CHI2, exploiting Martini 3.0.0 force field. Both proteins were immersed in a rectangular water box, respectively of 10Å of edge distance for CHI1 and 12.0 Å of edge distance for CHI2. The total charge was balanced adding ions of NaCl 0.15 M. For each protein, two different CG models were created, respectively at temperature 303.15 K and 310.15 K, in order to study the effects of thermal fluctuations on the system with the molecular dynamics simulation.

MD simulations were performed using GROMACS 2022.1. Two energy minimizations followed by an equilibration and a final production run were performed for each CG protein.

The *minimization* steps were performed using the Steepest Descent algorithm. It always converged to the maximum force value `Fmax` within 3000 steps despite the final value was above the requested tolerance of 10 kJ/mol/nm.

The following *equilibration* step exploited a leap-frog algorithm to integrate Newton's equations of motion. A time step of `dt` = 0.02 ps was used and the total equilibration time was chosen to be a fraction of the total duration of the production run. The isotropic pressure coupling was considered through the C-rescale barostat.

The leap-frog integrator was exploited also in the *production run* with a time step of `dt` = 0.02 ps. In this case, pressure coupling has been taken into account using a Parrinello-Rahman barostat. Two different simulation lasting 20 ns and 100 ns respectively were performed for each CG protein at different temperatures.

Finally, each protein in the correspondent output trajectory file was centered with the command `gmx trjconv` and aligned to the first frame.

### Backmapping

In order to gain insight about the conformational changes of both chimeras after the MD simulations of their CG structures, obtained exploiting the Martini 3.0.0 force field, a procedure known as *backmapping* was performed.

To obtain a reliable result, the backmapping - or reverse coarse-graining - requires a correspondence between the atomistic and CG models, meaning that a backmapping protocol needs to know at least which atoms contribute to which bead. This correspondence is defined in the mapping file, already implemented in the software used. It is worth noticing that this procedure could have been done just because the Martini CG was performed

with the same platform. If none, no known correspondence between beads and atoms was known and the backmapping procedure would have been much more difficult to compute. Files containing the AA structure of CHI1 and CHI2, resulting from the simulation of 100 ns, were generated exploiting the tool *Martini Maker - All-Atom Converter* in the Input Generator selection of the web-based platform CHARMM-GUI [20].

In particular, the PDB file corresponding to the last frame of the MD simulation was given in input to the tool and after a series of steps involving solvation, introduction of ions and generation of the pbc setup, a prototype of AA structure was given in output [21].

Of course, a series of steps of minimization of energy, equilibration and production run are needed to obtain a final structure at equilibrium. Thus, GROMACS was used to perform these various steps. However, the complete procedure suggested by the developers of the tool required to perform a scheme of at least 5 equilibration steps after the minimization one. Since lot of computational effort was needed, only one equilibration was performed as well as a shorter production run compared to the one suggested.

## 4 Results and discussion

In attempt to reproduce the misfolding process of the prion protein, a structural analysis of the data acquired from the MD simulations of both CHI1 and CHI2 was performed. The Python library *MDAnalysis* [22, 23] was used to analyse data obtained from the MD simulation.

Different structural parameters were considered to study the behaviour of the structure and understand whether an equilibrium conformation was reached or not.

Firstly, the radius of gyration  $R_g$  was studied to gain information on the average radius of the protein as if it was thought to be with an approximately spherical geometry over time. It is defined according to:

$$R_g^2 = \frac{1}{N+1} \sum_{i=0}^N \langle (\mathbf{r}_i - \mathbf{r}_{\text{cm}})^2 \rangle \quad (2)$$

where  $\mathbf{r}_{\text{cm}}$  is the position of the center of mass of the protein and  $\mathbf{r}_i$  the position of each residue.

An alignment of each chimera to the first frame of the trajectory was performed in order to minimize both the RMSD (root-mean-square deviation) and the RMSF (root-mean-square fluctuation) parameters.

The RMSD, as a function of time, is a parameter useful to determine the changes in conformation of a protein evolving in time from a reference structure. It is calculated according to:

$$\text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2} \quad (3)$$

in which the difference is carried out between the position of each residue at time  $t$  from their position in the reference structure ( $t = 0$ s).

In order to gain insight into the conformational convergence of the simulation, the pairwise RMSD of the trajectory to itself was also computed.

Furthermore, in order to study the fluctuations that characterise each residue over whole

trajectory, the RMSF of each residue was computed. It is defined as the time average of the deviation of the position of each residue to its mean value

$$\text{RMSF}_i = \sqrt{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle} \quad (4)$$

For the AA simulations, a Ramachandran plot of both the initial and final frame was computed in order to verify whether a conformational change occurred in the structure.

