Molecular simulations of intrinsically disordered proteins

Elizabeth Rhoades,
Chemistry, UPenn

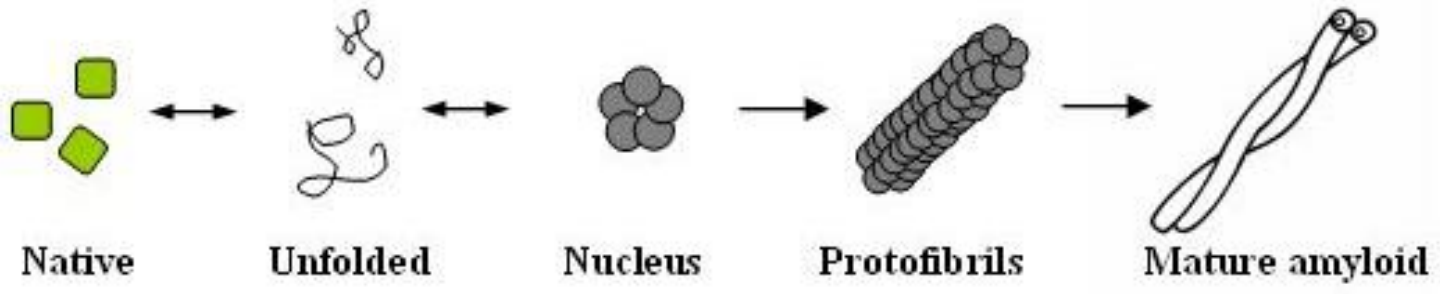
Wendell Smith
Physics

W. W. Smith, P.-Y. Ho, and CSO, "Calibrated Langevin-dynamics simulations of intrinsically disordered proteins," Phys. Rev. E 90 (2014) 042709.

The Raymond and Beverly Sackler Institute
for Biological, Physical and Engineering Sciences
Yale University

NIH NSF

Protein Aggregation



- Self-assembled structures that grow into large, insoluble aggregates
- Aggregation more likely for unfolded/partially folded protein; exposed hydrophobic regions on separate monomers bind
- Aggregation caused by protein overproduction, stress, mutation
- Variables that affect aggregation: amino acid sequence, environmental factors such as pH, temperature, protein concentration, chaperones
- Types of aggregates: amyloid fibrils (*in vivo*, *in vitro*, rich in beta-sheet, ordered, 10-nm diameter, origin-amylose), inclusion bodies (*in vivo* disordered aggregates, 1μm), disordered aggregates (*in vitro*)
- Amyloid fibril formation ubiquitous in polypeptides in nonnative conditions

Amyloid Fibrils

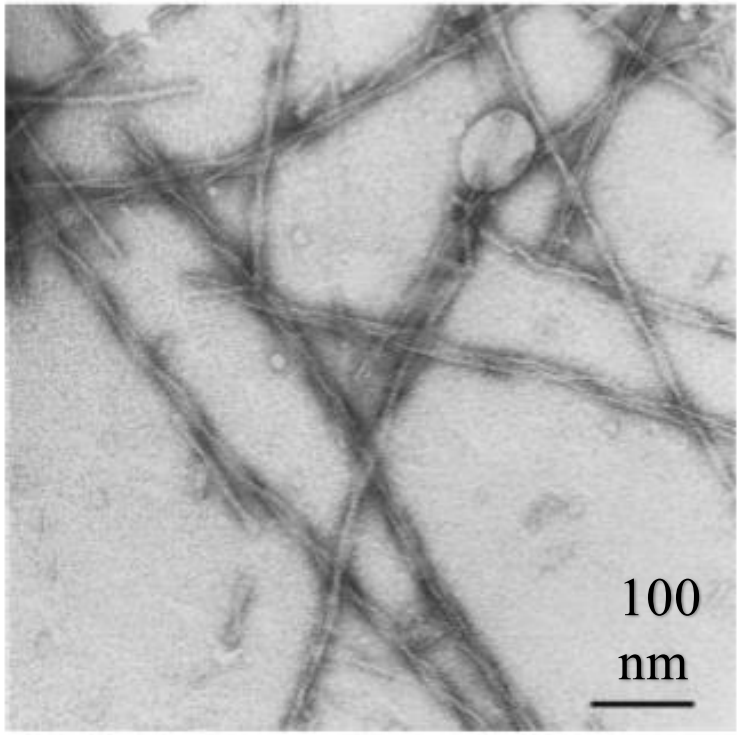


Figure 2

Electron micrograph of fibrils formed from an SH3 domain by incubation of a solution containing the protein at low pH (Ref. 19). Under these solution conditions the protein is partially unfolded and slowly aggregates to form a gel that contains the fibrils. Fibrils associated with the various amyloid diseases have a highly similar appearance to these fibrils formed under laboratory conditions[10]. The scale bar is 100 nm. Reproduced, with permission, from Ref. 19.

SH3 domain

Ribbon diagram of the SH3 domain, alpha spectrin, from chicken (PDB accession code 1SHG), colored from blue (N-terminus) to red (C-terminus).

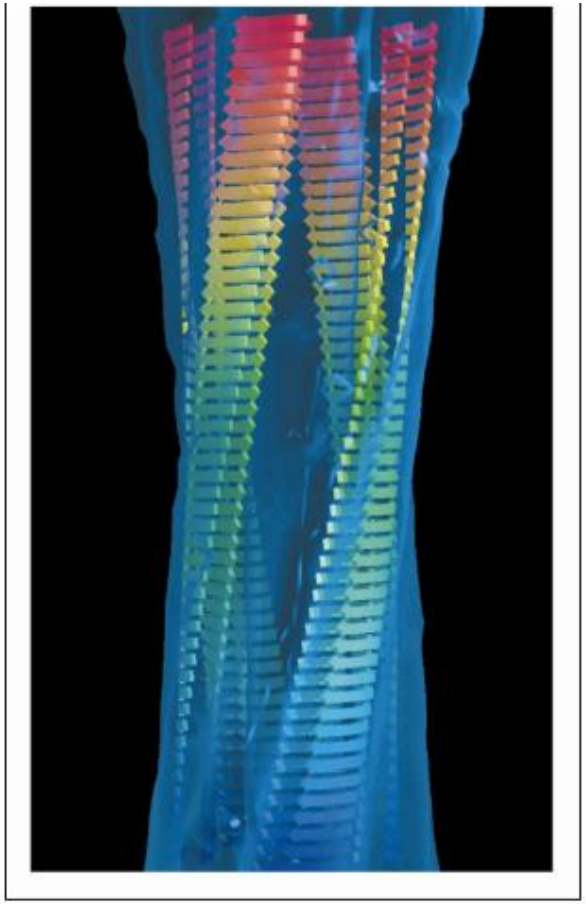
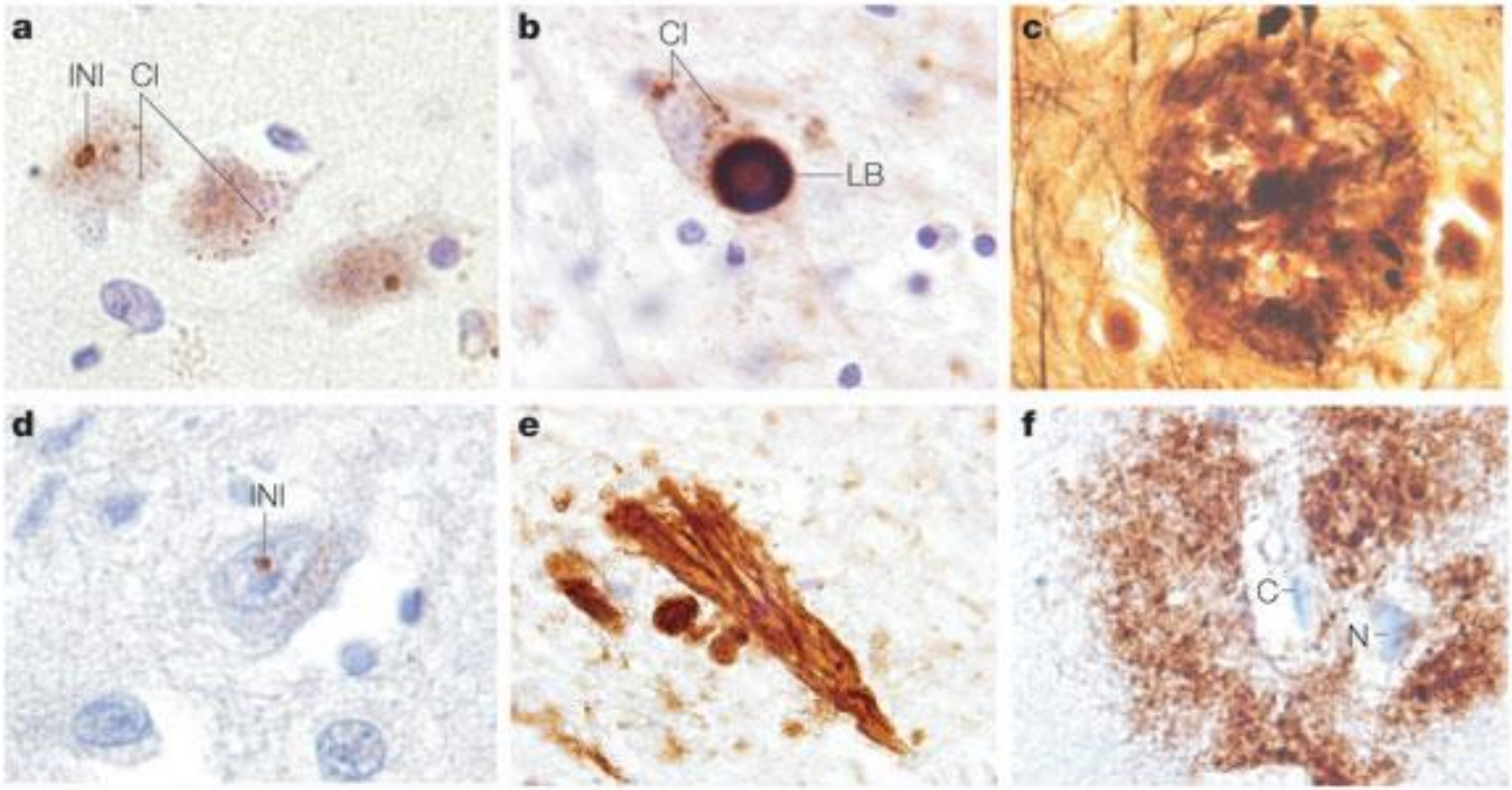


