POLITECNICO DI TORINO

Master Degree in Biomedical Engineering

Master Thesis

Molecular modelling to investigate relationships between protein conformation and aggregation tendency in neurodegenerative diseases

Thesis Supervisor:
Prof. M. A. Deriu

Co-Supervisors:
Prof. U. Morbiducci
Prof. J.A. Tuszynski

Candidate:
Luigi Leanza

ACADEMIC YEAR 2018-2019

Abstract

Proteins are complex molecules that play many fundamental roles in human body and their conformational structure is strictly related to their function. Sometimes proteins fail to adopt its functional conformational state, thus leading to pathological conditions, generally identified as protein misfolding diseases. The largest group of misfolding diseases is associated with the conversion of specific peptides or proteins from their soluble functional states ultimately into highly organized insoluble fibrillar aggregates.

Alzheimer's disease (AD) is the most common and fatal neurodegenerative disorder directly associated with the deposition of such aggregates, particularly amyloid beta ($A\beta$) aggregates, in the brain of patients. Two variants of $A\beta$ peptides exist in human: $A\beta_{40}$ and $A\beta_{42}$. The former is the most abundant in the plaques and assumes a U-shaped structure only, while the latter is the most toxic, forms aggregates more rapidly and assumes a S-shaped structure. It has been recently demonstrated in literature that the S-shaped $A\beta_{42}$ specie is more stable and compact.

Moreover, mutations in the $A\beta$ peptide alter the toxicity, oligomerization pathways and rate of fibril formation and can be grouped in pathogenic mutations and synthetic mutations.

However, the precise mechanism of amyloid aggregation remains unclear. Computational Molecular Modelling may help to clarify mechanisms behind aggregation and oligomerization of $A\beta$. The present study focuses on conformational stability of two different models of S-shaped wild-type (WT), G33L and G37L mutants of amyloid $A\beta_{11-42}$.

Overall, computational results here presented highlight an enhanced stability for the WT when compared to mutant types of amyloid $A\beta_{11-42}$, suggesting that specific aminoacidic substitution may affect $A\beta_{42}$ aggregation propensity.

Estratto

Le proteine sono delle molecole complesse che svolgono diverse funzioni fondamentali nell'organismo e la struttura terziaria che assumono è strettamente correlata alla loro funzione. A volte, le proteine ripiegano in maniera scorretta causando l'insorgere di condizioni patologiche generalmente denominate patologie da errato ripiegamento proteico. La maggior parte di queste malattie è associata alla conversione di determinati peptidi o proteine solubili in aggregati fibrillari insolubili altamente organizzati.

La malattia di Alzheimer è il disturbo neurodegenerativo associato alla deposizione di tali aggregati più diffuso ed invalidante, ed in particolare è causato dall'aggregazione di beta-amiloidi ($A\beta$) nel cervello dei pazienti. Esistono due varianti di $A\beta$, nell'uomo: $A\beta_{40}$ e $A\beta_{42}$. La prima è la più abbondante nelle placche e assume solo una forma ad U, mentre la seconda è la più tossica, forma aggregati più rapidamente e può assumere sia forma ad U che forma ad S. Di recente, è stato dimostrato che il peptide $A\beta_{42}$ con forma ad S è più stabile e compatto.

Inoltre, mutazioni nel peptide $A\beta$ alterano la tossicità, la via di oligomerizzazione e il tasso di formazione di fibrille e possono essere raggruppate in mutazioni patogeniche e mutazioni sintetiche. Ad ogni modo, risulta ancora mancante una chiara e completa descrizione del meccanismo di aggregazione di amiloidi. La modellazione molecolare potrebbe aiutare a chiarire i meccanismi dietro i processi di aggregazione e oligomerizzazione di $A\beta$. Questo lavoro si concentra sulla stabilità conformazionale di due diversi modelli di $A\beta_{11-42}$ della specie wild-type (WT) e dei mutanti G33L e G37L.

Complessivamente, i risultati computazionali qui riportati evidenziano una maggiore stabilità della specie WT rispetto alle due specie mutate, suggerendo che specifiche mutazioni amminoacidiche possono influenzare la tendenza all'aggregazione dell'$A\beta_{42}$.

1. Introduction

The present chapter is devoted to a general introduction of the present Master Thesis work, elucidating aims and objectives.

The assembly of peptides into fibrillar aggregates is one of the major event in a broad range of human diseases. In general, changes of pH, temperature or ionic strength[1,2] may lead almost any protein and peptide to aggregate. Many of these form amyloid-like fibrils.
Although the cause of Alzheimer's disease (AD) is not really well understood, compelling evidence indicates that deposition of such amyloid aggregates is the central event in the disease pathogenesis.
Understanding the precise mechanism behind the protein aggregation is essential for the development of novel pharmaceutical approaches, but still unclear.
The aim of the present work is to investigate two S-shaped models of $A\beta_{42}$ and the effects of specific aminoacidic substitutions on $A\beta_{42}$. Computational molecular methods allow to shed light on the molecular reasons responsible for the conformational stability of polymorphs.

The work presented is divided into chapters, briefly summarized below:

Chapter 1 is the present introductory part.

Chapter 2 provides a general view of Material and Methods applied in the present work. The first section introduces computational molecular modeling for investigating biological mechanisms. The second section presents the "Molecular Mechanics" approach, with an overview on physical basis behind molecular modelling. In section 2.3, Molecular Dynamics (MD) is presented, with a brief reference to statistical ensembles and implementation schemes. The following paragraph (2.4) is devoted to a general description of dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Functional Mode Analysis (FMA).

Chapter 3 provides an insight into conformational dynamics and structural stability of two different S-shaped models of $A\beta_{11-42}$. MD results show that both models assume stable arrangements, especially 2MXU, probably due to the higher β-strand content. Functional Mode Analysis (FMA) highlighted 2MXU and 5KK3 are subjected to a torsion and a bending mode, respectively.

Chapter 4 yields a detailed conformational study of G33L and G37L mutants of two S-shaped models of $A\beta_{11-42}$. Computational results emphasize that the wild types (WTs) are characterized by higher

stability, when compared to mutant types. Mutations cause an increased flexibility of the central core and C-terminal region, a marked loss of β-strand content in those regions and a noticeable distortion of the structure. FMA clearly revealed a prevalent torsional mode, especially in the 2MXU model and a tendency to open from the lateral chain of G33L in the 5KK3 model.

2. Materials and Methods

2.1 Computational Modelling of Biomolecular Systems

Molecular Modelling includes a set of theoretical and computational techniques that provides insight into molecular systems at atomic level by solving the equations of quantum and classical physics. It is used to represent, simulate and analyse the properties and the behaviour of molecules and complex molecular structures (e.g., proteins, nucleic acids, polymers). To date, Molecular Modelling is rapidly growing, supported by the increasing computer capacity. However, molecular systems, and in particular biological systems, consist of huge number of molecules. Thus, estimating thermodynamic or kinetic properties of these systems still represents a challenge. Molecular Dynamics (MD) is a powerful tool for understanding molecular mechanisms of protein folding or unfolding, docking of molecules, transport of drugs, multiscale modelling, cellular membranes and much more.

The most detailed analysis of a biological system starts at the quantum level, by solving the Schrodinger equation (known as an ab initio calculation)[3]. However, Schrödinger equations can be solved only for simple systems.

At the molecular level, Molecular Mechanics (MM) allows to model molecular systems with thousands of atoms: the result of MD is a trajectory that specifies how position and velocity of particles vary with time[3].

In the next sections, MM and MD will be discussed in detail.

2.2 Molecular Mechanics

Molecular Mechanics methods use Newtonian mechanics to model molecular systems with thousands of atoms interacting through a potential energy function. A molecular force field (FF) is a set of equations that allows to estimate the potential energy of the system as a function of the atoms' position. According to the Born-Oppenheimer approximation[4], the Molecular Mechanics method treats atoms as spheres whose mass depends on the elements and treats bonds as springs whose stiffness depends on the elements bonded together.

2.2.1 Potential Energy Function

The energy of a molecule in its ground electronic state can be considered as a function of the nuclear coordinates only. Any changes in the system (e.g., the torsion of an angle, the displacement of an atom) cause the variations on a multidimensional 'surface' called energy surface. The potential

energy surface for a molecular system of N atoms in a given conformation is the sum of four energy contributions:

$$V = V_{bonded} + V_{non-bonded} \tag{1}$$

Where the two terms are given by the following equation:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} \tag{2}$$

$$V_{non-bonded} = V_{electrostatic} + V_{Van\ der\ Waals} \tag{3}$$

2.2.2 Treatment of Bond and Non-Bond interactions

The central core of the MM method is the definition of a potential energy function. As the atoms are treated as spheres and bonds are treated as springs, the potential energy function can be described as[5]:

$$\begin{aligned} V(r_1, r_2, \dots, r_N) &= \sum_{bonds} \frac{1}{2} k_l [l - l_0]^2 + \sum_{angles} \frac{1}{2} k_\theta [\theta - \theta_0]^2 \\ &+ \sum_{dihedrals} k_\varphi [1 + \cos(n\varphi - \delta)] + 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}^{12}}{r_{i,j}} \right) - \left(\frac{\sigma_{i,j}^{6}}{r_{i,j}} \right) \right] + \frac{Q_i Q_j}{4\pi\varepsilon_0 r_{i,j}} \end{aligned} \tag{4}$$

In Eq.4, the first term describes the interactions between bonded atoms, modelled with a harmonic potential, where k_l is the force constant, l_0 is the reference bond length and *l* is the bond length;
The second term of the equation 2 encompasses the angle among three atoms covalently bound together and the potential energy is expressed as follow, where k_θ is the stiffness, θ_0 is the equilibrium angle and ϑ is the bond angle;
Last term in Eq.2 describes the dihedral bond interactions, that usually includes a torsional potential describing bond rotates: k_φ is the energetic barrier related to the angle deformation, δ is the phase that determines the minimum position for the torsional angle and n is the multiplicity.
Non-bond interactions are usually modelled as functions inversely proportional to the distance between two atoms and includes two components: Van der Waals forces and Electrostatic interactions.

The Van der Waals interactions are relatively weak when compared to covalent bonds, but strongly affect properties of organic compounds, such as their solubility. Once atoms are close enough, their electronic clouds overlap and they will repel each other with repulsive forces. The Lennard-Jones equation[6] is often used to represent the Van der Waals potential:

$$V_{Lennard-Jones} = 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{i,j}^{12}}{r_{i,j}}\right) - \left(\frac{\sigma_{i,j}^{6}}{r_{i,j}}\right)\right] \quad (5)$$

The first term in Eq.5 describes the repulsion forces; the second term represents the attraction energy. $\sigma_{i,j}$ and $\varepsilon_{i,j}$ represent the collision diameter and the depth of the potential well, respectively.

The electrostatic interactions are defined as long-range interactions, because the energy decrease with the distance and can be described by using the Coulomb's law:

$$V_e = \frac{Q_i Q_j}{4\pi\varepsilon_0 r_{i,j}} \quad (6)$$

The number of non-bond parameters increase as the square of the number of atoms in the system, therefore the computational cost required is very high. In order to decrease such computational expense, the Ewald summations is applied. The Ewald summation replace the interaction energies summation in real space with an equivalent summation in Fourier space and it is employed in the long-range interactions' calculation.

2.2.3 Periodic Boundary Conditions

The treatment of boundaries effects is a crucial issue in simulations, since boundary effects strongly influence the properties of the whole system. The Periodic Boundary Conditions (PBC)(Figure 1) allows to perform simulations with smaller number of particles. All the atoms are put in a box, usually filled with implicit or explicit models of water, surrounded by copies of itself. The presence of Periodic Boundary Conditions causes imprecisions, but still less severe than the error resulting from artificial boundary with vacuum.

Each particle in the box interacts with any other particles in adjacent boxes, which are repeated infinitely. Applying a Lennard Jones cut-off distance leads to neglectable errors, since it decreases very rapidly; On the other hands, the long-range interaction may require more accurate methods, with a view to avoid discontinuities in the potential energy calculation: Particle Mesh Ewalds[7],

Multipole Cells[8] and Reaction Field[9] are some methods proposed to overcome these kinds of problems.

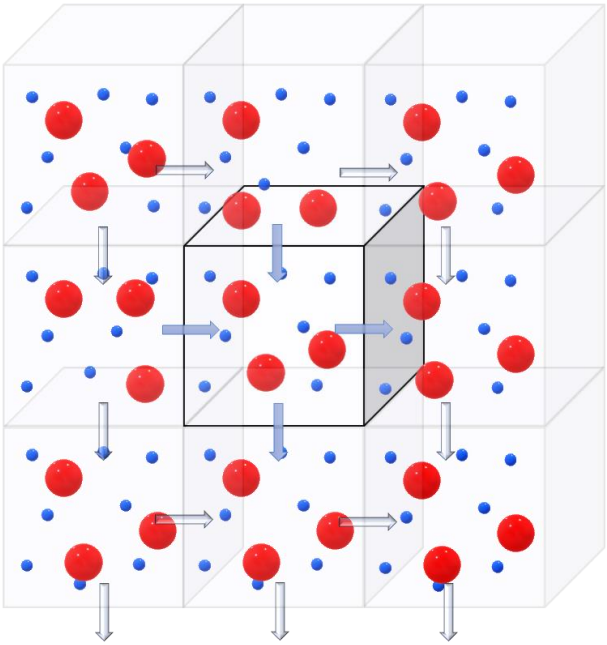

Figure 1: Scheme of periodic boundary conditions (PBC); the central box is filled with water and repeated infinitely.

2.2.4 Potential Energy Minimization

The Potential Energy function is a very complex multidimensional function: in an *N*-atoms system it is defined by *3N* Cartesian coordinates and *3N-6* internal coordinates (bonds, angles and torsional angles). There are a very large number of stationary points on the energy surface (Figure 2), where the forces on all the atoms are equal to zero and the first derivative of the energy with respect to the Cartesian or internal coordinates is zero. Minimum points correspond to stable structures. Any changes from the minimum configuration is characterized by higher energy. The energy minimization is able to reduce the potential energy of the system and two different methods are employed for locating stationary points: derivative and non-derivative methods. First order derivative methods change atom coordinates moving to lower energies and the starting point for each interaction is the molecular configuration obtained from the previous one (Steepest Descent[10], Conjugate Gradient[11]). Second order derivative methods provide information about the curvature of the function and calculate the inverse Hessian matrix of second derivatives (Newton-Raphson[12], L-BFGS[13]). These methods may be computationally demanding and can also require a significant amount of storage.

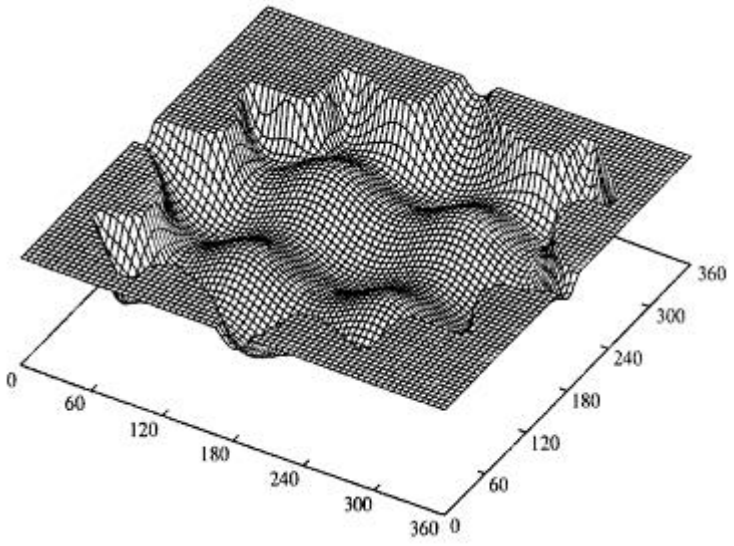

Figure 2: Potential Energy Surface[3] (PES), it is a multi-dimensional surface. It is characterized by locally and globally minima.

