Molecular Dynamics Simulations of α-Synuclein NAC Domain Fragment with ff14IDPSFF IDP-specific Force Field Suggest β-Sheet Intermediate State for Fibrillation

Cristian Privat, Sergio Madurga*, Francesc Mas and Jaime Rubio-Martinez*

Department of Material Science and Physical Chemistry & Research Institute of Theoretical and Computational Chemistry (IQTCUB), University of Barcelona, C/ Martí i Franquès 1, 08028, Barcelona, Spain

*Correspondence: jaime.rubio@ub.edu; s.madurga@ub.edu

Keywords: α-Synuclein NAC domain; Intrinsically Disordered Proteins; Force Fields; ff14IDPSFF; ff14SB; Molecular Dynamics simulations; β-sheet

ABSTRACT

For the discovery of treatments against synucleinopathies, it is necessary to unravel and fully understand the mechanism of fibrillation of the proteins involved. Among them, α-synuclein (αS) plays a key role in the development of these diseases through its aggregation into oligomers found in Lewy bodies. However, its structural disorder as an intrinsically disordered protein (IDP) makes its characterization by experimental techniques arduously difficult. Atomistic simulations aim to provide insights into this blank canvas and, fortunately, some studies have already suggested promising mechanisms. Still, it is urgent to consider the IDP features in simulations, so recently a lot of force fields designed to deal with IDPs have been developed. In this study we have carried out a total of 12 μs simulation of the αS core fragment using the popular ff14SB AMBER force field, and the ff14IDPSFF variation that includes a grid-based energy correction maps (CMAP) method. The predicted chemical shifts from the simulations and those measured from the αS protein in NMR solution indicate that ff14IDPSFF reproduces more accurately the experimental data. Moreover, the structural analysis exhibits opposite trends between secondary structure propensities. The ff14SB force field preserves the α-helices found in the micelle-bound αS structure, which is used as initial conformation, while the ff14IDPSFF stands out with increased structural disorder and the formation of β-sheet, which suggests that the IDP-specific force field can capture more suitable conformations representing possible intermediate states of the fibrillation process.

1. INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder in the world population, especially in the population over 60 years of age.[1,2] The development of PD is mainly attributed to the aggregation of misfolded α-synuclein (αS) protein in Lewis bodies that ultimately leads to the loss of dopaminergic neurons.[3–5] To date, no definitive cure for the disease has been found. The first stages of the mechanism of formation of these fibrils from αS monomers are still unrevealed, which in turn hinder drug design to treat the disease. So far, micelle-bound αS monomers[6,7] or fibril oligomers[8] have been characterized, but structures of the protein freely in solution or intermediate conformations of the fibrillation process have not yet been reported. The difficulty of identifying the intermediates of fibrillation is due to the transient nature of αS as an intrinsically disordered protein (IDP). IDPs are notable for their structural disorder and their ability to perform interconversion between conformational states over time, which is key to their biological functions. Some experimental techniques, such as nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS) or far-UV dichroism[8–13], can capture the structural properties, but restricting only to properties obtained from time-averaged conformations of the protein. Then, computational studies come into play as a resource capable of providing insights of the fibrillation process by atomistic simulation methods. However, αS fibrils occur on a time scale computationally inaccessible for classical simulation methods. Some studies have attempted to overcome this limitation using ingenious approaches that can range from enhanced-sampling techniques[14–17], coarse-grained description levels[18,19], simulation of specific fragments of αS[18,20] or guiding the simulations by experimental data bias.[21,22] Although these efforts have resulted in very promising mechanisms, further research on αS is still essential.

αS is a 140-aa presynaptic protein found mainly in nervous tissue and whose function is not exactly known to this day.[23] It has been associated with some processes, such as synaptic vesicle recycling, regulation of DNA repair or involvement in neuronal apoptosis.[24] Primary structure of αS is divided into (i) the N-terminal domain (1-60), (ii) the non-amyloid-β component (NAC) domain (61-95) and (iii) the C-terminal domain (96-140). The first domain consists of 7 imperfect repetitions of 11 amino acids that give it an amphipathic character and an overall positive charge. These repeats contain abundant KTKEGV segments, which have a propensity to adopt α-helices conformations and allow αS to bind to membranes.[6,25] The NAC domain plays as a hydrophobic core for fibrillation.[26] Finally, the C-terminal domain is highly charged and mobile because it is rich in acidic amino acids in its chain. Some studies suggest that C-terminal tail is responsible for inhibiting fibril formation by burying the NAC domain, which prevents interaction between monomers and thus oligomerization.[17,18] Interestingly, while the membrane-bound monomeric αS structures have a high α-helix content in the N-terminal and NAC domains, it is precisely the fibril conformation that is characterized by agglomerating with β-sheets. Apparently β-sheet structure may be critical in early stages of fibrillation process although another study suggests that an

intermediate α-strand/sheet intermediate may be necessary to the fibrillation mechanism.[27]

So far, most atomistic simulation studies use classical force fields that generally circumvent the transient nature of IDPs and lead to conformational biases. Historically, classical force fields have been parameterized to accurately reproduce well-defined, experimentally observable three-dimensional structures. Fortunately, force fields that include disorder structure in their parameterization have been developed in recent years. In fact, previous studies on differences between standard and modern force fields have demonstrated a high sensitivity in obtaining IDP conformational ensembles[28,29]. Approaches to include intrinsic disorder can range from adjusting the dihedral parameters (CHARMM22*[30], RSFF2[31] or OPLS-AA/M[32] force fields), adding a grid-based energy correction term to the ϕ/ψ dihedral energy surface called CMAP method (CHARMM36[33], ff14IDPSFF[34], ESFF1[35]) or refining the protein-water interactions (a99SB-disp[36], ff03ws[37], CHARMM36m[38]). Among these new force fields, ff14IDPSFF has been developed as a promising force field capable of correcting the dihedral distributions of all 20 amino acids from the popular ff14SB[39] force field but assisted with the CMAP method. For further confidence, this IDP-specific force field has been shown to improve the description of chemical shifts in alpha-synuclein protein.[34]

In this work we have performed ff14SB and ff14IDPSFF force field simulations of a fragment of the αS protein, which we referred as E-αSNAC^{SB} and E-αSNAC^{IDP} respectively, to analyse the structural differences present as a function of parameterization and hopefully provide insight into the intermediate states of fibrillation. This fragment is defined by amino acids 35-97, which constitutes the fibril core of the Greek-key topology that adopts αS fibril according to Tuttle et al.[8] and includes the NAC domain and a segment of the N-terminal domain. Furthermore, its conformational space was previously explored by 18 μs T-REMD simulation with the CHARMM27* force field.[20] The study concluded that there is a α-helix content at room temperature while β-sheets are formed at high temperatures. Here, we have performed 6 μs-length simulations for each force field system and observed a significant bias in the secondary structure propensities after analysing the conformational properties and contact maps. ff14SB simulations preserve α-helices of the micelle-bound αS structure while conformational sampling of ff14IDPSFF stands out with random coil and β-sheets structures, the latter present in the αS fibrils. Moreover, we have verified by determining the linear equations and the Pearson correlation coefficients that ff14IDPSFF exhibits a closer agreement to four sets of chemical shifts obtained from NMR in solution, demonstrating that the ff14IDPSFF force field indeed reproduces more accurately this experimental observable. Thus, the IDP-specific force field can adopt intermediate states with β-sheet conformations not normally found in classical force fields with only 6 μs simulations, which introduces promising studies on the conformational space of the αS and its role in the fibrillation mechanism.

