

Single-molecule spectroscopy of the unexpected collapse of an unfolded protein at low pH

Hagen Hofmann,^{a)} Daniel Nettels, and Benjamin Schuler^{a)}

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

(Received 4 July 2013; accepted 23 August 2013; published online 23 September 2013)

The dimensions of intrinsically disordered and unfolded proteins critically depend on the solution conditions, such as temperature, pH, ionic strength, and osmolyte or denaturant concentration. However, a quantitative understanding of how the complex combination of chain-chain and chain-solvent interactions is affected by the solvent is still missing. Here, we take a step towards this goal by investigating the combined effect of pH and denaturants on the dimensions of an unfolded protein. We use single-molecule fluorescence spectroscopy to extract the dimensions of unfolded cold shock protein (CspTm) in mixtures of the denaturants urea and guanidinium chloride (GdmCl) at neutral and acidic pH. Surprisingly, even though a change in pH from 7 to 2.9 increases the net charge of CspTm from -3.8 to $+10.2$, the radius of gyration of the chain is very similar under both conditions, indicating that protonation of acidic side chains at low pH results in additional hydrophobic interactions. We use a simple shared binding site model that describes the joint effect of urea and GdmCl, together with polyampholyte theory and an ion cloud model that includes the chemical free energy of counterion interactions and side chain protonation, to quantify this effect. © 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4820490>]

INTRODUCTION

Understanding protein folding^{1,2} and the functional properties of intrinsically disordered proteins (IDPs)^{3–5} requires detailed knowledge of the forces that act in unstructured polypeptide chains. These forces determine the dimensions of unfolded and disordered proteins and have been suggested to impact processes such as the coupled binding and folding of IDPs,⁵ or the rate of protein folding reactions.^{6–8} However, the dimensions of unfolded proteins and IDPs also depend critically on solution conditions such as ionic strength,^{4,9} temperature,¹⁰ osmolytes,¹¹ or denaturants,^{2,12–14} implying that the complex interplay between chain-chain and chain-solvent interactions has to be unraveled before quantitative predictions of the size of unstructured polypeptide chains and their effect on folding or binding will be possible.

In contrast to protein folding, the compaction or expansion of unfolded proteins with a change in solvent conditions does not occur in an all-or-none manner, but is a continuous process reminiscent to the coil-to-globule transition in homopolymers.¹⁵ In the past, the continuous collapse of polymers has been successfully treated by mean-field theories^{16–19} that relate the free energy of the chain to its density. Since the first theoretical description of homopolymer collapse by Flory,¹⁶ coil-to-globule theories have been developed that describe the expanded coil in good solvent as well as the compact globule in poor solvent.^{18–21} Experimentally, the observation of unfolded proteins over a broad range of solvent conditions by means of single-molecule Förster resonance energy transfer (FRET) allowed the quantitative application of these

theories to unfolded proteins.^{4,12–14} However, a prerequisite for their application is knowledge of the Θ -state, which serves as a reference state for the coil-to-globule transition. At the Θ -point, chain-chain and chain-solvent interactions balance, and the chain obeys the length scaling of an ideal chain. Based on the length scaling of unfolded proteins obtained from small-angle X-ray scattering (SAXS),²² and the length scaling of folded globular proteins,^{14,23} we recently obtained an approximate expression for the radius of gyration of unfolded proteins at the Θ -point,¹⁴ which allows the calculation of mean-field interaction free energies between the amino acids in unfolded proteins.¹⁴ Although we found a close correspondence of these mean-field free energies with the free energy of solvation of the individual amino acids,^{14,24} the description of unfolded proteins by one mean-field interaction parameter remains incomplete. Proteins are heteropolymers of complex composition, consisting of 20 different types of amino acids. Averaging over all possible types of interactions by extracting one mean-field interaction as done in the past^{14,25} is a crucial first step for understanding the behavior of polypeptide chains but will not be able to provide structural details.

In the past, heteropolymer theories based on replica approaches,²⁶ the random energy approximation (REM) in spin-glass theory,²⁷ and Flory-Huggins theory of mixing^{28,29} have been developed. However, different from the theories of homopolymer collapse, the corresponding heteropolymer theories obviously require more parameters to account for the chemical variability of the monomers in the chain, which complicates their application to experimental data. Here, we take an experimental route to overcome these limitations by using single-molecule FRET to determine the dimensions of unfolded cold shock protein (CspTm) over a wide range of conditions, in mixtures of two denaturants, urea and

^{a)}Authors to whom correspondence should be addressed. Electronic addresses: h.hofmann@bioc.uzh.ch and schuler@bioc.uzh.ch.

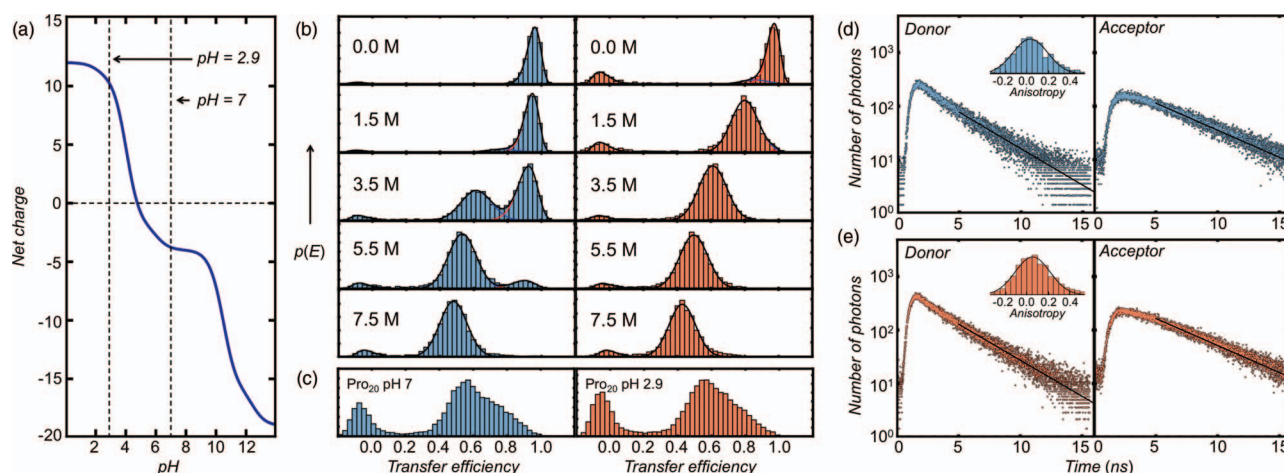


FIG. 1. (a) Net charge of unfolded CspTm as a function of pH calculated from the pK-values of the free amino acids. (b)–(e) Single-molecule fluorescence spectroscopy measurements for CspTm (Csp 66) and Polyproline 20 (Pro20) at pH 7 (blue) and pH 2.9 (red). (b) Single-molecule FRET efficiency histograms of CspTm at different concentrations of urea. Solid lines are fits to a sum of two log-normal and one Gaussian distribution used to determine the mean transfer efficiency, $\langle E \rangle$, of the peaks at intermediate transfer efficiencies, which correspond to the unfolded state. (c) Single-molecule FRET efficiency histograms for Polyproline 20 at pH 7 and pH 2.9 (in both cases $\langle E \rangle = 0.61$ for $E > 0.2$) show that the quantum efficiencies of the fluorophores and thus the FRET process itself are not affected significantly by the change in pH. (d) and (e) Subpopulation-specific time-correlated single photon counting (TCSPC) histograms of donor (left) and acceptor (right) fluorophores for unfolded CspTm at 5.5 M urea at pH 7 (d) and pH 2.9 (e). Only photons arising from unfolded molecules were used for the analysis. The solid line is a single exponential tail fit of the decay resulting in a donor fluorescence lifetime of 3.3 ± 0.1 ns at pH 7 and 3.21 ± 0.01 ns at pH 2.9. The fluorescence lifetimes of the acceptor after donor excitation are found to be 4.29 ± 0.05 ns at pH 7 and 4.30 ± 0.01 ns at pH 2.9. Inset: Subpopulation-specific donor fluorescence anisotropy histograms of unfolded CspTm, resulting in mean anisotropies of 0.040 ± 0.002 at pH 7 and 0.07 ± 0.02 at pH 2.9. Solid lines are fits to a Gaussian distribution.

guanidinium chloride (GdmCl), at two different pH-values (Fig. 1(a)); 258 different conditions in total. This extensive data set allowed us to use polyelectrolyte theory in combination with an empirical denaturant binding model to separate long-range interactions resulting from the electrostatics from short-ranged intra-chain interactions. In contrast to the chemical heterogeneity of proteins, the effect of charges on the size of polymers can be treated in a self-consistent manner,^{30–33} i.e., without additional parameters, and thus serves as a starting point for a more elaborate experimental investigation of the subtle balance of forces that determine the size of unfolded proteins and charged IDPs.³⁴

