Check for updates

OPEN

Long-range Regulation of Partially Folded Amyloidogenic Peptides

Shayon Bhattacharya, Liang Xu & Damien Thompson ✉

Neurodegeneration involves abnormal aggregation of intrinsically disordered amyloidogenic peptides (IDPs), usually mediated by hydrophobic protein-protein interactions. There is mounting evidence that formation of α-helical intermediates is an early event during self-assembly of amyloid-β42 (Aβ42) and α-synuclein (αS) IDPs in Alzheimer's and Parkinson's disease pathogenesis, respectively. However, the driving force behind on-pathway molecular assembly of partially folded helical monomers into helical oligomers assembly remains unknown. Here, we employ extensive molecular dynamics simulations to sample the helical conformational sub-spaces of monomeric peptides of both Aβ42 and αS. Our computed free energies, population shifts, and dynamic cross-correlation network analyses reveal a common feature of long-range intra-peptide modulation of partial helical folds of the amyloidogenic central hydrophobic domains *via* concerted coupling with their charged terminal tails (N-terminus of Aβ42 and C-terminus of αS). The absence of such inter-domain fluctuations in both fully helical and completely unfolded (disordered) states suggests that long-range coupling regulates the dynamicity of partially folded helices, in both Aβ42 and αS peptides. The inter-domain coupling suggests a form of intra-molecular allosteric regulation of the aggregation trigger in partially folded helical monomers. This approach could be applied to study the broad range of amyloidogenic peptides, which could provide a new path to curbing pathogenic aggregation of partially folded conformers into oligomers, by inhibition of sites far from the hydrophobic core.

Protein conformational disorders including Alzheimer's (AD) and Parkinson's disease (PD) present the hallmark features of misfolding, self-assembly, and accumulation of monomeric precursor peptides known as intrinsically disordered proteins (IDPs) or amyloidogenic peptides such as amyloid-β, Aβ (implicated in AD) and α-synuclein, αS (in PD)[1]. Low molecular weight soluble oligomers are thought to be neurotoxic in the self-assembly pathway[2], and a number of experiments suggest their formation may proceed *via* α-helical oligomeric intermediates[3–5]. Partially folded helical conformers (monomers) have been reported to be on-pathway to fibril formation[6,7]. Consequently, the tendency to form helical oligomers (through helix-helix associations) may be imprinted within the minor populations of aggregation-prone partially folded helical monomers[3,6–8] that both Aβ and αS display in membrane-free aqueous solution[9,10], and which arise *via* conformational exchanges between non-aggregating helically folded and completely unfolded states. Formation of partially folded intermediates has also long been correlated with aggregation and is identified as a key step in fibrillation[11,12]. A powerful mathematical model based on kinetic state transitions for Aβ structural evolution during fibrillogenesis has been able to reproduce the experimentally observed helical intermediates[13].

The energy landscapes of globular protein folding and binding have been well documented[14]. In addition, aggregation has been linked to an intricate balance between folding and binding[15]. Especially, the ruggedness and 'frustrations' of an IDP energy landscape have to a large extent been attributed to fluctuating allosteric transitions[16]. Allostery[17], or regulation at a distance, plays a central role in mediating protein recognition, signal transduction and promiscuous interactions of IDPs[18,19]. Nuclear magnetic resonance (NMR) relaxation measurements demonstrate correlated collective internal dynamics within IDP ensembles[20–23], where perturbation (such as folding or ligand binding) within one domain may result in a regulatory effect on a distant, but energetically coupled, domain[24,25]. As disorder is a functional advantage for regulation at long-range[26], conformational shifts may accommodate molecular recognition of IDPs through hydrophobic interactions[27] and promote folding-induced binding or conformational selections[28] during self-assembly in many neurodegenerative diseases[29,30]. Thus, the perturbations in the disordered regions of the helical monomers of Aβ42 and αS may also regulate the tendency to bind other helical monomers and form helical oligomers.

Department of Physics, Bernal Institute, University of Limerick, Limerick, V94 T9PX, Ireland. ✉e-mail: damien.thompson@ul.ie
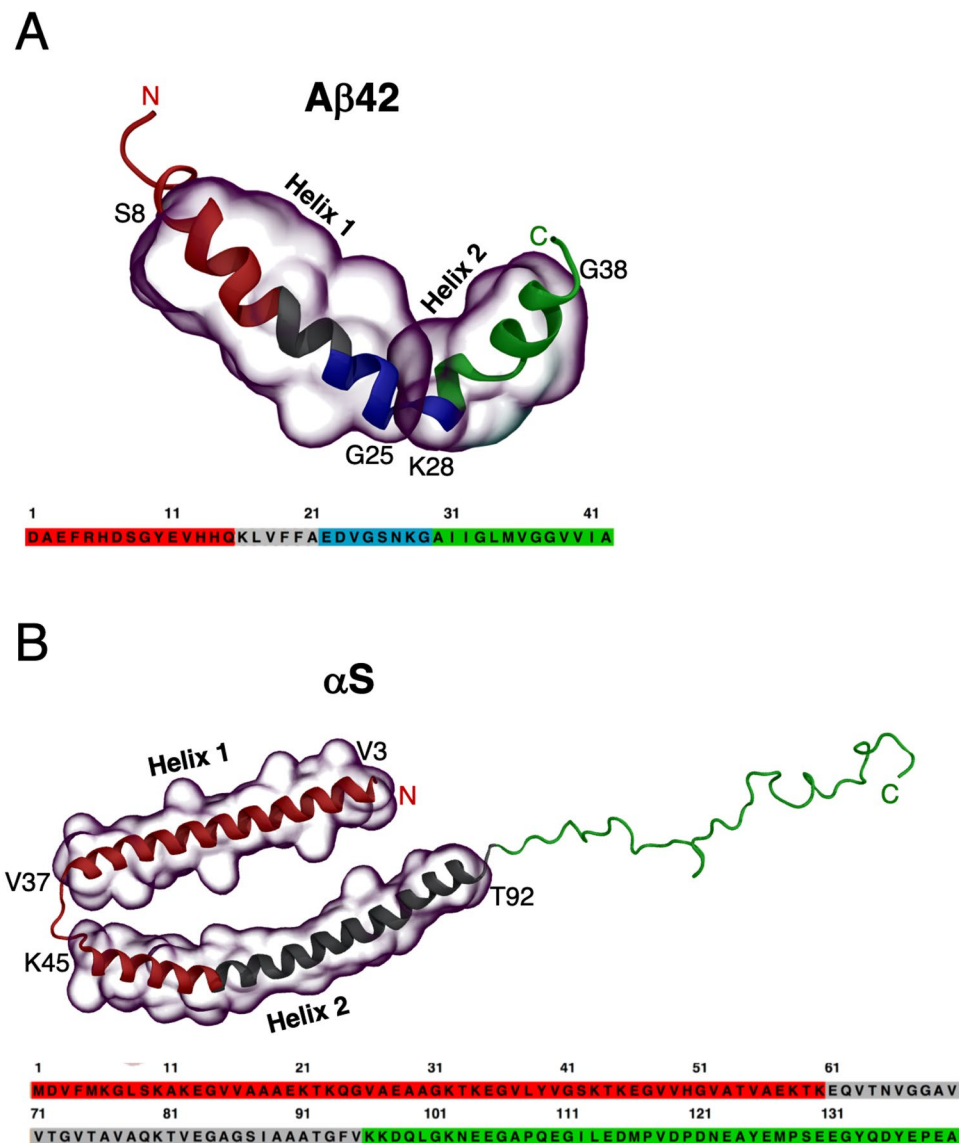
Figure 1. Helix-turn-helix molecular models solved by experimental NMR structures of **(A)** Aβ42 in 20% water/80% deuterated-hexafluoroisopropanol (PDB 1IYT[32]) and **(B)** micelle-bound αS in aqueous solution with sodium dodecyl sulfate (PDB 1XQ8[33]). The amino acid sequences are shown underneath each structure. The regions encompassing Helix 1 and Helix 2 are overlaid with transparent surfaces in both Aβ42 (S8–G25 and K28–G38) and αS (V3–V37 and K45–T92). The colour scheme for the subdomains in the structures and their sequences are, red: N-terminal region, green: C-terminal region, gray: central hydrophobic region (central hydrophobic cluster in Aβ42 or non-amyloid-β component in αS), and blue: turn region.

It is extremely challenging to identify and measure these subtle variations in the degree of helicity using spectroscopy measurements, due to the conformational diversity in the helical subspaces[8]. Identification of the difference between the aggregation propensities of helical intermediates is usually impossible due to the very short lifetimes of monomer partially folded states. Thus, little is known about what causes partially folded helical Aβ42 and αS to be so aggregation-prone, compared to the fully folded helical or completely unfolded conformations[6].

We have very recently shown from extensive molecular dynamics (MD) simulations that terminal groups in both Aβ42 and αS frequently make direct steric contacts with the central hydrophobic domains that help stabilize folded conformations[31]. Here, we look beyond just direct contacts in folded states and comprehensively map internal long-range communications across helically folded, partially folded and unfolded states of Aβ42 and αS peptides by computing dynamic cross-correlation networks from extensive microseconds-scale MD simulation data. We identify functionally relevant sub-domains involving the charged termini ends that remain $10-20$ Å from the hydrophobic domains yet play a critical role in modulating their helical folds, through long-range coupling. Starting from solved helix-turn-helix structures of both Aβ42 and αS (see Fig. 1)[32,33], we employ a range of physical models with diverse helix-coil transition propensities to generate a broad distribution of conformational

states (see Methods). No single physical model has been able to produce consistent results for both the folded and the disordered states of IDPs (or the transitions in-between)[30,34–37], and so we take advantage of model inhomogeneity to sample extensively in the rich IDP conformational sub-spaces. In detail, the traditional CHARMM36[38] protein model with CHARMM-modified TIP3P water[39] (TIP3P) and CHARMM36 with TIP4P[40] parametrized for globular proteins gave as expected predominantly fully-helical folded states while the more IDP-specific CHARMM22*[41] with TIP4P-D[35] and Amber ff03ws[34] with the scaled TIP4P/2005[34,42], gave partial helical folded states together with some completely unfolded (disordered) states (see Methods). We focus here on characterising the atomic-scale interactions that stabilise helix formation in the smallest, monomeric species, as the simplest starting point in their oligomerisation *via* conformational selection to helical oligomers. The specific scientific question is, what causes partially folded Aβ42 and αS to be so aggregation-prone, compared to the fully folded helical and unfolded conformations? In the present study, we monitor long-range coupling between domains that may influence and possibly direct, helix-helix associations of amyloidogenic peptides.

Results

Free energy surfaces identify and characterize partially folded helical states.
Initial classification of the helical subspaces was carried out by evaluating the temporal evolution of the model helical propensities (see Fig. S1A under Supplementary Information). We observe that the computed rates of helix unfolding (and thus the complete unfolding time) differ according to the choice of force field/water model, such that a fast unfolding identifies the short-lived partially folded helical states and a slow unfolding sample long-lived partially folded helical state. This enabled us to generate an extensive conformational ensemble of helical states for both Aβ42 and αS. We compute the helix (α-helix + π-helix + 3_{10}-helix) unfolding patterns and the corresponding population distributions of Aβ42 and αS from 2D free energy maps using the residual helical percentages and radius of gyration (R_g) as order parameters. The convergence of the equilibrium MD simulations (EMD) was assessed by the time cumulative average helical content (Fig. S1B). We also performed Hamiltonian replica exchange simulations with solute scaling (H-REST; see Methods for details) to gauge the extensiveness of our helical conformational sampling with EMD. Fig. S1C shows significant overlaps of the free energy surfaces (tested for αS with CHARMM22*/TIP4P-D) of partially folded states with EMD and refolded helices from completely disordered states with H-REST confirming the robustness of our EMD simulation. The complete helically folded states in our EMD simulations record a high percentage of helices. While such high helical content events might be extremely rare under physiological conditions in water, we reiterate that the idea here is to use molecular models to sample extensively in this regime to capture differences in helical stability that could be highly consequential for pathology[5,6,43] but remain imperceptible at experimentally accessible timescales.

