
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Nath, Abhinav; Sammalkorpi, Maria; DeWitt, David C.; Trexler, Adam J.; Elbaum-Garfinkle, Shana; O'Hern, Corey S.; Rhoades, Elizabeth

The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations

Published in:
Biophysical Journal

DOI:
[10.1016/j.bpj.2012.09.032](https://doi.org/10.1016/j.bpj.2012.09.032)

Published: 01/01/2012

Document Version
Early version, also known as pre-print

Please cite the original version:
Nath, A., Sammalkorpi, M., DeWitt, D. C., Trexler, A. J., Elbaum-Garfinkle, S., O'Hern, C. S., & Rhoades, E. (2012). The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophysical Journal*, 103(9), 1940-1947. <https://doi.org/10.1016/j.bpj.2012.09.032>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations

Abhinav Nath^{1*}, Maria Sammalkorpi^{2,3}, David C. DeWitt¹, Adam J. Trexler¹, Shana Elbaum-Garfinkle¹, Corey S. O'Hern^{4,5*}, and Elizabeth Rhoades^{1,5*}

¹ Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, P. O. Box 208114, New Haven, CT 06520-8114

² Department of Chemical & Environmental Engineering, Yale University, P. O. Box 208286, New Haven, CT 06520-8286

³ Department of Chemistry, Aalto University, P.O. Box 16100, FI-00076 Aalto, Finland

⁴ Department of Mechanical Engineering & Materials Science, Yale University, P. O. Box 208286, New Haven, CT 06520-8286

⁵ Department of Physics, Yale University, P. O. Box 208120, New Haven, CT 06520-8120

Running Title: Conformational ensembles of α S and tau

Abstract

Intrinsically disordered proteins (IDPs) are increasingly recognized for their important roles in a range of biological contexts, both in normal physiological function and in a variety of devastating human diseases. However, their structural characterization by traditional biophysical methods, for the purposes of understanding their function and dysfunction, has proved challenging. Here, we investigate the model IDPs α -Synuclein (α S) and tau, that are involved in major neurodegenerative conditions including Parkinson's and Alzheimer's diseases, using excluded-volume Monte Carlo simulations constrained by pairwise distance distributions from single-molecule fluorescence measurements. Using this novel approach, we find that a relatively small number of intermolecular distance constraints are sufficient to accurately determine the dimensions and polymer conformational statistics of α S and tau in solution. Moreover, this method can detect local changes in α S and tau conformations that correlate with enhanced aggregation. Constrained Monte Carlo simulations produce ensembles that are in excellent agreement both with experimental measurements on α S and tau and with all-atom, explicit solvent molecular dynamics simulations of α S, with much lower configurational sampling requirements and computational expense.

Introduction

A growing number of mammalian proteins are classified as natively unstructured or intrinsically disordered, because they lack in whole or part stable native secondary structure and tertiary interactions. Many in this class of proteins are involved in signaling functions, where they may act as hubs for information flow within the cell (1). A number of intrinsically disordered proteins (IDPs) are also directly implicated in a variety of human diseases, ranging from cancer to Type II diabetes as well as a large spectrum of neurodegenerative disorders (2).

In solution, IDPs populate an ensemble of rapidly-interconverting structures, with the relative populations of the various structures determined by the balance between residue-residue interaction energies and chain entropy. IDPs thus span a continuum of dynamic behavior between well-folded proteins that assume a single native state representing a global minimum in the conformational energy landscape (3), and ideal random coils which lack spatial or temporal correlations between residues (4). Characterizing the partially-ordered and disordered structures sampled by IDPs remains a challenging problem.

A variety of biophysical techniques have been used to provide indirect structural insight into IDPs (5): circular dichroism and infra-red spectroscopic techniques can probe the fractions of IDP residues involved in transient α -helical or β -strand secondary structures (6). Nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) can provide information on the local chemical environment around particular residues, as well as on mean inter-residue distances and relative orientations (5, 7). Fluorescence spectroscopy can probe similar inter-residue distances, as well as the dynamics of conformational change and reconfiguration within the solution-state ensemble (8). Small-angle X-ray or neutron scattering (SAXS or SANS, respectively) (7) and diffusion-based measurements (9, 10) can provide measures of global dimension, such as the radius of gyration (R_g) or hydrodynamic radius (R_h).

Computational approaches have also been frequently employed to study IDP conformational ensembles. Most attempts to simulate IDPs involve one of two approaches: 1) molecular dynamics (MD) simulations and enhanced-sampling variants (11-13), in which a physical representation of a protein chain evolves over time according to Newtonian equations of motion; and 2) stochastic conformational searches in which Monte Carlo (MC)-type algorithms are used to efficiently sample conformations of the protein chain (14). Both types of approaches can be applied to all-atom representations of proteins and solvent, or to models coarse-grained at varying levels of spatial and temporal resolution. The solvent can be represented either as a set of implicit corrections to protein-protein interactions, or as explicit molecules (15, 16). Simulated ensembles are also sensitive to the choice of the protein force-field. An important caveat is that these force-fields are typically parameterized to reproduce the behavior of folded proteins rather than IDPs, and thus they may fail to capture important aspects of IDP ensembles (17-20).

α -Synuclein (α S; Figure 1a) is a 140-residue protein whose aggregation is linked directly to the development and pathogenesis of Parkinson's disease (PD). Tau (Figure 1b) is a microtubule-associated protein up to 441 residues long (subject to alternative splicing) that forms characteristic neuronal aggregates in several human disorders collectively called tauopathies, including Alzheimer's disease (AD), Pick's disease and chronic traumatic encephalopathy (CTE). Both α S and tau are intrinsically disordered in their unbound monomeric states, but can self-assemble into a variety of partially or highly ordered oligomeric and fibrillar aggregates.

