Biophysical Journal, Volume *119*

Supplemental Information

Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins

Xiangze Zeng, Alex S. Holehouse, Ashutosh Chilkoti, Tanja Mittag, and Rohit V. Pappu

GaussianClusterV2.0.zip is a zip archive of the commented MATLAB code and the data files used to extract system-specific phase diagrams using our adaptation of the Gaussian Cluster Theory. All relevant details are found within the README file in the archive.

Figure S1: Pseudo code for extracting *B,* and *w*

Two approaches were developed to extract *B,* and *w* based on analysis of $\sqrt{\langle R_g^2 \rangle}$ versus *T* that we obtain from simulations. In Approach 1, we set $\frac{\left(T^* - T_\theta\right)}{T^*} Bn_K^{0.5} = -0.35$ irrespective of the value of *w.* In Approach, the choice of $\frac{\left(T^* - T_\theta\right)}{T^*} Bn_K^{0.5}$ is coupled to the choice made for *w*, although the equality used in Approach 1 provides bounds on the value for $\frac{\left(T^* - T_\theta\right)}{T^*} Bn_K^{0.5}$. The relation between $\frac{\left(T^* - T_\theta\right)}{T^*} Bn_K^{0.5}$ and *w* is obtained by numerically calculating the saddle point of the contraction ratio profile from GCT at different values for *w*. These two approaches give very similar results and the choice of approaches does not affect our conclusions. The pseudo code for approaches 1 and 2 is referenced as **Figure S1** in the main text.

A. Pseudo code for Approach 1:

for T^* in $[T_{min}, T_\theta]$, where T_{min} is the lowest simulation temperature:

do

get *B* based on $\frac{\left(T^* - T_\theta\right)}{T^*} Bn_K^{0.5} = -0.35$;

convert the simulated of $\sqrt{\langle R_g^2 \rangle}$ versus *T* curve to a contraction ratio profile as a function $\tau Bn_K^{0.5}$ and denote it as $\alpha_s^{sim}\left(\tau Bn_K^{0.5}\right)$;

Next, perform a parameter scan for *w*;

For *w* in $[w_{min}, w_{max}]$ ([0.001, 0.04] has proven to be sufficient)

do

Get the contraction ratio profile $\alpha_s^{GCT}\left(\tau Bn_K^{0.5}\right)$

Compute the difference between $\alpha_s^{sim}\left(\tau Bn_K^{0.5}\right)$ and $\alpha_s^{GCT}\left(\tau Bn_K^{0.5}\right)$;

Since the minimum of $\alpha_s^{GCT}\left(\tau Bn_K^{0.5}\right)$ always larger than the minimum of $\alpha_s^{sim}\left(\tau Bn_K^{0.5}\right)$, a scaling factor λ is used to best $\alpha_s^{sim}\left(\tau Bn_K^{0.5}\right)$ and $\alpha_s^{GCT}\left(\tau Bn_K^{0.5}\right)$ for λ in [0.3, 0.9]

do

compute difference $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\lambda\left[\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)-1\right]+1$

done

The minimum difference between $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\lambda\left[\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)-1\right]+1$ is chosen as the difference between $\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$

done

done

The parameters (B, w) that yield the minimal difference between $\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ is used for the system.

B. Pseudo code for Approach 2:

w in $[w_{\min}, w_{\max}]$

do

Get $\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ for the given w;

Get the position for the saddle point, namely the value of $\frac{\left(T^*-T_\theta\right)}{T^*}Bn_{\mathrm{K}}^{0.5}=a$ for T^* in $[T_{\min}, T_\theta]$,

do

Compute B based the value of $\frac{\left(T^*-T_\theta\right)}{T^*}Bn_{\mathrm{K}}^{0.5}=a$

Compute $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ for a given B and w;

Use the method in Approach 1 to calculate the difference between $\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$;

done

done

Again, the parameters (B, w) that yield the minimal difference between $\alpha_s^{\mathrm{GCT}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ and $\alpha_s^{\mathrm{sim}}\left(\tau Bn_{\mathrm{K}}^{0.5}\right)$ is used for the system.

