Politecnico di Torino

Master's degree in Biomedical Engineering

Conformational Dynamics and Molecular Characterization of MORN motifs to shed light on homophilic interactions driving macromolecular protein superassemblies

Supervisor

Prof. Marco Agostino Deriu

Candidate

Riccardo Tortarolo

Co-supervisor

Marcello Miceli

Academic year 2020/2021

Summary

Abstract

Membrane Occupation Recognition Nexus (MORN) motifs are protein domains poorly characterized despite the fact of being widely diffuse among different species in both eucaryotic and procaryotic organisms. These motifs are characterized by several repetitions of an highly conserved β-harpins modules. Experimental evidence is often conflicting showing a possible lipids binding activity as well as a protein-protein interaction capability. In mammalian MORN repetitions are found in Alsin, a multiple domain protein of 1657 amino acids transcripted from Amyotrophic Lateral Sclerosis type 2 (ALS2) gene, whose mutations are associated with Infantile-onset Ascending Hereditary Spastic Paralysis (IAHSP) disease. IAHSP is a rare neurodegenerative condition whose symptoms start to occur from the first years of life manifesting lower limbs spasticity and usually worsening also affecting the upper limbs and reaching tetraplegia. Literature evidence suggest that ALS2 MORN domains are involved in homophilic interactions to form an Alsin tetrameric form and binding activity of the guanosine triphosphatase Rab5. However, the relationship between Alsin protein biological functions and its three-dimensional structure is still unknown since, to our knowledge, no 3D models are available. In this work we assessed lack of such 3D model for ALS2 MORN and investigated its structural properties and potential dimeric state. We started by characterizing, through Molecular Dynamics (MD) methods, the geometry of three MORN dimeric assemblies from *Trypanosoma brucei*, *Plasmodium falciparum* and *Toxoplasma gondii* whose 3D structures are known from crystallography data of a recent study. Then we developed a first homology model of ALS2 MORN domain trough Computational Molecular Modelling methods and we validated it by comparing our best model with the best one generated from the ITasser suite. Based on the previous results, we finally built homology models of two possible dimeric conformations: a linear assembly and a V-shaped one. Their validation was based on comparison with the results of the previous three Results suggest that Alsin MORN domains are stable and seems to favour a V-shaped homophilic interaction rather than a more linear one giving a new insight into conformational characterization for the dimerization process of the Alsin.

1 Introduction

In this first chapter a general introduction of the master's thesis work is presented, summarizing the organization of the manuscript while briefly introducing the aims of the work and the results.

Among highly structured proteins the Membrane Occupation Recognition Nexus motifs are far from being understood despite their presence in various different organisms of both biological domains. Their assumed biological function to localise cellular membranes through interactions with the lipidic components is questioned because no direct evidence of this actually happening have been observed. Additionally, different studies reported involvement of MORN repetitions in protein-binding activities as well as homotypic interactions. This apparently ambiguous behaviour is still reason for debate. Recently the first crystallographic structures of MORN homodimers proteins have been presented supporting the idea of these repetitions being dimerization modules. From published 3D structure models it emerges that, even though secondary and tertiary structures conservation occurs, the macro molecular arrangement vary from domain to domain. Molecular assembly conformations are often important for protein activity and for this reason their investigation could help to understand to understand related biological functions.

To make a first step towards the characterization of these interesting domains, the aim of this M.Sc. thesis work is to shed light on their dynamical behavior focusing on the study of the mechanisms that drives their macromolecular assembly formation.

This work is organized as follows:

Chapter 1 is the present introduction.

Chapter 2 provides a biological background about the Membrane Occupation Recognition Nexus. The hypothesized biological functions are reported through the presentation of different MORN-conataining proteins found in various organisms.

Chapter 3 is a description of the materials and methods used in the present work. After the presentation of molecular modelling and Homology Modelling concepts, a theoretical explanation for molecular mechanics and dynamics will be reported.

Chapter 4 will concern about the first section of the present master' thesis work, describing the characterization of molecular assemblies mediated by MORN homophilic interactions.

Chapter 5 describes the final part of this work showing the application of Homology Modelling to reconstruct the MORN domain of structural-unknown Alsin protein and two hypothesized assembly conformations. Characterization of these novel structures is finally reported and discussed.

2 Biological and technological background

2.1 MORN domains overview

The Membrane Occupation and Recognition Nexus (MORN) are protein domains formed by several repetitions found ubiquitously in both eukaryotes and prokaryotes [1]. The name derives from the originally proposed role of mediating plasma membrane association, hypothesized when this assembly was first discovered in 2000 as a domain of 8 repetitions in junctophilins [2]. Bioinformatic analysis show that the number of MORN repetitions vary greatly among different proteins (from 2 to 20) and that they are found in combination of several different domains. Different recent studies suggest that each repetition is composed by 23-24 amino acids [3] [4] [5], however some groups favour a 14 amino acids composition as it was initially thought [2] [6]. The secondary structure is highly conserved and is characterized by an all-β secondary structure, with each repetition expected to form a β-hairpin (strand-loop-strand) pattern.

Even though these domains are generally assumed to be lipid-binding sites [2] [5], there is no direct evidence of this biological function and the experimental data are often conflicting and ambiguous. This leads to the hypothesis that MORN motifs could rather function as a protein-protein interaction domain [7].

In the following a report of the evidence supporting the two supposed functions is presented together with an introduction to a specific protein MORN domain.

2.2 The lipid-binding function hypothesis

As previously said, the first proposed function of the MORN domains was mediating the association of the plasma membrane with the protein containing the motifs. The hypothesis is that the interactions targets have to be lipidic components of the membrane. This idea is possibly supported by various experimental evidence on proteins like Junctophilins [2] [6] [8] [9] and a subfamily of phosphatidylinositol-4-phosphate (PtdInsP)2 5-kinases (PIPKs) [5][10] [11].

Junctophilins are a family of integral endoplasmic reticulum proteins that are components of the junctional complexes in excitable cells like muscle and neural ones. Their function is supposedly to provide stabilization of the junctional membrane complexes by anchoring the endoplasmic or sarcoplasmic reticulum to the plasma membrane and thus providing a structural basis for ion channels and cellular surface cross-talk. In the mammalian genome 4 proteins of the

junctophilins family have been discovered (JP-1, JP-2, JP-3, JP-4) [8] [12] [13]. JP-1 is expressed in skeletal muscle [2], JP-2 in skeletal and cardiac muscle [6], and JP3 and JP4 in the brain [13]. All of them share a common structural organization, based on 8 conserved MORN motifs at the N-terminus. The first 6 MORN repetitions are separated by the last 2 by a region of about 120 amino acids, referred to as the "joining region" [9].

JP-1 MORN repetitions are motifs of 14 amino acids each. Genomic sequences predictions found additional hypothetical protein that contain MORN motifs like those of JP-1 expressed by *C. elegans*, *A. thaliana* and *Cyanobacterium*. Moreover, sequence comparison shows a very conserved motif sequence "Tyr-Gln/Glu-Gly-Glu/Gln-Trp-x-Asn-Gly-Lys-x-His-Gly-Tyr-Gly" [2]. These suggests that MORN is a protein-folding module shared by functionally different proteins. Experimental results shows that MORN domain mediate binding to the sarcolemma via an interaction with PI(4,5)P2, a phosphoinositide concentrated in this membrane, [9] and that contribute to JP-1 binding capacity to the plasma membrane targets (supposedly sphingomyelin and phosphatidylcholine) depending on the number of the expressed repetitions [2]. However, interactions occur along the entire sarcolemma and not only at T-tubules indicating that MORN domain does not confer selectivity. Moreover, it has been demonstrated that partial or even complete deletion of MORN motifs does not affect the selective localization at triads and suggests that this is likely dependent on signals provided by regions of the protein other than MORN [9].

JP-2 is found in cardiac myocytes as a linker protein in the microdomain of the dyad spanning and holding the myocyte transverse-tubules (t-t) and the junctional sarcoplasmic reticulum (jSR) membranes systems in a precise geometry to facilitate calcium-induced calcium release process (CICR) which regulates cardiac contraction. Its linking function is thought to be possible due to the N-terminal portion containing the MORN domain which is supposedly able to bind to the t-t membrane while the C-terminus is anchored within the sarcoplasmic reticulum membrane [6]. Experimental analysis on a truncated form of JP-2 containing only the amino-terminal region (aa 1-452) show interactions with PtdIns(3,4,5)P3 PtdIns(3,5)P2 and PS, very weak binding to PtdIns(4,5)P2 and no interaction with PA.

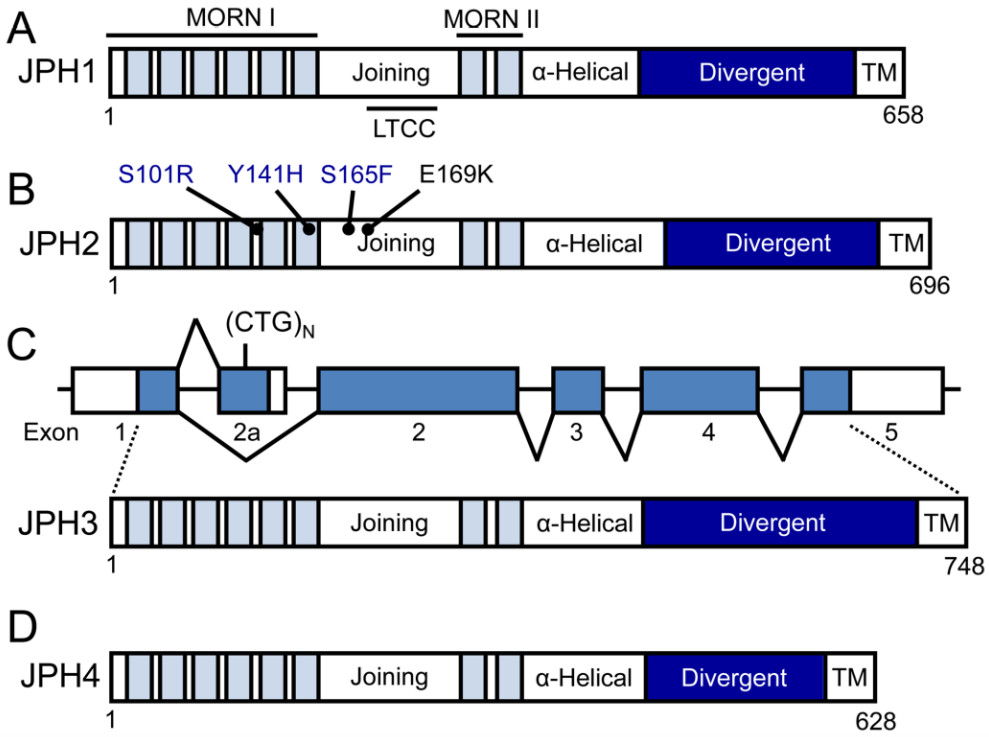

Figure 1. Graphical representation of the four mammalian junctophilins. MORN domains are highly conserved [12].

Other evidences of the lipid-binding activity comes from PIPKs, a group of enzymatic catalysts that synthesize phosphatidylinositol-(4,5)-bisphosphate (PtdIns(4,5)P2), a key component in phosphoinositide (PI) signalling that regulates many cellular processes.

In *Arabidopsis*, a small plant popular as model organism in plant biology and genetics, 11 isoforms of PIPKs are expressed. Among these, AtPIPK1-9 composing the subfamily B (the subfamily A includes PIPK10-11) contain repeated 23-amino acid MORN motifs in the N-terminal region (amino acids 1–251) that do not overlap with the catalytic domain [5]. More specifically, AtPIPK1-3 possess 7 MORN repetitions and AtPIPK4-9 contains 8 motifs. In AtPIPK1, each of the 7 MORN motifs is slightly different; however, they all contain a consensus sequence of hydrophobic and glycine residues, $YXGX(W/F)(X)_6GXG(X)_6G(X)_2$ joined with few amino acids. Experimental results show AtPIPKs MORN domain is essential for phosphatidic acid (PtdOH) activation and that its expression increases enzymatic activity *in vivo*. Data analyses support a model where MORN domain regulate accessibility of lipids to the catalytic site of the PIPKs proteins in a PtdOH-sensitive manner [5].

Another evidence comes from rice PIPK (OsPIPK1) formed by 9 MORN repeat motifs located in the N-terminus, 8 of wich containing a conserved peptide of 14 amino acids in length (Y-

Q/E-G-E/Q-T-X-N-G-K-X-H-G-Y-G) and the ninth one (YEREYVQGVLIMEQ) connected to the other by a 104 amino-acids linker sequence. In fact, fat Western blot analysis of the recombinant MORN region polypeptide containing the whole nine MORN motifs and the linker sequence showed that it could bind phosphatidic acid (PA), PI4P, and PI(4,5)P2 [10].

2.3 The protein-protein interactions hypothesis

Hypotheses considering MORN-lipid binding are supported by some scientific evidence. However, the topic is still debated. Moreover, alternative hypotheses concerning protein binding activity have been considered in literature. A first evidence comes from a study conducted on MORN4, a protein containing 4 MORN repetitions of 14 amino acid sequence followed by a C-terminal α-helix, contained in human retinophilin and its orthologue found in *Drosophila*.

have been reported to be a binding partner/cargo of class III myosins which are key proteins for skeletal muscle contraction [14].

In a recent paper of Sajko et al. show that MORN domains of *Trypanosoma brucei* (TbMORN) exhibits a contrasting behaviour towards lipid-binding. TbMORN is a protein composed solely from 15 consecutive 23 aa MORN repetitions with barely any intervening sequence whose function might be involved in endocytosis and regulating macromolecular cargo flow through the neck of the flagellar pocket. While this protein binds to PI(4,5)P2 with micromolar affinity via the C-terminal MORN repetitions 13 and 14, principally due to an interaction with longer aliphatic chains rather than the lipid headgroup, and to PE requiring motifs 1–9, they reports no association with liposomes and no evidence that MORN repeats can associate with phospholipid membranes *in vivo*. On the other hand they observed that TbMORN and its homologues *Toxoplasma gondii* (TgMORN) and *Plasmodium falciparum* (PfMORN) exists in dimeric form interacting in an anti-parallel fashion via their C-termini. This evidence suggests that MORN domains could be involved in protein-protein interactions rather than bind lipids.

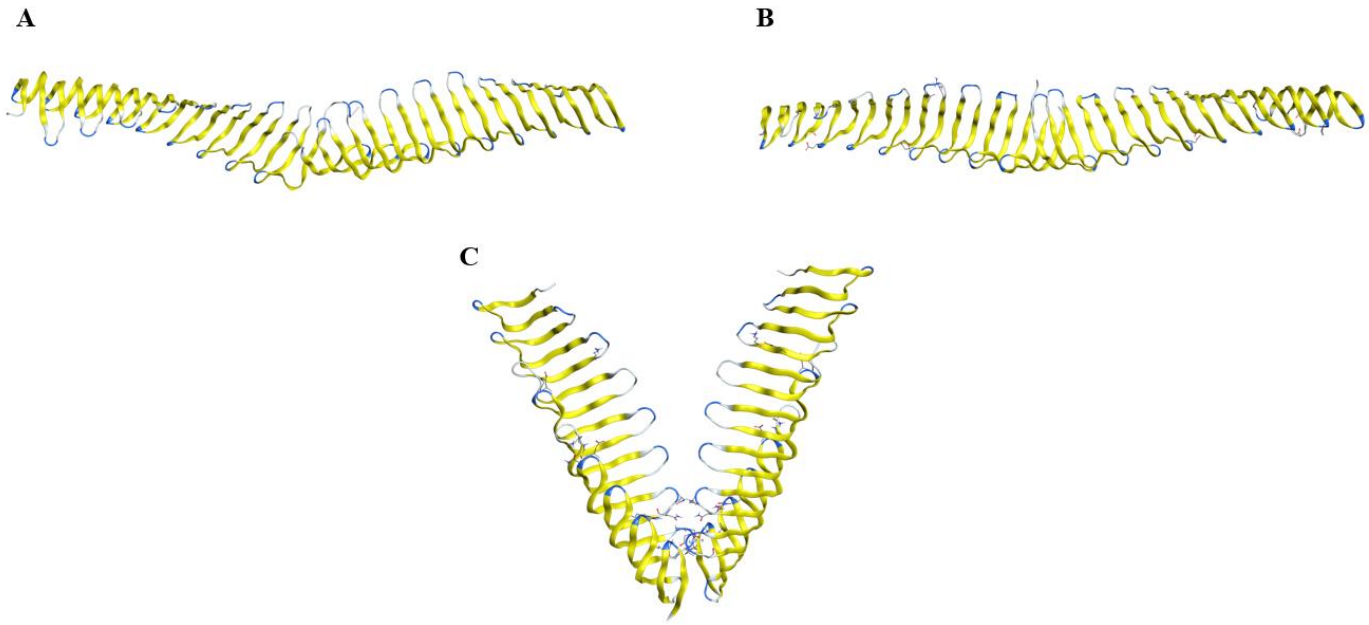

Figure 2. Cristallographic 3D structures of the dimeric form of TbMORN (A), TgMORN (B) and PfMORN (C)

Another evidence of the protein binding capability of the motifs comes from AtPIP5K2 which is reported to interacts with plant RAB-E proteins [11]. More specifically, it has been shown that a truncated form of AtPIP5K2 composed only by its MORN domain interacts with RAB-E1d more strongly than the full-length PIP5K2 protein whereas deletion of the MORN domain eliminates the interaction between the kinase and RAB-E proteins. Furthermore, the MORN domain exhibited the same specificity as the full-length protein towards the RAB-E subclass. This demonstrate that the MORN domain is necessary and sufficient for the specific interaction of PIP5K2 with the GTP-bound form of RAB-E proteins in yeast. Moreover, an experiment based on a GFP-tagged form of the amino-terminal domain containing the PIP5K2 MORN repeats shows that the PIP5K2 MORN domain is apparently insufficient for membrane localisation, as, failed to localise to the plasma membrane questioning the supposed lipids interaction capability of the domain.

2.4 Alsin MORN

In this work the focus is on the MORN domain of the Alsin protein (ALS2, Amyotrophic Lateral Sclerosis Type 2). This protein functions are still under investigation, but several studies suggest its involvement in the regulation of neuronal morphogenesis and maintenance, neuronal anticytotoxicity, signalling cascades, microtubule assembly, cytoskeleton organization,

endocytosis, vacuolar protein sorting, membrane transport and endosomal and mitochondrial trafficking [15]. Mutations in ALS2 protein are associated with Infantile-onset Ascending Hereditary Spastic Paralysis (IAHSP), Juvenile Primary Lateral Sclerosis (JPLS) and Juvenile Amyotrophic Lateral Sclerosis (JALS) diseases which are a family of rare hereditary motoneuron degenerative diseases called Hereditary Spastic Paralysis (HSP). IAHSP symptoms start to occur from the first years of life through lower limbs spasticity and typically worsen in the following years involving also the upper limbs reaching quadriplegia.

ALS2 sequence is composed by 1657 amino acids and, besides the MORN motifs, contains 4 putative GEF domains: RLD (aa 59-627) [16] [17] , DH/PH (aa 690-1007) [18][19], and VPS9 (aa 1513-1657) [20]. GEFs are known to promote the activation of the small GTPases by the cycle from inactive form (GDP-bound) and active form (GTP-bound), allowing cells to modulate various cellular and molecular processes.

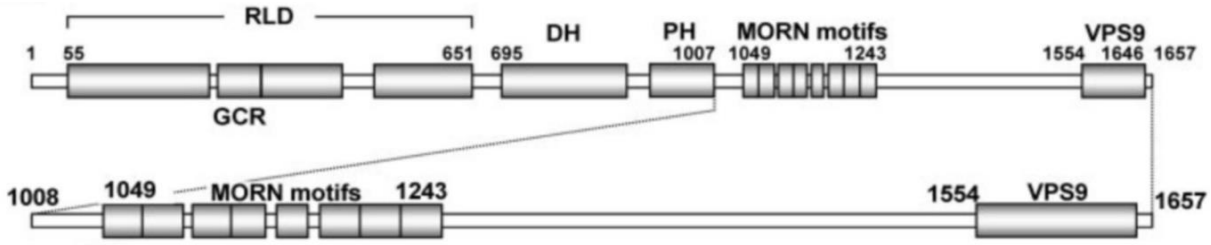

Figure 3. Graphical representation of Alsin amino acid sequence and its domains [16]

The ALS2 MORN motif is composed by 8 repetitions of 23 amino acids each [21]. Its function is unknown, but the repeats suggest an involvement in Rab5-GEF activity through an association of ALS2 with intracellular membranes [22]. Additionally, MORN motifs is involved in endosomal localization for ALS2 and together with the VPS9 domain are involved in the self-interactions of ALS2 by two C-terminal regions found within these domains (aa 1233–1351 and aa 1351–1454 in yeast cells) [23].

