Special Issue on "Theoretical and Computational Chemistry"

J. Indian Chem. Soc.,
Vol. 96, July 2019, pp. 991-1002

Characterization of conformational ensemble of C-terminally truncated α-synuclein in aqueous solution by atomistic molecular dynamics simulations

Moutusi Manna* and Rajesh K. Murarka*

Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhopal Bypass Road, Bhopal-462 066, Madhya Pradesh, India

E-mail: moutusi.manna.chem@gmail.com, rkm@iiserb.ac.in

Manuscript received online 30 April 2019, revised and accepted 28 May 2019

α-Synuclein is an intrinsically disordered protein, whose aggregation into amyloid fibril is involved in the pathology of Parkinson's disease. C-terminally truncated α-synuclein variants, which are naturally generated from the full-length α-synuclein, are enriched in the pathological α-synuclein aggregates and known to significantly enhance the aggregation process. In this work, we have performed extensive Gaussian accelerated molecular dynamics simulations (3.8 μs in total) to characterize the equilibrium conformational ensemble of a physiologically relevant C-terminally truncated α-synuclein monomer, AS(1-103), in aqueous solution. The nearly flat free-energy landscape of AS(1-103) suggests coexistence of multiple distinct conformational states, including disordered conformations and conformations with transient secondary and tertiary structures. We found that the predominant conformation of AS(1-103) in aqueous solution exhibit enrichment of β-sheet structure at different parts of the N-terminal and hydrophobic NAC domain, including the aggregation-prone NAC-core. We identified moderate intra- and inter-domain interactions within the N-terminal and the NAC domain that enhance and stabilize the partially folded β-sheet rich structures of AS(1-103) in solution. Such stable structural regions formed in the AS(1-103) ensemble, may serve as potential molecular signature that encodes the aggregation seed. The presented work enables characterization of conformational states of monomeric AS(1-103) in solution, and provide a plausible explanation for the experimentally determined enhanced aggregation propensity of C-terminally truncated variants of α-synuclein *in vivo*.

Keywords: α-Synuclein, C-terminal truncation, conformational ensemble, free-energy landscape, molecular dynamics simulations, enhanced sampling.

1. Introduction

Intrinsically disordered proteins (IDPs) are proteins that lack well-defined secondary and tertiary structure. IDPs are exceptions to the widely accepted Anfinsen's sequence-structure-function paradigm[1], according to which the primary sequence of a protein defines its unique three-dimensional structure that is inherently linked to its function. Instead, IDPs exist as dynamic ensembles of disordered or partially disordered conformations under physiological conditions. Due to conformational heterogeneity, IDPs are capable to bind a range of cellular targets and adopt folded structures complementary to the binding partners[2,3]. IDPs are abundant in eukaryotes[4] and are involved in critical cellular functions, including transcription, translation and cellular signaling[5,6]. Misfolding and aggregation of numerous IDPs are resultant in some of the most widespread neurodegenerative disorders, such as Alzheimer's disease, Parkinson's disease and Huntington's disease[7]. Comprehensive understanding of the mechanism underlying the pathogenic aggregation of an IDP remains as a major unsolved problem, despite decades of intense research.

The α-synuclein, an IDP, is the major component of the intracellular proteinaceous aggregates, termed as Lewy bodies, which are pathological hallmark of neurodegenerative disorders, such as Parkinson's disease and other synucleinopathies[8,9]. α-Synuclein is a neuronal protein localized primarily at the presynaptic terminals in the central nervous system and proposed to be physiologically involved in synaptic plasticity, vesicle fusion, and neurotransmitter release[10]. The primary sequence of α-synuclein (14.46 kDa) is usually

divided into three distinct regions: an amphipathic N-terminal (residue 1-60) crucial for membrane binding, a hydrophobic central region (residue 61-95) acting as building block of α-synuclein fibril, and a negatively charged C-terminal (residue 96-140) with chaperone-like activity (Fig. 1A)[11,12]. It is well documented that the C-terminal of α-synuclein is highly disordered[11–14], whereas the rest of the protein lives in a conformational limbo under different conditions. The 1-95 amino acids are known to fold predominantly into an α-helix in presence of lipid membrane[13–15] or into a β-sheet structure in amyloid fibril[16–18]. The most hydrophobic central region, termed as non-amyloid-β component (NAC), has the highest propensity to fold into a β-sheet secondary structure and play pivotal role in the aggregation process.

Transformation of natively unfolded α-synuclein into amyloid fibrils through oligomeric intermediates is considered as the disease-causative toxic mechanism. A vast literature supports higher toxicity of small oligomeric intermediates compared to the mature fibril[19,20]. Conformational transition of unfolded monomeric α-synuclein into partially folded intermediates could be accountable for aggregation and fibrillization. Accumulating evidences suggest that the monomeric α-synuclein populates ensemble of unfolded and partially folded conformations in solution[21–23]. Detailed characterization of the conformational states of α-synuclein monomer as well as identification of transient structural epitomes within the conformation ensemble are thus appears to be crucial, as those might have functional implications or disease relevance.

Various intrinsic and extrinsic factors, such as temperature, pH, metal ions and salts, point mutations, post-translational modification, membrane binding, etc., influence the misfolding and oligomerization pathways of α-synuclein[24-27]. C-terminal truncation of α-synuclein is one of the most prominent modifications which promote the aggregation of α-synuclein[28–30]. C-terminally truncated variants of α-synuclein are generated by normal cellular processing of full-length α-synuclein independent of α-synuclein aggregation in brains and in cultured cells[28]. This lower molecular mass α-synuclein species are enriched in α-synuclein aggregates and represent nearly 15% of all α-synuclein molecules found in Lewy bodies[29–31]. The pathological significance of the C-terminally truncated variants is further supported by their enhanced accumulation with familial Parkinson's disease mutants[28]. A better understanding of the structural states of the C-terminally truncated α-synuclein variants may, therefore, prove useful for gaining deeper insights into the disease mechanism.

The detailed structural characterization of monomeric α-synuclein in solution using conventional experimental technique is very challenging, due to the conformational heterogeneity, rapid conformational change and fast aggregation rate of α-synuclein. Most experimental techniques often suffer from limitations in their scale and time resolution, and may provide average information on the conformational ensemble of an IDP, in general[21,32]. Computational techniques including molecular dynamics (MD) simulations, with or without experimental data as input, have become a popular tool to elucidate the conformational ensemble of IDP of interest[22,33–36]. Standard empirical force fields used in MD simulations, which were originally purposed to study folded globular proteins, however, often fail to provide accurate description of IDP and tend to produce more compact structures[37–40]. Over the years, significant research efforts have been made to reassess and reparameterize the standard force fields to improve their ability of predicting structures of disorder proteins without losing their accuracy for folded proteins[38,41]. Due to the complexity of their conformational space with multiple minima, IDP demands high computational cost to sample functionally relevant conformations which are otherwise kinetically inaccessible to conventional MD simulations. Enhanced sampling techniques are therefore routinely applied to ensure adequate sampling of IDP's conformational landscape[33,39,42,43].