Both proteins are subjects of significant interest, not only because of their clinical relevance but also as models for other IDPs that self-assemble into amyloid aggregates. The solution-state ensembles of α S and tau have been studied by NMR (21-25), fluorescence (26-32), SAXS (33-39), and a variety of computational methods (13, 23, 40-43). Despite this extensive effort, there is significant disagreement about even the most fundamental structural properties of these proteins in solution (*e.g.*, reported R_g values for α S vary nearly 2-fold, from 2.6 nm (34) to 5.0 nm (36)), possibly due to aggregation or oligomerization in solution at the concentrations (30 μ M – 800 μ M) necessary for SAXS and NMR experiments. Fluorescence techniques such as single-molecule F \ddot{o} rster resonance energy transfer (smFRET) circumvent these complications by utilizing much lower protein concentrations (\sim 50 pM), but provide qualitatively lower-resolution data. Given that IDPs are increasingly recognized as viable pharmacological targets (44-46), smFRET may also be particularly well-suited to drug candidate screening: the low protein concentrations used should enable the accurate identification of compounds that alter the target conformation by binding with very high affinity (K_{DS} as low as \sim 500 pM). On the computational side, MD studies of α S in solution have typically required non-physiological temperatures (40, 41) to generate R_g and R_h values that agree with experiments. While fragments of tau have been studied by all-atom MD (47, 48), the full-length protein is too long (441 residues) for meaningful simulations to be run at reasonable computational expense with current resources. More consistent results have been obtained for both α S and tau by *post facto* filtering of an unconstrained ensemble to fit experimental data (23, 43) and using coarse-grained Monte Carlo approaches (49).

In this work, we describe computational studies based on our extensive experimental smFRET measurements that provide a quantitative description of α S and tau solution-state conformations. We first develop an experimentally-constrained Monte Carlo (ECMC) method to incorporate long-range pairwise distance constraints from smFRET into MC simulations, and find that remarkably few constraints are sufficient to reproduce the global dimensions of α S and tau found experimentally under various conditions. For α S, we find that all-atom MD (AAMD) simulations in explicit solvent accurately capture the average pairwise inter-residue distances of α S in solution, albeit at significantly higher computational cost than ECMC. The larger size of tau makes the computational cost for analogous AAMD simulations impractically high. Finally, unconstrained MC (UMC) simulations that include dihedral angle potentials, hydrogen bonding, effective residue-residue and residue-environment interactions tuned to match the global dimensions of α S also matched the AAMD ensemble and smFRET observations in terms of polymer scaling behavior and mean inter-residue distances. Both UMC and AAMD show large fluctuations about these mean distances that may be important in determining the populations of aggregation-prone states. The ECMC technique can be generalized to other unstructured proteins and will enable efficient characterization of the native-state ensembles of IDPs, and of the conformational changes produced by environmental factors, covalent modifications, and small-molecule modulators of function and dysfunction.

Materials & Methods

Details on smFRET data collection, constraint generation and resampling-based error estimation procedures are provided in the Supporting Material.

Monte Carlo (MC) Simulations

All MC calculations were implemented using the PyRosetta (50) and Scientific Python libraries for Python on Yale University High Performance Computing clusters. α S and tau were modeled with all backbone atoms represented explicitly, and each side-chain represented by a single bead located at the centroid of the appropriate atoms. For the ECMC simulations, the effective potential consisted of purely repulsive Lennard Jones (LJ) interactions and harmonic restraints with mean distances and variances derived from smFRET measurements (27, 51) as detailed in the Supporting Material. For the UMC simulations, the potential included the *pair* (statistical residue-residue interaction), *env* (statistical residue-environment interaction), *rama* (backbone dihedral propensities), *hbond_sr_bb* and *hbond_lr_bb* (short- and long-range backbone hydrogen bonding, respectively) terms of the Rosetta force-field, each with equal weights. The conformational sampling involved $\sim 10^6$ backbone moves for each model with the effective temperature decreased from 10 k_BT to the final target temperature (0.1 k_BT for ECMC, varied for UMC) over the course of the refinement. Ensembles of at least 400 different structures were generated for each set of parameters.

All-atom molecular dynamics (AAMD) simulations

We performed AAMD simulations of α S in explicit water using the fixed NPT ensemble (also known as the isobaric-isothermal ensemble, i.e., with the number of particles, pressure and temperature held constant) at temperature $T=293\text{K}$ and pressure $P=1$ bar, conditions which closely match the smFRET experiments, to measure the solution-state ensemble of α S. These simulations were implemented in Gromacs 4.5.1 (52) using the Amber99SB (53, 54) force-field with the TIP4P-Ew water model (55). This parameter set produced markedly better agreement with ECMC R_g values in preliminary simulations than other combinations of force-field and solvent model (not shown), and has also been reported to provide superior accuracy in protein simulations (17-19, 56, 57), especially in describing flexible protein regions (56). Pressure was controlled by the weak coupling method (58) with a 1.0 ps coupling time and temperature by a stochastic velocity distribution thermostat of Bussi *et al.* that yields the canonical distribution (59) with a 0.1 ps coupling time. The temperature of the protein was controlled independently of that of the solute and ions. Consistent with the original Amber99sb force-field parameterization (53), non-bonded interactions were cut-off at 0.8 nm and the particle-mesh Ewald method was employed to calculate the long-ranged electrostatic interactions. A time step of 2 fs and a cubic simulation box with volume $(11\text{ nm})^3$ were used for all simulations. The simulation volume is very large, but it is possible even with this box size for a completely extended protein configuration to interact with itself under periodic boundary conditions. However, such events are rare and were not included in the analysis. Ten different initial configurations randomly selected from the end states of ECMC simulations were run for a total of 574 ns. This ensured independent simulation trajectories for improved configurational sampling. The initial 10 ns of each trajectory were omitted from the analysis to allow for structural relaxation, leaving a total of 474 ns of simulation time used to generate the AAMD ensemble.

