Temperature-dependent solvation modulates the dimensions of disordered proteins

René Wuttke[a], Hagen Hofmann[a], Daniel Nettels[a], Madeleine B. Borgia[a], Jeetain Mittal[b], Robert B. Best[c,1], and Benjamin Schuler[a,1]

[a]Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland; [b]Department of Chemical Engineering, Lehigh University, Bethlehem, PA 18015; and [c]Laboratory of Chemical Physics, National Institute of Digestive and Diabetes and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

Edited by José N. Onuchic, Rice University, Houston, TX, and approved February 25, 2014 (received for review July 10, 2013)

For disordered proteins, the dimensions of the chain are an important property that is sensitive to environmental conditions. We have used single-molecule Förster resonance energy transfer to probe the temperature-induced chain collapse of five unfolded or intrinsically disordered proteins. Because this behavior is sensitive to the details of intrachain and chain–solvent interactions, the collapse allows us to probe the physical interactions governing the dimensions of disordered proteins. We find that each of the proteins undergoes a collapse with increasing temperature, with the most hydrophobic one, λ-repressor, undergoing a reexpansion at the highest temperatures. Although such a collapse might be expected due to the temperature dependence of the classical "hydrophobic effect," remarkably we find that the largest collapse occurs for the most hydrophilic, charged sequences. Using a combination of theory and simulation, we show that this result can be rationalized in terms of the temperature-dependent solvation free energies of the constituent amino acids, with the solvation properties of the most hydrophilic residues playing a large part in determining the collapse.

Sanchez theory | ABSINTH | HIV integrase | cold shock protein | prothymosin α

The properties of unfolded proteins have recently attracted renewed interest (1), triggered in particular by the realization that a large fraction of naturally occurring polypeptides are unstructured under physiological conditions (2, 3). Some of them fold into well-defined structures upon interaction with a ligand or binding partner, whereas others may remain unstructured under all conditions. Many of these "intrinsically disordered proteins" (IDPs) are involved in cellular signaling networks and are thus of great medical interest (4). Given the presence of varying degrees of disorder in unbound and bound states (5), a general framework for the description of the physicochemical properties of IDPs will aid our understanding of the molecular mechanisms underlying their function. Such a framework is beginning to emerge from recent work in which concepts from polymer physics have been found to capture very successfully key aspects of the global conformational and dynamic properties of IDPs and unfolded proteins in general (6). These include the role of charge interactions (7, 8), protein–solvent interactions (9–13), scaling laws (14–16), reconfiguration dynamics (17), and the effect of internal friction (18–21).

An aspect that is less well understood is the effect of temperature on unfolded and intrinsically disordered proteins. Recent single-molecule Förster resonance energy transfer (FRET) experiments showed that the small cold shock protein from *Thermotoga maritima* (Csp*Tm*) and the IDP prothymosin α (ProTα) become more compact with increasing temperature (22), even after the effect of denaturant present in solution (23) is taken into account. The results were in good agreement with dynamic light-scattering experiments on unlabeled protein, demonstrating that the effect is independent of the presence of the fluorophores (22). This result is in line with previous laser temperature jump experiments on acid-denatured BBL protein (24) and recent light- and small-angle X-ray–scattering results on the disordered N-terminal part of p53 (25) and several other IDPs (26). All of these observations are in contrast to the behavior expected for a polymer chain with a temperature-independent monomer–monomer interaction energy, which will expand with increasing temperature owing to the increasing entropic contribution to the free energy (27, 28). The observation of temperature-induced collapse in proteins thus implies the existence of temperature-dependent interactions, presumably with contributions from the hydrophobic effect (29–33) or, more generally, changes in solvation free energy as a function of temperature. A critical role for the solvent contribution is supported by molecular simulations of unfolded proteins with different water models (22, 34), and even simulations of hydrophobic homopolymers (35–39) and simple heteropolymers (40) in explicit water models exhibit a similar behavior. However, the detailed origin of the temperature-induced compaction has remained elusive. To address this question, two key ingredients are required: a larger dataset from experiments on different proteins that enables us to probe the effect of sequence composition more systematically (and in the absence of effects from denaturants), and a simulation model that provides realistic chain dimensions and allows us to investigate the role of solvation free energies.

Here, we investigate the generality and origin of the temperature-induced collapse of unfolded polypeptides by studying five natural proteins with very different sequence compositions, ranging from

Significance

A large number of naturally occurring proteins are now known to be unstructured under physiological conditions. Many of these intrinsically disordered proteins (IDPs) bind other biological macromolecules or ligands and are involved in important regulatory processes in the cell. For understanding the structural basis of these functional properties, it is essential to quantify the balance of interactions that modulate the heterogeneous conformational distributions of IDPs and unfolded proteins in general. In contrast to the behavior expected for simple polymers with temperature-independent intramolecular interactions, unfolded proteins become more compact when the temperature is raised. Here, we show that the temperature dependence of the interactions of the constituent amino acid residues with the aqueous solvent have a dominant effect on this behavior.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: schuler@bioc.uzh.ch or robertbe@helix.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1313006111/-/DCSupplemental.

very hydrophobic foldable sequences to very hydrophilic IDPs. A key feature of our study is the choice of proteins with sufficiently low conformational stability to ensure that the properties of the unfolded state can be observed directly under near-physiological conditions, without using additional denaturants (22) or nonneutral pH (24). Using single-molecule FRET experiments, we are able to specifically monitor the dimensions of the unfolded state as a function of temperature. For all of the sequences, we observe a decrease in chain dimensions with increasing temperature. Remarkably, the largest amplitude of collapse is observed for the most hydrophilic, charged sequences. We use a combination of polymer theory, empirical solvation free-energy data, and implicit solvent simulations to rationalize the observations in terms of the different properties of the underlying sequences. We show that the differences in collapse can be correlated with average solvation free energies of the residues, and that by parameterizing an implicit solvent model using these solvation free energies, we can semiquantitatively reproduce both the absolute radii of gyration and the extent of collapse with temperature. The unexpected result from our work is that, even though the classical hydrophobic effect undoubtedly plays a role in our observations, we find that large variations in solvation free energy with temperature for polar and charged residues also play a very important role. This has significant implications for the properties of IDPs, which are enriched in these residue types.

Results and Discussion

To probe the temperature-dependent protein collapse, we used single-molecule FRET, a technique that has recently been used very successfully for investigating the distance distributions and dynamics of unfolded proteins and IDPs (41–43). A particular advantage of this approach is the ability to separate subpopulations in heterogeneous systems. In our case, this means that we can quantify the properties of the unfolded subpopulation even in the presence of a majority of folded molecules (Fig. 1). Each protein was labeled with Alexa 488 and Alexa 594 as FRET donor and acceptor fluorophores, respectively, via maleimide derivatives that react with cysteine residues introduced by site-directed mutagenesis (*SI Materials and Methods*). Single-molecule observations were made on freely diffusing molecules in a confocal instrument with accurate temperature control of the sample (22).

An example of the results of such an experiment is shown in Fig. 1 for the helical N-terminal λ-repressor domain, a popular model system for protein-folding studies (44–46). At low temperature, two relevant populations are observed: a folded subpopulation with a mean transfer efficiency $\langle E_F \rangle$ of ∼0.9 and an unfolded subpopulation with a mean transfer efficiency $\langle E_U \rangle$ between ∼0.6 and ∼0.8; a third population at $E \sim 0$ originates from molecules that do not contain an active acceptor chromophore. With increasing temperature, two effects are observed: the folded population decreases, and the population of unfolded molecules increases accordingly, as expected from the typical temperature-induced unfolding of proteins; at 319 K and above, all molecules are unfolded. More importantly for the present study, however, the mean transfer efficiency of the unfolded state changes with temperature (in contrast to the folded state, whose transfer efficiency remains constant). Up to 319 K, $\langle E_U \rangle$ increases, corresponding to a compaction of the unfolded molecules. Above 319 K, no further increase in $\langle E_U \rangle$ is observed; in fact, a slight decrease in $\langle E_U \rangle$ even suggests a reexpansion of unfolded λ-repressor. To exclude an effect on our analysis from the incomplete separation of unfolded and folded subpopulations between ∼300 and ∼320 K, we additionally used recurrence analysis (47), a method that allows us to enrich individual subpopulations in a model-free manner, and obtained very similar results (Fig. 1*B*). Fig. 2*A* shows the resulting values of $\langle E_U \rangle$ as a function of temperature.