Table S1. Parameters for the temperature dependent free energy of solvation

Unit of interest	Reference temperature T_0 (K)	Reference free energy of solvation $\Delta G_{FES}(T_0)$ kcal/mol	Enthalpy of solvation ΔH_{FES} kcal/mol	Heat capacity of solvation Δc_P cal /mol-K
Backbone	298	-10.1	-17.0	4.00
Thr	298	-5.0	-12.5	4.66
Pro	298	2.0	-5.4	7.05
Lys	298	-100.9	-111.8	5.08
Ala	298	1.9	-3.3	4.96
Met	298	1.4	-10.0	5.59

Table S2. Details of all-atom simulation parameters for each system

System	Number of replicas	Temperature schedule	Number of independent simulations
Q_{20}	18	280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450 K	4
Q_{30}			
Q_{40}			
Q_{50}			
Q_{60}			
Q_{70}			
$(QGQSPYG)_9$			
$(TPKAMAP)_9$	18	300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 420, 440, 460, 480, 500, 520, 540 K	3

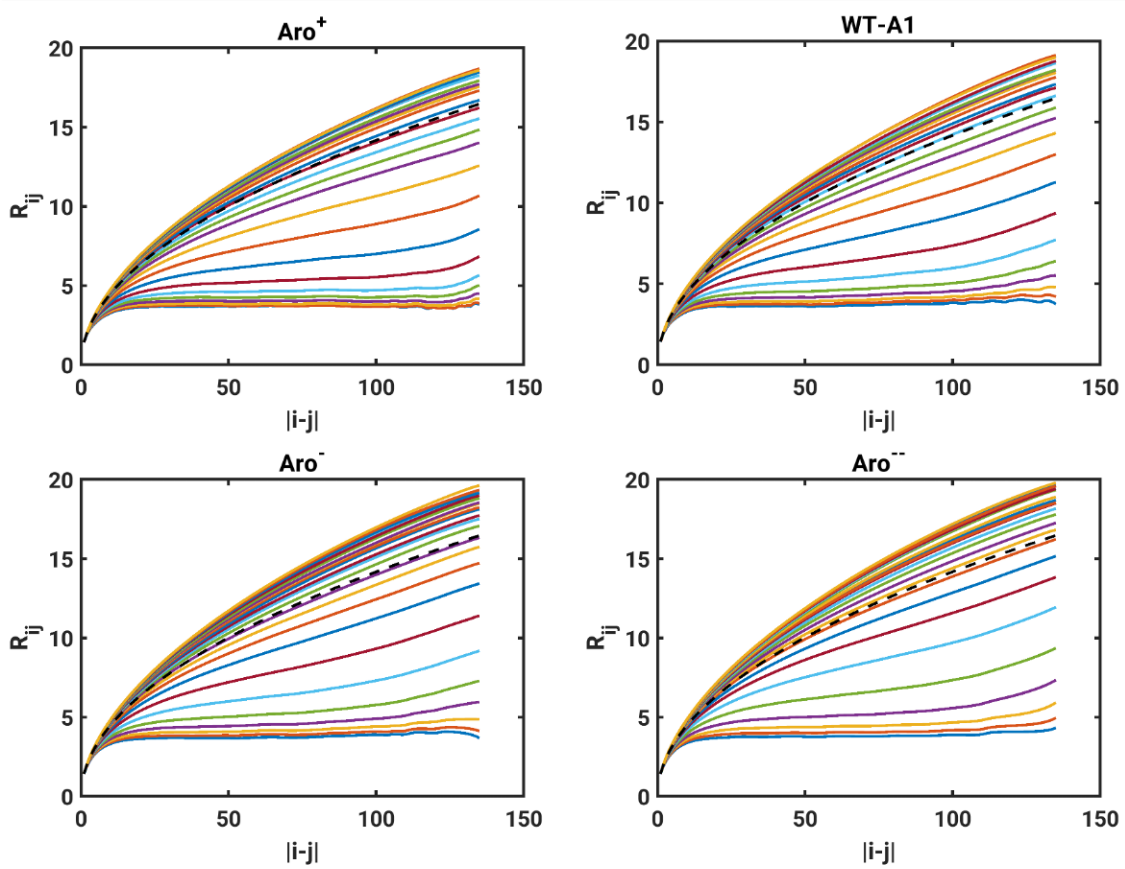

Figure S2. Internal scaling distance for Aro^{+}, WT-A1, Aro^{-}, and Aro^{--} variants. Dashed line denotes the reference internal scaling profile: $\left\langle\left\langle R_{ij}\right\rangle\right\rangle = R_0\left|j-i\right|^{0.5}$. Here, R_0 is the average value obtained across all the temperatures when we set |j–i| = 1.

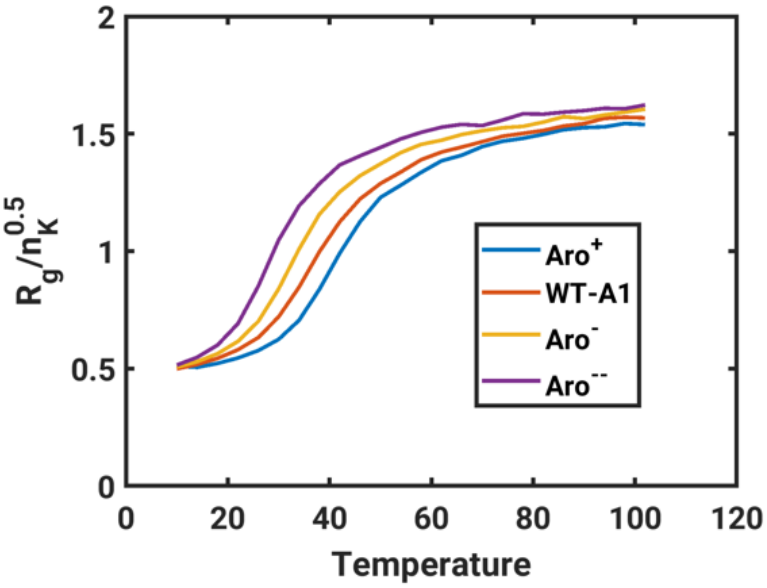

Figure S3. Plot of normalized R_g for Aro$^+$, WT-A1, Aro$^-$, and Aro^{--}.

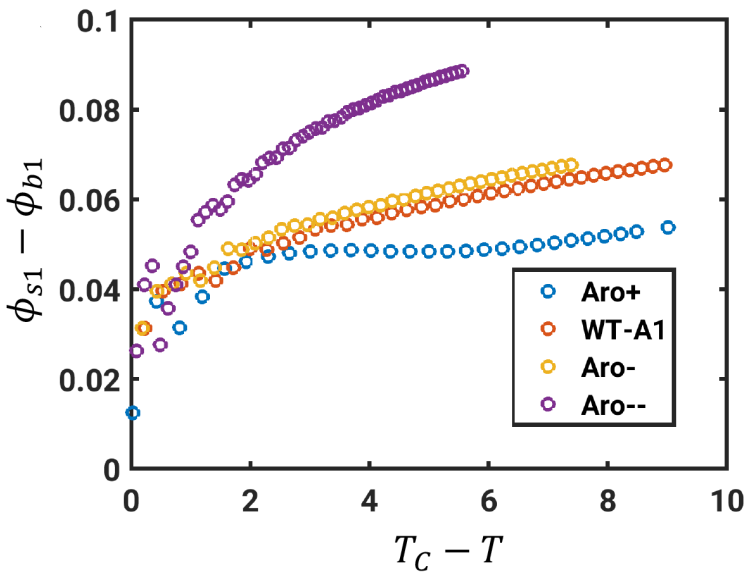

Figure S4: Analysis of how the width of the metastable regime $\Delta\phi_m(T)$ varies with $(T - T_c)$ for each of the four sequences, Aro$^+$, WT-A1, Aro$^-$, and Aro^{--}.

