Evolution of large A β 16–22 aggregates at atomic details and potential of mean force associated to peptide unbinding and fragmentation events

Antonio Iorio, Štěpán Timr, Letizia Chiodo, Philippe Derreumaux, Fabio Sterpone

Evolution of Large Aβ16-22 Aggregates at Atomic Details and Potential of Mean force Associated to Peptide Unbinding and Fragmentation Events

Antonio Iorio[a,b], Štěpán Timr[e], Letizia Chiodo[d],Philippe Derreumaux[a,b,c], and Fabio Sterpone[a,b,*]

[a]Laboratoire de Biochimie Théorique (UPR 9080), CNRS, Université Paris Cité,
13 rue Pierre et Marie Curie, Paris, 75005, France
[b]Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild,
13 rue Pierre et Marie Curie, Paris, 75005, France
[c]Institut Universitaire de France, 75005 Paris, France
[d]Research Unit in Non Linear Physics and Mathematical Modeling
Engineering Department Campus Bio-Medico University of Rome
Via Álvaro del Portillo 21, 00128, Rome, Italy
[e]J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences
Dolejškova 3, Prague 8, 18223, Czech Republic

Atomic characterization of large non-fibrillar aggregates of amyloid polypeptides cannot be determined by experimental means. Starting from β-rich aggregates of Y and elongated topologies predicted by coarse-grained simulations and consisting of more than 100 Aβ16-22 peptides, we performed atomistic molecular dynamics (MD), replica exchange with solute scaling (REST2), and umbrella sampling simulations using the CHARMM36m force field in explicit solvent. Here we explored the dynamics within 3 microseconds, the free energy landscape, and the potential of mean force associated with either the unbinding of one single peptide in different configurations within the aggregate or fragmentation events of a large number of peptides. Within the time scale of MD and REST2, we find that the aggregates experience slow global conformational plasticity, and remain essentially random coil though we observe slow beta-strand structuring with a dominance of antiparallel beta-sheets over parallel beta-sheets. Enhanced REST2 simulation is able to capture fragmentation events, and the free energy of fragmentation of a large block of peptides is found to be similar to the free energy associated with fibril depolymerization by one chain for longer Aβ sequences.

1 Introduction

A hallmark of amyloid-beta and tau linked to Alzheimer's disease, alpha-synuclein linked to Parkinson's disease and transthyretin linked to cardiac and systemic amyloidosis is that these polypeptides self-assembly to form amyloid fibrils with a cross-beta structure (1), and their oligomers are toxic species (2).

Experimentally trapping these oligomers is difficult as they are very transient and it is not possible to visualize their dynamics at the micro-second time scale (3). Computationally, this is an extremely difficult problem as the free energy landscape increases in complexity as the oligomer size augments. For this reason, most computational studies, with the exception of the simulation of 1000 transthyretin TTR(105-115) fragments (4) and 192 Aβ16-22 peptides (5), focused on small oligomer sizes < 36 peptides (6–8). To accelerate the sampling, they even resort to coarse-grained models and an implicit solvent representation and enhanced sampling techniques, such as umbrella and path sampling, replica exchange molecular dynamics (REMD), replica exchange with solute scaling (REST2), and metadynamics.

Recently, we explored the aggregation dynamics of 100 and 1000 Aβ16-22 peptides using the coarse-grained OPEP protein representation (9) coupled to the surrounding fluid described via the lattice Boltzmann equations. Using Lattice-Boltzmann molecular dynamics (LBMD) simulations (10), we found that hydrodynamics, included naturally in LBMD, increases the fluctuations of the oligomer size within 1 microsecond compared to Langevin dynamics (11) On the largest-size system, we were able to capture the formation and growth of a large elongated aggregate and its slow beta-sheet structure content (reaching 30%). Elongation proceeds through the fusion of oligomers either at the extremities of a nascent amorphous protofibril or at multiple lateral surface points, reminiscent of secondary nucleation (surface-catalyzed) mechanism (12).

While most of the simulations computed free energy landscapes of the dominant oligomers (8, 13), the aggregation pathways (14–17), the kinetic network analysis (18–20) and the disconnectivity graphs (21) between different oligomers, a few studies focused on the potential of mean force (PMF) associated with the microscopic aggregation events, namely primary nucleation, fragmentation and surface-catalyzed secondary nucleation, and fibril growth and depolymerization (22–27). Notably, Monte Carlo simulations of a mesoscopic model with few degrees of freedom in implicit solvent (28) determined a PMF on the order of 7-10 kcal/mol for Aβ42 primary nucleation at concentrations varying between 60 μM and 4 mM. Free energy calculation based on an implicit solvent predictive coarse-grained force field estimated a PMF of a few kcal/mol for Aβ42 primary nucleation (29). Atomistic umbrella sampling of Aβ9-40 surface activated secondary nucleation found that the major driving forces are the release of a large number of hydration water molecules and the formation of hydrophobic interface contacts (27). Discontinuous molecular dynamics (DMD) simulations of surface-catalysed secondary nucleation of Aβ40 by the Aβ16-22 peptide were also performed (30). Finally, path sampling simulations using a coarse-grained protein implicit solvent representation revealed multiple barriers up to 5 kcal/mol between the amorphous aggregates and cross-beta structures

of 12 Aβ16-22 peptides (31).

In this study, we explore by atomistic simulation the behaviour of large Aβ16-22 β-rich aggregates. We take advantage of previous multi-scale simulations of a very large system (1000 peptides) based on the OPEP CG model (12). We extracted two representative aggregate states and back-mapped them to the atomic resolution. The two OPEP-LBMD derived aggregates consisting of 139 and 106 peptides were simulated using the CHARMM36m protein force field with explicit solvent. We first investigated the MD time evolution of the two aggregates within 2.0 microseconds. For one system the MD simulation was extended by one microsecond. Next, starting from the structure used for MD, we explored the free energy landscape of one of the aggregates using REST2. Finally, using umbrella sampling we calculated the PMF associated with the dissociation of an individual peptide from the aggregate, and with the fragmentation of large groups of peptides from the aggregate.

2 Methods

2.1 Mapping Procedure and MD Simulations

A coarse-grained (CG) geometry of an Aβ cluster sampled in an LBMD simulation (10) using the OPEP force field (9) was used as a starting point of the mapping procedure. For each coarse-grained Aβ monomer contained in the cluster, we selected the best-fitting all-atom (AA) conformation from an ensemble of geometries sampled in an MD simulation of an Aβ monomer in dilute conditions. The selection was done by minimizing a cost function which combined the root-mean-square deviation (RMSD) of backbone atoms relative to the CG backbone, the RMSD of AA side chain centers of mass relative to the CG side-chain beads, the number of clashes (i.e., distances below < 1 Å) with the surrounding CG geometries, and the clashes with surrounding monomers that have already been mapped to all-atom resolution. The formula of the cost function was

$$f(\text{geometry}) = \text{RMSD}^2_{\text{backbone}} + 5/9 * \text{RMSD}^2_{\text{sidechain}} + \text{NC}^2_{\text{CG}} + \text{NC}^2_{\text{AA}} \qquad (1)$$

In this formula, $\text{RMSD}_{\text{backbone}}$ denotes the a-dimensional RMSD value (standard RMSD formally divided by a constant of value 1 Å) of the AA backbone atoms relative to the CG backbone, $\text{RMSD}_{\text{sidechain}}$ stands for the RMSD of the AA side-chain centers of mass relative to the CG side-chain beads, NC_{CG} is the number of clashes with the surrounding CG geometries, and NC_{AA} corresponds to the number clashes with surrounding monomers that have already been mapped to the all-atom resolution. To strongly penalise very close overlaps, NC_{AA} was set to 10^3 (10^4, 10^5, or 10^6) if the distance was below 0.9 Å (0.8, 0.7, or 0.6 Å, respectively).