To determine whether the temperature dependence of the unfolded state dimensions is a general phenomenon, and to relate its characteristics to the amino acid composition of the chain,

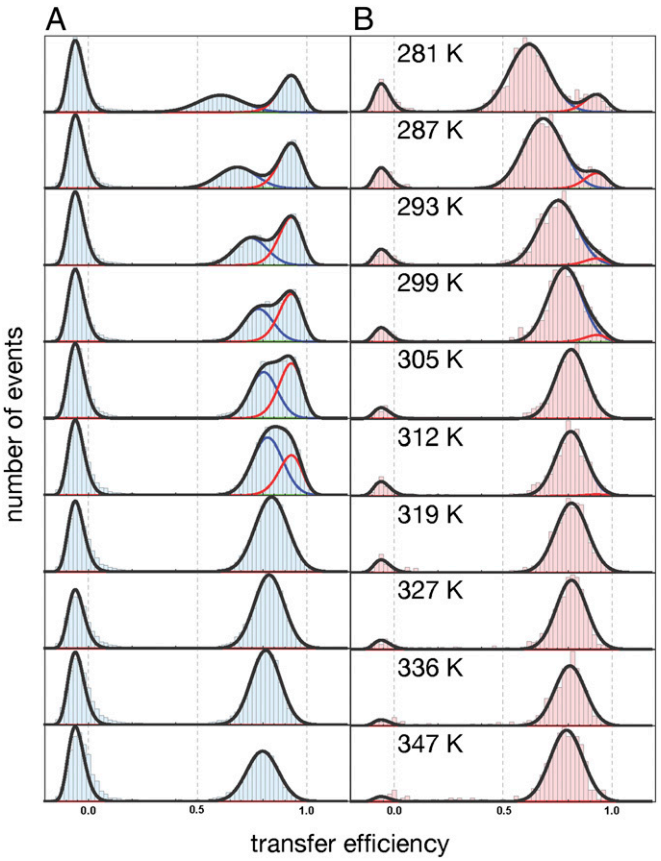

Fig. 1. Transfer efficiency histograms obtained in confocal single-molecule FRET experiments on λ-repressor as a function of temperature (representative dataset). (*A*) Histograms were analyzed in terms of three populations, with the fits indicated as solid lines (red: native; blue: unfolded; black: sum of all populations). The population at $E \sim 0$ corresponds to molecules without an active acceptor chromophore. (*B*) The same data were processed with recurrence analysis (47) before fitting the subpopulations to enrich the population of the unfolded state and minimize uncertainty from peak overlap.

we extended the study to other proteins (Table 1 and *SI Materials and Methods*): two variants of the highly charged IDP ProTα, which allow us to probe the N- and C-terminal halves of the polypeptide (ProTαN and ProTαC, respectively), whose charge content is very different (8); the N-terminal domain of HIV integrase (8), an IDP in which the folded structure is formed upon binding of Zn^{2+} (48); and a 34-aa fragment of Csp*Tm*, which is not folding competent (18) (CspM34). The average hydrophobicities of the sequences according to the Kyte–Doolittle score (49) are −2.44 for ProTαC, −1.5 for ProTαN, −0.6 for integrase and CspM34, and −0.25 for λ-repressor. In all cases, we can investigate the unfolded state under near-physiological conditions in the absence of denaturants, in contrast to the previous experiments, where extrapolation to zero denaturant was required (22). Even though these proteins vary considerably in amino acid composition, average hydrophobicity, and charge distribution, they all exhibit an increase in $\langle E_U \rangle$ with increasing temperature (Fig. 2*A*), corresponding to a compaction of the unfolded state.

For a quantitative analysis of the measured values of $\langle E_U \rangle$ in terms of distance distributions in the unfolded state and the magnitude of intramolecular interactions, we used the mean-field theory of Sanchez (27), as first applied to unfolded proteins by Haran and co-workers (11, 12), and with the radius of gyration at the θ state of 0.22 nm $N^{1/2}$ as a reference point (where N is the number of peptide bonds in the chain segment probed), as determined by Hofmann et al. (14) (see *SI Materials and Methods*

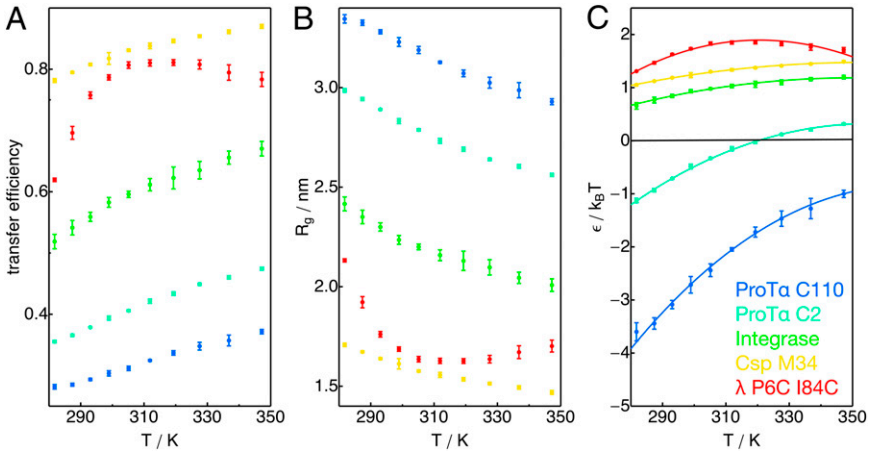

Fig. 2. Average transfer efficiencies of the unfolded proteins (*A*), their radii of gyration (*B*), and the effective intrachain interaction energies (*C*), as determined from the transfer efficiencies using Sanchez theory, as a function of temperature. Uncertainties in *C* are SDs estimated from two to three independent measurements. Fits (solid lines) are used to extract enthalpic and entropic components of ε, whose temperature dependences are determined by a heat capacity term (Table 1). Note that $\varepsilon > 0$ corresponds to attractive intrachain interactions, resulting in chains more compact than an excluded volume chain without additional interactions ($\varepsilon = 0$), and $\varepsilon < 0$ corresponds to repulsive intrachain interactions, resulting in chains more expanded than an excluded volume chain.

for details). Briefly, the theory treats the dye-to-dye distance distribution in terms of a Flory–Fisk distribution weighted by a Boltzmann factor whose value depends on the average effective interaction free energy, ε, between the monomers. A variation of ε as a function of solution conditions or temperature can then be used to account for the observed continuous changes in chain dimensions (11, 12, 14, 27).

Fig. 2 shows the resulting values for the average radii of gyration, R_g, and ε for all protein variants investigated as a function of temperature. All unfolded proteins exhibit temperature-induced collapse. Only λ-repressor shows a slight reexpansion of the chain at high temperature; all other proteins show a monotonic compaction. Correspondingly, the average interaction energy between the amino acids in the chain becomes more favorable with increasing temperature, but with pronounced curvature. For λ-repressor, the temperature dependence of ε exhibits a maximum at ∼320 K. This type of turnover in the temperature dependence is reminiscent of the hydrophobic effect (30, 50–55), suggesting that it plays an important role in the compaction of the chain. However, the pronounced collapse and large change in ε of the IDPs, in particular the extremely hydrophilic ProTα variants, indicates that the collapse does not arise from the classical hydrophobic effect alone. We can describe the temperature dependence by treating $-\varepsilon$ as a free energy of interaction with enthalpic and entropic contributions, whose temperature dependences are determined by a heat capacity term (56) (Fig. 2*C* and Table 1). This analysis shows that the monomer association leading to chain collapse is favored by entropy, as expected for the classical hydrophobic effect. However, there is also a large unfavorable enthalpy of association for the hydrophilic sequences, which is not expected for hydrophobic solutes. The decreasing solvation free energy of the chain with increasing temperature, which collapses the unfolded state, might also be expected to stabilize the folded protein. Indeed, the resultant destabilization as the temperature is lowered results in the "cold unfolding," which can be observed for certain proteins (35, 37, 38, 57–60). However, at higher temperatures, unfolding is driven by the large increase in chain entropy on unfolding (61); at these temperatures, the unfolded chain may nonetheless continue to collapse (driven by unfavorable solvation free energy) because the variation in configurational entropy for a reduction in chain dimensions is much smaller than that for folding.