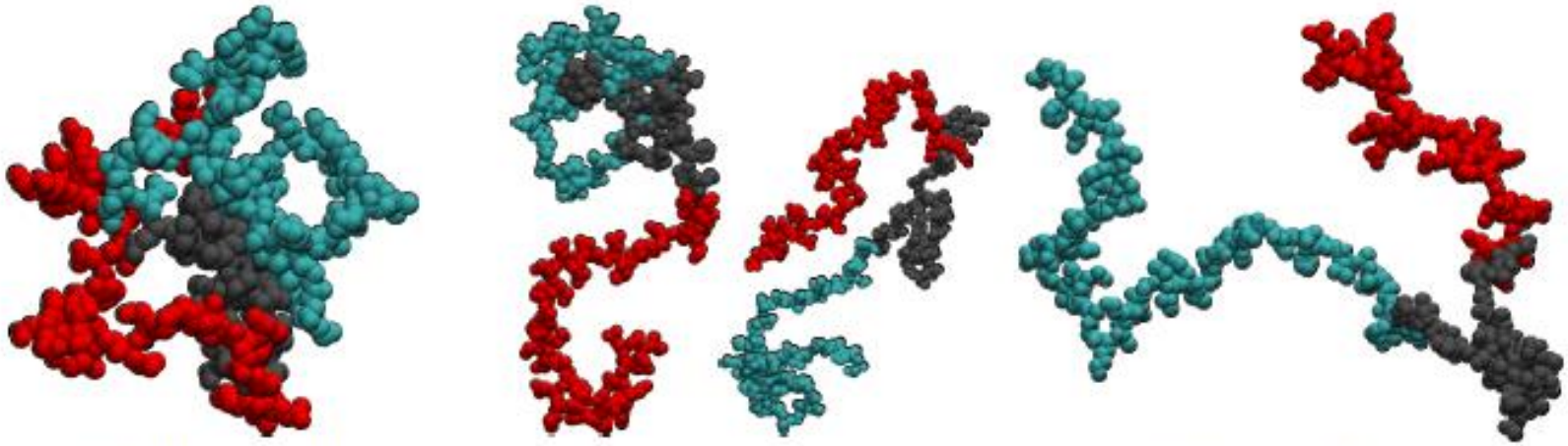
Figure 3

Molecular model of an amyloid fibril derived from cryo-EM analysis of fibrils grown from an SH3 domain. The fibril consists of four 'protofilaments' that twist around one another to form a hollow tube with a diameter of approximately 60 Å (Ref. 23). The model shown here represents one way in which regions of the polypeptide chain involved in β-sheet structure could be assembled within the fibrils.

Plaques and Inclusions in Diseased Brains



(a) Intranuclear (INI) and cytoplasmic inclusions (CI) in motor cortex of Huntington's disease brain recognized with 1C2 antibody. (b) Lewy body (LB) and other cytoplasmic inclusions (CI) that contain alpha-synuclein within a neuron of the substantia nigra of Parkinson's disease brain. (c) Neuritic plaque of Alzheimer's disease in cerebral cortex. Hirano silver stain identifies intracellular and extracellular protein aggregates. (d) Intranuclear inclusion in frontal cortex of Huntington's disease brain recognized with anti-ubiquitin antibody. (e) Neurofibrillary tangles of Alzheimer's disease in hippocampus immunostained with antibody specific for phosphorylated tau. (f) Diffuse plaque of Alzheimer's disease in cerebral cortex. Amyloid beta (Abeta)-specific antibody recognizes extracellular deposits of Abeta (surrounding a neuron (N) and a capillary (C)).