Aβ42 displays dense minima for the helically folded states (Fig. 2A) in the left high-helicity/small-radius energy basin. A sharp energy barrier separates conformations with less than ~50% helical content sampled at low R_g, which reflects the difficulty in initiating helical unfolding from a predominantly compact state. The MD trajectories show that a helical turn spanning residues G9–V12 in the N-terminal abruptly unfolds causing a sharp drop in overall helicity. This accounts for the energy barrier of ~10 kJ/mol (whitespace) at 50% helicity in the maps, which agrees well with the previous estimate of a free energy barrier of 19 ± 10 kJ/mol from simulated unfolding of two α-helical turns of surfactant-associated polypeptide C[44]. The regions in the Aβ42 peptide can be designated as the N-terminal region (NTR, spanning residues 1–16), central hydrophobic cluster (CHC, 17–21), the turn region (22–29)[45], and the C-terminal region (CTR, 30–42) (refer to Fig. 1A). The secondary structure (Fig. S2A) shows helix-rich (>80%) distal NTR (residues Y10–K16), full CHC, proximal CTR (A30–G38), and proximal turn region (E22–S26) for the folded states, with no instance of extended structures, while the remaining residues form random coils and turns.

The partially folded helical states populate the minima of low-helicity/large-R_g basins (Fig. 2B) as Aβ42 forms a more open structure. In contrast to the folded states, no such sharp energy barrier is observed for the partially folded states. The significantly lowered helix population of partial folds in the distal NTR, CHC and CTR is compensated by turn regions, demonstrating the conformational plasticity (Fig. S2A). More extended structures (β-strand + β-bridge) are extremely rare and transient, with helices being the predominant structural motif.

For αS, stable compact folded helices and elongated helices (Fig. 2C) are both observed covering narrow ranges of small and large R_g, respectively, any with ~35–60% helicity. Significant barriers identify distinct helical breaks at Gly residues (G31, G36, G67, G68, G73, G84 and G86) leading to transition from a higher helicity (smaller R_g) to a lower helicity (larger R_g). The specific regions in the peptide can be identified as the NTR (spanning residues 1–60), the non-amyloid-β component (NAC, 61–95), and the CTR (96–140) (refer to Fig. 1B). Persistent helices in the NTR and distal NAC (V70–V95) can be seen in the folded states of αS (Fig. S2B), with negligible extended structures, except A90 and T92 in the CTR, which forms transient β-strands. The CTR remains mostly random coil, with some small regions forming turn motifs (D98, Q99, E114 and D115), alongside small parts of NTR (K23 and Q24) and NAC (V63–V66, G84, A85, A90 and A91).

The dense population of the partially folded states (Fig. 2D) sit mainly in the lower-left quadrant of the free energy landscape. The R_g is densely populated between 15–35 Å, which identifies a heterogeneous population, but is at the same time fairly compact given the length of αS. Sparsely populated large R_g basins demonstrate that formation of extended helical states might be a minor event during complete unfolding, and not necessarily a requirement for unfolding helices as was seen for Aβ42. The helical population in the partially folded states is dominant in the NAC region (Fig. S2B). Overall, the unfolding pattern of αS involves a more compact helix unravelling than Aβ42, with low occurrence of helical refolding. Conformational heterogeneity is most prevalent in the partially folded states for both peptides.

The fold propensities of the peptides can also be estimated from their effective dimensions D, which is a measure of the distribution of residues over different length scales and can be used to complement the R_g. The effective
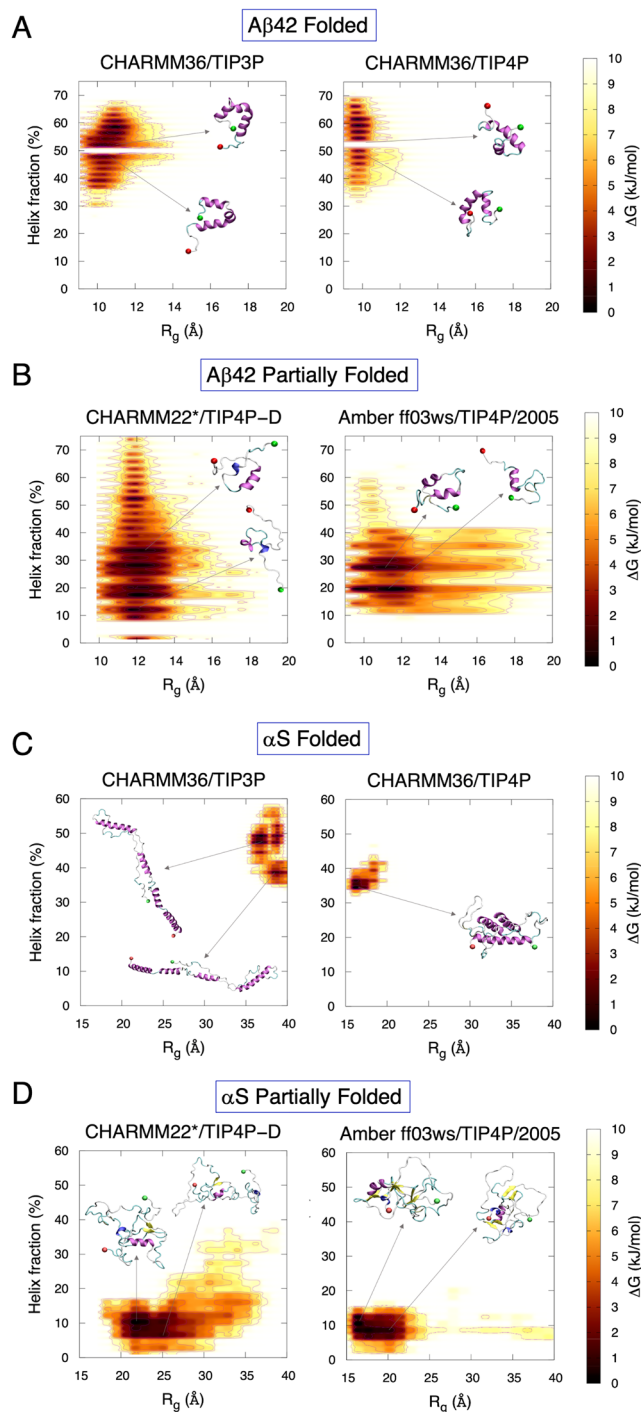
Figure 2. Free energy surfaces of fraction of helix content (%) and radius of gyration (R$_g$) for folded (**A,C**) and partially folded states (**B,D**) of Aβ42 and αS. The representative conformations corresponding to energy minima basins are marked. The N- and the C-terminus are represented as red and green spheres, respectively.

dimensions can be obtained by power-law scaling of the structure factor *S(q)* as a function of the wave vector *q*, given by,

$$S(q) = \langle \frac{1}{N} | \sum_{j=1}^{N} e^{-I\vec{q} \cdot r_j} |^2 \rangle_{|\vec{q}|}$$

(1)

where $|q| = 2\pi/\lambda$ is the wave vector of wavelength λ, and r_j is the residue position. We calculated the variation of the structure factor with wave vector for different fold propensities (Fig. S3), employing the smooth Particle Mesh
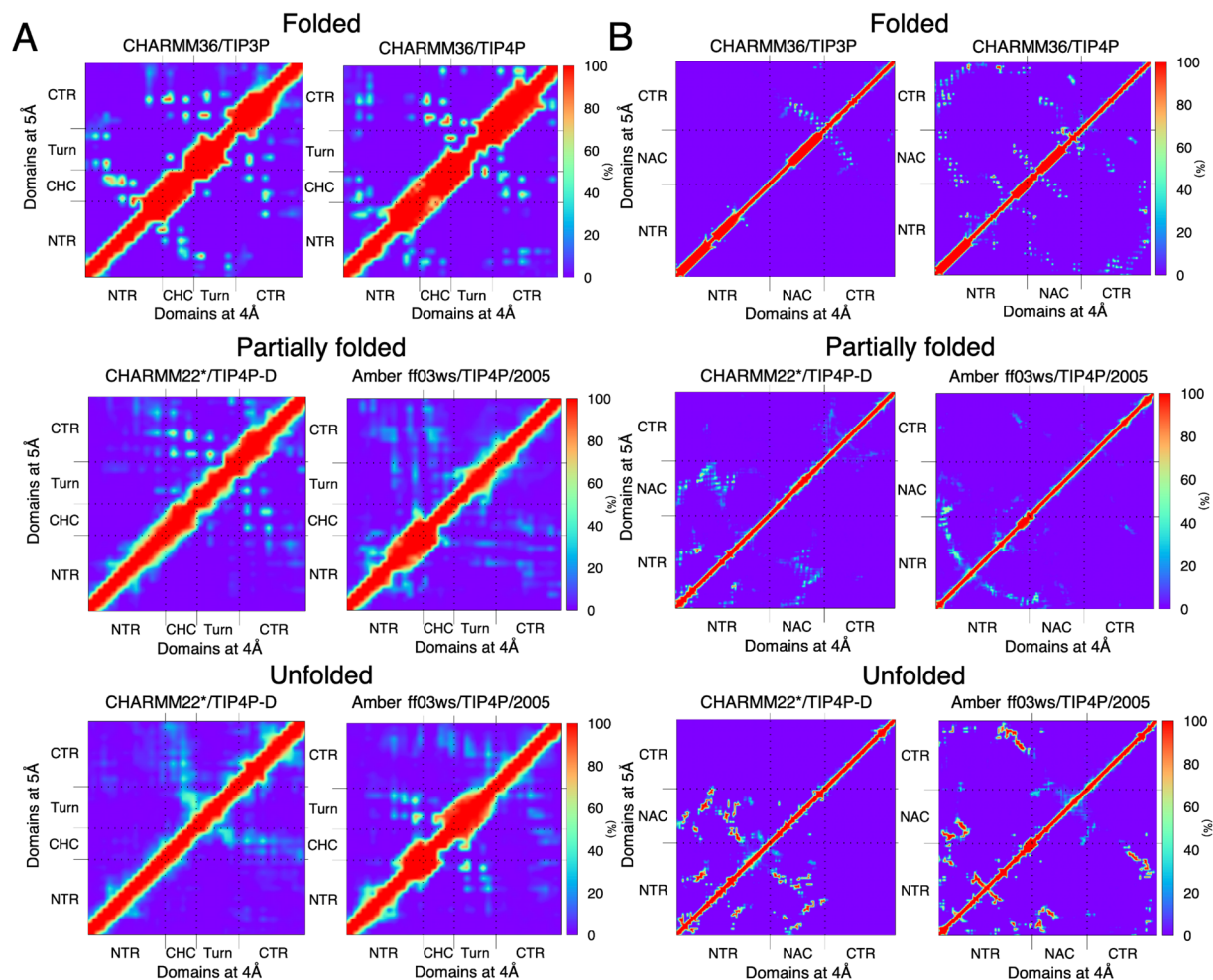
Figure 3. Computed contact maps (frequency, %) between residues in helically folded, partially folded helical, and unfolded states for **(A)** Aβ42, and **(B)** αS. A contact occurs if the minimum average distance between heavy atoms of each residue is within 5 Å (upper left triangle). Contact analyses performed with a 4 Å cut-off (lower right triangle) gave near-identical results. The respective domains in both the peptides are annotated in the map axes.

