



HAL
open science

Structures of Intrinsically Disordered Proteins in Aqueous Solution from Computer Simulations

Phuong H Nguyen, Philippe Derreumaux

► **To cite this version:**

Phuong H Nguyen, Philippe Derreumaux. Structures of Intrinsically Disordered Proteins in Aqueous Solution from Computer Simulations. Biophysical Chemistry, 2020. hal-03431142

HAL Id: hal-03431142

<https://hal.science/hal-03431142>

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structures of Intrinsically Disordered Proteins in Aqueous Solution from Computer Simulations

Phuong H. Nguyen,^{1,2} Philippe Derreumaux^{3,4*}

¹ CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005, Paris, France

² Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, PSL Research University, Paris, 75000 France

³ Laboratory of Theoretical Chemistry, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁴ Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Abstract. Intrinsically disordered proteins (IDPs) play many biological roles in the human proteome ranging from vesicular transport, signal transduction to neurodegenerative diseases. The A β and tau proteins, and the α -synuclein protein, key players in Alzheimer's and Parkinson diseases, respectively are fully disordered at the monomer level. The structural heterogeneity of the monomeric and oligomeric states and the high self-assembly propensity of these three IDPs have precluded experimental structural determination. Simulations have been used to determine the atomic structures of these IDPs. In this article, we review recent computer models to capture the equilibrium ensemble of long IDPs and notably of A β , tau and α -synuclein proteins at different association steps in aqueous solution and present new results of the PEP-FOLD framework on α -synuclein and monomer.

Key-words: amyloid, aggregation, computer simulations, intrinsically disordered proteins, aqueous solution, A β , tau and α -synuclein.

Highlights: * IDPs are involved in neurodegenerative diseases * Structure determination of the A β , tau and α -synuclein proteins is a challenge * Computer simulation results free or guided by experimental data * PEP-fold results on tau construct and α -synuclein monomers

1. Introduction

Intrinsically disordered proteins (IDPs) play key roles in many cellular processes, such as vesicular transport, signal transduction, and neurodegenerative diseases. While some IDPs have disordered and flexible regions important for protein-protein, protein-RNA and protein-DNA functions, others do not adopt a well-defined three-dimensional (3D) structure with a funnel-like free energy landscape [1]. Rather they have multiple distinct conformations at the monomer level and higher association steps, conformations being separated by free energy barriers of several to many times $k_B T$ at room temperature [2].

The inherent flexibility of IDPs associated with most degenerative diseases has prevented high-resolution structural determination by solution nuclear magnetic resonance (NMR) and X-ray diffraction experiments. Local information can be, however, obtained by chemical shifts, residual coupling constants, and J-couplings from NMR, and secondary structure from fast Fourier infrared spectroscopy (FTIR) or circular dichroism (CD). Long-range tertiary contacts can be deduced from single molecule Förster resonance energy transfer (FRET) or paramagnetic relaxation enhancement (PRE) NMR spectroscopy, and short-range distance contacts can be extracted by cross linked residues determined by mass spectrometry (MS). Low-resolution 3D information of monomers and oligomers can be obtained by dynamic light scattering (DLS), fluorescence correlation spectroscopy (FCS) or pulse field gradient NMR spectroscopy providing hydrodynamics radius, ion-mobility mass-spectrometry data (IM/MS), small-angle X-ray scattering (SAXS), atomic force microscopy (AFM), and cryo-electron microscopy (cryo-EM). Most of amyloid fibril models were mainly proposed based on solid-state NMR and cryo-EM experiments [2-5].

It is well established that, with aging, the plaques (fibrils) and the transient aggregates of the A β , tau and α -synuclein proteins are no longer degraded by the lysosomal and proteasomal machineries, and lead to neuronal death [6,7]. The A β protein consists of a charged hydrophilic N-terminus (residues 1-16), the central hydrophobic core (CHC, residues 17-21), a charged region (residues 22-29) and a hydrophobic C-terminus (residues 30-40/42) [8]. Full-length human tau (2N4R, 441 residues) has a charged N-terminal region, a proline-rich region PRR spanning the residues 208-244 followed by four repeats (labelled R1 to R4 spanning the residues 244-368) that under physiological conditions bind to the axonal microtubule, and a C-terminal region [9]. The α -synuclein consists of a N-terminal domain (residues 1-66) which binds to membranes, the hydrophobic nonamyloid component (NAC, residues 67-96), and a highly charged C-terminal

domain (residues 97-140) with many Asp and Glu residues [10].

In vitro, these three IDP's form cross- β structures with an aggregation kinetics profile displaying a sigmoidal curve where the proteins assemble into oligomers (lag-phase) prior to fibril elongation (growth phase), and a plateau where the fibrils and free monomers are in equilibrium (saturation phase) [2]. The aggregation kinetics and the life-times of the heterogeneous conformations of all species along the amyloid fibril formation pathways are very sensitive to three factors. (1) Amino acid length (e.g., A β 42 vs. A β 40), (2) genetic risks : mutations A2V, H6R, D7N, A21G, E22K, E22Q, E22G and D23N in A β [8], mutations A30P, E46K, H50Q, G51D and A53T in α -synuclein [11], and the pattern and the level of hyperphosphorylation in tau [12], and (3) experimental conditions such as pH, temperature, peptide concentration, external force applications resulting from agitation, electric field and shear forces, and the presence of membrane, metal ions and crowding [2,13-18].

At the monomer level, CD and NMR experiments reveal that A β 40 and A β 42 peptides are turn/coil with transient secondary structures in solution, [19] and both the htau441 and α -synuclein proteins are unstructured in solution and mammalian cells [3,10]. In vitro, A β 40 fibrils have an intra-peptide U-shape with two β -strands at residues 10-23 and 30-38. In contrast, A β 42 fibrils display an intra-peptide S-shape with three β -strands at residues 12-18, 24-33 and 36-40 with a disordered N-terminus. On the other hand, fibrils of AD-brain-derived A β 40 peptides have a distorted U-shape with a more ordered N-terminus [20]. All these A β fibril structures were determined by solid-state NMR experiments [5].

The fibrils of htau441 from the brain of an individual with AD [21] and of recombinant α -synuclein [11] were determined by cryo-EM at a resolution of 0.30-0.35 nm. The cores of tau fibrils consist of two identical protofilaments comprising only the R3-R4 domain (residues 306-378) with a cross- β / β -helix structure, while the rest of the protein is disordered. Note that paired helical and straight filaments differing in their inter-protofilament packing exist [21]. The core of the α -synuclein fibril consists of residues 37-99 adopting a Greek-key-like topology with high β -strand content, and the rest of the protein is highly flexible [11]. All together, these solid-state NMR and cryo-EM results demonstrate the diversity of topologies that can form amyloid fibrils with an intermolecular hydrogen-bond (H-bond) network parallel to the fibril axis.

High-resolution 3D oligomeric models are lacking experimentally [2]. Yet, A β oligomers up to

12-mers are believed to be the main culprits of AD, initiating a series of events leading to neuronal death (amyloid cascade hypothesis) [22], and soluble tau and α -synuclein oligomers are also the most toxic species [7]. It is important to note that albeit A β 40 and A β 42 peptides are the main component of senile plaques, there is no correlation between number of amyloid deposits and the severity of the disease.

Generating representative monomeric solvated structures of IDPs ranging from 40 (A β 40) to 441 (tau) amino acids free of or guided by available low-resolution local and global information is a challenge for computer simulations due to structural heterogeneity. The computational task becomes more difficult at an oligomer level. In all cases we face the problem of identifying the most relevant microstates and having a very accurate force field to represent intramolecular and intermolecular forces and the interactions between the solvent and the solute. Sections 2 and 3 review the current methods to capture the conformational ensemble of long IDPs associated with neurodegenerative diseases at various association steps in aqueous solution. We also present new results of the PEP-FOLD framework on monomeric α -synuclein.

2. Sampling the most relevant monomeric microstates of long IDPs

2A. Simulations free of experimental data

Atomistic molecular dynamics (MD) simulations in explicit environment offer the most detailed picture of protein folding. The longest trajectory on the fastest computer (Anton) reached 1 millisecond for a globular protein [23]. This time is sufficient for sampling the monomeric state of amyloid proteins, but is clearly insufficient for capturing all association-dissociation events during the lag phase, each event taking place with a timescale of hundreds of ns and ms [24,25]. In the past few years, there has been a constant effort in improving atomistic force fields and notably the interaction parameters between the solute and the solvent, and the torsional parameters of some side chains. The latest versions of CHARMM and AMBER, namely, CHARMM36m-TIP3P modified [26] and AMBERsb99-disp where disp is a variant of TIP4P water model [27], were designed to reproduce available experimental data on folded and disordered proteins, but other potentials have emerged.

