

Molecular simulations of the fluctuating conformational dynamics of intrinsically disordered proteins

W. Wendell Smith,¹ Carl F. Schreck,¹ Nabeem Hashem,¹ Sherwin Soltani,¹ Abhinav Nath,² Elizabeth Rhoades,^{1,2} and Corey S. O'Hern^{1,3}

¹*Department of Physics, Yale University, New Haven, Connecticut 06520-8120, USA*

²*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA*

³*Department of Mechanical Engineering and Materials Science, Yale University, New Haven, Connecticut 06520-8286, USA*

(Received 9 August 2012; published 12 October 2012)

Intrinsically disordered proteins (IDPs) do not possess well-defined three-dimensional structures in solution under physiological conditions. We develop all-atom, united-atom, and coarse-grained Langevin dynamics simulations for the IDP α -synuclein that include geometric, attractive hydrophobic, and screened electrostatic interactions and are calibrated to the inter-residue separations measured in recent single-molecule fluorescence energy transfer (smFRET) experiments. We find that α -synuclein is disordered, with conformational statistics that are intermediate between random walk and collapsed globule behavior. An advantage of *calibrated* molecular simulations over constraint methods is that physical forces act on all residues, not only on residue pairs that are monitored experimentally, and these simulations can be used to study oligomerization and aggregation of multiple α -synuclein proteins that may precede amyloid formation.

DOI: [10.1103/PhysRevE.86.041910](https://doi.org/10.1103/PhysRevE.86.041910)

PACS number(s): 87.15.H-, 87.14.E-, 87.15.ap

I. INTRODUCTION

Intrinsically disordered proteins (IDPs) do not possess well-defined three-dimensional structures in physiological conditions. Instead, IDPs can range from collapsed globules to extended chains with highly fluctuating conformations in aqueous solution [1]. IDPs play a significant role in cellular signaling and control, since they can interact with a wide variety of binding targets [2]. In addition, their propensity to aggregate to form oligomers and fibers has been linked to the onset of amyloid diseases [3]. The conformational and dynamic heterogeneity of IDPs makes their structural characterization by traditional biophysical approaches challenging. Also, force fields employed in all-atom molecular dynamics simulations, which are typically calibrated for folded proteins, can yield results that differ significantly from experiments [4].

In this manuscript, we focus on the IDP α -synuclein, which is a 140-residue neuronal protein linked to Parkinson's disease and Lewy body dementia [5]. Previous NMR studies have found that α -synuclein is largely unfolded in solution, but more compact than a random coil of the same length [4,6,7]. The precise mechanism for aggregation in α -synuclein has not been identified, although it is known that aggregation is enhanced at low pH [7–9], possibly due to the loss of long-range contacts between the N and C termini of the protein [10].

Quantitative structural information has been obtained for α -synuclein using single-molecule fluorescence resonance energy transfer (smFRET) between 12 donor and acceptor pairs [11]. These experimental studies have measured inter-residue separations for both the neutral- and low- pH ensembles. Prior studies have implemented the inter-residue separations from smFRET as constraints in Monte Carlo simulations with only geometric (e.g., bond length and bond angle) and repulsive Lennard-Jones interactions to investigate the natively disordered ensemble of conformations for monomeric α -synuclein [12]. In contrast, we develop all-atom, united-atom, and coarse-grained Langevin dynamics simulations of α -synuclein that include geometric, attractive hydrophobic, and screened

electrostatic interactions. The simulations are calibrated to closely match the inter-residue separations from the smFRET experiments. An advantage of this method over constrained simulations is that physical forces, which act on all residues in the protein, are tuned so that the inter-residue separations from experiments and simulations agree. In future studies, we will employ these calibrated Langevin dynamics simulations to study oligomerization and aggregation of multiple α -synuclein proteins over a range of solvent conditions.

II. METHODS

The 140-residue IDP α -synuclein includes a negatively charged N-terminal region, a hydrophobic central region, and a positively charged C-terminal region (Fig. 1) at neutral pH . We study three models for α -synuclein with different levels of geometric complexity: (a) all atom, (b) united atom, and (c) coarse grained, as shown in Fig. 2.

A. All-atom model

The all-atom model (including hydrogen atoms) matches closely the geometric properties of proteins. The average bond lengths $\langle l_{ij} \rangle$, bond angles $\langle \theta_{ijk} \rangle$, and backbone dihedral angle ω between atoms C_α -C-N- C_α on successive residues were obtained from the Dunbrack database of 850 high-resolution protein crystal structures [13]. The 242 distinct bonds and 440 distinct bond angles in α -synuclein were fixed using the following spring potentials:

$$V^{bl} = \frac{k_l}{2} \sum_{ij} (r_{ij} - \langle l_{ij} \rangle)^2, \quad (1)$$

where k_l is the bond-length stiffness and r_{ij} is the center-to-center separation between bonded atoms i and j , and

$$V^{ba} = \frac{k_\theta}{2} \sum_{ijk} (\theta_{ijk} - \langle \theta_{ijk} \rangle)^2, \quad (2)$$

MET ASP VAL PHE MET LYS GLY LEU SER LYS ALA LYS GLU GLY VAL VAL ALA ALA ALA GLU 20
 LYS THR LYS GLN GLY VAL ALA GLU ALA ALA GLY LYS THR LYS GLU GLY VAL LEU TYR VAL 40
 GLY SER LYS THR LYS ASN GLU GLY VAL HIS GLY VAL ALA THR VAL ALA GLU LYS THR LYS 60
 GLU GLN VAL THR ASN VAL GLY GLY ALA VAL VAL THR GLY VAL THR ALA VAL ALA GLN LYS 80
 THR VAL GLU GLY ALA GLY SER ILE ALA ALA ALA THR GLY PHE VAL LYS LYS ASP GLN LEU 100
 GLY LYS ASN GLU GLU GLY ALA PRO GLN GLU GLY ILE LEU GLU ASP MET PRO VAL ASP PRO 120
 ASP ASN GLU ALA TYR GLU MET PRO SER GLU GLU GLY TYR GLN ASP TYR GLU PRO GLU ALA 140

FIG. 1. (Color online) The three main regions of the 140-residue protein α -synuclein. Residues 1–60 form the highly basic N-terminal region (bold, blue), residues 61–95 form the hydrophobic central region (plain text), and residues 96–140 form the acidic C-terminal region (italics, red) [10,11].

where k_θ is the bond-angle stiffness and θ_{ijk} is the angle between bonded atoms i , j , and k . The average backbone dihedral angle between the C_α -C-N- C_α atoms was constrained to zero using

$$V^{da} = \frac{k_\omega}{2} \sum_{ijkl} \omega_{ijkl}^2. \quad (3)$$

We chose $k_l = 5 \times 10^3 k_b T_0 / \text{\AA}^2$ and $k_\theta = k_\omega = 2 \times 10^5 k_b T_0 / \text{rad}^2$ (with $T_0 = 293$ K) so that the rms fluctuations in the bond lengths, bond angles, and dihedral angles were below 0.05 \AA and 0.008 rad, respectively. These rms values occur in the protein crystal structures from the Dunbrack database. Note that no explicit interaction potentials were used to constrain the backbone dihedral angles ϕ and ψ and side-chain dihedral angles. However, the bond lengths, bond angles, and sizes of the atoms were calibrated so that they take on physical values (see Appendix A).