Sensitivity analysis for estimates of *B* and *w*

WT-A1 is used to test the robustness of the fitting result. There are 11 sampled temperatures below T_θ for A1-WT. The robustness is tested as following:

a) A subset of the 11 sampled temperatures which contains N randomly non-repeating temperatures is chosen at random.
b) This subset is used to fit *B* and *w*.
c) For each N in [5, 10], repeat a) and b) 10 times, and then get the mean value of *B* and *w*: B_N^{mean} and w_N^{mean}, and the corresponding standard deviation B_N^{std} and w_N^{std}.
d) Plot B_N^{mean} and w_N^{mean} as a function of N. B_N^{std} and w_N^{std} are plotted as the error bar. A complete analysis from a) to d) is denoted as one trial.
e) Repeat a)-d) 4 times, namely, perform 4 trial analysis.

Results are shown in **Figure S5**. Black dashed line indicates *B* or *w* fitted by using all the data. The fitting result becomes convergent when N>6.

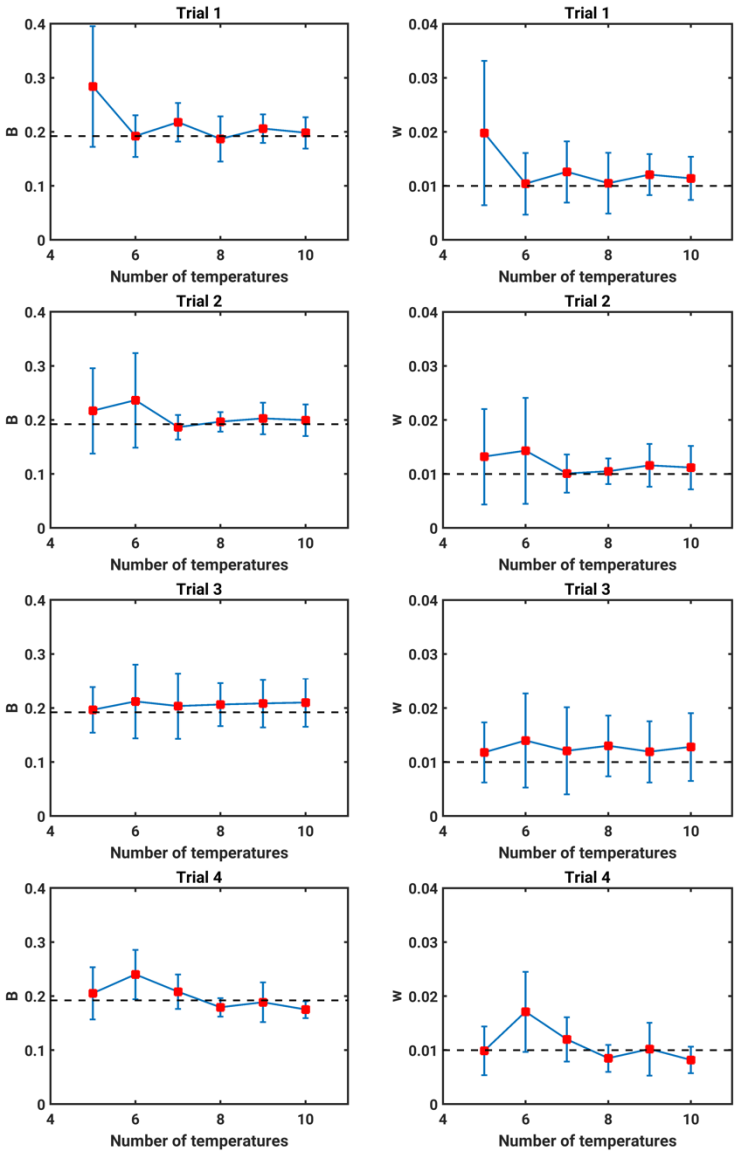

Figure S5: Sensitivity analysis of for estimates of *B* and *w*. Red square indicates the mean values of *B* or *w* for the subset with same number of sampled temperatures. Error bar indicates the stand deviation. Black dashed line indicates *B* or *w* fitted by using all the data.

