Supporting Information

for

Confinement and Crowding Effects on Folding of a Multi-domain Y-family DNA Polymerase

Xiakun Chu[1], Zucai Suo[2] and Jin Wang[1*]

[1] Department of Chemistry
State University of New York at Stony Brook, Stony Brook, New York 11794, USA

[2]Department of Biomedical Sciences, College of Medicine
Florida State University, Tallahassee, Florida 32306, USA

[*] jin.wang.1@stonybrook.edu

Contents

1 Native Contacts in DPO4

	F domain	P domain	T domain	LF domain	Linker
F domain	144	36	0	0	0
P domain	36	256	32	2	32
T domain	0	32	130	22	9
LF domain	0	2	22	257	11
Linker	0	32	9	11	2

Table S1: The native contact numbers of the intra- and inter-domains, as well as the flexible linker in DPO4. The total native contact number of DPO4 is 933, among which the intra-domain native contact number is 787 and the number of inter-domain native contacts that are mostly formed by the sequential neighboring domains, is 90. Only 2 inter-domain contacts are formed by non-sequential neighbor domains (the P-LF inter-domain interface). The number of contacts formed by the flexible linker is 54.

2 DPO4 folding in bulk

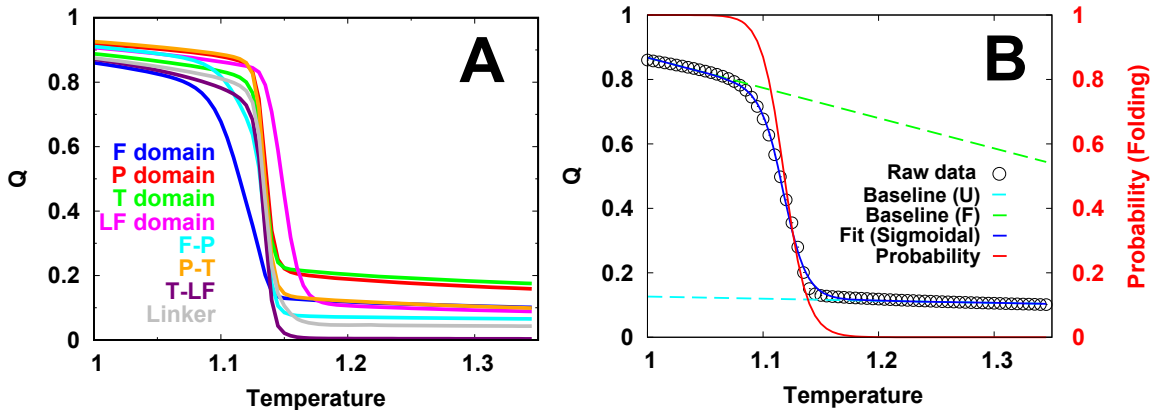

Figure S1: (A) Melting curves of Q for each intra- and inter-domain of DPO4 in bulk. (B) Melting curve for the F domain of DPO4 and its fitting to two-state sigmoidal transition to obtain the folding probability curve $p^I(T)$, where I is the index of intra- or inter-domain and T is the temperature. The data for the F domain shown here are chosen for illustrating and similar results can be observed for other intra- and inter-domains under different conditions. The fitting procedure follows the standard protocol instructed by Sborgi et al.[1].

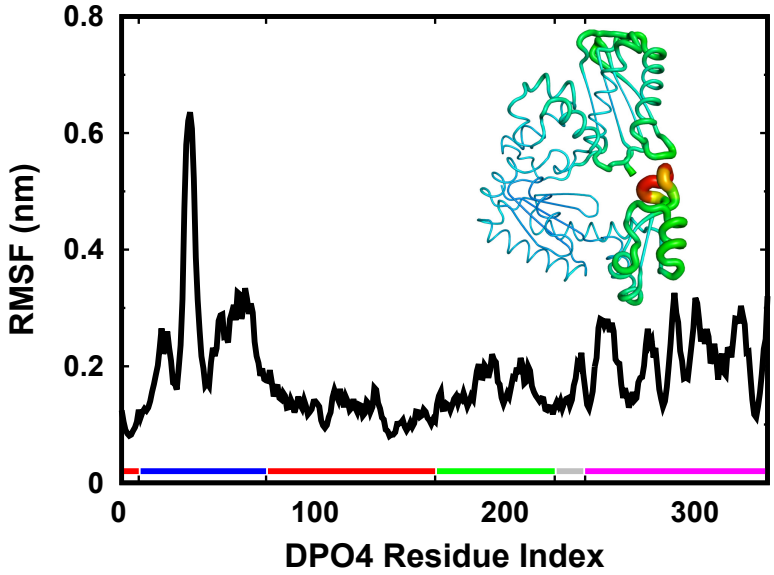

Figure S2: Root mean square fluctuation (RMSF) of DPO4 in bulk. RMSF is calculated based on the constant temperature simulations performed at room temperature, of which the value is estimated by the linear interpolation from rough connection between simulation and experiments: $T_{Room}(Sim) = T_f(Sim)/T_f(Exp) \times T_{Room}(Exp)$. As folding temperature of DPO4 in simulation $T_f(Sim)$ is 1.135 (reduced energy unit) from the heat capacity curve (Figure 2 in the main text) and the experimental folding temperature $T_f(Exp)$ of DPO4 is 369±1 K from thermal denaturation measurements[2], we can then deduce $T_{Room}(Sim) = 0.90$ (if assumes $T_{Room}(Exp) = 293$ K). Different domains of DPO4 are marked near to the x-axis with same color scheme used in Figure 1 in the main text. The native structure of DPO4 is shown insert with coloring (drawing) from blue (thin) to red (thick), corresponding to the RMSF values from low to high.

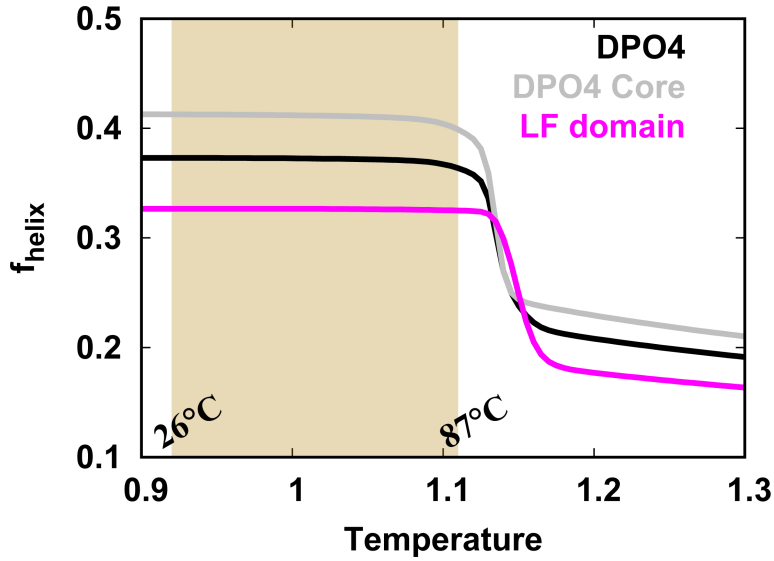

Figure S3: The fractions of helical content f_{helix} of DPO4, DPO4 Core and LF domain as a function of temperature. DPO4 Core is made up by the F, T and P domains. The f_{helix} as the function of temperature are reminiscent of the CD melting curves measured in[2]. The brown shadow region represents the experimental temperature range from 26 to 87 °C, where the fractions of helical content for DPO4, DPO4 Core and LF domain remain almost unchanged. The estimation on mapping of experimental temperature to simulation temperature is same as described in the Figure S2. The calculation of helical formation for a C_α-level coarse-grained structure is based on the algorithm suggested by de Sancho and Best[3].