Ewald summation[46] method for obtaining high resolution $S(q)$ values. Several previous studies have estimated the effective dimension of proteins by employing coarse-grain Monte Carlo (MC) simulations to predict changes in structure factor over a range of temperatures, where the wavelength λ was given in MC lattice units[47–49]. Here, we applied the *B*-spline interpolation-based approach for the paired interatomic potential lattice sums generated in our atomistic MD simulations. The scaling exponent of $S(q)$ with q from $S(q) \propto q^{-1/\gamma}$ can be used to estimate the effective dimension D, by fitting the $S(q)$ at different λ values within the R_g, such that $D \approx 1/\gamma$. Inset in Fig. S3A shows a rise in the average R_g values as Aβ42 unfolds. Estimates of $-1/\gamma$ (annotated alongside plots) reveals $D \approx 1.2$ for the fully folded state (CHARMM36/TIP3P) and $D \approx 1$ for the fully unfolded state (CHARMM22*/TIP4P-D), which compares well with the R_g.

For typical large globular proteins, a value of $D \geq 3$ is estimated to be a compact structure, whereas $D \leq 2$ is usually a random coil structure[47–49]. The effective dimension in both cases for Aβ42 is closer to 2 (than 3) suggesting a random coil-like behaviour which may indicate that Aβ42 is too small (only 42 residues) for accurate estimation of residue distributions with various length scales. By contrast, the $S(q)$ *vs.* q plots for the much larger αS (Fig. S3B) shows that the fully folded state (CHARMM36/TIP4P) is extremely compact and globular in nature ($D \approx 8$), while the fully unfolded state (Amber 03 ws/TIP4P/2005) is more open, approaching random coil ($D \approx 2.7$).

Charged terminal tails do not interact with the hydrophobic hotspots of aggregation in the partially folded helical states. We have previously demonstrated that the stability of helices in both Aβ42 and αS are governed predominantly by terminal-mediated short-range contacts with the central hydrophobic domains[31]. Here we check if short-range interactions persist also in partially folded helical and unfolded states within a distance of 5 Å. This cut-off distance was chosen in order to accommodate most short-range non-covalent side chain interactions including hydrogen bonds[50]. This cut-off distance was chosen in order to accommodate most short-range non-covalent side chain interactions including hydrogen bonds[50] and van der
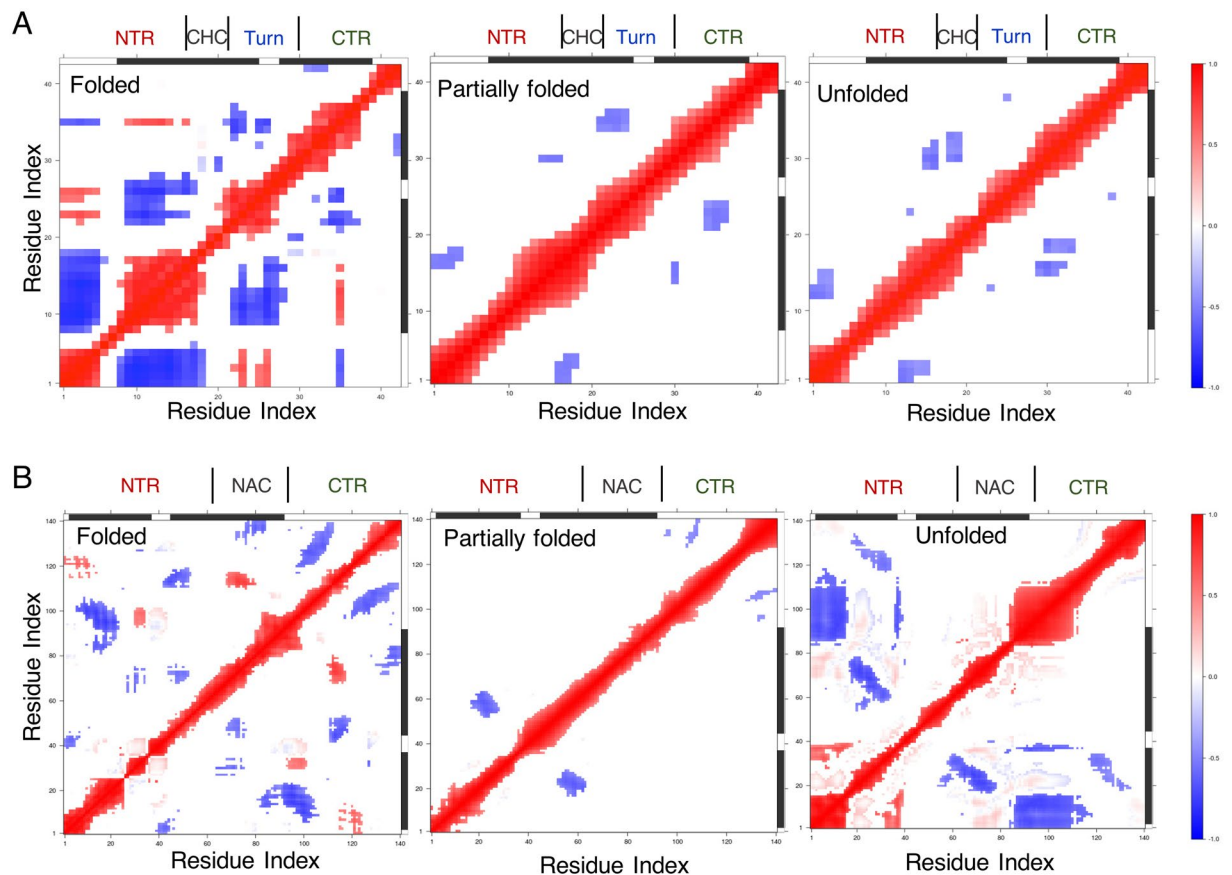
Figure 4. Consensus dynamic cross-correlation maps (DCCMs) with a correlation filter of 0.4 and above showing positively (red), negatively (blue) correlated and uncorrelated (white) fluctuating regions for the helically folded, partially folded and unfolded states of (**A**) Aβ42 and (**B**) αS.

Waals forces[51]. Additionally, protein network properties can be robustly evaluated at an optimal intramolecular distance of 5 Å, capturing all important short-range interactions irrespective of changes in force field and/or protein models[52]. We also computed contact frequencies within a tighter cut-off of 4 Å, which may potentially distinguish short-range hydrogen bonds from other non-covalent interactions. Contact probabilities between heavy atoms at a cut-off distance of either 5 or 4 Å across all domain pairs show anti-diagonal contacts (left and right triangles, respectively) for Aβ42 involving the CHC and CTR, corresponding to a helical hairpin-like conformation (Fig. 3A). In addition, frequently sampled contacts (>60%) between the distal NTR (R5–V12) and the CHC are found in the folded states. These contacts are extremely sparse or completely missing in the partially folded helical and unfolded states; instead the CHC-CTR contacts are more prominent. The first four NTR residues are disordered, and do not interact with CHC, but instead with the turn region in the helically folded states. These same interactions are also revealed within the shorter distance cut-off of 4 Å (Fig. 3B, right triangle) indicating no change in contact frequencies from the 5 Å cut-off. In order to identify the dominant force behind these interactions, we computed the interaction types from contacts with frequencies >50%. These interaction maps identify H-bonds, salt bridges and hydrophobic networks between the domains. We show that for all the states of Aβ42, the types of inter-domain interactions are very similar within a cut-off of 5 Å or 4 Å (Fig. S4A). Hydrophobic interactions and salt bridges are the dominant forces for the folded and partially folded states, mainly because of the charged residues in the NTR and the turn region and hydrophobic residues in the CHC. For the unfolded states, we see some remaining H-bonds between the NTR and the turn region.

αS also exhibits anti-diagonal contacts between the domains NAC and the proximal CTR (K96–Q134 away from the CTR tail) in the helically folded states (Fig. 3B), which are sparse (<20%) or totally absent in the partially folded and unfolded states for both 5 Å (left triangle) and 4 Å (right triangle) distance cut-offs. For CHARMM36/TIP4P model, the CTR end (135–140) interactions with NTR, in addition to NTR-NTR and NTR-NAC contacts, stabilize the folded state. By contrast, strong contacts (>70%) between NTR and NAC form in the unfolded states, with the CTR mostly extended into solution. The interaction maps (Fig. S4B) identify predominantly hydrophobic interactions between NAC and CTR, and salt bridge and H-bonds between NTR and CTR for the folded states, and hydrophobic contacts and H-bonds for the partially folded and the unfolded states.

Cross-correlation maps identify anti-correlated sub-domains that propagate disorder from helical states. To characterize the partially folded states and probe the origin of their higher conformational heterogeneity compared with unfolded states, we performed cross-correlated network analyses of Cα
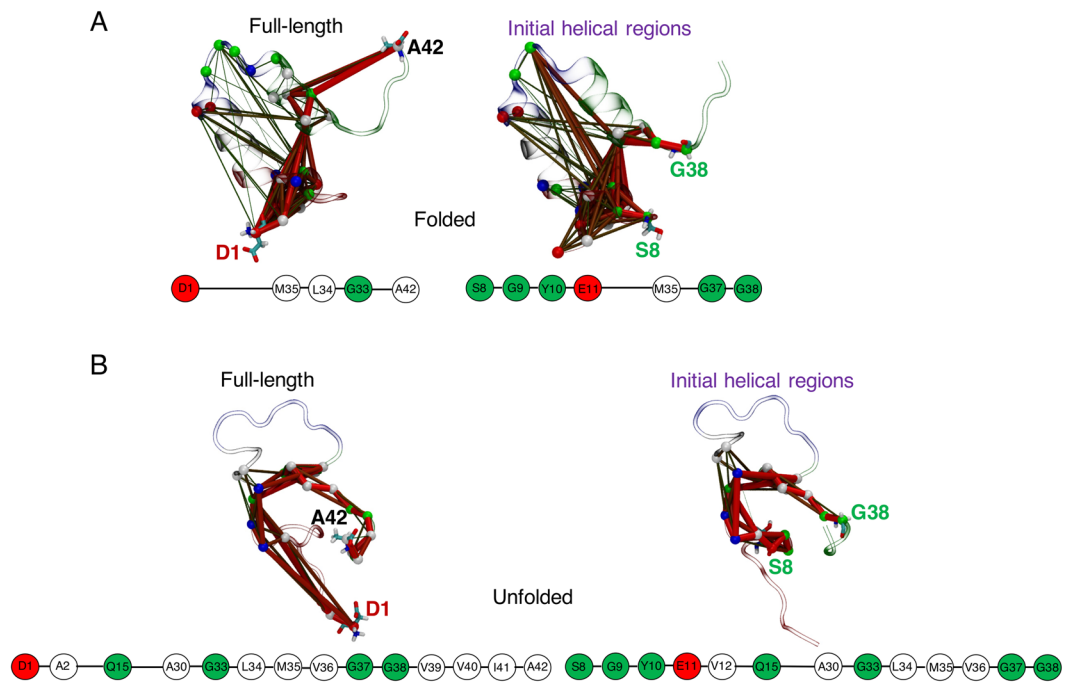
Figure 5. Consensus networks of Aβ42 using a $|c_{ij}|$ filter of 0.4, showing sub-optimal paths for propagation of disorder from **(A)** helically folded and **(B)** unfolded states, for full-length (source D1 – sink V42) and initial helical regions (source S8 – sink G38) in the folded structure. Network structures are overlaid on representative folded and disordered conformations with domains coloured by region: red = NTR, grey = CHC, blue = turn, and green = CTR. Path thickness reflects inter-residue (intra-domain) coupling strengths. Optimal paths are shown below each network, with residues coloured by hydropathy: white = hydrophobic, green = polar, red = negatively charged, and blue = positively charged. Broken lines between two nodes represent coupling including all residues in between.

atomic fluctuations using the R package, Bio3D[53,54] (see Methods) across all states of Aβ42 and αS. Dynamic cross-correlation maps (DCCMs) of Aβ42 filtered at Pearson's correlation coefficient (PCC or $|c_{ij}|$) ≥ 0.4 (see Methods for selection criteria of this threshold value) identify positive correlation (red) in the helically folded states between the closely interacting (Fig. 3A) proximal (D1–R5) NTR and the turn region (22–29), and between the distal NTR and M35 in the CTR (Fig. 4A, left panel), which helps preserve the 'foldedness' of the helices. However, the overall tendency to unfold helices to random coil (herein referred to as 'propagation of disorder' or 'disorder propagation') comes from dominant off-diagonal negatively correlated (anti-correlated, blue) sub-domains (cluster of residues within a domain) involving the NTR, turn region and CTR.