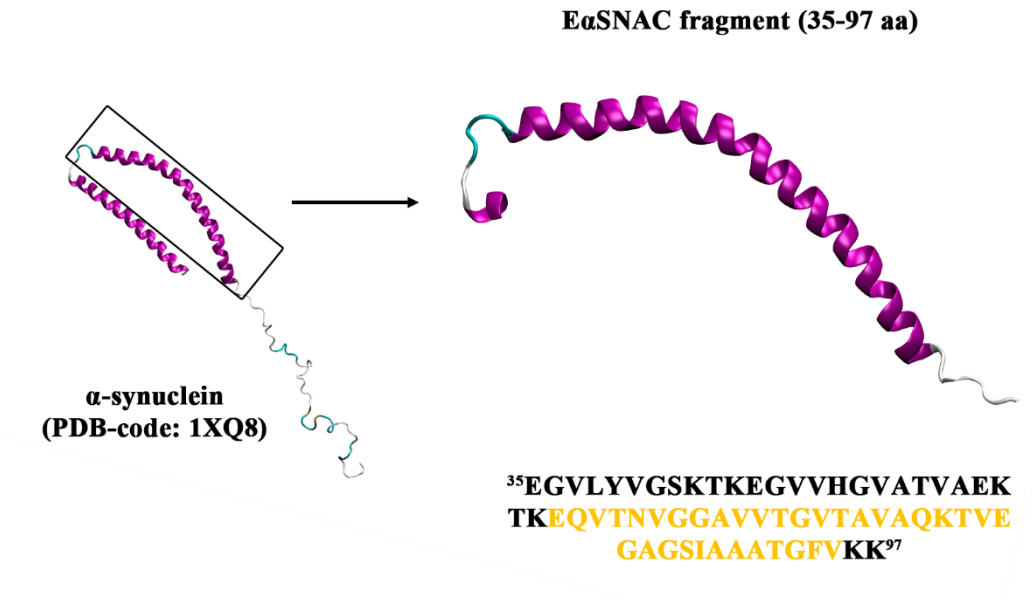

Figure 1. NMR structure of the human-micelle bound α-synuclein (PDB-code: 1XQ8) in the left and structure and amino acid sequence of the E-αSNAC fragment in the right. Orange color refers to the NAC domain amino acids (residues 61-95).

2. MATERIALS AND METHODS

2.1. Structures of Human α-Synuclein Protein. The Protein Data Bank database (https://www.rcsb.org/) provides many αS structures available to the scientific community. For our study we selected the structure characterized by solution NMR with PDB-code 1XQ8 (human-micelle bound αS).[6] The 1XQ8 structure was cleaved, retaining the core amino acids of the protein (35-97) that comprise the NAC domain and a segment of the N-terminal domain. We have renamed this selection of amino acids as the extended α-synuclein NAC domain (E-αSNAC) fragment.

2.2. Structure Preparation and Simulation Setup. E-αSNAC was oriented according to its principal moments of inertia (I_x, I_y, I_z) through an internal script of our research group. Then, we used the LEaP module of AMBER18[40] to parameterize the amino acids with the force field ff14SB[39] and define a box of dimensions 140x140x140 $Å^3$ as the simulation system. The system box was filled with TIP3P water molecules, leaving a space of 1.0 Å between any amino acid in the protein and the water molecules. The net charge of the system was neutralized by the addition of Na^+ or Cl^- counterions. The hydrogen mass was distributed among the amino acid atoms using the ParmEd module in order to increase the time step from 2 fs to 4 fs.[41] Next, a second simulation system was generated with a new topology including the parameterization of the IDPs-specific force field ff14IDPSFF.[34]

Subsequently, we carried out a three-phase protocol to minimize the systems. This protocol consisted of applying steepest descent (SD) method with three levels of restriction on the protein to relax the internal tensions of the system after adding the solvent molecules. In the first minimization phase, 5000 SD steps were performed restricting the entire protein and to relax only the surrounding water molecules. Second, protein and solvent were slowly relaxed during 5000 SD steps but applying restraints on the backbone atoms. Finally, the system was minimized for another 5,000 SD steps without applying any restraints. All the above restraints were defined by a force constant of 5 kcal·mol^{-1}·Å $^{-2}$.

The simulation box was heated with a linear increase of 1 K·ps^{-1} in the NVT ensemble for 300 ps until a final temperature of 300K was reached. A second equilibration step was then performed for 300 ps in the NPT ensemble to adjust the volume of the system box to a pressure of 1.0 atm. In the heating and equilibration steps, restraints were applied to the protein backbone with a force constant of 5 kcal·mol^{-1}·Å^{-2}. In order to increase the conformational sampling, three replicas of the system were generated[42], assigning them random initial velocities that follow a Maxwell-Boltzmann distribution. For each replica, production runs of 2 μs were carried out. The trajectory coordinates were written every 20 fs and the output data every 40 fs. The SHAKE algorithm constrained the hydrogen-involving bonds during the simulation and the temperature was maintained at 300K with the Langevin thermostat and a collision frequency of 3.0 ps^{-1}. A 9.0 Å cutoff and periodic boundary conditions (PBC) were applied.

2.3. Analysis of Simulations. The simulations were studied through various conformational properties as well as compared with NMR experiments through predicted chemical shifts. Visualization of the trajectories and the illustration of the conformations were performed with VMD software.[43] The plots were designed with Gnuplot (version 4.6).[44]

Conformational properties were calculated with the CPPTRAJ module[45] of AMBER18. The root mean square deviation (RMSD) of the protein backbone (Cα, C, N, O atoms) was calculated with respect to the reference structure, which corresponds to the 1XQ8 structure determined by NMR in solution. We also calculated the root mean squared fluctuation (RMSF) and radius of gyration (R_g) of the Cα atoms to provide insight into the flexibility and compactness of the conformations obtained during the simulation. The secondary structure (SS) of the protein was estimated with the DSSP method.[46] They were categorized into five classes: β-strand (isolated β-bridges and extended strands), helices (3_{10} helix, α-helix, and π-helix), coil (no secondary structure assignment), β-turn (hydrogen-bonded turn) and bend. To provide further insight into secondary structure conformations, intramolecular contacts were calculated using CPPTRAJ, for which we accepted as a contact any atom (except hydrogen) with a distance lesser than 8.0 Å to another amino acid atom. The contacts that are defined in the initial structure are called native contacts while those that appear during the simulation are referred to as non-native contact.

The conformational space was represented by principal components analysis (PCA). The covariance of the distance between CA atoms was used as a metric. The conformations of the trajectories were projected onto the first two principal components (PC1 and PC2), which represent a 33% of the covariance. For clustering, first the RMSD of the CA atoms was calculated as a distance metric with a 5-frame sieve. Next, starting from the conformations separately as individual clusters, the clusters were merged according to the average distance between the members of the two clusters until all conformations were assembled into 15 clusters. The RMSD of the centroid conformations of the five most-populated clusters was calculated and plotted on the 2D-RMSD plot. Next, an in-house script estimated the Gibbs free energy through the populations of PC1 and PC2 following the equation Eq1, where n_i and n_{max} refer to the population in bin *i* and n_{max}; the bin of maximum occupation, to build energy maps in the PCA space.

$$\Delta G = -k_b T ln(n_i/n_{max}) \quad \text{(Eq1)}$$

The chemical shifts of 15N, 13C, $13C^{\alpha}$ and $13C^{\beta}$ atoms of one-fifth of the conformations in the simulations were estimated using SPARTA+.[47] The averages of the chemical shifts of each atom were calculated and, subsequently, linear regression and Pearson correlation coefficients between the simulated and experimental chemical shifts data sets were computed using Gnuplot and SciPy module, respectively.[48] The NMR-measured chemical shifts were obtained from the Biological Magnetic Resonance Data Bank (BMRB) database (https://bmrb.io/) with the following IDs: 18857[49], 19337[50], 25527[51] and 6968[52].

3. RESULTS

We performed 6μs of extended αS NAC (E-αSNAC) domain fragment using the ff14SB ($E\text{-}\alpha SNAC^{SB}$) and ff14IDPSFF ($E\text{-}\alpha SNAC^{IDP}$) force field by three 2 μs-length simulations with random initial velocities. A total simulation time of 12 μs was carried out in this study, independently of the parametrization.