The resulting all-atom structure was solvated in a 150 mM NaCl solution. To avoid insertion of water molecules inside the cluster and allow for a natural hydration of the cluster during the subsequent steps of the mapping protocol, water molecules inserted within 2.5 Å from the peptides were removed. Subsequently, a short energy minimization was performed

to reduce the forces below 1000 kJ mol^{-1} nm^{-1}. The resulting geometries were checked for intertwined aromatic rings. Where present, this artefact was repaired manually, and the short energy minimization was repeated.

To further align the positions of the AA backbone atoms with the CG reference, a targeted MD protocol was conducted in which a harmonic bias was progressively imposed on the backbone RMSD from the CG structure. Specifically, during the first nanosecond the bias was increased progressively from 0 to 5.56$\cdot 10^4$ kJ mol nm^{-2}, and it was maintained at this value for the following 2 ns.

Subsequently, to relax the geometries of the side chains while keeping the backbone atoms aligned with the CG reference, a 10 ns simulated annealing run was performed. The peptide temperature was gradually increased from 300 K to 400 K during the initial 5 ns and then decreased back to 300 K during the final 5 ns. In contrast, the solvent was kept at 300 K for the entire simulated annealing run. Throughout the simulated annealing, the positions of the backbone atoms were restrained by a harmonic bias (with a force constant of 1000 kJ mol nm^{-2}). Finally, the backbone restraints were gradually released in 1 ns steps (500$-$200$-$100$-$50 kJ mol nm^{-2}), and a 1 ns non-restrained equilibration simulation was performed before the production run. It is worth noting that because of the different space filling of the CG vs the atomistic resolution, when back-mapped we lose the secondary structure in β-strand, that, as we will discuss in the Result section will be recovered in time.

MD simulations involved in the mapping procedure were all performed using GROMACS 2018.7 (32). Subsequent production MD simulations were done using the GROMACS 2020.1. The mapping procedure was performed using the a99SB-disp force field (33), however the obtained systems had the tendency to dis-aggregate at longer time when simulated at ambient conditions because of the too strong interactions with water as pointed out in (34). We then used the obtained mapped structures at atomistic resolution to perform production simulations using the CHARMM36m (35) force field. The water model was the three points modified TIP3P from CHARMM36 force field and the systems contained Na^+ and Cl^- ions that were described with the default parameters of the CHARMM36m force field. A 1.2 nm cutoff was applied to short-range non-bonded interactions. Long-range electrostatic interactions were calculated using the particle mesh Ewald method (36). The lengths of all bonds involving hydrogen in the peptides were constrained by the LINCS algorithm (37), and water molecules were kept rigid by the SETTLE algorithm (38). The simulations were conducted in the NpT ensemble at a temperature of 300 K, unless stated otherwise. The temperature was maintained by the velocity rescaling thermostat with a stochastic term (39) using a time constant of 1 ps and coupled separately to the peptides and the solvent. The pressure was maintained at 1.01 bar by the Parrinello-Rahman algorithm (40) with the time constant set to 1 ps.

We back-mapped two large aggregates sampled in a previous multi-scale simulation. We chose two different structures: a large aggregate having a compact Y shape and containing 139 peptides, and a smaller aggregate having a more elongated geometry formed by 106 peptides. Both systems were simulated by MD for 2 μs. The simulation of the largest system was extended for 1 μs to explore further local and global rearrangement. Also, the larger system was

explored using the REST2 enhanced sampling technique that we detailed in the next section.

2.2 REST2 simulations

We performed a REST2 simulation of one of the β-rich aggregates, the one containing 139 peptides. REST2 is a variation of the REMD technique that allows to enhance the sampling of the conformational landscape by exchanging molecular coordinates between replicas of the same system with rescaled potential energy terms involving the solute particles. Since in REST2 method only the degrees of freedom of the solute molecules are heated, a lower number of replicas than in REMD is sufficient to efficiently explore the conformational landscape. In this work we performed 400 ns long all-atom REST2 simulation using GROMACS 2018.3 patched with PLUMED 2.5b. A total of 23 replicas was employed, spanning a temperature range from 300 K to 380 K (41) (see Table 1). Configurations exchange between adjacent replicas was attempted every 5 ps. Temperatures for each replica have been chosen to ensure an exchange probability between adjacent replicas of about $P_{ex} \simeq 0.2$. (42). In the REST2 simulation, each replica started from the same aggregate configuration obtained from the back-mapping procedure that was used also for MD simulation.

Replica	0	1	2	3	4	5	6	7
T [K]	300.00	303.36	306.74	310.16	313.61	317.08	320.33	323.85
Replica	8	9	10	11	12	13	14	15
T [K]	327.40	330.99	334.60	338.23	341.90	345.60	349.33	353.09
Replica	16	17	18	19	20	21	22	
T[K]	356.88	360.70	364.55	368.43	372.34	376.28	380.00	

Table 1: Value of the temperature for each replica. The temperature associated to each scaling factor has been evaluated according to the approach described in (41)

2.3 Free energy calculations

We performed Umbrella Sampling (US) calculations to reconstruct the potential of mean force associated to the separation of a peptide, or a group of peptides from the β-rich aggregate.

For the unbinding of an individual peptide, the chosen reaction coordinate ξ was the distance between the peptide and the aggregate. Namely, once individuated the peptide to pull apart (the more external in the aggregate) we selected the peptide to which is originally bound, and define ξ as the distance between the centers of mass (COM) of the two peptides. In order to generate the initial configuration for the US calculations we performed steered MD simulations. The pulling rate was set to 0.01 nm/ps and the pulling force, of elastic constant $k = 1000$ kJ mol^{-1} nm^{-2}, was applied for 500 ps. The second peptide, closer to the aggregate, was kept in place by positional restraints. From these short trajectories, 30 equi-spaced

configurations, spanning a COM distance of 0.5 to 3.0 nm, were extracted. For each window, five 100 ps MD simulations were performed to create five slightly overlapping replicas for each window. Next, 2 ns simulations were performed for each replica of the 30 windows, for a total of 300 ns of simulation to use for the extraction of the energy profile. The PMF was computed using the Weighted Histogram Analysis Method (WHAM) (43, 44). For error estimates, we used the Bootstrap technique, considering each histogram as a data point and resampling the set of histograms used to compute the PMF 100 times.

We have computed a total of 6 PMF profiles: 3 profiles were calculated for the unbinding of one peptide that forms in the aggregate an anti-parallel β-sheet with 6 hydrogen bonds and 3 other profiles for the unbinding of a peptide forming in the aggregate a parallel β-sheet with 6 hydrogen bonds. To distinguish between parallel and anti-parallel orientations in the β-sheets we used, as a criterion, the distances between Cα atoms at the extremities of the two peptides in contact. A parallel configuration was assigned if $d_{head-head} < d_{head-tail}$, an anti-parallel configuration otherwise. For the anti-parallel case we were able to individuate three different peptides that localise in different parts of the aggregate. On the contrary for the parallel case we were able to individuate only one exposed peptide, and the three PMFs were calculated starting from different configurations of the aggregate that were extracted along the trajectory.

A similar procedure allowed us to compute the PMF of fragmentation, namely the process to separate a large cluster of peptides from the aggregate. Two small groups of peptides have been chosen, the first one composed of 18 peptides, the second of 6 peptides. The two groups were separated from the main aggregate applying a pulling force, of elastic constant $k = 8000$ kJ mol^{-1} nm^{-2}, and a pulling rate of 0.01 nm/ps for 500 ps. As reaction coordinate we used the distance between the centres of mass of the main aggregate and the small cluster of peptides. No positional restraints were applied during the pulling procedure.