Results

ECMC Simulations

SmFRET uses the efficiency of energy transfer between two labeled residues in a protein chain as a sensitive probe of protein conformation on the nm scale. We previously reported ET_{eff} values between twelve pairs of residues in α S (27, 51), and have also recently measured

analogous ET_{eff} between twelve residue pairs in tau (60). These datasets each represent the most extensive set of inter-residue distances measured on α S and tau, respectively, at the single-molecule level. In the case of dynamic random coils like α S and tau, the use of polymer models which reflect the ensemble of conformations sampled by the protein on the timescale of smFRET measurements (~ 1 ms) has been shown to lead to more accurate calculations of distances from ET_{eff} than the simple application of the Förster equation (61-63). We used results from excluded-volume MC simulations to convert the smFRET measurements of ET_{eff} into distance constraints, as detailed in the Supporting Material.

For all MC simulations, the proteins were modeled using the ‘centroid’ representation of Rosetta (64), which consists of an all-atom peptide backbone and a single-bead pseudo-atom for each residue side chain (accounting for side-chain size, and the location of the average center of mass of favorable side-chain conformations). We performed a MC conformational search implemented in the PyRosetta software package (50, 64), with only smFRET-derived constraints and purely repulsive LJ interactions included in the potential, using up to 5×10^6 backbone moves. From this ensemble of structures produced by the ECMC simulations, we generated R_g distributions, inter-residue distance maps, and the scaling behavior of the inter-residue distance with sequence separation. Totals of 10^6 and 5×10^6 backbone moves were sufficient to ensure convergence of these three structural characteristics for α S and tau, respectively. These protocols were used for all subsequent simulations.

To study the effects of harmonic constraints on α S ensembles, we performed ECMC with various subsets of the constraints derived from smFRET measurements at pH 7.4. The mean R_g decreased from 5.1 ± 0.9 nm without constraints (the excluded-volume random-coil limit for the α S chain), to 3.3 ± 0.3 nm for the fully-constrained ensemble (Figure 2a). This compaction relative to a random coil involves enhanced local contacts between the N- and C-terminal regions of α S (Figure 2b). Polymer conformational statistics are frequently described in terms of the relation between mean inter-residue distance r and sequence separation N :

$$r = CN^{\nu}, \quad (1)$$

where ν is the Flory scaling exponent, and the pre-factor C depends roughly on the size of the monomer unit. The α S ensemble obtained by ECMC possesses a mean scaling exponent of 0.39 (Figure 2c), intermediate between the exponents expected for an excluded-volume random coil (0.59) and a compact globule (0.33), which suggests the presence of attractive interactions within the disordered α S chain. An important caveat is that the scaling laws in polymer theories apply in the long chain limit, while α S and tau have chain lengths of 140 residues and 441 residues respectively. We are using the scaling form of the inter-residue separation in Equation 1 as an empirical characterization of the size of α S and tau.

Major sources of uncertainty in smFRET are experimental error in the determination of ET_{eff} (estimated to be ~ 1 -2% for our instrumentation based on repeated smFRET measurements of a standard sample), and variation in the Förster radius due to anisotropic dye tumbling and/or changes in quantum yield (estimated to be 7-8% based on fluorescence lifetime and anisotropy measurements). We implemented a resampling error estimation protocol described in the Supporting Material to quantify the resulting uncertainties in ECMC results, which are

represented by the shaded area in Figure 2c. Due to the redundant nature of smFRET-derived constraints, the standard error of the mean in ensemble properties such as R_g or ν is $\sim 1.5\%$.

ECMC using smFRET measurements of α S at pH 3.0, a condition that is known to accelerate aggregation (27, 67), indicates a slight compaction in terms of the mean R_g (3.2 ± 0.3 nm) relative to the neutral-pH ensemble (Figure 2d). However, the low-pH ensemble shows significant changes in the pattern of mean inter-residue distances, with decreased contact between the N-terminus and the central region, and increased contact between the C-terminus and much of the rest of the protein (Figure 2e).

Similarly, we compared smFRET distance measurements between 12 residue pairs (60) for tau, in the presence and absence of the polyanion heparin, which is known to dramatically enhance the aggregation rate of tau in solution. Like α S, tau displays a progressive compaction relative to an excluded-volume random-coil of equal length ($R_g = 10.9 \pm 1.9$ nm) as the number of constraints is increased (Figure S2). The mean R_g of tau increases from 5.1 ± 0.5 nm to 6.0 ± 0.6 nm upon the addition of heparin (Figure 3b), while the corresponding R_h values calculated by Hydropro (68) increase from 5.3 ± 0.3 nm to 5.9 ± 0.5 nm. In the absence of heparin, tau demonstrates polymer scaling intermediate between the random coil and globule limits at small and medium sequence separations, but is markedly more compact at larger separations, so that the overall mean scaling exponent is 0.27. Heparin eliminates this long-range compaction, increasing the mean value of ν to 0.44 (Figure 3a). This expansion, particularly evident between residues 1-50 and the span from residues 180-441, is partially offset by compaction between residues 140-220 and residues 340-420 (Figure 3c).