In the present work, we have carried out extensive atomistic MD simulations to determine the conformational ensemble of a physiologically relevant C-terminally truncated α-synuclein(1-103) monomer in solution. α-Synuclein(1-103), referred here as AS(1-103), has virtually all of the C-terminal domain truncated and is reported to readily convert to mature fibril, compared to full length α-synuclein or α-synuclein variants with partly removed C-terminus[29]. Here we have employed Gaussian accelerated molecular dynamics (GaMD)[44], an unconstrained enhanced sampling technique, to characterize the conformational substates explored by AS(1-103) in solution and determine the nature of the

underlying free-energy landscape that may serve as the basis for understanding the misfolding and oligomerization pathways.

2. Method

2.1. System preparation

The initial structures of α-synuclein were taken from the protein ensemble database (PeDB: 9AAC). We have selected four different full-length α-synuclein structures with radius of gyration (R_g) 31.75 Å, 32.42 Å, 33.18 Å and 34.84 Å. The selection criterion is based on 27-35 Å estimates of R_g obtained from NMR[45] and SAXS[46] measurements and also consistent with the selection criterion used in a previous simulation study[47]. From the full-length α-synuclein, the residues 104-140 were removed, resulting in C-terminally truncated AS(1-103) structures with R_g values 28.62 Å, 30.23 Å, 26.30 Å and 25.99 Å (Fig. 1). The truncated C-terminus was capped with amide group. The N-terminus was capped with acetyl group, as human α-synuclein has been shown to be N-terminally acetylated in its physiological state[48]. The protein was then solvated with water; counter ions were added to maintain the electroneutrality of the system and NaCl salt was added to achieve the physiological salt concentration of 150 mM. The simulations were then performed with all-atom CHARMM36m force field parameters[41] for protein and

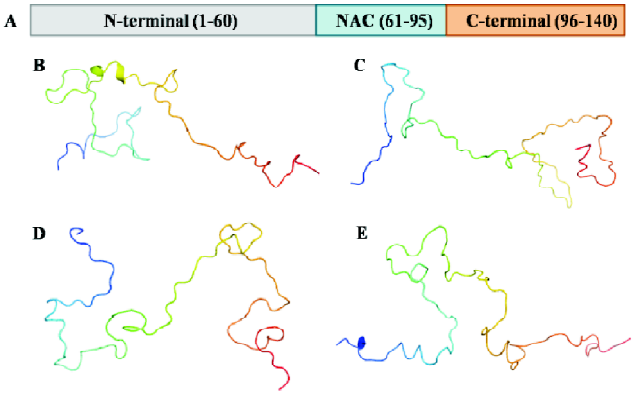

Fig. 1. (A) A schematic description of α-synuclein structure. N-terminal, NAC and C-terminal domains are colored gray, cyan and orange, respectively. (B-E) Four initial structures of C-terminally truncated α-synuclein variant, AS(1-103), used as stating structures of our conventional MD simulations, which are derived from NMR ensemble of full length α-synuclein. Structures are colored from blue to red, from the N- to truncated C-terminus; the NAC region falls in the orange to red section.

CHARMM-modified TIP3P water model[49] that includes specific modification of water hydrogen LJ parameters to scale up the protein-water dispersion interactions in order to overcome the problem of overly compact IDP ensemble. The CHARMM36m force field involves refinement of CHARMM36 force filed, with improved accuracy for generating conformational ensembles for IDP[41]. We have performed four independent conventional molecular dynamics (cMD) simulations, starting with different disordered structures of AS(1-103) (Fig. 1B-E). For cMD simulations, the protein was solvated in a 116×116×116 $Å^3$ cubic box filled with ~48500 water molecules, 138 Na^+ and 143 Cl^- ions. From the final snapshots of cMD simulations, we have performed four independent Gaussian accelerated molecular dynamics (GaMD) simulations, after truncating the simulation box to 106×106×106 $Å^3$ dimension that contains ~39000 water molecules, 111 Na^+ and 116 Cl^- ions.

2.2. Molecular dynamics simulations

All cMD simulations were performed using Groningen Machine for Chemical Simulations (GROMACS) package version 5.0.4[50,51]. A 2 fs time step was used for integrating the equations of motion. The Linear Constraint Solver (LINCS) algorithm[52] was applied to preserve hydrogen covalent bond lengths. Simulations were performed under isothermal-isobaric (NVT) ensemble. The temperature of the system was maintained at 310 K by employing the v-rescale (stochastic velocity rescaling) thermostat[53] with a time constant of 0.5 ps. The temperatures of protein and solvent (water and ions) were controlled independently. We have chosen to perform our simulations at 310 K, as the physiological temperature (human body temperature) is about 310 K. Periodic boundary conditions were applied in all three directions. The particle mesh Ewald method[54] was employed for long-range electrostatic interactions with a real space cutoff of 10 Å. The van der Waals interactions were treated using the Lennard-Jones potential with a cutoff distance of 10 Å. The neighbour lists were updated every 20 steps with a cutoff of 10 Å. Each of the four systems was first energy minimized. The systems were then equilibrated for 5 ns with position restraints on protein heavy atoms, followed by another 5 ns equilibration with position restraints only on protein backbone. After equilibration, all restraints were released and each sys-

tem was subjected to a production simulation of 100 ns (i.e. cumulative 4×100 ns of cMD).

We have employed recently developed novel GaMD method, to achieve both unconstrained enhanced sampling and free energy calculations[44]. A major advantage of GaMD is that it does not need to set any predefined reaction coordinates. GaMD enhances the conformational sampling of biomolecules by adding a harmonic boost potential to smooth the potential energy surface[44]. As the boost potential follows a Gaussian distribution, the original free energy profile of the system is achieved via cumulant expansion to the second order[44]. The details of the method will be found elsewhere[44,55]. Recent studies have shown that GaMD greatly reduces the energy barrier and accelerates the conformational transition of a protein by orders of magnitude[55,56].