## 4.1 All-atom simulations

For what concerns all-atom simulations, trajectories lasting only 1 ns were produced due to the high computational costs and lack of resources. As a consequence, important changes in the structure of the protein could have not been observed because of the limited interval of time over which data were obtained.

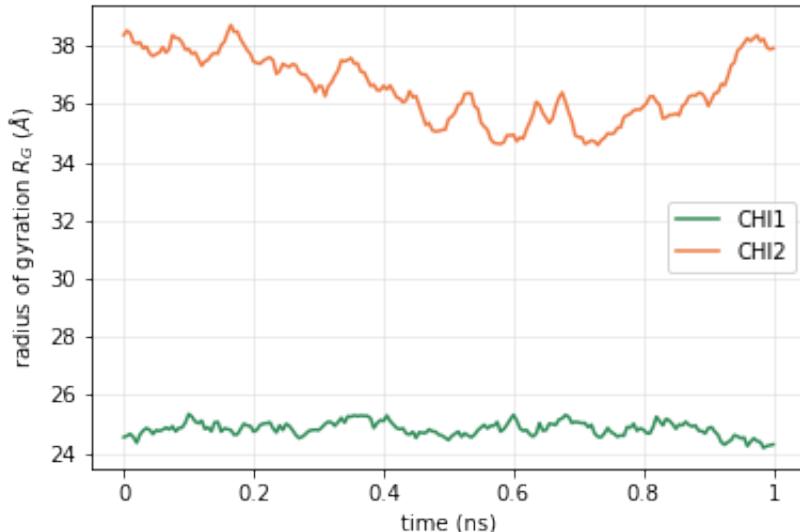


Figure 6: Radius of gyration of CHI1 (green) and CHI2 (orange)

Despite this limit, an analysis based on the available data for the two chimeras was performed anyway, in order to highlight possible structural differences between the two. Starting from the radius of gyration, as can be seen in Fig. 6 it is clear that the  $R_g$  of CHI2 fluctuates around an equilibrium value which is much higher in comparison with average value of CHI1.

This result is justified by the shape of the two chimeras: as can be seen in Fig. 4 and Fig. 5, CHI1 is more compact than CHI2.

Since the process under observation was an unfolding one, an increasing value for  $R_g$  was expected to be observed as both chimeras evolved in time. However, since the AA simulations were only 1 ns long, only fluctuations around a constant value were observed. For what concerns the RMSD, an increasing trend is observed for both chimeras. This result is justified by the fact that in the first moments of the simulations both structures needed to accommodate in space; the growing trend of RMSD and the absence of a *plateau* are due to a short simulation time (1 ns) that is not enough to observe any relevant process.

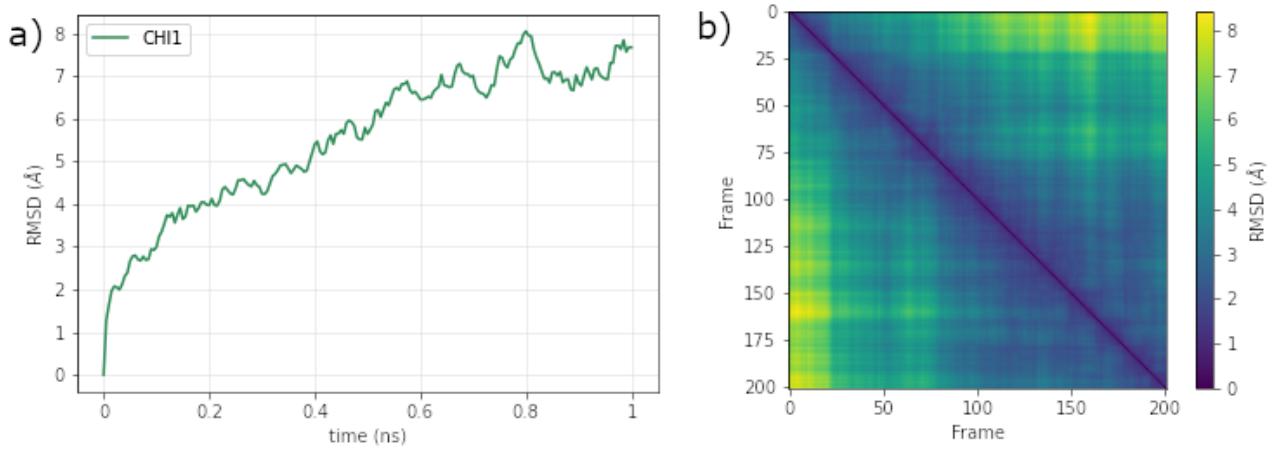


Figure 7: Pictures representing the RMSD for CHI1 (a) and the corresponding pairwise RMSD matrix (b)

