Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure

Wenli Meng[a,1], Nicholas Lyle[b,1], Bowu Luan[a], Daniel P. Raleigh[a,c,2], and Rohit V. Pappu[b,2]

[a]Department of Chemistry, State University of New York at Stony Brook, Stony Brook, NY 11794-3400; [b]Department of Biomedical Engineering, Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130; and [c]Graduate Program in Biochemistry and Structural Biology and Graduate Program in Biophysics, State University of New York at Stony Brook, Stony Brook, NY 11794

Edited by William A. Eaton, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, and approved December 20, 2012 (received for review September 28, 2012)

The sizes of unfolded proteins under highly denaturing conditions scale as $N^{0.59}$ with chain length. This suggests that denaturing conditions mimic good solvents, whereby the preference for favorable chain–solvent interactions causes intrachain interactions to be repulsive, on average. Beyond this generic inference, the broader implications of $N^{0.59}$ scaling for quantitative descriptions of denatured state ensembles (DSEs) remain unresolved. Of particular interest is the degree to which $N^{0.59}$ scaling can simultaneously accommodate intrachain attractions and detectable long-range contacts. Here we present data showing that the DSE of the N-terminal domain of the L9 (NTL9) ribosomal protein in 8.3 M urea lacks detectable secondary structure and forms expanded conformations in accord with the expected $N^{0.59}$ scaling behavior. Paramagnetic relaxation enhancements, however, indicate the presence of detectable long-range contacts in the denatured-state ensemble of NTL9. To explain these observations we used atomistic thermal unfolding simulations to identify ensembles whose properties are consistent with all of the experimental observations, thus serving as useful proxies for the DSE of NTL9 in 8.3 M urea. Analysis of these ensembles shows that residual attractions are present under mimics of good solvent conditions, and for NTL9 they result from low-likelihood, medium/long-range contacts between hydrophobic residues. Our analysis provides a quantitative framework for the simultaneous observation of $N^{0.59}$ scaling and low-likelihood long-range contacts for the DSE of NTL9. We propose that such low-likelihood intramolecular hydrophobic clusters might be a generic feature of DSEs that play a gatekeeping role to protect against aggregation during protein folding.

atomistic simulations | denatured proteins | paramagnetic relaxation

Quantitative descriptions of unfolded proteins are important for understanding collapse transitions (1), protein folding mechanisms (2), misfolding, aggregation (3, 4), and the effects of macromolecular crowding on protein stability (5, 6). Expanded unfolded states are sampled in high concentrations of chemical denaturants such as urea and guanidinium chloride. The sizes of these denatured proteins, quantified using hydrodynamic radii ($\langle R_h \rangle$) or radii of gyration ($\langle R_g \rangle$), scale as $N^{0.59}$ with chain length (7–9). This $N^{0.59}$ scaling arises because denatured proteins expand to make favorable contacts with the surrounding solvent, implying that high concentrations of denaturants are good solvents for generic proteins.

In good solvents, the ensemble-averaged interresidue pair interaction coefficient is positive, suggesting that the preference for favorable chain–solvent interactions leads to intrachain interactions being repulsive on average (10). The validity of this inference for denatured proteins has been demonstrated recently using a combination of single-molecule experiments and polymer theory (11). Quantitative descriptions of chain statistics for polymers in good solvents rely on the so-called excluded volume (EV) limit as an important reference state, and this is true for denatured state ensembles (DSEs) as well. EV limit ensembles are typically generated using atomistic descriptions of proteins and ignoring all nonbonded interactions excepting steric repulsions (12–15). Hence, the ensemble-averaged interresidue interaction coefficient (related to the second virial coefficient) is, by construction, positive in the EV limit, thus affording the reproduction of $N^{0.59}$ scaling (10). Descriptions of chain statistics based on EV limit ensembles are routinely used as reference states for intrinsically disordered proteins (IDPs) and denatured proteins and have proven useful in interpreting the results of NMR and small angle X-ray scattering (SAXS) measurements for these systems (12–21).

If measured properties of DSEs can be adequately explained using EV limit ensembles, then the implication is as follows: in a typical Flory-like mean field description, the energy for each polymer configuration can be written as $U = U_{EV} + U_{non\text{-}EV\text{–}intrachain} + U_{chain\text{–}solvent}$ (22). The EV limit is reached if the non-EV–intrachain interactions are exactly counterbalanced by chain–solvent interactions. Deviations from the EV limit arise if there is imperfect compensation between non-EV–intrachain interactions and chain–solvent interactions. If imperfect compensation results from attractive non-EV–intrachain interactions, then deviations from the EV limit can lead to persistent local structure and global compaction with deviation from $N^{0.59}$ scaling, as has been the observed for proteins under mild denaturing conditions (23–33) or for proteins under folding [i.e., poor solvent (34)] conditions (35, 36). Imperfect compensation that results from long-range intramolecular electrostatic repulsions and/or favorable chain–solvent interactions will also cause deviations from $N^{0.59}$ scaling. In these cases, the deviation causes increased chain expansion, as has been observed recently for highly charged IDPs in aqueous solutions and in the presence of denaturant (18, 37).

Of interest is the possibility that imperfect compensation can be achieved between non-EV–intrachain interactions and chain–solvent interactions without causing deviations from $N^{0.59}$ scaling.

