Supporting Information

Müller-Späth et al. 10.1073/pnas.1001743107

SI Text

SI Methods. ***Preparation and labeling of proteins.*** Cysteine residues were introduced by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes as described previously (1, 2). The truncated variant of Csp*Tm* was expressed with a cleavable hexahistidine tag to allow for rapid purification. The gene was cloned from vector pET21a (1, 2) into pET47b(+) and the sequence coding for the two C-terminal amino acids was deleted by site-directed mutagenesis. The protein was expressed in LB medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted and DNA was digested. The supernatant was cleared by centrifugation and loaded on a HisTrap column (GE Healthcare, BioSciences AB) in 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, 4 M GdmCl, pH 8.0. After the 280 nm UV absorption signal reached the baseline, the column was washed with two column volumes 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 8.0, and a gradient from 10 to 500 mM imidazole was used to elute the His-tagged protein. HRV 3C protease (containing a His-tag) was added to a final concentration of 0.3 mg/mL, and after 12 h at room temperature, the cleavage reaction was dialyzed against 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 8.0 and applied to a HisTrap column. The cleaved Csp*Tm* without His-tag was collected in the flow-through and concentrated. Labeling was performed as described previously (2).

The N-terminal domain of HIV1-integrase (IN) was expressed in the vector pET15b. Cysteine residues were introduced at positions 0 and 56 (residue numbering is starting from Phe in the protein sequence; Table S1). The protein was expressed in LB medium with carbenicillin and 1 mM IPTG at 37 °C. Harvested cells were lysed and the DNA was digested. The supernatant was cleared by centrifugation and loaded on a HisTrap HP column, equilibrated with 20 mM HEPES, 1 M sodium chloride, 20 mM imidazole, 2 mM β-mercaptoethanol, pH 7.5. The column was washed with 60 mM imidazole and IN was eluted using 10% glycerol and a gradient from 20 to 500 mM imidazole maintaining all other buffer conditions. Fractions were identified via SDS-PAGE, combined and dialyzed against 20 mM HEPES, 100 mM sodium chloride, 5 mM EDTA, 1 mM DTT, pH 7.5. The His-tag was cleaved using 10 units thrombin from bovine plasma (SERVA Electrophoresis, Heidelberg, Germany) per mg protein at an IN concentration of 0.13 mg/mL at room temperature for 30 minutes. Cleavage was verified by SDS-PAGE, and the protein was loaded on the HisTrap HP column under the same conditions as before. Fractions containing IN were combined, adjusted to 6 M GdmCl and 3 mM TCEP, and concentrated by ultrafiltration. Gel filtration was done with a HiLoad Superdex 75 prep grade column (GE Healthcare, BioSciences AB) under refolding conditions in 25 mM Tris, 250 mM sodium chloride, 10% glycerol, 0.1 mM $ZnCl_2$, 1 mM DTT, pH 7.4. For fluorophore labeling, IN was reduced with 3 mM TCEP and desalted with a HiTrap desalting column (GE Healthcare, BioSciences AB) in 50 mM sodium phosphate, 0.1 mM $ZnCl_2$ pH 7.0. Fractions were collected under argon atmosphere. The protein was incubated with Alexa Fluor 488 maleimide at a 1:1 molar ratio. Singly labeled protein was separated from unlabeled and doubly labeled protein using ion exchange chromatography with a MonoQ column (GE Healthcare, BioSciences AB). Fractions containing singly labeled protein, confirmed by matrix-assisted laser desorption/ionization mass spectrometry, were incubated with a 2:1 molar excess of Alexa Fluor 594 maleimide. Doubly labeled protein was separated as before and the correct molecular mass of the labeled protein was confirmed by mass spectrometry.