It has been found that ALS2 with only MORN/VPS9 region is still capable of homo-dimerizing or homo-oligomerizing in mammalian cells, demonstrating that the C-terminal region is essential for oligomerization [24]. At least six repeats in MORN are not necessary for oligomerization, but they are essential for the Rab5GEF activity [25].

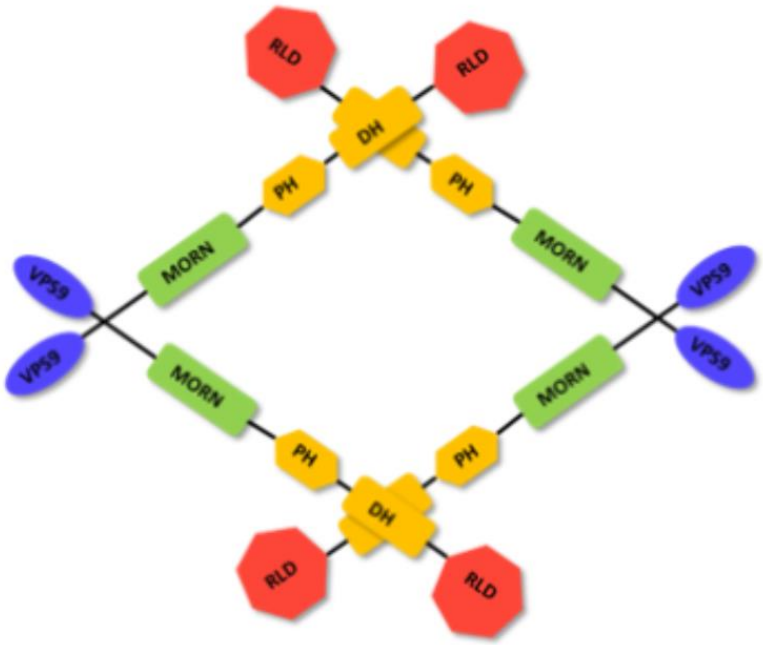

Figure 4. Graphical representation of the tetrameric form of ALS2 protein. MORN amino acid region 1233–1244 is crucial for the tetramer stability.

3 Materials and methods

The purpose of this chapter is to describe the core principles and methodologies of Molecular Modelling, i.e. the physical and theoretical foundations of the methods used in the present work. After a brief introduction, Homology Modelling and Molecular Docking are summarized, followed by a more detailed description of the underlying theoretical principles of Molecular Mechanics and their application in Molecular Dynamics.

3.1 Molecular Modelling

Molecular modelling is an interdisciplinary field that gather biology, physic, engineering and chemistry with the aim of study the physicochemical and mechanical behaviour of molecules, specifically proteins. Since biological systems are composed by a large number of particles and they are quite often complex, numerical methods are needed and implemented through computation. There are various tools in the Molecular Modelling field such as molecular mechanics, molecular dynamics, molecular docking, structure-activity relationships (SAR or QSAR), homology modelling, quantum mechanics and *ab initio* modelling. Each of those have their strengths and limitations and are often used in combination to increase the robustness of the model and/or have a better understanding of the results.

3.2 Homology modelling

Homology modelling is a technique trough which is possible to model the unknown 3D structure of a protein based on its amino acid sequence and a template. The template consists in another protein whose 3D structure model exist from experimental methods, like crystallography or NMR, and which share a certain degree of sequence identity and similarity. These structures can be found on publicly-available databases such as the Protein Data Bank [26]. To get a meaningful homology model two main requisites need to be matched: the identity/similarity percentage after sequence alignment [27] has to be the highest possible [28] and the resolution of the crystallographic/NMR model of the template must be sufficiently high. The rationale for the first requirement is due to the hypothesis that two proteins with almost the same primary structure share also a very similar 3D structure. However, since these percentages will never be 100% there are some non-overlapping regions that must be modelled by *de novo*. The second requirement is quite self-explanatory since if the template model is not quite accurate

itself it is not reasonable to think that the homology model obtained from it could be meaningful. Finally, the new 3D homology model obtained has to be optimized (ex. loops refinement, steric clashes) and validated. The entire process can be outlined as follow:

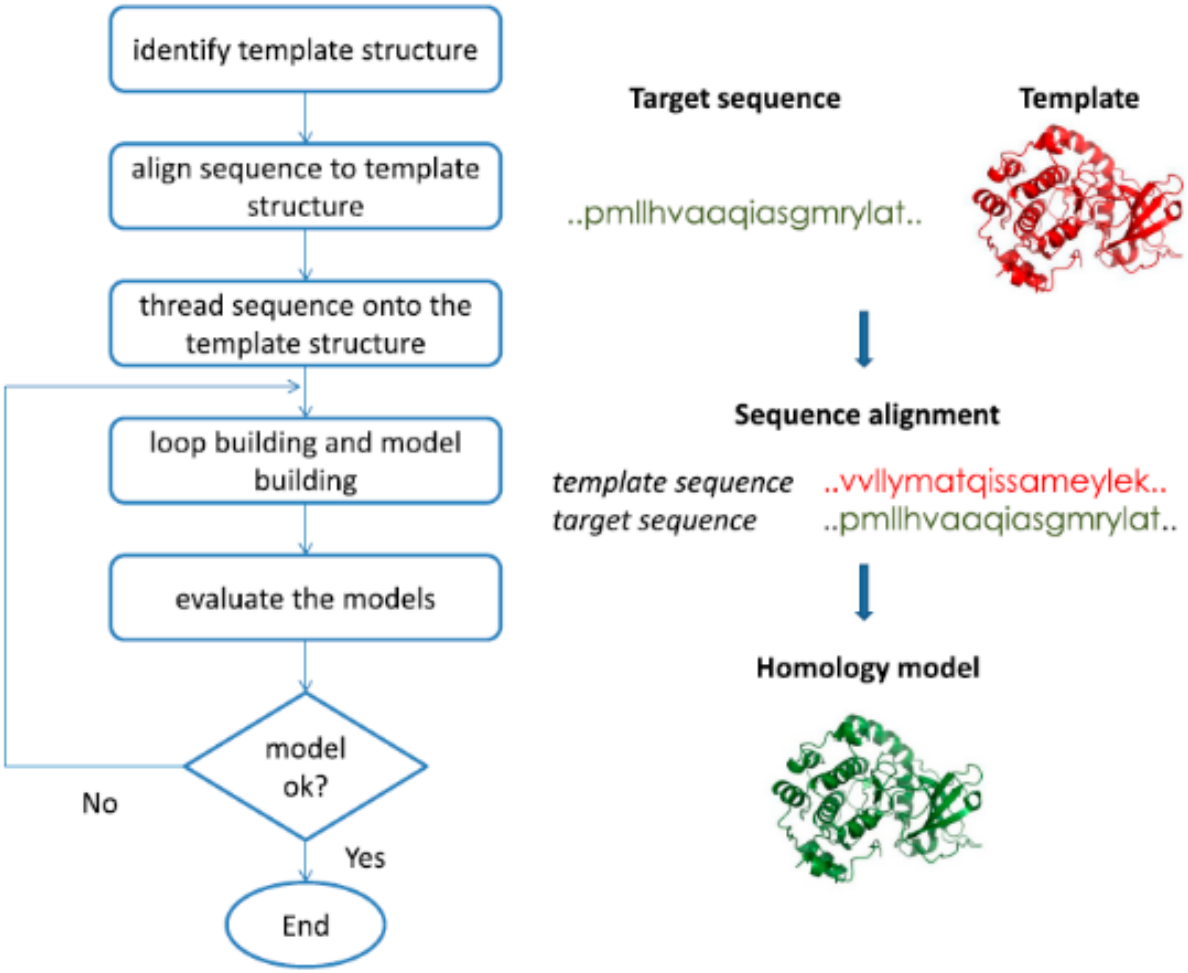

Figure 5: Flow-chart of homology modelling [29]

The optimization step usually consists in checking the stereochemistry of the structure, for example trough a Ramachandran plot is possible to see if the backbone angles phi and psi fall within favoured, or at least, allowed regions and not disallowed regions. Another additional optimization process can be the loop refinement. This can be important especially for drug design since active binding site of proteins is often quite flexible and affects the predicted physical and biological properties of a protein.

For what concern software implementation for modelling different environments are available, such as MOE [30], UCSF Chimera [31], Modeller [32] or SWISS-MODEL [33], which provide

numerous options to set up optimally the model generation. Some packages (ex. MOE, UCSF Chimera) integrate optimization tools beside the modelling ones. Additionally, optimization softwares are available, for example PROCHECK [32] and WHATCHECK, to assess the quality of the obtained model

3.3 Molecular Mechanics (MM)

Molecular mechanics is a method of studying molecular systems applying Newton's laws to every atom under the approximation of ignoring electrons and considering the nuclei as classical particles. The term was introduced in 1970 and was based on the pre-existing method called forcefield method.

The fundamentals to build a MM model are the system geometry, the definition of atom types (that includes information about each atom mass, charge, ibridation, etc.) and the description of interactions between atoms as masses interconnected by springs. This allows to characterize the evolution in time of the energy of the molecular system through the potential energy function (PEF) whose parameters differs according to the implemented mathematical model known as force field (FF) [34]. The general expression of the PEF takes in consideration two classes of interactions that are function of atoms position and can be describe as follows:

$$\mathcal{V}(r^N) = \mathcal{V}_{bond}(r^N) + \mathcal{V}_{non-bond}(r^N) \quad (1)$$

A central concept is the potential energy surface (PES) which expresses the relationship between the energy of a molecular system and its geometry (atoms positions). This multidimensional surface is "explored" during simulations to find minima which correspond to more favorable, and thus probable, structural conformations.

3.3.1 Bonded interactions

Bonded interactions arise from stretching or shortening of atomic bonds, angle variations and torsional motion around single bonds (also called dihedral). Their contribution to the PEF can be mathematically described as:

$$V_{bond}(r^N) = V_{bonds}(r^N) + V_{angles}(r^N) + V_{torsion}(r^N) \quad (2)$$

Each of the right-side terms can be modelled in different ways. The first one describes the variation of the PEF based on the length of covalent bonds between atom pairs and it is often described using Hook's law:

$$V_{bonds} = \sum_{angles} \frac{1}{2} K_l (l - l_0)^2 \quad (3)$$

K_l is the elastic constant which represents the strength of the bond expressed in kcal mol^{-1} Å^{-2}, l is the actual bond length and l_0 is the reference bond length for that specific couple of atoms.

Second term account for variations in angles formed by a triplet of atoms i-j-k, where j is bonded to i and k, and it is, as the previous term, modelled using harmonic relationship:

$$V_{angle} = \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (4)$$

K_θ is the bond stiffness constant measured in kcal mol^{-1} deg^{-1}, θ is the actual angle value and $\theta_0$is the reference bond angle.

The last term represents the contribution of torsion angle of a dihedral (a group of four bonded atoms) defined as the angle between the plane identified by three atoms and the fourth atom. This contribute is mathematically expressed as:

$$V_{dihedral} = \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi - \delta)] \quad (5)$$

K_φ is the barrier depth which represent the energetic cost related to the angle deformation, n is the multiplicity, that correspond to the number of energetic minima along a 360°, and δ is the phase that determines the minimum position for the torsional angle.

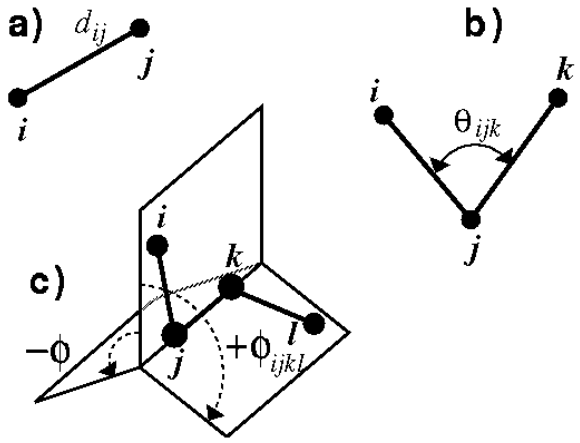

Figure 6. Representation of (a) bond length between atom i and j; (b) bond angle between atoms i, j and k; (c) dihedral angle between the four atoms i,j,k,l.. Source: ccl.net/cca/documents/molecular-modeling/node4.html

3.3.2 Non-Bonded interactions

Non bonded interactions describe the modality through which single atoms interact with each other without chemical bonds. These are usually modelled as functions inversely proportional to the distance between two atoms and can be expressed as the sum of two separate functions:

$$V_{bond}(r^N) = V_{van-der-Waals}(r^N) + V_{electrostatic}(r^N) \quad (6)$$

Van-der-Waals (VdW) term account for the physical properties of small molecules and atoms of attracting each other at short distance up to a certain range where the electron clouds start to overlap and attractive forces become highly repulsive ones. These interactions are modelled by Lennard-Jones 12-6 function:

$$V_{LJ}(r_{i,j}) = 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] \quad (7)$$

$\varepsilon_{i,j}$ is the is the VdW force minimum and $\sigma_{i,j}$ collision diameter. Since these two constants are dependent on the specific atom for which are experimentally calculated, a system with N different atom types require a set of N(N-1)/2 parameters to evaluate the cross-interactions. The calculation of these cross-interaction is very time consuming and it is often done using mixing rules for the two different parameters:

$$\varepsilon_{i,j} = \sqrt{\varepsilon_i \varepsilon_j} \quad (8)$$

$$\sigma_{i,j} = \frac{\sigma_i + \sigma_j}{2} \quad (9)$$

The electrostatic term describes interactions of charged non-bonded atoms. In particular, the difference in electronegativity causes the less negatively charged atoms to be attracted to more negatively charged ones causing a charge distribution. This phenomenon is modelled by partial atomic charges placed conveniently on the nucleus and its contribution to the PEF is calculated by the Coulomb's law:

$$V(r_{i,j})_{electrostatic} = \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (10)$$

q_i and q_j are the partial charges on the i-th and j-th atoms, ϵ is the dielectric constant and r_{ij} is the distance between atoms i and j.

Considering the great number of interactions to calculate (N^2 interactions for a N atoms system), non-bonded interactions are very expensive from a computational standpoint. For this reason, several different methods to reduce and speed up the process have been proposed. The most basic idea is to use the cut-off method which consist in calculate this type of interactions only within a selected radius of each atoms and ignoring the rest. The choice of the cut-off threshold

has to be made carefully since improper values leads to system instability. To reduce this possibility other methods have been proposed such as switch cut-off, shifted cut-off or particle mesh Ewalds.

3.3.3 Periodic boundary conditions

To correctly simulate an atomic system, interactions have to be calculated as if particles where in an infinite space but for obvious reason this is not the case. In fact, biological systems are simulated in a box of different shapes and sizes filled with water molecules and so there is a spatial limit. This problem is solved by implementing periodic boundary conditions (PBC) which allow to avoid edge effects by periodically repeat the box and its content in all directions. With PBC atoms see always other atoms periodic replica from neighbour boxes and, additionally, when one or more atoms move out of the central box their replica move in so that the number of particles in the box remain constant in time.

To implement this technique, the minimum image convention must be consider meaning that it is necessary to impede that a particle interacts with itself through its periodic replica. This means that when using a cut-off scheme to treat non-bonded interactions (see previous chapter) there is an upper limit to the threshold value. In fact, cut-off radius Rc must be less than half of the minor side of the box.

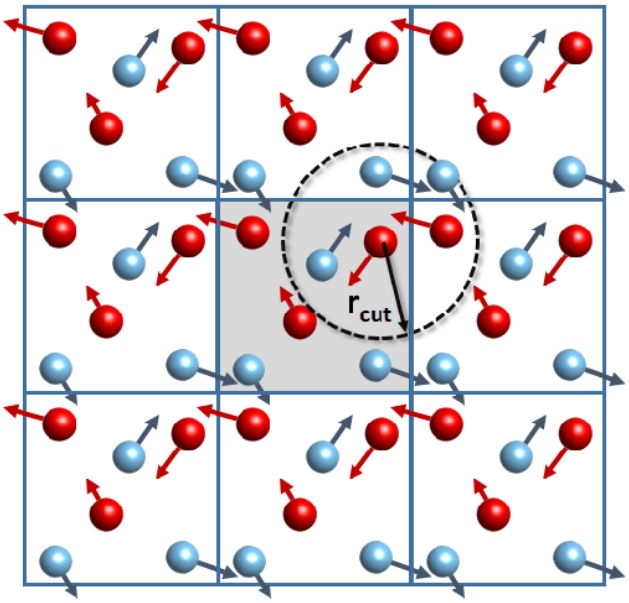

Figure 7: representation of PBC, the central box is surrounded by his replica [35]

3.3.4 Potential energy minimization

One of the main goals of molecular modelling techniques is to find favourable conformations of biological structures which correspond to minimum-energy arrangements. Moreover, this is a crucial step before every molecular dynamic simulation since high-energy structures often cause simulation crash and failure. The PEF is a multidimensional function and, more precisely, for a system of N atoms described by Cartesian coordinates, it is function of 3N coordinates and 3N-6 internal coordinates for the degrees of freedom. For this reason, its graphical representation through PES can be done only for a restricted set of parameters. The PES presents different *local minima*, that correspond to low energy configurations, and a *global minimum*, which corresponds to the lowest energy arrangement that the structure can assume. Thus, the process of energy minimization (also called optimization) is the search of one or more of these points in the multidimensional potential energy surface which are characterized by a zero value for the first derivative with respect to each variable and a positive value for the second derivative in respect to each variable:

$$\frac{\partial f}{\partial x_i} = 0 \ ; \ \frac{\partial^2 f}{\partial x_i^2} > 0 \tag{11}$$

Different local search algorithms have been proposed to investigate the PES, starting from the point of the initial structure, and they can be categorized either as non-derivative approaches or derivative approaches. In the former group are included *simplex method* and *sequential univariate method*. The latter group can be divided once more in two classes*: first order derivative methods* and *second order derivative methods*. The first one moves along the direction of the gradient giving information about the position of the minimum. This group includes *steepest decrescent method* [36]*, one-dimensional line search*, *arbitrary step approach*, and *conjugate gradients minimisation* [37]. The second group account methods that can foresee where there will be a change in the direction of the PES trough second derivate calculations. *Newton-Raphson Method* [38] and *Broyden-Fletcher-Goldfarb-Shann* (L-BFGS) [39] are part of this

class. Since each of these methods has its own advantages and drawbacks the choice of which one to implement depends on the specific application and on the available computational power.

3.4 Molecular Dynamics

To fully characterize biological systems potential energy minimization is only a first step because does not describe all the dynamic processes that leads to their thermodynamics properties. Molecular dynamics (MD) is the computational approach that, using Newton's laws and MM principles, try to retrieve these characteristics by generating a trajectory of the system as a function of atoms position and velocities. To properly understand MD simulations, statistical mechanics concept has to be introduced.

The *phase space* used to describe a N atom structure is defined as the collection of every possible configuration of the system and each point of this space, which correspond to a specific state, is described by 3N position and 3N momenta for a total of 6N parameters. Each point in the phase space represents a *Microstate* and each collection of Microstate with the same macroscopic properties is defined as a *Macrostate*. A collection of phase space points with the same Macrostate is defined as a *Statistical ensemble*. In MD simulations different ensemble are used:

The Micro-canonical Ensemble (NVE) describes an isolated system, it is characterized by a fixed number of particles (N), an assigned volume (V) and a constant Energy.

The Canonical Ensemble (NVT): corresponds to a closed system coupled in temperature with a thermostat, so fixed number of particle (N), assigned volume (V), and constant temperature (T).

The Isothermal-Isobaric ensemble (NPT) describes an isolated system coupled in temperature and pressure, with fixed number of particles (N), constant temperature (T) and pressure (P).

The Gran-Canonical Ensemble (μVT) correspond to an open system with fixed volume (V), temperature (T) and chemical potential (μ).