3 Effects of confinements on DPO4 folding

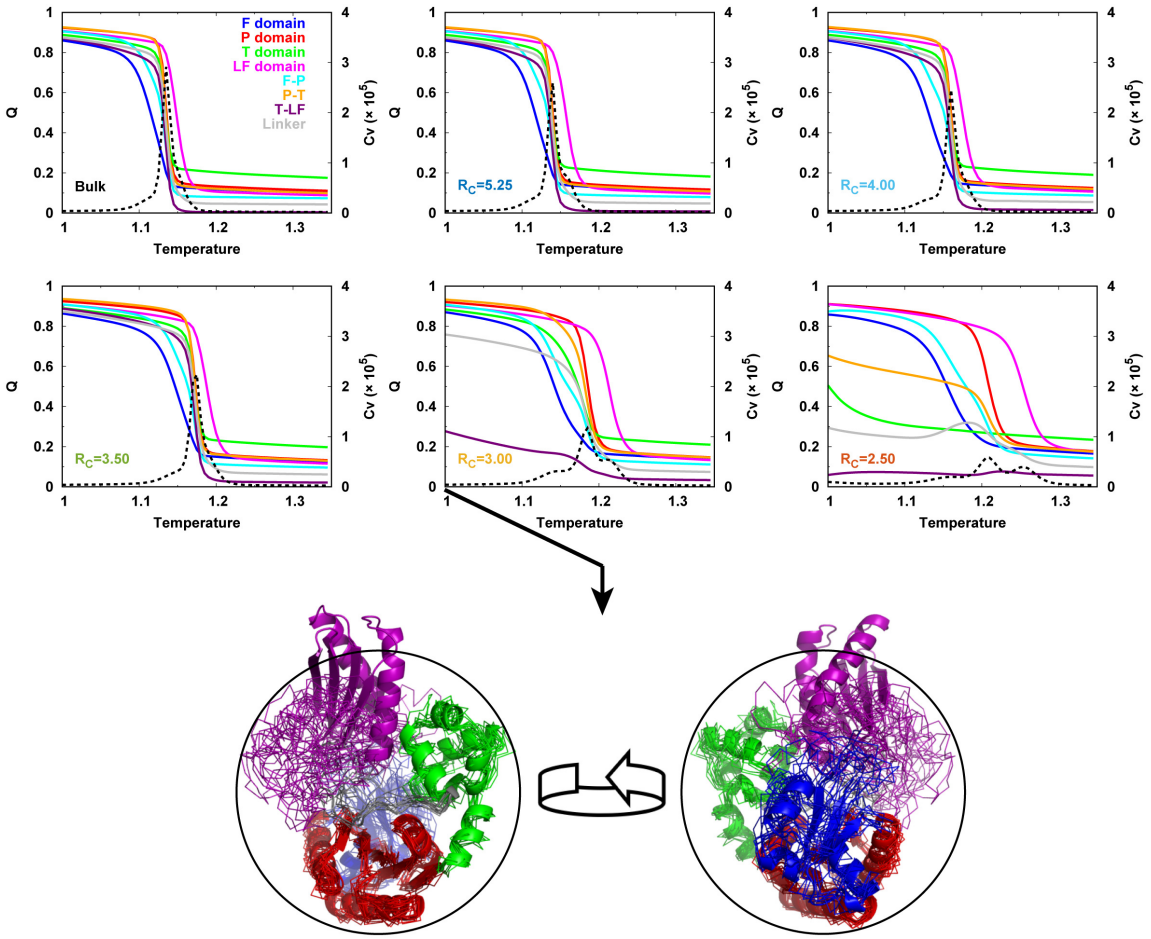

Figure S4: Melting curves of Q for each intra- and inter-domain of DPO4 under different strengths of confinement. The DPO4 structures at the bottom illustrate DPO4 distorted by the strong spherical confinement. 10 representative DPO4 structures shown in ribbons were extracted from the simulation trajectory at low temperature and superimposed to the native DPO4 structure, which is shown in cartoon. The circle represents the spherical confinement (R_C=3.00 nm).

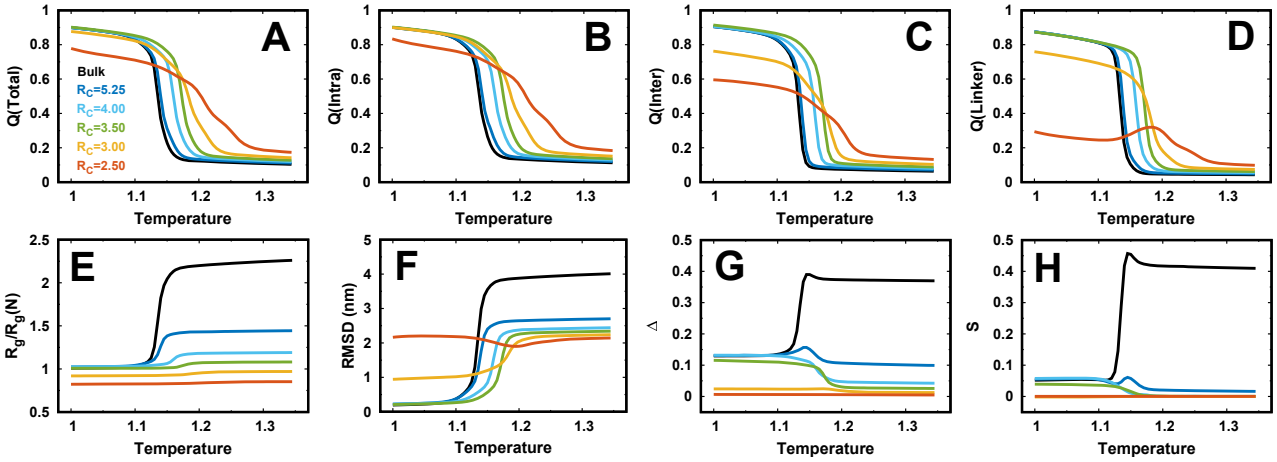

Figure S5: Melting curves of DPO4 for different order and shape parameters under different confinements.

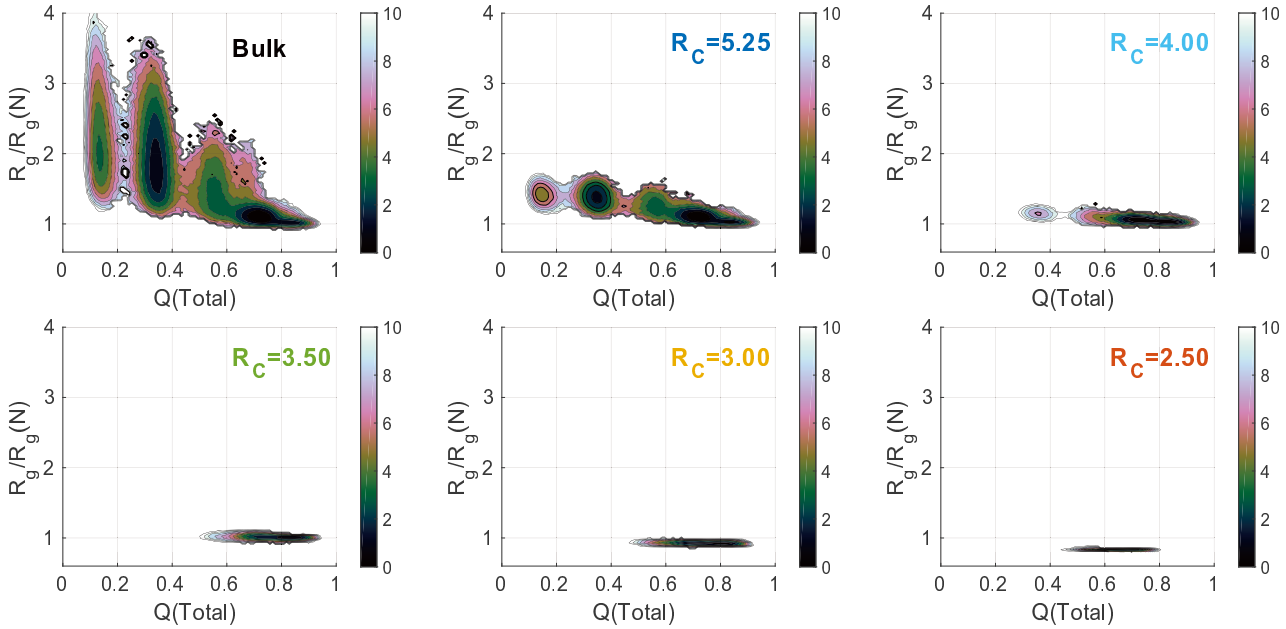

Figure S6: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g under different confinements at the bulk folding temperature T_f^{bulk}. $R_g(N)$ is the R_g of DPO4 at native PDB structure[4]. Free energy is in the unit of kT_f^{bulk}.

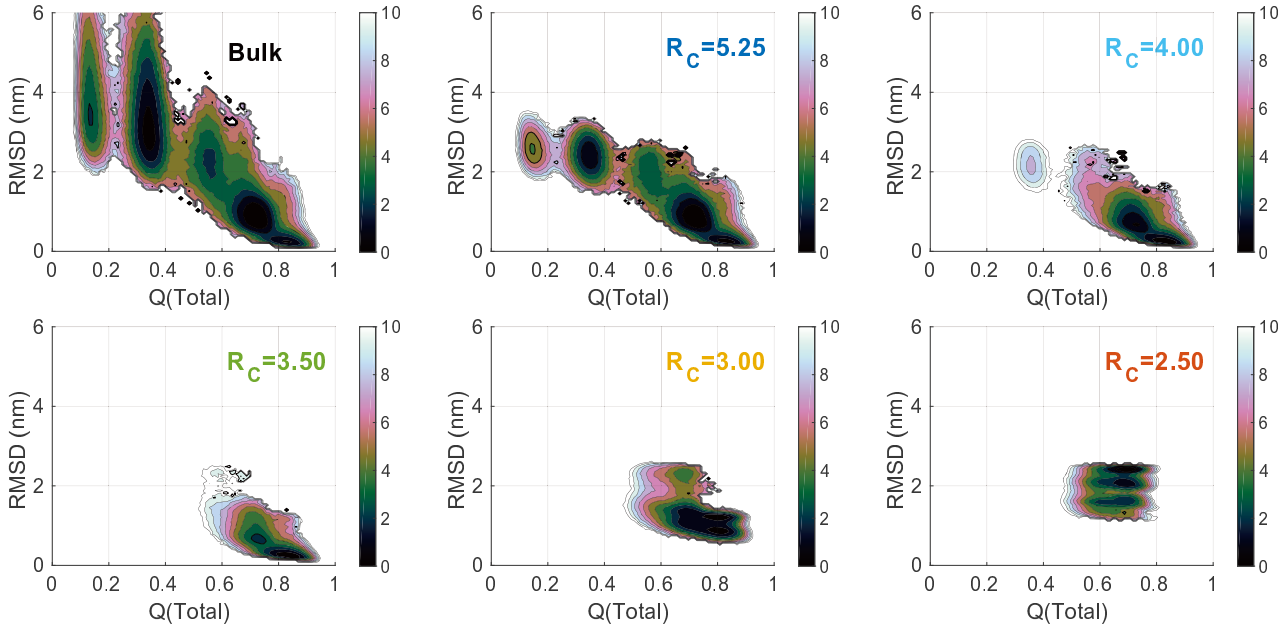

Figure S7: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ under different confinements at the bulk folding temperature T_f^{bulk}. Free energy is in the unit of kT_f^{bulk}.