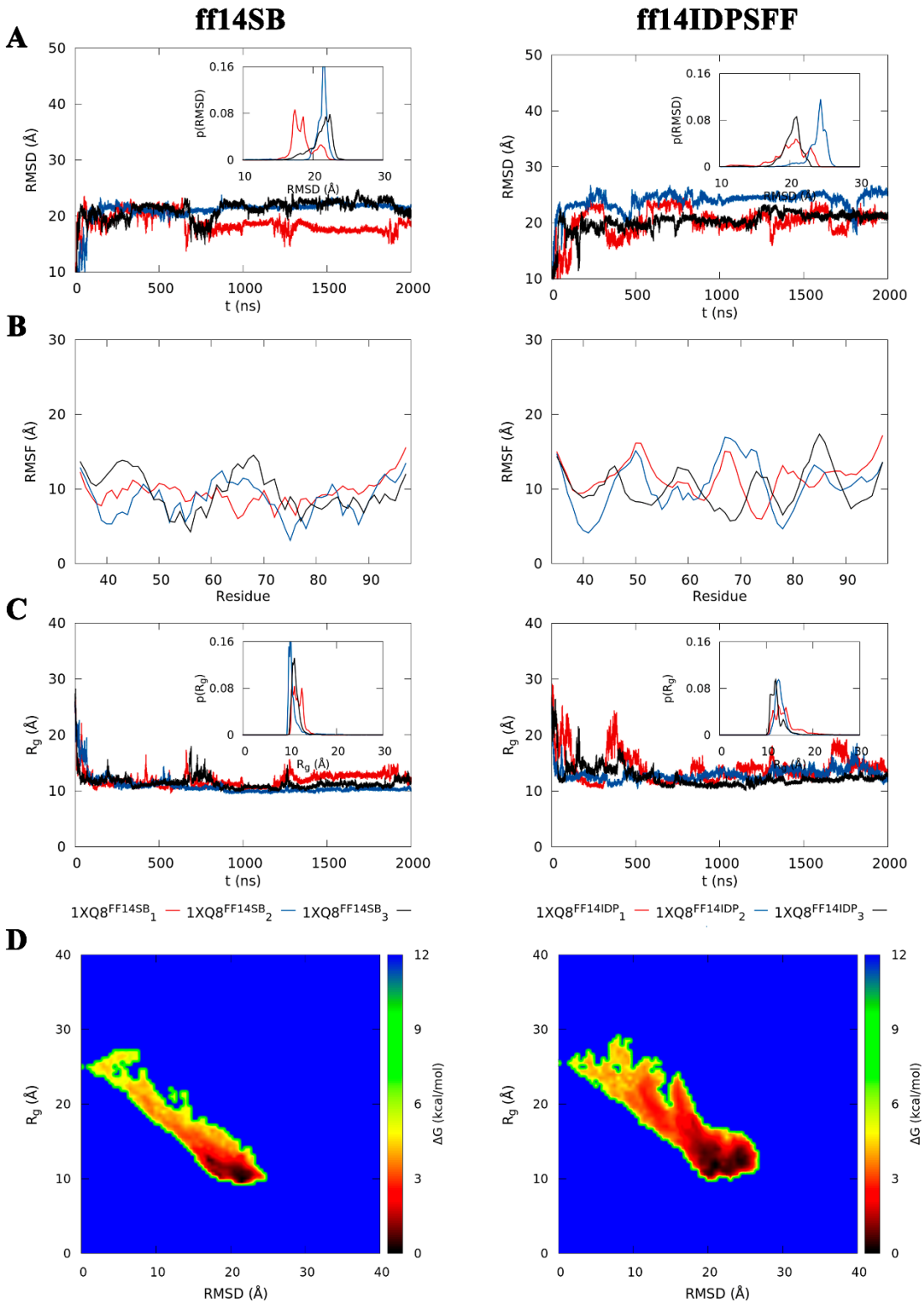

Figure 2. Conformational properties of ff14SB (left) and ff14IDPSFF (right) simulations of E-αSNAC. (A) Root mean squared displacement (RMSD), (B) fluctuations (RMSF) and (C) radius of gyration of each replica are in red, blue, and black lines. The probability distributions of RMSD (p(RMSD)) and radius of gyration ($p(R_g)$) are also illustrated inside each graph. (D) Energy map of the RMSD and radius of gyration expressed in kcal·mol^{-1}.

Conformational Properties. First, we calculated the conformational properties (RMSD, RMSF and radius of gyration) of the ff14SB and ff14IDPSFF simulations. After superimposing the trajectory conformations to the NMR-characterized 1XQ8 structure, the RMSD of protein backbone (Cα, C, N, O atoms) was calculated for each simulation and illustrated in Figure 2A. The fluctuations the amino acids during the simulation are illustrated in Figure 2B.

Distinct RMSD and RMSF are observed depending on the replica, which emphasize the importance of carrying out multiple production runs of the system to widely explore the conformational space. The time dependent RMSD shows that ff14SB simulations exhibit values below 21 Å in Figure 2A. Interestingly, replica 1 yield conformations with RMSD values around 15Å and the distribution shows conformations mostly in two RMSD ranges (~17Å and ~21Å). Furthermore, replica 2 presents a large peak in the distribution as well as stable time dependent RMSD profile, suggesting that the protein conformation does not change significantly during the trajectory. On the other hand, the time dependent RMSD of ff14IDPSFF apparently fluctuates more compared to ff14SB simulations. In fact, replica 2 stands out with values above 21 Å, even reaching conformations with RMSD around 25Å. The peaks of the distributions of ff14IDPSFF are around 20Å and 24Å. Besides, the fluctuations of the residues according to each force field simulation are illustrated in Figure 2B, in which it is observed that the RMSF profiles of ff14SB exhibit different behaviours in residues 35-47 and 65-75 between replicas, while in the rest of the protein is similar. In contrast, FF14IDPSFF shows different RMSF profiles in almost the entire E-αSNAC fragment, even exhibiting Δ(RMSF) of ~10 Å at residues 65-72. The RMSD of ff14IDPSFF shows larger distances with respect to the ff14SB counterpart, indicating conformational sampling further away from the α-helix-rich initial structure, while the RMSF profiles show less pronounced discrepancies between the replicas of the ff14SB force field. This could indicate that the IDP-specific force field samples regions in the conformational space where amino acid mobility varies significantly.

To study the compactness of E-αSNAC, the time-dependent radius of gyration and their distributions are illustrated in Figure 2C, and the energy maps built by RMSD/R_g distributions in Figure 2D. The averages of R_g are (11.4 ± 1.5) Å in E-αSNAC^{SB} and (13.2 ± 2.1) Å in E-αSNAC^{IDP}. The ff14SB simulations show a stable R_g, being consistent with the narrow distributions at ~11Å. In contrast, the R_g of ff14IDPSFF exhibits more significant fluctuations than the ff14SB force field along with broad distributions located at ~13Å. Although the R_g averages of each force fields are close, the conformations of ff14SB simulations appear modestly more compact compared to ff14IDPSFF. On the other hand, the energy maps in Figure 2D show that ff14SB have a thin populated region depicted in black within the graph. Simulations of ff14IDPSFF enlarge this black region as well as sample conformations with higher values of RMSD and R_g, which is in line within reported observations suggesting that the IDP-specific force field can capture conformations away from the initial α-helix-rich structure.