The sole cross-correlated regions are actually anti-correlated in the partially folded helical states of Aβ42, which involve communications between the sub-domains of NTR: D1–E3 and CHC: L17–V18, and turn: E22–V24 and CTR: V36–G38, and to a small extent sub-domain of NTR: H14 – K16 and A30 in CTR (Fig. 4A, middle panel). DCCM from completely unfolded states of Aβ42 is further reduced to intra-NTR and CHC-CTR negative correlations (Fig. 4A, right panel). In addition, two off-diagonal weakly anti-correlated residue pairs are observed: one between G25 and G38, and the other between V12 and E22, identifying that these two regions cycle between folding, unfolding and refolding, forming a transient, short β-hairpin motif from the disordered states.

Similar to Aβ42, DCCMs of the folded states of αS retain positively correlated areas at $|c_{ij}| \geq 0.4$ involving close-range (Fig. 3B) couplings between sub-domains of NTR and CTR, and NAC and CTR sub-domains (Fig. 4B, left panel), while a larger number of anti-correlated sub-domains ensure general propensity to unfold helices. As such, the partial folds only retain anti-correlated sub-domains (Fig. 4B, middle panel) encompassing NTR (V15–A27 and V52–K60) and NAC (E61–T64), NAC (F94–K96) and CTR (Y133–D135), and intra-CTR (E104–E110 and S129–A140). Due to the short-range NTR-NAC contacts (Fig. 3B) in the unfolded states, intra-NTR positive correlations are revealed (Fig. 4B, right panel), which are further counteracted by NTR-NAC and NTR-CTR anti-correlations. These off-diagonal NTR-NAC anti-correlated regions (between E20–E35 and K58–A78) identify a β-hairpin region folding from completely disordered states, similar to the anti-correlated residue pairs seen for Aβ42 in the unfolded states. As the purpose of the present work is to characterise the helical ensemble, the potential long-range intra-peptide residue couplings for a monomeric β-hairpin structure will be examined elsewhere. Overall, the general trend of propagation of disorder for both Aβ42 and αS involves anti-correlated motions between sub-domains for all the states.

Optimal paths of disorder propagation in both helically folded and fully unfolded states skip the central hydrophobic domains. Concerted residue fluctuations were modelled by consensus correlation networks with four chosen source-sink residue pairs denoting regions encompassing the full-length peptide,
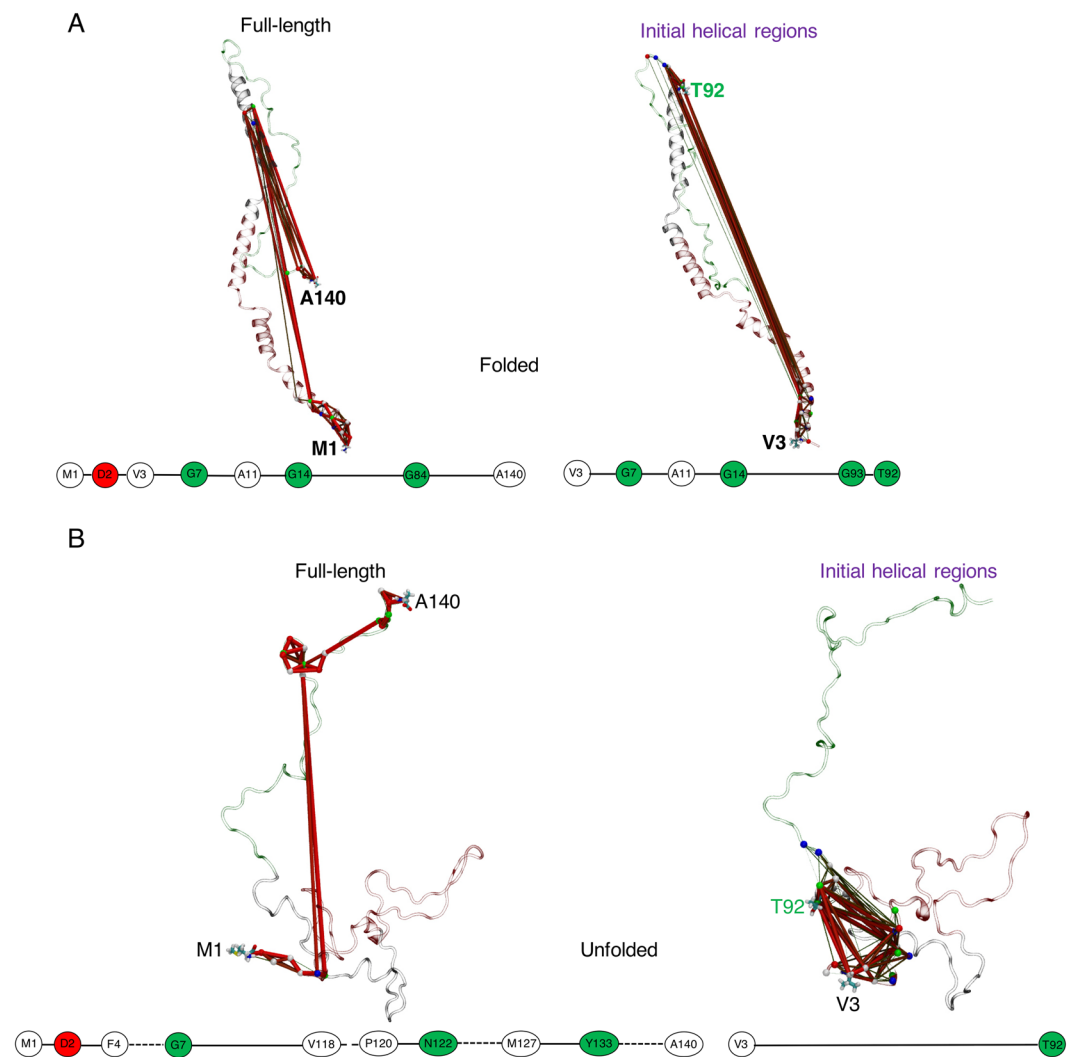
Figure 6. Consensus correlation networks of αS using a $|c_{ij}|$ filter of 0.4. Network paths for the **(A)** helically folded, and the **(B)** unfolded states. Legend is as for Fig. 5, except domains are coloured as red = NTR, grey = NAC and green = CTR, with different source-sink residue pairs annotating the full-length (source M1 – sink A140) and the initial helical regions (source S8 – sink G38).

initial helices, helix 1 and helix 2 for both Aβ42 and αS (see Methods) at $|c_{ij}| \geq 0.4$. These consensus networks were computed from combinations of the two physical models (force field and water model) designated for each of the folded, partially folded and completely unfolded states, thus providing robust sampling across both fast (short-lived) and slow (long-lived) unfolding helical states. The path length distributions for the helically folded states of Aβ42 (Fig. S5A under Supplementary Information) show that the overall path of helix 1 is relatively short, while helix 2 path is nearly as long as the initial helical regions. Cooperativity between helices in unfolding is observed as the shortest path of helix 2 (Fig. S5C) from source K28 traces back to S26 (adjacent to end of helix 1) before reaching M35 in the CTR, providing a potential explanation for the high energy barriers faced in unfolding the initial helices of the folded states (Fig. 2A). The optimal path in full-length helically folded Aβ42 (Fig. 5A) shows that disorder propagates through direct coupling of the NTR tip D1 and site M35 of CTR without populating other (sub)domain-specific nodes, which may confer a protective effect on the core helices. A similar trend (path from E11 in NTR to M35 in CTR) is seen when the initial helical regions are considered. Node degeneracy maps (Fig. S5B) reveal M35 forming a centralized hub (Fig. S5D).

The completely unfolded states show non-overlapping path-length distributions, with helix 2 paths being the shortest, but the helix 1 paths are less dense than full-length or initial helical region paths (Fig. S5E). The ideal path on helix 1 skips the CHC region (Fig. S5G, see also minor paths in Fig. S5F and network hubs in Fig. S5H, which mostly involves NTR and CTR), as do the optimal paths of full-length peptide and initial helical regions (Fig. 5B, showing direct coupling between NTR Q15 and CTR A30). However, helix 1 does not include the CTR, which contributes significantly to the path density of the full-length Aβ42 (Fig. 5B).

In contrast to the Aβ42 CHC region, DCCM of αS (Fig. 4B) shows that sub-domains of NAC are correlated with other regions in folded and unfolded states. However, the optimal paths of disorder propagation in these
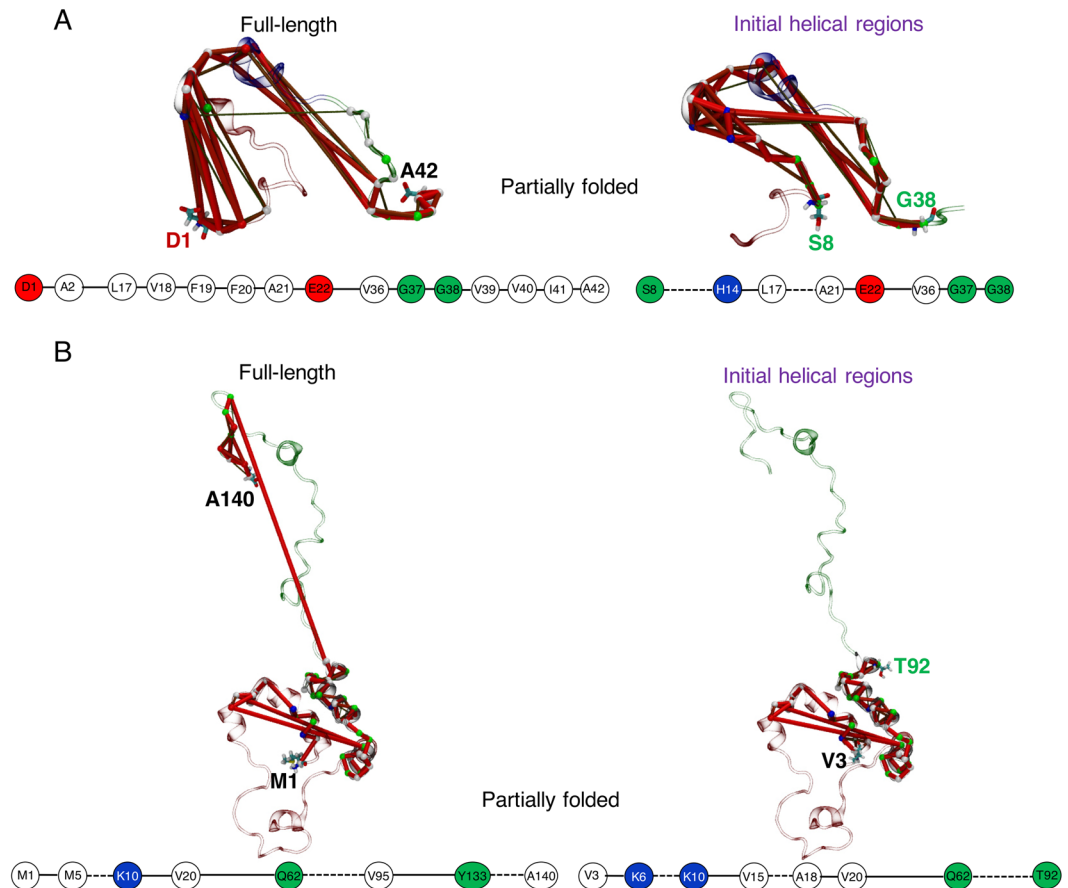
Figure 7. Consensus networks of the partially folded helical states of **(A)** Aβ42, and **(B)** αS following the same scheme of source-sink residue pairs used in Figs. 5, 6.