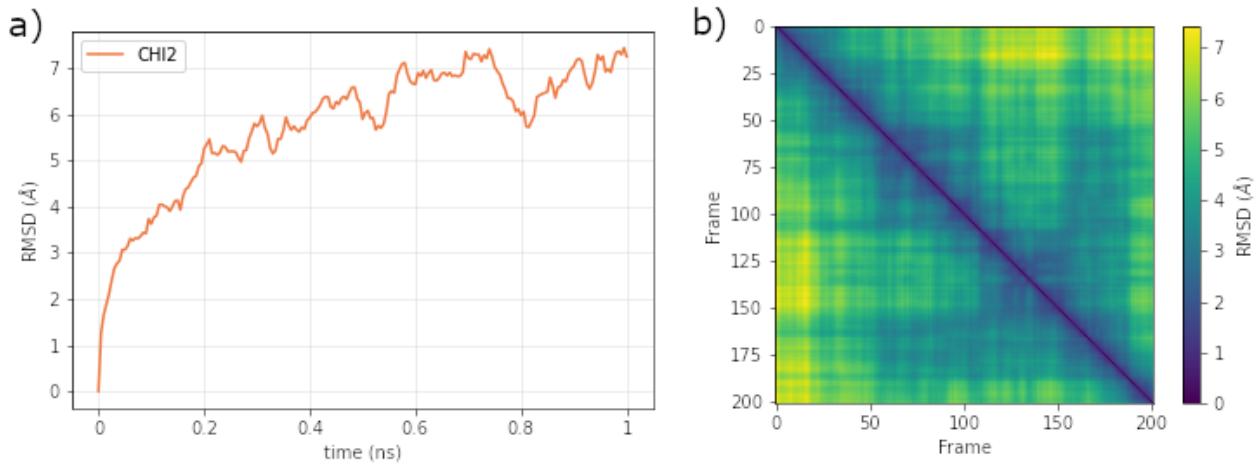


Figure 8: Pictures representing the RMSD for CHI2 (a) and the corresponding pairwise RMSD matrix (b)

More information can be gained by combining the obtained results with pairwise RMSD matrices. The diagonal of the plot is zero since the graph represents the RMSD of a structure with itself. In the case of CHI1 (Fig. 7) small blocks of low RMSD values are observed along the diagonal. It indicates that similar structures are explored all over the trajectory.

A different behaviour characterizes CHI2 (Fig. 8) since less blocks of low RMSD value are observed along the diagonal while more low RMSD off diagonal blocks are present. This suggests that the protein re-explores some earlier states. Again, this behaviour can be justified by the fact that a protein with a more relaxed structure can visit a broader region of the conformational space. We can relate the variation between the RMSD at around  $t = 0.8$  ns with the block placed at around frames  $150 \div 190$ , indicating the presence of a conformation state with a smaller  $R_g$  and RMSD. In Fig. 9 the RMSF of both CHI1 and CHI2 are reported. Fluctuations of residues of both proteins are of the order of 10 Å. In both cases, maxima are observed in correspondence of central residues, meaning that loops were forming in those regions.

All-atom simulations allowed the computation of Ramachandran plots for both chimeras

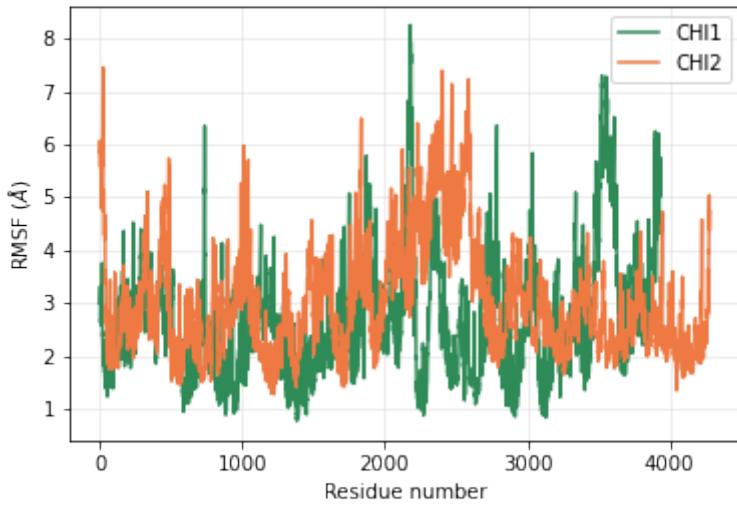


Figure 9: Pictures representing the RMSF for CHI1 (a, green) and CHI2 (b, orange)