Using Anton, the AMBER99sb-disp force field and 30 μ s MD simulation, α -synuclein was found mainly coil/turn with a fractional helix of 0.3 between residues 90 and 100 and a radius gyration, R_g , distribution displaying two peaks centered at 2.5 and 4.0 nm and a mean value at 3.7

vs. 3.1 ± 5 nm by one experimental technique [27]. Using CHARMM36m force field and the same MD simulation time, the mean Rg value is much smaller, 1.84 nm [27]. Using Anton α -synuclein trajectories, Mandaci et al. showed that the identification of epitope region varies substantially with the force field [28].

Based on 30 μ s MD, the monomeric A β 40 was found to have a mean Rg value of 1.39 nm with the new AMBER force field and 1.47 nm with the CHARMM36m force field vs. 1.20 ± 0.13 nm based on hydrodynamics radius, Rh, measurements. Amber99sb-disp4 simulations reveal fractional β -strand of 0.4 and 0.05 at the CHC and N-terminal residues of the C-terminal residues, and residues 10-13, respectively while CHARMM36 simulations reveal a fractional β -strand of 0.3 at the first two regions, and a peak of 0.15 at residues 4-5. [27] Computed free energies and interaction maps based on different protein and water force fields and a total of 36 μ s MD simulations showed that the flexible terminal groups, namely the N-terminus of monomeric A β 42 and C-terminus of monomeric α -synuclein, have a tendency to stabilize α -helices by hydrophobic interactions with the central hydrophobic domains, and then secondary salt bridges with other domains [29].

Chen et al. introduced a two-dimensional grid-based energy correction map (CMAP) potential into the OPLS/AA force field combined to the disp water potential to improve the performances on the torsional properties of well-folded and disordered peptides. By deriving a residue-specific force field, namely OPLSIDPSFF, they were able by MD simulations to show that the predicted chemical shifts and J-coupling of 11 monomeric IDPs ranging from 24 to 83 amino acids were in quantitative agreement with NMR observables. Applied to A β 40 and A β 42 monomers, OPLSIDPSFF still underestimates the β -sheet content with respect to the α -helix content [30]. In another study, the general purpose water OPC model and the AMBER ff99sb force field show promise for the A β 42 peptide though convergence of secondary structures likely requires a much longer MD simulation time than 10 μ s [31].

Using a hybrid-resolution model protein with atomic details in coarse-grained environment (PACE), Schulten et al. performed 180 μ s MD simulations to characterize the structural ensembles of wild-type (WT) and mutants monomeric α -synuclein in aqueous solution. The simulations reproduce well secondary structure content, long-range contacts, chemical shifts, and $^3J(\text{H}_\text{N}\text{H}_{\text{C}\alpha})$ -coupling constants of the WT protein. They also reveal that a short fragment encompassing the amino acids 38-53, adjacent to the NAC region, has a high probability to form a β -hairpin. The two disease-prone mutations, A30P and A53T, accelerate the formation of this β -hairpin, while the non-

amyloid G47V mutation decreases its probability, leading to the conclusion that this event in region 38-53 is the key driving force during α -synuclein aggregation. [32]

To explore longer IDPs sequences with amino acid specificity free of any local and global experimental and without an exponential increase of CPU time, one can resort to two generic approaches.

The first generic approach considers the full sequence at a time and is based on coarse-grained (CG) representations where each amino acid is represented by a varying number of particles for the backbone and each side-chain [33-37]. For instance, Ramis et al. performed μ s replica exchange with solute scaling (REST2) simulations of α -synuclein with the CG SIRAH model in water [38]. While REMD simulations involve exchange between neighboring replicas at consecutive temperatures and the total energy overlap of replicas scaling with the number of particles as $N^{-1/2}$, optimal exchange rate requires many MD trajectories in parallel. Potential energy rescaling and Hamiltonian replica exchange such as REST2 and H-REMD reduce the number of replicas four to five times, while preserving the number of unique conformational states. SIRAH, as shown in Figure 1A, is based on the positions of the backbone N, $C\alpha$, and O atoms and a CG side chain-dependent representation: hydrophobic residues (Val, Ile, Leu, and Met) are represented by one single bead, aromatic side chains are mapped to three (Phe, His, and Tyr) or five beads on a plane (Trp), and the beads of polar and charged side chains are mapped on charged groups or acceptors/donors of hydrogen bonds [37]. Water is represented by four molecules and beads. SIRAH REST2 simulations with an increase of 30% in the default protein-water interactions display a R_g histogram between 2.0 and 5.0 nm with a peak at 3.00 nm leading to a mean R_g of 3.22 ± 0.69 nm consistent with Rh experiment. Experimental SAXS profiles and NMR chemical shifts were also well reproduced. The α -synuclein protein is found mainly coil/turn ($73.0 \pm 6.8\%$) with some β -sheet structure ($26.8 \pm 6.8\%$) and a small amount of α -helix ($0.2 \pm 0.4\%$), consistent with CD in solution at physiological pH and room temperature (<2% α -helix, 30% β -sheet, and 68% random coil/turn). The most significant feature of the conformational ensemble is the greater number of contacts within the N-terminal domain than between the N-terminal and the other two domains, and the fact that the NAC domain has a propensity to form a β -hairpin loop with either the beginning of the C-terminus or the end of the N-terminus (less often).

Thirumalai et al. developed a self-organized polymer (SOP) CG model for IDPs (SOP-IDP)

based on a two-residue representation ($C\alpha$, one side-chain bead, Figure 1B) and an energy function consisting of a finitely extensible nonlinear potential for the bonds, an excluded potential between nonbonded interactions, a screened Debye-Hückel potential, and backbone - backbone (BB), backbone - side-chain (B-SC) and side-chain - side chain potentials (SC-SC). The initial values of the BB, BS and SC-SC energy parameters were optimized by using the SAXS data (R_g and $I(q)$) of three short IDPs (residue number, N , between 24 and 131) and the experimental R_g values of three long IDPs ($N > 202$). The structural characteristics of 12 monomeric IDPs between 24 and 273 residues and 8 monomeric tau constructs between 99 (K19 including the R1, R2 and R4 repeats) and 441 (2N4R) residues were determined by long underdamped Langevin dynamics ($5 \cdot 10^8$ steps) with a time step of 30 fs [39]. The simulations reproduced the experimental R_g , and the full scattering intensity profile obtained by small-angle X-ray scattering of all IDPs. R_g values of IDPs are well-described by the standard Flory scaling law, $R_g = 0.20 N^{\nu}$ (nm) with $\nu = 0.588$. For α -synuclein, the R_g distribution varies between 2.3 and 5 nm and shows a peak at 3.1 nm consistent with one experimental study and the SIRAH simulation. The configurational ensemble consists of a large population of compact structures (43.9%, mean R_g of 2.8 nm), fully extended (28.6%, mean R_g of 4.5 nm) and semi-extended (27.5%, mean R_g of 3.7 nm) conformations and reveals long-range interactions between the acidic-terminal region and the NAC domain, and tertiary contacts between the amino acids 74 and 94, consistent with electron transfer experiments [39].

The SOP-IDP model was also applied to study the differences in the free energies between the ground (G) and aggregation-prone (N^*) states of A β 40 and A β 42 monomers [40]. It is found that all the residues have a high propensity (>97%) to be turn/coil consistent with a few simulations [41], but not with most of them [2,42-45], and the free energy gap between the G and N^* states is smaller for A β 42 than A β 40. The SOP-IDP simulations also determined that the less stable N^* state of A β 2 coding for the U-shape fibril conformation should form earlier than the other N^* state coding for the S-shape fibril conformation [40]. The predicted 3D SOP-IDP structures also include a low population of helical and extended sub-states in agreement with other computational studies [44,46-47].

The second generic approach for long IDPs is hierarchical in nature and considers the assembly of fragments. Building on the "flexible-meccano" model [48], Grubmuller et al. developed a hierarchical chain-growth approach, where the structures of all fragments of the protein are exhaustively sampled by all-atom REMD simulations in TIP3P water. Then, they merged the structures of fragments overlapping along the sequence by eliminating the clashes to explore the

stereo-chemically possible global structures of IDPs [49]. For α -synuclein, 46 penta-peptide fragments were used, and the conformational ensemble reproduces the local structure probed by NMR J-couplings and chemical shifts, and predicts a global structure with a mean Rg value of 4.0 nm consistent with three SAXS values of 3.6, 4.0 and 4.5 nm, a difference between the measured values that can be related to different experimental conditions. All predicted models show little well-defined helical and β -strand structure differing therefore from CD experiment [49].