The coding sequence for human ProTα was cloned from vector pHP12 (3) into pET47b(+). Cysteine residues were introduced in positions 2 and 56 in the variant ProTαN and in positions 56 and 110 in the variant ProTαC by site-directed mutagenesis (residue numbering is from Met in the protein sequence, excluding the 19 residue N-terminal purification tag; Table S1). The protein was expressed in Terrific Broth medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were lysed and DNA was digested. The cleared supernatant was loaded on a HisTrap HP column (GE Healthcare, BioSciences AB) in 20 mM Tris, 100 mM sodium chloride and 2 mM β-mercaptoethanol, pH 7.0. A gradient from 40 to 500 mM imidazole was used to elute the His-tagged protein. Fractions were identified via SDS-PAGE, combined, extracted by butanol, and precipitated with ethanol (3). Pellets were dissolved in 50 mM sodium phosphate, pH 7.0, reduced with 5 mM TCEP, and purified on a Superdex75 gel filtration column (GE Healthcare, BioSciences AB) in 100 mM sodium phosphate, 2 mM β-mercaptoethanol, 0.01% Tween, pH 7.0. Fractions containing the full-length protein were combined, extracted with butanol, and precipitated with ethanol. The pellet was dissolved in 4 M GdmCl, 50 mM sodium phosphate, pH 7.0, and the protein concentration was determined with a bicinchoninic acid assay (BCA Protein Assay Kit, Pierce), because ProTα contains no aromatic residues. Fluorophore labeling was performed at a protein concentration of approximately 0.1 mg/mL with a threefold molar excess of the dyes Alexa Fluor 488 and Alexa Fluor 594 maleimide; gel filtration was used to remove the free dye. The correct molecular mass of the labeled protein was confirmed by matrix-assisted laser desorption/ionization mass spectroscopy. The donor-only labeled and acceptor-only labeled ProTα resulting from such random labeling does not interfere with single-molecule FRET measurements.

Single-molecule fluorescence spectroscopy. Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) equipped with a continuous wave 488 nm diode laser (Sapphire 488-100 CDRH, Coherent) and an Olympus UplanApo 60x/1.20W objective. Sample fluorescence was separated into donor and acceptor components using a dichroic mirror (585DCXR, Chroma), and two final filters (Chroma ET525/50M, HQ650/100). Each component was focused onto an avalanche photodiode (SPCM-AQR-15, PerkinElmer Optoelectronics), and the arrival time of every detected photon was recorded. Samples of labeled protein were diluted to a concentration of approximately 20 pM in 50 mM Tris buffer at the appropriate GdmCl (Pierce) concentration, and individually adjusted to pH 7.4. 0.001% Tween 20 (Pierce) was added to prevent surface adhesion of the protein (1). To minimize damage to the chromophores, the photo-protective additive β-mercaptoethanol (200 mM) was included. To eliminate zinc from its complex with IN, 1 mM EDTA was added. The measurements were performed at a laser power of 110 μW at the sample with an acquisition time of 1 h (for 7000 to 15000 identified bursts). Successive photons detected in either channel separated by less than 100 μs were combined into one burst. Identified bursts were corrected for background, differences in quantum yields of donor and acceptor, the different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation as described previously (4). A burst was retained as a significant event if the total number of counts exceeded 50.

The Förster radius R_0 was corrected for the changes in solution conditions, which were dominated by the change in refractive index with GdmCl (2, 5) or urea (6, 7) concentration. The overlap integral (8) of Alexa 488 emission and Alexa 594 absorption was found to be independent of denaturant concentration.

Determination of the radius of gyration. We determined the dimensions of the unfolded protein chains from the measured transfer efficiency with two different approaches: the Gaussian chain (1, 2, 9–11) and a variation of Sanchez theory previously used by Haran and coworkers (10, 12, 13).

The end-to-end distance distribution of a Gaussian chain is

$$P_{\text{Gaussian}}(r) = 4\pi r^2 \left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2} \exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right), \qquad \textbf{[S1]}$$

and its mean-square radius of gyration is given by $\langle R_g^2\rangle = \langle r^2\rangle/6$.

