Rationalizing the Relative Aggregability of Alpha-Synuclein Variants via Molecular Simulations

A Thesis

Submitted To
Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Sachin Chaudhary
(Reg. No. 20201085)

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

Jun 2024 - May 2025

Supervisor: Dr. Atanu Das
(CSIR- National Chemical Laboratory, Pune)

Certificate

This is to certify that this dissertation entitled Rationalizing the Relative Aggregability of Alpha-Synuclein Variants via Molecular Simulations, towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Sachin Chaudhary at CSIR-National Chemical Laboratory under the supervision of Dr. Atanu Das, Senior Scientist, Physical and Materials Chemistry Division, during the academic year 2024-2025.

Dr. Atanu Das

Prof. Anirban Hazra

Committee:

Dr. Atanu Das
Prof. Anirban Hazra

This thesis is dedicated to my Parents.

Declaration

I hereby declare that the matter embodied in the report entitled " Rationalizing the Relative Aggregability of Alpha-Synuclein Variants via Molecular Simulations " are the results of the work carried out by me at the Department of Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory Pune, under the supervision of Dr. Atanu Das, and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Signature of Student

Sachin Chaudhary

(Roll No. 20201085)

Name of Student

Acknowledgement

On this journey of my MS thesis, I have received tremendous support from many people. Without their help, I would not have been able to complete this work, In this section, I would like to thank all those who helped me during this journey.

First and foremost, I would like to thank my thesis supervisor, Dr**.** Atanu Das, for allowing me to work in his lab. His guidance, encouragement, and support throughout my MS thesis have been invaluable.

I am also thankful to Prof. Anirban Hazra for accepting to be my thesis expert and for providing valuable inputs that helped shape my work.

I want to give a special thanks to my lab mate, Vaishnavi**,** for always being there to help me throughout my thesis. I would also like to thank Sowmomita for bringing the fun element into the lab environment.

I would also like to thank the Director of NCL and IISER Pune for providing me with this opportunity, and I appreciate IISER Pune, NIT Trichy, and NCL PUNE for making supercomputing facilities available for my research.

A big thank you to my friends Gaurang, Ritesh, Debarghya, and Dhruv for their support during our undergraduate years.

Thank you to everyone who has been a part of this journey. Your help and encouragement have meant a lot to me.

Abstract

The aberrant aggregation of α-Synuclein (αS), an intrinsically disordered protein, leads to a fatal neurodegenerative disease – Parkinson's disease (PD). This study investigates the kinetics and thermodynamics of αS aggregation using extensive coarse-grained (CG) and atomistic simulations, implementing both conventional and enhanced sampling techniques. We examined the self-assembly of wild-type (WT) αS and its five abundant deleterious mutants (A30P, A53T, E46K, G51D, and H50Q) to understand their aggregation behaviour. αS mutants that alter charge (E46K, G51D, and H50Q) exhibit distinct trends in their relative rates of self-association and association with the WT chain. Among the neutral mutations, A30P dimerizes (either with itself or with WT αS) at a faster rate than A53T. The backbone contributes to the intra-chain hydrogen bond framework stability and the side chains dictate the inter-chain and protein-water interactions, possibly being more exposed. Whereas, the electrostatic and van der Waals forces specifically stabilize the intra-chain interactions over the inter-chain and protein-water talks – the former being the major protagonist. However, the pattern changes when the inter-chain binding affinity of the matured protofibrillar state is considered, compared to the initially aggregated constructs. One of the neutral mutations (A53T) and one of the charge-altering mutations (H50Q) show the highest and the lowest protofilament binding affinities (respectively) in homo-dimers, whereas E46K (a different charge-altering mutation) displays the highest binding affinity in hetero-dimers. Interestingly, the rate of association does not correlate with the protofilament binding affinity pattern, suggesting that faster aggregation does not necessarily lead to stable aggregates.

Contents

1: Introduction

Proteins are polymers made up of amino acids. A specific class of proteins, known as globular proteins, folds into a distinct three-dimensional structure, either independently or with the help of chaperone molecules. This folded state is crucial for maintaining their biological activity and function. In contrast, some proteins, called intrinsically disordered proteins, lack a defined three-dimensional structure. Changes in cellular conditions like pH, temperature and mutation can lead to changes in this defined structure which are protein misfolding or protein unfolding.[1] Protein misfolding occurs when proteins do not fold into their proper three-dimensional shapes, which is essential for biological functions. These misfolded proteins form aggregates because misfolding exposes hydrophobic regions that should typically remain hidden. Misfolded proteins tend to undergo a process called protein aggregation, where proteins self-assemble to form macromolecular assemblies called protein aggregates. Protein aggregation is implicated in several diseases, such as Alzheimer's disease (AD), Parkinson's disease (PD), and some forms of cancer.[2]

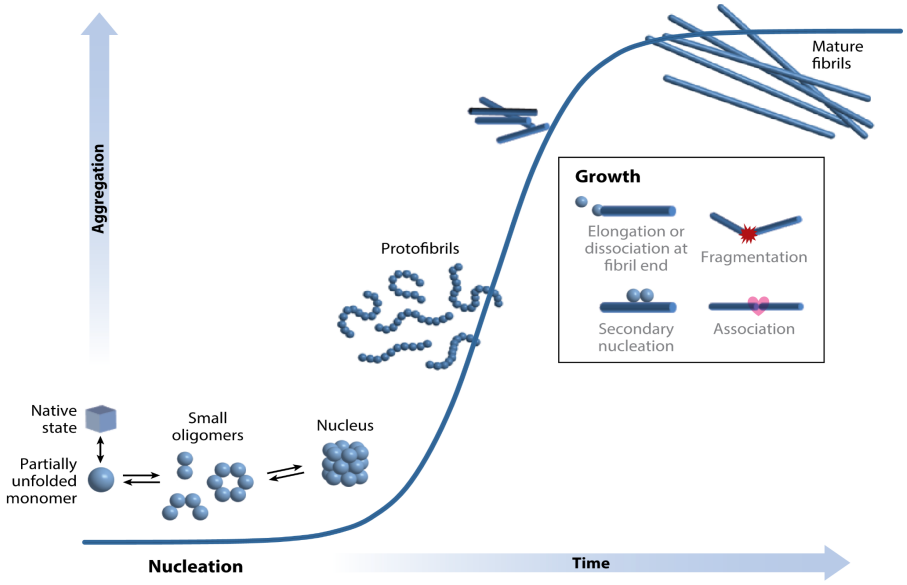


Figure 1. The nucleation growth mechanism of fibril formation.[3]

Experimental studies on protein aggregation demonstrate that fibril formation follows a characteristic sigmoidal curve, which represents a standard nucleation mechanism. In this mechanism, partially folded monomers associate with one another and form a critical nucleus, which is termed the “nucleation phase”. Once the critical nucleus forms, a small fibril starts to grow. After the nucleation phase, the growth phase begins, where fibrils elongate by adding more monomers. The elongation of the fibril occurs through one or many of the growth mechanisms, as highlighted in Figure 1. [4–7] After completion of the growth phase, i.e. the log phase, we reach a saturation phase, which is characterised by the presence of the matured aggregates.

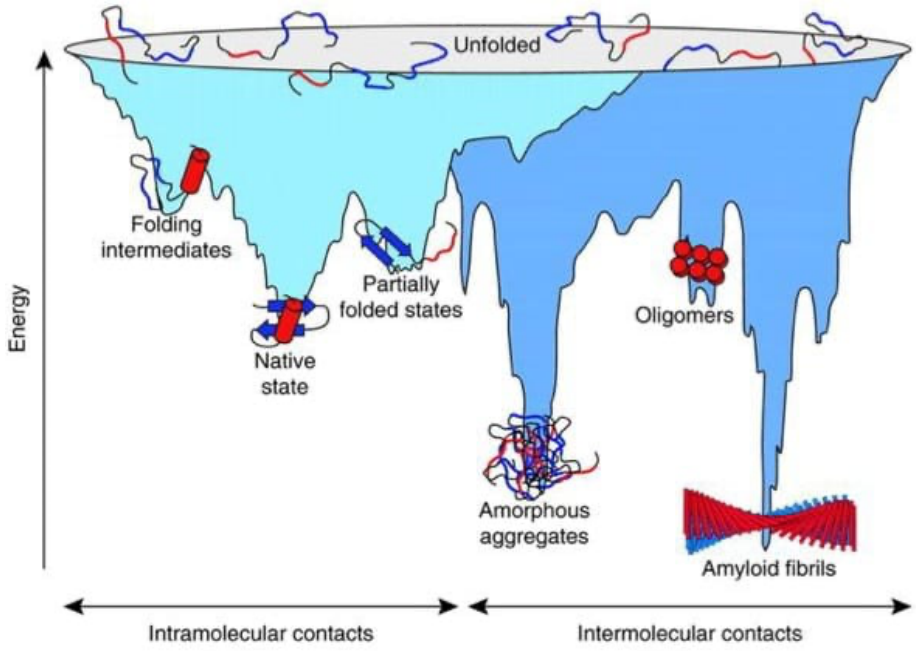


Figure 2. The free-energy landscape of protein folding, misfolding, and aggregation.[8]