RESULTS

To probe the dimensions of unfolded CspTm, we attached AlexaFluor 488 as a donor and AlexaFluor 594 as an acceptor fluorophore at the termini of the CspTm-variant Csp66 (for the sequence of the construct, see Ref. 14). Freely diffusing molecules were investigated with confocal single-molecule FRET to obtain transfer efficiency histograms (Fig. 1(b)) at different concentrations of urea, GdmCl, and at two different pH-values (pH 7 and pH 2.9). Up to three peaks are observed in the transfer efficiency histograms: The peak at very high transfer efficiency (E) results from folded molecules, and the peak at $E \approx 0$ results from molecules lacking an active acceptor dye (Fig. 1(b)). We focus on the peak at intermediate transfer efficiencies, which results from unfolded molecules (Fig. 1(b)). The advantage of single-molecule FRET to discriminate between folded and unfolded molecules allows us to observe unfolded molecules even at low concentrations of denaturant, where folded molecules dominate the population (Fig. 1(b)). With increasing concentration of denatu-

rant, the transfer efficiency distribution of unfolded CspTm molecules continuously shifts to lower transfer efficiencies, indicating an increase in the average inter-dye distance, and thus an expansion of the molecules (Fig. 1(b)). This effect has been observed for a broad variety of proteins and peptides^{2,4,8,9,12–14,35} and can be described as a continuous expansion of the chain on transfer to good solvent.²⁵ The interesting question here is how mixtures of denaturants³⁶ and drastic changes in pH will affect the dimensions of the unfolded state. Importantly, using polyproline 20 as a control, we confirmed that the relative quantum yields of the attached fluorophores are not altered at low pH (Fig. 1(c)). In addition, we found the fluorescence lifetimes of donor and acceptor fluorophores unaltered by the pH when attached to CspTm (Figs. 1(d) and 1(e)), indicating that a change of the protonation state of unfolded CspTm has no effect on the photophysical properties of the dyes. Finally, the steady-state anisotropies of the donor attached to CspTm are similarly low at both pH-values, indicating that the rotational averaging of the fluorophore is fast compared to its fluorescence lifetime, thus allowing a quantitative analysis of the determined transfer efficiencies without explicitly taking into account the orientational distribution of donor and acceptor transition dipole moments.³⁷

Calculation of radii of gyration from mean transfer efficiencies

In order to quantify the dimensions of unfolded CspTm, we determined the mean transfer efficiency ($\langle E \rangle$) of the unfolded subpopulation from the transfer efficiency histograms (Fig. 1(a)). Since the average interphoton time during the observation time from a single unfolded CspTm molecule

diffusing through the confocal spot is $\sim 10 \mu\text{s}$, and the inter-dye distance dynamics are in the 100 ns regime,⁶ the value of $\langle E \rangle$ obtained from transfer efficiency histograms results from a complete averaging of the distance distribution of unfolded CspTm. In order to extract average distances from $\langle E \rangle$, we model the distance distribution of unfolded CspTm using the coil-to-globule theory of Sanchez.¹⁸ The advantage of this theory is its ability to describe the dimensions of a chain under all solvent conditions by explicitly taking effects such as excluded volume, intrachain interactions, and multibody interactions into account.^{12,13,18} The theory provides an expression for the probability density function of the radius of gyration r_G as a Boltzmann-weighted Flory-Fisk distribution:^{13,14,18,28}

$$P(r_G, \varepsilon, R_{G\Theta}) = Q^{-1} r_G^6 \exp \left\{ -\frac{7r_G^2}{2R_{G\Theta}^2} + n \left[\frac{1}{2} \varepsilon \phi - \frac{1-\phi}{\phi} \ln(1-\phi) \right] \right\}. \quad (1)$$

Here, $n = N + 1$ is the number of monomers in the chain, N is the number of peptide bonds, ε is the mean-field interaction energy between two monomers in the chain, $\phi = (R_C/r_G)^3$ is the volume fraction of the chain with respect to the most compact state $R_C = [(3/4\pi)v_{aa}n]^{1/3}$, with the average volume of one amino acid being $v_{aa} = 0.13 \text{ nm}^3$, as shown in Ref. 14; Q is a normalization constant. The mean-field interaction energy ε determines the dimensions of the chain. A realistic value for the radius of gyration of the chain at the Θ -state has recently been estimated using a generalized expression for the length-scaling of polymers suggested by Hammouda:³⁸

$$R_{G\Theta} = \sqrt{\frac{2l_p b}{(2\nu + 1)(2\nu + 2)}} N_{bonds}^\nu. \quad (2)$$

In Eq. (2), $b = 0.38 \text{ nm}$ is the distance between two C_α -atoms, $\nu = \frac{1}{2}$, $l_p = 0.4 \text{ nm}$ is the persistence length of unfolded proteins in good solvent obtained from SAXS-results,^{14,22} and $N_{bonds} = N + 9$ takes into account the length of the fluorophore linker.^{14,39} In order to relate the distribution of radii of gyration $P(r_G, \varepsilon, R_{G\Theta})$ to a distance distribution $P(r, \varepsilon, R_{G\Theta})$, which is required to describe the mean transfer efficiencies $\langle E \rangle$, we used the conditional probability function $P(r|r_G)$,¹³ which relates distance distribution of two random points inside a sphere with the radius of the sphere $\delta \cdot r_G$:

$$P(r|r_G) = \frac{1}{\delta \cdot r_G} \left[3 \left(\frac{r}{\delta \cdot r_G} \right)^2 - \frac{9}{4} \left(\frac{r}{\delta \cdot r_G} \right) + \frac{3}{16} \left(\frac{r}{\delta \cdot r_G} \right)^5 \right] \quad 0 \leq r < 2\delta \cdot r_G. \quad (3)$$

Here, $\delta = \sqrt{5}$ was obtained from the condition that $6\langle R_G^2 \rangle = \langle r^2 \rangle$ at the Θ -state. The mean transfer efficiency is

finally given by

$$\begin{aligned} \langle E \rangle &= \int_0^L E(r) P(r, \varepsilon, R_{G\Theta}) \\ &= \int_0^L \frac{R_0^6}{R_0^6 + r^6} \int_{R_C}^{L/2} P(r|r_G) P(r_G, \varepsilon, R_{G\Theta}) dr_G dr, \end{aligned} \quad (4)$$

with the Förster distance of the two fluorophores, $R_0 = 5.4 \text{ nm}^2$, and the mean radius of gyration,

$$R_G = \sqrt{\int_{R_C}^{L/2} r_G^2 P(r_G, \varepsilon, R_{G\Theta}) dr_G}. \quad (5)$$

In Eq. (4), the integration limits for the radius of gyration are given by the radius of gyration of the most compact state (R_C) and the maximal possible radius of gyration, which is given by half the contour length ($L/2$). The mean radii of gyration obtained from the measured transfer efficiencies by Eq. (1)–(5) are accurate within an uncertainty of 6% for chains in good solvent,¹⁴ i.e., for conditions at which $R_G \geq R_{G\Theta}$. The situation is different for chains in poor solvent ($R_G < R_{G\Theta}$). Although R_G will still increase with $N^{1/3}$ for a globule in poor solvent, the end-to-end distance $\langle r^2 \rangle^{1/2}$ will not obey a scaling law with the number of bonds.^{23,40} Correspondingly, the measured transfer efficiency will not be a good measure for the chain size. However, under almost all of our experimental conditions, we found $R_G \geq R_{G\Theta}$, implying that unfolded CspTm is always at or above the Θ -state.

Denaturant binding in mixed denaturants

We calculated the radii of gyration of unfolded CspTm from the measured mean transfer efficiencies for a broad concentration range of mixtures of the two denaturants urea and GdmCl (Figs. 2(a) and 2(b)). At neutral pH, where the net charge of CspTm is -3.8 , R_G of unfolded CspTm increases monotonically with increasing concentrations of urea (c_u) and GdmCl (c_g), indicating the presence of pronounced preferential interactions between denaturant and unfolded CspTm (Fig. 2(a)).⁴¹ The precise type of interactions with which denaturants interact with polypeptide chains is still under debate,^{42–46} and an empirical binding model is commonly used to describe their effect on the free energy of solvation (Δg_{sol}) of the chain⁴⁷ via $\Delta g_{sol} = -m' \ln(1 + K_i c_i)/n$, where m_i' corresponds to the effective number of binding sites for denaturant molecules, and K_i is the apparent equilibrium constant in denaturant i . We note, however, that the assumption of defined binding sites for urea and GdmCl on polypeptide chains is an approximation, since molecular dynamics (MD) simulations have revealed a broad variety of interactions between polypeptide chains, denaturants and water⁴⁶ that might affect the dimensions of unfolded proteins. Correspondingly, our approach of describing protein-denaturant interactions by binding models is largely empirical and does not capture the detailed physics that underlie the effect of urea and GdmCl on polypeptide chains. However, the binding model is a useful concept for describing the expansion of unfolded proteins with increasing concentrations of urea and GdmCl based on an increase in the free energy of solvation, $|\Delta g_{sol}|$, of the