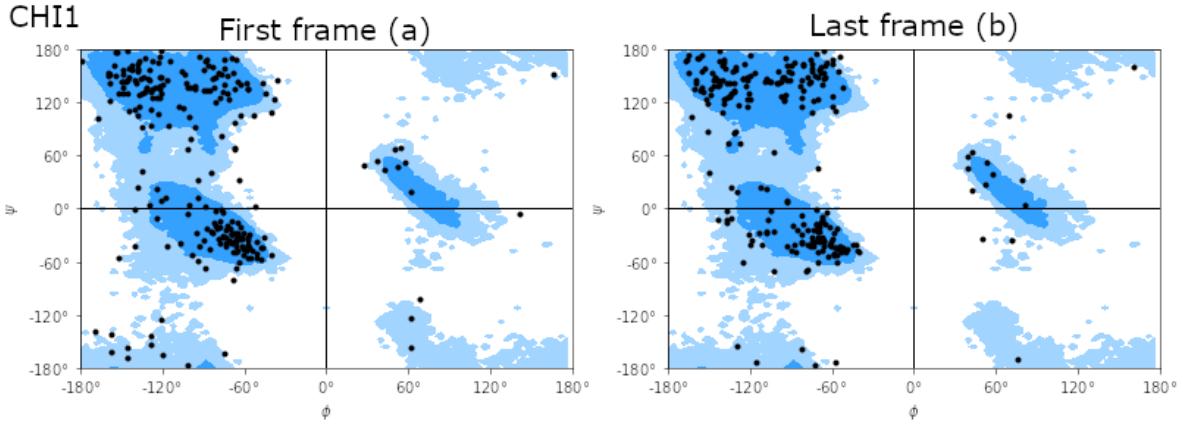


Figure 10: Ramachandran plots for CHI1 at first frame (a) and last frame (b)

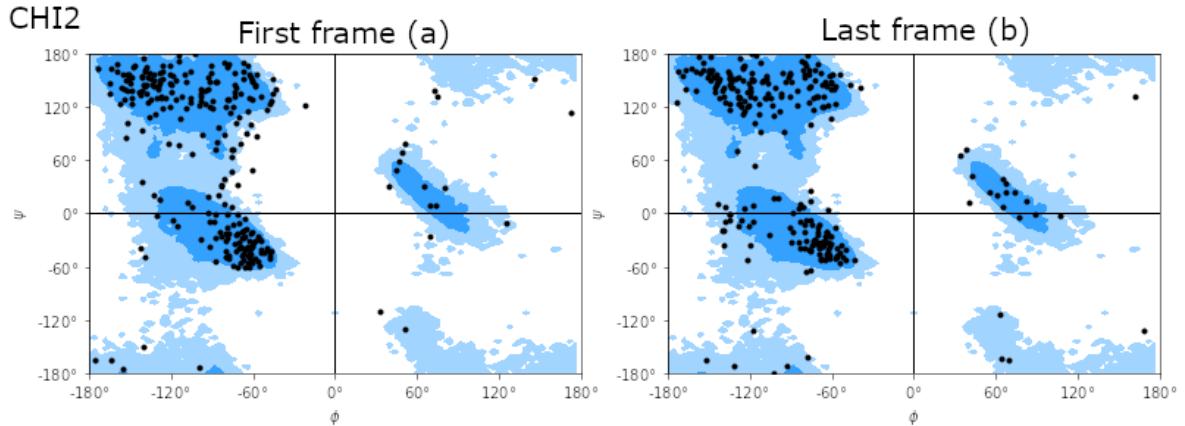


Figure 11: Ramachandran plots for CHI2 at first frame (a) and last frame (b)

(Fig. 10, 11). Plots confirm the hypothesis for which there are no relevant processes happening within the structure in the chosen simulation time. In fact the comparison

between the first frame and the final frame in both cases does not show any noticeable difference. There are just small rearrangements that cannot account for any transition.

## 4.2 Coarse grained simulations

### Gō model

Gō model coarse graining was the first method that was explored in this project to understand the transition between the  $\alpha$ -helix and  $\beta$ -sheet. It is important to remark that the evolution of the model is based on native contacts between fragments but often the interactions involved in protein folding are of other kind and cannot be simplified. These concepts come out remarkably when observing the differences between the radius of gyration of CHI1 and CHI2 (Fig. 12).

