Published in final edited form as:

ACS Chem Neurosci. 2018 May 16; 9(5): 1051–1065. doi:10.1021/acscchemneuro.7b00446.

From a highly-disordered to a metastable state: Uncovering insights of α -synuclein

Yoann Cote^{1,2}, Patrice Delarue², Harold A. Scheraga³, Patrick Senet^{2,3,*}, and Gia G. Maisuradze^{3,*}

¹Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR 7104 - INSERM U 964, Université de Strasbourg, 1 rue Laurent Fries, 67400 Illkirch-Graffenstaden, France

²Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS - Univ. Bourgogne Franche-Comté, 9 Av. Alain Savary, BP 47 870, F-21078 Dijon Cedex, France

³Baker Laboratory of Chemistry and Chemical Biology, Cornell University Ithaca, New York 14853-1301

Abstract

α -synuclein (α S) is a major constituent of Lewy bodies, the insoluble aggregates that are the hallmark of one of the most prevalent neurodegenerative disorders - Parkinson's disease (PD). The vast majority of experiments *in vitro* and *in vivo* provide extensive evidence that a disordered monomeric form is the predominant state of α S in water solution, and it undergoes a large-scale disorder-to-helix transition upon binding to vesicles of different types. Recently, another form, tetrameric, of α S with a stable helical structure was identified experimentally. It has been shown that a dynamic intracellular population of metastable α S tetramers and monomers coexists normally; and the tetramer plays an essential role in maintaining α S homeostasis. Therefore, it is of interest to know whether the tetramer can serve as a means of preventing or delaying the start of PD. Before answering this very important question, it is, first, necessary to find out, on an atomistic level, a correlation between tetramers and monomers; what mediates tetramer formation and what makes a tetramer stable. We address these questions here by investigating both monomeric and tetrameric forms of α S. In particular, by examining correlations between the motions of the side chains and the main chain, steric parameters along the amino-acid sequence, and one- and two-dimensional free-energy landscapes along the coarse-grained dihedral angles γ and δ and principal components, respectively, in monomeric and tetrameric α S, we were able to

*To whom correspondence may be addressed: psenet@u-bourgogne.fr; gm56@cornell.edu.

Author contributions: P.S. and G.G.M. designed the research; Y.C., P.D., P.S., and G.G.M. performed the research; Y.C., P.D., P.S., and G.G.M. analyzed the data; and Y.C., P.D., H.A.S., P.S., and G.G.M. wrote the paper.

The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the <http://pubs.acs.org>

Definition of the coarse-grained dihedral angles (γ and δ); Evolution of correlation coefficient and steric parameter over time for lys21; Secondary structure evolution as a function of time for the partially-disordered and micelle-bound monomeric α S; Free-energy profiles of the dihedral angles γ and δ for the partially-disordered monomeric α S; Evolution of the steric parameter and solvent-accessible surface area along the amino-acid sequence of the highly-disordered monomeric α S; 2-D maps for the distances between the C α s of the lysine residues and all amino acids of chains A, B, C, and D of tetrameric α S.

shed light on a fundamental relationship between monomers and tetramers, and the key residues involved in mediating formation of a tetramer. Also, the reasons for the stability of tetrameric α S, and inability of monomeric α S to fold are elucidated, here.

Graphical Abstract

Keywords

(i) monomeric alpha-synuclein; (ii) tetrameric alpha-synuclein; (iii) Parkinson's disease; (iv) all-atom molecular dynamics simulations; (v) free-energy landscape; (vi) dihedral principal component analysis

1. Introduction

α -synuclein (α S) is the principal protein component of the Lewy body and Lewy neurite deposits that are found in the brains of the victims of Parkinson's disease (PD)¹ – the second most common neurodegenerative disorder (after Alzheimer's disease) nowadays with a significant rise in the number of affected patients and cost of care.² α S is a 140 residue protein whose function in the healthy brain is still unclear, although recent studies revealed that α S acts as a molecular chaperone for the formation of SNARE complexes;³ it plays a key role in regulating the process of neurotransmitter release, which is associated with the mediation of synaptic vesicle interactions and assembly,⁴ and, also, has an inhibitory function on membrane fusion.⁵ α S is characterized by three domains: (i) an amphipathic N-terminal region dominated by four 11-residue repeats including the consensus sequence KTKEGV (residues 1–60);⁶ (ii) a central hydrophobic region, which includes the non-amyloid- β component (NAC) region and two additional repeats, involved in protein aggregation (residues 61–95);⁷ (iii) a highly acidic and proline-rich region which has no distinct structural propensity (residues 96–140).⁸

The vast majority of experiments provide extensive evidence that α S exists mainly as a disordered monomer [intrinsically disordered protein (IDP)^{9,10}] in solution, and undergoes a large-scale disorder-to-helix transition upon binding to vesicles of different types,^{11–16} which is a feature of proteins containing amphipathic helices, such as Hsp12¹⁷ and a subclass of proteins often referred to as amphitropic.¹⁸ However, some experiments suggest that α S can adopt a folded tetrameric state that has a relatively high helical content under physiological conditions.^{19,20} It should be noted that because of the difficulties of studying native, cellular α S outside of the cell with conventional biochemical approaches, the discovery of native tetramers and closely related multimers has received a lot of criticism

from some experimental groups^{13–16} who were defending the almost exclusive existence of disordered monomeric α S in cells; while other experimental groups^{21–23} observed folded multimers in human tissue. The monomer-tetramer controversy has been reconciled by observing that tetrameric α -synucleins, in contrast to other normal oligomeric proteins, dissociate rapidly to monomers upon conventional cell lysis but are retained partially as tetramers if cells are lysed at high protein concentrations.²⁴ Moreover, cellular membrane lipids are thought to play an important role in the regulation of tetrameric α S formation. Recently Kim et al.²⁵ provided mechanistic insights into how glucocerebrosidase 1 regulates the transition from monomeric α S to tetrameric and multimeric α S. Thus, α S exists natively as helical tetramers that are in dynamic equilibrium with disordered monomers, which is the dominant state of the ensemble. Similar conclusions were reached in theoretical studies by Gurry et al.,²⁶ although in that study apart from helical tetramers the ensemble also contained trimeric and tetrameric oligomers that are rich in β -strand content. The members of the synuclein family share conserved repeat motifs (consensus: KTKEGV) that occur at least six times in α S. Recently, Dettmer et al.²⁷ demonstrated that α S repeat motifs mediate physiological tetramerization, and perturbing them causes PD-like neurotoxicity. Because of its disordered nature the monomeric α S is prone to aggregation, whereas tetramers undergo little or no amyloid-like aggregation,¹⁹ suggesting the importance of the tetrameric form in maintaining α S homeostasis. The important question is how to deal with stabilizing the physiological tetramer as a means of preventing or delaying the start of PD.

There are several issues to be addressed before answering this important question. It is known that the mobile flexibility and structural instability of IDPs are encoded in their amino-acid sequences. Therefore, it is of interest to know (i) what governs the way that the monomeric α S lacks a fixed or ordered three-dimensional structure; (ii) what is the physics of the interactions between the main chain (MC) and side chains (SCs), which determines an ensemble of rapidly interconverting structures characteristic of the monomeric α S in solution;²⁸ (iii) what kind of relationship exists between disordered monomeric α S and stable tetrameric α S. These questions have been addressed in the present study by investigating both forms, monomeric¹² and tetrameric,²⁰ of α S.

The dynamics of disordered monomeric α S is examined by performing fifty (10 ns of each) all-atom molecular dynamics (MD) simulations at 310K starting from completely disordered structures generated by flexible-meccano software.²⁹ The stability of secondary structures in tetrameric α S is investigated by performing all-atom MD simulations at room temperature (300K) and at an artificially high temperature (500K) of a dynamic model for the tetramer structure.²⁰

There is strong evidence^{30–33} that protein folding is initiated and governed by local motions of the residues which, subsequently, result in global conformational changes and vice versa. In particular, the motions of the nonpolar SCs are one of the (hydrophobic) forces driving the folding of a protein: these motions orient the MC locally. Similarly, the formation of hydrogen bonds between the amide N–H bonds and the carbonyl C=O groups of the MC affects local motions of the SCs. Recently, studying the folding/unfolding dynamics of a model protein (Trp-cage), we have shown that, in the unfolded state, the SCs and the MC are strongly coupled to each other, unlike the SCs and MC of residues in secondary structures in

the native state.³⁴ The relevance of this fundamental property of the unfolded state to IDPs is examined here.