The final structures from the four cMD simulations, after truncating their box dimension as mentioned above, were used as the starting structures for four independent GaMD simulations. The *ParmEd* tool was used to convert the simulation files in AMBER format[57]. The GaMD module implemented in the GPU version of AMBER18[57] was then applied to perform GaMD simulations. To calculate the GaMD acceleration parameters, first we ran a 10 ns cMD simulation with no boot potential and collected the potential statistics e.g. the maximum (V_{max}), minimum (V_{min}), average (V_{ave}) and standard deviation (σ_V) values of system potential energies. Then we performed 10 ns equilibration by adding the boost potential, followed by GaMD production simulations with 0.5–1.5 μs durations (a total of 3.8 μs). All GaMD simulations were performed at "dual-boost" level with one boost potential applied to the dihedral energy term and another to the total potential energy term. In our calculations, the system threshold energy is set as $E = V_{max}$[44]. Further, σ_0, which is the upper limit of σ_V, was set to 6.0 kcal/mol for both dihedral and total energetic terms. The SHAKE algorithm[58] was used for restraining bonds contacting hydrogen atoms. The Langevin thermostat[59] with a damping constant of 2 ps^{-1} was used to maintain temperature of the system at 310 K, under NVT ensemble. The periodic boundary conditions were applied for all of the simulation systems. A 10 Å cut-off was used for the Lennard-Jones interactions. Coulombic interactions were evaluated using the PME algorithm[54] with real space cut-off of 10 Å. The *PyReweighting* toolkit[60] was used for reweighting. Analyses were performed averaged over all GaMD production trajectories, unless stated otherwise. VMD[61] and PyMol[62] were used for visualization.

2.3. Dihedral principal component analysis

In order to construct the free-energy landscape, we have performed the dihedral principal component analysis[63] (dPCA) on the trajectories sampled by GaMD simulations. The backbone dihedral angles are the main degrees of freedom that determine the conformational dynamics of a protein. In the dPCA method, the Cartesian coordinates are replaced by the dihedral angles (φ_i, ψ_i) of the protein backbone, where i = 1, 2,····, n (n is amino acid residues of the protein). To avoid potential problems due to periodicity of dihedral angles, instead of using (φ_i, ψ_i) angles directly in the dPCA, Mu *et al.*[64] used the variables

$$x_{4i-3} = \cos(\varphi_i),\ x_{4i-2} = \sin(\varphi_i)$$

$$x_{4i-1} = \cos(\psi_i),\ x_{4i} = \sin(\psi_i)$$

In this method, first a covariance matrix is constructed using sines and cosines of (φ, ψ) angles. The matrix is then diagonalized to obtain a set of orthogonal principal components (dPCs) and the corresponding eigenvalues. Finally, we projected our trajectory, with reweighted contributions, along the two main dPCs to construct the free-energy landscape in these collective coordinates. The free-energy surface is calculated as

$$\Delta F(\text{dPC1, dPC2}) = -k_B T\,[\ln P(\text{dPC1, dPC2}) - \ln P_{max}(\text{dPC1, dPC2})]$$

where, ln P(dPC1, dPC2) is the probability distribution, from which P_{max} (dPC1, dPC2) is subtracted to ensure that ΔF (dPC1, dPC2) = 0 for the lowest free-energy minimum.

3. Results and discussion

3.1. Secondary structure

We have calculated the secondary structure content of AS(1-103) in water using the widely used computer programme DSSP (Dictionary of Secondary Structure of Proteins)[65]. DSSP assigns α-helix, π-helix, 3_{10}-helix, β-sheet, β-bridge, turn and bend structures to each residue of a protein. If a residue does not satisfy any of the above structure types, it is assigned as coil[65]. Fig. 2A shows the average total secondary structure content of AS(1-103) in water. The

results indicate that AS(1-103) is predominantly disordered in solution with some residual β- and helical structures. The protein adopts 48% coil, 27% bend and 10% turn structures on an average. The average β-sheet and β-bridge contents of AS(1-103) are found to be 8 and 3%, respectively. The helical contents of the protein include only 2% α-helix and 1% 3_{10}-helix. This finding is in agreement with the single-molecule force spectroscopy (SMFS) characterization that predicted 54.5% random coil and 7.3% β-sheet like structure in monomeric α-synuclein in solution, without any interference of soluble oligomeric forms[21]. Formation of helical and β-sheet structures has been proposed to play crucial role in the aggregation of IDP[21,34,43].

Next we have calculated the secondary structure propensity of each residue of AS(1-103) and presented in Fig. 2B-D. We observed that the residues 7-16 and 53-61 located at N-terminal domain exhibit highest propensity to adopt α-helical structure (up to 10 and 14%, respectively) (Fig. 2B). Small residual α-helical content is also detected in the NAC region. The 1-95 residues of α-synuclein contain seven 11-residue repeats that include a highly conserved KTKEGV motif, which is proposed to be associated with the formation of α-helical structure[66]. However, the residual α-helical structure observed in our simulations of AS(1-103) in water is not nearly as extensive as the α-helical structure observed in presence of detergent micelle[14] and in lipid membrane[13]. In addition, our simulations indicate abundant β-sheet formation at the N-terminal domain (Fig. 2C). Resides 37-40 and 46-53 exhibit highest probability (~30% or above) of forming β-sheet. Formation of β-sheet at the N-terminal domain was proposed to be crucial for fibril formation of α-synuclein and the fibril structures determined by solid-state NMR[17] and cryo-electron microscopy[18] show existence of β-strands at N-terminal domain starting from residue 35 and 44, respectively. In our simulations, we also observed significant β-sheet formation at the NAC region (Fig. 2C). The residues 69-83, including the core of the NAC domain consists of residues 71-82, adopt β-sheet with an abundance of up to ~28% during our simulations. The NAC-core of α-synuclein exhibits reportedly the highest propensity to fold into β-sheet structure