Collapsed
Extended

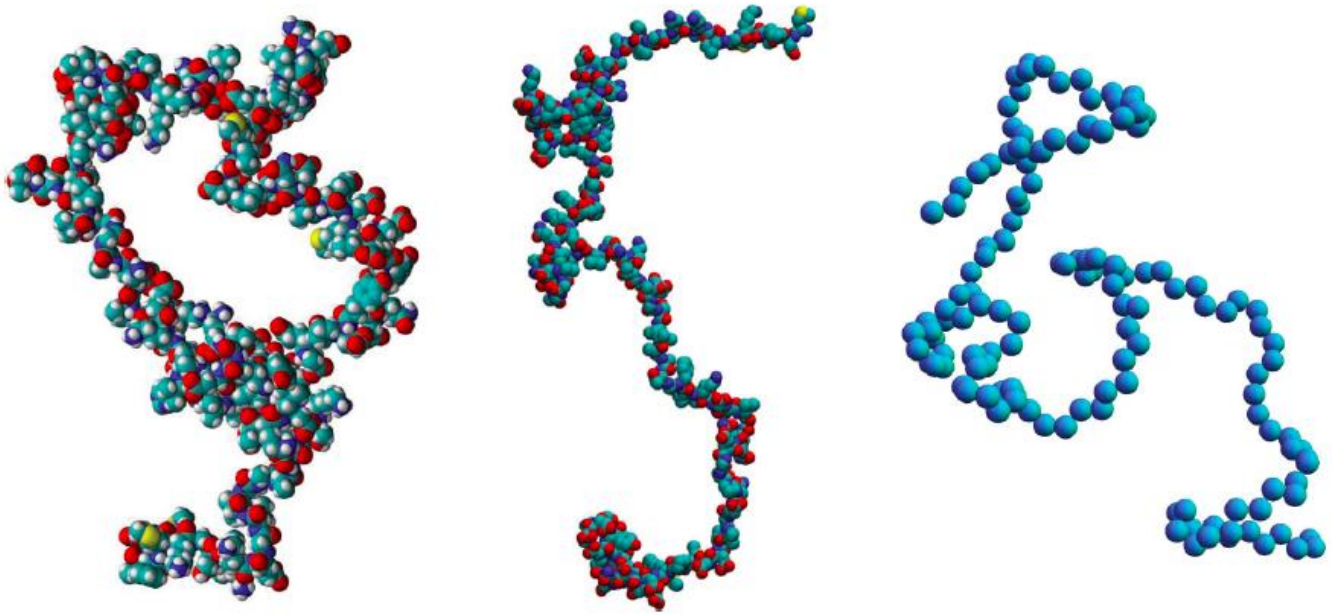
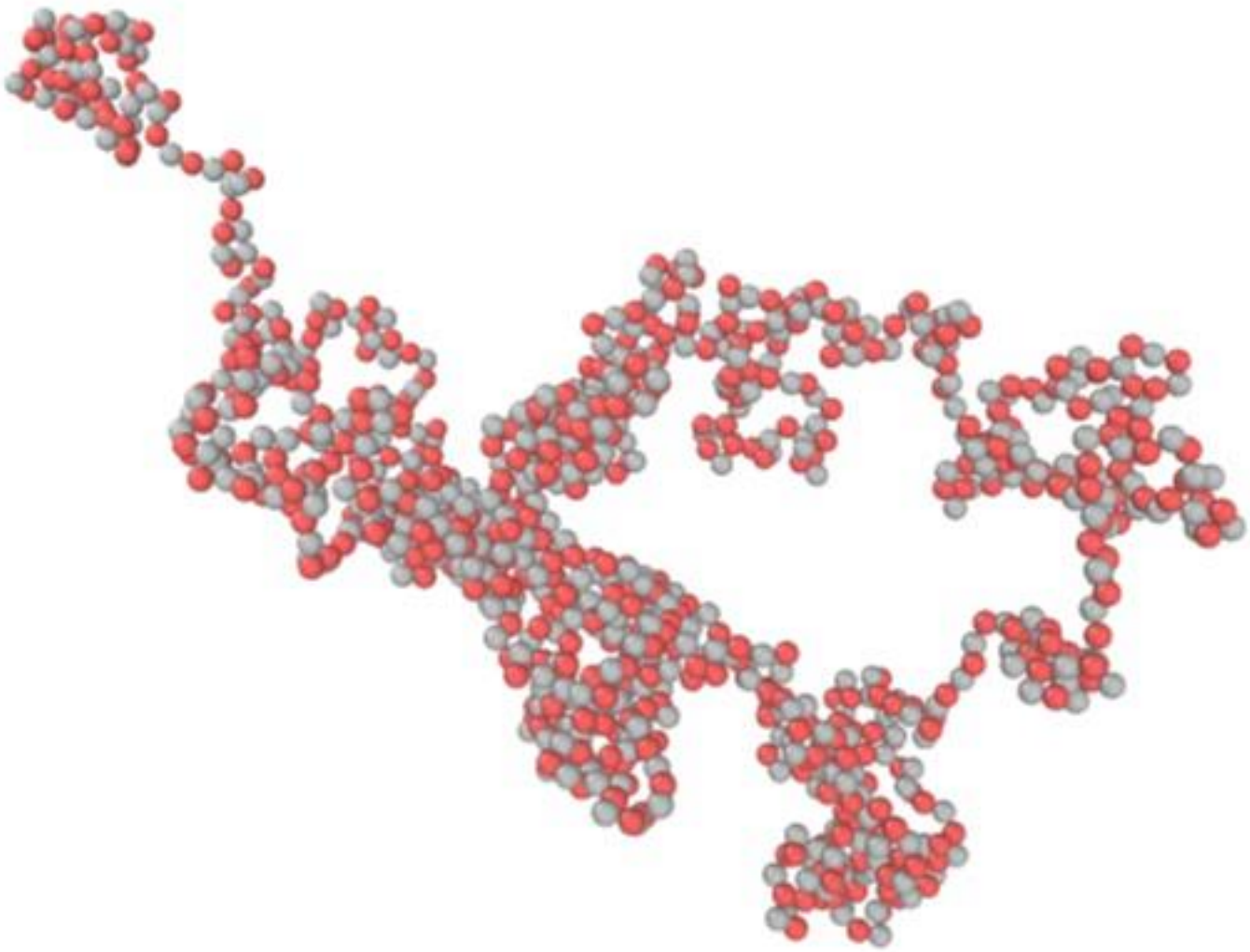
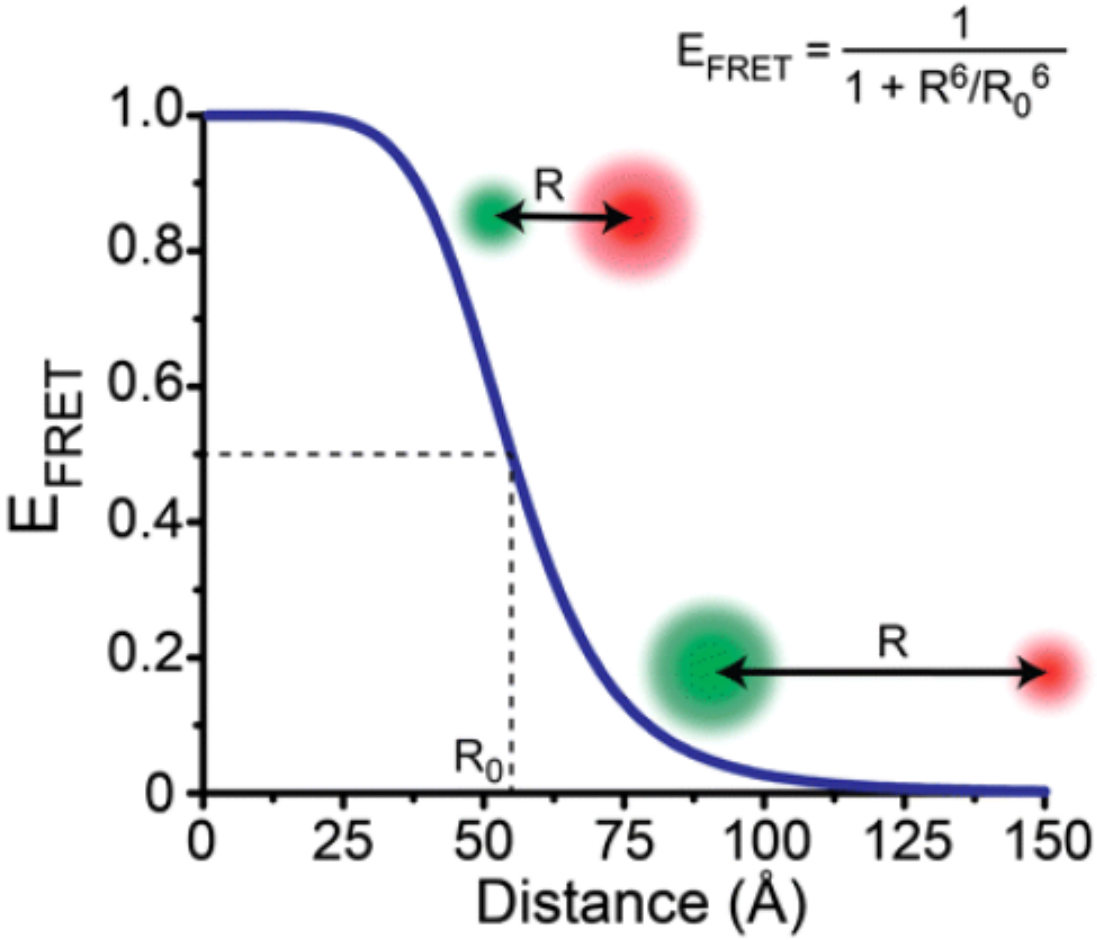


FIG. 2. (Color online) Snapshots of the (left) all-atom, (center) united-atom, and (right) coarse-grained representations of α-synuclein from Langevin dynamics simulations at temperature $T_0 = 293$ K, pH 7.4, and ratio of hydrophobic to electrostatic interactions $\alpha = 1.2$. For the all-atom and 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.