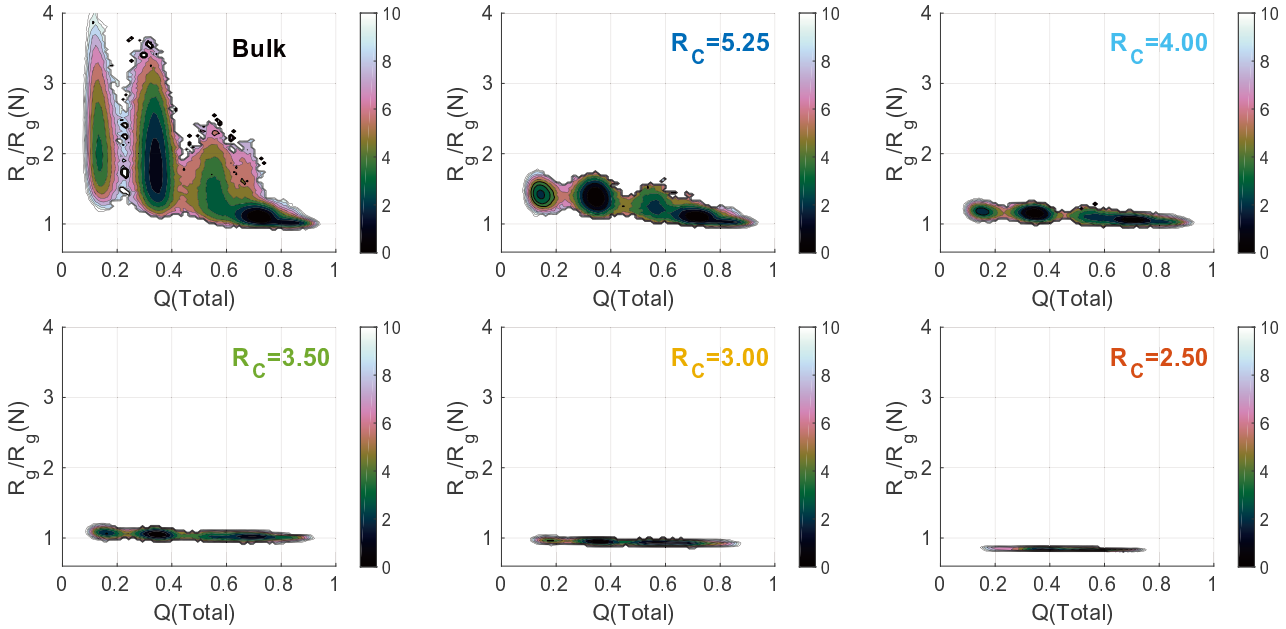

Figure S8: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g under different confinements at the corresponding folding temperatures. Free energy is in the unit of kT_f.

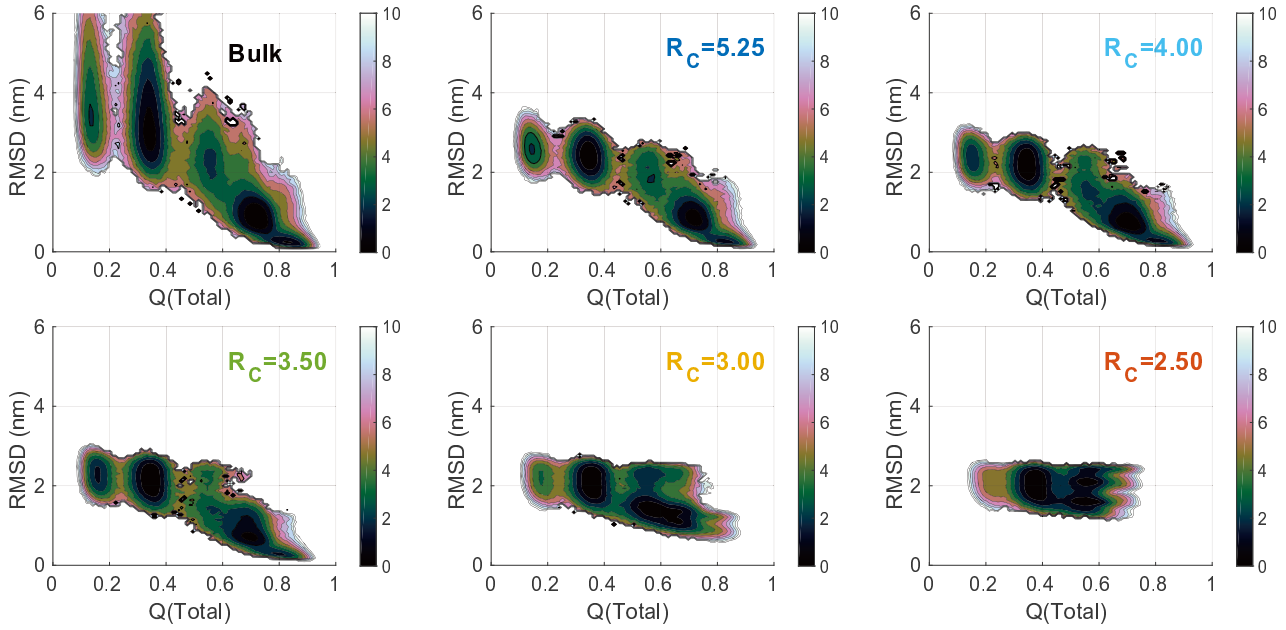

Figure S9: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ under different confinements at the corresponding folding temperatures. Free energy is in the unit of kT_f.

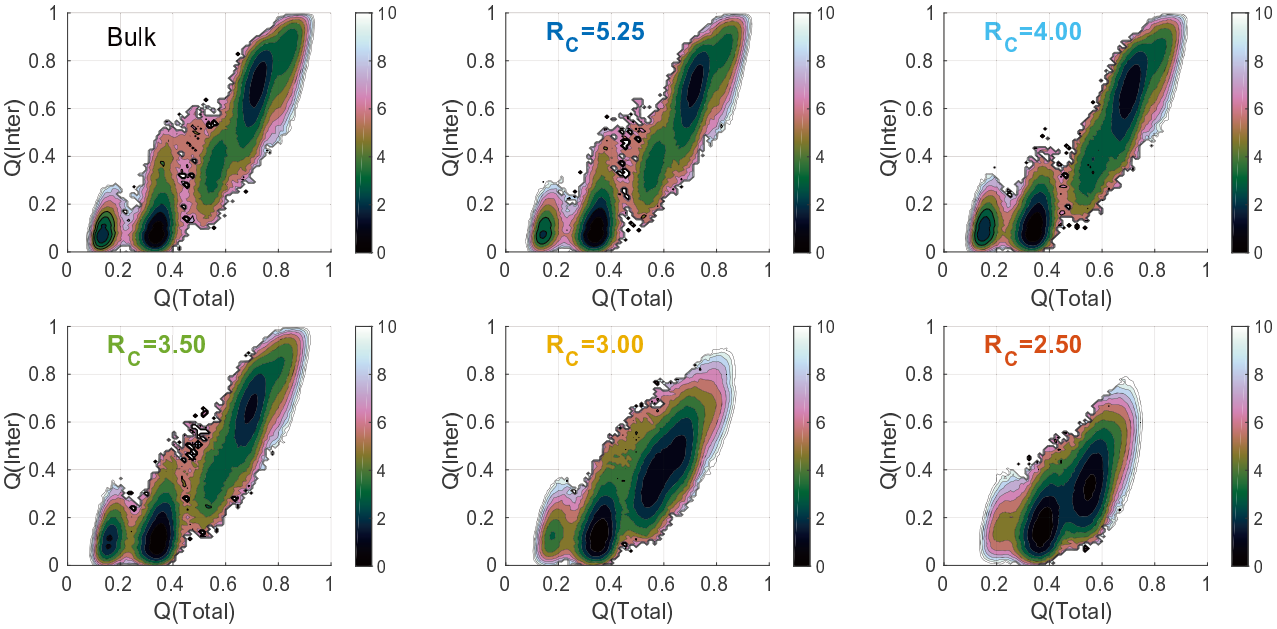

Figure S10: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $Q(Inter)$ under different confinements at the corresponding folding temperatures. Free energy is in the unit of kT_f. The two folding pathways from intermediate states I_3 to I_2 gradually merge to a single one that proceeds with transition state TS_1, as the strength of confinement increases (R_C decreases) (The pathways are illustrated in Figure 3 in the main text).

