Università della Svizzera italiana

Institute of Computational Science ICS

USI/SUPSI IDSIA

Modelling Protein-Protein Interactions to Elucidate Molecular Mechanisms Behind Neurodegenerative Diseases

Doctoral Dissertation submitted to the
Faculty of Informatics of the Università della Svizzera Italiana
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

presented by

Gianvito Grasso

Under the supervision of
Prof. Rolf Krause (Advisor)
Prof. Andrea Danani (Co-advisor)

October 2018

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

Gianvito Grasso

Lugano, October 2018

Dissertation Committee

Prof. Michele Parrinello

Prof. Luca Gambardella

Prof. Jack Tuszynski

Prof. Umberto Morbiducci

Abstract

The worldwide significant increase in life expectancy has recently drawn the attention of the scientific community to neurodegenerative pathologies of the elderly population. These neurodegenerative disorders arise from the abnormal protein aggregation in the nervous tissue leading to intracellular inclusions or extracellular aggregates in specific brain areas. A feasible strategy to prevent the resulting neurodegeneration is based on the development of anti-amyloid molecules, i.e., those capable of preventing the generation of toxic aggregates. To address this issue, it's extremely important to shed light on the molecular interactions responsible for protein aggregation. Despite substantial research efforts in this field, the fundamental mechanisms of protein misfolding and aggregation mechanisms remain somewhat unrevealed. In this context, computational molecular modelling represents a powerful tool in connecting macroscopic experimental findings to nanoscale molecular events.

The present PhD thesis focuses on the application of computational methodologies to investigate molecular features of protein-protein interactions responsible for two different pathologies: Spinocerebellar Ataxia Type-1 (SCA1) and Alzheimer's Disease (AD). To address this goal, molecular dynamics simulations have been employed to elucidate the early stages of protein aggregation mechanism at molecular level. From the computational point of view, insufficient sampling often limits the ability of computer simulations to investigate the conformational properties of biomacromolecules. The limitation mainly results from proteins' rough energy landscapes, with many local minima separated by high-energy barriers. Within this framework, one of the main challenges of MD simulations is the ability to sample experimentally relevant millisecond to second timescales. However, the time-scale of the classical MD simulations with atomic resolution is today limited to few μs. In this regard, enhanced sampling methods represent a powerful tool to improve the sampling efficiency of classical MD, including those that artificially add an external driving force to guide the protein from one structure to another. The present PhD work benefits from the application of enhanced sampling techniques and dimensionality reduction methodologies to elucidate the aggregation pathway of the Ataxin-1 and Amyloid Beta assembly, responsible for SCA1 and AD, respectively. Outcome of the present research represents an important piece of knowledge to design small molecules able to inhibit the protein-protein interactions leading to aggregation. On the other hand, fine tuning of the interatomic forces responsible for the intriguing mechanical properties of the amyloid fibrils is a crucial breakthrough to support the rational design of amyloid-inspired nanostructures as novel biomaterials.

Contents

Chapter 1

Computational Modelling of Biomolecular Systems

Molecular modelling provides theoretical methods and computational techniques used to describe complex chemical systems (e.g., molecules, protein, nucleic acids) in terms of a realistic atomistic description, aimed at understanding and predicting macroscopic properties of these systems. Molecular systems generally consist of a large number of molecules and for this reason it is difficult, if not impossible, to quantify the properties of such systems analytically. The previously mentioned analytical intractability is solved by using numerical methods. Molecular Dynamics (MD) is a powerful tool for understanding molecular processes because it represents a connection point between laboratory experiments and theory. MD method is based on physics, chemistry, information theory, statistical and molecular mechanics, but the simplest form of MD involves little more than Newton's second law. It is applied today mostly in materials science and modelling of biomolecules. The following list includes some examples in which MD is used: structure and dynamics of proteins, protein folding/unfolding, multiscale modelling, docking of molecules, analysis of polymers chains, transport and diffusion properties, and there is much more.

The present chapter provides the theoretical background for the present PhD thesis work, with the aim of explaining the physical basis behind the computational approach. More in depth, after a quick introduction to statistical mechanics (Section 1.1), the Sections 1.2 and 1.3 are devoted to the force field description and MD algorithm. Finally, the dimensionality reduction methodologies and enhanced sampling methods, able to overcome the limitation of the Classical MD, are presented in the Section 1.4 and 1.5.

1.1 Statistical Mechanics

There are two categories of macroscopic properties in a complex chemical system: static equilibrium properties (e.g., temperature, pressure, density, average potential energy, etc.) and dynamic or non-equilibrium properties (e.g., the dynamics of phase changes, diffusion processes). To compute macroscopic properties, it is necessary to generate a representative *statistical ensemble* at a given temperature, which defines all the accessible physical states of a molecular system. The space in which all possible physical states of a system are represented is known as *phase space* which often consists of all possible values of position and momentum variables. In general, every degree of freedom of the system is represented as an axis of a multidimensional space. In a system containing N atoms, 6N values are required to define the state of the system (three coordinates of position and three components of momentum) and each possible state corresponds to a specific point in the phase space. A single point in phase space describes the state of the system and the succession of this plotted points represents all the accessible system's microstates. There are different points in the phase space characterized by the same thermodynamic state. The collection of all possible system configurations which have different microscopic states, but an identical thermodynamic state is known as *statistical ensemble*.

Ensembles	Fixed variables
Thermally isolated equilibrium	N, V, E
Thermal equilibrium with a heat reservoir	N,V,T
Isobaric-isothermal	N, p, T
Grand canonical or Gibbs	μ, V, T

Table 1. The thermally isolated equilibrium ensemble, also known as microcanonical or NVE ensemble, number of particles N, volume V and total energy E are constant. The statistic ensemble in thermal equilibrium with a heat reservoir, also known as canonical or NVT ensemble, is characterized by fixed number of particles N, volume V and temperature T. The Isobaric-isothermal ensemble is characterized by fixed number of particles N, pressure p and temperature T. In the Grand canonical ensemble, there are variable number of particles, but fixed volume, temperature and chemical potential.

The ensemble average of property A is determined by integrating over all possible configurations of the system by:

$$< A >= \iint A(p^N, r^N)\rho(p^N, r^N)dp^N dr^N$$

where $A(p^N, r^N)$ is the observable of interest, r is the atomic positions, and p the momenta. The probabily density function $\rho(p^N, r^N)$ of the ensemble is given by:

$$\rho(p^N, r^N) = \frac{1}{Q} exp\left[\frac{-H(p^N, r^N)}{K_b T}\right]$$

where K_b is the Boltzmann factor, T is the temperature, and H is the Hamiltonian. The denominator Q in the expression is called the *partition function*, that is a dimensionless normalizing sum of Boltzmann factors over all microstates of the system:

$$Q = \iint exp\left[\frac{-H(p^N, r^N)}{K_b T}\right] dp^N dr^N$$

The partition function is extremely important because a lot of thermodynamics properties can be calculated from the previous equation. The partition function relates microscopic thermodynamic variables to macroscopic properties. However, in order to overcome the analytical difficulties in solving the above-reported equation, it can be used the *ergodic hypothesis*: over long periods of time, the time-average of a certain physical property, represents the ensemble-average of the same property.

$$< A >_{ensemble} = < A >_{time}$$

The time-average $<A>_{time}$ can be computed by:

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

where t is the simulation time, M is the number of time steps in the simulation and $A(p^N, r^N)$ is the instantaneous value of the calculated property. Therefore, from an efficient sampling of microstates in a specific ensemble sequentially in time, can be computed an ensemble-average properties of the system under investigation[1].

1.2 The Molecular Mechanics Force Field

A molecular dynamics simulation requires the definition of a potential energy function in order to solve the Newton's equation of motion. The Molecular Mechanics (MM) method uses Newtonian

mechanics to model molecular systems, neglecting the electronic motions and analyzing the system as a set of atoms interacting through a potential energy function. The most important theoretical basis of MM is founded on the results produced by the great names of analytical mechanics: Hamilton, Euler, Lagrange, Newton. The core of the MM approach is the set of the equation and parameters used to describe the potential energy function *V* of a molecular system, also known as *force field*.

The potential energy function

The force field allows to compute the molecular system's potential energy in a given conformation as a sum of individual energy contribution:

$$V = V_{bonded} + V_{non-bonded}$$

where the components of the covalent and non-covalent terms are given by the following equations:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals}$$

$$V_{non-bonded} = V_{electrostatic} + V_{Van\ Der\ Vaals}$$

Each of the terms mentioned above, can be modelled in a different way, depending on the particular simulation settings being used [1]. The implementation of the potential energy function plays a central role in the MM approach. The potential energy function can be described as:

$$V(r_1, r_2, \dots, r_n) = \sum_{bonds} \frac{1}{2} k_l [l - l_0]^2 + \sum_{angles} \frac{1}{2} k_\theta [\theta - \theta_0]^2$$
$$+ \sum_{dihedrals} k_\varphi [1 + \cos(n\varphi - \delta)] + \sum_{angles} \frac{1}{2} k_\zeta [\zeta]^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\frac{q_i q_j}{4\pi\varepsilon_1 \varepsilon_0 r_{ij}} + \frac{A(i,j)}{r_{ij}^{12}} + \frac{C(i,j)}{r_{ij}^{6}} \right)$$

The first term in the previous equation models the interaction between bonded atoms, described with a harmonic potential inducing potential energy increase when the bond length *l* departs from the reference value l_0. The second term describes the angle among three atoms again modelled by using a

harmonic potential. The third contribution is a torsional potential that describes bond rotates, and the fourth term represents the non-bond interactions. The *force field* equation previously shown consists of two distinct components: the set of equations used to generate the potential energies (also known as the potential function), and the parameters used in this set of equations. The *force field* parameters usually derived empirically or by means of quantistic approach. The non-bond interactions represent a very important component of the MD force field, being the dominant type of interaction between molecules, critical in maintaining the three-dimensional structure of proteins and nucleic acids. The non-bond terms are usually modelled as a function of an inverse power of the distance. The non-bond interactions (the latest two terms in the previous equation) are usually built up from two components: Van Der Waals forces and electrostatic interactions.

Van Der Waals interactions

The interactions between two atoms is also characterized by the *Van Der Waals* forces, representing the sum of the attractive or repulsive forces between molecules (or between parts of the same molecule), caused by correlations in the fluctuating polarizations of nearby particles. The *Van Der Waals* interactions are relatively weak compared to covalent bonds but play a fundamental role in defining many properties of organic compounds, including their solubility in polar and non-polar media. The *Van Der Waals* potential is often represented with the *Lennard-Jones* equation, an approximate model for the isotropic part of a total (repulsion plus attraction) van der Waals force as a function of distance (as shown below).

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

The previous equation contains two parameters, the collision diameters $\sigma_{i,j}$ and the well depth $\varepsilon_{i,j}$. The Lennard-Jones potential is characterized by a repulsive component (the first term in the previous equation) and an attractive component (the second term in the previous equation). The calculation of non-bond forces is extremely expensive in terms of computational effort, because the number of the non-bond interactions increases as the square of the number of atoms in the system. To properly reduce the computational effort, the non-bond interactions are computed by applying the *distance cutoff*. With the application of the cutoff distance, every interaction between two atoms is computed only if their distance is smaller than the cutoff.

Electrostatic interactions

The electrostatic interaction can be described by using the Coulomb's law:

$$V = \frac{Q_i Q_j}{4\pi\varepsilon_0\varepsilon_r r_{i,j}}$$

The factor $4\pi\varepsilon_0\varepsilon_r r_{i,j}$ is the electric conversion factor, where ε_0 is the free space permittivity and ε_r is the relative permittivity. This type of interaction is defined as long-range interaction, because the energy decreases with the distance between two atoms as $\frac{1}{r}$.

However, in order to efficiently compute the electrostatic interactions in biomolecular simulations, it is widely applied the Ewald summation, a methodological approach for computing long-range interactions in periodic systems. The Ewald summation is a special case of the Poisson summation formula, replacing the summation of interaction energies in real space with an equivalent summation in Fourier space. The long-range interaction is divided into two parts: a short-range contribution, and a long-range contribution. The short-range value is computed in real space, whereas the long-range term is estimated using a Fourier transform. The main advantage of this approach is the fast convergence of the energy compared with that of a direct summation. The practical application of this method requires charge neutrality of the molecular system in order to compute accurately the electrostatic interaction.

Periodic Boundary Conditions

The computational model of the molecular system is characterized by *Periodic Boundary Conditions* (PBC), in order to minimize the edge effects in a finite system. All the atoms are put in a space-filling box, usually filled with water (implicitly or explicitly modelled), surrounded by translated copies of itself (Figure 1), with the aim of removing boundaries of the system. The inaccuracies resulting from the presence of PBC are expected to be less severe than the errors resulting from an artificial boundary with vacuum [1]. In the minimum image convention, each individual particle in the simulation interacts with the closest image of the remaining particles in the system, which is repeated infinitely if PBC are settled on. Applying a cutoff distance is not a problem for short range interactions as the Lennard-Jones potential which decreases very rapidly, but the long-range interaction model requires the use of more accurate methods (e.g., shift function, and switch functions) with the aim of avoiding discontinuities in the potential energy calculation. In addition to the cutoff method, there are other

approaches to overcome these kind of problems and to properly compute the electrostatic interaction: Particle Mesch Ewalds , Reaction Field, Multipole Cells [1].

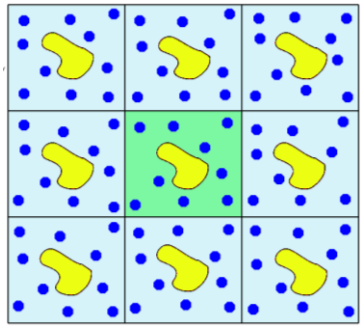


Figure 1. The central box, filled with water, is replicated in copies of itself.

Potential energy minimization

The Potential Energy function is a complex multidimensional function of molecular system coordinates. A minimum points of the Potential Energy Surface (PES) correspond to stable states of the system: any movement away from this configuration, is characterized by higher energy. However, there are also a lot of *local minima* in the PES, and the minimum with the lowest energy is known as *global minimum*. Starting from the MM description of the molecular system, in term of *force field*, energy minimization is able to reduce the potential energy of the system. To identify the minimum point of the PES there are two different approaches to the minimization problem: *derivative methods* and non-*derivative methods*. The set of derivative minimization methods is built up from *first order methods* (e.g., Steepest Descent, Conjugate Gradient) and *second order methods* (e.g., Newton-Raphson, L-BFGS). In the first order method, the direction of the first derivative of the energy (the gradient) indicates where the minimum lies. Second derivative indicates the curvature of the function, showing where the energy function change direction. Energy minimization is widely present before a molecular dynamics simulation, especially for simulation of complex system, such as macromolecules.

1.3 Molecular Dynamics

Ordinary differential equations of Newtonian dynamics

The basic idea of the MD method is to solve Newton's equations of motion for a system of N interacting atoms (or particles). For each atom, the acceleration a_i is given as the derivative of the potential energy with respect to the position, r:

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} = -\frac{1}{m_i}\frac{dV}{dx}$$

The Newton's laws of motion may be applied to different types of problem. In the simplest case, no force acts on each particle between collisions. From one collision to the next, the position of the particle thus changes by $v_i \partial t$, where v_i is the (constant) velocity and ∂t is the time between collisions. In the second case, the atoms are subjected to a constant force between collision, as in case of charged particle moving in a uniform electric field. In the last case, the force acting on the particle depends on its position relative to the other particles. In the latter case, because of the complex nature of the particles' trajectory, the motion is very difficult to describe analytically. The solution of this equation can be obtained by using some numerical integration schemes. Finite difference techniques are applied to generate molecular dynamics trajectories with continuous potential models, which we will assume to be pairwise additive. The fundamental idea is that the integration is decomposed into many small steps, each separated in time by a fixed time ∂t. The total force on each particle in the configuration at a time t is computed as the sum of its interactions with other particles.

Verlet integration methods

There are several algorithms for integrating the equations of motion using finite difference methods in molecular dynamics calculations. All algorithms assume that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions.

$$x(t+\partial t) = x(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \frac{1}{6}\partial t^3 b(t) + \frac{1}{24}\partial t^4 c(t) + \cdots$$

$$v(t+\partial t) = v(t) + \partial t a(t) + \frac{1}{2}\partial t^2 b(t) + \frac{1}{6}\partial t^3 c(t) + \cdots$$

$$a(t+\partial t) = a(t) + \partial t b(t) + \frac{1}{2}\partial t^2 b(t) + \cdots$$

Where v is the velocity, a is the acceleration, b is the third derivative of the position, and so on. The *Verlet algorithm* [2] is probably the most widely used method for integrating the equations of motion in a molecular dynamics simulation. The Verlet algorithm uses the positions and accelerations at time t, and the positions from the previous step at time $t - \partial t$ to calculate the new positions at time $t + \partial t, r(t+\partial t)$. In detail:

$$r(t+\partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \cdots$$

$$r(t-\partial t) = r(t) - \partial t v(t) + \frac{1}{2}\partial t^2 a(t) + \cdots$$

Considering the above-reported equations:

$$r(t+\partial t) = 2r(t) - r(t-\partial t) + \partial t^2 a(t) + \cdots$$

The velocities do not explicitly appear in the Verlet integration algorithm and can be computed in a variety of ways; a simple method is to divide the difference in position at times $t + \partial t$ and $t - \partial t$ by $2\partial t$. The Verlet implementation is straightforward and the storage requirements are modest. The main drawback is that the positions are obtained by adding a small term $\partial t^2 a(t)$ to the difference of two much larger terms, $2r(t)$ and $r(t - \partial t)$. This may lead to a loss of precision. Moreover, it is not a self-starting algorithm, considering that the new positions are calculated from the current position $r(t)$ and the positions from the previous time step, $r(t - \partial t)$. At $t = 0$ it's necessary to employ some other means to obtain position at $(t - \partial t)$.

A very important choice is the integration time-step (usually fs) in order to avoid instability and to sample correctly the phase-space. There are various integration methods for MD, for example a variation of the Verlet algorithm, called the Leap Frog Algorithm.

Velocity Verlet integration methods

The *velocity Verlet* method [3] provides positions, velocities and accelerations at the same time and does not affects precision:

$$r(t + \partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t)$$

$$v(t + \partial t) = v(t) + \partial t[a(t) + a(t + \partial t)]$$

This method is implemented as a three-steps process because the calculation of new velocities requires the accelerations at both t and $(t + \partial t)$. Thus, in the first stage the positions at $(t + \partial t)$ are computed using the velocities and the accelerations at time t. The velocities at time $\left(t + \frac{1}{2}\partial t\right)$ are then estimated applying:

$$v\left(t + \frac{1}{2}\partial t\right) = v(t) + \frac{1}{2}\partial t[a(t)]$$

New forces are then calculated from the current positions, thus providing $a(t + \partial t)$. In the last stage, the velocities at time $(t + \partial t)$ are determined using:

$$v(t + \partial t) = v\left(t + \frac{1}{2}\partial t\right) + \frac{1}{2}\partial t[a(t + \partial t)]$$

Molecular Dynamics implementation scheme

The MD flowchart is represented in Figure 2, starting from the molecular system input data. The initial atoms velocities v_0 chosen randomly from a Maxwell-Boltzmann distribution at a given temperature

and the initial atomic position are known (e.g., from Protein Data Bank). Following the MD flowchart, the set of atoms coordinates *x(t)* and velocities *v(t)* is generated step by step, giving the trajectory that describes positions, velocities and accelerations of the particles as functions of time. The MD method is deterministic: once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. After initial changes, the molecular system reaches an equilibrium state: this can be interpreted as a statistical ensemble that will provide a macroscopic description of the behavior of the system. Using the output trajectory of the MD, the macroscopic thermodynamic properties (e.g., temperature, energy, pressure) can be calculated as time averages.

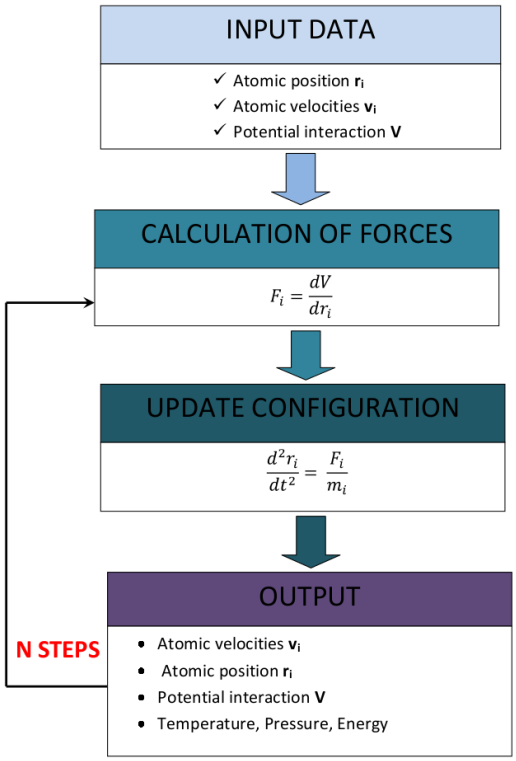


Figure 2. In the MD algorithm scheme, the initial positions and velocities are provided as input data. Starting from the atomic position r_i, the potential energy V, which models the interaction between atoms, is calculated. The scheme continues with the calculation of the forces F_i acting on each atom, by deriving the potential energy function. Then the integration of the equation of motion leads to the calculation of new position r_i and velocities v_i. The cycle goes on for a number of steps until the equilibrium is reached (convergence of the computed equilibrium property).

MD Software packages

There is a wide variety of MD codes for biomolecular simulation: AMBER, CHARMM, GROMACS, NAMD, etc. GROMACS (GROningen MAchine for Chemical Simulations) is a molecular dynamics simulation package originally developed in the University of Groningen, now maintained and extended at different places and university. GROMACS is written for Unix-like operating systems and the entire package is available under the GNU General Public License. GROMACS is a versatile tool to perform molecular dynamics simulations and energy minimization. It is primarily designed for biochemical

molecules like proteins, lipids and nucleic acids that have a lot of bonded interactions, but it is very fast in the calculation of non-bonded interactions, and for this reason many research groups are also using it for non-biological systems, e.g. polymers [4]. GROMACS 4.6 software offers GROMOS96 force-fields in its distribution; it is characterized by accurate description of the interaction potential energy function and a relatively simple computational form, with the aim of reducing the computational effort of the calculation.

1.4 Dimensionality Reduction Techniques

The protein dynamics is a stochastic search of the stable conformational state among all the accessible protein arrangements. The time scales of current molecular simulations (hundreds/thousands of nanoseconds) are at least one order of magnitude smaller than the time required for protein folding. In this context, many methodological approaches have been developed to address the folding problem. The polypeptide backbone and in particular the side chains are constantly moving due to thermal motion or kinetic energy of the atoms. For this reason, a molecular dynamics simulation produces very high-dimensional data-sets with millions of sampled protein arrangements. Out of thousands of modes in proteins, only a few degrees of freedom contain more than half of the total fluctuations of the system; therefore, a strategy is needed to identify the most important (slow) modes. For this purpose, Principal Component Analysis[5–8] and Functional Mode Analysis [9,10] are among the most efficient methods.

Principal Components Analysis

Principal Components Analysis is a powerful approach to determine a small set of collective vectors with the largest contribution to the mean square fluctuations (MSF) of the atomic coordinates. Starting from the 3N cartesian coordinates $x_i(t)$ (i=1, …, 3N), the elements of the covariance matrix C of the atomic positions are given by:

$$C_{ij} = < (x_i - < x_i >)(x_j - < x_j >) >$$

Before computing C, the simulation trajectory is superimposed onto a reference structure to remove translation and rotation of the entire biomolecule. Diagonalization of C yields a set of 3N orthonormal eigenvectors (e_j) with corresponding eigenvalues (σ_j^2). The eigenvectors are ordered according to descending eigenvalues and referred to as PCA vectors. The projection $p_j(t)$ is called j^{th} principal component (PC) and represents the position of the protein along the j^{th} PCA vector as reported below:

$$p\ \ (t) = \ \ [x(t) - < x >] * e_j$$

The protein motion along a simulation trajectory, quantified as MSF, can be decomposed into contributions from different PCs:

$$< (x - < x >)^2 >= \sum_{j=1}^{3N} var(p_j) = \sum_{i=1}^{3N} \sigma_i^2$$

In general, the first PCs represents a large fraction (80-90%) of the atomic MSF. Hence, if the molecular event under investigation is characterized by large protein motions, the first few PCs are reasonable basis set to analyse the biomolecular simulation. However, a functionally relevant mode is in most cases not identical to a specific normal or PCA mode but may be distributed over a number of PCA modes. To address this issue, it was recently developed the so-called Functional Mode Analysis (FMA), as described below.

Functional Mode Analysis

Starting from a set of protein conformations, together with an arbitrary "functional quantity" that quantifies the functional state of the protein (e.g., Radius of Gyration, Solvent Accessible Surface), the FMA approach seeks the collective motion that is maximally correlated to the functional quantity f(t) [9]. Assuming the function f as linearly correlated with the first d principal components, the collective vector MCM (Maximally Correlated Motion) is computed as:

$$MCM = \alpha_j PC_j \quad \text{with} \quad \text{j=1, …, d.}$$

Where α_i are obtained by maximizing the Pearson's correlation coefficient (R), reported below.

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

It has been demonstrated[9] that the maximum absolute value of R can be found by numerically solving the coupled linear set of equations:

$$\sum_{j=1}^{d} \alpha_i \, cov(p_j, p_k) = cov(f, p_k) \quad \text{with} \quad k=1, \ldots, d.$$

Considering that MD simulations of proteins are characterized by long autocorrelations, the maximization of R can lead to overfitting if too many free parameters α_i are used in the optimization process. Thus, it is extremely important to cross-validate the derived model with an independent set of simulation frames. Accordingly, the simulation can be divided into frames for model building and for cross-validation. The MCM is optimized applying the model building set only, and subsequently the obtained model is validated by predicting the function f(t) using the cross-validation set only.

1.5 Enhanced Sampling

Computer simulations, employing classical mechanics, have grown rapidly over the past few years, allowing the study of a broad range of biological systems, from small molecules to large protein complexes in a solvated environment. However, the results of an MD simulation are meaningful only if the run is long enough for the system to visit all the energetically relevant configurations or, in other words, if the system is ergodic in the timescale of the simulation. In fact, long computer time is required to simulate short phenomena from the life of biomolecules. Moreover, biological molecules are known to have rough energy landscapes, with many local minima frequently separated by high-energy barriers. Such limitations can affect the correct estimation of the macroscopic thermodynamic properties under investigation. Classical MD simulation sampling is in this case insufficient, because when the system is stuck in a relative free energy minimum, thermal fluctuations might never be able to overcome the energy barriers. To overcome this limitation, enhanced sampling techniques were developed in the past to improve the sampling efficiency of classical MD [11–17], including those that artificially add external driving force to guide the protein from one structure to another [14,18], such as Steered Molecular Dynamics (SMD), Metadynamics and Replica Exchange Molecular Dynamics (REMD).

2.5.1 Replica Exchange Molecular Dynamics

REMD has been firstly introduced in 1999 by Sugita and Okamoto [19] and nowadays is a firmly established enhanced sampling method. The general idea of REMD is to simulate N replicas of the original system of interest, each replica typically in the canonical ensemble, at a different temperature. The high temperature systems are generally able to sample large volumes of phase space, whereas low

temperature systems, whilst having precise sampling in a local region of phase space, may become trapped in local energy minima during the timescale of a typical computer simulation. REMD achieves good sampling by allowing the systems at different temperatures to exchange complete configurations, thus, the inclusion of higher temperature systems ensures that the lower temperature systems can access a representative set of low temperature regions of phase space [19]. Operatively, one constructs the ensemble of replicas equilibrated at their own temperatures, then starts a simulation for each one of them, and let the conformations of adjacent replicas to exchange with an adequate rate (Figure 3).

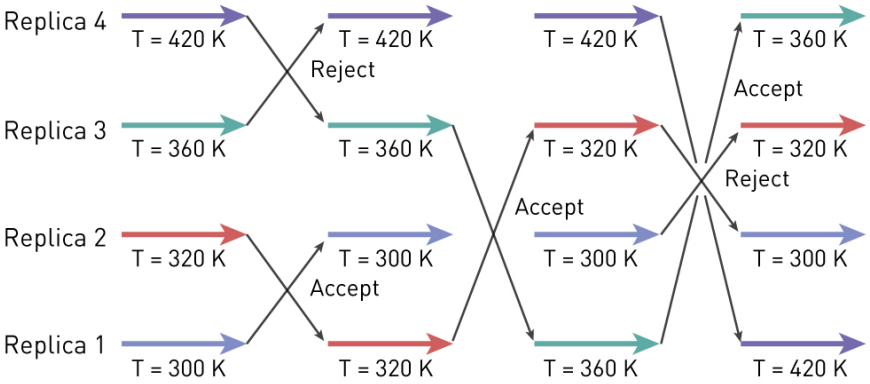


Figure 3. Schematic representation of REMD simulation and temperature exchange.

The transition probability from state X to X' in the REMD process is given by the Metropolis criterion:

$$w(X \to X') = \begin{cases} 1 & for\ \Delta \leq 0 \\ e^{-\Delta} & for\ \Delta > 0 \end{cases}$$

with

$$\Delta = (\beta_m - \beta_n)\{E(q^j) - E(q^i)\}$$

where E is the potential energy, q are the positions of atoms, m and n are the temperature indexes, i and j are the replica indexes, β is the inverse temperature defined by $\beta = 1/k_bT$. After the replica exchange, atomic momenta are rescaled as follows:

$$p^{i'} = \sqrt{\frac{T_n}{T_m}} p^i \quad p^{j'} = \sqrt{\frac{T_m}{T_n}} p^j$$

where p are the momenta of atoms [20,21] . As it can be seen by the above-mentioned equations, the probability depends on the difference of potential energy of the two adjacent replicas, so to have swaps between them they must have an adequate overlap in their potential distribution. If this doesn't occur, the system could get kinetically blocked. This leads to the choice of temperatures: the highest must be sufficiently high so as to ensure that no replicas become trapped in local energy minima, while the number of replicas used must be large enough to ensure that swapping occurs between all adjacent replicas.

2.5.2 Steered molecular dynamics

Steered molecular dynamics (SMD) applies an external steering force, applying a constraint (e.g. a harmonic potential), that drives along one or more reaction coordinates a system, in the configuration space, from one thermodynamic state to another. SMD employs a pulling force to cause a change in the structure during a molecular dynamics simulation. While the simulation run, all atoms adjust to the forced change in the structure so that conformations may be sampled along a pathway. In this way, the process, that usually is too slow due to energy barriers, is accelerated.

According to the stiff-spring approximation theory, the force constant k must be sufficiently large in order to ensure that the reaction coordinate closely follows the constraint positions:

$$F(t) = 2k(vt - s(t))$$

Where k is the spring constant of the constraint, v is the pulling velocity, and s(t) is the position of the ligand at time t. The work W is calculated by integrating the force over the pulled trajectory:

$$W\big(x(t)\big) = \int_0^{x(t)} F(t)dx(t)$$

The reaction path, in a typical investigation of a biomolecular process, is identified or hypothesized and the progress of the process is described by the reaction coordinate. The potential of mean force (PMF) plays an important role in such investigations. PMF is the free energy profile along the reaction

coordinate and is determined through the Boltzmann-weighted average over the other degrees of freedom. PMF captures the energy of the process studied and provides an essential ingredient for further modelling of the process; with all the other degrees of freedom averaged out, the motion along the reaction coordinate is well approximated as a diffusive motion on the effective potential identified as the PMF [22,23]. Since SMD is an optimal method to explore molecular processes, it is useful to calculate within its framework the PMFs. However, a SMD simulation is a non-equilibrium process, while PMF is an equilibrium property. Therefore, a theory, that connects equilibrium and non-equilibrium, has become available through recent advances in non-equilibrium statistical mechanics, especially through the discovery of Jarzynski's equality.

Imagine considering a finite classical system in contact with a heat reservoir. A central concept in thermodynamics is that a work is performed on a system when some external parameters of the system are changed over time. When the parameters are changed infinitely slowly along some path from an initial point A to a final point B in parameter space, then the total work W performed on the system is equal to the Helmholtz free energy difference ΔG between the initial and final configurations:

$$W = \Delta G = G^B - G^A$$

where G^A and G^B refer to the equilibrium free energy of the system, with the parameters held fixed at A or B. By contrast, when the parameters are switched along the path at a finite rate, then W will depend on the microscopic initial conditions of the system and reservoir, and will, on average, exceed ΔG:

$$\hat{W} \geq \Delta G$$

in which the overbar denotes an average over an ensemble of measurements of W. The difference $W_{diff} = \hat{W} - \Delta G$ is just the dissipated work associated with the increase of entropy during an irreversible process. By contrast, Jarzynski's equality states that:

$$e^{-\beta \hat{W}} = e^{-\beta \Delta G}$$

or, equivalently,

$$\Delta G = \frac{1}{\beta}\ln(e^{-\beta\hat{W}})$$

Where $\beta = 1/k_bT$. This result, which is independent of both the path from A to B, and the rate at which the parameters are switched along the path, is surprising: it says that we can extract equilibrium information ΔG from the ensemble of non-equilibrium measurements described above [24,25].

2.5.3 Metadynamics

Metadynamics belongs to a class of methods in which sampling is facilitated by the introduction of an additional bias potential (or force) that acts on a selected number of degrees of freedom, often referred to as collective variables (CVs). In metadynamics, an external history-dependent bias potential which is a function of the CVs is added to the Hamiltonian of the system. This potential can be written as a sum of Gaussians deposited along the system trajectory in the CVs space to discourage the system from revisiting configurations that have already been sampled. Let *S* be a set of *d* functions of the microscopic coordinates *R* of the system:

$$S(R) = \big(S_1(R), \dots, S_d(R)\big)$$

At the time t, the metadynamics potential can be written as:

$$V_G(S,t) = \int_0^t dt' \omega exp\left(-\sum_{i=1}^{d} \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2}\right)$$

where ω is an energy rate and σ_i is the width of the Gaussian for the *i*th CV.

The energy rate is constant and usually expressed in terms of a Gaussian height *W* and a deposition stride τ_G:

$$\omega = \frac{W}{\tau_G}$$

Metadynamics has several advantages: (i) it accelerates the sampling of rare events by pushing the system away from local free- energy minima, (ii) it allows exploring new reaction pathways as the system tends to escape the minima passing through the lowest free-energy saddle point, (iii) no *a priori* knowledge of the landscape is required. After a transient, the bias potential V_G provides an unbiased estimate of the underlying free energy:

$$V_G(S, t \to \infty) = -F(S) + C$$

where C is an irrelevant additive constant and the free energy $F(S)$ is defined as

$$F(S) = -\frac{1}{\beta}\ln\left(\int dR\, \delta\big(S - S(R)\big)e^{-\beta U(R)}\right)$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T the temperature of the system, and $U(R)$ the potential energy function. There are some disadvantages of this method: (i) The free energy landscape does not converge to a definite value but fluctuates around the correct result, leading to an average error which is proportional to the square root of the bias potential deposition rate. Thus, it may be difficult understand when to stop the simulation. It should be stopped when the motion of the CVs becomes diffusive in the region of interest. The other disadvantage is that correctly identifying a set of CVs appropriate for describing complex processes is far from trivial and represents a challenging task for a computer simulation. In order to address the first issues, it was developed the well-tempered metadynamics, as described in the following.

Well-tempered Metadynamics

In well-tempered metadynamics approach, the bias deposition rate decreases over simulation time, in order to control the regions of FES that are physically meaningful to explore. This is achieved by applying an history-dependent potential:

$$V(s, t) = \Delta T \ln(1 + \frac{\omega N(S, t)}{k_b \Delta T})$$

where ω has the dimension of an energy rate, ΔT is an input parameter with the same dimension of the temperature, and N(s, t) is the histogram of the S variables sampled during the simulation. Since V is a monotonic function of N, such a bias potential disfavors the more frequently visited configurations. It is worth mentioning that V (s, t) changes with the rate of:

$$\dot{V}(s,t) = \frac{\omega \Delta T \delta_{s,s(t)}}{\Delta T + \omega N(s,t)} = \omega e^{-[V(s,t)/\Delta T]} \delta_{s,s(t)}$$

The connection with metadynamics is evident if we examine the previous equation and replace $\delta_{s,s(t)}$ with a finite width Gaussian. The height of each Gaussian is determined by $w = \omega e^{-[V(s,t)/\Delta T]} \tau_G$, where τ_G is the time interval at which Gaussians are deposited. Thus, ω represents the initial bias deposition rate. It is worth highlighting that while the bias deposition rate decreases as 1/t, the dynamics of all the microscopic variables becomes progressively closer to thermodynamic equilibrium as the simulation proceeds. The bias potential converges to:

$$V(S, t \to \infty) = \frac{-\Delta T}{T + \Delta T} F + C$$

Where C is an immaterial constant. It is worth highlighting that in the long-time limit, the probability distribution of the CVs become:

$$P(S) \propto e^{\frac{-F(S)}{k_b(T+\Delta T)}}$$

While ordinary MD corresponds to the limit ΔT→0, the standard metadynamics is obtained for ΔT→ ∞. Fine tuning of ΔT allows to regulate the extent of FES exploration. This is a useful procedure to avoid overfilling of the free energy landscape and save computational time when many CVs are used. The introduction of a history-dependent potential alters the probability distribution sampled along the simulation. Several approaches have been developed in the past to reweight a metadynamics run and reconstruct the unbiased distribution for variables other than the CVs assuming an adiabatic evolution for the bias potential [26–28].

In conclusion, well-tempered metadynamics solves the convergence problems of metadynamics and allows the computational effort to be focused on the physically relevant regions of the conformational space [29]. The latter property makes it possible to use adaptive bias methods in higher dimensionality cases, thus paving the way for the study of complex systems where it is difficult to select a priori a very small number of relevant degrees of freedom

Chapter 2

Biological Background

Proteins are fascinating biochemical machines that undergo a huge number of conformational changes, in order to accomplish a wide range of physiological functions. Proteins are constituted by linear amino acids chains bonded together by peptide bonds between the carboxyl and amino groups of adjacent amino acid. Multiple structural protein level can be observed: the primary, secondary, tertiary, and sometimes quaternary structure. The protein physiological and/or pathological functions are closely related to its conformational state, demonstrating the importance of structural investigations in understanding molecular reasons underneath several biological phenomena. The folding process is a stochastic search of the protein native state among all the accessible conformations. External factors such as pH, temperature, and interaction with surfaces characterized by specific physical and chemical properties, may be responsible for the failure of the folding mechanism, with consequent achievement of an improper protein arrangement. The misfolding process triggers a cascade of events that culminates in the formation of fibrillar protein aggregates responsible for several human diseases. The presence of stable fibrillar aggregates in the organs of patients suffering from protein deposition diseases led initially to the reasonable postulate that mature fibrils were the main cause of the diseases. However, recent findings have raised the possibility that fibril precursors, such as soluble oligomers, represent the pathogenic species responsible for the disease onset and severity.

The present chapter describes the biological context of the scientific problem under investigation. In detail, protein folding process is presented in Sections 2.1. The fundamental interactions responsible for the protein folding are discussed in Section 2.2. The Sections 2.3 and 2.4 are devoted to the mechanism of protein misfolding and amyloid aggregation. The last two Sections 2.5 and 2.6 are focused on Spinocerebellar Ataxia Type 1 and Alzheimer's Disease, respectively.

2.1 Protein Folding and Energy Landscapes

Proteins, the most abundant molecules in biology other than water, participate in every process within cells. Proteins are synthesized in cells by ribosomes starting from the information contained within the cellular DNA. Following their biosynthesis, the majority of proteins must be converted into tightly folded compact structures in order to perform its physiological function. This mechanism, known as Folding Process, depends on the environment in which folding takes place. It is well known that within the cells of living organisms there are large numbers of auxiliary factors that assist in the folding process, including folding catalysts and molecular chaperones. These factors serve to enable polypeptide chains to fold efficiently in the complex, but they do not determine their native structures. Given the enormous complexity of the folding process, it is not surprising that failures to fold correctly sometimes occur. During the folding process, proteins organize themselves into specific three-dimensional structures from random coil through a lot of conformational changes, with the aim of reaching its native state. A correct folding process is closely related to the protein physiological function performed in living organism. The folding mechanism is a stochastic process starting from a quasi-infinite number of accessible states for protein configuration [30]. Following biosynthesized on a ribosome, a polypeptide chain is initially unfolded. Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein. Each of these conformational states and their intermediates are carefully regulated in physiological environment by enzymes activity, molecular chaperones, degradation system and quality control process.

Proteins carry out their function in living organism through a correct folding process, which leads proteins to a well-structured configuration, representing their native state. The protein conformational stability is reached during the folding process and is determined by a free energy balance (Figure 4). The main idea behind the *energy landscape theory* is that the energy landscape acts as a guide for protein native structure formation. Because of the interaction leading to the native state formation, the polypeptide chain is able to reach the lowest energy structure by a trial and error process. The unfolded protein undergoes specific kinetically preferred steps on the way to the native state. For large proteins, as well as for some smaller ones, there are kinetic barriers to reaching the native state. Thus, when left to fold on their own, these proteins become “kinetically trapped” in local energetic minima (Figure 4). It is such intermediate, partly folded conformations that can associate to form aggregates. [31]. These ideas have stimulated a huge number of investigations focused on the folding process by means of both experimental and theoretical procedures.

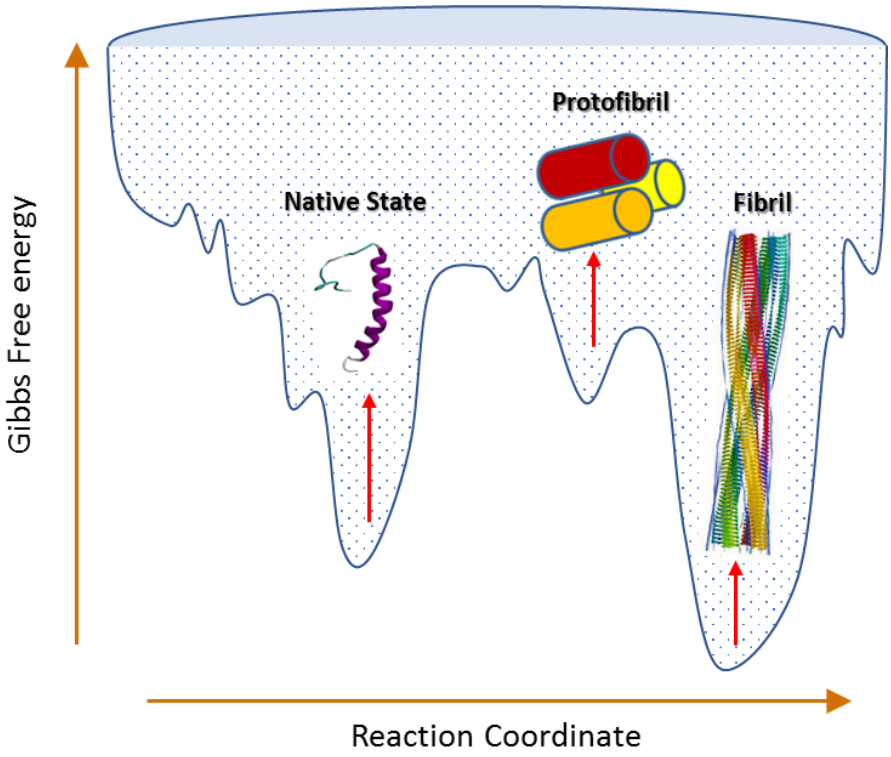


Figure 4. The energy landscape in the folding process. It shows all possible conformations of a molecular entity, or the spatial positions of interacting molecules in a system, and their corresponding energy levels, typically Gibbs free energy, on a three-dimensional Cartesian coordinate system. It is modelled as a funnel with traps [31].