Building on the chain-growth PEP-FOLD framework extensively tested on well-folded protein [33,50-54] and protein-peptide structures [55], we present for the first time its application on α -synuclein. We recall that PEP-FOLD performs a rigid assembly of fragments of 4-residue length (described by structural alphabet letters) by superimposing the first three α -carbons of the new fragment onto the last three of the previously built structure, and start the building from any randomly chosen amino acid to add one residue at each side of the growing structure. At each position, the algorithm keeps 3000 conformations, the 1000 energetically best OPEP states and 2000 randomly selected ones in the pool of the remaining generated conformations. When the full-length CG structure is built, the conformation is subject to Monte Carlo simulations of 50000 steps at 300 K, and the PEP-FOLD procedure delivers an all-atom model. One step of Monte Carlo consists of generating a new conformation either randomly using simple motion rules such as the torsional change of one amino acid or guided by some torsional distribution of one or several residues. Then, the new conformation is accepted if the Metropolis criterion is satisfied. The OPEP (optimized potential for efficient peptide structure prediction) model consists of one bead per side-chain, the exception being proline where all-atoms are considered, and the N, H, C α , C and O atoms for the backbone (Figure 1C). A particularity of OPEP is that the potential energy function includes two-body and four-body hydrogen-bond interaction terms between the backbone amide bonds in addition to the SC-SC interactions, energy terms controlling the Ramachandran plot of each residue and does not include any electrostatic interactions. Note that albeit OPEP was designed on small proteins in solution [33,56-58], this CG force field coupled to various sampling approaches was able to identify key metastable states of oligomers of short amyloid fragments [59-63].

Using PEP-FOLD framework and a total of 600 simulations, the α -synuclein protein display 30% α -helix, 5% β -sheet and 65% turn/coil. Though the predicted secondary structure differs from one CD experiment (α -helix 3%, β -sheet $23 \pm 8\%$ and random coil $74 \pm 10\%$), there is a non-negligible β -strand signal at positions 38-50, adjacent to the NAC region as reported by PACE simulations [32] and in the NAC region (residues 67-74). The three structures of lowest OPEP

energies, obtained from three distinct simulations, have R_g values of 1.50, 1.80 and 1.53 nm and different α -helix locations (Figure 2A). Using the full ensemble of simulations, the R_g distribution varies between 1.5 and 2.5 nm with a peak at 1.7 nm (Figure 2B). Taking the PEP-FOLD structure of lowest-energy, 50 ns MD simulation was performed at 300 K and 140 mM NaCl using AMBER99sb-disp. The RMSD and R_g evolution as a function of time are shown in Figure 3A. This conformation is very dynamic with R_g fluctuating between 1.5 and 1.9 nm R_g . The final 50 ns structure is shown in Figure 3B. It is important to note that the R_g determined by different experimental techniques vary between 2.3 to 5 nm [10,38,64,65]. While our structures are more compact than those predicted by most simulations performed free or guided by most experimental data, they have however a compaction similar to those determined by Dokhloyan et al. using structural proteomics data as described below [66].

2B. Simulations guided by experimental data

Numerous simulations of long monomeric IDPs were performed using low-resolution experimental data. The α -synuclein conformation was determined from Monte Carlo simulations, a CG model with only repulsive Lennard-Jones interactions, and inter-residue distances obtained from single molecule fluorescence resonance energy transfer. The conformations at 0 M TMAO span a R_g distribution between 2.2 and 3.3 nm and agree well with independent measurements from FCS, NMR, PRE, and SAXS data. There are rare contacts between the C-terminal and the NAC and N-terminus [67].

Using the same CG model and FRET-derived data, Nath et al. reported that Monte Carlo simulations of α -synuclein at pH 3.0, a condition that is known to accelerate aggregation, show a slight compaction in terms of the mean R_g (3.2 ± 0.3 nm) relative to the neutral-pH ensemble [64]. However, the low-pH ensemble shows significant inter-residue distance variations, with a reduction of the number of contacts between the N-terminus and the central region, and an increase of the number of contacts between the C-terminus and much of the rest of the protein. Using the same protocol and FRET-derived data, they found that the mean R_g of monomeric tau441 increases from 5.1 ± 0.5 nm to 6.0 ± 0.6 nm upon the addition of heparin [64].

Dedmon et al. designed a PRE experiment on α -synuclein where cysteine mutations facilitating attachment of the spin label, MTSL, were made at positions Q24, S42, Q62, S87, and N103. PRE-derived distance restraints were incorporated into ensemble MD simulations using the

CHARMM force field and 20 replicas. The R_g probability distribution is broad varying between 1.8 and 4.2 nm and a mean R_g of 2.5 nm, matching the experimental value. The compaction of the structures relative to a random coil model results from long-range contacts between the residues 120-140 of the negatively charged C-terminal and the residues 30-100 [68].

Dokhloyan et al. determined the unstructured conformation of the tau441 and α -synuclein monomers by atomistic discontinuous MD (DMD) simulations guided by structural proteomics data resulting from cross-linkers (reagents) and MS experiments [66,69]. DMD simulations are event-driven, keep track of the positions and velocities of all particles at collision times, and require spherically symmetric potentials consisting of square wells between particles.

The predicted structures of tau441 monomer are consistent with surface modification and cross-linking data and consist of an ensemble of rather compact globular conformations with transient α -helices and β -strands and very different topologies, with the C-terminal and N-terminal regions confined on both sides on the core. Interestingly, the R3-R4 domain spanning residues 306-378 is found to have seven out of eight β -strands (exception being the β 2 strand spanning residues 313-322) of the fibril core reported by cryo-EM and this structure helps explain why phosphorylation of Ser356 is beneficial in opening the structure of the monomer [69]. Note that the DMD-generated conformational ensemble differs substantially from that obtained by atomistic REMD simulations of tau K18 (repeats R1 to R4) and K19 (three repeats R1, R3 and R4) monomers free of any experimental data [70].

The DMD-generated structures of α -synuclein are in agreement with hydrogen-deuterium exchange and distance crosslinking data. The protein consists of an ensemble of globular conformations with different topologies [66]. The NAC and C-terminal regions are consistently made of β strands and β -hairpins consistent with the SIRAH simulations, and the predicted secondary structure is consistent with CD experiment. The four lowest energy-cluster centroids, which represent 95% of the full conformational ensemble, and with a mean R_g of 1.4 nm are more compact than those based on PRE-NMR and FRET ensembles, SAXS data or SIRAH simulations and a compaction similar to the PEP-FOLD simulation results. Note that the higher values of the radius of gyration determined by these experimental techniques may be caused by an equilibrium between the monomeric and multimeric states of the synuclein protein under the conditions used, and SAXS data are very sensitive to the more-extended conformations of the ensemble. Based on their centroids, Dokhloyan et al. provide some insights into the impact of the familial-disease mutations

on α -synuclein aggregation [66].

3. Sampling the oligomeric microstates of long IDPs.

3A. A β in aqueous solution

The A β 42 dimer was subject to atomistic REMD using OPLS-AA, CHARMM22*, AMBER99sb-ildn and AMBERSb14 for a total of 144 μ s [71]. The configurations are predicted to be mainly turn/coil with the calculated cross-collision sections (CCSs), hydrodynamics radius, and SAXS profiles independent of the force field. However, the secondary, tertiary and quaternary conformations differ between the force fields. The α -helix content varies between 3 and 20%, and the population of the intramolecular β -hairpin spanning residues 17-21 and 30-36 varies between 1.5 and 13%. Interestingly, AFM-based force clamp was applied for dissociating A β 42 dimer and identified two transient states having lifetimes of 188 ± 52 and 317 ± 67 ms [24].

For the A β 40 dimer, Nguyen et al. performed REMD on the WT sequence, the A2V/A2V mutant which leads to early-onset AD and the WT/A2V mutant which protects from AD [72,73]. The secondary structure content remains constant at 18% β -strand and 10% α -helix, but the intramolecular and intermolecular conformations change upon single and double A2V mutation, with a high content of intrapeptide-stabilized conformations in heterozygous dimers. These dynamic dimers are also stabilized by interactions within the CHC regions and the C-terminal regions and interactions between the CHC and C-terminal regions consistent with long-time MD simulations [74]. Similar results were obtained for the simulations of A β 40 WT/A2T and A β 42 A2T/A2T dimers, A2T mutation being known to reduce A β aggregation kinetics and the risk of AD [75,76].

In all these dimer simulations, both the antiparallel orientation with β -hairpin formation and the perpendicular peptide orientation are preferred over the parallel U-shape organization. This result is confirmed by a Markov state model to characterize the transition pathways and associated kinetics of the formation of A β 40 dimers obtained by the PACE model and the adaptive sampling technique. The antiparallel β -sheet structures is found to form by direct binding of A β in β -hairpin conformation with a time scale of 200 μ s; while the parallel U-shape structure results of the association of A β monomers in arbitrary conformations with a time scale of 25.8 ms [77]. Intermediate A β dimers can be hypothesized from simulations of A β 16-35 where dimerization populates equally the antiparallel and parallel orientations of the CHC regions with β -sheet content, and populates the parallel orientation of the unstructured residues 31-35 [78]. Overall, this

preference for antiparallel A β dimers, discussed so many times in the early steps of aggregation [79,80], was recently evidenced to induce cognitive impairments with increased toxicity. Indeed, the two FDA-approved drugs, oxytetracycline and sunitinib, were found to dissociate A β dimers and plaques to monomers in 5XFAD transgenic mice [81].