Since generally macroscopic properties depend on the microstate of the system, each property A can be described as a function dependent on positions (r) and momenta (p):

$$A = A(p^N, r^N) \tag{12}$$

Given that, it is possible to obtain the *ensemble average* of the property <A> by integrating over all possible configuration of the system:

$$< A >_{ensemble} = \iint A(p^N, r^N)\rho(p^N, r^N)dp^N dr^N \tag{13}$$

Where $\rho(p^N, r^N)$ is the probability density of the ensemble, which depends upon the chosen statistical ensemble. The crucial problem of this step is that generally it is impossible to know every possible state of the system because that would imply the knowledge of the entire PES. For this reason, it is necessary to assume valid the ergodicity hypothesis which states that ensemble averages can be replaced by time averages for long enough sampling [40] of a certain property:

$$< A >_{ensemble} = < A >_{time} \tag{14}$$

$$< A >_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t))dt \tag{15}$$

This allow to use MD trajectories to calculate macroscopic properties under the assumption that the M steps sampling of the simulation is large enough to obtain a sufficient description of the property of interest from the phase space. In fact, by numerically integrating over the M steps it is possible to approximate as follows:

$$< A >_{time} \approx \frac{1}{M} \sum_{i=1}^{M} A(p^N, r^N) \tag{16}$$

Since the generation of the trajectories is based on solving the Newton's laws and that each point of the phase space is calculated from the previous one, MD simulations can be defined as a deterministic method. At each step acceleration a of each atom i, is obtained by the derivative of the potential surface V in respect to its position r:

$$a = \frac{d^2r}{dt^2} = \frac{1}{m}\frac{dV}{dr} \tag{17}$$

The high dimensionality of the PES makes analytical solution of this differential equation impossible so that the equations of motions are integrated using *finite difference method* which divide integration in small stages, each separated from a fixed time-step *δt* [41]. These methods include algorithms such as *Verlet scheme*, *Velocity Verlet* and *Leap frog*. The choice of the integration algorithm is crucial as well as the correct setting of its parameters. In particular, the time step δt must be tuned carefully since it has to be quite short to sampling correctly the phase space and to avoid instability but not so small to make the simulation time unfeasible. A first approximation criterion would be to set δt as ten times smaller than the faster oscillation in the system [42], that in an all-atom simulation is usually related to Hydrogen atoms. A general scheme of a simulation is:

- Solvation inside a box (cubic, dodecahedral, …)
- Addition of neutralizing counterions
- Addition of ions to reach physiological salt concentration
- Energy Minimization of the system
- Equilibration at constant temperature (NVT)
- Equilibration at constant pressure (NPT)
- Molecular dynamic simulation and Trajectory production
- Removal of PBC artifacts
- Analysis of results

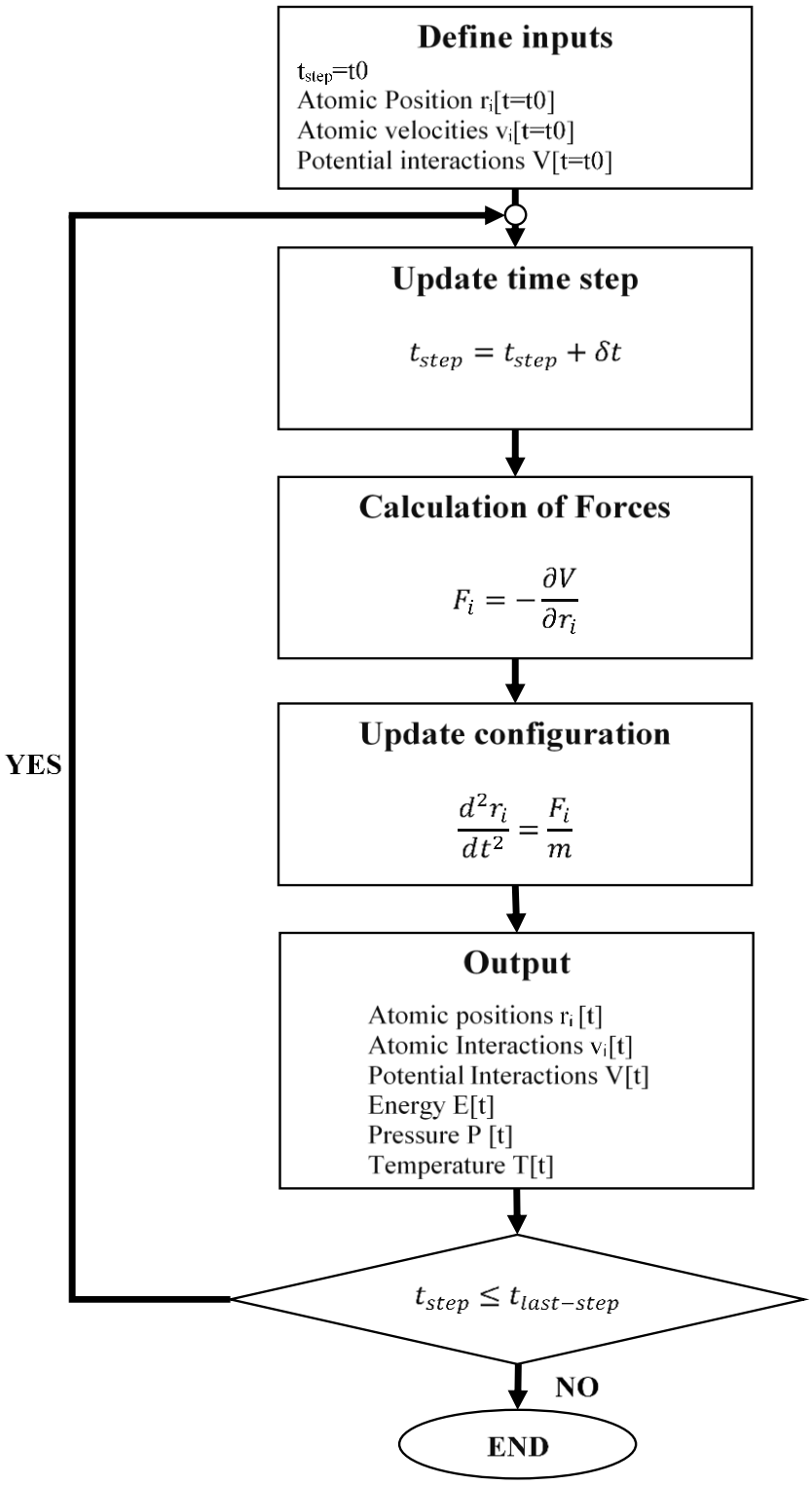

Figure 8: General scheme of Molecular Dynamics algorithm, first input parameters must be defined from an initial configuration. Then a loop on the chose number of steps is performed, for each step force are calculated as the derivative of potential energy, and then position and velocity are calculated integrating Newton's laws, the output

for each step is new atomics positions, velocities, energy, pressure and temperature at the current step. The loop is repeated until total number of steps are performed.

3.5 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a technique useful to reduce the high dimensionality of MD trajectories, which makes otherwise harder analyse them, while preserving as much of the data's variation as possible.

It is a statistical method that define a covariance matrix of the atoms position and, after diagonalizing it, proceed to calculate 3N orthonormal eigenvectors and eigenvalues. After that, the eigenvectors (e_j) are listed according to descending eigenvalues (σ_j^2). The eigenvectors correspond to the principal component of the motion and the firsts ones are the components that mainly describe the complex motion of the system.

3.6 APBS

Electrostatic properties and solvation are fundamental for the correct characterization of biological molecules and structures which are always surrounded by the physiologic environment. The methods which are used for modelling these two characteristics can be divided in two classes:

- *Explicit solvent methods* reproduce water properties by modelling H_2O molecules through its atomic component. For each molecule, its coordinate and some degree of freedom are considered to describe interactions. This allows to describe with great accuracy solvent characteristics but require large computational resources.
- *Implicit solvent methods* describe water behaviour approximating it by a homogeneous polarizable continuum medium with the closest possible properties. This allows to describe solvent characteristics with only a small number of parameters, but it is not as accurate as the explicit description.

Apbs is a software that implement implicit solvent models both polar and non-polar [43] [44] [45]. This is based on the combination of a non-polar model, to describe attractive solute-solvent dispersion terms, and the Poisson-Boltzmann equation (polar) which is a nonlinear

elliptic partial differential equation that is solved for the electrostatic potential within and around the biomolecule. The PB equation has the following formulation:

$$-\nabla \cdot \epsilon \nabla \phi - \sum_{i}^{M} c_i q_i e^{-\beta(q_i \phi + V_i)} = \rho \tag{18}$$

ϵ: dielectric coefficient function

ϕ: electrostatic potential

M: number of different mobile ion species

c_i: concentration of the i-th ion species

q_i: charge of the i-th ion species

V_i: steric ion-solute interaction potential

ρ: charge distribution function

$\beta = (kT)^{-1}$

The integration in the same framework of these two models allows to obtain a more accurate electrostatic potential calculation. The Apbs output can be easily analysed using molecular visualization software such as Visual Molecular Dynamic environment [46] or PyMOL [47] where the potential map can be integrated with the corresponding biomolecular structure for a better understanding of its properties.

.

4 Structural and conformational characterization of three MORN protein assemblies

4.1 Abstract

Characterization of MORN domains is the first step to a better understanding of their structural properties and might clarify their debated biological functions whether it a lipid binding module or a mediator for protein-protein interactions. In this work we investigated the second hypothesis starting from previous reported crystallographic MORN proteins which shows homophilic interaction through the formation of molecular superassemblies. Implementing Molecular Dynamics methods we evaluated the behaviour of these models in physiological conditions. Subsequent analyses were made including conformational characterization by feature angles and Principal Component Analysis revealing common aspects between each structure. In addition, electrostatic potential maps of the surfaces were generated to further investigate the interaction interfaces. Interesting results were obtained from one of the three structures which present a Zn ion that stabilize the assembly, where some the previously reported amino acids of the binding region unveil an expected behavior under dynamic conditions. Finally, a comparison of the assemblies was made by performing Boltzmann Inversion analysis to evaluate their respective Potential Mean Force.

4.1.1 Introduction

Lacking structural and biological information about MORN domains is the first obstacle for the characterization of these ubiquitous motif found in both eukaryotes and prokaryotes [1]. The number of repetitions vary greatly for each motif, and they are almost always found in combination with several different domains making it difficult to understand their role in biological activities. The original hypothesis of lipid-binding functions is debated since experimental data are ambiguous and only indirect evidence of this supposed capability are reported [2], [6], [10], Recent studies [7] unearth a new hypothesis which suggests that MORN motifs are essentially protein modules mediating homophilic interactions. Supporting this idea crystallographic assemblies have been obtained for three different all-MORN proteins found in the parasite organisms of *Trypanosoma brucei*, *Toxoplasma gondii* and *Plasmodium falciparum*. These 3D structures consist in dimers which present two different spatial arrangements: a linear conformation for the first two and a V-shaped one for the latter.

The use of computational methods to investigate the structural features and the interactions between proteins may help to further investigate and support the new biological function hypothesized, laying the way to a more in depth understanding of such peculiar domains.

In the present work conformational characterization of the above-mentioned crystallographic MORN assemblies is presented through the application of Molecular Dynamics methods, with the specific focus on structural stability, motion analysis and on the stabilization role of the Zn ion present in the V-shaped assembly is reported.

4.2 Materials and methods

4.2.1 Crystallographic structures

Three crystallographic structures form the Protein Data Bank databased were selected being the only ones presenting protein assemblies.

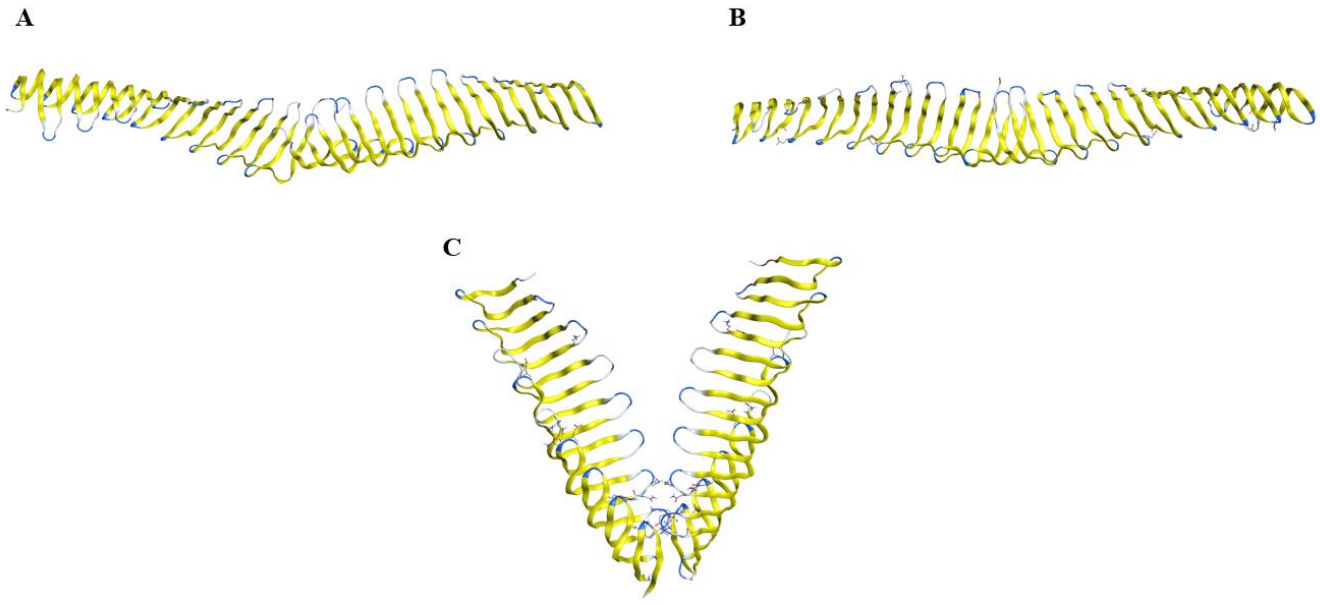

Figure 9. Crystallographic structures of TbMORN (A), TgMORN (B) and PfMORN (C). All three present the conserved secondary and tertiary structure composed by repetition of β-harpin subunits.

In this work models are referred to as follows: TbMORN (PDB id:6T4R , [7]) for the *Trypanosoma brucei*, TgMORN (PDB id:6T6Q, [7]) for the *Toxoplasma gondii* and PfMORN (PDB id:6T4D, [7]) for the *Plasmodium falciparum*. They do not present missing residues into their

3D structures and had been obtained using X-ray diffraction technique with a resolution of 2.352 Å, 2.902 Å and 2.14 Å respectively. Quality checking of each model was performed using PROCHECK software to produce Ramachandran plots describing the phi and psy angles distribution of each peptide bond in the sequences.

Before Molecular Dynamics simulations structures were prepared using the QuickPrep function of MOE 2020.09 [30] software package under the conditions of Amber10:EHT forcefield [48].

4.2.2 Molecular Dynamics

Three Molecular Dynamics simulations were performed, one for each MORN protein. GROMACS 2020.4 [49] engine was employed for MD simulations. Topologies have been build employing AMBER ff99SB-ILDN force field [50] The general system was configured in GROMACS in a dodecahedral box with periodic boundary conditions and a minimum distance between the protein and the box edge of 1.5 nm. All systems have been solvated using explicit TIP3P water model [51]. Additionally, to reach physiological salt concentration of 0.15M, an appropriate amount of Na^+ and Cl^- counterions were added to the solvent and to neutralize the total charge of the systems. Energy minimization step was carried out implementing the steepest descent algorithm for 50000 steps. Subsequently, two position restrain simulations were performed for each system, limiting C-alpha carbons positions, in two different ensembles: first an Constant volume and temperature (NVT) equilibration of 100 ps was carried out after which an Constant number of particles, pressure and temperature (NPT) simulation for another 100 ps was performed. The reference temperature and reference pressure for both NVT and NPT equilibrations was T=300K and P=1.0 bar. For the NVT ensemble, the modified the Berendsen thermostat [52] was used (τ=0.1), while for NPT equilibration the Parrinello-Rahman coupling (τ=2.0) with isotropic coupling was selected. Finally, an MD production in NVT ensemble was performed for 200 ns with integrating the equation of motion by the leap-frog algorithm with a timestep of 2 fs. Electrostatic have been treated using Particle Mesh Ewald [53], with an interpolation order of 4 and an FFT grid spacing of 0.16 nm and a cutoff of 1 nm for both Coulomb and Van der Waals interactions. This workflow was applied to each of the three MORN assemblies. After the simulation Route Mean Squared Deviation (RMSD), Route Mean Squared Fluctuation (RMSF) and radius of gyration (Rg) analysis were carried out to assess the stability of the proteins. For the RMSF, the calculation was made for the single chain to prevent artifacts due to global structure movements.

Visual inspection of the systems and their trajectories was made by using The Visual Molecular Dynamics (VMD) [46] environment.

4.2.3 Analysis

To assess the stability of the three structures, GROMACS integrated tools were used. First of all, the Root-mean-square deviation (RMSD) of the protein backbone atomic positions was analyzed throughout the entire trajectory.

After that, each single chain of the models was investigated calculating the Root Mean Square Fluctuations (RMSF) for each residue. Additionally, oscillations in the radius of gyration were measured to further evaluate the dimensional stability of the models.

Characterization of the hinge angle between the two chains and torsional angle of the single subunit were carried out to investigate dynamical behavior of the conformations. For this analysis a custom script using Python 3 numpy libraries was implemented trough which it was possible to calculate both the angles' values for each frame of the GROMACS produced trajectory. For each angle both probability distribution and variation in time function were calculated by using center of masses of visually selected residues groups. For hinge angle measurements, the following amino acids where grouped: TbMORN {0-26, 185-211, 213-239}, PfMORN {0-26, 175-201, 203-229}, TgMORN {0-26, 182-208, 211-237}. Groups for torsion angle were: TbMORN {4-20, 9-13, 187-204, 194-198}, PfMORN {4-20, 9-13, 165-182, 171-175}, TgMORN {3-20, 9-13, 188-205, 194-198}.

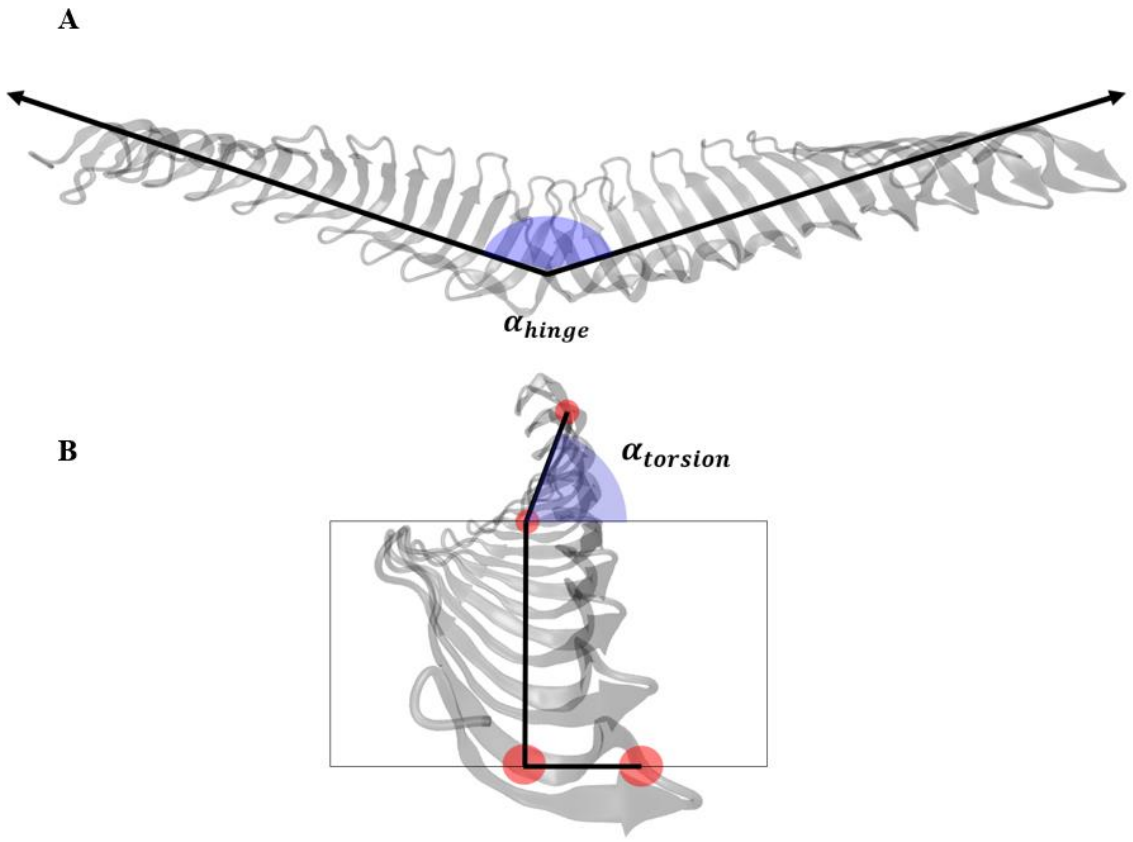

Figure 10. Description of the investigate angles: hinge angle between chains (A) and torsional angle of the single chain (B).

As an additional analysis, Principal Component Analysis was implemented using GROMACS. Moreover, calculations of the electrostatic potential maps were performed using a custom-made script which integrate a .pqr file generator and the APBS software [43]. Zinc ion distance analysis for the PfMORN structure were evaluated by using the GROMCS tool 'mindist'.