2. Results

2.1. Monomeric α -Synuclein

2.1.1. Correlation between the motions of the SCs and the MC—We ran fifty MD trajectories starting from completely disordered conformations generated by flexible-meccano software.²⁹ All fifty MD trajectories were monitored by computing the root-mean-square deviation (RMSD) from the initial structures (Fig. 1A), and the secondary structure evolution (Fig. 1B) as functions of time. The type of secondary structure conformation of each residue of the monomeric α S was determined for each frame of the entire MD trajectories at 310K. The results of RMSDs, and the secondary structure evolution indicate that the monomeric α S in all fifty MD trajectories retains highly disordered structure.

The local coupling between the SCs and the MC in monomeric α S was expressed in terms of the Pearson correlation coefficient R ,^{34,36,37} between the time series $\gamma_H(t)$ and $\delta_H(t)$, averaged over time, extracted from MD simulations of a protein at 310K.

The correlation coefficient R for highly disordered monomeric α S at 310 K (Fig. 2A) varies in an oscillatory manner and exhibits high values for coarse-grained dihedral angles (CGDAs) along the entire amino-acid sequence ($R_{\text{ave}} \sim 0.7$; $R_{\text{min}} = 0.60$ and $R_{\text{max}} = 0.83$). The amplitude and frequency of oscillation are lower in the membrane-binding and nonamyloid component domains than in the acidic domain. High values of R indicate a high flexibility of the MC and SCs, and point out that monomeric α S is an IDP. An increase of an average value of R (by ~ 0.2) in the acidic domain, caused mainly by proline residues, indicates the increase of disorder in this domain. It should be pointed out that most of the minima in the correlation coefficient curve coincide with KTKEGV repeats, important players in which are lysine and threonine residues (KTK, KT), lysine and alanine residues (KAK), or just lysine residues (KK) (red diamonds in Fig. 2A). It has been shown earlier³⁸ that the KTKEGV repeat motifs, and specifically, the positively charged lysine residues play a critical role in mediating the binding of α S to membrane phospholipids, and in regulating α S oligomer formation and α S aggregation. Also, as was mentioned above in the Introduction, Dettmer et al.²⁷ demonstrated that KTKEGV repeat motifs mediate α S physiological tetramerization, and perturbing them causes PD-like neurotoxicity. The reasons for low values of R at particular residues are addressed here. We have shown before³⁴ that the decrease of the R values is related to folding events during which the MC of the concerned residues acquire a specific secondary structure. However, by comparing the evolutions of $R(t)$'s for the regions formed by lysine residues and for the residues forming the secondary structure with smallest R values, we found that the decrease of $R(t)$ in the regions formed by lysine residues is not caused by the formation of the secondary structure. In order to understand why the values of $R(t)$ decrease for the lysine residues during small time windows along the full length of the MD simulation, we have compared the evolutions of $R(t)$ and $s_H(t)$ for one of the lysine residues - lys21 (Fig. S2).

It should be noted that the steric parameter, s_n , was introduced, in our recent study,³⁴ to characterize structural fluctuations of the Trp-cage. It measures the distance of the n^{th} SC relative to the rest of the protein. A low value of s_n means that the SC_{*n*} is far from the rest of the protein. On the contrary, a high value of s_n means that the SC_{*n*} occurs at short distances from the rest of the protein. Here, s_n was employed to study the dynamics of the α S.

The results showed that fluctuations of $R(t)$ and $s_n(t)$ are correlated (Fig. S2). Indeed, an increase of $s_n(t)$ is related with a decrease of $R(t)$ and vice versa. In other words, when SC of lysine is exposed to water [low values of $s_n(t)$], it is more flexible and strongly-correlated with the MC [high values of $R(t)$]; and when SC of lysine is partially buried [high values of $s_n(t)$], it interacts with other residues (Fig. S3) and becomes less flexible and weakly-correlated with the MC [low values of $R(t)$]. This type of dual behavior of lysine is rooted in its large size and amphipathic nature. Other minima in the correlation coefficient curve are mainly formed by tyrosine (tyr39, tyr125, tyr133) and phenylalanine (phe4, phe94) residues, which can be explained by their size and nature. In particular, because both tyrosine and phenylalanine are very large, aromatic, hydrophobic residues, they prefer to be buried in protein hydrophobic cores, and are involved in stacking interactions, hence they are less flexible. We also examined the maxima in the correlation coefficient curve. As was expected, most of the maxima in the membrane-binding and nonamyloid component domains are formed by the smallest residue glycine (green diamonds in Fig. 2A). Thus, the obtained results revealed important sites and features of monomeric α S in the highly-disordered state.

In order to find out whether these findings are characteristic of monomeric α S only in highly-disordered state or they are an intrinsic property of monomeric α S, we have performed (i) six all-atom MD simulations at 310K (10 ns of each) with initial structures extracted from 60 ns unfolding MD trajectory of the micelle-bound α S¹² at 500K (each structure was extracted after every 10 ns time interval); and (ii) one 100 ns MD trajectory of the micelle-bound α S¹² at 300K. The secondary-structure evolution as a function of time (Fig. S4) indicates that monomeric α S in all six MD trajectories at 310K is partially disordered (panel A), while in the 100 ns MD trajectory at 300K it retains a secondary structure in the membrane-binding domain and in a large portion of a nonamyloid component domain during the entire trajectory (panel B).

The correlation coefficients for partially disordered monomeric α S at 310 K and the micelle-bound α S¹² at 300K are illustrated in panels B and C of Fig. 2, respectively. Oscillatory behavior persisted more or less in both curves; however, R values are lower and the amplitudes of oscillations are larger than in highly disordered monomeric α S ($R_{\text{ave}} \sim 0.58$, $R_{\text{min}} = 0.38$ and $R_{\text{max}} = 0.75$ for partially disordered monomeric α S; and $R_{\text{ave}} \sim 0.56$, $R_{\text{min}} = 0.34$ and $R_{\text{max}} = 0.78$ for the micelle-bound monomeric α S). These results indicate that flexibility of the MC and SCs decreases in partially disordered and micelle-bound monomeric α S, and some regions (with very low values of R) are very stable and rigid. Moreover, as in highly disordered monomeric α S, most of the minima in the R curve correspond to KTK, KT, KAK and KK combinations, although there are some changes, especially in micelle-bound monomeric α S, related to formation of secondary structures.

These results indicate that important sites revealed in the highly-disordered state remain in partially-disordered and stable states.

2.1.2. Free-energy profiles along the CGDAs—To relate the variation of R to the behavior of the side chains and the main chain in highly disordered monomeric α S, the effective one-dimensional free-energy profiles (FEPs), $V(\delta_n)$ and $V(\gamma_n)$,^{34,36,39,40} along the amino-acid sequence were computed (Fig. 3). The FEPs were computed by using $V(\delta_n) = -kT \ln[P(\delta_n)]$ and $V(\gamma_n) = -kT \ln[P(\gamma_n)]$, where k and T are the Boltzmann constant and temperature, respectively; and $P(\delta_n)$ and $P(\gamma_n)$ are the probability distribution functions (PDFs) of δ_n and γ_n angles, respectively.

The results show that most of the dihedral angles have a flat 1-D FEP with a small barrier centered around -50° , which reveal that dihedral angles γ_n and δ_n probe all the conformations between $[-180^\circ:180^\circ]$ during the full MD runs. A few exceptions appear for the dihedral angles $n = 107, 116, 119, 127$ and 137 , in which γ_n and δ_n explore a smaller conformational space. In particular, the $[-50^\circ:120^\circ]$ and $[-10^\circ:30^\circ]$ regions are unexplored by $n = 116, 127, 137$ and $n = 107, 119$ residues, respectively, and two minima are exhibited in the FEPs of latter angles. Such a behavior of these five angles can be explained by the presence of proline residues in each angle. The point is that, because of its unique structure, proline is unable to occupy many of the MC conformations easily adopted by all other amino acids. The results for the FEPs $V(\gamma_n)$ and $V(\delta_n)$, overall, demonstrate complete flexibility of the backbone and the SCs of the monomeric α S in a highly-disordered state. Some minor exceptions (the residues $n = 107, 116, 119, 127$ and 137) do not change the general picture, plus, these residues pertain to the acidic domain – the most disordered part of the monomeric α S. Although the R value oscillates along the amino acid sequence (Fig. 2A), the amplitude of oscillation is small, which maintains the R value high (> 0.6); consequently, variations of R are not reflected on the shapes of FEPs.