The Sanchez-type theory (10, 12, 13) employs the modified Flory–Fisk equation (12, 13)

$$P(R_g) = P_0(R_g)\exp(-Ng(\phi,\varepsilon)/\mathrm{k_B T}), \qquad \textbf{[S2]}$$

where $P_0(R_g) \propto R_g^6 \exp(-7/2\ R_g^2/\langle R_{g\theta}^2\rangle)$ is the Flory–Fisk distribution for the radius of gyration of an ideal polymer chain (with a normalization chosen such that the integral of $P(R_g)$ equals unity), which is weighted by the expansion free energy per monomer, $g(\phi,\varepsilon) = -\frac{1}{2}\phi\varepsilon + \mathrm{k_B T}\frac{1-\phi}{\phi}\log(1-\phi)$. The function $g(\phi,\varepsilon)$ is defined according to eq. 26 of ref. 14, neglecting the constant term that can be included in the normalization factor. $\phi = R_{g,N}^3/R_g^3$ is the volume fraction occupied by the chain ($R_{g,N}$ is the radius of gyration of the fully compact/native state), and ε is a mean field interaction relative to the most collapsed state and is a measure of the two-body interactions within the chain. The radius of gyration at the θ-point can be estimated using the argument of Sanchez (13, 14) based on Landau's theory of phase transitions (13). Accordingly, the volume fraction in the θ-state depends on the number of amino acids, N, as $\phi_\theta = \sqrt{19/27}N^{-1/2}$. With the radius of gyration of a (hypothetical) native state corresponding to protein segments of the size investigated here (calculated by scaling the radius of gyration of Csp*Tm*, $R_{g,N0} \approx 1.2$ nm, for the number of amino acids N in the segment as $R_{g,N}^3 = R_{g,N_0}^3 N/N_0$), the radius of gyration of the θ-state is approximately 2.4 nm.

We convert the distribution for the radius of gyration into a distribution of the end-to-end distance using the approximation (13)

$$P_{\text{Sanchez}}(r) = \int p(r|R_g)P_{\text{Sanchez}}(R_g)dR_g, \qquad \textbf{[S3]}$$

where $p(r|R_g)$ is the conditional probability for a distribution of end-to-end distances given a value of R_g. $p(r|R_g)$ is taken as the distance distribution of two random points inside a sphere with a corrected radius of gyration $\delta \cdot R_g$:

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

where the phenomenological factor δ is chosen such that $\langle R_g^2\rangle = \langle r^2\rangle/6$ at the θ-point with Eq. **S3** and $P_{\text{Sanchez}}(r) = P_0(R_g)$, resulting in $\delta = 2.26$ in our case. Finally, $P_{\text{Sanchez}}(r)$ is used to fit the experimental mean FRET efficiencies by adjusting ε (Fig. S2). The average radius of gyration is then calculated from the resulting $P(R_g)$.

A comparison between the values of radius of gyration obtained with the two different treatments is reported in Fig. S1: significant differences are only observed in the case of the two variants of ProTα at very low ionic strength, where the chain is expanded far beyond the estimated θ-point, but this does not affect our conclusions significantly.

Polyelectrolyte theory. The simplest description that takes into account electrostatic repulsion considers the unfolded protein a polyelectrolyte, i.e. a polymer with only one type of charge (15–18). Ha and Thirumalai (18) showed that the effect of charges on the conformations of a polyelectrolyte chain can be described in terms of an effective excluded volume. Calculating the free energy and the dimensions of the chain by a standard self-consistent variational treatment, Ha and Thirumalai described the expansion factor α as

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} - f_{\text{el}}(\kappa)\alpha^2 = X, \qquad \textbf{[S5]}$$

where y provides an estimate of the three-body interaction, $f_{\text{el}}(\kappa)$ describes the electrostatic interactions as a function of the Debye screening length κ^{-1}, and X is related to the monomer excluded volume vb^3 through

$$X = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} vN^{0.5}. \qquad \textbf{[S6]}$$

In the limit of $\kappa b\alpha \gg 0$ in the presence of salt, the equation can be rewritten as

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} vN^{0.5} + 2\left(\frac{6}{\pi}\right)^{0.5} uZ^2N^{0.5}\frac{1}{(\kappa b)^2}, \qquad \textbf{[S7]}$$

where $u = l_B/b$ is the ratio of the Bjerrum length and the monomer length, and Z is the charge per monomer. Interpreting Z as the density of net charge over the chain and considering f and g as the fractions of positive and negative charges per monomer, $Z = f - g$, we obtain