Table 1. Thermodynamic parameters describing the interaction free energy $-\varepsilon$ between monomers

Protein	$\Delta H_0/k_BT$	$\Delta S_0/k_B$	$\Delta C_p/k_B$	T_M/K
ProTαC	19.6	0.056	−0.27	368.5 ± 3.6
ProTαN	9.8	0.031	−0.18	355.7 ± 1.6
Integrase	2.4	0.011	−0.07	350.8 ± 4.8
Csp M34	1.4	0.09	−0.05	354.9 ± 3.8
λ-Repressor	3.5	0.017	−0.24	320.5 ± 0.8

Data in Fig. 2*C* were fitted to a thermodynamic model, $-\varepsilon(T) = \Delta H_0 + \Delta C_p(T - T_0) - T(\Delta S_0 + \Delta C_p \log(T/T_0))$, where ΔH_0 and ΔS_0 are the enthalpic and entropic contributions to the collapse process, respectively, at the reference temperature T_0 = 298 K, and assuming a temperature-independent heat capacity ΔC_p. T_M is the temperature where $\varepsilon(T)$ is a maximum.

Can we rationalize the temperature-dependent ε in terms of the polypeptide sequence composition? The interaction free energy ε between two isolated residues can in general be divided into the direct interaction of those residues in the gas phase and a solvation free energy resulting from transferring the residues to water. The contribution from the solvation free energy to ε is the difference between the solvation free energy of the associated and dissociated residues, assuming the chain is sufficiently expanded that many-body effects can be neglected.

We explore this aspect by first considering the empirical solvation free energies for amino acid analogs and for the peptide group (62, 63), which fall approximately into five classes with different temperature dependences (Fig. 3*A*; see *SI Materials and Methods* for details), with the hydration of the aliphatic side chains being unfavorable and hydration of other residues being favorable. Interestingly, the hydrophobic aliphatic amino acids all exhibit a turnover in free energy because at low temperature the solvation entropy is unfavorable, but it becomes less so with increasing temperature. However, for almost all of the other amino acids, the solvation free energy becomes monotonically less favorable with increasing temperature, and the amplitude of this change is most pronounced for the most hydrophilic amino acids. These trends are clearly reminiscent of the differences in the temperature dependences of ε observed for the proteins with different mean hydrophobicity (Fig. 2*C*). In this connection, we note that, as all of the amino acid side chains have a positive solvation heat capacity (64), the sign of the heat capacity for

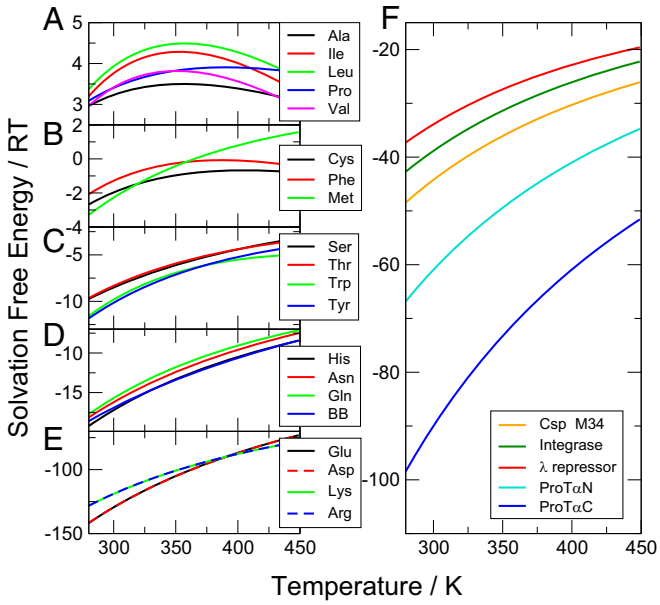

Fig. 3. Solvation free energies derived from small molecules. (*A–E*) Group solvation free energies of each amino acid side chain, and of the peptide backbone (BB), grouped by magnitude of solvation free energy. (*F*) Average residue solvation energies (Eq. S7) for each protein sequence.

contact formation is expected to be negative, as also observed in all cases (Table 1). This confirms that the sign of the heat capacity alone need not be an indication of hydrophobic effects (65).

To connect these individual residue solvation free energies to the effective intrachain interactions as determined from Sanchez theory, it would be necessary to know also the solvation free energy of the associated residue pairs. In a first approach, we approximate this effect by assuming that the solvation free energy of a buried residue is reduced (in magnitude) in proportion to the volume excluded by its neighbors (66) [as is done in the ABSINTH (67) and EEF1 models (68)]. Therefore, the change in solvation free energy upon forming a single residue–residue contact should be approximately proportional to the sum of the solvation free energies of the corresponding isolated residues, assuming each to exclude a similar volume from the other upon contact, and neglecting many-body effects. Following this reasoning, a simple approximation is that the solvation contribution to ε is proportional to the average solvation free energy per residue ΔG_{res}. This approach accounts for sequence composition, but not sequence order. Remarkably, the resulting temperature dependencies of the average residue solvation free energies (Fig. 3*B*) already resemble some of the key aspects observed in the temperature dependencies of the mean-field interaction energies fitted to the experimental data (Fig. 2*C*), in particular the pronounced curvature, the larger slope for more charged sequences, and the approximate rank order of the protein variants (only the adjacent Csp M34 and HIV integrase are switched).

Despite the qualitative success of this approach, it does not capture the maximum in ε seen for λ-repressor, and the large differences in the amplitudes of the change in ε with temperature for the different sequences seen in experiment (Fig. 2*C*). To make a closer connection between empirical solvation free energies and chain dimensions, we have used molecular simulations with the ABSINTH force field (67), thus capturing the effects of chain connectivity and sequence correlations, many-body solvation effects, as well as an explicit model of electrostatic interactions. The ABSINTH energy function includes an implicit solvent model in which the short-range contribution to the solvation free energy, W_{solv}, is written as a sum over group contributions, i.e., $W_{solv} = \sum_{i=1}^{N_{SG}} f_i \Delta G_i^{solv}$, where the sum runs over N_{SG} "solvation groups", each with solvation free energy ΔG_i^{solv} when fully solvent-exposed, and where f_i is the degree of solvent exposure of residue i as defined in ref. 67. The motivation for this expression is that the solvation free energy of a group of atoms will be approximately reduced in proportion to the volume excluded by neighboring residues. The solvation groups are subsets of atoms in each residue, which may be identified with small model compounds, and the solvation free energies are taken unmodified (for the most part) from experimental data for these compounds. We used the standard ABSINTH model, but with two changes: (*i*) we computed temperature-dependent solvation free energies using published thermodynamic data, as described in *SI Materials and Methods*, and (*ii*) we additionally considered the effect of the temperature dependence of the dielectric constant. In principle, explicit solvent simulations should be the most accurate method for treating unfolded proteins. However, unfolded states in all current force fields are too collapsed, having comparable, or sometimes smaller, radii of gyration than the folded protein (69, 70). However, ABSINTH results in a good correlation with experimental radii of gyration (Fig. 4*A*), as discussed below.