Fitting result by using different T^*

We assess the impact of different values of T^* on the fitting procedure used to obtain B and w. The set of parameters $\{T^*, B, w]$ that best fits the contraction ratio profile from simulation (α_S^{sim}) is chosen as the final parameter to calculate the phase diagram. The searching range for T^* is $\left[T|_{\alpha_S^{sim}=0.5(1+\min(\alpha_S^{sim}))} - 10, T|_{\alpha_S^{sim}=0.5(1+\min(\alpha_S^{sim}))} + 10\right]$. Here, $\min\left(\alpha_s^{sim}\right)$ is the minimum value of the contraction ratio from the simulations. For the UCST system, $\min\left(\alpha_s^{sim}\right)$ is the contraction ratio at the lowest sampled temperature. For the LCST system, it is the value of the contraction ratio at the highest sampled temperature. **Figure S6** shows the fitting result by using different T^* for WT-A1 system. Approach 2 in Figure S1 is used for the fitting process.

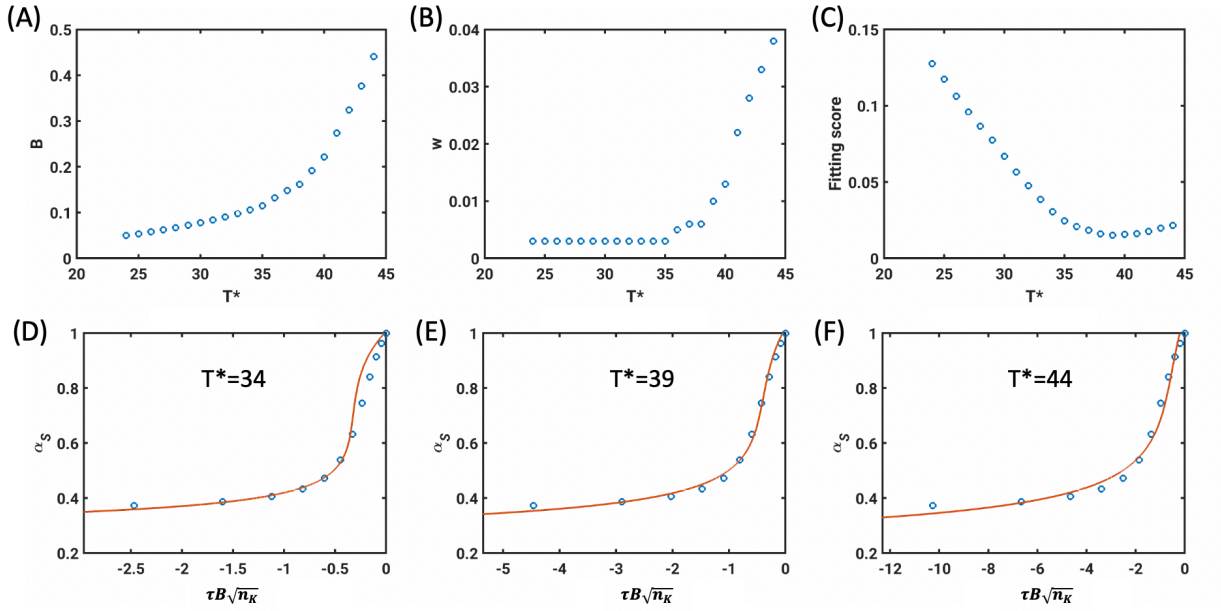

Figure S6: Fitting result by using different T^* for WT-A1 system. (A)-(C) B, w and the fitting score as a function of T^*. The fitting function is defined as the mean absolute difference between α_S^{simul} and α_S^{GCT}, detailed information about calculating the difference between α_S^{simul} and α_S^{GCT} can be found in **Figure S1**. $T^* = 39$ has the minimum fitting score, thus B and w fitted from $T^* = 39$ are chosen as the final parameter to calculate the phase diagram.**(D)-(F)** Comparison of α_S^{simul} and α_S^{GCT} at different T^*.