2.3 Molecular Dynamics

Molecular Dynamics (MD) is a computational approach that can predict the movement of large proteins in solutions in form of a trajectory, as a function of the time and allows to calculate average properties of a system by solving the Newton's equations of motion.

2.3.1 Statistical Ensemble

There are two types of macroscopic properties in a chemical system: static equilibrium properties (e.g., radial distribution function, averaged potential energy) and non-equilibrium properties (e.g., diffusion processes, dynamics of phase changes).

To compute such properties Boltzmann and Gibbs introduced the concept of *statistical ensemble: "*an ensemble is a collection of all possible systems which have different microscopic states but have an identical macroscopic or thermodynamic state"[14]. There are different ensembles:

- The Micro-Canonical Ensemble (NVE), it corresponds to an isolated system and it is characterized by a fixed volume, energy and number of atoms;
- The Canonical Ensemble (NVT), it corresponds to a closed system and it is characterized by a fixed volume, temperature and number of atoms;
- The Grand Canonical Ensemble (μVT), it corresponds to an open system and it is characterized by a fixed volume, temperature and chemical potential;

- The Isobaric-Ensemble (NPT), it is characterized by a fixed pressure, temperature and number of atoms.

The ensemble average of a property is calculated as follow:

$$\langle A \rangle_{ensemble} = \iint d\mathrm{p}^N d\mathrm{r}^N A(\mathrm{p}^N, \mathrm{r}^N) \rho(\mathrm{p}^N, \mathrm{r}^N) \tag{7}$$

Where r is the atomic positions, p is the momenta, $A(\mathrm{p}^N, \mathrm{r}^N)$ is the property of interest and $\rho(\mathrm{p}^N, \mathrm{r}^N)$ is the probability density function. More in deep, the probability density function is given by:

$$\rho(\mathrm{p}^N, \mathrm{r}^N) = \frac{1}{Q} \exp[-H(\mathrm{p}^N, \mathrm{r}^N)/k_b T] \tag{8}$$

In the equation above, k_b is the Boltzmann factor, T is the temperature and H is the Hamiltonian. Q is the partition function, expressed by:

$$Q = \iint d\mathrm{p}^N d\mathrm{r}^N exp[-H((\mathrm{p}^N, \mathrm{r}^N)/k_b T] \tag{9}$$

The partition function is a sum of Boltzmann factor over all microstates of the system and it relates microscopic thermodynamic variables to macroscopic properties.
However, the integral is generally extremely difficult to calculate because it is necessary to calculate all possible states of the system. According to the ergodic hypothesis, over long periods of time, the time average of the property is equal to the ensemble average of the same property and the time average can be computed as follow:

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\mathrm{p}^N(t), \mathrm{r}^N(t)) dt \simeq \frac{1}{M} \sum_{t=1}^{M} A(\mathrm{p}^N, \mathrm{r}^N) \tag{10}$$

In Eq.10, t is the simulation time, M is the time steps in the simulation and $A(\mathrm{p}^N, \mathrm{r}^N)$ is the instantaneous of A.

2.3.2 Molecular Dynamics Implementation Scheme

Molecular Dynamics calculates the dynamics of the system by sampling microstates in a specific ensemble. It is a deterministic method, that means the state of the system at any time step can be predicted from its previous state. The central idea in MD method is solving Newton's equation of

motion for a set of atoms which interact via potential energy function, known as force field. According to the Newton's second law, the acceleration for each atom is:

$$a = -\frac{1}{m}\frac{dV}{dr} \tag{11}$$

As the potential energy is a complex function of the atomic position, this equation cannot be solved analytically, but must be solved numerically. There are different integration methods, for example the Verlet algorithm[15], the Leap-frog algorithm[16], the Velocity Verlet[17] and many more.

The flow chart in Figure 3 describes the MD: the initial structure is provided as input data, usually extracted from experimental data (e.g., Protein Data Bank). Starting from the input, the potential energy is calculated; Afterwards, the algorithm continues with the calculation of the forces F_i, by deriving the potential energy function; Then, the integration method calculates a new set of atomic positions r_i and velocities v_i and the cycle goes on for the number of steps required until the equilibrium is reached. Lastly, the output trajectory is defined, and the macroscopic thermodynamic properties can be calculated as time averages.

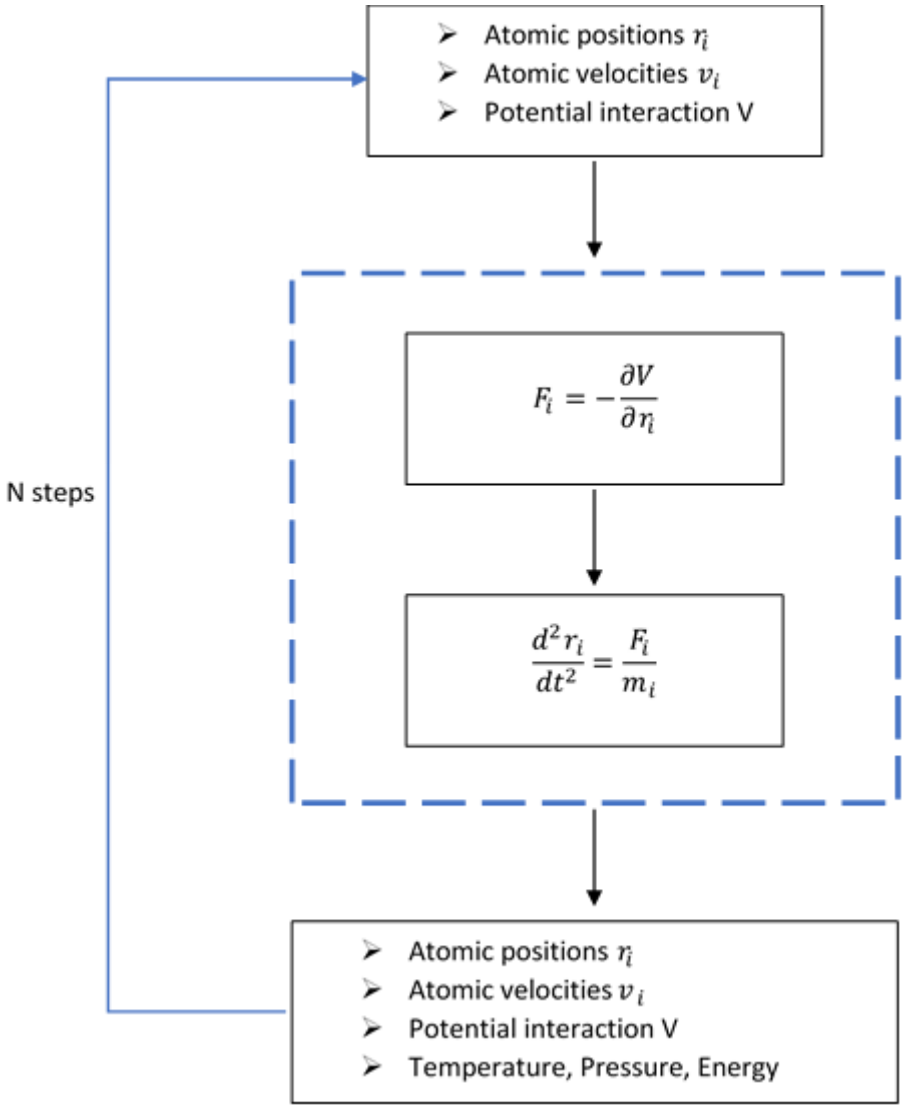

Figure 3: MD flow-chart: starting from the initial positions and velocities provided as input, the potential interaction is calculated; the algorithm goes on with the calculation of forces and, through integration methods, new positions and velocities are generated; the cycle continues for a certain number of steps, until the equilibrium is reached.

2.4 Dimensionality Reduction Techniques

Conformational structure of protein is strictly related to its function. However, proteins are not static molecules, but populate ensembles of conformations, in order to find the most stable one. For this reason, high-dimensional data-sets are produced by MD simulations, but just a few of them are important. Principal Component Analysis (PCA)[18] and Functional Mode Analysis (FMA)[19] are two efficient technique usually employed for reducing the degrees of freedom and simplify the analysis of protein dynamics.

2.4.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure that can be used to reduce a large set of variables into a small set of linearly uncorrelated variables, called principal components. This technique is based on the covariance matrix, which is diagonalized and the eigenvectors are ordered according to descending eigenvalues (σ_i^2). The motion of the protein is quantified as Mean Square Fluctuations (MSF) is the sum of *3N* contributions from different PCS:

$$\langle (x - \langle x \rangle^2 \rangle = \sum_{j=1}^{3N} var(p_j) = \sum_{i=1}^{3N} \sigma_i^2 \tag{12}$$

Where p_j is the projection of the motion along the j^{th}.

Generally, a large part of the fluctuations of the system can be described in terms of only few principal components[20]. However, in some cases, a functionally relevant mode is not identical to a PCA mode because it is distributed over a number of PCA modes. Therefore, Functional Mode Analysis was recently developed.

2.4.2 Functional Mode Analysis

Functional Mode Analysis (FMA)[19] is a technique that allows to identify collective motion of a system, that is maximally correlated to a quantity of interest, such as the radius of gyration, the number of hydrogen bonds or any quantity that is important for the dynamics of the proteins. The central idea of FMA[19] is the research of a vector **a** that is maximally correlated to the change in the functional quantity f_i. The collective vector Maximally Correlated Motion (MCM), if f is linearly correlated with the first n principal components (PC), is obtained as:

$$MCM = \alpha_i PC_i \tag{13}$$

α_i is obtained by maximizing the Pearson correlation coefficient (R):

$$R = \frac{cov(f, MCM)}{\sigma_f \sigma_{MCM}} \tag{14}$$

Where $cov(f, MCM)$ and σ represent covariance and standard deviation, respectively.

The maximization of R with large basis set may induce overfitting, therefore cross-validation is fundamental. An approach is to divide simulations into frames for model building and frames for cross-validation[19].

3. Conformational Polymorphism of S-Shaped Amyloid Beta Fibrils

Abstract

A broad range of human diseases is characterized by the deposition and aggregation of proteins. For example, neurodegenerative diseases like Parkinson's Disease, Amyotrophic Lateral Sclerosis and Alzheimer's Disease. Alzheimer's disease (AD) is the most common presenile neurodegenerative disease, characterized by the deposition and aggregation of amyloid-beta (Aβ). Two principal variants of Aβ exist in humans: $A\beta_{40}$, that is most abundant, and $A\beta_{42}$, that is the most toxic. Fibrils of $A\beta_{40}$ can only assume a U-shaped conformation while $A\beta_{42}$ can also arrange as S-shaped. Interestingly, the S-shaped is most compact and stable. Solid-State NMR analysis revealed several structures for the S-shaped $A\beta_{42}$ fibrils. In this study, Molecular Dynamics (MD) simulations were carried out for characterizing and comparing the conformational dynamics and structural stability of two different S-shaped $A\beta_{42}$ fibril structures. The outcome of this work shows the importance of different arrangements for the conformational dynamics in physiological environment, that may affect the aggregation propensity.

3.1 Introduction

The transition of native proteins to a partially unfolded and aggregated state is implicated in several human illnesses. The aggregation of peptides is usually promoted by changes in temperature, pH, salt concentration and many more[21]. Such peptides are able to aggregate into highly ordered fibrils and fibers[22], generally known as amyloid fibrils or plaques. Amyloid β is the major constituent of fibrils deposited in the brain of patients with Alzheimer's disease (AD). Two variants of $A\beta$ peptides, which derive from the transmembrane amyloid precursor protein (APP) via endoproteolytic cleavage by β- and γ- secretase[22], exist in human: $A\beta_{40}$ and $A\beta_{42}$. $A\beta_{40}$ is most abundant (90%) in the brain[23], but $A\beta_{42}$ tends to aggregate more rapidly[24] and represent a more toxic specie. On the basis of solid-state NMR restraints, different structural models of $A\beta_{40}$ and $A\beta_{42}$ fibrils have been proposed[25–32], since fibrils are characterized by polymorphism[33]. In detail, $A\beta_{40}$ can adopt only a U-shaped structure[34] characterized by β-strands (Y10-D23, A30-G38), connected by a bend (G25-G29)[35] and stabilized by interchain hydrogen bonds and salt-bridges (D23-K28). The $A\beta_{42}$ can arrange also in a S-shaped structure[34]. Recent studies[36] demonstrated that the S-shaped $A\beta_{42}$ is the most stable and compact specie. The S-shaped $A\beta_{42}$ species consist of three β-strands: β_1 strand is made of residues V12-V18, β_2 strand of residues V24-G33 and β_3 strand of residues V36-V40,

connected by coil and turn regions and stabilized by intermolecular hydrogen bonds and salt-bridges. Different hypotheses have been formulated in order to explain why the S-shaped structure is unstable in $A\beta_{40}$ species: Xiao et al[32] proposed that the S-arrangements in $A\beta_{42}$ is stabilized by the salt bridge between K28 and A42, absent in $A\beta_{40}$. Recently, it has been proposed that its stability depends on hydrophobic contacts between I41 and A42[34]. Amyloid fibrils have been studied in a huge number of both experimental and computational works, given the intimate relationship between conformational structure and disease onset. Computational approaches can be a powerful tool able to yield an insight into molecular level and provide detailed studies of molecular mechanisms, especially for relatively simple systems like short peptides[37]. Within this chapter, the conformational dynamics and stability of two different models of S-shaped $A\beta_{11-42}$ peptides have been investigated by performing Molecular Dynamics (MD) simulations.

3.3 Materials and Methods

3.3.1 Models

Two different 9-mer models of S-shaped $A\beta_{11-42}$ specie were studied: one extracted from PDB ID: 2MXU[38] and one from PDB ID: 5KK3[39] (Figure 4).

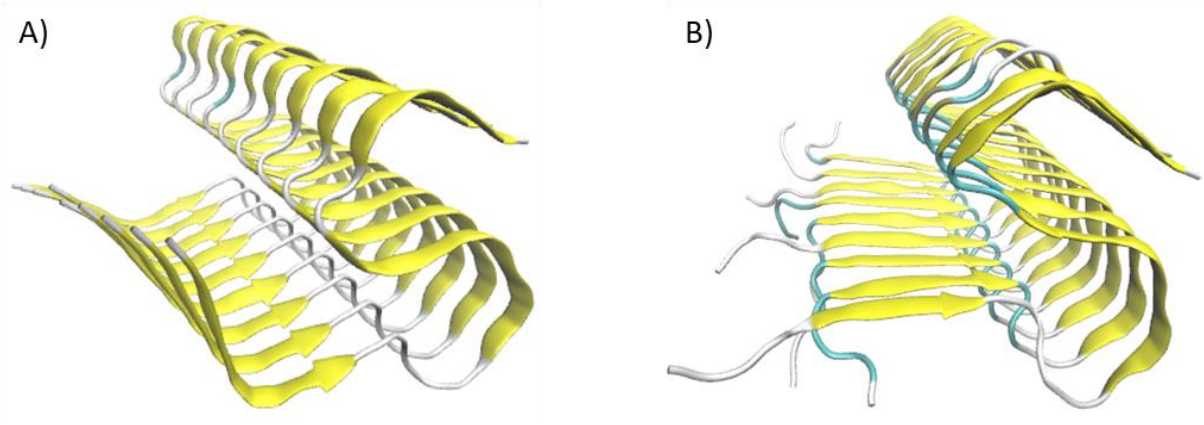

Figure 4:The two different models studied: A) 2MXU[32] B) 5KK3[31]

The models were solvated in an 8 nm cubic box and neutralized by counterions. Each system consisted of about 50,000 particles.