3 Results

3.1 Evolution of the β-rich aggregates by MD simulations

After the equilibration procedure described in the Section 2.1, we first investigate how the two back mapped aggregates structurally evolve within two microseconds. In Figure 1 we report, in form of scattered plots, the time evolution of the radius of gyration and the inter-peptide hydrogen bond connectivity for the two systems, on the left column (panels a,c,e) data refer to the largest cluster in Y shape, on the right column (panels b,d,f) to the smaller system of elongated shape. Here, hydrogen bonds are defined according to the electrostatic interaction energy between the donor and the acceptor atoms on the backbone of the peptides (45). At the microsecond timescale, we observe a progressive compaction of the structures. The largest effect is visible for the smaller system, see right Panel b in Figure 1: the gyration radius decreases by a factor of 1.3, passing from 4.6 nm to 3.5 nm. We also notice that the systems stabilise the inter-molecular backbone hydrogen-bonds. In both systems, the inter-molecular connectivity

increases as a function of time, and helps structuring the β-rich aggregates. The systems increase the per-peptide hydrogen bond connectivity from < 1 to $2.5 - 3$. To further investigate the structure of the amyloid aggregates, we used the Define Secondary Structure of Proteins (DSSP) algorithm to assign a secondary structure to each amino acid (45). Over time the percentage of β-sheet structures found in the aggregates increases up to 25-30%, depending on the system.

After 2 μs, the first system maintains a sort of Y topology, but is more compact and shows large stretch of peptides (4 to 7) in continuous β-strands that are localised at the three lobes of the aggregate. In these parts, because of enhanced mobility, the peptides are able to rearrange toward the β-structure. The second system loses its elongated shape, and show the presence of a small pore in the central part. To be noted that the presence of a pore was already sampled in our CG simulations of Aβ_{16-22} large system containing 1000 peptides. In both systems the β-sheet structures are mainly in anti-parallel configuration, $\sim$60%.

In order to further explore the structural evolution, we extended the simulation of the largest system for 1 extra μs. The results are reported in figure 2. We can appreciate that the system slowly increases the gyration radius and the number of inter-molecular hydrogen bond per peptide. Also the content of β structures increases slightly reaching a maximum of about 35%. The aggregate looses at this time scale the Y shape and moves towards a more extended structure. This is shown in the bottom part of the figure where we represented the time evolution of the β-rich aggregates. This MD extension confirms that overall structural relaxation of an aggregate of this size, 10^2 peptides, must occur at time scale of tenth of microseconds and beyond.

3.2 Exploring enhanced sampling by REST2

Exploring the conformational landscape of large aggregates is a computational challenge by brute force MD. To increase the number of configurations of the aggregates and, eventually, capture partial dis-aggregation, we performed REST2 simulation of one of the β-rich aggregate. Each, of the 23 replicas, was evolved for 400 ns. Fig. 3 reports, as scatter plots for four selected temperatures, the density probability of configurations sampled as a function of the rescaled radius of gyration (R_g normalized with respect to the value at t=0 $R_g = 4.2\,nm$) and the number of inter-peptide hydrogen bonds per peptide. The data are obtained considering the full length of the simulation (400 ns). At low temperature, 300 K, the aggregate remains in a compact state close to its initial configuration, even though it explores rescaled Rg values up to 1.2. At 300 K the minimum of the reconstructed free energy landscape localises at a value of $R_g \simeq 3.8$ nm. The minimum shifts at higher R_g values at the higher temperatures, as expected. Over the simulation time, the amino acids adopting β-structure reach a value of 25% that is comparable to what was obtained over the same time scale in the brute force simulation. It seems that in terms of restructuring, at the sub-microsecond time scale, the REST2 does not provide a meaningful enhancement of the conformation sampling for this large aggregate at room temperature. However, especially when focusing at replicas at higher temperature, we can identify a variety of dissociation events associated with the fragmentation of the aggregate and

leading to rescaled Rg values between 1.2 and 1.8. This fragmentation is visible even at 300 K and it increases with temperature. For example in the indicated characteristic configuration we see three aggregates of size 104, 25 and 7 peptides and a few isolated monomers, and even at 306 K some isolated events are populated. To better quantify the process we computed the distribution of the size of the cluster formed in the system during the REST2 simulation for four reference temperatures, see Fig. 4. At the lowest temperatures the aggregate is mainly intact but REST2 allows to sample the detachment of small clusters (dimer to decamer). This result must be contrasted with the brute force MD simulation at the same temperature where only isolated peptides detach from the main aggregate for very short time, see Fig. 4. Increasing the temperature we start observe fragmentation events involving larger aggregates composed by 20 to 30 peptides. As a consequence the main aggregate decreases in size.

3.3 Energetics of depolymerization and fragmentation

The PMFs of single peptide dissociation from the β-rich aggregate, Figs. 5 and 6, show two slightly different behaviours for the anti-parallel and parallel conformations. The minima of the three PMFs for the peptide in the antiparallel β-sheet are located at a value of the reaction coordinate ξ in the 0.66-0.71 nm range, see Fig. 5. The plateau values are reached around $\xi \simeq 1.5$ nm and the final dissociation energy is about 8 kcal/mol in all cases. On the other hand, the minima of the PMFs for the peptide in the parallel β-sheet are close 0.60-0.63 nm, a shorter equilibrium distance when compared to the anti-parallel case, see Fig. 6. The plateau values are reached around 1.7 nm. It is interesting to note that two of the three profiles show a dissociation energy that is almost identical, around 7 kcal$/$mol, that is 1 kcal/mol smaller than for the anti-parallel peptide. For one of the profile we even observe at $\xi \sim 1$ nm a free-energy barrier of 1 kcal/mol separating the bound and unbound states. The third profile reaches a plateau value of about 4 kcal$/$mol. This value is pathological and somehow incorrect. In fact in this latter case the peptide that has been separated from the main aggregate was still able to form hydrogen bonds with other peptides even at large value of the reaction coordinate. In fact, the chosen reaction coordinate restraints the distance between the centres of mass of the moving peptide with respect to a reference one in theaggregate, but still allows the pulled peptide to move orthogonally to it, on a spherical surface of given radius ξ. The peptide is thus able to move closer to other peptides in the aggregate and form hydrogen bonds with them. This points out that our reaction coordinate in some cases do not allow to describe properly the unbinding process because of the complex and dynamic shape of the aggregate.

We also calculated the free energy of fragmentation of the β-rich aggregate by applying the same umbrella sampling procedure but considering the separation of a small group of peptide from the main aggregate. In this case the reaction coordinate ξ is the distance between the centres of mass of the aggregate and of a group of bonded peptides that we want to separate from it. Starting from two different configurations of the system composed of 139 peptides, we selected two groups of 6 and 18 bonded peptides, respectively. The results are reported in Fig.

7. The PMF calculated for the separation of the smaller group of peptides, green curve in Fig. 7, reaches a plateau value of about 25 kcal/mol at around 6 nm, at a distance of 1.3 nm from the minimum. The fragmentation energy of the bigger group of peptides is about 18 kcal/mol, with the plateau value reached at around 6 nm, 1.7 nm away from the minimum. As it is possible to see in the molecular representation of the separating clusters given in the top layer of the Fig. 7, the larger fragment has in the starting configuration already a limited contact surface with the aggregate, in fact a pool of water is present in between. This loose initial connectivity helps to rationalise the smaller free energy variation upon separation when compared to the small fragment.

There are few computational studies reporting PMF calculations based on umbrella sampling on amyloid aggregates. De Pablo et al. determined the PMF of dissociation of the GNNQQQNY peptide starting from an aggregate forming a parallel β-sheet. Using atomistic force field and an explicit solvent representation, they found a PMF of 1.9 kcal/mol for this mostly hydrophilic system (46). Using the all-atom CHARMM27 force field, a dissociation PMF varying between 2 and 4 kcal/mol was obtained for the IAPP fragment spanning residues 20-29 as a function of concentration (47). The dissociation of the dimer of Aβ1-42 was also determined by Uline *et al.* at pH 8 and 150 mM NaCl based on self-consistent field theory with MD simulations (48). The aggregated forms displays a conformation of residues 17-42 matching the U-shape fibril conformation observed by solid-state NMR experiments. It was found a dissociation energy of 12 kcal/mol with a positive contribution of the free energy of solvatation and a negative contribution of the translational free energy of the water.