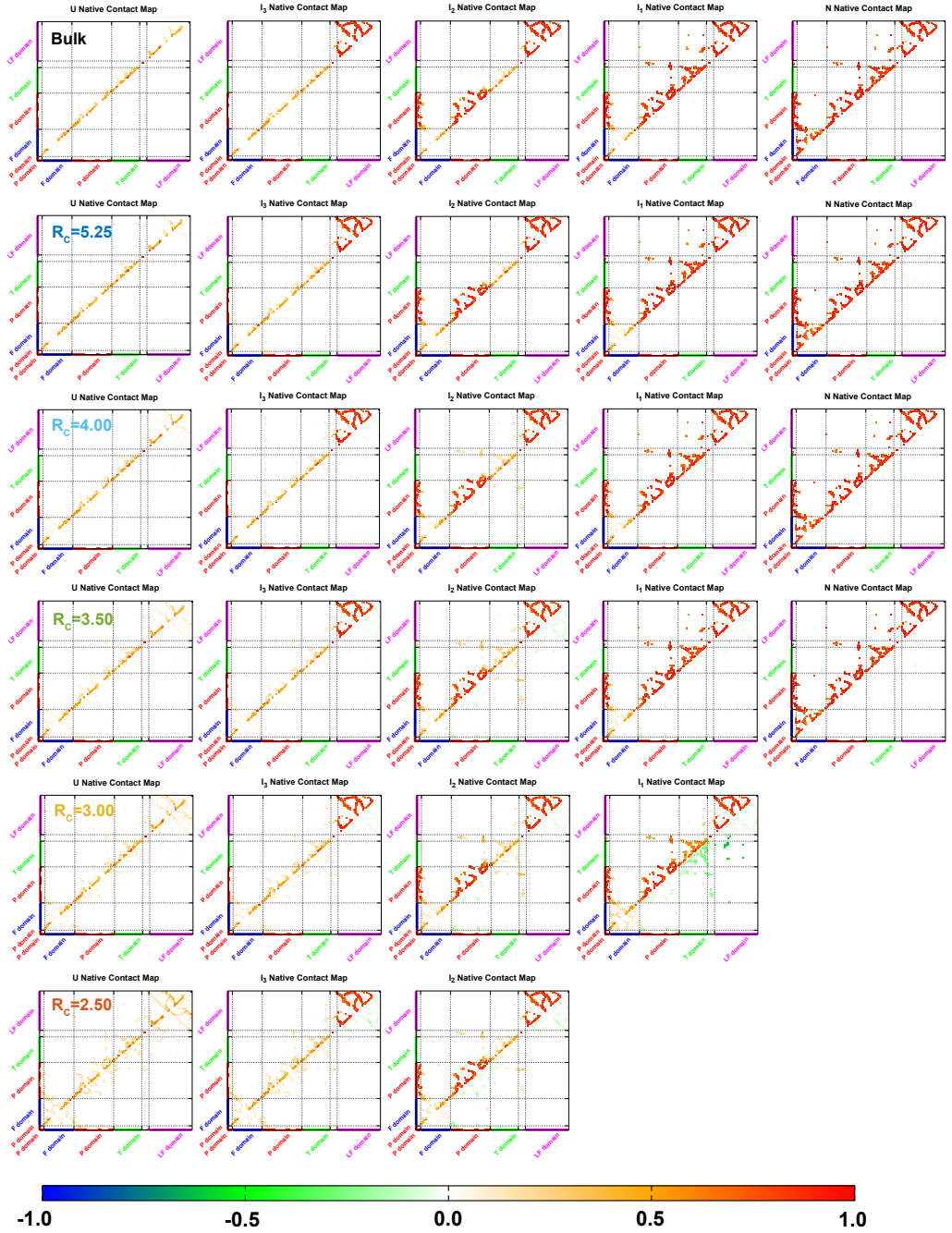

Figure S11: Native contact maps of DPO4 in each state identified from 1D free energy landscape shown in Figure 2 in the main text under differet confinements. In each sub-figure, top left shows the native contact probability and bottom right shows the differences to that in bulk. For each state, DPO4 preserves quite similar structural characteristics in respect of native contact formation for different confinements: (1) DPO4 in N states stays and fluctuates around the native PDB structure, associated with apparent dynamics within the F domain, which has been modelled very flexible within residues 34 to 39; (2) DPO4 in I_1 states has unfolded F domain, with the T-LF interface and linker starting to show fluctuations. (3) DPO4 in I_2 shows unfolded F and T domains, as well as the interfaces involved by these two domains, which are the F-P and P-T, and at the same time the native contacts in the T-LF and linker are entirely broken; (4) DPO4 in I_3 has only folded LF domain, while the other domains/interfaces are unfolded; (5) DPO4 in U is completely unfolded with very little native contact formation. Under strong confinements, DPO4 can not fully explore the whole states as those in bulk, so the N state at $R_C = 3.00\ nm$ and N, I_1 states at $R_c = 2.50\ nm$ are not shown.

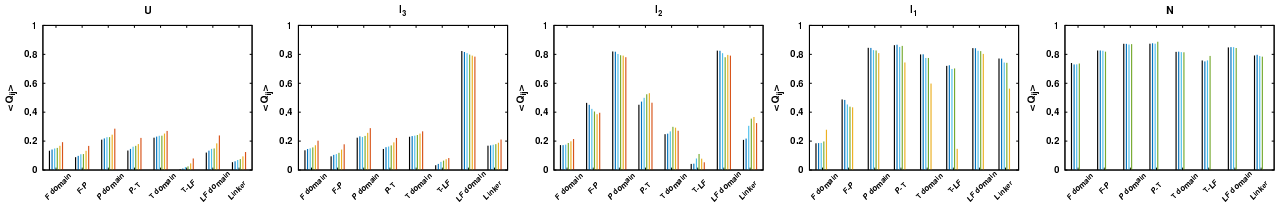

Figure S12: Native contact formations for each domain and interface of DPO4 during folding under different confinements.

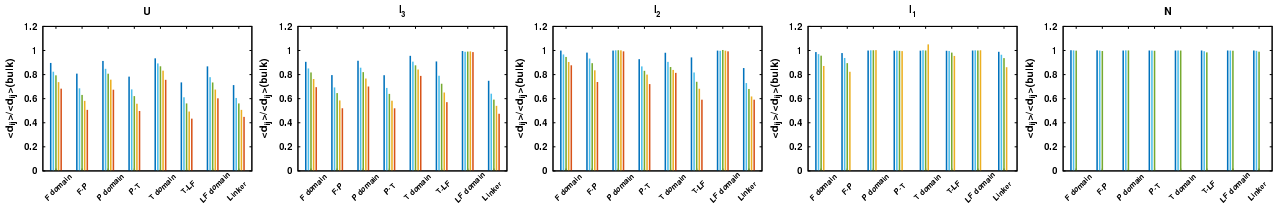

Figure S13: Spatial distance formations for the pairs in each domain and interface of DPO4 during folding under different confinements. d_{ij} is the spatial distance between residue i and j.

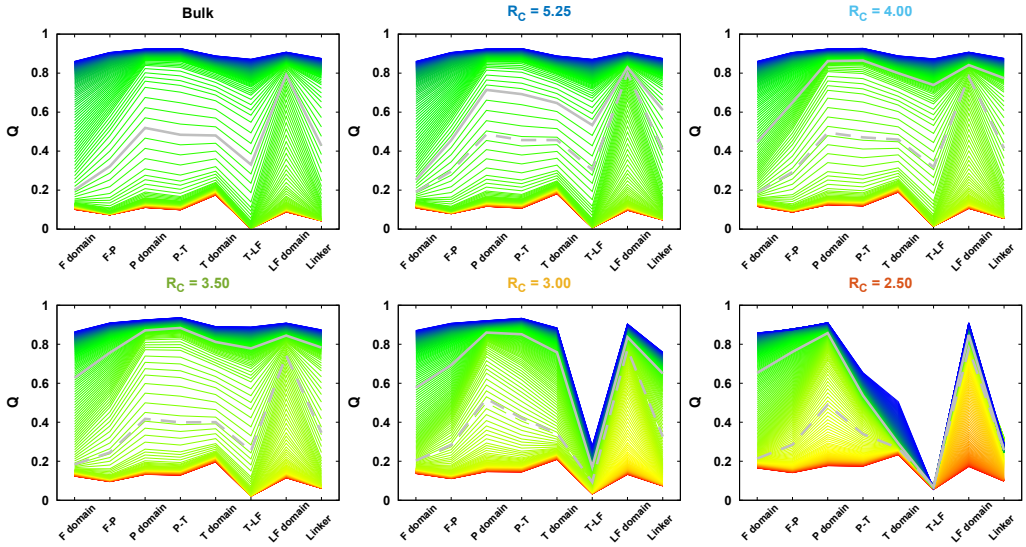

Figure S14: Native contact evolution for each domain and interface of DPO4 along with temperature under different confinements. One line represents the mean value of Q for each domain and interface in DPO4 at one temperature. Temperature increases from 1.00 to 1.35 coloring from blue to red. The grey lines correspond to the folding temperatures and the dashed lines correspond to the bulk folding temperature.