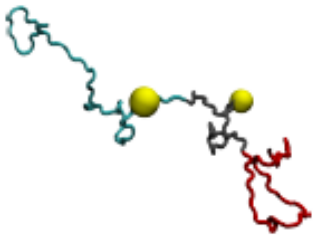
smFRET

Single-Molecule Förster Resonance Energy Transfer

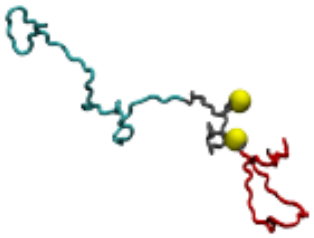


Tinoco I, Gonzalez R L Genes Dev. 2011; 25:1205-1231

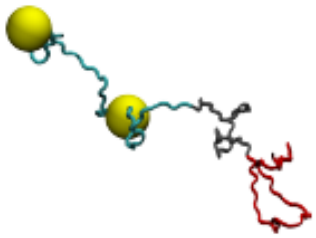
smFRET Pairs for α-Synuclein



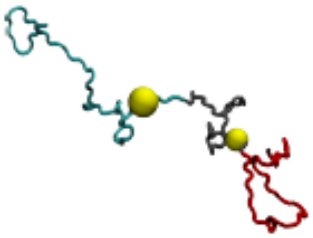
54–72



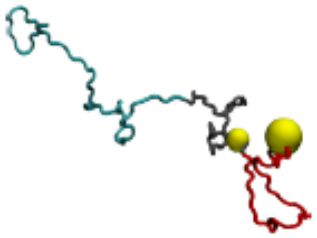
72–92



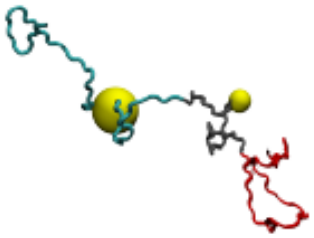
9–33



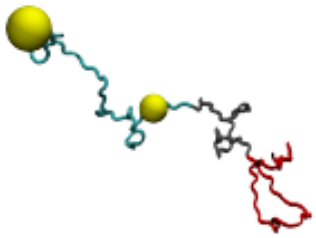
54–92



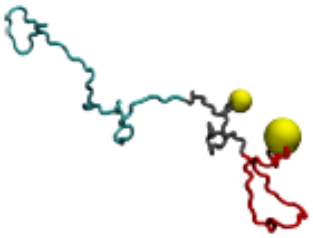
92–130



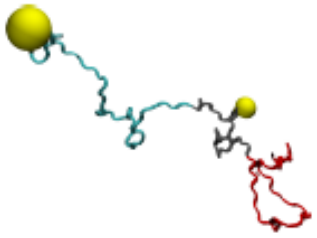
33–72



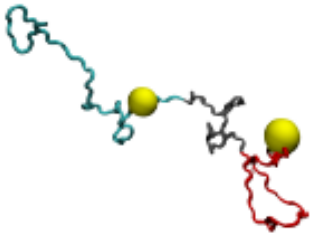
9–54



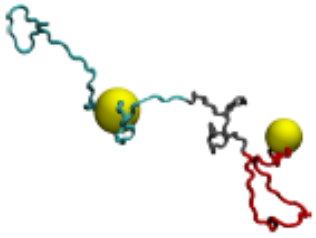
72–130



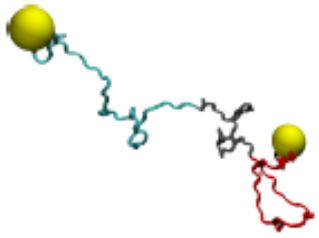
9–72



54–130



33–130



9–130

$$Q = N_p^{-1} \left| \sum_{i=1}^{N_p} Q_i \right|$$

$$H = N_p^{-1} \sum_{i=1}^{N_p} e_i$$

$$\varepsilon_i > 0$$

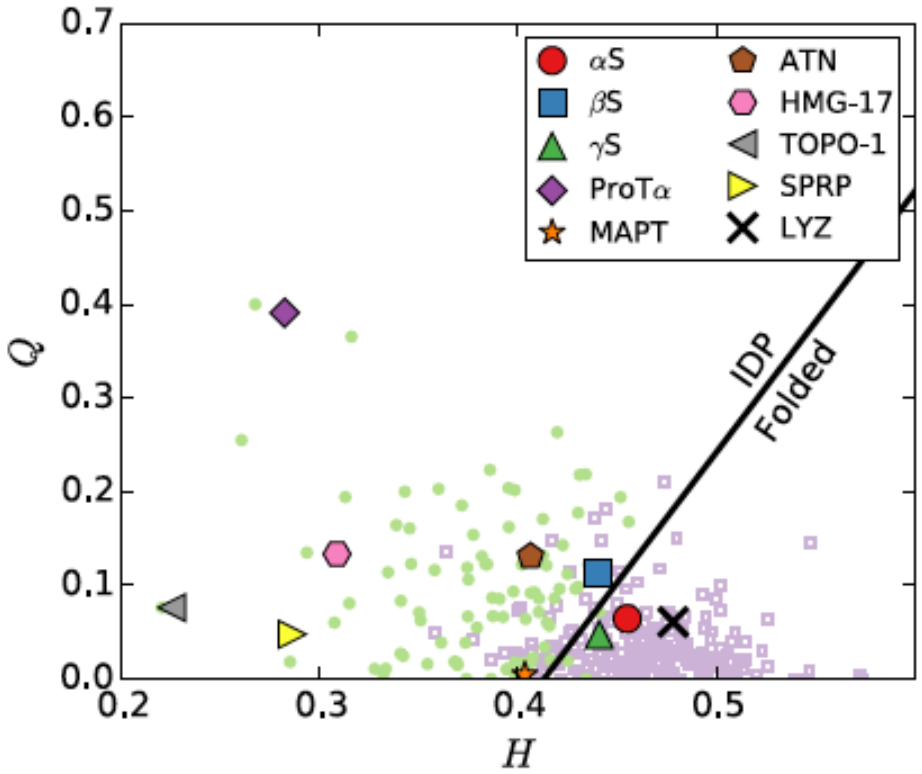


FIG. 1. (Color online) Absolute value of the electric charge per residue Q versus the hydrophobicity per residue H (using the shifted and normalized Monera hydrophobicity scale) for known IDPs (small circles) and 221 folded proteins [32] (small open squares). The IDPs αS (large circle), βS (large square), γS (upward triangle), ProTα (diamond), MAPT (star), ATN (pentagon), HMG-17 (hexagon), TOPO-1 (leftward triangle), SPRP (rightward triangle), and the folded protein lysozyme C (X) are highlighted. The line $Q = 2.785H - 1.151$ represents the dividing line between IDPs (above the line) and natively folded proteins (below the line) given in Ref. [32].

TABLE I. Numbers of each amino acid type in αS, βS, γS, MAPT, and Pro$T\alpha$. "+" and "−" denote positively and negatively charged residues, respectively (Table III). "a" and "r" indicate highly hydrophobic ($\epsilon_i \sim 1$) and hydrophilic ($\epsilon_i \sim 0$) residues using the scaled and shifted Monera hydrophobicity scale described in Sec. II.

Amino acid type	αS	βS	γS	MAPT	Pro$T\alpha$
ALA	19	18	16	34	11
ARG^{+}	0	2	2	14	2
ASN	3	1	4	11	6
ASP^{-r}	6	3	3	29	19
CYS	0	0	0	2	0
GLN	6	6	6	19	2
GLU^{-}	18	25	20	27	34
GLY	18	13	10	49	9
HIS^{+}	1	1	0	12	0
ILE^{a}	2	2	2	15	1
LEU^{a}	4	7	1	21	1
LYS^{+}	15	11	15	44	8
MET	4	4	2	6	1
PHE^{a}	2	3	2	3	0
PRO^{r}	5	8	2	43	1
SER	4	6	10	45	4
THR	10	7	10	35	6
TRP^{a}	0	0	0	0	0
TYR	4	4	1	5	0
VAL	19	13	21	27	5
Total	140	134	127	441	110