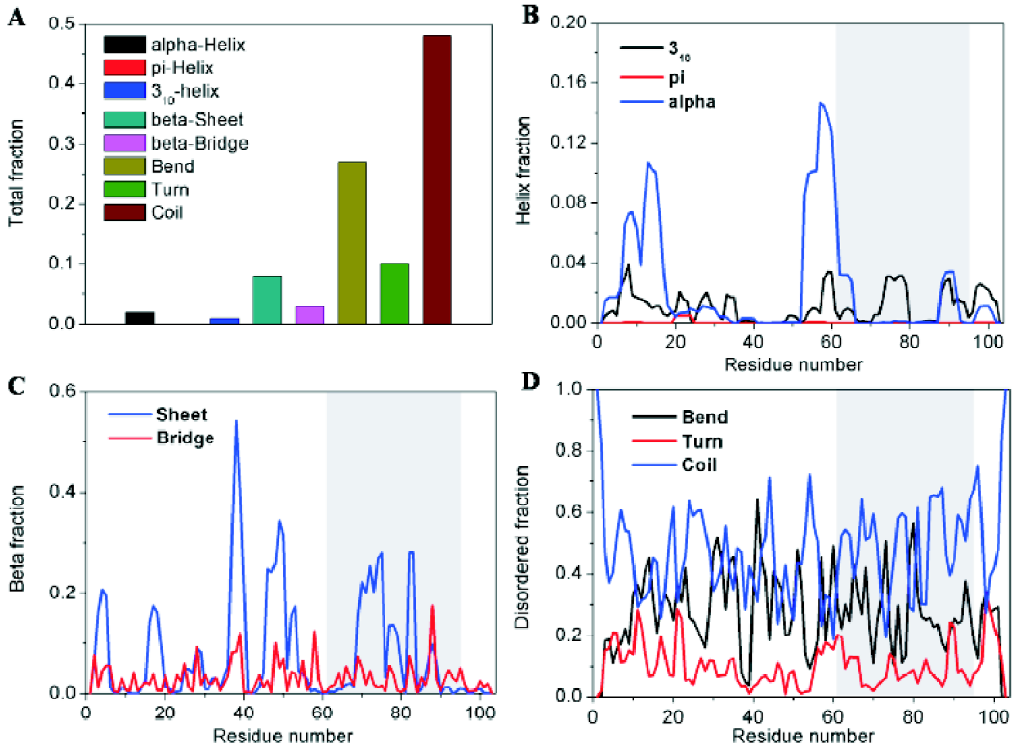

Fig. 2. The secondary structure propensity of AS(1-103) in water. (A) Total fraction of secondary structure components sampled and (B-D) per residue secondary structure propensities, including helix (α-helix, π-helix and 3_{10}-helix), β-(sheets and bridge) and disordered (everything else) secondary structures. In (B-D) the NAC region of the protein is marked as the gray area.

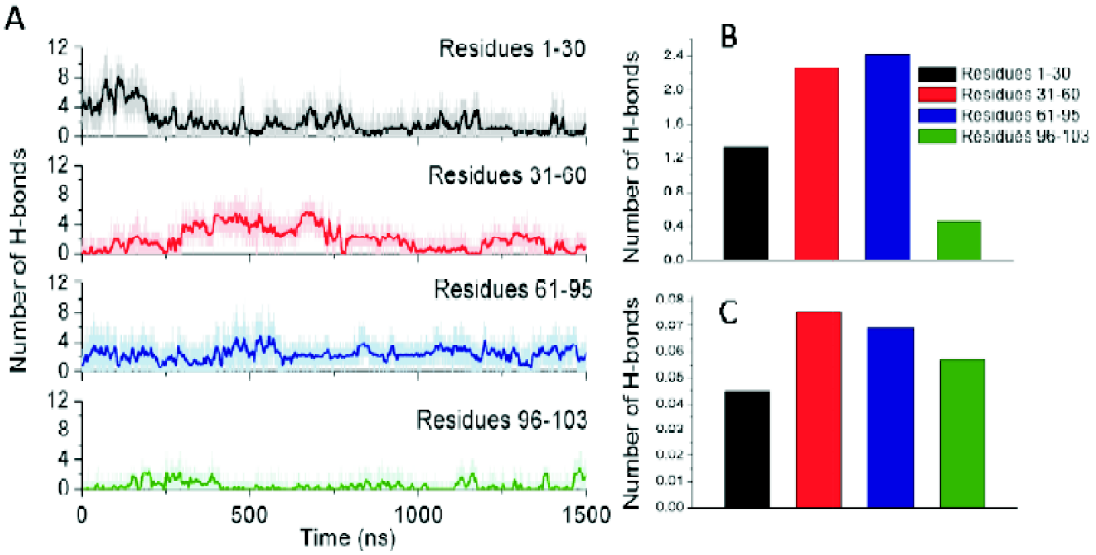

Fig. 3. (A) The time evolution of the number of backbone hydrogen bonds formed in different domains of AS(1-103) along a simulation trajectory (light colours) and the corresponding 500-point running averages are shown in dark colours (B) the time-averaged hydrogen bonds formed by different domains over last 1250 ns of the simulation trajectory. (C) The time-averaged hydrogen bonds divided by the number of residues of the corresponding domain.

and is critical for aggregation of α-synuclein[34,67]. Consistent with our finding, a recent crosslinking guided computational study reported existence of transient secondary structure elements, α-helix and β-sheet, in the conformational ensembles of monomeric α-synuclein in solution[22].

Next we have examined how the hydrogen bond dynamics of the protein backbone is changing with time. For those calculations, we have divided the protein into four domains: N-terminal residues 1-30, residues 31-60 (referred as pre-NAC domain), residues 61-95 (NAC domain) and residues 96-103 (the truncated C-terminus), and calculated the number of backbone hydrogen bonds formed within each domain (Fig. 3). In addition, we have also calculated the pair radial distribution function, $g(r)$, between the backbone heavy atoms of each domain of AS(1-103) and the oxygen atoms of the water molecules (Fig. 4). Here we have presented the results of a single GaMD trajectory; whereas similar results are obtained from other trajectories. As discussed in the above section, the pre-NAC and NAC domain exhibit high secondary structure propensities. In consistent with the secondary structure analysis, we observed that the maximum numbers of intra-domain hydrogen bonds are formed in the pre-NAC and NAC domain (Fig. 3). Accordingly we observed depletion of water from the solvation shell around these two domains, compared to other domains (Fig. 4). We have also observed significant inter-domain interactions between pre-NAC and NAC domains and also between N-terminal and the NAC domain, as discussed in details in the later sections. Such inter-domain interactions also contribute in depleting water molecules from the solvation shells of these

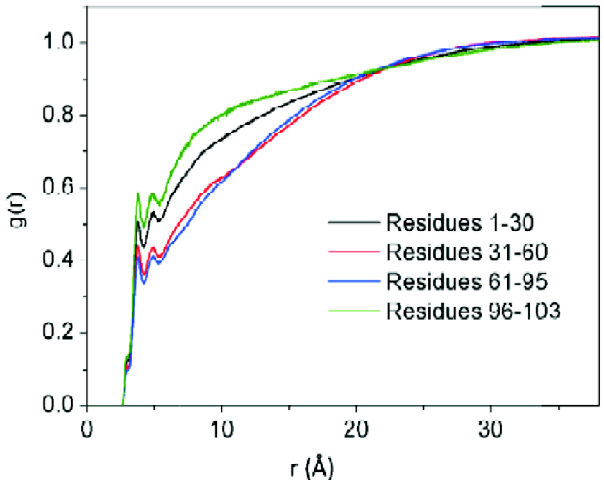

Fig. 4. The pair radial distribution function, $g(r)$, between the backbone heavy atoms of each domain of AS(1-103) and the oxygen atoms of the water molecules.

domains. The N-terminal 1-30 residues exhibit the intermediate behaviour (Figs. 3 and 4). Whereas, the minimum intra-domain hydrogen bonds are formed in the C-terminal domain (Fig. 3), which is the most disordered region of AS(1-103) observed in our simulations. Accordingly the corresponding $g(r)$ plot shows that this domain is highly solvated, with maximum number of water molecules present in its solvation shell (Fig. 4). Summarising, we observed that the domain of AS(1-103) which exhibits the high secondary structure propensity, forms larger number of intra-domain hydrogen bonds and consequently, is less solvated by water molecules and vice-versa.