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} N^{0.5}\left(\nu + 2\pi\frac{l_B(f-g)^2}{\kappa^2 b^3}\right). \qquad \textbf{[S8]}$$

As pointed out by Ha and Thirumalai, the right side of the equation can be considered an effective excluded volume term. This result is formally equivalent to neglecting attractive interactions in Eq. 4, such that the excluded volume can be expressed as

$$\nu^* b^3 = \nu b^3 + \frac{4\pi l_B(f-g)^2}{\kappa^2}. \qquad \textbf{[S9]}$$

The difference by a factor of 2 in the electrostatic term derives from slightly different approximations adopted in the calculations. The factor of 4 reported by Higgs and Joanny is also found by Muthukumar (19). We adopt this value for simplifying the comparison between polyelectrolyte and polyampholyte theory.

With this formalism, the dimensions of a polyelectrolyte chain can be analyzed analogous to polyampholyte theory in combination with Eqs. 3 and 5. The fits show that Eq. **S1** is a good approximation for ProTα, with its large proportion of glutamate and aspartate residues, and captures the rollover of R_g at low ionic strength. It still correctly predicts a small expansion of the chain for IN at low GdmCl concentrations. To fit the data of Csp*Tm*, however, it is necessary to compensate the effect of the electrostatic term by adjusting the solvent quality. In Fig. S3, we report the predicted rollover when a change in solvent quality is neglected, i.e. if we assume a constant value for the parameter ν. As pointed out in the *Discussion* of the main text, the KCl addition experiments (Fig. 5) indicate that in the case of Csp*Tm* (where there is not a clear dominance of one type of charge), attractive interactions between opposite charges within the polypeptide contribute to chain collapse. The sequences and charge distribution of all proteins investigated here are given in Table S1.

Results and Discussion. The values of all parameters obtained from fits with Eq. 5 to the results obtained with the Gaussian chain model and the modified Sanchez theory (Fig. S1) are given in Table S2, Table S3, Table S4, and Table S5. The effective binding constants for GdmCl of the protein variants investigated (Table S2, Table S3, Table S4, and Table S5) are between 0.2 and 1.3, in the range of binding constants reported previously (20–22), with the value for Csp*Tm* being slightly higher than for IN and ProTα. Similar results are obtained from fits of Eq. 2 to the urea data (Table S4 and Table S5). Considering that the preferential interaction of denaturants with the polypeptide is expected to depend on the nature of the side chains (21–28), e.g. through the complexation of guanidinium ions by acidic side chains (29), some variation in the effective binding constant with sequence composition may not be surprising. The values for the excluded volume νb^3 are similar for Csp*Tm* and IN; the larger value for ProTα is probably at least partially due to the presence of the large unlabeled segments in the polypeptide, which will exert an excluded volume effect on the labeled part.

1. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
2. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
3. Evstafieva AG, et al. (1995) Overproduction in *Escherichia coli*, purification and properties of human prothymosin alpha. *Eur J Biochem* 231:639–643.
4. Schuler B (2007) Application of single-molecule Förster resonance energy transfer to protein folding. *Methods Mol Biol* 350:115–138.
5. Nozaki Y (1972) The preparation of guanidine hydrochloride. *Methods Enzymol* 26 PtC:43–50.
6. Warren JR, Gordon JA (1966) On the refractive indices of aqueous solutions of urea. *J Phys Chem* 70:297–300.
7. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
8. Van Der Meer BW, Coker G III, Chen S-YS (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).
9. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
10. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
11. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
12. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11:83–93.
13. Ziv G, Haran G (2009) Protein folding, protein collapse, and Tanford's transfer model: Lessons from single-molecule FRET. *J Am Chem Soc* 131:2942–2947.
14. Sanchez IC (1979) Phase-transition behavior of the isolated polymer-chain. *Macromolecules* 12:980–988.
15. Odijk T (1977) Polyelectrolytes near the rod limit. *J Polym Sci Pol Phys* 15:477–483.
16. Skolnick J, Fixman M (1977) Electrostatic persistence length of a worm-like polyelectrolyte. *Macromolecules* 10:944–948.
17. Ha BY, Thirumalai D (1999) Persistence length of flexible polyelectrolyte chains. *J Chem Phys* 110:7533–7541.
18. Ha BY, Thirumalai D (1992) Conformations of a polyelectrolyte chain. *Phys Rev A* 46:R3012–R3015.
19. Muthukumar M (1987) Adsorption of a polyelectrolyte chain to a charged surface. *J Chem Phys* 86:7230–7235.
20. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226:491–505.
21. Tanford C (1970) Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
22. Schellman JA (2002) Fifty years of solvent denaturation. *Biophys Chem* 96:91–101.
23. Wallqvist A, Covell DG, Thirumalai D (1998) Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *J Am Chem Soc* 120:427–428.
24. Caflisch A, Karplus M (1999) Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure* 7:477–488.
25. Bennion BJ, Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci USA* 100:5142–5147.
26. Stumpe MC, Grubmuller H (2007) Interaction of urea with amino acids: Implications for urea-induced protein denaturation. *J Am Chem Soc* 129:16126–16131.
27. Auton M, Holthauzen LM, Bolen DW (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA* 104:15317–15322.
28. Nodet G, et al. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc* 131:17908–17918.
29. O'Brien EP, Dima RI, Brooks B, Thirumalai D (2007) Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: Lessons for protein denaturation mechanism. *J Am Chem Soc* 129:7346–7353.