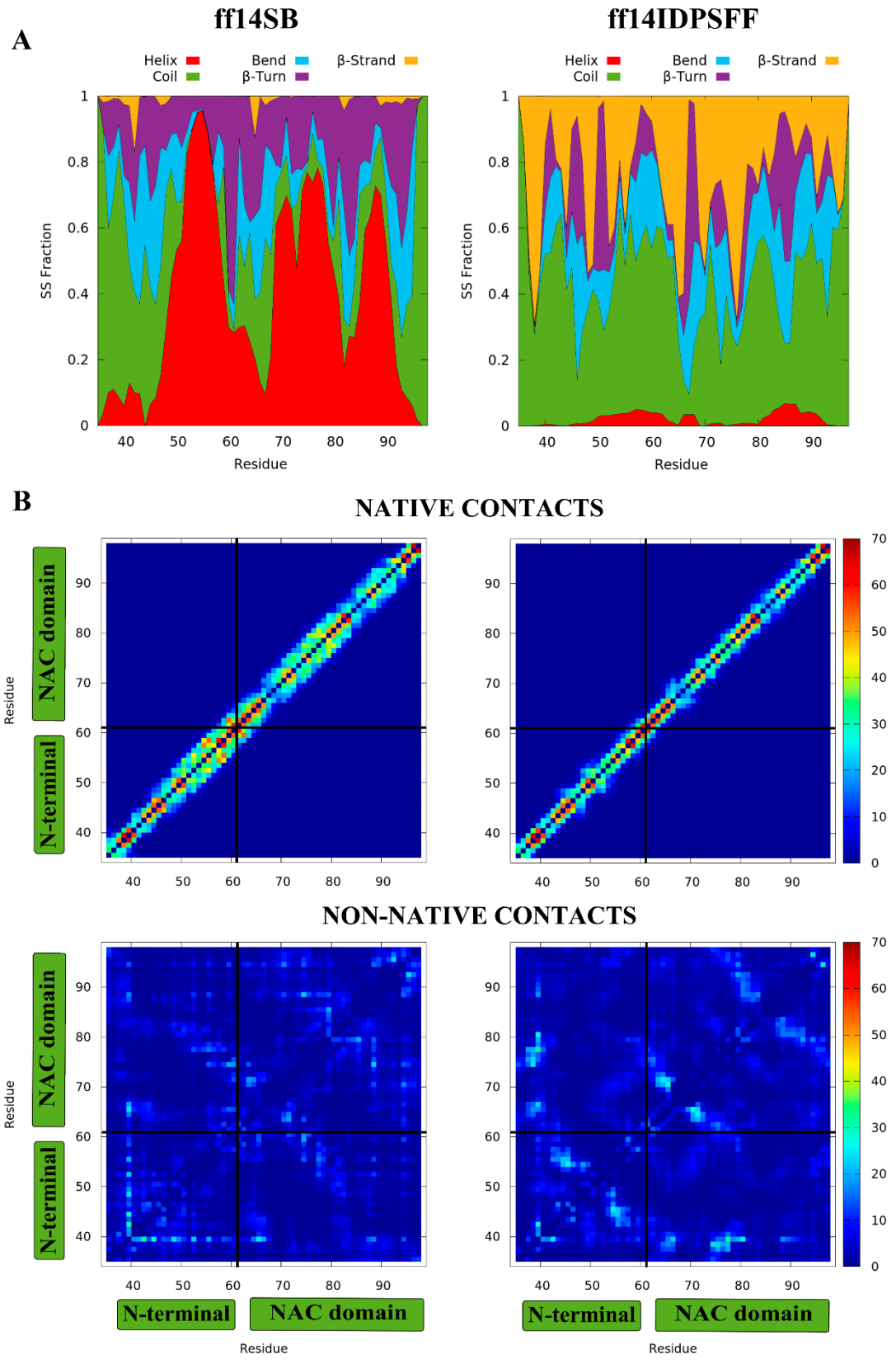

Figure 3. Structural analysis of ff14SB (left) and ff14IDPSFF (right) simulations in E-αSNAC. (A) Fractions of the secondary structure propensity of extended αS NAC domain fragment. Helix class (red) includes α-helix, π-helix and 3_{10}-helix, and β-strand

class (orange) includes anti- and parallel β-sheet. β-turn, bend and coil class are coloured in purple, cyan and green, respectively. (B) Native (middle) and non-native (bottom) contact maps of the extended αS NAC domain fragment. Sum of the contacts normalized by the total number of conformations of each atom pair inside a residue is represented by a colour scale (maximum and minimum number of contacts in red and blue, respectively). N-terminal and NAC domains are separated with black lines in the plot.

Secondary Structure and Contact Maps. We performed the DSSP method of Kabsch and Sander[46] on all conformations to calculate the secondary structure propensities. Subsequently, we averaged them to obtain the secondary structure propensity factors (fpSS), which are illustrated in Figure 3A. For convenience, the α-helix, π-helix, and 3_{10}-helix were grouped into the helix class and the parallel and antiparallel β-sheets into β-strand. The other SS propensities keep the conventional names (β-turn, bend, and coil). The fpSS in E-αSNAC^{SB} indicates a predominant helix (39%) content that coexist with coil (25%), β-turn (20%) and bend (14%) conformations. In the case of ff14IDPSFF, fpSS(E-αSNAC^{IDP}) exhibits a large random coil (44%) content along with β-strand (26%), bend (16%) and β-turn (12%) conformations.

DSSP analysis evidences that the parameterization of the E-αSNAC fragment leads to significant changes in the conformations adopted and thus in the fpSS. ff14SB preserves the helices found in the NAC and N-terminal domains of the 1XQR structure for a substantial simulation time, especially at residues 40-50, 50-70 and 85-90. Even so, bends and β-turn as well as the random coil conformations become present in the simulation, thus breaking with the high helicity of the native structure. Surprisingly, ff14IDPSFF shows a near absence of helical conformations that are instead mostly replaced by random coils, as well as conformations in smaller presence of β-strand, bend, and β-turn. The large random coil content together with the breakdown of native helices could explain larger flexibility indicated by the RMSD and R_g observables.

In addition, the residue contacts were tracked using an 8 Å distance cutoff. For each force field, two contact maps are illustrated in Figure 3B. Native contacts are those identified in the 1XQ8 structure and conserved during conformational sampling. Non-native contacts are those that are originated in the simulation and do not appear in the NMR reference structure. The contacts of each individual atom pair in a residue are normalized between 0 to 1.0 by the total number of conformations and summed to its corresponding residue. Values greater than 1.0 are expected, meaning that each residue have, in average, the indicated number of contacts with all the possible atom pairs that can establish in the protein within the cut-off during the simulation.

The ff14SB simulations show strong contacts on the diagonal of the native contact map, which are related with the formation of α-helices (Figure 3B). Residues with high population of contacts on the map diagonal also exhibit large helix content in the DSSP map. In the non-native contact maps of ff14SB, the fragment presents spurious antidiagonal contacts, which are identified with antiparallel β-sheets. Besides, 35-50 region is rich in contacts, apparently promoting random coil conformations as

suggest the DSSP map. On the other hand, native contacts of ff14IDPSFF are less abundant in the contact map, a fact consistent with the lack of helix content in DSSP plot. In fact, diagonal and antidiagonal contacts in the non-native contact map are notable for the E-αSNAC^{IDP} fragment. Several amino acids show antidiagonal contacts with each other (residues 43-58, 58-75, 78-95). In addition, the contact map also exhibits parallel contacts between amino acids 37-43 and 75-81 and, with less population, amino acids 53-59 and 69-75. In both force fields, Tyr39, the unique tyrosine in the N-terminal domain, establishes strong contacts with few residues. In ff14SB simulations, Tyr39 interacts especially with the amino acids S42, T44, V48, V52, V55, A65, and V66. This number of contacts is reduced for ff14IDPSFF, in which Tyr39 interacts with S42, V66, V77, and L80 but allows Leu38 to interact with residues V74-V77. Many regions show overlapping contacts between different amino acids, which would indicate dynamic folding and unfolding of distinct β-strand conformations. Thus, the formation of β-sheet is apparently not found at specific amino acids, but it is formed in different, and even shared, regions of the E-αSNAC fragment.