3.2. Intra-molecular interactions and chain dimension

We have generated the contact map by calculating the fraction of AS(1-103) structures where pairs of residues had at least one pair of heavy atoms within 5 Å of each other (Fig. 5). The contact map shows existence of substantial intra-molecular interactions between different parts of the protein. We have detected strong contacts between residues 34-54 at the N-terminal domain with residues 65-88 at the NAC domain of AS(1-103), with an abundance of ~30-40%. The N-terminus residues 1-8 also found to interact with residues 73-81 of NAC domain. We noted prominent contacts between residues 53-68 and 83-95, located within the pre-NAC and NAC domains. There are intra-domain contacts within the N-terminal domain between residues 13-20 and 33-42 and between residues 36-39 and 41-51. Thus we find moderate contacts between the N-terminal and NAC domains. Such intra-molecular interactions contribute in determining the tertiary structure of a protein.

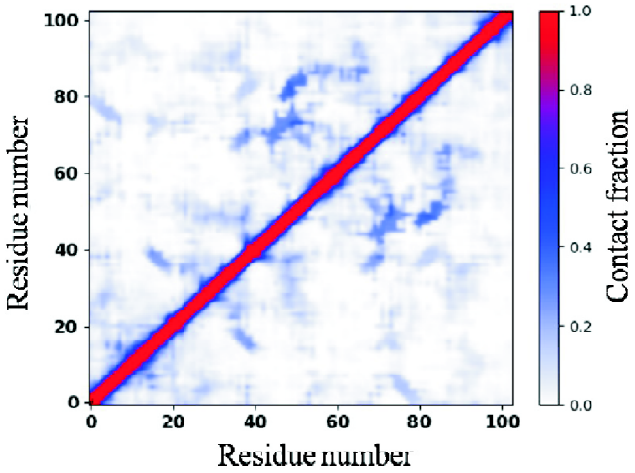

Fig. 5. Residue-residue contact map of AS(1-103).

In order to estimate the dimension of the polypeptide chain, we have calculated the radius of gyration (R_g). Fig. 6 represents the probability distribution of R_g. The probability of structures having R_g values between 13.5 Å to 18.5 Å is large, with a prominent peak at 15 Å. There is relatively much smaller number of structures with larger R_g values, 25-33 Å. These results suggest that the protein predominantly have collapsed polypeptide chain, with small population of extended structures in solution (Fig. 6). A recent computational study combined with experimentally derived short-range crosslinking constraints, reported ~15 Å R_g for the full-length α-synuclein in water[22]. In our simulation the R_g values, 13.5-18.5 Å, adopted by the majority of the AS(1-103) structures are much smaller than the R_g values, 27-35 Å, obtained from NMR[45] and SAXS[46] experiments for the full length α-synuclein in solution. However, we should mention here that in the present work, instead of full length α-synuclein, we have AS(1-103), where residue 104-140, that is nearly all the C-terminus domain is truncated. The highly acidic C-terminal domain is the most disordered part of the protein with highly extended structures, which contributes to the larger R_g value of full length α-synuclein obtained from experiment. Moreover, the conventional experiments, such as NMR and SAXS, provide ensemble averaged information of the protein. Thus the larger R_g values determined by these tech-

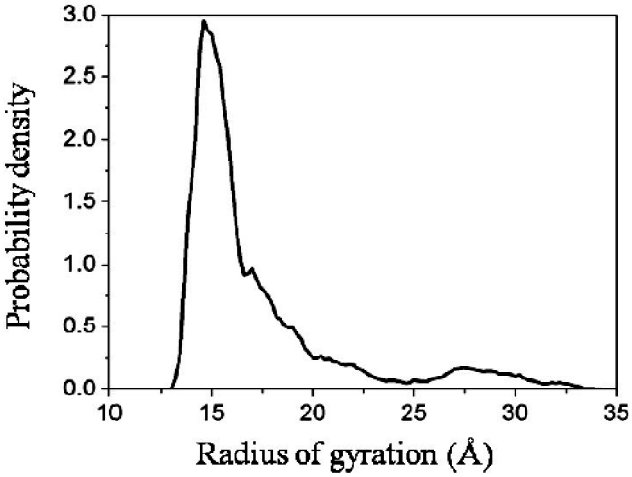

Fig. 6. Probability distribution of the radius of gyration (R_g).

niques, to some extent may also be caused by the co-existing monomeric and oligomeric states of α-synuclein in solution under the experimental conditions[22]. Standard empirical force fields used in MD simulations, often produced overly compact IDP structures. This issue has been taken care in the present work. We have performed extensive atomistic simulations (3.8 μs of GaMD simulations in total) using CHARMM36m force filed[41] with CHARMM-modified TIP3P water model[49], which is reported to be an appropriate force field for IDPs[39,41].

3.3. Free-energy landscape

A protein composed of N atoms has a 3 N dimensional conformational space. Here we have employed dPCA to reduce the dimensionality of this complex conformational space and to study the important collective motions of the protein. The contributions of first five dPCs to the internal coordinates are 15.07, 10.89, 6.43, 4.75 and 4.09%, in the sequential order. Therefore the collective contribution from the first two slowest modes is ~26%. We constructed the two-dimensional free-energy landscape along the first two dPC modes corresponding to the largest eigenvalues (Fig. 7). The free-energy landscape of AS(1-103) contains several energy minima of varying depths. We have clustered the structures populated in the dominant basins and a representative structure from each basin is shown in Fig. 7.