Bevan et al. employed MD center of mass pulling and umbrella sampling to study the thermodynamics of peptide dissociation from the cross-beta core of a model proto-fibril made of residues 17-42 at physiological temperature. Using GROMOS96 53a6 force field they found a dissociation energy of 50 kcal/mol for the wild-type sequence that decreases to 37.4 kcal/mol upon K28A mutation (22). The dissociation energy of a single Aβ15-40 peptide from the surface of a filament along its most probable pathway was revisited more recently using CHARMM36 force field leading to a value of 8.7 kcal/mol for the free energy barrier separating the bound and unbound state (49). Interestingly, the separating energy from the minima to the full unbounded state is about 20 kcal/mol. The employed method consists of conducting (slow) steered molecular dynamics and relaxing the dissociated peptide at each step of the unbinding pathway. Consistent with previous MD simulation which assessed the existence of dock and lock phases for binding (50) they are multiple energy barriers. It is worth noting that their values are consistent with two experiments leading to an elongation free energy barrier of 10-11 kcal/mol for Aβ40 and Aβ42 (51). The complexity of the Aβ fibril elongation reaction, with growth incompetent states and kinetic traps at the microscopic level, was also revealed by experimental and computational means (26, 52). Finally, Zacharias et al. investigated the fragmentation of a Aβ9-40 protofilament, revealing multiple minima along the pathways and a dissociation free energy of 30 kcal/mol (27).

Overall compared to past simulations, on one hand, our PMF on detaching one single peptide from the aggregate is independent of the parallel or antiparallel orientation of the peptide,

and is only twice compared to dimer PMF dissociation of other amyloid polypeptides. On the other hand, the PMF associated to the detachment of an oligomer of 6-18 peptides is on the order of magnitude of the PMF associated to fibril depolymerization.

4 Conclusion

In this work we presented atomistic simulations of two large β-rich aggregates of the peptide Aβ16-22 which aggregation and structure were previously sampled using the CG OPEP model. After performing the back-map procedure from the CG to the atomistic resolution, we explored the time relaxation of the aggregtes, the conformational landscape and energetics via a combination of approaches including brute force MD, REST2 enhanced sampling and umbrella free energy calculations. At the atomistic resolution the β-rich aggregates tend to assume a more compact structure. Their overall shapes evolve in the time scale of microseconds, and their secondary structure contents somehow saturate at $\sim$ 30%. Portion of β-sheets involving 6-8 peptides are observed in the aggregates. Evolution toward a much larger content of β structures should involve, probably, a massive rearrangement of the aggregates configuration beyond the length of our brute force and REST2 simulations. However, in real systems, the presence of structured patches in amorphous aggregates could play the role of local templates for the elongation of structured parts.

When heating the system (see REST2 simulations), de-polymerisation and fragmentation events are observed. However, the energetics of these events is very difficult to estimate since it depends on the specific locations. When a single peptide is considered, the unbinding free energy from the β-rich aggregate is about 7-8 kcal/mol. When larger blocks of peptides are separated from the main aggregate, much higher free energies of 25 kcal/mol are determined that in fact compare with the values estimated for fibril depolymerization by one chain for longer Aβ sequences. Overall, the obtained results confirm the good insights that OPEP CG simulation provided for Aβ aggregation, and our work explores the possibility to use atomistic modelling in explicit solvent to complement the CG approach for detailing intriguing aspects like the energetics of depolymerization and fragmentation in complex non-symmetrical aggregate geometries.

Acknowledgments

We acknowledge the financial support by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01, and "CACSICE", ANR-11-EQPX-0008). Part of this work was performed using HPC resources from GENCI [CINES, TGCC, IDRIS] (grant x20226818) and LBT. Part of the work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme; in particular, the author gratefully acknowledges the support of LC at the Engineering Department Campus Bio-Medico University of Rome and the computer

resources and technical support provided by CINECA HPC Italy. We thank Geoffrey Letessier for technical support at the LBT.

References

[1] CM Dobson, Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[2] PH Nguyen, A Ramamoorthy, BR Sahoo, J Zheng, P Faller, JE Straub, L Dominguez, JE Shea, NV Dokholyan, A De Simone, B Ma, R Nussinov, S Najafi, ST Ngo, A Loquet, M Chiricotto, P Ganguly, J McCarty, MS Li, C Hall, Y Wang, Y Miller, S Melchionna, B Habenstein, S Timr, J Chen, B Hnath, B Strodel, R Kayed, S Lesné, G Wei, F Sterpone, AJ Doig, P Derreumaux, Amyloid oligomers: A joint experimental/computational perspective on alzheimer's disease, parkinson's disease, type ii diabetes, and amyotrophic lateral sclerosis. *Chem. Rev*. **121**, 2545–2647 (2021).

[3] J Jeon, C Blake Wilson, WM Yau, KR Thurber, R Tycko, Time-resolved solid state nmr of biomolecular processes with millisecond time resolution. *J. Magn. Reson*. **342**, 107285 (2022).

[4] E Scalone, L Broggini, C Visentin, D Erba, FB Toplek, K Peqini, S Pellegrino, S Ricagno, C Paissoni, C Camilloni, Multi-*e*GO: An in silico lens to look into protein aggregation kinetics at atomic resolution. *Proc. Natl. Acad. Sci*. **119**, e2203181119 (2022).

[5] SJ Bunce, Y Wang, SE Radford, AJ Wilson, CK Hall, Structural insights into peptide self-assembly using photo-induced crosslinking experiments and discontinuous molecular dynamics. *AIChE J*. **67**, e17101 (2021).

[6] J Nasica-Labouze, M Meli, P Derreumaux, G Colombo, N Mousseau, A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide gnnqqny from the yeast prion sup-35. *PLOS Comput. Biol*. **7**, 1–18 (2011).

[7] DW Li, S Mohanty, A Irbäck, S Huo, Formation and growth of oligomers: A monte carlo study of an amyloid tau fragment. *PLOS Comput. Biol*. **4**, 1–12 (2008).

[8] PH Nguyen, F Sterpone, P Derreumaux, Self-assembly of amyloid-beta (aβ) peptides from solution to near in vivo conditions. *J. Phys. Chem. B* **126**, 10317–10326 (2022).

[9] F Sterpone, S Melchionna, P Tuffery, S Pasquali, N Mousseau, T Cragnolini, Y Chebaro, JF St-Pierre, M Kalimeri, A Barducci, Y Laurin, A Tek, M Baaden, PH Nguyen, P Derreumaux, The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev*. **43**, 4871–4893 (2014).

[10] F Sterpone, P Derreumaux, S Melchionna, Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics. *J. Chem. Theory Comput*. **11**, 1843–1853 (2015).

[11] M Chiricotto, S Melchionna, P Derreumaux, F Sterpone, Hydrodynamic effects on β-amyloid (16-22) peptide aggregation. *J. Chem. Phys*. **145** (2016).

[12] M Chiricotto, S Melchionna, P Derreumaux, F Sterpone, Multiscale Aggregation of the Amyloid Aβ16–22 Peptide: From Disordered Coagulation and Lateral Branching to Amorphous Prefibrils. *J. Phys. Chem. Lett*. **10**, 1594–1599 (2019).

[13] HL Nguyen, HQ Linh, P Matteini, G La Penna, MS Li, Emergence of barrel motif in amyloid-trimer: A computational study. *J. Phys. Chem. B* **124**, 10617–10631 (2020).

[14] Y Lu, P Derreumaux, Z Guo, N Mousseau, G Wei, Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. *Proteins: Struct. Funct. Bioinforma*. **75**, 954–963 (2009).

[15] X Ge, Y Sun, and F Ding, Structures and dynamics of β-barrel oligomer intermediates of amyloid-beta16-22 aggregation. *Biochim. Biophys. Acta Biomembr.* **1860**, 1687–1697 (2018).

[16] B Barz, Q Liao, B Strodel, Pathways of amyloid-aggregation depend on oligomer shape. *J. Am. Chem. Soc*. **140**, 319–327 (2018).

[17] Y Liu, Y Wang, Y Zhang, Y Zou, G Wei, F Ding, Y Sun, Structural perturbation of monomers determines the amyloid aggregation propensity of calcitonin variants. *J. Chem. Inf. Model*. (2022).