The FEPs results for partially disordered α S (Fig. S5), computed from initial structures extracted from unfolding trajectories of micelle-bound α S, show that the dihedral angles γ_n and δ_n explore a smaller conformational space, in general, than γ_n and δ_n angles for highly disordered α S shown in Fig. 3. Three different shapes of FEPs can be distinguished in Fig. S5: (i) γ_n and δ_n angles explore a smaller conformational space with the minimum at 60° ; (ii) γ_n and δ_n angles explore a larger conformational space having multiple local minima; and (iii) γ_n and δ_n angles probe entire conformational space without having any distinguishable minimum. The first shape of the FEPs corresponds to an α -helix conformation, which indicates the secondary structure regions in partially disordered α S, and R values for these angles are usually low (Fig. 2B). The second and third shape of the FEPs indicate the flexible and disordered regions in partially disordered α S; the R values for these angles are usually high (Fig. 2B).

Interestingly, the FEPs with the first shape in Fig. S5 correspond to the FEPs with a small barrier centered around -50° in Fig. 3, and the FEPs with the second and third shapes in Fig. S5 correspond to the FEPs without any barrier in Fig. 3. This finding indicates that the formation of helical conformations is encoded in an α S sequence, some “hints” of which can be observed even in highly-disordered state.

2.1.3. Characterization of protein fluctuations by the steric parameter—Figure 4A illustrates the values of s_n for highly disordered monomeric α S averaged over the entire duration of all 50 MD trajectories at 310 K. The value of s_n for the majority of the residues varies between ~ 65 and 86 (average $s_n \sim 77$) (black diamonds in Fig. 4A); however, for some residues the value of s_n is comparably lower (varies between ~ 42 and 64 , average $s_n \sim 60$), and surprisingly most of these residues are alanine (red diamonds in Fig. 4A). Because alanine is one of the smallest amino acids, according to the definition of s_n , the values of s_n for alanine residues should be high. In order to find out the reason for low s_n values for alanine, we have computed solvent-accessible surface area (SASA) of all residues of monomeric α S averaged over the entire duration of all 50 MD trajectories. The results (in percentages), illustrated in Fig. S6 (red diamonds), correspond to the values of the SASA computed by Gromacs for each residue and divided by the tabular SASA values of the corresponding residues. For comparison, the values of s_n (blue diamonds) are also depicted on the same panel. As the results show, most of the alanine residues are fully exposed to the solvent during the simulations, which explains the low values of s_n . Moreover, visual monitoring of the MD trajectories have shown that the chain of α S is well kinked at the positions of the alanine (and the glycine) residues, which is another explanation for low values of s_n at alanine residues.

In order to find out whether this behavior of s_n is characteristic of monomeric α S in a highly-disordered state or it is an intrinsic property of monomeric α S, we computed the values of s_n for the partially disordered monomeric α S averaged over the entire duration of all six MD trajectories at 310 K, and for the stable micelle-bound monomeric α S averaged over the entire duration of one 100 ns MD trajectory at 300K. The results illustrate that the low values of s_n at alanine residues persist for partially disordered monomeric α S (Fig. 4B) and for stable micellebound monomeric α S (Fig. 4C), which indicate that the low values of s_n at alanine residues is an intrinsic property of monomeric α S.

2.2. Tetrameric α -Synuclein

2.2.1. Correlation between the motions of the SCs and the MC—Figure 5 shows the evolution of the correlation coefficient along the amino-acid sequence of each chain at 300K and 500K. The results indicate that important sites (KTK, KT, KAK, KK), revealed in the highly-disordered state (Fig. 2A), remained in the correlation coefficient curves of tetrameric α S (red diamonds); however, because of secondary structures they are not as prominent as in highly disordered monomeric α S. As in monomeric α S, many maxima in the membrane-binding and nonamyloid component domains of tetrameric α S are formed by the smallest residue glycine (green diamonds). Moreover, the correlation coefficient in tetrameric α S follows the same trend in all four chains. As in monomeric α S, the correlation coefficient of tetrameric α S at 300K “oscillates” along the amino-acid sequence, and exhibits both low and high values for CGDAs located at both the secondary structure [$R_{\min} = 0.28$ for $n = 80$ of the chain A (panel A) and $R_{\max} = 0.70$ for $n = 67$ of the chain D (panel D)] and flexible [$R_{\min} = 0.31$ for $n = 4$ of the chain C (panel C) and $R_{\max} = 0.82$ for $n = 118$ of the chain C (panel C)] regions. However, the average value of R of the secondary structures in tetrameric α S is lower by ~ 0.1 than one in micelle-bound monomeric α S,

which indicates lower flexibility of MC and SCs and larger stability of secondary structures in tetrameric α S compared to micelle-bound monomeric α S.

The values of the correlation coefficient of tetrameric α S increase with temperature and “oscillate” along the amino-acid sequence (Fig. 5). However, the average values of R of secondary-structure regions at 500K are still low. This indicates that tetrameric α S is stable and retains the secondary structures (partially) at higher temperatures. The obtained results are in agreement with experiment.²⁰ It should be mentioned that the R values of some residues from the acidic domain are surprisingly dropping down with the increase of temperature lowering average values of R of the acidic domains of chains A, B (in a less extent), C and D.

2.2.2. FEPs along γ_n angles—The FEPs of the dihedral angles γ_n ($2 < n < 138$) for each chain of the tetramer were computed from the MD runs at 300K (Fig. 6A) and at 500K (Fig. 6B). The results show that both α -helices are well conserved in the MD run at 300K in all four chains of the tetramer. However, some portions ($n = 12 - 17, 30, 34, 63, 66 - 71, 75, 79$) of both helices in some chains exhibit instability by having extra minima or wide single wells, which is manifested in high values of the correlation coefficients corresponding to these angles. Interestingly, the loop regions of the tetrameric α S show local conformations which are different in function of the chains as is found for the residues of the C-terminal region ($n > 97$) and for the loops joining the two α -helices ($n = 37 - 51, 84 - 90$).

The conformational space explored by the dihedral angles γ_n of the tetrameric α S at 500K (Fig. 6B) is significantly smaller than one in the highly disordered monomeric α S at 310K (Fig. 3), and slightly larger than one in the partially disordered monomeric α S at 310K (Fig. S5). The conformational space explored by tetramer chains differ from each other. In particular, the backbone of chain C explores a wider conformational space than the other three chains of the tetramer. The backbone of chain A explores the smallest conformational space. In particular, the residues located in the first α -helix ($n = 8 - 23$) of chain A (Fig. 6B) explore the smallest conformational space. The FEPs $V(\gamma_n)$ of the residues of chain D located in the second α -helix ($n = 54 - 83$) are in a single well and their minima are centered at 60° whereas several residues of chains A, B and C ($n = 66 - 73$) explore the largest conformational space and have wide FEPs with one deep well and a flat region. These results are another proof that tetrameric α S is a stable protein, and retains secondary structures (partially) at higher temperatures.

2.2.3. Steric constraints in tetrameric α S—The steric parameter, s_n , was computed for each residue of tetrameric α S from the MD runs carried out at 300K and 500K, and its value was averaged over the full length of the MD runs (Fig. 7). It can first be seen that the values of s_n of the residues of the tetramer are ~ 2.5 times higher than the values of s_n in monomeric α S (Fig. 4). This result can be attributed to the presence of four chains in the system. The residues located in secondary structures have the highest value imposed by the constraint of the helical conformation and the residues located in loop regions show the smallest value of s_n . Nevertheless, small values of s_n for the residues of the first α -helix of chain A are observed at 300K (Fig. 7A), which could be the consequence of a conformational rearrangement in the loop joining α -helices 1 and 2 of chain A, and also a

rearrangement of the unstructured C-terminal regions of each chain of the tetramer. In particular, by monitoring the MD run at 300K, we could observe the following: C-terminal regions ($n > 97$) of each chain start the MD simulation with an extended conformation; however, chains B, C and D [Fig. 7(B, C, D), left panels] adopt a compact conformation around the stable core of the tetramer formed by the α -helical regions of chains B, C and D during the simulation. In the MD trajectory at 300K, the first α -helix of chain A adopts a lever motion starting from $t > 85$ ns (Fig. S7). The lever motion results in a drift of this α -helix away from the stable core of the tetramer.