In Figure 1, we have seen a generic overview of protein aggregation through the lens of kinetics. The thermodynamic perspective of protein aggregation can be understood by the energy landscape theory of protein folding proposed by Ken A. Dill in 1997. Figure 2 shows a representative free energy landscape of protein aggregation. Mainly, proteins have two states: the unfolded state with broad conformational space and the folded state with narrow conformational space for local fluctuations. Under cellular conditions, a protein's native state corresponds to its lowest free energy state.[9] Proteins that are

partially unfolded can be trapped in non-native local minima. A folding funnel, which creates deep free energy minima, can lead to a native state with a single well-defined tertiary structure. Typically, the aggregation-prone regions of the protein are within the hydrophobic core, so proteins may need to unfold to some extent before they can proceed to aggregation. Protein folding is primarily driven by intramolecular contacts, whereas protein aggregation is driven by intermolecular contacts. That's why basins representing protein aggregation are separated from folded states in the energy landscape.[8]

The need to study protein aggregates appears due to its role in several fatal diseases and one of the diseases is PD. Dr. James Parkinson first described PD in 1817 in his work "An Essay on the Shaking Palsy."[10] It is a neurodegenerative disorder. According to global estimates from 2019, more than 8.5 million people were living with PD.[11] PD affects movement, causing symptoms such as tremors, rigidity, and bradykinesia.[12]

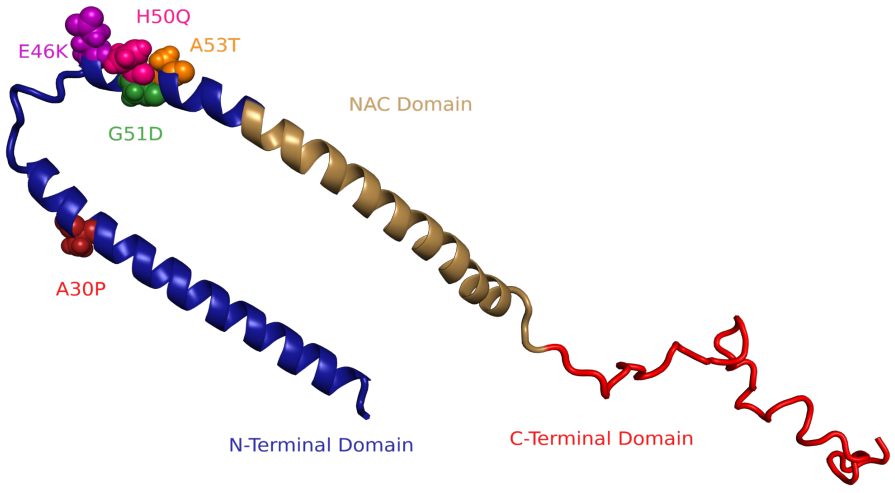


Figure 3. Human micelle-bound NMR structure of α-synuclein (αS).

The protein found to form Lewy bodies, a hallmark of PD, is αS. It is abundant in the human brain, and controls synaptic function and neurotransmitter release. αS is an intrinsically disordered protein that has three main regions, each responsible for different molecular and biological functions.[13] The N-terminus region is from amino acid residues

1-60. This region contains amphipathic repeats that form a α-helix structure. This structural feature is crucial for αS's ability to interact with cell membranes.[14] The non-amyloid-ß component (NAC region) is from amino acid residues 61-95. The C-terminus is from amino acid residues 96-140. This region is negatively charged and involved in Ca2+ binding. [15]

Since this protein has been implicated to a life-threatening disease, it is essential to understand the process of its aggregation. While the natural variant, i.e., the wild-type (WT) αS is known to aggregate in the elderly population and lead to disease, there are also known mutations to αS which are familial and more prone to aggregation which sometimes leads to early-onset of the disease, some of which are shown in table 1.

Mutation	Amino Acid Change
A30P	Alanine → Proline (30)
A53T	Alanine → Threonine (53)
E46K	Glutamic Acid → Lysine (46)
G51D	Glycine → Aspartic Acid (51)
H50Q	Histidine → Glutamine (50)

Table 1. Five abundant deleterious mutations in αS. [16,17]

As both WT and a specific mutant (A30P and A53T are neutral mutations, whereas E46K, G51D, and H50Q are charge-altering mutations) can be found together in an individual, the focus of our project was to study the kinetics and thermodynamics of both the self- and cross-assembly of WT and its mutants to better understand the molecular mechanism of PD pathogenesis. To investigate protein aggregation, we performed molecular dynamics (MD) simulations of protein dimers, which is a model to study protein

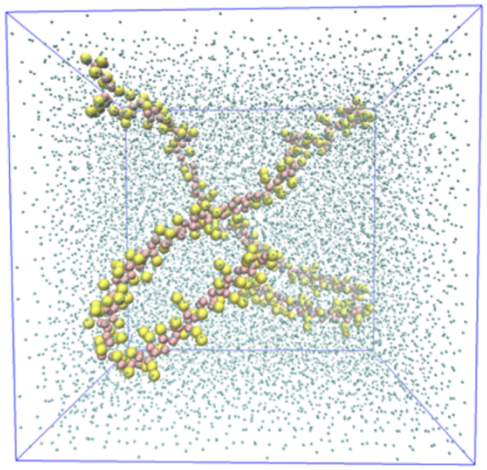
aggregation. We used extensive CG and atomistic simulations, along with conventional and enhanced sampling techniques, to study the kinetics and thermodynamics of aggregation.

Because of the huge size of our system, we used CG simulations to investigate the self- and cross-assembly processes over long time scales. However, CG simulations have limitations, as they cannot capture hydrogen bonding, electrostatic interactions, and van der Waals forces. To compensate for these shortcomings, we employed atomistic simulations. Additionally, we used enhanced sampling techniques with CG description of the protein systems to gain deeper insights into the thermodynamics of aggregation.

2: Methodology

2.1. Unconstrained CG simulations of two randomly placed protein chains to quantify the self-assembly process.

A



B

Homo-dimer	Hetero-dimer
WT-WT	WT-A30P
A30P-A30P	WT-A53T
A53T-A53T	WT-E46K
E46K-E46K	WT-G51D
G51D-G51D	WT-H50Q
H50Q-H50Q	

Figure 4. The simulation box with proteins placed at least 1.5 nm away for one another. Table showcases all the 11 systems that we simulated in our study.

GROMACS software version 2022.1[18] And MARTINI force field version 2.2.[19] Were used to perform all MD simulations. We used the PyMOL software package[20] for the mutation of the WT structure of αS (PDB ID: 1XQ8 [21]). MARTINI python script *martinize.py* was used to convert the all-atom system to a MARTINI coarse-grain system. Subsequently, two monomers were placed in a simulation box using the Packmol package.[22] Where two monomers were placed randomly in a box that was at least 1.5 nm away from one another. Afterwards, the system was solvated using the MARTINI representation of water molecules.[23] Then the system was minimized by using the steepest descent algorithm. After minimization, each system was equilibrated for 200 ps under an *NPT* ensemble at 300 K temperature and 1 bar pressure using a Berendsen thermostat [24] and barostat [24] For faster convergence. Final production was done for 200 μs at 300K and 1 bar pressure using the modified Berendsen thermostat[25] and Parrinello–Rahman barostat.[26] A standard minimum image convention with PBC was applied in all three directions.

2.2. CG conformations back traced to all-atom for the simulations of the dominant dimeric states.

Figure 5. CG to all-atom conversion scheme.

From all the 11 self-associated systems the most possible states were identified using Principal Component Analysis (PCA) [27,28]. Each of these 11 CG configurations was reverse-mapped using the Martini Backward program. [29], where the CG structures were back-traced to atomistic configurations. Each of these 11 back-traced atomistic structures was used for atomistic simulations.

The system was parameterized using the CHARMM36m force field.[30] Subsequently, each system was solvated with the TIP4P-D explicit water model [31] in a triclinic box. Then the genion module was used to make the system electroneutral by randomly placing Na^+ and Cl^- ions in the place of water molecules and the salt concentration was fixed at 100 mM. The steepest descent method of energy minimization was used to remove steric clashes in the system. To equilibrate the water molecules around the protein, an *NVT* equilibration was performed for 100 ps at 300 K temperature with the modified Berendsen thermostat.[25] Then, an *NPT* equilibration was performed for 100 ps at 300 K temperature and 1 bar pressure using the modified Berendsen thermostat[25] and Parrinello–Rahman barostat.[26] The final simulation was performed for 100 ns at 300 K temperature and 1 bar pressure using the modified Berendsen thermostat[25] and Parrinello–Rahman barostat[26] . The standard minimum image convention with periodic boundary conditions (PBC) was applied in all three directions; the Particle Mesh Ewald (PME)[32] method was applied to deal with long-range electrostatic interactions having a real-space cutoff of 10 Å; and a

distance cutoff of 10 Å was used for Lennard-Jones interactions having a Fourier spacing of 1.6 Å.[33]