states (Fig. 6) generally avoid the NAC sub-domains, with only G84 of the NAC falling on the ideal path traversing the NTR (M1 – G14) and CTR for the full-length helically folded states (Fig. 6A). This is because nodes within the NAC that could potentially act as hubs are isolated by long non-populated sequences (Figs. S6D, S6H). For the folded states, the shortest path on helix 1 travels until S42 (close to the start of helix 2) and traces back to the end of helix 1, while for helix 2 it travels all the way back to A18 (belonging to helix 1) before ending with V95 – T92 (Fig. S6C). This indicates a prevalent inter-helical cooperative effect in unfolding the initial stable helices, as also seen in Aβ42. The NTR shows the maximum number of paths (Fig. S6B). The shortest path along the full-length of the disordered states communicates between NTR (M1–G7) and CTR (V118–A140) (Fig. 6B), which for the initial helical regions do not involve any intermediary nodes (direct coupling between V3 and V92; see also Figs. S6E–G).

Dynamic coupling between terminal tips and central hydrophobic domains confers long-range regulation of partially folded helical states. Networks of partially folded Aβ42 also sample distinct path-length distributions much like unfolded states, but helix 1 paths are densest (Fig. S7A). This is because the CHC region not only traces the shortest path on helix 1 (Fig. S7C), but also allows a greater number of paths to traverse than in the full-length or initial helical paths. The optimal path of disorder propagation from full-length partial helical folds involves couplings between the N-terminal tip (D1–A2) and transiently folded CHC (L17–E22), before terminating at CTR (V36–A42) (Fig. 7A). In the initial helical region, the shortest path skips from the distal sub-domain of NTR (S8–H14) to the CHC region. The paths identify extensive communication between sub-domains of NTR and CHC (Fig. 7A), which do not form short-range (<5 Å) geometrical contacts in the partially folded states. Thus, the CHC residues are extremely degenerate with all source-sink paired paths (except helix 2, which does not include the CHC region) (Figs. S7B, S7D).

In contrast to folded and unfolded states, the optimal path of partially folded states of αS reveal significant coupling between the full NAC domain (Q62–V95) and CTR sub-domain (Y133–A140) spanning the source-sink pair of full-length αS (Fig. 7B). The shortest path for helix 1 identifies similar node connectivity when calculated across full-length or initial helices in the NTR, which for helix 2 is continuous from K45 to V82 mapping the KTKEGV imperfect repeats (Fig. S7G). The path lengths are distributed according to the size of the fragments and are very distinct (Fig. S7E). The dense population of several paths crossing the NAC residues (Fig. S7B) clearly distinguishes partially folded states from the node degeneracy of helically folded or unfolded states (Fig. S6B,
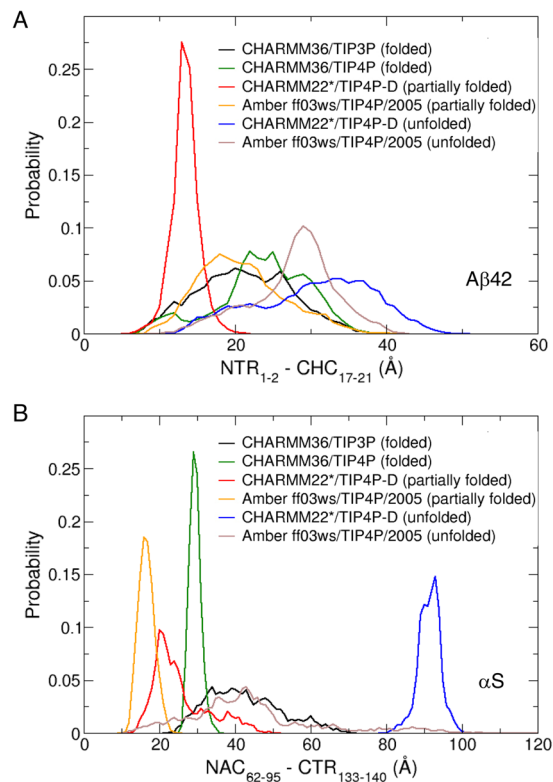
Figure 8. Probability distribution of distances from the **(A)** centre of mass (COM) of D1 – A2 in the NTR or NTR$_{1-2}$ to the COM of the hydrophobic CHC (L17 – A21) or CHC$_{17-21}$ for Aβ42, and **(B)** from the COM of the hydrophobic NAC (Q62 – V95) or NAC$_{62-95}$ to the COM of Y133 – A140 in the CTR or CTR$_{133-140}$ of αS, for folded, partially folded and unfolded states. These regions are shown to be highly correlated in the partially folded helical states (Fig. 7).

S6F) with large gaps on NAC nodes. Finally, the NAC domain is the most prominent network hub, followed by the CTR (Fig. S7H).

Long-range coupling between charged terminal ends and central hydrophobic domains are modulated at distances up to 20 Å.   We calculated separations between sub-domains involved in concerted long-range coupling in the partially folded helical states, by monitoring the distribution of distances from the centre of mass (COM) of the N-terminal end (D1–A2) of Aβ42 to the COM of CHC (L17–A21), and the COM of C-terminal end (Y133–A140) of αS to the COM of NAC (Q62–V95) (Fig. 8). For Aβ42, concerted fluctuations peak at separations of ~15–18 Å (Fig. 8A) in partially folded states, and for αS at ~18–20 Å (Fig. 8B). By contrast, all helically folded and unfolded states sample longer and/or broader distributions, which indicates that partially folded helical states are modulated by the (predominantly disordered) charged termini tails through long-range couplings up to a distance of 20 Å, beyond which the long-range effect fades out. Sharper peaks in distance distributions within the partially folded helical states of αS identifies that the disordered CTR in αS is able to regulate a comparatively more stable and longer helical NAC with low fluctuations in NAC-CTR distance distributions than shorter and less stable helical CHC, and fluctuations in NTR-CHC distances of Aβ42. It was previously shown that the β-hairpin formed by the Aβ(17–34) fragment is allosterically stabilized by the N-terminal around 10 Å from the centre of the fragment, by the release of entropy to the surrounding water[55]. Our analyses of water dynamics within 10 Å of the most hydrophobic CHC (Fig. S8A) and the fibril-forming core of NAC (68–78)[56] (Fig. S8B) show subtle differences in temporal water retention between states. Evidently, the cores of the partially folded structures retain more water molecules than either of the disordered or the folded states for both IDPs, which supports the hypothesis that entropically favourable release of retained water may drive inter-molecular oligomeric associations[57]. Note that direct modelling of oligomerization remains extremely challenging even with high-performance computing due to the degeneracy of the free energy surfaces for assembly.

Discussion

In light of the experimental evidences that early oligomerization of some amyloidogenic IDPs may be linked to formation of helical intermediates by helix-helix associations[3,43,58], we have extensively modelled the helical subspaces of Aβ42 and αS in order to account for the driving force behind aggregation-prone nature of partially folded helical states, and compared them to helically folded and unfolded states. The sampling strategy employed here combines two different force fields/water models for each of the three states, where differences in
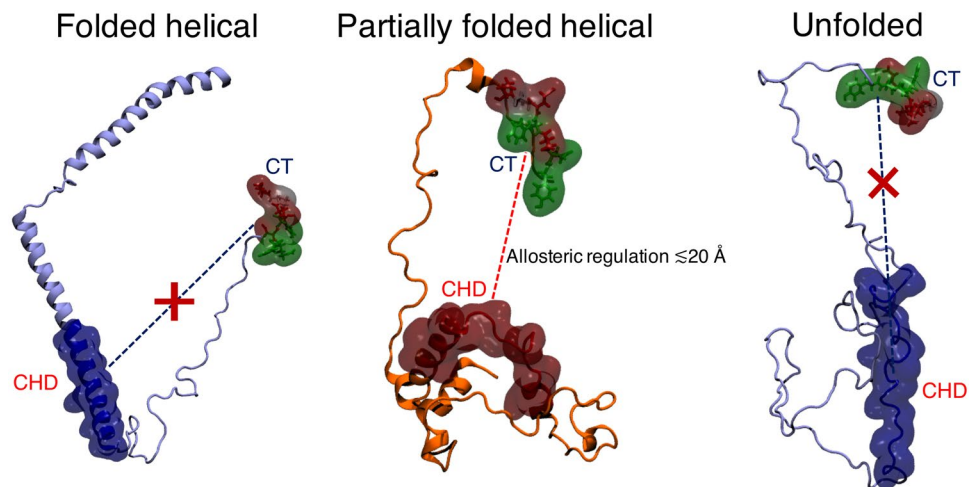
Figure 9. Schematic of proposed mechanism of long-range (up to ~20 Å) regulation between the tip of the charged terminus (CT) and the central hydrophobic domain (CHD) of aggregation-prone partially folded helical states. This coupling is lost at distances >20 Å and is absent for the non-aggregating helically folded and the unfolded states.

helix unfolding rates are used to sample short-lived and long-lived conformations, thus generating an expansive ensemble of folded, partially folded and unfolded states pertinent to the cellular environment. However, the possible existence of other physiologically relevant states cannot be ruled out. For example, both the helical and disordered states sampling efficiency could further be evaluated with the recently parametrized Amber ff99SB-disp[59] force field, as it is primarily developed to sample both folded and disordered proteins accurately. Many factors will modulate the helical sampling space in a cellular milieu[60]. For example, in a membrane-mimicking environment, Aβ42 has shown predominance of monomeric folded α-helical structures, which resist toxic aggregation[61,62]. The composition of the lipid membrane may also influence the rate of Aβ42 fibrillation, as the oligomers are rich in β-sheet content[63,64]. On the other hand, αS may sample several different helical states on binding to lipid membranes, which could further their aggregation to fibrils[65]. Both extended and broken helical structures have been reported for monomeric αS interacting with phospholipid bilayers[66,67]. Acidic pH may in general facilitate aggregation of amyloid peptides[60]. The conformational ensembles of αS monomers have been shown to form more compact and ordered structures at low pH[68], while Aβ42 may sample α-helical structures in the acidic environment[69]. Other conditions such as ionic strength and metal ion concentrations, and temperature may have an impact on aggregation propensity and the helical subspace sampling, as summarized in a very recent comprehensive review[60]. The present work focusses on helical conformational sampling monomers in their native state in aqueous solution, which the chosen model components and parameters mimic as closely as possible.

In our previous work, we have highlighted on the role of short-range intra-peptide modulation of the central hydrophobic domains by the flexible termini in stabilizing their aggregation-impeding folded helical states[31]. In this work, we provide an analytical insight into the regulation of aggregation-competent Aβ42 and αS partial helical folds *via* dynamic control of long-range intra-peptide signalling between two distal sub-domains encompassing the charged termini (C-terminus in αS and N-terminus in Aβ42) and the central hydrophobic hotspots of amyloidogenesis.