2.2 Dominant Interactions in Protein Folding

The folding process strongly depends on the environmental conditions: the solvent (water or lipid bilayer), the concentration of salts, the temperature, and the presence of molecular chaperones. For these reasons it becomes important to describe the possible interactions regulating the protein folding process, such as the hydrogen bond, the hydrophobic/hydrophilic interactions, the electrostatic Coulombic interactions and the Van Der Waals forces.

Hydrogen Bond

The hydrogen bond is an attractive interaction of a hydrogen atom with an electronegative atom, that comes from another molecule or chemical group. The hydrogen atom must be covalently bonded to another electronegative atom (oxygen, fluorine, or nitrogen) to create the bond. Intra-molecular hydrogen bonding is partially responsible for the secondary, tertiary, and quaternary structures of proteins and nucleic acids. It also plays a pivotal role in the structure of polymers, both synthetic and natural. The strength of the hydrogen bond is dependent on the distance between the donor and

acceptor atoms, which is itself dependent on their electronegativity. The standard hydrogen bond between the donor and acceptor atoms is of the order of 0.26-0.35 nm.

Hydrophobic interactions

The hydrophobic effect is the observed tendency of non-polar substances to aggregate in aqueous solution and exclude water molecules. This kind of interactions are the main responsible of the linear polypeptide folding in water and appear when non-polar compounds, as alkyl groups, are put into water and they tend to come into contact. Hydrophobic interaction is an entropy-driven process and it is certainly of great importance in maintaining the structural integrity of proteins and biological membranes. Structures of water-soluble proteins have a hydrophobic core in which side chains are buried from water, which stabilizes the folded state, and charged and polar side chains are situated on the solvent-exposed surface where they interact with surrounding water molecules. Minimizing the number of hydrophobic side chains exposed to water is the principal driving force behind the folding process, although formation of hydrogen bonds within the protein also stabilizes protein structure. In the case of protein folding, the hydrophobic effect is important to understand the structure of proteins that have hydrophobic amino acids clustered together within the protein.

Electrostatic interactions and Van Der Waals forces

The protein folding pathway is also related to the non-covalent Electrostatic and Van Der Waals interactions [30]. Electrostatic contribution plays a pivotal role in the conformational structure of proteins, since that many proteins contain polar and charged groups. Van Der Waals interactions, sum of the attractive or repulsive forces between molecules (or between parts of the same molecule) other than those due to covalent bonds or to the electrostatic interaction of ions with one another or with neutral molecules. Van der Waals forces are relatively weak compared to normal chemical bonds, but play a fundamental role in the structural biology, and folding process. The attraction between atoms and molecules, involved in the Van Der Waals interaction, depends on the distance among them. Differently by covalent and ionic bonds, Van der Waals forces are strictly related to polarization variation of nearby particles. The role of such forces can be understood thinking to the cohesion of ordinary substances.

2.3 Protein Misfolding

The protein folding is a stochastic research of the stable native-state among the accessible conformations. During this process, external factor, such as changing in pH, temperatures, hydrophobicity/hydrophilicity properties, can lead to the folding mechanisms failure, with consequent

achievement of an improper protein spatial configuration. Misfolded proteins are generally characterized by the exposition to protein environment of chain tracts buried in the inner part of the structure in the native globular state. *Amyloidosis* occurs when soluble and functional polypeptide are converted into insoluble ordered fibrils, named amyloids, characterized by dimensions from tens nanometer to microns (Figure 5). Proteins or peptides exposed under some stress condition convert from their native-state into highly ordered fibrillar aggregate responsible for pathological conditions ranging from neurodegenerative disorders to systemic amyloidosis. The failure to fold correctly, will give rise to the malfunctioning of living systems and therefore to disease. Indeed, a large amount of recent evidences has suggested that a wide range of human diseases is characterized by aberrations in the folding process. Some of these diseases (e.g. cystic fibrosis) result from the simple fact that if proteins do not fold correctly they will not be able to exercise their proper functions. In other cases, misfolded proteins results from the formation of intractable aggregates within cells or in the extracellular space. An increasing number of pathologies, including Alzheimer, Parkinson and Spinocerebellar Ataxia, are known to be directly associated with the deposition of such aggregates in tissue [32]. The protein stability depends in the free energy change between the folded and unfolded states. The folded conformation of a domain is apparently in a relatively narrow free energy minimum, and substantial perturbations of that folded conformation require a significant increase in free energy. The main causes which can lead to a protein misfolding and consequent unfolding are changes in solvent conditions (such as concentration, pH and components) [33], high stresses as increasing temperature [34], long time incubation [35] and interaction with surfaces characterized by specific physical and chemical properties [36]. Normally, temperature increasing is used to accelerate the folding and misfolding pathway, favoring faster and numerous experimental sessions. The unfolding free energy is temperature dependent because the heat capacity of the unfolded state is significantly greater than that of the folded state.

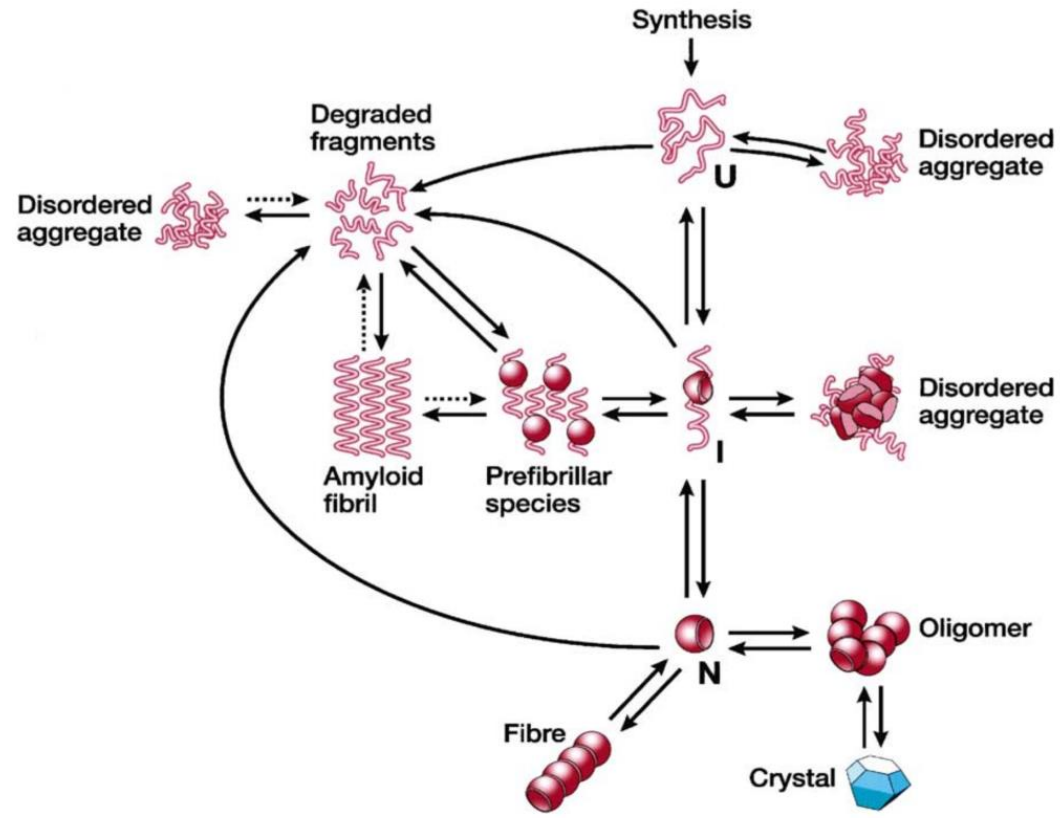


Figure 5. Schematic representation of some of the states accessible to a polypeptide chain following its synthesis on a ribosome [32].

2.4 Fibrillogenesis Pathway and Amyloid Formation

During the aggregation process, the polypeptide chains can adopt a lot of conformational states and interconvert among them on a wide range of time-scale (Figure 6). Amyloid nucleation is a complicated process due to the competition of the factors involved. In particular, it was shown that globular proteins need at least partial unfolding to be able to aggregate and form amyloid fibrils [37]. In some cases the presence of structured oligomers [38] or unstructured aggregates [39] has been reported as fibrils precursors. Aggregation pathway consist of a nucleation-dependent process and the results is the accumulation of rich in β-sheet fibrillar aggregates. To date, it is unclear whether aggregation proceeds via linear addition of single molecules or whether there are oligomeric intermediates, but the structural conversion to β-sheet results in the formation of prefibrillar intermediates. Protofibril assembly is followed by fibril formation, resulting in the characteristic inclusions observed in the amyloid-like diseases. Protofilaments and fibrils represent the last step of protein misfolding. The process is irreversible at this step since that mature fibrils structure is stable. Insoluble aggregates, as well as soluble oligomers, are characterized by the exposition of the hydrophobic core of the partially

unfolded protein, which attests the tendency of the protein to self-aggregate. The aggregation is prevented if protein concentration is below its critical threshold, while, if concentration is higher than critical threshold, a lag time exists before protein polymerization in order to favor nuclei formation. The presence of the stable fibrillar deposits in the organs of patients suffering from protein deposition diseases led initially to the reasonable postulate that amyloid fibrils were the main cause of the diseases. However, recent findings have raised the possibility that precursors to amyloid fibrils, such as low molecular weight oligomers and/or structured protofibrils, are the real pathogenic species, at least in neuropathic diseases [40,41].

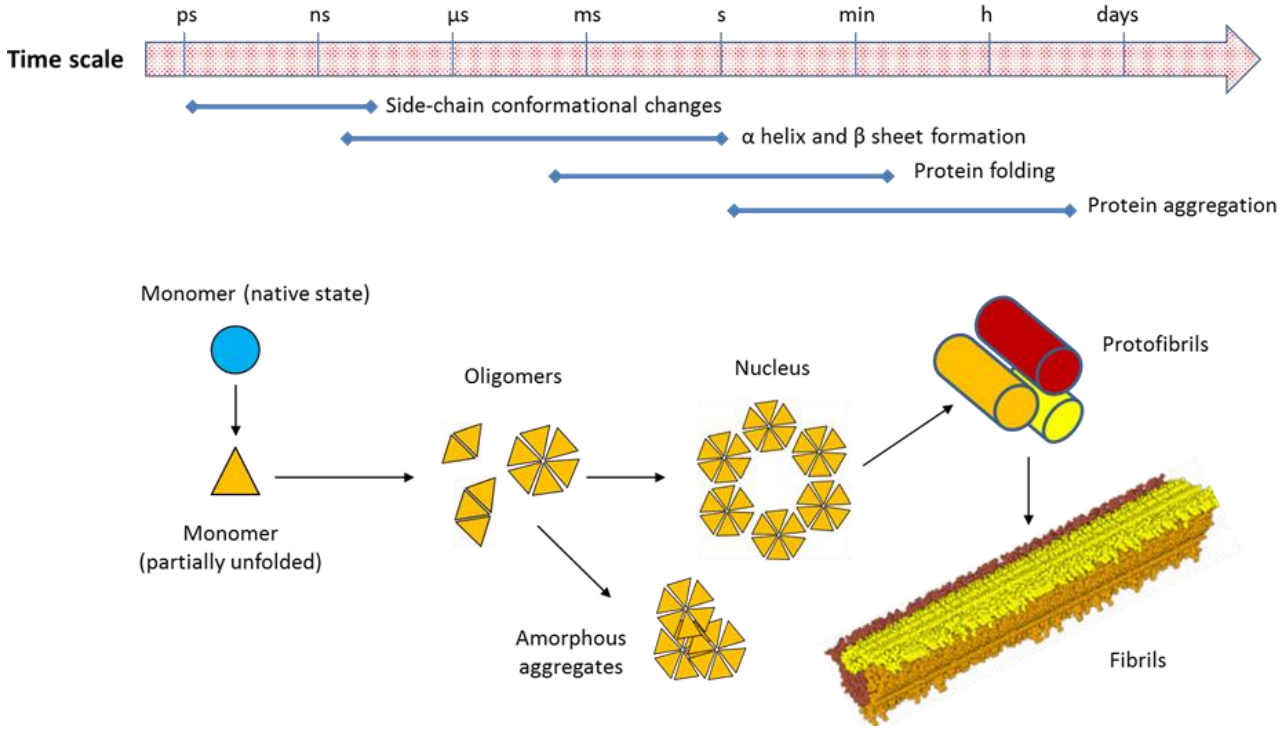


Figure 6. Some of the many conformational states adopted by polypeptide chains. The transition from beta-structured aggregates to amyloid fibrils can occur by addition of either monomers or protofibrils (depending on the protein) to preformed beta-aggregates.

2.5 Spinocerebellar Ataxia Type-1 and AXH domain

Several neurological disorders arise from the previously explained protein aggregation and amyloid formation mechanisms. In case of polyglutamine (polyQ) diseases, the neurodegenerative disorders are caused by an unstable expansion of genomic cytosine-adenine-guanine (CAG) trinucleotide repeats beyond a specific threshold. In the healthy population, the number of CAG repeats lies between 10 to 51, while 55-87 CAG repeats is associated with the disease [42]. The CAG tracts are translated into expanded glutamine regions in the respective proteins, promoting their self-assembly and amyloid structure formation. A large amount of experimental evidence has indicated a correlation between poly-Q containing inclusions and disease progression [43].

Within this context, Spinocerebellar ataxia type 1 (SCA1) is an inherited neurodegenerative disease belonging to the class of polyglutamine (polyQ) diseases. In fact, it is well established that SCA1 is mainly caused by polyQ expansion in Ataxin-1 (Figure 7a), a nuclear protein constituted by 820 residues [44]. Despite the well-established correlation between polyQ expansion and disease onset and severity [45], other findings have demonstrated the importance of the AXH domain (Figure 7b) of Ataxin-1 in both modulating the development of the disease and influencing the aggregation process [46–49].

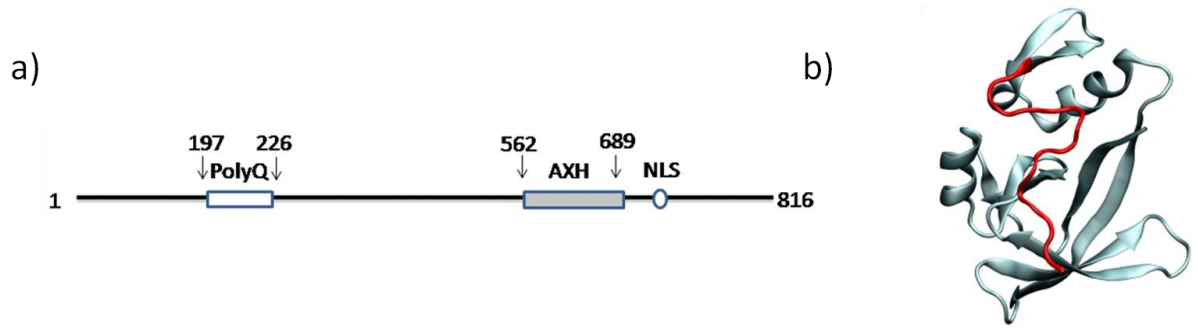


Figure 7. a) Ataxin-1 primary amino acid sequence. The structured AXH domain is also represented. b) Visual inspection of the AXH monomer. The N-terminal tail is highlighted in red.

It was reported that the isolated AXH has the tendency to misfold and aggregate into fibers without any destabilizing factor, such as temperature increase or chemical denaturants [48]. Moreover, recent findings have shown that Ataxin-1 aggregation is strongly reduced by replacement of AXH domain with the homologous sequence of the transcription factor HBP1 [48], confirming the important role of AXH in modulating the aggregation mechanism. Despite the importance of the AXH domain of Ataxin-1 in driving the Ataxin-1 aggregation process [46–49], the AXH self-association mechanism, is not yet clarified and several crucial questions remain open. The present PhD work employs enhanced sampling techniques to fully characterized the AXH aggregation pathway from monomer to tetramer, identifying several protein mutations responsible for the destabilization of the monomer/dimer/tetramer equilibrium (Chapter 3).

2.6 Alzheimer's Disease and Amyloid Hypothesis

Among several theories proposed to explain the cause of Alzheimer's Disease (AD), the amyloid hypothesis represents one of the most likely scenarios [50,51]. More in detail, the amyloid hypothesis is based on the idea that a mutation on an amyloid precursor protein (APP) induces the aggregation of Aβ peptides, whose deposition into senile plaques is followed by the formation of neurofibrillary tangles and neuronal cell death [51]. However, it is still not clarified if the formation of these fibrils is the cause or a secondary effect of the disease [52]. The major components of AD-associated amyloid

plaques are $A\beta_{1-40}$ peptides but also the more toxic $A\beta_{1-42}$ species [53], characterized by two additional amino acids and generated through a sequential cleavage of the amyloid precursor protein (APP) by β and γ secretases [54]. In general, these peptides are able to oligomerize and then the resulting oligomers can further aggregate giving rise to ordered fibrils and fibers [55]. Several experimental studies have been focused on the molecular characterization of amyloid fibrils, given the intimate relationship between molecular structure and disease onset and severity [56].

At present, all the $A\beta_{1-40}$ species resolved by NMR, share a U-shaped motif, where the peptide chains form two β-strands connected by a loop region [57–61]. In case of more toxic $A\beta_{1-42}$ species, despite earlier NMR models exhibited the same U-shaped motif [58], more recent investigations demonstrated the possibility of S-shaped arrangements [62–67], characterized by three β strands (Figure 8). Interestingly, the S-shaped arrangement is not stable in case of $A\beta_{1-40}$ species [68]. Within this framework, the higher toxicity of $A\beta_{1-42}$ species compared to $A\beta_{1-40}$ may be explained by their ability to form S-shaped assembly. Such a correlation could arise if the S-shaped model i) was characterized by a more stable molecular architecture per se, or ii) was able to assemble into structures that are not possible by considering the U-shaped Aβ chains, as recently suggested [68]. The present PhD thesis provides a molecular level understanding of the interactions governing the structural arrangement in $A\beta_{1-42}$ species (Chapter 4). Outcomes of the study suggest that the molecular architecture of the protein aggregates might play a pivotal role in formation and conformational stability of the resulting fibrils.

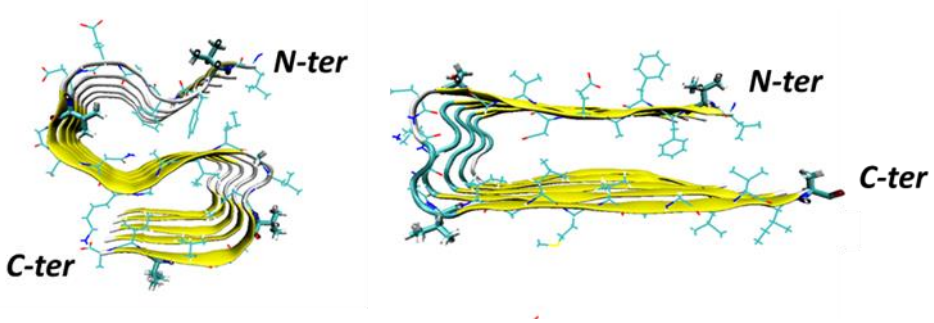


Figure 8. The S-shaped and U-shaped models of $A\beta_{17-42}$ pentamer are represented on left and right, respectively.

Chapter 3

AXH Aggregation Pathway: From Monomer to Tetramer

Ataxin-1 is the protein responsible for the Spinocerebellar ataxia type 1, an incurable neurodegenerative disease caused by polyglutamine expansion. The AXH domain of Ataxin-1 plays a pivotal role in the misfolding and aggregation pathway responsible for the pathology onset. The present work employs classical and enhanced sampling techniques to fully characterized the AXH aggregation pathway from monomer to tetramer, identifying several protein mutations responsible for the destabilization of the monomer/dimer/tetramer equilibrium.

The Section 3.1 is focused on the description of the AXH monomer, revealing its substantial conformational fluctuations in physiological environment, especially in the N-terminal region. These fluctuations can be generated by thermal noise since the free energy barriers between conformations are small enough for thermally-stimulated transitions.

The Section 3.2 is devoted to the characterization of the AXH inter-monomer interface and the molecular description of the AXH monomer-monomer interaction mode. The computational results support previous experimental findings suggesting that interactions leading to the dimer formation can stabilize the N-terminal region of the AXH. Moreover, the importance of the I580 residue in mediating the AXH monomer-monomer interaction dynamics was fully explained with atomic resolution.

The Section 3.3 reports a computational study to identify the structural and energetic basis of AXH tetramer stability. Results of the study point the attention, for the first time, on R638 protein residue, which shown to play a key role in AXH tetramer stability. Therefore, R638 might be also implicated in AXH oligomerization pathway and stands out as a target for future experimental studies focused on self-association mechanisms and fibril formation of full-length ATX1.

(i) ***G. Grasso****, M.A. Deriu, J.A. Tuszynski, D. Gallo, U. Morbiducci, A. Danani, Conformational fluctuations of the AXH monomer of Ataxin-1, Proteins Struct. Funct. Bioinforma. 84 (2016) 52–59. doi:10.1002/prot.24954.*

(ii) *M.A. Deriu,* ***G. Grasso****, J.A. Tuszynski, D. Massai, D. Gallo, U. Morbiducci, A. Danani, Characterization of the AXH domain of Ataxin-1 using enhanced sampling and functional mode analysis, Proteins Struct. Funct. Bioinforma. 84 (2016) 666–673. doi:10.1002/prot.25017.*

(iii) ***G. Grasso****, U. Morbiducci, D. Massai, J. Tuszynki, A. Danani, M. Deriu, Destabilizing the AXH Tetramer by Mutations: Mechanisms and Potential Antiaggregation Strategies, Biophys. J. 114 (2018) 323–330. doi:10.1016/j.bpj.2017.11.025.*

3.1 Conformational Fluctuation of AXH Monomer of Ataxin-1

Introduction

The structure of the AXH domain, the only globular region so far identified in Ataxin-1, is an asymmetric homodimer [44], whose interface is characterized by 20-residue motifs, which mutually adapt, but are different [69]. AXH is an unusual chameleon protein representing the first known example of the existence of alternative conformations [70]. For the previously-mentioned reason AXH domain was the subject of several experimental studies aimed at characterizing its behavior in solution under different conditions [44,69,71–74]. The emerging picture is that the AXH domain shows different structural properties leading to asymmetric dimeric interfaces [70]. More in depth, despite the well-preserved tertiary and quaternary structure, the AXH domain is characterized by local structural plasticity at the N-terminal regions of each monomer, which is involved in the dimer interface [70]. Moreover, it was reported that the dimer appears to be the predominant AXH species in solution [71], unlike the homologous sequence HBP1, which is a monomer [72]. While the existence of the AXH monomeric species has already been demonstrated, the conformational stability of the AXH monomer (hereinafter called AXH^m) in solution is a debated issue [70]. Previous works have also highlighted that the monomer is unstable in water with a tendency to aggregate [70]. Recently, it was demonstrated that the extraordinary structural plasticity of the AXH^m plays an essential role in determining the equilibrium between the physiological and pathological interactions of ATX1 [75]. Despite the large number of experimental studies focused on the AXH domain, no molecular modeling effort has been carried out to investigate structural and dynamics of ATH^m with an atomic resolution till now.

This study reports on the first Molecular Dynamics investigation of AXH monomer conformational properties and dynamics. To address this issue, both biased and unbiased MD techniques have been applied. Whereas classical MD simulation have revealed the extensive conformational fluctuation of the N-terminal tail of the AXH monomer in solution, Replica Exchange Molecular Dynamics (REMD) provided a better sampling of the AXH^m state space and, the estimation of the free energy landscape characterizing the AXH^m conformational changes. In the present study it has been paid attention to the dimensionality reduction of the MD trajectories, applying the widely used Principal Component Analysis (PCA) to elucidate the large-scale molecular phenomena characterizing the AXH^m folding pathway. Surprisingly with respect to previous hypothesis, the results presented here depict the AXH monomer as a relatively stable structure with, in general, highly conserved domains except for the N-terminal tail switching among several arrangements.

Materials and Methods

Classical Molecular Dynamics

The AXH monomer model coordinates were extracted from the recently refined solution structure of the AXH domain [70] (for consistency with previous literature [70], residue numbering refers to the UniProtKB/Swiss-Prot entry No. P54253+1). The system was fully solvated in explicitly modelled water [76] where the minimum distance between the protein and the edge of the box was 1.2 nm. The net charge of the system was neutralized by the addition of Cl− and Na+ ions. The above-mentioned system (20,000 interacting particles) was first minimized by the steepest descent energy minimization algorithm followed by two preliminary position restraint MD simulations of 100 ps and 50 ps in NPT and NVT ensemble, respectively. The production run was performed for 1μs in the NVT ensemble. GROMACS 4.6 package has been used for all MD simulations and data analysis [4]. AMBER99-ILDN force-field [77,78] was used to define the protein topology, since its ability to provide an accurate description of protein conformational ensembles was widely demonstrated [77,79]. The TIP3P model [80] was used for representing water molecules. Electrostatic interactions were calculated at every step with the Particle-Mesh Ewald method with a short-range electrostatic interaction cut-off of 1.2 nm. A cut-off of 1.2 nm was also applied to Lennard-Jones interactions. The LINCS algorithm [81] approach has allowed an integration time step of 2 fs. The GROMACS *v-rescale* thermostat [82] was used to keep simulation temperature constant at 310 K. The Visual Molecular Dynamics (VMD) [83] package was used for the visual inspection of the simulated systems. Dedicated GROMACS tools were used for a quantitative analysis in terms of Root-Mean-Square Deviation (RMSD) and Root-Mean-Square Fluctuation (RMSF). Analysis of secondary structure (SS) dynamics was performed by applying the STRIDE software [84,85]. Ramachandran plots, produced using PROCHECK [86], were used for better insight into the atomistic changes of the AXH residues.

Replica Exchange Molecular Dynamics and Cluster Analysis

REMD [19] was applied to explore the conformational ensembles of the AXH Domain. In particular, 64 replicas were simulated for temperatures ranging from 310 K to 480 K. Temperatures were distributed across the replicas in a geometric progression, i.e. with the same ratio used to scale each temperature from the one below it, keeping the overlap of the potential energy distributions constant across the temperature space (Appendix A-1). The resulting exchange probability was 0.4. The exchange of replicas was attempted every 1 ps. This time interval was chosen to be large enough compared to the coupling time of the heat bath (τ=0.1 ps). Each replica was simulated for 50 ns, obtaining a cumulative simulation time of 3.2 μs. To obtain the canonical average of each physical quantity at a specific

temperature, the computational data were analyzed as time average over all trajectory steps corresponding to the chosen temperature, as usually done in REMD [87]. Structure based clustering approach has been carried out to get insight into likeliness of AXH N-terminal arrangements. GROMOS approach [88], implemented in GROMACS package, was applied to the Replica Exchange Molecular Dynamics trajectory at 310 K.

Principal Component Analysis and Free energy Reconstruction

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the system, elucidating large-scale and low-frequency modes, respectively, thus yielding collective motions directly related to a specific molecular event [89]. After the alignment of the AXH^m C-alpha Cartesian coordinates, the covariance matrix was calculated and diagonalized. To estimate the free energy landscape of the AXH^m, the first two eigenvectors derived from the PCA were used as reaction coordinates. The corresponding free energy surface along the 2-dimensional space *V*=(*pca1,pca2*) is given by:

$$\Delta G(V) = -K_b T * [lnP(V) - \ln(P_{max})]$$

where *P*(*V*) is the probability distribution obtained from the histogram of the computational data taken from the MD trajectory, T is the temperature, and K_b is the Boltzmann constant. P_{max}, representing the maximum of the distribution, was used to ensure that $\Delta G=0$ for the lowest free energy minimum.

Results

Classical MD

A MD simulation of 1 μs was performed to analyze the AXH^m structural properties in solution. The C-alpha-C-alpha RMSD reported in Figure 9a, shows that the AXH^m undergoes several conformational transitions until the stability is reached (from 600 to 1000 ns). Analysing the RMSF of the AXH^m residues (Figure 9b), it is possible to identify the N-terminal domain (residues Ala565-Ile580) as mainly responsible for the protein conformational fluctuations mentioned above (RMSF=1.5 nm). The visual inspection of the AXH^m conformational states corresponding to the stability region of the RMSD plot (reported in Figure 9a) are shown in c, confirming the leading role of the N-terminal domain as responsible for the conformational fluctuation of the AXH^m. In greater detail, once the RMSD stability

is reached (from 600 to 1000 ns), the N-terminal tail (Ala565-Pro568) is packed against Gly680-Leu688 and Asp608-Glu620 (Figure 9c, IV).

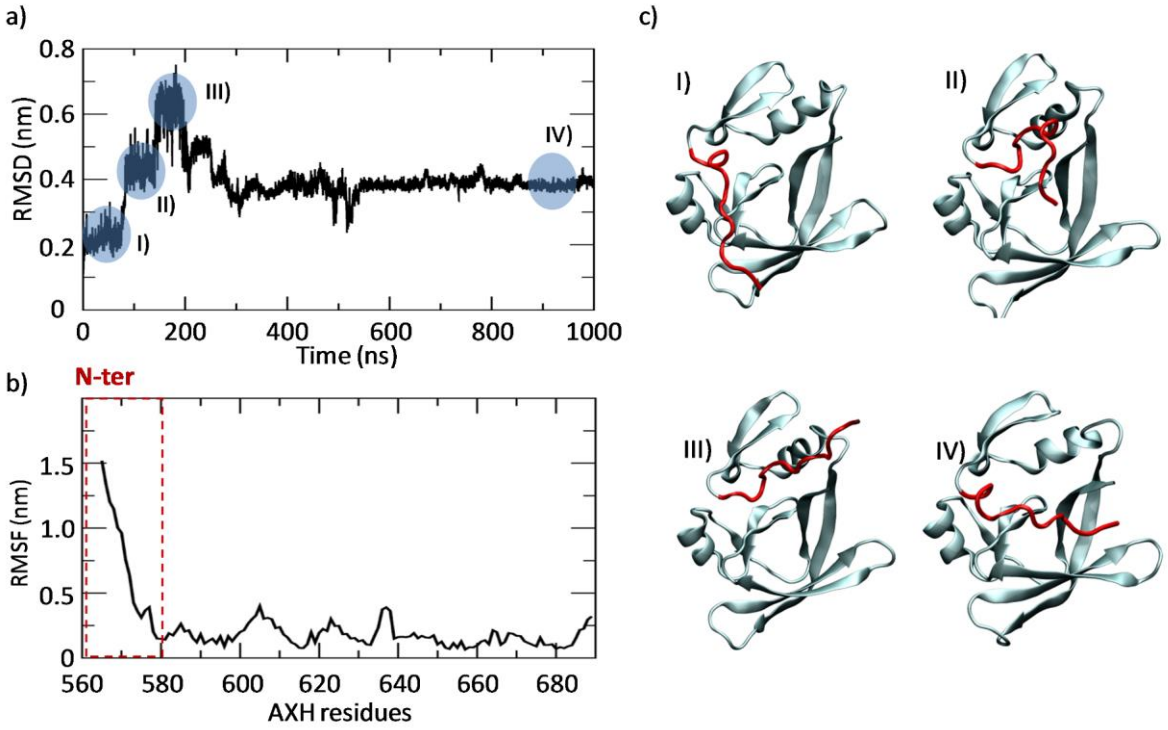


Figure 9. a) C-alpha/C-alpha Root Mean Square Deviation (RMSD) calculated over the classical MD trajectory. The stability regions of the RMSD plot are highlighted by circles. For each stability point (I, II, III and IV), the corresponding AXH monomer conformation is represented on panel c. b) Root Mean Square Fluctuation (RMSF) calculated over the classical MD trajectory of the AXH monomer in water.

Figure 10 shows the AXH^m secondary structure probability, calculated by averaging the 4 monomer configurations taken from 4APT [70] and 400 configurations taken every 1000 ps from the MD equilibrium trajectories (from 600 to 1000 ns). The protein secondary structure resulted to be highly conserved, with the exception of α2 and partially α3, both characterized by an intrinsic tendency of helix-coil transition in water environment. More than 99% of the protein residues were found in the most-favoured and additional-allowed regions of the Ramachandran Plots (Appendix A-1), providing support for the model's correctness.

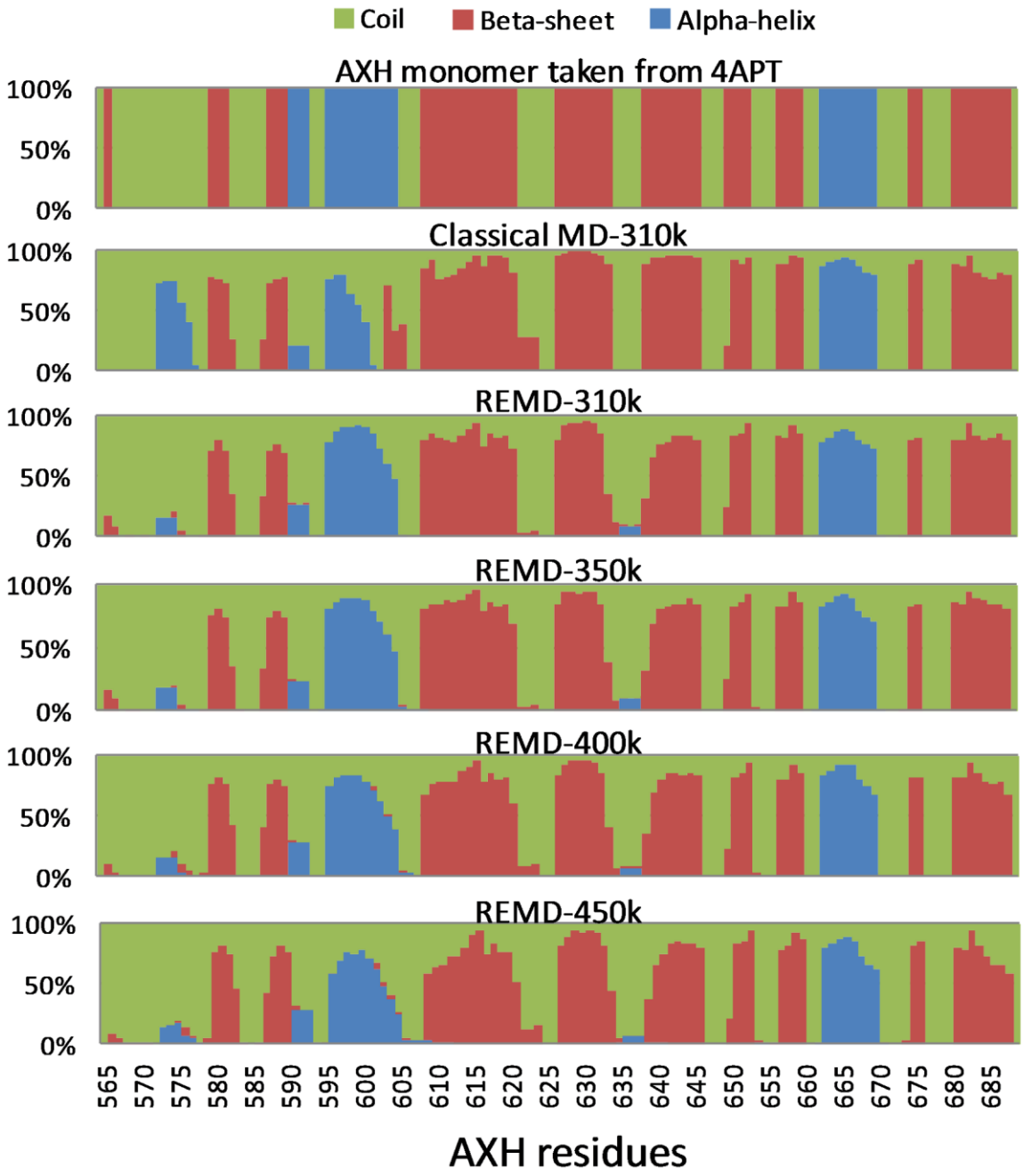


Figure 10. Secondary structure probability calculated over the monomer conformation taken from APT (as average), the classical MD simulations at the equilibrium, and REMD ensamble at different temperature.

Replica Exchange Molecular Dynamics

From the REMD simulations, i) all the temperatures were explored numerous times by each replica and ii) acceptance ratios of more than 0.4 were obtained. The computational data were analyzed as time averages over all trajectory steps corresponding to a specific temperature, as is usually done in REMD [87], to study the conformational variation of the AXH^m at a specific temperature. Interestingly, it was observed that the AXH^m secondary structure is highly conserved in the temperature range from

310 K to 450 K, with the exception of α2 and (partially) α3, confirming data coming from the classical MD trajectory.

The application of the Principal Component Analysis (PCA) allowed to highlight the large-scale and low-frequency modes mainly related to a specific protein conformational change. After the alignment of the AXH^m C-alpha atoms, the MD trajectory at 310 K was filtered to show only the motion along the first and second eigenvector, calculated by the covariance matrix diagonalization.

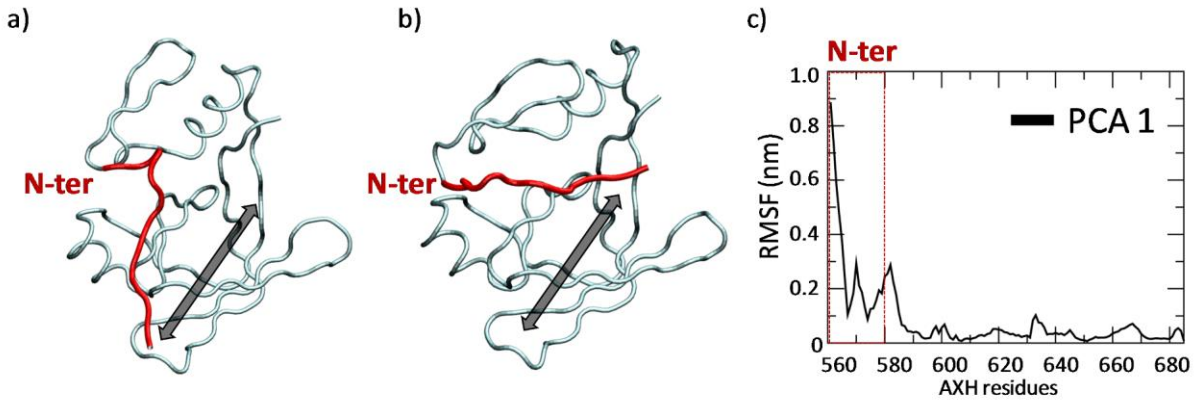


Figure 11. a) and b) Visual inspection of the extreme extensions along the first Principal Component eigenvector; c) Root Mean Square Fluctuation calculated over the Classical MD trajectory filtered on the first Principal Component eigenvector.

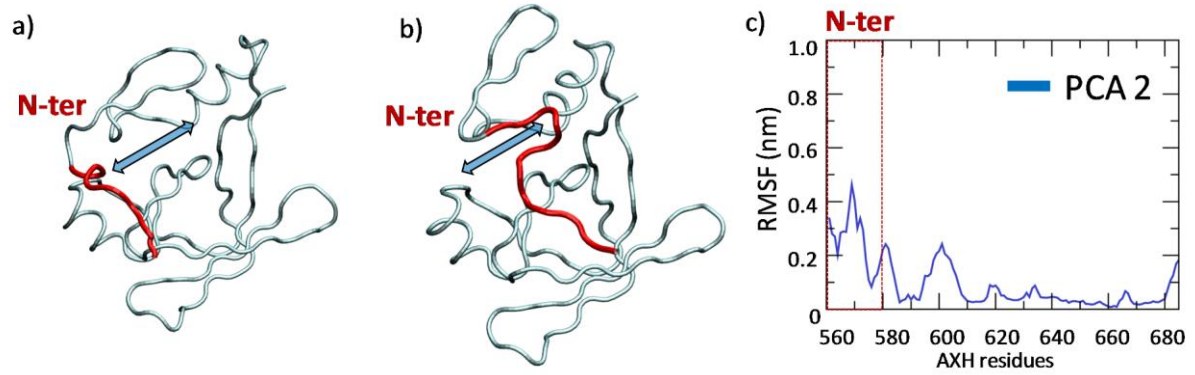


Figure 12. a) and b) Visual inspection of the extreme extensions along the second Principal Component eigenvector. c) Root Mean Square Fluctuation calculated over the Classical MD trajectory filtered on the second Principal Component eigenvector.

The AXH^m RMSF calculated over the filtered trajectory (Figure 11 and Figure 12) clearly shows that, as expected, the first and second PCA eigenvectors are able to effectively describe the conformational changes of the N-terminal region ($RMSF_{PCA1}$=0.9 nm and $RMSF_{PCA2}$=0.5 nm). The free energy landscape of AXH^m was calculated by applying the Boltzmann inversion to the probability distribution of the projections along PCA1 and PCA2 taken from the REMD trajectory at 310 K.

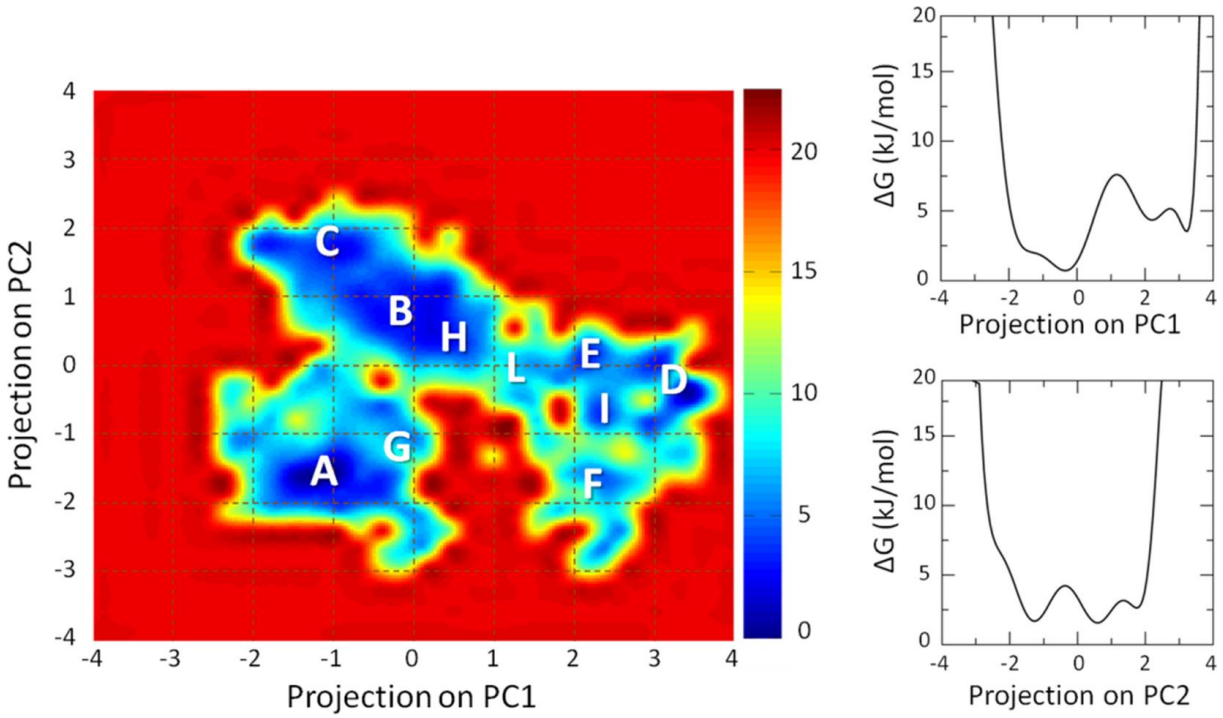


Figure 13. On the left the free energy profile (kJ/mol) of the AXH^m are represented as function of the first and second Principal Component projection. Labels indicate the position on the free energy landscape (in term of first and second Principal Component projection values) of cluster centroids obtained by structure based clustering approach on the REMD trajectory at 310 K. The 1D free energy profiles as function of each Principal Component projection are reported on the right panels.