We included three types of interactions between nonbonded atoms: (1) the purely repulsive Lennard-Jones potential V^r to model steric interactions, (2) attractive Lennard-Jones interactions V^a between C_α atoms on each residue to model hydrophobicity, and (3) screened electrostatic interactions V^{es} between atoms in the charged residues LYS, ARG, HIS, ASP, and GLU (i.e. lysine, arginine, histidine, aspartic acid, and glutamic acid, respectively). Thus, the total interaction energy is $V = V^{bl} + V^{ba} + V^{da} + V^r + V^a + V^{es}$ (see Fig. 3.)

The purely repulsive Lennard-Jones potential is

$$V^r = \epsilon_r \left(4 \left[\left(\frac{\sigma_{ij}^r}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}^r}{r_{ij}} \right)^6 \right] + 1 \right) \Theta(2^{1/6} \sigma_{ij}^r - r_{ij}), \quad (4)$$

where $\Theta(x)$ is the Heaviside step function that sets $V^r = 0$ for $r_{ij} \geq 2^{1/6} \sigma_{ij}^r$, $\epsilon_r / k_b T_0 = 1$, and $\sigma_{ij}^r = (\sigma_i^r + \sigma_j^r) / 2$ is the average diameter of atoms i and j . We used the atom sizes (for hydrogen, carbon, oxygen, nitrogen, and sulfur) from Ref. [14] after verifying that the backbone dihedral angles for the all-atom model sample the sterically allowed ϕ and ψ values in the Ramachandran map [15] when $V = V^{bl} + V^{ba} + V^{da} + V^r$

TABLE I. Hydrophobicity indices h_i that range from 0 (hydrophilic) to 1 (hydrophobic) for residues in α -synuclein at pH 7.4 [16].

ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE
0.735	0.37	0.295	0.41	0.76	0.41	0.54	0.5	0.29	1
LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
0.985	0.385	0.87	1	0.27	0.475	0.565	0.985	0.815	0.88

and the system behaves as an excluded volume random walk (see Appendix A).

The hydrophobic interactions between residues were modeled using the attractive Lennard-Jones potential

$$V^a = \epsilon_a \sum_{ij} \left[\lambda_{ij} \left(4 \left[\left(\frac{\sigma^a}{R_{ij}} \right)^{12} - \left(\frac{\sigma^a}{R_{ij}} \right)^6 \right] + 1 \right) \times \Theta(R_{ij} - 2^{1/6} \sigma^a) - \lambda_{ij} \right], \quad (5)$$

where ϵ_a is the attraction strength, R_{ij} is the center-to-center separation between C_α atoms on residues i and j ,

$$\lambda_{ij} = \sqrt{h_i h_j}, \quad (6)$$

h_i is the hydrophobicity index for residue i that ranges from 0 (hydrophilic) to 1 (hydrophobic) in Table I, and $\sigma^a \approx 4.8 \text{\AA}$ is the typical separation between centers of mass of neighboring residues. We find that the results for the conformational statistics for α -synuclein are not sensitive to small changes in σ^a and h_i (Appendix A).

The screened Coulomb potential was used to model the electrostatic interactions between atoms i and j for α -synuclein in water:

$$V^{es} = \epsilon_{es} \sum_{ij} \frac{q_i q_j \sigma^a}{e^2 r_{ij}} e^{-\frac{r_{ij}}{\ell}}, \quad (7)$$

where e is the fundamental charge, $\epsilon_{es} = e^2 / 4\pi\epsilon_0\epsilon\sigma^a$, ϵ_0 is the vacuum permittivity, $\epsilon = 80$ is the permittivity of water, and $\ell = 9 \text{\AA}$ is the Coulomb screening length in an aqueous solution with a 150-mM salt concentration. The partial charge

TABLE II. Partial charges q_i on atom i (left) and total charge Q_i on residue i (right) for the charged residues LYS, ARG, HIS, ASP, and GLU at pH 7.4 [17]. The total partial charge $q = \sum_i q_i$ for the N-terminal, central, and C-terminal regions are 4.1, -1.0 , and -12.0 , respectively.

Residue	Atom	Atom charge q_i	Residue charge Q_i
LYS	N_ζ	1	1
	ARG	$N_{\eta 1}$	
ARG	$N_{\eta 2}$	0.39	1
	N_ϵ	0.22	
	HIS	$N_{\delta 1}$	
HIS	$N_{\epsilon 2}$	0.05	0.1
	ASP	$O^{\delta 1}$	
ASP	$O_{\delta 2}$	-0.5	-1
	GLU	$O_{\epsilon 1}$	
GLU	$O_{\epsilon 2}$	-0.5	-1

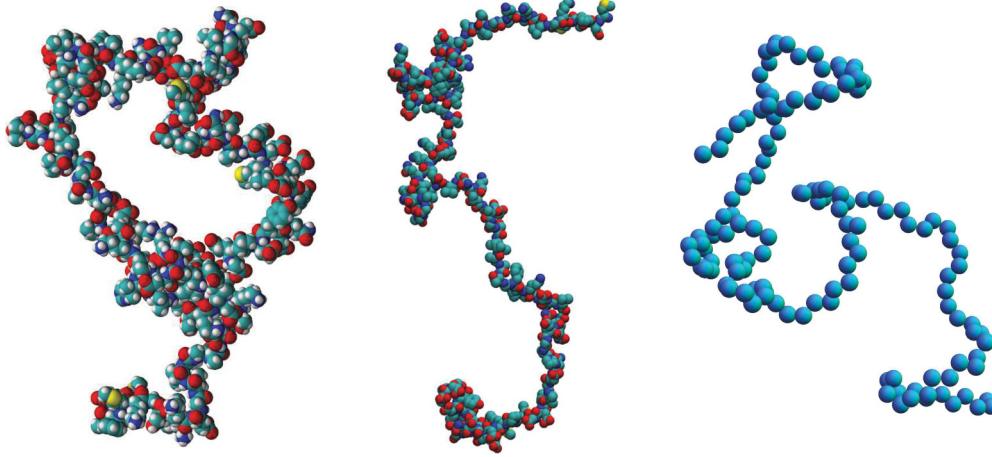


FIG. 2. (Color online) Snapshots of the (left) all-atom, (center) united-atom, and (right) coarse-grained representations of α -synuclein from Langevin dynamics simulations at temperature $T_0 = 293$ K, pH 7.4, and ratio of hydrophobic to electrostatic interactions $\alpha = 1.2$. For the all-atom and united-atom models, hydrogen, carbon, oxygen, nitrogen, and sulfur atoms are colored white (small, light), cyan (gray), red (large, dark), blue (small, dark), and yellow (large, light), respectively. For the coarse-grained model, each monomer represents an amino acid.

q_i on atom i in one of the charged residues LYS, ARG, HIS, ASP, and GLU is given in Table II.