Free energy maps show that the partially folded helical conformers (intermediate between folded and completely disordered states) in both Aβ42 and αS display a relatively flat surface that samples heterogeneously interconverting states with expansive populations in comparison to the folded helical or unfolded conformations. Contact maps identify that the charged terminal (CT) tips in both peptides do not make short-range (≤5 Å or ≤4 Å) contacts with the central hydrophobic domains (CHD) in the partially folded helical states. However, the dynamic cross-correlation maps reveal that sub-domains of CT and CHD frequently partake in long distance anti-correlated motions, which are absent in unfolded states and counterbalanced by positively correlated sub-domains in the folded states. Computed cross-correlated network analyses of both helically folded and unfolded states reveal that the optimal paths of disorder propagation do not traverse through the CHD in both Aβ42 and αS, but display long-range dynamic coupling between the termini and CHD in the partially folded helical states, conferring that the flexible termini may regulate the local helical folds in CHD.

Our calculations reveal that the susceptibility to form partially folded helices in the CHD is upregulated by long-range coupling with charged residues in predominantly disordered terminal ends of Aβ42 and αS. This long-range regulatory effect manifests across a distance of up to ~20 Å and might play a major role in promoting helix-helix associations. Our predictive model is consistent with hydrophobic protein self-assembly[70] modulated by immobilized ions[71]. Hydrophobic interactions can be switched on or off according to spatial arrangements of charged groups[72], as indicated also from our analyses, providing a potential explanation for why partial helices may be more prone to oligomerization[6,7,43,73] than predominantly folded helical or unfolded states (Fig. 9). This long-range effect is reminiscent of allosterically coupled sites that may facilitate initial assembly of the peptides. In turn, our predictive models are in strong agreement with the proposed reaction pathways and state parameters

of initial aggregation, such that increasing concentrations of solvent-induced α-helical content leads to increased rate of fibrillation up to a certain fraction of helical content (partially folded helices), beyond which the rate of fibrillation slows down considerably (fully folded helices)[13].

In a more recent study involving an engineered coiled-coil helical dodecamer (12-mer) of the small model peptide ccβ[74], the same authors used extensive REMD simulations to map the intermediate steps for helix →β-hairpin/β-strand→β-sheet transition during the formation of amyloid-like protofibrils. The identified transitions involved thermally induced conformational straightening of the β-hairpin structures facilitated by breaking of intrapeptide H-bonds, and consequent formation of interpeptide H-bonds with other β-strands to form β-sheets. Furthermore, the authors highlight that the chain additions are a highly cooperative process, and that hydrophobic interactions may play a secondary role in directing the site-specificity of the aligned assemblies. Our predictive models confirm the importance of swapping between intra and inter peptide contacts during structural transitions, involving rearrangements in both H-bond patterns and hydrophobic contacts. Our analysis of helix-helix associations identifies that aggregation may be triggered by decay of intrapeptide H-bonds as the peptides transition from fully to partially folded helices, and also that interpeptide hydrophobic interactions could be modulated by charged terminal residues, which may change the aggregation pathway.

Moreover, our predictions are consistent with those drawn from a master-equation based formalism that was used in conjunction with REMD simulations to map kinetics of dimer formation in the short amyloidogenic peptide NNQQ[75]. Using monomer-monomer closest contact distances as the reaction coordinate, the authors characterized different binding modes that could potentially distinguish fast interconverting binding that forms unstable encounter complexes from slow binding that creates stable dimers. Along similar lines, our analyses reveal that the diversely distributed population of partially folded helical states may form stable helical dimers when there is concerted coupling between the CT and CHD only at distances ≤20 Å. The long range coupling effect is lost above this distance, and may lead to just formation of encounter complexes, in the terminology of ref.[75].

There is mounting evidence that the N-terminus of Aβ oligomers are implicated in neurotoxicity[76–78] and should be considered as promising targets for treatment of AD[79]. In addition to the N-terminus in Aβ42, our unprecedentedly detailed models suggest that the C-terminus in αS may be considered an important target in PD. A feature of the dynamic cross-correlated network analyses method is the difficulty in picking up signals of correlated motions that may be functionally relevant, but which may appear only fleetingly in MD simulations. Attempts made to circumvent this limitation includes use of a Bayesian-based pattern recognition technique named multi-modal DCC[80]. Future tests of this method on intra-peptide long-range communication in IDPs could be useful. In turn, other computational methods for detecting pathways from conformational ensembles generated from MD simulations would include (but are not limited to) normal mode analyses, perturbation-response scanning method, and mutual information analyses[81]. In addition, the information theory measure of entropy transfer can be effectively employed to complement dynamically coupled pathways between residue pairs or domains forming the source-sink pairs in order to assess the causality of correlated fluctuations[82]. Recent works have estimated transfer entropy from MD data to identify driver-responder residues that are involved in unfolding and refolding of α-helix to β-sheet in C-terminal domain of transcription antiterminator RfaH[83], helix unfolding in Ets-1 protein following binding to the target DNA sequence[84], folding/unfolding of small proteins[85], and allosteric activity that governs recognition of Ubiquitin in Ubiquitin-Proteasome binding[86]. We wish to emphasize that a very large computational effort would be necessary to explicitly model the multitude of paths for self-assembly of multiple helical monomers into oligomers – our current dataset provide the required and representative populations of different types of starting helical conformations for such future studies that may become more feasible with advances in high-performance computing and advanced sampling methods. We envisage that the modelling strategy employed in the present study of Aβ42 and αS could be applied to study monomeric helical intermediates of other IDPs[87] and help find ways to rationally reduce their tendency to promote helical oligomerization[43].

Methods

Molecular dynamics simulations.   The initial structures for equilibrium atomistic molecular dynamics (MD) simulations were obtained from the NMR-derived helix-kink-helix structures of Aβ42 solved in 20% water/80% deuterated hexafluoro isopropanol (PDB code 1IYT[32]) and micelle bound human αS solved in aqueous solution with sodium dodecyl sulphate (PDB code 1XQ8[33]). We wanted to monitor the unfolding of helices, so we chose the aforementioned experimental models with predominant helical conformations. We note that other experimental helical structures of Aβ42 (PDB ID: 1Z0Q[88]) and αS (PDB ID: 2KKW[89]) exist. However, 1Z0Q has lower helical content than that of 1IYT, and 2KKW has similar helical regions to 1XQ8, thus explaining our rationale for selecting the helical models as our starting geometry. The two helices for Aβ42 span Ser8–Gly25 (helix 1) and Lys28–Gly38 (helix 2), while for αS they span Val3–Val37 (helix 1) and Lys45–Tyr92 (helix 2). All MD simulations were performed using the Gromacs 5.1.4[90] package. The starting structures were represented by three different protein force fields – CHARMM36[38], CHARMM22*[41] and Amber ff03ws[34], and solvated in boxes containing four different water models – CHARMM-modified TIP3P[39], TIP4P, protein-water interaction scaled TIP4P/2005[34,42] and TIP4P-D[35] resulting in four distinct combinations of force fields and water models: CHARMM36/TIP3P, CHARMM36/TIP4P, CHARMM22*/TIP4P-D and Amber ff03ws/TIP4P/2005. Additionally, we performed 3.4 μs of Hamiltonian replica exchange (70 ns * 48 replicas simulated at different Hamiltonians). These simulations used solute tempering (scaling)[91,92] (H-REST) with CHARMM22*/TIP4P-D (αS) starting from a completely disordered state. The steady states of the equilibrium MD (EMD) and H-REST simulations were assessed by the convergence of temporal cumulative average helix percentages (Fig. S1B under Supplementary Information). The converged helical content for CHARMM22*/TIP4P-D with αS for both EMD and H-REST show steady state populations of ~20%, indicating similar helical subspace sampling for the two different methods. The last 200 ns was used for analysis of the EMD simulations of different states (folded, partially

States	Simulation time (ns)	Complete unfolding time (ns)	α-helical content (%mean ± SD)
Aβ42			
Folded (CHARMM36/TIP3P)	2014	N/A	50 ± 4
Folded (CHARMM36/TIP4P)	2018	N/A	52 ± 4
Partially folded and Unfolded (CHARMM22*/TIP4P-D)	2011	206	18 ± 6
Partially folded and Unfolded (Amber ff03ws/TIP4P/2005)	2020	1527	20 ± 4
αS			
Folded (CHARMM36/TIP3P)	1540	N/A	44 ± 8
Folded (CHARMM36/TIP4P)	1411	N/A	37 ± 5
Partially folded and Unfolded (CHARMM22*/TIP4P-D)	1413	625	8 ± 6
Partially folded and Unfolded (Amber ff03ws/TIP4P/2005)	1488	1160	6 ± 6

Table 1. Details of equilibrium MD simulations (EMD) for different states of Aβ42 and αS.

folded and unfolded), while the last 20 ns (out of the 70 ns/replica) was used for analysis of the H-REMD simulation. We show using free energy maps of percent helicity and radius of gyration, that the conformational spaces explored for the partially folded helical states in our EMD simulations matches up with H-REST simulations. H-REST sampled conformational spaces of transiently formed helices that are refolded from a completely disordered state, confirming that our sampling of helical subspaces is statistically robust (Fig. S1C).

Physiological concentrations (0.15 M) of NaCl were added to neutralize the peptide formal charge. Na$^+$ counterions were added to neutralize the peptide formal charge and 0.15 M of NaCl was added to mimic physiological ionic strength. 3 Na$^+$ counter-ions and 25 NaCl ion pairs were added for Aβ42, while 9 Na$^+$ counter-ions and 133 NaCl ion pairs were added for the more highly-charged and larger αS simulation cell. All Histidine residues (H6, H13 and H14 in Aβ42, and H50 in αS) were modelled in their neutral states with the ε-nitrogen protonated, as deduced from His contacts to neighbouring residues in the starting structures. Both termini of the two monomers were kept charged. The minimum distance between a protein atom and the boundary of the boxes was set at 20 Å. The protein and water molecule bond lengths to hydrogen were constrained using the LINCS[93] and the SETTLE[94] algorithms, respectively. An integration step of 2 fs was chosen for running the simulations employing the velocity Verlet integrator[95], and coordinates were saved every 20 ps. Long-range electrostatics were treated by the Particle mesh Ewald (PME) method[96]. All systems were minimized for 5000 steps, and equilibrated for 1 ns in NVT ensemble followed by another 1 ns of NPT equilibration with the reference pressure at 1 bar and a time constant of 4 ps using the Berendsen barostat[97]. Protein and non-protein (water and ions) molecules were coupled separately to an external heat bath with a coupling time constant of 1 ps using the velocity rescaling method[98]. The production runs were carried out in the constant pressure and temperature NPT ensemble.

Simulations were run for ~2 μs for Aβ42 and ~1.5 μs for αS with each of the four combinations of protein force fields and waters models (for details of these simulations, see Table 1), amounting to a total run time of ~14 μs. For simulations using CHARMM36/TIP3P and CHARMM36/TIP4P, complete unfolding of the helices does not occur within the simulation time, with both Aβ42 and αS, retaining ~1/3 helicity (α-helix + π-helix + 3₁₀-helix, Fig. S1A). These states are hence regarded as folded states. For CHARMM22*/TIP4P-D, fast and complete unfolding occurred by ~0.2 μs for Aβ42 and ~0.6 μs for αS, while for Amber ff03ws/TIP4P/2005 a slower complete unfolding occurred by ~1.5 μs for Aβ42 and ~1.2 μs for αS (Fig. S1A). Hence the last 200 ns before complete unfolding of CHARMM22*/TIP4P-D and Amber ff03ws/TIP4P/2005 are regarded as partially folded states, and the final 200 ns of the trajectories with CHARMM22*/TIP4P-D and Amber ff03ws/TIP4P/2005 are regarded as completely unfolded states. All analyses were carried out using the Gromacs 5.1.4 analysis tools. Interdomain contact probability maps and interaction maps were generated using the CONAN[99] contact analysis tool and water dynamics was monitored using VMD[100] software.