[18] AM Illig, B Strodel, Performance of markov state models and transition networks on characterizing amyloid aggregation pathways from md data. *J. Chem. Theo. Comput*. **16**, 7825–7839 (2020).

[19] S Samantray, W Schumann, AM Illig, M Carballo-Pacheco, A Paul, B Barz, B Strodel, Molecular dynamics simulations of protein aggregation: Protocols for simulation setup and analysis with markov state models and transition networks. *Methods Mol. Biol.* **2340**, 235–279 (2022).

[20] M. Ur Rahaman, K. Song, L.T. Da, and H.F. Chen, Early aggregation mechanism of ab1622 revealed by markov state models. *Int. J. Biol. Macromol.* **204**, 606–616 (2022).

[21] B Barz, DJ Wales, B Strodel, A kinetic approach to the sequence–aggregation relationship in disease-related protein assembly. *J. Phys. Chem. B* **118**, 1003–1011 (2014).

[22] JA Lemkul, DR Bevan, Assessing the stability of alzheimer's amyloid protofibrils using molecular dynamics. *J. Phys. Chem. B* **114**, 1652–1660 (2010).

[23] W Han, K Schulten, Fibril elongation by aβ17–42: Kinetic network analysis of hybrid-resolution molecular dynamics simulations. *J. Am. Chem. Soc*. **136**, 12450–12460 (2014).

[24] Z Wang, W Huang, M Liu, SJ Kennel, JS Wall, X Cheng, Computational investigation of the binding of a designed peptide to light chain amyloid fibril. *Phys. Chem. Chem. Phys.* **23**, 20634–20644 (2021).

[25] IM Ilie, A Caflisch, Simulation studies of amyloidogenic polypeptides and their aggregates. *Chem. Rev*. **119**, 6956–6993 (2019).

[26] N Schwierz, CV Frost, PL Geissler, M Zacharias, Dynamics of seeded aβ40-fibril growth from atomistic molecular dynamics simulations: Kinetic trapping and reduced water mobility in the locking step. *J. Am. Chem. Soc*. **138**, 527–539 (2016).

[27] N Schwierz, CV Frost, PL Geissler, M Zacharias, From afilament to fibril: Molecular mechanism of surface-activated secondary nucleation from all-atom md simulations. *J. Phys. Chem. B* **121**, 671–682 (2017).

[28] A Šarić, YC Chebaro, TPJ Knowles, D Frenkel, Crucial role of nonspecific interactions in amyloid nucleation. *Proc. Natl. Acad. Sci*. **111**, 17869–17874 (2014).

[29] W Zheng, MY Tsai, PG Wolynes, Comparing the aggregation free energy landscapes of amyloid beta(1–42) and amyloid beta(1–40). *J. Am. Chem. Soc*. **139**, 16666–16676 (2017).

[30] SJ Bunce, Y Wang, KL Stewart, AE Ashcroft, SE Radford, CK Hall, AJ Wilson, Molecular insights into the surface-catalyzed secondary nucleation of amyloid-β_{40} by the peptide fragment aβ_{16-22}. *Sci. Adv*. **5**, eaav8216 (2019).

[31] JA Luiken, PG Bolhuis, Primary nucleation kinetics of short fibril-forming amyloidogenic peptides. *J. Phys. Chem. B* **119**, 12568–12579 (2015).

[32] MJ Abraham, T Murtola, R Schulz, S Páll, JC Smith, B Hess, E Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).

[33] P Robustelli, S Piana, DE Shaw, Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115**, E4758–E4766 (2018).

[34] BK S. Samantray, F. Yin, B Strodel, Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model*. **60**, 6462–6475 (2020).

[35] J Huang, S Rauscher, G Nawrocki, T Ran, M Feig, BL de Groot, H Grubmüller, AD MacKerell, Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

[36] T Darden, D York, L Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys*. **98**, 10089–10092 (1993).

[37] B Hess, H Bekker, HJC Berendsen, J Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem*. **18**, 1463–1472 (1997).

[38] S Miyamoto, PA Kollman, Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem*. **13**, 952–962 (1992).

[39] G Bussi, D Donadio, M Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys*. **126**, 14101 (2007).

[40] M Parrinello, A Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*. **52**, 7182–7190 (1981).

[41] G Stirnemann, F Sterpone, Recovering protein thermal stability using all-atom hamiltonian replica-exchange simulations in explicit solvent. *J. Chem. Theory Comput*. **11**, 5573–5577 (2015).

[42] A Patriksson, D van der Spoel, A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **10**, 2073–2077 (2008).

[43] S Kumar, JM Rosenberg, D Bouzida, RH Swendsen, PA Kollman, ThE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem*. **13**, 1011–1021 (1992).

[44] JS Hub, BL de Groot, D van der Spoel, g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput*. **6**, 3713–3720 (2010).

[45] W Kabsch, C Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

[46] AS Reddy, L Wang, S Singh, YL Ling, L Buchanan, MT Zanni, JL Skinner, JJ de Pablo, Stable and metastable states of human amylin in solution. *Biophys. J.* **99**, 2208–16 (2010).

[47] TD Rivera E, Straub J, Sequence and crowding effects in the aggregation of a 10-residue fragment derived from islet amyloid polypeptide. *Biophys. J.* **96**, 4552–4560 (2009).

[48] NP van der Munnik, MSJ Sajib, MA Moss, T Wei, MJ Uline, Determining the potential of mean force for amyloid-dimerization: Combining self-consistent field theory with molecular dynamics simulation. *J. Chem. Theory Comput*. **14**, 2696–2704 (2018).

[49] RA Rodriguez, LY Chen, G Plascencia-Villa, G Perry, Thermodynamics of amyloid-fibril elongation: Atomistic details of the transition state. *ACS Chem. Neurosci*. **9**, 783–789 (2018).

[50] PH Nguyen, MS Li, G Stock, JE Straub, D Thirumalai, Monomer adds to preformed structured oligomers of aβ-peptides by a two-stage dock–lock mechanism. *PNAS* **104**, 111–116 (2007).

[51] K Hasegawa, K Ono, M Yamada, H Naiki, Kinetic modeling and determination of reaction constants of alzheimer's beta-amyloid fibril extension and dissociation using surface plasmon resonance. *Biochemistry* **41**, 13489–13498 (2002).

[52] KC Young LJ, Kaminski Schierle GS, Imaging aβ(1-42) fibril elongation reveals strongly polarised growth and growth incompetent states. *Phys Chem Chem Phys* **19**, 27987–27996 (2017).