The most dominant basin is referred as "a". β-Sheet conformations are enriched in this basin (Fig. 7A). We observed formation of several β-strands at different parts of the N-terminal and NAC domain. Some of the β-strands are loosely packed together, resulting in a relatively compact structure with R_g 15.01±0.01 Å. We observed formation of a β-hairpin motif in the NAC domain involving residues 69-71 and 77-80. Another β-hairpin formed at the pre-NAC region (defined as residue 31-65 in Fig. 7) of the N-terminal domain between

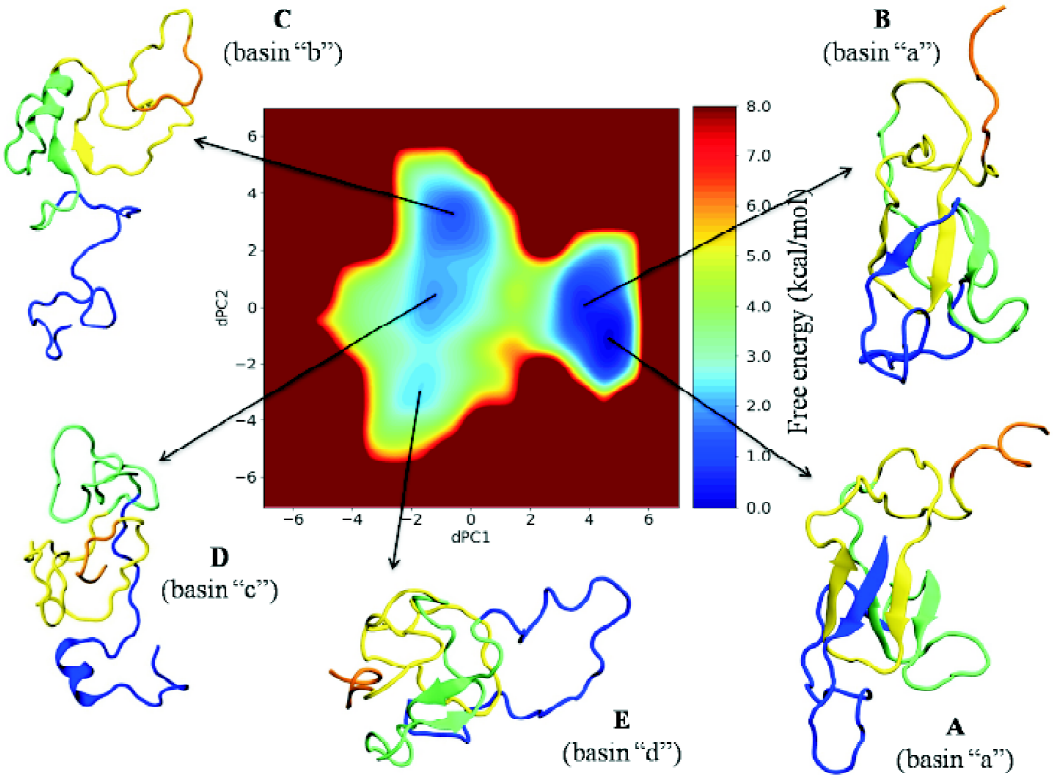

Fig. 7. The free-energy landscape of AS(1-103) in aqueous solution obtained from GaMD simulations as a function of the first two principal components dPC1 and dPC2. The representative structures from different basins are displayed as snapshots, at the periphery of the free-energy landscape. In these snapshots, the N-terminal residues 1-30, pre-NAC region of N-terminal domain (residues 31-60), the NAC domain (residues 61-95) and the C-terminal residues 96-103 are shown in blue, green, yellow and orange, respectively.

residue 37-40 and 51-53. Among which, residue 37-40 is found to interact with above mentioned residue 69-71 of the NAC domain to form a parallel β-sheet, which is the longest persistent β-sheet formed in this cluster (Fig. 7A). The N-terminal 2-5 residues also interact with 77-80 of the NAC domain to form an antiparallel β-sheet. There is also a short β-stand formed by residues 28-29 of the N-terminal domain. The structures populated in basin "a", contain an average 13% β-sheet and 5% β-bridge contents. There is no helical conformation detected in this basin. The structures have predominant disordered regions, with 47% coil, 27% bend and 9% turn contents on an average.

We have also shown a lower energy representative conformation from another cluster of this broad basin "a" in Fig. 7B. Here again, we observed formation of a β-hairpin at the NAC domain by residues 69-72 and 77-79. Residues 37-40 in the pre-NAC region of the N-terminal domain interact with residues 69-72 of NAC domain to form a parallel β-sheet. An antiparallel β-sheet motif has also been formed between the N-terminal 3-5 residues and residues 77-79 of the NAC domain. However, as compared to the lowest energy representative (Fig. 7A), the antiparallel β-sheet is found to be persistently more stable than the parallel one. The structures populated in this cluster have slightly lesser β-content (11% β-sheet and 3% β-bridge) and are more disordered (with 50% coil, 29% bend and 7% turn) regions. An average R_g value of this conformational subpopulation is 15.46±0.02 Å.

Structures in the other two energy basins "b" and "c" are predominantly disordered with mostly partial α-helical structures (Fig. 7C, D). In basin "b" a short α-helix is persistently formed at residues 50-53 of the N-terminal domain. There is also short β-sheet motif formed by residues 38-39 of N-terminal domain and 82-83 of NAC domain (Fig. 7C). Overall, this conformational subpopulation contains 6% α- and 1% 3_{10}-helical contents, 4% β-sheet and 2% β-bridge structures. Whereas, structures of basin "c" exhibit presence of a α-helix involving residues 11-16 (Fig. 7D) and have a total of 7% α-helical, 3% 3_{10}-helical, 2% β-sheet and 4% β-bridge secondary structures. We have identified another minor basin "d", which contains structures with short β-sheet involving residues 35-36 and 49-50 (Fig. 7E), with an average of 3% β-sheet and 2% β-bridge contents and barely any α-helical structures. These three basins, "b", "c", and "d", populated relatively extended polypeptide chains with R_g values 16.85±0.04, 17.11±0.05 and 17.05±0.05 Å, respectively.

Taken together, the free-energy landscape of AS(1-103) reveals coexistence of distinct conformational subpopulations, suggesting that the protein populates an heterogeneous ensemble of conformations in aqueous solution owing to its intrinsically disordered nature. We found that the predominant conformational ensemble of AS(1-103) is enriched in β-structures consists of multiple β-hairpin and β-sheet motifs (Fig. 7). In agreement with our findings, an earlier coarse-grained simulation study had also reported transient β-hairpin formation in full-length α-synuclein monomer in solution[68]. Accumulating evidences from previous experimental studies support the importance of the β-hairpin states in α-synuclein aggregation. In fact, the designed peptides that presumably recognize β-hairpins were able to capture the monomeric state of α-synuclein and block the aggregation process[69,70]. On the other hand, the reagents that are known to stabilize β-hairpins, can promote α-synuclein aggregation[71]. Moreover, the β-sheet structure formation is frequently linked to the aggregation process of α-synuclein[43,72,73]. The increased formation of β-sheet structures in different familial α-synuclein mutants is often associated with greater oligomerization and/or aggregation propensities[25,72,73]. Therefore, the formation of transient β-sheet structures observed in our simulations could promote the formation of early oligomeric states, and might account for the experimentally determined higher aggregation propensity of this C-terminally truncated variant of α-synuclein[28–30].