Replica exchange Monte Carlo simulations were run for each of the five proteins using the modified ABSINTH model, over a wide range of temperatures. In addition, we considered two reference models: the original ABSINTH model (temperature-

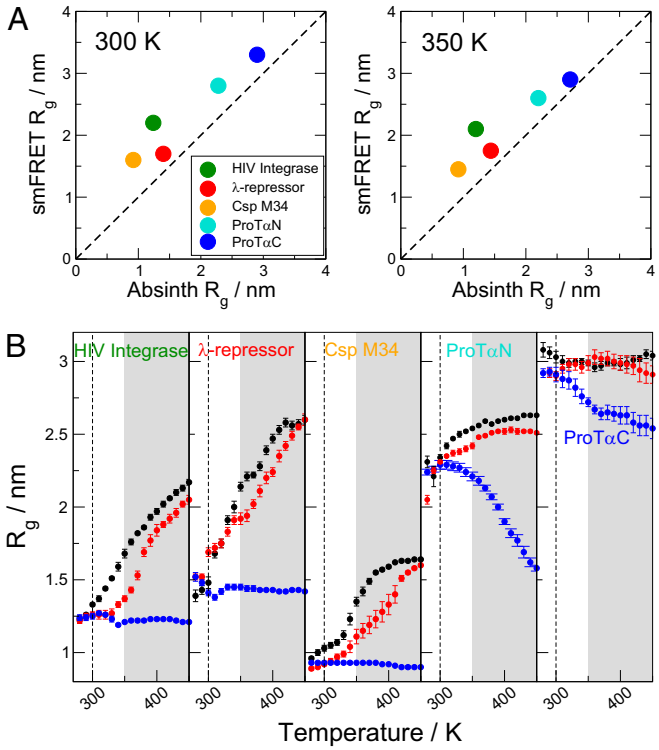

Fig. 4. Results from simulations with the OPLS/ABSINTH force field. (*A*) Correlation between experimental single-molecule FRET (smFRET) estimate of radius of gyration with that calculated from ABSINTH calculations. Linear correlation coefficients are 0.95 and 0.92 at 300 and 350 K, respectively. (*B*) For each protein, the average temperature-dependent radius of gyration is plotted for simulations in which (*i*) the solvation free energies and solvent dielectric constants were fixed to those at 300 K (black symbols); (*ii*) the dielectric constant varied with temperature (red symbols); (*iii*) both the dielectric constant and solvation free energies were made temperature dependent (blue symbols). The unshaded region represents the temperature range probed in experiments.

independent solvation free energies and dielectric constant) and a model in which only the dielectric constant varied with temperature. The radius of gyration of each peptide is shown as a function of temperature in Fig. 4*B*. The original ABSINTH model, although capturing very well the dimensions near 300 K, shows a large expansion with temperature in all cases, as expected from entropic considerations. The less pronounced expansion of the prothymosin variants with increasing temperature is expected in view of the large electrostatic repulsion present in these molecules (8). The modest influence of a temperature-dependent dielectric indicates that this is not an important effect. However, the model including temperature-dependent solvation free energies results in a dramatic shift in observed properties relative to the former models. Instead of rapidly expanding at low temperatures, the radius of gyration initially shows only a weak temperature dependence, or modest collapse for HIV integrase, Csp M34, and λ-repressor, thereby linking the temperature dependence of the solvation free energy with temperature-induced chain collapse. For both ProTαN and ProTαC, there is a marked collapse, reflecting the larger amplitude reduction in radius of gyration for these sequences in experiment. In Fig. 4*A*, we show that the average radii of gyration for each sequence correlate well with the experimental values at both 300 and 350 K when temperature-dependent solvation free energies are used.

Although the change in radius of gyration in the simulations is relatively modest for HIV integrase, Csp M34, and λ-repressor, the two prothymosin fragments appear to collapse monotonically until very high temperature. This difference appears to correlate with overall sequence hydrophobicity and is consistent with the different trends in solvation free energies of hydrophobic and polar groups in Fig. 3*A*. The properties of the more hydrophobic chains, and observation of hydrophobic clusters for some unfolded proteins by NMR (71, 72), suggest that a description similar to a classic hydrophobic collapse mechanism may be appropriate in these cases (30, 31, 73). A second effect that needs to be taken into account is that the amplitudes of chain collapse with temperature will be affected by the sign of the interactions within the polypeptide. For a chain with overall attractive interactions between the monomers, an increase in temperature will favor chain expansion (assuming the interactions are temperature independent). However, for a chain with overall repulsive interactions, as in the case of ProTα (14), an increase in temperature will favor chain compaction, because in the limit of high temperature, only the excluded volume part of the interactions remains important; therefore, the effect of temperature-dependent solvation free energy will be amplified.

In summary, we have shown with a combination of advanced single-molecule methods that temperature-induced collapse is a common feature of five intrinsically disordered or unfolded proteins with very different sequences. The corresponding variation in sequence composition allows us to reveal details of the temperature-dependent interactions within the chains. By analyzing the data with the polymer model of Sanchez, we can interpret intrachain interactions in terms of temperature-dependent free energies and thus link experiment, theory, and simulations. Although the hydrophobic effect is clearly a contributing factor to the collapse, an unexpected finding is the pronounced compaction of the most hydrophilic chains with increasing temperature. Ultimately, it should be possible to address all of these aspects quantitatively in explicit-solvent molecular dynamics simulations. However, only recently are force fields and water models emerging that provide a reliable description of unfolded and disordered proteins (34, 74–78). We have shown that the temperature effects on unfolded state dimensions can be understood at a semiquantitative level by means of a molecular model with implicit solvent by including a temperature-dependent solvation free energy for the constituent amino acid residues. This combination should enable a wide range of simulations for which explicit solvent models are either not accurate enough or prohibitively expensive. Finally, data of the type presented here will be an important benchmark for further improving simulations and our understanding of solvation effects on the structure and dynamics of unfolded and intrinsically disordered proteins.

Materials and Methods

Proteins were expressed recombinantly, purified, and labeled with fluorophores as described in *SI Materials and Methods*. Single-molecule measurements were performed using a MicroTime 200 confocal instrument (PicoQuant). For details of the simulations and their parameterization, see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Andrea Soranno for discussion, Brigitte Buchli for labeled CspM34, Andreas Vitalis for helpful advice on the use of the CAMPARI code, and Rohit Pappu and Andreas Vitalis for discussions on solvation free energies. This work was supported by the Swiss National Science Foundation (B.S.), the Swiss National Center of Competence in Research for Structural Biology (B.S.), a Starting Investigator grant of the European Research Council (to B.S.), the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health (R.B.B.), and National Science Foundation Grant CBET-1120399 (to J.M.). This study used the high-performance computing capabilities of the Biowulf cluster at the National Institutes of Health (R.B.B.) and the Extreme Science and Engineering Discovery Environment, which is supported by National Science Foundation Grant MCB-120014 (to J.M.).

1. Babu MM, Kriwacki RW, Pappu RV (2012) Structural biology. Versatility from protein disorder. *Science* 337(6101):1460–1461.
2. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3):197–208.
3. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18(6):756–764.
4. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annu Rev Biophys* 37:215–246.
5. Marsh JA, Teichmann SA, Forman-Kay JD (2012) Probing the diverse landscape of protein flexibility and binding. *Curr Opin Struct Biol* 22(5):643–650.
6. Mao AH, Lyle N, Pappu RV (2013) Describing sequence-ensemble relationships for intrinsically disordered proteins. *Biochem J* 449(2):307–318.
7. 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.
8. Müller-Späth S, et al. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(33):14609–14614.
9. 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.
10. Haran G (2012) How, when and why proteins collapse: The relation to folding. *Curr Opin Struct Biol* 22(1):14–20.
11. 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.
12. Ziv G, Haran G (2009) Protein folding, protein collapse, and tanford's transfer model: Lessons from single-molecule FRET. *J Am Chem Soc* 131(8):2942–2947.
13. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104(1):105–110.
14. 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.
15. Kohn JE, et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101(34):12491–12496.
16. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci USA* 103(45):16764–16769.
17. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104(8):2655–2660.
18. Soranno A, et al. (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109(44):17800–17806.
19. Borgia A, et al. (2012) Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat Commun* 3:1195.
20. Cheng RR, Hawk AT, Makarov DE (2013) Exploring the role of internal friction in the dynamics of unfolded proteins using simple polymer models. *J Chem Phys* 138(7):074112.