From the script results, evaluation of free energy changes as a function of the angle variation were made through the calculation of the potential of mean force (PMF) starting from the angle probability distribution obtained after calculating the angle for each frame of the trajectory. The angle probability distribution was calculated as:

$$p(\alpha_{hinge}, \alpha_{torsion}) = \frac{H(\alpha_{hinge}, \alpha_{torsion})}{\sum_{\alpha_{hinge}} \sum_{\alpha_{torsion}} H(\alpha_{hinge}, \alpha_{torsion})} \qquad (192021)$$

Where α_{hinge} is the angle between the two chains, $\alpha_{torsion}$ the torsional angle of the single chains. Discretization of the state space along the two coordinates was made by using bins of 1° for both the angles. Boltzmann Inversion (BI) is used to calculate angle-associated potential, starting from the angle probability distribution. The corresponding PMF function was then calculated as:

$$PMF(\alpha_{hinge}, \alpha_{torsion}) = -k_B T ln\,[p(\alpha_{hinge}, \alpha_{torsion})] \quad (22)$$

Where k_B is the Boltzmann constant and T is the system temperature assumed to be 300 K.

4.2.4 Plots and Figures

Three-dimensional representations of the MORN assemblies were rendered using the VMD environment. RMSD, RMSF, radius of gyration and angles data plots as well as contact probability maps and PMF figures were generated using Python 3 custom made scripts using numpy library functions. PCA plots were generated in Microsoft Excel.

4.3 Results

4.3.1 Molecular Dynamics

System stability during MD simulations of the different protein assemblies was checked through examination of the RMSD over the 200 ns trajectories. The three proteins reached equilibrium at different times with the faster being the TbMORN (20 ns) followed by the TgMORN (less than 90 ns) and PfMORN (100 ns). Further analysis on the RMSF, calculated on the last 50 ns of the single chains showed several peaks in each structures in the positions corresponding to the loops connecting each β-strand and major peaks in the N-termini. These last fluctuations have to be attributed to the highly fluctuating nature of these regions. Nonetheless, each chains reported almost identical behavior to their counterpart. The most stable is the TgMORN with a maximum fluctuation of about 0.3 nm at the amino-terminal region, while for the PfMORN and TbMORN the highest peaks are respectively of little less than 0.6 nm and almost 0.8 nm. Additional confirmations of the stability of the structures are the gyration radius' behavior which, after the equilibrium is reached, shows only little fluctuations (TbMORN=4.66 ± 0.04 nm, PfMORN=2.91 ± 0.09 nm, TgMORN=4.64 ± 0.04 nm).

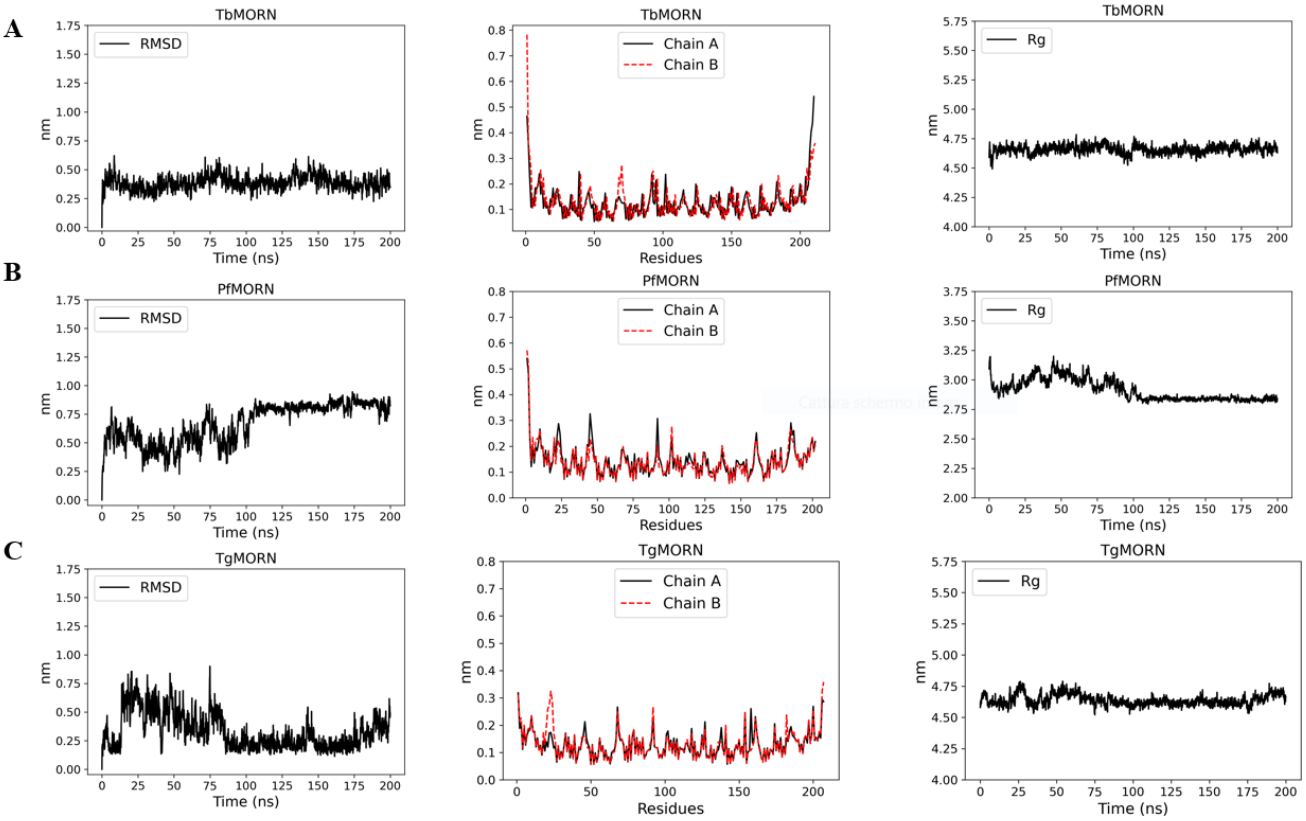

Figure 11 Post MD analyses for TbMORN (A), PfMORN (B) and TgMORN (C). For each protein, plots are in the following order RMSD (left), RMSF (center) and Rg (right).

4.3.2 Characterization of angles and motions

Results obtained from the custom-made script for the hinge angle show that structures possess specific bending values which remain stable after equilibrium. The two linear dimers, TbMORN and TgMORN, possess very similar angles, 150-155˚ and 165-170˚ respectively, generating an almost straight configuration. The PfMORN on the other hand shows a different conformation with an angle value of 25-30˚ which produce a very closed V shape.

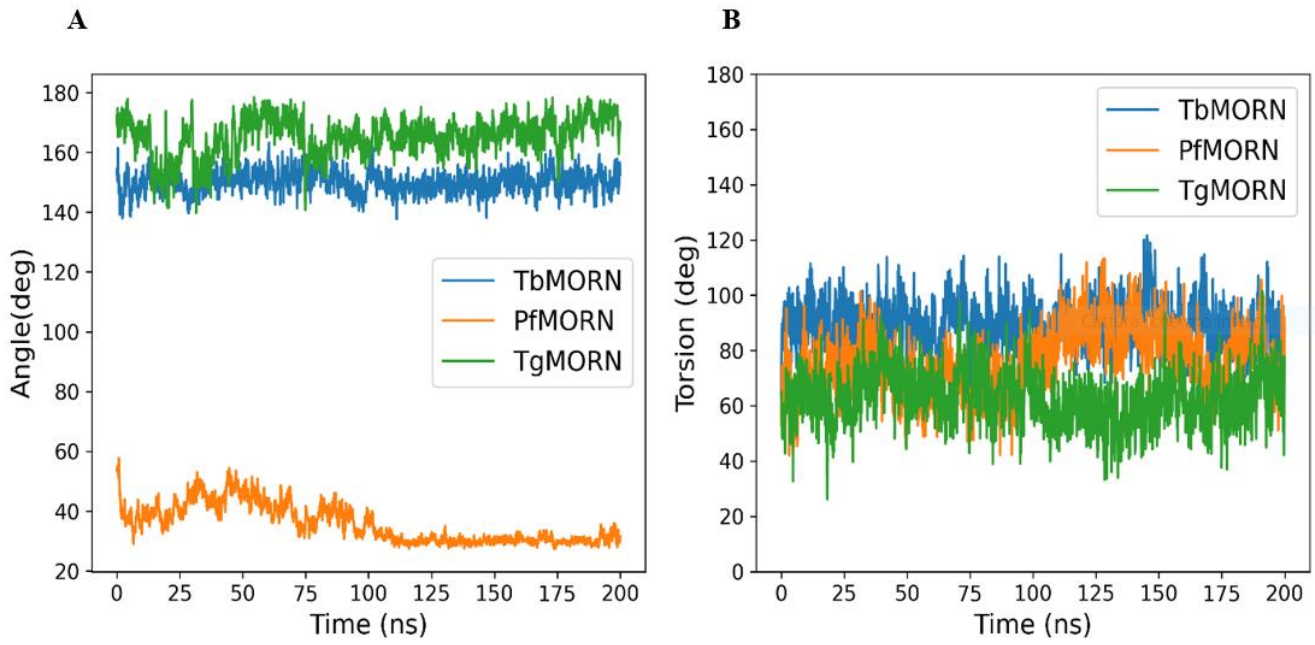

Figure 12. Hinge (A) and torsional (B) angles variations plotted as function of time. Despite the differences in the former angle every structure shows torsional motion in its chains.

Further characterization was made by investigating the internal torsional angles of each assemblies' subunit. Chains appear to be subjected to an internal torsional motion which is quite similar for every chain of the three considered proteins. In fact, the highest value is associated with the TbMORN which torsional angle fluctuate around 90˚, followed by the PfMORN and then the TgMORN, with measurements around 80˚ and 65˚ each.

Principal Component Analysis performed on the MD trajectories revealed movement pattern that appear strictly correlated to the arrangements of the assemblies and the angle measurement. As a matter of fact, TbMORN and TgMORN show motions mainly described by the first two principal components (42.5% and 29% for the former, 48% and 41% for the latter) which in both cases are bending motions. Additionally, TbMORN is characterized by a torsional motion of the chains as its third principal component account for 15.6% of its total dynamic behaviour. In a similar fashion, PfMORN is dominated by its first component, accounting for about 80% of the total motion, which represents an opening motion of the angle with a slight rotation of the two chains.

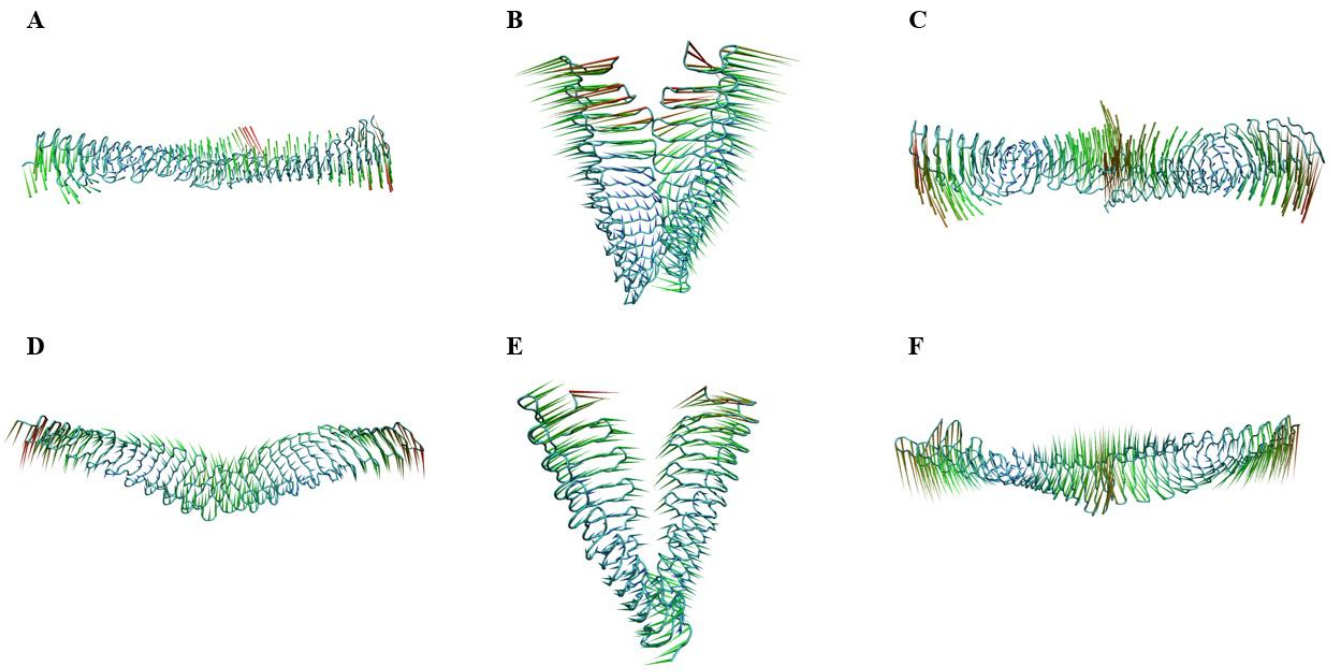

Figure 13. Porcupine plots representations of first two principal motional components for TbMORN (A-D), PfMORN (B-E) and TgMORN (C-F).

Finally, since the investigation of principal motion components showed that conformation changes are mainly caused by angular variation in time, the two angles α_{hinge} and $\alpha_{torsion}$ were chosen as coordinates to evaluate the associated PMF through Boltzmann Inversion of their probability distribution. Colormaps representation confirms our previous results on structural stability showing limited continuous regions. Specifically, the two linear motif MORNs possess very similar PMFs with larger potential energy minima than the PfMORN which is expected since they possess grater α_{hinge} fluctuations. On the other hand, the V-shaped assembly reported a more restrictive minimum, which develops more along the $\alpha_{torsion}$ axis between 80 ˚ and 100˚ with a specific hinge angle of about 30˚. Overall, the lowest energy configurations are the ones of the TgMORN and PfMORN which reported the very similar energy minima with the former one reaching the lowest value.

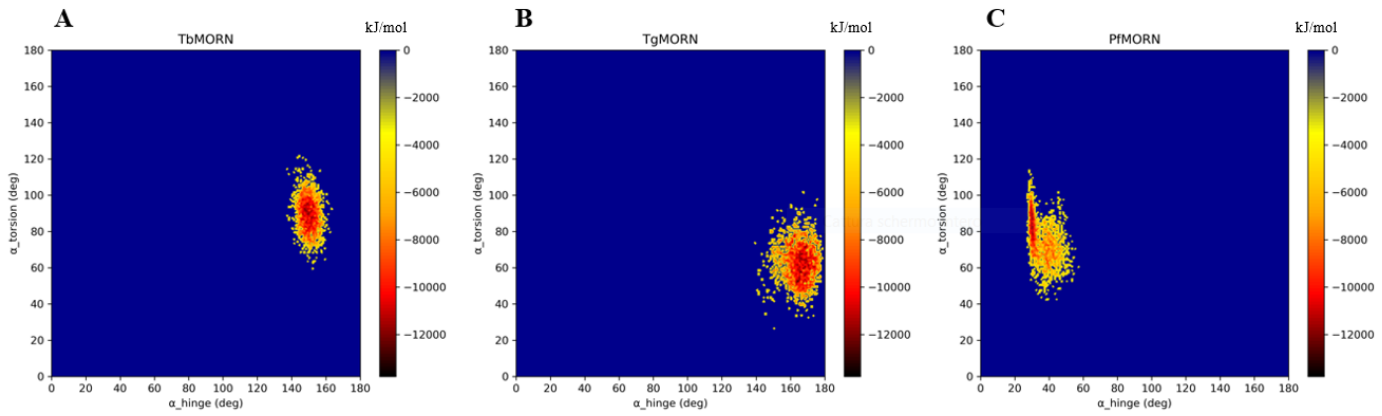

Figure 14. Colorimetric maps of the three structures PMFs. As expected TbMORN (A) and TgMORN (B) show very similar maps with larger low energy regions while PfMORN (C) report a smaller minimum region which suggest less conformational changes.

4.3.3 Electrostatic potential mapping and Zn ion investigation

To extend the investigation of these structures an electrostatic potential map was generated for each of them. The representations shows that every model shares a net negative potential on the surface and that the distribution of charges is not homogeneous. At the dimerization interface the potential of the TbMORN and PfMORN shows a common motif with a central positive patch surrounded by a negative area, whereas the TgMORN present a prevalently positive region.

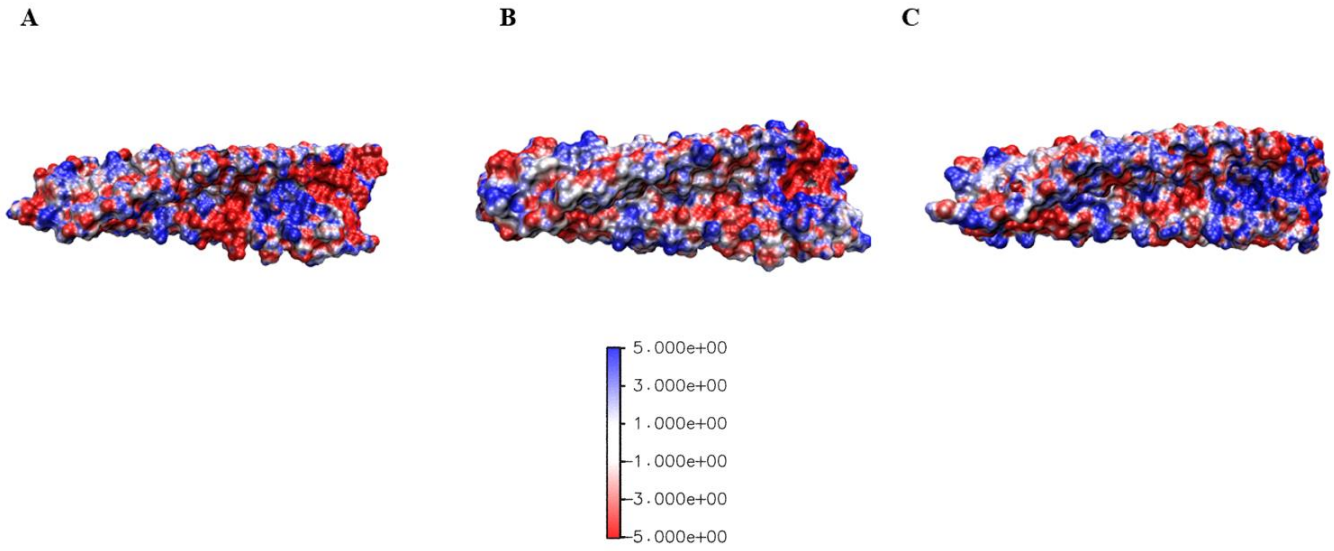

Figure 15. Electrostatic potential maps of the interaction (inner) side of the models. TgMORN (C) is the only structure characterized from a positive potential interface whereas TbMORN (A) and PfMORN (C) reported negative interaction surfaces.

Since the dimeric assembly of the PfMORN contains a Zn^{2+} ion between the interaction regions, which is suggested to be important for its conformational stability, and that previous experimental evidence reported that the same ion was found in a TgMORN assembly with a V-shape conformation, a further investigation was carried out. Sajko et al. [7] suggested that the interaction with the Zn ion is mediated by a conserved quadruplet-sequence found in both proteins formed by C-x-E-D amino acids. To study these interactions, analysis of the minimum distance of the ion from these specific amino acids was performed. Interestingly, distances for the Cys306 and Glu308 at the equilibrium are 3 to 4 time greater than the expected values (0.231 nm and 0.199 nm respectively) for this type of interaction [54]. Nonetheless, the distance of Glu seems to be stable whereas the one of Cys has greater fluctuations. Asp309 residue, on the other hand, reported a distance which is around the predicted value and shows an exceptional stability.