Although this is the intuitive and canonical expectation given the documented denatured state effects on protein folding (6), no quantitative evidence has been offered in support of imperfect compensation being achieved while preserving $N^{0.59}$ scaling. Here, we report an archetypal dataset for the DSE of the N-terminal domain of the L9 (NTL9) ribosomal protein in 8.3 M urea, showing chain expansion consistent with $N^{0.59}$ scaling and the presence of detectable long-range contacts despite negligible secondary structure. Paramagnetic relaxation enhancements

Author contributions: W.M., N.L., D.P.R., and R.V.P. designed research; W.M. and N.L. performed research; B.L. contributed new reagents/analytic tools; W.M., N.L., D.P.R., and R.V.P. analyzed data; and N.L., D.P.R., and R.V.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The assignments reported in this paper have been deposited in the Biological Magnetic Resonance Data Bank, www.bmrb.wisc.edu [accession nos. 17506 (native state) and 17761 (urea-induced denatured state ensemble)].

[1]W.M. and N.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: draleigh@notes.cc.sunysb.edu or pappu@wustl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216979110/-/DCSupplemental.

(PREs) provide support for the presence of transient long-range contacts in the DSE of NTL9. The PRE data cannot be explained using EV limit ensembles and hence highlight the presence of imperfect compensation with $N^{0.59}$ scaling. To understand the observations for the DSE of NTL9 we used atomistic thermal unfolding simulations of NTL9. In the simulations, the temperature acts as a proxy for the effect of denaturant (which changes solvent quality) by modulating the balance between the preference for compact conformations (seen at low temperatures) and expanded conformations (seen at high temperatures). We identified a set of high temperatures whose ensembles yield averages that are congruent with experimental data for the DSE of NTL9 in 8.3 M urea. Analysis of these ensembles provides a quantitative reasoning for the observed properties of NTL9's DSE in 8.3 M urea. Our simulation approach does not impose any a priori assumptions of equivalence between thermal and urea denaturation. Instead, we use it to identify unfolded ensembles that yield properties consistent with the full panel of measurements, thus yielding a quantitative, albeit computationally tractable framework for interpreting the experimental observations.

Results

NTL9 Samples Predominantly Expanded Conformations with Negligible Secondary Structure in Its DSE in 8.3 M Urea. The values of R_h and R_g were measured using NMR pulsed field gradient diffusion and SAXS experiments, respectively. The measured R_h for NTL9 DSE in 8.3 M urea at 12 °C is 22.5 Å, and the measured R_g is 21.3 ± 1.5 Å at a protein concentration of 7.5 mg/mL (*SI Appendix*, Figs. S1 and S2). At a similar concentration, the measured R_g value for the native state is 12.2 Å (*SI Appendix*, Fig. S3). The measured R_h and R_g values are consistent with reported scaling laws for highly unfolded polypeptide chains (8, 9).

NMR spectra for the native and urea-induced denatured states of NTL9 are shown in *SI Appendix*, Fig. S4. Although the spectrum is less well resolved in 8.3 M urea, we obtained complete backbone ^{1}H, ^{13}C, and ^{15}N assignments as well as side chain $^{13}C_\beta$ assignments using standard approaches (*SI Appendix*, p. S20). In 8.3 M urea, 93% of the molecules are unfolded. Although native state resonances were detected at lower contour levels, these were in slow exchange with the denatured state. Secondary shifts (38) were used to quantify the deviation between measured and random coil chemical shifts. *SI Appendix*, Fig. S5 shows a plot of the $^{13}C_\alpha$, $^{13}C_\beta$, $\Delta\delta^{13}C_\alpha - \Delta\delta^{13}C_\beta$, ^{13}CO, and $^{1}H_\alpha$ secondary shifts for the native state and for the NTL9 DSE in 8.3 M urea. The secondary shifts for the DSE are significantly smaller than those in the native state, with all of the ^{13}C secondary shifts being less than 1.0 ppm and the ^{1}H secondary shifts being less than or equal to 0.2 ppm. Residues 42–47 exhibit small positive $^{13}C_\alpha$ and ^{13}CO secondary shifts (0.31–0.37 ppm) as well as small positive values for $\Delta\delta^{13}C^\alpha - \Delta\delta^{13}C^\beta$ (maximum of 0.34 ppm). This suggests the presence of residual, albeit very small, α-helical propensity. In Fig. 1 we present a summary of the chemical shifts using secondary structure propensity (SSP) scores (39). A score of 1 indicates fully formed α-helical structure, and a score of −1 represents fully formed β-strands. The SSP scores for the NTL9 DSE in 8.3 M urea are close to zero and support the conclusion of negligible preference for α-helical and β-strand structure.