3.3.2 Simulation Setups

The CHARMM36 force field[40] was used to define protein topologies and the TIP3P[41] model was used for the explicit solvent. GROMACS 5.1.4[42] version was employed for simulations. Steepest descent algorithm[10] was utilized to minimize the system. Then, a 50 ps simulation in NVT ensemble was

conducted, by applying the V-rescale algorithm[43] to keep temperature at 300 K and time constant of 0.1 ps. In order to increase the statistic, five replicas for each $A\beta_{11-42}$ model were generated. We ran a 50 ps simulation in NPT ensemble for each replica. Berendsen[44] and V-rescale[43] coupling methods were used as pressure and temperature coupling. Finally, we performed an MD simulation of 100 ns duration. Nose-Hoover[45] and Parrinello-Rahman[44] were used as temperature and pressure coupling; The LINCS algorithm[46] was used to constrain the length of h-bonds. Periodic boundary conditions were applied in all three dimensions, the PME algorithm[47] was employed to calculate electrostatic interactions and Van der Waals were defined within a cut-off of 1.0 nm. Trajectories were extracted every 50 ps of simulation and the Visual Molecular Dynamics (VMD) package was employed to display the simulated systems. The secondary structures were obtained by using STRIDE web server, by pulling out the first and the last frame of the MD trajectory of each model.

3.3.3 Order Parameter and Functional Mode Analisys (FMA)

With the aim of estimating the structural stability of the models and therefore the alignment among protein chains, an order parameter was calculated as below:

$$ordP = \frac{1}{N}\sum_{r=11}^{42}\frac{\langle v_r, x\rangle}{\|v_r\| \cdot \|x\|} = \frac{1}{N}\sum_{r=11}^{42}\cos\alpha \tag{15}$$

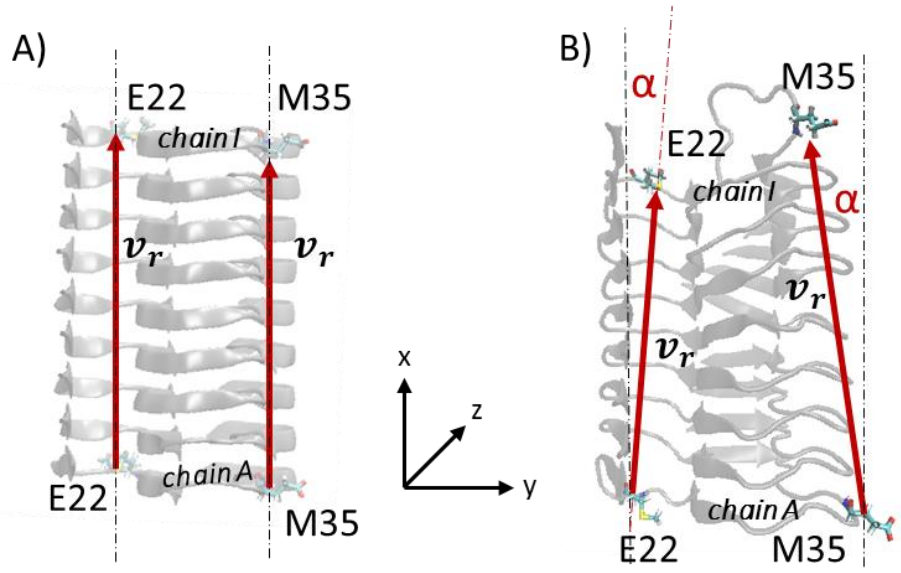

Figure 5: Schematic representation for the calculation of order parameter value. The order parameter is equal to 1 in a completely aligned fibril (A) and lower than 1 in a distorted fiber (B).

Where v_r is the vector joining C_α-atom of each residue *r* of chain A with the corresponding C_α-atom of the same residue of chain I and *x* is the fibril axis. If ordP assumes values close to 1, the chains maintain an alignment to the initial structure; if the values of ordP are lower than 1, the overall structural order decreases. The above mentioned order parameter have been recently used with success in a computational study[48]. The functional mode analysis (FMA) was performed to reduce the functionality of the system. This technique detects a single collective motion that is maximally correlated with a quantity of interest[49], that is the order parameter. Assuming the order parameter is a linear function of the Principal components (PC), the Maximally Correlated vector is obtained by maximizing the Pearson correlation coefficient[49]. In FMA, data-set is divided into a subset of frames for model building and a subset for cross-validation.

3.4 Results

The conformational stability of each molecular system was analyzed by computing the Root Mean Square Deviation (RMSD) (Figure 4), showing that convergence is reached in the last 20 ns for all MD replicas. In particular, RMSD swings around 0,4-0,5 nm.

Each computational data here reported is thus calculated as a time average over the last 20 ns of all trajectories, unless otherwise specified.

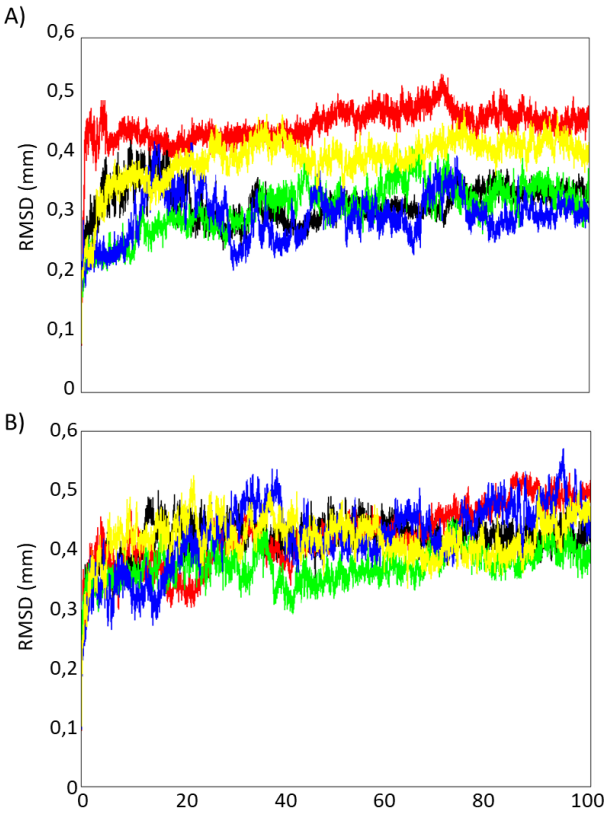

Figure 6: RMSD from the minimized structure, calculated on each simulation of the two models: A) 2MXU, B) 5KK3 showed reached convergence for all configurations.

The protein fluctuations were evaluated by Root Mean Square Fluctuation (RMSF) analysis (Figure 7) and shows how N-terminal tail is the most flexible region in both models (2MXU and 5KK3) because of the solvent exposure. It is worth mentioning that the 5KK3 model shows an increased value of RMSF in the region E11-K16 if compared with 2MXU model. In the central core of both models, the fluctuation values are low, except for residues E22 and D23 and almost identical. A remarkable difference is located at residue M35, that is much higher in 5KK3 ($\mathrm{RMSF}_{\mathrm{M35}}[2\mathrm{MXU}] = 0{,}085 \pm 0{,}009\ \mathrm{nm};\ \mathrm{RMSF}_{\mathrm{M35}}[5\mathrm{KK3}] = 0{,}125 \pm 0{,}006\ \mathrm{nm}$). The C-terminal region is quit rigid, since RMSF is lower than 0,1 nm.

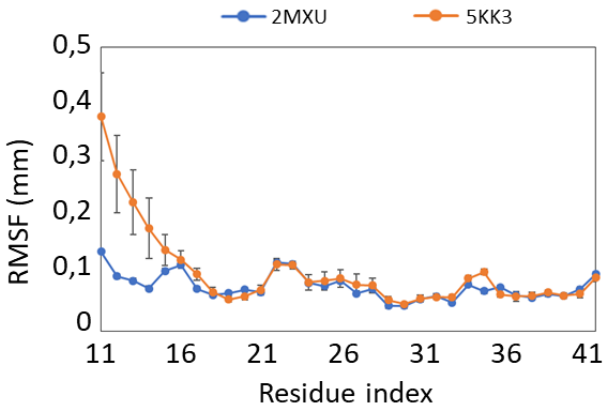

Figure 7: RMSF of structures as a function of residue and relative standard deviation, for the two models, calculated as the average over the last 20 ns of MD trajectory for each simulation.

The previously highlighted differences in fluctuation values of the N-terminal region can be explained analyzing the secondary structure probability of the models, displayed in Figure 8. Although rigid β-strands and flexible coils are the most predominant secondary structures in both models, several differences are located along chains. The structures differ in the region V12-H14 and G33-V36. Particularly, β-strand content of 5KK3 (20%) in V12-H14 region is deeply lower when compared to that of 2MXU (95%) and it is well known β-strands stabilize peptides, give strength and stability to the structure[50].

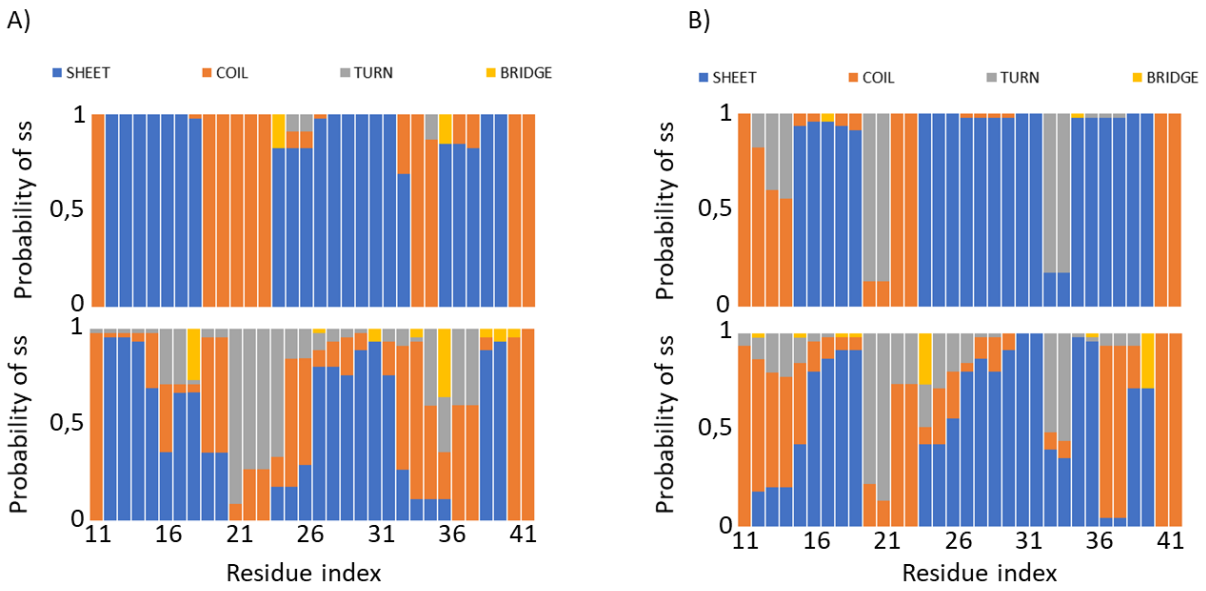

Figure 8: Probability of secondary structure averaged on the chains for the two models: A)2MXU and B)5KK3. The upper row in each panel shows the initial configuration and the lower panel shows the final configuration. The

secondary structures were obtained by pulling out the first and the last frame of each simulation, by using STRIDE web server.

The tendency of each protein residue of exposing itself to the solvent is evaluated by Solvent Accessible Surface Area (SASA)[51] per residue (Figure 9). The N-terminal is the most solvent exposed region in both model, but the values in 5KK3 are higher than 2MXU. This evidence is in agreement with previously reported RMSF (Figure 7) and secondary structure probability (Figure 8). In the central core, it is noticeable that whereas residues M35 is more exposed to the solvent in the 5KK3 model with respect to 2MXU ($\mathrm{SASA_{M35}[2MXU]} = 0{,}19 \pm 0{,}04\ \mathrm{nm^2}$; $\mathrm{SASA_{M35}[5KK3]} = 0{,}88 \pm 0{,}13\ \mathrm{nm^2}$), residue V36 exhibits an opposite behavior ($\mathrm{SASA_{V36}[2MXU]} = 0{,}70 \pm 0{,}10\ \mathrm{nm^2}$; $\mathrm{SASA_{V36}[5KK3]} = 0{,}17 \pm 0{,}05\ \mathrm{nm^2}$).

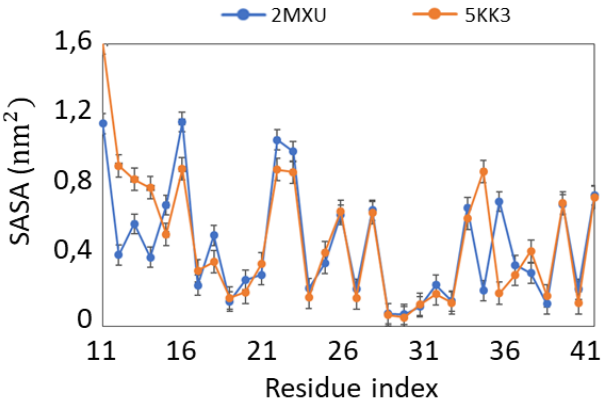

Figure 9: Solvent Accessible Surface Area per residue and relative standard deviation, calculated as the average over the last 20 ns of MD trajectory for each simulation.

Figure 10.A displays the total SASA for the two models and 5KK3 ($\mathrm{SASA[5KK3]} = 136{,}22 \pm 4{,}08\ \mathrm{nm^2}$) is slightly higher than that of 2MXU ($\mathrm{SASA[2MXU]} = 128{,}34 \pm 3{,}00\mathrm{nm^2}$). This result suggests that 2MXU reaches a more compact arrangement.

The same conclusions can be drawn by plotting the inter-chain contact area (Figure 10.B). It is interesting to note that the lateral chains are the most unstable because more surface is exposed to the solvent. The overall contact area between adjacent chains is higher in 2MXU than 5KK3, highlighting the ability of 2MXU model to maximize the inter-chain contacts.

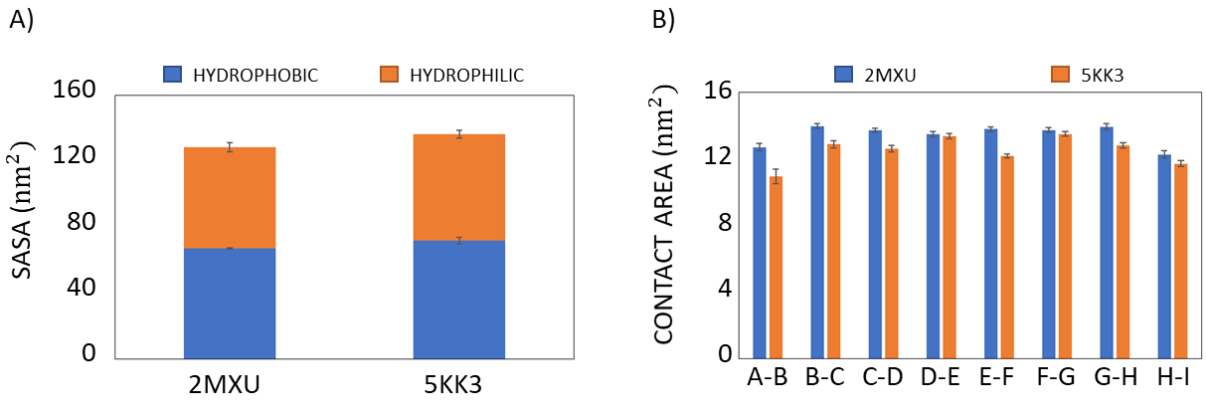

Figure 10: A) Total Solvent Accessible Surface Area and relative standard deviation, B) Contact area between adjacent chains. Both metrics was calculated as an average over the last 20 ns of each simulation.