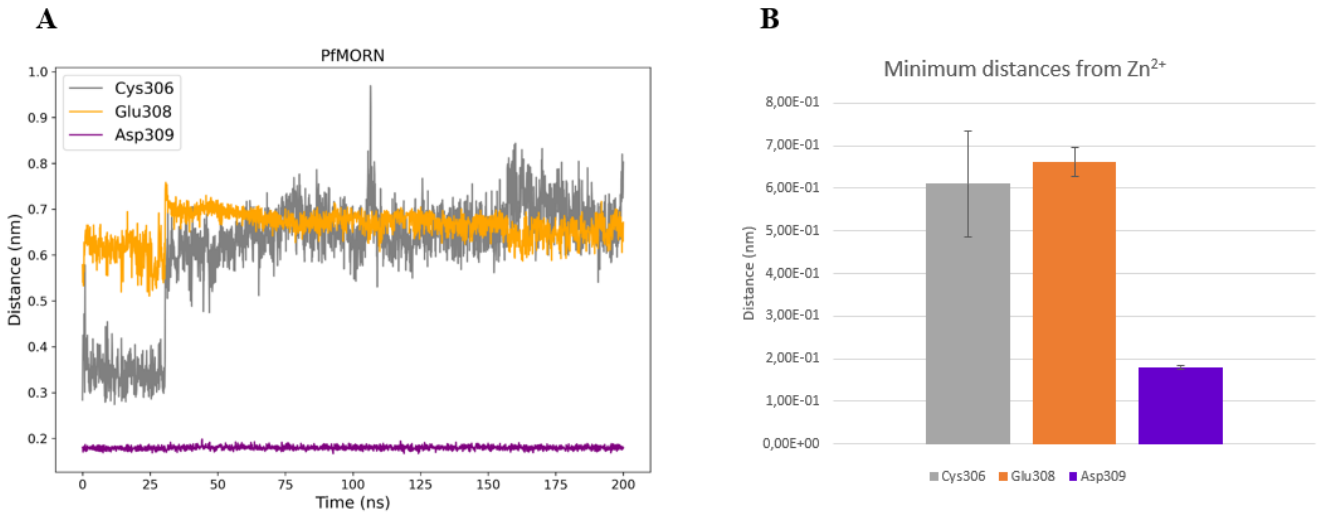

Figure 16. Graphical representation of the distances between literature reported coordination residues of the PfMORN and zinc ion (A). Mean values and standard deviations are reported as histogram bars (B).

4.4 Discussion

As stated earlier in this work, MORN domains are poorly characterized as only few data can be found on their structural features and biological functions. Moreover, as far as we know, no Molecular Dynamics studies involving these repetitions have been reported in the literature.

Thus, here we present a first work where MORN motifs were investigated through MD methods. Among the few possible structures available on the RCSB database we selected the MORN proteins from three different parasite organisms, namely *Trypanosoma brucei*, *Toxoplasma gondii* and *Plasmodium falciparum*, based on the fact that they are the only crystallographic models composed solely of MORN repetitions and their structures have no missing residues in the sequences.

To characterize the conformations of the three all-MORN assemblies and their dynamical behaviour different analysis were carried out starting from the produced MD trajectories. From a structural point of view all the structures show a very stable secondary and tertiary structure with high mobility in the N-terminal regions as reported by the RMSF results. For the conformational investigation we focused on the characteristic spatial arrangement which determines a different hinge angle in each assembly as well as torsional ones. The most peculiar one is the PfMORN which present a unique V-shaped motif with a very small angle between its chains. This conformation is stabilized by a Zn^{2+} ion at the dimerization interface which presumedly contribute to the internal torsional movement of the subunits stabilizing them while the amino-terminal region possess high mobility. Principal Component Analysis support this hypothesis showing that the motion is predominantly described by hinge angle variation due to internal rotation of the chains.

Even though the Zn^{2+} role in the assembly stabilization seems reasonably crucial, since the surface potential map shows an important negative C-terminal region, its electrostatic interactions seem to be slightly different from what was previously reported [7]. In this work we investigated this mechanism by analysing the minimum distances of this atom from the supposed amino acidic quadruplet. In fact, our results suggest that the main interaction partner of the Zn^{2+} is the Asp309 residue rather than the originally supposed Cys306 since the latter shows several times greater distance from literature reported values as well as high variations.

Interestingly, while at the interface the TbMORN possess a very similar electrostatic potential surface to the PfMORN, the one of the TgMORN, which is reported to be able to assume a V-shaped conformation in the presence of the Zn ion thanks to the conserved quadruplet, shows an extended positive region at the C-termini. This seems to suggest that the TgMORN assembly here investigated undergoes important charge redistribution in the presence of the Zn ion. Additional analysis on the TgMORN show that this protein possess similar dynamic properties to the other two assemblies. Its hinge angle resembles the one of the TbMORN with a slightly

more open configuration while its torsion values are the lowest among the three studied proteins but still quite similar. Despite sharing greater identity and similarity percentage with the PfMORN at the amino acid sequence level, PCA shows a distribution more similar to the TbMORN structures with opening and closing components motion of the hinge angle almost equally dominant. Nonetheless, every structures are characterized by almost the same principal components of opening-closing and torsional-rotational movements.

Finally, we reported the PMF calculated along the previously analysed angles for each of the three assemblies showing that their dynamical behaviours reflect their potential energy state. The result of this analysis shows that the most stable conformation is probably the V-shaped PfMORN whose minima are depicted to be very narrow around a specific region of the PMF surface.

Taken together these results represent a first step in the characterization of these largely unknown protein domains. The main limitation of this current work lays on the limited simulation time due to computational costs of the productions which limited the phase space sampling. Additionally, it would be of particular interest to recreate the alternative assembly conformation of the TgMORN to further investigate the Zn stabilizing properties as well as the dynamical behaviour of the conformation itself.

5 Structural and conformational characterization of human ALS2 MORN domain

5.1 Introduction

Among the few evidence of MORN homophilic interaction, ALS2 gene encoded Alsin protein is expected to form tetrameric assemblies trough the final region of its MORN domain [16]. Alsin is a 1657 amino acids protein composed by 5 different domains whose mutations are related to Infantile-onset Ascending Hereditary Spastic Paralysis (IAHSP) [17], [57], [49], [50], disease a rare neurodegenerative condition that affects children in their first years of life manifesting lower limbs spasticity and in most cases worsening as time passes reaching tetraplegia [58] [59]. Unfortunately, Alsin protein 3D structure is currently unknown making it difficult to

deeply understand its mechanical and biological functions in both physiological and pathological conditions.

In this work we present a first model of ALS2 MORN domain together with two possible dimeric assemblies conformation obtained through Homology modelling methodology. For each model conformational investigations were made by computational methods and comparison with dynamical behaviour of the three crystallographic previously analysed is here discussed.

5.2 Materials and methods

5.2.1 Homology modeling of ALS2 MORN and its superassemblies

Since no experimental solved structure of ALS2 MORN was found on the Protein Data Bank database, the three-dimensional model was generated starting from its primary sequence. Fasta file of the amino acid sequence was created by manually selecting the interested region from the complete amino acids list of Alsin protein available data on the UniProt database [60] (code Q96Q42). Since only few experimental structures were currently available from RCSB database, bioinformatic sequence comparison with 6 known structures, chosen among the highest resolution crystallographic structures for each MORN-bearing protein type, was made using MOE [30] software alignment tool. The best scores, in terms of identity and similarity percentages, come from *Plasmodium Falciparum* protein (referred in this work as PfMORN) obtained from X-ray refraction technique with a resolution of 2.14 Å and whose structure contains no missing residues and presents a unique Zn ion ligand. Subsequently, Homology Modelling was performed using MOE software package and the obtained model is here referred to as ALS2 MORN. This technique is an essential method for molecular modelling whose rationale is the hypothesis that proteins with the same amino acid sequences share the same three-dimensional structures. This allows to build an experimentally unsolved 3D protein structure starting from other known 3D structures of proteins which share high identity percentage. Obviously the greater the identity and similarity the more accurate the generated model will be. To validate the new built model additional structures were generated employing ITasser suite [61] [62]. Among the 5 additional structures produced the best one, evaluated based on its C-score and TM score, was compared to the previously obtained model. Further validation of the ALS2 MORN was made by evaluating model's stereochemical quality through Ramachandran plot that displays the pairs of angles ϕ and ψ (i.e., the two degrees of freedom of the backbone due

to the relative rotation of two bonded amino acids around the plane of the peptide bond). The plot was generated by the PROCHECK [32]software suite.

To build the two dimeric superassemblies conformations the above ALS2 MORN model was employed as three-dimensional structure of the two protein chains. Using MOE environment, we then recreate the targets spatial arrangements employing the superposition tool to align the model chains to the interfaces of already discussed linear and V-shaped structures of the PfMORN and TbMORN. For the V-shaped assembly, Zn^{2+} ion derived from the PfMORN, was maintained. Finally, each structure quality was checked using PROCHECK to inspect their Ramachandran plots.

5.2.2 Molecular Dynamics

Molecular Dynamics simulations were carried out for the ALS2 MORN homology model and for the two assemblies employing GROMACS 2020.4 [49]. AMBER ff99SB-ILDN force field [50] was selected, and the system was set up for the runs in a dodecahedral box with periodic boundary conditions imposing a minimum distance between the protein and the box edge of 1.5 nm. Each system was solvated by explicit TIP3P water model [51]. Moreover, physiological salt concentration conditions (0.15M) was recreated through the addition of Na^+ and Cl^- counterions in proper quantities to counterbalance the net charge of the systems making them globally neutral. Steepest descent method was implemented for the energy minimization step setting 50000 steps as a stop condition for the algorithm. After this phase, each system undergoes through two equilibration steps performed in two specific ensembles. The first step consists in an NVT simulation of 1000 ps followed by the second one carried out in under NPT conditions. The NVT and NPT simulations were performed using positions restraints on the heavy atoms of the system and with reference values for temperature and pressure if T=300K and P=1.0 bar respectively. Berendsen thermostat [52] was used (τ=0.1) for NVT equilibration whereas for NPT equilibration a Parrinello-Rahman coupling (τ=2.0) with isotropic coupling was selected. The last step consists in the production in an NVT ensemble of 200 ns Molecular Dynamics simulations using a 2 fs time step. Electrostatic interactions were treated using PME scheme [53] (fourth order interpolation with 0.16 nm for the FFT grid spacing) with a cutoff of 1 nm for both short and long range interactions. Post MD simulation analyses to assess the stability of the systems were performed by Root-mean-square deviation (RMSD), Root Mean Square Fluctuations (RMSF) and radius of gyration (Rg) calculations. In the case of the two

superassemblies, RMSF analysis was performed on each chain independently to avoid artifacts generation due to global structure movements.

Systems and their respective trajectories were inspected using Visual Molecular Dynamics (VMD) software environment [46].

5.2.3 Analysis

Evaluation of the structures stability was carried out using GROMACS integrated analysis tools. Root-mean-square deviation (RMSD) of the protein backbone atomic positions was investigated first along the entire trajectory. Subsequently, calculations of the Root Mean Square Fluctuations (RMSF) for the Cα of each residue were performed with singular chain measurements for the assembly structures. Variations in the radius of gyration were analysed as an additional indication of the dimensional stability of the models.

For the two superassembly models, investigation on their dynamical behaviours was made to characterize their spatial conformations by focusing on the hinge angle between the two chains and the torsional angle of the single subunit. A custom-made script, implementing Python 3 numpy libraries, was employed to perform measurements of both angles for each frame of the two trajectories. For hinge angle measurements, the following amino acids were grouped for both Linear assembly and V-shaped assembly {0-26, 170-196, 197-223}. Groups for torsion angle were: Linear motif {23:40, 30-33, 172-189, 178-182}, V-shaped motif {23:40, 29-33, 172-189, 179-183}.

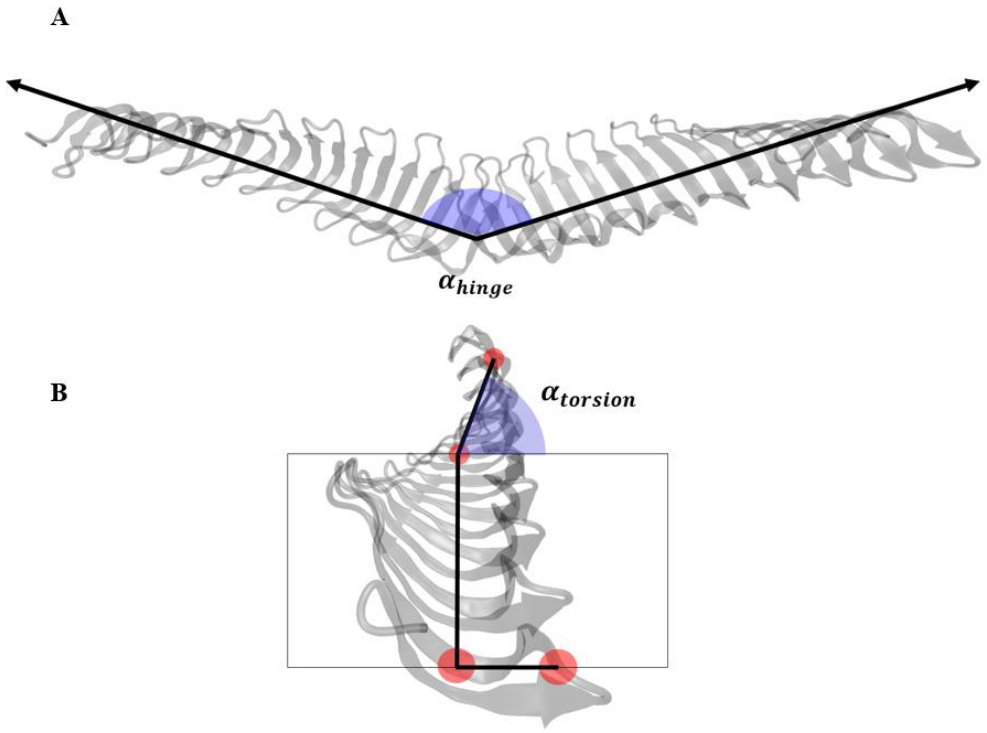

Figure 17 Description of the investigate angles: hinge angle between chains (A) and torsional angle of the single chain (B).

Further characterizations of all three models were carried out by means of the Principal Component Analysis (PCA) and another custom-made bash script, integrating a .pqr file generator and APBS software [43], to generate electrostatic potential maps. For the V-shaped motif dimer, zinc ion distance measurements were evaluated by using the GROMACS tool 'mindist'.

Starting from the results of PCA and angles variations analysis, calculation of the potential of mean force (PMF) was performed to investigate free energy changes along the two angular coordinates. A python code was employed to obtain probability distribution of the angles from the entire trajectories. This distribution was calculated as follow:

$$p(\alpha_{hinge}, \alpha_{torsion}) = \frac{H(\alpha_{hinge}, \alpha_{torsion})}{\sum_{\alpha_{hinge}} \sum_{\alpha_{torsion}} H(\alpha_{hinge}, \alpha_{torsion})} \quad (23)$$

Where α_{hinge} is the angle between the two chains, $\alpha_{torsion}$ the torsional angle of the single chains. Using the same script, we then performed Boltzmann Inversion (BI) on the obtained

distribution to evaluate angle-associated potential. Discretization of the state space along the two coordinates was made by using bins of 1°. Finally the associated PMF function was then calculated as:

$$PMF(\alpha_{hinge}, \alpha_{torsion}) = -k_B T ln\,[p(\alpha_{hinge}, \alpha_{torsion})] \quad (24)$$

Where k_B is the Boltzmann constant and T is the system temperature assumed to be 300 K.

5.2.4 Plots and Figures

Representations of the ALS2 MORN homology model and its assemblies were rendered using the VMD environment. RMSD, RMSF, radius of gyration and angles data plots as well as contact probability maps and PMF figures were generated using Python 3 custom made scripts. PCA plots were generated in Microsoft Excel.

5.3 Results

5.3.1 Bioinformatic analysis

Before starting the homology modelling process, bioinformatic analysis have been carried out. First, the alignment of each singular repetition sequence has been made (Figure 1) showing that the first seven repetitions are composed by 23 aa and the eighth one by 24. The consensus shows three completely conserved glycine in all repetitions respectively in position 3, 12 and 14 which are indicated as the most conserved aa in every MORN domain. Other conserved amino acids between repetitions are tyrosine in position 1 (6/8) and valine in position 15 (5/8) while some of the remaining shares at least similar chemical properties.

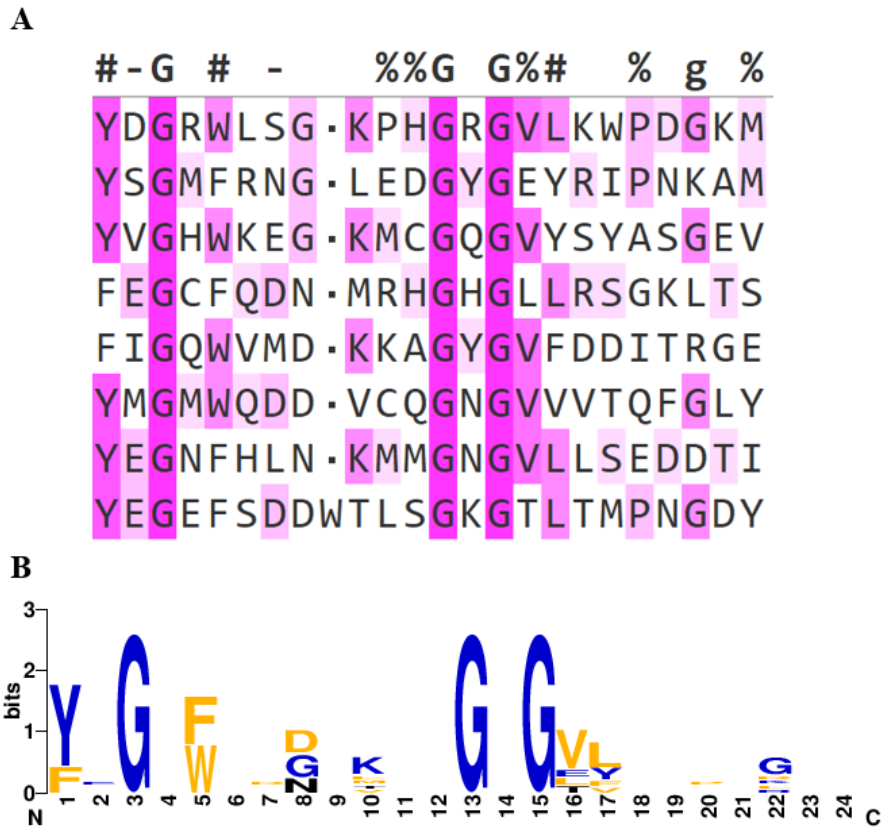

Figure 18. (A) Bioinformatic consensus analysis for the eight repetitions of the ALS2 MORN. # symbols indicates residues with conserved aromatic groups, - stands for conserved acidic residues, % represents conserved hydrophobic groups. Letters stands for totally conserved (capital letter) or partially (normal letter. (B) Logo representation of the consensus where larger letters indicate more conserved amino acids.

5.3.2 Homology model of ALS2 MORN

Alsin protein's MORN domain sequence (aa 1049-1244) was compared to other selected 6 derived from the highest resolution models MORN crystallographic structures available on RSCB database. Identity and similarity percentages of each structure are reported in Table 1. The best protein structure based on these parameters is the PfMORN (PDB id: 6T4D [7]) with 34.7% sequence identity.

Table 1. Sequence similarity and identity of other MORN compared to the ALS2 MORN sequence

PDB id	Identity %	Similarity %
6T4D [7]	34.7	49.5
6T6Q [7]	34.7	47.4
6T4R [7]	32.7	48
1H3I [63]	20.8	32.5
7JR9 [64]	9.5	18
6JLE [14]	21.9	39.7

Validation of the homology model obtained from this template was assessed in two ways. First, we employed ITasser suite to generate 5 prediction 3D models and then we compare the best predicted model with the previous one through RMSD calculation performed after structural superposition using MOE environment tools. The result shows a displacement of 1.392 Å which implies that the two models are almost identical.

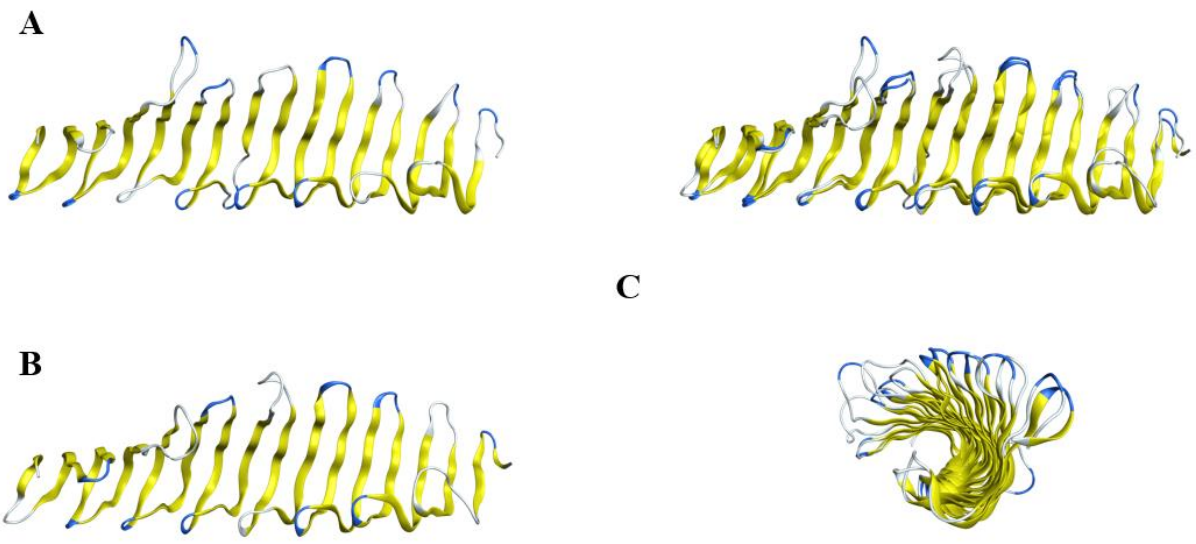

Figure 19 Graphical representation of 3D models for ALS2 MORN obtained through Homology Modelling (A), or ITasser software (B), and their superposition (C)

Finally, Ramachandran plot was generated through PROCHECK software showing that 78.2% of residues were set in the most favoured regions, while 17.9% and 2.6% of residues were located in additional allowed and generously allowed regions, respectively. Only a 1.3% of residues pertained to disallowed regions.