B. United-atom model

For the united-atom model, we do not explicitly model the hydrogen atoms. Instead, we use a set of 11 atom sizes σ_i^r from Ref. [18], where the hydrogens are subsumed into the heavy atoms: C ($\sigma_i^r/2 = 1.53$ Å), CH (1.80 Å), CH₂ (1.80 Å), CH₃ (1.80 Å), O (1.26 Å), OH (1.44 Å), N (1.53 Å), NH (1.53 Å), NH₂ (1.57 Å), NH₃ (1.80 Å), and S (1.62 Å). We optimized the atom sizes by characterizing the backbone dihedral angles ϕ and ψ as a function of σ_i^r in the united-atom simulations with $V = V^{bl} + V^{ba} + V^{da} + V^r$. The ϕ and ψ backbone dihedral angle distributions closely match those from the Ramachandran map (i.e., the α -helix and β -sheet regions) when we scale the atom sizes in Ref. [18] by 0.9 as shown in Appendix A. Otherwise, the all-atom and united-atom models use the same interaction potentials in Eqs. (1)–(7).

C. Coarse-grained model

For the coarse-grained model, we employed a backbone-only C_α representation of α -synuclein, where each residue i is represented by a spherical monomer i with size σ^a , mass M ,

hydrophobicity h_i , and charge Q_i . The average bond length between monomers i and j was fixed to $\langle l_{ij} \rangle = 4.0$ Å, which is the average separation between C_α atoms on neighboring residues, using Eq. (1) (with r_{ij} replaced by R_{ij}). The bond angle Θ (between three successive C_α atoms) and dihedral angle Φ (between four successive C_α atoms) potentials were calculated so that the Θ and Φ distributions matched those from the united-atom simulations with $V = V^{bl} + V^{ba} + V^{da} + V^r$. The Θ distributions from the united-atom model were approximately Gaussian with mean $\langle \Theta \rangle = 2.13$ rad and standard deviation $\sigma_\Theta = 0.345$ rad.

The dihedral angle potential V^{da} for the coarse-grained simulations was obtained by fitting the distribution $P(\Phi)$ from the united-atom simulations to a seventh-order Fourier series

$$V^{da}(\Phi) = \sum_{k=0}^6 a_k \cos(k\Phi) + b_k \sin(k\Phi),$$

where $a_k = -2k_b T_0 \langle \cos(k\Phi) \log P(\Phi) \rangle$, $b_k = -2k_b T_0 \langle \sin(k\Phi) \log P(\Phi) \rangle$, and the angle brackets indicate an average over time and dihedral angles along the protein backbone.

For steric interactions between residues, we used the purely repulsive Lennard-Jones potential in Eq. (4), with r_{ij} and σ_{ij}^r replaced by R_{ij} and σ^a , respectively. The hydrophobic

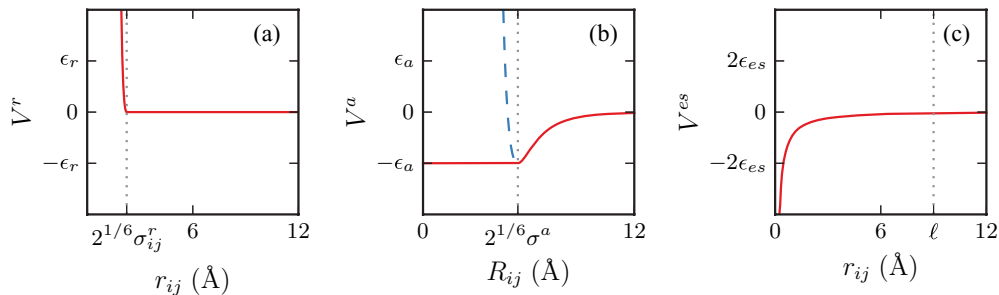


FIG. 3. (Color online) Schematics of (a) the purely repulsive Lennard-Jones potential V^r in Eq. (4) (solid line), (b) attractive Lennard-Jones potential V^a in Eq. (5) (solid line), and (c) screened Coulomb potential V^{es} in Eq. (7) (solid line). The dashed line in (b) represents the repulsive Lennard-Jones interaction between residues i and j in the coarse-grained model.

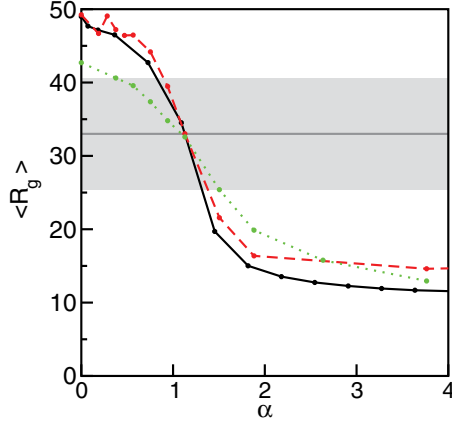


FIG. 4. (Color online) Average radius of gyration $\langle R_g \rangle$ versus the ratio of the attractive hydrophobic to electrostatic interactions α for the coarse-grained (black solid), united-atom (red dashed), and all-atom (green dotted) models at T_0 (or the temperature that gives $R_g \approx 33 \text{ \AA}$ in the coarse-grained simulations) and $pH 7.4$. The horizontal line and gray shaded region indicate the average and standard deviation over recent NMR, SAXS, and smFRET experimental measurements, $\langle R_g \rangle = 33.0 \pm 7.7 \text{ \AA}$, for monomeric α -synuclein near T_0 and neutral pH [4,7,9,12,21–24].

interactions are the same as those in Eqs. (5) and (6), with $\epsilon_a = \epsilon_r$. The electrostatic interactions between residues are given by Eq. (7), with q_i and r_{ij} replaced by Q_i and R_{ij} , respectively.