With the aim of elucidating large-scale and low-frequency modes associated to the MD trajectory, Principal Component Analysis (PCA) was performed. The amplitude of the eigenvalues corresponds to the variance of the protein motion, usually quantified as mean square fluctuations, along their relative eigenvectors yields by PCA. The sum of the first 5 eigenvalues in the 5KK3 (6,37 nm^2) are significantly higher than those of 2MXU ($3{,}96\ \mathrm{nm}^2$), confirming that the former seems to be less stable.

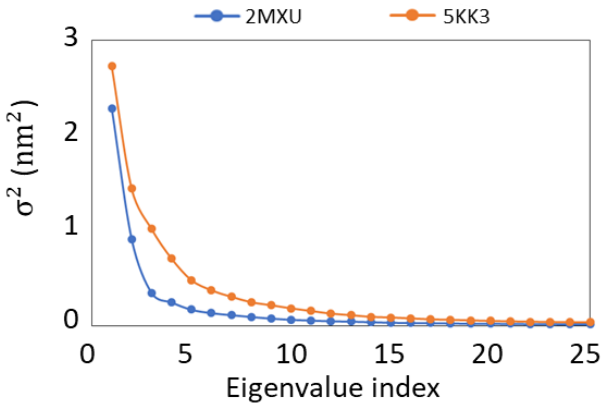

Figure 11: Eigenvalues and relative amplitudes.

Order Parameter and Functional Mode Analysis

The expected values and corresponding standard deviations of order parameter are displayed in Figure 12. Both models show a sharp distribution, peaked at 0,985 and 0,989, for 2MXU and 5KK3, respectively ($\sigma_{2\mathrm{MXU}} = 1{,}57\mathrm{e}^{-3}, \sigma_{5\mathrm{KK3}} = 2{,}62\mathrm{e}^{-3}$) indicating that the ordered structure is maintained, and chains remain aligned.

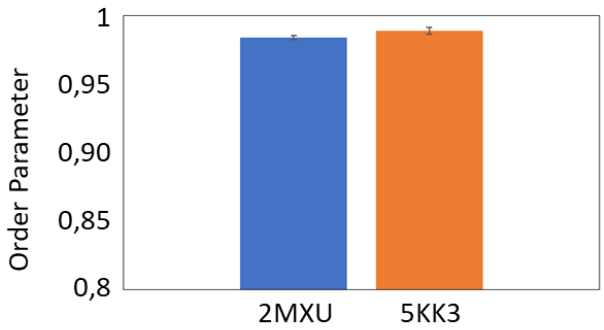

Figure 12: Expected value and corresponding standard deviation of probability of the order parameter, calculated over the last 20 ns of trajectories.

The Functional Mode Analysis is able to detect a single collective motion that is maximally correlated to a quantity of interest. In the case here studied, the quantity of interest is the order parameter as a function of the time.

The models were validated by predicting the quantity of interest with a Pearson correlation coefficient higher than 0,90 in 2MXU and 0,80 in 5KK3 (Figure 13).

Observing the initial and the final frames (Figure 14) extracted from trajectories along ewMCM vector clearly show 2MXU and 5KK3 models are subjected to a torsional mode.

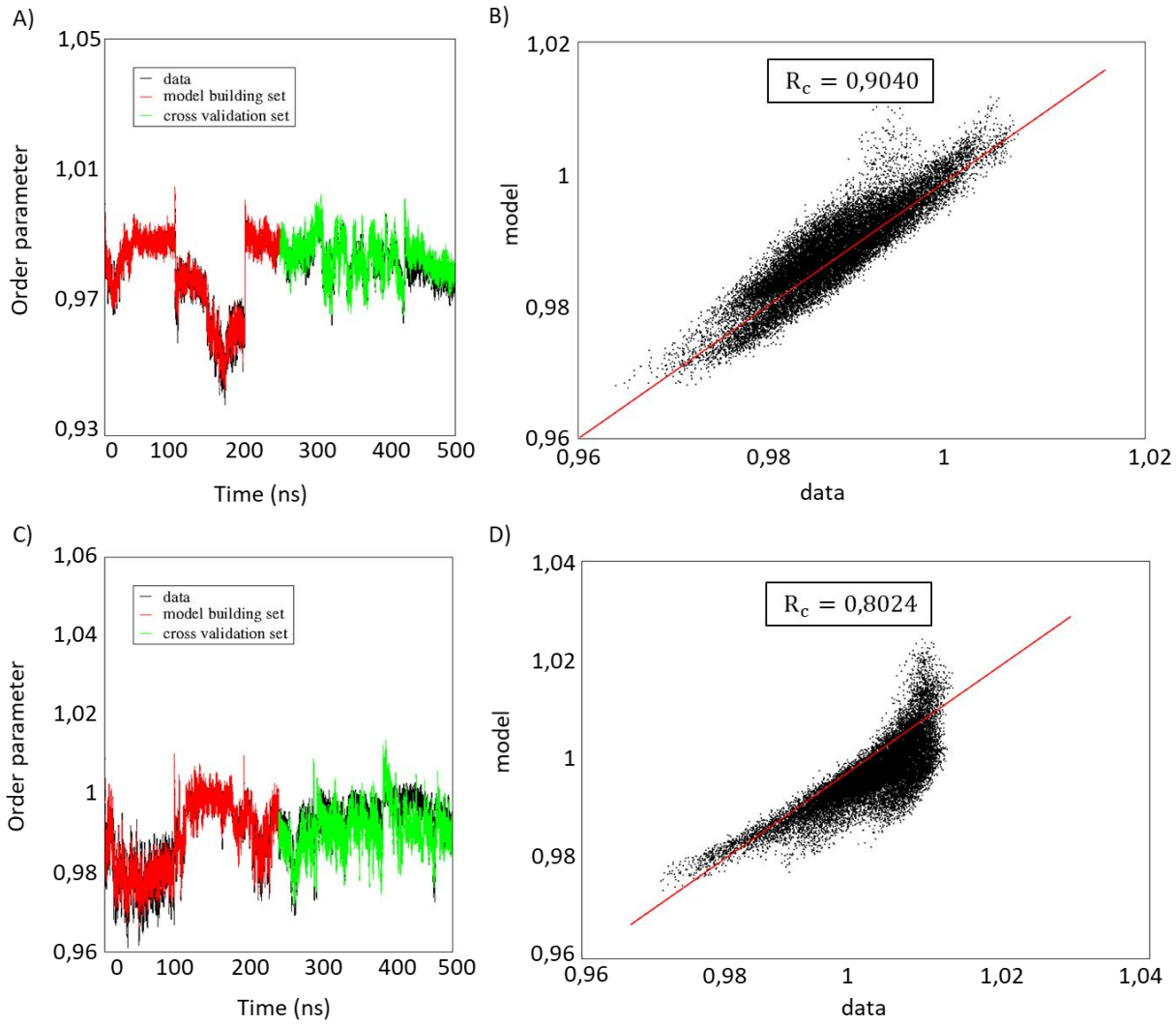

Figure 13: A) and C) Order parameter (black) along the five MD trajectories of 2MXU and 5KK3, respectively. The model in the building set (red) and in the validation set (green). B) and D) show the scatter plot of the cross validation set.

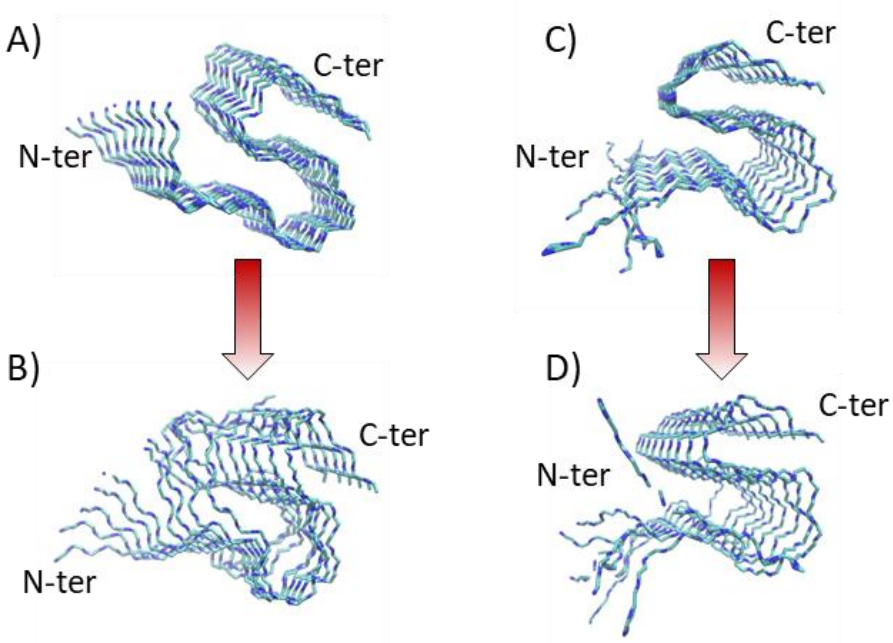

Figure 14: A) B) and C) D) show the conformational changes of 2MXU (left) and 5KK3 (right). Snapshots are obtained from ewMCM trajectories.

3.5 Discussion

Proteins are complex molecules that play many fundamental roles in human body and their conformational structure is strictly related to their function. Sometimes, peptides misfold or proteins fail to adopt, or remain in, its functional conformational state and thus pathological conditions arise. These pathological conditions are generally referred to as protein misfolding (or protein conformational) diseases.

The largest group of misfolding diseases is associated with the conversion of specific peptides or proteins from their soluble functional states ultimately into highly organized fibrillar aggregates. These structures are generally described as amyloid fibrils or plaques. Alzheimer's Disease is an example, which is characterized by cerebral deposits of amyloid-β peptides.

Two types of Aβ exist in the brain of patients with AD, the most abundant $A\beta_{40}$ and the more toxic $A\beta_{42}$. The difference between the two variants of peptide is the presence of two additional residues in $A\beta_{42}$: I41 and A42. It has also been proposed[34,38] that these two residues play a key role in the stabilization of S-shaped conformation of $A\beta_{42}$, that is unstable in the $A\beta_{40}$ variant. A conformational study revealed the S-shaped structure of $A\beta_{42}$ is the most stable variant[36] and that is why we focused on two different arrangements of such structure.

Despite the models here investigated share a S-shaped motif and the same peptide sequences, significant differences in the morphology and structure have been denotated. Particularly, the two

models differ in the starting secondary structure ($\beta_{2MXU} = 52\%, \beta_{5KK3} = 45\%$) and in β-strands sequences[31,32].

The changes in metrics here evaluated are in line with changes in the secondary structure (Figure 8). In detail, the higher beta-sheet content in the N-terminal makes 2MXU less flexible than 5KK3 in that region. On the other hands, β-strands reduction, associated with the increased solvent exposure (Figure 10.A), can explain the high fluctuation of region H14-K16 of 2MXU. It is important to note that such evidences are confirmed by literature, since it has been widely demonstrated that β-strands give stability and strength to proteins[50,52] and protein structure and solvent exposure are closely related[51] . Another meaningful change in fluctuations is located at residue M35. In 5KK3 arrangement, this residue is involved in a turn motif and, consequently, has a higher surface exposed to the solvent.

These findings suggest that the 2MXU globally assumes a more stable and compact conformation, as also showed by the total SASA values and by the contact area between adjacent chains (Figure 10) that highlight the ability of 2MXU to maximize the inter-chain contacts. This hypothesis seems being confirmed by the PCA analysis. The amplitude of the eigenvalue, representing the variance of the protein motion along the corresponding eigenvector[53] indicates that 5KK3 is subjected to a bigger deformation. However, despite various differences in the metrics here computed, the order parameter applied shows the alignment between chains and the structural stability of the models are substantially maintained.

In summary, both arrangements are quite stable in water and preserve the S-shape and chains remain aligned (Figure 12). Nevertheless, a partially destabilization can be noted, especially in 5KK3 model, in terms of secondary structures and fluctuations.

Based on our data, the importance of different arrangements on the conformational stability and dynamics behavior in physiological environment of S-shaped $A\beta_{11-42}$ has been showed. In this scenario, this work can be very relevant to shed the light behind mechanisms of aggregation and oligomerization of different amyloid-β species.

3.6 Conclusions

The aggregation of amyloid-β peptides plays a key role in the onset of Alzheimer's disease and $A\beta_{42}$ is the most toxic specie. In this chapter, two different models of the S-shaped $A\beta_{42}$ peptide were analyzed. Their conformational stability and dynamic behavior were evaluated employing MD simulations and correlated to the fluctuation of an order parameter, early reported in literature[36].

Our findings showed that both models maintain an ordered structure, in spite of meaningful changes in the secondary structure, although 2MXU seems to be able to assume a more stable and compact arrangements.

The next chapter will provide a detailed study about the impact of specific aminoacidic mutations on the two different models.

4. Impact of Specific Aminoacidic Substitutions on Amyloid Beta Aggregation Propensity

Abstract

The aggregation of amyloid-beta 42 ($A\beta_{42}$) peptides is associated with the pathogenesis of Alzheimer's Disease. Within the hydrophobic core of the Aβ sequence, there is a repeated GxxxG involving essential for assuring stability and promoting the process of fibril formation, called *glycine zipper*. Mutations in this motif lead to a completely different oligomerization pathway and rate of fibril formation. In this work, we have tested two variants containing substitutions in the glycine zipper. We found that G37L and G33L, mostly, undergo remarkable changes in the conformational stability, confirming that the disruption of glycine zipper can reduce the aggregation propensity of $A\beta_{42}$. If confirmed with further investigations, this hypothesis could pave the way for developing new Aβ inhibitors.

4.1 Introduction

Alzheimer's disease (AD) is the most common presenile neurodegenerative disease. The patient with AD has short-term memory loss, mood swings, problems with the language. Three main hypotheses have been proposed about the AD etiology[54] : Tau hypothesis[55], Cholinergic hypothesis[55] and Amyloid cascade hypothesis[56]. Amyloid hypothesis is the most widely accepted and supported hypothesis. Amyloid β derives from the transmembrane amyloid precursor protein (APP), via endoproteolytic cleavage by β- and γ- secretase[22].

The proteolytic cleavage of APP produces Aβ peptides of different amino acid length. Senile plaques are primarily composed of $A\beta_{40}$, the most abundant, and $A\beta_{42}$, the most toxic[23,24,57,58].

The exact structural knowledge of the fibril is essential for a complete understanding of the mechanism of fibril formation. For this purpose, several experimental and computational studies have been focused to gain insights into the structure of $A\beta_{42}$[59,60,69,61–68].

Recently, six residues were identified to be essential for the formation of $A\beta_{42}$ fibrils: H14, E22, D23, G33, G37 and G38[65]. Mutations of such residues modify the physico-chemistry and alter the toxicity of peptides.