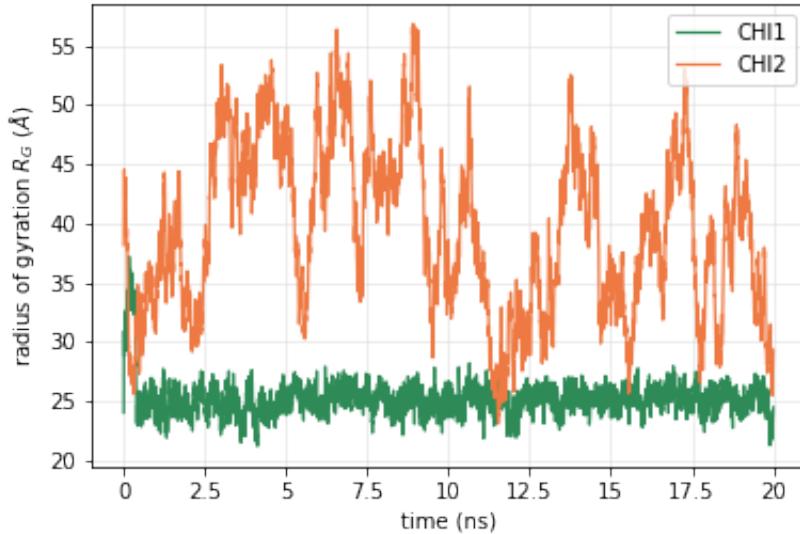


Figure 12: Radiiuses of gyration of CHI1 (green) and CHI2 (orange) obtained with the Gō model coarse graining

CHI1 shows a very stable  $R_g$ , in fact the structure is more compact and it is easier for the algorithm to recognize the fragments involved in native interactions and generate and maintain contacts. CHI2 shows a very different situation: by recalling the elongated structure (Fig. 5), it can be easily seen that the native contacts are more difficult to get in touch.

Moreover, for the latter protein the plot shows that the radius of gyration has a periodic trend, meaning that both attractive and repulsive interactions are involved. The latter ones could be due to the constraints imposed by the long free chain between the two secondary structures. This periodic elongation and contraction of CHI2 was also observed in the simulation film.

The same trends are highlighted by the RMSD plot (Fig. 13). RMSD for CHI1 fluctuates around 10 Å while CHI2 shows greater fluctuations.

Finally, the RMSF reported in Fig. 14 shows that CHI2 is subject to more fluctuations in comparison with CHI1, with the most fluctuating part corresponding roughly to the fragment connecting the  $\alpha$ -helix and the  $\beta$ -sheet.