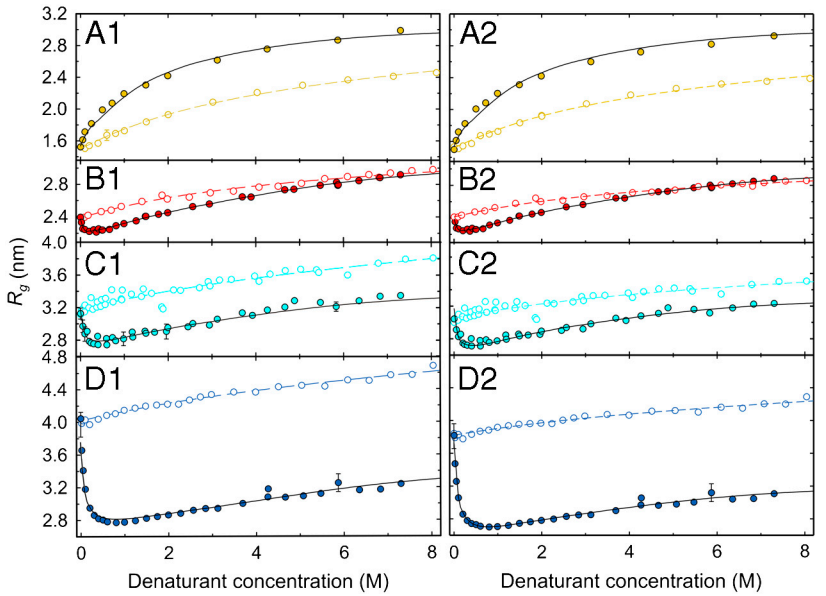

Fig. S1. Comparison of the apparent root-mean-square radii of gyration of the labeled protein segments using a Gaussian chain model (Eq. **S1**, *Left*, same fits as in Fig. 4) or the modified Sanchez theory (Eq. **S3**, *Right*). Data are reported as a function of GdmCl (filled circles) and urea (open circles) concentration, with (*A*) Csp*Tm* (yellow), (*B*) IN (red), (*C*) ProTαN (cyan), and (*D*) ProTαC (blue). Fits according to polyampholyte theory (Eq. **5**) are shown as lines. Note that the fits are performed based on thermodynamic activities, but plotted on a concentration scale.