The increase of temperature from 300K to 500K changed the values of s_n slightly, in general. This indicates that the quaternary structure of the system did not change significantly. Small deviations between the values of s_n of the chains are found in the results from the MD run carried out at 500K (Fig. 7, right panels). In fact, at 500K, the steric parameter of the residues of chain A located in the loop joining the first two α -helices, and of the second α -helix ($n = 34 - 70$), is smaller than the steric parameters of the same residues in chains B, C and D, except for residues $n = 34 - 38$ of chain B which have similar or lowest s_n values (Fig. 7). The results from the C-terminal region of the chains indicate that chains B and C have stronger constraints on residues $n = 67 - 117$ for chain B and residues $n = 85 - 128$ for chain C.

Moreover, we monitored the values of s_n at alanine residues at both temperatures (red diamonds). The low values of s_n at alanine residues are still dominant at both temperatures, which indicates on intrinsic nature of this property.

2.3. Thermal denaturation of tetrameric α S

Recent experimental studies²⁰ of tetrameric α S showed not only the stability of a tetramer in the absence of lipid bilayers or micelles, but also the presence of a hydrophobic core, which seems to be responsible for tetramer formation.

Here, we address these aspects of α S by analyzing the MD trajectories with different approaches. In the first approach, we build the free-energy landscapes (FELs) along the first two principal components (PCs) for the entire MD trajectories of tetrameric α S at 300K (Fig. 8A) and 500K (Fig. 8B).

As illustrated in Fig. 8, tetrameric α S appears to be stable at both temperatures. First, it is very stable at low, 300K, temperature (Fig. 8A); not only do the secondary structures but also the quaternary structure almost do not change during the 100 ns MD simulation (see representative structures of minima). The stability of the tetramer seems to be caused by the second α -helices of each chain forming the core of the complex. More precisely, the nonpolar residues of the second α -helices interact and keep the tetramer stable.

With the increase of temperature to 500K some parts of the helices lose secondary structure, and the stability of the tetramer decreases; however, some features of the tetramer are still present (Fig. 8B). In particular, by monitoring the MD run of the tetramer at 500K, we observed that the second α -helices of each chain quickly approach each other and the second α -helix of chain D pushes the second α -helix of chain C away ($t < 300$ ps). Therefore, in the

initial stage of the trajectory, chains A, B, C and D form a quaternary structure, similar to the structure at 300K, with chain C slightly separated from the core of the tetramer (see Fig. S8). The entire MD trajectory consists of fluctuations of the rest of the regions of the chains around the stable core formed by the second α -helices of chains A, B and D. For example, (i) the first α -helix of chain B participates in formation of the stable core by interacting with the second α -helices of chains A and D (see the representative structures of minima 6 – 13 of FEL in Fig. 8B); (ii) The C-terminal regions of each chain quickly contract and stay close to the core for the rest of the trajectory (compare the representative structures of minima 2 and 3 in Fig 8B); (iii) The region formed by the end of the first α -helix and the beginning of the second α -helix of chains A and C ($n = 30 - 60$) exhibits large fluctuations.

In order to justify and strengthen the results obtained by the FELs, we have studied the evolutions of secondary structures in the course of the entire MD trajectories. In particular, the type of secondary structure conformation of each residue of tetrameric α S (Fig. 9) was determined for each frame of the entire MD trajectories at 300K and at 500K. The color code used in Figs. 9 is the same as in Fig. 1B.

It is clear from Fig. 9 that the secondary structures are better conserved in the MD runs at 300K than in those performed at 500K. The results extracted from the MD run of the tetramer performed at 300K (Fig. 9A) show that the secondary structures are very stable in all four chains with partial unfolding occurring at the end of the first α -helix and at the beginning of the second α -helix of the four chains. In particular, (i) in chain A, the first α -helix is very well conserved during the entire MD run whereas the beginning of the second α -helix ($n = 49 - 53$) unfolds in the time-intervals $10 < t < 27$ ns and $70 < t < 85$ ns; (ii) in chain B, the residues $n = 34 - 38$ of the first α -helix show unfolding/folding events during the entire MD simulation whereas the beginning of the second α -helix is well conserved. Only a small break of the second α -helix conformation on residues $n = 67$ and 68 for $t > 5$ ns can be observed; (iii) in chain C, breaks of the first α -helix on residues $n = 33 - 38$ can be observed for the interval $20 < t < 75$ ns and for $t > 90$ ns. The second α -helix is well conserved during the entire MD trajectory; (iv) in chain D, a small portion of the first α -helix loses structure between $10 - 30$ ns. Moreover, a small α -helix is found between the two long α -helices in chain D at 300K and another small α -helix is also found in its flexible C-terminal part ($n = 122 - 126$). Small β -strand motifs appear in the C-terminal region of chain A ($n = 101, 107, 129$ and 132) and in the loop joining α -helices 1 and 2 ($n = 43$ and 48). Rare occurrences of β -strand motifs can be found in chains B ($n = 39, 41, 44, 105, 109, 124$, and 130) and C ($n = 45, 115$ and 118).

As expected, the secondary structures of the tetramer at 500K are less stable than at 300K due to the increase of the fluctuations (Fig. 9B). The loop connecting the first and second α -helices and some portions of the α -helices in the vicinity of the loop are observed to be more flexible than other portions of the system and consequently their secondary structures are lost. Moreover, partial unfolding events of both α -helices appear more frequently at this temperature. The behavior of the C-terminal region of each chain also changes at 500K. Indeed, the small α -helices of chains B ($n = 112 - 116$) and D ($n = 122 - 126$) are completely unfolded either from the beginning of the trajectory (chain B) or from $t > 35$ ns (chain D); 3/10-helix structures are formed in chains A, C and D ($n = 120 - 140$). Moreover,

extended β -strand conformation is found for residues $n = 94 - 97$ and $101 - 104$ of chain C during almost the entire MD run, and for different residues of the rest of the chains for short time intervals (Fig. 9B). In spite of all these “fluctuations”, the tetramer, in general, retains ~70% of its secondary structures in the MD run at high temperature.

In order to shed light on the reason for keeping the tetramer very stable at 300K and partially stable at 500K, we examined the behavior of the distances between the C α s of the nonpolar residues of each chain over the entire MD trajectories at both temperatures. In order to illustrate “the dynamics” of hydrophobic interactions between the chains over time, we built 2-D maps (Fig. 10) illustrating the distances between the C α s of nonpolar residues for the sets of representative structures of minima of FELs along PC1 and PC2 (Fig. 8).

As expected, nonpolar residues pertaining to the second α -helix of each chain are the main players in the hydrophobic interactions at 300K (Fig. 10A). In particular, (i) nonpolar residues pertaining to the second α -helix of chain A interact with nonpolar residues pertaining to the second α -helices of the rest of the chains [Ala66(A)/Ala66(B), Val70(A)/Ala69(B), Val77(A)/Ala76(B), Val63(A)/Ala69(C), Phe94(A)/Ala85(C), Ala89(A)/Ile88(D), Ala89(A)/Ala85(D) and Ala90(A)/Ile88(D)]; (ii) nonpolar residues of the second α -helix of chain B make strong hydrophobic contacts with the nonpolar residues pertaining to the second α -helix of chain C [Ala29(B)/Val55(C), Val52(B)/Val49(C) and Val52(B)/Val52(C)]. Moreover, nonpolar residues pertaining to the end of the first α -helix and the loop between the first and second α -helices (Val37 and Val40) of chain C interact with nonpolar residues pertaining to the beginning of the second α -helix (Val52, Ala53, Ala56 and Val74) and the loop between the first and second α -helices (Val48, Val49 and Ala76) of chain D. The stability of the tetramer is also ensured by a few hydrophobic contacts between the N-terminal region of chains A/C and the ends of the second α -helix of chains D/B, respectively [Met1(A)/Ala89(D), Leu8(C)/Ile88(B) and Leu8(C)/Ala89(B)]. Also, there are no contacts at all between the nonpolar residues of chains B and D during the entire MD simulation. Most of these hydrophobic contacts, especially between the nonpolar residues pertaining to the second α -helices of the chains, remain stable during the entire MD trajectory.