21. Schulz JC, Schmidt L, Best RB, Dzubiella J, Netz RR (2012) Peptide chain dynamics in light and heavy water: Zooming in on internal friction. *J Am Chem Soc* 134(14): 6273–6279.
22. Nettels D, et al. (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci USA* 106(49):20740–20745.
23. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226(2):491–505.
24. Sadqi M, Lapidus LJ, Muñoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100(21):12117–12122.
25. Langridge TD, Tarver MJ, Whitten ST (2013) Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins* 82(4):668–678.
26. Kjaergaard M, et al. (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci* 19:1555–1564.
27. Sanchez IC (1979) Phase-transition behavior of the isolated polymer-chain. *Macromolecules* 12(5):980–988.
28. Sun ST, Nishio I, Swislow G, Tanaka T (1980) The coil-globule transition—radius of gyration of polystyrene in cyclohexane. *J Chem Phys* 73(12):5971–5975.
29. Urbic T, Vlachy V, Kalyuzhnyi YV, Dill KA (2007) Theory for the solvation of nonpolar solutes in water. *J Chem Phys* 127(17):174505.
30. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* 437(7059):640–647.
31. Lum K, Chandler D, Weeks JD (1999) Hydrophobicity at small and large length scales. *J Phys Chem B* 103(22):4570–4577.
32. Hummer G, Garde S, García AE, Pohorille A, Pratt LR (1996) An information theory model of hydrophobic interactions. *Proc Natl Acad Sci USA* 93(17):8951–8955.
33. Lynden-Bell RM, Giovambattista N, Debenedetti PG, Head-Gordon T, Rossky PJ (2011) Hydrogen bond strength and network structure effects on hydration of non-polar molecules. *Phys Chem Chem Phys* 13(7):2748–2757.
34. Best RB, Mittal J (2010) Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J Phys Chem B* 114(46):14916–14923.
35. De Los Rios P, Caldarelli G (2001) Cold and warm swelling of hydrophobic polymers. *Phys Rev E Stat Nonlin Soft Matter Phys* 63(3 Pt 1):031802.
36. Paschek D, Nonn S, Geiger A (2005) Low-temperature and high-pressure induced swelling of a hydrophobic polymer-chain in aqueous solution. *Phys Chem Chem Phys* 7(14):2780–2786.
37. Dias CL, Ala-Nissila T, Karttunen M, Vattulainen I, Grant M (2008) Microscopic mechanism for cold denaturation. *Phys Rev Lett* 100(11):118101.
38. Matysiak S, Debenedetti PG, Rossky PJ (2012) Role of hydrophobic hydration in protein stability: A 3D water-explicit protein model exhibiting cold and heat denaturation. *J Phys Chem B* 116(28):8095–8104.
39. Athawale MV, Goel G, Ghosh T, Truskett TM, Garde S (2007) Effects of lengthscales and attractions on the collapse of hydrophobic polymers in water. *Proc Natl Acad Sci USA* 104(3):733–738.
40. Jamadagni SN, Bosoy C, Garde S (2010) Designing heteropolymers to fold into unique structures via water-mediated interactions. *J Phys Chem B* 114(42):13282–13288.
41. Schuler B, Müller-Späth S, Soranno A, Nettels D (2012) Application of confocal single-molecule FRET to intrinsically disordered proteins. *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, eds Uversky VN, Dunker AK (Springer, New York), Vol 896, pp 21–45.
42. Schuler B, Hofmann H (2013) Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr Opin Struct Biol* 23(1):36–47.
43. Ferreon AC, Moran CR, Gambin Y, Deniz AA (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol* 472:179–204.
44. Yang WY, Larios E, Gruebele M (2003) On the extended beta-conformation propensity of polypeptides at high temperature. *J Am Chem Soc* 125(52):16220–16227.
45. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423(6936):193–197.
46. Huang GS, Oas TG (1995) Submillisecond folding of monomeric lambda repressor. *Proc Natl Acad Sci USA* 92(15):6878–6882.
47. Hoffmann A, et al. (2011) Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: Recurrence analysis of single particles (RASP). *Phys Chem Chem Phys* 13(5):1857–1871.
48. Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* 93(24):13659–13664.
49. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105–132.
50. Baldwin RL (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc Natl Acad Sci USA* 83(21):8069–8072.
51. Schellman JA (1997) Temperature, stability, and the hydrophobic interaction. *Biophys J* 73(6):2960–2964.
52. Privalov PL, Gill SJ (1989) The hydrophobic effect: A reappraisal. *Pure Appl Chem* 61(6):1097–1104.
53. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29(31):7133–7155.
54. Ashbaugh HS, Truskett TM, Debenedetti PG (2002) A simple molecular thermodynamic theory of hydrophobic hydration. *J Chem Phys* 116:2907–2921.
55. Garde S, García AE, Pratt LR, Hummer G (1999) Temperature dependence of the solubility of non-polar gases in water. *Biophys Chem* 78(1-2):21–32.
56. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 20:447–490.
57. Kuhlman B, Raleigh DP (1998) Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H_2O and D_2O. Determination of the thermodynamic parameters, deltaH(o), deltaS(o), and deltaC(o)p and evaluation of solvent isotope effects. *Protein Sci* 7(11):2405–2412.
58. Pastore A, et al. (2007) Unbiased cold denaturation: Low- and high-temperature unfolding of yeast frataxin under physiological conditions. *J Am Chem Soc* 129(17): 5374–5375.
59. Graziano G (2010) On the molecular origin of cold denaturation of globular proteins. *Phys Chem Chem Phys* 12(42):14245–14252.
60. Aznauryan M, Nettels D, Holla A, Hofmann H, Schuler B (2013) Single-molecule spectroscopy of cold denaturation and the temperature-induced collapse of unfolded proteins. *J Am Chem Soc* 135(38):14040–14043.
61. Robertson AD, Murphy KP (1997) Protein structure and the energetics of protein stability. *Chem Rev* 97(5):1251–1268.
62. Wolfenden R, Andersson L, Cullis PM, Southgate CC (1981) Affinities of amino acid side chains for solvent water. *Biochemistry* 20(4):849–855.
63. Radzicka A, Pedersen L, Wolfenden R (1988) Influences of solvent water on protein folding: Free energies of solvation of cis and trans peptides are nearly identical. *Biochemistry* 27(12):4538–4541.
64. Makhatadze GI, Privalov PL (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J Mol Biol* 232(2):639–659.
65. Sedlmeier F, Netz RR (2013) Solvation thermodynamics and heat capacity of polar and charged solutes in water. *J Chem Phys* 138(11):115101.
66. Lazaridis T (1998) Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids. *J Phys Chem* 102:3542–3550.
67. 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.
68. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35(2):133–152.
69. Best RB, Hummer G, Eaton WA (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 110(44):17874–17879.
70. Best RB, Mittal J (2011) Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins* 79(4):1318–1328.
71. Klein-Seetharaman J, et al. (2002) Long-range interactions within a nonnative protein. *Science* 295(5560):1719–1722.
72. Hiller S, Wider G, Imbach LL, Wüthrich K (2008) Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX. *Angew Chem Int Ed Engl* 47(5):977–981.
73. Li ITS, Walker GC (2011) Signature of hydrophobic hydration in a single polymer. *Proc Natl Acad Sci USA* 108(40):16527–16532.
74. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004–9015.
75. Nerenberg PS, Head-Gordon T (2011) Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model proteins. *J Chem Theory Comput* 7:1220–1230.
76. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE (2012) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J Am Chem Soc* 134(8):3787–3791.
77. Knott M, Best RB (2012) A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: Evidence from molecular simulations. *PLoS Comput Biol* 8(7):e1002605.
78. Mittal J, Yoo TH, Georgiou G, Truskett TM (2013) Structural ensemble of an intrinsically disordered polypeptide. *J Phys Chem B* 117(1):118–124.

Supporting Information

Wuttke et al. 10.1073/pnas.1313006111

SI Materials and Methods

Protein Preparation and Labeling. The amino acid sequence of the pseudo wild-type λ-repressor, as found in the Protein Data Bank structure 1LMB, was used as a basis to generate a construct optimized for codon use in *Escherichia coli*, which was expressed in the Novagen vector pET-47b(+), containing a cleavable hexahistidine tag. Cysteine residues were placed at positions 6 and 84 to allow for covalent dye attachment. Sequence synthesis and subcloning were carried out by Celtek Bioscience LLC.