Computational studies, especially when associated with experimental evidences, have demonstrated to be helpful for investigating how the above-mentioned mutations affect

mechanisms of aggregation, oligomerization pathways and rate of fibril formation, as reported in the following.

Mutations of residues E22 and D23 lead to rare familial variants of AD, as Arctic (E22G), Italian (E22K), Dutch (E22Q), Osaka (E22Δ) and Iowa (D23N). It has been proposed that these mutations reduce the local charges and produce more toxic oligomers[65]. In particular, Italian E22K induces helix formation in the region 20-24 and this alteration leads to faster peptide aggregation[59]. The Osaka mutant lacks the glutamate at position 22 and this deletion leads to rapid fibril formation in solution. Studies on this mutant reveal it is even more stable than the WT and exhibits a rigid structure. The formation of a SB between E3 and K28 may explain the structural stability of mutant than the WT[60–62]. The Iowa mutation changes the backbone structure: a negatively charged aspartic acid residue is replaced with a neutral asparagine residue and the HB between residue 23 and the surrounding residues is disrupted[64]. Region A21-G29 is extremely flexible and the mutation increase the β-strand content of Aβ40 at the CHC and the C-terminal and it suggests important consequences on its dock-and-lock mechanism when it binds fibril edge[63]. D23N fibrils can form both parallel and antiparallel structures[61,70]. Both conformations are highly flexible at the ends and in the loop regions of the protein strands, and much less so in the β-strand region. On the other hand, the C-terminal-to-C-terminal interactions are more stable in the parallel conformation and it may be the responsible for the observed increased stability. Moreover, in the parallel conformation of the WT-dimer, K28 becomes more flexible and the SB with D23 contribute to the stability[62].

Besides, Hsu et al[65] hypothesized that H14 stabilizes the turn motif in region V12-Q15 and its effect also depends to the direction to which the side-chain points.

Three of the six key residues include glycine. Experimental studies demonstrated that glycine stabilizes the packing of β-sheets in the formation of amyloid fibrils[67]. This evidence can explain the importance of G33, G37 and G38 in aggregation process. G33 and G37 residue are also involved in a GxxxG motif, called glycine zipper[68]. This motif includes four glycine residues within the hydrophobic region (G25, G29, G33 and G37) and this motif facilitates the conversion of α-helices and random coils to β-sheet and thus promote the fibril formation[67,69,71]. Investigations of peptides with specific substitutions can provide insight into the mechanisms of aggregation and oligomerization. In $A\beta_{40}$ specie, the substitution of G33 and G37 with leucine destabilized the fibril structure by disrupting the glycine-zipper packing interface[68,69,72]. In vivo and in vitro studies[69,73,74] showed a reduced aggregation propensity in $A\beta_{42}$ by G33 and G37 to alanine, isoleucine or leucine substitutions. A computational analysis[69] illustrated that conformational changes of $A\beta_{42}$ are

induced by increased hydrophobicity. The effect of hydrophobicity on peptide aggregation has been previously demonstrated by Chiti et ai[75]. However, the precise mechanism of destabilization and how G-to-L substitutions interfere with structural stability remain still unclear. Therefore, a detailed investigation of the conformational dynamics of two glycine-to-leucine substitutions was carried out: G33L and G37L. Our MD results suggest that both mutations, in particular G33L mutation, dramatically reduces the stability of $A\beta_{11-42}$.

4.2 Material and Methods

4.2.1 Models

Fibrils are characterized by polymorphism and different structures of the same species have been proposed based on solid-state NMR restraints[25–32]. Two different 9-mer models of S-shaped $A\beta_{11-42}$ specie were studied: one extracted from PDB ID: 2MXU[38] and one from PDB ID: 5KK3[39] for comparing the impact of aminoacidic mutations on different structures. G33L and G37L mutants were obtained by using CHIMERA[76] mutation tool, from the corresponding wild-type specie of 2MXU and 5KK3. All the models here investigated are shown in Figure 15, highlighting also the mutated residues. All the models were solvated in an 8 nm cubic box and neutralized by counterions. Each system consisted of about 50,000 particles.

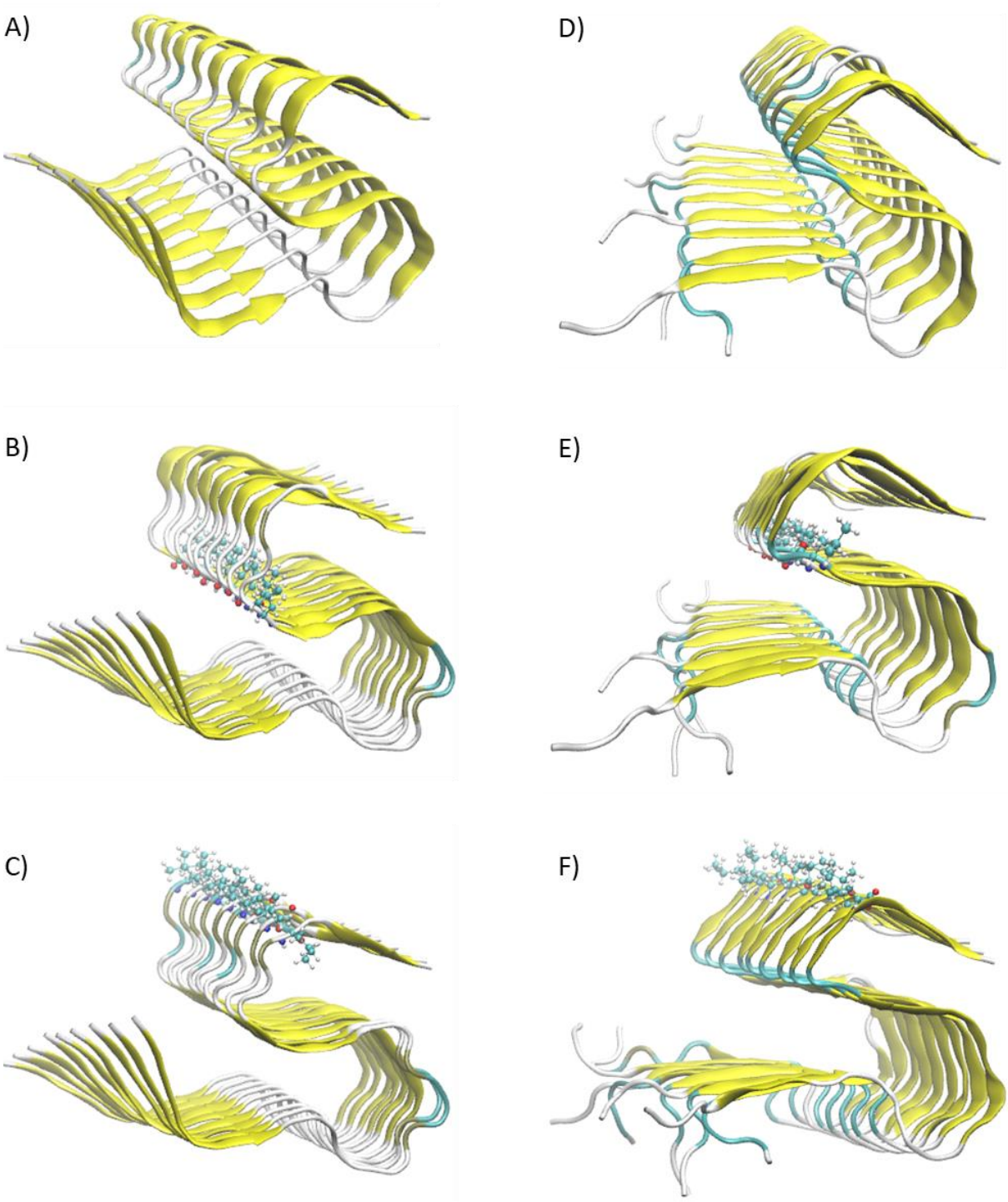

Figure 15: The six different models studied, the Wild-type specie (A), G33L mutant (B) and G37L mutant (C) of the 2MXU[32] model; The Wild-type specie (D), G33L mutant (E) and G37L mutant (F) of the 5KK3[31] model. All the mutants were obtained by using CHIMERA[76].

4.2.2 Simulations Setups

The CHARMM36 force field[40] was used to define protein topologies and the TIP3P [41] model was used for the explicit solvent. GROMACS 5.1.4[42] version was employed for simulations. Steepest descent

algorithm[10] was utilized to minimize the system. Then, a 50 ps simulation in NVT ensemble was conducted, by applying the V-rescale algorithm[43] to keep temperature at 300 K and time constant of 0.1 ps. In order to increase the statistic, five replicas for each $A\beta_{11-42}$ model were generated. We ran a 50 ps simulation in NPT ensemble for each replica. Berendsen[44] and V-rescale[43] coupling methods were used as pressure and temperature coupling. Finally, we performed an MD simulation of 100 ns duration. Nose-Hoover[45] and Parrinello-Rahman[44] were used as temperature and pressure coupling; The LINCS algorithm[46] was used to constrain the length of h-bonds, the PME algorithm[47] to the calculation of electrostatic interactions and Van der Waals were defined within a cut-off of 1.0 nm. Periodic boundary conditions were applied in all three dimensions. Trajectories were extracted every 50 ps of simulation and the Visual Molecular Dynamics (VMD) package was employed to display the simulated systems. The secondary structures were obtained by using STRIDE web server, by pulling out the first and the last frame of the MD trajectory of each model.

4.2.3 Order Parameter and Functional Mode Analisys (FMA)

With the aim of estimating the structural stability of the models and therefore the alignment among protein chains, an order parameter was defined and calculated as below:

$$ordP = \frac{1}{N}\sum_{r=11}^{42}\frac{\langle v_r, x\rangle}{\|v_r\| \cdot \|x\|} = \frac{1}{N}\sum_{r=11}^{42}\cos\alpha \tag{15}$$

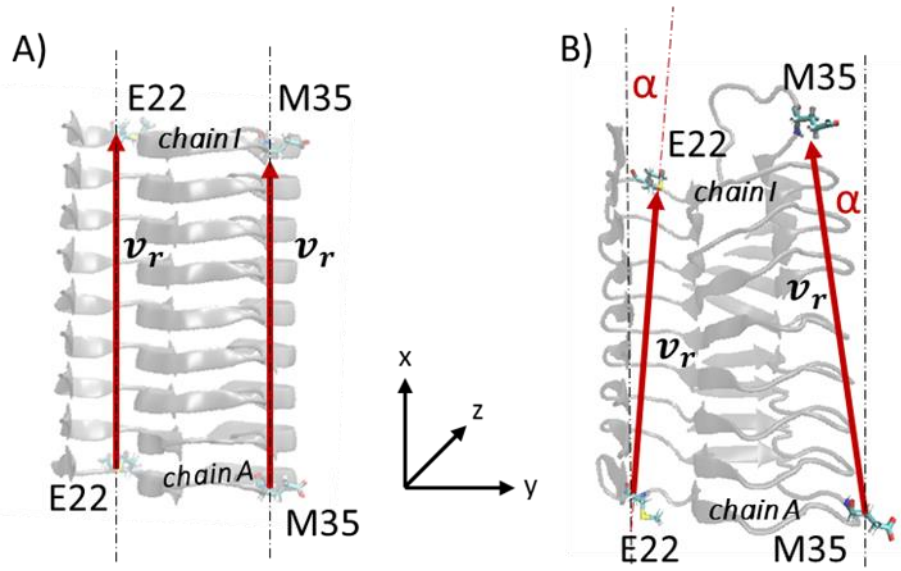

Figure 16: Schematic representation for the calculation of order parameter value. The order parameter is equal to 1 in a completely aligned fibril (A) and lower than 1 in a distorted fiber (B).

Where v_r is the vector joining C_α-atom of each residue *r* of chain A with the corresponding C_α-atom of the same residue of chain I and *x* is the fibril axis. If ordP assumes values close to 1, the chains maintain an alignment to the initial structure; if the values of ordP are lower than 1, the overall structural order decreases. The above mentioned order parameter have been recently used with success in a computational study[48]. The functional mode analysis (FMA) was performed to reduce the functionality of the system. This technique detects a single collective motion that is maximally correlated with a quantity of interest[49], that is the order parameter. Assuming the order parameter is a linear function of the Principal components (PC), the Maximally Correlated vector is obtained by maximizing the Pearson correlation coefficient[49]. In FMA, data-set is divided into a subset of frames for model building and a subset for cross-validation.

4.3 Results

The conformational stability of each molecular system was analyzed by computing the Root Mean Square Deviation (RMSD)(Figure 17) showing that convergence is reached in the last 20 ns for all MD replicas. In particular, RMSD swings around 0,4 and 0,5 nm in all simulations, except for one of the five replicas of G37L mutant in the 2MXU model, where RMSD reaches 0,7 nm.

Each computational data here reported is thus calculated as a time average over the last 20 ns of all trajectories, unless otherwise specified.

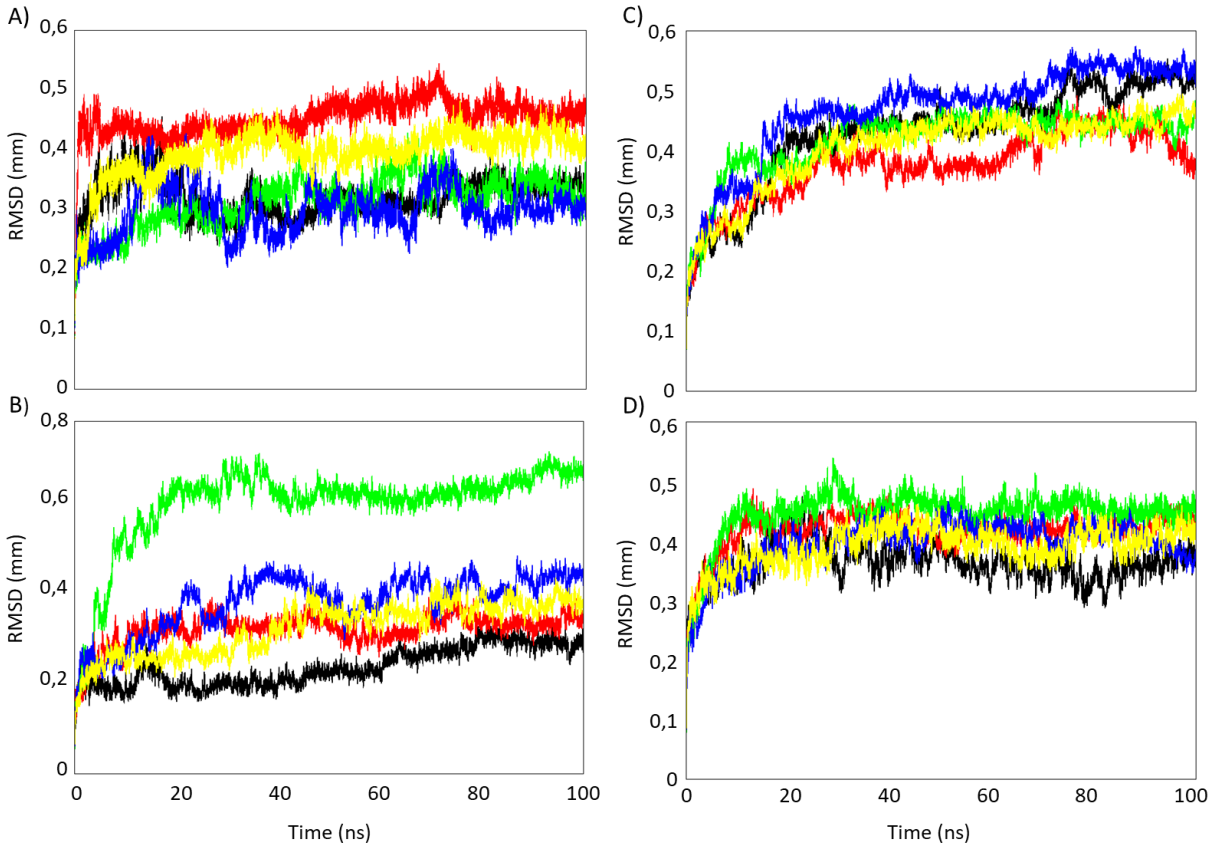

Figure 17: On the left, RMSD calculated on each simulation for G33L (A) and G37L (B) of 2MXU model; On the right, RMSD calculated on each simulation for G33L (C) and G37L (D) of 5KK3 model.

