Supporting Information for:

Protein folding, protein collapse and Tanford's transfer model:

Lessons from single-molecule FRET

Guy Ziv and Gilad Haran

This file contains:

Supporting Text.

Supporting Table 1.

Supporting Figure 1.

Supporting Figure 2.

Supporting Text

Calculation of the conditional probability $P(R_{ee}|S)$. This conditional probability represents the distribution of the end-to-end distance (R_{ee}) for a chain with a fixed radius of gyration (S). We are not aware of an analytical formula for this distribution. In the spirit of the mean-field Sanchez model, we adopt a simple model in which $P(R_{ee}|S)$ is taken as the distance distribution of two random points inside a sphere of radius $g \cdot S$:

$$P(R_{ee}|S) = \frac{1}{g \cdot S}\left(3\left(\frac{R}{g \cdot S} \right)^2 - \frac{9}{4}\left(\frac{R}{g \cdot S} \right)^3 + \frac{3}{16}\left(\frac{R}{g \cdot S} \right)^5 \right) \qquad [S1]$$

The phenomenological factor g corrects the statistics for the case of an ideal chain, in which $P(R_{ee})$ is a Gaussian obeying $\left\langle R_{ee}^{\,2} \right\rangle_0 = 6\left\langle S^2 \right\rangle_0$. To find the correction factor, we insert Equation S1 into the expression $P(R_{ee}) = \int P(R_{ee}|S)P(S)dS$ and evaluate this expression with $P(S) = P_0(S)$. We use a nonlinear regression algorithm to best match the Gaussian distribution and the model. It can be shown that the value of g should not depend on n.

Finite-size correction to CG transition point. The coil-globule transition occurs at α^2=1 only for very long polymers**.** The number of amino acids (corresponding to polymer length in the Sanchez theory) of the proteins studied in this work ranges between 64 to 155, and therefore the position of the CG transition may differ from

α^2=1. Following Sanchez, one can use the Landau theory of phase transitions, demanding that the first three derivatives of the free energy $G(\varepsilon)$ vanish to get the following equation**:**

$$\left(\alpha_{CG}{}^3 - \phi_0\right)^2\left(1 - \frac{35}{54}\alpha_{CG}{}^2\right) = \frac{1}{2}n\phi_0{}^2$$

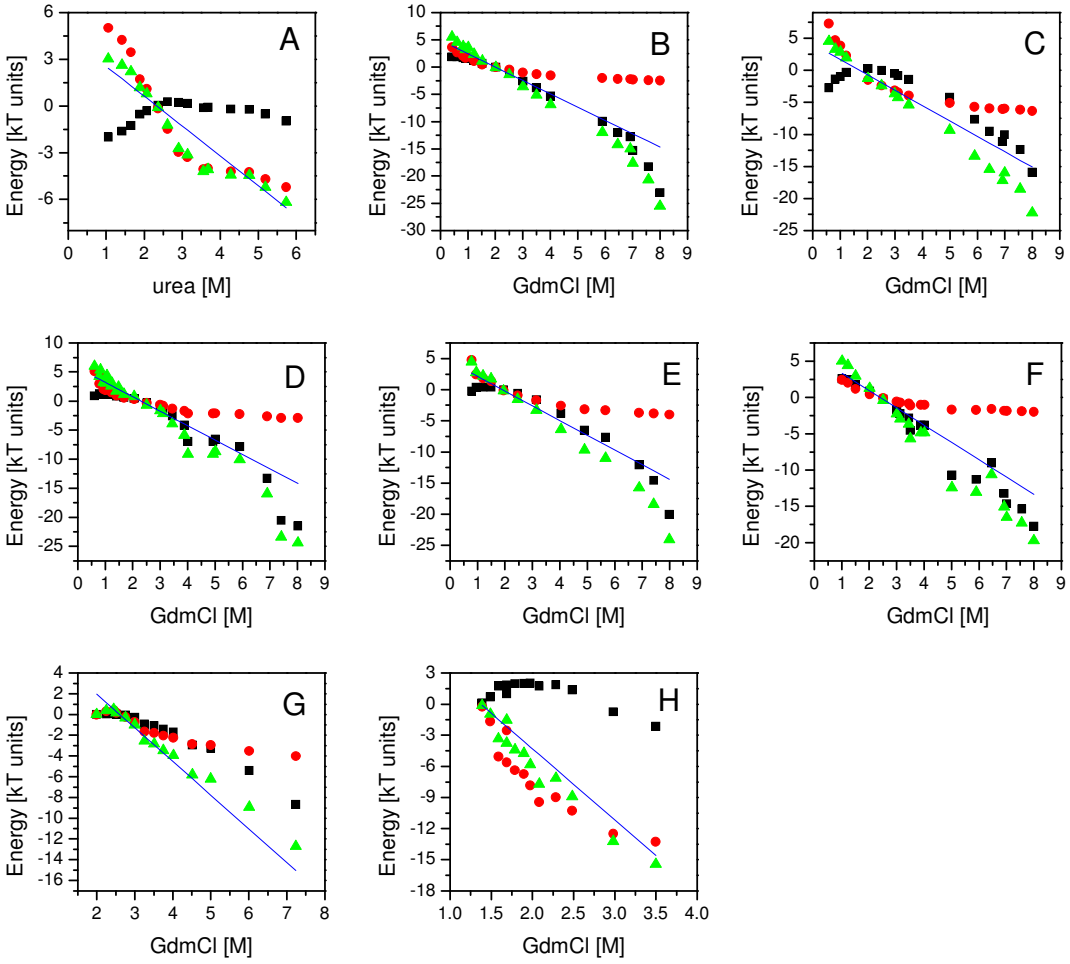
Expanding this equation to first order in ϕ_0 we find that the CG transition occurs at α_{CG}=1+19/22·ϕ_0. Using the expression for ϕ_0 in the main text, the CG transition point is at α^2≈1.2 for the set of proteins studied. The corresponding denaturant concentrations, D_{CG}, were found by extrapolation of α^2 using the linear fits to ε and are listed in Table 1 of the main text.