Unconstrained MC and All-Atom MD Simulations

To further test the relation between R_g and inter-residue distances, we investigated how accurately unconstrained simulations that reproduce the global dimensions of α S could predict the mean smFRET-derived distances. UMC simulations using a more complete Rosetta potential were performed over a range of effective temperatures to identify conditions that best reproduced the mean R_g obtained by ECMC. The potential used for UMC simulations included statistically-obtained residue-environment and residue-residue interaction terms, backbone dihedral angle propensities, backbone hydrogen-bonding, and repulsive LJ interaction terms, reflecting charge interactions, hydrophobicity and polarity, and the sterics of the protein backbone. The UMC ensemble at $1.2 k_B T$ most accurately captured the mean R_g of α S at neutral pH, and the mean inter-residue distances extracted from this ensemble were in excellent agreement with the corresponding distances measured by smFRET (Figure 4a). We further performed all-atom, explicit-solvent MD (AAMD) on α S using the Amber99SB (53, 54) force-field with the TIP4P-Ew water model (55) at 293 K. This combination of parameters produces an ensemble with a R_g of 3.1 ± 0.9 nm, in close agreement with mean experimental values, and reproduces the mean inter-residue distances measured by smFRET with high accuracy (Figure 4a). The ensemble generated by AAMD resembles that produced by UMC simulations (Figure 4b, c) but the computational cost is significantly higher. Properties of the ensembles generated by AAMD, UMC and ECMC are compared in greater detail in the Supporting Material.

Discussion

ECMC Simulations

The ECMC protocol we have developed here is an efficient, powerful probe of α S and tau global dimensions, polymer scaling behavior, and local contacts. ECMC accurately reproduces the global dimensions of α S given pairwise distances from smFRET experiments: the mean R_g decreases smoothly and asymptotes to the experimental value as the number of constraints increases, which indicates that the smFRET data are internally self-consistent. Remarkably, as few as two or three constraints are enough to capture much of the compaction of α S, even though the only other term in the effective potential for these simulations is the repulsive LJ interaction. The measurement of R_g by ECMC (3.3 nm) lies within the broad range (2.6 nm to 5.0 nm) suggested by previous SAXS experiments (33-36), but is substantially more precise than these ensemble measurements (see details of error estimation in Supporting Material). We recently measured the R_h value of α S to be 2.74 ± 0.04 nm using pulsed field gradient NMR (69). The R_g/R_h ratio of α S is therefore ~ 1.2 , in excellent agreement with the analogous ratio measured for unfolded protein L using smFRET and fluorescence correlation spectroscopy (FCS) (62).

The polymer scaling behavior of α S as determined by ECMC also recapitulates very closely the results of tryptophan triplet-state lifetime (65) and time-resolved FRET (29) experiments, even at small sequence separations where the inter-residue distances are too short for direct measurement by smFRET. We find that the scaling behavior depends on sequence separation, with the Flory exponent ν changing from ~ 0.59 for $N < 40$ to ~ 0.34 for $N > 40$. The intriguing observation that the scaling exponent for α S shifts from nearly the self-avoiding random-coil value to the globule value as sequence separation increases may reflect the fact that the intrinsic stiffness and excluded-volume properties of the polypeptide backbone predominate at small N while intramolecular attractive interactions predominate at large N . The long-range contacts that cause this compaction in the ECMC ensembles resemble those observed in paramagnetic relaxation enhancement (PRE) NMR experiments (21, 23) and numerical simulations (43), particularly in terms of the enhanced contacts between residues 1-30 and 100-140 (23, 43). Approximately 11% of the structures in the ECMC ensemble display end-to-end distances below 5 nm. The inter-residue distance distributions obtained from ECMC simulations appear largely unimodal (red histograms in Figure 4a).

The polymer scaling of tau in the absence of heparin is qualitatively similar to that of α S, with a progressive compaction with increasing N : the exponent ν changes from ~ 0.60 for $N < 40$, to ~ 0.39 for $40 < N < 140$, and then to ~ 0.09 for $N > 140$. This marked deviation from excluded-volume power-law scaling behavior at high sequence separations to scaling that is below the theoretical globule limit may reflect specific interactions between the N- and C-termini of tau as suggested by Mandelkow and co-workers (31, 32) and also observed by PRE NMR (25). Heparin has a dramatic effect on this behavior: the exponent ν changes from ~ 0.62 for $N < 40$, to ~ 0.43 for $N > 40$, while the hypercompaction at high N is eliminated. This expansion is also reflected in the R_g and calculated R_h values of tau, which increase by 19% and 11% respectively in the presence of heparin. The latter change is in excellent agreement with FCS experiments which show a 14% increase in the diffusion time upon binding of heparin (60).

ECMC is also sensitive to local changes in structure that do not have a major effect on global dimensions. While decreasing the pH from 7.4 to 3.0 results in only small changes in α S R_g (3.3 nm vs. 3.2 nm), the two ECMC ensembles are significantly different in local structure. The decreased charge on the C-terminus at low pH leads to a significant collapse of this region, and

enhanced interactions with residues 25-100. This span includes the aggregation prone region (residues 60-95) of α S, and so this altered interaction may relate to the increased aggregation propensity of α S at low pH. There is a compensatory expansion of the N-terminus that may be due to the increased net positive charge of this region in acidic environments. This observation agrees very well with PRE NMR experiments on α S at low and neutral pH (23, 41), which also show a relative collapse of the C-terminus and slight expansion of the N-terminus. Similarly, although the addition of heparin caused a moderate increase in the global dimensions of tau, this masks a compaction of the central microtubule-binding region that forms the core of tau-derived amyloid aggregates. Heparin also eliminates the long-range contacts between the termini of tau that have been previously observed by ensemble FRET (31, 32). These conformational changes may be driven by neutralization of the net positive charge on the proline-rich and microtubule-binding regions of tau by the polyanion heparin. Our results illustrate the ability of ECMC to effectively identify local changes in structure, even in the absence of major global changes, which could provide crucial insight into the environmental factors and binding partners that affect the tendency of α S and tau to pathologically self-assemble.