The formation of A β tetramers was investigated by two computational studies starting from random conformations. Firstly, the multiscale simulation performed by Li et al. combining the CG UNRES force field and REMD simulations followed by atomistic MD simulations reveals that the A β 42 tetramer is characterized by coil/turn (78%), α -helix (3%) and β -sheet (18%) with high β -strand propensity of the residues 6-14, 16-19, 30-34 and 38-40. The five dominant clusters have CCS values consistent with IM/MS values. They do not any particular topology and are more compact than the fibrillar state due to the role of water [82]. Secondly, the predictive coarse-grained protein AWSEM force field, developed by Wolynes et al. and based on an associative memory term determined by 20 fragment memories for each segment with length 9, was coupled to MD and umbrella sampling simulations. The free energy landscapes at 40 μ M and 300 K lead essentially to U-shape intramolecular A β 40 and A β 42 conformations with perpendicular orientation of double layers of β -sheets, disordered orientations of β -sheets and cylindrin-shaped structures with variations from perfection [83]. There is therefore a significant difference between the conformational ensembles obtained by Wolynes and Li. Notably, in the AWSEM simulations extended to octamers, the A β 42 peptide is found to accelerate much more the conversion from trimers to fibrillar tetramers, providing one explanation of the enhanced aggregation kinetics of A β 42 relative to A β 40.

It is interesting that a well-defined tetrameric β -barrel consisting of two distinct β -hairpins and asymmetric arrangement of eight antiparallel β -strands was found to exist transiently for A β 42 and not for A β 40 in aqueous solution using extensive REMD simulations and the two last atomistic force fields developed for IDPs, namely the CHARMM36m-TIP3P modified and AMBER99sb-disp models [84]. Forming such a precise and ideal β -barrel is clearly unfeasible from ab initio or de novo atomistic simulations as it is a rare event requiring minutes to hours to fold into vesicles. Though out of the scope of this chapter, this tetrameric β -barrel has a much higher probability in a membrane mimicking environment for A β 42WT, A β 42 A2T and A β 42 D23N by using REMD simulations [85,86].

Finally, the early steps of the assembly of 20 A β disordered peptides were investigated by atomistic MD simulations in implicit solvent on the microsecond time scale followed by transition networks analysis. The simulations showed that the A β 40 and A β 42 aggregation pathways depend on oligomer shapes with compact and extended configurations, and A β 40 primarily populates dimers, trimers and tetramers vs. dimers, tetramers and hexamers for A β 42 [87]. We must recall, however, that implicit solvent models neglect hydrodynamics interactions known to play significant effects in modelling the dynamics of amyloids [88-90]. Based on simulations of A β 16-22 peptides, the free energy surface and the kinetics of association and dissociation of the dimer is very sensitive to the force field [91,92] and the hexamerization kinetics generated by four force fields is found to differ more from each other than the kinetics between amyloid and non-amyloid peptides simulated using a single force field [93]. Lastly, it has been demonstrated by surface enhanced Raman scattering (SERS) and computer simulations that oligomeric aggregates exist with an alternate distribution of partially ordered and disordered monomers, providing therefore evidence of the considerable conformational heterogeneity of the transient aggregates [94].

3B. Tau constructs and α -synuclein oligomers in aqueous solution

In comparison to the A β protein, the number of computational studies on α -synuclein and long tau constructs oligomers is fairly small.

The dimeric structure of α -synuclein with high helical content was investigated by MD simulations with the GROMOS53A6 and SPC water force fields starting from eight different interfaces between the chains. Zhang et al. found that the dimers produce antiparallel β -hairpin motifs, adjacent to the NAC region, and some intermolecular β -sheets with out-of-register of H-bonds compared to the fibril structure [95]. The structural dynamics of α -synuclein dimers was also revealed by single-molecule time-lapse high-speed AFM and DMD simulations. The study highlights three types of dimers with distinct interfaces. In one case, one chain is extended and the other chain is compact [96].

Wei et al. performed MD simulations on the Greek-key core (residues 44-96) starting from the fibril cryo-EM configuration using the AMBER99SB-ILDN force field. Based on multiple trajectories of 300 ns, they found that the trimer is the critical nucleus for α -synuclein44-96 fibril formation and the tetramer is the minimal stable nucleus [97]. In the same line, Xu et al. studied the stability of the Greek-key core fibril consisting of five monomers spanning residues 20-110 using the

CHARMM27 force field with CMAP corrections and MD simulations of 200 ns of the WT sequence and the five familial mutations. Their results show that the mutations do not modify the overall fibril, but shift their relative stabilities [98]. In these two studies, it remains to be determined whether longer simulation times and the full sequence impact the results.

Because α -synuclein can coexist as a tetramer and a monomer experimentally and the tetramer plays an essential role in maintaining α S homeostasis, simulations were performed on the monomer and tetramer. The simulations shed light on a fundamental relationship between monomers and tetramers, and the key residues involved in mediating formation of a tetramer [99].

Finally, looking at oligomers of long tau constructs, and apart from MD simulations examining the cross-seeding and conformational selection between 3R and 4R human tau proteins [100], and the stability of C-shaped octamers of the R3-R4 and R1-R2 domains [101], we performed atomistic REMD simulations on the tau R3-R4 domain starting from the fibril C-shape. It is found that the WT tau R3-R4 dimer explores elongated, U-shaped, V-shaped and globular topologies. Phosphorylation of Ser356, pSer356, is known to block the interaction between the tau protein and the A β 42 peptide. MD simulations of pSer356 for a total of 5 μ s compared to its WT counterpart reveal a variation of the population of accessible topologies, and a decrease of intermediates near the fibril like conformers [9]. Our computational result is consistent with in vitro experiments reporting that WT K18 develops seeding activity more rapidly than pSer356 K18 [102].

4. Conclusions

We have reviewed recent results of computer simulations aimed at determining the structure of the A β , tau, and α -synuclein protein in aqueous solution at different association steps. Significant progress has been made using more sophisticated force fields, atomistic and coarse-grained models coupled to much efficient sampling approaches and longer simulation times at the monomer level. We also have presented for the first time the application of the PEP-FOLD framework for the α -synuclein monomer.

At the oligomer level, the exception being the A β protein, the simulation results are very scarce for the full-length tau and α -synuclein and this represents a real challenge for molecular modelling approaches in aqueous solution [103]. Developing simulations able to explore the most

relevant microstates of monomers and oligomers by treating in vivo conditions, that is in the presence of membranes, crowding and confinement, is becoming urgent and is a critical issue to cure Alzheimer's and Parkinson diseases [104-109].

AUTHOR INFORMATION

Corresponding Author

E-mail: philippe.derreumaux@tdtu.edu.vn

ORCID

Phuong H. Nguyen: 0000-0003-1284-967X

Philippe Derreumaux: 0000-0001-9110-5585

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge support by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01, and "CACISICE", ANR-11-EQPX-0008). PD thanks Université de Paris, CNRS and PSL.

REFERENCES

- (1) Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* 2016, 44, 1185–1200.
- (2) Nasica-Labouze J, Nguyen PH, Sterpone F et al. Amyloid beta protein and Alzheimer's disease: when computer simulations complement experimental studies. *Chem Rev* 2015, 115, 3518–3563.
- (3) Theillet FX, Binolki A, Bekei B, Martorana A, Rose HM, Stuijver M, Verzini S, Lorenz D, van Rossum M, Goldfarb D, Selenko P. Structural disorder of alpha-synuclein persists in mammalian cells. *Nature* 2016, 530: 45-50.