hydrophobic

hydrophilic

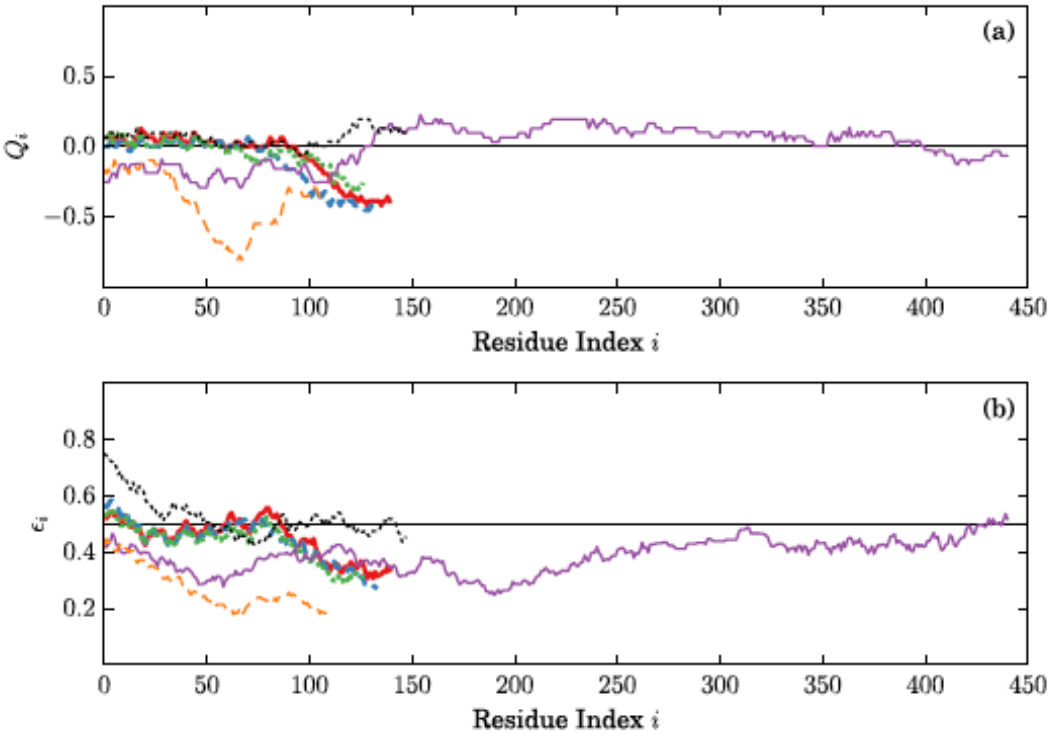
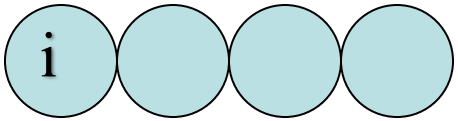


FIG. 2. (Color online) (a) Electric charge Q_i (in units of the electron charge q_e) and (b) hydrophobicity ϵ_i as a function of the residue index i originating from the N-terminus for the IDPs αS (thick, solid red line), βS (thick, dashed blue line), γS (thick, dotted green line), MAPT (thin, solid purple line), ProTα (thin, dashed orange line), and the folded protein lysozyme C (thin, dotted black line). We quote the normalized and shifted Monera hydrophobicity scale [34], where 0 is the least and 1 is the most hydrophobic [see Eq. (8)]. Data for each i is averaged over 31 nearby residues, with data at the endpoints reflected beyond the endpoints to reduce edge effects. This averaging is employed to visualize the general differences between biologically important regions of the proteins. Note that the curves for Q_i and ϵ_i are not strongly sensitive to the averaging length.

Coarse-grained model



M_i, σ_i, ε_i, Q_i

l=3.9Å

θ_0=2.12 rad

λ=9Å

$\alpha_{CG}=\varepsilon_A/\varepsilon_{es}$

$e_{ij}=\sqrt{e_i e_j}$

$$V^{\mathrm{bl}} = \frac{k_\ell}{2}\sum_{\langle ij\rangle}(r_{ij}-\ell)^2$$

$$V^{\mathrm{ba}} = \frac{k_\theta}{2}\sum_{\langle ijk\rangle}(\theta_{ijk}-\theta_0)^2$$

$$V^{\mathrm{da}} = \sum_{\langle ijkl\rangle}\sum_{s=1}^{4} A_s\cos(s\phi_{ijkl}) + B_s\sin(s\phi_{ijkl})$$

$$V^{r} = \epsilon_r \sum_{ij}\left\{4\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]+1\right\} \times \Theta(2^{\frac{1}{6}}\sigma - r_{ij})$$

$$V^{a} = \epsilon_a \sum_{ij}\left(\epsilon_{ij}\left\{4\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]+1\right\} \times \Theta(r_{ij} - 2^{\frac{1}{6}}\sigma) - \epsilon_{ij}\right)$$

$$V^{\mathrm{es}} = \epsilon_{\mathrm{es}}\sum_{ij}\frac{Q_i Q_j}{q_e^2}\frac{\sigma}{r_{ij}}e^{-\frac{r_{ij}}{\lambda}}$$

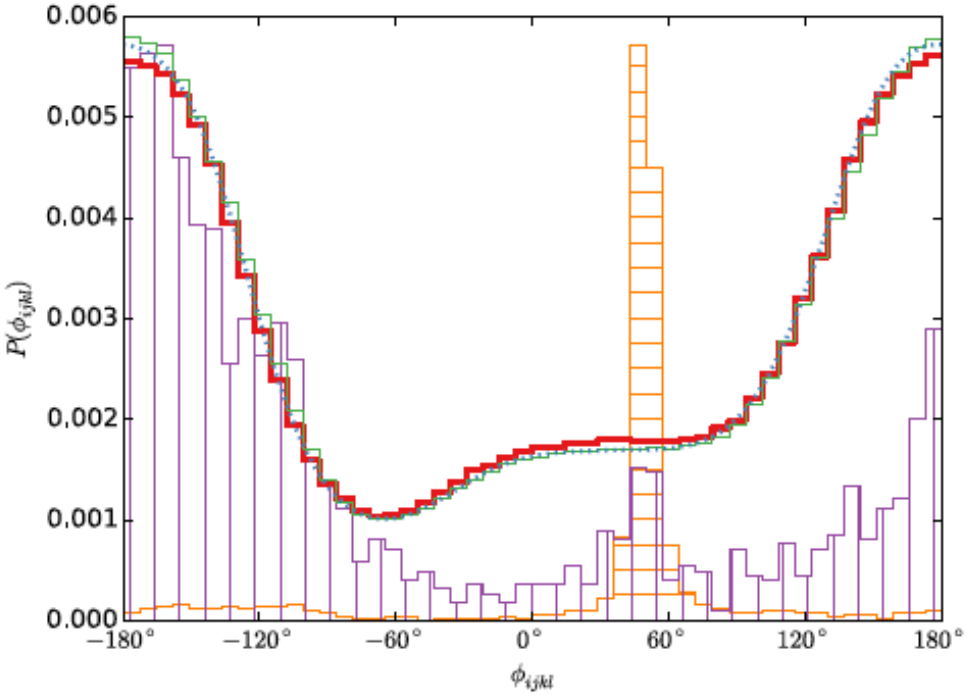


FIG. 3. (Color online) The backbone dihedral angle distribution $P^{\mathrm{UA}}(\phi_{ijkl})$ obtained from the UA description of αS (light green solid line) with only hard-sphere atomic interactions plus stereochemical constraints obtained from the Dunbrack database of high-resolution protein crystal structures. We fit $P^{\mathrm{UA}}(\phi_{ijkl})$ for the UA model using four coefficients (Table II) in the Fourier series in Eq. (4) (blue dotted line). We show that $P^{\mathrm{CG}}(\phi_{ijkl})$ from Langevin dynamics simulations of the CG model for αS with only bond-length, bond-angle, and dihedral angle interactions in Eqs. (2), (3), and (4) (thick solid red line) matches that from the hard-sphere UA model for α-synuclein. $P(\phi_{ijkl})$ from stretches of α helices (orange horizontal lines) and β sheets (purple vertical lines) that are longer than 10 residues in the Dunbrack database of high-resolution protein crystal structures are also shown for comparison. For ease of visual comparison, the dihedral angle distributions from α-helical and β-sheet structures were not normalized.