2.3. Steered molecular dynamics (SMD) and umbrella simulations of the protofibrils.

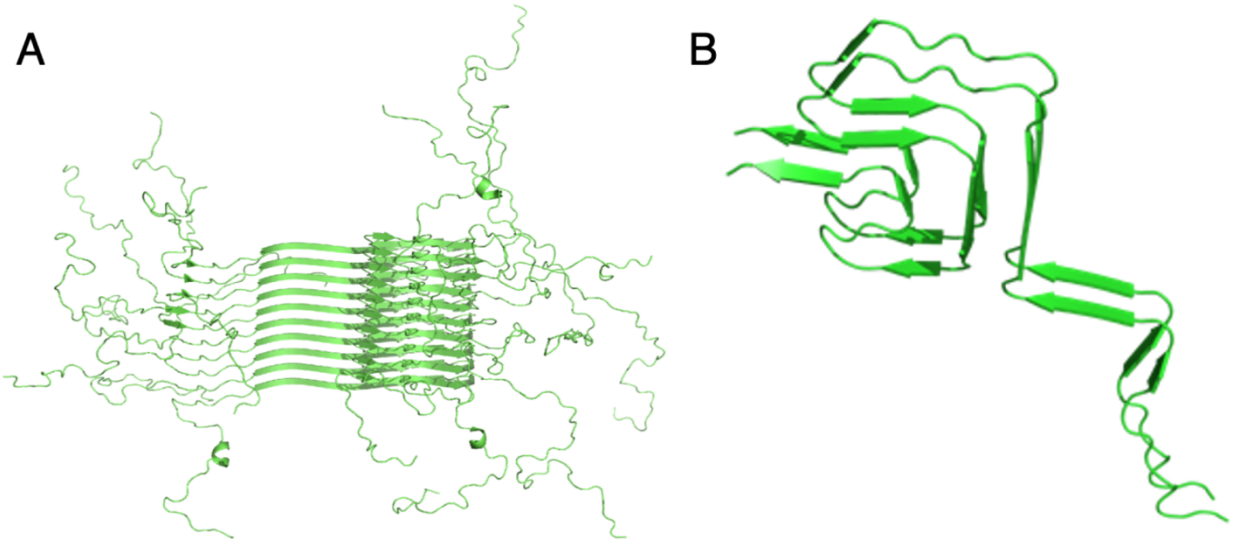


Figure 6. (A) The complete fibril structure of αS (PDB ID: 2N0A [34]). (B) The protofibrillar state of αS is used for SMD and umbrella sampling.

The structure shown in Figure 6B was used for SMD and umbrella sampling. From the fibrillar structure of αS Figure 6A two chains F&G with residues 28-98 were extracted using PyMOL software Package[20] for umbrella sampling, which is shown in Figure 6B. This was done to remove the disordered regions in the structure to ensure better convergence and avoid any artefacts they may introduce. The MARTINI python script was used to convert the all-atom system to a MARTINI coarse-grain system. The system was solvated with the MARTINI representation of water molecules. The steepest descent method of energy minimization was used to remove steric clashes in the system. Then, *NPT* equilibration was performed for 200 ps and at 300 K temperature and 1 bar pressure using the modified Berendsen thermostat. [25] and Parrinello–Rahman barostat [26]. For the

entire duration of equilibration, the protofibril was kept at one of the edges of the box using a harmonic force constant of 1000 kJ mol^{-1} nm^{-2}. Finally, the centre of mass of one chain was held fixed, and that of the other chain was pulled along the *Y*-axis, having a pulling rate of 0.1 Å/ps over a time of ~2000 ps, utilizing a harmonic spring with a spring constant of 1000 kJ mol^{-1} nm^{-2}. After that, configurations were taken for centre of mass (COM) distances from 0.5 nm to 3 nm along the y-axis using 0.1 nm spacing and for 3 nm to 6 nm along the y-axis using 0.2 nm spacing. NPT equilibration was performed for 200 ns for each configuration using Berendsen thermostat[24] and barostat [24] at 300K temperature and 1 bar pressure. A final simulation run of 1 μs was performed for each configuration, using a berendsen thermostat[24] and barostat [24]. Then Weighted Histogram Analysis Method (WHAM)[35] was used to extract the potential of mean force (PMF).

3: Results

3.1. Kinetics of association.

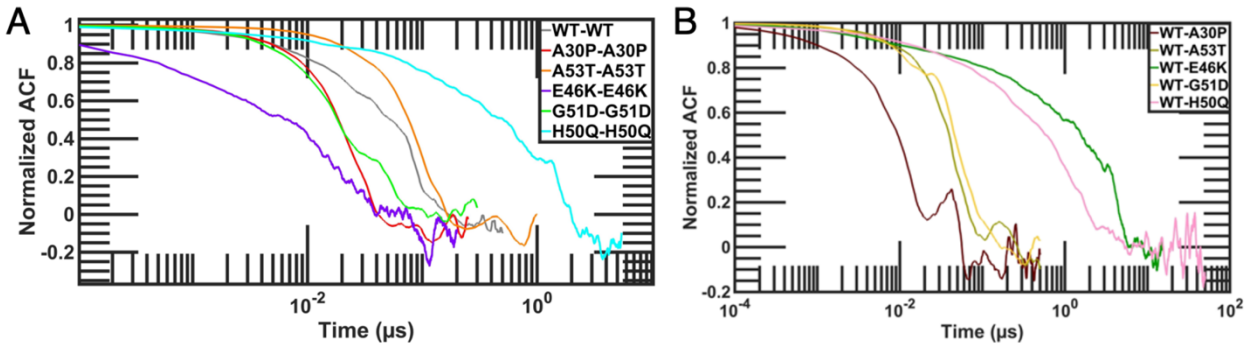


Figure 7. Normalized ACF of inter-chain distances of – (A) homo-dimers and (B) hetero-dimers.

The data presented in Results sections 3.1 to 3.7 were obtained from the coarse-grained simulations described in Methodology section 2.1.

The distance between the centres of masses of two protein chains of a dimeric species was calculated as a function of time, and the normalized auto-correlation function(ACF) of the distance was calculated for each of the eleven systems (Figure 7). The timescale of association are as follows: the homo- (E46K-E46K< A30P-A30P < G51D-G51D < WT-WT < A53T-A53T < H50Q-H50Q, Figure 7A) and hetero-dimeric (WT-A30P < WT-A53T < WT-G51D < WT-H50Q < WT-E46K, Figure 7B) cases – (1) among the mutations that alter charge density of the protein chain, the rate of association decreases from E46K to G51D to H50Q in homo-dimers and G51D to H50Q to E46K in heterodimers (2) between the neutral mutations, A30P shows faster rate of association over A53T.

3.2. The overall trend of the global compactness of the charged mutations retains its identity in homo- and hetero-dimers.

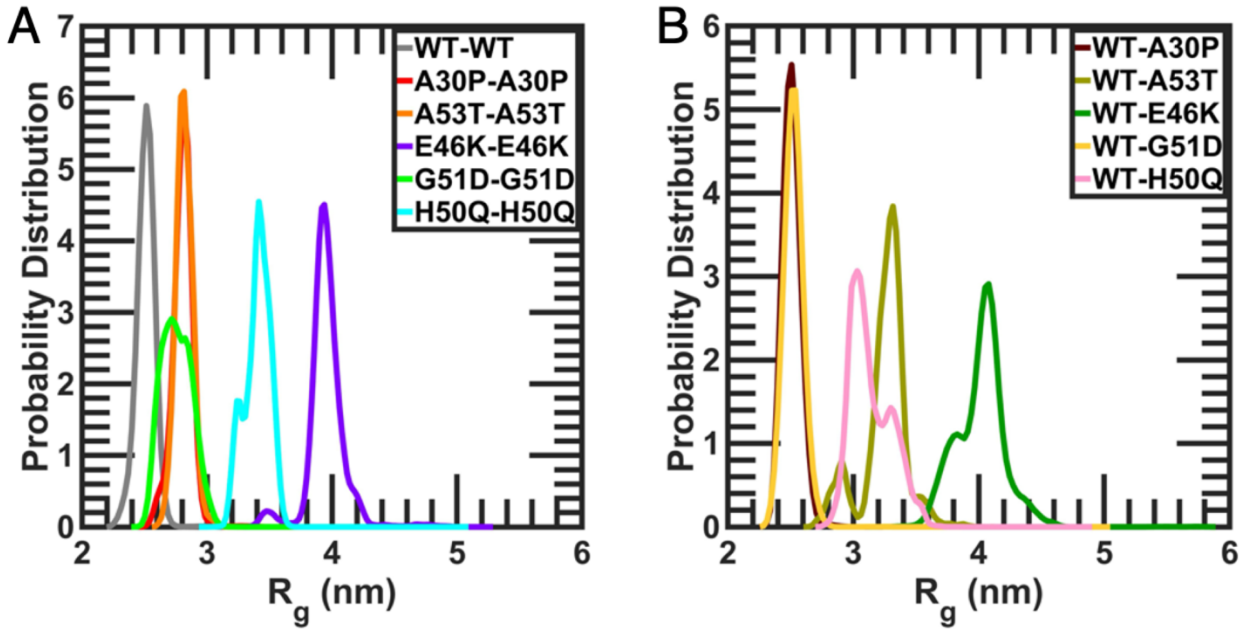


Figure 8. Normalized probability distributions of R_g – (A) homo-dimers and (B) hetero-dimers.