Thus, the free energy landscape shown in Figure 13 is reported in a 2-dimensional space *E*=(*pca1,pca2*). Several free energy wells can be identified (Figure 13) at PCA projection values of: i) PCA1=-1.2, PCA2=-1.5; ii) PCA1=-1.8, PCA2=-1.8; iii) PCA1=0.2, PCA2=-0.7; and iv) PCA1=3.4, PCA2=0.3.

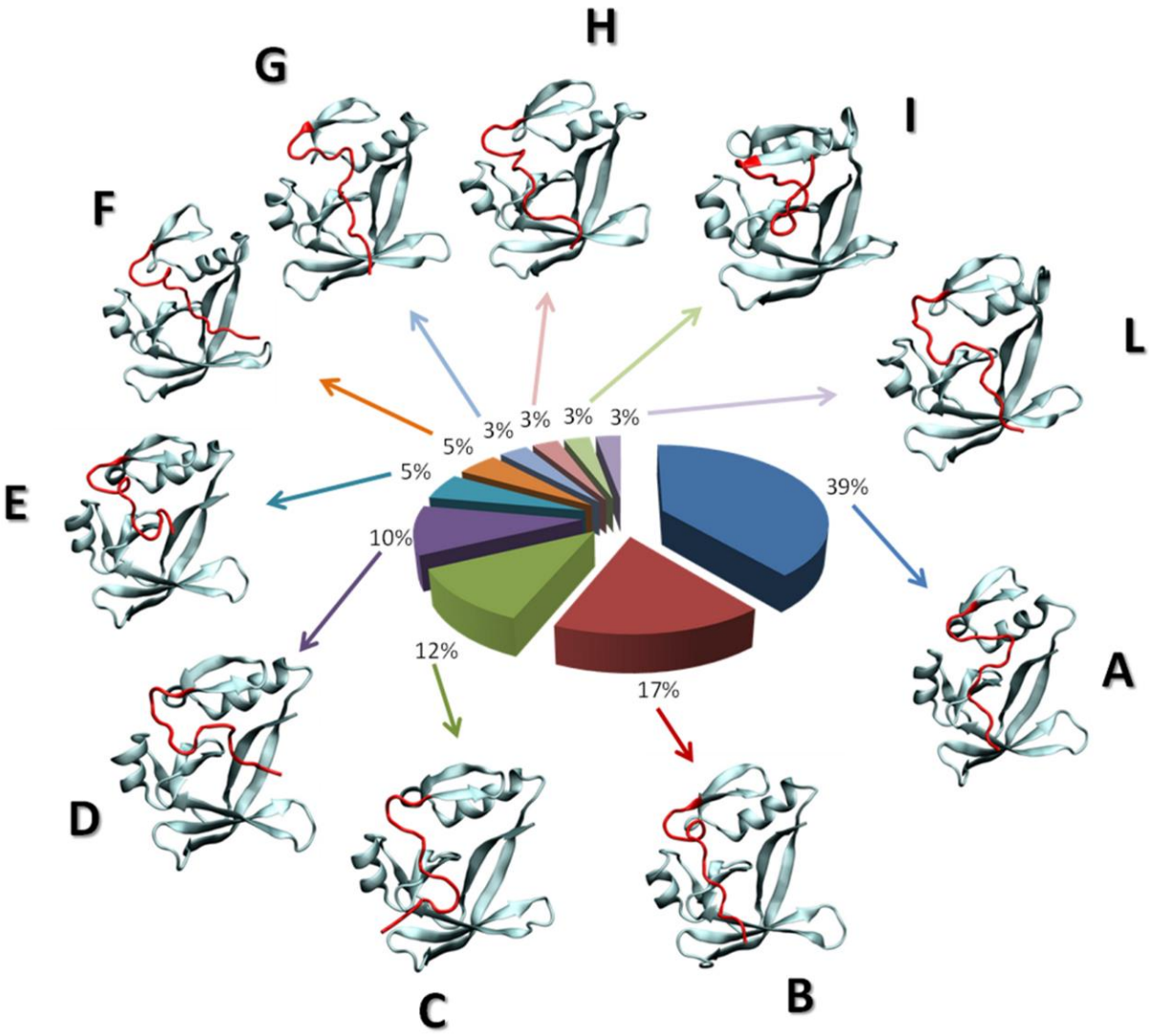


Figure 14. Structure based clustering analysis on REMD trajectory at 310 K. Cluster centroids are also shown and labeled. Clusters not shown are characterized by a percentage population lower than 2% over the whole trajectory steps.

A structure-based clustering approach to get insight into other possible AXH^{m} conformational states, was applied to the REMD trajectory at 310 K (Figure 14). The most populated clusters are represented in Figure 14 whereas clusters characterized by a percentage population lower than 2% over the whole trajectory steps have not been considered. Cluster A, B, C and D are the most populated and refer to 4 N-terminal different arrangements.

Moreover, cluster centroids have been also represented (in term of first and second Principal Component projection values) on the free energy landscape in Figure 13. It is worth noticing how, cluster centroids are positioned on energy wells covering mostly the whole free energy landscape. As expected, centroids representing the most populated clusters (A, B, C and D) fall in deepest energy minima.

The free energy barrier that separate the previously mentioned minima lie in range 5-10 kJ/mol demonstrating that the AXH^m may overcome the energy barriers under the thermal fluctuation, moving from one state to another. These findings suggest that, despite the well-defined and conserved secondary structure (Figure 10), the AXH^m alone in water environment is characterized by extensive conformational fluctuation of the N-terminal region, as previously reported for Classical MD simulation.

In view of already known AXH crystal structure [44,70], it is interesting to highlight how the experimentally solved conformations fall in the estimated free energy landscape (Appendix A-1).

The three available Protein Data Bank (PDB) models containing AXH tetramers (1OA8, 4APT and 4AQP) have been represented in term of first and second Principal Component projection values for each monomer, called a,b,c, and d (Appendix A-1). In all cases, PDB models fall in free energy areas close to those covered by clusters B and H, suggesting that results in term of free energy landscape are consistent with experimental observations. Of interest is the finding that none of the PDB models fall into energy wells close to cluster A (the most populated REMD cluster) indicating that this specific N-terminal arrangement is not achieved in a dimeric or tetrameric form.

Discussion

The crystal structure of the isolated AXH domain contains two globular dimers, (i.e. four AXH molecules), with the protomers characterized by asymmetric dimer interfaces. Despite the sharing of the secondary structure, the N-terminal region of the AXH domain is characterized by a conformational heterogeneity already described by experimental works [69]. These findings have raised the interesting issue of whether the conformational flexibility is i) an intrinsic feature of the AXH monomer, ii) an artifact of crystallization or, iii) dependent on the monomer-monomer interaction [70]. A recent experimental study has demonstrated that the AXH aggregation and misfolding are impaired by the interaction between AXH^m and its cellular partner, i.e. the transcriptional regulator CIC peptide [75]. This finding has led to suggest a strategy for preventing AXH aggregation by using CIC peptide as a template to design peptide-mimetic molecules able to stabilize the AXH^m [75]. In detail, the AXH physiological interactions are able to shift the equilibrium between monomeric and dimeric species towards the monomer. A leading role in these physiological interactions is played by the AXH domain N-terminal, that is drastically stabilized.

Here, it was presented the first computational description of the AXH^m conformational fluctuations through a characterization at atomistic level of the AXH monomer in water by means of both biased and unbiased MD simulations. An approach that combines the extraction of collective motions with the sampling of its free energy surface was applied, as previously done in literature [87,90]. This method

brings together the efficient sampling of REMD with a dimensionality reduction technique. In particular, PCA was used to elucidate the collective motion maximally related to the conformational fluctuation of the N-terminal region of the AXH^m, and the REMD simulation was carried out to deeply sample the state space and to calculate the corresponding free-energy landscape.

Recently, it has been suggested that the AXH^m might be unstable because its sickle-cell shape leaves the protein hydrophobic core exposed [70]. Here, the AXH^m in water environment was found to be stable for temperature ranging from 310 to 450K, with the exception of the N-terminal region (Figure 9 and Figure 13). In fact, one of the most important result of the present work is that, although the AXH^m is characterized by a well-preserved secondary structure (Figure 10), the N-terminal region has shown an intrinsic tendency to accomplish a wide range of conformational changes (Figure 9 and Figure 13). These structural fluctuations are well described analyzing the MD trajectory along the first two Principal Components (Figure 11 and Figure 12), used in this work to estimate the corresponding free energy landscape. The identified free energy wells (Figure 13) are separated by an energy barrier of about 5-10 kJ/mol demonstrating that the AXH^m may overcome this barrier under the thermal fluctuation, moving from one low-energy state to another. These findings suggest that, despite the well-defined and conserved secondary structure (Figure 10), the AXH^m alone in water environment is characterized by extensive conformational fluctuations of the N-terminal region (clusters A to H, Figure 13, and Figure 14). Moreover, the available PDB models fall only in free energy areas corresponding to clusters B and H (Appendix A-1). Therefore the specific N-terminal arrangement characterizing cluster A (the most likely for the AXH^m) has never been observed in a dimeric or tetrameric form [44,70]. It is plausible that the interactions leading to the dimer formation might be able to stabilize the N-terminal region of the AXH^m, which is involved in the dimeric interface or, *vice versa*, that this arrangement favors the dimerization by affecting the kinetics.

Starting from the assumption that the equilibrium of the monomer/dimer/tetramer species of the AXH represents an important point of the aggregation pathway of the Ataxin-1 [91], further work is needed to better clarify the importance of the monomer-monomer interaction in the structural and dynamics properties of the AXH domain. Together with previous experimental findings [75], the computational results presented here can be considered as a basis for a future design of ataxin aggregation inhibitors that will require several key conformations identified in the present study as molecular targets for ligand binding.

Conclusions

The present work points out the attention on N-terminal fluctuations of Ataxin-1 providing evidences on several possible conformations which may be accessible in the AXH monomeric form. Those

monomeric conformations fall into energy minima separated by low energy barriers, indicating that the AHX^m N-terminal arrangement can switch continuously from a configuration to another under the effect of thermal fluctuations. Nevertheless, experimental data on AHX dimers and tetramers suggest that few of the above shown N-terminal conformational states are actually accessible in aggregated AHX. Therefore, a relationship may possibly exist between AXH dimerization kinetics and N-terminal flexibility. Within this context, the present work provides a number of AXH models for further investigations on AXH dimers to shed light on molecular mechanism behind AXH aggregation. Moreover, on the basis of insights coming from these further studies, novel aggregation inhibitors might be designed to drive N-terminal conformational changes in order to reduce the monomer-monomer binding affinity and destabilize dimeric and tetrameric aberrant aggregates.

3.2 Characterization of the AXH Dimer by Functional Mode Analysis

Introduction

The AXH domain, the only globular region identified in ATXN1[70] so far, is an unusual chameleon protein constituted by an asymmetric homo-dimer [44], and with an interface characterized by 20-residue motifs mutually adapting, although different [69]. Despite the well-preserved tertiary and quaternary structure, the N-terminal (Nter) region of each AXH monomer, which is involved in the dimer interface [70], adopts different topologies leading to asymmetric dimeric interfaces [69].

In the previous section of the present PhD thesis (Section 3.1), classical and biased molecular dynamics (MD) has been applied to investigate the extensive conformational fluctuation of the Nter tail of the AXH monomer (AXH^m) in solution [92]. It has been demonstrated that AXH^m is characterized by a relatively stable structure with, in general, highly conserved domains, except for the Nter tail switching between several conformations.

Based on those findings, and in accordance with previous observations [70], it has been suggested that interactions leading to the dimer formation might be able to stabilize the Nter region of the AXH^m, which is involved in the dimeric interface [92]. However, several issues concerning the protein-protein interactions characterizing AXH conformational stability and, more generally, a dynamic description of protein conformational changes at dimer interface, are still unknown. Computational approaches, and in particular MD, have often demonstrated to be a powerful tool to explore the protein-protein interactions and conformational changes with atomistic resolution [93–100]. In particular, the dimensionality reduction of the trajectory obtained from Replica Exchange Molecular Dynamics (REMD) simulations can be an effective instrument in the identification of the protein-protein dominant binding modes. In this regard, Functional Mode Analysis (FMA) has been recently developed to identify collective atomic motions maximally related to the fluctuation of specific variables of interest [9].

Here, REMD [19] is applied to investigate protein-protein interactions leading to the AXH dimer stability, focusing on: i) the characterization of the AXH monomer-monomer interface, and ii) the molecular description of the monomer-monomer interaction dynamics for the AXH dimer. A coupled approach, based on FMA applied to REMD trajectories to describe protein conformational changes at the dimer interface, is employed. The findings of the study, in agreement with previous experimental results [70], demonstrate the pivotal role of the I580 residue in mediating the AXH monomer-monomer interaction dynamics. Finally, the computational results here presented might also be seen as a starting point for further *ad hoc* experiments focused on evaluating the role of residues here identified in the stability of AXH dimer.

Materials and Methods

Replica Exchange Molecular Dynamics and Mode Analysis

The crystal structure of the AXH dimer (PDB code 4APT [70]), already employed in a previous study [92], was selected here as a starting structure. The 4APT [70] model contains two interacting AXH dimers, AB and CD, (where A, B, C and D are monomers with identical sequence). For the sake of consistency with the previous literature [70], residue numbering refers to the UniProtKB/Swiss-Prot entry No. P54253 +1. Each dimer structure AB and CD was placed in a dodecahedron box where the minimum distance between the protein and the edge of the box was set up to 1.2 nm. The box was then fully solvated in explicitly modeled water, and the total charge was neutralized by the addition of Cl− and Na+ ions. Each system consisted of about 40,000 interacting particles. These systems were first minimized by applying the steepest descent energy minimization algorithm, followed by preliminary position restrained MD of 100 ps (NPT, 1 atm and 310K) and 50 ps (NVT, 310K) duration. Then, 64 replicas were generated with temperatures ranging from 300 K to 500 K, and distributed applying the exponential spacing strategy, as previously suggested [101,102]. A first NVT position restrained MD was run on each replica for 100 ps. Finally a 50 ns of production NVT-REMD was carried out on each replica at its own temperature, according to previous works [87,103,104].

The replicas' exchange interval of 1 ps was considered as being large enough compared to the coupling time of the heat bath (τ=0.1 ps). The resulting exchange probability was 0.4.

To obtain the canonical average of each physical quantity, the computational data were analyzed as it is usually done in REMD[87], i.e., by time-averaging data over all trajectory steps corresponding to the chosen temperature (in this work 310 K).

As a control, for each AXH dimer, a 500 ns MD production run was performed in the NVT ensemble at 310 K.

AMBER99-ILDN force-field [77,78] was used to define the protein topology [77,79] and the TIP3P model [80] was used to represent the water molecules. The GROMACS 4.6 package was applied for MD simulations and data analysis [4].

The residues most responsible for monomer-monomer interaction were identified by contact probability plots. Contact probability for each residue was calculated, as reported elsewhere [105], using the following procedure. For each snapshot extracted from the REMD replica at 310 K, the distance between a residue in one monomer and all residues pertaining to the other monomer was calculated. If at least one distance value, among the residue-residue distances, was equal or less than a chosen threshold (0.2 nm), the residue of the first monomer was considered in contact with the interacting one

in that snapshot. The number of "contact snapshots" divided by the number of total snapshots taken out from the MD trajectories was defined as the contact probability associated with the residue [105].

FMA was applied to the REMD trajectory at 310 K to reduce the dimensionality of the system. This allowed the elucidation of large-scale and low-frequency modes, respectively, thus yielding collective motions directly related to a specific molecular event [9]. The applied method detects a collective motion maximally correlated to the fluctuation of the quantity of interest, i.e., in the case under study the monomer-monomer distance. Assuming that the variable of interest is a linear function of the Principal Components, the maximally correlated vector can be derived by maximizing the Pearson coefficient [9] to quantify the contributions of the individual PCA vectors to the fluctuations of the variables of interest. This approach yields a single collective mode, which drives the phenomenon under investigation, referred to as ensemble-weighted Maximally Correlated Motion (ewMCM).

In applying FMA, it is crucial to cross-validate the derived model for an independent set of simulation frames. The established approach applied for cross-validating the obtained results is to divide the simulation into a subset of frames for model building and a subset of frames for cross-validation. In this work, the obtained maximally-correlated motion was validated by predicting the function of interest, in the cross-validation subset, setting the Pearson correlation coefficient higher than 0.99. Further details of the FMA calculation are provided in Appendix (A-2).

Results

Along the REMD trajectories, the temperature axis was widely explored by each replica and acceptance ratios of more than 0.4 were obtained. The computational data were analyzed as time averages over all trajectory steps corresponding to 310K, to study: i) protein-protein interactions characterizing the AXH dimer, and ii) protein conformational changes at the dimer interface with atomistic resolution.

More data regarding the structural integrity of the individual monomeric units at 310 K are reported in Appendix (A-2). It is important to remark that the REMD trajectory at 310 K did not sample any dissociation event in AB and CD dimers. Data coming from classical MD simulations on AB and CD systems were in close agreement with the REMD trajectory at 310 K (Appendix A-2).

The dimerization interface in both AB and CD dimers buried 14.5 nm^2 of solvent accessible surface. More than 55% of the contact area is composed of hydrophobic contacts (Figure 15).

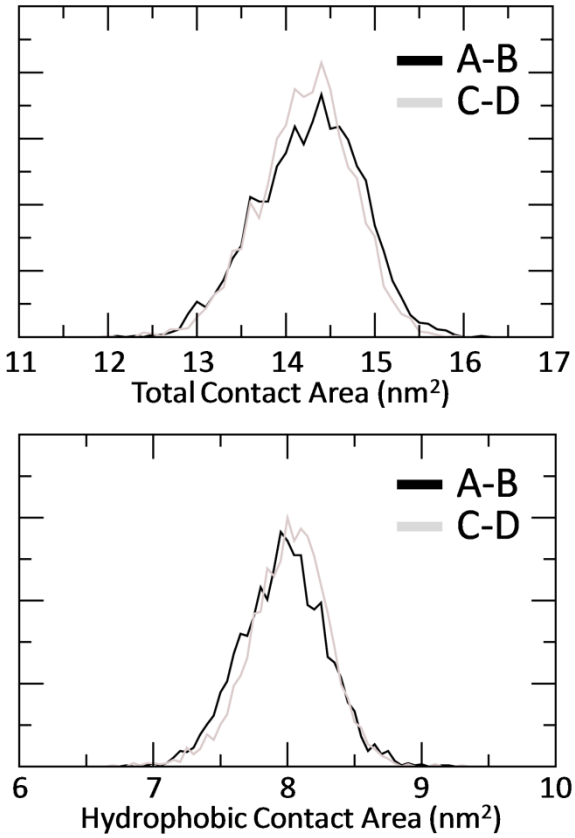


Figure 15. Distribution plot of the monomer-monomer total contact area calculated for each dimer (AB and CD), along the REMD trajectory at 310 K (top). Distribution plot of the monomer-monomer hydrophobic contact area calculated for each dimer (AB and CD), along the REMD trajectory at 310 K (bottom).

The monomers' secondary structure [105] was evaluated by averaging the 1000 configurations taken from the REMD trajectory at 310 K (Figure 16).

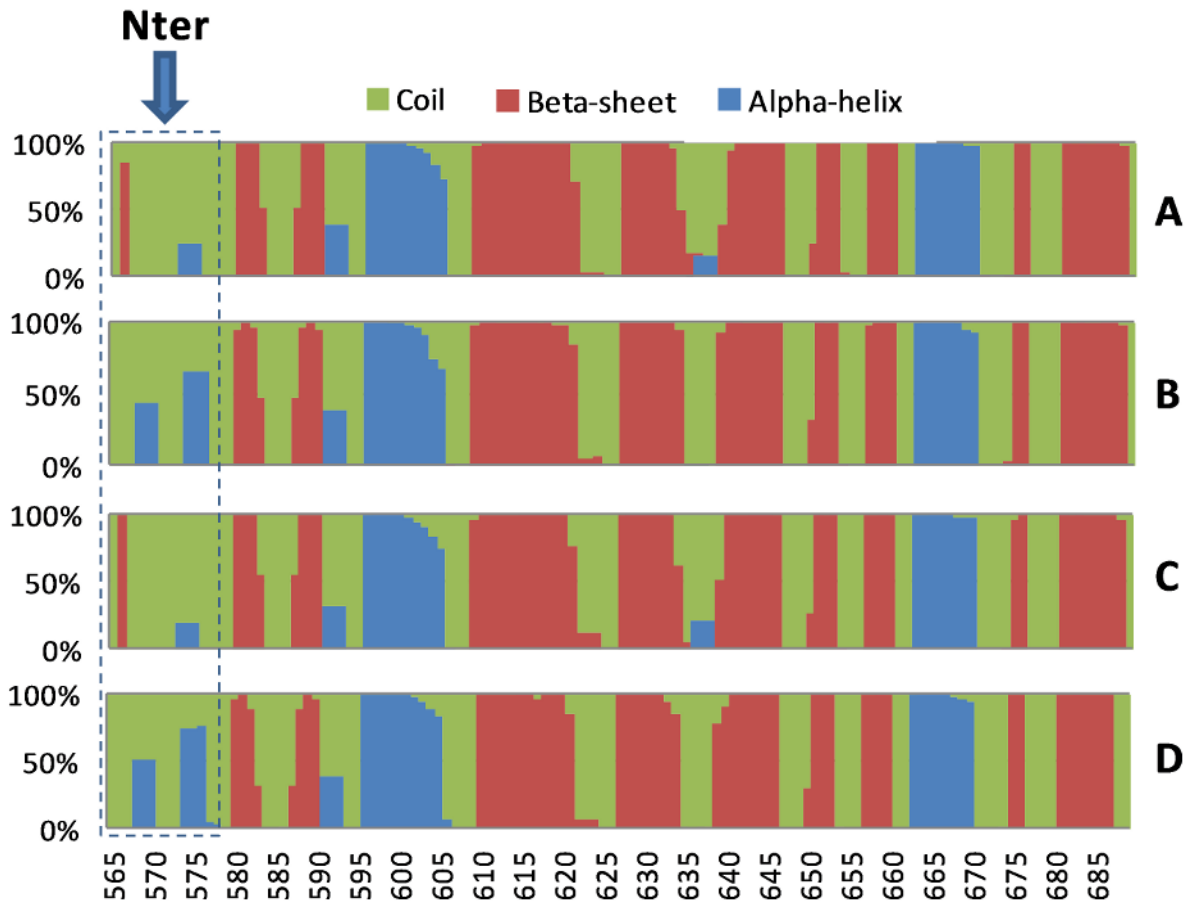


Figure 16. Secondary structure probability calculated for each monomer (A,B,C and D) over the REMD trajectory at 310 K in case of AB and CD simulations. The Nter tail is also highlighted.

It was found that the AXH secondary structure was highly conserved among A, B, C and D monomers, with some exceptions. The main observed differences involve the Nter region, which promotes the asymmetry of the dimerization interface by achieving several different conformations. In agreement with previous investigations [70], it was found that the protomers B-D differ from A-C, forming a structured helix between residues P568-T570 and P573-F575.

The conformational heterogeneity, or plasticity, of the Nter region also resulted in different Root Mean Square Fluctuation (RMSF) values (Figure 17). In particular, monomers A and C exhibited higher RMSF values, corresponding to L577 (0.3 nm), with respect to other monomers (0.15 nm). On the contrary, higher fluctuation values (higher than 0.3 nm) for residue Y574 were found in the case of monomer B and D. Another RMSF difference between monomers was located at K638. The cause of the high fluctuation value (0.45 nm) calculated in the case of monomer A was identified in the partial unfolding of the site when compared to other monomers (Figure 16).

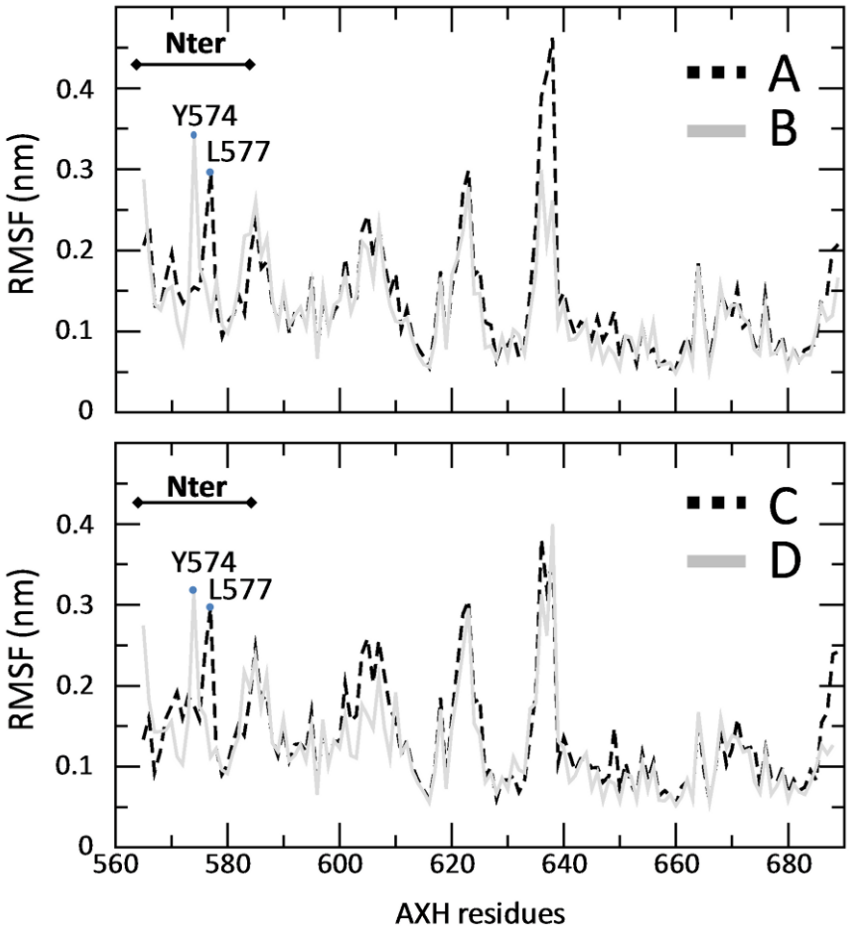


Figure 17. Root Mean Square Fluctuation (RMSF) calculated for each monomer (A, B, C and D) over the REMD trajectory at 310 K in case of AB (top) and CD (bottom). The N-terminal tail is also highlighted in figure.

The monomer-monomer residues' contact probability plot is presented in Figure 18. Notably, the dimerization interface was mainly characterized by the interaction between the Nter and C-terminal tails (Figure 18). Moreover, interacting interfaces were essentially characterized by non-charged residues, in particular I580, the residue most frequently involved in the AB dimerization interface, being part of the contact area over the 95% of the total sampled configurations. Additional residues M566, L571, S642, S685 and S686 were characterized by lower contact probability values (lower than 50%). In the case of the CD dimer, also R665 and L688 were markedly involved in the dimer interface, as demonstrated by contact probability values of 95% and 98%, respectively.

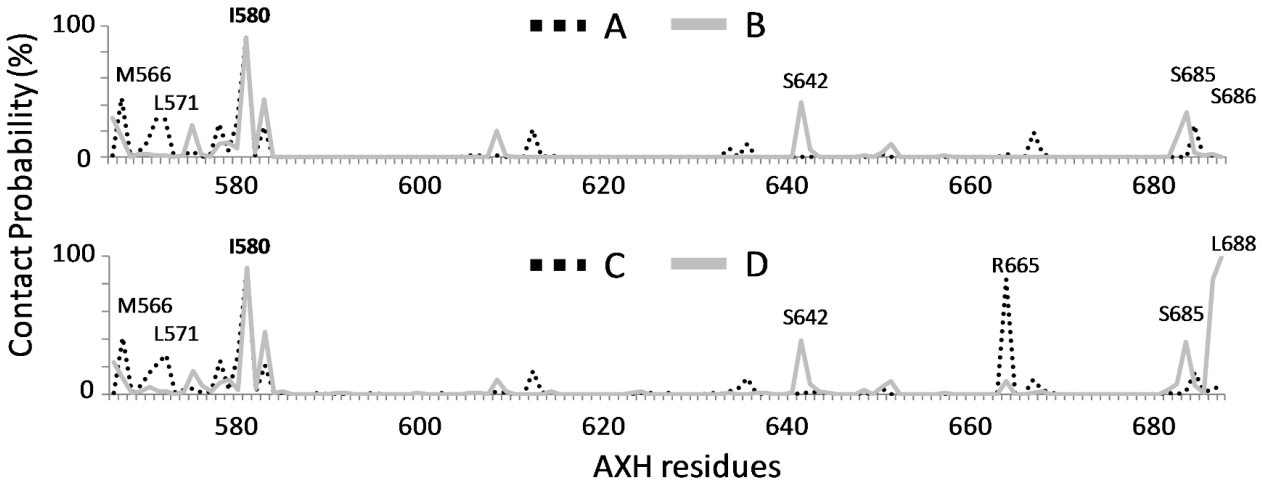


Figure 18. AXH inter-monomer contact probability plot calculated for AB and CD dimers by applying a distance cut-off of 0.2 nm. Residues mainly involved in the monomer-monomer contact area are reported in black.

In view of the already known free energy landscape of the AXH monomer in solution [92], it is interesting to notice how each Nter conformation sampled for AXH monomers (A, B, C and D) by REMD at 310 K falls in the previously estimated free energy landscape.

It is worth mentioning that the free energy landscape of the AXH monomer was calculated in the previous section (Section 3.1) by applying the Boltzmann Inversion procedure to the probability distribution of the projections along first and second Principal Components taken from the REMD trajectory at 310 K for a single monomer in water [92] (Figure 19). Black dots in Figure 19 indicate the position on the free energy landscape [92] in term of first and second Principal Components of AHX monomers extracted from the REMD trajectory at 310 K, where monomers are bound in a dimeric form.

Findings presented in Figure 19 describe how monomer A and C conformational state is influenced by the presence of the other interacting monomer B and D, and *vice versa*. In particular, part of the accessible Nter arrangements previously observed for the AXH monomer [92], were never observed throughout the AXH dimer dynamics. These findings support the hypothesis that the interactions leading to the dimer formation are able to stabilize the Nter region of the AXH, which is involved in the dimeric interface, as recently proposed [70,92].

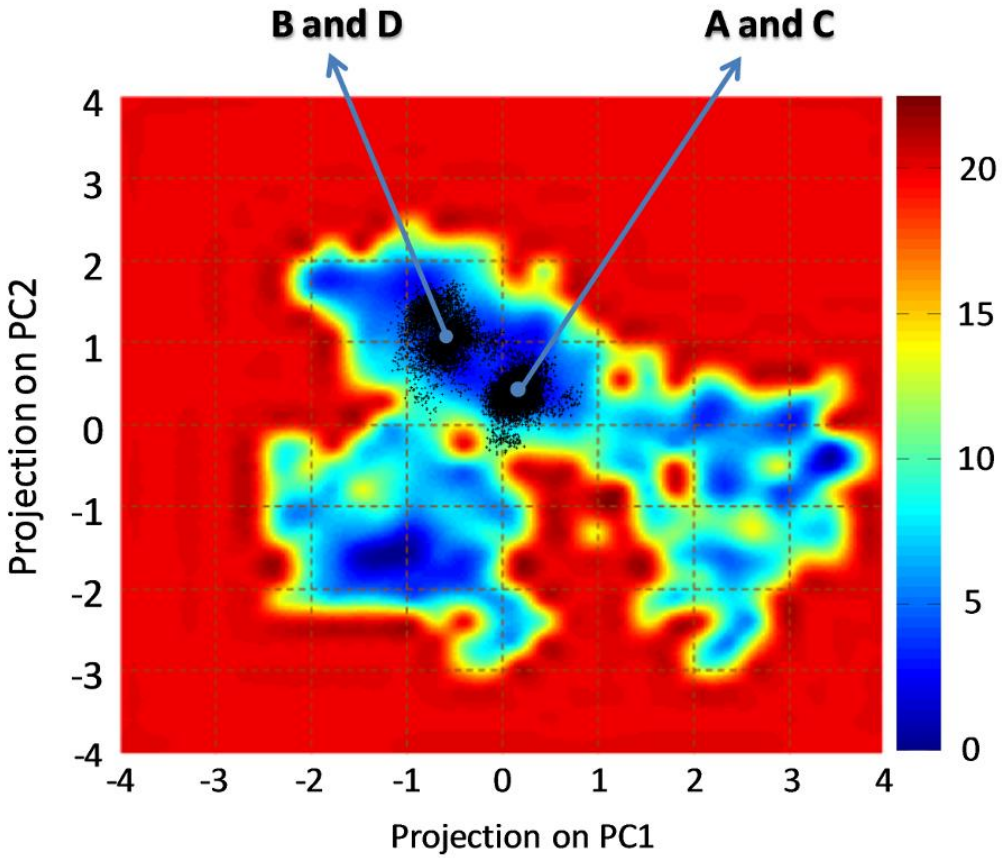


Figure 19. Free energy profile (kJ/mol) of the AXH monomer represented as function of the first and second Principal Component projection previously computed (Section 3.1). Labels indicate the position on the free energy landscape in term of first and second Principal Component projection values of the A, B, C and D monomer conformational states sampled in this work along the REMD trajectory at 310K.

FMA was applied to describe the fluctuation of the inter-monomer distance in terms of internal collective motions of the protein. From FMA it was possible to characterize the contribution of individual PCA vectors to the fluctuations in the monomer-monomer distance, yielding a single vector which drives the monomer-monomer interaction mode, referred to as ewMCM. The analysis of the REMD trajectory at 310 K filtered on the ewMCM, allowed the identification of those residues responsible for the fluctuation of the AB and CD distances (Figure 20).

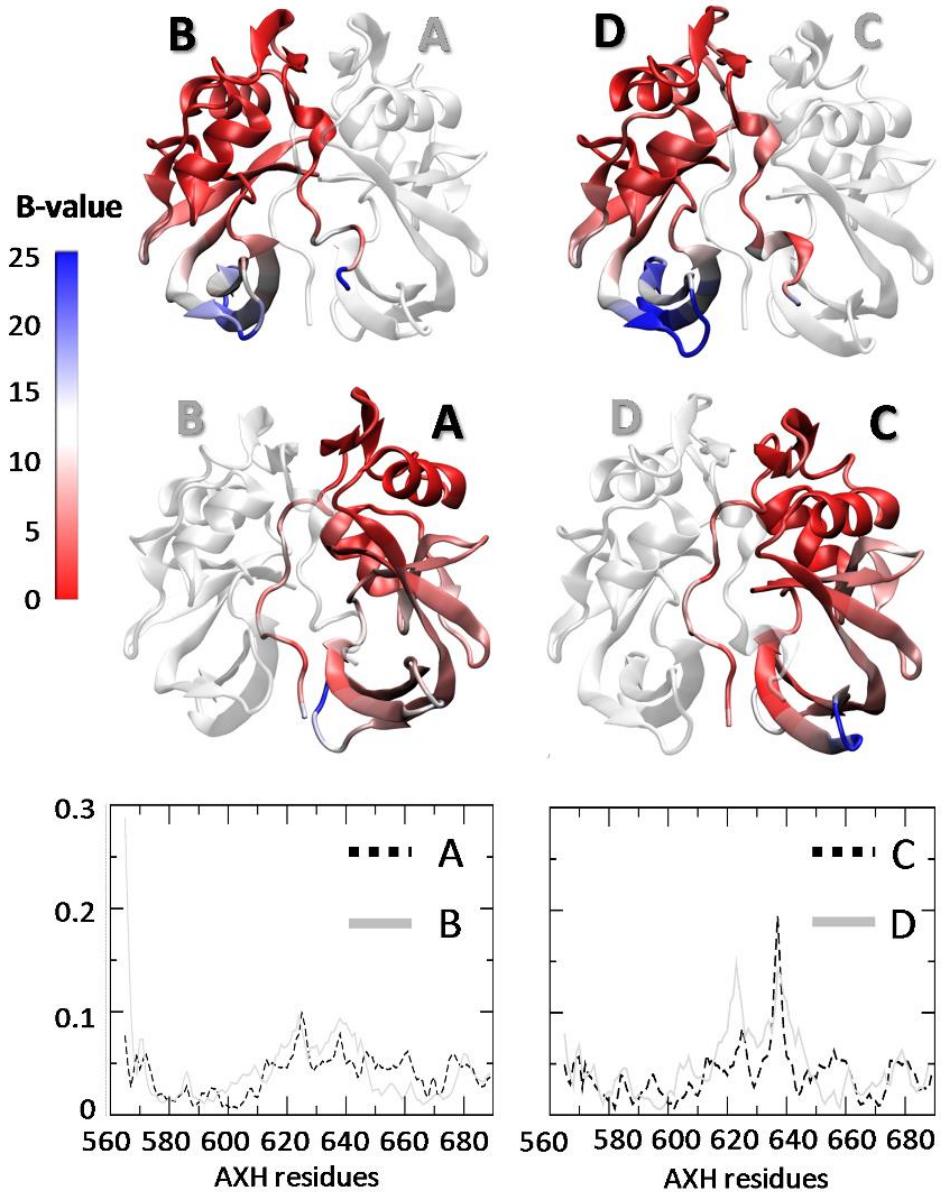


Figure 20. Color map of the B-values reported on the AB and CD dimers (top). The B-values are quantified starting from the RMSF plot calculated over the REMD trajectory at 310 K filtered on the ewMCM (bottom).

In particular, protein domain Q620-P650 was characterized by the highest fluctuation along the filtered trajectory (Figure 20), suggesting a possible key role in affecting monomer-monomer distance. Also, the Nter tail of monomer A showed high RMSF values. Molecular arrangements of dimers AB and CD are shown in Figure 20, where B-values correspond to the RMSF plot mentioned above. The trajectory filtered on the ewMCM is shown in Figure 21. To provide a more explanatory visual inspection of the functional motion that makes the inter-monomers distance (d) fluctuating within the range n x σ_d (where σ_d is the standard deviation of d), the extension of the motion was artificially emphasized (Figure 21, obtained imposing n = 6).

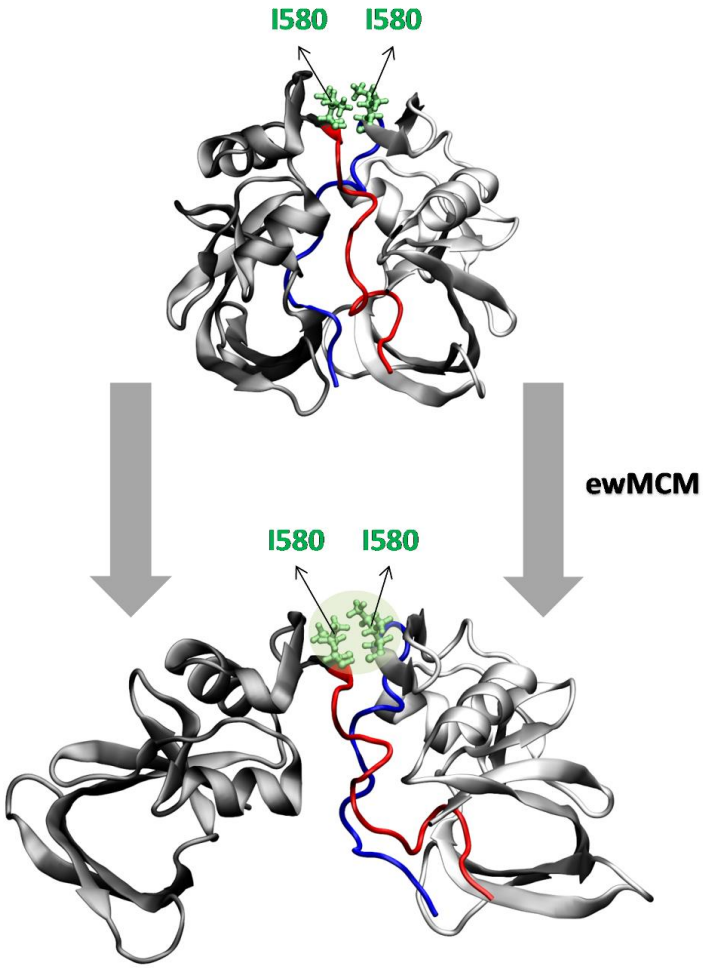


Figure 21. Visual inspection of the essential trajectory performed along the ewMCM. The Nter region is highlighted in red (monomer B) and blue (monomer A). Amino acid I580 is highlighted in green. The extension of the motion has been artificially emphasized.

It can be observed that along the monomer-monomer interaction mode, the Nter tail of one monomer (B/D) remains in contact with the other one (A/C). Most importantly, the monomer-monomer interaction mode is far from a simple rigid translation between monomers, exhibiting instead a roto-translational dynamic around a well-defined fulcrum positioned at the Nter basis, where the I580 residues are positioned. The proposed AXH monomer-monomer binding motion confirms the leading role of I580 in affecting the dimer stability as suggested by recent experimental findings [70,92].

Discussion

The present section is mainly focused on the AXH dimer, which is thought to play a key role in the physiological function of the Ataxin-1, and which has been demonstrated to form amyloid fibers *in vitro* without the presence of any destabilizing conditions [48]. The extraordinary structural plasticity, together with the unusual chameleon folding, have made the AXH domain conformational properties the subject of several studies [44,69,71–74,92].

Starting from the assumption that the equilibrium among monomer, dimer and tetramer species of the AXH represents an important point of the aggregation pathway for Ataxin-1 [91], the main aim of this study is to better identify those inter-monomer interactions which are mainly involved in the AXH dimer stability. Results have been obtained by bridging together two approaches: REMD, which provides an enhanced conformational sampling throughout the AHX dimer dynamics, and FMA, which allows dimensionality reduction of the REMD trajectory, thus revealing the motion mode related to the monomer-monomer binding.

Hence, the main novelty of this work resides in the use of molecular simulations and enhanced sampling methods to provide: *i*) the characterization of the AXH monomer-monomer interface, and *ii*) the molecular description of the AXH monomer-monomer interaction mode.

Supported by very recently estimated AXH monomer free energy landscape [92], the findings of this study have highlighted how several accessible Nter arrangements, achievable for free AHX monomers in solution, are never observed in the dimeric form. Thus, the presented findings (1) support the already suggested [70,92] hypothesis that interactions leading to the dimer formation are able to stabilize the Nter region of the AXH, which is involved in the dimeric interface, and (2) confirm that the protein's Nter region is characterized by conformational heterogeneity in terms of its secondary structure (Figure 16). In particular, it has been observed that the protomers B-D differ from A-C and form a structured helix between residues P568-T570 and P573-F575.

Interacting interfaces are essentially characterized by non-charged residues differently from what was observed, e.g., for another member of the polyQ family, the Ataxin-3 (ATX3) [105,106], where the Josephin Domain (JD) plays a role in modulating the ATX3 fibrillization pathway. In details, recent investigations[105,106] have indicated that the JD-JD binding might be mainly driven by electrostatic interactions involving charged residues such as Arginine. In the case of AXH dimers, I580 has been found most frequently involved in the dimerization interface, as being part of the contact area involving over the 95% of the total sampled configurations (Figure 18).