4. Conclusion

C-terminal truncation is an important modification of full-length α-synuclein, which occurs naturally in cellular environment and has shown to promote the pathogenic aggregation of α-synuclein[28–30]. An earlier experimental study has shown that a physiologically relevant C-terminally truncated α-synuclein AS(1-103), but not the full-length α-synuclein, has the capacity to readily convert into amyloid fibrils at physiological pH[29]. To identify the conformational signature that promotes the misfolding and aggregation process, in this work we have carried out atomistic MD simulations to characterize the conformational ensemble of AS(1-103) monomer in aqueous solution. For an adequate sampling of the confor-

mational space of the protein, we have employed extensive GaMD simulations (a total of 3.8 μs), using CHARMM36m force field[41] that has been successful in reproducing the conformational behaviour of IDPs. Our results show that the protein adopts multiple distinct conformations in solution with varying degree of transient secondary and tertiary structures. Interestingly, we observed spontaneous formation of β-sheet structures at residues 37-53 of the N-terminal and residues 69-82 of the aggregation-prone NAC domain. Formation of such partially folded β-sheet enriched conformations could be the key factor for aggregation proliferation of AS(1-103) observed in experiment[29]. It has been proposed that the negatively charged carboxyl terminus (residues 96-140) of the full-length α-synuclein counteracts protein aggregation through electrostatic interactions with the positively charged N-terminus and also through its interactions with the NAC domain[74,75]. In absence of charged C-terminal segment (residues 104-140), we found that the N-terminal domain freely interacts with the NAC domain, which could be responsible in stabilizing the partially folded β-rich conformations of AS(1-103). The insights obtained from our simulations may help in understanding the molecular basis of the pathogenic involvement of the C-terminally truncated AS(1-103) in Parkinson's disease and prove useful for designing new therapeutics to combat the disease.

Acknowledgement

MM thanks the Department of Biotechnology, Government of India, for financial support under Bio-CARe 2016 program for Women Scientists (No. BT/PR17981/BIC/101/576/2016). This work was performed using the high-performance computing resources of IISER Bhopal.

References

1. C. B. Anfinsen, *Science*, 1973, **181**, 223.
2. A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, *FEBS J*., 2005, **272**, 5129.
3. M. Arai, K. Sugase, H. J. Dyson and P. E. Wright, *Proc. Natl. Acad. Sci. USA*, 2015, **112**, 9614.
4. T. L. Gall, P. R. Romero, M. S. Cortese, V. N. Uversky and A. K. Dunker, *J. Biomol. Struct. Dyn*., 2007, **24**, 325.
5. H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol*., 2005, **6**, 197.
6. P. E. Wright and H. J. Dyson, *Nat. Rev. Mol. Cell Biol*., 2015, **16**, 18.
7. F. Chiti and C. M. Dobson, *Annu. Rev. Biochem*., 2006, **75**, 333.
8. M. G. Spillantini, M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert, *Nature*, 1997, **388**, 839.
9. M. Goedert, *Clin. Chem. Lab. Med*., 2001, **39**, 308.
10. D. F. Clayton and J. M. George, *Trends Neurosci*., 1998, **21**, 249.
11. D. P. Hong, W. Xiong, J. Y. Chang and C. Jiang, *FEBS Lett*., 2011, **585**, 561.
12. T. D. Kim, S. R. Paik and C. H. Yang, *Biochemistry*, 2002, **41**, 13782.
13. G. Fusco, M. Sanz-Hernandez and A. De Simone, *Curr. Opin. Struct. Biol*., 2018, **48**, 49.
14. T. S. Ulmer, A. Bax, N. B. Cole and R. L. Nussbaum, *J. Biol. Chem*., 2005, **280**, 9595.
15. M. A. Fakhree, I. S. Nolten, C. Blum and M. M. Claessens, *J. Phys. Chem. Lett*., 2018, **9**, 1249.
16. J. A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang and L. Jiang, *Nature*, 2015, **525**, 486.
17. M. D. Tuttle, G. Comellas, A. J. Nieuwkoop, D. J. Covell, D. A. Berthold, K. D. Kloepper, J. M. Courtney, J. K. Kim, A. M. Barclay, A. Kendall and W. Wan, *Nat. Struct. Mol. Biol*., 2016, **23**, 409.
18. Y. Li, C. Zhao, F. Luo, Z. Liu, X. Gui, Z. Luo, X. Zhang, D. Li, C. Liu and X. Li, *Cell Res*., 2018, **28**, 897.
19. B. Winner, R. Jappelli, S. K. Maji, P. A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar, S. Campioni and C. Tzitzilonis, *Proc. Natl. Acad. Sci. USA*, 2011, **108**, 4194.
20. R. F. Roberts, R. Wade-Martins and J. Alegre-Abarrategui, *Brain*, 2015, **138**, 1642.
21. M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco and B. Samorì, *PLoS Biol*., 2008, **6**, e6.
22. N. I. Brodie, K. I. Popov, E. V. Petrotchenko, N. V. Dokholyan and C. H. Borchers, *PLoS Comput. Biol*., 2019, **15**, e1006859.
23. S. L. Bernstein, D. Liu, T. Wyttenbach, M. T. Bowers, J. C. Lee, H. B. Gray and J. R. Winkler, *J. Am. Soc. Mass. Spectr*., 2004, **15**, 1435.
24. L. Breydo, J. W. Wu and V. N. Uversky, *Biochim. Biophys. Acta Mol. Basis. Dis*., 2012, **1822**, 261.
25. P. Flagmeier, G. Meisl, M. Vendruscolo, T. P. Knowles, C. M. Dobson, A. K. Buell and C. Galvagnion, *Proc. Natl. Acad. Sci. USA*, 2016, **113**, 10328.
26. V. N. Uversky and D. Eliezer, *Curr. Protein Pept. Sci*., 2009, **10**, 483.
27. M. Manna and C. Mukhopadhyay, *PloS One*, 2013, **8**, e71308.
28. W. Li, N. West, E. Colla, O. Pletnikova, J. C. Troncoso, L. Marsh, T. M. Dawson, P. Jäkälä, T. Hartmann, D. L. Price

and M. K. Lee, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 2162.