D. Langevin dynamics

The all-atom, united-atom, and coarse-grained models were simulated at fixed number of particles, volume, and

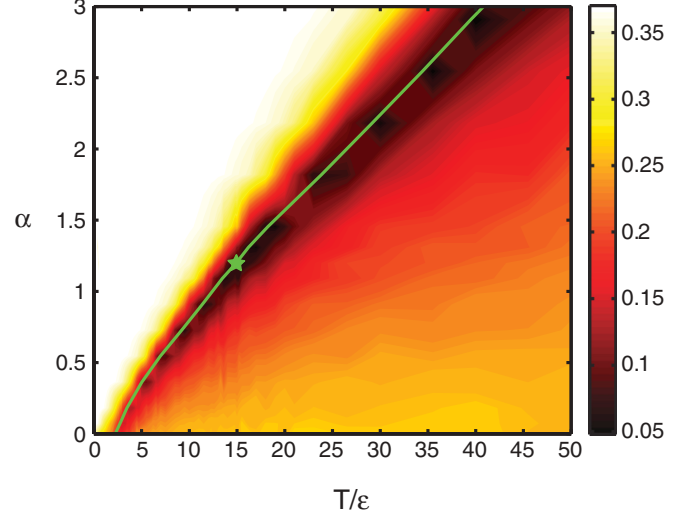


FIG. 6. (Color online) rms deviations Δ between the coarse-grained and experimental FRET efficiencies for the 12 residue pairs considered in Ref. [12] as a function of α and $k_b T / \epsilon_r$. The solid line indicates systems that give $\langle R_g \rangle \approx 33 \text{ \AA}$. Note that this line coincides with the minimum values for the rms deviations.

temperature (NVT) using a Langevin thermostat [19], modified velocity Verlet integration scheme, and free boundary conditions. We set the time step $\Delta t = 10^{-2} t_0$ and damping coefficient $\gamma = 10^{-3} t_0^{-1}$, where $t_0 = \sqrt{m \langle \sigma_{ij}^2 \rangle} / \epsilon_r$ and m is the hydrogen mass for the all-atom and united-atom models and $t_0 = \sqrt{M \sigma^a} / \epsilon_r$ for the coarse-grained model. The initial atomic positions were obtained from a micelle-bound NMR structure (protein data bank identifier 1XQ8) for α -synuclein

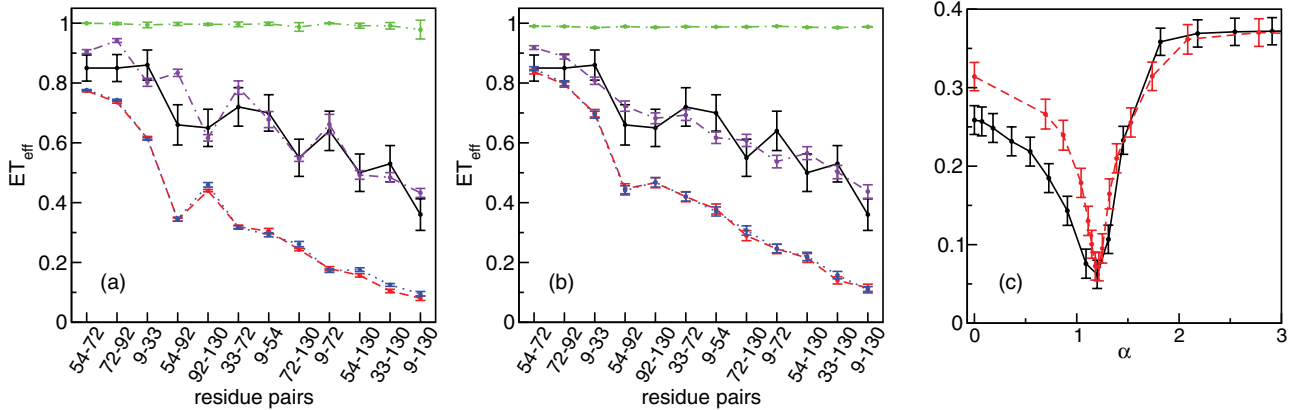


FIG. 5. (Color online) A comparison of FRET efficiencies ET_{eff} for 12 residue pairs from simulations and experiments of α -synuclein. In (a) the data includes FRET efficiencies from united-atom simulations of a random walk (red dashed), collapsed globule (green dot-dot-dashed), only electrostatic interactions at temperature T_0 (blue dotted), and ratio of attractive hydrophobic to electrostatic interactions $\alpha = 1.2$ (purple dot-dashed) at T_0 and recent smFRET experiments [12] (black solid). The error bars for the united-atom simulations were obtained from the standard deviation over eight independent simulations and are comparable to the size of the symbols. Error bars from experiments were calculated using a resampling method that accounts for uncertainty in the determination of ET_{eff} (1%–2%) and variations in R_0 (7%–8%) due to the effects of the protein environment on the smFRET fluorophores. In (b) we compare the FRET efficiencies from recent smFRET experiments [12] to the coarse-grained simulations of a random walk (red dashed), collapsed globule (green dot-dot-dashed), only electrostatics interactions (blue dotted), and both attractive hydrophobic and electrostatic interactions with $\alpha = 1.2$ at a temperature that yields $\langle R_g \rangle \approx 33 \text{ \AA}$ (purple dot-dashed). Error bars for the coarse-grained simulations were obtained from the standard deviation over single runs for times much longer than the decay of the radius of gyration autocorrelation function. (c) The rms deviation Δ between the FRET efficiencies from the united-atom simulations and smFRET experiments (red dashed) and the coarse-grained simulations and smFRET experiments (black solid) versus α . The minimum rms $\Delta_{\text{min}} \approx 0.07$ occurs near $\alpha = 1.2$ for both the united-atom and coarse-grained simulations.