The key role of I580 is also highlighted by the monomer-monomer interaction mode proposed in this work. The monomer-monomer interaction mode is characterized by a roto-translational dynamic around a well-defined domain including the I580 residue, which acts as a fulcrum. This finding highlights the leading role of I580 in modulating the AXH monomer-monomer interaction, in agreement with a recent experimental study [70], reporting that the population of the monomeric, dimeric and tetrameric species can be strongly modified by a mutation on I580. Specifically, I580 mutations destabilize the dimer structure, by increasing the monomeric species in the AXH population [70]. Together with the above-mentioned previous pieces of evidence, the results of this work confirm the pivotal role of the I580 residue in stabilizing the AXH dimer.

Finally, additional residues M566, L571, S642, R665, S685 and S686 and L688 were identified here as a potentially important regions involved in the AXH inter-monomer interface (Figure 18). In this connection, the results here presented can be considered as a starting point for further *ad hoc* experimental and computational studies aimed at validating the direct involvement of these residues in the AXH dimer stability.

3.3 AXH tetramer: Molecular Mechanisms and Potential Anti-Aggregation Strategies

Introduction

The present in silico study (Section 3.1 and 3.2) together with previous experimental [69,70] findings have highlighted the presence of a well-preserved tertiary and quaternary structure for the AXH domain, apart from the N-terminal (*Nter*) region of each AXH monomer (AXH^m), which is involved in the dimer interface. Experimental studies have also suggested the possibility for the AXH domain to organize itself in tetramer [44,70]. All the previously mentioned tetramer models are very similar in terms of unit cell dimensions, crystal packing, asymmetric unit content, and quaternary and monomeric backbone structures, as demonstrated in recent literature [70]. A number of AXH mutations have been also experimentally tested to identify residues responsible for the destabilization of the monomer/dimer/tetramer equilibrium [70]. In detail, the tetramer population was noticeably reduced by residue I580 mutation, thus identifying the above mentioned residue as a key player in mediating the AXH self-association [70].

However, the AXH self-association mechanism is not yet clarified and several questions remain open. For example, it is unclear how the residue I580, located in the monomer-monomer interface (intra-dimer), is able to drive the tetramer assembly [70]. Unanswered questions like this make of great interest the investigation of non-local molecular effects consequence of single point mutations, and closely related to modifications in AXH self-assembly tendency.

In this study, molecular modeling was applied to fully characterize structural and energetics of the AXH tetramer assembly. Classical and enhanced molecular dynamics (MD) have allowed to elucidate those non-local molecular events triggered by a protein mutation which lately affect the conformational feature of the whole AHX tetramer assembly. Moreover, first in literature, the free energy landscape associated to the AXH tetramer formation process was estimated in presence of amino acid substitutions. Such a picture of the free energy landscape is crucial for a clear understanding of the molecular processes driving protein assembly and aggregation.

Our data confirm the key role of I580 and suggest R638 as a further residue mainly determining AXH tetramer stability. Outcome of the present work may be of interest for both (1) a better understanding of the self-association mechanism characterizing AXH dimers, and (2) identifying novel modulators of the AXH protein-protein interaction predicted in the present work as important for AXH tetramer stability.

Materials and Methods

Classical Molecular Dynamics on AHX Dimers

The AXH dimer model was extracted from the 4APT [70] model, as in previous studies [92,107]. For consistency with previous literature, residue numbering refers to the UniProtKB/Swiss-Prot entry No. P54253 +1. Here, two different AXH dimer molecular systems were investigated, the wild type AXH dimer (AXHd-WT simulation) and the mutated AXH dimer (AXHd-I580A). Based on recent experimental evidences [74], the alanine mutation on residue I580 was considered. Simulations were carried out by applying the computational scheme described below.

Each AXH dimer was placed in a dodecahedron box where the minimum distance between the protein and the edge of the box was set to 1.2 nm. The box was then fully solvated in explicitly modeled water and the total charge was neutralized by the addition of Cl− and Na+ ions. Each dimer system in water consisted of about 40,000 interacting particles. Following an initial energy minimization of 1000 steps of steepest descent, two preliminary position restraint MD simulations (each one of the duration of 100 ps) were carried out in NVT and NPT ensemble, respectively, where the heavy atoms of the proteins were restrained using a force constant of 1000 kJ $mol^{-1}nm^{-2}$. During the first restrained MD simulation, involving MD in the NVT ensemble, protein and non-protein atoms were coupled separately to temperature baths using a *v-rescale* coupling algorithm [82] with a coupling time of 1 ps. Subsequently, the second restrained MD simulation was performed keeping the pressure at 1 bar by applying Berendsen's weak coupling method [76] with a time constant of 5 ps. Finally, 600 ns of production MD was simulated in NPT ensemble. The *v-rescale* coupling algorithm [82] was applied again to maintain the system's temperature at 310 K (time constant of τ_T=0.1 ps) and the pressure was maintained at 1 bar using the Parrinello-Rahman [108,109] barostat (time constant of τ_P=2 ps) in the isobaric-isochoric ensemble with long-range dispersion correction applied for both the energy and pressure terms. AMBER99-ILDN force-field [77,78] and TIP3P model [80] were employed to define the protein topology [77,79] and water molecules, respectively. Gromacs-5.1.2 [110] with the PLUMEDv2.3 patch [111] were employed for MD simulations and data analysis [4]. Visual Molecular Dynamics (VMD) [112] package was employed for the visual inspection of the simulated systems. The computational data taken from the Classical MD trajectory were analyzed to characterize the AXH dimer-dimer interface. Principal Component Analysis (PCA) was applied to reduce the dimensionality of the system, elucidating large-scale and low-frequency modes, respectively, thus yielding collective motions directly related to a specific molecular event [89]. After the alignment of the AXH^m C-alpha (C_α) Cartesian coordinates, the covariance matrix was calculated and diagonalized.

AHX tetramers free energy landscape estimation

Also, in this case, two different AXH tetramer molecular systems (constituted by monomer A, B, C and D) were investigated, the wild type AXH tetramer (AXHt-WT simulation) and the mutated AXH tetramer (AXHt-I580A). The AXH tetramer structure (PDB code 4APT[70]) was fully solvated in a dodecahedron box where the minimum distance between the protein and the edge of the box was set up to 2.5 nm. The total charge was neutralized by the addition of Cl− and Na+ ions. The resulting system placed in water consisted of about 220,000 interacting particles. Classical 300 ns MD equilibration were carried out on both systems, applying the same scheme as for the 4APT dimer models.

Starting from the equilibrated system obtained by classical MD simulation, Metadynamics was applied for the estimation of the dimer-dimer binding free energy. Metadynamics [113,114] is a powerful technique able to enhance sampling in MD simulations by adding a history dependent bias potential on the subspace identified by a set of user-defined Collective Variables (CVs). In detail, Gaussian deposition rate of 0.2 kJ/mol·ps was initially applied on the AXH inter-dimer distance and gradually decreased on the basis of an adaptive scheme. A Gaussian width of 0.2 Å was used following a well-established scheme [15,115]. In particular, the Gaussian width value was of the same order of magnitude as the standard deviation of the distance CV, calculated during unbiased simulations. Each system was simulated (2 fs time step) until convergence was reached. The reconstruction of the free-energy surface was performed by the reweighting algorithm procedure [28], allowing the estimation of the free energy landscape as a function of two CVs : a) distance between dimer AB and dimer CD, and b) the cosine of the angle between vectors connecting the center of mass of two distinct AXH regions per dimer (identified as "*cosine CV*" in the following). Exhaustive details on the definition of the CVs and the convergence of the Metadynamics simulations, are presented in Appendix A-3.

Results

Wild type and Mutated AXH Molecular Dynamics: Dimer Level

Protein structural stability of systems AXHd-WT and AXHd-I580A was analyzed by monitoring the time evolution of the Root Mean Square Deviation (RMSD). It can be observed that protein conformational stability was satisfactorily reached in the last 100 ns of the simulations (Appendix A-3). The dimerization interface in both AXH-WT and AXH-I580A simulations buried 14.5 nm^2 of solvent accessible surface, in agreement with previous literature [107]. The application of PCA highlighted the large-scale and low-frequency modes mainly related to a specific protein

conformational change. After the alignment of the AXH C-alpha atoms, the covariance matrix was calculated and diagonalized for each system (AXHd-WT and AXHd-I580A).

The MD trajectory of both AXHd-WT and AXHd-I580A system was filtered along the first and second PCA eigenvector. The Root Mean Square Fluctuation was calculated over the filtered MD trajectory (Figure 22a, AXHd-WT system; Figure 22b, AXHd-I580A system) to highlight fluctuations related to major motion modes of the dimer.

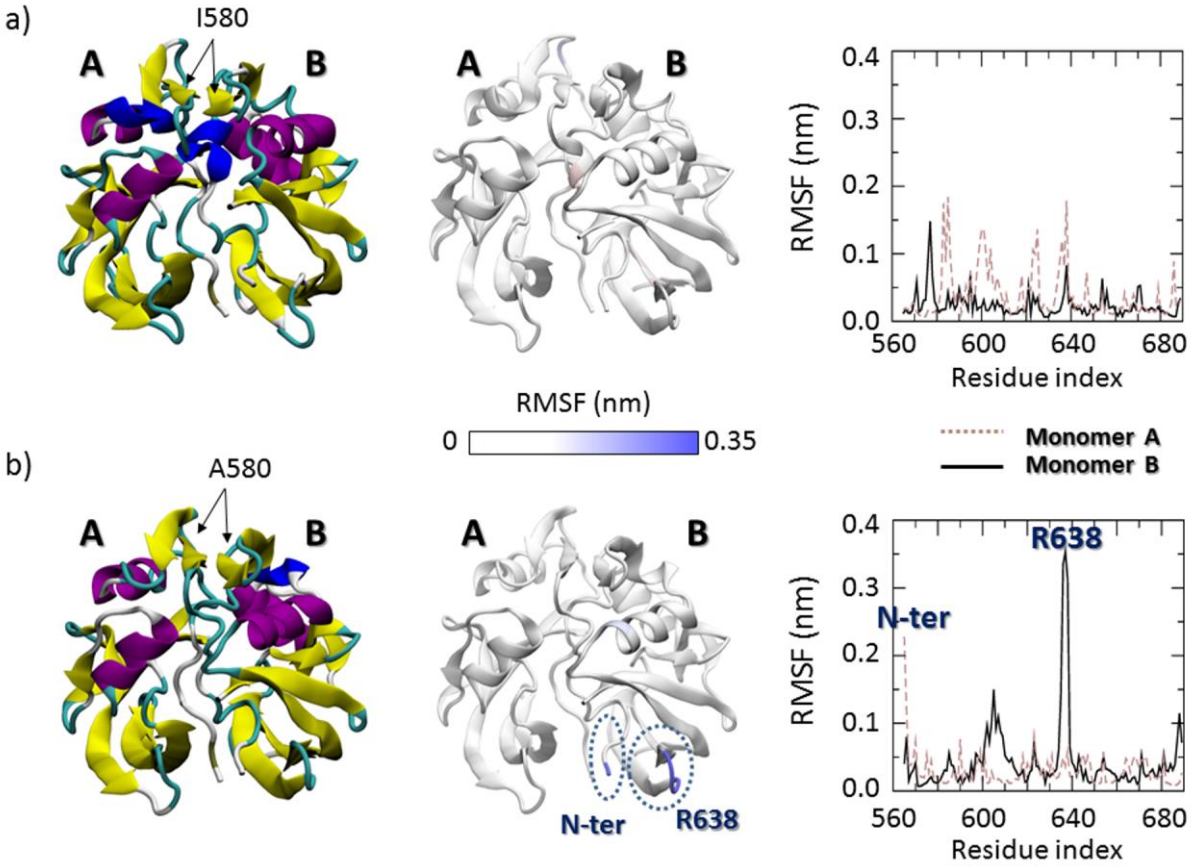


Figure 22. a) Visual inspection of the AXH-WT conformational arrangement at the equilibrium (left). The Root Mean Square Fluctuation (RMSF) is reported in the right panel and represented on the protein 3D structure (center). b) Visual inspection of the AXH-I580A conformational arrangement at the equilibrium (left). The Root Mean Square Fluctuation (RMSF) is reported in the right panel and represented on the protein 3D structure (center).

By comparing the RMSF curves (Figure 22a and Figure 22b, right), it can be noticed that the I580 mutation may lead to a destabilization of the AXH N-terminal tail (AXH d-I580A $RMSF_{Nter}$=0.23 nm, Figure 22b, pink dashed line) on one monomer (e.g., monomer A in Figure 22b). Notably, the RMSF profile of the interacting monomer (Figure 22b, monomer B), is instead characterized by a relevant fluctuation peak of residue R638 (AXHd-I580A $RMSF_{R638}$= ~ 0.35 nm). Interestingly, the R638 on monomer B is spatially close to the N-terminal tail of monomer A and relatively far from residue I580 (Figure 22). These findings suggest a non-local correlation between residue I580 in one monomer and residue R638 in the interacting one, mediated by the N-terminal tail. In a greater detail, the increased

fluctuation of the N-terminal tail, as induced by I580 mutation, drives a conformational modification of residue R638 on the interacting monomer. Interestingly R638 participates to dimer-dimer interaction in AXH tetramerization. Therefore, it is meaningful pointing the attention on the conformational behavior of residue R638 in AXH tetramer dynamics as done in the following.

Wild type and Mutated AXH Molecular Dynamics: Tetramer level

Using the same protein model (PDB code 4APT[70]), two different AXH tetramer systems were considered, the wild type AXH tetramer (AXHt-WT simulation) and the mutated AXH tetramer (AXHt-I580A). Classical 300 ns long MD equilibration was performed on both systems. The two AXH tetramer systems are initially characterized by the same interdimer contact regions: T596-I600, V653-T658 and G680-K689 of monomer B and T596-T615, G635-Q640, V653-T658 and G680-K689 of monomer C. Dimer-dimer conformational arrangement is quite conserved during the AXHt-WT simulation, as also demonstrated by the stable RMSD plot observed from 50 to 300 ns (Figure 23a, top). The previously mentioned arrangement is characterized by residues R638B and R638C non-covalently bound by two hydrogen bonds between the guanidinium group on one residue and the backbone carbonyl oxygen on the other one (Figure 23b-I). In the AXHt-I580 system hydrogen bonds are lost, weakening the interdimer interaction (Figure 23b-II).

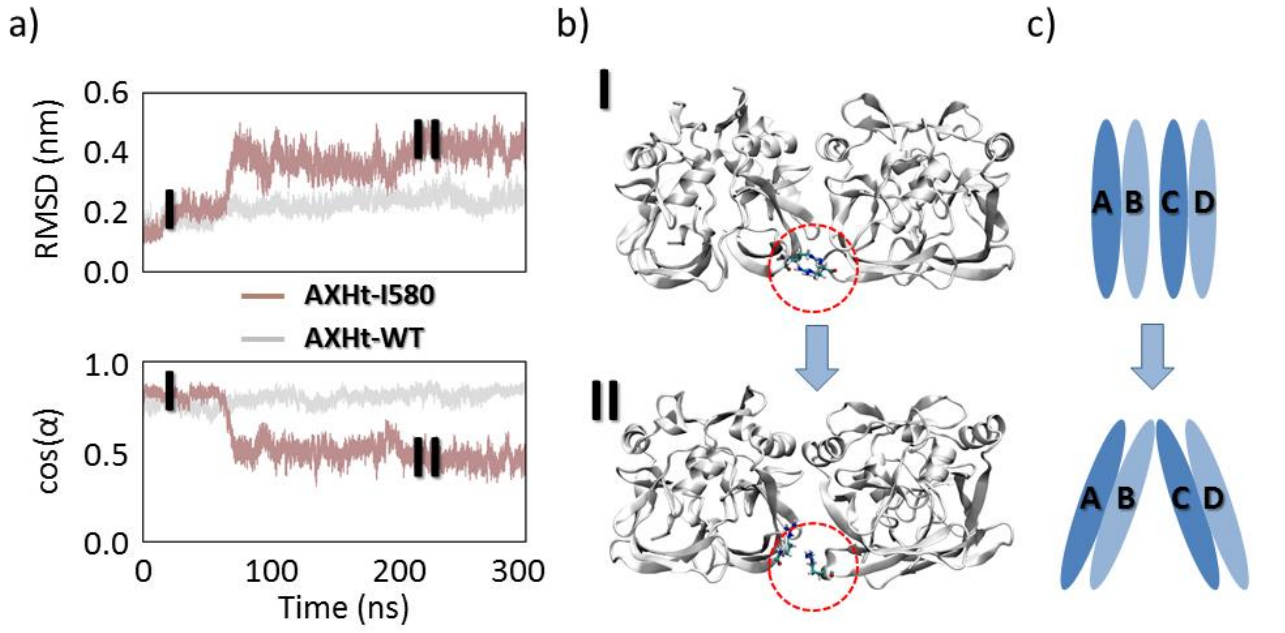


Figure 23. Backbone RMSD throughout the overall MD simulation for wild type (grey) and mutated protein (brown). b) Visual inspection of the molecular systems at the equilibrium in case of AXHt-WT (I) and AXHt-I580A system (II). c) Scheme illustrating the main structural modifications characterized by a relative rotation between dimers AB and CD.

This result is also confirmed by the Radial Distribution Function (RDF) of the guanidinium group amine nitrogen atoms with respect the backbone carbonyl oxygens of the interacting arginine (Figure 24a). The peak at 2.8 Å characterizing the AXHt-WT trajectory (Figure 24a, black curve) does indicate the presence of a strong primary interaction (hydrogen bond) between R638B and R638C. In summary, the I580 mutation destabilizes the protein region containing the R638 residues, as already observed (Figure 22), thus leading to a less stable interaction between R638B and R638C, as highlighted by the RDF analysis (Figure 24a, brown curve). Consequently, it was observed a relative rotation between dimers AB and CD in the AXHt-I580A, inducing a marked decrease of dimer-dimer contact area, when compared to AXHt-WT ($AXHt\text{-}WT_{buried}$=14.7±0.8 nm^2; $AXHt\text{-}I580A_{buried}$=11.9±0.5 nm^2). The rotation between dimers AB and CD can be described computing the cosine of the angle between vectors connecting the center of mass of two distinct AXH regions per dimer (Figure 23a, bottom).

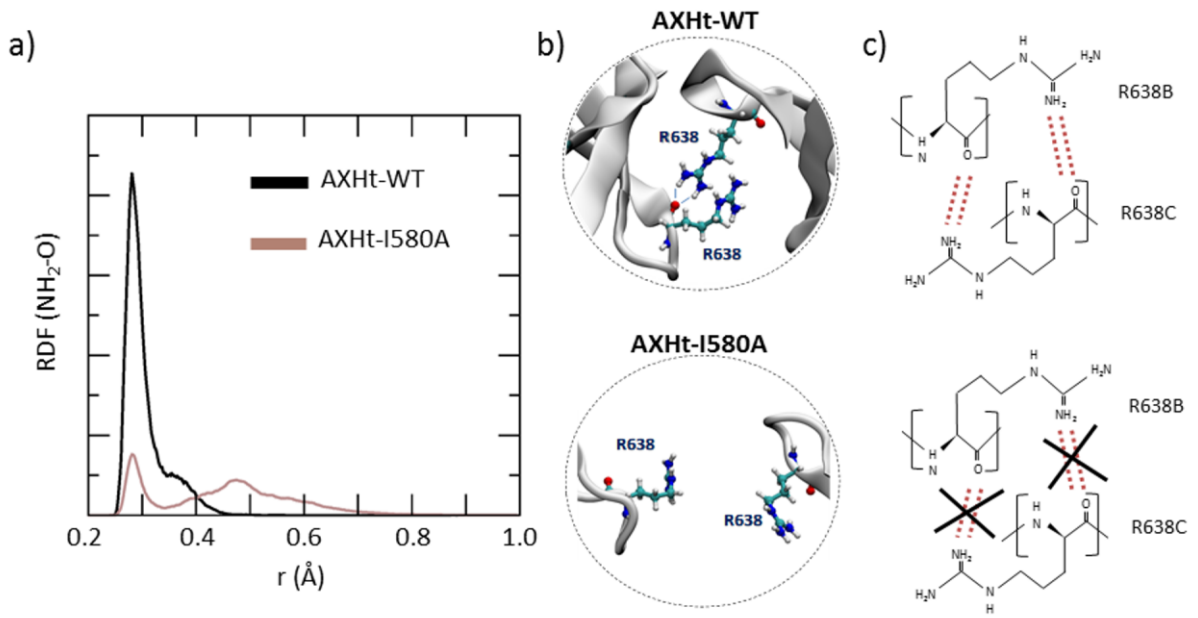


Figure 24. (a) RDF of the amine nitrogen atoms of guanidinium group with respect the backbone carbonyl oxygens of the interacting arginine residue. Only R638B and R638C are considered for the calculation. Two different peaks at 2.8 Å and 4.7 Å can be identified. The first peak is related to the primary interaction resulting from the hydrogen bond between the amine hydrogen atoms and the backbone oxygen, while the second peak indicates the water-mediated hydrogen bonding [116]. The hydrogen bonds between the R638B and R638C residues are also represented in (b) and (c).

Wild type and mutated AXH tetramer free energy landscape and unbinding pathway

Starting from the equilibrated systems obtained by classical MD simulations, the process of dimer-dimer unbinding was investigated by metadynamics. An overall picture of the dimer-dimer free energy landscape of AXHt-WT and AXHt-I580A systems is provided in Figure 25, where the free energy

surface was expressed as a function of two collective variables: the AXH dimer-dimer distance, and the cosine CV.

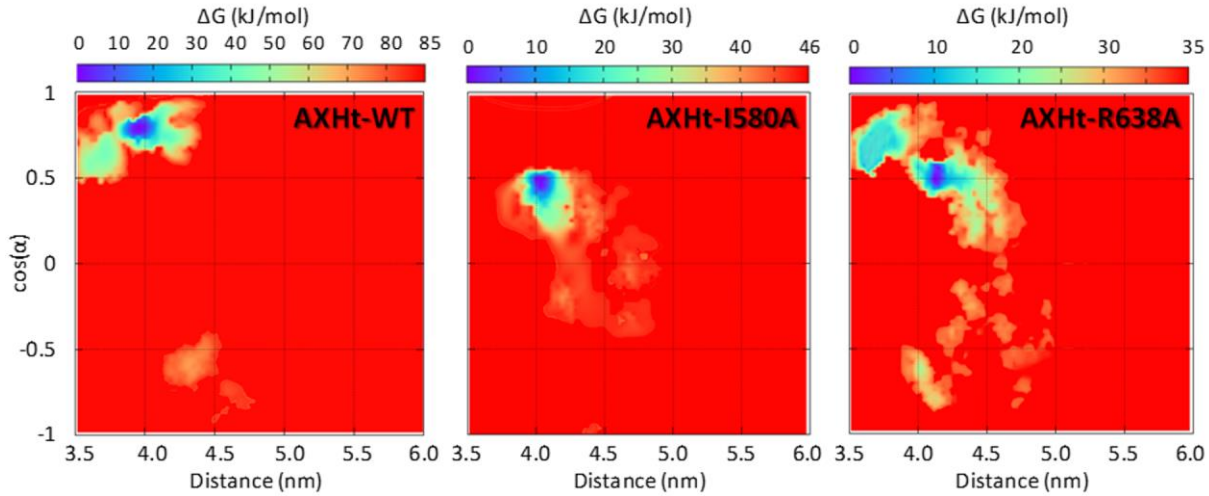


Figure 25. a) Free energy profile (kJ/mol) of the AXHt-WT, AXHt-I580A, and AXHt-R638A molecular systems (b), represented as function of two collective variables: the AXH dimer-dimer distance and the cosine of the angle between vectors connecting two distinct AXH regions per dimer.

A first difference between the two systems concerns the dimer-dimer binding free energy. In detail, AXHt-WT is characterized by the highest binding affinity ($\Delta G_{AXHt\text{-}WT}$=85 kJ/mol; $\Delta G_{AXHt\text{-}I580A}$=46 kJ/mol, Table 2).

Table 2. Tetramer binding affinity values estimated in the present work and compared with a recent experimental work reporting the experimental tetramer percentage in solution [70].

Molecular System	Tetramer ΔG [kJ/mol]	Exp.Tetramer %
AXH-WT	85 ± 3	26 % [70]
AXH-I580A	46 ± 2	8 % [70]
AXH-R638A	35 ± 3	//

Therefore, the I580 alanine substitution results in a lower energetic cost for the dimer-dimer unbinding. This interesting result is in agreement with recent experimental findings [70] demonstrating that the I580 mutation leads to a significant reduction of AXH tetramer percentage in solution (Table 2). Moreover, a difference between the first binding pose can be noticed comparing the cosine CV value

corresponding to the deepest energy well in case of AXHt-WT ($\cos(\alpha)_{WT}$=0.8) and AXHt-I580A ($\cos(\alpha)_{I580A}$=0.5). This result, in agreement with Classical MD outcome (Figure 23, Figure 24), indicates a relative rotation between dimers AB and CD as a consequence of I580 mutation, lowering the energetic cost of the dimer-dimer unbinding (Table 2).

In agreement with a recent study [70], there is evidence that the I580 mutation affects the dimer-dimer free energy landscape in terms of binding affinity and binding pose. The results of this study let suppose this mechanism significantly driven by the destabilization of the R638 residue. To further test this hypothesis, a classical MD and metadynamics simulation was performed by replacing the native R638 residues with Alanine (AXHt-R638A), instead of mutating the I580. The results, reported in Figure 25, clearly show that Arginine substitution of R638 dramatically reduces the binding affinity ($\Delta G_{AXHt\text{-}R638A}$=35 kJ/mol) if compared with AXHt-WT and AXHt-I580A (Table 2). Moreover, the comparison between the cosine CV value corresponding to the deepest energy well in case of AXHt-I580A and AXHt-R638A simulation ($\cos(\alpha)_{I580A}$=$\cos(\alpha)_{R638A}$=0.5) demonstrate that both molecular systems are characterized by similar binding poses (Figure 25). This outcome provides a convincing evidence on the importance of R638 as the key player in AXH tetramer stability.

Discussion

The present study is mainly focused on the protein-protein interactions driving the AXH tetramer stability, which is thought to play a role in the self-association process of full-length ATX1 and amyloid fibril formation [48]. Previous experimental studies have demonstrated the presence of monomers, dimers, and tetramers in solution [70]. Moreover, the structural consequence of selected mutations, designed to disrupt specific intermolecular interfaces, was analyzed in order to identify residues, which are mainly responsible for determining the monomer/dimer/tetramer equilibrium [70]. Experimental studies have shown that tetramer population was noticeably reduced by I580 mutation, demonstrating that I580 plays an important role in mediating the AXH self-association [70]. However, it was unclear how a centrally located residue in the monomer-monomer interface can affect the tetramer formation mechanism. In this regard, the lack of clear knowledge of the AXH species present in solution makes experimental, more quantitative analysis (i.e., an estimate of the dissociation constants) of the systems very challenging [70]. In this connection, computational approaches have demonstrated to be effective in predicting physical-chemical properties of biomolecular systems [97,99,106,117–127], in particular the effect of amino acid mutations on protein-protein [105,107,128] and protein-substrate [100,129] interactions, allowing the free energy estimation corresponding to individual residue substitutions.

In the present work classical MD and metadynamics were applied to investigate the structural and dynamic consequences of the I580 mutation, by applying the Principal Component Analysis (PCA)

which has already demonstrated to be able to elucidate the large-scale and low-frequency modes mainly related to the AXH conformational changes in monomeric [92] and dimeric [107] form. Here, it was demonstrated that I580 mutation on AXH dimer leads to the destabilization of the N-terminal tail. Consequently, the residue R638, spatially close to the fluctuating N-terminal tail, is characterized by a high degree of conformational flexibility (Figure 22). This is an important molecular event, considering that the MD simulation of the wild-type AXH tetramer has demonstrated the non-covalent interaction between the guanidinium groups and the backbone carbonyl oxygens of residues R638B and R638C (Figure 24). Moreover, a less stable interaction between R638B and R638C was observed in case of I580 mutation on AXH tetramer, suggesting a possible explanation for the role played by a centrally located residue in the monomer-monomer interface in reducing the AXH tetramer stability. The well-established observation that the AXH domain of ATX1 adopts different topologies leading to the asymmetry of monomer-monomer interface [44,69,70,91,92,107], let suppose that the overall tetramer topology may be dependent on the structural fluctuation of the AXH dimer caused by single-point mutation.

To provide further insight into this issue, the metadynamics approach was employed to estimate the free energy landscape governing the tetramer formation process in case of wild-type and mutated AXH tetramer system. As a result, the I580 alanine substitution affected the free energy landscape, lowering the energetic cost for the dimer-dimer unbinding. Those findings are in agreement with experimental observations [70] that I580 mutation leads to a significant reduction of AXH tetramer percentage in solution (Table 2). More interestingly, the simulations performed in this study shed light on the importance that the hydrogen bond between R638B and R638C has in driving the dimer-dimer unbinding event. The effect of R638 conformational instability because of I580 protein mutation leads to a different protein-protein detachment pathway, lowering the energetic cost of the dimer-dimer unbinding (Figure 25). A proof of evidence for this mechanism was provided by the dramatic reduction of binding affinity caused by the replacement of R638 residues with Alanine, and hence demonstrating the importance of R638 residue in driving the AXH tetramer stability (Figure 25).

Taken together, these findings suggest that residue R638 plays a key role in AXH tetramer stability, and candidates to be the target for future experimental designs of ATX1 mutations able to reduce the phenomenon of self-association of full-length ATX1, that leads to amyloid fibril formation and, ultimately, to neurodegenerative disease.

Chapter 4

Structural Polymorphism of Alzheimer's Beta Amyloid Fibrils

Alzheimer's disease is the most fatal neurodegenerative disorder characterized by the aggregation and deposition of Amyloid beta (Aβ) fibrils in the brain of patients. Two principal variants of Aβ exist in human: $Aβ_{1-40}$ and $Aβ_{1-42}$. The former is the most abundant in the plaques, while the latter is the most toxic species and forms fibrils more rapidly. Interestingly, fibrils of $Aβ_{1-40}$ peptides can only assume U-shaped conformations while $Aβ_{1-42}$ can also arrange as S-shaped three-stranded chains, as recently discovered. As alterations in protein conformational arrangement correlate with cell toxicity and speed of disease progression, it is important to characterize, at molecular level, the conformational dynamics of amyloid fibrils. Within this framework, the higher toxicity of $Aβ_{1–42}$ species compared to $Aβ_{1–40}$ may be explained by their ability to form a more stable S-shaped assembly. The Section 4.1 of this chapter is focused on the conformational dynamics of U-shaped and S-shaped Aβ small fibrils. The computational results provide evidences of the stability of the recently proposed S-shaped model due to the to the maximized interactions involving the C-terminal residues, suggesting that molecular architecture of the protein aggregates might play a pivotal role in formation and conformational stability of the resulting fibrils.

The Section 4.2 shed light on the mechanical performance of the U-shaped and S-shaped Aβ fibrils. The computational study here presented highlights the superior mechanical behaviour of the S-architecture, characterized by a Young's modulus markedly higher than the U-shaped architecture. The S-architecture showed a higher mechanical resistance to the enforced deformation along the fibril axis, consequence of a better interchain hydrogen bonds' distribution. This study suggests the S-shaped $Aβ_{1-42}$ species as a target of election in computational screen/design/optimization of effective aggregation inhibitors.

(i) G. Grasso, M. Rebella, S. Muscat, U. Morbiducci, J. Tuszynski, A. Danani, M. Deriu, Conformational Dynamics and Stability of U-Shaped and S-Shaped Amyloid β Assemblies, Int. J. Mol. Sci. 19 (2018) 571. doi:10.3390/ijms19020571.

4.1 Elucidating the Conformational Stability of U- and S-Shaped Amyloid Beta Fibrils

Introduction

The major components of AD-associated amyloid plaques are $A\beta_{1-40}$ peptides but also the more toxic $A\beta_{1-42}$ species [53], characterized by two additional amino acids and generated through a sequential cleavage of the amyloid precursor protein (APP) by β and γ secretases [54]. In general, these peptides are able to oligomerize and then the resulting oligomers can further aggregate giving rise to ordered fibrils and fibers [55]. Several experimental studies have been focused on the molecular characterization of amyloid fibrils, given the intimate relationship between molecular structure and disease onset and severity [56]. At present, all the $A\beta_{1-40}$ species resolved by NMR, share a U-shaped motif, where the peptide chains form two β-strands connected by a loop region [57–61]. In case of more toxic $A\beta_{1-42}$ species, earlier NMR models exhibited the same U-shaped motif [58]. The above mentioned molecular assembly is constituted by two β strands (involving residues V18-S26 and I31-A40) connected by a central loop region and stabilized by interchain' H-bonds and salt bridges between residues D23–K28 [58]. More recent investigations demonstrated the possibility of S-shaped arrangements [62–67], characterized by three β strands: the N-terminal strand β1 made of residues V12–V18, the central strand β2 of residues V24–G33, and the C-terminal strand β3 of residues V36–V40. The three β strands are connected by major coil and turn regions. Interestingly, the S-shaped arrangement is not stable in case of $A\beta_{1-40}$ species [68]. Initially, this was explained through the intrachain salt bridge linking the side chain of K28 with the main chain of A42, which does not exist in $A\beta_{1-40}$ [63]. Recently it has been proposed that the cause is the lack of hydrophobic contacts in $A\beta_{1-40}$ generated by the C-terminal residues I41 and A42 in $A\beta_{1-42}$ peptides [130]. Within this framework, the higher toxicity of $A\beta_{1-42}$ species compared to $A\beta_{1-40}$ may be explained by their ability to form S-shaped assembly. Such a correlation could arise if the S-shaped model *i)* was characterized by a more stable molecular architecture *per se*, or *ii)* was able to assemble into structures that are not possible by considering the U-shaped Aβ chains, as recently suggested [68]. In this connection, a molecular level understanding of the interactions governing the structural arrangement in $A\beta_{1-42}$ species represent an important research advance. Computational approaches such as Replica Exchange Molecular Dynamics (REMD) can be used as a powerful tool to elucidate the molecular mechanisms responsible for protein hierarchical organization. In fact, computer simulations have been widely demonstrated to be helpful in capturing mechanisms of protein folding [92,93,131] and protein-protein aggregation [132,133]. Recent computational works investigated the stability of the U-Shaped fibril models of $A\beta_{1-42}$ and $A\beta_{1-40}$ species [134,135]. Those studies highlighted the importance of inter-sheet side chain contacts, hydrophobic contacts among the strands and salt bridges in stabilizing U-shaped protein aggregates [134]. A further

development of the above-mentioned studies might be an investigation of U-shaped and S-shaped assemblies, with the aim of comparing the structural stability and dynamics. Here, REMD was carried out to yield novel insights into the above-mentioned issue by providing a detailed conformational study of S-shaped and U-shaped $A\beta_{17-42}$ pentamer fibril models. Several differences have been found, which clearly highlighted the S-shaped fibril as the most stable architecture due to a maximization of inter-chain hydrophobic contacts and H-bonds involving the C-terminal residues I41 and A42, in agreement with previously published reports [134]. Moreover, data concerning the U-shaped model indicated non-negligible distortions and a tendency to arrange in a more disordered fashion with respect to the S-shaped assembly.

Material and methods

Replica exchange molecular dynamics (REMD)

Two different models () for the $A\beta_{1-42}$ species were considered: the U-shaped $A\beta_{17-42}$ (PDB ID: 2BEG [58]) and the recently resolved S-shaped $A\beta_{11-42}$ (PDB ID: 2MXU [63]).

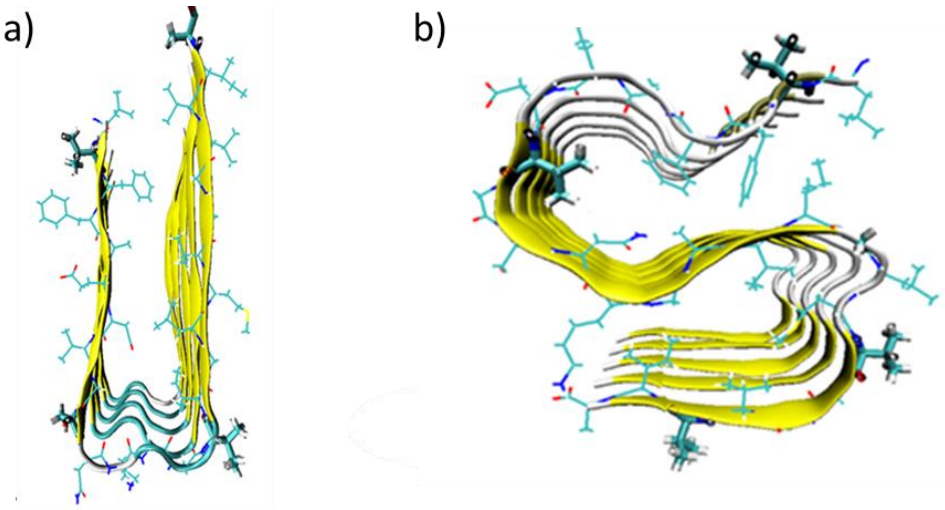


Figure 26. A) U-shaped and B) S-shaped models

A pentamer of $A\beta_{17-42}$ was extracted from each PDB structure. In the manuscript the considered pentamers are called, the U-shaped model ($A\beta_{17-42}$ pentamer extracted from 2BEG.pdb [58]) and the S-shaped model ($A\beta_{17-42}$ pentamer extracted from 2MXU.pdb [63]).

The U-shaped and the S-shaped models were solvated in a cubic box with each side equal to 6 nm and neutralized by counterions. Each system consisted of about 21,000 interacting particles.

The AMBER99SB-ILDN force field [78] was used to define protein topologies and the TIP3P model [80] was used to represent the water molecules. The systems were first minimized by applying the steepest descent energy minimization algorithm, followed by preliminary simulation in NVT ensemble (constant Number, Volume, and Temperature) of 50 ps duration. V-rescale thermostat was applied to keep temperature at 300 K with a time constant of 0.1 ps [82]. An additional simulation in NPT ensemble (constant Number, Pressure, and Temperature) of 50 ps duration was carried out at 300 K (τ=1 ps) and 1 atm (τ=5 ps). V-rescale [82] and Berendsen [76] coupling methods were used as temperature and pressure coupling. Then, 100 replicas were generated with temperatures ranging from 280 K to 458 K and distributed applying the exponential spacing strategy, as previously done in literature [101,102]. A first NVT position restrained MD was run on each replica for 50 ps. Finally, a 60 ns of production NVT-REMD was carried out on each replica at its own temperature, according to previous works [136]. The replica exchange interval was set equal to 1 ps, large enough if compared to the time constant of the heath bath (τ=0.1 ps). The resulting exchange probability was 0.3. The computational data were time-averaged over all trajectory steps corresponding to the chosen temperature, 300 K in this work. The LINCS algorithm [81] was used to constrain the lengths of all bonds. The integration time step was 2 fs. Periodic boundary conditions were applied along *xyz*. The short-range Van der Waals (VDW) and electrostatics interactions were cut off after 1 nm; the Particle Mesh Ewald (PME) method [137] was employed for long-range electrostatics. GROMACS 5 was used for REMD simulations and data analysis [138]. The inter-chain protein contacts were identified by contact probability plots. Contact probability for each residue was calculated as already described in a previous work [136].

Order Parameter and Functional Mode Analysis (FMA)

With the purpose of estimating the structural order of the two models and therefore how much protein chains are aligned; an order parameter was calculated for each REMD snapshot as follows:

$$ordP = \frac{1}{N_r}\sum_{r=17}^{42}\frac{\langle v_r, z\rangle}{\|v_r\|\cdot\|z\|} = \frac{1}{N_r}\sum_{r=17}^{42}\cos\ \alpha$$

Where v_r is the vector joining each of the N_r C_α-atoms pertaining to chain A with the corresponding C_α-atom (same residue number) of chain E and *z* is the fibril axis. Values of *ordP* close to 1 indicated an alignment close to the initial structure, i.e., aligned fibre along the fibril axis *z*. Values of *ordP* lower than 1 indicated a structure distortion (also refer to Appendix B-1).

Functional mode analysis (FMA) was applied to the REMD trajectory at 300 K [9] to elucidate collective motions directly related to fibre distortion. The applied method detects a collective motion maximally correlated to the fluctuation of the quantity of interest, that is, in the case under study, the above-mentioned order parameter. Assuming that the variable of interest is a linear function of the Principal Components, the maximally correlated vector can be derived by maximizing the Pearson coefficient [9] to quantify the contributions of the individual PCA vectors to the fluctuations of the variables of interest. This approach yields a single collective mode, which drives the phenomenon under investigation, referred to as ensemble-weighted Maximally Correlated Motion (*ewMCM*). In applying FMA, it is crucial to cross-validate the derived model for an independent set of simulation frames. The established approach applied for cross-validating the obtained results is to divide the simulation into a subset of frames for model building and a subset of frames for cross-validation. In this work, the obtained maximally correlated motion was validated by predicting the function of interest, in the cross-validation subset, with Pearson correlation coefficient higher than 0.93 for U-shaped and 0.97 for S-shaped models. Further details of the FMA calculation are provided in Appendix B-1.

Results

REMD simulations were carried out on the U-shaped model ($A\beta_{17-42}$ pentamer extracted from 2BEG.pdb file [58]) and the S-shaped model ($A\beta_{17-42}$ pentamer extracted from 2MXU.pdb file [63]) surrounded by explicitly modelled water and ions. Data analysis have been performed on the conformational ensemble at 300 K. More detailed information on simulation set up and analysis are provided in the Method Section.

Characterization of the Aβ conformational arrangements

The Root Mean Square Fluctuation (RMSF) plot shows the atomic fluctuations averaged on each protein residue (Figure 27a). In both cases, as expected, terminal regions are characterized by larger fluctuations with respect to the central region due to a higher solvent exposure (Figure 27a). By comparing the two different U-shaped and S-shaped models, it is worth noticing that the main difference is located at the C-terminal tail. In particular, the protein region V36-A42 is characterized by larger fluctuations in case of the U-shaped model ($RMSF_{A42}$=0.87 ± 0.10 nm), differently from what has been observed in the S-shaped model ($RMSF_{A42}$=0.56 ± 0.17 nm). A visual inspection of the above-mentioned fluctuations is provided in Figure 27b. In case of the U-shaped model, peptide chains most exposed to the solvent are also characterized by higher conformational instability. Also, regions V24-N27 and V36-G38 are characterized by high fluctuation peaks located on V24 ($RMSF_{V24}$=0.43 ± 0.07 nm) and G37 ($RMSF_{G37}$=0.52 ± 0.07 nm), respectively.

The previously highlighted conformational instability of residues V24 and G37 in the U-shaped model can be explained by analysing the secondary structure probability of the two simulated systems (Figure 27c). For each model the secondary structure has been calculated as a probability along all chains and all considered frames, as done in previous works [132]. The secondary structure probability along the REMD ensemble at 300 K (Figure 27c, lower row) was compared with the same probability in the PDB model (Figure 27c, upper row) for both U-shaped (Figure 27c, left) and S-shaped (Figure 27c, right) architectures.