- (4) Metskas LA, Rhoades E. Single-Molecule FRET of Intrinsically Disordered Proteins. *Annu Rev Phys Chem.* 2020, 71:391-414.
- (5) Kreuzer AG, Nowick JS. Elucidating the Structures of Amyloid Oligomers with Macrocyclic beta-Hairpin peptides: Insights into Alzheimer's Disease and Other Amyloid Diseases. *Acc. Chem. Res* 2018, 51: 706-718.
- (6) Cliffe R, Sang JC, Kundel F, Finley D, Klenerman D, Ye Y. Filamentous Aggregates are Fragmented by the Proteasome Holoenzyme. *Cell Rep.* 2019, 26: 2140-2149.e3.
- (7) Deger JM, Gerson JE, Kaye R. The interrelationship of proteasome impairment and oligomeric intermediates in neurodegeneration. *Aging cell.* 2015, 14: 715-724.
- (8) Viet MH, Nguyen PH, Ngo ST et al. Effect of the Tottori familial disease mutation (D7N) on the monomers and dimers of A β 40 and A β 42. *ACS Chem Neurosci* 2013, 4:1446-1457.
- (9) Derreumaux P, Man VH, Wang J, Nguyen PH. Tau R3-R4 Domain Dimer of the Wild Type and Phosphorylated Ser356 Sequences. I. In Solution by Atomistic Simulations. *J Phys Chem B.* 2020, 124(15):2975-2983.
- (10) Schwalbe M, Ozenne V, Bibow S, Jaremko M, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, Zweckstetter M, Blackledge M. Predictive atomic resolution descriptions of intrinsically disordered hTau40 and alpha-synuclein in solution from NMR and small angle scattering. *Structure* 2014, 22: 238-249.
- (11) Li Y, Zhao C, Luo F et al. Amyloid fibril structure of α -synuclein determined by cryo-electron microscopy. *Cell Res* 2018, 28:897-903.
- (12) Fichou, Y.; Al-Hilaly, Y.K.; Devred, F.; Smet-Nocca, C.; Tsvetkov, P.O. ; Verelst, J. ; Winderickx, J.; Geukens, N. ; Vanmechelen, E. ; Perrotin, A. et al. The Elusive Tau Molecular Structures: Can We Translate the Recent Breakthroughs into New Targets for Intervention? *Acta Neuropathol Commun.* 2019, 7, 31-47.
- (13) Owen MC, Gnutt D, Gao M et al. Effects of in vivo conditions on amyloid aggregation. *Chem Soc Rev* 2019, 48:3946-3996.
- (14) Lu Y, Shi XF, Salisbury FR Jr, Derreumaux P. Small static electric field strength promotes aggregation-prone structures in amyloid- β (29-42). *J Chem Phys.* 2017, 146(14): 145101.
- (15) Xu W, Zhang C, Derreumaux P, Gräslund A, Morozova-Roche L, Mu Y. Intrinsic determinants of A β (12-24) pH-dependent self-assembly revealed by combined computational and experimental studies. *PLoS One.* 2011;6(9):e24329.
- (16) Jang H, Arce FT, Ramachandran S et al. Disordered amyloidogenic peptides may insert into the membrane and assemble into common cyclic structural motifs. *Chem Soc Rev* 2014, 43: 6750-6764.

- (17) Kinoshita M, Lin Y, Dai I, Okumura M, Markova N, Ladbury JE, Sterpone F, Lee YH. Energy landscape of polymorphic amyloid generation of β 2-microglobulin revealed by calorimetry. *Chem Commun (Camb)*. 2018, 54(57):7995-7998.
- (18) Dobson J, Kumar A, Willis LF, et al. Inducing protein aggregation by extensional flow. *Proc Natl Acad Sci U S A*. 2017;114(18):4673-4678.
- (19) Roche J, Shen Y, Lee JH et al. Monomeric A β (1-40) and A β (1-42) Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* 2016, 55:762-765.
- (20) Qiang W, Yau WM, Lu JX, Collinge J, Tycko R. Structural variation in amyloid- β fibrils from Alzheimer's disease clinical subtypes. *Nature*. 2017;541(7636):217-221.
- (21) Fitzpatrick AWP, Falcon B, He S et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* 2017, 547:185-190.
- (22) Selkoe DG, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med*. 2016, 8:595-608.
- (23) Lindorff-Larsen K, Maragakis P, Piana S et al. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J Phys Chem B*. 2016, 120:8313-820.
- (24) Maity S, Lyubchenko YL. Force clamp approach for characterization of nano-assembly in amyloid beta 42 dimer. *Nanoscale*. 2019;11(25):12259-12265.
- (25) Michaels TCT, Šarić A, Curk S, et al. Dynamics of oligomer populations formed during the aggregation of Alzheimer's A β 42 peptide. *Nat Chem*. 2020;12(5):445-451.
- (26) Huang J, Rauscher S, Nawrocki G, Ran T et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2017;14:71-73.
- (27) Robustelli P, Piana S, Shaw DE Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci U S A*. 2018;115:E4758-E4766.
- (28) Mandaci SY, Caliskan M, Sariaslan MF, Uversky VN, Coskuner-Weber O. Epitope region identification challenges of intrinsically disordered proteins in neurodegenerative diseases: Secondary structure dependence of α -synuclein on simulation techniques and force field parameters. *Chem Biol Drug Des*. 2020;10.1111/cbdd.13662. doi:10.1111/cbdd.13662
- (29) Bhattacharya S, Xu L, Thompson D. Molecular Simulations Reveal Terminal Group Mediated Stabilization of Helical Conformers in Both Amyloid- β 42 and α -Synuclein. *ACS Chem Neurosci*. 2019;10(6):2830-2842.
- (30) Song D, Liu H, Luo R, Chen HF. Environment-Specific Force Field for Intrinsically Disordered and Ordered Proteins. *J Chem Inf Model*. 2020;60(4):2257-2267.

- (31)Shabane PS, Izadi S, Onufriev AV. General Purpose Water Model Can Improve Atomistic Simulations of Intrinsically Disordered Proteins. *J Chem Theory Comput.* 2019;15(4):2620-2634.
- (32)Yu H, Han W, Ma W, Schulten K. Transient β -hairpin formation in α -synuclein monomer revealed by coarse-grained molecular dynamics simulation. *J Chem Phys.* 2015;143(24):243142.
- (33)Sterpone F, Melchionna S, Tuffery P. et al. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA Systems. *Chem. Soc. Rev* 2014; 43: 4871-4893.
- (34)Rojas AV, Maisuradze GG, Scheraga HA. Dependence of the Formation of Tau and A β Peptide Mixed Aggregates on the Secondary Structure of the N-Terminal Region of A β . *J Phys Chem B.* 2018; 122 : 7049-7056.
- (35)Bunce SJ, Wang Y, Stewart KL et al. Molecular insights into the surface-catalyzed secondary nucleation of amyloid- β_{40} (A β_{40}) by the peptide fragment A β_{16-22} . *Sci Adv.*2019 ; 5:eaav8216.
- (36)Urbanc B, Betnel M, Cruz, L et al. Elucidation of amyloid β -protein oligomerization mechanisms: discrete molecular dynamics study. *J. Am. Chem. Soc.* 2010; 132: 4266-4280.
- (37)Darré L, Machado MR, Brandner AF, González HC, Ferreira S, Pantano S. SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J Chem Theory Comput.* 2015;11(2):723-739.
- (38)Ramis R, Ortega-Castro J, Casasnovas R, et al. A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein α -Synuclein. *J Chem Inf Model.* 2019;59(4):1458-1471.
- (39)Baul U, Chakraborty D, Mugnai ML, Straub JE, Thirumalai D. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. *J Phys Chem B.* 2019;123(16):3462-3474.
- (40)Chakraborty D, Straub JE, Thirumalai D. Differences in the free energies between the excited states of A β_{40} and A β_{42} monomers encode their distinct aggregation propensities. *bioRxiv* 2020, 2020.02.09.940676.
- (41)Meng F, Bellaiche MMJ, Kim J-Y, Zerze GH, Best RB, Chung H.S. Highly Disordered Amyloid- β Monomer Probed by Single-Molecule FRET and MD Simulation. *Biophys J* 2018, 114, 870-884.
- (42)Rosenman DJ, Wang C, García AE. Characterization of A β Monomers through the Convergence of Ensemble Properties among Simulations with Multiple Force Fields. *J Phys Chem B.* 2016;120(2):259-277.
- (43)Côté S, Derreumaux P, Mousseau N. Distinct Morphologies for Amyloid Beta Protein Monomer: A β_{1-40} , A β_{1-42} , and A β_{1-40} (D23N). *J Chem Theory Comput.* 2011;7(8):2584-2592.
- (44)Granata D, Baftizadeh F, Habchi J et al. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci Rep* 2015; 5:15449.