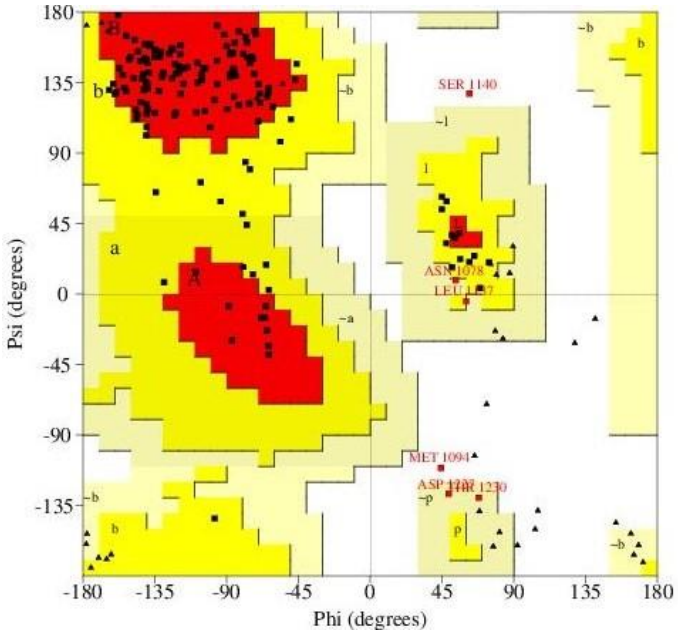

Figure 20. Ramachandran plot of the ALS2 MORN homology model

5.3.3 Molecular Dynamic simulation of ALS2 MORN

MD simulation stability along the 200 ns trajectory of the aforementioned ALS2 MORN homology model was assessed by the evaluation of the root-mean squared deviation (RMSD). Structural equilibrium was reached in about 50 ns. At this stage we investigated residue fluctuations by analysing the RMSF which shows three main peaks where the first two (0.3 nm 0.4 nm respectively) corresponds to random coil structures each connecting two consecutive β-strands while the last one, which present the greatest fluctuation value of the entire protein (1 nm), is due to the last residues of the C-terminal region. Radius of gyration measurements revealed that the structure is globally stable with a mean value of 2.34 nm and a standard deviation of 0.02 nm.

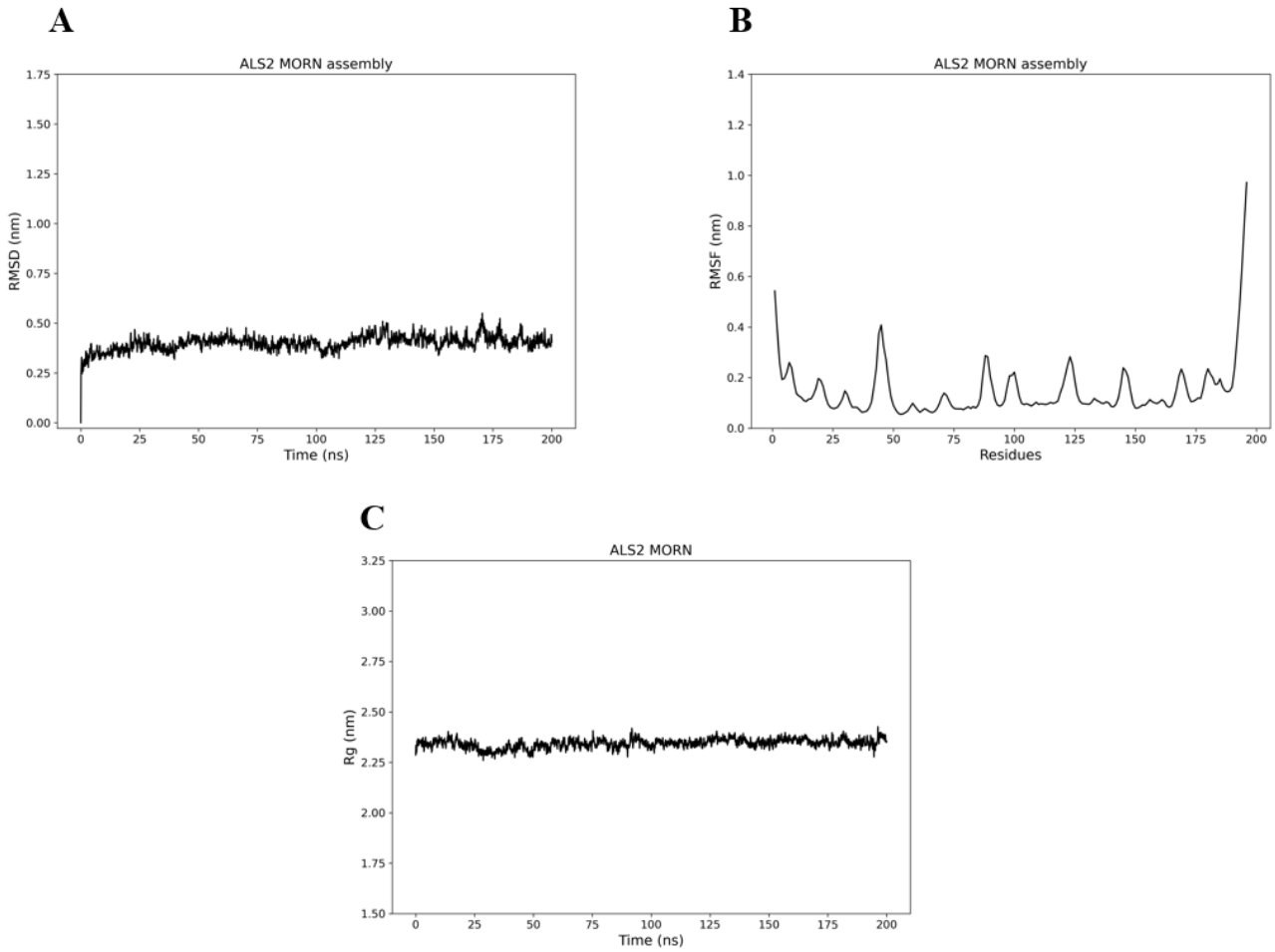

Figure 21. Representations of the post simulation analyses RMSD (A), RMSF (B) and Rg (C).

5.3.4 PCA and electrostatic analyses

To further characterize the ALS2 MORN principal component analysis was carried out. Results report that more than 50 % of the total movements described by the first two principal components are to attribute to the highly fluctuating last residues of the carboxy-terminal region, as highlighted from RMSF calculation, which tend distend and "lift up" inducing an opening motion to the near area. An additional 21.6 % of the movement is due to internal torsional movements which, in the light of the results presented in chapter 3, seems to be a common feature of MORN domains.

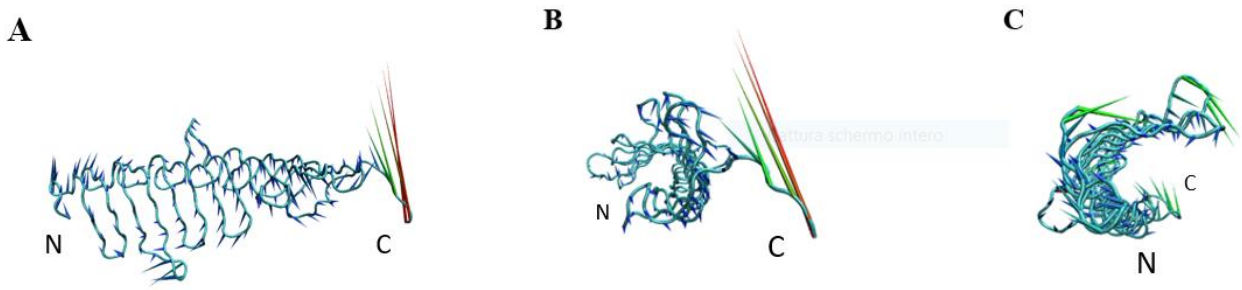

Figure 22. Porcupine plot representation of the ALS2 MORN principal component analysis results. High mobility of the C-terminus last residues is shown in the two first components (A and B) while torsional motion is accounted in the third one (C).

Finally, since electrostatic characteristics are known to be relevant for possible interactions, analysis of the related surface potential was calculated. The colorimetric map revealed a net negative charge, red surface, with a non-homogeneous distribution. While the N-terminus is predominantly positive, blue, with only a central positive patch in the central region of the inner side, the amino terminal region electrostatic potential is highly negative resembling the ones of the TbMORN and PfMORN discussed in the previous chapter.

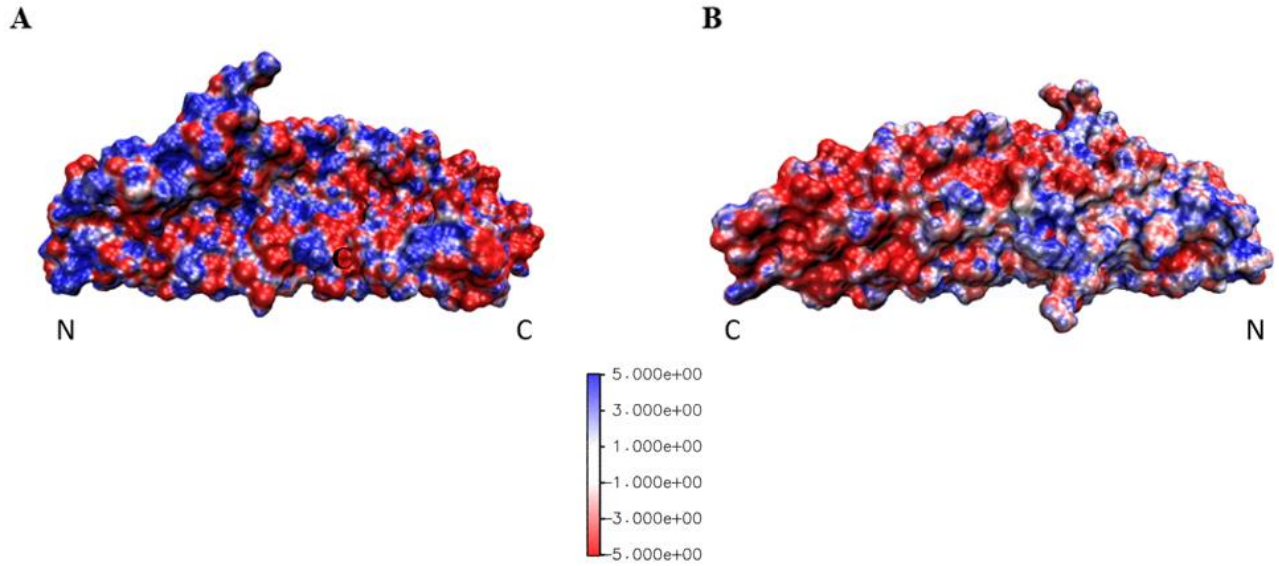

Figure 23. Colorimetric maps of electrostatic potential on the inner (A) and outer (B) surface of ALS2 MORN. Blue color express positive potential values whereas red negative ones. Grey coloration indicates neutral areas

5.3.5 ALS2 MORN superassemblies

Starting from the homology model of the ALS2 MORN described above, we constructed two different superassemblies based on the only two conformations, V-shaped and linear, whose structures are experimentally solved and analysed in Chapter 3. After reproducing the wanted spatial arrangements validation of the new models was carried out by performing Ramachandran plot calculation using PROCHECK. The plot (Figure 6C-6D) reports the following results:

- For the Linear MORN, 79.5% of aa in favourable regions, 17% and 2.2% in additional allowed and generously allowed regions, 1.3% in disallowed regions.
- For the Linear MORN, 78.8% of aa in favourable regions, 17.9% and 1.9% in additional allowed and generously allowed regions, 1.3% in disallowed regions.

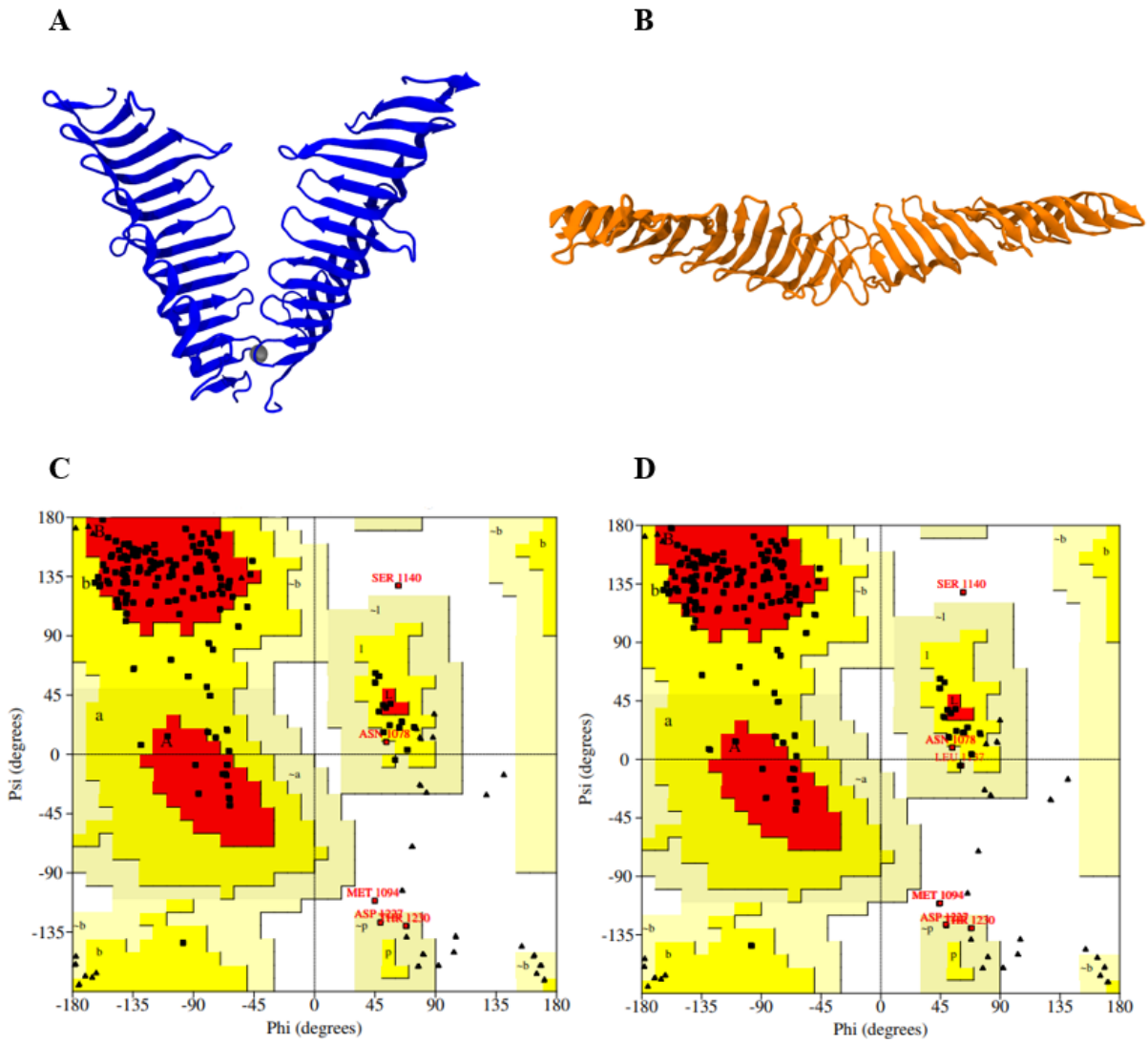

Figure 24. New cartoon representation of V-shaped ALS2 MORN assembly (A) and Linear assembly (B) and their respective Ramachandran plots, V-shaped (C) and Linear (D)

5.3.6 Molecular Dynamics simulation of Linear and V-shaped assemblies

RMSD analyses was carried out to evaluate the stability of the entire 200 ns trajectories produced by the simulations. The V-shaped assembly reached conformational equilibrium in about half the time needed for the Linear one, showing a plateau after 25 ns rather than 50 ns. Root-mean squared fluctuations were calculated for each chain of both protein motifs showing an almost conserved behaviour for the Linear subunits compared with the monomeric form, with high mobility in the C-terminal residues and others common fluctuating regions, whereas for the chains of the V-shaped assembly an increased fluctuation in the terminals regions were reported. Additional stability data was obtained by the radius of gyration analysis which shows

expected major dimensional changes in the Linear motif assembly (Rg=3.92 ± 0.19 nm) while the V-shaped one confirm its stability (Rg=2.83 ± 0.04 nm).

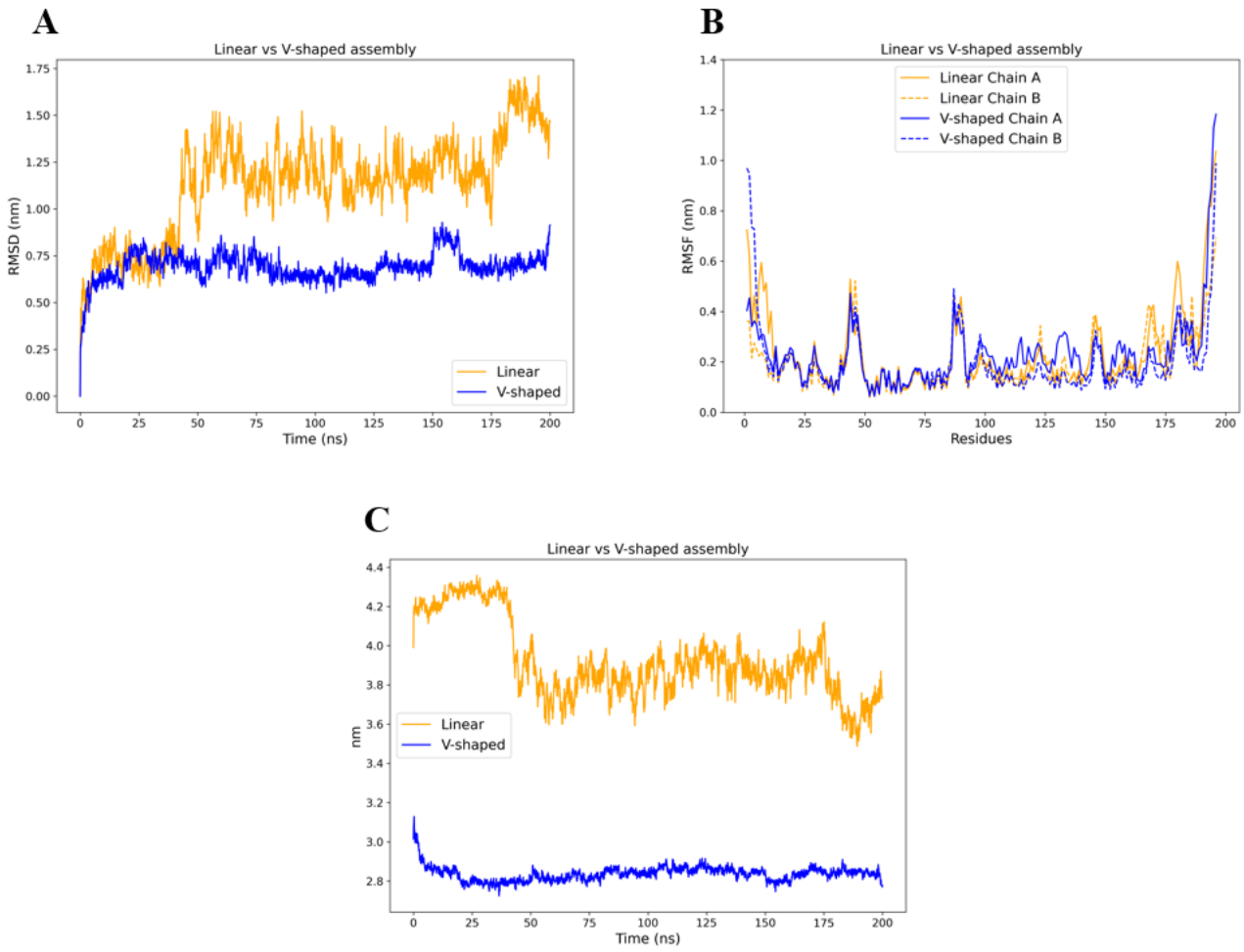

Figure 25. Post MD analyses of the Linear and V-shaped assemblies. RMSD e Gr plots (A and C) show a greater stability of the V-shaped conformation even though Linear motif reveals lower RMSF fluctuations (B).