UMC and AAMD Simulations

Having demonstrated that ECMC accurately reproduces global dimensions from pairwise constraints, we addressed the inverse problem of whether it is possible to derive pairwise inter-residue separations from global dimensions of the chain. UMC simulations with a single free parameter (effective temperature) tuned to match the mean R_g accurately reproduced mean inter-residue distances that span nearly the complete sequence of α S. The polymer scaling behavior and inter-residue distance map of the UMC ensemble also resemble those of the ECMC ensemble at neutral pH. This remarkable result indicates that it is possible to infer computationally much of the detailed conformational properties of α S given only the mean R_g of the ensemble and the robust knowledge-based potential implemented in Rosetta.

To investigate whether this result is specific to Rosetta, or also applies to other modern simulation techniques, we performed extensive AAMD simulations on α S which faithfully reflected α S solution state behavior at physiological temperature, without any bias or *post facto* filtering. The combination of Amber99SB and TIP4P-Ew solvent appears to simulate the α S conformational ensemble with high accuracy, in good agreement with the extensive results from other groups (17-20, 56, 57) successfully applying this parameter set to disordered proteins and peptides. The high accuracy in reproducing the mean inter-residue separations, as measured by smFRET, is especially noteworthy given that previous MD studies (with other force-fields and solvent treatments) had to resort to temperatures 150-300 K above physiological values to achieve similar agreement with experiment (40, 41). Recent corrections to the Amber99SB forcefield (70) used with TIP3P solvent have shown improved accuracy in predicting the folding behavior of several model peptides and proteins (71, 72), and in future work we will test whether this parameter set also accurately reproduces α S solution-state behavior.

Conclusion

In this work, we have developed and validated a new approach to effectively investigate the conformational behavior of disordered proteins, using Monte Carlo simulations constrained by single-molecule FRET experiments. In particular, we have shown that the global dimensions and

polymer conformational statistics of the model protein α S and tau can be accurately determined from a relatively small number of inter-residue distance measurements, and produce ensembles in excellent agreement with previous experimental studies, as well as with the AAMD simulations in the case of α S. Conversely, we have shown that recently-developed protein energy functions make it possible to predict detailed conformational properties and mean inter-residue distances of α S from its global dimensions, and that transient charge-mediated contacts between residues distant in sequence may underlie the dynamic conformational fluctuations of α S and tau suggested by our findings and others (30, 65, 73, 74).

The ECMC approach is much less computationally-intensive than simulation methods that predict IDP conformational ensembles with comparable accuracy, which become rapidly more demanding as the size of the protein of interest increases. For example, comparable AAMD studies on tau would require simulation of a system at least an order of magnitude larger in volume (to accommodate the greater dimensions of tau), for a substantially longer period of time (to sample the increased number of degrees of freedom of the tau chain with a similar level of detail to α S). Given typically available resources, the consequent computational expense makes routine AAMD simulations of large IDPs infeasible. ECMC could therefore serve as an efficient alternative strategy to characterize the detailed conformations of diverse IDPs. ECMC based on smFRET measurements is also well-suited to the identification of high-affinity binding partners of IDPs; this is potentially of broad pharmacological relevance, given that IDPs may be effective drug targets for a range of diseases (46). ECMC could therefore prove to be a versatile probe of the effects of environmental factors, drug-like small molecules, and biological ligands on IDP conformations in health and disease.

Author Information

Corresponding Authors

*Elizabeth Rhoades, Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, P. O. Box 208114, New Haven, CT 06520-8114, phone: +1(203)432 5342, email: elizabeth.rhoades@yale.edu; Corey O'Hern, Department of Mechanical Engineering & Materials Science, Yale University, P. O. Box 208286, New Haven, CT 06520-8286, phone: +1(203)432 4258, email: corey.ohern@yale.edu; Abhinav Nath, Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, P.O. Box 208114, New Haven, CT 06520-8114, phone: +1(203)432-5647, email: abhinav.nath@yale.edu.

Acknowledgment

This research was supported by the National Science Foundation under Grant Nos. DMS-0835742 (to CSO) and MCB-0919853 (to ER), The Ellison Medical Foundation (to ER), an American Heart Association Postdoctoral Fellowship (to AN), the Academy of Finland (to MS), an institutional training grant from the National Institute of Health GM007223 (to DCD, AJT, and SE-G), and a seed grant from the Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences (to CSO and ER). The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement #293861 (to MS). This work also benefited from the facilities and staff of

the Yale University Faculty of Arts and Sciences High Performance Computing Center, CSC IT Center for Science in Finland (MS), and NSF Grant No. CNS-0821132 that partially funded acquisition of the computational facilities.

Abbreviations

AAMD, all-atom molecular dynamics; ECMC, experimentally-constrained Monte Carlo; EPR, electron paramagnetic resonance; ET_{eff} , energy transfer efficiency; (sm)FRET, (single molecule) Förster resonance energy transfer; IDP, intrinsically disordered protein; LJ, Lennard-Jones; MC, Monte Carlo; MD, molecular dynamics; NMR, nuclear magnetic resonance; PRE, paramagnetic relaxation enhancement; SAX(N)S, small-angle X-ray (neutron) scattering; UMC, unconstrained Monte Carlo.

Supporting Citations

References (75-78) appear in the Supporting Material.