- (45) Daggett V. Alpha-sheet: The toxic conformer in amyloid diseases?. *Acc Chem Res.* 2006;39(9):594-602.
- (46) Bhattacharya S, Xu L, Thompson D. Long-range Regulation of Partially Folded Amyloidogenic Peptides. *Sci Rep.* 2020;10(1):7597.
- (47) Nguyen PH, Tarus B, Derreumaux P. Familial Alzheimer A2 V mutation reduces the intrinsic disorder and completely changes the free energy landscape of the A β 1-28 monomer. *J Phys Chem B.* 2014;118(2):501-510.
- (48) Ozenne V, Bauer F, Salmon L, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 2012;28(11):1463-1470.
- (49) Pietrek LM, Stelzl LS, Hummer G. Hierarchical Ensembles of Intrinsically Disordered Proteins at Atomic Resolution in Molecular Dynamics Simulations. *J Chem Theory Comput.* 2020;16(1):725-737.
- (50) Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* 2009;37(Web Server issue):W498-W503.
- (51) Maupetit J, Derreumaux P, Tufféry P. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J Comput Chem.* 2010;31(4):726-738.
- (52) Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* 2012;40(Web Server issue):W288-W293.
- (53) Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J Chem Theory Comput.* 2014;10(10):4745-4758.
- (54) Sutherland GA, Grayson KJ, Adams NBP, et al. Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase with folding variants of a de novo-designed heme protein. *J Biol Chem.* 2018;293(18):6672-6681.
- (55) Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 2016;44(W1):W449-W454.
- (56) Derreumaux P. Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations. *Phys Rev Lett* 2001;85:206-209.
- (57) Wei G, Mousseau N, Derreumaux P. Complex folding pathways in a simple beta-hairpin. *Proteins.* 2004;56(3):464-474.
- (58) Sterpone F, Nguyen PH, Kalimeri M, Derreumaux P. Importance of the ion-pair interactions in the OPEP coarse-grained force field: parametrization and validation. *J Chem Theory Comput.* 2013;9(10):4574-4584.
- (59) Melquiond A, Boucher G, Mousseau N, Derreumaux P. Following the aggregation of amyloid-forming peptides by computer simulations. *J Chem Phys.* 2005;122(17):174904.
- (60) Melquiond A, Mousseau N, Derreumaux P. Structures of soluble amyloid oligomers from computer simulations. *Proteins.* 2006;65(1):180-191.
- (61) Santini S, Wei G, Mousseau N, Derreumaux P. Pathway complexity of Alzheimer's beta-amyloid A β 16-22 peptide assembly. *Structure.* 2004;12(7):1245-1255.
- (62) Song W, Wei G, Mousseau N, Derreumaux P. Self-assembly of the beta2-microglobulin NHVTL β peptide using a coarse-grained protein model reveals a beta-barrel species. *J Phys Chem B.* 2008;112(14):4410-4418.

- (63) Nasica-Labouze J, Meli M, Derreumaux P, Colombo G, Mousseau N. A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *PLoS Comput Biol*. 2011;7(5):e1002051.
- (64) Nath A, Sammalkorpi M, DeWitt DC, et al. The conformational ensembles of α -synuclein and tau: combining single-molecule FRET and simulations. *Biophys J*. 2012;103(9):1940-1949.
- (65) Dibenedetto D, Rossetti G, Caliandro R, Carloni P. A molecular dynamics simulation-based interpretation of nuclear magnetic resonance multidimensional heteronuclear spectra of α -synuclein-dopamine adducts. *Biochemistry*. 2013;52(38):6672-6683.
- (66) Brodie NI, Popov KI, Petrotchenko EV, Dokholyan NV, Borchers CH. Conformational ensemble of native α -synuclein in solution as determined by short-distance crosslinking constraint-guided discrete molecular dynamics simulations. *PLoS Comput Biol*. 2019;15(3):e1006859.
- (67) Ferrie JJ, Haney CM, Yoon J, et al. Using a FRET Library with Multiple Probe Pairs To Drive Monte Carlo Simulations of α -Synuclein. *Biophys J*. 2018;114(1):53-64.
- (68) Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc*. 2005;127(2):476-477.
- (69) Popov KI, Makepeace KAT, Petrotchenko EV, Dokholyan NV, Borchers CH. Insight into the Structure of the "Unstructured" Tau Protein. *Structure*. 2019;27(11):1710-1715.e4
- (70) Luo Y, Ma B, Nussinov R, Wei G. Structural Insight into Tau Protein's Paradox of Intrinsically Disordered Behavior, Self-Acetylation Activity, and Aggregation. *J Phys Chem Lett*. 2014;5(17):3026-3031.
- (71) Man VH, Nguyen PH, Derreumaux P. High-Resolution Structures of the Amyloid- β 1-42 Dimers from the Comparison of Four Atomistic Force Fields. *J Phys Chem B* 2017, 121: 5977 – 5987.
- (72) Tarus B, Tran TT, Nasica-Labouze J. et al. Structures of the Alzheimer's wild-type A β 1-40 dimer from atomistic simulations. *J Phys Chem B*. 2015, 119 : 10478-10487.
- (73) Nguyen PH, Sterpone F, Campanera JM et al. Impact of the A2V mutation on the heterozygous and homozygous A β 1-40 dimer structures from atomistic simulations. *ACS Chem Neurosci* 2016, 7: 823-832.
- (74) Zhang Y, Hashemi M, Lv Z, Lyubchenko YL. Self-assembly of the full-length amyloid A β 42 protein in dimers. *Nanoscale*. 2016;8(45):18928-18937.
- (75) Nguyen PH, Sterpone F, Pouplana R, Derreumaux P, Campanera JM. Dimerization Mechanism of Alzheimer A β ₄₀ Peptides: The High Content of Intra-peptide-Stabilized Conformations in A2V and A2T Heterozygous Dimers Retards Amyloid Fibril Formation. *J Phys Chem B*. 2016;120(47):12111-12126.
- (76) Das P, Chacko AR, Belfort G. Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters A β ₄₂ Dimer Structure. *ACS Chem Neurosci*. 2017;8(3):606-618.
- (77) Cao Y, Jiang X, Han W. Self-Assembly Pathways of β -Sheet-Rich Amyloid- β (1-40) Dimers: Markov State Model Analysis on Millisecond Hybrid-Resolution Simulations. *J Chem Theory Comput*. 2017;13(11):5731-5744.
- (78) Chebaro Y, Mousseau N, Derreumaux P. Structures and thermodynamics of Alzheimer's amyloid-beta A β (16-35) monomer and dimer by replica exchange molecular dynamics simulations: implication for full-length A β fibrillation. *J Phys Chem B*. 2009;113(21):7668-7675.

- (79) Chebaro Y, Jiang P, Zang T, et al. Structures of A β 17-42 trimers in isolation and with five small-molecule drugs using a hierarchical computational procedure. *J Phys Chem B*. 2012;116(29):8412-8422.
- (80) Man VH, Nguyen PH, Derreumaux P. Conformational Ensembles of the Wild-Type and S8C A β 1-42 Dimers. *J Phys Chem B*. 2017;121(11):2434-2442.
- (81) Lee JC, Kim HY, Lee S, et al. Discovery of Chemicals to Either Clear or Indicate Amyloid Aggregates by Targeting Memory-Impairing Anti-Parallel A β Dimers. *Angew Chem Int Ed Engl*. 2020;10.1002/anie.202002574.
- (82) Nguyen HL, Krupa P, Hai NM, Linh HQ, Li MS. Structure and Physicochemical Properties of the A β 42 Tetramer: Multiscale Molecular Dynamics Simulations. *J Phys Chem B*. 2019;123(34):7253-7269.
- (83) Zheng W, Tsai MY, Wolynes PG. Comparing the Aggregation Free Energy Landscapes of Amyloid Beta(1-42) and Amyloid Beta(1-40). *J Am Chem Soc*. 2017;139(46):16666-16676
- (84) Nguyen PH, Campanera JM, Ngo ST, Loquet A, Derreumaux P. Tetrameric A β 40 and A β 42 β -Barrel Structures by Extensive Atomistic Simulations. II. In Aqueous Solution. *J Phys Chem B*. 2019;123(31):6750-6756.
- (85) Nguyen PH, Campanera JM, Ngo ST, Loquet A, Derreumaux P. Tetrameric A β 40 and A β 42 β -Barrel Structures by Extensive Atomistic Simulations. I. In a Bilayer Mimicking a Neuronal Membrane. *J Phys Chem B*. 2019;123(17):3643-3648.
- (86) Ngo ST, Nguyen PH, Derreumaux P. Impact of A2T and D23N Mutations on Tetrameric A β 42 Barrel within a Dipalmitoylphosphatidylcholine Lipid Bilayer Membrane by Replica Exchange Molecular Dynamics. *J Phys Chem B*. 2020;124(7):1175-1182.
- (87) Barz B, Liao Q, Strodel B. Pathways of Amyloid- β Aggregation Depend on Oligomer Shape. *J Am Chem Soc*. 2018;140(1):319-327.
- (88) Sterpone F, Derreumaux P, Melchionna S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *J Chem Theory Comput*. 2015;11(4):1843-1853.
- (89) Chiricotto M, Melchionna S, Derreumaux P, Sterpone F. Hydrodynamic effects on β -amyloid (16-22) peptide aggregation. *J Chem Phys*. 2016;145(3):035102.
- (90) Chiricotto M, Melchionna S, Derreumaux P, Sterpone F. Multiscale Aggregation of the Amyloid A β ₁₆₋₂₂ Peptide: From Disordered Coagulation and Lateral Branching to Amorphous Prefibrils. *J Phys Chem Lett*. 2019;10(7):1594-1599.
- (91) Nguyen PH, Li MS, Derreumaux P. Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the A β (16-22) dimer and trimer. *Phys Chem Chem Phys*. 2011;13(20):9778-9788.
- (92) Man VH, He X, Derreumaux P, et al. Effects of All-Atom Molecular Mechanics Force Fields on Amyloid Peptide Assembly: The Case of A β ₁₆₋₂₂ Dimer. *J Chem Theory Comput*. 2019;15(2):1440-1452.
- (93) Carballo-Pacheco M, Ismail AE, Strodel B. On the Applicability of Force Fields To Study the Aggregation of Amyloidogenic Peptides Using Molecular Dynamics Simulations. *J Chem Theory Comput*. 2018;14(11):6063-6075.
- (94) La Rosa C, Condorelli M, Compagnini G, et al. Symmetry-breaking transitions in the early steps of protein self-assembly. *Eur Biophys J*. 2020;49(2):175-191.
- (95) Zhang T, Tian Y, Li Z, et al. Molecular Dynamics Study to Investigate the Dimeric Structure of the Full-Length α -Synuclein in Aqueous Solution. *J Chem Inf Model*. 2017;57(9):2281-2293.