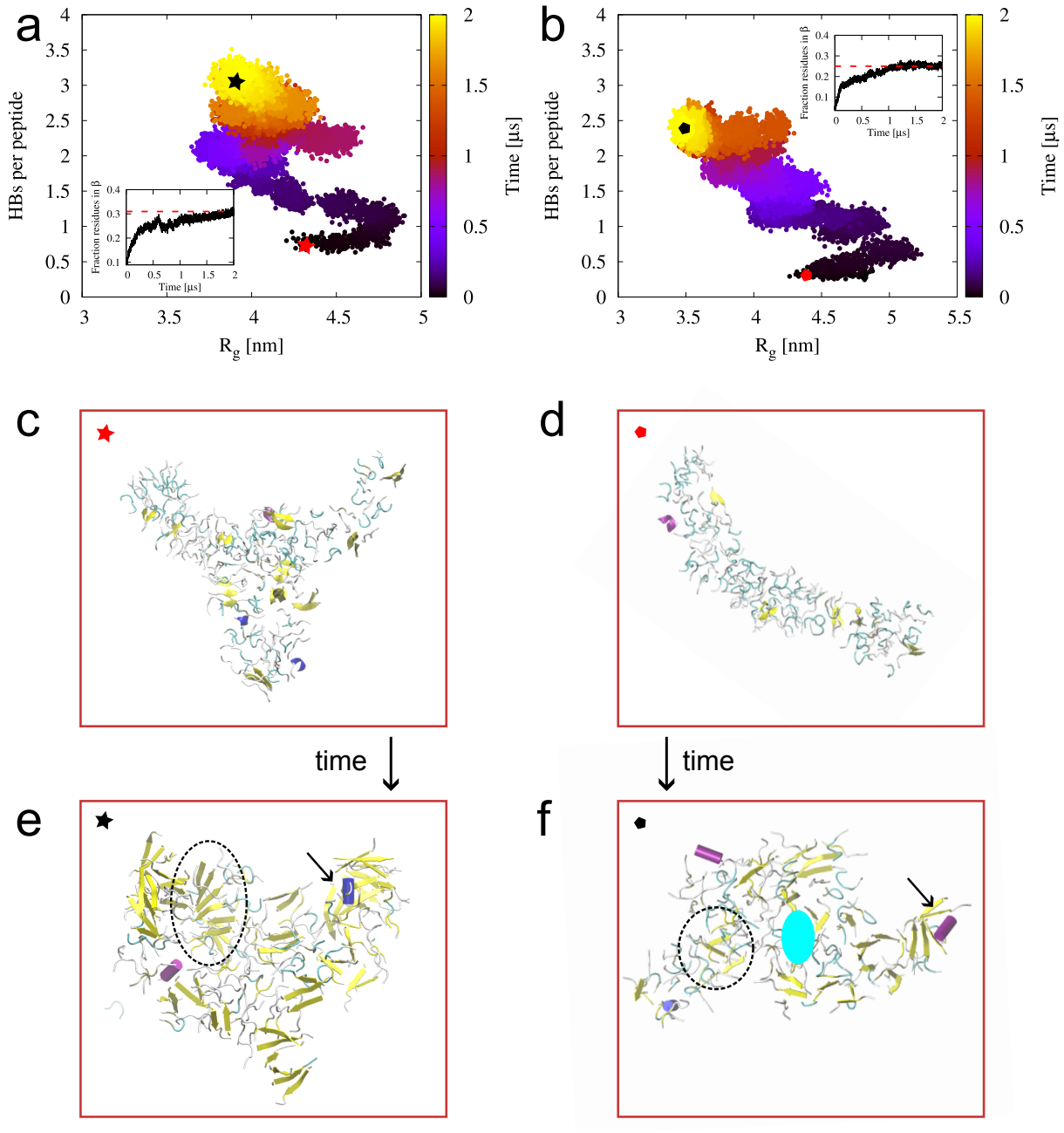

Figure 1: Scatter plots of the radius of gyration vs number of backbone hydrogen bonds per peptide in two systems. Panel a refers to the largest β-rich aggregate composed of 139 Aβ_{16-22} peptides while panel b refers to the smallest one with 106 Aβ_{16-22} peptides. In each panel, the inset reports the fractions of residues in β conformation as a function of time, dashed red lines are average values computed during the last 100 ns of the simulations. Panel c and d show a molecular representation of the two β-rich aggregate at the beginning of the simulations. In panels e and f we report the final configuration sampled after 2 μs of the two fibrils, the largest one (left) and the smallest one (right). The aggregates are represented according to the secondary-structure (with β-strands coloured in yellow). Dashed circles highlight the localisation of some extended β-sheet structures in the aggregates; arrows point to some localised α-helices. In panel f we indicate the presence of a small pore by a cyan ellipse.

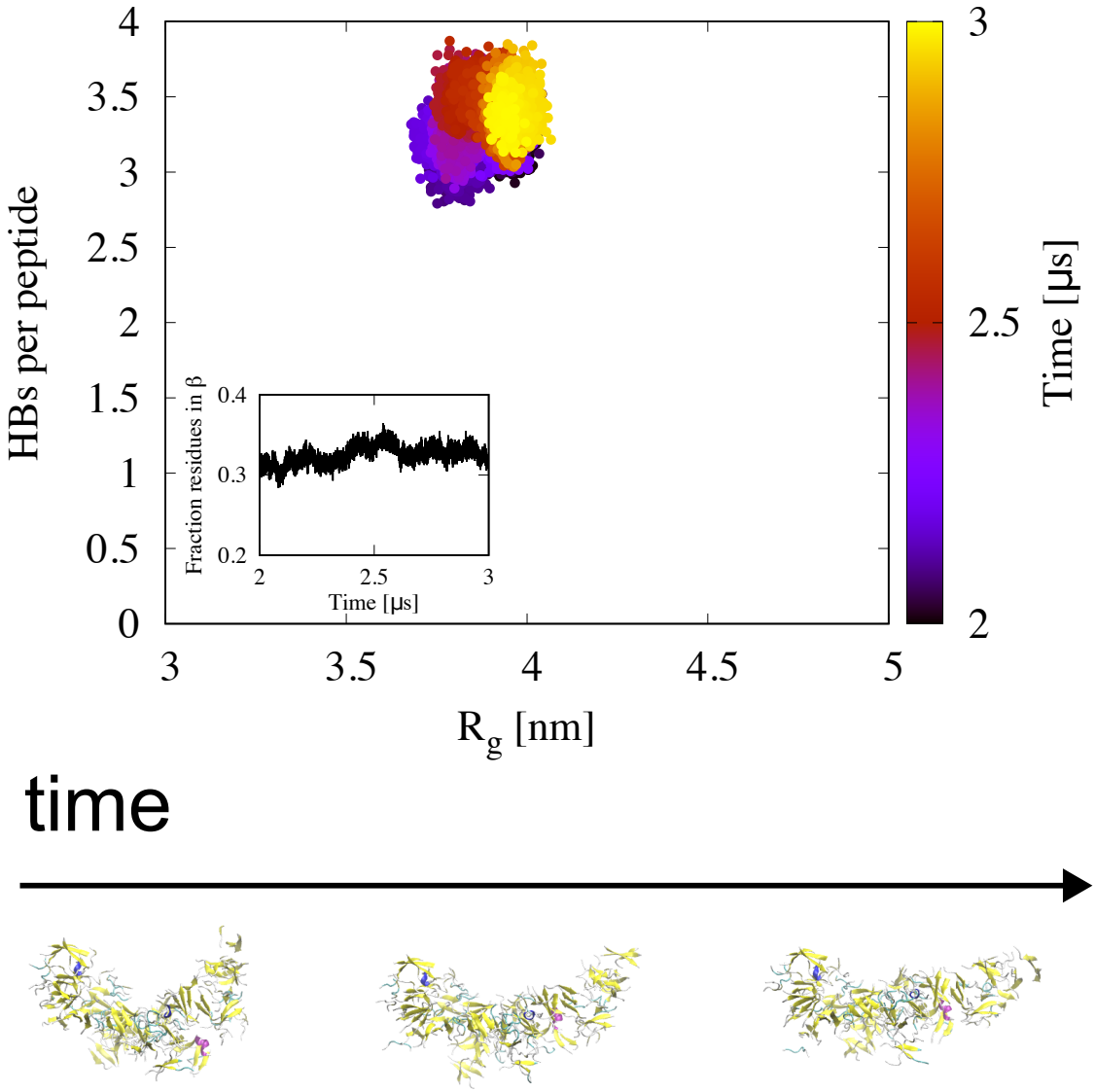

Figure 2: Scatter plots of the radius of gyration vs number of backbone hydrogen bonds for the extended 1 μs MD simulation. Inset reports the fractions of residues in β conformation as a function of time. In the bottom part of the figure three molecular representations of the β-rich aggregate at different times along the extended trajectory are shown.