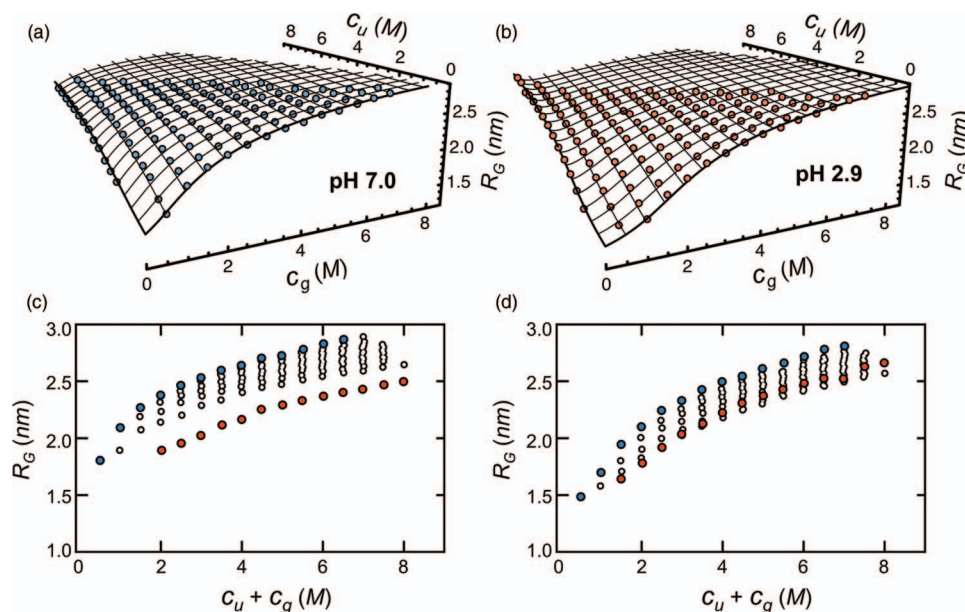


FIG. 2. Mean radii of gyration, R_G , of unfolded CspTm in mixtures of urea (c_u) and GdmCl (c_g) at pH 7 (a) and pH 2.9 (b). Mean radii of gyration of unfolded CspTm in pure urea (red), pure GdmCl (blue), and mixtures of both denaturants (white) at pH 7 (c) and pH 2.9 (d).

single amino acids, which diminishes the mean-field interaction free energy of the chain.⁴⁸

Starting with the simple Flory-Huggins treatment of mixing, and only considering chain-chain (w_{pp}) and chain-denaturant (w_{pd}) interactions, the mean-field interaction free energy is given by $\varepsilon = -w_{pp} + 2w_{pd} = \varepsilon_0 + \Delta g_{sol}$, thus leading to

$$\varepsilon = \varepsilon_0 - \frac{m'_i}{n} \ln(1 + K_i c_i), \quad (6)$$

where ε_0 is the mean-field interaction free energy between two amino acids in water. Indeed, this model has been successfully used to describe the expansion of unfolded protein L with increasing concentrations of GdmCl.¹² Additionally, we showed recently for a variety of proteins that $\varepsilon - \varepsilon_0$ correlates with the free energy of transferring the chain from water to GdmCl.^{14,24} However, in mixtures of urea and GdmCl, the situation is more complex because binding of urea might compete with the binding of GdmCl. If we neglect such competition to a first approximation, i.e., if we assume that urea and GdmCl bind to different sites on the polypeptide chain, the free energy of solvation in a mixture of urea and GdmCl is the sum of the contributions from the two denaturant, i.e., $\Delta g_{sol} = \Delta g_u + \Delta g_g$. It is known that the free energy of solvation in GdmCl is always larger than or equal to that in urea ($|\Delta g_g| \geq |\Delta g_u|$)²⁴ for identical molar concentrations of the two denaturants. Hence, according to Eq. (6), the mean interaction free energies within the chain at a given GdmCl-concentration, ε_g , are expected to be always less than those at the same urea concentration, ε_u , with intermediate values for mixtures of the denaturants, ε_{ug} , i.e., $\varepsilon_g \leq \varepsilon_{ug} \leq \varepsilon_u$. Since R_G decreases with increasing ε , it also follows that $R_{G,u} \leq R_{G,ug} \leq R_{G,g}$. Indeed, at neutral pH, all R_G values of unfolded CspTm in mixtures of the denaturants range between the values observed in pure urea and pure GdmCl (Fig. 2(c)). Thus, it is reasonable to approximate the mean-field interaction en-

ergies by $\varepsilon_{ug} = \varepsilon_0 + \Delta g_u + \Delta g_g$ (Fig. 3(a)), with Δg_u and Δg_g resulting from global fits of the mean-field free energies in pure urea and GdmCl with Eq. (6). The remaining difference is small on average (rmsd = 0.1 kT), but can be as high as 0.25 kT in mixtures of urea and GdmCl with high equimolar concentrations of both denaturants, suggesting a more complex binding model.

In a more realistic model, urea and GdmCl compete for binding sites on the polypeptide chain. Assuming m_u specific binding sites for urea, m_g specific binding sites for GdmCl, and m_{ug} binding sites that can be occupied by both denaturants, the free energy change upon solvation is given by $\Delta g_{sol} = \Delta g_u + \Delta g_g + \Delta g_{ug}$. From the fits of Eq. (6) to the data in the pure denaturants, we obtain $m'_u = m_u + m_{ug}$ total binding sites for urea and $m'_g = m_g + m_{ug}$ total binding sites for GdmCl. With these relations, we finally obtain for ε in mixed denaturants

$$\varepsilon = \varepsilon_0 - \frac{m'_u - m_{ug}}{n} \ln(1 + K_u c_u) - \frac{m'_g - m_{ug}}{n} \times \ln(1 + K_g c_g) - \frac{m_{ug}}{n} \ln(1 + K_u c_u + K_g c_g). \quad (7)$$

Equation (7) contains only one free parameter: the number of binding sites shared between urea and GdmCl (m_{ug}). All other parameters are determined from the data in pure urea and GdmCl, respectively. A fit with Eq. (7) yields a better description of the data⁴⁹ (rmsd = 0.044 kT), with a maximal deviation of 0.1 kT (Fig. 3(c) and Table I). The results suggest that 90 ± 21 binding sites are accessible to the denaturants. Of these, 19 ± 6 (21%) are specific for GdmCl, 27 ± 20 (30%) are specific for urea, and 43 ± 2 (49%) are shared between urea and GdmCl. The large uncertainty in the number of urea binding sites mainly results from the low affinity of urea for the polypeptide chain.⁵⁰ Indeed, the binding constants obtained for the two denaturants ($K_u = 0.16 \pm 0.07 \text{ M}^{-1}$ and $K_g = 0.58 \pm 0.13 \text{ M}^{-1}$) indicate that GdmCl

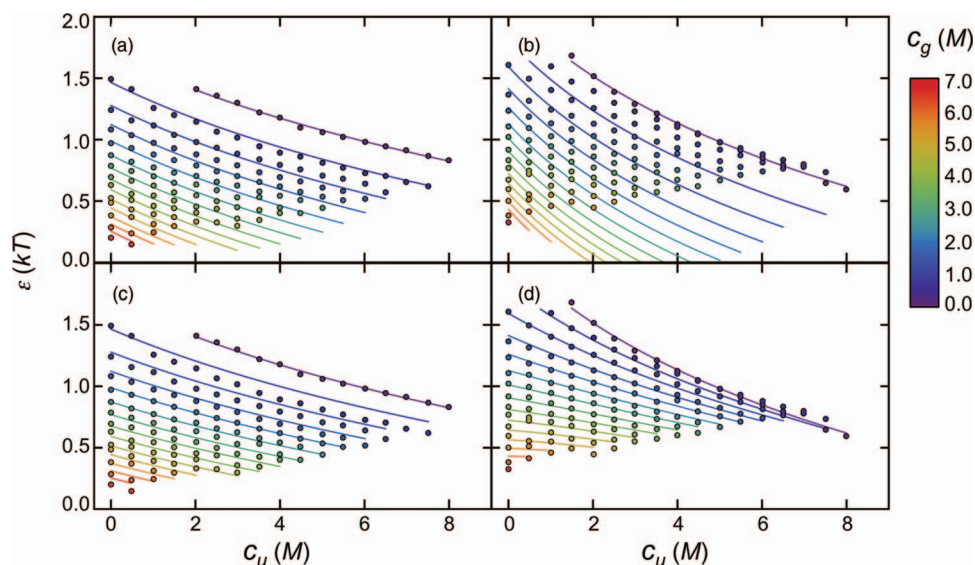


FIG. 3. Mean-field interaction free energies, ε , in the unfolded state as a function of urea for different GdmCl concentrations at pH 7 (left) and pH 2.9 (right). Solid lines are the best fits of the ε values at pH 7 (a) and pH 2.9 (b) with the independent site model for denaturant binding. Fits with the shared binding site model (Eq. (7)) for pH 7 (c) and pH 2.9 (d) provide a much better description of the data. The color indicates the concentration of GdmCl for the data and the fit (see color bar).

interacts with unfolded CspTm with significantly higher affinity than urea. These affinities are close to values obtained from calorimetric measurements,⁵¹ $K_u = 0.06 \pm 0.01 \text{ M}^{-1}$ and $K_g = 0.6 \pm 0.01 \text{ M}^{-1}$, supporting the validity of the analysis.

To investigate the effect of charges in the polypeptide chain on the binding of denaturants to unfolded CspTm, we increased the net charge of CspTm to +10.2 by lowering the pH to 2.9. Importantly, the protonation states of the two denaturants remain largely unchanged at this low pH, e.g., urea is only 0.3% protonated at pH 2.9,⁵² and the spectroscopic properties of the FRET dyes do not change significantly in this pH range (Figs. 1(c)–1(e)). As observed at neutral pH, the R_G at pH 2.9 increases monotonically with increasing concentration of urea and GdmCl (Fig. 2(b)), indicating a continuous expansion of unfolded CspTm driven by interactions of the two denaturants with the polypeptide, with chain dimensions comparable to those obtained at pH 7. This result is surprising, since recent experiments^{4,9,14} and Monte-Carlo simulations⁵³ suggest that the dimensions of unfolded proteins increase with increasing net charge of the chain; according to this effect, the high net charge of unfolded CspTm at pH 2.9 should lead to a pronounced expansion of the chain at low ionic strength. Our results imply that a change in the protonation state of the polypeptide by lowering the pH does not only increase the net charge, but must also alter other properties of the polypeptide chain. Additionally, in contrast to pH 7, the condition $R_{G,u} \leq R_{G,ug} \leq R_{G,g}$ is not met at pH 2.9 under all conditions (Fig. 2(d)). Instead, the R_G values above 4 M urea are

slightly smaller in the presence of 0.5 M GdmCl salt than in its absence, an effect most likely originating from a screening of charge interactions in the polypeptide chain. Consequently, the independent site model for denaturant binding fails to describe the data at pH 2.9 (Fig. 3(b)). In contrast, the shared binding site model (Eq. (7)) provides a good fit of the data (Fig. 3(d)), however, with unphysical binding parameters (Table I). For example, negative values for the number of urea and GdmCl binding sites are obtained ($m_u = -17 \pm 13$, $m_g = -26 \pm 7$), indicating that this model also fails to describe the chain-denaturant interactions at pH 2.9.