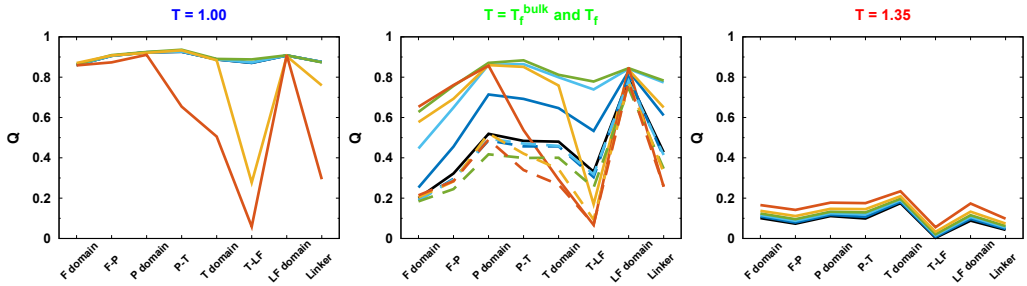

Figure S15: The mean Q for each domain and interface of DPO4 at the low, high and folding temperatures under different confinements.

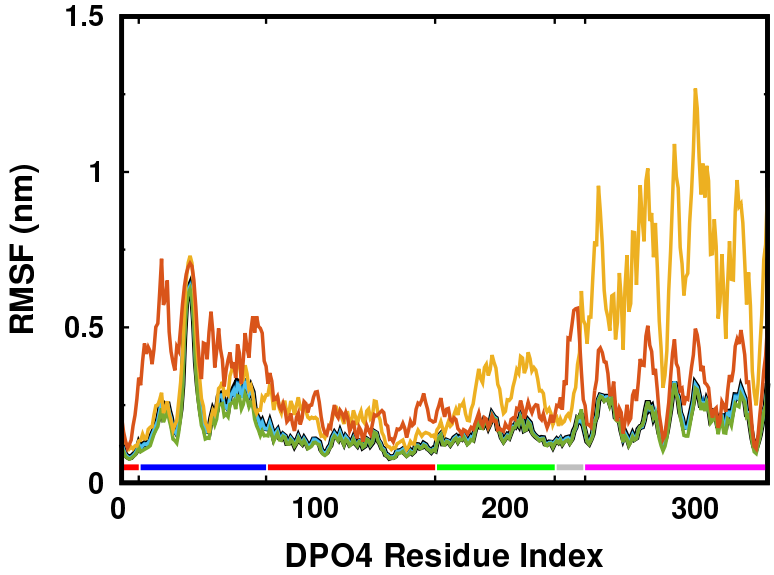

Figure S16: RMSF of DPO4 under different confinements at room temperature.

4 Effects of repulsive crowders on DPO4 folding

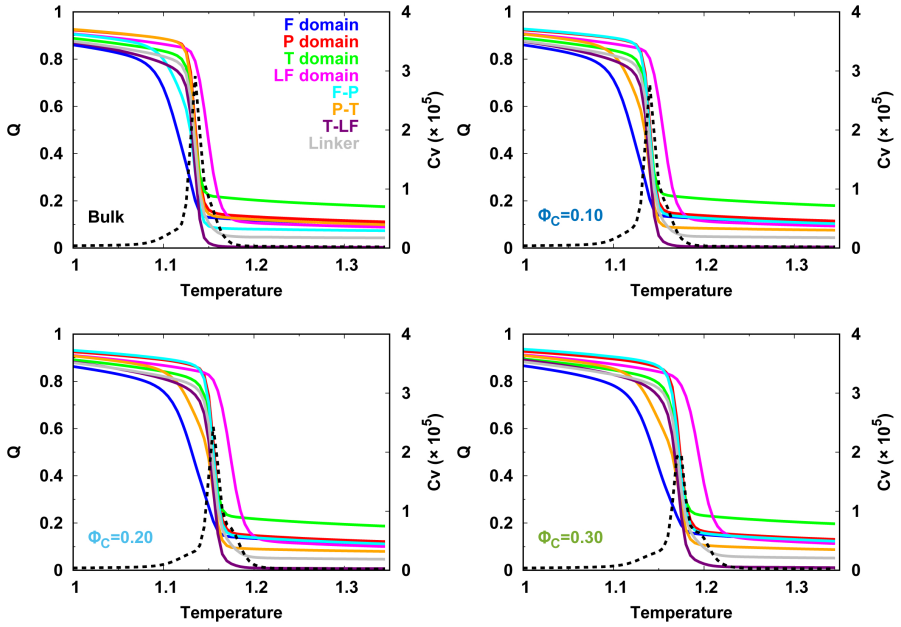

Figure S17: Melting curves of Q for each intra- and inter-domain of DPO4 under different concentrations of repulsive crowders.

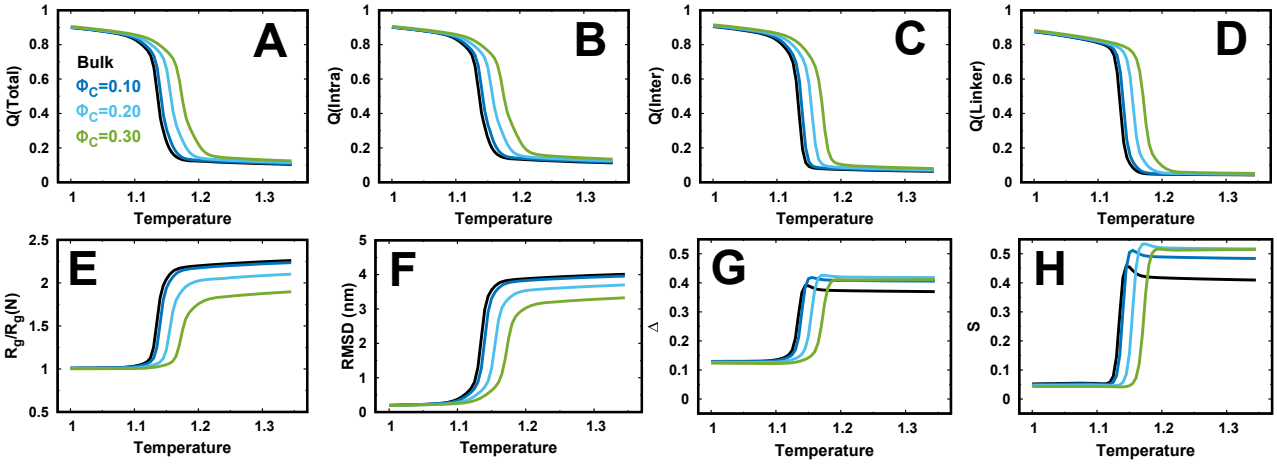

Figure S18: Melting curves of DPO4 for different order and shape parameters under different concentrations of repulsive crowders.

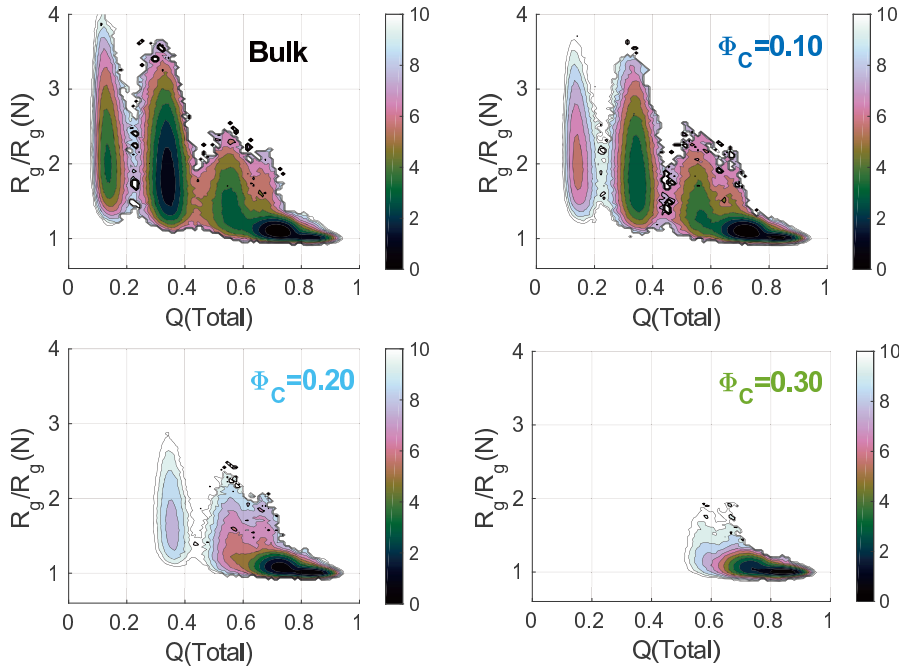

Figure S19: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g under different concentrations of repulsive crowders at the bulk folding temperature T_f^{bulk}. Free energy is in the unit of kT_f^{bulk}.