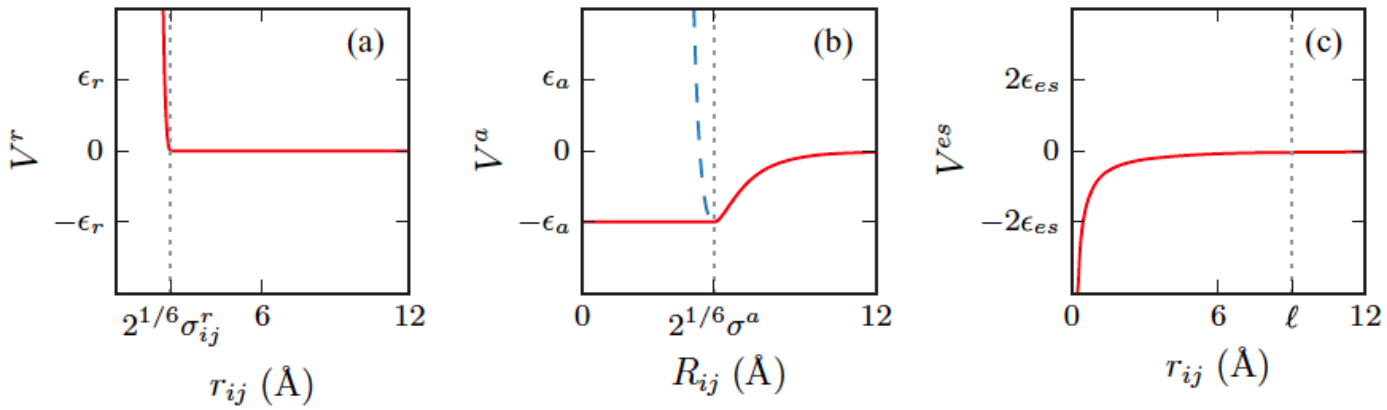


FIG. 3. (Color online) Schematics of (a) the purely repulsive Lennard-Jones potential V^r in Eq. (4) (solid line), (b) attractive Lennard-Jones potential V^a in Eq. (5) (solid line), and (c) screened Coulomb potential V^{es} in Eq. (7) (solid line). The dashed line in (b) represents the repulsive Lennard-Jones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

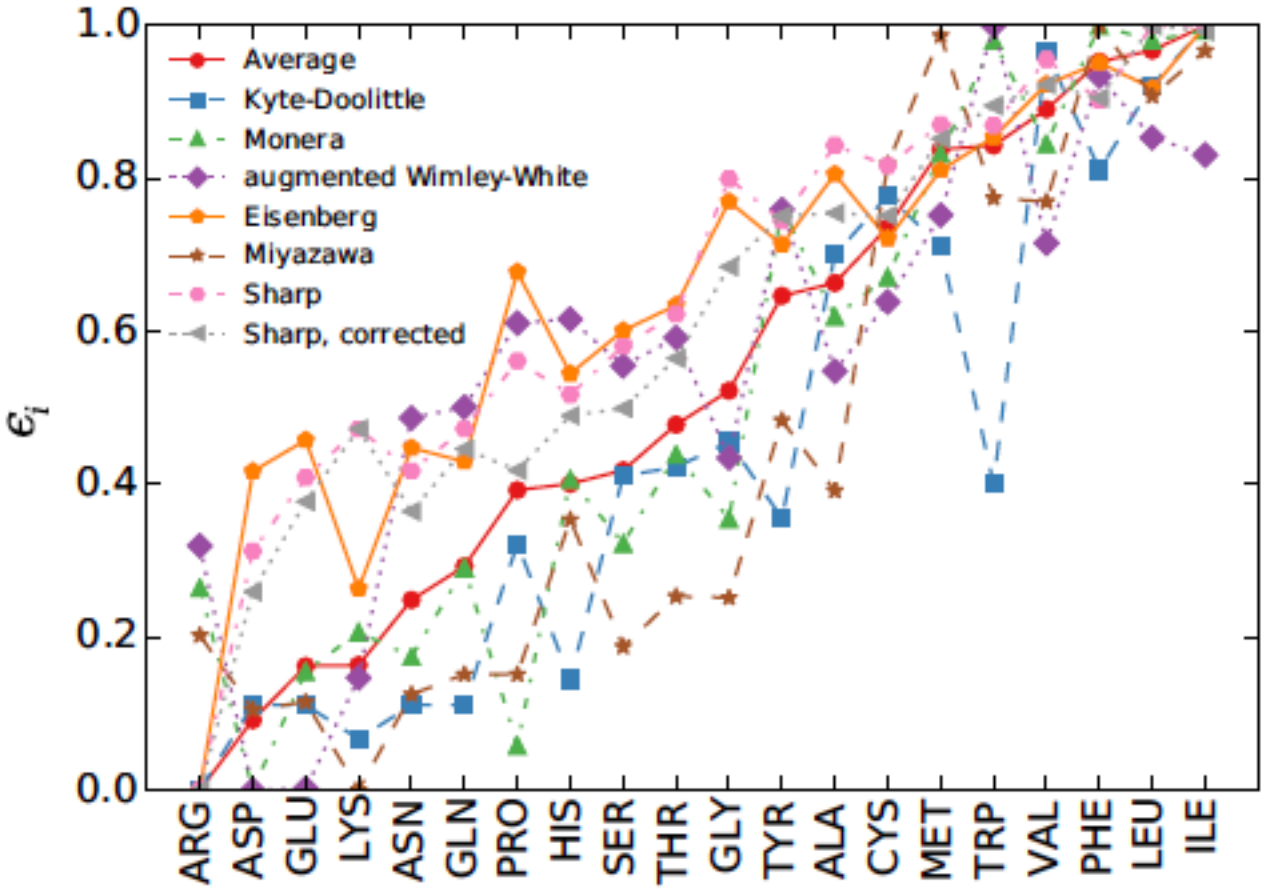
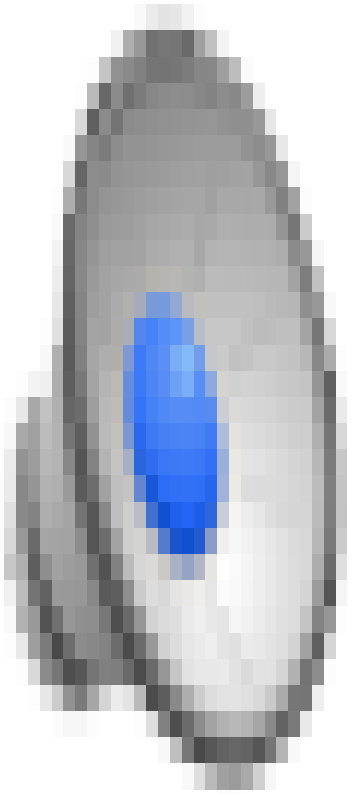


FIG. 4. (Color online) Seven commonly used hydrophobicity scales (Kyte-Doolittle [47], Monera [34], augmented Wimley-White [48,49], Eisenberg [50], Miyazawa [51], Sharp, and Sharp with solvent-solute size difference corrections [52]) for each amino acid type that have been shifted and normalized so that $0 \leqslant \epsilon_i \leqslant 1$. The "average" value for each residue indicates the shifted and normalized average over the seven shifted and normalized hydrophobicity scales. The residues are ordered according to their average ϵ_i.