Free energy calculations. The values for the Gibbs free energy (kcal/mol) using two different order parameters were obtained using the equation:

$$\Delta G = -k_B \cdot T \cdot (lnP_i - lnP_{max}) \tag{2}$$

where P_i is the probability distribution for pairs of order parameters, and P_{max} is its maximum, such that $lnP_i - lnP_{max}$ identifies the lowest free energy point at $\Delta G = 0$.

Dynamic cross-correlation network analysis. Dynamic network analyses of cross-correlated motions were performed using the Bio3D R package[53,54]. Correlated residue pairs were identified using linear Cα atomic displacements, as opposed to linear mutual information[101]. Consensus networks were generated for each of the designated folded, partially folded, and unfolded states of WT Aβ42 and αS, using the method similar to Luthey-Schulten[102], but with network edges weighted based on the Cα-Cα cross-correlation values to include long-range interactions between residues that are not in direct contact. In a mass-weighted network graph, the

residues represent nodes i and j, and the weights represent the edges with Pearson's cross-correlated (PCC) values[103] c_{ij} connecting the nodes.

A community is defined as a cluster of nodes (residues) that are highly connected through correlated motions. The value of c_{ij} ranges between -1 and 1. c_{ij} is closer to -1 when the two coupled residues are negatively correlated or anti-correlated, is 0 when they uncorrelated, and closer to 1 when they are positively correlated. All networks were constructed with a $|c_{ij}| \geq 0.4$ (the modulus represents absolute value taken for both positive and negative correlations) based on the following criteria: *(1)* False positives or false negatives using too low or high values of $|c_{ij}|$, respectively, could be circumvented by using a near-midpoint value of cross-correlated PCC. *(2)* The network structure remains consistent between 0.35, 0.4 and 0.45 allowing similar community partitioning. Dynamic cross-correlation maps (DCCMs) were further plotted at a PCC threshold of 0.4, from matrices representing $|c_{ij}| \geq 0.4$.

Network path analyses were carried out by considering four different combinations of 'source' and 'sink' residue pairs (start and end nodes) in both Aβ42 and αS: *(1)* N-terminal end and C-terminal end, *(2)* Start of initial helical regions (in the original PDB structures) and end of helix, *(3)* Start of helix 1 and end of helix 1, and *(4)* Start of helix 2 and end of helix 2. Consequently, the optimal path (shortest path) and sub-optimal paths were deduced out of a total of 500 distinct paths constructed for each source and sink pair, using the shortest loopless path algorithm[104] implemented in Bio3D. Path length (summation over the weights of traversed edges and not representative of an actual geometrical length) distributions with their probability density were compared, which measures the coupling strength between residue pairs. Calculated non-normalized node degeneracies quantify the number of paths traversing each node, highlighting densely connected nodes. In addition, frequently crossed 'hub' nodes were identified from measured betweenness centrality (or node centrality) of the number of unique shortest paths crossing each node.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

References

1. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1332–1340, https://doi.org/10.1038/s41593-018-0235-9 (2018).
2. Nagel-Steger, L., Owen, M. C. & Strodel, B. An Account of Amyloid Oligomers: Facts and Figures Obtained from Experiments and Simulations. *Chembiochem* **17**, 657–676, https://doi.org/10.1002/cbic.201500623 (2016).
3. Abedini, A. & Raleigh, D. P. A role for helical intermediates in amyloid formation by natively unfolded polypeptides? *Phys. Biol.* **6**, 015005, https://doi.org/10.1088/1478-3975/6/1/015005 (2009).
4. Kirkitadze, M. D., Condron, M. M. & Teplow, D. B. Identification and characterization of key kinetic intermediates in amyloid beta-protein fibrillogenesis. *J. Mol. Biol.* **312**, 1103–1119, https://doi.org/10.1006/jmbi.2001.4970 (2001).
5. Ghosh, D. *et al.* Structure based aggregation studies reveal the presence of helix-rich intermediate during alpha-Synuclein aggregation. *Sci. Rep.* **5**, 9228, https://doi.org/10.1038/srep09228 (2015).
6. Fezoui, Y. & Teplow, D. B. Kinetic studies of amyloid beta-protein fibril assembly. *Differential Eff. alpha-helix stabilization. J. Biol. Chem.* **277**, 36948–36954, https://doi.org/10.1074/jbc.M204168200 (2002).
7. Anderson, V. L., Ramlall, T. F., Rospigliosi, C. C., Webb, W. W. & Eliezer, D. Identification of a helical intermediate in trifluoroethanol-induced alpha-synuclein aggregation. *Proc. Natl Acad. Sci. USA* **107**, 18850–18855, https://doi.org/10.1073/pnas.1012336107 (2010).
8. Abedini, A., Cao, P. & Raleigh, D. P. Detection of Helical Intermediates During Amyloid Formation by Intrinsically Disordered Polypeptides and Proteins. *Methods Mol. Biol.* **1345**, 55–66, https://doi.org/10.1007/978-1-4939-2978-8_4 (2016).
9. Rovere, M., Sanderson, J. B., Fonseca-Ornelas, L., Patel, D. S. & Bartels, T. Refolding of helical soluble α-synuclein through transient interaction with lipid interfaces. FEBS Letters, https://doi.org/10.1002/1873-3468.13047 (2018).
10. Vivekanandan, S., Brender, J. R., Lee, S. Y. & Ramamoorthy, A. A partially folded structure of amyloid-beta(1-40) in an aqueous environment. *Biochem. Biophys. Res. Commun.* **411**, 312–316, https://doi.org/10.1016/j.bbrc.2011.06.133 (2011).
11. Khurana, R. *et al.* Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* **40**, 3525–3535 (2001).
12. Ahmad, A., Millett, I. S., Doniach, S., Uversky, V. N. & Fink, A. L. Partially folded intermediates in insulin fibrillation. *Biochemistry* **42**, 11404–11416, https://doi.org/10.1021/bi034868o (2003).
13. Steckmann, T., Awan, Z., Gerstman, B. S. & Chapagain, P. P. Kinetics of peptide secondary structure conversion during amyloid beta-protein fibrillogenesis. *J. Theor. Biol.* **301**, 95–102, https://doi.org/10.1016/j.jtbi.2012.02.012 (2012).
14. Wolynes, P. G. Recent successes of the energy landscape theory of protein folding and function. *Q. Rev. biophysics* **38**, 405–410 (2005).
15. Budyak, I. L. *et al.* Early folding events protect aggregation-prone regions of a beta-rich protein. *Structure* **21**, 476–485, https://doi.org/10.1016/j.str.2013.01.013 (2013).
16. Wei, G., Xi, W., Nussinov, R. & Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.* **116**, 6516–6551, https://doi.org/10.1021/acs.chemrev.5b00562 (2016).
17. Wodak, S. J. *et al.* Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27**, 566–578, https://doi.org/10.1016/j.str.2019.01.003 (2019).
18. Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390–394, https://doi.org/10.1038/nature12294 (2013).
19. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433–443, https://doi.org/10.1002/prot.20232 (2004).
20. Salvi, N., Abyzov, A. & Blackledge, M. Analytical Description of NMR Relaxation Highlights Correlated Dynamics in Intrinsically Disordered Proteins. *Angew. Chem. Int. Ed. Engl.* **56**, 14020–14024, https://doi.org/10.1002/anie.201706740 (2017).
21. Kurzbach, D. *et al.* Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference. *Phys. Chem. Chem Phys* **18**, 5753–5758, https://doi.org/10.1039/c5cp04858c (2016).

22. Kurzbach, D. *et al.* NMR probing and visualization of correlated structural fluctuations in intrinsically disordered proteins. *Phys. Chem. Chem Phys* **19**, 10651–10656, https://doi.org/10.1039/c7cp00430c (2017).
23. Parigi, G. *et al.* Long-range correlated dynamics in intrinsically disordered proteins. *J. Am. Chem. Soc.* **136**, 16201–16209, https://doi.org/10.1021/ja506820r (2014).
24. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339, https://doi.org/10.1038/nature13001 (2014).
25. Berlow, R. B., Dyson, H. J. & Wright, P. E. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J. Mol. Biol.* **430**, 2309–2320, https://doi.org/10.1016/j.jmb.2018.04.003 (2018).
26. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl Acad. Sci. USA* **104**, 8311–8315, https://doi.org/10.1073/pnas.0700329104 (2007).
27. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396, https://doi.org/10.1038/nrm3810 (2014).
28. Nussinov, R. & Tsai, C. J. Allostery in disease and in drug discovery. *Cell* **153**, 293–305, https://doi.org/10.1016/j.cell.2013.03.034 (2013).
29. Hsieh, M. C., Liang, C., Mehta, A. K., Lynn, D. G. & Grover, M. A. Multistep Conformation Selection in Amyloid Assembly. *J. Am. Chem. Soc.* **139**, 17007–17010, https://doi.org/10.1021/jacs.7b09362 (2017).
30. Bhattacharya, S., Xu, L. & Thompson, D. Revisiting the earliest signatures of amyloidogenesis: Roadmaps emerging from computational modeling and experiment. *Wiley Interdiscip. Reviews: Computational Mol. Sci.* **8**, e1359, https://doi.org/10.1002/wcms.1359 (2018).
31. Bhattacharya, S., Xu, L. & Thompson, D. Molecular simulations reveal terminal group mediated stabilization of helical conformers in both amyloid-β42 and α-synuclein. ACS chemical neuroscience (2019).
32. Crescenzi, O. *et al.* Solution structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain. *Eur. J. Biochem.* **269**, 5642–5648 (2002).
33. Ulmer, T. S., Bax, A., Cole, N. B. & Nussbaum, R. L. Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.* **280**, 9595–9603, https://doi.org/10.1074/jbc.M411805200 (2005).
34. Best, R. B., Zheng, W. & Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Computation* **10**, 5113–5124, https://doi.org/10.1021/ct500569b (2014).
35. Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. The. *J. Phys. Chem. B* **119**, 5113–5123, https://doi.org/10.1021/jp508971m (2015).
36. Rauscher, S. *et al.* Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **11**, 5513–5524, https://doi.org/10.1021/acs.jctc.5b00736 (2015).
37. Carballo-Pacheco, M. & Strodel, B. Comparison of force fields for Alzheimer's A β42: A case study for intrinsically disordered proteins. *Protein Sci.* **26**, 174–185, https://doi.org/10.1002/pro.3064 (2017).
38. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273, https://doi.org/10.1021/ct300400x (2012).
39. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616, https://doi.org/10.1021/jp973084f (1998).
40. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935, https://doi.org/10.1063/1.445869 (1983).
41. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–49, https://doi.org/10.1016/j.bpj.2011.03.051 (2011).
42. Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505, https://doi.org/10.1063/1.2121687 (2005).
43. Abedini, A. & Raleigh, D. P. A critical assessment of the role of helical intermediates in amyloid formation by natively unfolded proteins and polypeptides. *Protein Eng. Des. Sel.* **22**, 453–459, https://doi.org/10.1093/protein/gzp036 (2009).
44. Zangi, R., Kovacs, H., van Gunsteren, W. F., Johansson, J. & Mark, A. E. Free energy barrier estimation of unfolding the alpha-helical surfactant-associated polypeptide C. *Proteins* **43**, 395–402, https://doi.org/10.1002/prot.1052 (2001).
45. Petkova, A. T., Yau, W. M. & Tycko, R. Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry* **45**, 498–512, https://doi.org/10.1021/bi051952q (2006).
46. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
47. Pokhrel, R. *et al.* Domain rearrangement and denaturation in Ebola virus protein VP40. *AIP Adv.* **8**, 125129 (2018).
48. Fritsche, M., Pandey, R. B., Farmer, B. L. & Heermann, D. W. Conformational temperature-dependent behavior of a histone H2AX: a coarse-grained Monte Carlo approach via knowledge-based interaction potentials. *PLoS One* **7**, e32075, https://doi.org/10.1371/journal.pone.0032075 (2012).
49. Mirau, P., Farmer, B. L. & Pandey, R. B. Structural variation of alpha-synuclein with temperature by a coarse-grained approach with knowledge-based interactions. *AIP Adv.* **5**, 092504 (2015).
50. Jeffrey, G. A. An introduction to hydrogen bonding. Vol. 12 (Oxford university press New York, 1997).
51. Sun, W. & He, J. From isotropic to anisotropic side chain representations: comparison of three models for residue contact estimation. *PLoS One* **6**, e19238, https://doi.org/10.1371/journal.pone.0019238 (2011).
52. Salamanca Viloria, J., Allega, M. F., Lambrughi, M. & Papaleo, E. An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass. *Sci. Rep.* **7**, 2838, https://doi.org/10.1038/s41598-017-01498-6 (2017).
53. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696, https://doi.org/10.1093/bioinformatics/btl461 (2006).
54. Skjaerven, L., Yao, X. Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinforma.* **15**, 399, https://doi.org/10.1186/s12859-014-0399-6 (2014).
55. Xu, L., Nussinov, R. & Ma, B. Allosteric stabilization of the amyloid-β peptide hairpin by the fluctuating N-terminal. *Chem. Commun.* **52**, 1733–1736, https://doi.org/10.1039/c5cc08107f (2016).
56. Rodriguez, J. A. *et al.* Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature* **525**, 486–490, https://doi.org/10.1038/nature15368 (2015).
57. Arya, S. *et al.* Femtosecond Hydration Map of Intrinsically Disordered α -Synuclein. *Biophysical J.* **114**, 2540–2551, https://doi.org/10.1016/j.bpj.2018.04.028 (2018).
58. De Carufel, C. A., Quittot, N., Nguyen, P. T. & Bourgault, S. Delineating the Role of Helical Intermediates in Natively Unfolded Polypeptide Amyloid Assembly and Cytotoxicity. *Angew. Chem. Int. Ed.* **54**, 14383–14387, https://doi.org/10.1002/anie.201507092 (2015).
59. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl Acad. Sci. USA* **115**, E4758–E4766, https://doi.org/10.1073/pnas.1800690115 (2018).
60. Owen, M. C. *et al.* Effects of *in vivo* conditions on amyloid aggregation. *Chem. Soc. Rev.* **48**, 3946–3996, https://doi.org/10.1039/c8cs00034d (2019).
61. Shao, H., Jao, S., Ma, K. & Zagorski, M. G. Solution structures of micelle-bound amyloid beta-(1−40) and beta-(1−42) peptides of Alzheimer's disease. *J. Mol. Biol.* **285**, 755–773, https://doi.org/10.1006/jmbi.1998.2348 (1999).