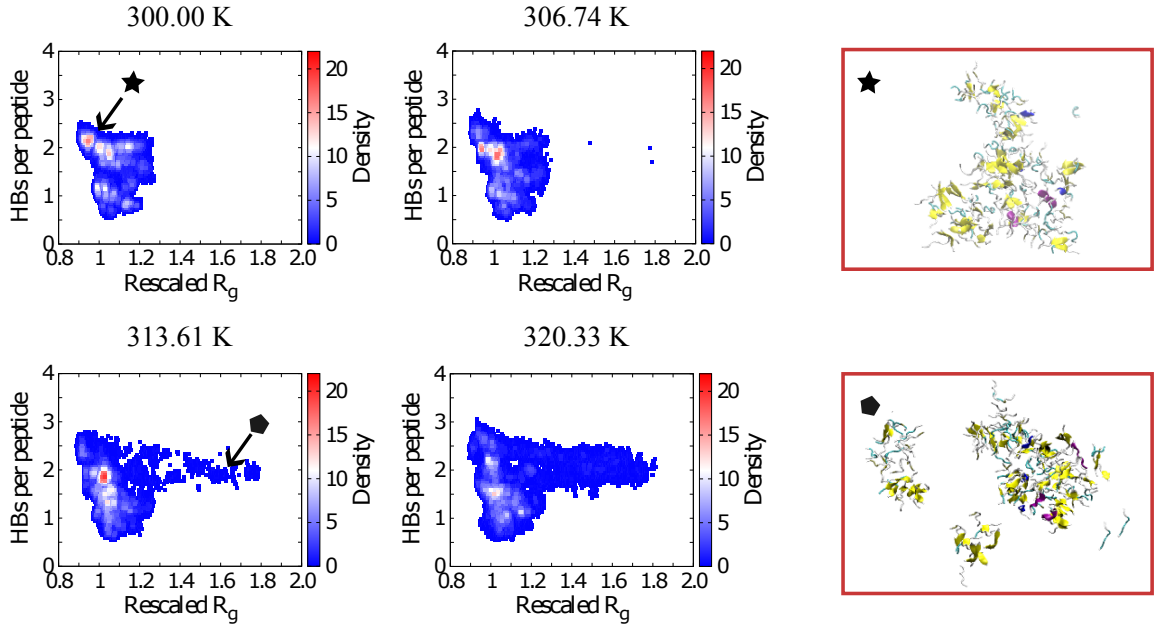

Figure 3: Scatter plots of the rescaled radius of gyration vs number of backbone hydrogen bonds per peptide in 4 replicas of the REST2 simulation (see manuscript for details). The rescaled value of the radius of gyration is defined with respect to the value computed at $t = 0$ and is the same for all replicas, $R_g = 4.2$ nm. In the right part of the figure we report two representative structures of the aggregates, (top) fully formed aggregate extracted from the 300 K replica (see symbol star), (bottom) partially fragmented aggregate extracted from the 313.6 K replica (see symbol pentagon).

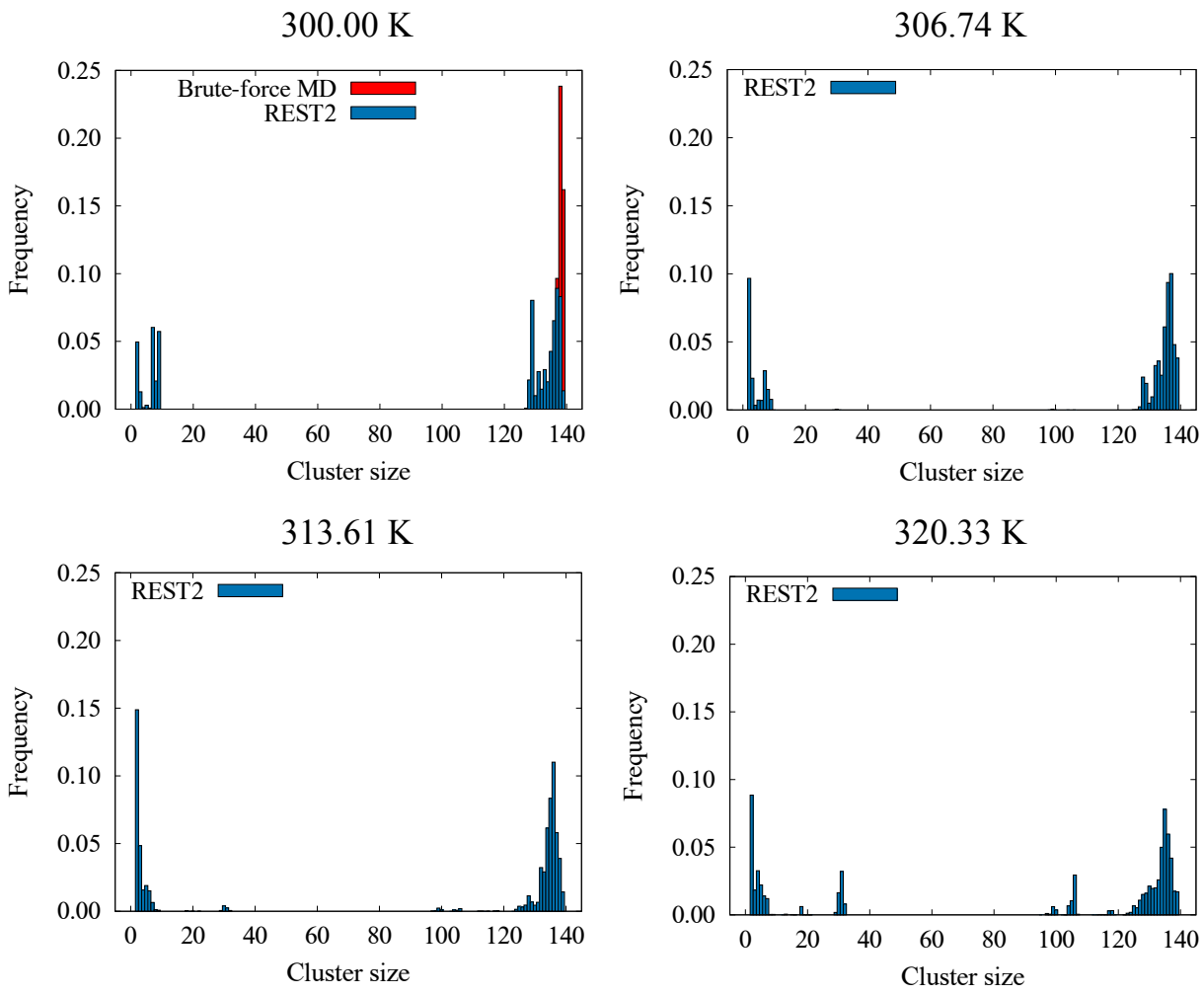

Figure 4: Distribution of the sizes of the clusters formed in the REST2 simulations at four different temperatures. In the 300 K panel we also report the data extracted from brute force MD.