R_g (radius of gyration) is a measure of the distribution of mass around the centre of mass of a molecule and describes how compact a system is. This parameter was utilized to rank order the global compactness of the six homo- and five hetero-dimeric species evolved in the kinetic assembly simulations.

For homo-dimers (Figure 8A), the relative order of R_g displays the following trend: WT-WT < G51D-G51D < A53T-A53T ≈ A30P-A30P < H50Q-H50Q < E46K-E46K. Here we see that two out of three mutants that alter charges (E46K and H50Q) show the least compact and 2nd least compact structures, respectively, while the WT homo-dimer produces the most compact state. Interestingly, the neutral homo-dimers generate ensembles with comparable compactness.

For the hetero-dimers (Figure 8B), the relative order of R_g appears as the following: WT-A30P ≈ WT-G51D < WT-H50Q < WT-A53T < WT-E46K. Interestingly, the relative order

of global compactness among the hetero-dimers involving mutations that alter charges remains the same as observed for homo-dimers, i.e., the ability to generate compact ensembles decreases from G51D to H50Q to E46K. Additionally, as observed for the homo-dimers, E46K mutations tend to create the least compact hetero-dimer as well. However, the neutral mutations break the comparable relative compactness trend as the A30P hetero-dimer produces a much more compact ensemble than the A53T hetero-dimer.

3.3. The variability in the solvent exposure among the αS variants mimics the trend of global compaction.

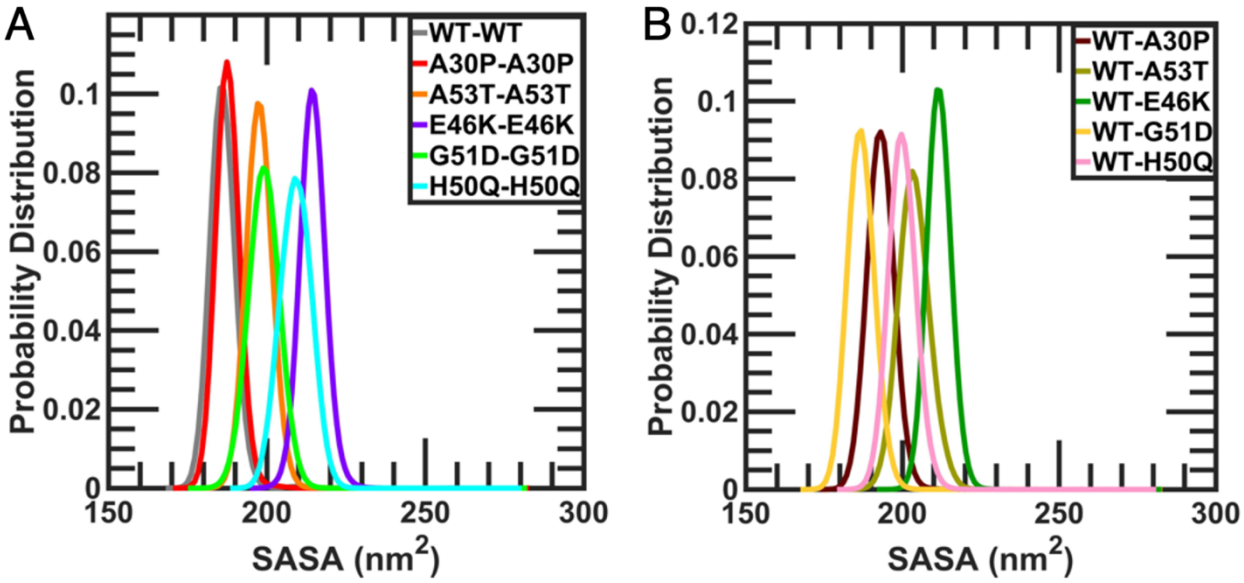


Figure 9. Normalized probability distributions of SASA – (A) homo-dimers and (B) hetero-dimers.

A broadly correlated parameter to the compactness of a protein is its solvent exposure which is measured in terms of solvent-accessible surface area (SASA), the higher the compaction, the lower the solvent exposure.

The relative order of solvent exposure of the homo-dimers (WT-WT ≈ A30P-A30P < A53T-A53T ≈ G51D-G51D <H50Q-H50Q < E46K-E46K, Figure 9A) matches nicely with

the compaction trend in multiple counts: (1) WT and A30P have the least exposure, (2) A53T and G51D have comparable exposures, (3) solvent exposure increases from G51D to H50Q to E46K, and (4) E46K has the maximum solvent exposure.

Similarly, the hetero-dimers display a solvent exposure trend which is very similar to their compactness order: WT-G51D < WT-A30P < WT-H50Q < WT-A53T < WT-E46K (Figure 9B). Moreover, the relative trend of solvent exposure of the homo-dimeric species retains its signature in the hetero-dimers, specifically for mutants which change the charge density of the protein chain.

3.4. Compaction signature of the dimer-forming individual chains persists in the global compactness.

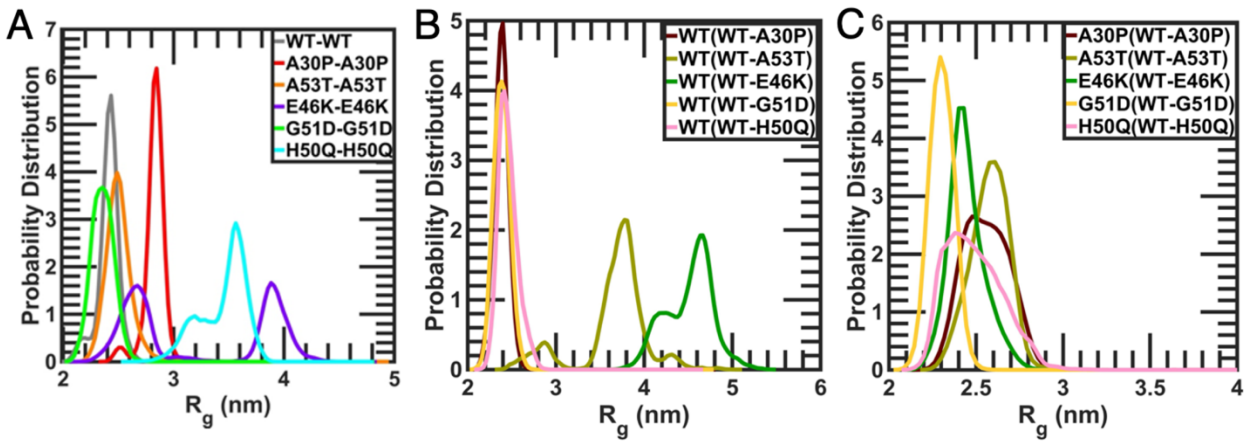


Figure 10. Normalized probability distributions of R_g of individual chains – (A) homo-dimers, (B) WT chain of hetero-dimers, and (C) mutant chain of hetero-dimers.

For the homo-dimers, we concatenated the R_g values of the individual chains (Figure 10A). The relative order of R_g displays a trend (G51D-G51D < WT-WT < A53T-A53T < A30P-A30P < H50Q-H50Q < E46K-E46K) which is very similar to the global compactness of the homo-dimers. This consistency in the relative trend of compactness between the global and chain-wise compactness also appears in the relative compactness order of the WT chain of the hetero-dimer (WT-A30P ≈ WT-G51D ≈ WT-H50Q < WT-A53T < WT-

E46K, Figure 10B), though majorly for A53T and E46K mutant-based hetero-systems. However, the mutant chains (G51D < H50Q < E46K < A30P < A53T) fail to replicate the relative compactness trend of the hetero-dimers. Specifically, the E46K chain does not create the least compact ensemble. However, the relative compactness among the mutants that alter the charge density of the protein remains consistent. Moreover, A53T creates the least compact ensemble which mimics the global compactness of the WT-A53T hetero-dimer. So, overall, the global compactness of a dimeric state is dictated by the chain-specific compactness. The extent of contribution toward the hetero-dimeric compaction shits from the WT to the mutant chain, depending on the sequence identity of the mutant.

3.5. A consistent appearance of specific thermally labile regions in αS variants.

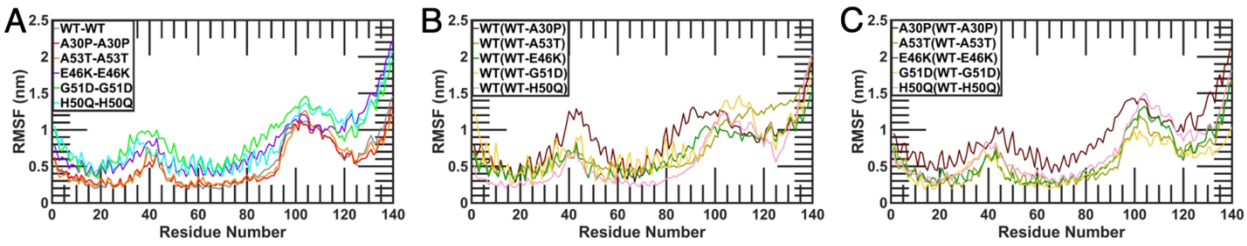


Figure 11. RMSF as a function of residue index – (A) averaged over the two chains of the homo-dimers, (B) WT chains of the hetero-dimers, and (C) mutant chains of the hetero-dimers.