The protein fluctuations were evaluated by Root Mean Square Fluctuation (RMSF) analysis. In the 2MXU model, both mutations lead to an overall increase in fluctuation values (Figure 18), indicating the destabilization of the fibril conformation in presence of protein mutations. In detail, G33L mutation strongly affects the flexibility of the central core and C-terminal region, unlike the G37L substitution, that affects the C-terminal region only. In particular, E22-K28 region of G33L mutant is

much more unstable than the WT specie. It is also worth mentioning the differences located at the mutated sites:

$$\mathrm{RMSF}_{33}[\mathrm{WT}] = \ 0{,}060 \pm 0{,}008\ \mathrm{nm};\ \mathrm{RMSF}_{33}[\mathrm{G33L}] = 0{,}108\ \pm 0{,}020\ \mathrm{nm};$$

$$\mathrm{RMSF}_{37}[\mathrm{WT}] = \ 0{,}075 \pm 0{,}006\ \mathrm{nm};\ \mathrm{RMSF}_{37}[\mathrm{G37L}] = 0{,}145\ \pm 0{,}022\ \mathrm{nm}$$

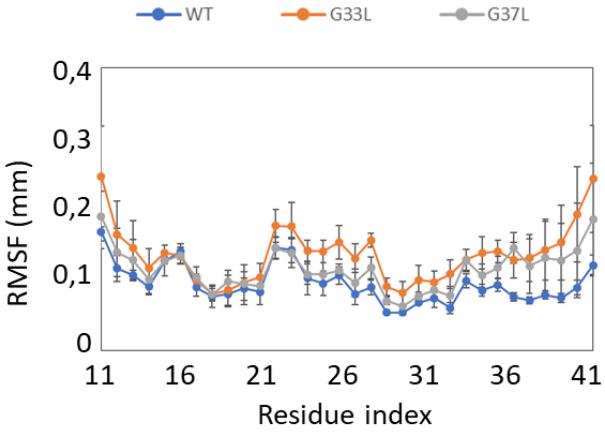

Figure 18: RMSF of WT and the two mutants in the 2MXU model, calculated over the last 20ns of each simulation and averaged on the nine chains.

The effects of G33L and G37L mutations on 5KK3 model are less severe, as shown in Figure 19. The N-terminal region flexibility is almost identical in all the three species. Differences are mainly located along the central core and C-terminal of WT and G33L. In detail, the flexibility of the chain slightly

increase between residues S26-K28 ($\mathrm{RMSF}_{28}[\mathrm{WT}] = 0{,}096 \pm 0{,}015$ nm; $\mathrm{RMSF}_{28}[\mathrm{G33L}] = 0{,}170 \pm 0{,}034$ nm).

On the other hands, G37L does not lead to meaningful variations in RMSF, except for the mutated site ($\mathrm{RMSF}_{37}[\mathrm{G37L}] = 0{,}117 \pm 0{,}017$ nm, $\mathrm{RMSF}_{37}[\mathrm{WT}] = 0{,}073 \pm 0{,}010$ nm).

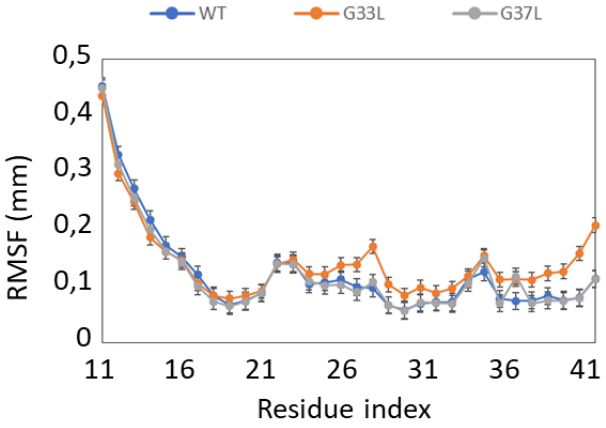

Figure 19: RMSF of WT and the two mutants in the 5KK3 model, calculated over the last 20ns of each simulation and averaged on the nine chains.

As widely demonstrated in scientific literature[50,52], the secondary structure is responsible for the conformational stability of proteins. For this purpose, the secondary structures probability was calculated. Figure 20 shows a marked loss of β-strand content for both G33L (A) and G37L (B), in the 2MXU model. Globally, the β-content dramatically reduces in G33L (from 64% to 39%). The β-strand content loss is a little less severe in G37L but still considerable (from 57% to 45%). In detail, the central core (V24-S26) and C-terminal region (V36-V40) of both mutants are subjected to the biggest reduction of secondary structure. This evidence is in line with changes in fluctuations and it is probably the main factor that affects the stability of such structures.

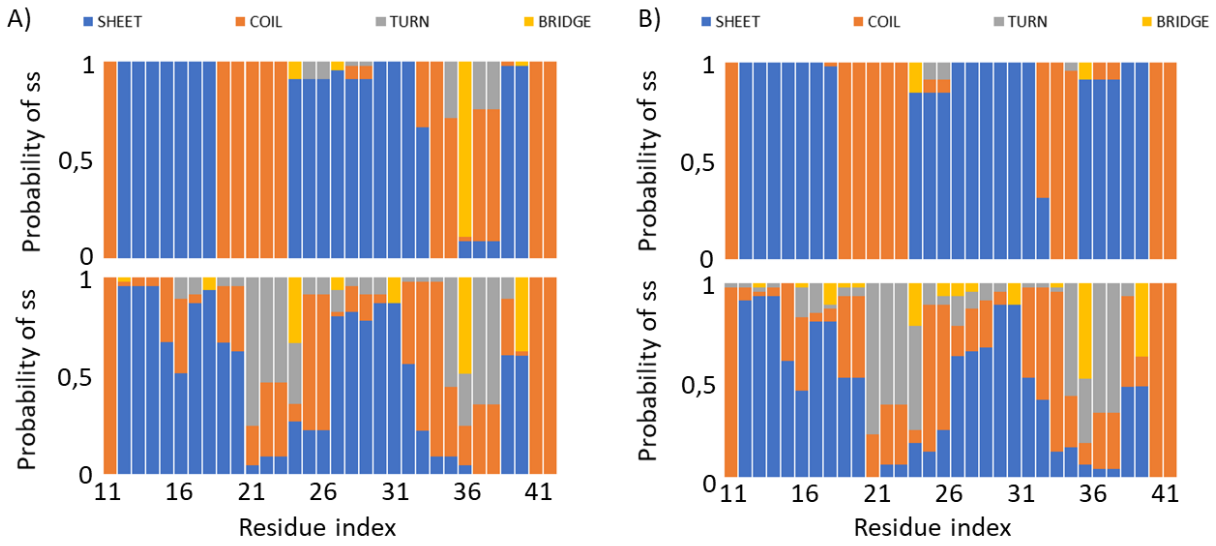

Figure 20: Probability of secondary structure averaged on the chains for the two mutants in the 2MXU model: A) G33L and B) G37L. The first row in each panel shows the initial configuration and the second panel shows the final configuration.

The same trend can be shown in Figure 21, showing the secondary structure probability in the mutant types of 5KK3. Destabilization of the secondary structure is noticeable in the central core of both mutants. In region V24-A30 of G33L mutant the β-strand content deeply decreases from 94% to 37%. The same evidence, but less severe (from 94% to 59%) can be noticed from G37L. In G37L, more remarkable differences are located around the mutated site. In particular, the region comprising residues G38, V39 and V40 almost totally loses its β-strand content.

Interestingly, the G-to-L substitution at position 33 also affects the C-terminal. This result can explain the increased flexibility of that region displayed in Figure 19.

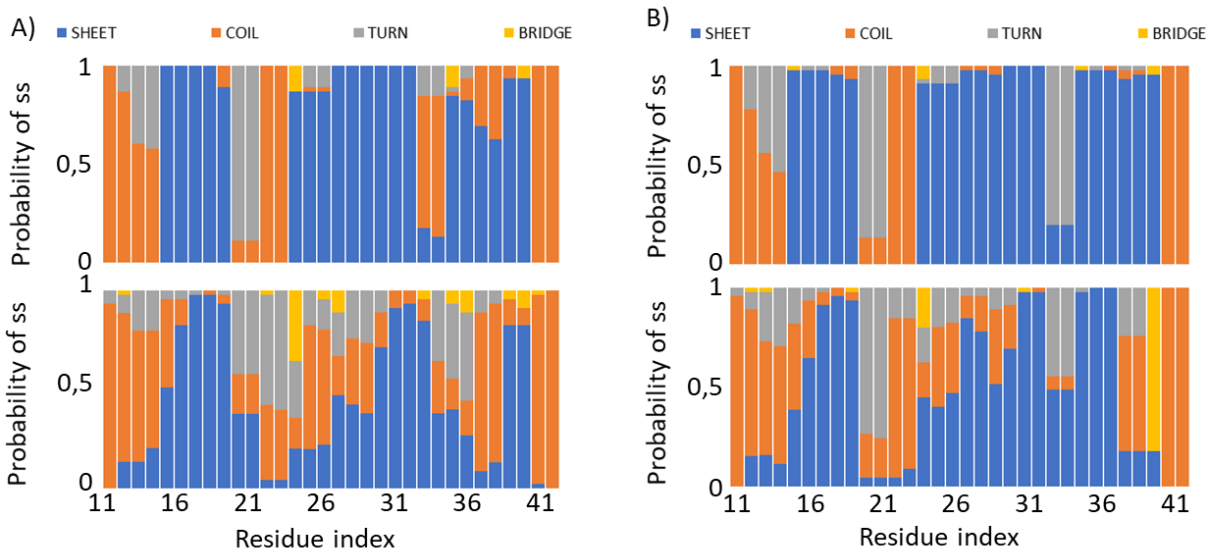

Figure 21: Probability of secondary structure averaged on the chains for the two mutants in the 5KK3 model: A) G33L and B) G37L. The first row in each panel shows the initial configuration and the second panel shows the final configuration.

The total and per residue Solvent Accessible Surface Area (SASA) were calculated as a measure of the tendency of structures to reach a stable and compact arrangement. The SASA value per residue of the 2MXU model is displayed in Figure 22. It shows a global increased tendency for exposing residues to solvent, due to the substitution, as noticeable for residues E22, D23, K28 and in the mutated site of G37L. These results support the changes in RMSF (Figure 18) and secondary structure (Figure 20).

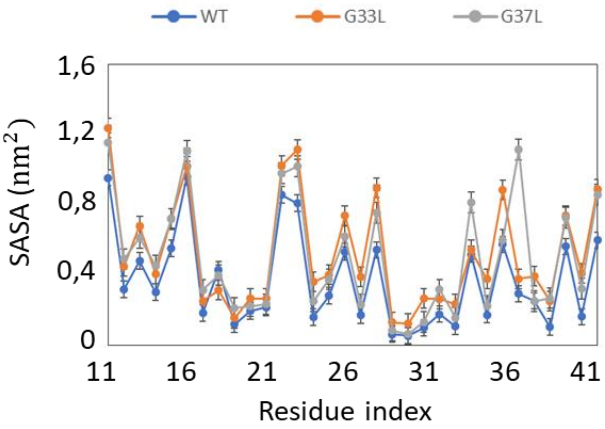

Figure 22: Solvent Accessible Surface Area per residue of wild type and G33L and G37L, in the 2MXU model, calculated as an average over the last 20 ns of each simulation.

The 5KK3 mutant types do not show the same enhanced tendency to expose residues E22 and D23 to solvent. However, also in this case, G33L and G37L mutations affect the solvent exposure profile of the fibril. In detail, differences in residue K28 ($\mathrm{SASA_{WT}} = 0{,}638 \pm 0{,}125\ \mathrm{nm^2}$ and $\mathrm{SASA_{G33L}} = 1{,}091 \pm 0{,}197\ \mathrm{nm^2}$) and G37 ($\mathrm{SASA_{WT}} = 0{,}279 \pm 0{,}057\ \mathrm{nm^2}$ and $\mathrm{SASA_{G37L}} = 0{,}696 \pm 0{,}150\ \mathrm{nm^2}$) are worth to be mentioned (Figure 23).

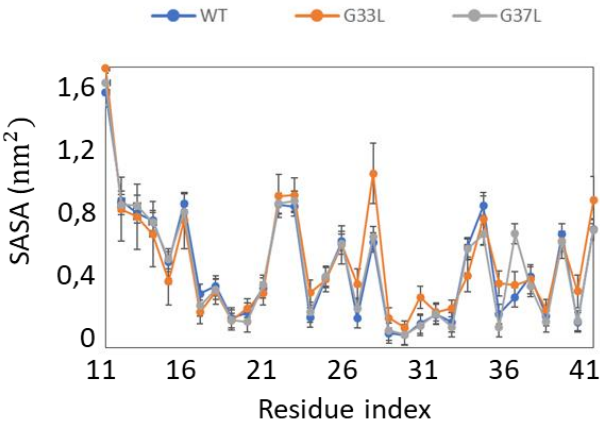

Figure 23: Solvent Accessible Surface Area per residue of wild type and G33L and G37L mutants, in the 5KK3 model, calculated as an average over the last 20 ns of each simulation.

The hydrophobic Solvent Accessible Surface Area (blue column in Figure 24) slightly increases in all the mutated species. This is an expected result, considering that the glycine substitution with leucine increases the hydrophobicity[77]. The mutants also increase the total exposed surface, suggesting less stable and compact arrangements. In a greater detail, the protein Solvent Accessible Surface Area in case of G33L mutation ($\mathrm{SASA_{2MXU}} = 145{,}06 \pm 4{,}23\ \mathrm{nm^2}$ and $\mathrm{SASA_{5KK3}} = 148{,}01 \pm 5{,}75\ \mathrm{nm^2}$) is higher than G37L ($\mathrm{SASA_{2MXU}} = 140{,}55 \pm 1{,}99\ \mathrm{nm^2}$ and $\mathrm{SASA_{5KK3}} = 136{,}32 \pm 4{,}93\ \mathrm{nm^2}$) and WT ($\mathrm{SASA_{2MXU}} = 128{,}35 \pm 3{,}11\ \mathrm{nm^2}$ and $\mathrm{SASA_{5KK3}} = 136{,}22 \pm 4{,}08\ \mathrm{nm^2}$) in both 2MXU and 5KK3 molecular system.