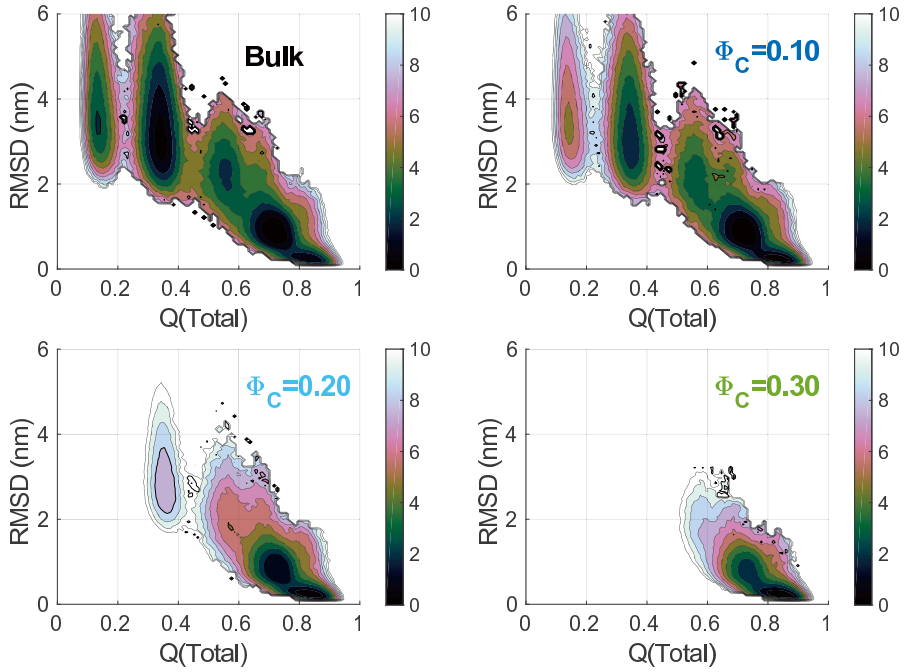

Figure S20: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ under different concentrations of repulsive crowders at the bulk folding temperature T_f^{bulk}. Free energy is in the unit of kT_f^{bulk}.

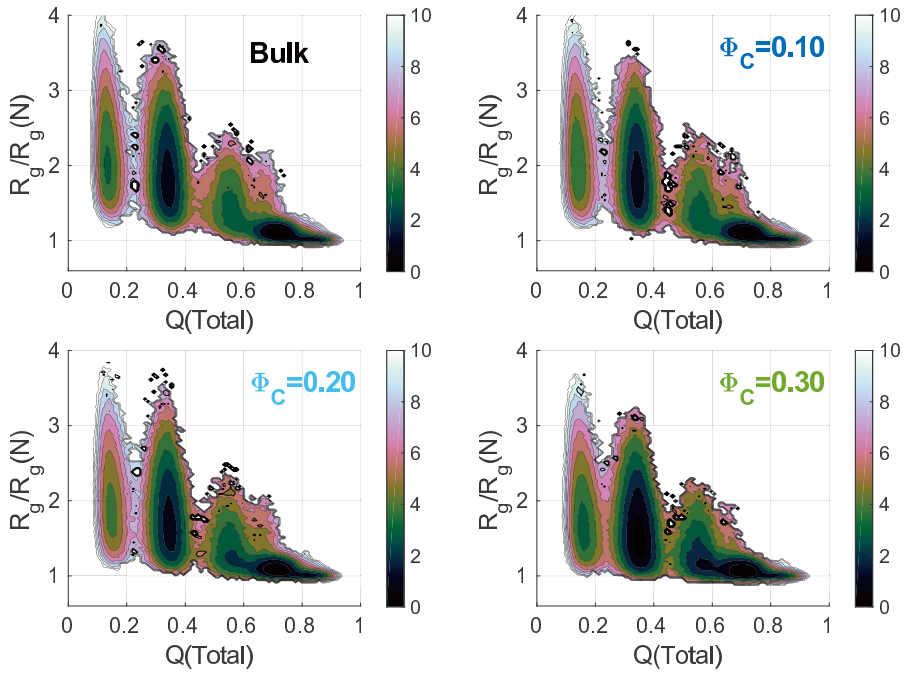

Figure S21: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g under different concentrations of repulsive crowders at the corresponding folding temperature. Free energy is in the unit of kT_f.

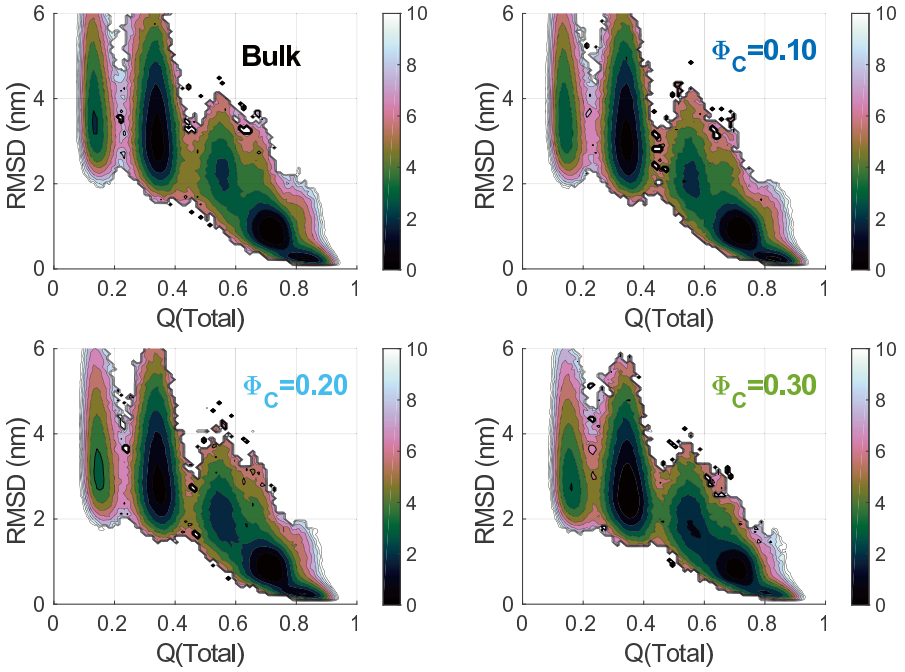

Figure S22: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ under different concentrations of repulsive crowders at the corresponding folding temperature. Free energy is in the unit of kT_f.

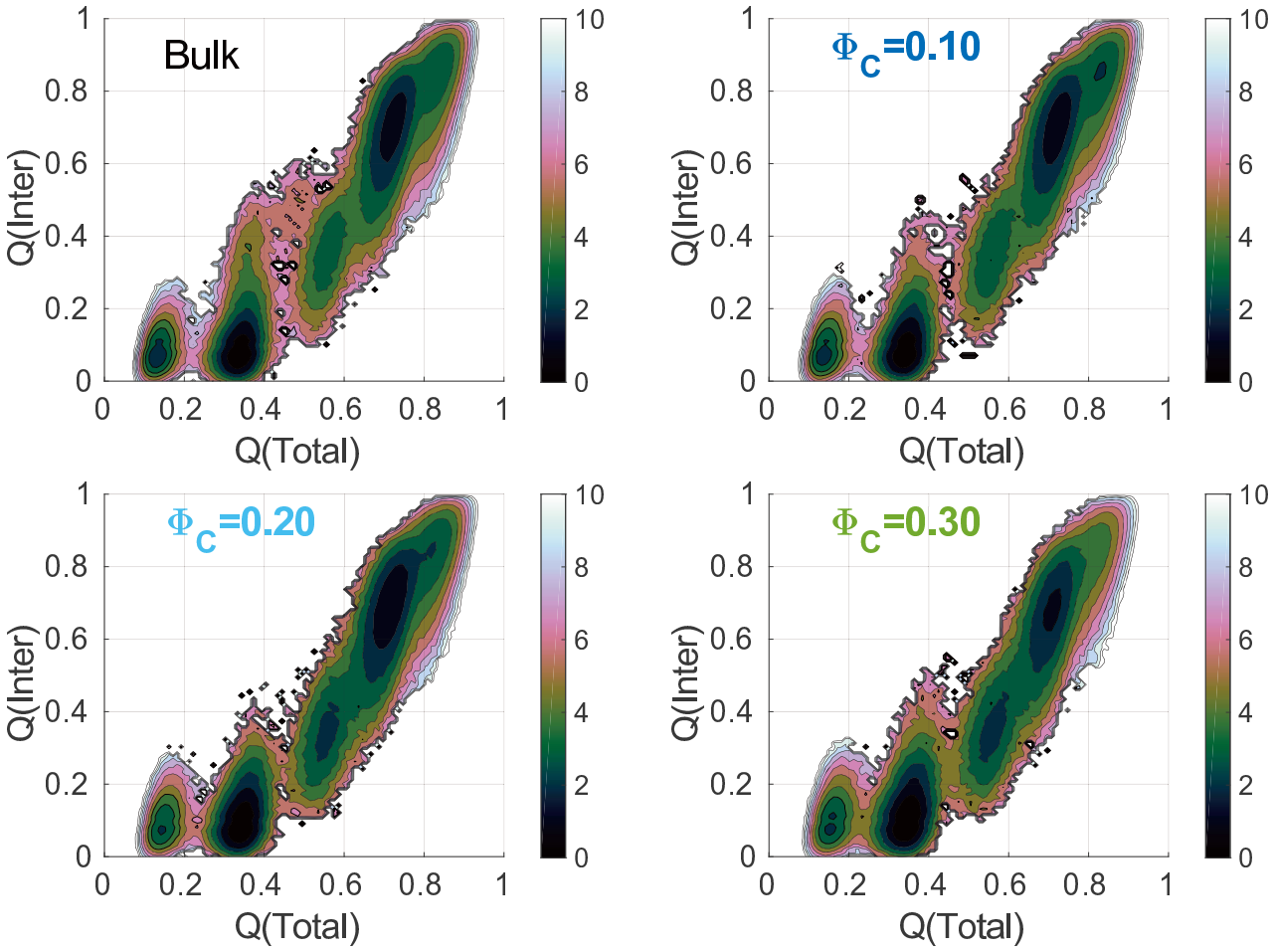

Figure S23: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $Q(Inter)$ under different concen-

trations of repulsive crowders at the corresponding folding temperatures. Free energy is in the unit of kT_f.