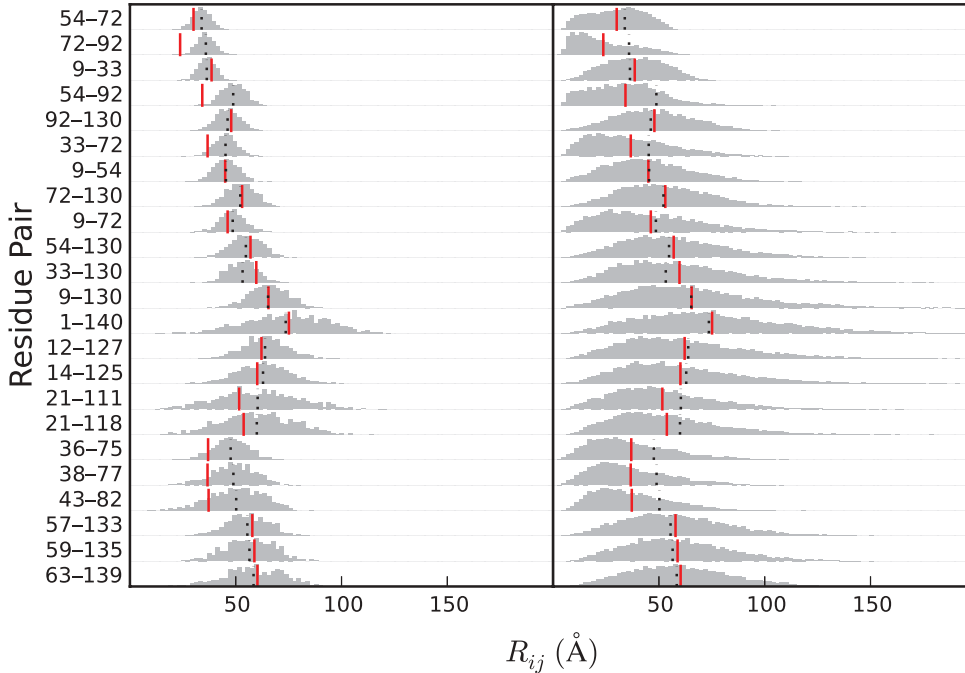


FIG. 7. (Color online) Probability distributions for the inter-residue separations $P(R_{ij})$ for the 12 residue pairs considered in Ref. [12] and 11 additional pairs for experimentally constrained Monte Carlo (ECMC) [12] (left) and united-atom (with $\alpha = 1.2$; right) simulations. The average inter-residue separations $\langle R_{ij} \rangle$ for the united-atom and ECMC simulations are shown with solid and dashed lines, respectively.

at pH 7.4 and temperature 298 K [20]. The initial positions for the coarse-grained model were obtained from simulations at high temperature with only bond-length, bond-angle, and dihedral-angle potentials and repulsive Lennard-Jones interactions. The simulations were run for times much longer than the characteristic relaxation times from the decay of the radius of gyration autocorrelation function.

In the results below, we will study the radius of gyration R_g and distribution of inter-residue separations $P(R_{ij})$ as a function of the ratio of the attractive hydrophobic and electrostatic energy scales $\alpha = \epsilon_a/\epsilon_{es}$ and quantitatively compare the results from smFRET experiments and all-atom, united-atom, and coarse-grained simulations.

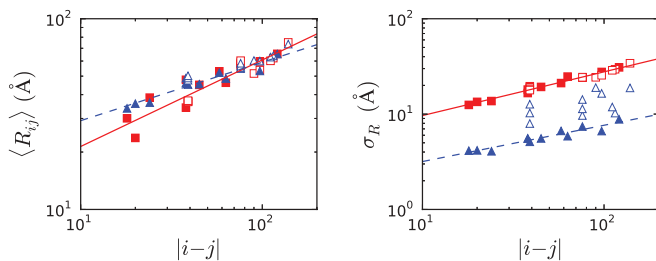


FIG. 8. (Color online) Average $\langle R_{ij} \rangle$ and standard deviation σ_R of the inter-residue separation distributions in Fig. 7 for the united-atom (squares) and ECMC (triangles) simulations versus chemical distance between residues $|i - j|$. The filled symbols indicate residue pairs that were considered in smFRET experiments [12] and open symbols indicate other pairs. The solid and dashed lines have slopes of 0.45 and 0.31 (left panel) and 0.47 and 0.38 (right panel), respectively.

III. RESULTS

In Fig. 4, we show the radius of gyration that characterizes the overall protein shape for the all-atom, united-atom, and coarse-grained models,

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i - \langle \vec{r}_i \rangle)^2}, \quad (8)$$

where \vec{r}_i is the position of atom or monomer i , as a function of the ratio of the attractive hydrophobic to electrostatic interactions α at temperature T_0 and pH 7.4. For $\alpha \gg 1$, the protein forms a collapsed globule with $\langle R_g \rangle \approx 12\text{--}15 \text{ \AA}$, whereas for $\alpha \ll 1$, the models only include electrostatics interactions, and $\langle R_g \rangle$ is similar to the random walk values for the three models (all atom: 42.8 \AA , united atom: 48.6 \AA , coarse grained: 48.2 \AA). The crossover between random walk and collapsed globule behavior for $\langle R_g \rangle$ occurs near $\alpha \approx 1$.

A number of recent small-angle x-ray scattering (SAXS), NMR, and smFRET experiments have measured the radius of gyration for monomeric α -synuclein near T_0 and neutral pH [4,7,9,12,21–24]. As shown in Fig. 4, the average over these experimental measurements is $\langle R_g \rangle = 33.0 \pm 7.7 \text{ \AA}$, and thus the $\langle R_g \rangle$ for α -synuclein falls in between the random walk and collapsed globule values.

We can more quantitatively compare simulation and experimental studies of α -synuclein by calculating the distributions of inter-residue distances or, equivalently, the FRET efficiencies. FRET efficiencies between residues i and

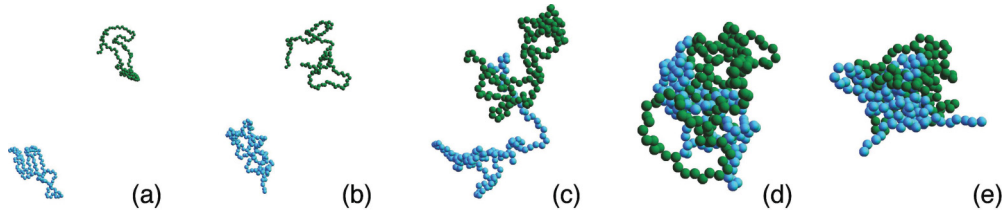


FIG. 9. (Color online) Snapshots from preliminary aggregation studies of two monomeric α -synuclein proteins (dark green and light blue) using coarse-grained simulations with the temperature set so that $\langle R_g \rangle \approx 33 \text{ \AA}$ at $\alpha = 1.2$ (for individual protein monomers) for (a) $\alpha = 0.7$, (b) 1.1, (c) 1.3, (d) 1.5, and (e) 1.8.

j are obtained from

$$ET_{\text{eff}} = \left\langle \frac{1}{1 + \left(\frac{R_{ij}}{R_0}\right)^6} \right\rangle, \quad (9)$$

where $R_0 = 54 \text{ \AA}$ is the Förster distance for the fluorophore pair in Refs. [11,25] and the angle brackets indicate an average over time. To calculate $\langle R_{ij} \rangle$ from the FRET efficiencies, one must invert Eq. (9) using the distribution of inter-residue separations $P(R_{ij})$.