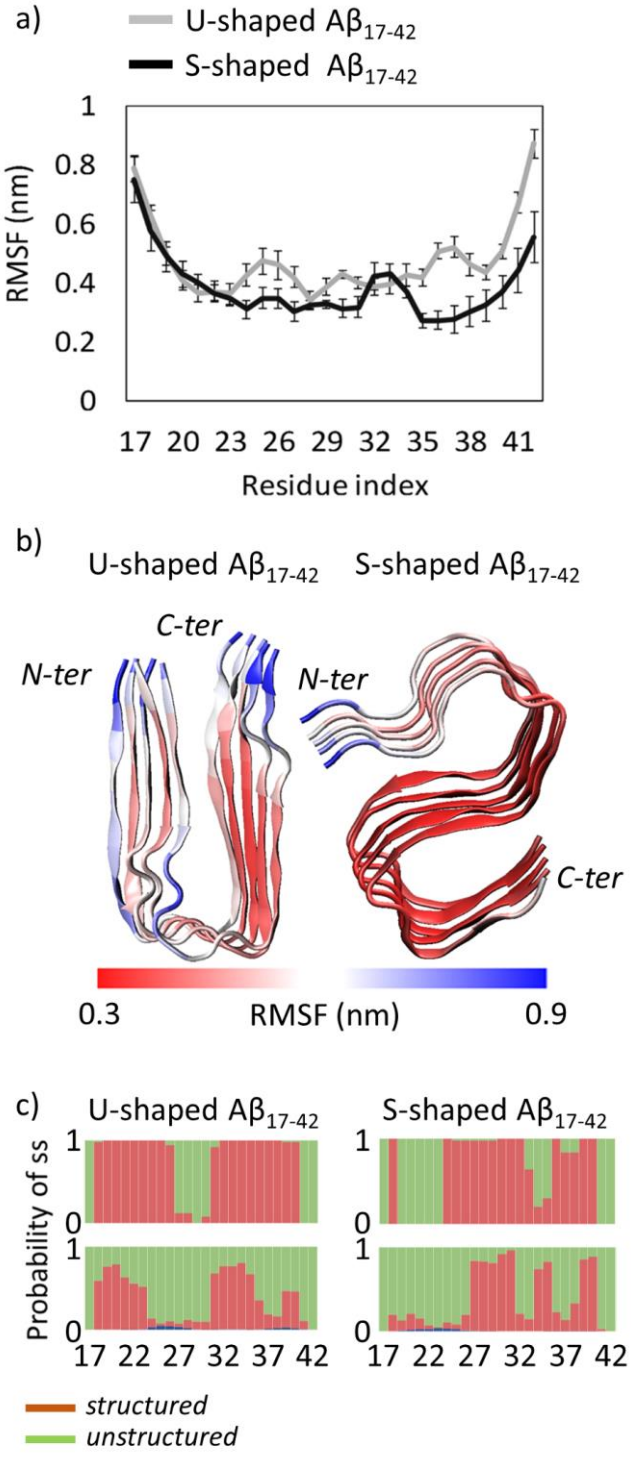


Figure 27. a) U-shaped and S-shaped Root Mean Square Fluctuation (RMSF) of atomic positions averaged on each protein residue. Each average value and relative standard deviation was obtained by mediating the RMSF on the five considered protein chains (A-E). b) U-shaped and S-shaped structural models coloured on the basis of RMSF values. The scale bar moves from red (RMSF=0.3 nm) to blue (RMSF=0.9 nm). c) U-shaped and S-shaped residue secondary structure probability, calculated over 5 considered chains (A-E) in the pdb models (upper row) and on the REMD ensemble at 300 K (lower row). For the sake of clarity, the secondary structures are classified in structured (red) and unstructured (green). Moreover, the structured class does not contain helices (shown in blue) being their contribution negligible throughout the overall REMD ensemble at 300 K.

Although in both models the two predominant structures are rigid beta-sheets and flexible coils, the secondary structures are differently distributed along the peptide chain. In detail, beta structures are mainly located in regions V18-D23, I31-M35 and V39-V40 for the U-shaped model and N27-I31, L34-M35 and V39-V40 for the S-model. A marked loss of beta-sheets was observed in both cases (U-shaped and S-shaped fibrils) if compared with the original NMR models. In detail, residues V18, V24-S26, I32, G33 and V36-G38 are characterized by a spontaneous beta-coil transition in the S-shaped fibrils whereas a reduction of beta-sheets was located at residues V24-S26 and V36-V40 in case of U-shaped models. It is worth mentioning that the loss of beta-sheets here observed in the U-model is consistent with a previous computational study [139]. Moreover, this evidence is in line with the conformational fluctuations of residues V24 and G37 highlighted in Figure 27a. The loop domains of both U-shaped and S-shaped fibrils remain largely unstructured along the simulation trajectory, in line with the NMR starting model (Figure 27c). The only difference is located at the turn region connecting β2 (residues V24-G33) and β3 (residues V36-V40) of the S-shaped fibril. In this case, it was observed an increased tendency to form a structured beta strand of residues L34-M35. Interestingly, the total Solvent Accessible Surface Area (SASA) of the U-shaped model (76.07 ± 4.17 nm^2) is slightly higher than that of the S-shaped model (70.62 ± 3.71 nm^2). This result might be related to the ability of the S-shaped model to reach a more compact arrangement. The above-mentioned observation suggests that the S-model is better able to maximize intra- and inter-chain contacts. More detailed information on SASA and RG is reported in Appendix B-1.

Characterization of the Aβ interatomic interactions

The detailed intra/inter chain interatomic interactions at an atomistic level were analysed in order to provide a deeper understanding of the interactions leading to the above-mentioned conformational properties. An overall view of regions mainly involved in the interchain non-covalent bonds are provided by contact probability plots (Figure 28a). A lack of inter-chain interactions can be detected in different regions of both models. Regarding the S-shaped model, a slight decrease in interatomic interactions may be observed at residues L17-V18, G37 and a marked one in range A21-G25. In case of the U-shaped model, lower contact probability was observed in protein regions V24-A30 and G37-A42.

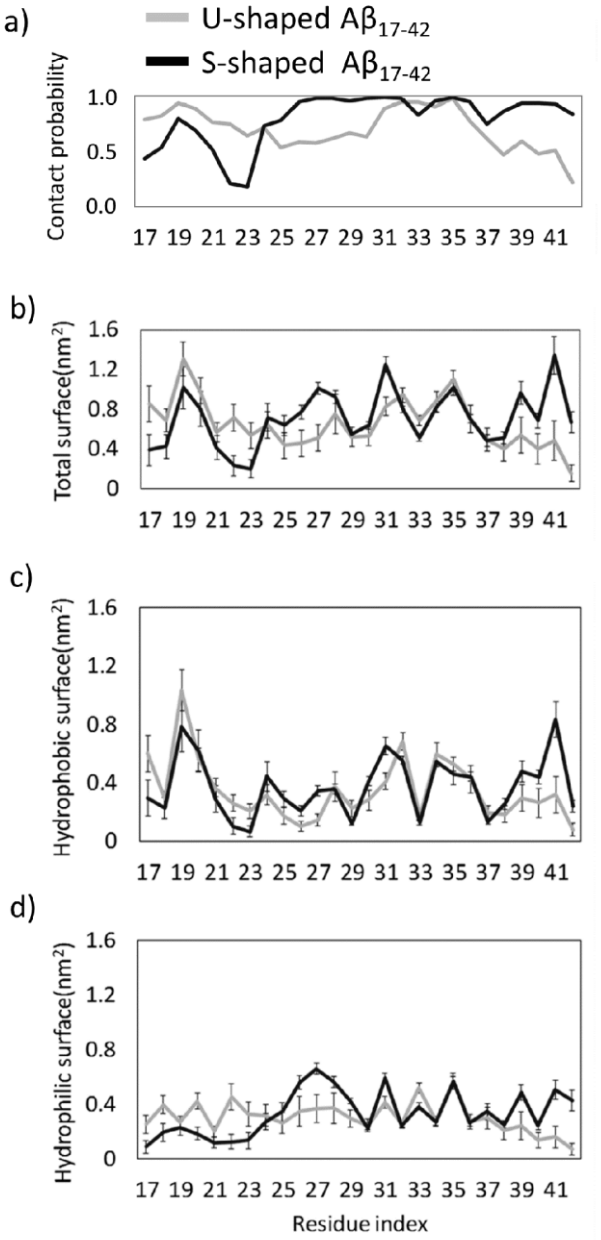


Figure 28. a) U-shaped and S-shaped per-residue inter-chain contact probability plot. b) U-shaped and S-shaped per-residue inter-chain total interacting surface. c) U-shaped and S-shaped per-residue inter-chain hydrophobic interacting surface. d) U-shaped and S-shaped per-residue inter-chain hydrophilic interacting surface. In all plots, contacts between chains B-C and C-D were considered and averaged on the REMD ensemble at 300K.

A noticeable difference between the two models is found at the C-terminal residues V39-A42, showing lower contact probability values in case of the U-shaped model (probability=0.22) when compared with the S-shaped model (probability=0.85). The lack of interactions indicates the presence of defects in the fibril structure. These defects, which are localized in both central and C-terminal regions in the U-shaped model, may be related to a higher conformational instability with respect to the S-shaped where inter-chain contact defects are mainly localized only in region L17-D23.

Another picture of the presence of the above-mentioned local defects in inter-chain contacts is provided by a detailed analysis of the inter-chain total, hydrophobic and hydrophilic interaction surface (Figure 28b, c, and d, respectively).

In a greater detail the U-model presented a reduced inter-chain interaction surface in both the core and C-terminal regions, whereas the S-shaped model showed a lower total surface only in the L17-D23 region. It may be of interest to decompose the total interaction surface in its hydrophobic (Figure 28c) and hydrophilic (Figure 28d) components. In the core and C-terminal region, the S-shaped model showed to maximize both hydrophobic and hydrophilic inter-chain interaction surfaces with respect to the U-shaped model, whereas the latter showed only a significantly higher hydrophilic interaction in the L17-D23 region. Hydrophilic interactions and interchain contacts provide an indication of hydrogen bond presence, strongly related to the conformational stability of the amyloid oligomers and fibrils as indicated by literature in this field [57,140,141].

Figure 29a focuses on inter-chain hydrogen bonds, calculated using a cut-off of 0.35 nm [142]. The protein domain L17-D23 of the U-shaped model, in line with hydrophilic character of the buried surface, shows the highest probability of inter-chain hydrogen bonds (Figure 29a). Instead, the S-shaped arrangement showed a high probability contact in the central domain and C-terminal region (V39-A42).

In addition to the inter-chain H-bonds, it is worth investigating the intra-chain H-bonds calculated within the same chain C using a cut-off of 0.35 nm. The highest probability of finding intra-chain H-bonds for the U-shaped model is between the side chain of residue D23 and backbone of G25 and sides chains of residues D23 and K28 (Figure 29b). The result is in agreement with previous literature indicating, in the U-shaped model, a salt bridge able to stabilize the loop region connecting two β-sheets preventing larger backbone motions [59,143,144].

It is worth mentioning that, in the central core, also the S-architecture presents two intra-chain contacts. The first one between the side chain of N27 and the backbone of G29, and a second one between side chains of A42 and K28 (Figure 29b), the latter identified earlier in the literature [63,130].

To get an overall view of inter-chain contacts, a map of all non-bonded interactions inside a cut-off=0.45 nm is shown in Figure 29c. The map clearly indicates how non-bonded interactions among same residues in an adjacent chain stabilize the S-shaped model particularly in the core and C-terminal region. In a greater detail, interactions between residues I41 and K28, G29 and A42 and K28 occurred only in the S-shaped model map.

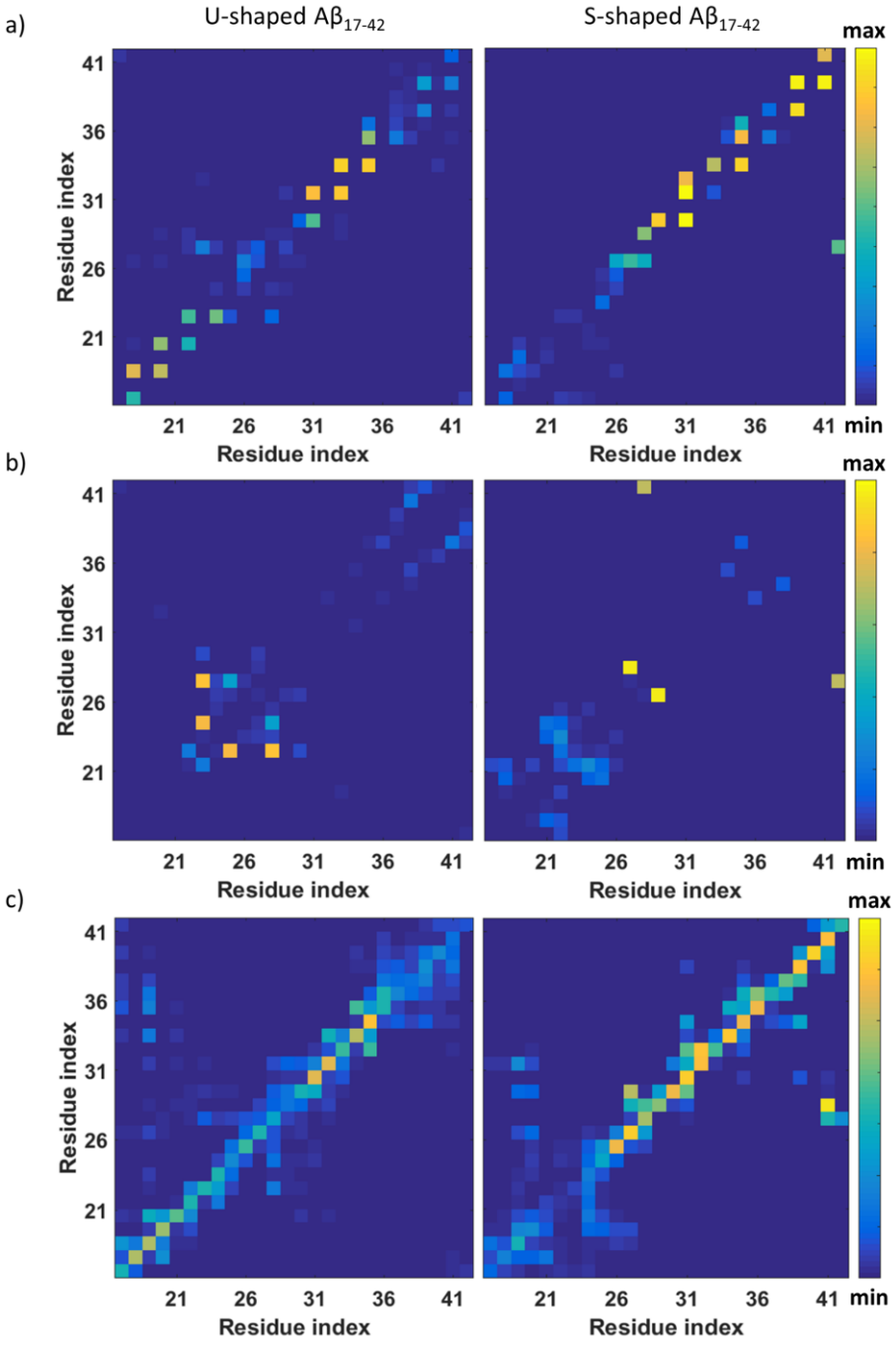


Figure 29. a) U-shaped and S-shaped probability contact maps of inter-chain hydrogen bonds. b) U-shaped and S-shaped probability contact maps of intra-chain hydrogen bonds. c) U-shaped and S-shaped probability contact maps of inter-chain non-bonded contacts. The hydrogen bonds (a, b) have been calculated considering a distance cutoff of 0.35 nm between donors and acceptors. The non-bonded contacts (c) have been calculated considering a distance cutoff of 0.45 nm between inter-chain heavy atoms.

Order parameter and Functional Mode Analysis

The probability distribution of the order parameter, *ordP* (Appendix B-1), calculated throughout the REMD trajectory at 300 K, is shown in Figure 30a. The S-shaped *ordP* along the overall 300 K REMD showed a sharp distribution with an average value and peak close to 0.95, thus indicating that the S-model maintains an intrinsic order of the fibre with chains aligned along the fibril axis. Instead, the U-shaped *ordP* has a spread distribution with a peak value around 0.8.

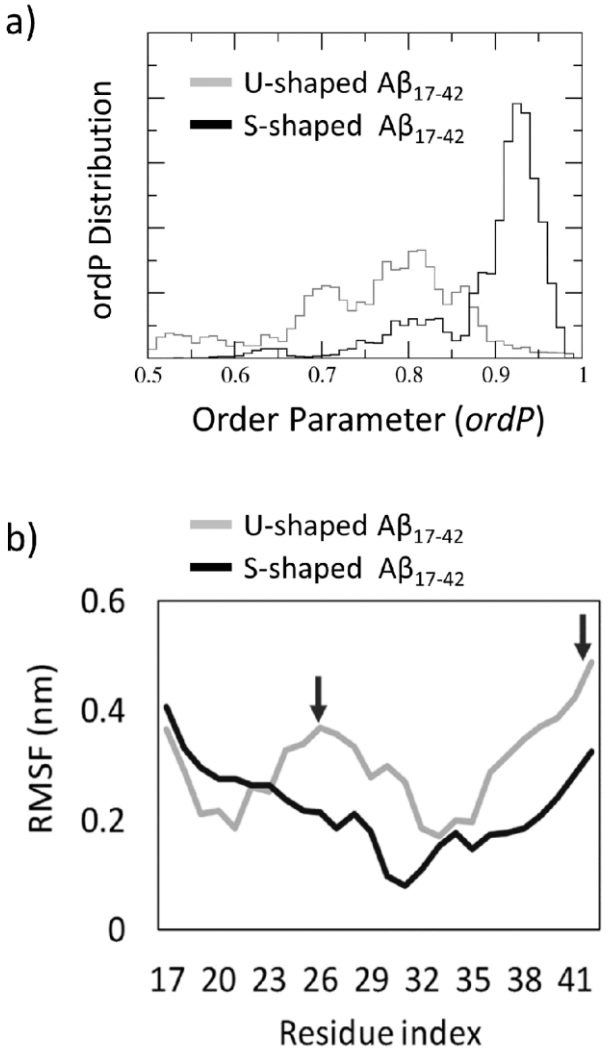


Figure 30. a) Order Parameter, ordP, distribution calculated throughout the overall REMD trajectory at 300 K. The same number of snapshots was considered for both U-shaped (grey line) and S-shaped (black line) models. The ordP value provides a quantitative estimation of the fibril order. Values close to 1 indicate an alignment close to the starting structure, i.e., aligned fibre along the fibril axis. Values lower than 1 indicated a structure distortion. b) Root Mean Squared Fluctuation (RMSF) plot calculated over the REMD trajectory at 300 K filtered on the ensemble weighted Maximally Correlated Motion (ewMCM) vector. Arrows indicate the area with highest fluctuations.

The Functional Mode Analysis (FMA) allowed to characterize the contribution of individual Principal Components Analysis (PCA) vectors to the fluctuation of *ordP*, yielding a single vector (the so-called *ensemble weighted Maximally Correlated Motion, ewMCM*), which drives the fibril structural destabilization (Appendix B-1). Observing the residues RMSF (Figure 30b) calculated over the *ewMCM* trajectories (starting and final snapshots shown in Figure 31) a significantly different conformational behaviour can be observed for the U-shaped and the S-shaped models.

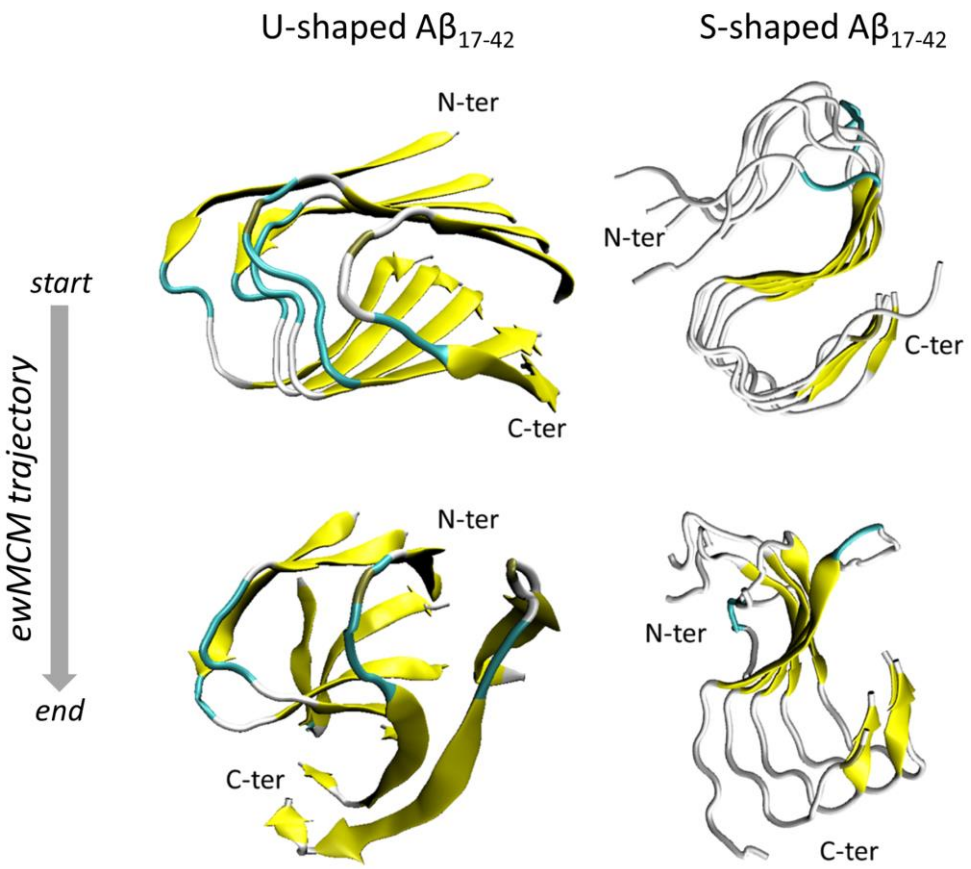


Figure 31 : Snapshots representing the starting and the final configuration extracted from the ensemble weighted Maximally Correlated Motion (ewMCM) for U-shaped and S-shaped models after performing the Functional Mode Analysis (FMA) on the REMD ensemble at 300 K.

Overall, the U-shaped model fluctuates much more than the S-shaped model, in particular for what concerns core and C-terminal regions (highlighted by arrows in Figure 30b). The C-terminal region is characterized by the highest fluctuation ($RMSF_{A42}$= 0.49 nm^2) followed by the central loop area D23-I31 ($RMSF_{S26}$=0.37 nm^2). Instead, the S-shaped model, showed higher fluctuations of the N-terminal region, in agreement with inter-chain contact analysis (Figure 29) indicating a lack of inter-chain H-bonds in this region.

In summary, the *ordP* shape factor and RMSFs calculated on *ewMCM* trajectories provided an interesting indication of the higher order maintained by the S-model under thermal motion, whereas the U-shaped model appeared to be more unstable and characterized by an overall disruptive conformational distortion (Figure 31). Furthermore, in agreement with previous data, provided by structural (Figure 27) and inter-chain analysis (Figure 28 and Figure 29), the S-shaped model seems to be subjected to a partial distortion only in the N-terminal region (L17-D23 region), whereas the U-shaped model assumed a more disordered configuration with a tendency to break in the central region losing almost completely the original conformation.

Discussion

In the brain of patients affected by AD, those peptides build up, layer by layer, hierarchically organized assemblies. This molecular phenomenon is related to a progressive loss of brain function, especially memory loss and cognitive deficit, that becomes ultimately fatal. Amyloid fibrils exist in an equilibrium of interchanging structures of monomers and oligomers characterized by polymorphism [56–59,61,145].

Several structural models exist for the $A\beta_{1–40}$ species, all sharing a U-shaped motif, made of two β strands (residues V18-S26 and residues I31-A40) linked by a central loop domain. Instead, the $A\beta_{1–42}$ species can also assume a S-shaped conformation [62–67], where three β strands (residues V12-V18, residues V24-G33, and residues V36-V40) are connected by major coil and turn region. It has been recently demonstrated that the S-shaped arrangement is not stable in case of $A\beta_{1–40}$ [130]. Recently, the higher toxicity of $A\beta_{1–42}$ species has been associated with its ability to assemble into ring-like N-fold models starting from the S-shaped fibril [68]. In this scenario, it is interesting to focus the attention on the two possible arrangements proposed for $A\beta_{1-42}$ species.

Results of the present research study highlighted the S-shaped assembly as more stable when compared with the U-shaped model. Moreover, the U-shaped model showed a high degree of conformational plasticity, especially considering the high fluctuations of residues I41 and A42 (Figure 27). The presented data are in apparent contradiction to literature published over the last decade classifying the U-shaped architecture as conformationally rather stable [146–148]. However, it is worth mentioning that the conformational sampling performed in previous computational studies was in general restricted to classical MD simulations and/or limited simulated time (from tens to hundreds of ns). In this view, the present study should not be seen in contrast with previous literature, but as an improvement of the protein conformational sampling provided by REMD coupled with dimensionality reduction methods. Interestingly, the S-shaped model showed to maximize the formation of intra- and inter-chain hydrophobic contacts within the fibril model, especially on residues I41 and A42, characterized by a

higher hydrophobic buried surface than the U-model. Nevertheless, the most significant contribution to the stability of S-model is attributed to the inter-chain hydrophobic (Figure 28c) and hydrophilic (Figure 28d) interaction surface, especially in the C-terminal region. More in depth, in agreement with a recent computational study [130], the following inter-chain hydrophobic contacts are present only in the S-shaped model: I41-K28, I41-G29, and A42-K28. The fundamental role played by those residues was also confirmed by analysing the inter-chain contacts (Figure 28a). Even the FMA analysis highlighted the importance of the C-terminal region, which showed to be much more stable, than the N-terminal one. Instead the U-shaped model was affected by a higher distortion, which started from the core region related to inter-chain contacts disruption. The above-mentioned observations (Figure 31) were quantified by RMSFs profiles (Figure 30b).

It is important to remark, that, for the sake of a meaningful comparative analysis, the same protein region (residues L17-A42) has been considered for both U-shaped and S-shaped models. The neglected domain is known to be unstructured and not present in the U-shaped pdb file (2BEG [58]). On the other hand, the S-shaped pdb model (2MXU[63]) contains an additional structured region between residues E11-K16. It is rational to consider that neglecting the above-mentioned region may somehow affect the whole S-shaped arrangement toward a higher or a lower stability. In this connection, Figure 28a and Figure 4b provide a first indication on possible effects on the overall protein assembly stability. In a greater detail, whereas the U-architecture instability is related with the central area and C-terminals regions, the N-terminal tail (residues L17-V24) represents the weakest area of the S-architecture. The presence of the E11-K16 structured domain is reasonably expected to strengthen and further stabilize the interchain hydrogen bonds of protein region L17-V24. A convincing demonstration of the above-mentioned hypothesis is shown in Appendix B-1. As expected, the S-shaped$_{11\text{-}42}$ model, showed a higher intrinsic order with respect to the S-shaped$_{17\text{-}42}$ (Figure 59).

Summarizing, the existence of U-shaped and S-shaped assemblies for the A$\beta_{1\text{-}42}$ species has been already demonstrated by several previous studies [58,62–67] and not under discussion in the present work. Instead, outcomes of the present comparative study, provided clear information on the tendency of a specific conformational state to explore and eventually get out of the free energy minimum identified by the correspondent experimental model. These data, based on 6 μs of enhanced conformational sampling for each model, clearly suggest the U-shaped much less stable than S-shaped model, at least for what concern a A$\beta_{17\text{-}42}$ pentamer.

Nonetheless, previous computational studies have focused on U-shaped models to investigate the ligand driven destabilization of A$\beta_{1\text{-}42}$ species [148–151]. Thus, in the specific case of A$\beta_{1\text{-}42}$ species, the S-shaped model should also be considered as a target for rational design/discovery/optimization of effective compounds.

4.2 Estimating the Mechanical Performance of U- and S-shaped Amyloid Beta Fibrils

Introduction

The mechanical tests of amyloid fibrils are of high importance, considering the well-known interplay between mechanical performance of the fibril architecture and amyloid proliferation [152–158]. Measurements of the rigidity of these β-sheet-rich protein aggregates have provided conflicting results ranging from soft Young Modulus (0.1 GPa [159]) to extremely hard (10 GPa [160]) depending on the employed methodology, fiber geometry, and assembly characteristics [160–164]. Several hypotheses have been proposed in literature to unveil the connection between fibrils stiffness and amyloid pathological proliferation. The mechanical failure of amyloid fibrils was proposed as one of the key factors behind the amyloid proliferation, given the correlation between the fracture properties of amyloid fibrils and the propagation of amyloid diseases [165]. Other investigations have highlighted the exceptional rigidity of the amyloid material as a key factor to distort cell membranes and disrupt their function [157].

In the past, atomic force microscopy (AFM) [166–169] was applied to probe in vitro the mechanical properties of amyloid assemblies such as their elastic modulus [170–174] or their bending rigidity [160]. In the field of $A\beta_{1\text{-}42}$ fibril, however, nanoindentation or force spectroscopy experiments do not allow to distinguish among U- or S-shaped architecture, structural transitions in polypeptide chains, and interactions governing the mechanical strength of amyloid structures [175].

Molecular modelling techniques [92,96,123,176–178], e.g., steered molecular dynamics (SMD) [100,132,179], can be applied to address this issue with the final aim of (1) increasing knowledge and (2) rationalizing available experimental data. In particular, SMD allows to precisely control the manner in which the force is virtually applied to the system and to clarify the nature of the relationships linking mechanical properties of the system to interatomic interactions [180,181]. Recently, SMD simulations have been employed to evaluate the force needed to pull away a single peptide from U-shaped Aβ fibrils [175]. In silico approaches have also been applied to estimate the Young's modulus of amyloid fibrils [182], and to examine the dependence of their mechanical stability on the amino acid sequence [181].

From the body of available literature, it emerges the need to clearly identify those factors governing fibril material features at molecular level [164], as amyloid fibrils are characterized by a multiscale nature in which nanoscale phenomena determine macroscale properties [160].

Here, steered molecular dynamics (SMD) simulations were carried out to provide a detailed mechanical characterization of U- and S-shaped $A\beta_{17\text{-}42}$ small fibrils. Technically, a computational workflow similar to the one successfully applied elsewhere to get insight on the stability of small amyloid fibrils [162,181], was here applied.

Marked differences were observed between U- and S-shaped $A\beta_{17-42}$ architectures in term of mechanical properties along the fibril axis, in particular for what concerns the mean peaks force, stiffness constants and Young's modulus. Notably, the findings of this study provide clear evidence about the stability of the recently proposed S-shaped model of the $A\beta_{1-42}$ species.

The computational approach applied in this work can be further considered to design strategies aimed at destabilizing amyloid fibril for therapeutic applications.

Moreover, investigating the polymorphic character of amyloid assemblies will offer new insights into the selection criteria that should be considered in the design/development of new bio-inspired material.

Materials and Methods

Two different models for the $A\beta_{1-42}$ species were considered in this work: the U-shaped $A\beta_{17-42}$ (PDB ID: 2BEG [58]) and the recently resolved S-shaped $A\beta_{11-42}$ (PDB ID: 2MXU [63]). A pentamer of $A\beta_{17-42}$ was extracted from each PDB structure.

For each molecular structure, we will refer to S-shaped model and U-shaped model in the following.

Molecular Dynamics Simulations

Each molecular system was solvated in a 7 nm side dimension cubic box and neutralized by counterions. AMBER99-ILDN force-field [77,78] and TIP3P model [80] were employed to define protein topology [77,79] and water molecules, respectively. Each system consisted of about 35,000 interacting particles. The models were firstly minimized by applying the steepest descent energy minimization algorithm. Then, in order to increase the statistics of MD data, five independent replicas of each molecular system, differing in initial atom velocities, were obtained from the minimized system. In detail, for each replica a random velocity taken from a Maxwell-Boltzmann distribution at 300 K was assigned to every atom of the system during a preliminary position restrained MD simulation (100 ps time duration) in NVT ensemble. V-rescale thermostat was applied to keep temperature at 300 K with a time constant of 0.1 ps [82]. Then, a second position restrained MD (100 ps time duration) in NPT ensemble was carried out for each system at 300 K (τ=1 ps) and 1 atm (τ=5 ps). V-rescale [82] and Berendsen [76] coupling methods were used as temperature and pressure coupling. Then, an unrestrained MD simulation (100 ns time duration) was run for each replica of both S- and U- fibril models. During each production run, the V-rescale [82] and Parrinello-Rahman [109] approaches were employed for temperature and pressure coupling, respectively. The LINCS algorithm [81] was used to constrain the lengths of all bonds, so as to extend the integration time step to 2 fs. Periodic boundary conditions were applied to all (3) dimensions. The short range Van der Waals (VDW) interactions were cut-off after 1 nm, and long range electrostatic forces were calculated with the Particle Mesh Ewald (PME)

method [137]. GROMACS 5.1.2 package was used for all MD simulations and data analysis [138]. The overall structural stability of each model was determined through the calculation of the Root Mean Square Deviation (RMSD) of the C-alpha atoms with respect to the initial configuration of the minimized structure (Appendix B-2).

Steered Molecular Dynamics Simulations

To compare the mechanical response of S-shaped and U-shaped fibrils, several SMD simulations were performed, differing for the prescribed initial condition represented by a MD output configuration (5 MD replicas for U- and S-shaped models), following the procedure described below (Figure 32).
The fibril axis was aligned with the *Z*-axis in the 3D cartesian space (represented in Figure 32, top row). A defined set of atoms was pulled toward a certain direction while a second group of atoms was position-restrained, as recently done in literature to study the stability of small amyloid fibrils [162,181,183].

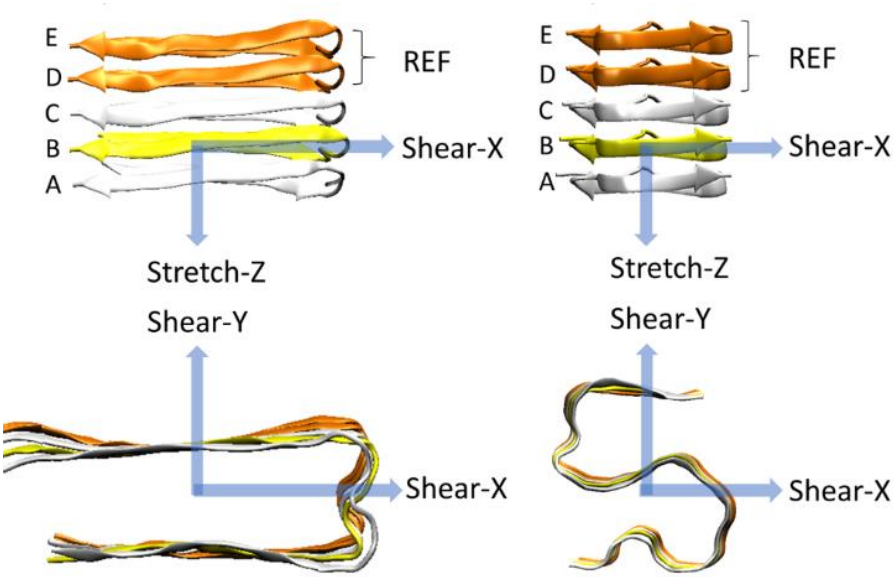


Figure 32. Representation of the deformation protocols applied to the fibril model in case of U-shaped (left) and S-shaped (right). Each deformation is defined by a pulling direction, highlighted by the blue arrow. For each deformation, the constrained pulled group and the restrained reference group (both composed only by C-alpha atoms) are evidenced in yellow and orange respectively. All the other atoms of the peptides are free to move.

In a greater detail, the C-alpha atoms of the chain B were pulled away from the position-restrained reference group (chain D and chain E). A harmonic force corresponding to the spring constant of 3000 kJ/(mol nm^2) was used to pull the chain B away with a velocity of 0.01 Å/ps. Other two pulling velocities (0.1 Å/ps and 0.001 Å/ps) were considered to carefully check the dependence of mechanical response on the choice of pulling velocity (Appendix B-2).

The V-rescale [82] thermostat was used as temperature coupling. SMD simulations were performed using GROMACS 5 [138]. In summary, 90 SMD simulations have been carried out, i.e., 2 models (S- and U-fibrils), per 5 replicas (configurations taken from production dynamics), 3 pulling directions (*X*, *Y*, and *Z* as shown in *Figure 32*) per replica, and 3 pulling velocities (0.1 Å/ps, 0.01 Å/ps, and 0.001 Å/ps) per each pulling direction.

Results

Steered Molecular Dynamics of the Aβ$_{17-42}$ Architectures

SMD simulations were carried out on the U-shaped model (Aβ$_{17-42}$ pentamer extracted from 2BEG.pdb file [58]) and the S-shaped model (Aβ$_{17-42}$ pentamer extracted from 2MXU.pdb file [63]) surrounded by explicitly modelled water and ions. All the presented data have been calculated as average ± standard deviation over the 5 SMD replicas for each simulated architecture considering the pulling velocity of 0.01 Å/ps. Other two pulling velocities (0.1 Å/ps and 0.001 Å/ps) were considered to carefully check the dependence of mechanical response on the choice of pulling velocity (Appendix B-2). More detailed information on simulation set up and analysis are provided in the Method Section. Explanatory snapshots taken from representative SMD trajectories, together with the simulated force-time profiles, are presented in Figure 33. The conformational state corresponding to the fibril break and the force [pN] peak indicating the mechanical failure can be clearly identified (Figure 33, B). The mechanical resistance of U- and S-shape small fibrils were evaluated in terms of the produced peaks shear (X and Y direction) and stretch (Z direction) force values, averaged over the five replicas (Figure 34A).
A significant difference was observed between stretch force peaks in the two configurations ($F_{stretchU}$ =1795±186 pN *vs.* $F_{stretchS}$ =2073±204 pN). On the contrary, data from the other two deformation protocols (*X* and *Y* pulling direction), indicated similar shear force peak values for U- and S-architectures (Figure 34A). It is worth mentioning that the magnitude of forces estimated by SMD is dependent on the imposed pulling velocity (Appendix B-2), as highlighted by the most current research [184,185]. However, the observed difference in the mechanical response to stretching of the two-different fibril U- and S-shaped configurations is also conserved lowering the pulling velocity (Appendix B-2).

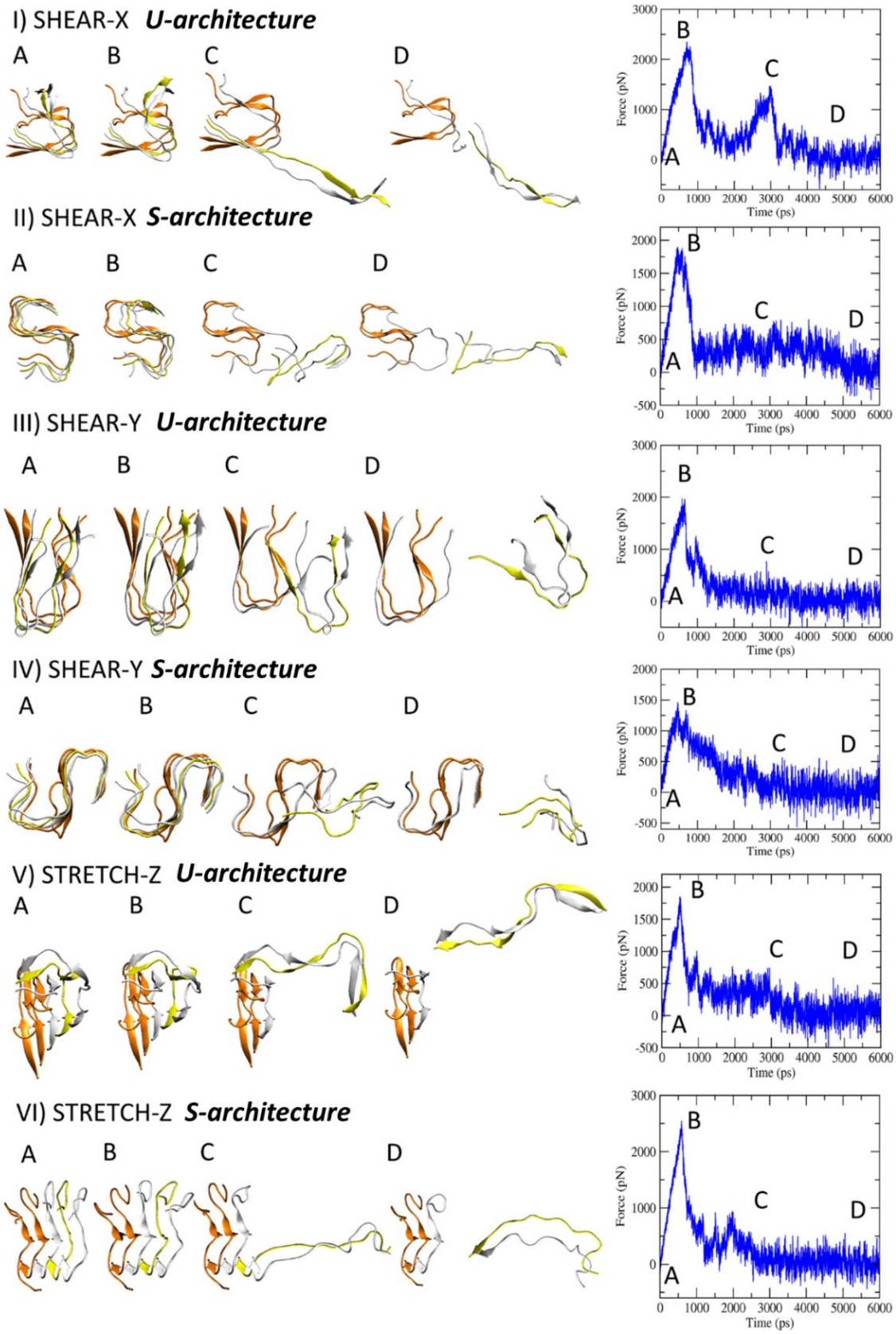


Figure 33. Representative snapshots of U- and S-models evolving in time when the three deformation protocols are applied. On the right side the corresponding force-time profile is plotted.

Analysis of SMD simulations suggested an anisotropic behavior characterizing the fibril mechanics, here quantified in terms of stiffness (Figure 34B), i.e., the gradient of the force versus displacement curves (Appendix B-2) in the linear response regime (small deformations region). Interestingly, a marked difference between U and S-architectures was observed in the value of the stretching stiffness k ($k_S \approx 3k_U$), as shown in Figure 34B. The observed shear stiffness (Figure 34B), lower than stretching stiffness in both U- and S-architectures, confirmed the higher mechanical stability and resistance of the $A\beta_{17\text{-}42}$ assembly along the fibril direction.

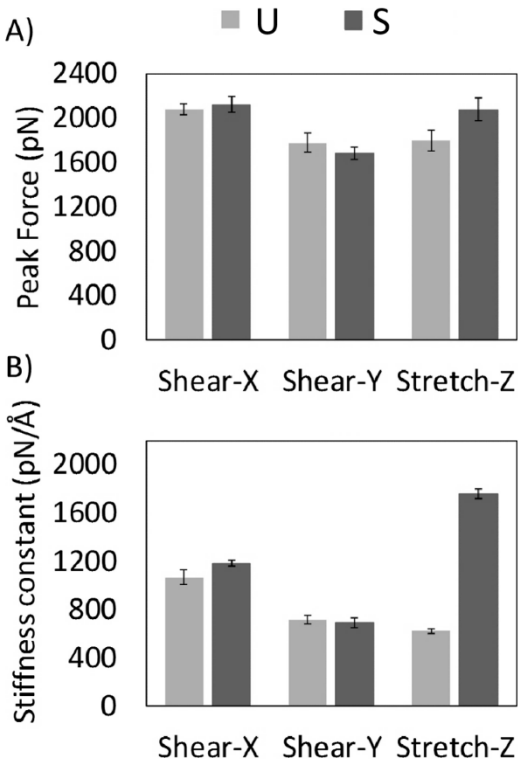


Figure 34. A) Average values with related standard error of the peak forces recorded for U- and S-models during each type of SMD protocol. B) Average values with relative standard error of stiffness constants extracted from the gradient of the force versus displacement graphs in the linear region.