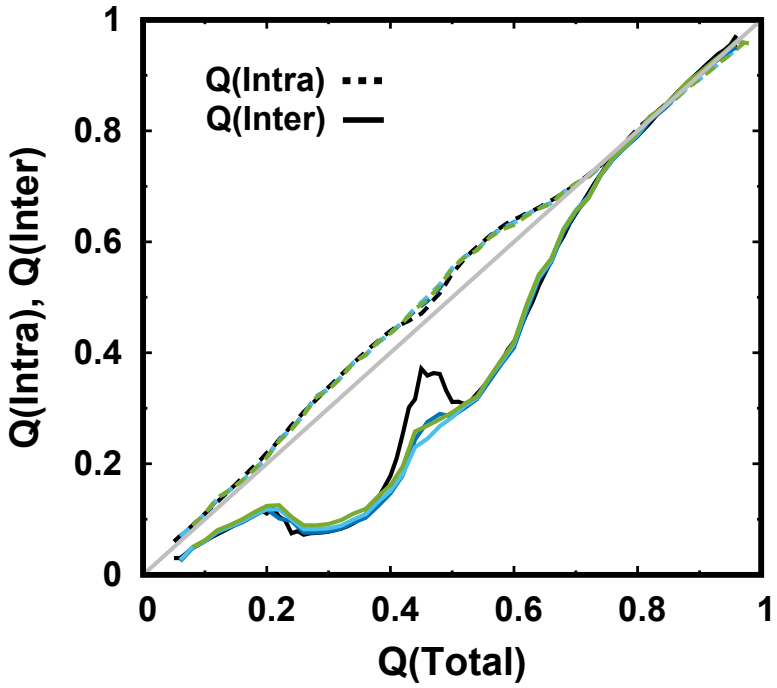

Figure S24: Native contact formation during DPO4 folding under different concentrations of repulsive crowders.

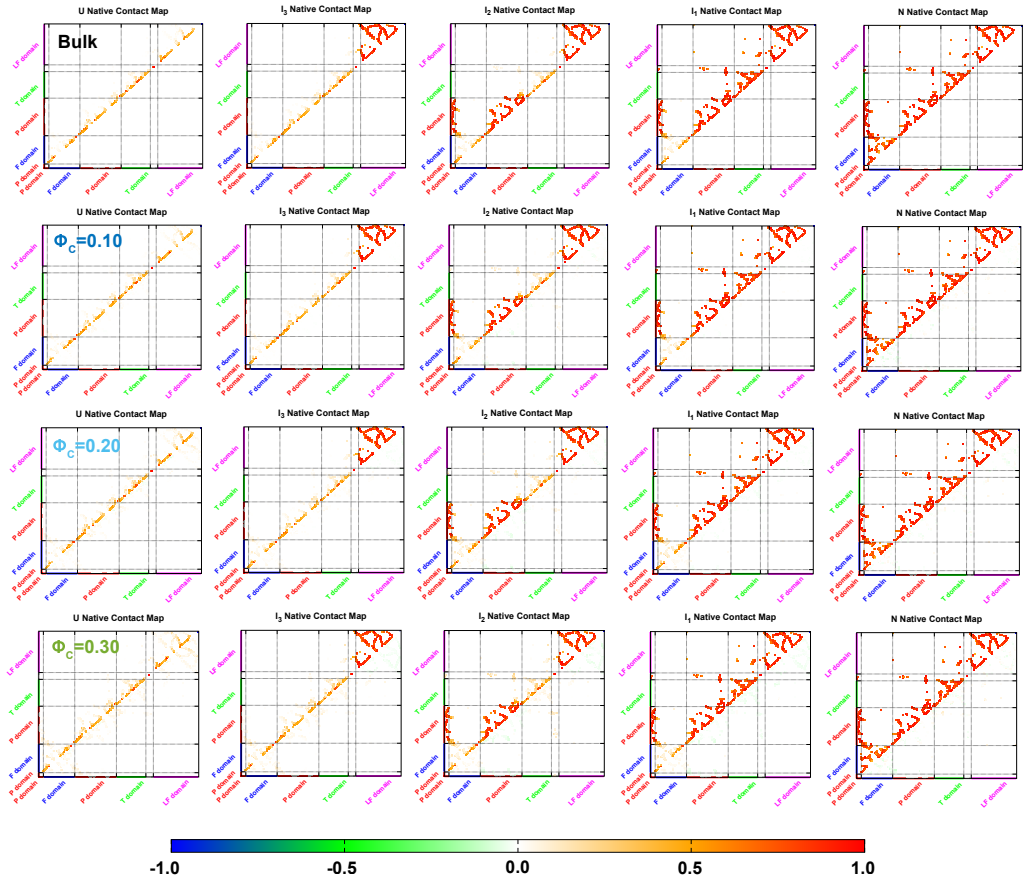

Figure S25: Native contact maps of DPO4 in each state identified from 1D free energy landscape shown in Figure 4 in the main text under different concentrations of repulsive crowders. In each sub-figure, top left shows the native contact

probability and bottom right shows the differences to that in bulk. For each state, DPO4 preserves quite similar structural characteristics in respect of native contact formation for different concentrations of repulsive crowders.

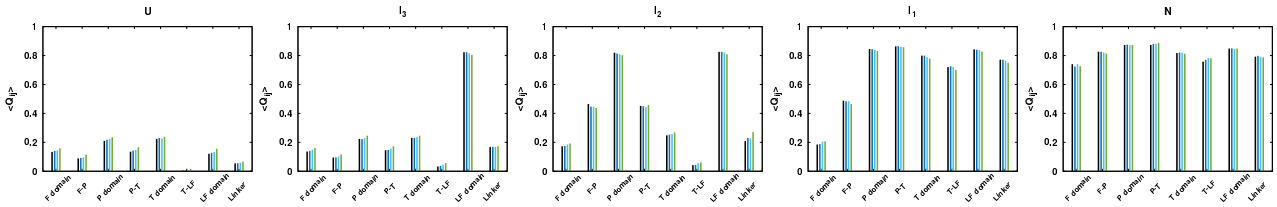

Figure S26: Native contact formations for each domain and interface of DPO4 during folding under different concentrations of repulsive crowders.

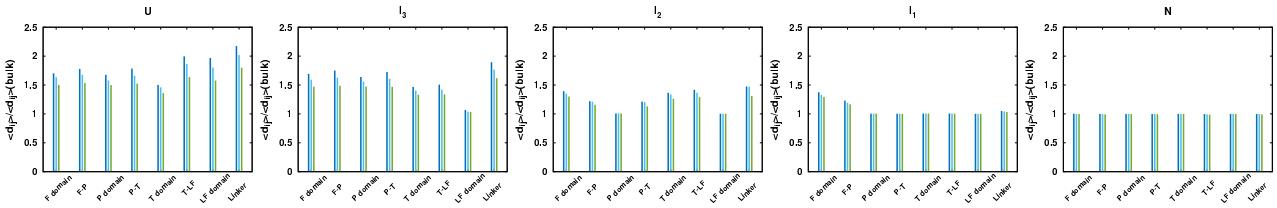

Figure S27: Spatial distance formations for the pairs in each domain and interface of DPO4 during folding under different concentrations of repulsive crowders.

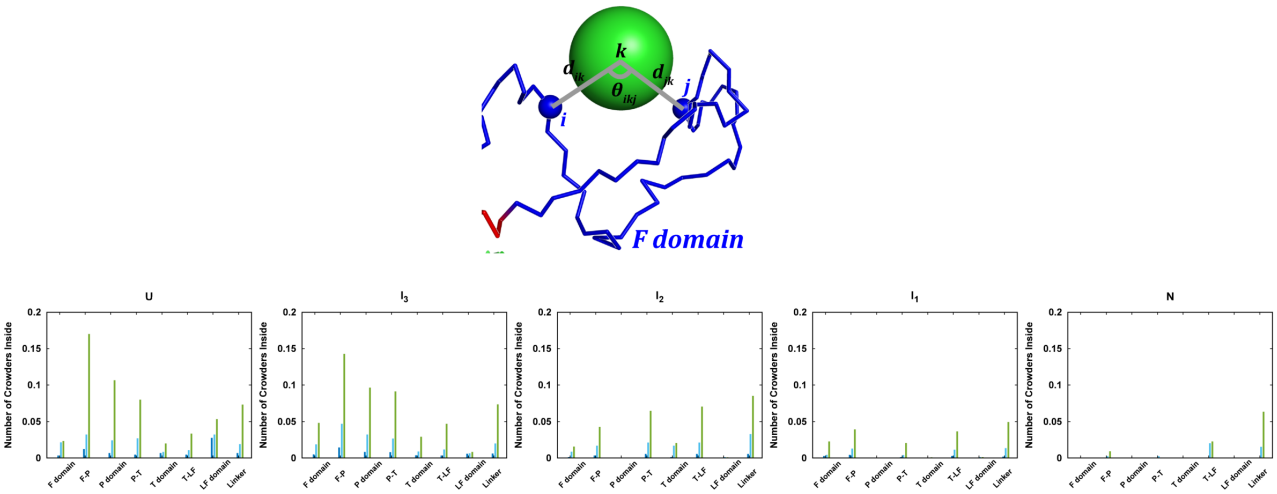

Figure S28: Number of crowders inside each domain/interface of DPO4 during folding under different concentrations of repulsive crowders. The top figure illustrate the definition of the crowder inside. The crowder is inside only to fulfill the following criteria: (1) the distances between crowders and the two residues of DPO4 are shorter than 1.2 *nm*; (2) the angle formed by the crowders and the two residue of DPO4 are larger than 150°.