The protein was mainly expressed in inclusion bodies of *E. coli* BL21 cells in LB medium with 1 mM kanamycin at 37 °C overnight. Harvested cells were resuspended in 100 mM Tris buffer, 1 mM EDTA, pH 8, and subjected to disruption in the presence of Complete Protease Inhibitor Mixture (Roche) and benzonase for DNA/RNA digestion. The resulting suspension was mixed with 0.5 vol of 60 mM EDTA, 6% (wt/vol) Triton X-100, and 1.5 M NaCl, pH 8, and vigorously stirred for at least 2 h at 4 °C. The inclusion bodies were pelleted by centrifugation and washed two to four times with the same buffer. The pellet was then dissolved in 20 mM Tris buffer, 0.5 M NaCl, 6 M guanidinium chloride (GdmCl), 20 mM imidazole, and 2 mM β-mercaptoethanol at pH 8 (IMAC buffer). This solution was directly applied to a 5-mL HisTrap HP column (GE Healthcare Biosciences). An imidazole gradient from 20 to 300 mM was used to elute the protein. Protein-containing fractions were pooled and concentrated via ultrafiltration (Amicon Centricon 3,000-Da molecular weight cutoff). To cleave off the histidine tag, the protein was rapidly diluted into the appropriate buffer for cleavage with HRV 3C protease (50 mM Tris buffer, 150 mM NaCl, pH 7.5) and digested for at least 2 h at room temperature. After ultrafiltration and buffer exchange to IMAC buffer, uncleaved protein and protease where removed by another HisTrap chromatography. The flow-through fractions were collected and again concentrated by ultrafiltration.

Just before labeling, the cysteine residues were reduced by a large excess of DTT (150 mM). Monomeric reduced protein was separated from DDT by size exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare Biosciences) in 50 mM sodium phosphate buffer, 6 M GdmCl, and 150 mM NaCl. The correct molecular weight of the unlabeled protein was confirmed by electron spray ionization mass spectrometry. The concentration was determined by UV absorption spectroscopy. The fluorescent dyes, Alexa Fluor 488 C5-malemide (donor) and Alexa Fluor 594 C5-malemide (acceptor; Invitrogen), were each dissolved in dimethyl sulfoxide to a concentration of 20 μg/μL and sonicated for at least 20 min. The protein was incubated with the donor in a 1:0.6 stoichiometric ratio for 2 h at room temperature under nitrogen atmosphere. Then the acceptor dye was added at a ratio of protein to dye of 1:10, and the reaction was continued overnight. Protein and unreacted dye were separated by size exclusion chromatography. Correct labeling was confirmed by electron spray ionization mass spectrometry. All other proteins were produced and labeled essentially as described previously (1, 2) with modifications described in ref. 3 for prothymosin α.

Single-Molecule Measurements. Samples were stored at −80 °C in 8 M GdmCl, and freshly and rapidly diluted in 50 mM sodium phosphate buffer, pH 7.0, with 0.001% Tween 20 and 150 mM β-mercaptoethanol to a final protein concentration of ∼15 pM. The single-molecule measurements were performed using a custom-built temperature-controlled sample holder (4) in a MicroTime 200 confocal single-molecule instrument as described previously (4). The laser power for the measurements was set to 120 μW at 485 nm. To minimize sample evaporation during the measurements, the protein solution of 100 μL was overlaid with 40 μL of mineral oil (Sigma). Data at each temperature were collected for 30 min. At temperatures above 60 °C, samples were frequently exchanged by aliquots from the same stock solution. The sequences of all proteins used are given in Table S1.

Analysis of Single-Molecule Data. Corrections for the different quantum efficiencies, detection efficiencies, signal cross talk, and the effect of temperature on the Förster distance were taken into account (4). Ten thousand to 40,000 fluorescence bursts were obtained in a 30-min interval using a threshold of 50 photons for burst identification and a maximum time between successive photons of <100 μs in a burst. To obtain the mean transfer efficiencies of the unfolded populations represented in the transfer efficiency histograms, we fitted the histograms with normal distributions. Native-state populations and donor-only distributions were fitted with log-normal distributions. The positions, widths, and asymmetries of the latter were fixed to the values obtained at the lowest temperature. In cases in which populations were overlapping, the contribution of the unfolded state to the overall histogram was enriched using recurrence analysis (5). Briefly, the time at which the same-molecule probability drops below 60% was calculated from every measurement. Histograms from bursts of recurring molecules in this time span and in an appropriate transfer efficiency range where most of the folded and donor only population is excluded were generated. These histograms were then fitted as described above. Note that, within the range of temperatures used here, the effect of the change in reconfiguration rate of the chain on the observed FRET efficiency is negligible (4). The calculation of the overall interaction energies from transfer efficiencies was performed essentially as described previously (1, 6) using the mean-field theory of Sanchez (7–9).

Polymer Theory and Analysis. The calculation of the overall interaction energies from transfer efficiencies was performed using a method similar to our previous work (1, 6). We used the mean-field theory of Sanchez (7–9) to obtain an expression for the end-to-end distance distribution $P(r)$ of a self-avoiding chain. The theory gives an expression for the probability density function of the radius of gyration, $P(r_g)$, as a function of the intrachain interaction energy ε and the root-mean-squared radius of gyration of the protein at the Θ-state, $R_{g\Theta}$:

$$P_{Sanchez}(r_g) = P_0(r_g)\exp[(N+1)q(\phi_s,\varepsilon)]. \qquad \textbf{[S1]}$$

The Θ-state is defined as the state of a polymer in which attractive and repulsive forces within the chain and with the solvent balance and the polymer obeys the length scaling of an ideal chain. Eq. **S1** consists of two components: first, a distribution of radii of gyration, r_g, for an ideal chain, $P_0(r_g)$, which is approximated by the Flory–Fisk distribution (10):

$$P_0(r_g) \propto r_g^6 \exp\left(-\frac{7r_g^2}{2R_{g\Theta}^2}\right). \qquad \textbf{[S2]}$$

$R_{g\Theta}$ is set to $R_{g\Theta} = 0.22\text{nm}\ N_{bonds}^{1/2}$ (6), where $N_{bonds} = N + 9$ is the number of peptide bonds between the dyes, taking the fluorophore linkers into account (11). The second component

is a Boltzmann factor, whose value depends on the excess free energy per monomer with respect to the ideal chain:

$$q(\phi_s, \varepsilon) = \frac{1}{2}\varepsilon\phi_s - (1-\phi_s)\ln(1-\phi_s)/\phi_s, \qquad \textbf{[S3]}$$

where ε (in k_BT per residue) is the mean interaction energy between amino acids, and $\phi_s = R_c^3/r_g^3$ is the volume fraction of the chain. $R_c = [3(N+1)\upsilon/4\pi\ ]^{1/3}$ is the radius of gyration of the most compact state, where υ is the weighted mean volume of an amino acid (0.13 nm^3).

To relate the distribution of radii of gyration obtained from the Sanchez theory to the end-to-end distance distribution that is used to describe the single-molecule data, we use a conditional probability distribution of the distance between two random points in a sphere of a given radius of gyration, as suggested by Ziv and Haran (9):

$$P(r|r_g) = \frac{1}{\delta\cdot r_g}\left(3\left(\frac{r}{\delta\cdot r_g}\right)^2 - \frac{9}{4}\left(\frac{r}{\delta\cdot r_g}\right)^3 + \frac{3}{16}\left(\frac{r}{\delta\cdot r_g}\right)^5\right), \qquad \textbf{[S4]}$$

with $\delta = \sqrt{5}$, which was determined from the condition that $<r^2> = 6R_g^2$ at the Θ-point, and $R_g^2 = <r_g^2>$. With the Förster equation, $E(r) = R_0^6/\ (R_0^6 + r^6)$, where R_0 is the Förster distance, and Eqs. **S1** and **S4**, we can describe the mean transfer efficiencies obtained from the subpopulation of unfolded molecules in the transfer efficiency histograms:

$$E = \int_0^L \left(\frac{R_0^6}{R_0^6 + r^6}\right) \int_{R_c}^{L/2} P(r|r_g) P_{Sanchez}(r_g) dr_g dr, \qquad \textbf{[S5]}$$

where L is the contour length of the chain (or chain segment). We solve this equation for ε numerically to obtain the interaction energies in the chain.

Calculation of Temperature-Dependent Amino Acid Solvation Free Energies. Solvation free energies were calculated as a function of temperature for each amino acid side chain and the peptide group. To determine these free energies, we used the solvation free energies at 298 K for these groups in the ABSINTH force field (12), together with enthalpies and heat capacities for the corresponding groups [principally taken from data tabulated by Makhatadze and Privalov (13)], listed in Table S2 (model 2), together with data sources. The only difference from the ABSINTH parameters was that the 30 kcal/mol offset for solvation free energies and enthalpies introduced in the ABSINTH parameterization (12) was removed (i.e., 30 kcal/mol was added to the values in Table S2). This empirical offset is needed in the ABSINTH implementation, but undesirable when computing mean solvation free energies for each sequence.