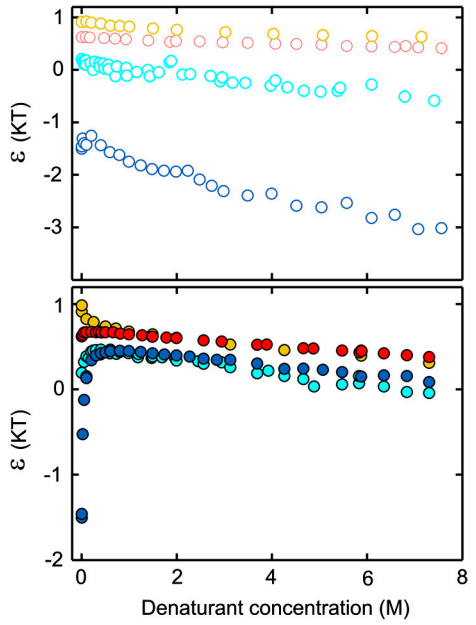

Fig. S2. Dependence of the effective interaction parameter ε in the modified Sanchez theory [Eq. **S2**] on the concentration of GdmCl (*Lower*) and urea (*Upper*) for Csp*Tm* (yellow), IN (red), ProTαN (cyan), and ProTαC (blue).

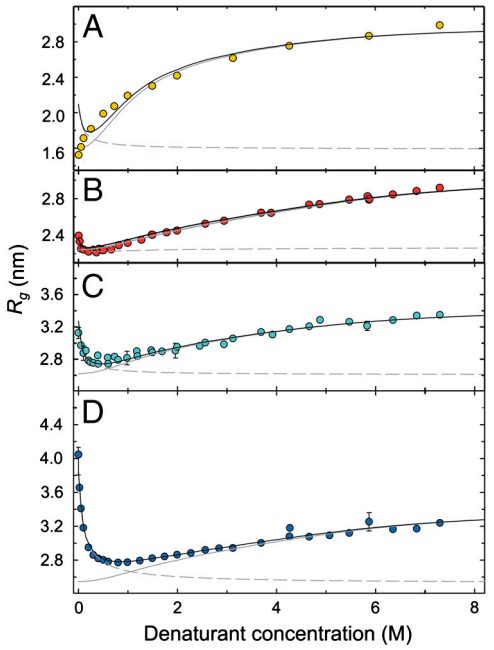

Fig. S3. Dependence of the apparent radii of gyration (R_g) of the labeled protein segments on the concentration of GdmCl, with (*A*) Csp*Tm* (yellow), (*B*) IN (red), (*C*) ProTαN (cyan), and (*D*) ProTαC (blue). In contrast to Fig. 4, fits according to polyelectrolyte theory with a fixed excluded volume (Eq. **S2**, black solid line) are shown. The two components of the fit to Eq. **S2**, corresponding to GdmCl binding and electrostatic repulsion, are indicated as continuous and dashed gray lines, respectively. The rollover in the fit of Csp*Tm* could be eliminated by a compensating pronounced variation in the excluded volume term ν [Eq. **S9**], corresponding to a strong change in solvent conditions over this range of GdmCl concentration. Note that the fits are performed based on thermodynamic activities, but plotted on a concentration scale.

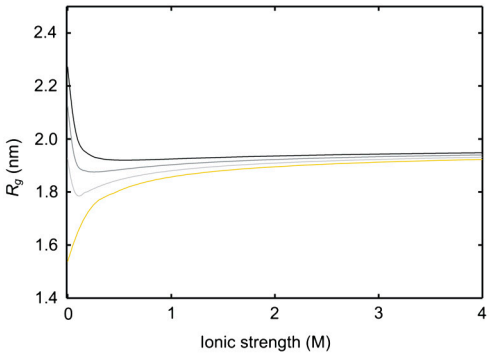

Fig. S4. Transition from collapse to expansion at low ionic strength with changing charge balance, illustrating the pronounced sensitivity for charge composition of the chain. The dependencies are calculated according to Eq. 5 with the following parameters: excluded volume $\nu b^3 = 0.2$ nm^3, number of segments $N = 53$, number of negatively charged segments $N_- = 14$, number of positively charged segments $N_+ = 8$ (Csp*Tm*, yellow line), $N_+ = 7$ (light gray line), $N_+ = 6$ (dark gray line), $N_+ = 5$ (black line).