Paramagnetic Relaxation Enhancement Experiments Indicate the Presence of Long-Range Contacts for the NTL9 DSE in 8.3 M Urea. Nitroxide spin labels cause significant broadening of NMR resonances of spins that are within 20 Å of the spin label (40–42). These effects provide a useful probe for long-range contacts (15, 19, 30, 43, 44). Spin labels were attached at positions 2, 10, 32, 49, and 51, respectively (*SI Appendix*, Fig. S6). Under native conditions all of the spin-labeled mutants are folded as judged by NMR (*SI Appendix*, Fig. S7). We used the peak intensity ratios from the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of the paramagnetic and diamagnetic proteins. Using the methods of Iwahara et al. (45) we also measured the paramagnetic relaxation enhancement (PRE) $^{1}H_N$ transverse relaxation rates (Γ_2) for the

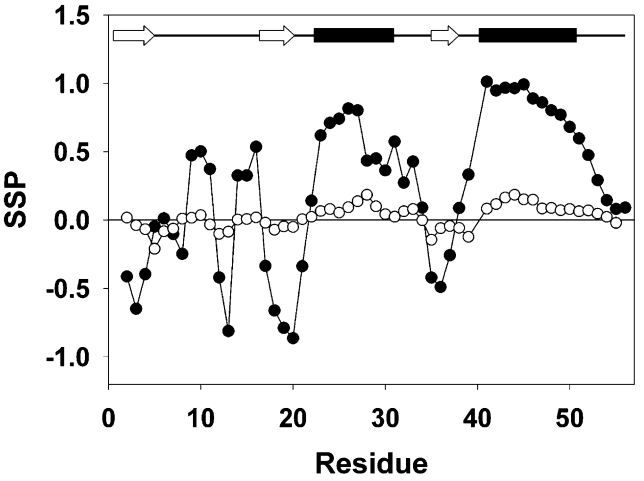

Fig. 1. SSP scores for the folded (●) and the DSE (○) of NTL9 in 8.3 M urea. A schematic diagram of the elements of secondary structure of the native state is shown at top.

NTL9 DSE in 8.3 M urea. Both measurements yield similar trends, as shown in Fig. 2.

The PRE results for the DSE (Fig. 2) show deviations from the values predicted by standard random coil models (40) for all positions. The observed PRE effects are not due to contributions from a small native fraction, because the two states are in slow exchange. We observed similar albeit reduced effects for the variants labeled at residue 49 and 51. There is an asymmetry in the PREs observed for the labels near the N terminus (K2) compared with the labels near the C terminus (A49 and K51). We propose that this reflects the differences in contact patterns between the two termini (see simulation results below) and the fact that the spin label is attached to a long side chain. For example, insertion of the spin labeled side chain of K2C into a hydrophobic cluster can enhance the relaxation of amide protons in the C terminus; conversely, the spin labeled side chains of the A49C and K51C mutants might project away from the clusters, leading to decreased relaxation enhancement for amide protons in N terminus.

The PRE data show significant deviations from the standard Gaussian chain random coil model used to benchmark PREs (40). The Gaussian random coil model does not account for EV effects nor does it account for the size and flexibility of the attached spin label, and it is possible that the observed deviations might reflect limitations of the Gaussian random coil model. To address these issues we performed two sets (with and without spin labels) of atomistic EV limit simulations (16) of NTL9. Results from both simulations differ from the Gaussian chain model, and there are small but detectable differences between the EV simulations with and without spin labels (*SI Appendix*, Fig. S9). Although the EV limit models predict more extensive PRE effects than the Gaussian chain model, neither model can reproduce the experimentally observed PREs. The observed PREs cannot be attributed solely to the simplicity of reference models used to calibrate these effects. Instead they suggest the presence of long-range contacts that need quantitative characterization.

Atomistic Simulations Help Identify Ensembles That Are Consistent with the Properties of NTL9 DSEs in 8.3 M Urea. We performed Metropolis Monte Carlo (MC) simulations based on atomistic descriptions of NTL9, the ABSINTH implicit solvation model and underlying force field paradigm (46), and parameters from the OPLSS-AA/L molecular mechanics force field (47). For each simulation temperature between 240 K and 500 K we performed multiple independent MC simulations, each based on a different random seed and the native state as the starting conformation. We calculated ensemble averages for R_g, secondary structure propensities, and PRE data as a function of the simulation temperature. These

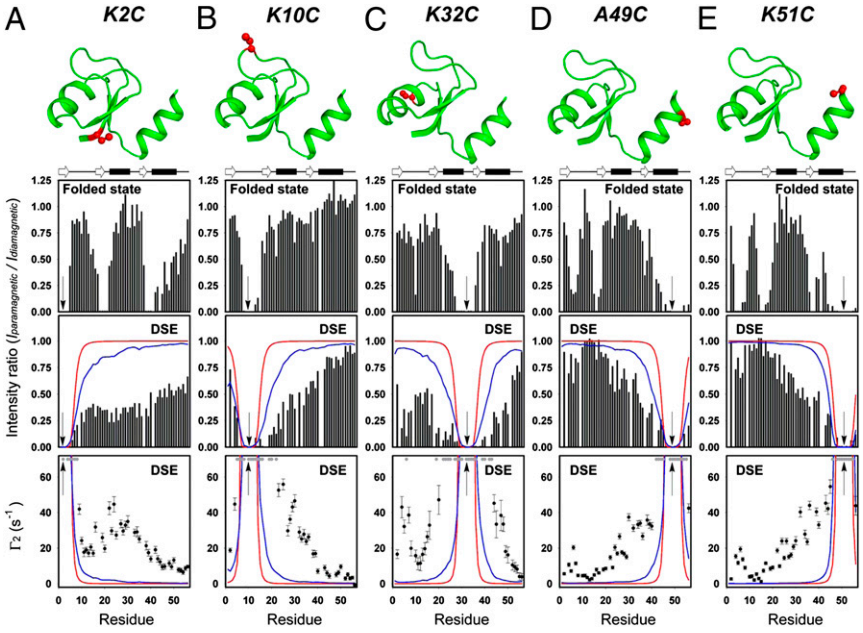

Fig. 2. Paramagnetic relaxation data for NTL9 under native state conditions and for the DSE in 8.3 M urea for (*A*) K2C, (*B*) K10C, (*C*) K32C, (*D*) A49C, and (*E*) K51C. The histograms display the intensity ratio of the ^{1}H-^{15}N cross-peaks in the HSQC spectra in the folded and urea denatured state. (*Bottom*) PRE $^{1}H_N$-Γ_2 rates (●) for urea denatured state are plotted. Residues for which the peaks disappeared in the paramagnetic form are indicated by gray dots. Red lines represent the values expected from the Gaussian random-coil model. The blue line represents the values calculated using simulations of atomistic ensembles in the EV limit with explicit incorporation of the spin label. Arrows indicate the location of each spin label.