62. Shabestari, M. H., Meeuwenoord, N. J., Filippov, D. V. & Huber, M. Interaction of the amyloid beta peptide with sodium dodecyl sulfate as a membrane-mimicking detergent. *J. Biol. Phys.* **42**, 299–315, https://doi.org/10.1007/s10867-016-9408-5 (2016).

63. Brown, A. M. & Bevan, D. R. Molecular Dynamics Simulations of Amyloid beta-Peptide (1-42): Tetramer Formation and Membrane Interactions. *Biophys. J.* **111**, 937–949, https://doi.org/10.1016/j.bpj.2016.08.001 (2016).

64. Niu, Z., Zhang, Z., Zhao, W. & Yang, J. Interactions between amyloid beta peptide and lipid membranes. Biochim Biophys Acta Biomembr, https://doi.org/10.1016/j.bbamem.2018.04.004 (2018).

65. Alza, N. P., Iglesias González, P. A., Conde, M. A., Uranga, R. M. & Salvador, G. A. Lipids at the Crossroad of α-Synuclein Function and Dysfunction: Biological and Pathological Implications. Frontiers in Cellular Neuroscience 13, https://doi.org/10.3389/fncel.2019.00175 (2019).

66. Lokappa, S. B. & Ulmer, T. S. Alpha-synuclein populates both elongated and broken helix states on small unilamellar vesicles. *J. Biol. Chem.* **286**, 21450–21457, https://doi.org/10.1074/jbc.M111.224055 (2011).

67. Pfefferkorn, C. M., Jiang, Z. & Lee, J. C. Biophysics of alpha-synuclein membrane interactions. *Biochim. Biophys. Acta* **1818**, 162–171, https://doi.org/10.1016/j.bbamem.2011.07.032 (2012).

68. Wu, K. P., Weinstock, D. S., Narayanan, C., Levy, R. M. & Baum, J. Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J. Mol. Biol.* **391**, 784–796, https://doi.org/10.1016/j.jmb.2009.06.063 (2009).

69. Khandogin, J. & Brooks, C. L. 3rd Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proc. Natl Acad. Sci. USA* **104**, 16880–16885, https://doi.org/10.1073/pnas.0703832104 (2007).

70. Chong, S. H. & Ham, S. Impact of chemical heterogeneity on protein self-assembly in water. *Proc. Natl. Acad. Sci. USA* **109**, 7636–7641, https://doi.org/10.1073/pnas.1120646109 (2012).

71. Ma, C. D., Wang, C., Acevedo-Velez, C., Gellman, S. H. & Abbott, N. L. Modulation of hydrophobic interactions by proximally immobilized ions. *Nature* **517**, 347–350, https://doi.org/10.1038/nature14018 (2015).

72. Huang, K., Gast, S., Ma, C. D., Abbott, N. L. & Szlufarska, I. Comparison between Free and Immobilized Ion Effects on Hydrophobic Interactions: A Molecular Dynamics Study. *J. Phys. Chem. B* **119**, 13152–13159, https://doi.org/10.1021/acs.jpcb.5b05220 (2015).

73. Uversky, V. N., Li, J. & Fink, A. L. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **276**, 10737–10744, https://doi.org/10.1074/jbc.M010907200 (2001).

74. Steckmann, T., Bhandari, Y. R., Chapagain, P. P. & Gerstman, B. S. Cooperative structural transitions in amyloid-like aggregation. *J. Chem. Phys.* **146**, 135103, https://doi.org/10.1063/1.4979516 (2017).

75. Leahy, C. T., Murphy, R. D., Hummer, G., Rosta, E. & Buchete, N. V. Coarse Master Equations for Binding Kinetics of Amyloid Peptide Dimers. *J. Phys. Chem. Lett.* **7**, 2676–2682, https://doi.org/10.1021/acs.jpclett.6b00518 (2016).

76. Lin, T.-W. *et al.* Alzheimer's amyloid-β A2T variant and its N-terminal peptides inhibit amyloid-β fibrillization and rescue the induced cytotoxicity. *PLoS one* **12**, e0174561 (2017).

77. Chakraborty, S. & Das, P. Emergence of Alternative Structures in Amyloid Beta 1−42 Monomeric Landscape by N-terminal Hexapeptide Amyloid Inhibitors. *Sci. Rep.* **7**, 9941, https://doi.org/10.1038/s41598-017-10212-5 (2017).

78. Sharma, B., Ranganathan, S. V. & Belfort, G. Weaker N-Terminal Interactions for the Protective over the Causative Aβ Peptide Dimer Mutants. *ACS Chem. Neurosci.* **9**, 1247–1253 (2018).

79. Murray, B., Sharma, B. & Belfort, G. (ACS Publications, 2017).

80. Kasahara, K., Fukuda, I. & Nakamura, H. A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer-DNA complex. *PLoS One* **9**, e112419, https://doi.org/10.1371/journal.pone.0112419 (2014).

81. Feher, V. A., Durrant, J. D., Van Wart, A. T. & Amaro, R. E. Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **25**, 98–103, https://doi.org/10.1016/j.sbi.2014.02.004 (2014).

82. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464, https://doi.org/10.1103/PhysRevLett.85.461 (2000).

83. Gc, J. B., Bhandari, Y. R., Gerstman, B. S. & Chapagain, P. P. Molecular dynamics investigations of the alpha-helix to beta-barrel conformational transformation in the RfaH transcription factor. *J. Phys. Chem. B* **118**, 5101–5108, https://doi.org/10.1021/jp502193v (2014).

84. Kamberaj, H. & van der Vaart, A. Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys. J.* **97**, 1747–1755, https://doi.org/10.1016/j.bpj.2009.07.019 (2009).

85. Qi, Y. & Im, W. Quantification of Drive-Response Relationships Between Residues During Protein Folding. J Chem Theory Comput 9, https://doi.org/10.1021/ct4002784 (2013).

86. Hacisuleyman, A. & Erman, B. Entropy transfer between residue pairs and allostery in Proteins: Quantifying allosteric communication in Ubiquitin. *PLoS computational Biol.* **13**, e1005319 (2017).

87. Williamson, J. A. & Miranker, A. D. Direct detection of transient alpha-helical states in islet amyloid polypeptide. *Protein Sci.* **16**, 110–117, https://doi.org/10.1110/ps.062486907 (2007).

88. Tomaselli, S. *et al.* The alpha-to-beta conformational transition of Alzheimer's Abeta-(1−42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of beta conformation seeding. *Chembiochem* 7, 257–267, https://doi.org/10.1002/cbic.200500223 (2006).

89. Rao, J. N., Jao, C. C., Hegde, B. G., Langen, R. & Ulmer, T. S. A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *J. Am. Chem. Soc.* **132**, 8657–8668, https://doi.org/10.1021/ja100646t (2010).

90. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Phys. Commun.* **91**, 43–56 (1995).

91. Yang, M., Huang, J. & MacKerell, A. D. Jr. Enhanced conformational sampling using replica exchange with concurrent solute scaling and hamiltonian biasing realized in one dimension. *J. Chem. Theory Comput.* **11**, 2855–2867, https://doi.org/10.1021/acs.jctc.5b00243 (2015).

92. Wang, L., Friesner, R. A. & Berne, B. J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438, https://doi.org/10.1021/jp204407d (2011).

93. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18, 1463-1472, https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H (1997).

94. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Computational Chem.* **13**, 952–962, https://doi.org/10.1002/jcc.540130805 (1992).

95. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).

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

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

98. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101, https://doi.org/10.1063/1.2408420 (2007).

99. Mercadante, D., Grater, F. & Daday, C. CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophys. J.* **114**, 1267–1273, https://doi.org/10.1016/j.bpj.2018.01.033 (2018).

100. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

101. Lange, O. F. & Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, Bioinforma.* **62**, 1053–1061 (2006).
102. Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl Acad. Sci. USA* **106**, 6620–6625, https://doi.org/10.1073/pnas.0810961106 (2009).
103. Ichiye, T. & Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205–217, https://doi.org/10.1002/prot.340110305 (1991).
104. Yen, J. Y. Finding the k shortest loopless paths in a network. *Manag. Sci.* **17**, 712–716 (1971).

Acknowledgements

Author contributions

S.B., L.X. and D.T. designed the research; S.B. performed the simulations and analysed the data; S.B. and D.T. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-64303-x.

Correspondence and requests for materials should be addressed to D.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.