In summary, the chain dimensions of unfolded CspTm at low pH are difficult to reconcile with our current understanding of how charges and denaturant binding affect the dimension of unfolded protein chains. To address this issue and decouple charge effects from other intra-chain interactions, we proceed with a detailed investigation of the electrostatics of unfolded CspTm.

Treating charge-charge interactions in unfolded CspTm with polyampholyte theory

It has recently been shown that electrostatic interactions can dominate the dimensions of highly charged unfolded proteins and IDPs, especially at low ionic strength.^{4,9} In accord with physical intuition, unfolded polypeptide chains have been shown to expand with increasing net charge.^{4,14,30,31,33,53} Based on these results, an increase in net charge, $|Z|$, of unfolded CspTm from 3.8 at pH 7 to 10.2 at pH 2.9 would therefore be expected to lead to an expansion of the chain, in contrast to our experimental observation. In the past, the effect of charges on the dimensions of polymers has often been treated as an additional contribution to the effective excluded volume of the chain, and several theories have been developed based on this concept to describe the

TABLE I. Binding parameters obtained from the fit of ε with the shared binding site model (Eq. (7)).

pH	ε_0 (kT)	m_u	m_g	m_{ug}	K_u (M^{-1})	K_g (M^{-1})
7	1.7 ± 0.1	27 ± 20	19 ± 6	43 ± 2	0.16 ± 0.07	0.58 ± 0.13
2.9	2.1 ± 0.1	-17 ± 13	-26 ± 7	94 ± 1	0.32 ± 0.12	0.64 ± 0.19

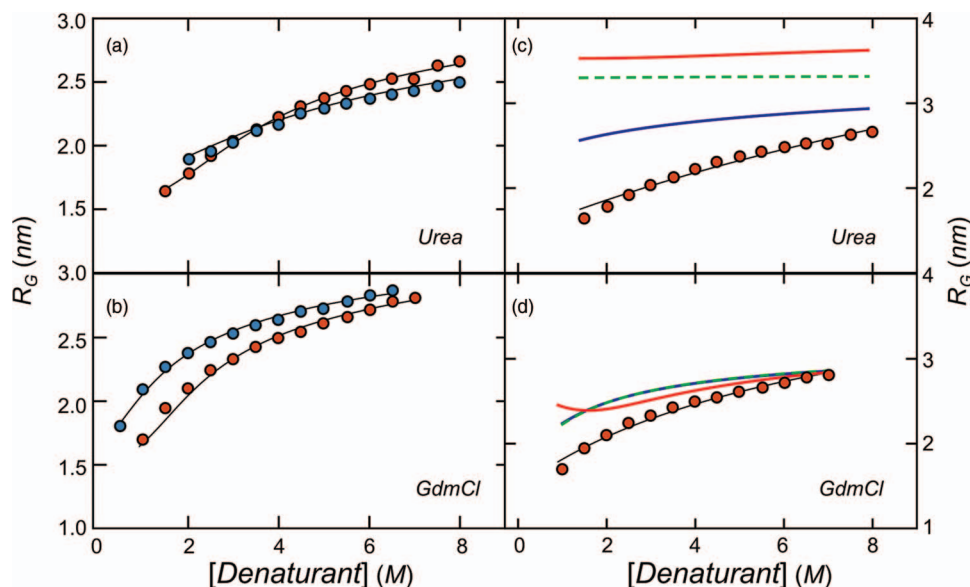


FIG. 4. The effect of pH on the R_G of unfolded CspTm in GdmCl or urea. Comparison of R_G of CspTm at pH 7 (blue) and pH 2.9 (red) in pure urea (a) and GdmCl (b). (c) and (d) A fit of the data at pH 2.9 with polyampholyte theory (Eq. (A3)) (black line) describes the data well with $\omega b\phi^3 = -1.64 \text{ nm}^3$, $\gamma_g = 1.85$, $K_g = 0.14 \text{ M}^{-1}$ for GdmCl (d) and $\omega b\phi^3 = -2.39 \text{ nm}^3$, $\gamma_u = 3.03$, $K_u = 0.05 \text{ M}^{-1}$ for urea (c). Predictions for R_G at pH 2.9 in urea (c) and GdmCl (d) are shown as calculated from polyampholyte theory with the excluded-volume term obtained from a fit of polyampholyte theory to the data at pH 7 (red line), and for the ion-cloud model including the chemical free energy of the electrical double layer and proton binding (solid blue line), and taking only repulsive charge interaction into account (dashed green line).

behavior of polyelectrolytes and polyampholytes^{30,31,54} (all of which are conceptually similar). In order to estimate the effect of charge interactions on the radius of gyration of CspTm, we used the polyampholyte theory developed by Higgs and Joanny,³⁰ which includes repulsive electrostatic interactions and a second-order correction to account for attractive electrostatic interactions of charges with opposite sign. The theory treats repulsive and attractive electrostatic interactions in a self-consistent manner based on the probability of occurrence of positive ($f = n_+/N_{\text{bonds}}$) and negative charges ($g = n_-/N_{\text{bonds}}$) in the chain. Polyampholyte theory can be used to describe the data for both pH values investigated here, but at pH 2.9 only with negative values for the excluded volume (-1.64 nm^3 in GdmCl, and -2.39 nm^3 in urea) (Figs. 4(c) and 4(d)), indicative of attractive intra-chain interactions that counteract charge repulsion. To get an idea of the magnitude of the effect of the excluded volume on R_G , we use the theory in a predictive manner by assuming that the difference between the R_G values at pH 7 and pH 2.9 (Figs. 4(a) and 4(b)) is caused solely by the difference in the protonation states of the amino acids. Thus, we first compute the effective excluded volume parameter of the chain (ω) for the data at pH 7 from Eq. (A1) and (A2) (see the Appendix) and then use these values of ω to predict the radii of gyration at pH 2.9. A quantitative comparison of the R_G values predicted for pH 2.9 with the experimentally obtained data at pH 2.9 reveals a pronounced discrepancy (Figs. 4(c) and 4(d)), especially at low ionic strength (Fig. 4(c)). However, at high concentrations of GdmCl, where charge interactions are expected to be negligible, the prediction is close to the experimentally determined radii of gyration, but at low concentrations of denaturant, the deviation increases (Fig. 4(d)). In summary, polyampholyte theory would predict a much more expanded chain at pH 2.9

than observed experimentally (Figs. 4(c) and 4(d)). The result implies that either the treatment of the electrostatics of the chain by polyampholyte theory is incomplete, or the predicted electrostatic repulsions at pH 2.9 are indeed compensated by additional attractive chain-chain interactions. In order to distinguish these two possibilities, we used an alternative treatment of the electrostatic interactions that includes in more detail the preferential enrichment of counterions in the volume explored by the chain and changes in protonation states of the polypeptide.

Treating charge-charge interactions in unfolded CspTm with an ion-cloud model

Recent polymer theories include the effect of counterions for polyelectrolytes in good⁵⁵ or poor solvent,⁵⁶ but do not account for changes in the protonation state of the polypeptide chain. We thus adapted the model developed by Alonso, Stigter and Dill,^{32,33} which describes the unfolded protein as a porous sphere in which protein residues and buffer solution are randomly mixed.³³ In this simplified picture of the chain as an ion cloud confined to the volume of gyration of the chain, the electrostatic potential ψ completely determines the electrostatic free energy for a given spatial arrangement of the ion cloud, which is treated in terms of Poisson-Boltzmann (PB) theory. Such models have been used successfully to describe the titration of proteins⁵⁷ or the behavior of polyelectrolytes.⁵⁸ While this model does not take into account the connectivity of the chain, it includes the preferential enrichment of counter ions in the volume pervaded by the chain, in contrast to the polyampholyte theory used above, which includes charge screening only in terms of the Debye-Hückel approximation.