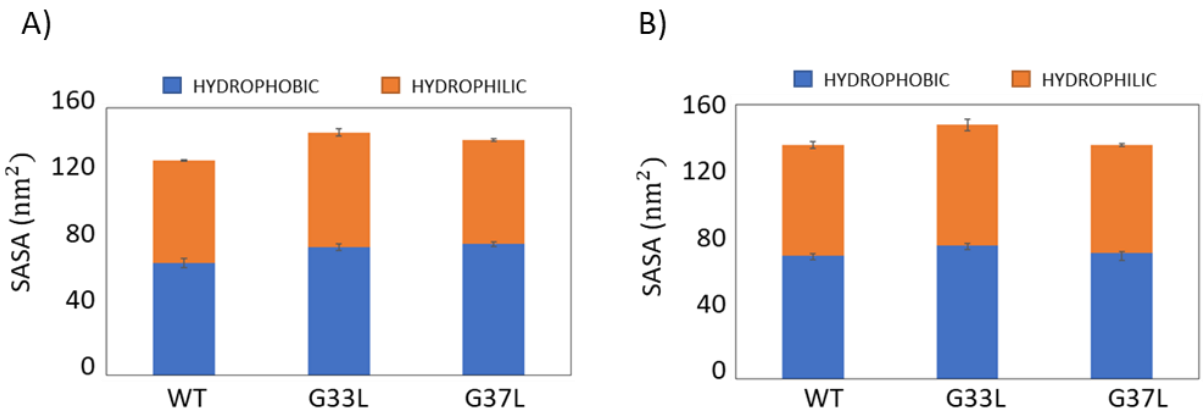

Figure 24: Solvent Accessible Surface Area for WT, G33L and G37L for 2MXU model (A) and 5KK3 (B), calculated as an average over the last 20 ns of each simulation.

In order to highlight the low frequency mode of motion corresponding to the highest amplitude, Principal Component Analysis was performed. The amplitude of eigenvalues corresponds to the variance of the protein motion, usually quantified as mean square fluctuations, along their relative eigenvectors yields by PCA. By observing Figure 25, it is clear the presence of a prevalent PC mode in G33L mutants of the two models (the amplitude of the first eigenvalue is 12,42 nm^2 in the 2MXU model, 10,12 nm^2 in the 5KK3 model). The same can be said for G37L mutant of 2MXU model. Here, the amplitude of the first eigenvalue is 4,38 nm^2, still much higher when compared to WT (2,28 nm^2). The previously mentioned results demonstrate how mutations strongly affects the conformational dynamics of investigated models when compared to WT.

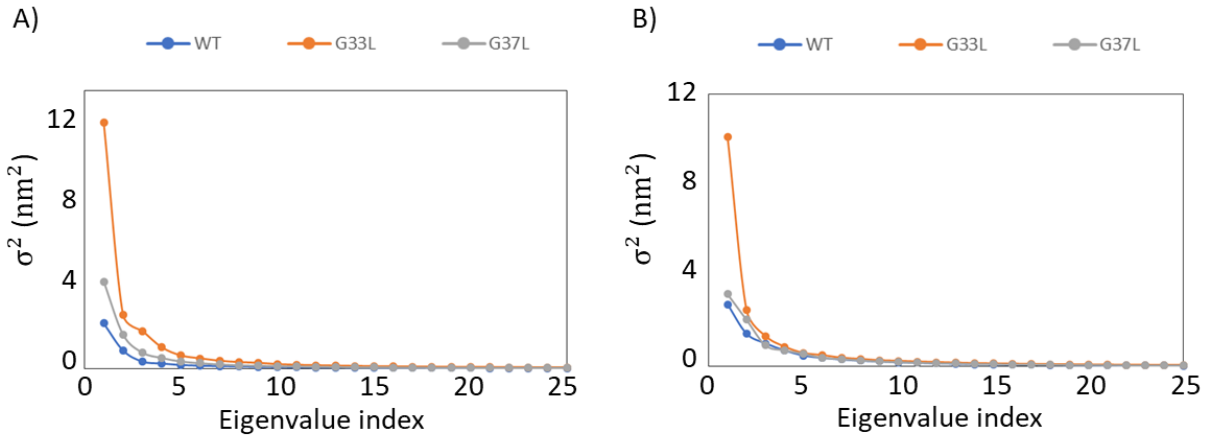

Figure 25: Eigenvalues and relative amplitudes.

With the aim of analyzing the structural stability of the proteins, the order parameter defined in Eq. 17 was calculated. The order parameter (ordP) gives a measure of the alignment among protein chains. If ordP assumes values close to 1, the chains maintain an alignment to the initial structure. On the contrary, if the values of ordP are lower than 1, the overall structural order decreases. The value of the order parameter (Figure 26) emphasizes the effects of G-to-L substitutions on structural stability. It is evident that G33L ($\mathrm{ordP}_{2\mathrm{MXU}} = 0{,}902$, $\sigma_{2\mathrm{MXU}} = 2{,}82\mathrm{e}^{-2}$ and $\mathrm{ordP}_{5\mathrm{KK3}} = 0{,}928$, $\sigma_{5\mathrm{KK3}} = 1{,}70\mathrm{e}^{-2}$) strongly destabilizes the ordered structure of WTs ($\mathrm{ordP}_{2\mathrm{MXU}} = 0{,}984$, $\sigma_{2\mathrm{MXU}} = 1{,}57\mathrm{e}^{-3}$ and $\mathrm{ordP}_{5\mathrm{KK3}} = 0{,}989$, $\sigma_{5\mathrm{KK3}} = 2{,}62\mathrm{e}^{-3}$). The impact of G37L is also relevant in the 2MXU model ($\mathrm{ordP}_{2\mathrm{MXU}} = 0{,}961$, $\sigma_{2\mathrm{MXU}} = 1{,}07\mathrm{e}^{-2}$), unlike the 5KK3 one ($\mathrm{ordP}_{5\mathrm{KK3}} = 0{,}981$, $\sigma_{5\mathrm{KK3}} = 2{,}82\mathrm{e}^{-2}$).

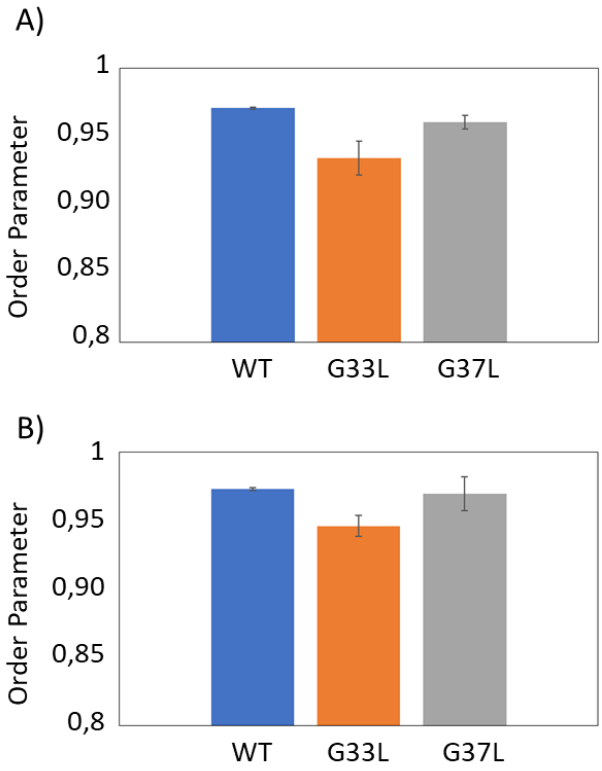

Figure 26: Expected value and corresponding standard deviation of the order parameter, calculated over the last 20 ns of trajectories. The panel A) shows the three species of 2MXU model, the panel B) shows the three species of 5KK3.

With the aim of highlighting the per-residue alignment along the fibril structure, the order parameter is computed as a function of the residue index (Figure 27). Overall, N-terminal is the most disordered region in all the six structures evaluated. Besides, the central core including D23-S26 of both G33L mutant types are subjected to a notable distortion. In particular, the order parameter

swings around 0,90-0,91 in the 2MXU model and 0,94-0,95 in the 5KK3 model (Figure 27). The same evidence is noticed in the C-terminal regions ($\mathrm{ordP}_{\mathrm{2MXU}} \simeq 0{,}8,; \mathrm{ordP}_{\mathrm{5KK3}} \simeq 0{,}85$). On the other hands, the effect of G37L substitution is less remarkable. In fact, the order parameter along the chain is almost identical when compared to WT and close to 1, especially in 5KK3 model. All these results are consistent with fluctuations (Figure 18 and Figure 19) and changes in the secondary structure (Figure 20 and Figure 21), confirming that the most flexible regions assumes a more disordered arrangement.

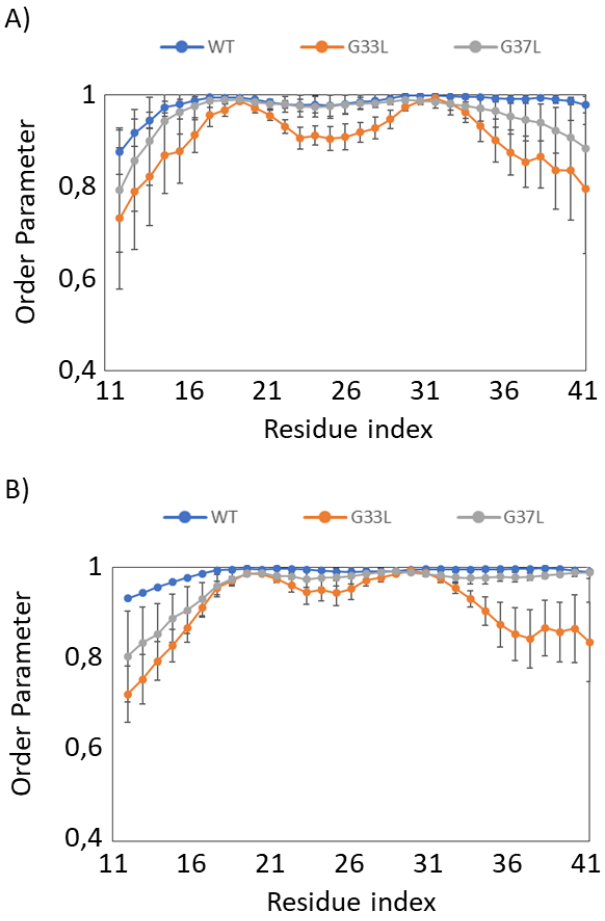

Figure 27: Order parameter as a function of the residue index in the 2MXU model (A) and 5KK3 model (B).

Lastly, a Functional Mode Analysis (FMA) on the order parameter as a function of the time was performed, in order to detect a single collective mode maximally correlated to fluctuations in such parameter. The models were validated by predicting the quantity of interest with a Pearson correlation coefficient higher than 0,80. The initial and final frames extracted from trajectories along ewMCM vectors clearly shows all the investigated structures are subjected to a torsional mode, in

particular the mutant types of 2MXU (Figure 28) and 5KK3 models. G33L mutation in 5KK3 model (Figure 29, panel A) and B)) is characterized by a tendency to open from the lateral chain.

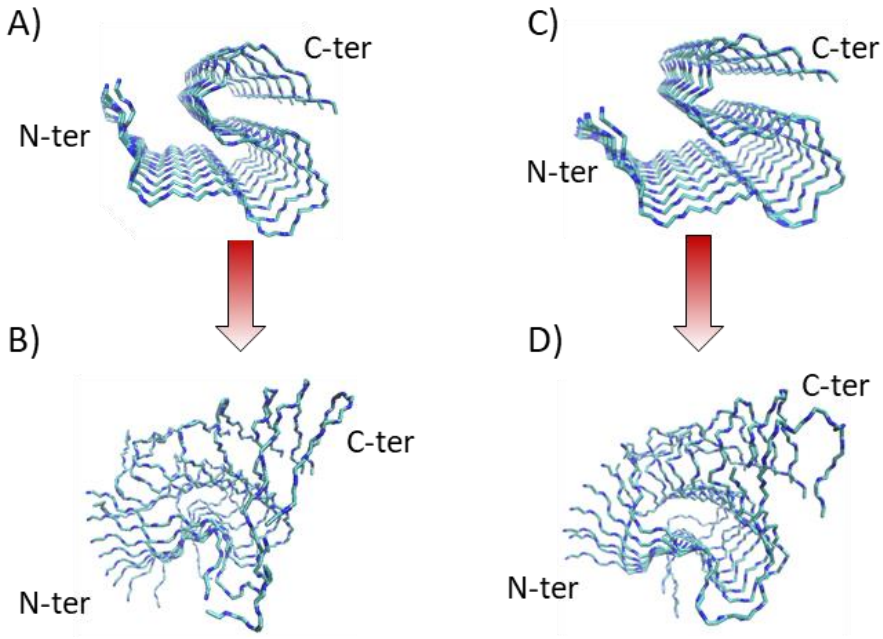

Figure 28: A) B) and C) D) show the conformational changes of G33L (left) and G37L (right) in the 2MXU model. Snapshots are obtained from ewMCM trajectories.

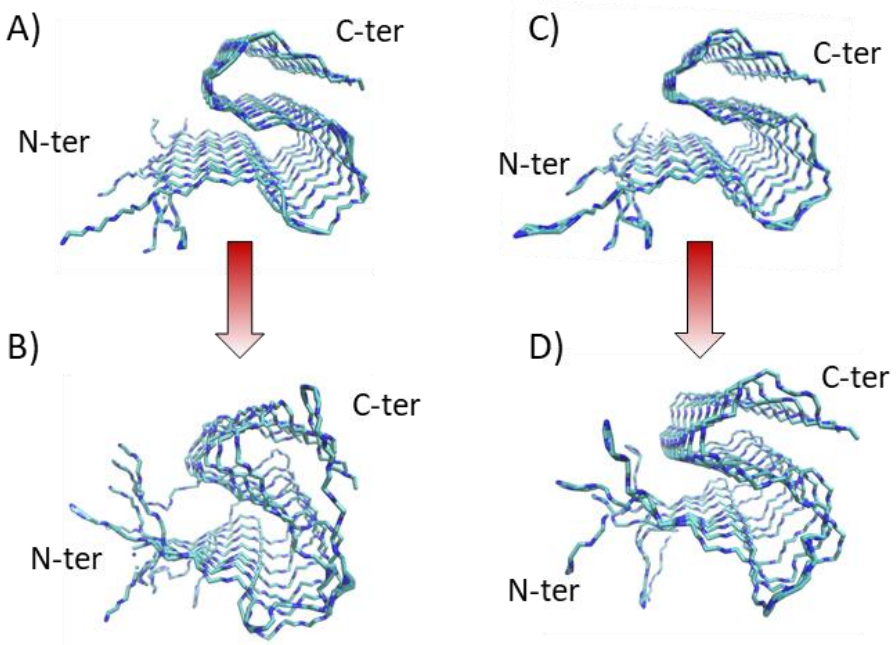

Figure 29: A) B) and C) D) show the conformational changes of G33L (left) and G37L (right) in the 5KK3 model. Snapshots are obtained from ewMCM trajectories.

4.4 Discussion

Protein aggregation is implicated into a broad range of human illnesses. Alzheimer's Disease, for example, is characterized by aggregates of two types amyloid-beta peptides in the brain of patients: $A\beta_{40}$ and $A\beta_{42}$. The precise mechanisms of aggregation remain unclear, but six residues were identified to be essential for the formation of such aggregates: H14, E22, D23, G33, G37 and G38[65]. Two of the six residue proposed, G33 and G37, are involved in a GxxxG motif called glycine zipper[68]. Because it has been shown that the GxxxG motif has a key role in the fibril formation[67,69,71], investigations of peptides with specific aminoacidic substitutions may help to clarify mechanisms behind aggregation and oligomerization of Aβ. For this purpose, two glycine-to-leucine substitutions on two different models of $A\beta_{42}$ (2MXU and 5KK3) have been investigated: G33L and G37L. Leucine and not alanine was used to replace glycine, because AxxxG motif and GxxxA are implicated in protein-protein interactions[78], as previously proposed by Hung et al.[73] In this scenario, this work is aimed to be complementary to the evidences proposed by Harmeier et ai.[69], Hung et al.[73] and by Fonte et ai.[74].