The FRET efficiencies for the 12 residue pairs from recent smFRET experiments on α -synuclein [11] and the united-atom and coarse-grained simulations are shown in Figs. 5(a) and 5(b). Errors in the inter-residue separation distributions from experiments can occur in both the directly measured ET_{eff} values and R_0 . To estimate the errors, we generated ten decoy sets of inter-residue separations using ET_{eff} and R_0 values drawn from distributions accounting for the individual uncertainties. We then calculated the rms deviation over each decoy set assuming that we know R_0 precisely. Errors in ET_{eff} from the united-atom simulations are obtained by calculating the standard deviation over eight independent runs. Errors from

the coarse-grained simulations are obtained from the standard deviation over single runs for times much longer than the decay of the radius of gyration autocorrelation function.

We identify several important features in the comparison of the FRET efficiencies from experiments and simulations in Figs. 5(a) and 5(b): (1) the united-atom and coarse-grained models yield qualitatively similar results for the FRET efficiencies; (2) the FRET efficiencies for the random walk and pure electrostatics models are similar to each other and much lower than most of the residue pair FRET efficiencies from experiments; (3) the FRET efficiencies for the collapsed globule ≈ 1 and do not match those from experiments; and (4) by tuning α , we are able to match quantitatively the FRET efficiencies from the experiments and simulations.

As shown in Fig. 5(c), the rms deviations Δ between the FRET efficiencies from the united-atom simulations and smFRET experiments and between the FRET efficiencies from the coarse-grained simulations and smFRET experiments are minimized when $\alpha \approx 1.2$. For the united-atom model, $\alpha \approx 1.2$ gives $\langle R_g \rangle \approx 31.8 \pm 0.7 \text{ \AA}$, which is similar to that found in Ref. [12]. The largest deviations in the FRET efficiencies between the united-atom simulations and smFRET experiments occur for small inter-residue separations, which are likely caused by the finite size of the dye molecules. Note that the deviations at small inter-residue separations are much weaker for the coarse-grained simulations. Thus, we find that it is crucial to include both electrostatic and attractive hydrophobic interactions in modeling α -synuclein in solution.

For the coarse-grained simulations, we also studied the variation of the FRET efficiencies as a function of temperature (not only at $T = T_0$). In Fig. 6, we show the rms deviation between the FRET efficiencies for the coarse-grained simulations and smFRET experiments for the 12 residue pairs considered in Ref. [12] as a function of α and $k_b T / \epsilon_r$. We find that the line of α and $k_b T / \epsilon_r$ values that give $\langle R_g \rangle \approx 33 \text{ \AA}$ lies in the region where the rms deviations in the FRET efficiencies are minimized, which indicates that there is a class of polymeric structures with similar conformational statistics to that of α -synuclein.

In Fig. 7, we compare the inter-residue separation distributions $P(R_{ij})$ obtained from experimentally constrained Monte Carlo (ECMC) and united-atom (with $\alpha = 1.2$) simulations. For the ECMC simulations discussed in detail in Ref. [12], we assumed that $P(R_{ij})$ was similar to that for a random walk C_α model with only bond-length, bond-angle, and dihedral-angle (ω) constraints and repulsive Lennard-Jones interactions to obtain $\langle R_{ij} \rangle$ from the experimentally measured FRET efficiencies. We find that $\langle R_{ij} \rangle$ for the ECMC and

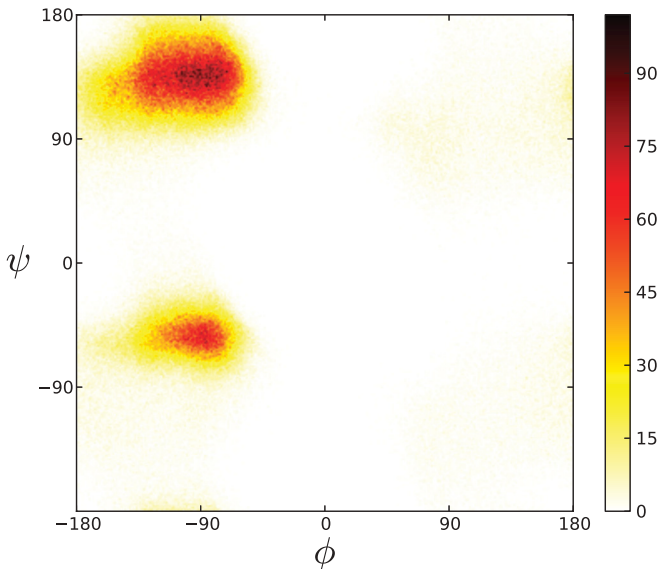


FIG. 10. (Color online) Ramachandran plot for the backbone dihedral angles ϕ and ψ obtained from the all-atom random walk simulations with no attractive hydrophobic and electrostatic interactions, and atom sizes given in Ref. [14]. The highly populated ϕ and ψ angles indicate β -sheet (upper left) and α -helix (lower left) conformations.