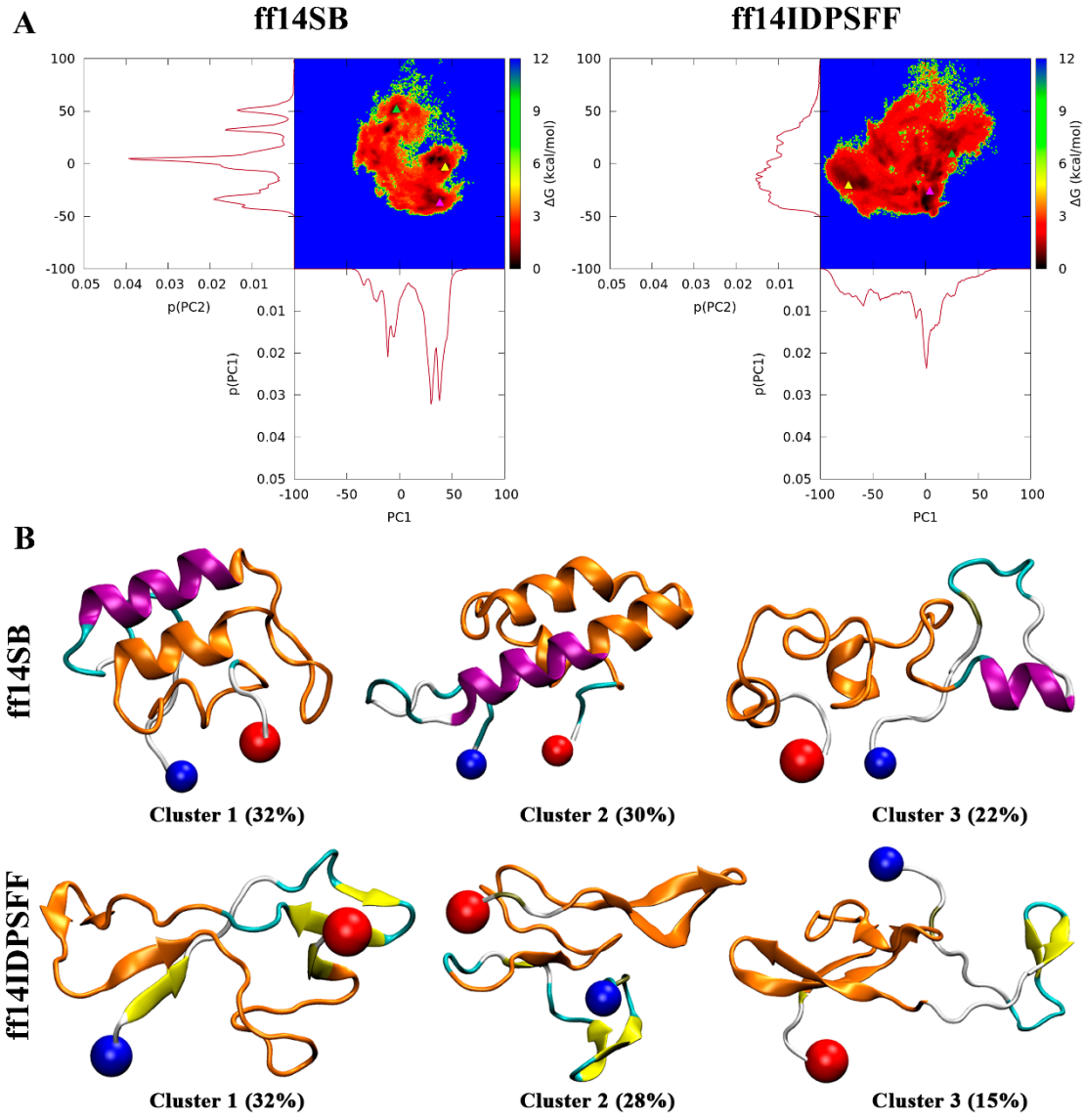

Figure 4. (A) Gibbs free energies maps in the Principal Components space of the extended αS NAC domain fragment using ff14SB (left) and ff14IDPSFF (right) force fields. Centroid conformations of clusters 1, 2 and 3 are marked with yellow, pink, and green triangles, respectively. Probability distributions of the PC1 and PC2 are also represented. (B) Top 3 clusters of ff14SB (top) and ff14IDPSFF (bottom) simulations. NAC domain (residues 61-95) is coloured in orange.

Conformational Sampling. Principal Component Analysis (PCA) was performed using all the conformations obtained from the simulations with ff14SB and ff14IDPSFF force fields. The conformations of each trajectory were projected to the PCA space and the two first eigenvalues (PC1 and PC2) were transformed to Gibbs free energies as detailed in Methods. The PCA energy maps and the distributions of PC1 and PC2 are illustrated in Figure 4A. The conformations of the ff14SB and ff14IDPSFF simulations were clustered into 15 clusters using the hierarchical agglomerative clustering method. The centroids of the three most-populated clusters are plotted in Figure 4B. The 2D-RMSD of the centroids of the five most-populated clusters is represented in Figure S1.

The PCA energy map of ff14SB distinguishes five low-energy regions in the conformational sampling. The distributions of the PC1 and PC2 show several peaks, suggesting that certain conformations are preferred and, conversely, the access to some regions inside the conformational space is apparently restricted. After clustering, the centroid conformations of the two most-populated clusters show a high proportion of helix content in their structure, while the third most-population cluster has a disordered structure in its centroid. In contrast to ff14SB, E-αSNAC^{IDP} shows a widespread conformational sampling as well as the population is more distributed within it, which is consistent with the broad PC1 and PC2 distributions. ff14IDPSFF samples region PC1 = [-100, -50], PC2 = [-50, 100], which does not appear in the ff14SB simulation. In fact, the centroid of the most-populated cluster in the E-αSNAC^{IDP} is in this region and exhibits a high proportion of β-strand. On the contrary, the region PC1 = [25,50], PC2 = [0,50] is not sampled in ff14IDPSFF, in which area the centroids of the most-populated clusters of ff14SB are located. Then, the exploration of certain regions in the PCA space is favored or restricted depending on the used force field.

The centroid conformations of the three most-populated clusters, comprising >75% of the population, are illustrated in Figure 4B. The populations of each cluster are indicated below. 84% of conformations in the ff14SB simulations are found within the first three clusters. In contrast, the three most-populated clusters account for 76% of the conformations in ff14IDPSFF, suggesting that the conformational sampling is more spread out in the IDP-specific force field. Furthermore, the centroid conformations also evidence distinct secondary structure propensities in the simulations. E-αSNAC^{SB} preserves the helices of the native structure, except for the centroid of cluster 3. Given the importance of the NAC domain for the fibril formation, this region (61-95 residues) is colored orange in the structure. Centroids 1 and 2 exhibit α-helix conformations, especially at amino acids 70-80 and 86-90. The 60-70 amino acids become loops or turns to allow the protein to change direction and, in

turn, makes possible the α-helix interaction of the cleaved N-terminal and the NAC domain to form α-hairpins. Indeed, cluster 2 centroid captures part of the helix formation (~30%) in the N-terminal domain interacting with the rest of the E-αSNAC fragment.

ff14IDPSFF does not exhibit any helix formation in their conformations and instead adopts β-strands, even β-hairpins, which are formed within the NAC domain itself or together with amino acids of the N-terminal domain. Cluster 1 centroid has β-strands (residues 1-4, 14-20, 40-42, 58-61) formed between NAC and N-terminal domain, as well as short β-hairpin (residues 29-35) in the NAC domain itself. The centroids of clusters 2 and 3 present larger β-hairpins that involved different amino acids of the NAC domain. These β-strands, whether forming β-hairpin or not, appear in different amino acids as indicated by the DSSP method and the contact maps.

We have also studied the goodness of clustering through the distance-to-centroid, Davies-Bouldin Index (DBI), pseudo-F statistic (psF) and SSR/SST magnitudes collected in Table S1. The distance-to-centroid calculates the mean of the distance of the conformations captured within a cluster with respect to the centroid, DBI measures whether the separation of clusters and the classification of conformations is correct, psF aims to capture the tightness of the clusters and SSR/SST indicates the percentage of variance captured in the clustering. The magnitudes presented in the table indicate that the DBI and SSR/SST values are similar between the clustering of the ff14SB and ff14IDPSFF simulations, thus noting that the conformations obtained from the simulation are captured and classified with the same degree of goodness. On the other hand, the distance-to-centroid and the psF magnitudes show lower and higher values in the ff14SB simulations, indicating a higher similarity between conformations within the narrower clusters. Conversely, although the clustering of the ff14IDPSFF is ranked with the same degree of goodness, the conformations within a cluster show greater dissimilarity and the clusters are broader in conformational space.

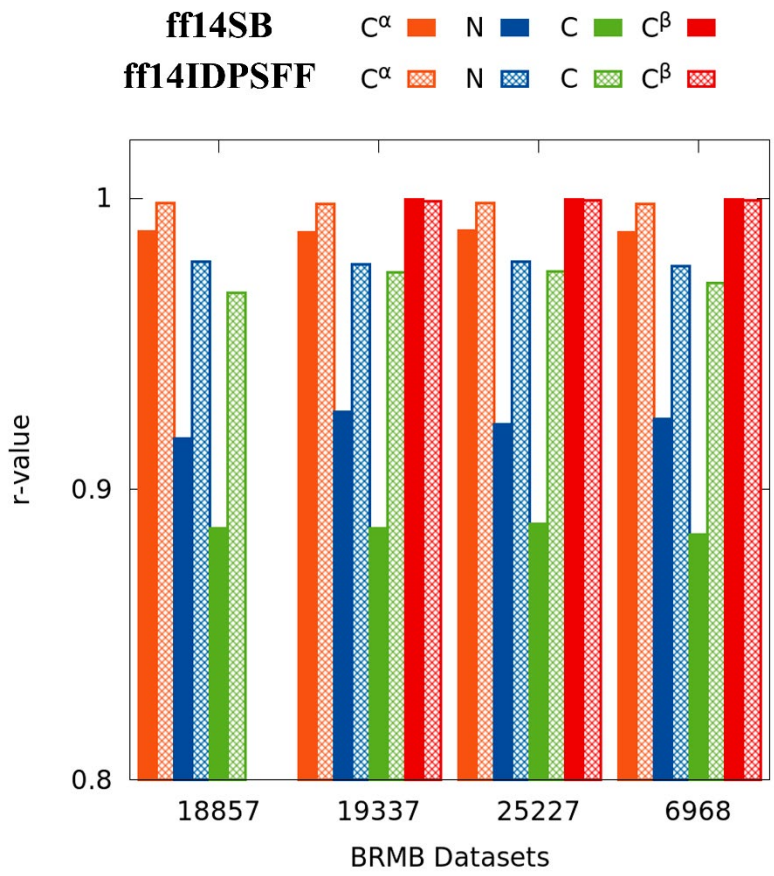

Figure 5. Pearson correlation coefficients between experimental and simulated chemical shifts using four BRMB data sets indicated in x axis (BRMB ID: 18857, 19337, 25227 and 6968) of the ff14SB (solid) and ff14IDPSFF (pattern) simulations. Orange, blue, green, and red boxes are C^{α}, N, C and C^{β} atoms. The C^{β} chemical shifts were not available in the BRMB 18857 data set.