Root mean square fluctuation (RMSF) analysis was performed to study the flexibility of the protein during the simulation. It helps in identifying regions with high or low movement, providing insights into structural dynamics.

For homo-dimers (Figure 11A), charged mutants (E46K-E46K, G51D-G51D, and H50Q-H50Q) show more fluctuations compared to the other variants. However, for hetero dimers, the WT chain of WT-A30P displays a greater extent of thermal fluctuations compared to the other WT chains (Figure 11B) and concomitantly, the mutant chain of WT-A30P (Figure 11C) shows more fluctuations compared to other variants.

Interestingly, irrespective of the nature of the dimeric species or the protein variant being considered (WT or mutant), two specific regions of the protein consistently display a higher degree of fluctuations compared to the rest of the sequence – residue 35-45 and residue 95-115.

3.6. Regions that participate in forming the dimer interface.

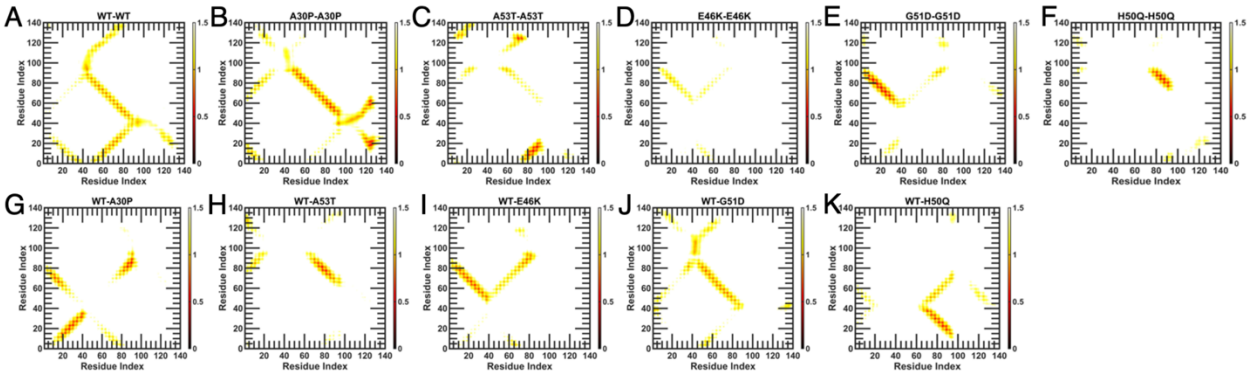


Figure 12. Inter-chain contact maps for – homo-dimers (panels A-F) where both x- and y-axes represent the two chains of the same variant and hetero-dimers (panels G-K) where the x- and y-axes represent the WT chain and the mutant chain, respectively.

Inter-chain backbone-based contact map for a protein dimer represents residue-residue interactions between two chains (similar or dissimilar). A contact was defined if the two backbone beads belonging to two different chains came within a distance cut-off of 0.6 nm.

For the homo-dimers (Figure 12, top panels), the mid-part of the protein chain appears as a major contact-forming segment that predominantly forms contact with the mid-part or N-terminal of the other chain, and they prefer an anti-parallel orientation (increasing residue number along x-direction with a concomitant decreasing residue number along the y-direction). However, the variability originates due to the extent of contact of formation: most prevalent for WT and A30P; moderately visible for A53T, G51D, and

H50Q; and sparsely evident for E46K. This observation justifies the variability in arrangements of two protein chains in the homo-dimers observed in the conformational ensemble analysis as the most dominant states are mostly overlapping for WT and A30P, partially overlapping for A53T, G51D, and H50Q, and barely overlapping for E46K.

In the case of the hetero-dimers (Figure 12, bottom panels), the prevalent contact formation is observed for A30P and G51D and moderate contact formation is observed for the other three cases which is justified because of the specific orientations of the protein chains as obtained from the conformational landscape analysis. Here too, the mid-part and the N-terminal segment of the protein chains participate in contact formation. The variability among the mutants appears due to the simultaneous and alternate participation of the above-mentioned segments.

3.7. Search for the dominant conformational ensembles of the aggregates.

PCA was performed to identify the most dominant conformations of the protein aggregates based on the two major collective motions. The emerged conformational landscapes can be broadly classified as the following: (1) narrow distribution with one major sharp peak (WT-WT and E46K-E46K) or a slight broadening of the single-peaked narrow distribution (A53T-A53T) and (2) a broad distribution with merged (A30P-A30P) or separated (G51D-G51D and H50Q-H50Q) multiple peaks. Homo dimers with a narrower distribution would hint toward a more definitive dominant conformational ensemble, while the others with a broader distribution would suggest a heterogeneous conformational ensemble, i.e., the presence of more than one dominant structure.

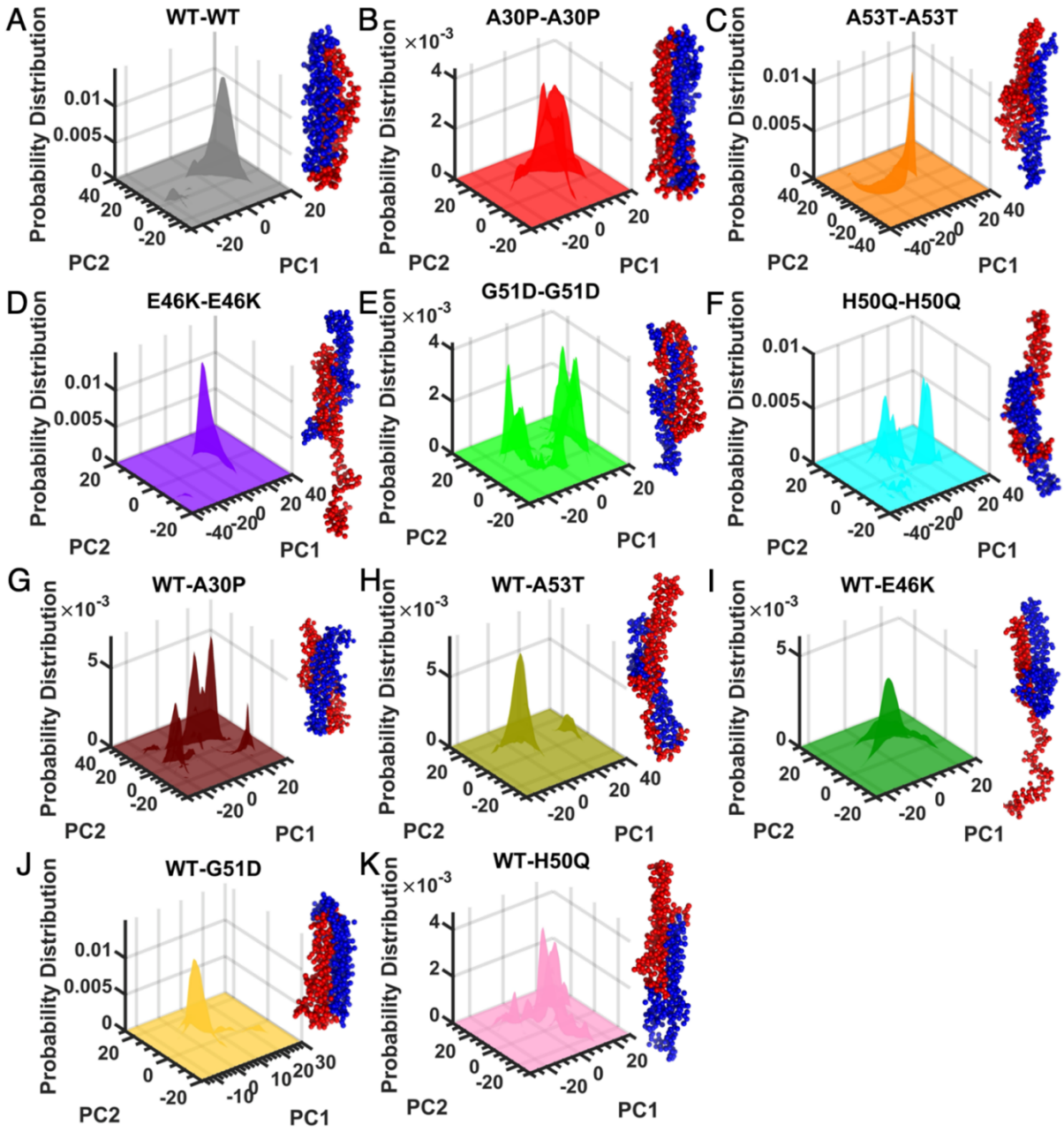


Figure 13. Conformational landscapes are generated using the first two principal components (PC1 and PC2).

The fundamental classifications of the conformational landscapes evolved for the homo-dimers reappear in the hetero-dimeric states, where the mutants chiefly retain the nature of the landscape – (1) a narrow distribution with a sharp peak (WT-A53T and WT-E46K) and (2) a broader distribution with multiple peaks (WT-A30P and WT-H50Q). Only the

G51D mutant switches from a broad to a narrower distribution while moving from homo- to hetero-dimer. This is possibly because of the presence of one WT chain as the homo-dimeric conformational landscape of the WT chain contains a narrower single-peaked distribution. So, in general, mutants retain their features irrespective of the nature of the dimeric state. And the G51D hetero-dimer can be seen as a perfect hybrid of WT and G51D homo-dimers.