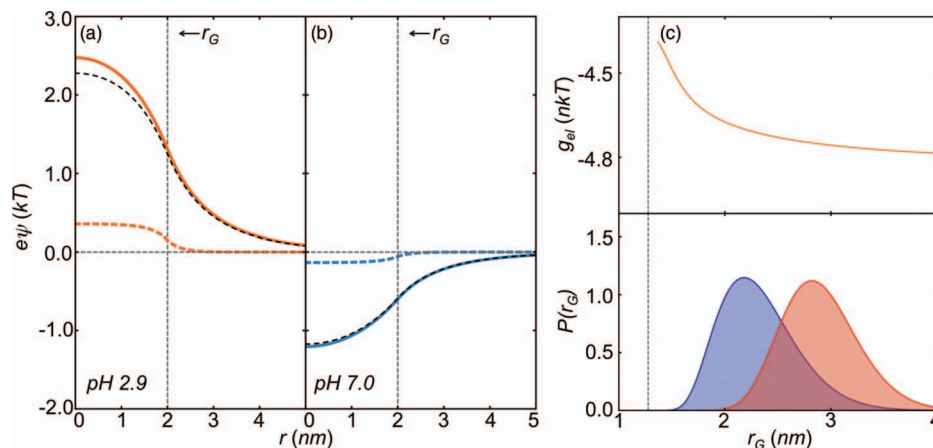


FIG. 5. The electrostatic potential ψ at pH 2.9 (a) and pH 7 (b) and the dependence of g_{el} on the radius of gyration (r_G) and its effect on $P(r_G)$ are depicted (c). (a) and (b) The electrostatic potential ψ is shown as function of the distance r from the center of the sphere with radius r_G for pH 2.9 (orange, (a)) and pH 7 (blue, (b)). The solution of the linearized PB equation (Debye-Hückel approximation) are shown at low (solid) and high (dashed) ionic strength. The solution of the exact PB equation at low ionic strength is shown for comparison (black dashed line). (c) The electrostatic free energy g_{el} as a function of r_G (orange), as obtained from the linearized PB equation (Debye-Hückel approximation) for CspTm at pH 2.9. The effect of g_{el} on the probability density function of radii of gyration (Eq. (1) and (9)) is illustrated for CspTm at pH 2.9 for a mean-field interaction energy of 1 kT resulting in the distribution function for the uncharged chain (blue) and the charged chain (red).

In the ion-cloud model, the n residues of the chain occupy n lattice sites, each with the volume v_{aa} and a radius $r_{aa} = (3v_{aa}/4\pi)^{1/3}$. A sphere with radius r_G and volume fraction ϕ occupies $n\phi^{-1}$ lattice sites, n of which are occupied by the protein residues, and the rest, $n(1 - \phi)^{-1}$, are filled with salt solution. For a given arrangement of charges inside the volume of this sphere, the electrostatic potential ψ determines the electrostatic free energy (g_{el}). Once the electrostatic potential ψ is known, all other quantities such as the free energy of electrostatic repulsion (g_{rep}), the chemical free energy due to the preferential enrichment of counterions inside the volume of the chain (g_{chem}), and the entropy of distributing the protons to the ionizable groups (g_α) can be computed.^{33,59,60} It should be noted that the ion-cloud model deviates from other polyelectrolyte models by its explicit incorporation of the chemical potential difference (g_{chem}) resulting from the preferential enrichment of counterions around the potential iso-surface defined by the volume average of the electrostatic potential, $\langle\psi\rangle$. The free energy of the electrical double layer resulting from protein charges and counterions is given by the repulsive electrostatic contribution of the protein charges (g_{rep}) and the chemical free energy of the counterion accumulation (g_{chem}). According to Alonso, Stigter and Dill,^{33,59} the total electrostatic free energy of the chain (g_{el}) additionally includes an entropic contribution resulting from the protonation equilibria of the ionizable amino acids in the chain (g_α), and is given by

$$\begin{aligned} g_{el}(r_G) &= g_{rep}(r_G) + g_{chem}(r_G) + g_\alpha(r_G) \\ &= (nkT)^{-1} \left[\int_0^Z e\langle\psi\rangle(r_G, z) dz - Ze\langle\psi\rangle(r_G, Z) \right. \\ &\quad \left. + kT \sum_{i=1}^t n_i \langle \ln(1 - \alpha_i) \rangle(r_G) \right]. \end{aligned} \quad (8)$$

Here, $\langle \dots \rangle$ denotes the spatial average over the volume of the sphere, Z is the net charge of the polypeptide, α_i is the proto-

nation state of amino acids of type i , and n_i is the number of amino acids of type i .

After solving the PB equation (Eq. (A4) in the Appendix) numerically (Figs. 5(a) and 5(b)), we computed g_{rep} , g_{chem} , and g_α for a chain with the charge content of CspTm (Eq. (8)). With decreasing radius of gyration (r_G), g_α increases, i.e., the entropy of distributing protons to the different charged groups decreases with increasing compaction of the chain. Concomitant with compaction, electrostatic repulsion and the corresponding free energy, g_{rep} , increase, which decreases the chemical free energy of the electrical double layer (g_{chem}) and counteracts the increase in g_α . In total, g_{el} increases with decreasing radius of gyration of the chain, leading to a repulsive contribution to the potential of mean force of the chain (Fig. 5(c)).

In order to estimate the effect of $g_{el}(r_G)$ on the dimensions of unfolded CspTm, we follow the work of Alonso, Stigter, and Dill,⁵⁹ and add $g_{el}(r_G)$ to the potential of mean force of the chain as given by the theory of Sanchez,¹⁸ leading to a modification of Eq. (1):

$$P_{el}(r_G, \varepsilon^*, R_{G\Theta}) = Q^{-1} r_G^6 \exp \left\{ -\frac{7r_G^2}{2R_{G\Theta}^2} + n \left[\frac{1}{2} \varepsilon^* \phi - \frac{1 - \phi}{\phi} \ln(1 - \phi) - g_{el}(r_G) \right] \right\}. \quad (9)$$

In contrast to Eq. (1), the mean-field interaction free energy (ε^*) in Eq. (9) does not include contributions from electrostatics. As with polyampholyte theory, we calculate the R_G values of unfolded CspTm expected at pH 2.9 based on the data at pH 7 by first determining ε^* at pH 7 from the mean transfer efficiencies by replacing $P(r_G, \varepsilon, R_{G\Theta})$ by $P_{el}(r_G, \varepsilon^*, R_{G\Theta})$ in Eq. (4). In a second step, we use ε^* at pH 7 in combination with g_{el} at pH 2.9 to predict the mean radius of gyration (R_G) at pH 2.9 using Eq. (5). Similar to polyampholyte theory, the ion-cloud model predicts an expansion of the chain on

lowering the pH to 2.9 (Figs. 4(c) and 4(d)), but to a smaller extent than polyampholyte theory. As for polyampholyte theory, the prediction of the ion-cloud model is close to the experimentally determined radii of gyration at high concentrations of GdmCl, i.e., high ionic strength (Fig. 4(d)), but the deviation increases for lower ionic strength (Figs. 4(c) and 4(d)). When neglecting the effect of counterions and proton binding, and considering only charge repulsion (g_{rep}), an even more expanded chain would be predicted by the ion-cloud model at low ionic strength (Fig. 4(c)), suggesting that the effects of counterion accumulation near the potential iso-surface of the porous sphere and proton binding seem to be essential for the better agreement of the ion-cloud model with the experimental data. However, the ion-cloud model is also not able to fully predict the dimensions of unfolded CspTm at low pH, indicative of additional short-range intra-chain interactions, such as H-bonds or contributions from the hydrophobic effect, present at pH 2.9 and contributing to ϵ^* . Correspondingly, whereas the difference between the uncorrected mean interaction energies at pH 2.9 and pH 7, $\Delta\epsilon = \epsilon_{pH2.9} - \epsilon_{pH7.0}$, is small at all concentrations of GdmCl and urea (Fig. 6(a)), the difference in the mean-field interaction energy for the corresponding uncharged chain, $\Delta\epsilon^*$, increases significantly with decreasing denaturant concentrations, i.e., increasing compactness of the chain. In the absence of denaturants, the intra-chain interaction energies per amino acid are by $\Delta\epsilon_0^* = 1.4 \pm 0.2$ kT higher at pH 2.9 than at pH 7.0 (Fig. 6(b) and Table II).

Once the electrostatic contribution to the free energy of unfolded CspTm is accounted for approximatively with the ion-cloud model, the binding of denaturant to the chain at both pH-values can be revisited (Figs. 6(c) and 6(d)). Surprisingly, when describing ϵ^* instead of ϵ with our denaturant binding model (Eq. (7)), substantially increased apparent binding constants for both denaturants are found,

TABLE II. Binding parameters obtained from the fit of ϵ^* with the shared binding site model (Eq. (7)).

pH	ϵ_0^* (kT)	m_u	m_g	m_{ug}	K_u (M ⁻¹)	K_g (M ⁻¹)
7	1.9 ± 0.1	45 ± 34	16 ± 5	36 ± 2	0.14 ± 0.10	1.0 ± 0.3
2.9	3.3 ± 0.2	15 ± 11	10 ± 3	45 ± 1	0.7 ± 0.5	4.0 ± 1.5

together with a shift towards more shared binding sites (Table II). This result suggests that with the increased chain-chain interactions inferred from $\Delta\epsilon^*$ (Fig. 6(b)), also chain-denaturant interactions increase at low pH, indicating that chain-chain and chain-denaturant interactions are closely related. If only chain-chain interactions were increased at low pH, the mean-field interaction free energy in water (ϵ_0^*) would be higher at pH 2.9 than at pH 7, but the denaturant binding parameters would remain unchanged. In contrast, if only the interaction of denaturant molecules with the polypeptide chain were increased on lowering the pH, the denaturant binding parameters would change, but ϵ_0^* would remain unchanged. Instead, we find that both quantities, chain-chain and chain-denaturant interactions, increase on lowering the pH from 7 to 2.9, suggesting that chain-chain and chain-denaturant interactions are of similar type.