References

1. Dunker, A.K., M.S. Cortese, P. Romero, L.M. Iakoucheva, and V.N. Uversky. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272: 5129-48.
2. Midic, U., C.J. Oldfield, A.K. Dunker, Z. Obradovic, and V.N. Uversky. 2009. Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. *Protein Peptide Lett.* 16: 1533-47.
3. Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science.* 181: 223-230.
4. Flory, P.J. 1942. Thermodynamics of high polymer solutions. *J. Chem. Phys.* 10: 51.
5. Eliezer, D. 2009. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struc. Biol.* 19: 23-30.
6. Receveur-Bréchet, V., J.-M. Bourhis, V.N. Uversky, B. Canard, and S. Longhi. 2006. Assessing protein disorder and induced folding. *Proteins.* 62: 24-45.
7. Mittag, T., and J.D. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struc. Biol.* 17: 3-14.
8. Chen, H., and E. Rhoades. 2008. Fluorescence characterization of denatured proteins. *Curr. Opin. Struc. Biol.* 18: 516-24.
9. Price, W.S. 1997. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. *Concepts in Magnetic Res.* 9: 299-336.

10. Haustein, E., and P. Schuille. 2007. Fluorescence correlation spectroscopy: novel variations of an established technique. *Annu. Rev. Biophys. Bio.* 36: 151-69.
11. Leone, V., F. Marinelli, P. Carloni, and M. Parrinello. 2010. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struc. Biol.* 20: 148-54.
12. Scheraga, H.A., M. Khalili, and A. Liwo. 2007. Protein-folding dynamics: overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* 58: 57-83.
13. Rauscher, S., and R. Pomès. 2010. Molecular simulations of protein disorder. *Biochem. Cell. Biol.* 88: 269-90.
14. Vitalis, A., and R.V. Pappu. 2009. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* 5: 49-76.
15. Dill, K.A., T.M. Truskett, V. Vlachy, and B. Hribar-Lee. 2005. Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Bio.* 34: 173-99.
16. Chen, J., C.L. Brooks, and J. Khandogin. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struc. Biol.* 18: 140-8.
17. Fawzi, N.L., A.H. Phillips, J.Z. Ruscio, M. Doucleff, D.E. Wemmer, et al. 2008. Structure and dynamics of the A β (21-30) peptide from the interplay of NMR experiments and molecular simulations. *J. Am. Chem. Soc.* 130: 6145-58.
18. Aliev, A.E., and D. Courtier-Murias. 2010. Experimental verification of force fields for molecular dynamics simulations using Gly-Pro-Gly-Gly. *J. Phys. Chem. B.* 114: 12358-75.
19. Lange, O.F., D. van der Spoel, and B.L. de Groot. 2010. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.* 99: 647-655.
20. Piana, S., K. Lindorff-Larsen, and D.E. Shaw. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100: L47-9.
21. Dedmon, M.M., K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C.M. Dobson. 2005. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* 127: 476-7.
22. Sung, Y.-H., and D. Eliezer. 2007. Residual structure, backbone dynamics, and interactions within the synuclein family. *J. Mol. Biol.* 372: 689-707.
23. Salmon, L., G. Nodet, V. Ozenne, G. Yin, M.R. Jensen, et al. 2010. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132: 8407-18.

24. Mukrasch, M.D., P. Markwick, J. Biernat, M. von Bergen, P. Bernadó, et al. 2007. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* 129: 5235-43.
25. Mukrasch, M.D., S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, et al. 2009. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* 7: e34.
26. Ferreon, A.C.M., C.R. Moran, J.C. Ferreon, and A.A. Deniz. 2010. Alteration of the alpha-synuclein folding landscape by a mutation related to Parkinson's disease. *Angew. Chem. Int. Ed.* 49: 3469-72.
27. Trexler, A.J., and E. Rhoades. 2010. Single Molecule Characterization of α -Synuclein in Aggregation-Prone States. *Biophys. J.* 99: 3048-3055.
28. Rhoades, E., T.F. Ramlall, W.W. Webb, and D. Eliezer. 2006. Quantification of alpha-synuclein binding to lipid vesicles using fluorescence correlation spectroscopy. *Biophys. J.* 90: 4692-700.
29. Grupi, A., and E. Haas. 2010. Segmental conformational disorder and dynamics in the intrinsically disordered protein α -Synuclein and its chain length dependence. *J. Mol. Biol.* 405: 1283-1267.
30. Grupi, A., and E. Haas. 2011. Time resolved FRET detection of subtle temperature induced conformational biases in ensembles of α -synuclein molecules. *J. Mol. Biol.* 411: 234-47.
31. Jeganathan, S., M. von Bergen, H. Brumlach, H.-J. Steinhoff, and E. Mandelkow. 2006. Global hairpin folding of tau in solution. *Biochemistry.* 45: 2283-93.
32. Jeganathan, S., A. Hascher, S. Chinnathambi, J. Biernat, E.-M. Mandelkow, et al. 2008. Proline-directed pseudo-phosphorylation at AT8 and PHF1 epitopes induces a compaction of the paperclip folding of Tau and generates a pathological (MC-1) conformation. *J. Biol. Chem.* 283: 32066-76.
33. Li, J. 2002. Conformational behavior of human α -synuclein is modulated by familial parkinson's disease point mutations A30P and A53T. *NeuroToxicology.* 23: 553-567.
34. Tashiro, M., M. Kojima, H. Kihara, K. Kasai, T. Kamiyoshihara, et al. 2008. Characterization of fibrillation process of alpha-synuclein at the initial stage. *Biochem. Biophys. Res. Commun.* 369: 910-4.
35. Rekas, A., R.B. Knott, A. Sokolova, K.J. Barnham, K.A. Perez, et al. 2010. The structure of dopamine induced alpha-synuclein oligomers. *Eur. Biophys. J.* 39: 1407-19.