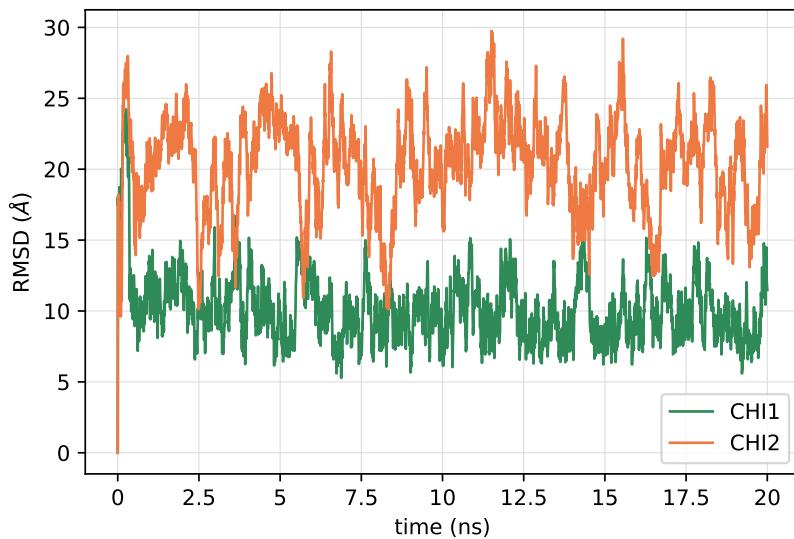


Figure 13: RMSD of CHI1 and CHI2 obtained with the Gō model coarse graining

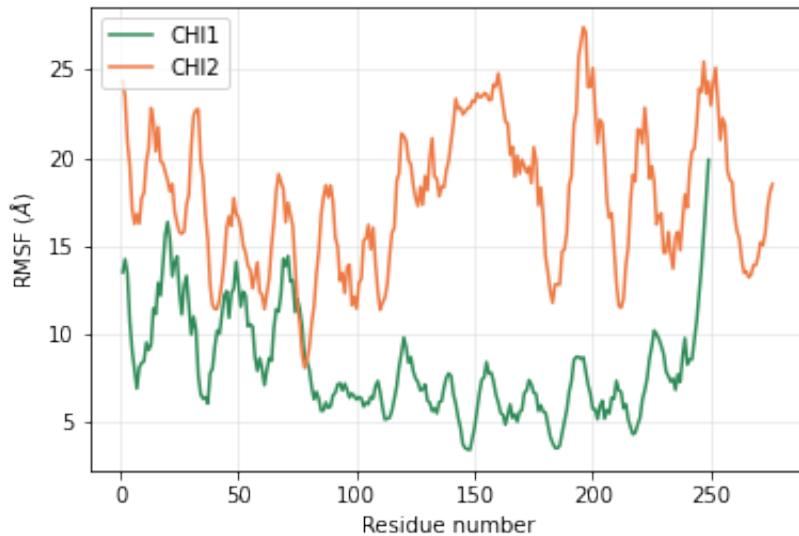


Figure 14: RMSF of CHI1 and CHI2 obtained with the Gō model coarse graining

### Martini 3

In the case of MD simulations performed with Martini 3 coarse-grained proteins, not only their different behaviour was analysed, but also the temperature dependence of the results. Indeed, for each chimera two simulations of 100 ns, respectively at temperature  $T_1 = 303.15$  K and  $T_2 = 310.15$  K, were performed.  $T_1$  was the default temperature given by CHARMM-GUI while  $T_2$  is the average internal temperature of the organisms considered to build both chimeras, and it is also the temperature at which supposedly the misfolding process occurs.

As in the previous analysis, the first parameter to be analysed is the radius of gyration.

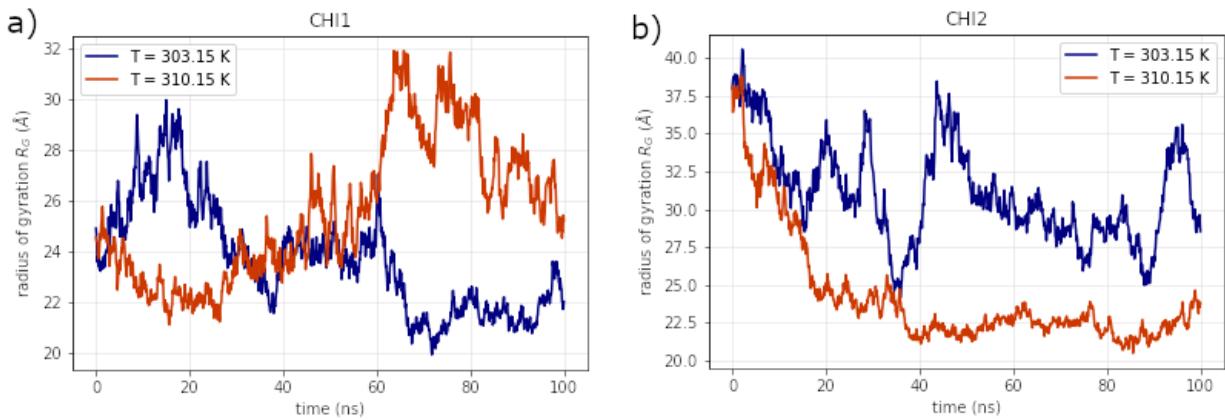


Figure 15: Radius of gyration for CHI1 (a) and CHI2 (b) at  $T_1$  (blue) and  $T_2$  (red)

From Fig. (15a) it is possible to see that CHI1 is more compact at high temperature in the first part of the plot while the simulation at lower temperature produces a more compact structure in the final part of the run. An opposite behaviour is observed for CHI2 (Fig.15b) which seems to reach an equilibrium configuration after about 40 ns at higher temperatures. The dependence of  $R_g$  on temperature can be explained by the fact that probably at higher temperature CHI2 reaches a conformation in which interactions are stronger, leading to a more compact structure. CHI2, although, presents a RMSD who suggests a different behaviour than  $R_g$ : RMSD at  $T_2$  is higher than RMSD at  $T_1$  (Fig. 16b). Both  $R_g$  and RMSD underline an equilibrium configuration for CHI2, the structure stays stable in time but it "moves around" in a local minimum in the energy landscape. The movement is due to thermal fluctuations induced by the higher temperature. CHI2 seems to be in a more unstable state at lower temperature, subject to more fluctuations, and transitions to a local minimum state at higher temperature.