29. I. M. van der Wateren, T. P. Knowles, A. K. Buell, C. M. Dobson and C. Galvagnion, *Chem. Sci*., 2018, **9**, 5506.
30. Z. A. Sorrentino, N. Vijayaraghavan, K. M. Gorion, C. J. Riffe, K. H. Strang, J. Caldwell and B. I. Giasson, *J. Biol. Chem*., 2018, **293**, 18914.
31. M. Baba, S. Nakajo, P. H. Tu, T. Tomita, K. Nakaya, V. M. Lee, J. Q. Trojanowski and T. Iwatsubo, *Am. J. Pathol*., 1998, **152**, 879.
32. D. Eliezer, *Curr. Opin. Struct. Biol*., 2009, **19**, 23.
33. P. Das, S. Matysiak and J. Mittal, *ACS Cent. Sci*., 2018, **4**, 534.
34. J. R. Allison, R. C. Rivers, J. C. Christodoulou, M. Vendruscolo and C. M. Dobson, *Biochemistry*, 2014, **53**, 7170.
35. S. H. Chong, P. Chatterjee and S. Ham, *Annual Rev. Phys. Chem*., 2017, **68**, 117.
36. N. Stanley, S. Esteban-Martin and G. De Fabritiis, *Prog. Biophys. Mol. Biol*., 2015, **119**, 47.
37. S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw, *J. Phys. Chem. B*, 2015, **119**, 5113.
38. P. Robustelli, S. Piana and D. E. Shaw, *Proc. Natl. Acad. Sci. USA*, 2018, **115**, E4758.
39. J. Lincoff, S. Sasmal and T. Head-Gordon, *J. Chem. Phys*., 2019, **150**, 104108.
40. G. H. Zerze, W. Zheng, R. B. Best and J. Mittal, *J. Phys. Chem. Lett*., 2019, DOI: 10.1021/acs.jpclett.9b00850.
41. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell (Jr.), *Nat. Methods*, 2017, **14**, 71.
42. N. Ahalawat, S. Arora and R. K. Murarka, *J. Phys. Chem. B*, 2015, **119**, 11229.
43. O. Wise-Scira, A. K. Aloglu, A. Dunn, I. T. Sakallioglu and O. Coskuner, *ACS Chem. Neurosci*., 2013, **4**, 486.
44. Y. Miao, V. A. Feher and J. A. McCammon, *J. Chem. Theory Comput*., 2015, **11**, 3584.
45. A. S. Morar, A. Olteanu, G. B. Young and G. J. Pielak, *Protein Sci*., 2001, **10**, 2195.
46. M. Schwalbe, V. Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M. R. Jensen, J. Biernat, S. Becker, E. Mandelkow and M. Zweckstetter, *Structure*, 2014, **22**, 238.
47. S. Arya, A. K. Singh, K. Bhasne, P. Dogra, A. Datta, P. Das and S. Mukhopadhyay, *Biophys. J*., 2018, **114**, 2540.
48. A. Iyer, S. J. Roeters, N. Schilderink, B. Hommersom, R. M. Heeren, S. Woutersen, M. M. Claessens and V. Subramaniam, *J. Biol. Chem*., 2016, **291**, 21110.
49. A. D. MacKerell (Jr.), D. Bashford, M. L. Bellott, R. L. Dunbrack (Jr.), J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha and D. Joseph-McCarthy, *J. Phys. Chem. B*, 1998, **102**, 3586.
50. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem*., 2005, **26**, 1701.
51. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19.
52. B. Hess, H. Bekker, H. J. Berendsen and J. G, Fraaije, *J. Comput. Chem*., 1997, **18**, 1463.
53. G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys*., 2007, **126**, 014101.
54. T. Darden, D. York and L. Pedersen, *J. Chem. Phys*., 1993, **98**, 10089.
55. Y. Miao and J. A. McCammon, *Annu. Rep. Comput. Chem*., 2017, **13**, 231.
56. Y. Miao and J. A. McCammon, *Proc. Natl. Acad. Sci. USA,* 2018, **115**, 3036.
57. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, AMBER 2018, University of California, San Francisco, 2018.
58. J. P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys*., 1977, **23**, 327.
59. G. S. Grest and K. Kremer, *Phys. Rev. A*, 1986, **33**, 3628.
60. Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker and J. A. McCammon, *J. Chem. Theory Comput*., 2014, **10**, 2677.
61. W. Humphrey, A. Dalke and K. Schulten, *J. Molec. Graphics*, 1996, **14**, 33.
62. W. L. DeLano, *CCP4 Newsletter on Protein Crystallography*, 2002, **40**, 82.
63. A. Altis, P. H. Nguyen, R. Hegger and G. Stock, *J. Chem. Phys*., 2007, **2126**, 244111.
64. Y. Mu, P. H. Nguyen and G. Stock, *Proteins: Struct., Funct., Bioinf*., 2005, **58**, 45.
65. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577.
66. U. Dettmer, A. J. Newman, V. E. von Saucken, T. Bartels and D. Selkoe, *Proc. Natl. Acad. Sci. USA*, 2015, **112**, 9596.
67. L. Xu, S. Bhattacharya and D. Thompson, *Phys. Chem. Chem. Phys*., 2018, **20**, 4502.
68. H. Yu, W. Han, W. Ma and K. Schulten, *J. Chem. Phys*., 2015, **143**, 12B623_1.
69. E. A. Mirecka, H. Shaykhalishahi, A. Gauhar, Ş. Akgül, J. Lecher, D. Willbold, M. Stoldt and W. Hoyer, *Angew. Chem. Int. Ed*., 2014, **53**, 4227.

70. K. N. Huggins, M. Bisaglia, L. Bubacco, M. Tatarek-Nossol, A. Kapurniotu and N. H. Andersen, *Biochemistry*, 2011, **50**, 8202.

71. F. Rose, M. Hodak and J. Bernholc, *Sci. Rep*., 2011, **1**, 11.

72. J. Li, V. N. Uversky and A. L. Fink, *Neurotoxicology*, 2002, **23**, 553.

73. J. Li, V. N. Uversky and A. L. Fink, *Biochemistry*, 2001, **40**, 11604.

74. I. V. Murray, B. I. Giasson, S. M. Quinn, V. Koppaka, P. H. Axelsen, H. Ischiropoulos, J. Q. Trojanowski and V. M. Lee, *Biochemistry*, 2003, **42**, 8530.

75. W. Hoyer, D. Cherny, V. Subramaniam and T. M. Jovin, *Biochemistry*, 2004, **43**, 16233.