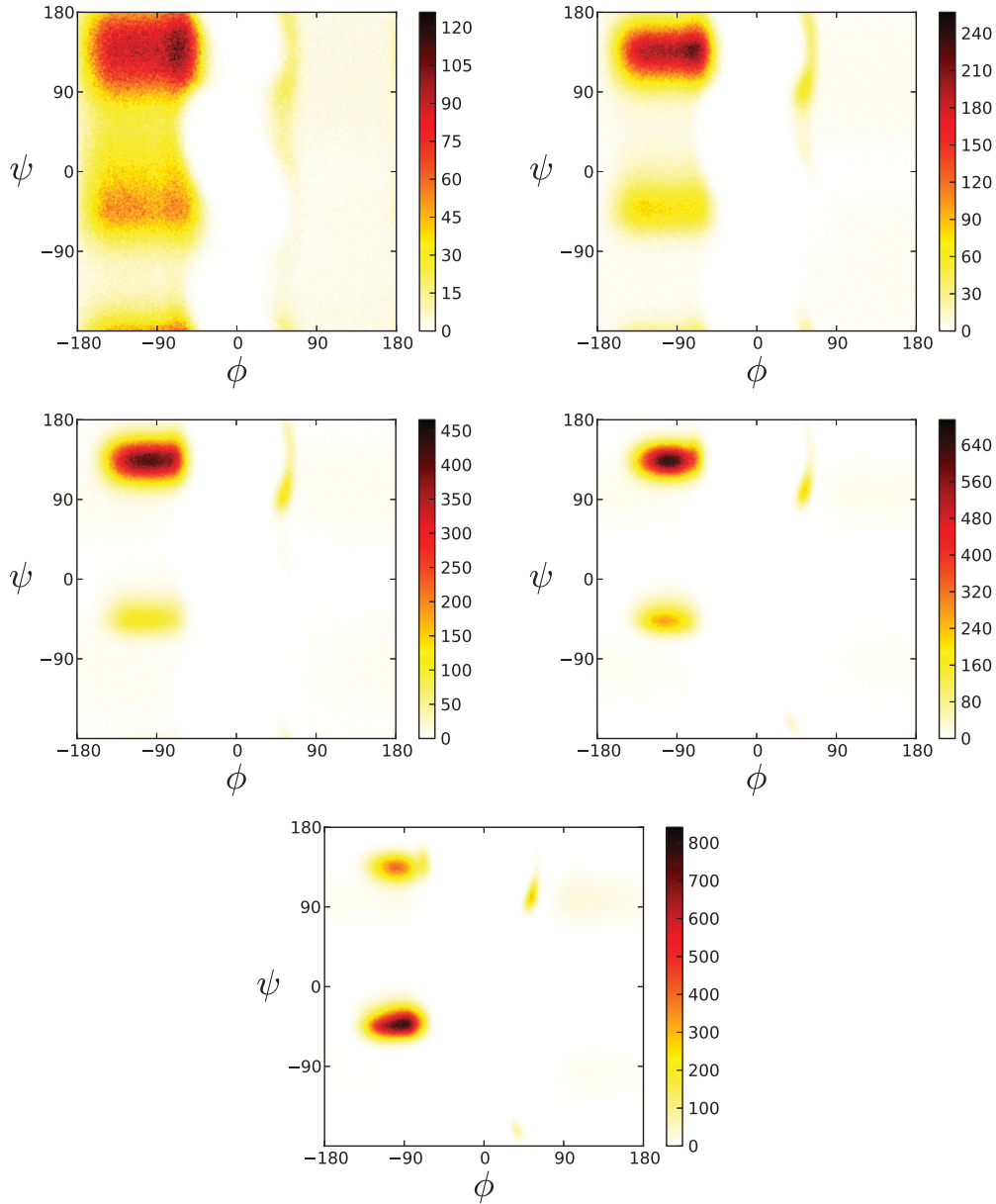


FIG. 11. (Color online) Ramachandran plots for the backbone dihedral angles ϕ and ψ obtained from the united-atom random walk simulations with no attractive hydrophobic and electrostatic interactions and atom sizes 0.8 (upper left), 0.85 (upper right), 0.9 (middle left), 0.95 (middle right), and 1.0 (bottom) times those given in Ref. [18].

united-atom simulations agree to within roughly 10% [Fig. 8 (left)], however, the standard deviations differ significantly, as shown in Fig. 8 (right). The standard deviation of $P(R_{ij})$ for the united-atom simulations is larger than that for the ECMC simulations for all residue pairs and scales as $\sigma_R \sim |i - j|^\delta$ with $\delta \sim 0.47$ (compared to the excluded volume random walk scaling exponent $\delta = 0.59$). Further, σ_R for residue pairs that are not constrained in ECMC do not obey the scaling behavior with $i - j$, as found for residue pairs that were constrained ($\sigma_R \sim |i - j|^\delta$ with $\delta \sim 0.4$ [12]).

IV. CONCLUSIONS AND FUTURE DIRECTIONS

We have shown that we are able to accurately model the conformational dynamics (i.e., the inter-residue separations) of the IDP α -synuclein at temperature $T_0 = 293$ K and neutral

pH using all-atom, united-atom, and coarse-grained Langevin dynamics simulations that include only geometric constraints and hydrophobic and electrostatic interactions. Our results show that the structure of α -synuclein is intermediate between that for random walks and collapsed globules. The $\langle R_g \rangle$ is intermediate between the high-temperature and collapsed values, and the rms separation σ_R between residues i and j scales as $|i - j|^\delta$ with $\delta \sim 0.47$, compared to $\delta \sim 0.59$ for the self-avoiding random walk. The *calibrated Langevin dynamics simulations* presented here have the advantage over constraint methods in that physical forces act on all residues, not only on residue pairs that are monitored experimentally, and can be tuned to match FRET efficiencies from experiments. In future work, we will employ calibrated Langevin dynamics simulations to study the conformational dynamics of α -synuclein at low pH and the interaction and association between two or

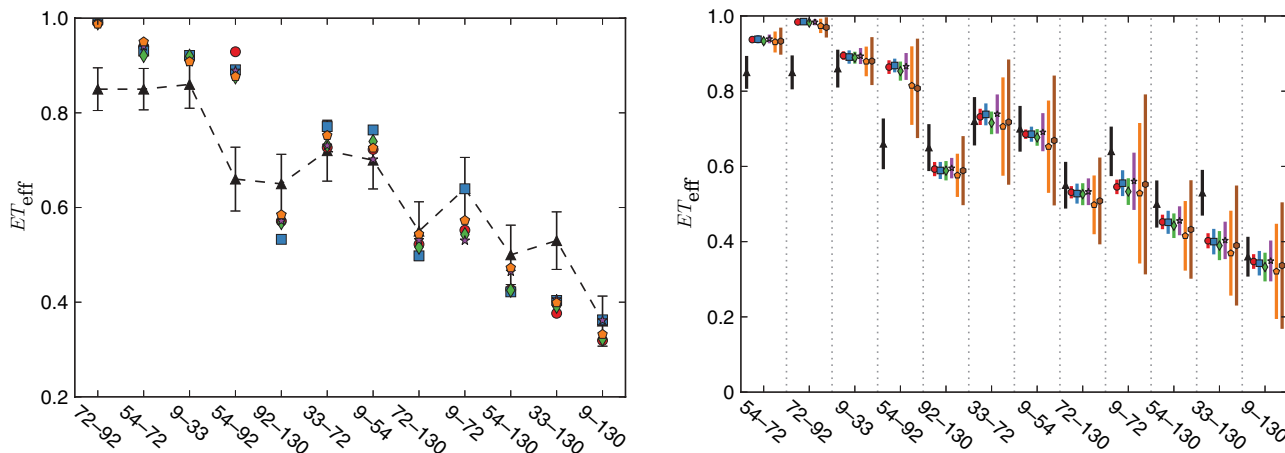


FIG. 12. (Color online) (Left) FRET efficiencies ET_{eff} for the 12 residue pairs considered in Ref. [12] from smFRET experiments (upward triangles) and united-atom simulations with α set so that $R_g \approx 33$ Å and $\sigma^a = 4.4$ Å (circles), 4.6 Å (squares), 4.8 Å (diamonds), 5.0 Å (stars), and 5.2 Å (pentagons). (Right) FRET efficiencies ET_{eff} for the 12 residue pairs considered in Ref. [12] for the united-atom simulations for $\alpha = 1.2$ with $\sigma^a = 5.2$ Å and varying hydrophobicity indices $h'_i = h_i + \Delta h$, where Δh is chosen from a zero-mean Gaussian distribution with standard deviation 0.0 (circles), 0.02 (squares), 0.05 (diamonds), 0.1 (stars), 0.3 (pentagons), and 0.5 (hexagons). The average ET_{eff} and its standard deviation for 32 samples are shown for each Δh .