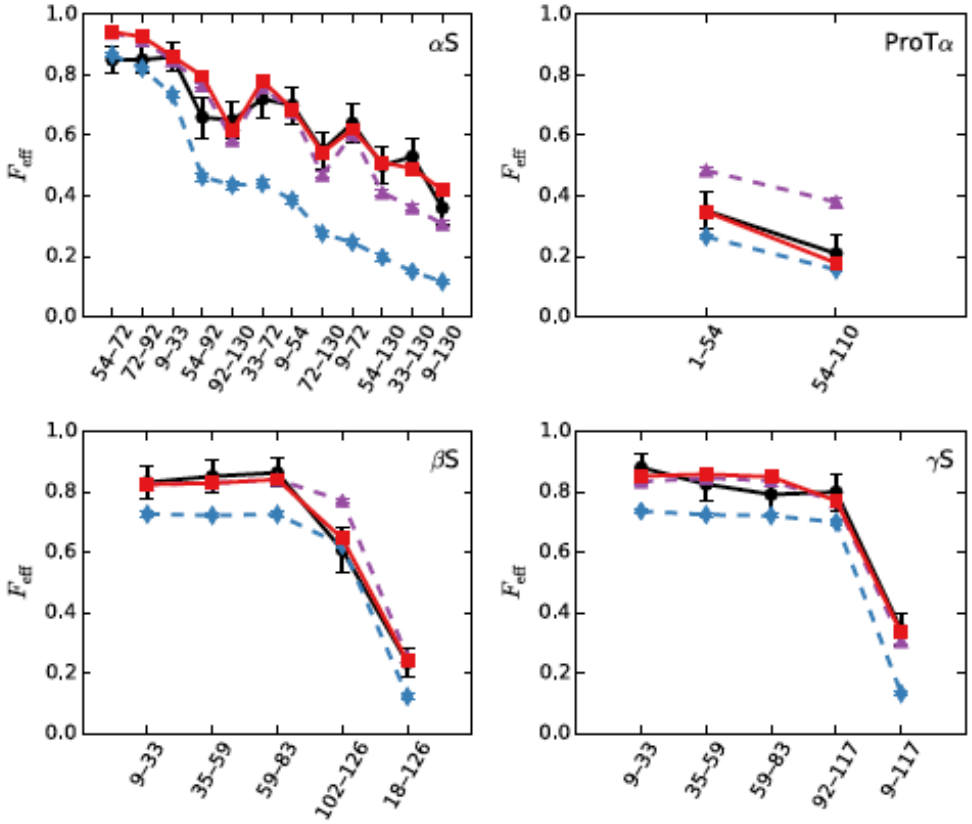


FIG. 5. (Color online) FRET efficiencies $F_{\rm eff}$ for αS (upper left), ProTα (upper right), βS (lower left), and γS (lower right) from experimental measurements (black circles) and CG Langevin dynamics simulations. We include data for three choices for the strength of the hydrophobic and electrostatic interactions $\overline{\epsilon}_a$ and $\overline{\epsilon}_{\rm es}$ for each IDP: (1) $\overline{\epsilon}_a = 0$ and $\overline{\epsilon}_{\rm es} = \kappa_{\rm es}$ such that the chains behave as extended coils (blue diamonds), (2) the optimal $\alpha_{\rm CG}$ for each protein with $\overline{\epsilon}_{\rm es} = \kappa_{\rm es}$, where the root-mean-square deviations between the experimental and simulation $F_{\rm eff}$ are minimized (red squares), and (3) the optimal $\alpha_{\rm CG}$ for each protein with no electrostatic interactions $\overline{\epsilon}_{\rm es} = 0$ (purple triangles). The error bars for $F_{\rm eff}$ from the simulations give the error in the mean. Error bars that are not visible are smaller than the symbols.

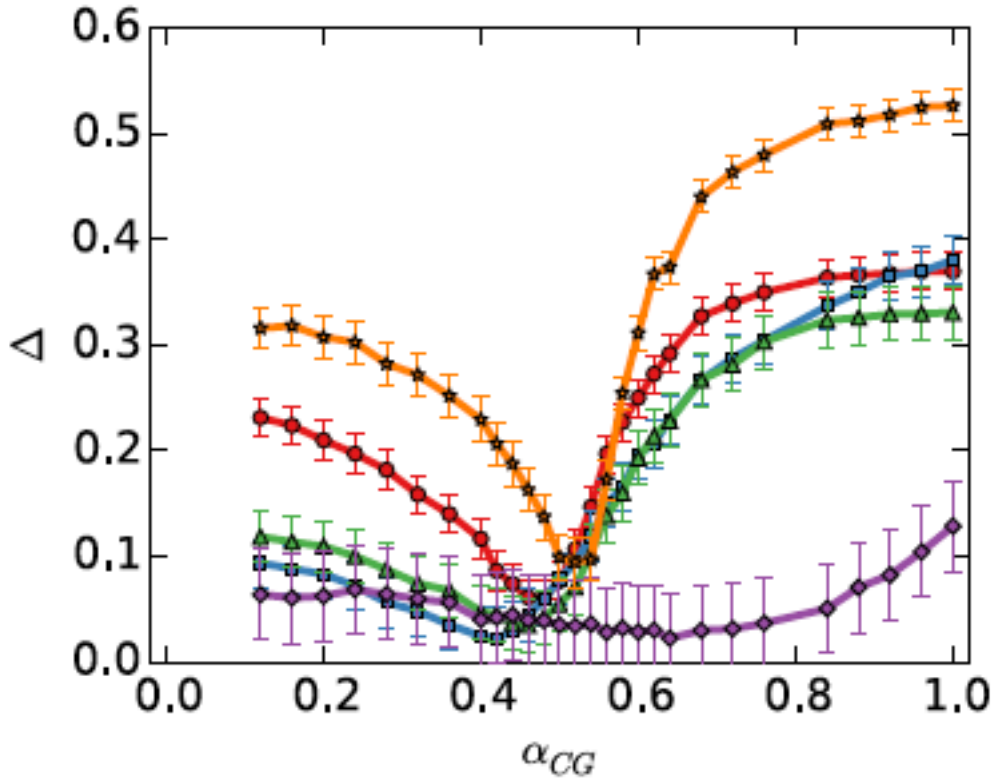


FIG. 6. (Color online) Root-mean-square deviation Δ in F_{eff} between experiments and simulations versus the ratio α_{CG} of the hydrophobicity and electrostatic interactions for αS (red circles), βS (blue squares), and γS (green triangles), MAPT (orange stars), and ProTα (purple diamonds).

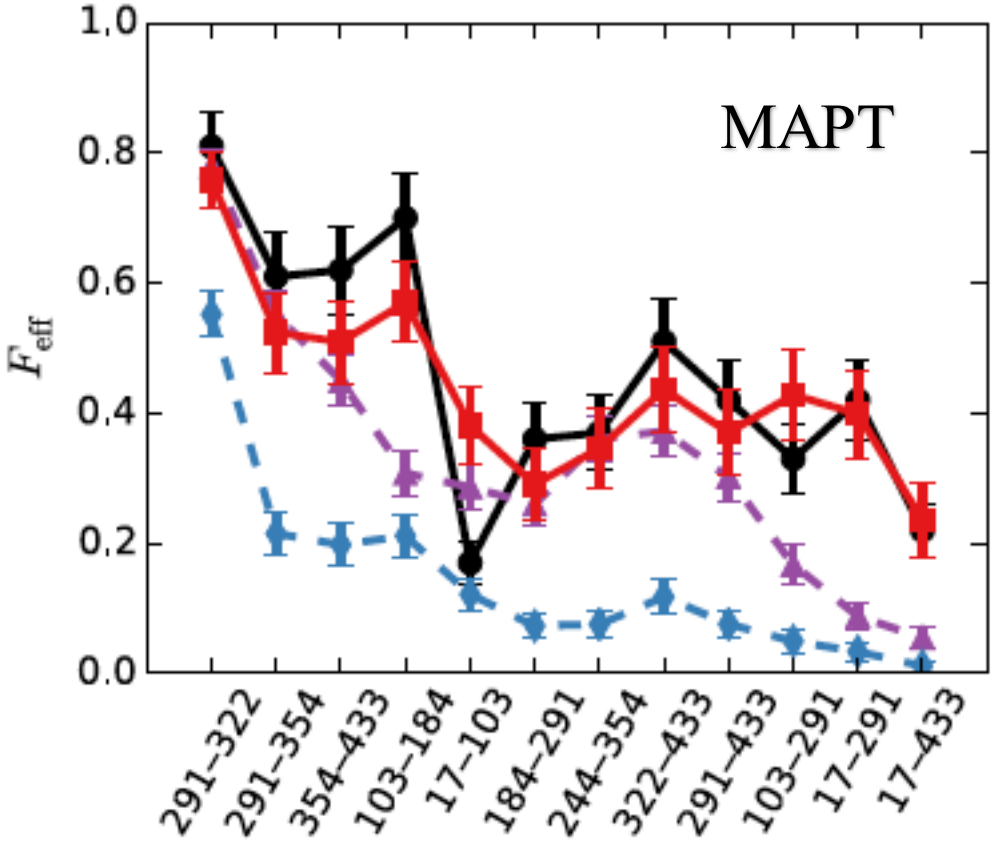


FIG. 7. (Color online) FRET efficiencies F_{eff} for MAPT from smFRET experiments (black solid line with circles) and three CG simulations: (1) $\overline{\epsilon}_a = 0$ and $\overline{\epsilon}_{\text{es}} = \kappa_{\text{es}}$ (blue diamonds), (2) the optimal $\alpha_{\text{CG}} = 0.52$ with $\overline{\epsilon}_{\text{es}} = \kappa_{\text{es}}$, where the root-mean-square deviations between the experimental and simulation F_{eff} are minimized (red squares), and (3) the optimal $\alpha_{\text{CG}} = 0.52$ with no electrostatic interactions $\overline{\epsilon}_{\text{es}} = 0$ (purple triangles).