Simulated and Experimental NMR Chemical Shifts. Finally, the protein backbone chemical shifts ($\delta^{15}N$, $\delta^{13}C$, $\delta^{13}C^{\alpha}$, $\delta^{13}C^{\beta}$) were predicted using the SPARTA+ program[47]. The results of the ff14SB and ff14IDPSFF simulations were compared with measured NMR data of full-length αS obtained from four data sets available in the Biological Magnetic Resonance Data Bank (BRMB). The relationships between predicted and experimental sets of chemical shifts were analysed with the Pearson correlation coefficient (r) in Figure 5. The linear regression of each set of measured and predicted chemical shifts for each atom and force field are illustrated in Figure S2-S9.

Linear regression of the C^{α} and C^{β} atoms shows a high reproducibility of the chemical shifts in both force fields. Only the slope of C^{α} atom linear equation shows a minor deviation in contrast to the rest of the dataset, which fits accurately. On the other hand, the linear regression of the chemical shifts of N presents a greater slope and intercept for the ff14IDPSFF force field with values around 0.95 and 4.5-7, respectively. Contrarily, the predicted chemical shifts of the N atom in ff14SB deviate significantly from the experimental ones thus fitting its linear equation to slopes

around 0.78. Finally, the chemical shifts of the C atoms show lower reproducibility with respect to the rest of the atoms independently of the force field. Even so, they are not negligible, presenting slopes ~0.88 and ~0.85 for ff14SB and ff14IDPSFF, respectively. Pearson correlation coefficient is often used to assess whether two variables, in this case the simulated and measured NMR, are correlated. The r-values obtained from the simulations are generally well predicted as they are >0.88 but with some differences in the N and C atoms depending on the force field. Of note are the C^{α} and C^{β} atoms, which have r-values very close to 1.00 for all simulations independently of the force field. The r-values of the chemical shifts of N and C atoms are also noteworthy with values around 0.97 in ff14IDPSFF simulations. On the other hand, the r-values of these atoms in ff14SB are around 0.91 and 0.88, respectively, being lower with respect to the ff14IDPSFF force field. Then, the comparison between the measured NMR and the predicted chemical shifts indicates a better correlation for the IDP-specific force field, suggesting that the conformations obtained from ff14IDPSFF simulations are more appropriated to study the conformational properties of αS.

4. DISCUSSION

The ff14SB and ff14IDPSFF simulations of the extended NAC domain fragment of αS exhibit different conformational sampling in the reported analysis, highlighting the preservation of the α-helix in the former while the IDP-specific force field coexist in random coils and β-strands. The exploration of the PCA space shows specific regions depending on the parameterization of the protein. Apparently, in these regions are located the most populated clusters of each force field. Although the RMSD and R_g observables have close average values, the ff14IDPSFF simulations show broader distribution as well as vaguely larger values. Considering the conformational analysis, the results point out that the force field ff14SB achieves more restricted and compacted conformations while the force field ff14IDPSFF explore more flexible conformations far from the reference structure. Furthermore, indicators of the goodness of clustering point out that the conformations obtained from ff14IDPSFF are grouped more broadly in clusters and with greater structural difference, such an observation may be expected given the greater structural diversity afforded by an IDP-specific force field.

According to Yu et al.[18], who collected experimental data in full-length αS under various measurements conditions reported in the literature, the α-helix, β-sheet and turn contents are within 10%-48%, 0%-20% and 17%-41%, respectively. These results agree with ff14SB simulations, which have a ~40% of α-helix content and a negligible propensity of β-strand (<1%). On the other hand, Chen et al.[17] report the secondary structure content from far-UV CD spectrum of monomeric αS in solution: 6% α-helix, 34% β-strand, 18% β-turns, and 42% other conformations (includes random coil, bends and other helices). These propensities are in good agreement with E-$\alpha SNAC^{IDP}$. Despite reproducing experimental data with simulations of E-αSNAC, it should be noted that the distribution of secondary structure in αS is heterogeneous. E-αSNAC fragment removes the disordered C-terminal domain (residues 95-140) and part of the N-

terminal domain (residues 1-37), in which extensive α-helix conformation is found in the structure of the membrane-bound αS. Neglect of these regions may result in variations in the secondary structure content if compared to the full-length protein experiments. Moreover, it must be considered that these secondary structure ratios determined by the experimental techniques are conditioned by the interactions that may be established between the multiple monomers of the protein in solution. In our simulations we work with a single αS fragment, in which intermolecular interactions are absent and may lead deviations from the experimental data.

Fortunately, computational studies focusing on fragments of αS have been performed [18,20,53], motivated by the prohibitive cost of carrying out long-time simulations of the full protein. Jain et al.[20] performed 18μs T-REMD simulations of the E-αSNAC fragment with CHARMM27* force field in which he determined a higher content of α-helix (~62%) at the expense of a lower content of random coil, β-turn and bend (~16%, ~11%, ~8%, respectively). Their results agree with our ff14SB simulations despite having a more distributed secondary structure content. Still, the contact map reported in Jain et al. work indicated that residues 70-85 have a 20-40% probability of establishing antidiagonal contacts, which are related to β-hairpin conformations. The DSSP map of ff14IDPSFF simulations shows this type of contacts and the β-sheet content is within the probability. On the other hand, Yu et al.[18] used a hybrid-resolution model to perform long-time simulations of αS and a short fragment (36-55 amino acids) parameterized with CHARMM27 force field and CMAP correction. In the 36-55 fragment they found β-hairpin formation. They suggest that strong interactions between the C-terminal and this β-hairpin region reduce the access to β-strand formation and increase the required folding time. The E-αSNAC structure includes this β-hairpin region, which are N-terminal residues close to NAC domain. The ff14IDPSFF simulations are in line with Yu et al. results since β-hairpins conformations are found in the centroids of clusters 2 and 3.

In addition, Chen et al. work finds β-strands in the NAC domain of some cluster centroids after performing trFRET-guided DMD simulations of full-length αS.[17] More importantly, this observation is supported by previous studies pointing out that hydrophobic central amino acids of αS form β-strands as a first step for oligomerization and fibril formation.[18,54–56] On the other hand, a recent study of Balupuri et al.[27] suggests that aggregation of αS could be accomplished by an intermediate with an α-strand/sheet conformation found particularly in the critical NAC region of amino acids 72-74. Indeed, we analyzed the conformations of these three amino acids, but did not detect any α-strand content, regardless of the force field. Interestingly, this critical NAC region in ff14IDPSFF simulations exhibits a large β-turn content between adjacent β-sheets, in addition to showing a higher density of contacts in the 63-65 region with respect to ff14SB. Similarly, other regions apparently crucial for αS fibrillation also show specific contacts only observed in ff14IDPSFF. Residues 74-81, which are part of the critical segment of the NAC domain for fibril formation[26], show significant contacts with amino acids 37-39. For these 74-81 amino acids, ff14SB shows only weak, sparse, and heterogeneous contact density with residues 48-67. On the other hand,

ff14IDPSFF shows another region with high contacts between residues 43-49 and 54-57 in which β-sheet propensity is observed in the DSSP map. In this case, amino acids 43-49 are part of one of the seven imperfect KTKEGV repetitions reported in the NAC-domain, probably involved in the association of the protein with membrane lipids although it has also been hypothesized that they may play a role in the tetramerization of αS[25,57]. However, this fragment is part of the compact hydrophobic β-sheet-rich structure in the fibrils. The map also exhibits a region with minor antidiagonal contacts between amino acids 74-83 and 86-97 that probably adopt the β-hairpin observed in the cluster 2 representative. To ensure the convergence of these β-sheet structures observed in the clusters, we have illustrated the β-sheet propensity of E-αSNAC^{IDP} in Figure S10. The convergence plot shows that the β-sheet content is stable after simulating 6 μs in the ff14IDPSFF force field. Finally, Meade et al.[58] review lists the most important mutations (A30P, E46K, H50Q, G51D, A53T/E) that affect αS fibrillation rate. We have studied whether these amino acids establish significant interactions but unfortunately any force field shows remarkable contacts.