were used to identify temperatures whose ensembles generated averages that are concordant with experimental data and thus serve as models for the DSE of NTL9 in 8.3 M urea.

We compared the measured PRE data with profiles calculated from simulated ensembles using the parameters Δ_1 and Δ_2 that quantify the temperature-dependent root mean square deviations from measured values of the intensity ratios and Γ_2, respectively (Fig. 3 *A* and *B* and *SI Appendix*, Eq. **S10**). The smallest Δ_j values were obtained for three simulation temperatures—380 K, 390 K, and 400 K—and we refer to these as T_D temperatures. Fig. 4 and *SI Appendix*, Fig. S10 show details of the quantitative agreement between PRE profiles calculated for each of the T_D temperatures and experimental data.

Fig. 3*C* plots the average α-helical and β-sheet contents calculated at each of the simulation temperatures. These contents are less than 3% at T_D temperatures in agreement with estimates from NMR experiments for the DSE of NTL9 in 8.3 M urea. Fig. 3*D* plots the temperature dependence of the ensemble-averaged R_g. We obtained $\langle R_g \rangle = 20.50 \pm 0.08$ Å, 21.21 ± 0.06 Å, and 21.74 ± 0.05 Å for 380 K, 390 K, and 400 K, respectively. These $\langle R_g \rangle$ values are in accord with expectations from scaling relations (9) and similar to the values obtained using SAXS measurements for the DSE of NTL9 in 8.3 M urea. As a reference, $\langle R_g \rangle = 28.31 \pm 0.02$ Å for NTL9 in the EV limit. We used the ensembles for T_D temperatures as proxies for NTL9 DSE in 8.3 M urea.

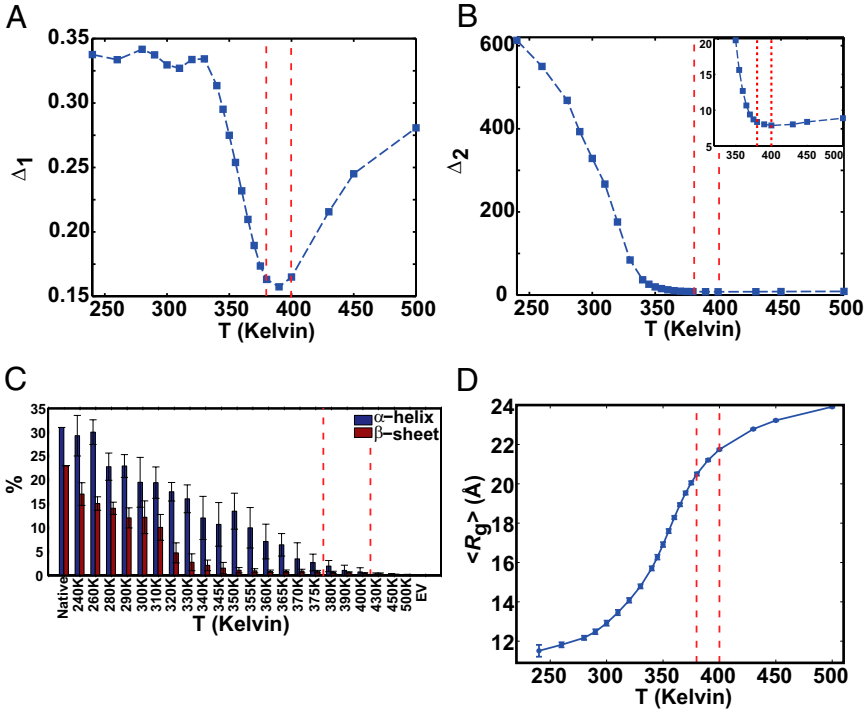

Fig. 3. Identification of the temperature interval for T_D ensembles. In all panels the dashed lines bracket the temperatures $T = 380$ K, 390 K, and 400 K, which are the T_D temperatures. (*A* and *B*) Plots of Δ_1 (*A*) and Δ_2 (*B*) vs. temperature. (*C*) Plot of the temperature dependence of secondary structure contents, which were calculated using the DSSP algorithm (51). (*D*) Plot of the temperature-dependent $\langle R_g \rangle$ values from simulation results.