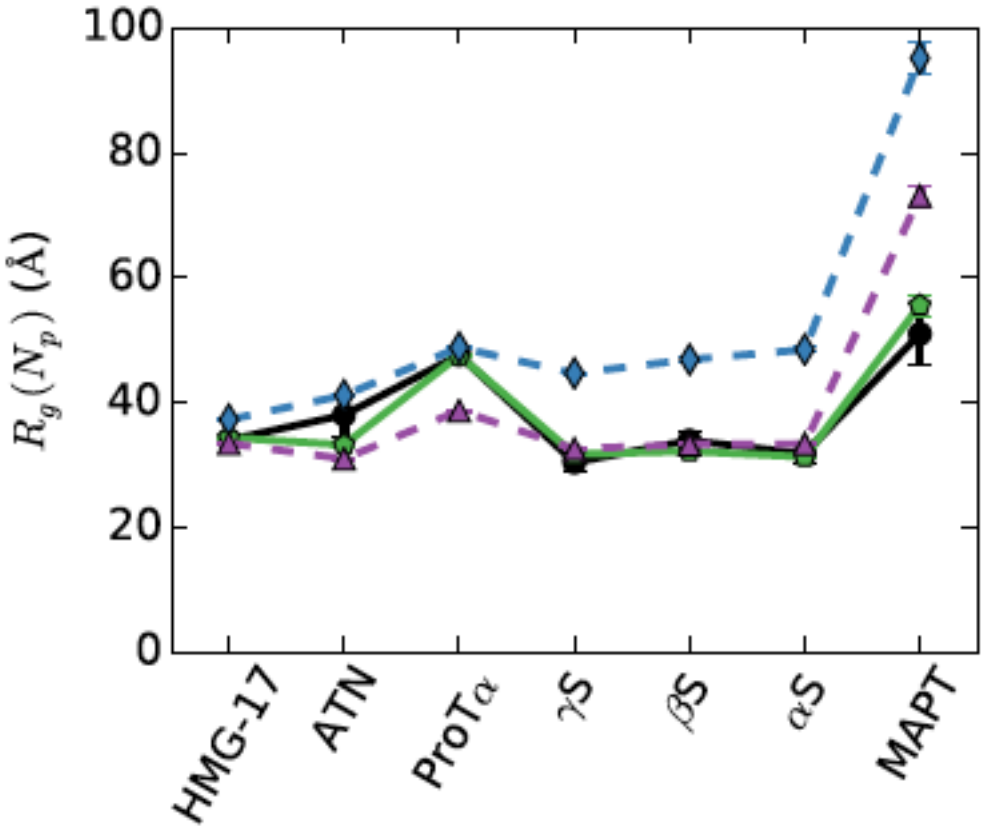


FIG. 9. (Color online) Radius of gyration $R_g(N_p)$ of seven IDPs from experiments [27,42,56–58] (black circles) and simulations of three CG models: (1) $\epsilon_a = 0$ and $\overline{\epsilon}_{es} = \kappa_{es}$ such that the chains behave as extended coils (blue diamonds), (2) $\alpha_{CG} = 0.50$ and $\overline{\epsilon}_{es} = \kappa_{es}$ (green pentagons), and (3) $\alpha_{CG} = 0.50$ and $\overline{\epsilon}_{es} = 0$ (purple triangles). The IDPs are ordered from shortest to longest (left to right).

$$R_g(n) = \frac{1}{N_p - n + 1} \sum_{i=1}^{N_p - n + 1} \langle R_g(i, i + n - 1) \rangle_t,$$

where $\langle . \rangle_t$ denotes a time average,

$$R_g(i,j) = \sqrt{\frac{1}{j - i + 1} \sum_{k=i}^{j} (\vec{r}_k - \langle \vec{r}_k \rangle)^2},$$

and

$$\langle \vec{r}_k \rangle = \frac{1}{j - i + 1} \sum_{k=i}^{j} \vec{r}_k,$$

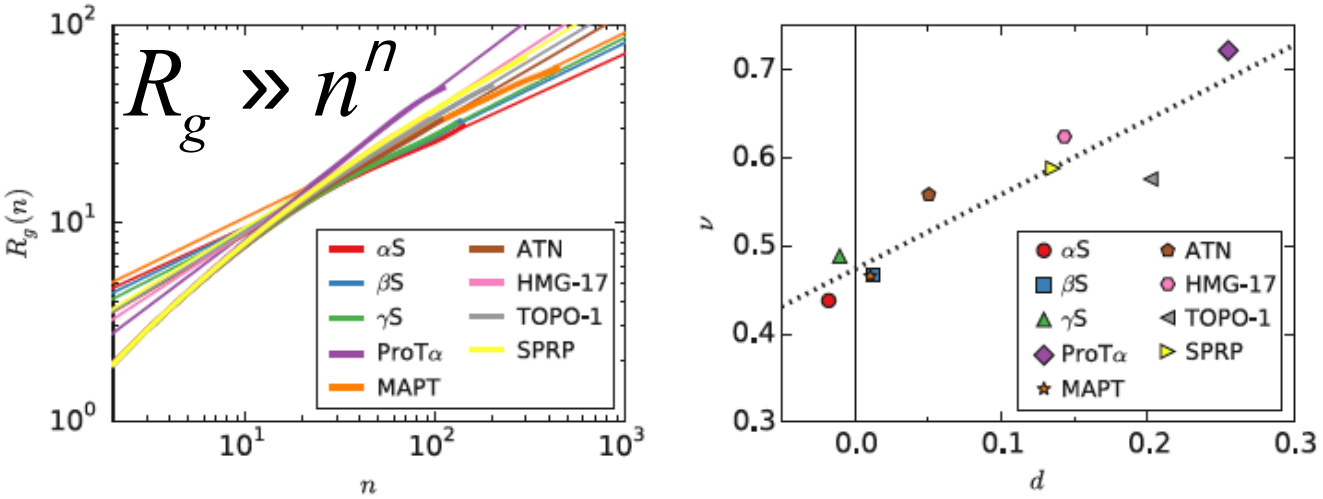


FIG. 10. (Color online) (Left) Radius of gyration $R_g(n)$ (thick lines) versus chemical distance n along the chain for several IDPs with $N_p \geqslant 90$ so that $R_g(n)$ is in the power-law scaling regime. Power-law fits of the data to $R_g = R^0 n^\nu$ for $n > 20$ are shown as thin lines. The error in R_g is comparable to the line thickness. (Right) Power-law scaling exponent ν as a function of the distance d from the dividing line between folded and intrinsically disordered proteins (Fig. 1). The dotted line follows $\nu = 0.47 + 0.85d$.

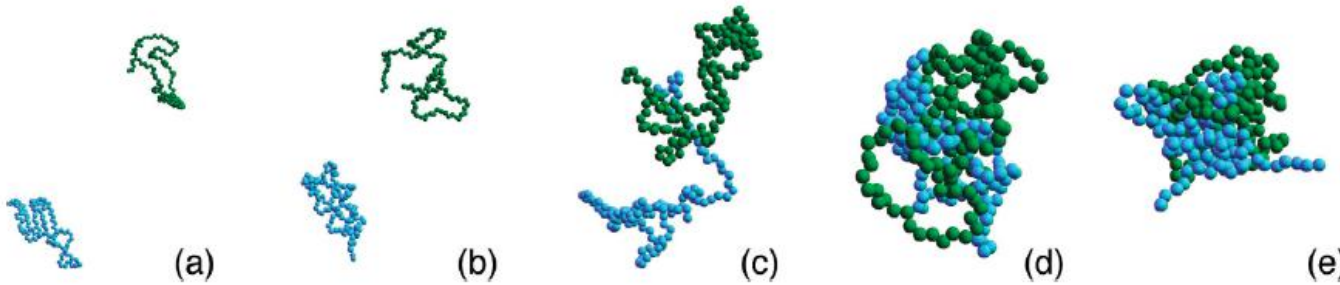


FIG. 9. (Color online) Snapshots from preliminary aggregation studies of two monomeric α-synuclein proteins (dark green and light blue) using coarse-grained simulations with the temperature set so that $\langle R_g \rangle \approx 33$ Å at $\alpha = 1.2$ (for individual protein monomers) for (a) $\alpha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

Conclusions

- Coarse-grained model that accurately captures Feff and Rg of IDPs
- MAPτ is somewhat different than other IDPs
- Studies of multiple IDPs