Supporting Table 1: m and $\tilde{m}$ values of analyzed proteins

Number	Protein	m/*n* [1] [kcal/mol/M]	$\tilde{m}$ [2] [kcal/mol/M]
1a-1b	Barstar	0.0207	0.0545
		0.0126	0.0166
2a	Csp*Tm*	0.0265	0.0561
2b-2f	Csp*Tm*	0.0216	0.0703
		0.0212	0.0714
		0.0219	0.0748
		0.0209	0.0711
		0.0210	0.0670
3	Im9	0.0117	0.0370
4a	Protein L	0.0233	0.0642
4b	Protein L	0.0295	0.0569
5	RNaseH	0.0261	0.0665

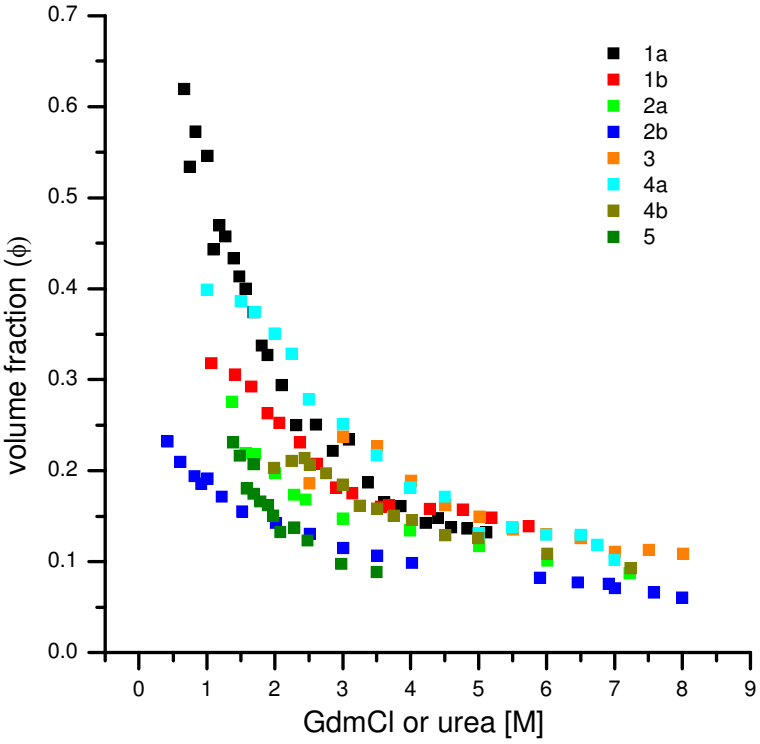
[1] m-values as reported in original references (see Table 1 of main text), or obtained from a fit of reported denaturation data to a two-state model, and divided by the number of residues, *n*.
[2] Obtained from the analysis delineated in the main text.

Supporting Figure 1



Thermodynamic functions of collapse calculated from smFRET results for those proteins not shown in Figure 2 of main text. Values of the free energy of collapse, $-\Delta G_{U\to C}$, as a function of denaturant concentration, are shown as green triangles. The enthalpic contributions $-\Delta E_{U\to C}$ and the entropic contributions $k_B T\cdot\Delta S_{U\to C}$ are shown as black squares and red circles, respectively. The interpolated value of each thermodynamic function at D_{50} is subtracted from all points. Solid lines show the denaturant dependence of the free energy of folding, as published in the literature. A. Barstar denatured in urea (labeled 1b in Table 1 of main text). B-F. CspTm denatured in GdmCl (2b-2f in Table 1). G. Protein L denatured in GdmCl (4b in Table 1). H. RNaseH denatured in GdmCl (5 in Table 1).

Supporting Figure 2



Volume fraction ϕ as a function of denaturant concentration. ϕ values for the series of proteins analyzed in the main text were obtained using the theory of CG transition in the denatured state, based on smFRET data. The numbers in the legend match the numbers in Table 1 of the main text.