A number of hydrophobic contacts between the chains increases with increase of temperature, especially between the nonpolar residues pertaining to the first and second α -helices; however, some rearrangements between interacting nonpolar residues are observed (Fig. 10B). In particular, more hydrophobic contacts between the nonpolar residues pertaining to the second α -helices of chains A and B [Ala69(A)/Ala56(B), Ala78(A)/Ala78(B), Val70(A)/Val71(B), Ala78(A)/Val82(B), Ala56(A)/Val66(B), Ala52(A)/Ala52(B) and Val55(A)/Val55(B)], A and D [Ala85(A)/Val63(D), Ala85(A)/Val66(D), Ala78(A)/Val66(D), Ala90(A)/Val49(D), Val63(A)/Val71(D), Val70(A)/Val63(D) and Val71(D)/Val70(D)], B and D [Ala85(B)/Ala85(D), Val71(B)/Ala69(D), Val70(B)/Val66(D), Ala78(B)/Val77(D), Val74(B)/Ala69(D) and Val70(B)/Ala69(D)], C and D [Ala78(C)/Ala91(D), Val82(C)/Ile88(D), Ala78(C)/Ile88(D), Val77(C)/Ala90(D), Val82(C)/Ala91(D) and Ala91(C)/Ile88(D)] are observed at 500K than at 300K. Strong interactions between chains A and C, and B and C, detected at 300K, decrease at 500K. As at 300K, the hydrophobic contacts between the nonpolar residues pertaining to the second α -helices of

the chains are more stable than those between the nonpolar residues pertaining to the first and second α -helices.

The hydrophobic contacts illustrated in Fig. 10 indicate that the nonpolar residues pertaining to the second α -helices of the chains play a crucial role in the stability of the tetrameric α S. This finding is in agreement with an earlier experiment.²⁰

Finally, because of a special role of KTKEGV repeat motifs, and specifically, the positively charged lysine residues, in mediating α S physiological tetramerization,²⁷ and in regulating α S oligomer formation and α S aggregation,³⁸ we have examined interactions between the lysine residues and all amino acids for the sets of representative structures of minima of FELs along PC1 and PC2 at 300K and 500K (Fig. 8).

It appears that, in all the structures (at both temperatures), the lysine residues interact with negatively charged residues (Asp or Glu) forming salt bridges, and thus participate in stabilization of the tetramer conformation (Fig. S9).

At 300K, because the initial tetrameric conformation is well-maintained during the entire duration of the MD trajectory (Fig. 8A), the interactions between lysine residues and negatively charged residues (Asp and Glu) mainly occur between the α -helices of the four chains (Fig. S9A). Indeed, chain A interacts with chains B and D through residues: Lys34(A)/Glu57(B), Asp121(A)/Lys97(B), Lys97(A)/Glu105(D) but no lysine interactions are found between chains A and C. However, chain C interacts with chains B and D via residues: Lys21(B)/Asp135(C), Lys80(C)/Glu83(B), Lys32(C)/Glu28(D), Lys60(C)/Glu57(D) and Lys21(C)/Glu13(D). Intra-chain lysine interactions also occur inside the α -helices, and between the α -helices and the flexible regions of the chains, as shown by the diagonal parts of the contact maps (Fig. S9A).

The results extracted from the structures at 500 K (Fig. 8B) revealed supplemental interactions between the lysine residues and the rest of the amino acids (Fig. S9B) in comparison to the results found at 300K (Fig. S9A). All four chains have undergone more conformational changes at 500K, which have generated new interactions; in particular, between the residues of chains A and C: Glu20(A)/Lys58(C), Lys23(A)/Glu57(C) and Glu139(A)/Lys97(C); and which have eliminated the interactions between chains B and C (Fig. S9B). Nevertheless, there are still interactions between chains A, B and D: Lys23(B)/Glu83(A), Lys23(A)/Glu46(D), Asp98(A)/Lys12(B) and Glu105(A)/Lys97(D). Also, a large number of interactions was observed between chains C and D: Lys34(C)/Glu35(D), Lys43(C)/Glu35(D), Lys43(C)/Glu61(D), Glu57(C)/Lys60(D), Lys58(C)/Glu57(D) and Lys80(C)/Glu83(D); and the chains B and D: Lys12(D)/Glu46(B), Lys58(D)/Asp97(B) and Lys97(D)/Glu137(B).

Thus, it was found that two types of interactions, hydrophobic and salt bridge, keep tetrameric α S stable. In particular, (i) at lower temperature (300K) both hydrophobic interactions between nonpolar residues and salt bridges, formed by positively charged lysine residues and negatively charged aspartate and glutamate residues, play a crucial role in the stability of tetramer; (ii) at higher temperature (500K), in spite of the increase of a number of hydrophobic contacts, salt bridges become main contributor to the stability of tetramer,

because the strength of hydrophobic interactions diminishes at high temperatures,⁴¹ whereas salt bridges are extremely resilient to temperature increases.⁴² These findings indicate that hydrophobic interactions and salt bridges along with KTKEGV repeat motifs govern tetramer formation.

3. Discussion and Conclusions

By analyzing the all-atom MD trajectories of monomeric α S at 310K, and tetrameric α S at 300K and 500K in terms of the local and global motions, correlations between the main-chain and the side-chain motions, and a recently introduced steric parameter, we have investigated: (i) the dynamics of α S as intrinsically disordered protein (monomer) and as metastable tetramer; (ii) a correlation between tetramers and monomers; (iii) what mediates tetramer formation and what makes tetramer stable.

A highly-disordered α S, a common form of monomeric α S in solution, was a main object of the scrutiny; however, a partially-disordered and micelle-bound (stable) monomeric α S were also studied to reveal how the physics of the interactions between the MC and SCs changes when a protein is “transitioning” from a completely disordered state to a partially-disordered state and to a folded state, although the latter has never been observed in solution experimentally without bound to a micelle. For tetrameric α S, we used the structure determined in Ref. 20.

As was expected, strong correlations between the motions of the SCs and the MC were observed along the entire sequence for a highly disordered monomeric α S. The R value varies in an oscillatory manner with small amplitude (Fig. 2A), which indicates a high-flexibility of MC and SCs along the entire sequence. The minima on the correlation coefficient curve are mainly associated with lysine and threonine (KTK, KT), and lysine and alanine (KAK) combinations, or with just lysine residues (KK), most of which are parts of KTKEGV repeats – key players in mediating α S physiological tetramerization²⁷ and the binding of α S to membrane phospholipids, and in regulating α S oligomer formation and aggregation.³⁸ Specifically, the positively charged lysine residues play a critical role in the latter three phenomena.³⁸ Interestingly, these minima remain on the correlation coefficient curves for partially-disordered (Fig. 2B) and micelle-bound (Fig. 2C) monomeric α S, which indicate that these sites are playing a crucial role in the stabilization of the monomeric α S. It was shown that such a behavior of lysine is associated with its large size and amphipathic nature.

The maxima on the correlation coefficient curve are mainly associated with glycine, the smallest residue, which also remain on the correlation coefficient curves for partially-disordered and micelle-bound monomeric α S (Fig. 2). These results indicate that glycines are responsible for high-flexibility and disorder of monomeric α S. It should be noted that monomeric α S contains eighteen glycine and nineteen alanine (most of which have very high values of R , not highlighted in the Figures) residues, which is ~27% of the entire sequence, whereas Trp-cage and VA3, regular folded proteins, contain only ~15% of glycine and alanine residues.

Another interesting finding is that lowest values of the steric parameter, s_n , are associated with alanine residues in highly disordered monomeric α S (Fig. 4A). The results have shown that most of the alanine residues are fully exposed to the solvent during the simulations (Fig. S6) and the chain of α S is well kinked at the positions of the alanine (and the glycine) residues. Similar behavior of s_n at alanine residues persists for partially disordered monomeric α S (Fig. 4B) and for the stable micelle-bound monomeric α S (Fig. 4C), although there are some changes in the latter related to formation of secondary structures. Interestingly, new minima on the s_n curve for micelle-bound monomeric α S are formed in the flexible regions of a protein (loop, C-terminal end, and broken part of second α -helix). These results indicate that this feature of s_n is an intrinsic property for the monomeric α S, and alanine residues are playing an important role in instability of monomeric α S.