more α -synuclein monomers as a function of pH to identify mechanisms for α -synuclein oligomerization. In preliminary calibrated coarse-grained Langevin dynamics simulations, we find that two monomeric α -synuclein proteins only associate for sufficiently strong attractive hydrophobic interactions ($\alpha \geq 1.2$), as shown in Fig. 9. Note that the quantitative agreement between ET_{eff} from smFRET experiments and simulations was obtained without including hydrogen bonding interactions. It is possible that hydrogen bonding interactions will be necessary to model τ , Intracisternal A particle-promoted polypeptide (IPP), and other intrinsically disordered protein monomers, as well as aggregation of multiple proteins.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under Grants No. DMR-1006537 (C.O., C.S.), No. PHY-1019147 (W.S.), and No. BIO-0919853 (E.R.), and the Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences (C.O., E.R.). This work also benefited from the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and NSF Grant No. CNS-0821132, which partially funded acquisition of the computational facilities.

APPENDIX A: CALIBRATION OF ATOM SIZES

In this Appendix, we test the choice of the atom sizes used in the all-atom and united-atom models by measuring the

Ramachandran plot [15] for the backbone dihedral angles ϕ and ψ . In Fig. 10, we show that the Ramachandran plot for the random walk all-atom model of α -synuclein with no attractive hydrophobic and electrostatic interactions and atom sizes from Ref. [14] closely resembles that for dipeptides with a highly populated α -helix and β -sheet regions, even without modeling backbone hydrogen bonding interactions. In Fig. 11, we show the Ramachandran plots for the backbone dihedral angles ϕ and ψ obtained from the random walk united-atom model of α -synuclein with no attractive hydrophobic and electrostatic interactions and atom sizes 0.8, 0.85, 0.9, 0.95, and 1.0 times those from Ref. [18]. We find that the Ramachandran plot for united-atom model with a factor of 0.9 for the atom sizes is similar to that for the all-atom model.

APPENDIX B: ROBUSTNESS OF THE HYDROPHOBIC INTERACTIONS

In this Appendix, we study the sensitivity of the FRET efficiencies for the united-atom simulations to small variations in the length scale σ^a , above which the attractive hydrophobic interactions are nonzero, and relative strengths h_i of the attractive hydrophobic interactions for different residues. In Fig. 12 (left), we show that the FRET efficiencies for the 12 residue pairs show only small variations with σ^a over the range from 4.4 Å to 5.2 Å. In Fig. 12 (right), we show that the FRET efficiencies for the 12 residue pairs are robust for $\Delta h < 0.5$.

- [1] S. Vucetic, C. J. Brown, A. K. Dunker, and Z. Obradovic, *Proteins: Struct., Funct., Bioinf.* **52**, 573 (2003).
 [2] K. Sugase, H. J. Dyson, and P. E. Wright, *Nature (London)* **447**, 1021 (2007).
 [3] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Annu. Rev. Biophys.* **37**, 215 (2008).

- [4] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, *J. Am. Chem. Soc.* **127**, 476 (2004).
 [5] M. Vilar, H.-T. Chou, T. Lührs, S. K. Maji, D. Riek-Loher, R. Verel, G. Manning, H. Stahlberg, and R. Riek, *Proc. Nat. Acad. Sci. USA* **105**, 8637 (2008).

- [6] D. Eliezer, E. Kutluay, R. Bussell, Jr., and G. Browne, *J. Mol. Biol.* **307**, 1061 (2001).
- [7] J. Li, V. N. Uversky, and A. L. Fink, *NeuroToxicology* **23**, 553 (2002).
- [8] I. F. Tsigelny, P. Bar-On, Y. Sharikov, L. Crews, M. Hashimoto, M. A. Miller, S. H. Keller, O. Platoshyn, J. X. J. Yuan, and E. Masliah, *FEBS J.* **274**, 1862 (2007).
- [9] V. N. Uversky, G. Yamin, L. A. Munishkina, M. A. Karymov, I. S. Millett, S. Doniach, Y. L. Lyubchenko, and A. L. Fink, *Mol. Brain Res.* **134**, 84 (2005).
- [10] O. Ullman, C. K. Fisher, and C. M. Stultz, *J. Am. Chem. Soc.* **133**, 19536 (2011).
- [11] A. J. Trexler and E. Rhoades, *Biophys. J.* **99**, 3048 (2010).
- [12] A. Nath, M. Sammalkorpi, D. C. De Witt, A. J. Trexler, S. Elbaum-Garfinkle, C. S. O'Hern, and E. Rhoades (unpublished).
- [13] R. L. Dunbrack, Jr., and F. E. Cohen, *Protein Sci.* **6**, 1661 (1997).
- [14] A. Q. Zhou, C. S. O'Hern, and L. Regan, *Biophys. J.* **102**, 2345 (2012).
- [15] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.* **7**, 95 (1963).
- [16] O. D. Monera, T. J. Sereda, N. E. Zhou, C. M. Kay, and R. S. Hodges, *J. Peptide Sci.* **1**, 319 (1995).
- [17] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, *J. Comput. Chem.* **25**, 1656 (2004).
- [18] F. Richards, *J. Mol. Biol.* **82**, 1 (1974).
- [19] D. L. Ermak and H. Buckholz, *J. Comput. Phys.* **35**, 169 (1980).
- [20] T. S. Ulmer, A. Bax, N. B. Cole, and R. L. Nussbaum, *J. Biol. Chem.* **280**, 9595 (2005).
- [21] V. N. Uversky, J. Li, and A. L. Fink, *FEBS Lett.* **509**, 31 (2001).
- [22] M. Tashiro, M. Kojima, H. Kihara, K. Kasai, T. Kamiyoshihara, K. Uéda, and S. Shimotakahara, *Biochem. Biophys. Res. Commun.* **369**, 910 (2008).
- [23] A. Rekas, R. Knott, A. Sokolova, K. Barnham, K. Perez, C. Masters, S. Drew, R. Cappai, C. Curtain, and C. Pham, *Eur. Biophys. J.* **39**, 1407 (2010).
- [24] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, and M. Blackledge, *J. Am. Chem. Soc.* **132**, 8407 (2010).
- [25] B. Schuler, E. A. Lipman, and W. A. Eaton, *Nature (London)* **419**, 743 (2002).