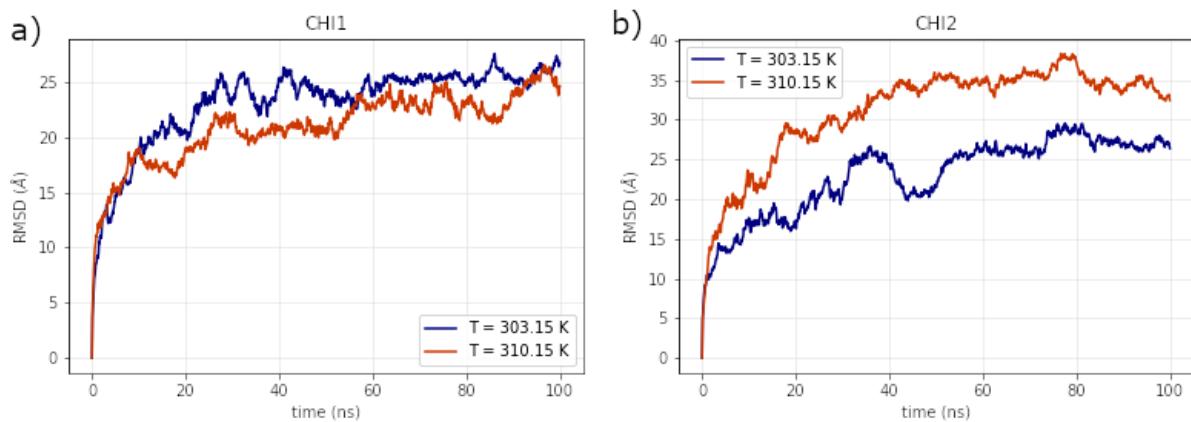


Figure 16: RMSD for CHI1 (a) and CHI2 (b) at  $T_1$  (blue) and  $T_2$  (red)

It is worth to note the differences in RMSF between the three simulation methods that have been used (Figs. 9, 14, 17) since this parameter does not rely on simulation time but on the number of residues, despite the different numeration. All RMSF plots show a periodic-like trend, with residues showing higher fluctuations in the central part of the molecule, representing the connecting fragment. The main difference lies in the order of magnitude of the fluctuations, that is way smaller for AA simulation with respect to the

coarse grained ones: this is due to a shorter simulation time (1 ns) and a lower temperature (300 K).

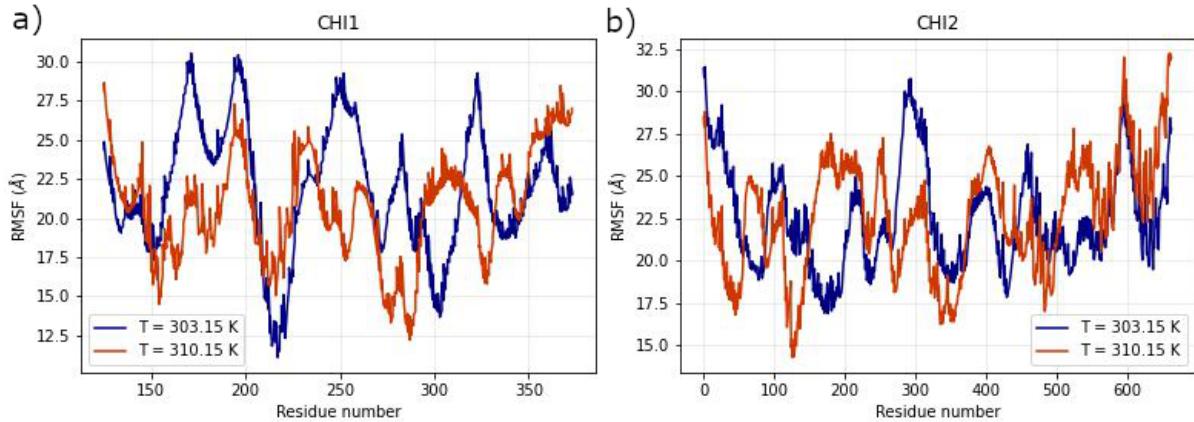


Figure 17: RMSF for CHI1 (a) and CHI2 (b) at  $T_1$  (blue) and  $T_2$  (red)

### 4.3 Backmapping

As a final step in the analysis, a backmapping of both CHI1 and CHI2 at 303.15 K was performed. It is worth mentioning that due to the high computational effort involved in performing a long AA simulation, only one equilibration step has been computed as well as a shorter production run compared to the one suggested by CHARMM-GUI.



Figure 18: Rendering of the backmapped structure of CHI1 at 303.15 K obtained with VMD

Accordingly, the resulting structures may not be completely equilibrated and still subject to external forces that may modify their final secondary structure.

In Fig. 18 the backmapped structure of CHI1 is reported. Interestingly, the  $\beta$ -sheets have completely unfolded, giving rise to random coils while the  $\alpha$ -helix has maintained its structure, not being particularly modified by the simulation. Since a similar structure was observed also for CHI2 and no ulterior relevant information could be extracted from it, it was not reported.