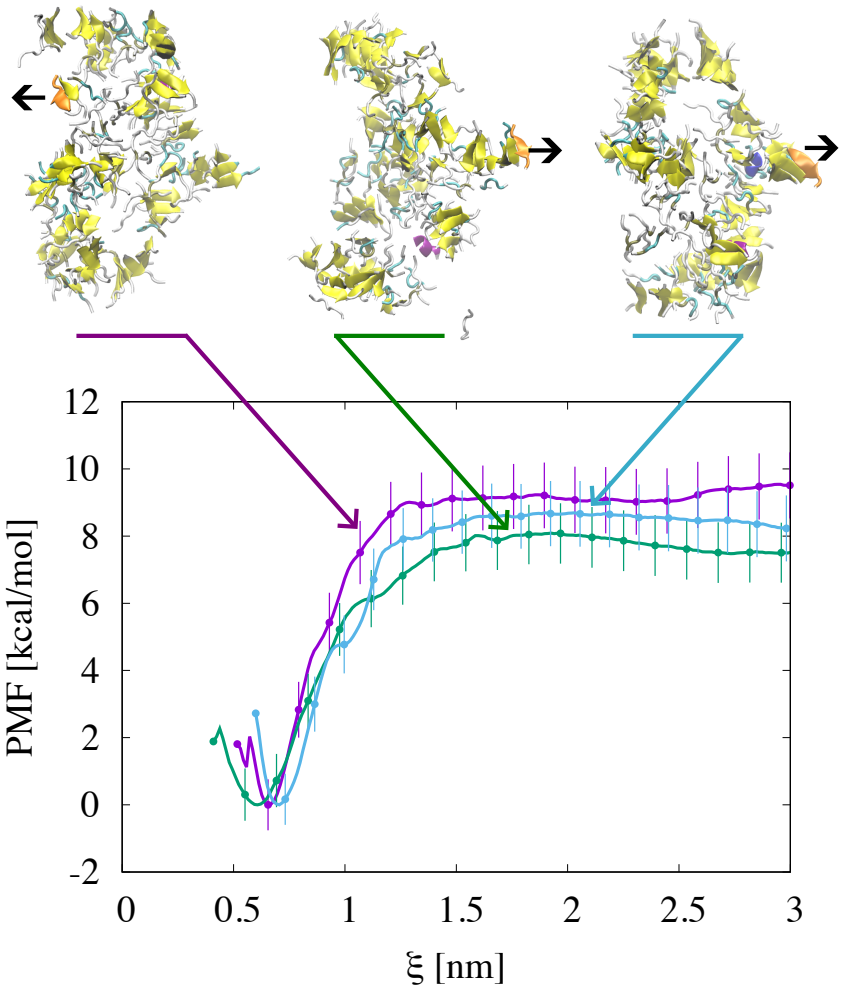

Figure 5: Potential of mean force profiles for the separation of one peptide from the β-rich aggregate with error bars reported every 10 points. The pulled peptides form an anti-parallel β-sheet in the aggregate. We selected three peptides that localise in different parts of the aggregate surface.The minima of the energy profiles have been placed to 0 for simplicity. The selected peptides are represented in the top layer of the figure and are coloured in orange along with a black arrow.

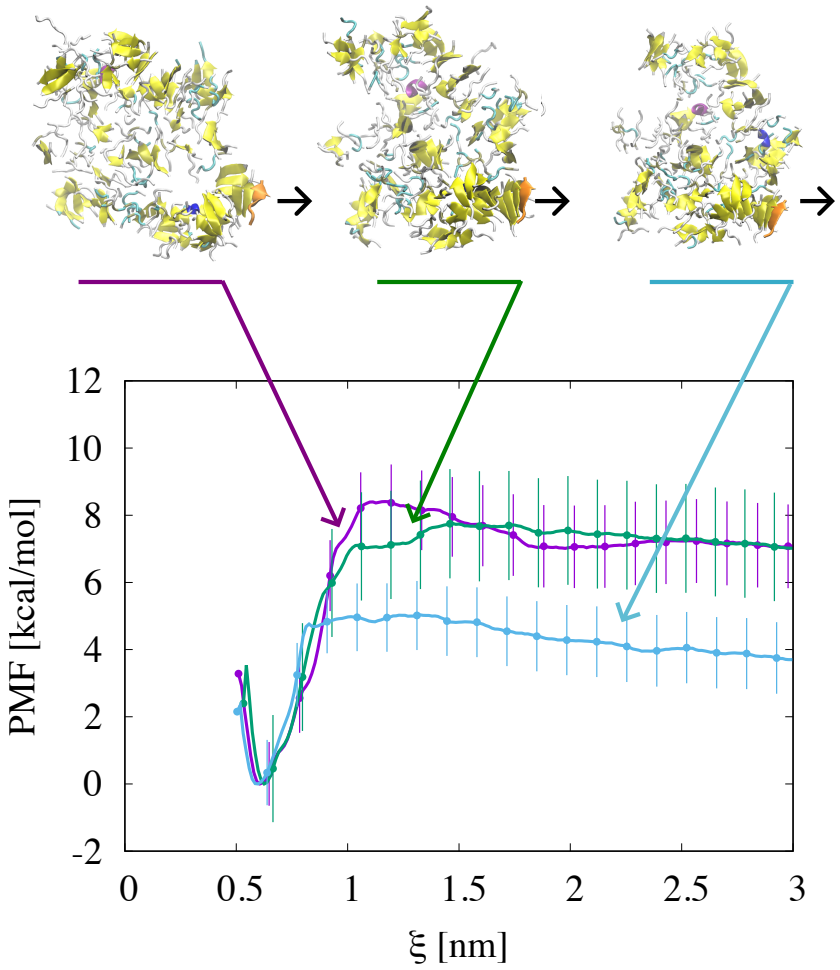

Figure 6: Potential of mean force profiles for the separation of one peptide from the β-rich aggregate with error bars reported every 10 points. The pulled peptides form a parallel β-sheet in the aggregate. We selected a unique peptide that localises at the surface of the aggregate. The PMF calculations were started from different configurations of the aggregate. The minima of the energy profiles have been placed to 0 for simplicity. The selected peptides are represented in the top layer of the figure and are coloured in orange along with a black arrow.

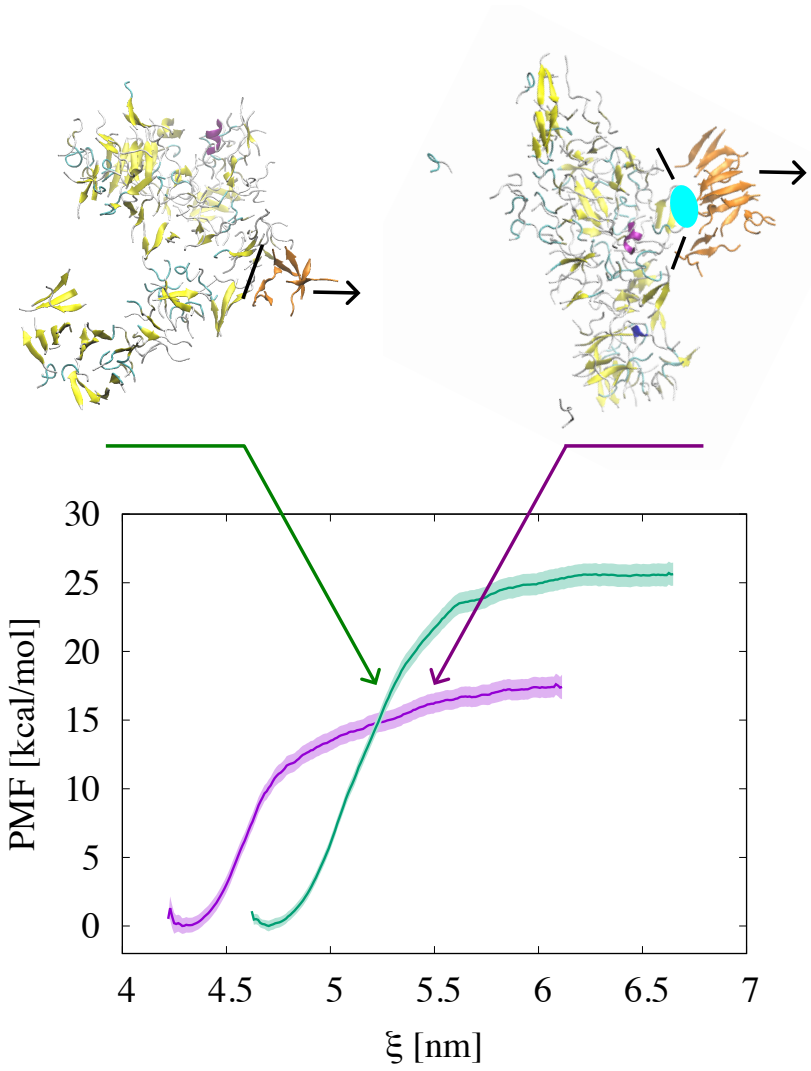

Figure 7: Fragmentation potential of mean force of two selected groups of peptides separating from the β-rich aggregate with error bars. The green profile refers to the small cluster, the violet to the bigger one. The minima of the energy profiles have been placed to 0 for sake of clarity. In the top layer we provide a molecular representation of the separating clusters that are colored in orange along with a pulling arrow. The black lines pictorially indicate the contact among the fragments and the aggregate. In the right panel the cyan ellipse indicates the presence of a pore filled by water and that separates the fragment from the main body of the aggregate.