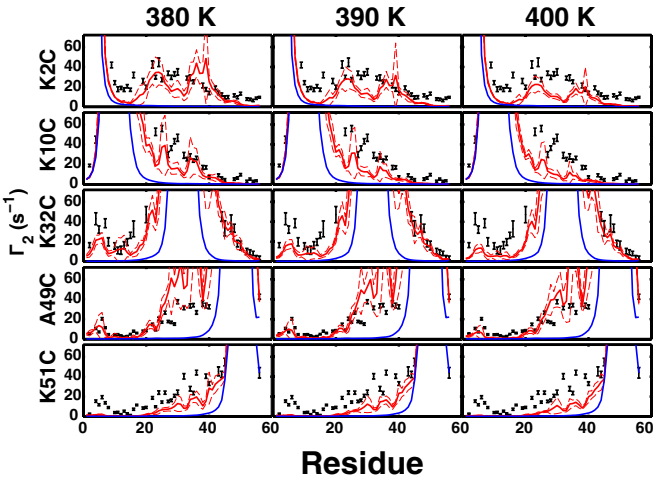

Fig. 4. Comparison between experimental data (in black symbols with error bars) and calculated PRE profiles, with comparisons shown in terms of Γ_2. The profiles shown in blue were obtained using conformations drawn from the EV ensemble, and the profiles shown in red were obtained using conformations for T_D ensembles, with each column corresponding to a specific T_D temperature. The dashed red curves indicate the confidence intervals on calculated profiles. We calculated the latter by partitioning the simulated ensembles into 10 blocks and using the deviations of block averages from the overall mean calculated these errors.

Comparative Analysis of Contact Probabilities. In the EV limit all pairwise interactions are purely repulsive, and as a result the average percent probability p_{ij} of realizing contacts between residues i and j decreases sharply with increasing sequence separation |i–j|, such that $0.001\% < p_{ij} < 0.05\%$ for |i–j| > 10 (Fig. 5). In contrast, the corresponding probabilities for conformations drawn from the T_D ensembles ($0.01\% < p_{ij} < 1\%$) are at least two orders of magnitude larger than the EV limit for sequence separations in the range $10 < |i–j| < 40$. This indicates the presence of detectable low likelihood medium/long-range contacts in the T_D ensembles.

Do T_D Temperatures Mimic Good Solvents for NTL9? The scaling exponent is $\nu \sim 0.59$ in the EV limit and for generic proteins in high concentrations of denaturants (10, 17). For expanded conformations the average distance $\langle R_{ij}\rangle$ between any pair of residues i and j should follow a power law relationship *viz.*, $\langle R_{ij}\rangle = R_o|j–i|^{\nu}$ (*SI Appendix*, Fig. S11) (10, 16, 20). If the T_D temperatures mimic good solvents, then ν should approach 0.59 for the scaling of $\langle R_{ij}\rangle$ as a function of |j–i|, providing the latter is long enough for scaling theory to apply (16, 20). For different combinations of pairs of residues (i,j) and (k,l) we used $\nu = \frac{\ln\langle R_{ij}\rangle - \ln\langle R_{kl}\rangle}{\ln(|j-i|) - \ln(|l-k|)}$, where $\langle R_{ij}\rangle$ and $\langle R_{kl}\rangle$ denote average distances, and |j–i| and |l–k| denote sequence separations between residues (i,j) and (k,l), respectively. We analyzed the EV limit ensembles of NTL9 for all combinations of (i,j), and (k,l) that satisfy the constraints |j–i| and |k–l| ≥ 25. This yields a distribution of values for ν (Fig. 6*A*) with an average value of $\langle\nu\rangle = 0.59 \pm 0.03$. A similar analysis was carried out for each T_D temperature, and we found $\langle\nu\rangle$ to be 0.59 ± 0.05, 0.58 ± 0.05, and 0.57 ± 0.04 for 380 K, 390 K, and 400 K, respectively. Therefore, T_D temperatures mimic good solvent conditions for NTL9.

Reconciling $\nu \sim 0.59$ with Deviations from the EV Limit. Dimensionless quantities such as $|j–i|^{0.59}$ are converted to distances using a multiplicative prefactor R_o. In a good solvent, this parameter can be used to quantify the average volume excluded per residue for favorable interactions with the surrounding solvent. If the net charge per residue is ≤0.3 (18) and intrachain electrostatic repulsions are screened (48), then $R_o \leq R_o^{EV}$ in a generic good solvent where R_o^{EV} denotes the value of R_o calculated in the EV limit. If $R_o < R_o^{EV}$, then residual intrachain attractions persist in the good solvent. This will influence the probabilities associated with long-range contacts.

We estimated the value of R_o for each T_D temperature and compared these with R_o^{EV}. We fit the values for $\langle R_{ij}\rangle$ to the equation $\ln(\langle R_{ij}\rangle) = \ln(R_o) + 0.59\ln(|j–i|)$ for all $|j–i| \geq 25$, where R_o is the free parameter. The results are shown in Fig. 6*B*. The maximal value of R_o is R_o^{EV}, and these values decrease as the T_D temperatures decrease. Our analysis provides quantitative evidence for imperfect compensation between non-EV–intrachain and chain–solvent interactions. For the DSE of NTL9 our model ensembles suggest that imperfect compensation results from the presence of residual attractions.