It should be noted that 11-residue repeats of N-terminal domain of α S are similar to ones found in apolipoproteins and are consistent with a class A₂ amphipathic α -helices.⁴³ It has been shown that the N-terminal domain of a monomeric lipid-free 243-residue apolipoprotein A-I (apoA-I), the main protein component of high density lipoproteins, which has only six repeats, is predominantly α -helical; and self-association is an inherent property of the lipid-free apoA-I.^{44,45} Therefore, the size of α S may play a role in its inability to fold without lipid association. Also, it has been illustrated that the N-terminus of α S serves as a nucleation center for membrane binding of the full-length protein and helix folding up to approximately residue 100.⁴⁶

As in monomeric α S, the correlation coefficient of tetrameric α S at 300K “oscillates” along the amino-acid sequence, and important sites (KTK, KT, KAK, KK), revealed in highly-disordered state, are remained in the correlation coefficient curves of tetrameric α S (Fig. 5, left panels); however, because of secondary structures they are not as prominent as in a highly disordered monomeric α S. Also, as in monomeric α S, many maxima in the membrane-binding and nonamyloid component domains of tetrameric α S are formed by the smallest residue glycine. Correlations between the motions of the SCs and the MC in tetrameric α S become stronger with increase of temperature (Fig. 5, right panels), which is understandable because, with the increase of temperature, the system become less stable and disordered; although they are still weaker (at both 300K and 500K) than ones in a highly disordered monomeric α S at 310K indicating stability of tetrameric α S. Moreover, the average value of R of the secondary structures in tetrameric α S at 300K is lower by ~ 0.1 than one in micelle-bound monomeric α S at 300K, which indicates lower flexibility of MC and SCs and larger stability of secondary structures in tetrameric α S compared to micelle-bound monomeric α S. The reason for the stability of secondary and tertiary structures in the tetrameric α S is a hydrophobic core formed by nonpolar residues pertaining to the second α -helix of each chain (Fig. 10), along with salt bridges formed by positively charged lysine residues and negatively charged aspartate and glutamate residues (Fig. S9).

Moreover, as in monomeric α S, the low values of s_n at alanine residues are still dominant at both temperatures in tetrameric α S (Fig. 7), which indicates on intrinsic nature of this property.

Finally, the results obtained in this study enabled us to elucidate a fundamental relationship between monomers and tetramers; the key residues involved in mediating tetramer formation; the reasons for the stability of tetrameric α S, and inability of monomeric α S to fold.

4. Materials and Methods

4.1. MD Simulations

Fifty all-atom MD trajectories (10 ns of each) starting from completely disordered conformations of the monomeric α S generated by flexible-meccano software,²⁹ for which the side chains were built by using the PDB Reader⁴⁷ in CHARMM GUI,^{48,49} and six all-atom MD trajectories (10 ns of each), with initial structures extracted from 60 ns unfolding MD trajectory (each structure was extracted after every 10 ns time interval) of the micelle-bound monomeric α S¹² at 500K, were performed at 310K by using the GROMACS package⁵⁰ and the CHARMM27 force field.^{51,52} In order to have robust results, we performed fifty-six all-atom MD trajectories additionally, with the same initial structures at the same 310K by using the AMBER99SB force field.⁵³ Initial results obtained from both sets of trajectories were very similar to each other; therefore, we selected the MD trajectories generated with the CHARMM27 force field for scrutiny. Moreover, one 100 ns all-atom MD trajectory starting from the micellebound monomeric α S¹² was carried out at 300K.

Two all-atom MD trajectories of tetrameric²⁰ α S (a structure generously provided by Dr. Thomas Pochapsky) were performed at 300K for 100 ns and at 500 K for 60 ns by using the same GROMACS package⁵⁰ and the CHARMM27 force field.^{51,52} Since both the micelle-bound monomeric α S¹² and tetrameric α S²⁰ experiments were performed at room temperature, we performed MD simulations for both systems at 300K.

The proteins were solvated in a cubic box with explicit water molecules (between 21291 and 221275 for completely disordered conformations of the monomeric α S, 214515 for the micelle-bound conformation of the monomeric α S, and 237790 for the tetrameric α S) by using the TIP3P force field⁵⁴ and with a salt (NaCl) concentration of 0.1 mol/l. The coordinates were saved every 1 ps. Periodic boundary conditions were applied. A distance of 1 nm was assigned between the protein and the sides of the unit supercell in order to avoid any interaction between the proteins of the neighboring supercells. The temperature of the MD simulations was kept at 310K and 300K for the monomer, and at 300K and 500K for the tetramer with a v-rescale thermostat,⁵⁵ and the pressure (Parrinello-Rahman barostat)⁵⁶ was kept at 1 bar for both systems. The different steps of energy minimization were performed by using the steepest descent algorithm and the conjugate gradient algorithm as implemented in the GROMACS package⁵⁰ with a tolerance of 100 kJ/mol/nm and a maximum step of 0.01 nm for the displacement. The particle-mesh-Ewald method^{57,58} was used for calculating long-range electrostatic interactions, and a distance of 1.1 nm was used for the van der Waals cutoff. After the desired temperature was reached, an equilibration of 0.6 ns duration was performed with random initial conditions generated by using a random seed for the initial velocities. It should be noted that because of the periodic boundary conditions used in MD simulations, water evaporation cannot occur. The use of an artificial

high temperature of 500 K with the periodic boundary conditions allows to simulate protein unfolding in a short computational time. Indeed, unfolding of large proteins, as the tetrameric α S, with all-atom MD simulations at the actual unfolding temperature (about 320–340 K usually) are out of reach with the present computers. We equilibrated the system under ambient conditions in an NPT ensemble to reach the proper density of water with periodic boundary conditions and subsequently performed MD simulations at 500 K in an NVT ensemble, in which the density is a constant.

4.2. Coarse-grained dihedral angles

To express the correlation between the main-chain and the side-chain motions quantitatively, two coarse-grained dihedral angles (CGDAs) were defined for each residue n (Fig. S1). The angles characterizing the fluctuations of the main chain are the CGDAs γ_n formed by the virtual bonds joining four consecutive C^α atoms ($n - 1$, n , $n + 1$ and $n + 2$) along the amino-acid sequence with $2 \leq n \leq N - 2$ and N being the number of residues. The angles γ_n are coordinates used to describe large changes of protein conformations and are part of coarse-grained models of proteins. The angles characterizing the fluctuations of the side chains are defined by the CGDAs δ_n ($1 \leq n \leq N - 1$) formed by the virtual bond joining two consecutive C^α atoms (n , $n + 1$) and the bonds joining these C^α atoms to their respective C^β atoms (Fig. S1).³⁶ For the residue glycine, the side-chain H atom was defined as a pseudo- C^β atom.

4.3. Steric parameter

To quantify the steric constraints imposed on the side chain of each n^{th} residue by the rest of the protein, a steric parameter s_n which involves the distance between all the atoms of the protein and all the atoms of the SC of the n^{th} residue, was defined³⁴ as

$$s_n = \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M \frac{1}{\|\mathbf{r}_{ij}\|} \quad (1)$$

where N corresponds to the number of atoms of the protein (excluding the atoms of the n^{th} residue considered), and M corresponds to the number of atoms of the SC of the n^{th} residue located after the C^β atom. The quantity $\|\mathbf{r}_{ij}\|$ is the distance between atom i of the protein and atom j of the SC of the n^{th} residue. The parameter s_n cannot be computed for Gly because it does not have a C^β atom. The average values of s_n computed over the entire duration of the trajectories are calculated by averaging the mean value of s_n in the different windows of 100 ps.

4.4. Pearson correlation coefficient

The Pearson correlation coefficient R computed between two functions $x_i(t)$ and $y_i(t)$ is given by^{34,36,37}

$$R = \frac{\sum_{t=t_0}^{t_{\max}} [\Delta x_i(t) \Delta y_i(t)]}{\sqrt{\sum_{t'=t_0}^{t_{\max}} [\Delta x_i(t')]^2 \sum_{t''=t_0}^{t_{\max}} [\Delta y_i(t'')]^2}} \quad (2)$$

where $x_i(t) = x_i(t) - \langle x_i \rangle$, $y_i(t) = y_i(t) - \langle y_i \rangle$ and $t_{\max} - t_0$ is the time interval considered. For the calculation of R between the trajectories of the CGDAs,^{34,36} we cannot use $x_i(t) = \gamma_n(t)$ and $y_i(t) = \delta_n(t)$; i.e., the CGDAs computed for each MD snapshot are defined in $[-\pi, \pi]$, and the average cannot be computed correctly (for example the simple average angle between $-\pi$ and $+\pi$ would give an angle of zero degrees). Therefore, to define the time averages $\langle \dots \rangle$ of the dihedral angles in Equation (2), we built two discrete time series S_γ and S_δ from the successive dihedral angle displacements (between two snapshots) γ and δ respectively. For example, $S_\gamma(t=0) = \gamma(t=0)$, $S_\gamma(t=1) = S_\gamma(t=0) + \gamma(t=1)$, ... $S_\gamma(t=m) = S_\gamma(t=m-1) + \gamma(t=m)$... (see Ref. 36). Use of the variables $S_\gamma(t)$ and $S_\delta(t)$ ensures that a jump from $-\pi$ to π is of length zero. The average values of R computed over the entire duration of the trajectories are calculated by averaging the mean value of R in the different windows of 100 ps.