5.3.7 Characterization of angles and motions

To assess conformational features the same method used for the three structures of the chapter 3 was employed. Hinge angle and torsional angle of both assemblies were evaluated as functions of time. Unexpectedly, the Linear assembly shows an initial decrease of the α_{hinge} in the first 50 ns of the simulation decreasing from the typical range of the TbMORN to a more V-like motif with an opening of about 120˚. On the other hand, V-shaped ALS2 MORN is characterized by a hinge angle whose value remain quite stable at around 45˚ resembling the

geometry of the PfMORN which possess a slightly closer angle. From a torsional angle stand point both assemblies present higher fluctuations respect to the ones of the three MORN protein previously analysed. Nonetheless, after reaching the equilibrium state, $\alpha_{torsion}$ values ranges between about 40˚ to 120 ˚ which falls in the same interval of the latter structures.

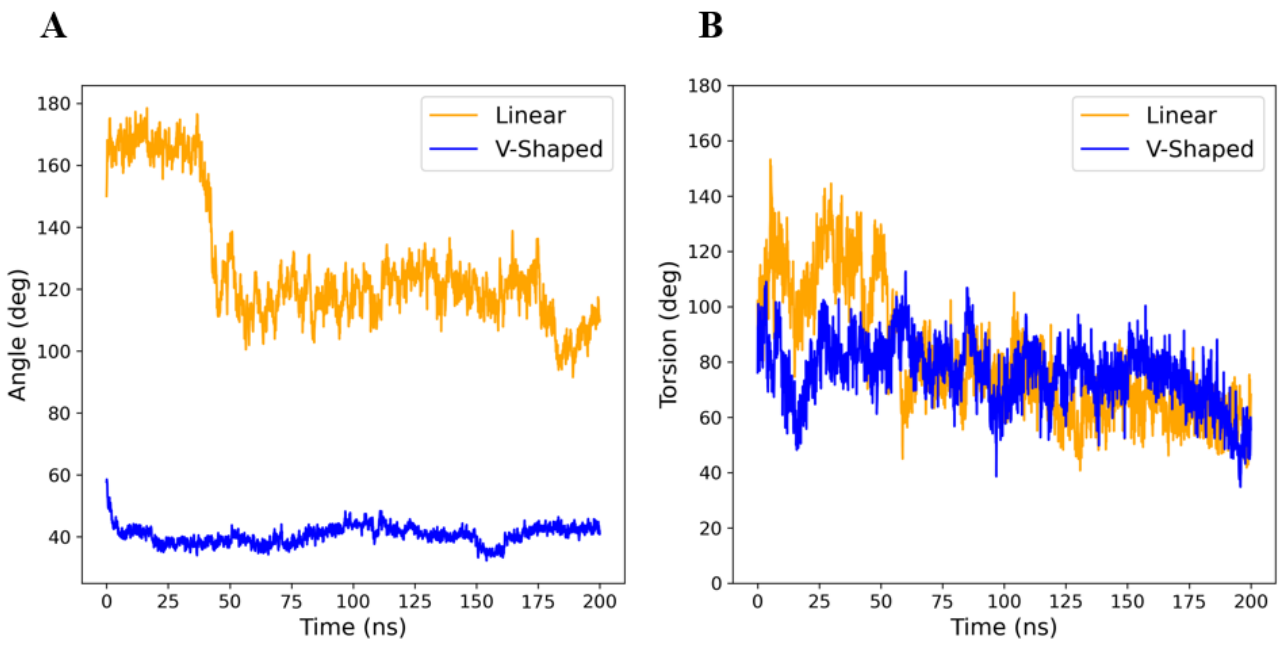

Figure 26. Hinge (A) and torsional (B) angles variations plotted as function of time. A very notable change in the hinge angle of the linear structure can be observed in contrast to the stable one of the V-shaped assembly (A). Torsion angle variations (B) seems greater in these structures rather than the ones discussed in the previous chapter.

To further characterize the dynamical behavior of these assemblies we carried out a Principal Component Analysis which report data consistent with the above observations. The dynamic of the Linear motif dimer is dominated by the hinging motion that tend to close the angle between the two chains which describe more than 85% of the total movement. V-shaped assembly reported a similar behavior to the PfMORN with a motion composed mainly by a variation of the α_{hinge} angle associated with rotational movements of the monomeric subunits. The main difference between the two structures is the fact that the former's mentioned motion is described form the first two principal components (48.7% and 19.2%) whereas the latter's is related by a dominant first component (79.4%). Together with the evaluation of Linear motif's α_{hinge} angle

important variation during the trajectory, this result seems to suggest that ALS2 MORN favours a V-shaped conformation rather than a linear one.

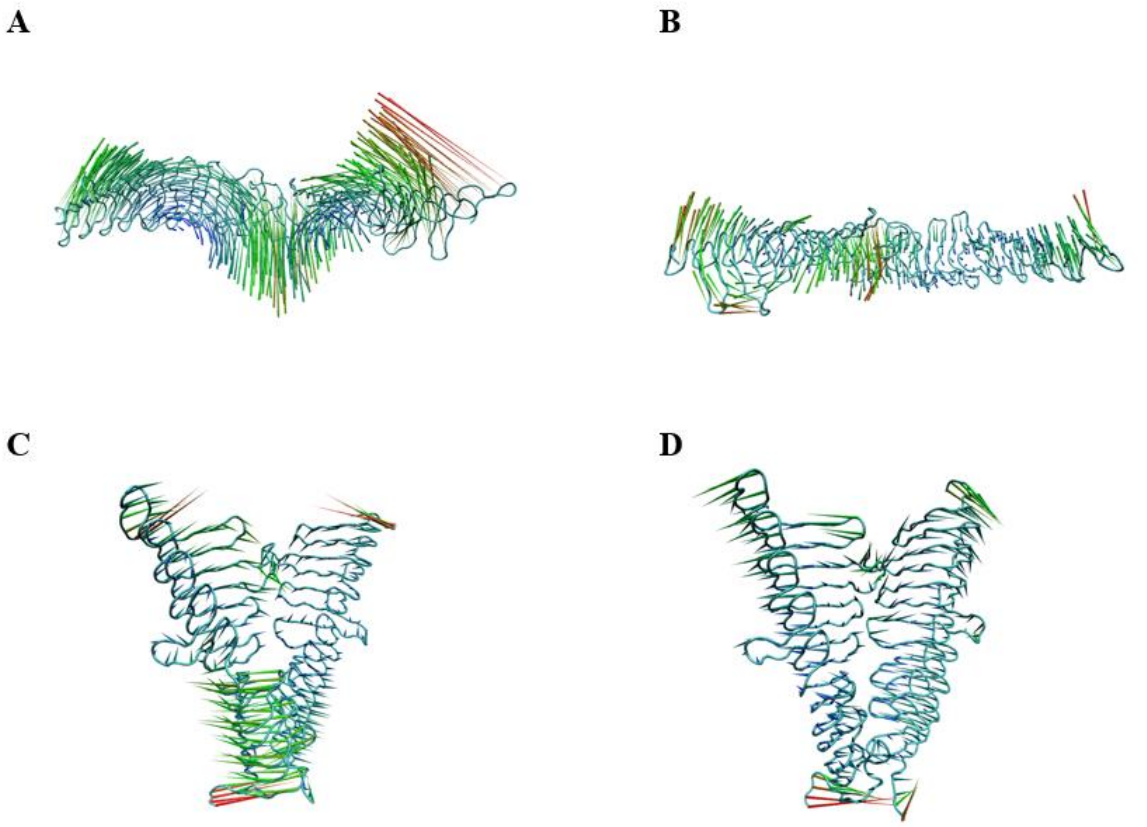

Figure 27. Representation of principal motion components through porcupine plots. Linear motif assembly first two components (A and B) suggest a conformational instability dominated by closing hinge angle motion with secondary torsional components. The first two components of the V-shaped structures instead, shows rotational movement which seems to not affect its geometry in an important way.

To further investigate the plausibility of this idea Potential Mean Force analysis was performed using the two angles α_{hinge} and $\alpha_{torsion}$ as coordinates for the investigation. The same custom-made script used in chapter 3 was employed to calculate angles distributions and their Boltzmann Inversion. As shown from the colorimetric maps V-shaped motif highly resembles the PMF of the PfMORN showing a very narrow minima with an elongated direction along the $\alpha_{torsion}$ axis while the corresponding α_{hinge} variations are very limited around 45° as expected from the previous results. Notably, the Linear assembly generated a map which shows two extended distinguishable regions. This is a clear instability evidence of the conformation which

further suggest our hypothesis about the most favourable assembly geometry for the ALS2 MORN dimer.

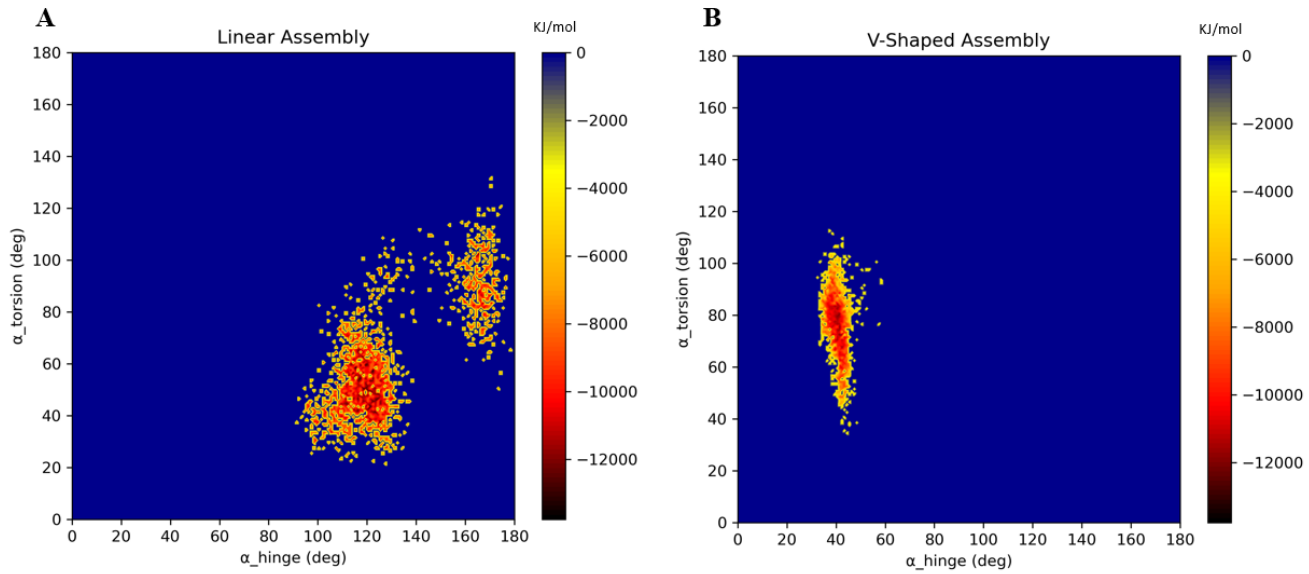

Figure 28 Colorimetric maps of the ALS2 MORN superassemblies PMFs. The two distinct regions identified by the Linear motif PMF strongly suggest a great conformation instability (A) while the V-shaped configuration report a PfMORN like narrow minimum region suggesting a stable and favourable geometry.

5.3.8 Electrostatic potential mapping and Zn ion investigation

The electrostatic potential at the supposed dimerization interface was calculated for both assemblies using APBS calculations integrated in a customized script. The correspondent colorimetric maps were visualized using VMD software environment.

Both structures shared very similar surface electrostatic potential resembling the one of the monomeric form reported for the ALS2 MORN. The charge distribution is still globally non-homogeneous, and the interfaces remain characterized by a negative potential. Very little variations are observed in the inner surface of the V-shaped assembly where apparently the potential change from almost neutral to positive probably due to charges redistribution caused by the Zn ion presence.

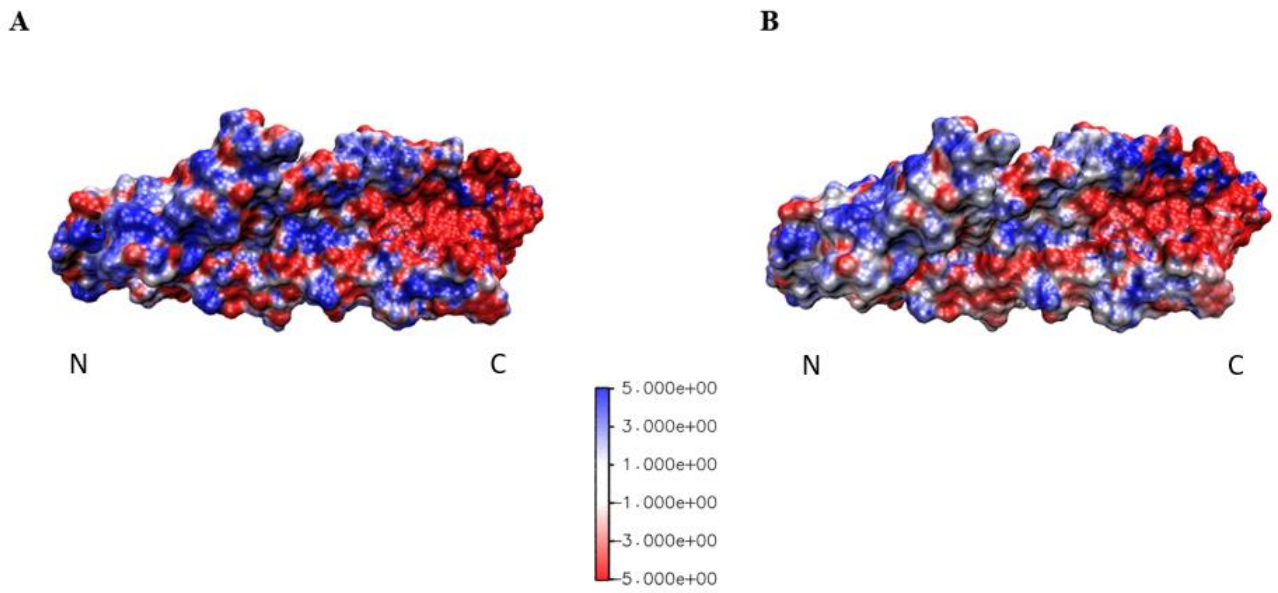

Figure 29 Colorimetric maps of electrostatic potential (KbT/e) of the interaction surfaces for the Linear conformation (A) and the V-shaped one (B)..

Finally, we further investigate the supposed stabilizing Zn interactions on the V-shaped conformations. In alignement step of the homology modelling process we observed that the sequence of ALS2 MORN presents a similar sequence to the one deputed to Zn cordination activity in the same region composed by S-E-D-D residues. For this reason we decided to keep the ion in the pdb file of the superassembly. As for the PfMORN mesurements of the Zn from the homologus amino acids of the quadruplet were performed. Results shows an arguably better interaction of the homologus residues than the PfMORN one. In fact, both the mesurements of Ser1215 and Glu1216 show closer and less fluctuating distance than the corispective Cys306 and Glu308, with the latter ALS2 MORN residue reporting the expected value from experimental evidence [54]. Interestingly, both Zn-containing assemblies shows identical values for the Asp residues involved in the interactions raising the idea that this residue among the four investigated is the most crucial for zinc coordination.

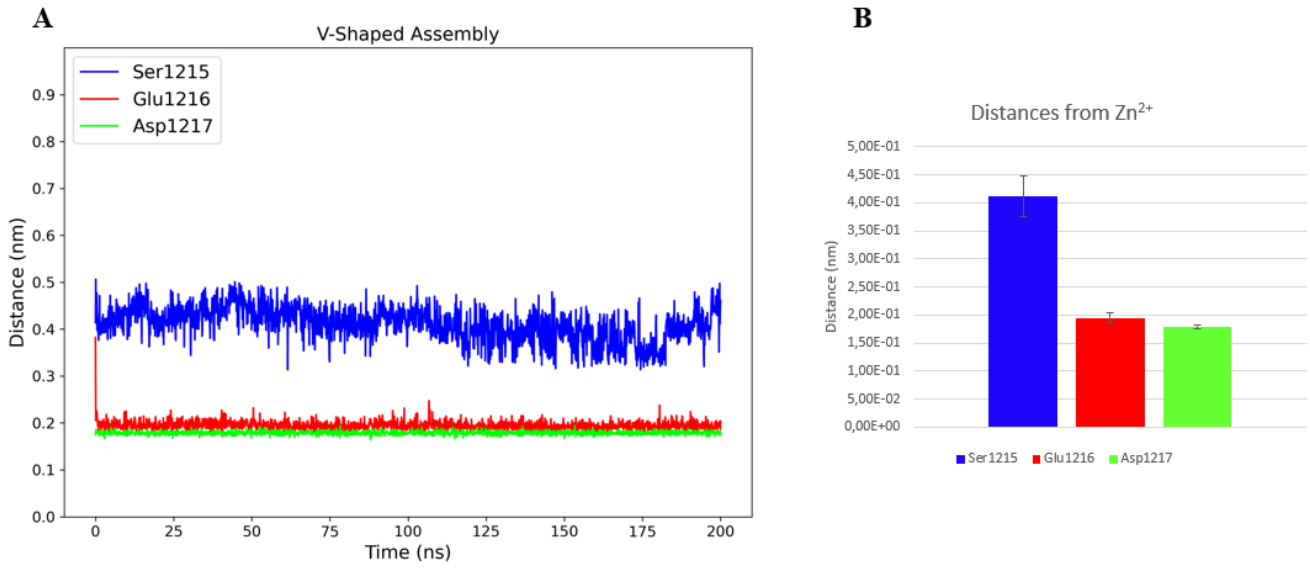

Figure 30. Graphical representation of the distances between ALS2 MORN homologous coordination residues quadruplet and zinc ion (A). Mean values and standard deviations are reported as histogram bars (B).

5.4 Discussion

Supporting the hypothesis of homofilic protein interactions as MORN biological functions literature evidence suggests that Alsin protein is expected to form an homotetramer by interactions trough, among other regions, the C-terminal region of its MORN domains. For this reason, in the present work we decided to study the possible conformation of the suggested dimeric superassembly. To do so, Homology Modeling tecnique was employed to reconstruct the unknown ALS2 MORN three-dimensional structure by selecting as a template the previously charactherized PfMORN protein. Validation steps showed that the obtained model is of satisfctory quality and Molecular Dynamic simulations and subsequential analyses reveal an overall stable structure with high mobility in the C-terminal region, uderlined from performed Principal Component Analysis. After the ALS2 MORN characterization we then built the superassemblies starting from its structure as homodimer subunits. The spatial arrangements were retrived by superposition from the only two known motifs of the linear TbMORN and the V-shaped PfMORN. MD simulations shows different structural stability suggesting that the Linear arrangement is a less favourable configuration. This data was further confirmed by hinge and torsional angle evaluation which reported a repentine decrease in the

former angle from the initial value around 165˚ to 120˚. PCA results underline this behaviour which can be found in the linear motif proteins TbMORN and TgMORN which however do not change their mean angle value along their trajectories. To support this observation PMF plot confrontations reveale a substantial difference among the three structures showing the last two protein are associated to an unique contained region of the plot whereas the former reveale a two separate regions, symptoms of structural instability. As a matter of fact, toward the end of the trajectory the angle of the Linear ALS2 MORN reaches a new minimum of about 100˚ and we suspect that longer simulations could underline the tendency of the structure to assume a PfMORN-like geometry. This hypotesis is reinforced by the conformational analysis carried out on the V-shaped assembly which shows a very stable conformation characterized by an hinge angle of 40-45˚, slightly more opened than the PfMORN one. Additionally, PCA revealed a very similar motion of the former assembly to the latter one formed mainly by rotational and hinging motions and PMF representation results in almost identical minimum shape and location.