DISCUSSION

We used single-molecule fluorescence spectroscopy in combination with the coil-to-globule transition theory developed by Sanchez¹⁸ to extract the radii of gyration and the mean-field interaction free energies for unfolded CspTm in urea, GdmCl, and mixtures of both denaturants at two different pH-values. Our most surprising observation is the lack of expansion of the unfolded protein when transferred from pH 7 to pH 2.9, even though the higher net charge of unfolded CspTm at pH 2.9 (+10.2) will strongly increase charge repulsion within the chain^{4,9,14,53} (Fig. 2): unfolded CspTm at pH 2.9 exhibits R_G values surprisingly close to those found at pH 7. Two types of theories,^{30,32,33} a polyampholyte theory and an ion-cloud model, both of which have previously been used successfully to describe electrostatic interactions in unfolded proteins,^{4,33,59} fail to correctly predict the dimensions of unfolded CspTm at low pH and indicate the presence of additional intra-chain interactions at pH 2.9. The only chemical difference between unfolded CspTm at pH 2.9 and pH 7 is the protonation of 12 acidic groups (8 Glu, 3 Asp, C-terminus). Hence, two possible candidates for additional attractive interactions are (1) an increased number of intra-chain H-bonds due to the protonated carboxy groups of Glu and Asp and (2) an increased hydrophobic effect, driven by the more hydrophobic side chains of protonated Glu and Asp compared to the unprotonated residues. Indeed, in hydrophobicity scales based on reversed-phase HPLC, Glu and Asp have been found to be significantly more hydrophobic at pH 2 than at neutral pH.⁶¹ Both types of interactions, H-bonds and the hydrophobic effect, are short-range interactions, and are expected to become more important with increasing compaction of the chain, in accord with our observation that the largest discrepancy in ϵ^* between pH 7 and pH 2.9 is observed

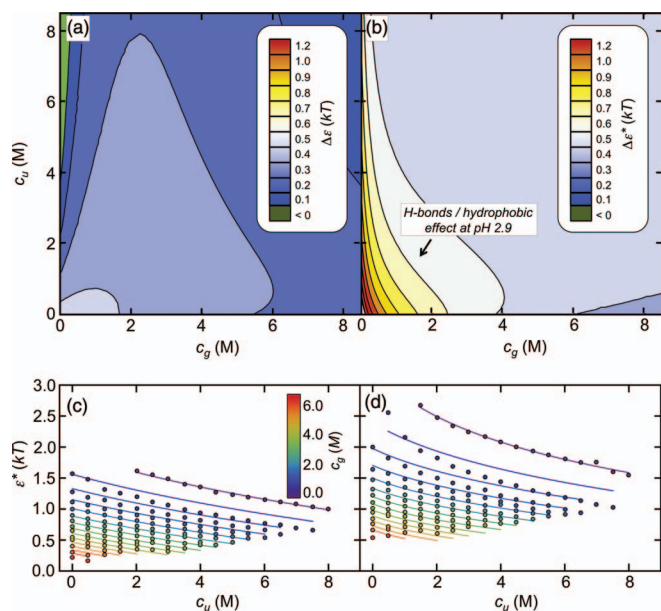


FIG. 6. Differences in the mean-field interaction energies at pH 2.9 and pH 7 for the charged chain, $\Delta\epsilon$, (a) and the uncharged chain, $\Delta\epsilon^*$, (b), and fits (solid lines) of the mean-field interaction free energies (ϵ^*) with the shared binding-site model (Eq. (7)) at pH 7 (c) and pH 2.9 (d).

at low concentrations of both denaturants, were the chains are most compact.

Another surprising observation of our analysis are the stronger effective interactions between the denaturants and the polypeptide chain found at pH 2.9 compared to pH 7.0. Again, the protonation of acidic side chains and the resulting increase in hydrophobicity might be responsible for this effect. The shift towards more binding sites shared by urea and GdmCl (Figs. 6(c) and 6(d), and Table II) indicates an increased role of interaction types that are common to urea and GdmCl, supporting the importance of dispersion (van-der-Waals) interactions and the hydrophobic effect. Both denaturants have been found to interact strongly with hydrophobic surfaces⁴⁵ and amino acids,⁴³ leading to a dewetting of non-polar surfaces, and suggesting that weakening the hydrophobic effect is one major driving force for denaturant-protein interactions.⁴⁶ In agreement with the preferential interaction with hydrophobic surfaces, urea has also been found to exhibit a moderate tendency to self-associate.⁴² Recent molecular dynamics simulations of protein L and lysozyme in mixtures of urea and GdmCl even suggest that this self-aggregation tendency might cause a collapse of unfolded proteins.³⁶ As the driving force for this collapse, a replacement of urea by guanidinium ions at the polypeptide chain surface has been suggested that results in an enhanced local crowding effect.³⁶ Such coil-to-globule transitions in mixtures of immiscible solvents have been predicted to occur close to the critical demixing point of the two solvents.⁶² Experimentally, however, we do not observe this effect for unfolded CspTm (Fig. 2), indicating that the critical demixing point for GdmCl and urea is not close to our experimental conditions.

In summary, we arrive at an unexpected, and at a first glance counterintuitive conclusion. The increased net charge of unfolded CspTm at low pH does not lead to the expansion of the chain expected from the increased electrostatic repulsion. Instead, a more complex interplay of electrostatic interactions, chemical free energy of counter ion binding, proton binding entropy, denaturant binding, and the hydrophobic effect appears to be involved, leading to a more compact chain at pH 2.9 than at pH 7. Both theoretical approaches we employed to describe the data imply the presence of such additional interactions. However, both theories neglect important aspects of the physics of charge interactions in polymers, and a comprehensive description might have to include concurrently the explicit repulsive and attractive interactions of polyampholyte theory,³⁰ the effects of protonation entropy and counter ion condensation of the ion-cloud model used here,^{32,33} ideally with retained chain connectivity⁵⁶ and allowing for a spatial redistribution of charged and uncharged residues, and possibly even correlations between counterions condensed on the chain, which has been shown to generate additional attractions in highly charged polyelectrolytes.⁶³ A combination of further developments in theoretical concepts and, in particular, molecular simulations of unfolded proteins will be important for obtaining a more detailed understanding of the underlying balance between charge-charge interactions and hydrophobic hydration effects.

MATERIALS AND METHODS

Single-molecule measurements were performed at 22 °C using a Micro Time 200 confocal microscope equipped with a HydraHarp 400 counting module (Picoquant, Berlin, Germany). The donor dye was excited with linear polarized light at 483 nm coming from a diode laser (LDH-D-C-485, PicoQuant) at an average power of 100 μ W. Single-molecule FRET efficiency histograms were acquired in samples with a protein concentration of about 20–50 pM, at a repetition rate of 64 MHz; photon counts were recorded with a resolution of 8 ps by the counting electronics (time resolution was thus limited by the timing jitter of the detectors). The photons were distributed according to their polarization relative to the polarization of the excitation light using a polarizing beam splitter cube. Subsequently, both components were further subdivided into donor and acceptor photons by a dichroic mirror (585DCXR, Chroma). Subpopulation-specific steady-state anisotropies were determined essentially as described by Sisamakris *et al.*,⁶⁴ with corrections obtained as described by Koshioka *et al.*⁶⁵ Fluorescence lifetimes were obtained by fitting the tail of the TCSPC histograms using a single-exponential decay. TCSPC histograms were constructed exclusively from photons originating from doubly-labeled unfolded CspTm molecules. The measurements were performed in 20 mM sodium phosphate buffer at pH 7.0 (ionic strength $I = 35$ mM) or pH 2.9 (adjusted to $I = 35$ mM with NaCl), 100 mM β -mercaptoethanol (Sigma), and 0.001% Tween 20 (Pierce) with varying concentrations of GdmCl (Pierce) or urea. The exact denaturant concentrations were determined by the refractive index of the solutions. The change in the Förster distance (R_0) due to changes in the refractive index was taken into account in the calculation of the mean radius of gyration with Eqs. (4) and (5).

For electrostatics calculations, we used the total amount of ionizable groups in CspTm, including the charges of the termini. The charge-state of the dyes was taken into account. Alexa Fluor 488 and Alexa Fluor 594 contain two sulfonic acid groups, for which we assumed the pK-value of benzenesulfonic acid ($pK = -2.5$), one carboxylic group for which we assumed the pK-value of benzoic acid ($pK = 4.2$), and one delocalized positive charge which we assumed not to be titratable in the pH range used here. For the calculation of the electrostatic free energy with the ion-cloud model, the potentials at ionic strengths greater than 1 M were approximated by the potential at 1 M, since very high ionic strength leads to numerical problems in solving the PB equation. Importantly, since charge effects are already negligible at 1 M salt in the mean-field PB framework, this approximation has no effect on the values of the mean-field interaction energies obtained from Eq. (9).

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation, the National Center of Competence in Research for Structural Biology, and a Starting Investigator Grant of the European Research Council.