3.8.The overall hydrogen bonding pattern of the dimeric state.

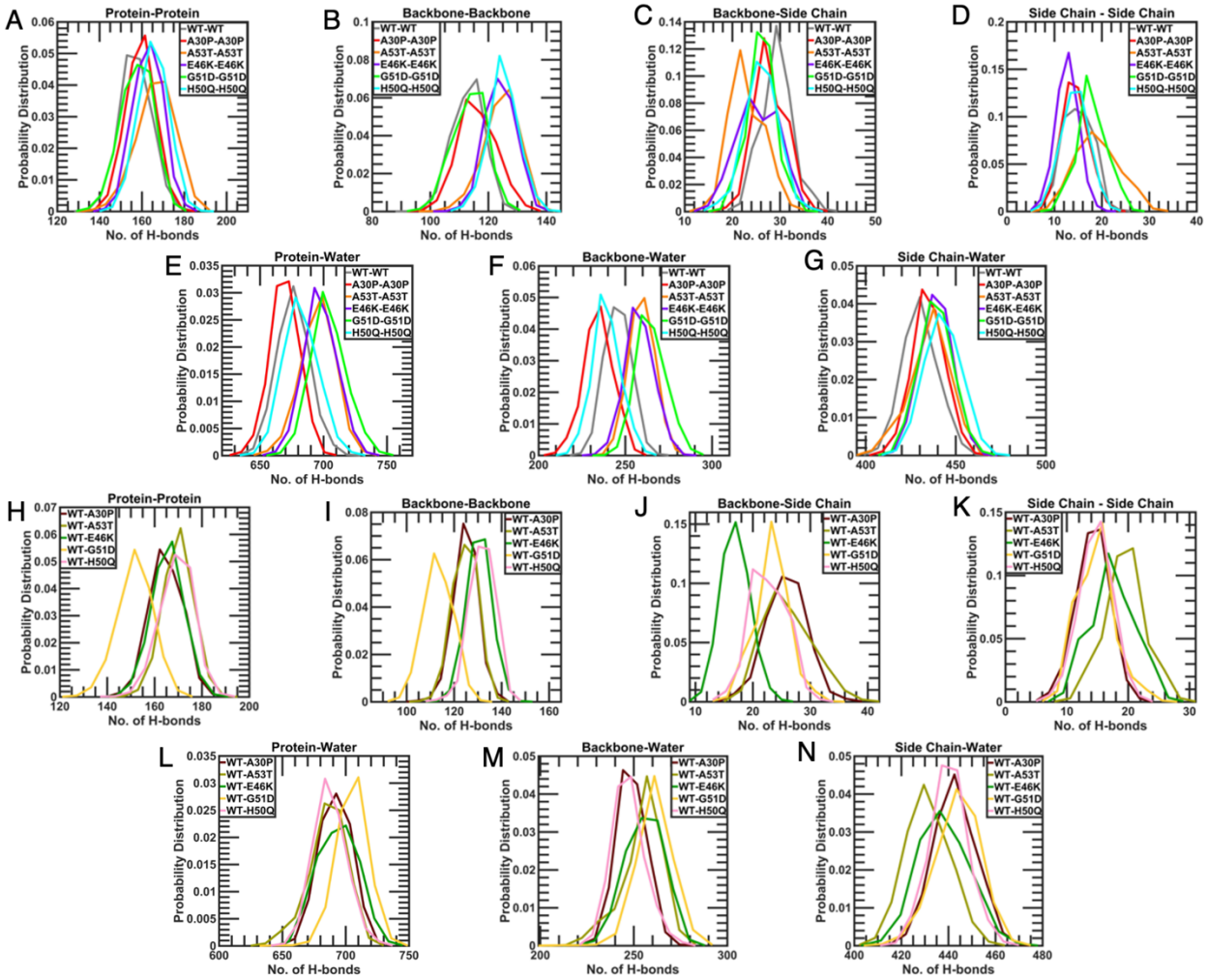


Figure 14. Normalized probability distributions of protein-protein and protein-water hydrogen bonds of the homo- and hetero-dimers.

The data presented in Results sections 3.8 to 3.10 were obtained from the atomistic simulations described in Methodology section 2.2.

We calculated the number of hydrogen bonds present within the dimeric species and between the dimer and surrounding water and deconvoluted them to the backbone- and side-chain based ones to identify the relative contributions (Figure 14). It appears that the major contribution to the overall hydrogen bonding network within the dimeric species originates from the backbone irrespective of the nature of the dimer (homo or hetero). Contrarily, the side chain appears as the dominant contributor to the protein-water hydrogen bonding pattern – be it homo- or hetero-dimers. So, there is a clear distinction between the roles of backbone and side chains in the hydrogen bonding network – the backbone holds the dimer together and side chains help to solvate the dimer. However, the variability among species majorly appears due to the varied extent of hydrogen bonds contributed by the backbone in both intra-dimer and dimer-water instances for both homo- and hetero-dimers.

3.9. The chain-wise intra- and inter-chain hydrogen bond network of the dimeric state.

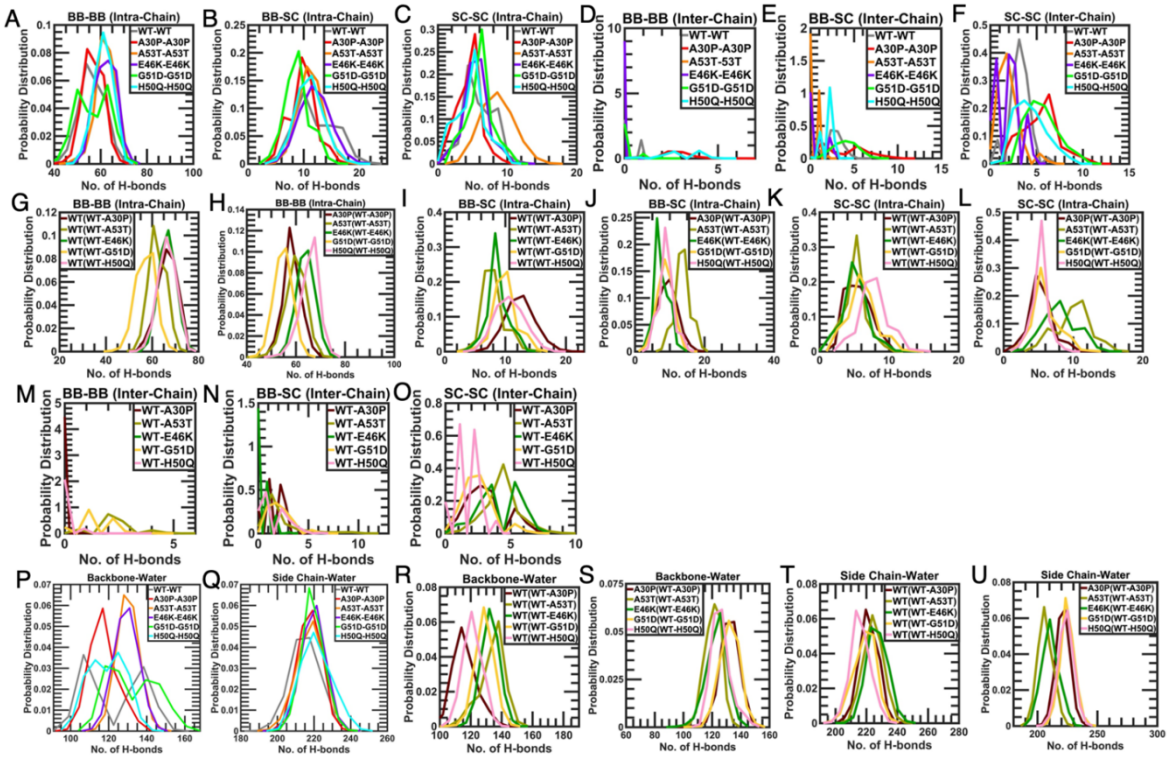


Figure 15. Normalized probability distributions of intra-chain, inter-chain and protein-water hydrogen bonds (per chain) were obtained for the homo- and hetero-dimers.

One of the crucial cohesive forces active in a protein is the hydrogen bond.[36] We calculated intra-chain and inter-chain hydrogen bonds. For both homo- and hetero-dimers, the contribution to the intra-chain hydrogen bonds gradually decreases from backbone-backbone to backbone-side chain to side chain-side chain. The intra-chain hydrogen bond formation is dominated by the backbone. In the inter-chain hydrogen bonds, the previously observed trend reverses. The contribution to the inter-chain hydrogen bonds decreases from side chain-side chain to backbone-side chain to backbone-backbone. In conclusion, intra-chain hydrogen bonds are dominated by backbone-backbone interaction while inter-chain hydrogen bonds are dominated by side chain-side chain interactions. The side chains majorly contribute to the formation of hydrogen bonds with the surrounding water molecules and this mimics the pattern obtained from the dimer-water hydrogen bond.