One of the key points in understanding the results presented is the conformational bias that exist in the force fields according to the studies reported so far. Duong et al.[59] suggests that there are secondary structures preferences in ff14SB and ff14IDPSFF force fields. Moreover, several studies also points out that ff14SB force field overestimate helix formation due to the usage of globular protein structures in the parameterization.[39,59,60] On the other hand, ff14IDPSFF is relatively recent within the set of force fields that attempt to include the IDPs flexibility by adding and optimizing the CMAP correction terms in the potential energy function of ff14SB. After studying short peptides and RNA-binding protein HIV-1 Rev, Duong et al.[59] concludes that ff14IDPSFF force field promotes random coil conformations and disordered secondary structures, which is consistent with experiments. An example of the capability of this force field are the works of Song et al.[34] and Dan et al.[61], in which the simulation of all-atom microtubule-associated Tau protein with ff14IDPSFF is able to capture β-sheet conformations that are also observed in experiments. In a comparative study between force fields, β-hairpin was found in conformational ensembles of β-amyloids proteins using IDP-specific force fields.[62] A similar trend of β-strand content is observed in E-αSNAC^{IDP} after performing conventional 6μ-length MD simulations with ff14IDPSFF. In addition to the conformational bias in force field, Yu et al.[18] demonstrate with extensive hybrid-model PACE simulations that C-terminal interactions affect in the β-hairpin formation of the 38-53 region. Therefore, E-αSNAC^{IDP} possesses two factors that facilitates random coil or β-strand conformations, i.e. (i) the promotion of random coil conformations by the ff14IDPSFF force field and (ii) the absence of the C-terminal domain and its interactions with the NAC domain.

Finally, correlations between predicted and measured NMR chemical shifts point out that ff14IDPSFF reproduces more accurately the experiments of αS in solution, especially for ^{13}C and ^{15}N atoms. Indeed, it is in line with previous works, which report that a promising feature of the ff14IDPSFF force field is the improved prediction of chemical shifts compared to ff14SB.[34,59,61,62] Our results with the IDP-

specific force field together with the conformational analysis of previous studies and contrasted with NMR experimental data point to a great success of exploring the αS conformational space within affordable MD simulation times lengths.

CONCLUSION

αS protein adopts a wide range of conformations during time evolution due to its structural disorder typical of IDPs. Given the difficulty in experimentally characterizing these conformations, atomistic simulations come into play, especially those that in recent years have attempted to include IDP features. In this study we have selected an αS fragment (residues 35-97) to reproduce its conformational space through 6 μs simulations using the classical ff14SB AMBER force field and the IDP-specific force field ff14IDPSFF developed by Song et al. [34]. The results point out that the classical force field preserves the conformations typically found in the micelle-bound α-synuclein structure, characterized by the high presence of α-helices. However, the force field ff14IDPSFF excels with a set of conformations that are apparently not easily accessible in conventional simulations, or at least at an affordable computational cost, which is dominated by higher structural disorder and the presence of β-sheets. Further study of these conformations could shed some light on the mechanism of protein fibrillation and, subsequently, better understand the group of neurodegenerative disorders derived from synucleinopathies. To validate these simulations, we have performed linear regression and reported the Pearson's correlation coefficients between the predicted chemical shifts and those measured from solution NMR of αS, demonstrating that ff14IDPSFF reproduces NMR data more accurately. Therefore, the results presented in this work suggest that ff14IDPSFF is reliable for exploring αS conformations that are not normally observed in generic force fields and adds new evidence to the body of work pointing to β-sheet formation as intermediate state to fibril formation in αS protein.

AUTHOR CONTRIBUTIONS

Conceptualization, S.M., F.M. and J.R.-M.; methodology, C.P.; formal analysis, C.P.; investigation, C.P.; resources, S.M., F.M. and J.R.-M.; data curation, C.P.; writing—original draft preparation, C.P. and J.R.-M.; writing—review, S.M. and F.M.; visualization, C.P.; supervision, S.M., F.M. and J.R.-M. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

There are no conflicts to declare.

ACKNOWLEDGMENTS

This research was funded by the Generalitat de Catalunya, grant number 2017SGR1033. We acknowledge the financial support from the grant MDM-2017-0767 of Spanish Structures and Excellence María de Maeztu program.

REFERENCES

1 M. Goedert, Alpha-synuclein and neurodegenerative diseases, *Nat. Rev. Neurosci. 2001 27*, 2001, **2**, 492–501.

2 L. Hirsch, N. Jette, A. Frolkis, T. Steeves and T. Pringsheim, The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis, *Neuroepidemiology*, 2016, **46**, 292–300.

3 M. G. Spillantini, M. L. Schmidt, V. M. Y. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert, α-Synuclein in Lewy bodies, *Nat. 1997 3886645*, 1997, **388**, 839–840.

4 T. M. Dawson and V. L. Dawson, Molecular Pathways of Neurodegeneration in Parkinson's Disease, *Science*, 2003, **302**, 819–822.

5 L. Breydo, J. W. Wu and V. N. Uversky, α-Synuclein misfolding and Parkinson's disease, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2012, **1822**, 261–285.

6 T. S. Ulmer, A. Bax, N. B. Cole and R. L. Nussbaum, Structure and dynamics of micelle-bound human α-synuclein, *J. Biol. Chem.*, 2005, **280**, 9595–9603.

7 J. N. Rao, C. C. Jao, B. G. Hegde, R. Langen and T. S. Ulmer, A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins, *J Am Chem Soc*, 2018, **132**, 8657–8668.

8 M. D. Tuttle, G. Comellas, A. J. Nieuwkoop, D. J. Covell, D. A. Berthold, K. D. Kloepper, J. M. Courtney, J. K. Kim, A. M. Barclay, A. Kendall, W. Wan, G. Stubbs, C. D. Schwieters, V. M. Y. Lee, J. M. George and C. M. Rienstra, Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein, *Nat. Struct. Mol. Biol. 2016 235*, 2016, **23**, 409–415.

9 V. N. Uversky, J. Li and A. L. Fink, Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation, *J. Biol. Chem.*, 2001, **276**, 10737-10744.

10 M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, and M. Vendruscolo and C. M. Dobson, Mapping Long-Range Interactions in α-Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations, *J. Am. Chem. Soc.*, 2004, **127**, 476–477.

11 Y.-H. Sung and D. Eliezer, Residual structure, backbone dynamics, and interactions within the synuclein family, *J. Mol. Biol.*, 2007, **372**, 689–707.

12 C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin and M. Zweckstetter, Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1430–1435.

13 M. Schwalbe, V. Ry Ozenne, S. Bibow, M. Jaremko, L. Jaremko, M. Gajda, M. R. Jensen, J. Biernat, S. Becker, E. Mandelkow, M. Zweckstetter and M. Blackledge, Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution from NMR and Small Angle Scattering, *Structure*, 2014, **22**, 238–249.

14 K.-P. Wu, D. S. Weinstock, C. Narayanan, R. M. Levy and J. Baum, Structural Reorganization of α-Synuclein at Low pH Observed by NMR and REMD Simulations, *J. Mol. Biol.*, 2009, **391**, 784–796.

15 I. M. Ilie, D. Nayar, W. K. Den Otter, N. F. A. Van Der Vegt and W. J. Briels, Intrinsic Conformational Preferences and Interactions in α-Synuclein Fibrils: Insights from Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2018, **14**, 3298–3310.

16 Y. Zhang, M. Hashemi and Z. Lv, High-speed atomic force microscopy reveals structural dynamics of-synuclein monomers and dimers, *J. Chem. Phys*, 2018, **148**, 123322.

17 J. Chen, S. Zaer, P. Drori, N. Kalisman, E. Lerner and N. V Dokholyan, The structural heterogeneity of a-synuclein is governed by several distinct subpopulations with interconversion times slower than milliseconds, *Struct. Des.*, 2021, **29**, 1048-1064.

18 H. Yu, W. Han, W. Ma and K. Schulten, Transient β-hairpin formation in α-synuclein monomer revealed by coarse-grained molecular dynamics simulation, *J. Chem. Phys.*, 2015, **143**, 10089.

19 R. Ramis, J. Ortega-Castro, R. Casasnovas, L. Mariñ, B. Bartolomévilanova, M. Adrover and J. Frau, A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein α-Synuclein, *J. Chem. Inf. Model.*, 2019, **59**, 1458–1471.

20 K. Jain, O. Ghribi and J. Delhommelle, Folding Free-Energy Landscape of α-Synuclein (35−97) Via Replica Exchange Molecular Dynamics, *J. Chem. Inf. Model.*, 2020, **61**, 432–443.

21 J. R. Allison, R. C. Rivers, J. C. Christodoulou, M. Vendruscolo and C. M. Dobson, A Relationship between the Transient Structure in the Monomeric State and the Aggregation Propensities of α-Synuclein and β-Synuclein, *Biochemistry*, 2014, **53**, 28.