The numerous similarities between the V-shaped ALS2 MORN and PfMORN suggests that even homophilic interaction mechanism could be the same. To assess this possibility we evaluated the Zn coordination capability of the literature reported PfMORN qudruplet C-x-E-D and its suggested ALS2 MORN homologues S-E-D-D. Minimum distances of the key residues for both structures from the Zn^{2+} ion were calculated as an indicative parameter of interaction. Results seem to attribute, surprisingly, a stronger interaction to the ALS2 MORN superassembly suggessting that Cys homologus Ser is arguably even better at coordinating the zinc and that Asp residue is probably the most important for this in this small ion-binding patch, being found at experimental expected distance in both structures with almost no flucuations.

Limitations of the current work is related to the limited phase space sampling due to high computational costs of the produced simulations. As previously stated, longer MD trajectories would be usefull to better investigate the dynamical behavior of these assemblies.

6 Conclusions

Recent scientific studies suggest that Membrane Occupation Recognition Nexus domains may interact with each other to provide a novel biological function as homo-dimerization modules, different from the lipid-binding function originally proposed. However, these domains remain largely unknown, and the mechanisms that regulate this kind of interactions is yet to be fully understood. In this study we investigated conformational dynamics of MORN homodimers with the aim of characterizing at molecular level the homophilic interactions that drives superassemblies formation. Our work was composed by two principal sections with the first one laying the foundations for the second. In the first part the structures of the TbMORN, TgMORN and PfMORN, the only superassembly whose crystallographic conformations had been reported, were studied through Molecular Dynamics applications and subsequent analysis of the main structural features. Results reveal that MORN domains are capable of interacting with each other in a stable way assuming macromolecular conformations characterized by specific angles between protein chains as well as common torsional internal motions. Spatial arrangement was shown to be determined by the amino acid sequence of the MORN repetitions as well as the presence or absence of specific ion species. Investigation on the interactions of the zinc ion, which is found in a coordination region in the PfMORN assembly as well as in an alternative TgMORN conformation reported in the literature, with three specific amino acids showed unexpected distances of two of them among which the supposedly most important residue is the farthest.

Using these results as the base for comparisons, we evaluate the plausibility of two superassembly conformation for the Alsin protein MORN domain. Since this protein three-dimensional structure is unknown, as a first step, model generation was made through Homology Modelling technique selecting PfMORN among the available structures as the target template. After model validation and subsequential characterization was carried out, reconstruction of two superassemblies inspired by the conformations of the protein discussed in the previous section was performed. Through software tools a Linear motif assembly and a V-shaped one were recreated. Using the same methods employed for the previous MORN proteins, characterization of the new assemblies was made. Linear ALS2 MORN revealed high conformational instability showing the tendency to abruptly decreasing its hinge angle and assuming a more V-like shape. Opposites results described the behaviour of the V-shaped assembly which resemble in almost

all aspects the PfMORN. Even for this structure, Zn distances calculation were used to study the interactions between the metal ion and an hypothesized homologus coordination region which compared to the above discussed one seems to perform better thus making more plausible the involvement of this ion species in the homophilic dimerization of Alsin protein.

This M.Sc. thesis present a first important step towards a better understanding of MORN motifs homophilic interactions and its superassemblies conformations. Despite the limitations of the study, results here reported will be the starting point for further computational and experimental analyses, to elucidate the biological function of these poorly characterized protein repetitions.

7 Acknowledgements

I would like to thank my supervisor Prof. Marco Agostino Deriu for the opportunity to work under his supervision to this project. The great enthusiasm, expertise and professionality showed during my master thesis period were both motivational and inspirational in my journey through this work maturation process. My gratitude also goes to Marcello Miceli whose guidance and support helped me to approach this new field of molecular mechanics and to produce the presented results.

I would like to thank Olivia's family, the Help Olly association, and the project Seed Grant Spring 2020 – IAHSP, since their effort to shed light into IAHSP allowed me to get in touch with this topic which captured my interest. I would thank the Telethon foundation, that gives the opportunity to study rare and complex pathologies and train students and researchers.

I also want to show my gratitude to my family for supporting me in my university studies and for giving me the chance of studying away from home. I would also thank my friends for confronting thoughs on everything and sharing memories. Thanks to Marco for working with me and for the constant exchange of ideas.

A special and hearted thanks to my life-partner Laura whose support and love have helped me going through my studies.

Riccardo Tortarolo

8 References

[1] S. El-Gebali *et al.*, "The Pfam protein families database in 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427–D432, 2019, doi: 10.1093/nar/gky995.

[2] H. Takeshima, S. Komazaki, M. Nishi, M. Iino, and K. Kangawa, "Junctophilins: A novel family of junctional membrane complex proteins," *Mol. Cell*, vol. 6, no. 1, pp. 11–22, 2000, doi: 10.1016/s1097-2765(05)00005-5.

[3] K. Hu *et al.*, "Cytoskeletal Components of an Invasion Machine—The Apical Complex of Toxoplasma gondii," *PLoS Pathog.*, vol. 2, no. 2, p. e13, Feb. 2006, doi: 10.1371/journal.ppat.0020013.

[4] M.-J. Gubbels, S. Vaishnava, N. Boot, J.-F. Dubremetz, and B. Striepen, "A MORN-repeat protein is a dynamic component of the Toxoplasma gondii cell division apparatus," *J. Cell Sci.*, vol. 119, no. 11, pp. 2236–2245, Jun. 2006, doi: 10.1242/jcs.02949.

[5] J. I. Yang, A. J. Davis, I. Y. Perera, E. Johannes, N. S. Allen, and W. F. Boss, "The N-terminal membrane occupation and recognition nexus domain of Arabidopsis phosphatidylinositol phosphate kinase 1 regulates enzyme activity," *J. Biol. Chem.*, vol. 282, no. 8, pp. 5443–5452, 2007, doi: 10.1074/jbc.M611342200.

[6] H. J. Bennett, J. B. Davenport, R. F. Collins, A. W. Trafford, C. Pinali, and A. Kitmitto, "Human junctophilin-2 undergoes a structural rearrangement upon binding PtdIns(3,4,5)P3 and the S101R mutation identified in hypertrophic cardiomyopathy obviates this response," *Biochem. J.*, vol. 456, no. 2, pp. 205–217, 2013, doi: 10.1042/BJ20130591.

[7] S. Sajko *et al.*, "Structures of three MORN repeat proteins and a re-evaluation of the proposed lipid-binding properties of MORN repeats," *PLoS One*, vol. 15, no. 12, p. e0242677, Dec. 2020, doi: 10.1371/journal.pone.0242677.

[8] T. Nakada, T. Kashihara, M. Komatsu, K. Kojima, T. Takeshita, and M. Yamada, "Physical interaction of junctophilin and the CaV1.1 C terminus is crucial for skeletal muscle contraction," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 17, pp. 4507–4512, 2018, doi: 10.1073/pnas.1716649115.

[9] D. Rossi *et al.*, “Molecular determinants of homo- And heteromeric interactions of Junctophilin-1 at triads in adult skeletal muscle fibers,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 31, pp. 15716–15724, 2019, doi: 10.1073/pnas.1820980116.

[10] H. Ma, Y. Lou, W. H. Lin, and H. W. Xue, “MORN motifs in plant PIPKs are involved in the regulation of subcellular localization and phospholipid binding,” *Cell Res.*, vol. 16, no. 5, pp. 466–478, 2006, doi: 10.1038/sj.cr.7310058.

[11] L. Camacho, A. P. Smertenko, J. Pérez-Gómez, P. J. Hussey, and I. Moore, “Arabidopsis Rab-E GTPases exhibit a novel interaction with a plasma-membrane phosphatidylinositol-4-phosphate 5-kinase,” *J. Cell Sci.*, vol. 122, no. 23, pp. 4383–4392, 2009, doi: 10.1242/jcs.053488.

[12] A. P. Landstrom, D. L. Beavers, and X. H. T. Wehrens, “The junctophilin family of proteins: from bench to bedside,” *Trends Mol. Med.*, vol. 20, no. 6, pp. 353–362, Jun. 2014, doi: 10.1016/j.molmed.2014.02.004.

[13] M. Nishi, H. Sakagami, S. Komazaki, H. Kondo, and H. Takeshima, “Coexpression of junctophilin type 3 and type 4 in brain,” *Mol. Brain Res.*, vol. 118, no. 1–2, pp. 102–110, 2003, doi: 10.1016/S0169-328X(03)00341-3.

[14] J. Li *et al.*, “Structure of the MORN4/Myo3a Tail Complex Reveals MORN Repeats as Protein Binding Modules,” *Structure*, vol. 27, no. 9, pp. 1366-1374.e3, 2019, doi: 10.1016/j.str.2019.06.004.

[15] S. Hadano, R. Kunita, A. Otomo, K. Suzuki-Utsunomiya, and J.-E. Ikeda, “Molecular and cellular function of ALS2/alsin: Implication of membrane dynamics in neuronal development and degeneration,” *Neurochem. Int.*, vol. 51, no. 2–4, pp. 74–84, Jul. 2007, doi: 10.1016/j.neuint.2007.04.010.

[16] R. Kunita *et al.*, “Homo-oligomerization of ALS2 through its unique carboxyl-terminal regions is essential for the ALS2-associated Rab5 guanine nucleotide exchange activity and its regulatory function on endosome trafficking,” *J. Biol. Chem.*, vol. 279, no. 37, pp. 38626–38635, 2004, doi: 10.1074/jbc.M406120200.

[17] S. Millecamps, B. J. Gentil, F. Gros-Louis, G. Rouleau, and J.-P. Julien, “Alsin is partially associated with centrosome in human cells,” *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1745, no. 1, pp. 84–100, Aug. 2005, doi: 10.1016/j.bbamcr.2004.12.008.

[18] S. M. Soisson, A. S. Nimnual, M. Uy, D. Bar-Sagi, and J. Kuriyan, “Crystal Structure of the Dbl and Pleckstrin Homology Domains from the Human Son of Sevenless Protein,” *Cell*, vol. 95, no. 2, pp. 259–268, Oct. 1998, doi: 10.1016/S0092-8674(00)81756-0.

[19] B. Aghazadeh *et al.*, “Structure and mutagenesis of the Dbl homology domain,” *Nat. Struct. Biol.*, vol. 5, no. 12, pp. 1098–1107, Dec. 1998, doi: 10.1038/4209.

[20] R. Kunita, A. Otomo, H. Mizumura, K. Suzuki-Utsunomiya, S. Hadano, and J. E. Ikeda, “The Rab5 activator ALS2/alsin acts as a novel Rac1 effector through Rac1-activated endocytosis,” *J. Biol. Chem.*, vol. 282, no. 22, pp. 16599–16611, 2007, doi: 10.1074/jbc.M610682200.

[21] K. Kanekura *et al.*, “A Rac1/Phosphatidylinositol 3-Kinase/Akt3 Anti-apoptotic Pathway, Triggered by AlsinLF, the Product of the ALS2 Gene, Antagonizes Cu/Zn-superoxide Dismutase (SOD1) Mutant-induced Motoneuronal Cell Death,” *J. Biol. Chem.*, vol. 280, no. 6, pp. 4532–4543, Feb. 2005, doi: 10.1074/jbc.M410508200.

[22] S. M. Wakil *et al.*, “Infantile-onset ascending hereditary spastic paraplegia with bulbar involvement due to the novel ALS2 mutation c.2761C > T,” *Gene*, vol. 536, no. 1, pp. 217–220, Feb. 2014, doi: 10.1016/j.gene.2013.11.043.

[23] M. Dasso, “Running on Ran,” *Cell*, vol. 104, no. 3, pp. 321–324, Feb. 2001, doi: 10.1016/S0092-8674(01)00218-5.

[24] R. S. Devon *et al.*, “Cross-species characterization of the ALS2 gene and analysis of its pattern of expression in development and adulthood,” *Neurobiol. Dis.*, vol. 18, no. 2, pp. 243–257, Mar. 2005, doi: 10.1016/j.nbd.2004.10.002.

[25] T. A. Shelkovnikova *et al.*, “Fused in Sarcoma (FUS) Protein Lacking Nuclear Localization Signal (NLS) and Major RNA Binding Motifs Triggers Proteinopathy and Severe Motor Phenotype in Transgenic Mice,” *J. Biol. Chem.*, vol. 288, no. 35, pp. 25266–25274, Aug. 2013, doi: 10.1074/jbc.M113.492017.

[26] J. Westbrook, “The Protein Data Bank: unifying the archive,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 245–248, Jan. 2002, doi: 10.1093/nar/30.1.245.

[27] W. Gans, A. Amann, and J. C. A. Boeyens, Eds., *Fundamental Principles of Molecular Modeling*. Boston, MA: Springer US, 1996.

[28] C. Chothia and A. M. Lesk, “The relation between the divergence of sequence and structure in proteins.,” *EMBO J.*, vol. 5, no. 4, pp. 823–6, Apr. 1986, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3709526.

[29] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, “Computational methods in drug discovery,” *Pharmacol. Rev.*, vol. 66, no. 1, pp. 334–395, 2014, doi: 10.1124/pr.112.007336.

[30] “Molecular Operating Environment (MOE).” 2016, [Online]. Available: https://www.chemcomp.com/Products.htm.

[31] E. F. Pettersen *et al.*, “UCSF Chimera?A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004, doi: 10.1002/jcc.20084.

[32] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, “PROCHECK: a program to check the stereochemical quality of protein structures,” *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, Apr. 1993, doi: 10.1107/S0021889892009944.

[33] L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey, and T. Schwede, “Protein structure homology modeling using SWISS-MODEL workspace,” *Nat. Protoc.*, vol. 4, no. 1, pp. 1–13, Jan. 2009, doi: 10.1038/nprot.2008.197.

[34] R. Berger, “Computational Chemistry. Introduction to the Theory and Applications of Molecular and Quantum Mechanics. By Errol G. Lewars.,” *Angew. Chemie Int. Ed.*, vol. 43, no. 38, pp. 4979–4980, Sep. 2004, doi: 10.1002/anie.200485057.

[35] Jacob Chapman, “Improving the Functional Control of Ferroelectrics using Insights from Atomistic modelling - Scientific Figure on ResearchGate.” .

[36] R. Fletcher and M. J. D. Powell, “A Rapidly Convergent Descent Method for Minimization,” *Comput. J.*, vol. 6, no. 2, pp. 163–168, Aug. 1963, doi: 10.1093/comjnl/6.2.163.

[37] J. F. McCarthy, “Block-conjugate-gradient method,” *Phys. Rev. D*, vol. 40, no. 6, pp. 2149–2152, Sep. 1989, doi: 10.1103/PhysRevD.40.2149.

[38] J. Verbeke and R. Cools, “The Newton-Raphson method,” *Int. J. Math. Educ. Sci. Technol.*, vol. 26, no. 2, pp. 177–193, Mar. 1995, doi: 10.1080/0020739950260202.

[39] J. D. Head and M. C. Zerner, “A Broyden—Fletcher—Goldfarb—Shanno optimization procedure for molecular geometries,” *Chem. Phys. Lett.*, vol. 122, no. 3, pp. 264–270, Dec. 1985, doi: 10.1016/0009-2614(85)80574-1.

[40] R. D. Skeel, “What Makes Molecular Dynamics Work?,” *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 1363–1378, Jan. 2009, doi: 10.1137/070683660.

[41] K. Reiss, U. Morzan, A. Grigas, and V. Batista, “Water Network Dynamics Next to the Oxygen-Evolving Complex of Photosystem II,” *Inorganics*, vol. 7, no. 3, p. 39, Mar. 2019, doi: 10.3390/inorganics7030039.

[42] S. Kim, “Issues on the Choice of a Proper Time Step in Molecular Dynamics,” *Phys. Procedia*, vol. 53, pp. 60–62, 2014, doi: 10.1016/j.phpro.2014.06.027.

[43] E. Jurrus *et al.*, “Improvements to the <scp>APBS</scp> biomolecular solvation software suite,” *Protein Sci.*, vol. 27, no. 1, pp. 112–128, Jan. 2018, doi: 10.1002/pro.3280.

[44] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, “Electrostatics of nanosystems: Application to microtubules and the ribosome,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 18, pp. 10037–10041, Aug. 2001, doi: 10.1073/pnas.181342398.

[45] M. J. Holst and F. Saied, “Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods,” *J. Comput. Chem.*, vol. 16, no. 3, pp. 337–364, Mar. 1995, doi: 10.1002/jcc.540160308.

[46] K. Umphrey, W.; Dalke, A.; Schulten, “VMD -Visual Molecular Dynamics,” *J. Mol. Graph.*, no. 14, pp. 33–38, 1996.

[47] “The PyMOL Molecular Graphics System, Version 2.4.1 Schrödinger, LLC.” [Online]. Available: https://pymol.org/.

[48] D. A. Case *et al.*, “The Amber biomolecular simulation programs,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, Dec. 2005, doi: 10.1002/jcc.20290.

[49] M. J. Abraham *et al.*, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1–2, pp. 19–25, Sep. 2015, doi: 10.1016/j.softx.2015.06.001.

[50] K. Lindorff-Larsen *et al.*, “Improved side-chain torsion potentials for the Amber ff99SB protein force field,” *Proteins Struct. Funct. Bioinforma.*, vol. 78, no. 8, pp.

1950–1958, Jun. 2010, doi: 10.1002/prot.22711.

[51] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, “A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field,” *J. Chem. Phys.*, vol. 108, no. 24, pp. 10220–10230, Jun. 1998, doi: 10.1063/1.476482.

[52] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, Oct. 1984, doi: 10.1063/1.448118.

[53] P. P. Ewald, “Die Berechnung optischer und elektrostatischer Gitterpotentiale,” *Ann. Phys.*, vol. 369, no. 3, pp. 253–287, 1921, doi: 10.1002/andp.19213690304.

[54] M. M. Harding, “Small revisions to predicted distances around metal sites in proteins,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 62, no. 6, pp. 678–682, 2006, doi: 10.1107/S0907444906014594.

[55] C. C. Verschuuren-Bemelmans, P. Winter, D. A. Sival, J.-W. Elting, O. F. Brouwer, and U. Müller, “Novel homozygous ALS2 nonsense mutation (p.Gln715X) in sibs with infantile-onset ascending spastic paralysis: the first cases from northwestern Europe,” *Eur. J. Hum. Genet.*, vol. 16, no. 11, pp. 1407–1411, Nov. 2008, doi: 10.1038/ejhg.2008.108.

[56] C. Lai *et al.*, “Amyotrophic Lateral Sclerosis 2-Deficiency Leads to Neuronal Degeneration in Amyotrophic Lateral Sclerosis through Altered AMPA Receptor Trafficking,” vol. 26, no. 45, pp. 11798–11806, 2006, doi: 10.1523/JNEUROSCI.2084-06.2006.

[57] E. Eymard-Pierre *et al.*, “Infantile-Onset Ascending Hereditary Spastic Paralysis Is Associated with Mutations in the Alsin Gene,” *Am. J. Hum. Genet.*, vol. 71, no. 3, pp. 518–527, Sep. 2002, doi: 10.1086/342359.

[58] M. Helal *et al.*, “Clinical presentation and natural history of infantile-onset ascending spastic paralysis from three families with an ALS2 founder variant,” *Neurol. Sci.*, vol. 39, no. 11, pp. 1917–1925, Nov. 2018, doi: 10.1007/s10072-018-3526-8.

[59] P. V. S. de Souza, W. B. V. de Rezende Pinto, G. N. de Rezende Batistella, T. Bortholin, and A. S. B. Oliveira, “Hereditary Spastic Paraplegia: Clinical and Genetic

Hallmarks," *The Cerebellum*, vol. 16, no. 2, pp. 525–551, Apr. 2017, doi: 10.1007/s12311-016-0803-z.

[60] "UniProt (universal protein resource)," in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Dordrecht: Springer Netherlands, pp. 2064–2064.

[61] A. Roy, A. Kucukural, and Y. Zhang, "I-TASSER: a unified platform for automated protein structure and function prediction," *Nat. Protoc.*, vol. 5, no. 4, pp. 725–738, Apr. 2010, doi: 10.1038/nprot.2010.5.

[62] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, "The I-TASSER Suite: protein structure and function prediction," *Nat. Methods*, vol. 12, no. 1, pp. 7–8, Jan. 2015, doi: 10.1038/nmeth.3213.

[63] J. R. Wilson *et al.*, "Crystal structure and functional analysis of the histone methyltransferase SET7/9," *Cell*, vol. 111, no. 1, pp. 105–115, 2002, doi: 10.1016/S0092-8674(02)00964-9.

[64] I. Grossman-Haham *et al.*, "Structure of the radial spoke head and insights into its role in mechanoregulation of ciliary beating," *Nat. Struct. Mol. Biol.*, vol. 28, no. 1, pp. 20–28, Jan. 2021, doi: 10.1038/s41594-020-00519-9.