4.5. Dihedral Principal Component Analysis (dPCA)

A detailed description of the dPCA method is available in our previous papers^{36,59} and in earlier references;^{60,61} therefore, only a brief outline of this approach is presented here. A dPCA, a covariance matrix-based mathematical technique, is an effective method for extracting important motions from MD simulations. In dPCA the internal (dihedral) coordinate space is rotated to a new space with new coordinates (PCs), a few of which are sufficient to describe a large part of the fluctuations of a protein. Here, structural fluctuations of the CGDAs (MSFs) can be decomposed into collective modes by dPCA. The modes have “frequencies” and directions corresponding to the eigenvalues and eigenvectors of the covariance matrix. The mode k with the largest eigenvalue (λ_k) corresponds to the mode that contributes the most to the structural fluctuations of the protein. The contribution of each angle (γ_i and δ_i) to mode k is called the influence, v_i^k .^{36,61}

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Gregory Petsko and Thomas Pochapsky for supplying the structure of tetrameric α S and for helpful discussions. This research was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB10-19767) and the PARI NANO2BIO (Région Bourgogne Franche-Comté), and was conducted by using the HPC resources from DSI-CCUB (Université de Bourgogne).

References

1. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. α -Synuclein in Lewy bodies. *Nature*. 1997; 388:839–840. [PubMed: 9278044]

2. Bertram L, Tanzi RE. The genetic epidemiology of neurodegenerative disease. *J Clin Invest*. 2005; 115:1449–1457. [PubMed: 15931380]
3. Chandra S, Gallardo G, Fernández-Chacón R, Schlüter OM, Südhof TC. α -synuclein cooperates with CSP α in preventing neurodegeneration. *Cell*. 2005; 123:383–396. [PubMed: 16269331]
4. Fusco G, Pape T, Stephens AD, Mahou P, Costa AR, Kaminski CF, Kaminski Schierle GS, Vendruscolo M, Veglia G, Dobson CM, De Simone A. Structural basis of synaptic vesicle assembly promoted by α -synuclein. *Nat Commun*. 2016; 7:12563. [PubMed: 27640673]
5. Kamp F, Exner N, Lutz AK, Wender N, Heggermann J, Brunner B, Nuscher B, Bartels T, Giese A, Beyer K, Eimer S, Winklhofer KF, Haass C. Inhibition of mitochondrial fusion by α -synuclein is rescued by PINK1, Parkin and DJ-1. *EMBO J*. 2010; 29:3571–3589. [PubMed: 20842103]
6. Clayton DF, George JM. The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. *Trends Neurosci*. 1998; 21:249–254. [PubMed: 9641537]
7. Uéda K, Fukushima H, Masliah E, Xia Y, Iwai A, Yoshimoto M, Otero DA, Kondo J, Ihara Y, Saitoh T. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci USA*. 1993; 90:11282–11286. [PubMed: 8248242]
8. Eliezer D. The mysterious C-terminal tail of alpha-synuclein: nanobody's guess. *J Mol Biol*. 2013; 425:2393–2396. [PubMed: 23541591]
9. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999; 293:321–31. [PubMed: 10550212]
10. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model*. 2001; 19:26–59. [PubMed: 11381529]
11. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT. NACP, A protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*. 1996; 35:13709–13715. [PubMed: 8901511]
12. Ulmer TS, Bax A, Cole NB, Nussbaum RL. Structure and dynamics of micelle-bound human α -synuclein. *J Biol Chem*. 2005; 280:9595–9603. [PubMed: 15615727]
13. Fauvet B, Mbefo MK, Fares MB, Desobry C, Michael S, Ardah MT, Tsika E, Coune P, Prudent M, Lion N, Eliezer D, Moore DJ, Schneider B, Aebischer P, El-Agnaf OM, Masliah E, Lashuel HA. α -Synuclein in central nervous system and from erythrocytes, mammalian cells, and *Escherichia coli* exists predominantly as disordered monomer. *J Biol Chem*. 2012; 287:15345–15364. [PubMed: 22315227]
14. Binolfi A, Theillet FX, Selenko P. Bacterial in-cell NMR of human α -synuclein: a disordered monomer by nature? *Biochem Soc Trans*. 2012; 40:950–954. [PubMed: 22988846]
15. Burré J, Vivona S, Diao J, Sharma M, Brunker AT, Südhof TC. Properties of native brain α -synuclein. *Nature*. 2013; 498:E4–E6. [PubMed: 23765500]
16. Theillet FX, Binolfi A, Bekei B, Martorana A, Rose HM, Stuijver M, Verzini S, Lorenz D, van Rossum M, Goldfarb D, Selenko P. Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature*. 2016; 530:45–50. [PubMed: 26808899]
17. Welker S, Rudolph B, Frenzel E, Hagn F, Liebisch G, Schmitz G, Scheuring J, Kerth A, Blume A, Weinkauff S, Haslbeck M, Kessler H, Buchner J. Hsp12 is an intrinsically unstructured stress protein that folds upon membrane association and modulates membrane function. *Mol Cell*. 2010; 39:507–520. [PubMed: 20797624]
18. Cornell RB, Taneva SG. Amphipathic helices as mediators of the membrane interaction of amphitropic proteins, and as modulators of bilayer physical properties. *Current Protein and Peptide Science*. 2006; 7:539–552. [PubMed: 17168787]
19. Bartels T, Choi JG, Selkoe DJ. α -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature*. 2011; 477:107–110. [PubMed: 21841800]
20. Wang W, Perovic I, Chittuluru J, Kaganovich A, Nguyen LT, Liao J, Auclair JR, Johnson D, Landeru A, Simorellis AK, Ju S, Cookson MR, Asturias FJ, Agar JN, Webb BN, Kang C, Ringe D, Petsko GA, Pochapsky TC, Hoang QQ. A soluble α -synuclein construct forms a dynamic tetramer. *Proc Natl Acad Sci USA*. 2011; 108:17797–17802. [PubMed: 22006323]

21. Trexler AJ, Rhoades E. N-terminal acetylation is critical for forming α -helical oligomer of α -synuclein. *Prot Sci*. 2012; 21:601–605.
22. Westphal CH, Chandra SS. Monomeric synucleins generate membrane curvature. *J Biol Chem*. 2013; 288:1829–1840. [PubMed: 23184946]
23. Gould N, Mor DE, Lightfoot R, Malkus K, Giasson B, Ischiropoulos H. Evidence of native α -synuclein conformers in the human brain. *J Biol Chem*. 2014; 289:7929–7934. [PubMed: 24474688]
24. Selkoe D, Dettmer U, Luth E, Kim N, Newman A, Bartels T. Defining the native state of α -synuclein. *Neurodegener Dis*. 2014; 13:114–117. [PubMed: 24192542]
25. Kim S, Yun SP, Lee S, Umanah GE, Bandaru VVR, Yin X, Rhee P, Karuppagounder SS, Kwon SH, Leea H, Mao X, Kim D, Pandey A, Lee G, Dawson VL, Dawson TM, Ko HS. GBA1 deficiency negatively affects physiological α -synuclein tetramers and related multimers. *Proc Natl Acad Sci USA*. 2018; 115:798–803. [PubMed: 29311330]
26. Gurry T, Ullman O, Fisher CK, Perovic I, Pochapsky T, Stultz CM. The dynamic structure of α -synuclein multimers. *J Am Chem Soc*. 2013; 135:3865–3872. [PubMed: 23398399]
27. Dettmer U, Newman AJ, von Saucken VE, Bartels T, Selkoe D. KTKGV repeat motifs are key mediators of normal α -synuclein tetramerization: their mutation causes excess monomers and neurotoxicity. *Proc Natl Acad Sci USA*. 2015; 112:9596–9601. [PubMed: 26153422]
28. Nath A, Sammakorpi M, DeWitt DC, Trexler AJ, Elbaum-Garfinkle S, O'Hern CS, Rhoades E. The conformational ensembles of α -synuclein and tau: combining single-molecule FRET and simulations. *Biophys J*. 2012; 103:1940–1949. [PubMed: 23199922]
29. Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*. 2012; 28:1463–1470. [PubMed: 22613562]
30. Matheson RR, Scheraga HA. A method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules*. 1978; 11:819–829.
31. Fersht AR. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci USA*. 1995; 92:10869–10873. [PubMed: 7479900]
32. Daggett V, Fersht AR. The Present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol*. 2003; 4:497–502. [PubMed: 12778129]
33. Maisuradze GG, Liwo A, Senet P, Scheraga HA. Local vs global motions in protein folding. *J Chem Theory Comput*. 2013; 9:2907–2921. [PubMed: 23914144]
34. Cote Y, Maisuradze GG, Delarue P, Scheraga HA, Senet P. New insights into protein (un)folding dynamics. *J Phys Chem Lett*. 2015; 6:1082–1086. [PubMed: 25866611]
35. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Molec Graphics*. 1996; 14:33–38.
36. Cote Y, Senet P, Delarue P, Maisuradze GG, Scheraga HA. Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state. *Proc Natl Acad Sci USA*. 2012; 109:10346–10351. [PubMed: 22689963]
37. Pearson K. Mathematical contributions to the theory of evolution. III Regression, heredity and panmixia. *Phil Trans R Soc Lond A*. 1896; 187:253–318.
38. Zarbiv Y, Simhi-Haham D, Israeli E, Elhadi SA, Grigoletto J, Sharon R. Lysine residues at the first and second KTKGV repeats mediate α -Synuclein binding to membrane phospholipids. *Neurobiol Dis*. 2014; 70:90–98. [PubMed: 24905915]
39. Senet P, Maisuradze GG, Foulie C, Delarue P, Scheraga HA. How main-chains of proteins explore the free-energy landscape in native states. *Proc Natl Acad Sci USA*. 2008; 105:19708–19713. [PubMed: 19073932]
40. Cote Y, Senet P, Delarue P, Maisuradze GG, Scheraga HA. Nonexponential decay of internal rotational correlation functions of native proteins and self-similar structural fluctuations. *Proc Natl Acad Sci USA*. 2010; 107:19844–19849. [PubMed: 21045133]
41. Némethy G, Scheraga HA. The structure of water and hydrophobic bonding in proteins. III The thermodynamic properties of hydrophobic bonds in proteins. *J Phys Chem*. 1962; 66:1773–1789.