36. Giehm, L., D.I. Svergun, D.E. Otzen, and B. Vestergaard. 2011. Low-resolution structure of a vesicle disrupting α -synuclein oligomer that accumulates during fibrillation. *Proc. Natl. Acad. Sci. USA.* 108: 3246-51.
37. Schweers, O., E. Schönbrunn-Hanebeck, A. Marx, and E. Mandelkow. 1994. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J. Biol. Chem.* 269: 24290-7.
38. Mylonas, E., A. Hascher, P. Bernadó, M. Blackledge, E. Mandelkow, et al. 2008. Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry.* 47: 10345-53.
39. Shkumatov, A.V., S. Chinnathambi, E. Mandelkow, and D.I. Svergun. 2011. Structural memory of natively unfolded tau protein detected by small-angle X-ray scattering. *Proteins.* 79: 2122-31.
40. Allison, J.R., P. Varnai, C.M. Dobson, and M. Vendruscolo. 2009. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* 131: 18314-26.
41. Wu, K.-P., D.S. Weinstock, C. Narayanan, R.M. Levy, and J. Baum. 2009. Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J. Mol. Biol.* 391: 784-96.
42. Tsigelny, I.F., Y. Sharikov, M.A. Miller, and E. Masliah. 2008. Simulation and modeling of synuclein-based 'protofibril' structures: as a means of understanding the molecular basis of Parkinson's disease. *J. Phys. Conf. Ser.* 125: 012056.
43. Ullman, O., C.K. Fisher, and C.M. Stultz. 2011. Explaining the structural plasticity of α -synuclein. *J. Am. Chem. Soc.* 133: 19536-46.
44. Cheng, Y., T. LeGall, C.J. Oldfield, J.P. Mueller, Y.-Y.J. Van, et al. 2006. Rational drug design via intrinsically disordered protein. *Trends. Biotechnol.* 24: 435-42.
45. Masuda, M., N. Suzuki, S. Taniguchi, T. Oikawa, T. Nonaka, et al. 2006. Small molecule inhibitors of alpha-synuclein filament assembly. *Biochemistry.* 45: 6085-94.
46. Metallo, S.J. 2010. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14: 481-8.
47. Zhao, J.-H., H.-L. Liu, C.-K. Chuang, K.-T. Liu, W.-B. Tsai, et al. 2010. Molecular dynamics simulations to investigate the stability and aggregation behaviour of the amyloid-forming peptide VQIVYK from tau protein. *Mol. Simulat.* 36: 1013-1024.

48. Miller, Y., B. Ma, and R. Nussinov. 2011. Synergistic interactions between repeats in tau protein and A β amyloids may be responsible for accelerated aggregation via polymorphic states. *Biochemistry*. 50: 5172-81.
49. Jho, Y.S., E.B. Zhulina, M.W. Kim, and P.A. Pincus. 2010. Monte carlo simulations of tau proteins: effect of phosphorylation. *Biophys. J.* 99: 2387-97.
50. Chaudhury, S., S. Lyskov, and J.J. Gray. 2010. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*. 26: 689-91.
51. Trexler, A.J., and E. Rhoades. 2009. Alpha-synuclein binds large unilamellar vesicles as an extended helix. *Biochemistry*. 48: 2304-6.
52. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4: 435-447.
53. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, et al. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 65: 712-25.
54. Sorin, E.J., and V.S. Pande. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* 88: 2472-93.
55. Horn, H.W., W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, et al. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120: 9665-78.
56. Showalter, S.A., E. Johnson, M. Rance, and R. Brüschweiler. 2007. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J. Am. Chem. Soc.* 129: 14146-7.
57. Wong, V., and D.A. Case. 2008. Evaluating rotational diffusion from protein MD simulations. *J. Phys. Chem. B.* 112: 6013-24.
58. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81: 3684.
59. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126: 014101.
60. Elbaum-Garfinkle, S., and E. Rhoades. 2012. Long-range interactions modulate aggregation of tau by altering the conformational ensemble. Manuscript submitted. .

61. O'Brien, E.P., G. Morrison, B.R. Brooks, and D. Thirumalai. 2009. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J. Chem. Phys.* 130: 124903.
62. Sherman, E., and G. Haran. 2006. Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA.* 103: 11539-43.
63. McCarney, E.R., J.H. Werner, S.L. Bernstein, I. Ruczinski, D.E. Makarov, et al. 2005. Site-specific dimensions across a highly denatured protein; a single molecule study. *J. Mol. Biol.* 352: 672-82.
64. Rohl, C.A., C.E.M. Strauss, K.M.S. Misura, and D. Baker. 2004. Protein Structure Prediction Using Rosetta. *Method. Enzymol.* 383: 66-93.
65. Lee, J.C., H.B. Gray, and J.R. Winkler. 2005. Tertiary contact formation in alpha-synuclein probed by electron transfer. *J. Am. Chem. Soc.* 127: 16388-9.
66. Grupi, A., and E. Haas. 2010. Segmental conformational disorder and dynamics in the intrinsically disordered protein α -Synuclein and its chain length dependence. *J. Mol. Biol.* 405: 1267-83.
67. Uversky, V.N., J. Li, and A.L. Fink. 2001. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* 276: 10737-44.
68. García De La Torre, J., M.L. Huertas, and B. Carrasco. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78: 719-30.
69. Sevcsik, E., A.J. Trexler, J.M. Dunn, and E. Rhoades. 2011. Allostery in a disordered protein: oxidative modifications to α -synuclein act distally to regulate membrane binding. *J. Am. Chem. Soc.* 133: 7152-8.
70. Best, R.B., and G. Hummer. 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B.* 113: 9004-15.
71. Mittal, J., and R.B. Best. 2010. Tackling force-field bias in protein folding simulations: Folding of villin HP35 and Pin WW domains in explicit water. *Biophys. J.* 99: L26-L28.
72. Best, R.B., and J. Mittal. 2010. Balance between alpha and beta Structures in Ab Initio Protein Folding. *J. Phys. Chem. B.* 114: 8790-8798.
73. Bernstein, S.L., D. Liu, T. Wyttenbach, M.T. Bowers, J.C. Lee, et al. 2004. Alpha-synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. *J. Am. Soc. Mass Spectr.* 15: 1435-43.