3.10. The role of electrostatic and van der Waals interactions in the dimeric structures of the αS variants.

We calculated intra-chain, inter-chain, and protein-water electrostatic[37] and van der Waals[38] Interactions (Figure 16). Intra-chain electrostatic interactions are stronger than inter-chain and protein-water electrostatic interactions. Similarly, intra-chain van der Waals interactions are stronger than inter-chain and protein-water van der Waals interactions. Overall, electrostatic interactions dominate over van der Waals by a large margin.

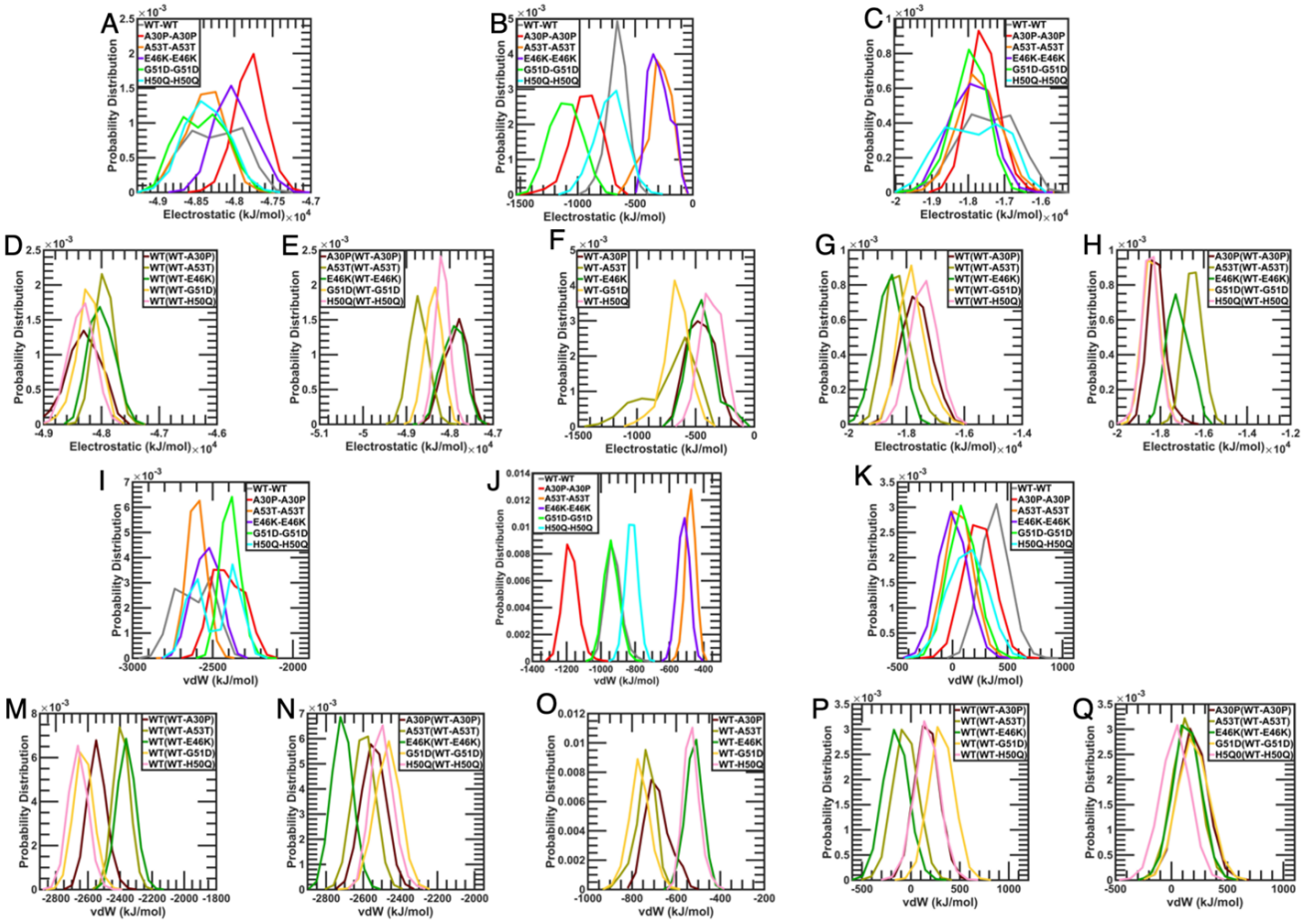


Figure 16. Normalized probability distributions of the electrostatic and van der Waals interactions operative in homo- (first and third rows) and hetero-dimers (second and fourth rows), respectively. (A) intra-chain, (B) inter-chain, and (C) protein-water electrostatics for homo-dimers; (D) intra-wildtype, (E) intra-mutant, (F) inter-chain, (G) WT-water, and (H) mutant-water electrostatics for hetero-dimers. The next two rows display the operative van der Waals interactions in the same order as shown for the electrostatics.

3.11. The inter-peptide binding free energy of the protofibrillar construct.

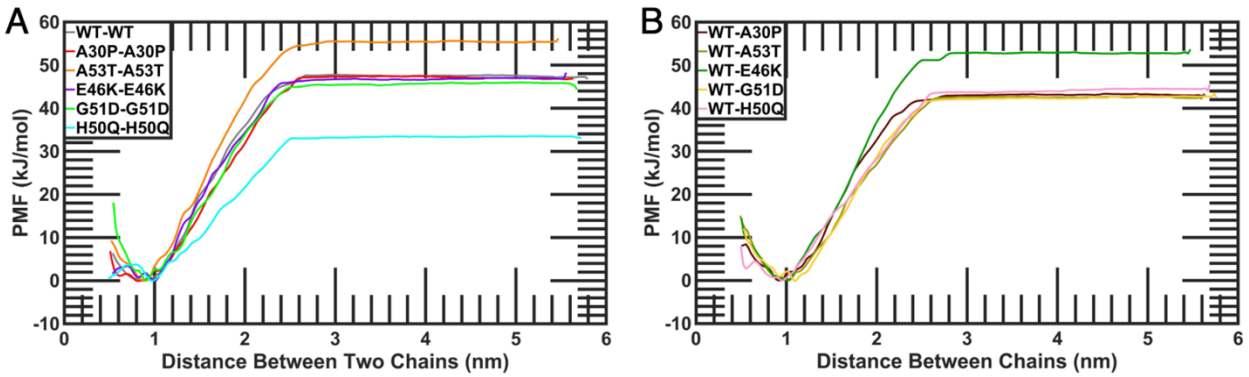


Figure 17. Potential of mean force profiles of – (A) homo-dimers and (B) hetero-dimers.

The PMF profiles generated using SMD and WHAM provide the free energy of binding between a couple of chains that form a specific protofibrillar construct, where the free energy change associated with the unbinding process of one protein chain from the other was estimated by taking the difference between the minimum value of the PMF profile (zero) and the asymptotic limit of the PMF profile.

In the homo-dimeric protofibrils, the rank ordering of binding affinities displays the following order: A53T (55.73 ± 2.19 kJ/mol) > WT (47.77 ± 1.96 kJ/mol) ≈ A30P (47.62 ± 2.24 kJ/mol) ≈ E46K (47.13 ± 2.43 kJ/mol) ≈ G51D (46.10 ± 2.11 kJ/mol) > H50Q (34.03 ± 1.82 kJ/mol). So majority of the homo-dimers display comparable protofilament binding affinities as observed for the WT, i.e., mutation, either neutral or charge-altering does not impact the protofibrillar stability. However, A53T and H50Q break the above-mentioned trend in two extreme ends – the former being a neutral mutation increases the affinity, while the latter being a charge-altering mutation decreases it, compared to the WT protofibril.

The hetero-dimers display a somewhat similar pattern (multiple variants with comparable protofilament binding affinities), but with a different relative order – E46K (53.37±2.22 kJ/mol) > H50Q (44.93±3.00 kJ/mol) ≈ A30P (43.53±2.17 kJ/mol) ≈ G51D (43.10±3.13

kJ/mol) ≈ A53T (43.06±1.99 kJ/mol). Here, the E46K mutant stands out as an outlier displaying the maximum extent of binding affinity.

A few conclusions can be made by comparing all the binding free energies:

1. Except for the E46K mutant, all the other mutants form a slightly less stable hetero-protofibril compared to the homo-protofibril of the WT variant.
2. The protofilament binding affinities are comparable between homo- and hetero-protofibrillar states of A30P and G51D along with the WT homo-protofibril.
3. The mutants with very high binding affinities change their identities from neutral (A53T) to charge-altering (E46K) as the protofibril changes its identity from homo to hetero.
4. No hetero-protofibrils show extremely decreased binding affinities as does the H50Q homo-protofibril.

4: Discussion

In this work, the entire aggregation landscape of αS was analyzed by characterizing its kinetic and thermodynamic aspects. The comparative analysis of the WT αS features compared to its five prevalent deleterious mutants provides a detailed understanding of the differences between protein variants.

αS mutants that alter charge exhibit distinct trends in their relative rates of self-association and association with the WT chain. Among the neutral mutations, A30P dimerizes (either with itself or with WT αS) at a faster rate than A53T (Figure 7). The homo- and hetero-dimeric structures originating from such associations always display lesser compactness and larger solvent exposure for the mutant-containing ones compared to the WT homo-dimer (Figures 8-9). Moreover, the ability to generate compact/exposed ensembles decreases/increases from G51D to H50Q to E46K (Figures 8-9). Additionally, the global compactness of a dimeric state is dictated by the chain-specific compactness (Figure 10).