## 5 Conclusion

The aim of the project was to observe the conformational change from  $\alpha$ -helix to  $\beta$ -sheet of two custom-generated prion protein chimeras. To do so, three simulation methods were used: all-atom simulations, that showed only thermal fluctuations due to a small simulation time and little computational resources; Gō model coarse graining, that gave no meaningful results and was proven to be ineffective for protein folding; and Martini 3 coarse graining, that was the method that gave some insights. From the performed simulations some information about the energy landscape of the two chimeras can be extracted, even if the runs were not directly comparable because of different methods, parameters, and simulation times. CHI1 is the most stable and compact one, CHI2 showed a behaviour that sometimes diverged from the expectations, minding that the  $\alpha$ -helix fragment was generated by AlphaFold, hence it has some degrees of uncertainty. Overall, the target transition did not occur in any of the two structures. The main factor is that simulations themselves are an oversimplification of highly complex processes, sometimes even an all-atom simulation is not detailed enough to represent them. Coarse-graining heavily approximates the interactions between residues, hence the information that could have been needed for the conformational change in the specific case of this project has been lost in the dimensionality reduction process. An important thing to note is that the majority of the studies on protein folding do not rely on *ab initio* molecular dynamics simulations, but instead "help" proteins transition to the correct target state. This is the case, for example, of the already cited paper from Spagnolli *et al.*[3].

The last thing worth mentioning is about the misfolded  $\beta$ -sheet structure that was here used as a mold: there are more recent papers where further experiments on PrP<sup>sc</sup> were performed [24, 25] using techniques at higher resolution. The PrP<sup>sc</sup> architecture used in this paper was a 4-rung  $\beta$ -solenoid, the newly detected architecture for the scrapie prion is called PIRIBS (parallel in register  $\beta$  stack), whose PDB has not been published yet. This shows that research in prions biochemistry is still strongly ongoing.

## References

1. Prusiner, S. B. Prions. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13363–13383 (1998).
2. Pan, K. *et al.* Conversion of  $\alpha$ -helices into  $\beta$ -sheets features in the formation of scrapie prion proteins. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10962–10966 (1993).
3. Spagnolli, G. *et al.* Full atomistic model of prion structure and conversion. *PLOS Pathogens* **15**, 1–18 (2019).
4. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128**, 244114 (2008).
5. Hiroshi, T., Yuzo, U. & Nobuhiro, G. Studies on Protein Folding, Unfolding and Fluctuations by Computer Simulation. *International Journal of Peptide and Protein Research* **7**, 445–459 (1975).
6. Clementi, C., Nymeyer, H. & Onuchic, J. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* **298**, 937–953 (2000).
7. Tozzini, V. Coarse-grained models for proteins. *Current Opinion in Structural Biology* **15**, 144–150 (2005).
8. Emanuele, P., Michele, V. & Martin, K. Validity of Gō Models: Comparison with a Solvent-Shielded Empirical Energy Decomposition. *Biophysical Journal* **83**, 3032–3038 (2002).
9. Souza, P., Alessandri, R. & Barnoud, J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods* **18**, 382–388 (2021).
10. Zhang, Y., Swietnicki, W., Zagorski, M. G., Surewicz, W. K. & Sönnichsen, F. D. Solution Structure of the E200K Variant of Human Prion Protein: IMPLICATIONS FOR THE MECHANISM OF PATHOGENESIS IN FAMILIAL PRION DISEASES\*. *Journal of Biological Chemistry* **275**, 33650–33654 (2000).
11. AlphaFold Protein Structure Database - PrPsc <https://alphafold.com/entry/G0Z735>. Accessed: August, 24, 2022.
12. Jumper, J., Evans, R. & Pritzel, A. e. a. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
13. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).
14. Pettersen, E. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004).
15. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
16. Humphrey, W., Dalke, A. & Schulter, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

17. *VMD* <http://www.ks.uiuc.edu/Research/vmd/>. Accessed: August, 29, 2022.
18. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
19. *SMOG - A Webtool for BioMolecular Simulations* <https://smog-server.org/cgi-bin/GenTopGro.pl>. Accessed: June, 23, 2022.
20. *CHARMM-GUI* <https://www.charmm-gui.org>. Accessed: August, 24, 2022.
21. *Martini- Backmapping* <http://www.cgmartini.nl/index.php/tutorials/37-tutorial2/314-tutorial-reverse-mapping>.
22. Gowers, R. J. *et al.* *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations* in *Proceedings of the 15th Python in Science Conference* (eds Benthall, S. & Rostrup, S.) (2016), 98–105.
23. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319–2327 (2011).
24. Martín-Pastor, M. *et al.* Solid state NMR reveals a parallel in register architecture for an infectious recombinant prion (2021).
25. Caughey, B., Standke, H., Artikis, E., Hoyt, F. & Kraus, A. Pathogenic prion structures at high resolution. *PLoS Pathogens* **18** (2022).