42. Thomas AS, Elcock AH. Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J Am Chem Soc.* 2004; 126:2208–2214. [PubMed: 14971956]
43. Dikiy I, Eliezer D. Folding and misfolding of alpha-synuclein on membranes. *Biochim Biophys Acta.* 2012; 1818:1013–1018. [PubMed: 21945884]
44. Jayaraman S, Abe-Dohmae S, Yokoyama S, Cavigliolo G. Impact of self-association on function of Apolipoprotein A-I. *J Biol Chem.* 2011; 286:35610–35623. [PubMed: 21835924]
45. Smith LE, Segrest JP, Davidson WS. Helical domains that mediate lipid solubilization and ABCA1-specific cholesterol efflux in apolipoproteins C-I and A-II. *J Lipid Res.* 2013; 54:1939–1948. [PubMed: 23620136]
46. Bartels T, Ahlstrom LS, Leftin A, Kamp F, Haass C, Brown MF, Beyer K. The N-terminus of the intrinsically disordered protein α -synuclein triggers membrane binding and helix folding. *Biophys J.* 2010; 99:2116–2124. [PubMed: 20923645]
47. Jo S, Cheng X, Islam SM, Huang L, Rui H, Zhu A, Lee HS, Qi Y, Han W, Vanommeslaeghe K, MacKerell AD Jr, Roux B, Im W. CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing non-standard residues. *Adv Protein Chem Struct Biol.* 2014; 96:235–265. [PubMed: 25443960]
48. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem.* 2008; 29:1859–1865. [PubMed: 18351591]
49. Brooks BR, Brooks CL 3rd, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. *J Comput Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
50. Hess B, Kutzner C, van der Spoel D, Lindahl EJ. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput.* 2008; 4:435–447. [PubMed: 26620784]
51. MacKerell AD Jr, Feig M, Brooks CL 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
52. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J Chem Theory Comput.* 2010; 6:459–466. [PubMed: 26617301]
53. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 2006; 65:712–725. [PubMed: 16981200]
54. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983; 79:926–935.
55. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007; 126:014101. [PubMed: 17212484]
56. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys.* 1981; 52:7182–7190.
57. Darden T, York D, Pedersen LG. Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J Chem Phys.* 1993; 98:10089–10092.
58. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys.* 1995; 103:8577–8593.
59. Maisuradze GG, Leitner DM. Free energy landscape of a biomolecule in dihedral principal component space: sampling convergence and correspondence between structures and minima. *Proteins.* 2007; 67:569–578. [PubMed: 17348026]
60. Mu Y, Nguyen PH, Stock G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins.* 2005; 58:45–52. [PubMed: 15521057]
61. Altis A, Nguyen PH, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys.* 2007; 126:244111. [PubMed: 17614541]

Figure 1.

Average values of RMSDs with standard deviations computed for fifty MD trajectories for highly disordered monomeric α S as functions of time at 310K (A). Secondary structure evolution as a function of time, computed by VMD program,³⁵ for fifty MD trajectories of highly disordered monomeric α S at 310K (B). The following color code was used in panel B: the turns are in cyan, the coils in green, the extended β -strands in yellow, the β -sheets in orange, the α -helices in red and the 3/10 helices in magenta.

Figure 2.

Evolution of the correlation coefficient, R , between the time series of the dihedral angles γ and δ computed from and averaged over the full duration of MD simulations along the amino-acid sequence of highly disordered monomeric α S (A) at 310K, partially disordered monomeric α S (B) at 310K, and micelle-bound monomeric α S (C) at 300K. The grey stripes (in panel C) indicate the position of the α -helices as they are defined in the experimental structure (PDB ID: 1XQ8¹²). KTK, KT, KAK, and KK combinations are highlighted in red, and locations of glycine residues are highlighted in green.

Figure 3.

Comparison of the free-energy profiles of the dihedral angles γ (black curves) and δ (red curves) along the amino-acid sequence of α S in the highly disordered monomeric form calculated from fifty 10 ns MD runs at 310K.

Figure 4.

Evolution of the steric parameter, s_n , computed from, and averaged over, the full duration of the MD simulations along the amino-acid sequence of the highly disordered monomeric α S (A) at 310K, partially disordered monomeric α S (B) at 310K, and micelle-bound monomeric α S (C) at 300K. The grey stripes (in panel C) indicate the position of the α -helices as they are defined in the experimental structure (PDB ID: 1XQ8¹²). Locations of alanine residues are highlighted in red.

Figure 5.

Evolution of the correlation coefficient, R , between the time series of the dihedral angles γ and δ computed from, and averaged over, the full duration of the MD simulations at 300K (panels A–D on the left side) and 500K (panels A–D on the right side) along the amino-acid sequence of chain A (panels A), B (panels B), C (panels C) and D (panels D) of the tetrameric α S. The grey stripes indicate the position of the α -helices defined in the initial PDB structure²⁰ of the tetramer. KTK, KT, KAK, and KK combinations are highlighted in red, and locations of glycine residues are highlighted in green.

Figure 6.

Comparison of the free-energy profiles of the dihedral angles γ of the chain A (black curve), the chain B (red curve), the chain C (green curve) and the chain D (blue curve) of tetrameric form of α S computed from the 100 ns MD run at 300K (A) and 60 ns MD run at 500K (B). The red boxes indicate the positions of the α -helices in the chains of the tetramer.

Figure 7.

Evolution of the steric parameter, s_n , along the amino-acid sequence of chain A (panels A), B (panels B), C (panels C) and D (panels D) of tetrameric α S computed from the MD run of the tetramer carried out at 300K (panels A–D on the left side) and 500K (panels A–D on the right side). The grey stripes indicate the position of the α -helices defined from the initial PDB structure²⁰ of the tetramer. Locations of alanine residues are highlighted in red.

Figure 8A.

Figure 8B.

Figure 8.

The free-energy landscapes (in $k_B T$) along the first two PCs with representative structures in the minima for tetrameric αS at 300K (A) and 500K (B). Chains A, B, C, and D in the

representative structures are highlighted in red, green, blue, and magenta, respectively. The nonpolar residues in the representative structures are highlighted in yellow.

Figure 9. Secondary structure evolution as a function of time, computed by the VMD program,³⁵ for tetrameric α S at 300K (A) and 500K (B). The color code in both panels is the same as in Fig. 1B.

Figure 10A.

Figure 10B.

Figure 10.

2-D maps for the distances between the C α s of nonpolar residues of chains A, B, C, and D of tetrameric α S for the sets of representative structures of minima of the FELs along PC1 and PC2 (Fig. 8) at 300K (A) and 500K (B). The numbers in the panels correspond to numbers of minima in FELs.