The mutant types were seen to undergo remarkable changes in the conformational stability. These evidences substantially confirm the key role of the glycine zipper in the mechanism of fibril formation. The overall RMSF in the 2MXU model (Figure 18) is sharply higher than that of the WT. Unlike the G33L, that also causes an enhanced flexibility of the central hydrophobic core, the G37L only affects the C-terminal region. It is interesting to note that changes in RMSF are in line with changes in the secondary structure (Figure 20 and Figure 21), since a marked loss of β-strand content is clearly visible in the most flexible regions. This evidence is in accordance with results shown by Perczel et al., which demonstrated that β-strands stabilize and give strength to peptides[50].

It is worth highlighting that the Solvent Accessible Surface Area per Residue plot, displayed in Figure 22 and Figure 23, shows a notable enhanced tendency of exposing residue K28 in both 2MXU and 5KK3 models (in the latter, this tendency is mainly noticeable in G33L mutant type). As reported from solid-state NMR, residue K28 of S-shaped structure of $A\beta_{42}$ is involved in an intramolecular salt bridge (SB) with residue A42 that stabilizes the triple β-motif[79]. This finding suggests that the increased solvent exposure of residue K28 can lead to the disruption of K28-A42 and destabilize the structure of the folded protein[80].

Moreover, changes in the exposed hydrophobic surface (Figure 24) are substantially supported by literature because, according to the scale of Kyte and Doolittle[77], the substitution of a glycine to a leucine increases the hydrophobicity by 0,15. The total SASA represents another significant

contribution to the reduced stability of mutant types[51]. The value of SASA is related to the tendency of a protein to reach a stable and compact arrangements and, in our cases decrease as follow, for both model: G33L>G37L>WT.

This hypothesis is confirmed by the distribution of probability of the order parameter (Figure 26), that shows the same tendency. To be specific, G33L causes a distortion in the N- and C-terminal and in the central hydrophobic core. On the other hands, G37L affects the N- and C- terminal region of the 2MXU model, mostly. Unlike G33L, the effect of such mutation is less notable in the central core. It is interesting to note that the misalignment among residues of chain A and the corresponding residues of chain I is in line with the tendency of residues for exposing themselves to the solvent and with changes in the secondary structure.

The above-mentioned instability of the mutant types with respect to the WTs is also confirmed by the Functional Mode Analysis performed on the order parameter. Specifically, the MD trajectories filtered on ewMCM clearly highlight the presence of a torsional mode in both G33L mutants and in the G37L mutant of the 2MXU model. Contrarily, G37L tends to maintain a more compact arrangement, even though a torsional mode is still present.

To summarize, our data confirm the importance of residues 33 and 37 for formation of $A\beta_{42}$ fibrils, since the substitutions here investigated notably affect the conformational dynamics and structural stability of the peptides. In particular, G33 of the central glycine zipper has been proposed to be essential for assuring stability and promoting the process of fibril formation. It has been proposed[69,73] that the aggregation of $A\beta_{42}$ is alleviated by enhancing the hydrophobicity at position G33 with G-to-A or G-to-I substitution. We also find that G37L mutant type of the 2MXU model appears less stable. These evidences, combined with the studies early mentioned[69,73,74], confirm that the disruption of the glycine zipper can reduce the aggregation propensity of $A\beta_{42}$.

Therefore, additional computational studies might yield a more detailed insight into the effects of mutations in such region.

On the basis of our data, further investigations might consider this key region as a binding site for inhibitors drugs aimed at reducing the stability of the Aβ fibrils.

4.5 Conclusions

In this chapter, two different glycine-to-leucine substitutions at position 33 and 37 were analyzed and compared in two models of $A\beta_{11-42}$. Molecular Dynamics simulation in explicit solvent were performed to study their dynamic behavior in physiological environment and how mutations affect

the aggregation tendency of $A\beta_{42}$. MD clearly showed that the disruption of the glycine zipper, induced by specific aminoacidic substitutions, causes conformational changes and leads to an enhanced instability of the structure. The results are line with earlier experimental and computational studies[69,73,74], indicating that binding glycine zipper might be helpful for developing new Aβ inhibitors.

Acknowledgments

I would first like to thank my supervisor Prof. Marco Agostino Deriu for accepting me on this fascinating research topic and giving me all the helps I needed. I am also very thankful to my supervisor for continuous motivation during my thesis work and for introducing me in the Computational Biophysics Group of the Dalle Molle Institute for Artificial Intelligence.

I would like to show my gratitude to my Co-Supervisors Prof. Umberto Morbiducci and Prof. Jacek Adam Tuszynski from Politecnico of Turin for the evaluation and critical review of my work.

I would also like to acknowledge Gianvito Grasso, I am gratefully indebted to him for his very valuable support, patience and encouragement in the whole period of this work.

Sincere acknowledgements to Prof. Andrea Danani of Computational Biophysics Group of the Dalle Molle Institute for Artificial Intelligence for accepting me in his research group.

Finally, I must express my gratitude to my family for having always believed in me and for providing me with unfailing support throughout my years of study and lifetime.

Luigi Leanza

References

1. Thirumalai, D., Klimov, D. K. & Dima, R. I. Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr. Opin. Struct. Biol.* **13,** 146–59 (2003).

2. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81,** 678–699 (2003).

3. Leach, A. R. *Molecular modelling : principles and applications*. (Prentice Hall, 2001).

4. Combes, J. M., Duclos, P. & Seiler, R. The Born-Oppenheimer Approximation. in *Rigorous Atomic and Molecular Physics* 185–213 (Springer US, 1981). doi:10.1007/978-1-4613-3350-0_5

5. Vanommeslaeghe, K., Guvench, O. & MacKerell, A. D. Molecular mechanics. *Curr. Pharm. Des.* **20,** 3281–92 (2014).

6. Johnson, J. K., Zollweg, J. A. & Gubbins, K. E. The Lennard-Jones equation of state revisited. *Mol. Phys.* **78,** 591–618 (1993).

7. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

8. Ding, H., Karasawa, N. & Goddard, W. A. Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97,** 4309–4315 (1992).

9. Tironi, I. G., Sperb, R., Smith, P. E. & van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102,** 5451–5459 (1995).

10. Fletcher, R. & Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Comput. J.* **6,** 163–168 (1963).

11. Conjugate Gradient Methods. in *Numerical Optimization* 101–134 (Springer New York, 2006). doi:10.1007/978-0-387-40065-5_5

12. Ben-Israel, A. A Newton-Raphson method for the solution of systems of equations. *J. Math. Anal. Appl.* **15,** 243–252 (1966).

13. Head, J. D. & Zerner, M. C. A Broyden—Fletcher—Goldfarb—Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* **122,** 264–270 (1985).

14. Zhang, J. *Molecular dynamics analyses of prion protein structures : the resistance to prion diseases down under*.

15. Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Mol. Simul.* **6,** 121–142 (1991).

16. Van Gunsteren, W. F. & Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1,** 173–185 (1988).

17. Martys, N. S. & Mountain, R. D. Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Phys. Rev. E* **59,** 3733–3736 (1999).

18. Jolliffe, I. Principal Component Analysis. in *International Encyclopedia of Statistical Science* 1094–1096 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-04898-2_455

19. Hub, J. S. & de Groot, B. L. Detection of Functional Modes in Protein Dynamics. *PLoS Comput. Biol.* **5,** e1000480 (2009).

20. Altis, A., Nguyen, P. H., Hegger, R. & Stock, G. Dihedral angle principal component analysis of

molecular dynamics simulations. *J. Chem. Phys.* **126,** 244111 (2007).

21. Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharm. Res.* **20,** 1325–1336 (2003).

22. Cummings, J. L. Alzheimer's Disease. *N. Engl. J. Med.* **351,** 56–67 (2004).

23. Lee, H. J. *et al.* Structural and Mechanistic Insights into Development of Chemical Tools to Control Individual and Inter-Related Pathological Features in Alzheimer's Disease. *Chemistry* **23,** 2706–2715 (2017).

24. Querfurth, H. W. & LaFerla, F. M. Alzheimer's Disease. *N. Engl. J. Med.* **362,** 329–344 (2010).

25. Schütz, A. K. *et al.* Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew. Chem. Int. Ed. Engl.* **54,** 331–5 (2015).

26. Lu, J.-X. *et al.* Molecular structure of β-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **154,** 1257–68 (2013).

27. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 18349–54 (2008).

28. Qiang, W., Yau, W.-M., Luo, Y., Mattson, M. P. & Tycko, R. Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4443–8 (2012).

29. Lührs, T. *et al.* 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17342–7 (2005).

30. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E4976-84 (2016).

31. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.* **138,** 9663–74 (2016).

32. Xiao, Y. *et al.* Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **22,** 499–505 (2015).

33. Fändrich, M., Meinhardt, J. & Grigorieff, N. Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion* **3,** 89–93

34. Xi, W., Wang, W., Abbott, G. & Hansmann, U. H. E. Stability of a Recently Found Triple-β-Stranded Aβ1–42 Fibril Motif. *J. Phys. Chem. B* **120,** 4548–4557 (2016).

35. Petkova, A. T. *et al.* A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16742–7 (2002).

36. Grasso, G. *et al.* Conformational Dynamics and Stability of U-Shaped and S-Shaped Amyloid β Assemblies. *Int. J. Mol. Sci.* **19,** (2018).

37. Flöck, D., Rossetti, G., Daidone, I., Amadei, A. & Nola, A. Di. Aggregation of small peptides studied by molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* **65,** 914–921 (2006).

38. Xiao, Y. *et al.* Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **22,** 499–505 (2015).

39. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic $A\beta_{42}$ Amyloid Fibrils. *J. Am. Chem. Soc.* **138,** 9663–9674 (2016).

40. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered

proteins. *Nat. Methods* **14,** 71–73 (2017).

41. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

42. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).

43. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).

44. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

45. Evans, D. J. & Holian, B. L. The Nose–Hoover thermostat. *J. Chem. Phys.* **83,** 4069–4074 (1985).

46. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

47. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

48. Grasso, G. *et al.* Conformational Dynamics and Stability of U-Shaped and S-Shaped Amyloid β Assemblies. *Int. J. Mol. Sci.* **19,** 571 (2018).

49. Hub, J. S. & de Groot, B. L. Detection of functional modes in protein dynamics. *PLoS Comput. Biol.* **5,** e1000480 (2009).

50. Perczel, A., Gáspári, Z. & Csizmadia, I. G. Structure and stability of β-pleated sheets. *J. Comput. Chem.* **26,** 1155–1168 (2005).

51. Chien, Y.-T., Hwang, J.-K. & Huang, S.-W. On the Relationship Between Residue Solvent Exposure and Thermal Fluctuations in Proteins. in *Protein Structure* (InTech, 2012). doi:10.5772/37148

52. Kung, V. M., Cornilescu, G. & Gellman, S. H. Impact of Strand Number on Parallel β-Sheet Stability. *Angew. Chem. Int. Ed. Engl.* **54,** 14336–9 (2015).

53. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A. Math. Phys. Eng. Sci.* **374,** 20150202 (2016).

54. Huet, A. & Derreumaux, P. Impact of the Mutation A21G (Flemish Variant) on Alzheimer's β-Amyloid Dimers by Molecular Dynamics Simulations. *Biophys. J.* **91,** 3829–3840 (2006).

55. Bartus, R. T. On Neurodegenerative Diseases, Models, and Treatment Strategies: Lessons Learned and Lessons Forgotten a Generation Following the Cholinergic Hypothesis. *Exp. Neurol.* **163,** 495–529 (2000).

56. Hardy, J. & Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (80-.).* **297,** 353–356 (2002).

57. Gravina, S. A. *et al.* Amyloid beta protein (A beta) in Alzheimer's disease brain. Biochemical and immunocytochemical analysis with antibodies specific for forms ending at A beta 40 or A beta 42(43). *J. Biol. Chem.* **270,** 7013–6 (1995).

58. Roher, A. E. *et al.* beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10836–40 (1993).

59. Lin, Y.-S., Bowman, G. R., Beauchamp, K. A. & Pande, V. S. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **102,**

315–324 (2012).

60. Coskuner, O., Wise-Scira, O., Perry, G. & Kitahara, T. The Structures of the E22Δ Mutant-Type Amyloid-β Alloforms and the Impact of E22Δ Mutation on the Structures of the Wild-Type Amyloid-β Alloforms. *ACS Chem. Neurosci.* **4,** 310–320 (2013).

61. Berhanu, W. M., Alred, E. J. & Hansmann, U. H. E. Stability of Osaka Mutant and Wild-Type Fibril Models. *J. Phys. Chem. B* **119,** 13063–13070 (2015).

62. Kassler, K., Horn, A. H. C. & Sticht, H. Effect of pathogenic mutations on the structure and dynamics of Alzheimer's Aβ42-amyloid oligomers. *J. Mol. Model.* **16,** 1011–1020 (2010).

63. Côté, S., Derreumaux, P. & Mousseau, N. Distinct Morphologies for Amyloid Beta Protein Monomer: $A\beta_{1-40}$, $A\beta_{1-42}$, and $A\beta_{1-40}$ (D23N). *J. Chem. Theory Comput.* **7,** 2584–2592 (2011).

64. Murray, M. M. *et al.* Amyloid β-Protein: Experiment and Theory on the 21−30 Fragment. *J. Phys. Chem. B* **113,** 6041–6046 (2009).

65. Hsu, F., Park, G. & Guo, Z. Key Residues for the Formation of Aβ42 Amyloid Fibrils. *ACS omega* **3,** 8401–8407 (2018).

66. Qiang, W., Yau, W.-M., Luo, Y., Mattson, M. P. & Tycko, R. Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4443–8 (2012).

67. Liu, W. *et al.* Structural Role of Glycine in Amyloid Fibrils Formed from Transmembrane α-Helices †. *Biochemistry* **44,** 3591–3597 (2005).

68. Kim, S. *et al.* Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14278–83 (2005).

69. Harmeier, A. *et al.* Role of amyloid-beta glycine 33 in oligomerization, toxicity, and neuronal plasticity. *J. Neurosci.* **29,** 7582–90 (2009).

70. Qiang, W., Yau, W.-M., Luo, Y., Mattson, M. P. & Tycko, R. Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4443–8 (2012).

71. Munter, L.-M. *et al.* GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of Abeta42. *EMBO J.* **26,** 1702–12 (2007).

72. Sato, T. *et al.* Inhibitors of Amyloid Toxicity Based on β-sheet Packing of Aβ40 and Aβ42 †. *Biochemistry* **45,** 5503–5516 (2006).

73. Hung, L. W. *et al.* Amyloid-beta peptide (Abeta) neurotoxicity is modulated by the rate of peptide aggregation: Abeta dimers and trimers correlate with neurotoxicity. *J. Neurosci.* **28,** 11950–8 (2008).

74. Fonte, V. *et al.* A glycine zipper motif mediates the formation of toxic β-amyloid oligomers in vitro and in vivo. *Mol. Neurodegener.* **6,** 61 (2011).

75. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide andprotein aggregation rates. *Nature* **424,** 805–808 (2003).

76. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

77. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

78. Russ, W. P. & Engelman, D. M. The GxxxG motif: A framework for transmembrane helix-helix association. *J. Mol. Biol.* **296,** 911–919 (2000).

79. Xiao, Y. *et al.* Aβ(1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **22,** 499–505 (2015).

80. Bosshard, H. R., Marti, D. N. & Jelesarov, I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **17,** 1–16 (2004).