Mechanical Performance of $A\beta_{17\text{-}42}$ Architectures

To provide further insight into the mechanical performance of U- and S-shaped models, the stress-strain curves were computed from SMD force-displacement results along the pulling directions (i.e., X-, Y-, and Z- axis). Technically, each stress value was obtained by dividing the force with the interaction surface between the pulled chain B and the chain C. The application of a linear fitting to the stress-strain data (as detailed in Appendix B-2), allowed the estimation of the Young and Shear moduli (E, G_{xy}, and G_{yx}) of both the two-fibril configuration. In detail, the estimated Young's modulus was E_U=1.4±0.3 GPa and E_S=2.7±0.3 GPa, for U- and S- architecture, respectively. An average value

of the shear modulus G along the xy plane (perpendicular to the fibril axis) can be obtained as $G=(1/G_{xy}+1/G_{yx})^{-1}$ and results in a value of about G_U=1.4±0.3 GPa and G_S=2.7±0.3 GPa. Noteworthy, the SMD-based estimates of the Young's moduli are in excellent agreement with recent AFM-based experimental data on $A\beta_{1-42}$ species (3.2 ± 0.8 GPa [161]) and in the range of estimated Young's moduli obtained from other amyloidogenic peptides [160–164], as detailed in Table 3. It's also important to highlight that two- and three-fold symmetry $A\beta_{1-40}$ fibrils exhibited, as expected, a higher Young Modulus due to a different conformational arrangement and fibril geometry [186]. It is worth mentioning that no marked variations (in both absolute and comparative values) of the quantified moduli were observed by lowering the SMD pulling rate (Appendix B-2). The close agreement with experimental data [160–164] and the low Young's modulus variation in SMD at lower velocities (0.001 Å/ps) confirm that the proposed simulation set-up is able to properly replicate the mechanical behavior of the investigated fibril configurations.

Table 3.Experimentally estimated Young's moduli of amyloidogenic peptides

Amyloid species	Young Modulus (GPa)
$A\beta_{1-42}$ species ([161])	3.2
β-lactoglobulin ([187]; [188])	3.3; 4
$A\beta_{1-40}$ ([189]; [182]; [190])	0.05-1.62; 2-18; 18-30
insulin fibrils ([174])	0.001-0.1
HypF-N protofibril ([191])	0.06-0.5
hIAPP amyloid fibril ([192]; [193])	0.4-0.6; 12-14
α-synuclein fibril [163]	1.2
prion fibrils [194]	0.5-1.3
HET-s prion fibril ([180]; [158])	1.5-9.8

Interatomic interactions drive the mechanical response along the fibril axis

There is ample evidence that the mechanical response of amyloid fibrils is driven by the collective rupture behavior of hydrogen bonds sustaining the cross-β structure of Aβ [180,181]. In particular, the Aβ fibrils are stabilized by a network of backbone hydrogen bonds that acts as a chemical glue allowing them to withstand mechanical forces [160]. Here, to provide a deeper understanding of how the interatomic interactions rule the overall mechanical properties of U- and S-shaped fibrils, the hydrogen bonds between the pulled chain B and the chain C were studied in detail. In Figure 35 it is presented how, in both the two configurations, the percentage number of H-bonds between the pulled chain B

and chain C along the pulling directions, i.e., X-, Y-, and Z-. The U- and S-shaped configurations during the simulated pulling reacted differently to the applied force (Figure 35). In particular, the S-shaped configuration exhibited a higher tendency than the U-shaped configuration to retain H-bonds as function of the interchain displacement.

The most relevant differences in inter-chain H-bonds between U- and S-shaped configurations are found along fibrils axis (Figure 35C), in agreement with the marked difference observed in the mechanical response during the stretching protocol.

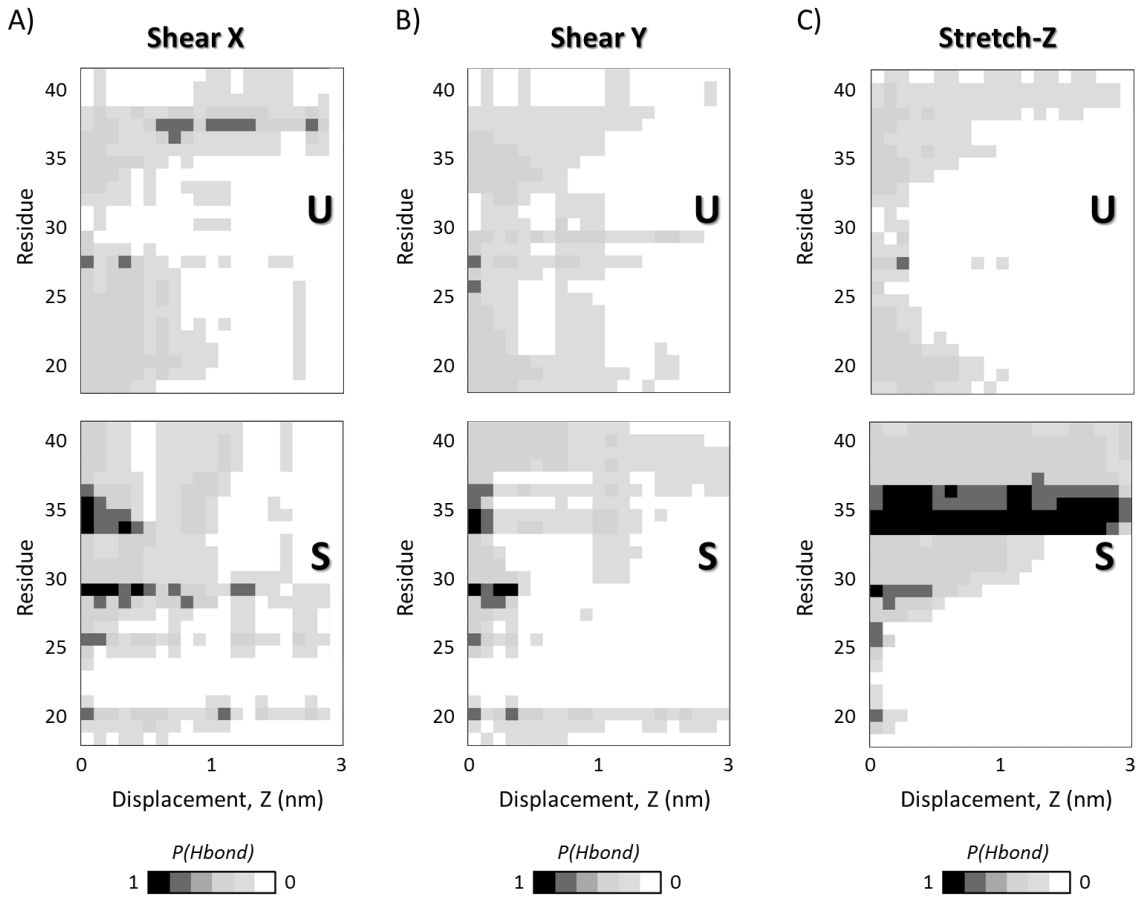


Figure 35. Interchain H-bonds (between chain B and chain C) as function of the interchain chain displacement in A) X-, B) Y-, and C) Z-direction along the SMD simulations at pulling rate v=0.01 Å/ps. Plots show the probability of H-bonds presence as function of interchain displacement in case of U- (top) and S-shaped (bottom) arrangements. For each residue in one chain (e.g., chain B) the presence of H-bonds with the adjacent chain (e.g., chain C) is quantified as a probability calculated over 5 SMD replicas. Black color indicates a probability of 1, whereas the white code represents a probability of 0, as reported in the legend.

Given that the two fibril configurations share the same sequence, these findings clearly suggest that the different architecture is responsible for eliciting a different chain detachment pathway. In case of U arrangement, the first rupture stage was associated with the disruption of the central H-bonds located at Ala30-Ile31 region. Instead, the S-architecture presented a completely different response, with the N-terminal residues initially (displacement value lower than 1 nm) detached due to the disruption of

β_1 (Ala21-Asp23). For a more clear interpretations of the data, it is worth considering that the per-residue H-bonds depicted in Figure 35 are presented in terms of probability distribution, calculated by sampling all the SMD replicas per architecture, to meaningfully improve the statistical significance of the performed analysis.

Discussion

Amyloid protein fibrils are a broad class of misfolded protein structures whose mechanical properties are driven by the non-covalent interactions among (primarily) β-sheets, causing the monomers to stack into fibrils [195]. In particular, amyloid beta (Aβ) assemblies, a subclass of amyloid fibrils, share a cross-β motif in which individual β-strands are oriented perpendicular to the fibril axis, resulting in the formation of a dense hydrogen bonded network which extends along the entire length of the fibril [196,197] and gives resistance to the structure.

In the last decade, the intriguing mechanical properties of amyloid fibrils have been the subject of a number of experimental and computational studies [161,162,166–174,181], motivated by the convincement that investigating the mechanical response of amyloid fibrils could provide crucial insight into many different research contexts.

For example, amyloids have been investigated as possible application in the field of functionalized nanowires, novel edible and biodegradable protein-based polymers [198], biological adhesives, and network-type materials [164]. Fine tuning of the molecular interactions leading to target mechanical properties of amyloid fibrils represents a pivotal point for further applications of amyloid-based nanostructures as novel biomaterials. Moreover, understanding the physico-chemical properties responsible for to the amyloid mechanical stability may yield an important piece of knowledge into the mechanisms of cellular toxicity. It is now well established that amyloid fibrils are generally polymorphic at the molecular level [199], with a given peptide or protein being capable of forming a variety of fibrils all having different structural features. This is also the case of the AD, where the $A\beta_{1-42}$ fibril may assume both a U-shaped and S-shaped motif [63,65–67].

Elucidating the connection between the molecular features of a specific fibril architecture and its mechanical performance is therefore crucial to rationally design therapeutic strategies aimed at destabilizing amyloid assemblies.

The different mechanical response of U- and S-shaped architectures was tested here by an extensive computational investigation based on SMD simulations. The resistance of the fibril models to mechanical deformation along specific directions was first analyzed by calculating the magnitude of the mean peak force (Figure 34A). Interestingly the S-shaped arrangement has highlighted better mechanical performances with respect to the U-shaped fibril arrangement. This is an interesting

finding, considering that stretching along the fibril axis directly probes the strength of inter-β-sheet hydrogen-bonds. To complete the picture, it was observed that in both U-shaped and S-shaped configuration the interchain hydrogen-bonding network provided a cooperative resistance. As detected by peak force analysis, the S-shaped configuration performed better than the U-shaped one, in terms of stiffness constant along the fibril axis (Figure 34B) and in terms of Young's modulus (E_S= 2.7±0.3 GPa, E_U= 1.4±0.3 GPa). Notably, these results are in excellent agreement with recent AFM experimental data on $A\beta_{1-42}$ species (3.2 ± 0.8 GPa [161]), which is exactly the one considered in this work, and in the range of Young's moduli from other amyloidogenic peptides [160–164], as detailed in Table 3. As expected, the two- and three-fold symmetry $A\beta_{1-40}$ fibrils exhibit a higher Young Modulus due to a different conformational arrangement and fibril geometry [200].

The previously mentioned results describe the importance of size effects in elucidating the mechanical properties of amyloid fibrils, as shown in case of different amyloid species [154,158,193,201,202]. This issue is of great importance for understanding the molecular mechanism behind amyloid growth and proliferation, and it represents a fundamental piece of knowledge to compare computational and experimental results concerning mechanical performance of biological material.

Starting from the well consolidated knowledge about the link between mechanical properties of fibrils and the collective rupture behavior of hydrogen bonds that sustain the cross-β structure of Aβ complexes [180], here, a picture of H-bond rupture as function of fibril deformation was also provided, in order to identify the molecular phenomena responsible for the greater mechanical stability of the S-architecture (Figure 35). The most relevant differences in inter-chain H-bonds between U- and S-shaped configurations are found along fibrils axis (Figure 35C). Although the two models are characterized by the same number of H-bonds in the starting configuration, a different reaction to the applied deformation was observed during the stretching dynamics, with the S-architecture H-bond distribution assuring a better distribution of the stress over the interchain surface, and consequently a higher mechanical resistance.

It worth mentioning that, in this work, for the sake of a meaningful comparative analysis, the same protein region (residues L17-A42) has been considered for both U-shaped and S-shaped models. The neglected domain is unstructured and not present in the U-shaped experimental model (2BEG [58]). Instead, the experimental S-shaped model (2MXU [63]) contains an additional structured region (residues E11-K16). However, the presence of the E11-K16 structured domain may be reasonably expected to stabilize the interchain hydrogen bonds of protein region L17-V24, further increasing its superior mechanical performance revealed in this manuscript. An evidence is given by Figure 35, indicating the hydrogen bond rupture as function on fibril strain. Whereas U-architecture rupture always begins in the central area (Ala30-Ile31), the S-architecture disruption starts at N-terminal tail (residues L17-

V24) and the unrolling endures progressively to the C-terminal region. Nevertheless, it was previously proved (Section 4.1) that the presence of the E11-K16 structured domain provide additional interchain hydrogen bonds in the protein region L17-V24 of the S-shaped assembly [203].

Outcomes of the present work indicate that different fibrils architecture have similar anisotropic mechanical behavior with a high mechanical resistance along the fibril axis direction. Moreover, these findings clearly point out the attention on the S-architecture, indicating that the mechanical performance is markedly superior to the U-shaped architecture, at least for the case of small fibrils.

Conclusions and Future Developments

Protein-protein interactions leading to protein aggregation represent a critical event in the onset of several neurodegenerative rare diseases. Elucidating the common structural features of protein aggregation diseases at molecular level, may provide novel opportunities for overarching therapeutic approaches such as inhibiting aggregation, or blocking common aggregation-induced cellular toxicity pathways. To address this issue, it's extremely important to shed light on the molecular interactions responsible for protein aggregation. Despite substantial research efforts in this field, the fundamental mechanisms of protein misfolding and aggregation mechanisms remain somewhat unrevealed.

In the present PhD thesis computational molecular modelling has been employed to shed light on the early stages of protein aggregation responsible for the onset of two different diseases: Spinocerebellar Ataxia Type-1 (SCA1) and Alzheimer's Disease (AD). Outcomes of the present study may be of relevance in the field of SCA1 research with focus on AXH aggregation mechanism. Classical and enhanced sampling techniques were applied to fully characterized the AXH aggregation pathway from monomer to tetramer, highlighting several protein mutations responsible for the destabilization of the monomer/dimer/tetramer equilibrium. For the first time in literature, the R638 residue, was identified as mainly responsible for dimer-dimer interaction and may be considered as an important target space for pharmacological intervention. Within this framework, further computational/experimental tests are needed to validate the results presented in this work and to elucidate disease-causing aggregation mechanism.

The present work has also investigated the conformational dynamics and stability of Aβ fibrils responsible for the Alzheimer's Disease. The computational results provide evidences about the stability of the recently proposed S-shaped model of Aβ fibrils due to interatomic interactions involving the C-terminal residues. Moreover, the S-shaped architecture shows superior mechanical performance with respect with the U-shaped motif due to a better distribution of the mechanical stress.

The presented data have demonstrated the importance of molecular architecture in modulating the conformational dynamics and stability of the protein assembly. Apart from the scientific interest in amyloid biomechanics, this work may also stimulate further research on amyloid aggregation modulators. The computational platform developed in this study could be further considered to test the efficiency of potential aggregation inhibitors aimed at destabilizing/reducing the stability of the Aβ fibrils.

Appendix-A

A-1 Conformational Fluctuation of AXH Monomer of Ataxin-1

Potential energy distribution among Replica Exchange Molecular Dynamics

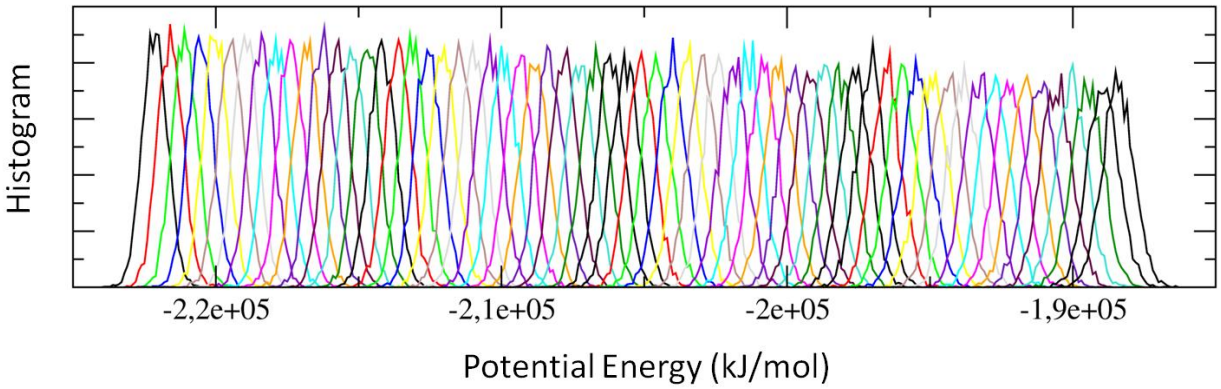


Figure 36. a) Potential energy distribution among Replica Exchange Molecular Dynamics (REMD) simulations. Temperatures were distributed across the replicas in a geometric progression, i.e. with the same ratio used to scale each temperature from the one below it, keeping the overlap of the potential energy distributions constant across the temperature space.

Ramachandran Plot

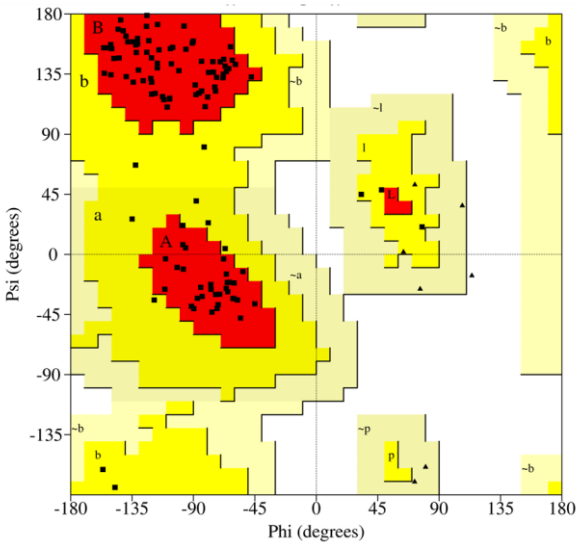


Figure 37. Ramachandran Plots show that all protein residues were found in the most-favoured and additional-allowed regions, providing support for the model's correctness.

REMD cluster vs. PDB models on Free Energy Landscape

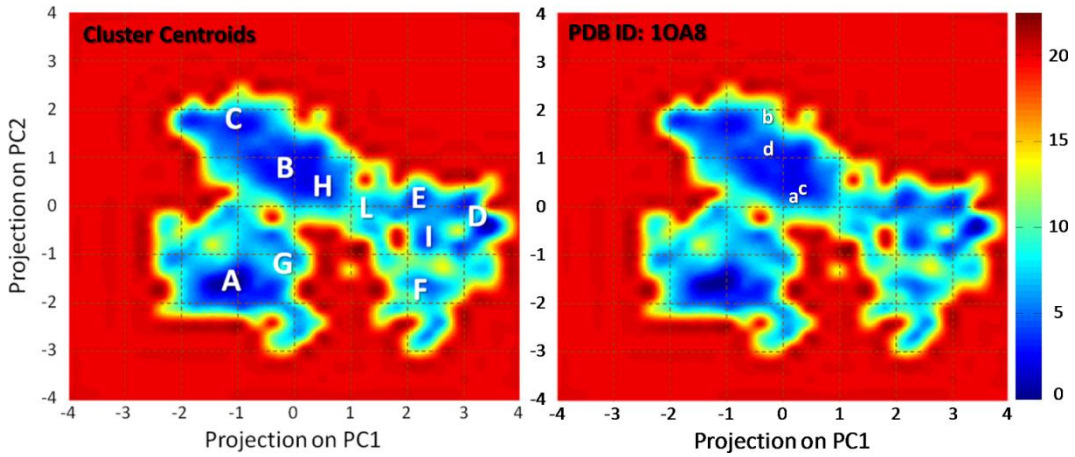


Figure 38. Free energy profile (kJ/mol) of the AXH^m represented as function of the first and second Principal Component projection. Labels indicate the position on the free energy landscape in term of first and second Principal Component projection values of: left) cluster centroids obtained (labels A to L) by structure based clustering approach on REMD trajectory and right) AHX^m atomic coordinates (label a,b,c,d) taken from 1OA8 PDB model[44].

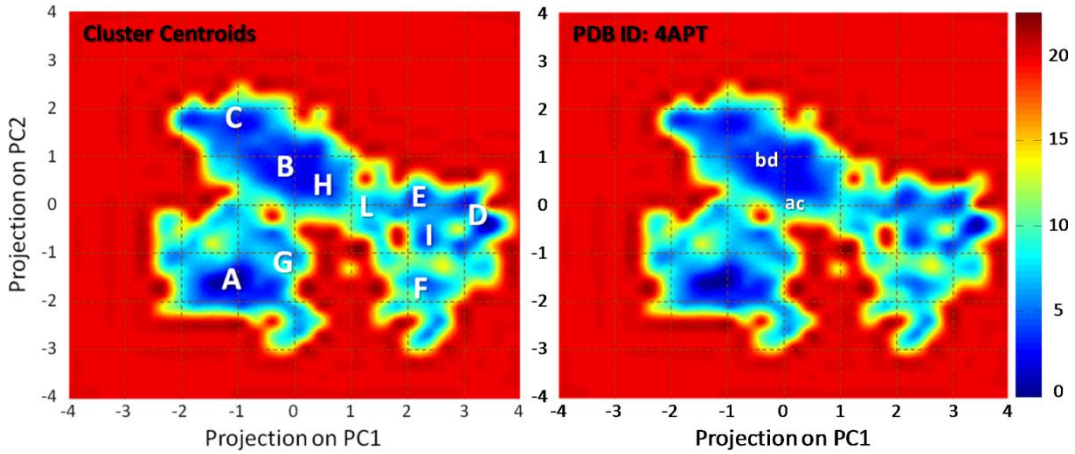


Figure 39. Free energy profile (kJ/mol) of the AXH[m] represented as function of the first and second Principal Component projection. Labels indicate the position on the free energy landscape in term of first and second Principal Component projection values of: left) cluster centroids obtained (labels A to L) by structure based clustering approach on REMD trajectory and right) AHX[m] atomic coordinates (label a,b,c,d) taken from 4APT PDB model[70].

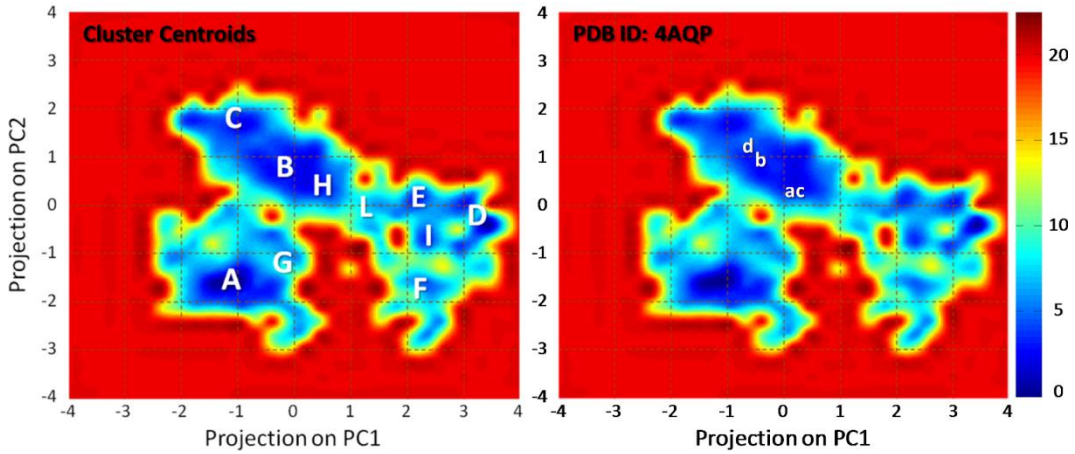


Figure 40. Free energy profile (kJ/mol) of the AXH[m] represented as function of the first and second Principal Component projection. Labels indicate the position on the free energy landscape in term of first and second Principal Component projection values of: left) cluster centroids obtained (labels A to L) by structure based clustering approach on REMD trajectory and right) AHX[m] atomic coordinates (label a,b,c,d) taken from 4AQP PDB model[70].

A-2 Characterization of the AXH Dimer by Functional Mode Analysis

Replica Exchange Molecular Dynamics

The latest refined crystal structure of the AXH dimer [70], was selected as a starting point for the present work. 64 replicas of each dimer structure (AB and CD) were simulated for 50 ns, achieving overall 6.5 μs of MD ensemble overall. Temperatures from 300 to 500 K were distributed across the replicas in a geometric progression, i.e. with the same ratio used to scale each temperature from the lower one, keeping the overlap of the potential energy distributions constant across the temperature axis. The

replicas' exchange interval of 1 ps was selected as being large enough compared to the coupling time of the heat bath (τ=0.1 ps). The resulting exchange probability was 0.4.

Functional Mode Analysis

Functional Mode Analysis was applied to REMD trajectories to reduce the dimensionality of the system, elucidating large-scale and low-frequency modes, respectively, thus yielding collective motions directly related to a specific molecular event [9]. Given a large set of conformational states taken from a REMD trajectory at 310 K, the method detects a collective motion maximally correlated to the fluctuation of the quantity of interest, in this case the inter-monomer distance. Assuming that the variable of interest is a linear function of the Principal Components, the maximally correlated vector can be derived by maximizing the Pearson coefficient [9] to quantify the contributions of the individual PCA vectors to the fluctuations of the variables of interest (Figure 41a). In Figure 41b is reported the scatter plot (data vs. model) of the model building set together with R_m, i.e. the correlation value between data and model in the model building set (R_m= 0.9948). FMA yields a single collective mode, which drives the phenomenon under investigation, referred to as ensemble-weighted Maximally Correlated Motion (ewMCM). The inter-monomer distance versus the projection p_a on the functional mode is reported in Figure 41c.

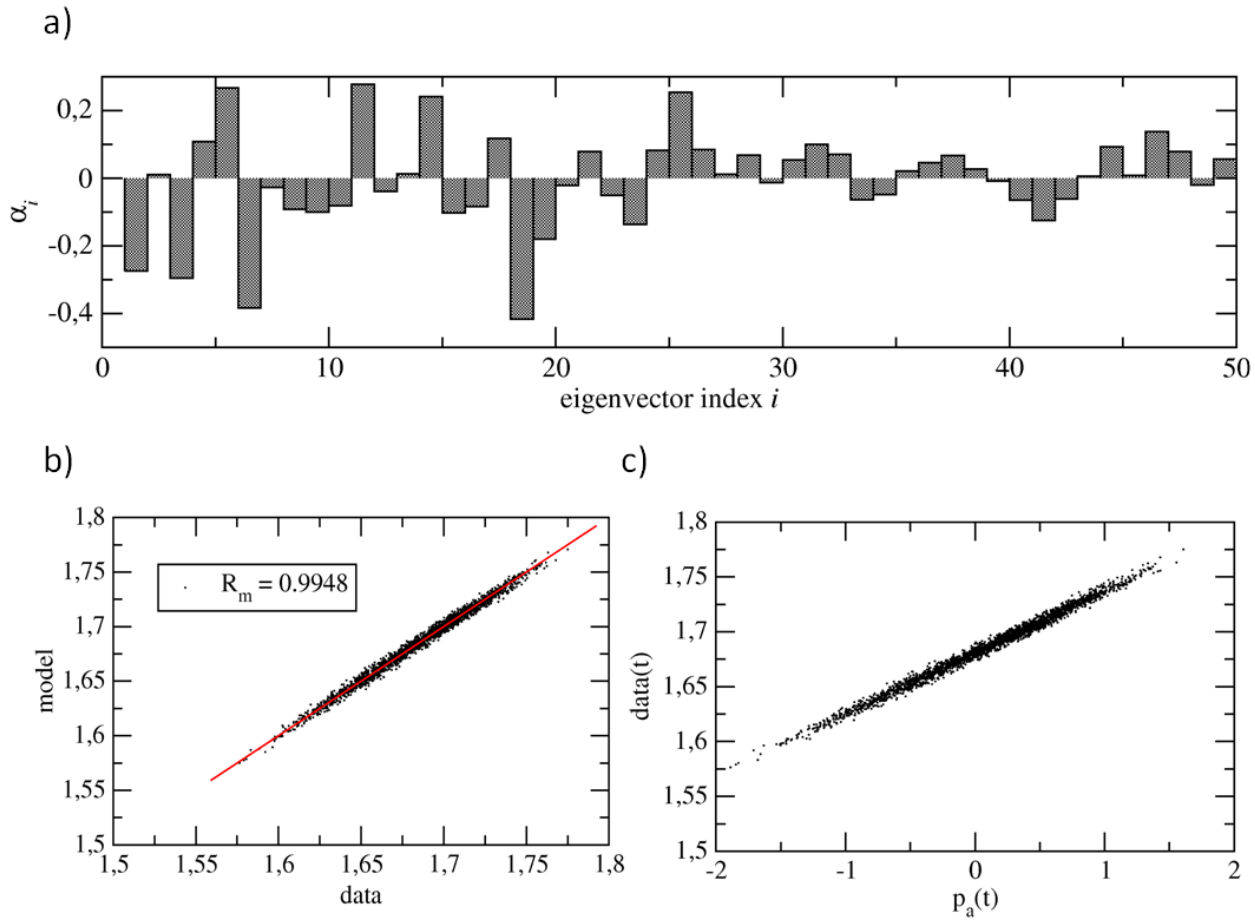


Figure 41. a) Coefficients α_i of the ewMCM with respect to the PCA vectors. b) scatter plot (data vs. model) of the model building set, also providing Rm, i.e. the correlation between data and model in the model building set. c) inter-monomer distance versus the projection p_a on the functional mode.

During the FMA calculation, it is essential to cross-validate the derived model for an independent set of simulation frames. The established approach to cross-validate the obtained results is to divide the simulation into frames for model building and for cross-validation (Figure 42a). The obtained maximally-correlated motion is validated in this work by predicting the function of interest in the cross-validation set with a Pearson correlation coefficient of R_c= 0.9941, as reported in Figure 42b.

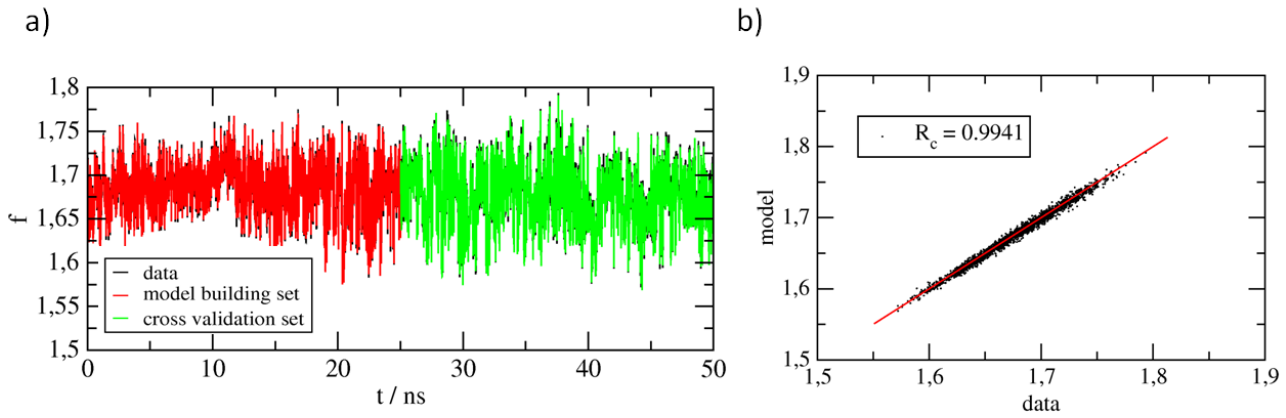


Figure 42. a) Inter-monomer fluctuation along the REMD trajectory at 310 K (black), the model in the model building set (red), and the model in the validation set as a function of time (green). b) Scatter plot (data vs. model) of the cross validation set, also providing the correlation Rc between data and model of the cross validation building set.

REMD structural integrity analysis at 310 K

Data regarding the structural integrity of the individual monomeric units are reported below.

Structure based GROMOS clustering approach[88], implemented in GROMACS package, was applied to the REMD trajectory at 310 K by using a cutoff of 0.18 nm. The centroid of the most populated cluster (Figure 43), representing over the 98% of the sampled REMD ensemble, was analyzed to gauge the quality of the ensemble generated by REMD approach. In detail, Ramachandran Plots of monomer A, B, C and D (Figure 44) show that protein residues were found in the most-favored and additional-allowed regions, providing support for the model's correctness. The C-alpha/C-alpha Root Mean Square Deviation for each step of the REMD trajectory at 310 K with respect the starting experimental conformational state, are represented as distribution plot in Figure 45.

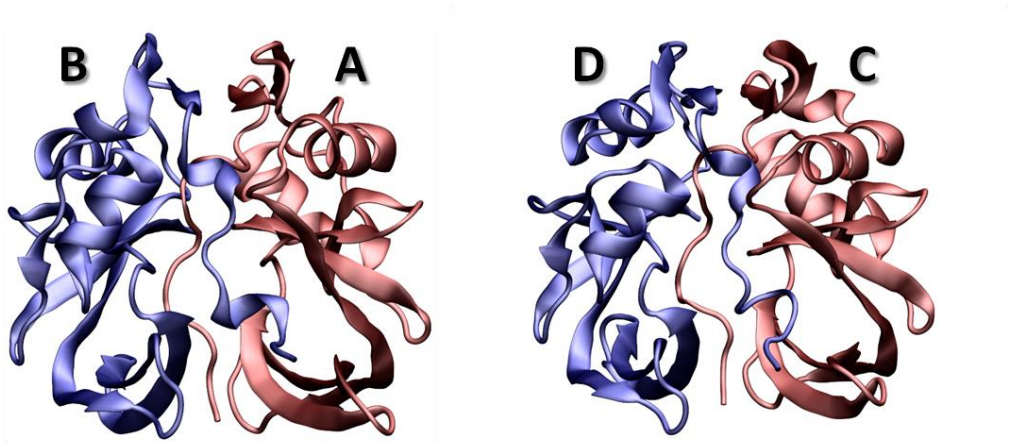
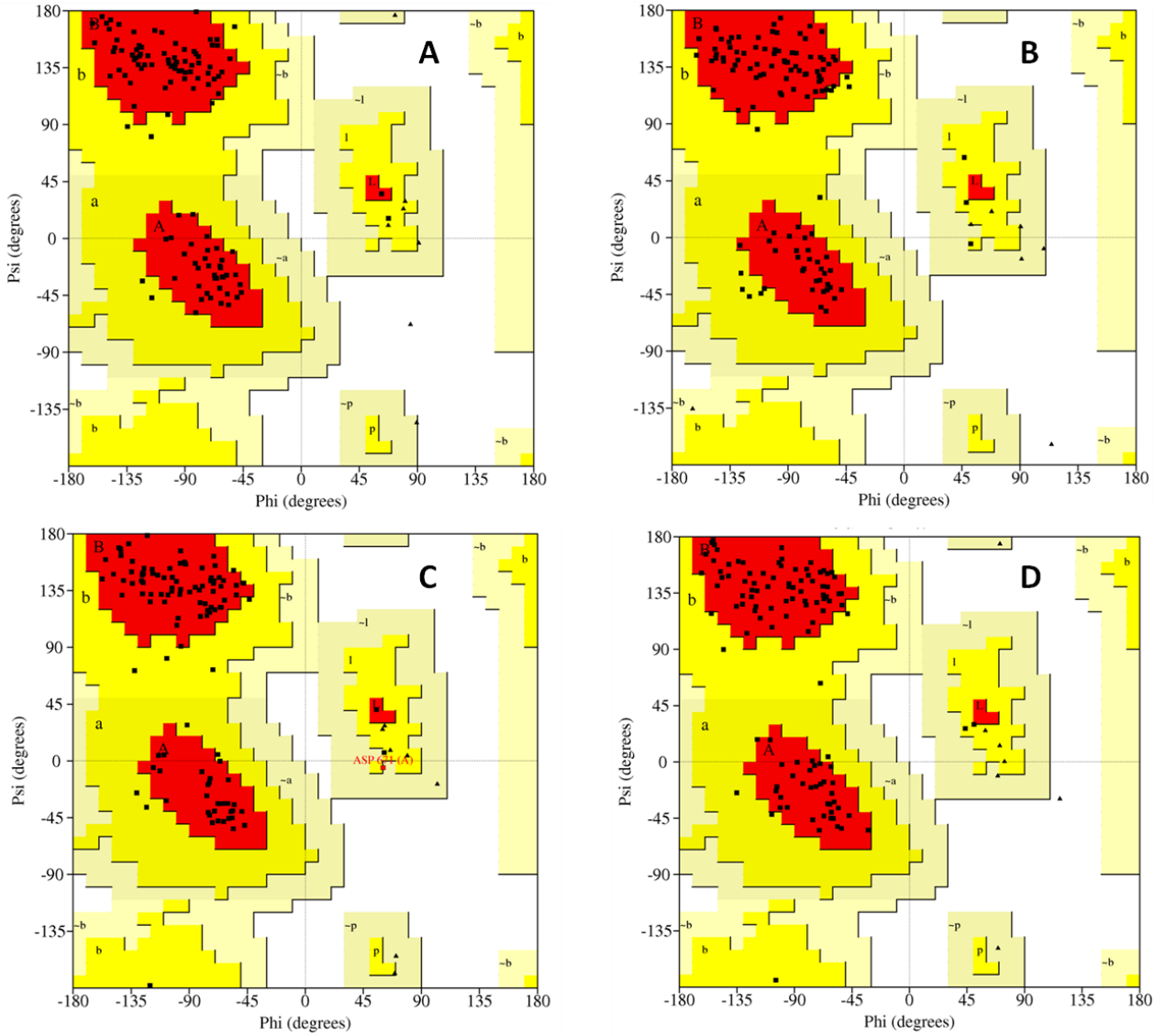


Figure 43. Visual inspection of the centroid of the most populated cluster, representing over the 98% of the sampled REMD ensemble at 310 K. The centroid was calculated by applying structure based GROMOS clustering approach[88], implemented in GROMACS package to the REMD trajectory at 310 K by using a cutoff of 0.18 nm.



	A	B	C	D
Residues in most favored regions	88%	93.5%	90.7%	91.6%
Residues in additional allowed regions	12%	6.5%	8.3%	8.4%
Residues in generously allowed regions	0%	0%	0%	0%
Residues in disallowed regions	0%	0%	0%	0%

Figure 44. Ramachandran plot of monomers A,B,C and D show that all protein residues were found in most-favored (red) and additional-allowed regions (yellow), providing support for the model's correctness. The centroid of the most populated cluster, representing over the 98% of the sampled REMD ensemble at 310 K, was considered as input for Procheck[86].

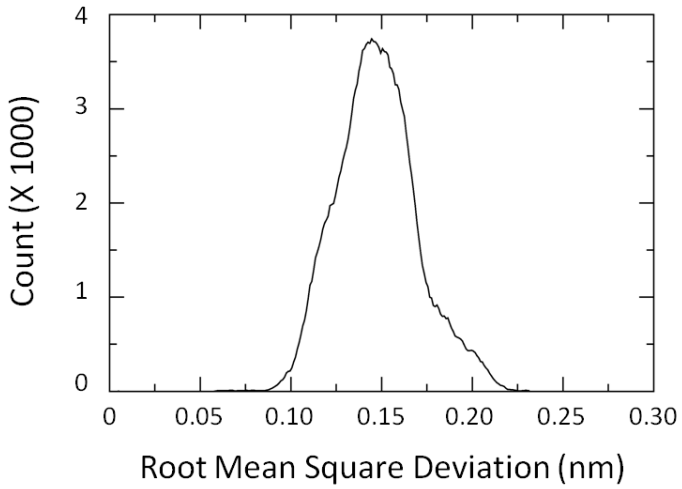


Figure 45. Distribution plot of the Root Mean Square Deviation calculated for each step of the REMD trajectory at 310 K with respect the starting experimental conformational state.

Classical Molecular Dynamics of AHX monomers

A classical MD simulation of 500 ns was carried out for each AHX dimer (AB and CD) in water as comparison to the REMD trajectory at 310 K, following the setup described in the main text. The visual inspection of the AXH dimers at the end of each MD simulation is reported in Figure 46. The dimerization interface in both AB and CD dimers buries 14.5 nm^2 of solvent accessible surface (Figure 48), in agreement with the computational data taken from the REMD trajectory at 310 K (Figure 15). More than 55% of the contact area is composed of hydrophobic contacts (Figure 48). C-alpha/C-alpha Root Mean Square Deviation for each considered dimer are reported in Figure 47.

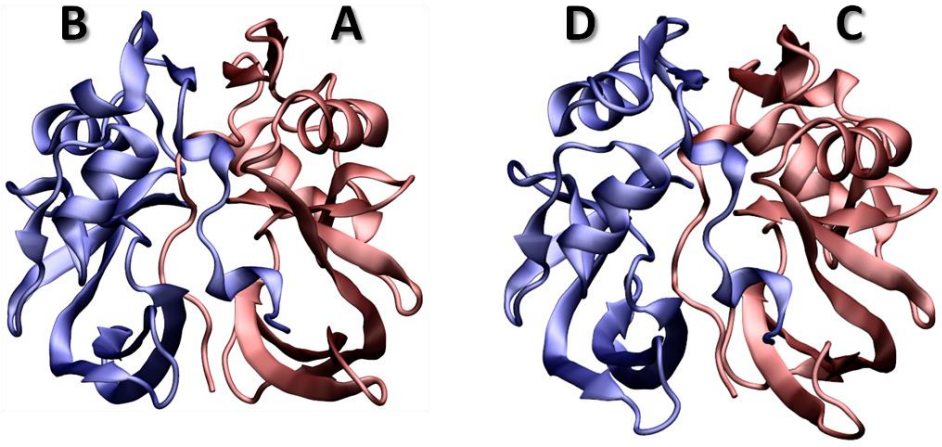


Figure 46. Picture of AHX dimer at the end of the simulation.

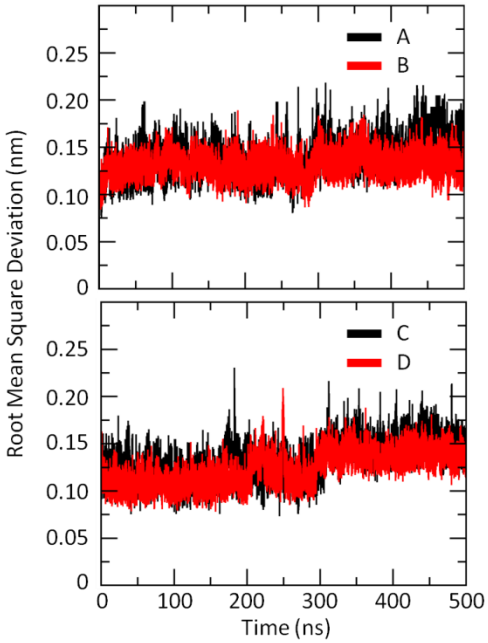


Figure 47. Root Mean Square Deviation calculated for each step of the REMD trajectory at 310 K with respect the starting experimental conformational state.