APPENDIX: THEORY

Polyampholyte theory

Charge screening is treated by the Debye-Hückel theory, with the Debye length $\kappa_w^{-1} = (8\pi l_B I)^{-1/2}$, where l_B is the ionic strength, and $l_B = e^2/(4\pi\epsilon_0\epsilon_r kT)$ is the Bjerrum length with the elementary charge e , the permittivity of vacuum ϵ_0 , and the dielectric constant ϵ_r . For a chain in good solvent, the electrostatic interactions act as an additional contribution to the effective excluded volume parameter of the monomer, ω^* , leading to

$$\alpha^5 - \alpha^3 = \frac{4}{3} \left(\frac{3}{2\pi} \right)^{3/2} N_{bonds}^{1/2} \omega^* \text{ with } \alpha = \frac{R_G}{R_{G\Theta}}, \quad (\text{A1})$$

and

$$\omega^* = \omega + \frac{4\pi l_B (f - g)^2}{b_\Theta^3 \kappa_w^2} - \frac{\pi l_B^2 (f + g)^2}{b_\Theta^3 \kappa_w}. \quad (\text{A2})$$

Here, Eq. (A1) is the Flory result for a chain in good solvent, which relates the expansion factor of the chain (α) to an effective excluded volume parameter (ω^*), which is related to the intra-chain interaction energy. In Eq. (A2), f is the probability of occurrence of a positive charge ($f = n_+/N_{bonds}$), g is the probability of occurrence of a negative charge ($g = n_-/N_{bonds}$) when neglecting unity compared to $N_{bonds} = 75$, and the segment length of a monomer at the Θ -state (b_Θ) is given by the prefactor in Eq. (2) for $\nu = \frac{1}{2}$, which yields $b_\Theta = (0.22 \text{ nm})^{6/2} = 0.54 \text{ nm}$. Equation (A1) has only one free parameter, the excluded volume parameter ω . For a fit of the experimental data, we followed our previous approach⁴ and combined Eq. (A1) and (A2) with an empirical denaturant binding model

$$R_G = N_{bonds}^{1/2} \alpha b_\Theta^3 \left(1 + \gamma_i \frac{K_i c_i}{1 + K_i c_i} \right) / \sqrt{6}. \quad (\text{A3})$$

Here, K_i is the denaturant binding constant for denaturant i and γ_i is an adjustable parameter that describes the amplitude of the denaturant-induced expansion.

Ion-cloud model

A detailed description of a similar ion-cloud model can be found in Ref. 33. Briefly, in the ion-cloud model adapted from Alonso, Stigter, and Dill,³³ the electrostatic potential for a given distribution of charges is given by the Poisson-Boltzmann equation:

$$\begin{aligned} \nabla^2 \psi &= \lambda \kappa_{sphere}^2 \sinh \left(\frac{e\psi}{kT} \right) - \frac{\rho_{el}}{\epsilon_0 \epsilon_{sphere}} \\ &\approx \kappa_{sphere}^2 \psi - \frac{\rho_{el}}{\epsilon_0 \epsilon_{sphere}} \text{ for } 0 < r < r_G. \end{aligned} \quad (\text{A4})$$

with $\lambda = kT/e$. Here, r is the distance from the center of the sphere with the radius of gyration r_G . In the ion-cloud model, only the radial component of the Laplace operator ($\nabla^2 = r^{-2} \partial(r^2 \partial/\partial r)/\partial r$) is required, since we assume a spherically symmetric arrangement of the charges inside the volume of the chain. In Eq. (A4), κ_{sphere}^{-1} is the Debye screening length inside the porous sphere, and ρ_{el} is the charge

density inside the sphere. The right-hand side of Eq. (A4) is the Debye-Hückel approximation (linearized PB equation), which is a good approximation for small potentials, i.e., at low net charges and higher ionic strengths. We used this approximation for all calculations but also checked its validity at low ionic strength by using the exact PB equation. The charge distribution in our ion-cloud model is uniform within the volume of gyration explored by the chain

$$\rho_{el}(r) = \frac{e}{V_G} \sum_{i=1}^t q_i n_i \alpha_i(r). \quad (\text{A5})$$

For the t different types of ionizable groups, $q_i = +1$ for basic and $q_i = -1$ for acidic groups, α_i is the degree of ionization, n_i is the number of ionizable groups of type i , $V_G = 3\pi r_G^3/4$ is the volume of gyration. The degree of ionization, α_i , of ionizable group i is given by

$$q_i \log \left(\frac{\alpha_i(r)}{1 - \alpha_i(r)} \right) = -(pH + pK_i) - 0.434 \frac{e\psi(r)}{kT}, \quad (\text{A6})$$

where the value of 0.434 is the decadic logarithm of Eulers number. Here, the pK_i values are those of the free amino acids. The porous-sphere model of Stigter and Dill³³ also accounts for the change in the dielectric constant on compaction of the chain according to

$$\epsilon_{sphere} = (1 - \phi)\epsilon_w + \phi\epsilon_p \text{ for } r < r_G, \quad (\text{A7})$$

where the dielectric constant of the protein interior is assumed to be $\epsilon_p = 3.5$.⁶⁶ Assuming an average number of salt ions, $(1 - \phi)n_s$, inside the sphere, with n_s being the concentration (number density) of salt ions of type s in free solution, the Debye length inside the sphere, κ_{sphere}^{-1} , is given by

$$\kappa_{sphere}^2 = \frac{(1 - \phi)\epsilon_w}{(1 - \phi)\epsilon_w + \phi\epsilon_p} \kappa_w^2. \quad (\text{A8})$$

Here, κ_w^{-1} is the Debye screening length in water as described in the main text. As a boundary condition, the potential ψ must be continuously differentiable at r_G , i.e., the gradient at $r = r_G$ on the outside of the sphere must be equal to the gradient at $r = r_G$ on the inside of the sphere. In addition, symmetry requires that $d\psi/dr = 0$ for $r = 0$ if we assume that no point charge is at the center of the sphere:³³

$$\left(\frac{d\psi}{dr} \right)_{r_G} = -\frac{1 + \kappa_w r_G}{r_G} \psi(r_G) \quad r = r_G^+, \quad (\text{A9})$$

$$\left(\frac{d\psi}{dr} \right)_{r_G} = -\frac{\epsilon_w(1 + \kappa_w r_G)}{\epsilon_{sphere} r_G} \psi(r_G) \quad r = r_G^-, \quad (\text{A10})$$

$$\frac{d\psi}{dr} = 0 \quad r = 0. \quad (\text{A11})$$

In Eq. (A9) and (A10), r_G^+ and r_G^- indicate the inside and the outside of the sphere at the position r_G . Equation (A4) is numerically solved to obtain the general solution $\psi'(r)$, which we normalized such that $\psi \rightarrow 0$ for $r \rightarrow \infty$, which gives

$$\psi(r) = \begin{cases} \psi'(r) - A_2 & r \leq r_G \\ \psi_{outside}(r) & r > r_G \end{cases}. \quad (\text{A12})$$

The potential outside the sphere is given by

$$\psi_{outside}(r) = A_1 \psi'(r_G) r_G \frac{\exp[-\kappa_w(r - r_G)]}{r}, \quad (\text{A13})$$

where

$$A_1 = -\frac{r_G(d\psi'/dr)_{r_G}}{(1 + \kappa_w r_G)\psi'(r_G)} \text{ and } A_2 = \psi'(r_G) + \frac{r_G(d\psi'/dr)_{r_G}}{1 + \kappa_w r_G} \quad (\text{A14})$$

are constants that result from the boundary conditions mentioned in the main text.

In the Debye-Hückel approximation, g_{rep} and g_{chem} are linearly related^{32,33} according to

$$g_{chem}(r_G) = -2g_{rep}(r_G), \quad (\text{A15})$$

and the electrostatic free energy simplifies to

$$g_{el}(r_G) = g_{\alpha}(r_G) - g_{rep}(r_G). \quad (\text{A16})$$

Equations (A15) and (A16) imply that the electrostatic free energy g_{el} is always dominated by the chemical free energy of the electrical double layer (g_{chem}), and by the entropy of distributing protons on the ionizable groups (g_{α}). The explicit expressions for g_{rep} , g_{chem} , and g_{α} are

$$g_{rep}(r_G) = (nkT)^{-1} \int_0^Z e\langle\psi\rangle(r_G) dz \\ = (nkT)^{-1} \sum_{i=1}^t \frac{q_i n_i}{V_G} \int_0^{r_G} 4\pi r^2 \frac{\alpha_i(r) e\psi(r)}{2} dr, \quad (\text{A17})$$

$$g_{chem}(r_G) = -Ze\langle\psi\rangle(r_G)/nkT \\ = (nkT)^{-1} \frac{-Ze}{V_G} \int_0^{r_G} 4\pi r^2 \psi(r) dr, \quad (\text{A18})$$

$$g_{\alpha}(r_G) = n^{-1} \sum_{i=1}^t n_i \langle \ln(1 - \alpha_i) \rangle \\ = n^{-1} \sum_{i=1}^t \frac{n_i}{V_G} \int_0^{r_G} 4\pi r^2 \ln(1 - \alpha_i(r)) dr. \quad (\text{A19})$$

The net charge of the chain (Z) is given by

$$Z(r_G) = \sum_{i=1}^t \frac{q_i n_i}{V_G} \int_0^{r_G} 4\pi r^2 \alpha_i(r) dr. \quad (\text{A20})$$

¹S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **93**(21), 11615 (1996); O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, *ibid.* **96**(17), 9597 (1999).

²B. Schuler, E. Lipman, and W. Eaton, *Nature (London)* **419**(6908), 743 (2002).

³H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **6**(3), 197 (2005); K. Sugase, H. Dyson, and P. E. Wright, *Nature (London)* **447**, 1021 (2007); V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Annu. Rev. Biophys.* **37**, 215 (2008).

⁴S. Müller-Späh, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rügger, L. Reymond, D. Nettels, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14609 (2010).

⁵B. Shoemaker, J. Portman, and P. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **97**(16), 8868 (2000).

⁶A. Soranno, B. Buchli, D. Nettels, R. Cheng, S. Müller-Späh, S. Pfeil, A. Hoffmann, E. Lipman, D. Makarov, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17800 (2012); D. Nettels, I. Gopich, A. Hoffmann, and B. Schuler, *ibid.* **104**(8), 2655 (2007).