- (96)Zhang Y, Hashemi M, Lv Z, et al. High-speed atomic force microscopy reveals structural dynamics of α -synuclein monomers and dimers. *J Chem Phys.* 2018;148(12):123322.
- (97)Zou Y , Qian Z , Gong Y , Tang Y , Wei G , Zhang Q . Critical nucleus of Greek-key-like core of α -synuclein protofibril and its disruption by dopamine and norepinephrine. *Phys Chem Chem Phys.* 2019;22(1):203-211.
- (98)Xu L, Ma B, Nussinov R, Thompson D. Familial Mutations May Switch Conformational Preferences in α -Synuclein Fibrils [published correction appears in *ACS Chem Neurosci.* 2018 Jul 18;9(7):1866-1867]. *ACS Chem Neurosci.* 2017;8(4):837-849
- (99)Cote Y, Delarue P, Scheraga HA, Senet P, Maisuradze GG. From a Highly Disordered to a Metastable State: Uncovering Insights of α -Synuclein. *ACS Chem Neurosci.* 2018;9(5):1051-1065.
- (100)Yu X, Luo Y, Dinkel P, et al. Cross-seeding and conformational selection between three- and four-repeat human Tau proteins. *J Biol Chem.* 2012;287(18):14950-14959.
- (101)Li X , Dong X , Wei G , Margittai M , Nussinov R , Ma B . The distinct structural preferences of tau protein repeat domains. *Chem Commun (Camb).* 2018;54(45):5700-5703.
- (102)Haj-Yahya M, Gopinath P, Rajasekhar K, Mirbaha H, Diamond MI, Lashuel HA. Site-Specific Hyperphosphorylation Inhibits, Rather than Promotes, Tau Fibrillization, Seeding Capacity, and Its Microtubule Binding. *Angew Chem Int Ed Engl.* 2020;59(10):4059-4067.
- (103)Tuffery P, Derreumaux P. Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface.* 2012;9(66):20-33.
- (104)Doig AJ, Derreumaux P. Inhibition of protein aggregation and amyloid formation by small molecules. *Curr Opin Struct Biol.* 2015;30:50-56.
- (105)Doig AJ, Del Castillo-Frias MP, Berthoumieu O, et al. Why Is Research on Amyloid- β Failing to Give New Drugs for Alzheimer's Disease?. *ACS Chem Neurosci.* 2017;8(7):1435-1437.
- (106)Hoang Viet M, Derreumaux P, Nguyen PH. Nonequilibrium all-atom molecular dynamics simulation of the bubble cavitation and application to dissociate amyloid fibrils. *J Chem Phys.* 2016;145(17):174113.
- (107)Nguyen P, Derreumaux P. Understanding amyloid fibril nucleation and $\alpha\beta$ oligomer/drug interactions from computer simulations. *Acc Chem Res.* 2014;47(2):603-611.
- (108)Pujols J, Peña-Díaz S, Pallarès I, Ventura S. Chemical Chaperones as Novel Drugs for Parkinson's Disease. *Trends Mol Med.* 2020;26(4):408-421.
- (109)Mahul-Mellier AL, Burtscher J, Maharjan N, et al. The process of Lewy body formation, rather than simply α -synuclein fibrillization, is one of the major drivers of neurodegeneration. *Proc Natl Acad Sci U S A.* 2020;117(9):4971-4982.

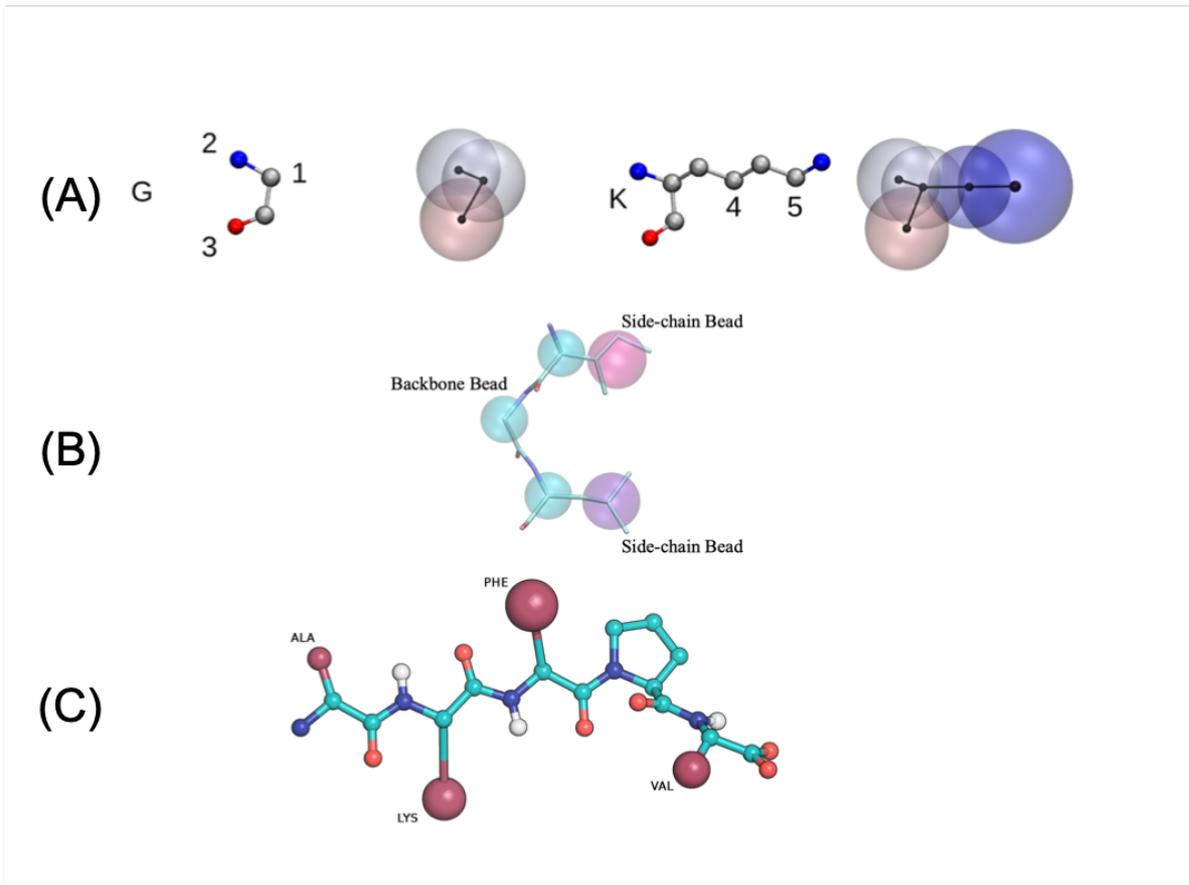
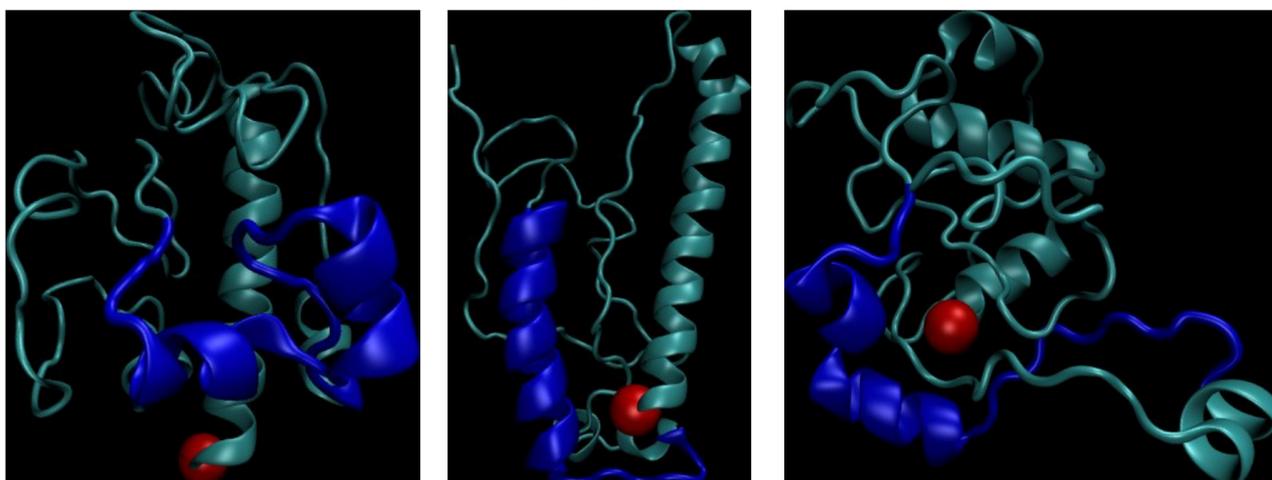


Figure 1. Coarse-grained models. (A) SIRAH for the glycine and lysine residues. (B) SOP-IDP. (C) OPEP for the A-K-F-P-V pentapeptide.