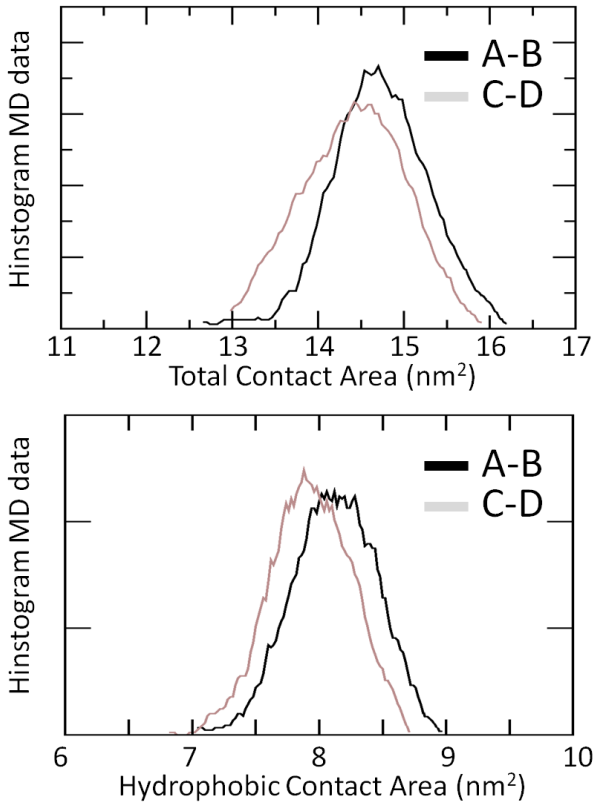


Figure 48. Distribution plot of the monomer-monomer total contact area calculated for each dimer (AB and CD), along the REMD trajectory at 310 K (top). Distribution plot of the monomer-monomer hydrophobic contact area calculated for each dimer (AB and CD), along the REMD trajectory at 310 K (bottom).

A-3 AXH tetramer: Molecular Mechanisms and Potential Anti-Aggregation Strategies

Protein Conformational Stability

The production MD simulation trajectory of 600 ns was analyzed in a structural stability region in order to explore the protein conformational properties of the AXHd-WT and AXHd-I580 simulations. Protein structural stability was analyzed by monitoring the time evolution of the Root Mean Square Deviation (Figure 49), demonstrating that a reasonable stability of the above-mentioned quantities has been reached in all cases in the last 100 ns of the simulation.

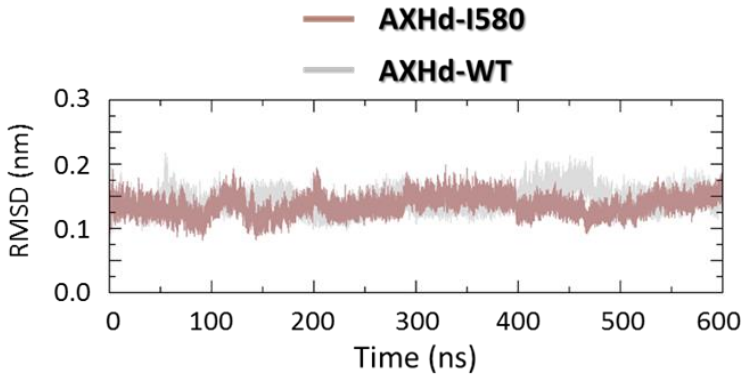


Figure 49. Time evolution of the Backbone/Backbone Root Mean Square Deviation (RMSD) calculated over the Classical MD trajectory in case of AXHd-WT, and AXHd-I580A.

To further test the conformational stability of the system, a cluster analysis was carried out on the last 400 ns of the MD trajectory (i.e., timer range 200-600 ns), with a threshold cutoff of 0.15 nm. Two clustering tests have been performed following single-linkage [138] and gromos [88]. In every performed test (*single-linkage* and *gromos*), one single cluster containing the 100% of explored conformations was detected for wild and mutated AXH simulations. It is worth mentioning that the RMSD cutoff threshold has been chosen based on experimental AXH dimer B-factors, which are directly related to RMSF of the atoms. The RMSD associated to the protein data bank B-factor for 4APT code was calculated equal to 0.15 nm.

Definition of the Collective Variables (CVs)

The free energy landscape representing the AXH dimer-dimer interaction was investigated by means of Metadynamics (metaD) [113,114]. In metaD, an external history-dependent bias potential which is a function of a few selected collective variables (CVs) is added to the Hamiltonian of the system. This potential is conceived as a sum of Gaussians deposited along the system trajectory in the CV space to

discourage the system from revisiting configurations that have already been sampled. In the widely used well-tempered procedure [29] the bias deposition rate decreases over simulation time as follows:

$F(s,t) = -\frac{T+\Delta T}{T}V(s,t)$, where $\gamma = -\frac{T+\Delta T}{T}$ is the bias factor

A bias factor of 25, 24, and 15 was used in case of AXHt-WT, AXHt-I580A, and AXHt-R638A simulation. Gaussian hills are added at a constant time interval in the position explored by the system in the space of selected Collective Variables (CVs).

Table 4 reports the CVs used for the reconstruction of the free energy landscape: a) distance between dimer AB and dimer CD, and b) the cosine of the angle between two vectors connecting the center of mass of distinct AXH regions per dimer (E636-R638 → T596-E604). The cosine CV is able to identify the relative rotation between dimers AB and CD, as explained in Figure 50.

Table 4. Collective Variables used in the reconstruction of the free energy landscape.

CV type	Protein Domains	Atom Type
Distance	DimerAB DimerCD	Cα
cos(α)	Vector1:E636-R638/T596-E604 (DimerAB) Vector2: E636-R638/T596-E604 (DimerCD)	Cα

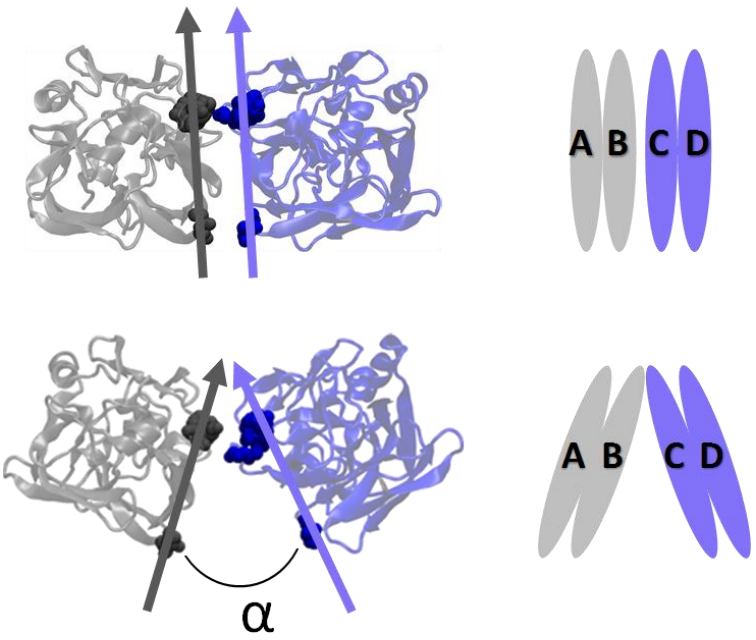


Figure 50. The representation of the cosine CV, which is able to identify the relative rotation between two dimers, AB and CD.

Convergence of the metadynamics free energy estimation

The convergence of the Metadynamics simulation was demonstrated by following a well-established computational procedure performed in a recent work [17], which require to check i) recrossing events between low energy states and ii) fluctuations of the free energy difference between the low-energy states around a specific value.

Several recrossing events between low energy states can be identified analyzing the time evolution of the CVs (Figure 51b and c, Figure 52b and c, Figure 53b and c). These events lead to the convergence in the estimation of the free energy landscape.

The free energy difference between low energy states at different times along the simulation (Figure 51d, Figure 52d, Figure 53d) was calculated to assess the convergence. The estimated free energy profile is reasonably stable in the last 50 ns of each simulation. It is worth mentioning that the uncertainty, calculated as the SD from the asymptotic value of the free energy obtained from the last part of the simulation, does not take into account the force-field inaccuracy.

As done in noteworthy reference works in the metadynamics research field [17] the diffusive behavior is deduced by checking the ability of the system to overcome the free energy barriers when the gaussian height is closed to zero. In all cases, molecular systems are able to get out of the bound state at the end the metadynamics simulation, when the energy added to the system is closed to zero (Figure 51b and c, Figure 52b and c, Figure 53b and c).

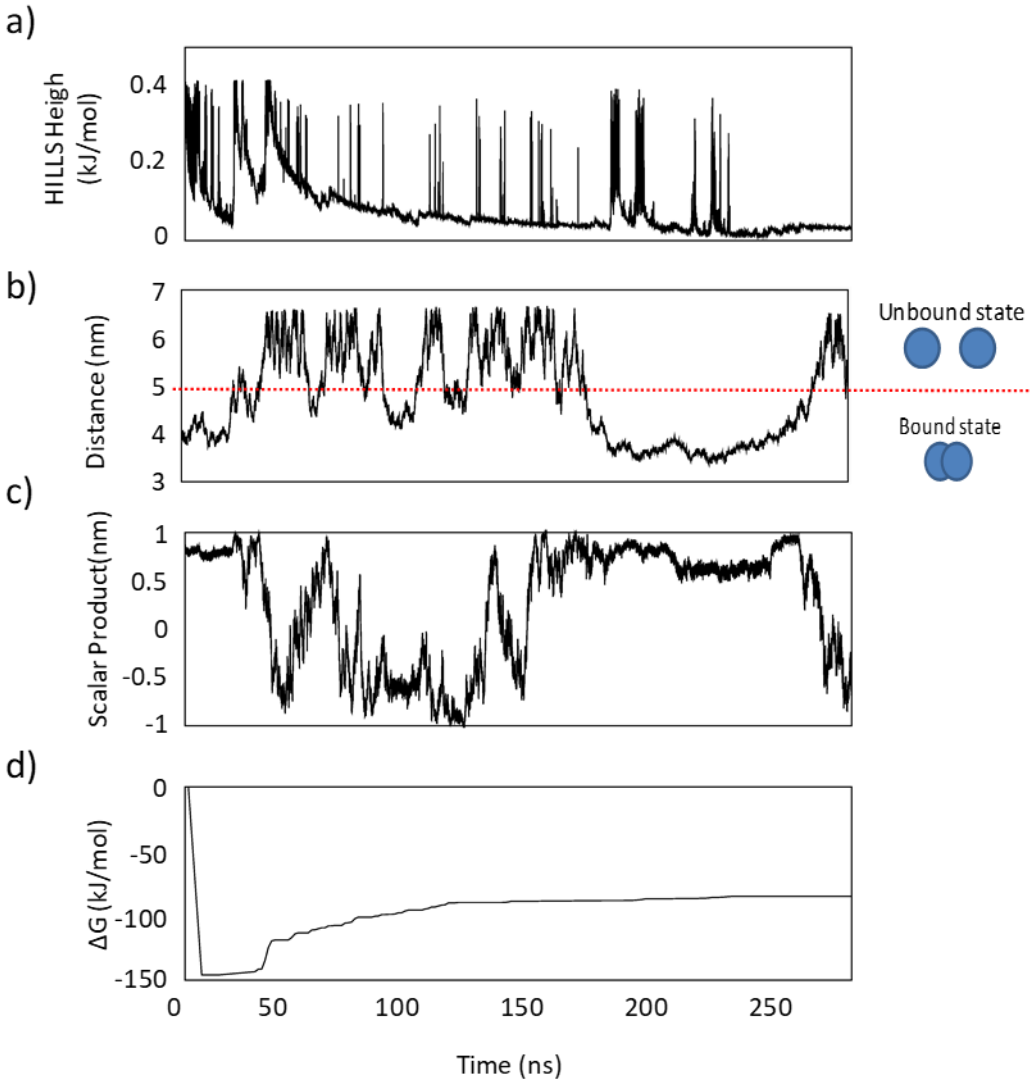


Figure 51. Convergence of the AXHt-WT metadynamics simulation. a) Plot of the Gaussian Height added to the system along the Metadynamics simulation. b) Plot of the dimer-dimer distance along the metadynamics simulation. c) Plot of the dimer-dimer scalar product along the metadynamics simulation. Several recrossing events between the low energy states can be identified. d) ΔG is calculated between state the first binding site and the unbound state to demonstrate the convergence of the free energy estimation. The uncertainty, calculated as the SD from the asymptotic value of the free energy obtained from the last part of the simulation is 4.0 kJ/mol.

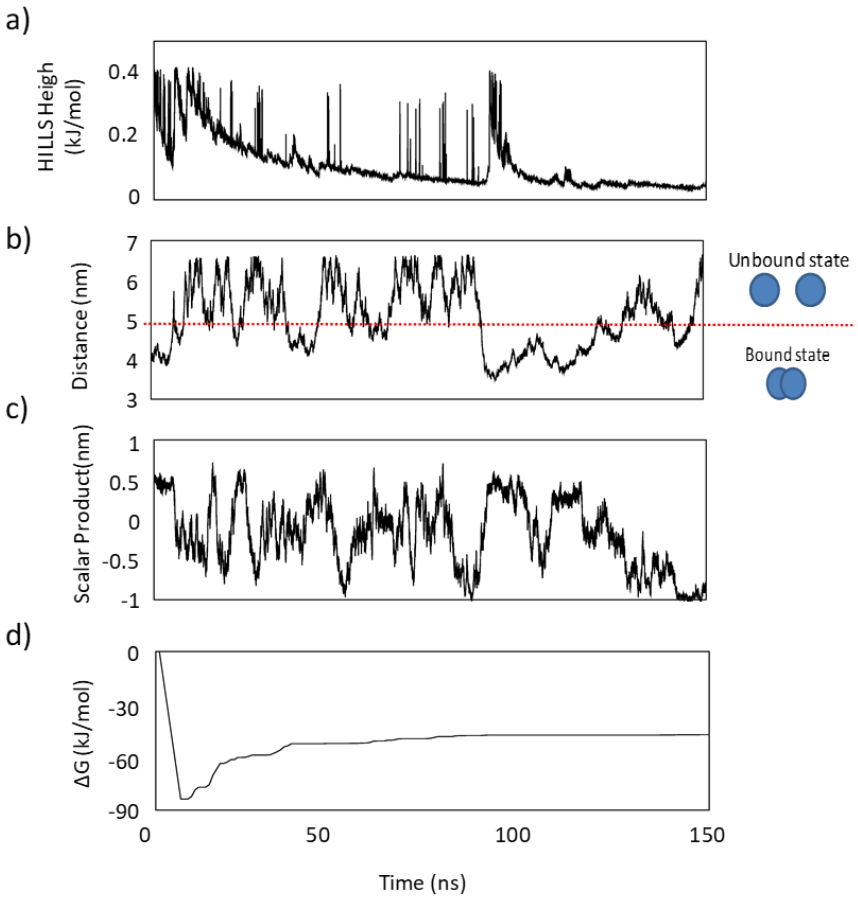


Figure 52. Convergence of the AXHt-I580A metadynamics simulation. a) Plot of the Gaussian Height added to the system along the Metadynamics simulation. b) Plot of the dimer-dimer distance along the metadynamics simulation. c) Plot of the dimer-dimer scalar product along the metadynamics simulation. Several recrossing events between the low energy states can be identified. d) ΔG is calculated between state the first binding site and the unbound state to demonstrate the convergence of the free energy estimation. The uncertainty, calculated as the SD from the asymptotic value of the free energy obtained from the last part of the simulation is 1.0 kJ/mol.

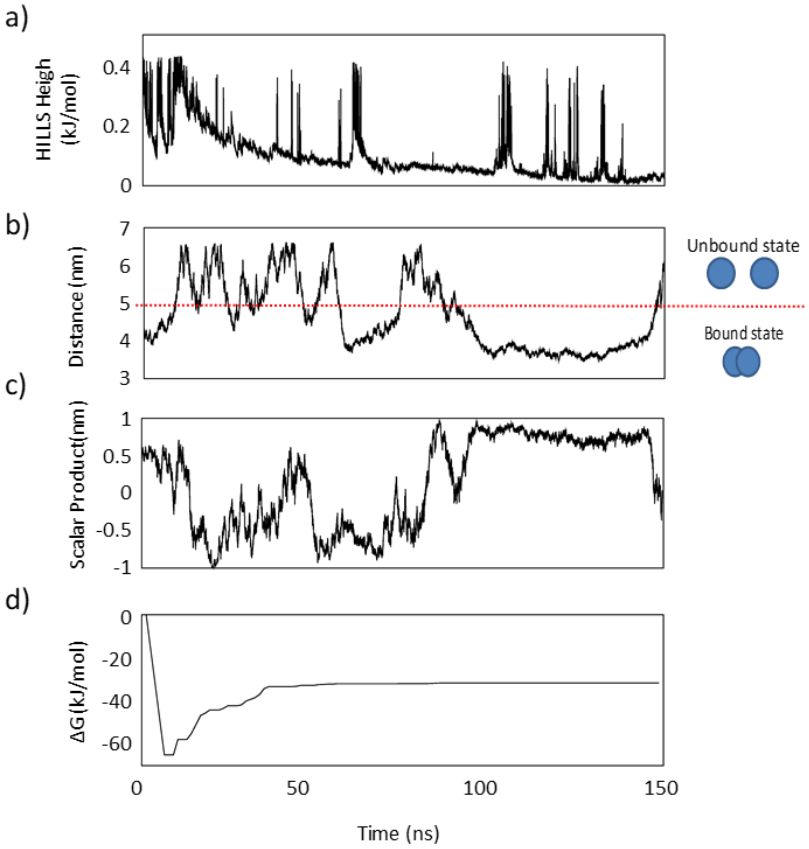


Figure 53. Convergence of the AXHt-R638A metadynamics simulation. a) Plot of the Gaussian Height added to the system along the Metadynamics simulation. b) Plot of the dimer-dimer distance along the metadynamics simulation. c) Plot of the dimer-dimer scalar product along the metadynamics simulation. Several recrossing events between the low energy states can be identified. d) ΔG is calculated between state the first binding site and the unbound state to demonstrate the convergence of the free energy estimation. The uncertainty, calculated as the SD from the asymptotic value of the free energy obtained from the last part of the simulation is 1.0 kJ/mol.

Consistency of classical MD and metadynamics simulations

It is generally expected a close agreement between the free energy minima (detected on metadynamics free energy landscape) and the system conformations sampled by the classical MD (i.e., structural equilibrium). To verify this hypothesis, each tetramer conformation sampled by Classical MD at the structural equilibrium with the metadynamics free energy maps were compared. Data clearly show (Figure 54) that free energy minima computed by metadynamics lie in regions regularly sampled by Classical MD at the structural equilibrium, confirming the reliability of the simulations.

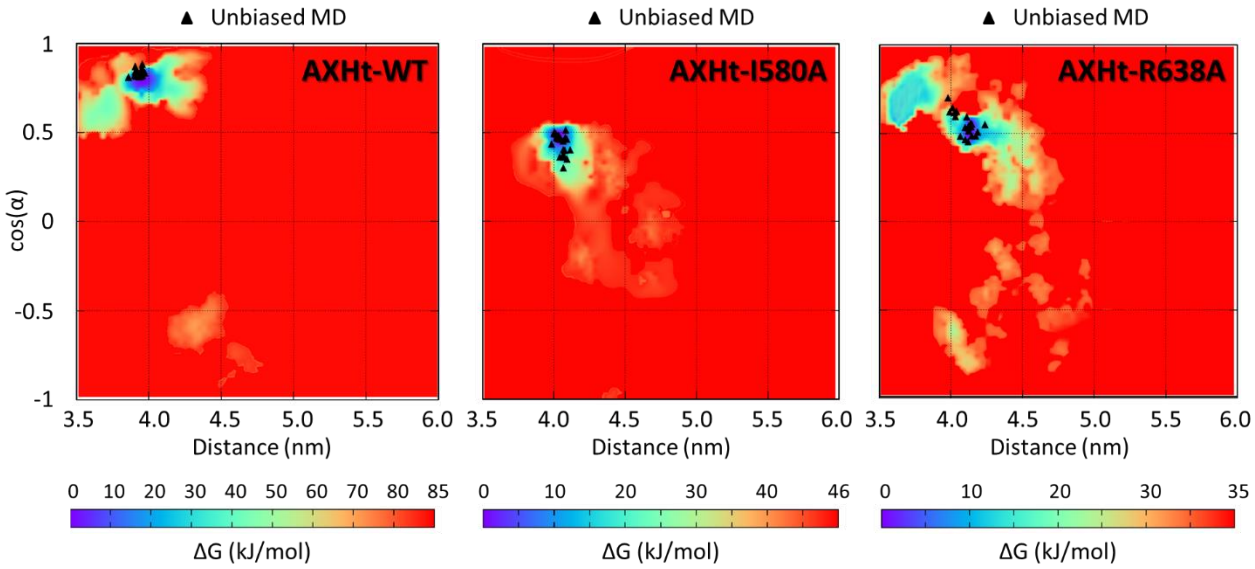


Figure 54. a) Free energy profile (kJ/mol) of the AXHt-WT, AXHt-I580A, and AXHt-R638A molecular systems (b), represented as function of two collective variables: the AXH dimer-dimer distance and the cosine of the angle between vectors connecting two distinct AXH regions per dimer. Black triangles show the sampled configurations throughout the last 100 ns of the unbiased (classical) MD.

Appendix-B

B-1 Elucidating the Conformational Stability of U- and S-Shaped Amyloid Beta Fibrils

Conformational Analysis of the REMD Ensemble at 300 K

The Solvent Accessible Surface Area (SASA) together with the estimation of its hydrophobic and hydrophilic component is shown in (Figure 55A). The total SASA of U-shaped model (76.07 ± 4.17 nm^2) is slightly higher than that of the S-shaped model (70.62 ± 3.71 nm^2). This result may be related to the S-shaped model ability to achieve a slightly more compact arrangement, as also shown by the Radius of Gyration (Figure 55B).

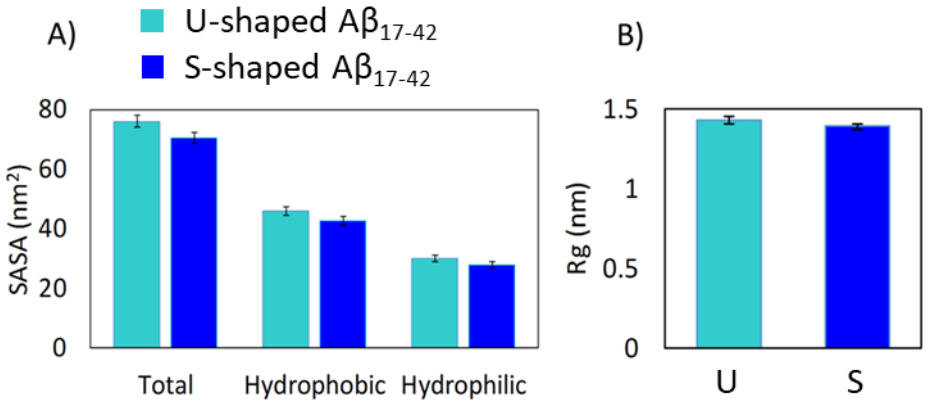


Figure 55. A) Total, Hydrophobic and Hydrophilic SASA of the two models compared. B) Radius of gyration of the two models.

Definition of the Order Parameter

With the purpose of comparing the structural order of the U-shaped and S-shaped models, an order parameter was defined.

$$ordP = \frac{1}{N_r}\sum_{r=17}^{42}\frac{\langle v_r, z\rangle}{\|v_r\|\cdot\|z\|} = \frac{1}{N_r}\sum_{r=17}^{42}\cos\ \alpha$$

Where r is the vector joining the N_r C_α-atoms pertaining to chain A with the corresponding C_α-atoms of chain E (i.e., on the same residue number) and z is the fibril axis. Values of *ordP* close to 1 indicate the amyloid-like shape alignment, whereas values of *ordP* much lower than 1 are typical of a distorted structure (Figure 56).

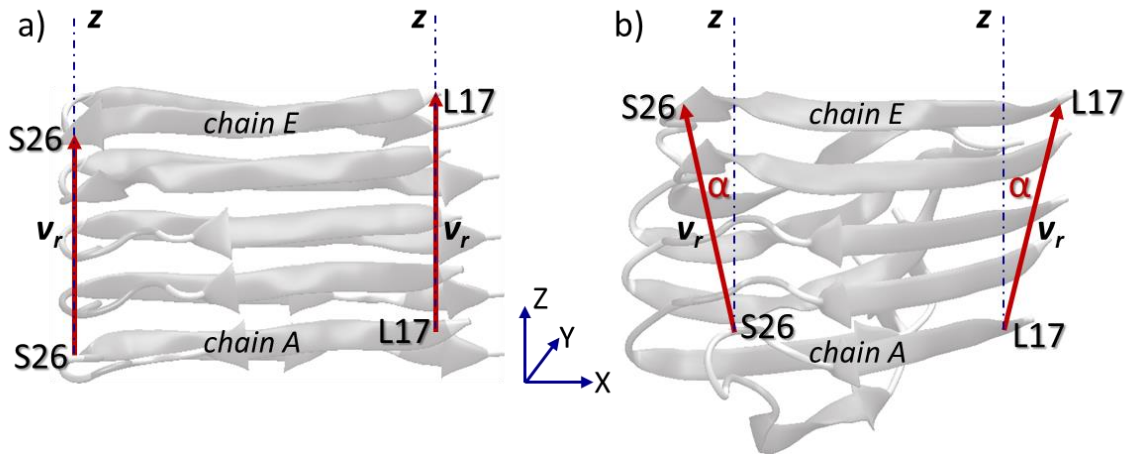


Figure 56. Schematic representation of the main players in determining the ordP value. The ordP value is equal to 1 in a completely aligned fiber (a), and lower than 1 for a distorted fiber (b).

Functional mode analysis

In Figure 57 the scatter plot of the data-model is reported for the two assemblies. The Pearson correlation coefficient is high for both models.

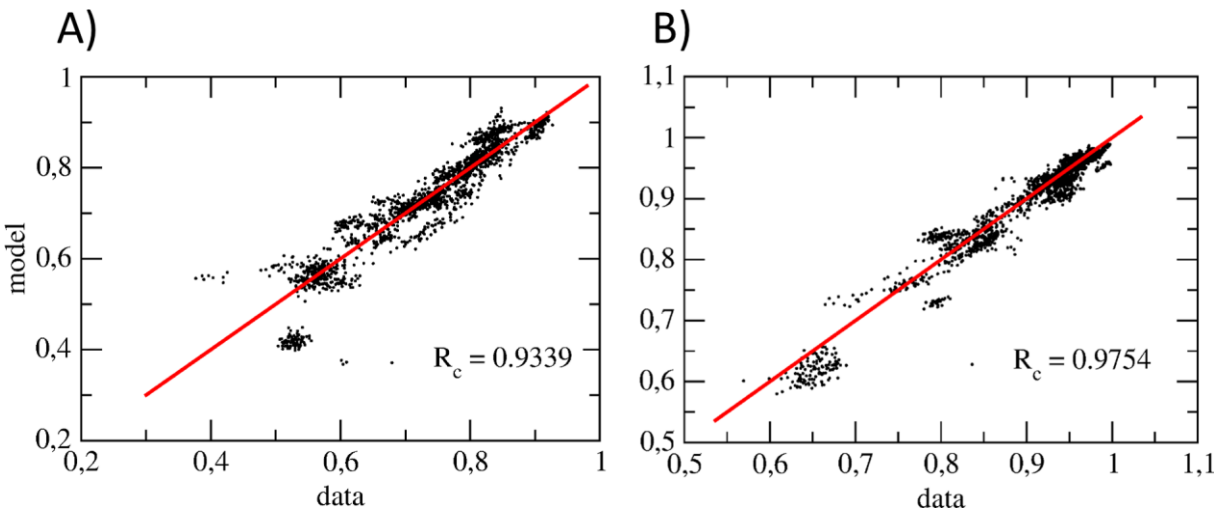


Figure 57. Scatter plots of the data-model for A) U-shaped and B) S-shaped model. In red the line corresponding to the cross-validation is reported.

The cross-validation of the maximally correlated motion was made by dividing the simulation in two parts: one for model building and the second for cross-validation as shown in Figure 58:

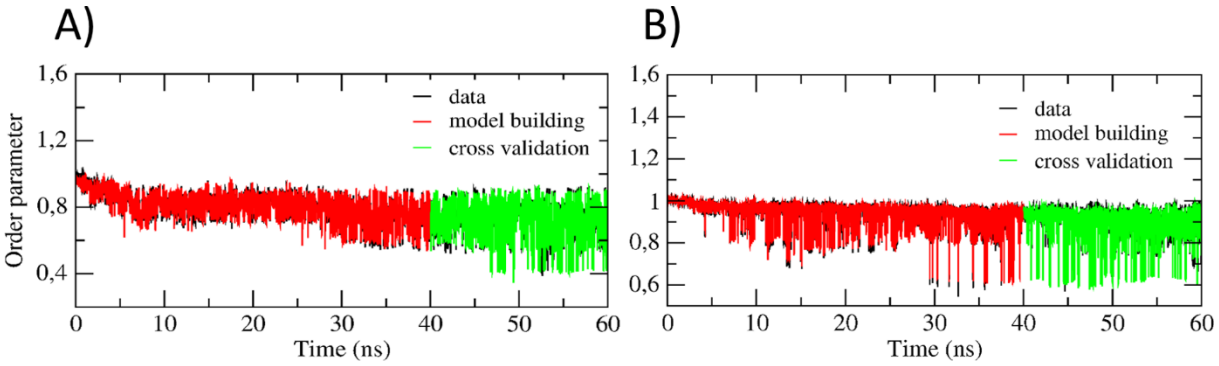


Figure 58. Order parameter as a function of time (black) obtained from REMD trajectories of U-shaped (A) and S-shaped (B) model. The model building curve (red) and the cross validation curve (green) are reported.

Stability of the S-shaped Aβ17-42 compared to $A\beta_{11-42}$

A pentamer of $A\beta_{11-42}$ extracted from 2MXU.pdb file was considered. As done for the $A\beta_{17-42}$ U-shaped and S-shaped models, the system was solvated by adding explicitly modelled water and ions. The simulation set-up in terms of system minimization, classical and replica exchange MD follows the same procedure described in the main text for the $A\beta_{17-42}$ U-shaped and S-shaped models. The ensemble at 300 K resulting from the REMD on the S-shaped $A\beta_{11-42}$ was considered to quantify the order parameter, *ordP*. The *ordP* distribution was then compared with the one estimated on the conformational ensemble at 300 K obtained by REMD simulations carried out on the S-shaped $A\beta_{17-42}$ model (Figure 59).

Results clearly highlight the stabilizing role played by residues E11-K16.

Hence, the comparison between the $A\beta_{17-42}$ U-shaped and S-shaped models presented in the main manuscript can be reasonable considered as not affected by the choice to neglect the residue range E11-K16 in the 2MXU.pdb file. On the other hand, that choice made meaningful the comparative analysis between the U-shaped and S-shaped models.

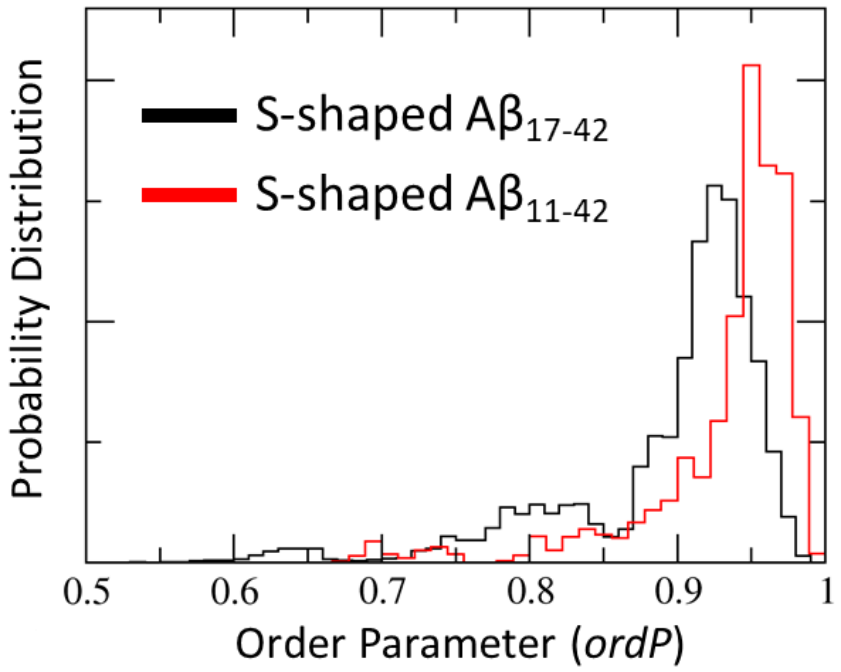


Figure 59. a) ordP distribution of the $A\beta_{17-42}$ S-shaped model and the $A\beta_{11-42}$ S-shaped model. The ordP distribution has been calculated throughout the overall REMD trajectory at 300 K taking the same number of snapshots for both models. It is worth mentioning how the structured protein portion E11-K16 provides support toward a higher fibril order.

B-2 Estimating the Mechanical Performance of U- and S-shaped Amyloid Beta Fibrils

Conformational Stability

Conformational Stability along all MD simulations was monitored through the Root Mean Square Deviation (RMSD) of C-alpha atoms of each model, as shown in Figure 60. Conformational stability is reasonably achieved in all cases in the last 25 ns.

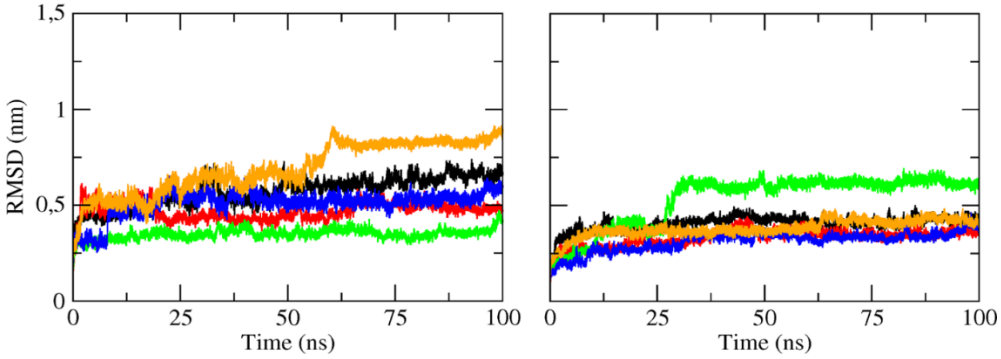


Figure 60: RMSD of C-alpha atoms for the five MD simulations carried out for U-shaped (left) and S-shaped (right) structures.

Mechanical properties of U- and S-architectures

In Figure 61 the mean peak force as a function of three different pulling velocities is reported for both U- and S-architectures, 2BEG and 2MXU models, respectively. Structures were subjected to three different deformation rates. Steered Molecular Dynamics were performed in 3 directions (X,Y,Z). For each direction, 5 replicas were run at three different rates for both models. Hence a total of 90 simulations were carried out.

The calculation of the mean peak force is dependent on the pulling velocities. In contrast, the difference between these values for each model subjected to the same deformation remains constant and is independent on the pulling velocities.

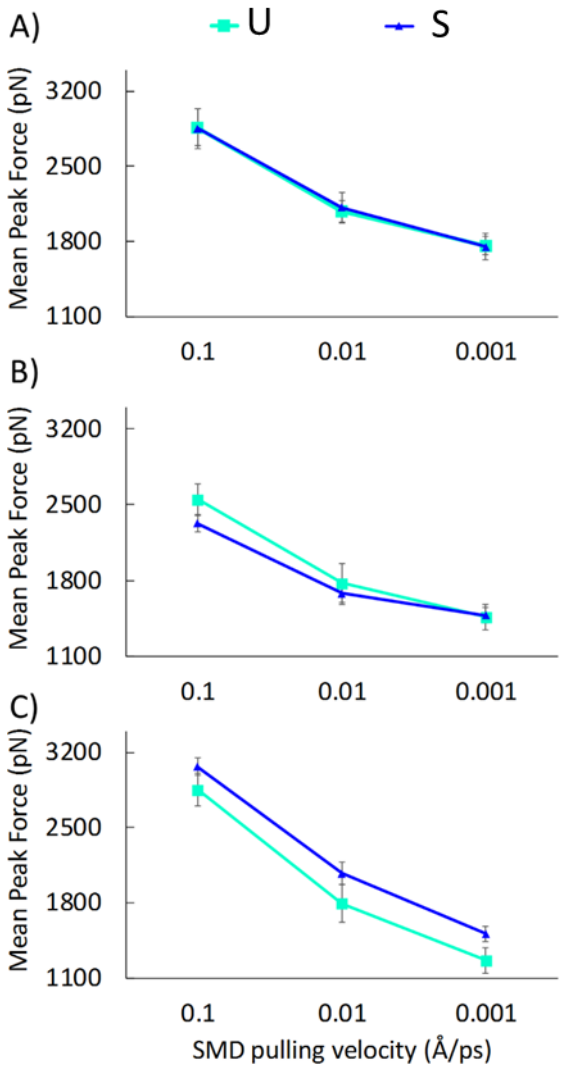


Figure 61: A) mean peak force for the U- and S-model calculated over the A) shear-X, B) shear-Y C) stretch-Z simulations run at three different pulling velocities, with five repeats in each case.

In Figure 62 the force-time profiles measured for U- and S-model in the three pulling directions X, Y and Z (pulling rate 0.01 Å/ps) are shown:

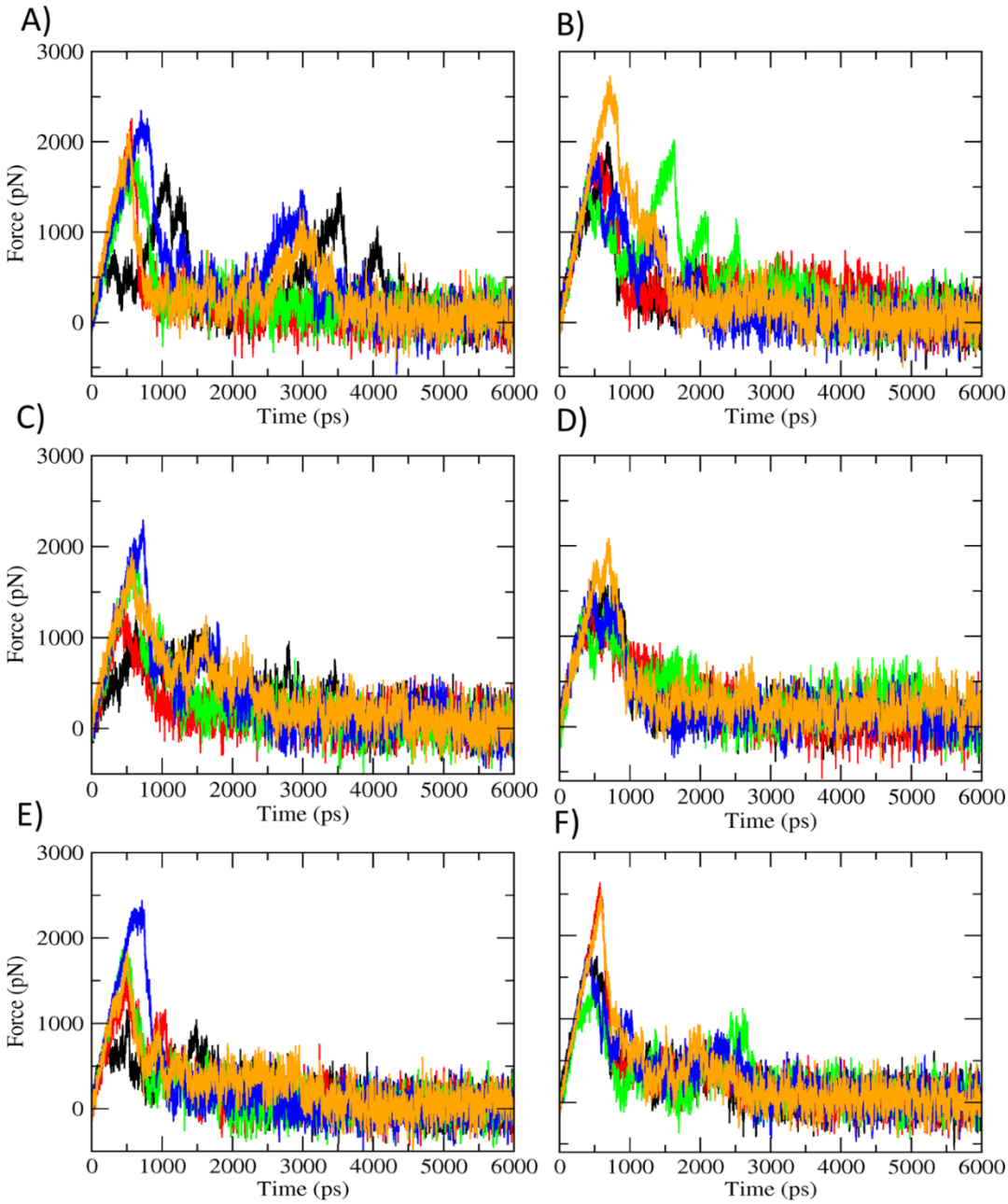


Figure 62: force-time profiles for the U- and S-models, on the left and right panel respectively, as the three different deformations are applied: A) and B) shear-X, C) and D) shear-Y, E) and F) stretch-Z. The colors black, red, blue, green, orange represent the five repeat trajectories carried out for each type of deformation.

Although, as expected, each replica follows a slightly different path through the conformational space, similar features are observed in all equivalent force-time profiles. In Figure 63 the representative force/displacement graphs are reported for each model subjected to the different deformations to show in which area the linear fit (identified with the red line) was performed in order to calculate the stiffness constant.