⁷C. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6369 (1993); S. Waldauer, O. Bakajin, and L. Lapidus, *ibid.* **107**(31), 13713 (2010).

⁸A. Möglich, K. Joder, and T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A.* **103**(33), 12394 (2006).

⁹H. Hofmann, R. Golbik, M. Ott, C. Hübner, and R. Ulbrich-Hofmann, *J. Mol. Biol.* **376**(2), 597 (2008).

¹⁰D. Nettels, S. Müller-Späh, F. Küster, H. Hofmann, D. Haenni, S. Rügger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. Best, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20740 (2009).

¹¹A. C. M. Ferreón, M. M. Moosa, Y. Gambin, and A. A. Deniz, *Proc. Natl. Acad. Sci. U.S.A.* **109**(44), 17826 (2012).

¹²E. Sherman and G. Haran, *Proc. Natl. Acad. Sci. U.S.A.* **103**(31), 11539 (2006).

¹³G. Ziv and G. Haran, *J. Am. Chem. Soc.* **131**(8), 2942 (2009).

¹⁴H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **109**(40), 16155 (2012).

¹⁵S. Sun, I. Nishio, G. Swislow, and T. Tanaka, *J. Chem. Phys.* **73**(12), 5971 (1980); C. Wu and S. Zhou, *Phys. Rev. Lett.* **77**(14), 3053 (1996).

¹⁶P. Flory, *J. Chem. Phys.* **17**(3), 303 (1949).

¹⁷P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979); A. Grosberg and D. Kuznetsov, *Macromolecules* **25**, 1980 (1992); **25**(7), 1991 (1992); **25**(7), 1996 (1992).

¹⁸I. Sanchez, *Macromolecules* **12**, 980 (1979).

¹⁹A. Grosberg and D. Kuznetsov, *Macromolecules* **25**(7), 1970 (1992).

²⁰P. de Gennes, *J. Phys. Lett.* **36**, 55 (1975).

²¹T. Birshtein and V. Pryamitsyn, *Macromolecules* **24**, 1554 (1991).

²²J. Kohn, I. Millett, J. Jacob, B. Zagrovic, T. Dillon, N. Cingel, R. Dethager, S. Seifert, P. Thiagarajan, T. Sosnick, M. Hasan, V. Pande, I. Ruczinski, S. Doniach, and K. Plaxco, *Proc. Natl. Acad. Sci. U.S.A.* **101**(34), 12491 (2004).

²³R. Dima and D. Thirumalai, *J. Phys. Chem. B* **108**(21), 6564 (2004).

²⁴Y. Nozaki and C. Tanford, *J. Biol. Chem.* **245**(7), 1648 (1970).

²⁵G. Haran, *Curr. Opin. Struct. Biol.* **22**(1), 14 (2012).

²⁶E. I. Shakhnovich and A. M. Gutin, *Biophys. Chem.* **34**(3), 187 (1989).

²⁷J. Bryngelson and P. Wolynes, *Biopolymers* **30**, 177 (1990).

²⁸K. A. Dill, *Biochemistry* **24**(6), 1501 (1985).

²⁹K. A. Dill, D. O. Alonso, and K. Hutchinson, *Biochemistry* **28**(13), 5439 (1989).

³⁰P. G. Higgs and J.-F. Joanny, *J. Chem. Phys.* **94**(2), 1543 (1991).

³¹B.-Y. Ha and D. Thirumalai, *Phys. Rev. A* **46**(6), R3012 (1992).

³²D. Stigter and K. A. Dill, *J. Phys. Chem.* **93**, 6737 (1989).

³³D. Stigter and K. A. Dill, *Biochemistry* **29**(5), 1262 (1990).

³⁴V. Uversky, J. Gillespie, and A. Fink, *Proteins* **41**, 415 (2000).

³⁵A. A. Deniz, T. A. Laurence, G. S. Belligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* **97**(10), 5179 (2000); E. Kuzmenkina, C. Heyes, and G. Nienhaus, *ibid.* **102**(43), 15471 (2005); K. Merchant, R. Best, J. Louis, I. Gopich, and W. Eaton, *ibid.* **104**(5), 1528 (2007).

³⁶Z. Xia, P. Das, E. I. Shakhnovich, and R. Zhou, *J. Am. Chem. Soc.* **134**(44), 18266 (2012).

³⁷B. W. Van Der Meer, G. Coker III, S. Y. S. Chen, *Resonance Energy Transfer: Theory and Data* (VCH Publishers, Inc., New York, 1994).

³⁸B. Hammouda, *Adv. Polym. Sci.* **106**, 87 (1993).

³⁹E. R. McCarney, J. H. Werner, S. L. Bernstein, I. Ruczinski, D. E. Makarov, P. M. Goodwin, and K. W. Plaxco, *J. Mol. Biol.* **352**(3), 672 (2005); G. F. Schröder, U. Alexiev, and H. Grubmüller, *Biophys. J.* **89**(6), 3757 (2005); A. Hoffmann, A. Kane, D. Nettels, D. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. Horsley, R. Seckler, O. Bakajin, and B. Schuler, *Proc. Natl. Acad. Sci. U.S.A.* **104**(1), 105 (2007).

⁴⁰H. T. Tran, A. Mao, and R. V. Pappu, *J. Am. Chem. Soc.* **130**(23), 7380 (2008).

⁴¹Even at the lowest concentration of urea ($c_u = 2$ M, $R_G = 1.9$ nm) and GdmCl ($c_g = 0.5$ M, $R_G = 1.8$ nm), the expansion factor $\alpha = R_G/R_{G\Theta}$ is close to 1 ($\alpha = 1$ in urea and $\alpha = 0.95$ in GdmCl), which justifies the use of Eqs. (1)–(5).

⁴²M. C. Stumpe and H. Grubmüller, *J. Phys. Chem. B* **111**(22), 6220 (2007).

⁴³M. C. Stumpe and H. Grubmüller, *J. Am. Chem. Soc.* **129**(51), 16126 (2007).

⁴⁴E. P. O'Brien, R. I. Dima, B. Brooks, and D. Thirumalai, *J. Am. Chem. Soc.* **129**(23), 7346 (2007).

⁴⁵J. L. England, V. S. Pande, and G. Haran, *J. Am. Chem. Soc.* **130**(36), 11854 (2008).

⁴⁶J. England and G. Haran, *Annu. Rev. Phys. Chem.* **62**, 257 (2011).

- ⁴⁷J. Schellman, *Biophys. Chem.* **96**(2–3), 91 (2002).
- ⁴⁸E. O'Brien, G. Ziv, G. Haran, B. Brooks, and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13403 (2008).
- ⁴⁹A free fit with Eq. (7) (i.e., if all five parameters are freely adjusted) does not provide a significant improvement of the fit quality. For comparison, $\text{rmsd} = 0.027\text{ }kT$ with a maximal deviation of $0.1\text{ }kT$.
- ⁵⁰In combination with the limited range of accessible urea concentrations, the lower affinity of urea for polypeptide chains results in an almost linear decrease of the experimentally determined interaction energies with increasing concentrations of urea, leading to a large uncertainty of the logarithmic term in Eq. (6).
- ⁵¹G. I. Makhatadze and P. L. Privalov, *J. Mol. Biol.* **226**(2), 491 (1992); G. Makhatadze, *J. Phys. Chem. B* **103**, 4781 (1999).
- ⁵²N. P. Wen and M. H. Brooker, *J. Phys. Chem.* **97**(33), 8608 (1993).
- ⁵³A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, *Proc. Natl. Acad. Sci. U.S.A.* **107**(18), 8183 (2010).
- ⁵⁴M. Muthukumar, *J. Chem. Phys.* **86**(12), 7230 (1987).
- ⁵⁵E. Y. Kramarenko, A. R. Khokhlov, and K. Yoshikawa, *Macromol. Theory Simul.* **9**(5), 249 (2000).
- ⁵⁶A. Kundagrami and M. Muthukumar, *Macromolecules* **43**(5), 2574 (2010).
- ⁵⁷K. Linderstrøm-Lang, *C. R. Trav. Lab. Carlsberg* **15**, 1 (1924).
- ⁵⁸J. J. Hermans and J. T. G. Overbeek, *Recl. Trav. Chim. Pays-Bas* **67**, 761 (1948).
- ⁵⁹D. O. Alonso, K. A. Dill, and D. Stigter, *Biopolymers* **31**(13), 1631 (1991).
- ⁶⁰D. Stigter, D. O. Alonso, and K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **88**(10), 4176 (1991).
- ⁶¹C. T. Mant, J. M. Kovacs, H.-M. Kim, D. D. Pollock, and R. S. Hodges, *Biopolymers* **92**(6), 573 (2009).
- ⁶²P.-G. de Gennes, *J. Phys. Lett.* **37**, 59 (1976).
- ⁶³G. C. L. Wong and L. Pollack, *Annu. Rev. Phys. Chem.* **61**, 171 (2010).
- ⁶⁴E. Sisamakias, A. Valeri, S. Kalinin, P. J. Rothwell, and C. A. M. Seidel, *Methods Enzymol.* **475**, 455 (2010).
- ⁶⁵M. Koshioka, K. Sasaki, and H. Masuhara, *Appl. Spectrosc.* **49**, 224 (1995).
- ⁶⁶M. Gilson and B. Honig, *Proteins* **4**, 7 (1988).