Parameterization of ABSINTH Enthalpies and Heat Capacities. The original ABSINTH implicit solvent model is only parameterized with solvation free energies at 300 K. However, because the experimental solvation free energies are used directly as parameters, there is a straightforward route to generalize to other temperatures, provided suitable model compound data can be identified. Our extension is mostly described in the main text, so here we focus on the three possibilities considered for deriving the parameters for charged residues.

Model 1. Both enthalpies and heat capacities are taken from Makhatadze and Privalov (13), and scaled by the ratio of the ABSINTH solvation free energy (12) to that of the neutral compound (13). This results in reasonable enthalpies, but heat capacities that are probably too large.

Model 2. Solvation enthalpies were taken from Marcus (14), and 30 kcal/mol was subtracted to match them to the original ABSINTH free energies, which had a 30 kcal/mol offset. (Note that using the solvation free energies and enthalpies of the ionizable groups without the 30 kcal/mol offset results in unfolded configurations that are universally much too compact, as observed in the original ABSINTH reference. We therefore did not pursue such variants further.) Solvation heat capacities were derived from the data of Abraham and Marcus (15).

Model 3. This model uses the same heat capacities as model 2, but with an alternative choice of enthalpies, based on expectations from a variety of charged model compounds.

Implicit Solvent Simulations. Simulations were carried out on each protein using the ABSINTH implicit solvent model with OPLS charges (12), modified as described below. In each case, we simulated the part of the sequence between the two cysteines, excluding the cysteines themselves (Table S1). The ABSINTH model includes a "direct mean-field interaction" in which the contribution to the solvation free energy W_{solv} of a protein chain from short-range interactions is written as a sum over group contributions:

$$W_{solv} = \sum_i^{N_{SG}} f_i \Delta G_i^{solv}. \qquad \textbf{[S6]}$$

This sum runs over N_{SG} "solvation groups," which are localized groups of atoms within the protein. The factor f_i is the fractional solvent accessibility of group i, computed according to a scheme described in ref. 12, whereas ΔG_i^{solv} is a reference free energy attained if the group were fully solvent exposed. These reference free energies are taken from small-molecule model compounds similar in chemical functionality to the solvation group in the protein. For example, formamide represents the peptide backbone and methane the alanine side chain. We have made a simple generalization of ABSINTH by computing temperature-dependent solvation free energies using experimental, or experimentally derived, enthalpies and heat capacities of solvation for the same, or similar, solvation groups. For the most part, enthalpies and heat capacities of the backbone and side-chain groups were taken from Privalov and Makhatadze (16), who derived these values from an analysis of model compounds. The exceptions to this were the data for naphthalene (representing the "hydrophobic" part of tryptophan), the sodium and chloride ions, and the ionizable side chains [the Privalov and Makhatadze data refer to the neutral forms in these cases (16)]. The numerical data used are summarized in Table S2, together with the sources. For the ionizable size chains (Asp, Glu, Arg, Lys), we tested several possible parameters, obtaining similar results in each case (see *Parametrization of ABSINTH Enthalpies and Heat Capacities* and Fig. S1).

In addition to varying the short-range solvation free energy, we also considered the effect of varying the dielectric constant. Reference data for the temperature dependence of the static dielectric of water at 1-bar pressure were taken from ref. 17 and found to fit very well to the following quadratic expression:

$$\epsilon(T) = 252.3 - 0.8074\ T/K + 7.52\times10^{-4}(T/K)^2. \qquad \textbf{[S7]}$$

Replica exchange Monte Carlo (REMC) simulations were carried out using a version of the Campari code (18) modified to allow for Hamiltonian replica exchange between different solvation free-energy models. Simulations were carried out in a spherical cavity of radius 100 Å in the presence of ~50 mM sodium chloride (represented by explicit ions), with between 2×10^7 and 5×10^7 steps per replica per run. In addition, two independent REMC runs were performed for each protein, one starting from a fully extended conformation and one from a randomly generated conformation. Data from the two runs was pooled in the final anal-

ysis, after discarding the first 1×10^7 steps of each as equilibration. REMC simulations were run with 32 replicas, spanning temperatures from 280 to 590 K in 10 K increments, with exchange attempts every 2,000 steps.

1. 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.
2. Soranno A, et al. (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109 (44):17800–17806.
3. Yi S, Brickenden A, Choy WY (2008) A new protocol for high-yield purification of recombinant human prothymosin alpha expressed in *Escherichia coli* for NMR studies. *Protein Expr Purif* 57(1):1–8.
4. Nettels D, et al. (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci USA* 106(49):20740–20745.
5. Hoffmann A, et al. (2011) Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: Recurrence analysis of single particles (RASP). *Phys Chem Chem Phys* 13(5):1857–1871.
6. 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.
7. Sanchez IC (1979) Phase-transition behavior of the isolated polymer-chain. *Macromolecules* 12(5):980–988.
8. 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.
9. Ziv G, Haran G (2009) Protein folding, protein collapse, and tanford's transfer model: Lessons from single-molecule FRET. *J Am Chem Soc* 131(8):2942–2947.
10. Flory PJ, Fisk S (1966) Effect of volume exclusion on dimensions of polymer chains. *J Chem Phys* 44(6):2243–2248.
11. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104(1):105–110.
12. 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.
13. Makhatadze GI, Privalov PL (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J Mol Biol* 232(2):639–659.
14. Marcus Y (1987) The thermodynamics of solvation of ions. *J Chem Soc Faraday Trans I* 83(2):339–349.
15. Abraham MH, Marcus Y (1986) The thermodynamics of solvation of ions. *J Chem Soc Faraday Trans I* 82(10):3255–3274.
16. Privalov PL, Makhatadze GI (1993) Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J Mol Biol* 232(2): 660–679.
17. Fernandez DP, Goodwin ARH, Sengers JMHL (1995) Measurements of the relative permittivity of liquid water at frequencies in the range of 0.1 to 10 kHz and at temperatures between 273.1 and 373.2 K at ambient pressure. *Int J Thermophys* 16 (1):929–955.
18. Vitalis A, Pappu RV (2009) Methods for Monte Carlo simulations of biomacromolecules. *Annu Rep Comput Chem* 5:49–76.

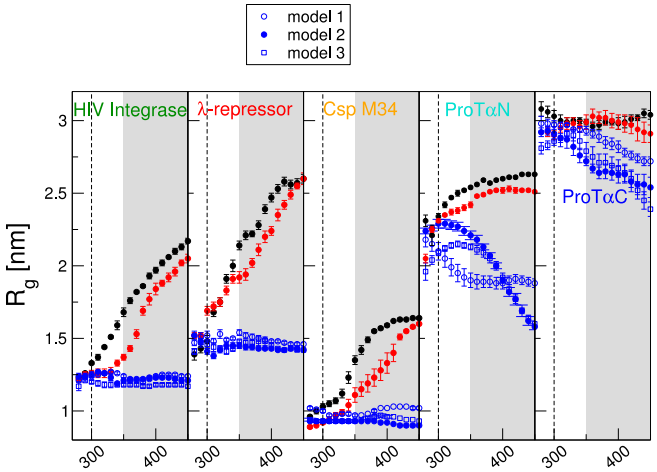

Fig. S1. Temperature-dependent radius of gyration for different ABSINTH solvation models. In addition to the data in Fig. 4 in main text, we show alternative versions of the model with temperature-dependent solvation free energies. Models 1–3 correspond to the alternative parameters listed in Table S2, with model 2 being the one shown in the main text.