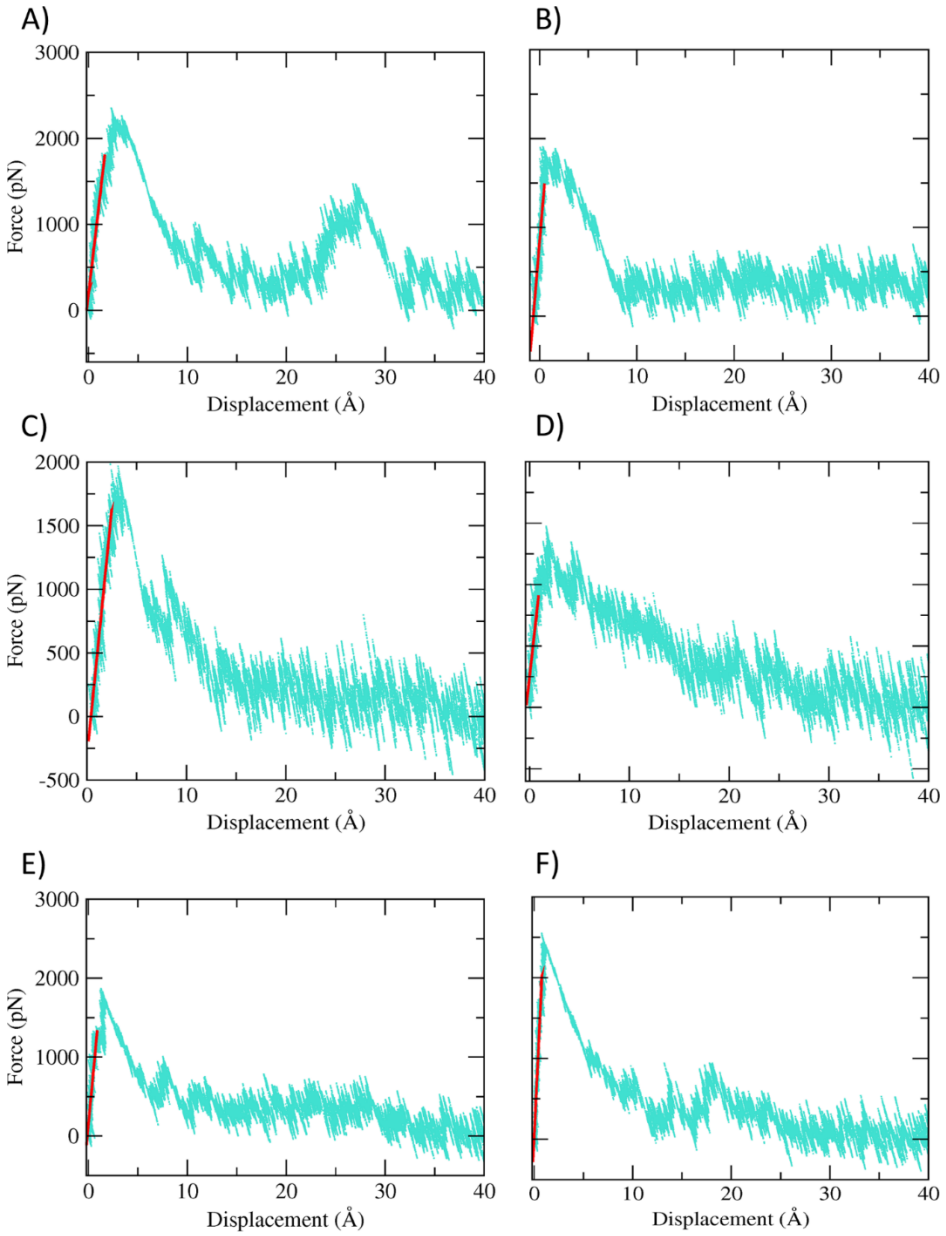


Figure 63: The representative graphs force versus displacement for U- and S-model are reported on the left and on the right side of the panel, respectively. In particular, A) and B) are obtained from the shear-X trajectories, C) and D) from the shear-Y trajectories and E) and F) from the stretch-Z trajectories. The red line represents the linear fit in the elastic regime whose slope is used to calculate the stiffness constant [pN/Å] of the fibril prior to failure.

The stress-strain curves presented in Figure 64 and Figure 65, are obtained from SMD force-displacement results along the pulling directions (i.e., X-, Y-, and Z- axis). Technically, each stress value was obtained by dividing the force with the interaction surface between the pulled chain B and the chain C. The application of a linear fitting to the stress-strain data in Figure 64 and Figure 65, allowed the estimation of the Young and Shear moduli (E, G_x, and G_y) of both the two-fibril configuration.

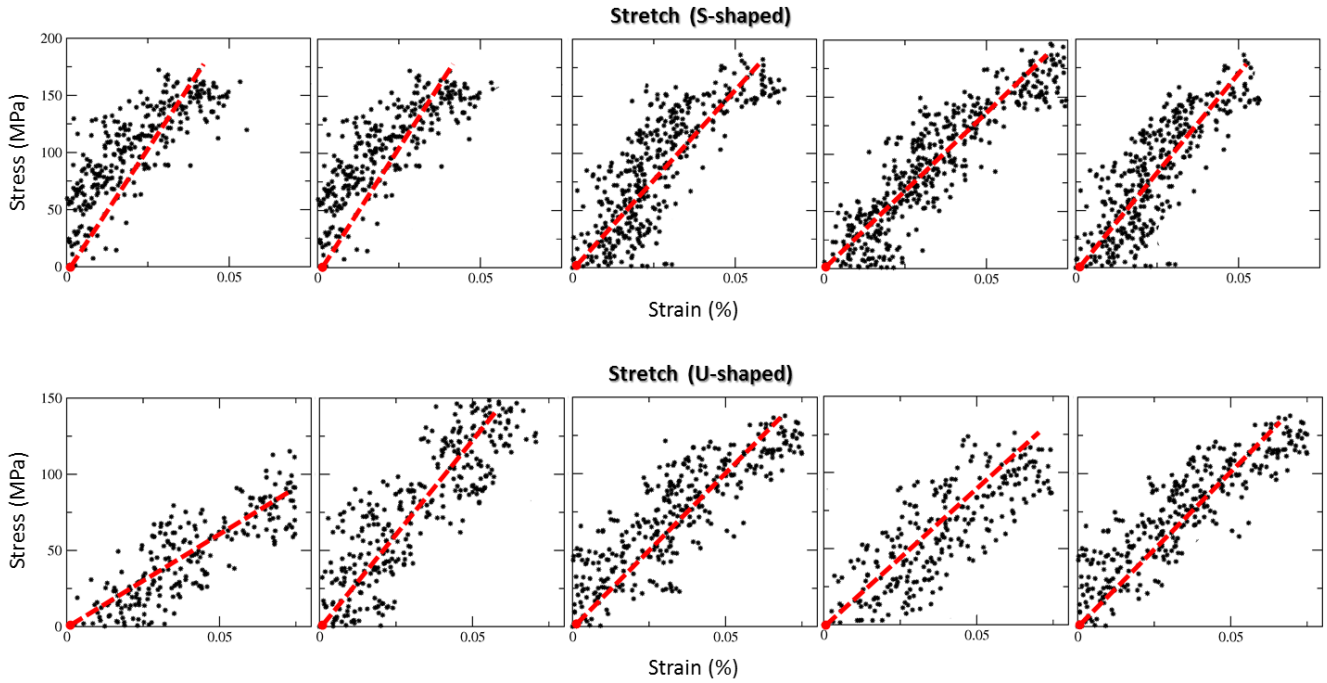


Figure 64. Stress-strain curve obtained from the force-displacement results of the two models subjected to the stretching deformation along the fibril axis (Z-direction). The stress is calculated as (force [pN])/(buried surface [nm^2]). The force is applied along the Z- axis. The buried surface is calculated between the chain B and chain C. The solid line represents the linear fit (correlation coefficient higher than 0.8 in all cases) in the elastic regime and small deformations. The linear fit slope is the Young's modulus [GPa] of the fibrils.

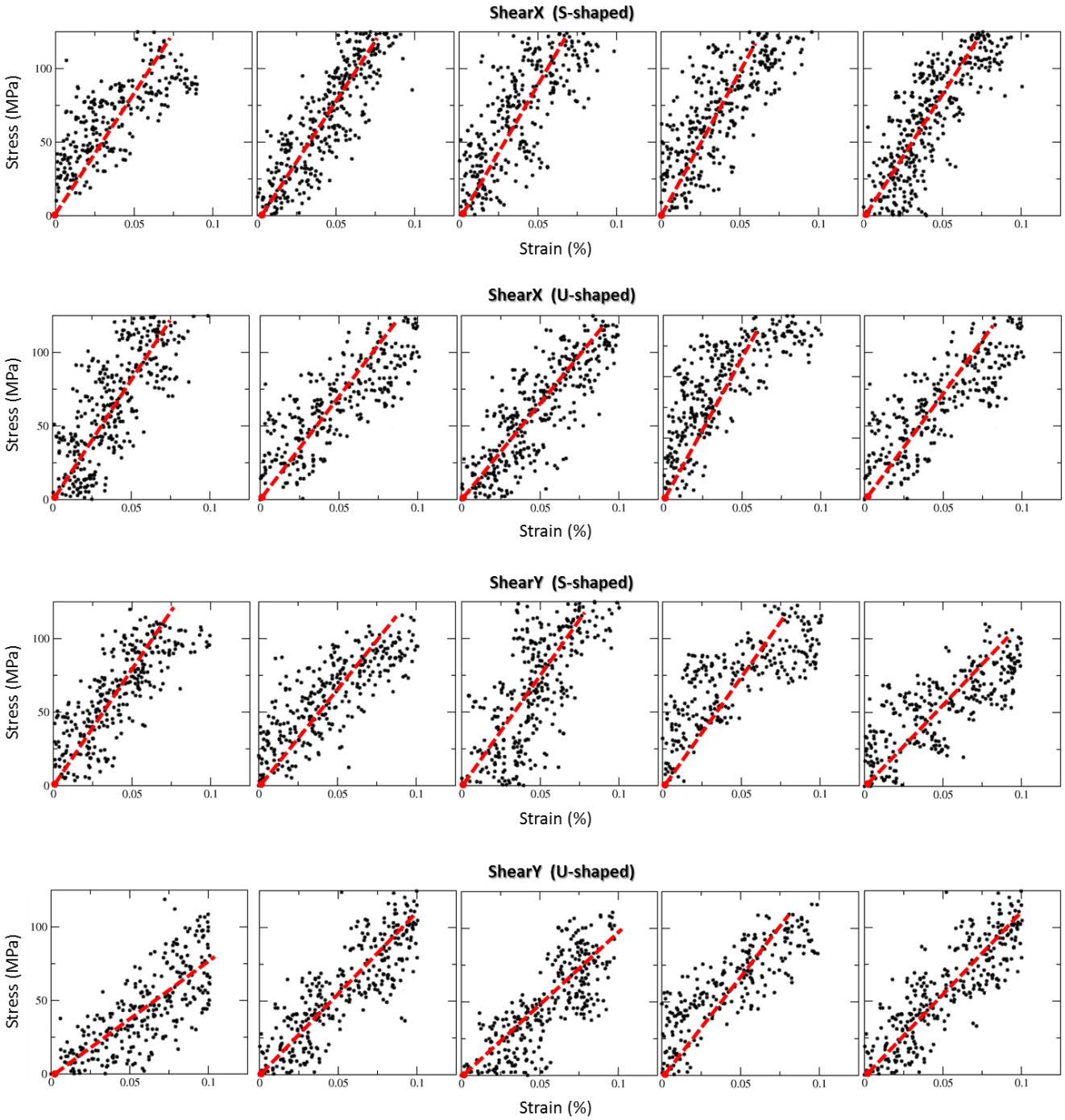


Figure 65. Stress-strain curve obtained from the force-displacement results of the two models subjected to the shear deformation along the X- and Y-direction. The stress is calculated as (force [pN])/(buried surface [nm^2]). The force is applied along the X- and Y- axis, respectively. The buried surface is calculated between the chain B and chain C. The solid line represents the linear fit (correlation coefficient higher than 0.8 in all cases) in the elastic regime and small deformations. The linear fit slope is the Young's modulus [GPa] of the fibrils.

In Figure 66 the Young's modulus as a function of three different pulling velocities is reported for both U- and S-models subjected to the stretching deformation. Lowering the pulling rate from 0.01 to 0.001 Å/ps, we did not detect, on the same model, statistically significant differences of Young's modulus.

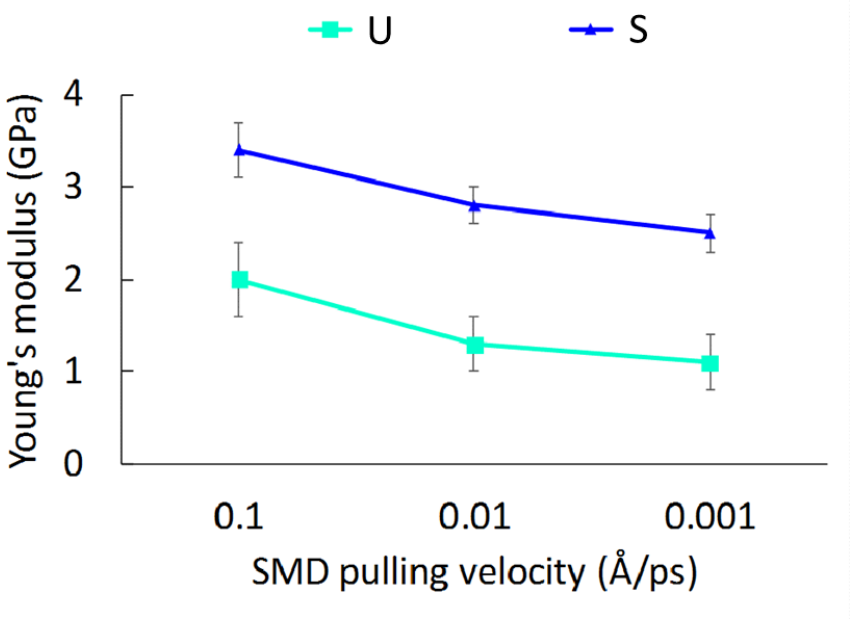


Figure 66. Young's modulus [GPa] for the U- and S-model calculated over the stretch-Z simulations run at three different pulling velocities, with five repeats.

References

1. Leach, A. R. *Molecular Modelling: Principles and Applications*. *Computers* **21,** (Prentice Hall, 2001).
2. Verlet, L. Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159,** 98–103 (1967).
3. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76,** 637–649 (1982).
4. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4,** 435–447 (2008).
5. Kitao, A., Hirata, F. & Gō, N. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **158,** 447–472 (1991).
6. Hayward, S. & Go, N. Collective Variable Description of Native Protein Dynamics. *Annu. Rev. Phys. Chem.* **46,** 223–250 (1995).
7. Balsera, M. a., Wriggers, W., Oono, Y. & Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **100,** 2567–2572 (1996).
8. Kitao, A., Hayward, S. & Go, N. Energy landscape of a native protein: Jumping-among-minima model. *Proteins Struct. Funct. Genet.* **33,** 496–517 (1998).
9. Hub, J. S. & de Groot, B. L. Detection of Functional Modes in Protein Dynamics. *PLoS Comput. Biol.* **5,** e1000480 (2009).
10. Krivobokova, T., Briones, R., Hub, J. S., Munk, A. & de Groot, B. L. Partial Least-Squares Functional Mode Analysis: Application to the Membrane Proteins AQP1, Aqy1, and CLC-ec1. *Biophys. J.* **103,** 786–796 (2012).
11. Narayanan, C., Weinstock, D. S., Wu, K.-P., Baum, J. & Levy, R. M. Investigation of the Polymeric Properties of α-Synuclein and Comparison with NMR Experiments: A Replica Exchange Molecular Dynamics Study. *J. Chem. Theory Comput.* **8,** 3929–3942 (2012).
12. Wu, K.-P., Weinstock, D. S., Narayanan, C., Levy, R. M. & Baum, J. Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J. Mol. Biol.* **391,** 784–96 (2009).
13. Wei, G. & Shea, J.-E. Effects of solvent on the structure of the Alzheimer amyloid-beta(25-35) peptide. *Biophys. J.* **91,** 1638–47 (2006).
14. Spiwok, V., Lipovová, P. & Králová, B. Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B* **111,** 3073–6 (2007).
15. Laio, A. & Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **71,** 126601 (2008).
16. Granata, D., Camilloni, C., Vendruscolo, M. & Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci.* **110,** 6817–6822 (2013).
17. Limongelli, V., Bonomi, M. & Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 6358–6363 (2013).
18. Sutto, L., Marsili, S. & Gervasio, F. L. New advances in metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2,** 771–779 (2012).
19. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314,** 141–151 (1999).
20. Mori, T., Miyashita, N., Im, W., Feig, M. & Sugita, Y. Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochim. Biophys. Acta - Biomembr.* **1858,** 1635–1651 (2016).
21. Schlick, T. Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000 Biol. Rep.* **1,** 51 (2009).
22. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **119,** 3559 (2003).
23. Park, S. & Sculten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120,** 5946–5961 (2004).
24. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78,** 2690–2693 (1997).
25. Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. Lett.* **56,** 5018–5035 (1997).
26. Tiana, G. Estimation of microscopic averages from metadynamics. *Eur. Phys. J. B* **63,** 235–238 (2008).
27. Marinelli, F., Pietrucci, F., Laio, A. & Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **5,** e1000452 (2009).
28. Bonomi, M., Barducci, A. & Parrinello, M. Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **30,** 1615–1621 (2009).
29. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100,** 020603 (2008).
30. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14,** 70–5 (2004).
31. Horwich, A. Protein aggregation in disease: A role for folding intermediates forming specific multimeric interactions. *Journal of Clinical Investigation* **110,** 1221–1232 (2002).
32. Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15,** 3–16 (2004).
33. Singh, S. M., Cabello-Villegas, J., Hutchings, R. L. & Mallela, K. M. G. Role of partial protein unfolding in alcohol-induced protein aggregation. *Proteins Struct. Funct. Bioinforma.* **78,** 2625–2637 (2010).
34. Day, R., Bennion, B. J., Ham, S. & Daggett, V. Increasing temperature accelerates protein unfolding without changing the

pathway of unfolding. *J. Mol. Biol.* **322,** 189–203 (2002).
35. Ellisdon, A. M., Pearce, M. C. & Bottomley, S. P. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J. Mol. Biol.* **368,** 595–605 (2007).
36. Sharma, S., Berne, B. J. & Kumar, S. K. Thermal and structural stability of adsorbed proteins. *Biophys. J.* **99,** 1157–1165 (2010).
37. Uversky, V. N. & Fink, A. L. Conformational constraints for amyloid fibrillation: The importance of being unfolded. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1698,** 131–153 (2004).
38. Kayed, R. *et al.* Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases. *J. Biol. Chem.* **279,** 46363–6 (2004).
39. Krishnan, R. & Lindquist, S. L. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435,** 765–772 (2005).
40. Chiti, F. & Dobson, C. M. C. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).
41. Chiti, F. & Dobson, C. M. C. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5,** 15–22 (2008).
42. Cummings, C. J. & Zoghbi, H. Y. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* **9,** 909–916 (2000).
43. Scherzinger, E. *et al.* Huntingtin encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90,** 549–558 (1997).
44. Chen, Y. W., Allen, M. D., Veprintsev, D. B., Löwe, J. & Bycroft, M. The Structure of the AXH Domain of Spinocerebellar Ataxin-1. *J. Biol. Chem.* **279,** 3758–3765 (2004).
45. Zoghbi, H. Y. & Orr, H. T. Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, Spinocerebellar ataxia type 1. *Journal of Biological Chemistry* **284,** 7425–7429 (2009).
46. Emamian, E. S. *et al.* Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. *Neuron* **38,** 375–387 (2003).
47. Klement, I. A. *et al.* Ataxin-1 nuclear localization and aggregation: Role in polyglutarnine- induced disease in SCA1 transgenic mice. *Cell* **95,** 41–53 (1998).
48. De Chiara, C., Menon, R. P., Dal Piaz, F., Calder, L. & Pastore, A. Polyglutamine is not all: The functional role of the AXH domain in the ataxin-1 protein. *J. Mol. Biol.* **354,** 883–893 (2005).
49. Saunders, H. M. & Bottomley, S. P. Multi-domain misfolding: Understanding the aggregation pathway of polyglutamine proteins. *Protein Engineering, Design and Selection* **22,** 447–451 (2009).
50. Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256,** 184–5 (1992).
51. Reitz, C. Alzheimer's Disease and the Amyloid Cascade Hypothesis: A Critical Review. *Int. J. Alzheimers. Dis.* **2012,** 1–11 (2012).
52. Aguzzi, A. & O'Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat. Rev. Drug Discov.* **9,** 237–248 (2010).
53. Querfurth, H. W. & LaFerla, F. M. Alzheimer's Disease. *N. Engl. J. Med.* **362,** 329–44 (2010).
54. Jarrett, J. T. & Lansbury, P. T. Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73,** 1055–1058 (1993).
55. Cummings, J. L. Alzheimer's Disease. *N. Engl. J. Med.* **351,** 56–67 (2004).
56. Schütz, A. K. *et al.* Atomic-Resolution Three-Dimensional Structure of Amyloid β Fibrils Bearing the Osaka Mutation. *Angew. Chemie Int. Ed.* **54,** 331–335 (2015).
57. Petkova, A. T. *et al.* A structural model for Alzheimer's -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci.* **99,** 16742–16747 (2002).
58. Luhrs, T. *et al.* 3D structure of Alzheimer's amyloid- (1-42) fibrils. *Proc. Natl. Acad. Sci.* **102,** 17342–17347 (2005).
59. Petkova, A. T., Yau, W.-M. & Tycko, R. Experimental Constraints on Quaternary Structure in Alzheimer's β-Amyloid Fibrils †. *Biochemistry* **45,** 498–512 (2006).
60. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's -amyloid fibrils. *Proc. Natl. Acad. Sci.* **105,** 18349–18354 (2008).
61. Lu, J.-X. *et al.* Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **154,** 1257–1268 (2013).
62. Gremer, L. *et al.* Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. *Science (80-.).* **358,** 116–119 (2017).
63. 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).
64. Gu, L., Tran, J., Jiang, L. & Guo, Z. A new structural model of Alzheimer's Aβ42 fibrils based on electron paramagnetic resonance data and Rosetta modeling. *J. Struct. Biol.* **194,** 61–67 (2016).
65. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic Aβ 42 Amyloid Fibrils. *J. Am. Chem. Soc.* **138,** 9663–9674 (2016).
66. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proc. Natl. Acad. Sci.* **113,** E4976–E4984 (2016).
67. Schmidt, M. *et al.* Peptide dimer structure in an Aβ(1–42) fibril visualized with cryo-EM. *Proc. Natl. Acad. Sci.* **112,** 11858–11863 (2015).
68. Xi, W. & Hansmann, U. H. E. Ring-like N-fold Models of Aβ42 fibrils. *Sci. Rep.* **7,** 6588 (2017).
69. de Chiara, C. *et al.* The AXH domain adopts alternative folds the solution structure of HBP1 AXH. *Structure* **13,** 743–53 (2005).
70. De Chiara, C. *et al.* Self-assembly and conformational heterogeneity of the AXH domain of ataxin-1: An unusual example of a chameleon fold. *Biophys. J.* **104,** 1304–1313 (2013).
71. Tsuda, H. *et al.* The AXH domain of ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/senseless proteins. *Cell* **122,** 633–644 (2005).

72. de Chiara, C. *et al.* The AXH module: an independently folded domain common to ataxin-1 and HBP1. *FEBS Lett.* **551,** 107–112 (2003).
73. de Chiara, C. & Pastore, A. in *Supramolecular Structure and Function 10* 87–99 (Springer Netherlands, 2011). doi:10.1007/978-94-007-0893-8_5
74. Menon, R. P. *et al.* Mapping the self-association domains of ataxin-1: identification of novel non overlapping motifs. *PeerJ* **2,** e323 (2014).
75. de Chiara, C., Menon, R. P., Kelly, G. & Pastore, A. Protein-Protein Interactions as a Strategy towards Protein-Specific Drug Design: The Example of Ataxin-1. *PLoS One* **8,** e76456 (2013).
76. 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).
77. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65,** 712–25 (2006).
78. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78,** 1950–8 (2010).
79. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS One* **7,** e32131 (2012).
80. 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 (1983).
81. 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).
82. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).
83. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–8, 27–8 (1996).
84. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23,** 566–79 (1995).
85. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32,** W500–W502 (2004).
86. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
87. Nguyen, P. H., Stock, G., Mittag, E., Hu, C. K. & Li, M. S. Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study. *Proteins Struct. Funct. Genet.* **61,** 795–808 (2005).
88. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **38,** 236–240 (1999).
89. Maisuradze, G. G., Liwo, A. & Scheraga, H. a. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **385,** 312–29 (2009).
90. Lazim, R., Mei, Y. & Zhang, D. Replica exchange molecular dynamics simulation of structure variation from α/4β-fold to 3α-fold protein. *J. Mol. Model.* **18,** 1087–1095 (2012).
91. de Chiara, C. & Pastore, A. Kaleidoscopic protein–protein interactions in the life and death of ataxin-1: new strategies against protein aggregation. *Trends Neurosci.* **37,** 211–218 (2014).
92. Grasso, G. *et al.* Conformational fluctuations of the AXH monomer of Ataxin-1. *Proteins Struct. Funct. Bioinforma.* **84,** 52–59 (2016).
93. Deriu, M. A. *et al.* Josephin Domain Structural Conformations Explored by Metadynamics in Essential Coordinates. *PLOS Comput. Biol.* **12,** e1004699 (2016).
94. Rakers, C., Bermudez, M., Keller, B. G., Mortier, J. & Wolber, G. Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations? *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5,** 345–359 (2015).
95. Invernizzi, G., Lambrughi, M., Regonesi, M. E., Tortora, P. & Papaleo, E. The conformational ensemble of the disordered and aggregation-protective 182-291 region of ataxin-3. *Biochim. Biophys. Acta* **1830,** 5236–47 (2013).
96. Deriu, M. A. *et al.* Multiscale modeling of cellular actin filaments: from atomistic molecular to coarse-grained dynamics. *Proteins* **80,** 1598–609 (2012).
97. Deriu, M. A. *et al.* Biomechanics of actin filaments: A computational multi-level study. *J. Biomech.* **44,** 630–636 (2011).
98. Enemark, S., Deriu, M. A., Soncini, M. & Redaelli, A. Mechanical Model of the Tubulin Dimer Based on Molecular Dynamics Simulations. *J. Biomech. Eng.* **130,** 041008 (2008).
99. Deriu, M. A. *et al.* Anisotropic Elastic Network Modeling of Entire Microtubules. *Biophys. J.* **99,** 2190–2199 (2010).
100. Grasso, G. *et al.* Cell Penetrating Peptide Adsorption on Magnetite and Silica Surfaces: A Computational Investigation. *J. Phys. Chem. B* **119,** 8239–8246 (2015).
101. Denschlag, R., Lingenheil, M. & Tavan, P. Optimal temperature ladders in replica exchange simulations. *Chem. Phys. Lett.* **473,** 193–195 (2009).
102. Lingenheil, M., Denschlag, R., Mathias, G. & Tavan, P. Efficiency of exchange schemes in replica exchange. *Chem. Phys. Lett.* **478,** 80–84 (2009).
103. Zerze, G. H., Miller, C. M., Granata, D. & Mittal, J. Free Energy Surface of an Intrinsically Disordered Protein: Comparison between Temperature Replica Exchange Molecular Dynamics and Bias-Exchange Metadynamics. *J. Chem. Theory Comput.* **11,** 2776–2782 (2015).
104. Bergonzo, C. *et al.* Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **10,** 492–499 (2014).
105. Deriu, M. A. *et al.* Investigation of the Josephin Domain Protein-Protein Interaction by Molecular Dynamics. *PLoS One* **9,** e108677 (2014).
106. Apicella, A. *et al.* A Hydrophobic Gold Surface Triggers Misfolding and Aggregation of the Amyloidogenic Josephin Domain in Monomeric Form, While Leaving the Oligomers Unaffected. *PLoS One* **8,** e58794 (2013).
107. Deriu, M. A. *et al.* Characterization of the AXH domain of Ataxin-1 using enhanced sampling and functional mode analysis. *Proteins Struct. Funct. Bioinforma.* **84,** 666–673 (2016).
108. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).

109. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
110. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180,** 1961–1972 (2009).
111. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).
112. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–8, 27–8 (1996).
113. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1,** 826–843 (2011).
114. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12562–12566 (2002).
115. Granata, D., Camilloni, C., Vendruscolo, M. & Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 6817–22 (2013).
116. Ziebarth, J. & Wang, Y. Molecular Dynamics Simulations of DNA-Polycation Complex Formation. *Biophys. J.* **97,** 1971–1983 (2009).
117. Deriu, M. A. *et al.* Multiscale modelling of cellular actin filaments: From atomistic molecular to coarse grained dynamics. *Proteins Struct. Funct. Bioinforma.* **80,** 1598–1609 (2012).
118. Soncini, M. *et al.* Mechanical response and conformational changes of alpha-actinin domains during unfolding: A molecular dynamics study. *Biomech. Model. Mechanobiol.* **6,** 399–407 (2007).
119. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9,** 646–652 (2002).
120. Kuzmanic, A. *et al.* Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *Elife* **6,** (2017).
121. Rui, H., Artigas, P. & Roux, B. The selectivity of the Na + /K + -pump is controlled by binding site protonation and self-correcting occlusion. *Elife* **5,** (2016).
122. Burnley, B. T., Afonine, P. V, Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* **1,** (2012).
123. Paciello, G., Acquaviva, A., Ficarra, E., Deriu, M. A. & MacIi, E. A molecular dynamics study of a miRNA:mRNA interaction. *J. Mol. Model.* **17,** 2895–906 (2011).
124. Gentile, F. *et al.* Structure based modeling of small molecules binding to the TLR7 by atomistic level simulations. *Molecules* **20,** 8316–8340 (2015).
125. Deriu, M. A., Popescu, L. M., Ottaviani, M. F., Danani, A. & Piticescu, R. M. Iron oxide/PAMAM nanostructured hybrids: combined computational and experimental studies. *J. Mater. Sci.* **51,** 1996–2007 (2016).
126. Havelka, D., Deriu, M. A., Cifra, M. & Kučera, O. O. Deformation pattern in vibrating microtubule: Structural mechanics study based on an atomistic approach. *Sci. Rep.* **7,** 4227 (2017).
127. Deriu, M. A., Enemark, S., Soncini, M., Montevecchi, F. M. & Redaelli, A. Tubulin: from atomistic structure to supramolecular mechanical properties. *J. Mater. Sci.* **42,** 8864–8872 (2007).
128. Joseph, P. R. B. *et al.* Proline substitution of dimer interface ??-strand residues as a strategy for the design of functional monomeric proteins. *Biophys. J.* **105,** 1491–1501 (2013).
129. Gu, Z. *et al.* Surface Curvature Relation to Protein Adsorption for Carbon-based Nanomaterials. *Sci. Rep.* **5,** 10886 (2015).
130. 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).
131. Grasso, G. *et al.* Thermodynamic and kinetic stability of the Josephin Domain closed arrangement: evidences from replica exchange molecular dynamics. *Biol. Direct* **12,** 2 (2017).
132. Janaszewska, A. *et al.* Multivalent interacting glycodendrimer to prevent amyloid-peptide fibril formation induced by Cu(II): A multidisciplinary approach. *Nano Res.* **11,** 1204–1226 (2018).
133. Grasso, G. *et al.* Destabilizing the AXH Tetramer by Mutations: Mechanisms and Potential Antiaggregation Strategies. *Biophys. J.* **114,** 323–330 (2018).
134. Berhanu, W. M. & Hansmann, U. H. E. Structure and dynamics of amyloid-β segmental polymorphisms. *PLoS One* **7,** e41479 (2012).
135. Alred, E. J., Scheele, E. G., Berhanu, W. M. & Hansmann, U. H. E. Stability of Iowa mutant and wild type A β -peptide aggregates. *J. Chem. Phys.* **141,** 175101 (2014).
136. Grasso, G; Deriu, MA; Patrulea, V; Borchard, G; Möller, M ; Danani, A. *et al.* Free energy landscape of siRNA-polycation complexation: elucidating the effect of molecular geometry, polymer flexibility, and charge neutralization. *PLoS One* **12,** e0186816 (2017).
137. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089 (1993).
138. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).
139. Cheon, M., Hall, C. K. & Chang, I. Structural Conversion of Aβ17–42 Peptides from Disordered Oligomers to U-Shape Protofilaments via Multiple Kinetic Pathways. *PLOS Comput. Biol.* **11,** e1004258 (2015).
140. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **273,** 729–739 (1997).
141. Inouye, H., Fraser, P. E. & Kirschner, D. A. Structure of beta-crystallite assemblies formed by Alzheimer beta-amyloid protein analogues: analysis by x-ray diffraction. *Biophys. J.* **64,** 502–519 (1993).
142. Ngo, S. T., Hung, H. M., Truong, D. T. & Nguyen, M. T. Replica exchange molecular dynamics study of the truncated amyloid beta (11–40) trimer in solution. *Phys. Chem. Chem. Phys.* **19,** 1909–1919 (2017).
143. Tarus, B., Straub, J. E. & Thirumalai, D. Dynamics of Asp23−Lys28 Salt-Bridge Formation in Aβ 10-35 Monomers. *J. Am. Chem. Soc.* **128,** 16159–16168 (2006).
144. Masman, M. F. *et al.* In Silico Study of Full-Length Amyloid β 1−42 Tri- and Penta-Oligomers in Solution. *J. Phys. Chem. B*

113, 11710–11719 (2009).
145. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci.* **105,** 18349–18354 (2008).
146. Khatua, P., Sinha, S. K. & Bandyopadhyay, S. Size-Dependent Conformational Features of Aβ 17–42 Protofilaments from Molecular Simulation Studies. *J. Chem. Inf. Model.* **57,** 2378–2392 (2017).
147. Zheng, J., Jang, H., Ma, B., Tsai, C.-J. & Nussinov, R. Modeling the Alzheimer Aβ17-42 Fibril Architecture: Tight Intermolecular Sheet-Sheet Association and Intramolecular Hydrated Cavities. *Biophys. J.* **93,** 3046–3057 (2007).
148. Fan, H.-M. *et al.* Destabilization of Alzheimer's Aβ42 Protofibrils with a Novel Drug Candidate wgx-50 by Molecular Dynamics Simulations. *J. Phys. Chem. B* **119,** 11196–11202 (2015).
149. Lemkul, J. A. & Bevan, D. R. Destabilizing Alzheimer's Aβ 42 Protofibrils with Morin: Mechanistic Insights from Molecular Dynamics Simulations. *pubs.acs.org/Biochemistry Biochem.* **49,** 3935–3946 (2010).
150. Shuaib, S., Saini, R. K., Goyal, D. & Goyal, B. Insights into the Inhibitory Mechanism of Dicyanovinyl-Substituted J147 Derivative against Aβ $_{42}$ Aggregation and Protofibril Destabilization: A Molecular Dynamics Simulation Study. *ChemistrySelect* **2,** 1645–1657 (2017).
151. Kumar, A. *et al.* Molecular insight into amyloid oligomer destabilizing mechanism of flavonoid derivative 2-(4′ benzyloxyphenyl)-3-hydroxy-chromen-4-one through docking and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **34,** 1252–1263 (2016).
152. Adamcik, J. & Mezzenga, R. Study of amyloid fibrils via atomic force microscopy. *Curr. Opin. Colloid Interface Sci.* **17,** 369–376 (2012).
153. Choi, B., Kim, T., Lee, S. W. & Eom, K. Nanomechanical Characterization of Amyloid Fibrils Using Single-Molecule Experiments and Computational Simulations. *J. Nanomater.* **2016,** 1–16 (2016).
154. Choi, B., Yoon, G., Lee, S. W. & Eom, K. Mechanical deformation mechanisms and properties of amyloid fibrils. *Phys. Chem. Chem. Phys.* **17,** 1379–1389 (2015).
155. Lee, M., Baek, I., Chang, H. J., Yoon, G. & Na, S. The bond survival time variation of polymorphic amyloid fibrils in the mechanical insight. *Chem. Phys. Lett.* **600,** 68–72 (2014).
156. Solar, M. & Buehler, M. J. Comparative analysis of nanomechanics of protein filaments under lateral loading. *Nanoscale* **4,** 1177–1183 (2012).
157. Fitzpatrick, A. W. P., Park, S. T. & Zewail, A. H. Exceptional rigidity and biomechanics of amyloid revealed by 4D electron microscopy. *Proc. Natl. Acad. Sci.* **110,** 10976–10981 (2013).
158. Yoon, G., Kab Kim, Y., Eom, K. & Na, S. Relationship between disease-specific structures of amyloid fibrils and their mechanical properties. *Appl. Phys. Lett.* **102,** 011914 (2013).
159. Sachse, C., Grigorieff, N. & Fändrich, M. Nanoscale Flexibility Parameters of Alzheimer Amyloid Fibrils Determined by Electron Cryo-Microscopy. *Angew. Chemie Int. Ed.* **49,** 1321–1323 (2010).
160. Knowles, T. P. *et al.* Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318,** 1900–3 (2007).
161. Adamcik, J. *et al.* Measurement of intrinsic properties of amyloid fibrils by the peak force QNM method. *Nanoscale* **4,** 4426 (2012).
162. Paul, T. J. *et al.* Structural and Mechanical Properties of Amyloid Beta Fibrils: A Combined Experimental and Theoretical Approach. *J. Phys. Chem. Lett.* **7,** 2758–2764 (2016).
163. Sweers, K., van der Werf, K., Bennink, M. & Subramaniam, V. Nanomechanical properties of α-synuclein amyloid fibrils: a comparative study by nanoindentation, harmonic force microscopy, and Peakforce QNM. *Nanoscale Res. Lett.* **6,** 270 (2011).
164. Sweers, K. K. M., Bennink, M. L. & Subramaniam, V. Nanomechanical properties of single amyloid fibrils. *J. Phys. Condens. Matter* **24,** 243101 (2012).
165. Knowles, T. P. J. & Buehler, M. J. Nanomechanics of functional and pathological amyloid materials. *Nat. Nanotechnol.* **6,** 469–479 (2011).
166. Carrion-Vazquez, M. *et al.* Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74,** 63–91 (2000).
167. Engel, A. & Gaub, H. E. Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* **77,** 127–48 (2008).
168. Hane, F., Tran, G., Attwood, S. J. & Leonenko, Z. Cu2+ Affects Amyloid-?? (1-42) Aggregation by Increasing Peptide-Peptide Binding Forces. *PLoS One* **8,** e59005 (2013).
169. Lv, Z., Roychaudhuri, R., Condron, M. M., Teplow, D. B. & Lyubchenko, Y. L. Mechanism of amyloid β−protein dimerization determined using single−molecule AFM force spectroscopy. *Sci. Rep.* **3,** 2880 (2013).
170. del Mercato, L. L. *et al.* Amyloid-like Fibrils in Elastin-Related Polypeptides: Structural Characterization and Elastic Properties. *Biomacromolecules* **9,** 796–803 (2008).
171. Smith, J. F., Knowles, T. P. J., Dobson, C. M., Macphee, C. E. & Welland, M. E. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 15806–11 (2006).
172. Kol, N. *et al.* Self-Assembled Peptide Nanotubes Are Uniquely Rigid Bioinspired Supramolecular Structures. *Nano Lett.* **5,** 1343–1346 (2005).
173. Graveland-Bikker, J. F., Schaap, I. A. T., Schmidt, C. F. & de Kruif, C. G. Structural and Mechanical Study of a Self-Assembling Protein Nanotube. *Nano Lett.* **6,** 616–621 (2006).
174. Guo, S. & Akhremitchev, B. B. Packing density and structural heterogeneity of insulin amyloid fibrils measured by AFM nanoindentation. *Biomacromolecules* **7,** 1630–6 (2006).
175. Raman, E. P., Takeda, T., Barsegov, V. & Klimov, D. K. Mechanical Unbinding of Aβ Peptides from Amyloid Fibrils. *J. Mol. Biol.* **373,** 785–800 (2007).
176. Havelka, D., Kučera, O., Deriu, M. A. & Cifra, M. Electro-acoustic behavior of the mitotic spindle: A semi-classical coarse-grained model. *PLoS One* **9,** e86501 (2014).
177. Bidone, T. C., Kim, T., Deriu, M. A., Morbiducci, U. & Kamm, R. D. Multiscale impact of nucleotides and cations on the

conformational equilibrium, elasticity and rheology of actin filaments and crosslinked networks. *Biomech. Model. Mechanobiol.* **14,** 1143–1155 (2015).
178. Paparcone, R. & Buehler, M. J. Failure of Aβ(1-40) amyloid fibrils under tensile loading. *Biomaterials* **32,** 3367–3374 (2011).
179. Liu, Z., Xu, Y. & Tang, P. Steered molecular dynamics simulations of Na+ permeation across the gramicidin A channel. *J. Phys. Chem. B* **110,** 12789–95 (2006).
180. Solar, M. & Buehler, M. J. Tensile deformation and failure of amyloid and amyloid-like protein fibrils. *Nanotechnology* **25,** 105703 (2014).
181. Ndlovu, H., Ashcroft, A. E., Radford, S. E. & Harris, S. A. Effect of Sequence Variation on the Mechanical Response of Amyloid Fibrils Probed by Steered Molecular Dynamics Simulation. *Biophys. J.* **102,** 587–596 (2012).
182. Paparcone, R., Keten, S. & Buehler, M. J. Atomistic simulation of nanomechanical properties of Alzheimer's AB(1-40) amyloid fibrils under compressive and tensile loading. *J. Biomech.* **43,** 1196–1201 (2010).
183. Kouza, M. *et al.* Kinetics and mechanical stability of the fibril state control fibril formation time of polypeptide chains: A computational study. *J. Chem. Phys.* **148,** 215106 (2018).
184. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. & Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* **19,** 13–25 (2001).
185. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11,** 224–230 (2001).
186. Xu, Z., Paparcone, R. & Buehler, M. J. Alzheimer's Aβ(1-40) Amyloid Fibrils Feature Size-Dependent Mechanical Properties. *Biophys. J.* **98,** 2053–2062 (2010).
187. Adamcik, J., Berquand, A. & Mezzenga, R. Single-step direct measurement of amyloid fibrils stiffness by peak force quantitative nanomechanical atomic force microscopy. *Appl. Phys. Lett.* **98,** 1–4 (2011).
188. Adamcik, J. *et al.* Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nat. Nanotechnol.* **5,** 423–428 (2010).
189. Sachse, C., Grigorieff, N. & Fändrich, M. Nanoscale Flexibility Parameters of Alzheimer Amyloid Fibrils Determined by Electron Cryo-Microscopy. *Angew. Chemie Int. Ed.* **49,** 1321–1323 (2010).
190. Paparcone, R. & Buehler, M. J. Failure of AB(1-40) amyloid fibrils under tensile loading. *Biomaterials* **32,** 3367–3374 (2011).
191. Relini, A. *et al.* Detection of populations of amyloid-like protofibrils with different physical properties. *Biophys. J.* **98,** 1277–1284 (2010).
192. Yoon, G., Lee, M., Kim, J. I., Na, S. & Eom, K. Role of sequence and structural polymorphism on the mechanical properties of amyloid fibrils. *PLoS One* **9,** 1–13 (2014).
193. Yoon, G., Kwak, J., Kim, J. I., Na, S. & Eom, K. Mechanical characterization of amyloid fibrils using coarse-grained normal mode analysis. *Adv. Funct. Mater.* **21,** 3454–3463 (2011).
194. Moduli, Y., Lamour, G., Yip, C. K. & Li, H. High Intrinsic Mechanical Flexibility of Mouse Prion Nano fi brils Revealed by Measurements of Axial and Radial. 3851–3861 (2014).
195. Schleeger, M. *et al.* Amyloids: From molecular structure to mechanical properties. *Polymer (Guildf).* **54,** 2473–2488 (2013).
196. Jaroniec, C. P. *et al.* High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci.* **101,** 711–716 (2004).
197. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–7 (2007).
198. Wei, G. *et al.* Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology. *Chem. Soc. Rev.* **46,** 4661–4708 (2017).
199. Tycko, R. Amyloid Polymorphism: Structural Basis and Neurobiological Relevance. *Neuron* **86,** 632–645 (2015).
200. Xu, Z., Paparcone, R. & Buehler, M. J. Alzheimer's Aβ(1-40) Amyloid Fibrils Feature Size-Dependent Mechanical Properties. *Biophys. J.* **98,** 2053–2062 (2010).
201. Xu, Z., Paparcone, R. & Buehler, M. J. Alzheimer's abeta(1-40) amyloid fibrils feature size-dependent mechanical properties. *Biophys. J.* **98,** 2053–62 (2010).
202. Pampaloni, F. *et al.* Thermal fluctuations of grafted microtubules provide evidence of a length-dependent persistence length. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10248–53 (2006).
203. Grasso, G. *et al.* Conformational Dynamics and Stability of U-Shaped and S-Shaped Amyloid β Assemblies. *Int. J. Mol. Sci.* **19,** 571 (2018).