All the dimeric species display two common highly fluctuating regions for their constituting chains (Figure 11). Interestingly, the inter-chain contacts are formed majorly by the N-terminal and mid-part of the αS chain (both WT and mutant, Figure 12), both of which belong to the least fluctuating regions of the protein chain. Despite the commonality in the fluctuation pattern, the dimeric conformational landscapes display a considerable diversity among variants. Still, except for G51D, the conformational landscape features remain consistent in homo- and hetero-dimers for a specific set of variants (narrow for E46K and A53T and broad for A30P and H50Q (Figure 13). The origin of multiple common and consistent features observed for the homo- and hetero-dimers of six considerably different variants could be attributed to the consistent nature of specific operative cohesive forces. In general, the backbone contributes to the intra-chain hydrogen bond framework stability and the side chains dictate the inter-chain and protein-water interactions (Figures 14-15), possibly being more exposed. Whereas, the electrostatic forces specifically stabilize the intra-chain interactions over the inter-chain and protein-

water talks (Figure 16). The dispersion interactions also contribute to all the possible protein-protein and protein-water cross-talks, but their contribution is always weaker compared to the electrostatic forces.

The above analyses talk about the dimeric states that appear during the course of aggregation. However, the dimeric constructs are not the final fibrillar states. To check whether there is any correlation between the initially associated dimeric structures and the fibrillar states, we shifted our focus to the protofibrillar constructs made up of two protein chains with the same or different sequence identities. Surprisingly, the distinguishing features of the charge-altering mutants do not appear prominently in the protofilament binding affinity estimations (Figure 17) where a neutral mutant (A53T) appears as the species with the highest binding affinity and one of the charge-altering mutants practically shows the lowest protofilament binding affinity (H50Q). Moreover, the charge-altering mutants switch their ability to form a stable protofibril from the highest among hetero-dimers (E46K, gain in positive charge) to the lowest among homo-dimers (H50Q, loss of positive charge). In addition, the rate of association also does not correlate with the protofilament binding affinity pattern, suggesting that faster aggregation does not necessarily lead to stable aggregates.

5: Conclusions

In this work, we have tried to answer a few relevant questions of αS aggregation. We have used extensive CG and atomistic simulations along with conventional and enhanced sampling techniques to answer the kinetics and thermodynamics of the aggregation process to uncover the complex mechanism. We found that binding free energy does not directly correlate with self-association, suggesting that other factors may be influencing the process. To gain further insight, I will study the literature to identify potential influencing factors and alternative mechanisms.

Still, some limitations exist in the present work: approximations of the CG model, the use of only two chains for aggregation studies, the consideration of only two chains of the fibril, and the inability to include the unstructured segments of the fibril. The reason behind all these limitations lies in the present hurdle of handling very large systems with adequate computing facilities within a realistic timescale to achieve viable quantitative measurements with trustable statistics. Future work would be directed toward the identification of the connection between disease phenotype and physical properties of the protein.

References

(1) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry*, 2006, *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

(2) Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (6), 384–396. https://doi.org/10.1038/nrm3810.

(3) Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annual Review of Physical Chemistry*, 2015, *66*, 643–666. https://doi.org/10.1146/annurev-physchem-040513-103738.

(4) Thirumalai, D.; Klimov, D.; Dima, R. Emerging Ideas on the Molecular Basis of Protein and Peptide Aggregation. *Curr. Opin. Struct. Biol.* **2003**, *13* (2), 146–159. https://doi.org/10.1016/S0959-440X(03)00032-0.

(5) Straub, J. E.; Thirumalai, D. Toward a Molecular Theory of Early and Late Events in Monomer to Amyloid Fibril Formation. *Annual Review of Physical Chemistry*, 2011, *62*, 437–463. https://doi.org/10.1146/annurev-physchem-032210-103526.

(6) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326* (5959), 1533–1537. https://doi.org/10.1126/science.1178250.

(7) Michaels, T. C. T.; Knowles, T. P. J. Role of Filament Annealing in the Kinetics and Thermodynamics of Nucleated Polymerization. *J. Chem. Phys.* **2014**, *140* (21), 214904. https://doi.org/10.1063/1.4880121.

(8) Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *Protein Seq. Struct. Warp Speed Alphafold Impacts Biol.* **2021**, *433* (20), 167182. https://doi.org/10.1016/j.jmb.2021.167182.

(9) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4* (1), 10–19. https://doi.org/10.1038/nsb0197-10.

(10) James Parkinson. *Wikipedia*; 2025.

(11) *Parkinson disease*. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease (accessed 2025-03-18).

(12) Kouli, A.; Torsney, K. M.; Kuan, W.-L. Parkinson's Disease: Etiology, Neuropathology, and Pathogenesis. In *Parkinson's Disease: Pathogenesis and Clinical Aspects*; Stoker, T. B., Greenland, J. C., Eds.; Codon Publications: Brisbane (AU), 2018.

(13) Fusco, G.; De Simone, A.; Gopinath, T.; Vostrikov, V.; Vendruscolo, M.; Dobson, C. M.; Veglia, G. Direct Observation of the Three Regions in α-Synuclein That

Determine Its Membrane-Bound Behaviour. *Nat. Commun.* **2014**, *5* (1), 3827. https://doi.org/10.1038/ncomms4827.
(14) Vamvaca, K.; Volles, M. J.; Lansbury, P. T. The First N-Terminal Amino Acids of α-Synuclein Are Essential for α-Helical Structure Formation In Vitro and Membrane Binding in Yeast. *J. Mol. Biol.* **2009**, *389* (2), 413–424. https://doi.org/10.1016/j.jmb.2009.03.021.
(15) Nielsen, M. S.; Vorum, H.; Lindersson, E.; Jensen, P. H. Ca2+ Binding to α-Synuclein Regulates Ligand Binding and Oligomerization*. *J. Biol. Chem.* **2001**, *276* (25), 22680–22684. https://doi.org/10.1074/jbc.M101181200.
(16) Ruggeri, F. S.; Flagmeier, P.; Kumita, J. R.; Meisl, G.; Chirgadze, D. Y.; Bongiovanni, M. N.; Knowles, T. P. J.; Dobson, C. M. The Influence of Pathogenic Mutations in α-Synuclein on Biophysical and Structural Characteristics of Amyloid Fibrils. *ACS Nano* **2020**, *14* (5), 5213–5222. https://doi.org/10.1021/acsnano.9b09676.
(17) Meade, R. M.; Fairlie, D. P.; Mason, J. M. Alpha-Synuclein Structure and Parkinson's Disease – Lessons and Emerging Principles. *Mol. Neurodegener.* **2019**, *14* (1), 29. https://doi.org/10.1186/s13024-019-0329-1.
(18) Bauer, P.; Hess, B.; Lindahl, E. GROMACS 2022.1 Manual. **2022**. https://doi.org/10.5281/zenodo.6451567.
(19) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9* (1), 687–697. https://doi.org/10.1021/ct300646g.
(20) The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
(21) Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L. Structure and Dynamics of Micelle-Bound Human α-Synuclein*. *J. Biol. Chem.* **2005**, *280* (10), 9595–9603. https://doi.org/10.1074/jbc.M411805200.
(22) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.
(23) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLOS Comput. Biol.* **2010**, *6* (6), e1000810. https://doi.org/10.1371/journal.pcbi.1000810.
(24) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.
(25) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.
(26) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(27) Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14* (2), 325–332. https://doi.org/10.1021/ma50003a019.

(28) Bahar, I.; Erman, B.; Haliloglu, T.; Jernigan, R. L. Efficient Characterization of Collective Motions and Interresidue Correlations in Proteins by Low-Resolution Simulations. *Biochemistry* **1997**, *36* (44), 13512–13523. https://doi.org/10.1021/bi971611f.
(29) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10* (2), 676–690. https://doi.org/10.1021/ct400617g.
(30) Huang, J.; MacKerell Jr, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.
(31) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119* (16), 5113–5123. https://doi.org/10.1021/jp508971m.
(32) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(33) Tammara, V.; Doke, A. A.; Jha, S. K.; Das, A. Deciphering the Monomeric and Dimeric Conformational Landscapes of the Full-Length TDP-43 and the Impact of the C-Terminal Domain. *ACS Chem. Neurosci.* **2024**, *15* (23), 4305–4321. https://doi.org/10.1021/acschemneuro.4c00557.
(34) Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M. Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-Synuclein. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 409–415. https://doi.org/10.1038/nsmb.3194.
(35) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/jcc.540130812.
(36) Pace, C. N.; Fu, H.; Lee Fryar, K.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; Sevcik, J.; Urbanikova, L.; Myers, J. K.; Takano, K.; Hebert, E. J.; Shirley, B. A.; Grimsley, G. R. Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.* **2014**, *23* (5), 652–661. https://doi.org/10.1002/pro.2449.
(37) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118* (4), 1691–1741. https://doi.org/10.1021/acs.chemrev.7b00305.
(38) Roth, C. M.; Neal, B. L.; Lenhoff, A. M. Van Der Waals Interactions Involving Proteins. *Biophys. J.* **1996**, *70* (2), 977–987. https://doi.org/10.1016/S0006-3495(96)79641-8.