Analysis of Contact Patterns for T_D Temperature Ensembles. *SI Appendix*, Fig. S12 quantifies the probabilities of interresidue contacts in the native state ensemble (240 K), the T_D temperatures, and the EV limit, respectively. Each cell in a contact map quantifies the probability that residues i and j are in contact. We define residues to be in contact if the interresidue distance between at least one pair of heavy atoms is ≤3.5Å. We observe patterns that include several low-probability native as well as nonnative contacts ($p_{ij} < 0.1$). We calculated two sets of difference contact maps to quantify the differences between the ensembles sampled at T_D temperatures and the native state and EV limit ensembles, respectively. Fig. 7 and *SI Appendix*, Fig. S13 show difference maps with respect to the native state ensemble and the EV limit as the reference. The colors of the cells are set by the magnitude and sign of the quantity $d_{ij} = p_{ij} - q_{ij}$, where q_{ij} denotes the probability that residues i and j are in contact in the reference ensemble (native or EV). If $d_{ij} < 0$, then the probability of realizing a contact between residues i and j is lower at the T_D temperatures compared with the reference ensembles, and the converse is true if $d_{ij} > 0$.

The prominent differences between T_D ensembles and the two reference ensembles are the higher probabilities associated with long- and intermediate-range contacts involving hydrophobic residues drawn from four specific groups labeled g_1–g_4, where $g_1 \equiv$ (M1, V3, I4, F5, L6), $g_2 \equiv$ (I18, K19, N20, V21, A22), $g_3 \equiv$ (G24, Y25, A26, N27, N28, F29, L30, F31), and $g_4 \equiv$ (G34, L35, A36, I37). These contacts have low probabilities (≤0.1), are predominantly nonnative (top row in Fig. 7 and *SI Appendix*, Fig. S13), and are either medium-range (contacts between residues from g_1 and g_2, g_2 and g_3, or g_3 and g_4) or long-range (contacts between residues from g_1 and g_3 or g_1 and g_4). *SI Appendix*, Fig. S14 displays a montage of conformations to illustrate how

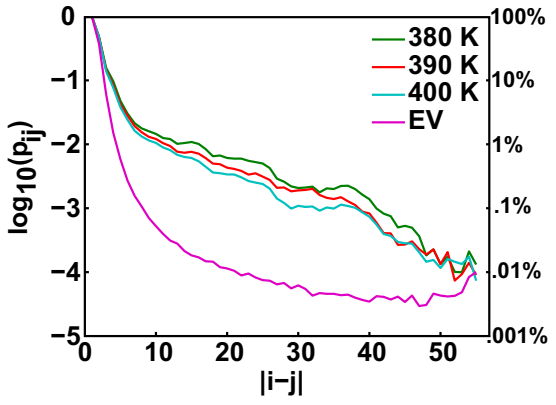

Fig. 5. Comparison of the probabilities associated with short-, intermediate-, and long-range contacts. We calculated the probability p_{ij} of realizing spatial contacts between residues i and j that are |i–j| apart in the linear sequence. The figure shows plots of $\log_{10}(p_{ij})$ plotted vs. linear sequence separation |i–j| for the three T_D temperatures and the EV ensemble. In all cases the p_{ij} values decrease with increasing sequence separation |j–i|, and the p_{ij} values are less than 1% for |j–i| >10.