Table S1. Sequences, charges, and labeling positions of the proteins investigated

A	B	C	D
Csp*Tm*	−2	−6	1 10 20 30 40 50 54 60 64 GPG **C**RGKVKWFDS KKGYGFITKD EGGDVFVHWS AIEMEGFKTL KEGQVVEFEI QEG**C**KGGQAA HVKV + + + − ++ +− − − − − + +− − − − + +
IN	−4	−8	0 1 10 20 30 40 50 56 GSH**C** FLDGIDKAQE EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVD**C** − −+ − − −+ + − +− −+ + − −
ProTαN (C2–C56)	−14	−18	2 10 20 30 40 50 MAHHHHHHS AALEVLFQGP M**S**DAAVDTSS EITTKDLKEK KEVVEEAENG RDAPANGNAN EENGEQEADN − − − − +− +−+ +− −− − +− −− − − −
ProTαC (C56–C110)	−27	−31	56 60 70 80 90 100 110 EVDEE**C**EEGG EEEEEEEEGD GEEEDGDEDE EAESATGKRA AEDDEDDDVD TKKQKTDED**D** − −−− −− −−−−−−−− − −−−− −−−− − − ++ −−−−−−− − ++ + −−−−

(A) Protein, (B) Net charge of interdye sequence, (C) Net charge of interdye sequence including the charges of the dyes, and (D) Amino acid sequence and positions where Cys residues were introduced for labeling (indicated in **bold**); charges at pH 7.4 are indicated below the respective amino acids.

Table S2. GdmCl concentration dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with polyampholyte theory (Eq. 5)

	K	νb^3 (nm^3)	ρ
Csp*Tm*	1.3 ± 0.2*	0.20 ± 0.01	0.56 ± 0.02
IN	0.34 ± 0.03	0.52 ± 0.01	0.48 ± 0.02
ProTαN	0.4 ± 0.2	1.14 ± 0.06	0.28 ± 0.04
ProTαC	0.2 ± 0.1	1.26 ± 0.08	0.29 ± 0.09

*The errors given represent only the uncertainty of the fit

Table S3. GdmCl concentration dependence obtained with Eq. S3 (Sanchez model, Fig. S1); parameters from fits with polyampholyte theory (Eq. 5)

	K	νb^3 (nm^3)	ρ
Csp*Tm*	1.3 ± 0.2	0.20 ± 0.01	0.56 ± 0.02
IN	0.43 ± 0.03	0.38 ± 0.01	0.32 ± 0.01
ProTαN	0.4 ± 0.1	1.02 ± 0.04	0.24 ± 0.02
ProTαC	0.5 ± 0.1	0.91 ± 0.05	0.25 ± 0.02

Table S4. Urea concentration dependence dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with binding model (Eq. 2)

	K	R_{g0} (nm)	ρ
Csp*Tm*	0.20 ± 0.02	1.47 ± 0.01	1.15 ± 0.06
IN	0.13 ± 0.02	2.39 ± 0.01	0.49 ± 0.04
ProTαN	0.10 ± 0.05	3.17 ± 0.02	0.5 ± 0.1
ProTαC	0.08 ± 0.03	4.02 ± 0.01	0.4 ± 0.1

Table S5. Urea concentration dependence obtained with Eq. S3 (Sanchez model, Fig. S1); parameters from fits with binding model (Eq. 2)

	K	R_{g0} (nm)	ρ
Csp*Tm*	0.20 ± 0.02	1.50 ± 0.01	1.02 ± 0.07
IN	0.18 ± 0.03	2.41 ± 0.01	0.32 ± 0.02
ProTαN	0.14 ± 0.07	3.17 ± 0.02	0.25 ± 0.06
ProTαC	0.10 ± 0.05	3.82 ± 0.01	0.29 ± 0.08

Table S6. GdmCl concentration dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with polyelectrolyte theory (Eq. S8)

	K	νb^3 (nm^3)	ρ
Csp*Tm*	2 ± 2 *	0.05 ± 0.01	0.8 ± 0.4
IN	0.39 ± 0.03	0.36 ± 0.01	0.36 ± 0.01
ProTαN	0.7 ± 0.2	0.86 ± 0.05	0.31 ± 0.02
ProTαC	0.5 ± 0.2	0.72 ± 0.08	0.34 ± 0.03

*The large error reflects the inadequacy of the model for Csp*Tm* with a constant value of the excluded volume (see Fig. S3).