Table S1. Sequences and labeling positions of the proteins investigated

Protein	Sequences and labeling positions
λ-Repressor	1 10 20 30 40 50 60 GP**C**LTQEQLE DARRLKAIYE KKKNELGLSQ ESVADKMGMG QSGVGALFNG INALNAYNAA 70 80 LLAKILKVSV EEFSPSIARE **C**R
CspM34	1 10 20 30 **C**EGFKTLKEG QVVEFEIQEG KKGGQAAHVK VVE**C**
IN	1 10 20 30 40 50 60 GSH**C**FLDGID KAQEEHEKYH SNWRAMASDF NLPPVVAKEI VASCDKCQLK GEAMHGQVD**C**
ProTαN (C1–C56)	1 10 20 30 40 50 GP **C**DAAVDTSSE ITTKDLKEKK EVVEEAENGR DAPANGNAEN EENGEQEADN 60 70 80 90 100 110 EVDEE**C**EEGG EEEEEEEEGD GEEEDGDEDE EAESATGKRA AEDDEDDDVD TKKQKTDEDD
ProTαC (C56–C110)	1 10 20 30 40 50 MAHHHHHHS AALEVLFQGP MSDAAVDTSS EITTKDLKEK KEVVEEAENG RDAPANGNAN EENGEQEADN 60 70 80 90 100 110 EVDEE**C**EEGG EEEEEEEEGD GEEEDGDEDE EAESATGKRA AEDDEDDDVD TKKQKTDED**C**

Table S2. Thermodynamic data and sources used to compute temperature-dependent group solvation free energies in ABSINTH

Model no.	Residue/unit	Model compound	ABSINTH keyword	ABSINTH ΔG, kcal/mol	ΔH, kcal/mol	ΔC_p, cal·mol^{-1}·K^{-1}	Source
	Polypeptide backbone	*N*-Methyl acetamide	PEP_BB	−10.1	−14.24 (1)	−15.11	Ref. 1
	Glycine		GLY	0	0.00	0.00	
	Alanine	Methane	ALA	1.9	−1.98 (1)	34.03	Ref. 1
	Valine	Propane	VAL	2	−3.28 (1)	62.12	Ref. 1
	Leucine	2-Methylpropane	LEU	2.3	−4.09 (1)	69.26	Ref. 1
	Isoleucine	Butane	ILE	2.2	−4.09 (1)	74.19	Ref. 1
	Proline	Propane	PRO	2	−2.43 (1)	26.41	Ref. 1
	Methionine (total)	Ethyl methyl thioether	MET	−1.4	−8.27 (1)	20.46	Ref. 1
	Methionine (nonpolar)	Butane	NLE	2.2	−4.09 (1)*	74.19	Ref. 1*
	Serine	Methanol	SER	−5.1	−10.27 (1)	10.49	Ref. 1
	Threonine (total)	Ethanol	THR	−5	−10.77 (1)	29.23	Ref. 1
	Threonine (polar part)	Methanol	SER	−5.1	See Ser above		
	Cysteine	Methanethiol	CYS	−1.2	−5.49 (1)	50.98	Ref. 1
	Asparagine	Acetamide	ASN	−9.7	−16.03 (1)	5.86	Ref. 1
	Glutamine (total)	Propionamide	GLN	−9.4	−16.84 (1)	22.23	Ref. 1
	Glutamine (polar part)	Acetamide	ASN	−9.7	See Asn above		
	Phenylalanine	Toluene	PHE	−0.8	−6.05 (1)	68.07	Ref. 1
	Tyrosine (total)	*p*-Cresol	TYR	−6.1	−13.71 (1)	41.66	Ref. 1
	Tyrosine (nonpolar)	Toluene	PHE	−0.8	See Phe above		
	Tryptophan (total)	3-Methyl indole	TRP	−5.9	−14.05 (1)	80.28	Ref. 1
	Tryptophan (nonpolar part)	Naphthalene	NAP[†]	−2.4	−11.20 (2)	82.07	Ref. 2
	Histidine	4-Methyl imidazole	HIS	−10.3	−16.64 (1)	13.46	Ref. 1
Model 1	Aspartate (−)	Acetic acid	ASP	−107.3[‡]	−110.83[§]	50.37	[§]
	Glutamate (−)	Propionic acid	GLU	−107.3[‡]	−127.50[§]	222.50	[§]
	Lysine (+)	1-Butylamine	LYS	−100.9[‡]	−123.58[§]	459.20	[§]
	Arginine (+)	*n*-Propyl guanidine	ARG	−100.9[‡]	−120.30[§]	204.94	[§]
Model 2	Aspartate (−)	Acetic acid	ASP	−107.3[‡]	−131.60[¶]	5.30	[‖]
	Glutamate (−)	Propionic acid	GLU	−107.3[‡]	−132.60[¶]	21.00	[‖]
	Lysine (+)	1-Butylamine	LYS	−100.9[‡]	−107.00**	54.50	[‖]
	Arginine (+)	*n*-Propyl guanidine	ARG	−100.9[‡]	−107.00**	54.50	[‖]
Model 3	Aspartate (−)	Acetic acid	ASP	−107.3[‡]	−120.00	5.30	[‖]
	Glutamate (−)	Propionic acid	GLU	−107.3[‡]	−120.00	21.00	[‖]
	Lysine (+)	1-Butylamine	LYS	−100.9[‡]	−109.00	54.50	[‖]
	Arginine (+)	*n*-Propyl guanidine	ARG	−100.9[‡]	−109.00	54.50	[‖]
	Sodium (+)	Na^+	NA+	−87.2	−93.45 (3)	−7.17	Ref. 4
	Chloride (−)	Cl^-	CL-	−74.6	−93.69 (3)	−18.40	Ref. 4
	Charged N-ter	Methylamine	PEP_CNT	−106.5[‡]	−275.90[§]	1,025.50	[§]
	Charged C-ter	Acetic acid	PEP_CCT	−107.3[‡]	−181.13[§]	82.30	[§]
	Aspartate (neutral)			−10.95 (5)	−11.31 (1)	5.14	Ref. 1
	Glutamate (neutral)			−10.2 (5)	−12.12 (1)	21.15	Ref. 1
	Lysine (neutral)			−9.52 (5)	−11.66 (1)	43.33	Ref. 1
	Arginine (neutral)			−19.92 (5)	−23.75 (1)	40.46	Ref. 1
	Neutral N-ter	Methylamine		−4.5 (2)	−11.66 (1)[††]	43.33	Ref. 1[††]
	Neutral C-ter	Acetate		−6.7 (2)	−11.31 (1)[††]	5.14	Ref. 1[††]

*Parameters for the nonpolar part of Met were taken from those for ILE in ref. 1, as both are modeled on butane in ABSINTH and should therefore be consistent.

[†]New keyword NAP introduced for naphthalene, which was previously hard-coded.

[‡]These values were adjusted in the calibration process in the original ABSINTH paper.

[§]ΔH and ΔC_p values for charged species obtained by scaling enthalpies and heat capacities for the uncharged species by the ratio of the charged:uncharged ΔG.

[¶]Aspartate and glutamate ΔH taken from acetic acid data from Marcus (6), with 30 kcal/mol offset to match that used for ΔG in ABSINTH.

[‖]ΔC_p values based on data from Abraham and Marcus (4).

**Lysine and arginine ΔH taken from ethyl amine data from Marcus (6), with 30 kcal/mol offset as above and 2 kcal/mol to account for additional methylene groups.

[††]ΔH and ΔC_p for the uncharged N and C termini were taken, respectively, from the ref. 1 values for lysine and aspartate.

1. Makhatadze GI, Privalov PL (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J Mol Biol* 232(2):639–659.
2. Cabani S, Gianni P, Mollica V, Lepori L (1981) Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous-solution. *J Solution Chem* 10(8):563–595.
3. Schmid R, Miah AM, Sapunov VN (2000) A new table of the thermodynamic quantities of ionic hydration: Values and some applications (enthalpy-entropy compensation and Born radii). *Phys Chem Chem Phys* 2(1):97–102.
4. Abraham MH, Marcus Y (1986) The thermodynamics of solvation of ions. 1. The heat-capacity of hydration at 298.15-K. *J Chem Soc Faraday Trans* 1(82):3255–3274.
5. Wolfenden R, Andersson L, Cullis PM, Southgate CC (1981) Affinities of amino acid side chains for solvent water. *Biochemistry* 20(4):849–855.
6. Marcus Y (1987) The Thermodynamics of Solvation of Ions Part 2.–The Enthalpy of Hydration at 298.15 K. *J Chem Soc Faraday Trans* 1(83):339–349.