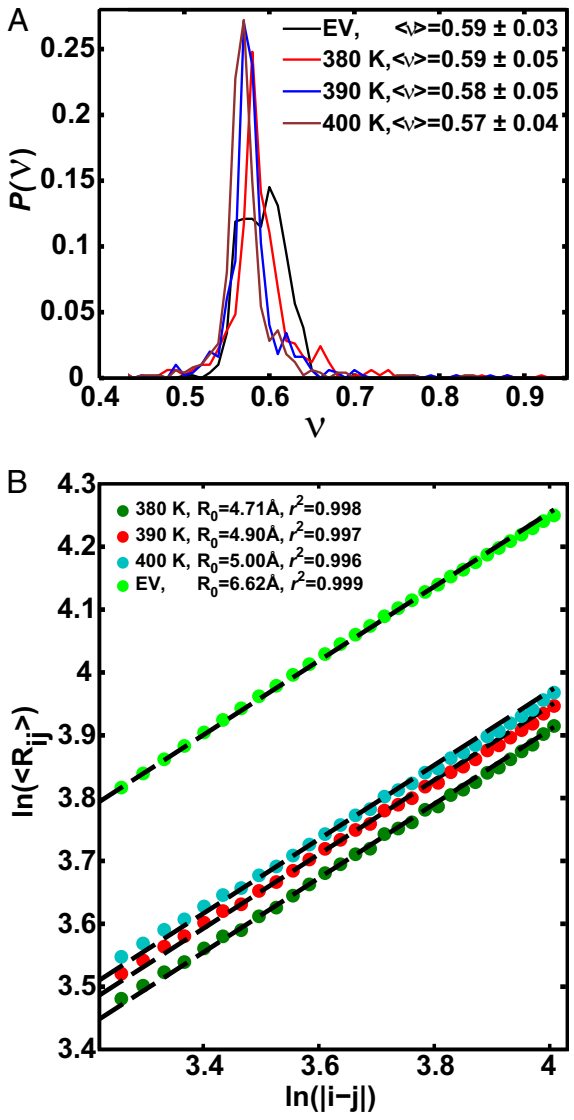

Fig. 6. Assessing the solvent quality of T_D ensembles. (*A*) Estimates of the scaling exponent ν. For each combination of i,j and k,l pairs, we obtained a distinct estimate for ν and the analysis over all combinations of pairs leads to a distribution of values, $P(\nu)$, for ν, which are shown here for each of the three T_D temperatures and the EV limit. In all of the cases the bin widths are 0.01. The constraints on |j–i| and |l–k| were chosen by requiring $P(\nu)$ to be a Gaussian distribution for the EV limit as assessed by the D'Agostino-Pearson omnibus test (52) and a P value of 0.05. (*B*) Estimates of R_o using linear regression to analyze results for the scaling of internal distances with sequence separation by fitting these results to the equation $\ln(\langle R_{ij}\rangle) = \ln(R_o) + 0.59\ln(|j-i|)$. The circles are simulation results for $\ln(\langle R_{ij}\rangle)$, and dashed lines are fits to the data. The legend shows the values obtained for R_o and the coefficient that measures the correlation between each linear fit and the corresponding data set. Our estimates for R_o will be larger than the intercept for the scaling of $\langle R_g\rangle$—it should be approximately $\sqrt{7/6}$ times the intercept for the scaling of internal distances. Multiplying our R_o values by $\sqrt{7/6}$ yields values for $\langle R_g\rangle$ that are similar to the estimate of Kohn et al. (9).

medium/long-range, low-probability contacts can form without requiring either persistent secondary structure or global compaction at T_D temperatures.

Discussion

We have shown that imperfect compensation between non-EV–intrachain and chain–solvent interactions is compatible with $N^{0.59}$ scaling for denatured proteins. This is manifest as detectable, low-probability ($p_{ij} < 0.1$), medium/long-range contacts, which our analysis ascribes to residual intrachain attractions between specific clusters of hydrophobic residues. Interestingly, the residues involved in many of the low-probability nonnative medium/long-range contacts are highly conserved among different NTL9 sequences (*SI Appendix*, Fig. S15). The presence of such contacts may therefore be a conserved feature among members of the NTL9 family. Furthermore, the $\langle R_g\rangle$ values for proteins in the EV limit (16, 17) are larger than those obtained for denatured proteins using SAXS (9). This implies the presence of residual intrachain attractions in most denatured proteins. Indeed, Wu et al. (2) have proposed that clusters of Ile, Leu, Val, and Phe are present in unfolded states because they form early during folding. Weak intrachain attractions can reshape the DSE vis-à-vis the EV limit and do so without altering the $N^{0.59}$ scaling behavior. These interactions might reduce the likelihoods associated with deleterious intermolecular interactions between unfolded proteins that lead to protein aggregation (4), modulate internal friction in the denatured state (49), and reshape barriers to protein folding (50).

Materials and Methods

Experiments and Analysis. Details regarding protein expression, purification, sample preparations, NMR experiments (including experimental settings, chemical shift measurements, pulsed field gradient NMR, and PRE measurements), NMR assignments, data analysis, SAXS measurements, and the calculation of PRE profiles from simulation results are described in detail in *SI Appendix*. Experiments for NMR assignments were conducted using a 500-MHz Bruker spectrometer with a cryoprobe. A Bruker 600 MHz spectrometer equipped with a cryoprobe was used for the pulsed-field gradient NMR diffusion experiments. The ^{1}H-^{15}N HSQC experiments were collected for both the paramagnetic form and diamagnetic form on a Varian Inova 600-MHz spectrometer with a conventional

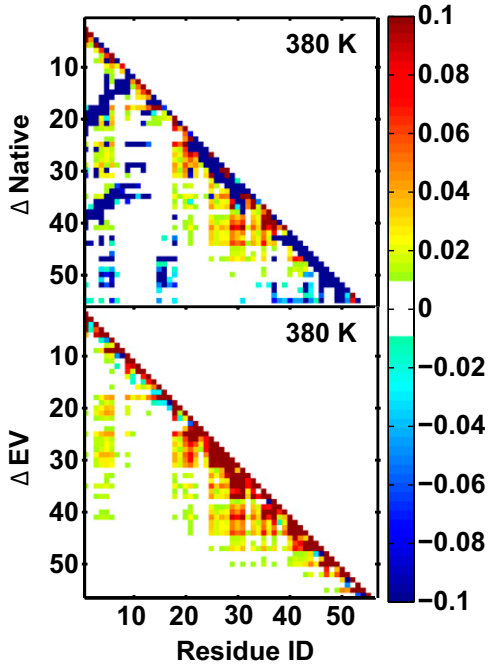

Fig. 7. Difference contact maps for 380 K. (*Upper*) Difference contact maps with respect to the native state ensemble (240 K); (*Lower*) similar maps with respect to the EV ensemble. The cooler colors imply that the contact in question has a higher probability in the reference ensemble (native or EV), whereas the warmer colors imply that the contact has a higher probability at 380 K compared with the reference ensemble. A value of 0 implies that the contacts either have similar probabilities in both the T_D and references ensembles or are missing in both.

probe at 12 °C. The PRE ^1H_N-Γ_2 rates measurement were performed in a two-time-point approach (45) on a Bruker 700 MHz spectrometer with a conventional probe.

MC Simulations. We used the CAMPARI software package (http://campari.sourceforge.net/), the ABSINTH implicit solvation model (46), and the underlying force field paradigm. Parameters were taken from the abs3.2_opls.prm set. The protein was modeled in atomic detail, and the ABSINTH implicit solvation model was used to model solvent-mediated interactions. Details regarding the simulation setup, analysis of conformational ensembles, solution conditions, and its impact on our results are discussed in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Jaehyun Cho and David Eliezer for helpful discussions and Vadim Patsalo for assistance with SAXS measurements. This work was supported by National Science Foundation Grants MCB-0919860 (to D.P.R.) and MCB-1121867 (to R.V.P.). Use of the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886.

1. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11(1):83–93.
2. Wu Y, Kondrashkina E, Kayatekin C, Matthews CR, Bilsel O (2008) Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc Natl Acad Sci USA* 105(36):13367–13372.
3. Dobson CM (2003) Protein folding and misfolding. *Nature* 426(6968):884–890.
4. Jahn TR, Radford SE (2008) Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys* 469(1):100–117.
5. Zhou HX, Rivas GN, Minton AP (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys* 37:375–397.
6. Cho JH, Raleigh DP (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: Effects of denatured state interactions on the analysis of transition state structure. *J Mol Biol* 359(5):1437–1446.
7. Tanford C (1968) Protein denaturation. *Adv Protein Chem* 23:121–282.
8. Wilkins DK, et al. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 38(50):16424–16431.
9. Kohn JE, et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101(34):12491–12496.
10. Schäfer L (1999) *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group* (Springer, Berlin).
11. Hofmann H, et al. (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109(40):16155–16160.
12. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102(37):13099–13104.
13. Bernadó P, et al. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci USA* 102(47):17002–17007.
14. Meier S, Blackledge M, Grzesiek S (2008) Conformational distributions of unfolded polypeptides from novel NMR techniques. *J Chem Phys* 128(5):052204.
15. Salmon L, et al. (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132(24):8407–8418.
16. Tran HT, Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophys J* 91(5):1868–1886.
17. Tran HT, Wang XL, Pappu RV (2005) Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* 44(34):11369–11380.
18. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(18):8183–8188.
19. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* 127(2):476–477.
20. Ding F, Jha RK, Dokholyan NV (2005) Scaling behavior and structure of denatured proteins. *Structure* 13(7):1047–1054.
21. Zhou HX (2004) Polymer models of protein stability, folding, and interactions. *Biochemistry* 43(8):2141–2154.
22. Rubinstein M, Colby RH (2003) *Polymer Physics* (Oxford Univ Press, New York).
23. Allison JR, Varnai P, Dobson CM, Vendruscolo M (2009) Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J Am Chem Soc* 131(51):18314–18326.
24. Francis CJ, Lindorff-Larsen K, Best RB, Vendruscolo M (2006) Characterization of the residual structure in the unfolded state of the Delta131Delta fragment of staphylococcal nuclease. *Proteins* 65(1):145–152.
25. Kristjansdottir S, et al. (2005) Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J Mol Biol* 347(5):1053–1062.
26. Lindorff-Larsen K, et al. (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J Am Chem Soc* 126(10):3291–3299.
27. Robustelli P, Kohlhoff K, Cavalli A, Vendruscolo M (2010) Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* 18(8):923–933.
28. Fieber W, Kristjansdottir S, Poulsen FM (2004) Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J Mol Biol* 339(5):1191–1199.
29. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR (2003) The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J Mol Biol* 333(5):977–991.
30. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA* 105(17):6278–6283.
31. Mohana-Borges R, Goto NK, Kroon GJA, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J Mol Biol* 340(5):1131–1142.
32. Marsh JA, et al. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J Mol Biol* 367(5):1494–1510.
33. Klein-Seetharaman J, et al. (2002) Long-range interactions within a nonnative protein. *Science* 295(5560):1719–1722.
34. Tran HT, Mao A, Pappu RV (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* 130(23):7380–7392.
35. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc* 132(5):1526–1528.
36. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103(31):11539–11543.
37. Müller-Späth S, et al. (2010) From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(33):14609–14614.
38. Schwarzinger S, Kroon GJA, Foss TR, Wright PE, Dyson HJ (2000) Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. *J Biomol NMR* 18(1):43–48.
39. Marsh JA, Singh VK, Jia ZC, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Sci* 15(12):2795–2804.
40. Lietzow MA, Jamin M, Dyson HJ, Wright PE (2002) Mapping long-range contacts in a highly unfolded protein. *J Mol Biol* 322(4):655–662.
41. Neri D, Billeter M, Wider G, Wüthrich K (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* 257(5076):1559–1563.
42. Schmidt PG, Kuntz ID (1984) Distance measurements in spin-labeled lysozyme. *Biochemistry* 23(18):4261–4266.
43. Gillespie JR, Shortle D (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* 268(1):158–169.
44. Xue Y, et al. (2009) Paramagnetic relaxation enhancements in unfolded proteins: Theory and application to drkN SH3 domain. *Protein Sci* 18(7):1401–1424.
45. Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson* 184(2):185–195.
46. Vitalis A, Pappu RV (2009) ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* 30(5):673–699.
47. Kaminski G, Friesner R, Tirado-Rives J, Jorgensen W (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6487.
48. Ha BY, Thirumalai D (1992) Conformations of a polyelectrolyte chain. *Phys Rev A* 46(6):R3012–R3015.
49. Cellmer T, Henry ER, Hofrichter J, Eaton WA (2008) Measuring internal friction of an ultrafast-folding protein. *Proc Natl Acad Sci USA* 105(47):18320–18325.
50. Clementi C, Plotkin SS (2004) The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci* 13(7):1750–1766.
51. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637.
52. Breton MD, Devore MD, Brown DE (2008) A tool for systematically comparing the power of tests for normality. *J Statist Comput Simulation* 78(7):623–638.