74. Natalello, A., F. Benetti, S.M. Doglia, G. Legname, and R. Grandori. 2011. Compact conformations of α -synuclein induced by alcohols and copper. *Proteins*. 79: 611-21.
75. Lee, J.C., B.T. Lai, J.J. Kozak, H.B. Gray, and J.R. Winkler. 2007. Alpha-synuclein tertiary contact dynamics. *The Journal of Physical Chemistry. B*. 111: 2107-12.
76. Wittkop, M., S. Kreitmeier, and D. Göritz. 1996. The distribution function of internal distances of a single polymer chain with excluded volume in two and three dimensions: A Monte Carlo study. *The Journal of Chemical Physics*. 104: 351.
77. Caracciolo, S., M.S. Causo, and A. Pelissetto. 2000. End-to-end distribution function for dilute polymers. *The Journal of Chemical Physics*. 112: 7693.
78. Dale, R., J. Eisinger, and W. Blumberg. 1979. The orientational freedom of molecular probes: The orientation factor in intermolecular energy transfer. *Biophysical Journal*. 26: 161-93.

Figure legends

Figure 1. (a) Schematic of 140-residue α S in solution. Residues 60-95 (white) form the core of amyloid fibrils, while residues 1-95 form an amphipathic helix in the presence of anionic lipid membranes. The N-terminal region (cyan) bears a net positive charge, while the C-terminus (yellow) is highly negatively charged at pH 7.4 but neutral at pH 3.0. (b) Schematic of the 441-residue isoform of tau. The proline-rich region (residues 151-243) is marked in orange, while the microtubule-binding region (residues 244-369; yellow) both forms the core of amyloid aggregates and is also the putative binding site for the molecular aggregation inducer heparin.

Figure 2. (a) Mean R_g values obtained by ECMC using between 0 and 12 randomly-selected smFRET-derived constraints of α S. Grey circles and associated vertical lines indicate means and standard deviations of distributions obtained from a single randomly selected subset of the full constraint set, with the number of constraints indicated on the horizontal axis. Black circles denote means over all subsets for a given number of constraints. (b) An inter-residue distance difference map shows the change in mean inter-residue distances between the unconstrained random coil ensemble and ECMC-derived α S ensemble at pH 7.4. Compaction (shown in red) relative to the random coil is particularly evident between residues 1-30 and 100-140. (c) The ECMC-derived α S ensemble (red solid line) demonstrates polymer scaling intermediate between the excluded-volume random-coil limit (obtained from unconstrained MC simulations of α S with purely repulsive LJ interactions and no smFRET constraints) and a compact globule ($C = 0.38$ nm, $\nu = 0.33$ for Equation 1 in the text), both shown as black dotted lines. The pink shaded area indicates the standard deviation in ECMC-derived r values, obtained by resampling error estimation as described in the Supporting Material. Points are discrete values obtained by smFRET (triangles), tryptophan triplet-state lifetime (65) (circles) and time-resolved FRET (66) (squares) measurements. Black and gray dashed lines are fits of Equation 1 to the ECMC data for $N < 40$ and $N > 40$, respectively. The inset shows data replotted on logarithmic axes. (d) R_g histograms obtained by ECMC of α S at pH 7.4 (red; $R_g = 3.3 \pm 0.3$ nm) and pH 3.0 (blue; 3.2 ± 0.3 nm), compared with the excluded-volume random coil model (black; 5.1 ± 0.9 nm). Low pH, which enhances α S aggregation, has a minor effect on R_g . (e) Distance difference map (pH 3.0–pH 7.4) showing the detailed effect of acidic environments on α S conformation. Compaction is indicated by orange-red, and expansion by blue.

Figure 3. (a) Tau in the absence of heparin (red) displays similar polymer scaling behavior to α S at short and medium sequence separations, between the random coil and globule limits (dotted lines), but displays marked compaction at long sequence separations. This long-range compaction is eliminated by heparin (blue), resulting in an increase in the mean scaling exponent from 0.28 to 0.43. (b) The global expansion induced by heparin is

reflected in an increase in R_g from 5.1 ± 0.5 nm (red) to 6.0 ± 0.6 nm (blue), compared to 10.9 ± 1.9 nm for an excluded-volume random coil (black). c) A distance difference map (heparin - buffer only) reveals that this global expansion (blue) is partially compensated for by a compaction (orange/red) of the proline-rich and microtubule-binding regions of tau in the presence of heparin.

Figure 4. (a) Comparisons between inter-residue distances from ECMC and corresponding distance distributions from UMC (left) and AAMD (right) of α S. Red histograms represent the experimentally-derived inter- C_α distances, while grey histograms represent the corresponding inter-residue distances from the unconstrained simulations; red and black lines represent means of the respective distributions. Both simulation techniques accurately capture the mean pairwise distances but predict broader distance distributions, given either no experimental information (in the case of AAMD), or only the R_g (in the case of UMC). Note that AAMD distance distributions appear to be multimodal, especially for residue pairs spanning the N- and C-terminal regions of α S. (b) R_g distributions of ensembles from UMC (orange) and AAMD simulations (blue) match the mean (3.3 nm) of the R_g distributions from ECMC, but have greater width (cf. Figure 2d). (c) Distance difference maps comparing a random coil simulation with UMC (top) and AAMD (bottom) ensembles closely resemble each other, and the analogous map for ECMC simulations (cf. Figure 2b).

Figures