22 M. C. Ahmed, L. K. Skaanning, A. Jussupow, E. A. Newcombe, B. B. Kragelund, C. Camilloni, A. E. Langkilde and K. Lindorff-Larsen, Refinement of α-Synuclein Ensembles Against SAXS Data: Comparison of Force Fields and Methods, *Front. Mol. Biosci.*, 2021, **8**, 216.

23 L. Maroteaux, J. T. Campanelli and R. H. Scheller, Synuclein: A Neuron-Specific Protein Localized to the Nucleus and Presynaptic Nerve Terminal, *J. Neurosci.*, 1988, **8**, 2804–2815.

24 J. T. Bendor, T. P. Logan and R. H. Edwards, The Function of α-Synuclein, *Neuron*,

2013, **79**, 1044–1066.

25 R. Bussell and D. Eliezer, A Structural and Functional Role for 11-mer Repeats in α-Synuclein and Other Exchangeable Lipid Binding Proteins, *J. Mol. Biol.*, 2003, **329**, 763–778.

26 B. I. Giasson, I. V. J. Murray, J. Q. Trojanowski and V. M. Y. Lee, A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of α-Synuclein Is Essential for Filament Assembly, *J. Biol. Chem.*, 2001, **276**, 2380–2386.

27 A. Balupuri, K.-E. Choi and N. S. Kang, Computational insights into the role of α-strand/sheet in aggregation of α-synuclein, *Sci. Rep.*, 2019, **9**.

28 J. O. Henriques, C. Cragnell and M. Skepö, Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment, *J. Chem. Theory Comput.*, 2015, **11**, 3420-3431.

29 W. Kang, F. Jiang and Y. D. Wu, How to strike a conformational balance in protein force fields for molecular dynamics simulations?, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, e1578.

30 S. Piana, K. Lindorff-Larsen, D. E. Shaw, † D E Shaw Research and N. York, How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization?, *Biophysj*, 2011, **100**, L47–L49.

31 S. Yang, H. Liu, Y. Zhang, H. Lu and H. Chen, Residue-Specific Force Field Improving the Sample of Intrinsically Disordered Proteins and Folded Proteins, *J. Chem. Inf. Model*, 2019, **59**, 52.

32 M. J. Robertson, J. Tirado-Rives and W. L. Jorgensen, Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field, *J. Chem. Theory Comput.*, 2015, **11**, 3499–3509.

33 J. Huang and A. D. Mackerell, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, *J. Comput. Chem.*, 2013, **34**, 2135–2145.

34 D. Song, R. Luo and H.-F. Chen, The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins, *J. Chem. Inf. Model.*, 2017, **57**, 1166–1178.

35 D. Song, H. Liu, R. Luo and H. F. Chen, Environment-Specific Force Field for Intrinsically Disordered and Ordered Proteins, *J. Chem. Inf. Model.*, 2020, **60**, 2257–2267.

36 D. E. Robustelli, Paul; Piana, Stefano; Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E4758–E4766.

37 R. B. Best, W. Zheng and J. Mittal, Balanced Protein−Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association, *J. Chem. Theory Comput.*, 2014, **10**, 5124.

38 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H.

Grubmüller and A. D. Mackerell, CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins HHS Public Access, *Nat Methods*, 2017, **14**, 71–73.

39 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput*, 2015, **11**, 3696–3713.

40 D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018, 2018, University of California, San Francisco.

41 C. W. Hopkins, S. Le Grand, R. C. Walker and A. E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, J. Chem. Theory Comput., 2015, **11**, 1864-1874.

42 J. J. Perez, M. Santos Tomas and J. Rubio-Martinez, Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory, *J. Chem. Inf. Model.*, 2016, **56**, 1950–1962.

43 W. Humphrey, A. Dalke and K. Schulten, VMD: Visual Molecular Dynamics, *J. Mol. Graph.*, 1996, **14**, 33–38.

44 T. Williams, C. Kelley, J. Campbell, R. Cunningham, D. Denholm, G. Elber, R. Fearick, C. Grammes, L. Hart, L. Hecking, T. Koenig, D. Kotz, E. Kubaitis, R. Lang, A. Lehmann, A. Mai, M. Bastian, E. a Merritt, P. Mikul, T. Tkacik, J. Van Der Woude, J. R. Van Zandt, A. Woo and J. Zellner, Gnuplot 4.6: an interactive plotting program, *Softw. Man.*, 2012, **238**.

45 D. R. Roe and T. E. Cheatham, PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

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

47 Y. Shen and A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, *J. Biomol. NMR*, 2010, **48**, 13–22.

48 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro,

F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat. Methods*, 2020, **17**, 261–272.

49 F. Vandova, Gergana; Tamiola, Kamil; Oktaviani, Nur; Mulder, Backbone 1H, 13C,and 15N chemical shift assignments for alpha-synuclein at different pH and temperature.

50 L. Kang, M. K. Janowska, G. M. Moriarty and J. Baum, Mechanistic Insight into the Relationship between N-Terminal Acetylation of α-Synuclein and Fibril Formation Rates by NMR and Fluorescence) Mechanistic Insight into the Relationship between N-Terminal Acetylation of α-Synuclein and Fibril Formation Rates by NMR and Fluorescence, *PLoS One*, 2013, **8**, 75018.

51 R. Porcari, C. Proukakis, C. A. Waudby, B. Bolognesi, P. Patrizia Mangione, J. F. S. Paton, S. Mullin, L. D. Cabrita, A. Penco, A. Relini, G. Verona, M. Vendruscolo, M. Stoppini, G. Gaetano Tartaglia, C. Camilloni, J. Christodoulou, A. H. V Schapira and V. Bellotti, The H50Q Mutation Induces a 10-fold Decrease in the Solubility of α-Synuclein*, *J. Biol. Chem.*, 2015, **290**, 2395–2404.

52 W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat and R. Pierattelli, Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins, *J. Am. Chem. Soc.*, 2006, **128**, 3918–3919.

53 J. Pujols, S. Peña-Díaz, D. F. Lázaro, F. Peccati, F. Pinheiro, D. González, A. Carija, S. Navarro, M. Conde-Giménez, J. García, S. Guardiola, E. Giralt, X. Salvatella, J. Sancho, M. Sodupe, T. F. Outeiro, E. Dalfó and S. Ventura, Small molecule inhibits α-synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 10481–10486.

54 N. I. Brodie, K. I. Popov, E. V Petrotchenko, N. V Dokholyan ID and C. H. Borchers ID, 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**, e1006859.

55 T. Graen, R. Klement, A. Grupi, E. Haas and H. Grubmüller, Transient Secondary and Tertiary Structure Formation Kinetics in the Intrinsically Disordered State of α-Synuclein from Atomistic Simulations, *ChemPhysChem*, 2018, **19**, 2507–2511.

56 M. A. Healey, M. T. Woodside and J. A. Tuszynski, Phase transitions and structure analysis in wild-type, A30P, E46K, and A53T mutants of α-synuclein, *Eur. Biophys. J.*, 2016, **45**, 355–364.

57 U. Dettmer, A. J. Newman, V. E. Von Saucken, T. Bartels and D. Selkoe, KTKEGV repeat motifs are key mediators of normal α-synuclein tetramerization: Their mutation causes excess monomers and neurotoxicity, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9596–9601.

58 R. M. Meade, D. P. Fairlie and J. M. Mason, Alpha-synuclein structure and Parkinson's disease-lessons and emerging principles, *Mol. Neurodegener.*, 2019, **14**.

59 R. Duong, Vy T.; Chen, Zihao; Thapa, Mahedra T., Luo, Computational Studies of Intrinsically Disordered Proteins, *J. Phys. Chem. B*, 2018, **122**, 10455–10469.

60 L. M. Reid, I. Guzzetti, T. Svensson, A.-C. Carlsson, W. Su, T. Leek, L. Von Sydow, W. Czechtizky, M. Miljak, C. Verma, L. De Maria and J. W. Essex, How well does molecular simulation reproduce environment-specific conformations of the intrinsically disordered peptides PLP, TP2 and ONEG?, *Chem. Sci.*, 2022, **13**, 1957–197.

61 A. Dan and H. F. Chen, Secondary structures transition of tau protein with intrinsically disordered proteins specific force field, *Chem. Biol. Drug Des.*, 2019, **93**, 242–253.

62 M. Ur Rahman, A. Ur Rehman, H. Liu and H.-F. Chen, Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins, *J. Chem. Inf. Model.*, 2020, **60**, 4912–4923.