(A)



(B)

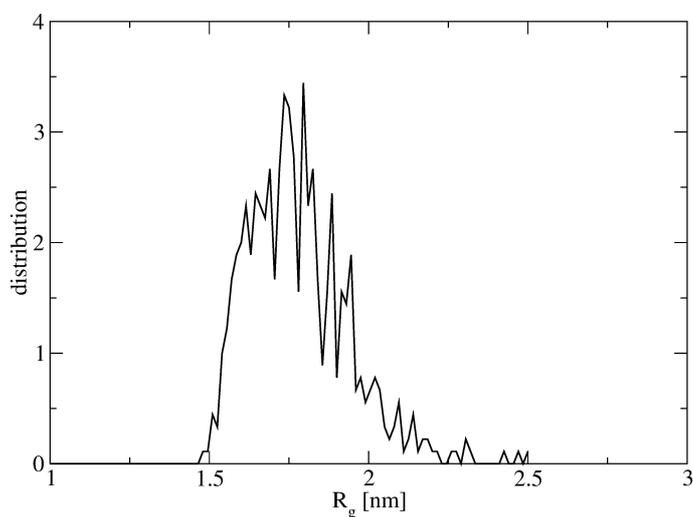
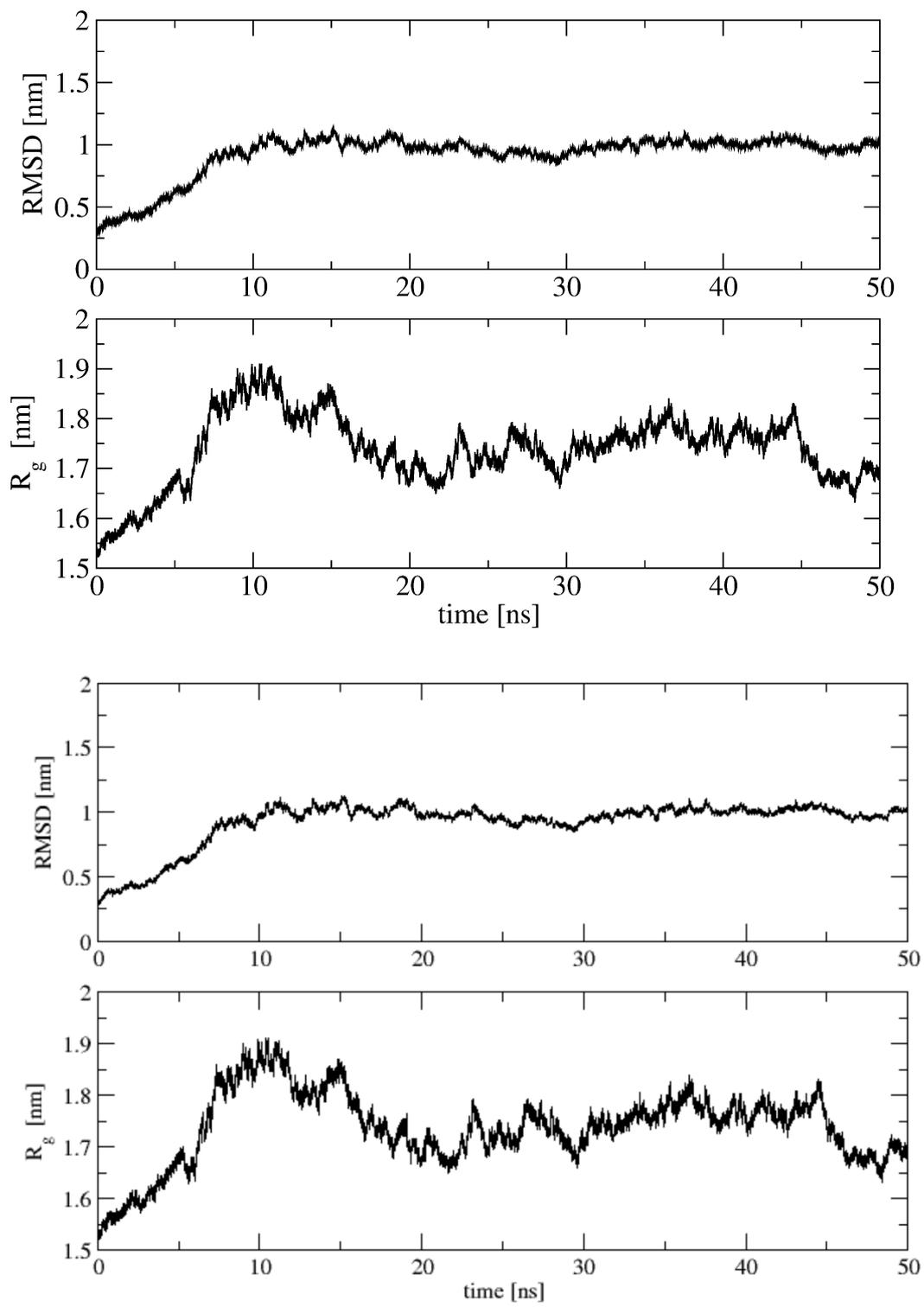


Figure 2. PEP-FOLD simulations on α -synuclein protein. (A) Three structures of lowest-energy from distinct simulations, the N-terminus is in red; Rg of 1.50 nm with α -helix position at residues 1-24, 60-66, 76-81, 85-91 (left), Rg of 1.80 nm with helix at residues 2-25, 27-32, 53-57, 62-66, 75-92 (middle) and Rg of 1.53 nm with helix at residues 1-10, 13-24, 27-31, 60-66, 77-86, 88-92 (right). (B) Radius gyration distribution obtained from 600 simulations.

(A)



(B)

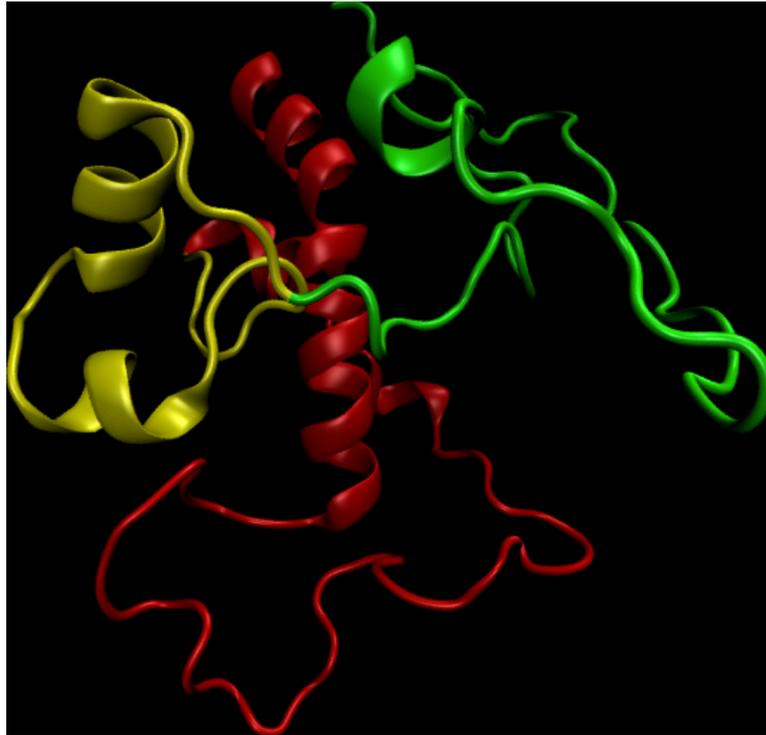


Figure 3. MD simulations on α -synuclein monomer. (A) RMSD and Rg evolution from the PEP-FOLD structure. (B) Final structure at 50 ns with the three domains with different colors.