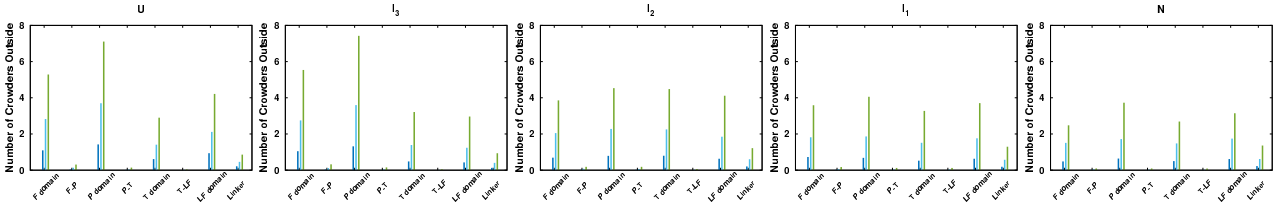

Figure S29: Number of crowders outside each domain/interface of DPO4 during folding under different concentrations of repulsive crowders. The crowder is considered to locate outside of domain/interface in DPO4 when it is closer to the residue in DPO4 than 1.2 *nm* but is not inside.

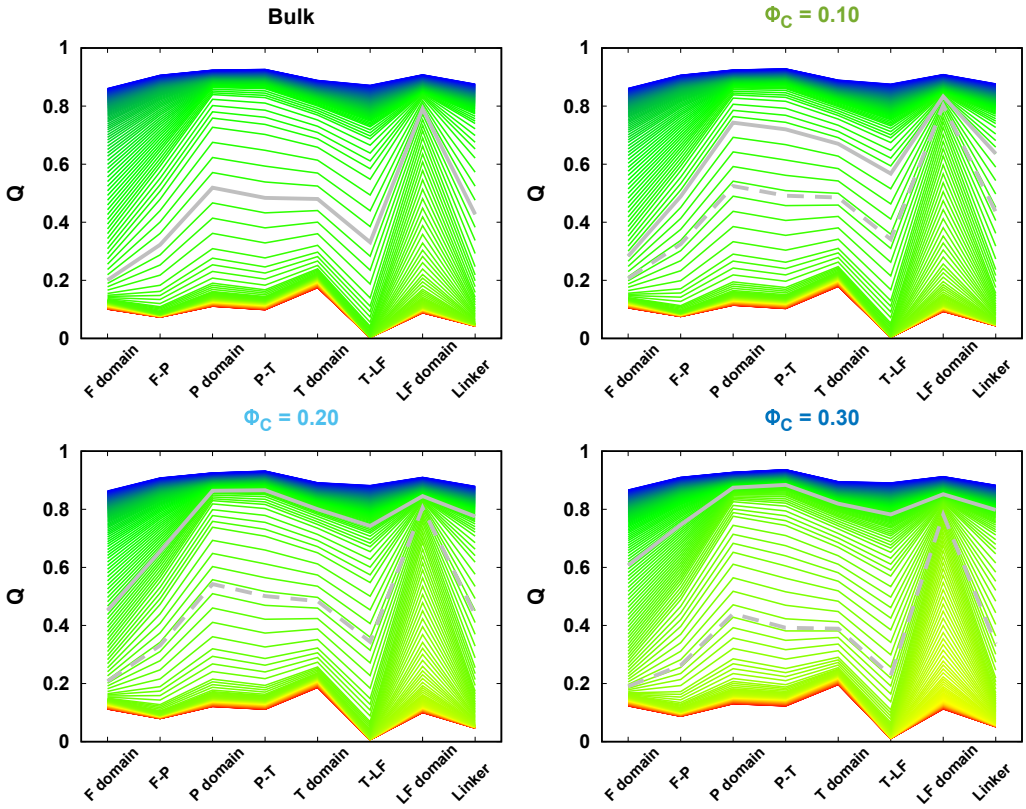

Figure S30: Native contact evolution for each domain and interface of DPO4 along with temperature under different concentrations of repulsive crowders.

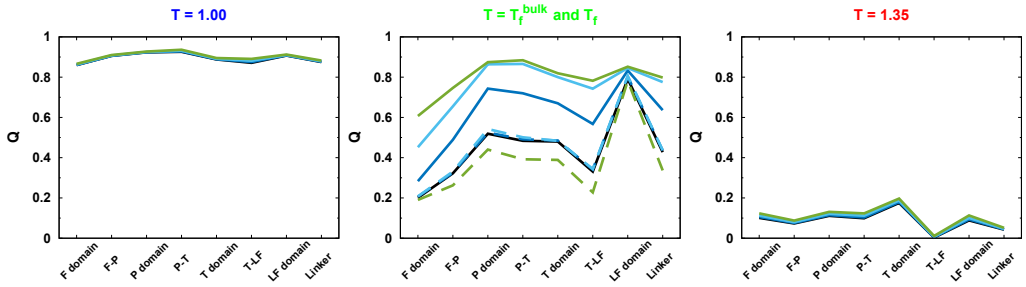

Figure S31: The mean Q for each domain and interface of DPO4 at the low, high and folding temperatures under different concentrations of repulsive crowders.

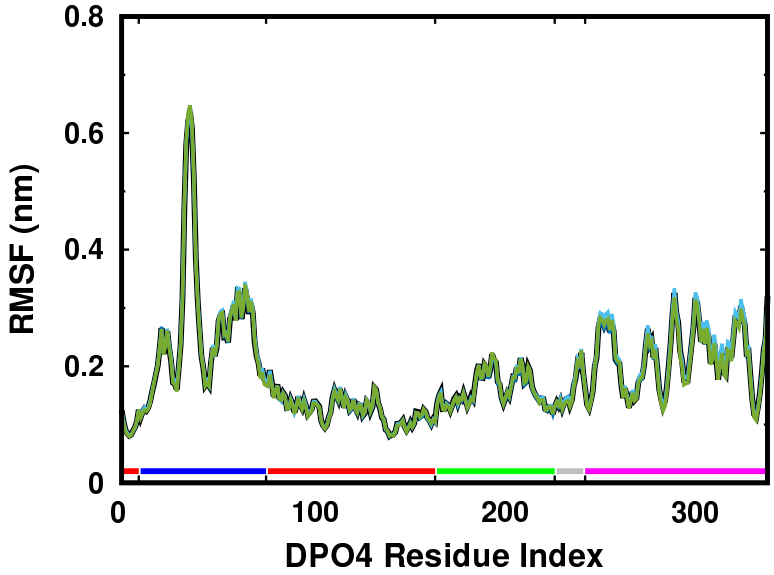

Figure S32: RMSF of DPO4 under different concentrations of repulsive crowders at room temperature.

5 Effects of attractive crowders on DPO4 folding

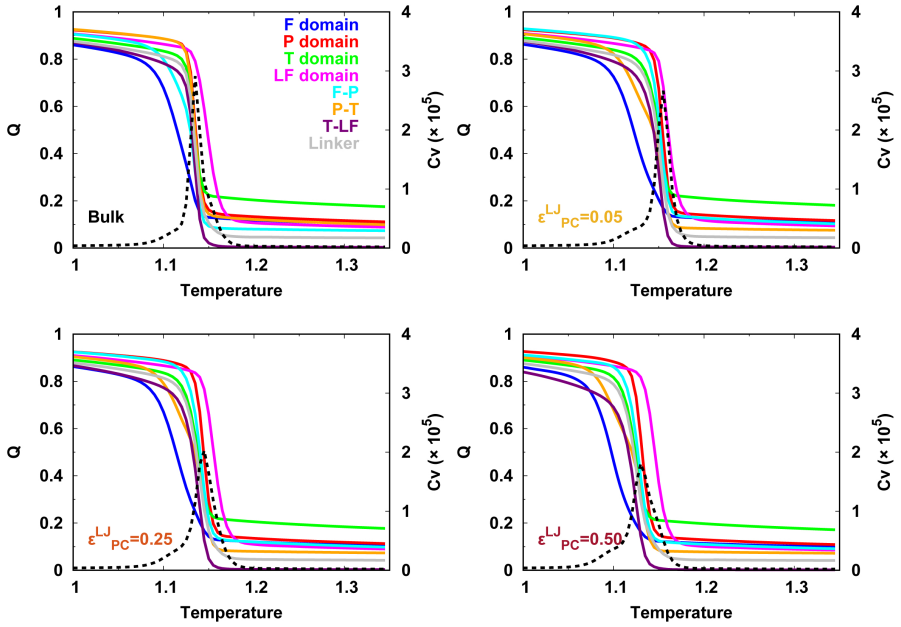

Figure S33: Melting curves of Q for each intra- and inter-domain of DPO4 with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

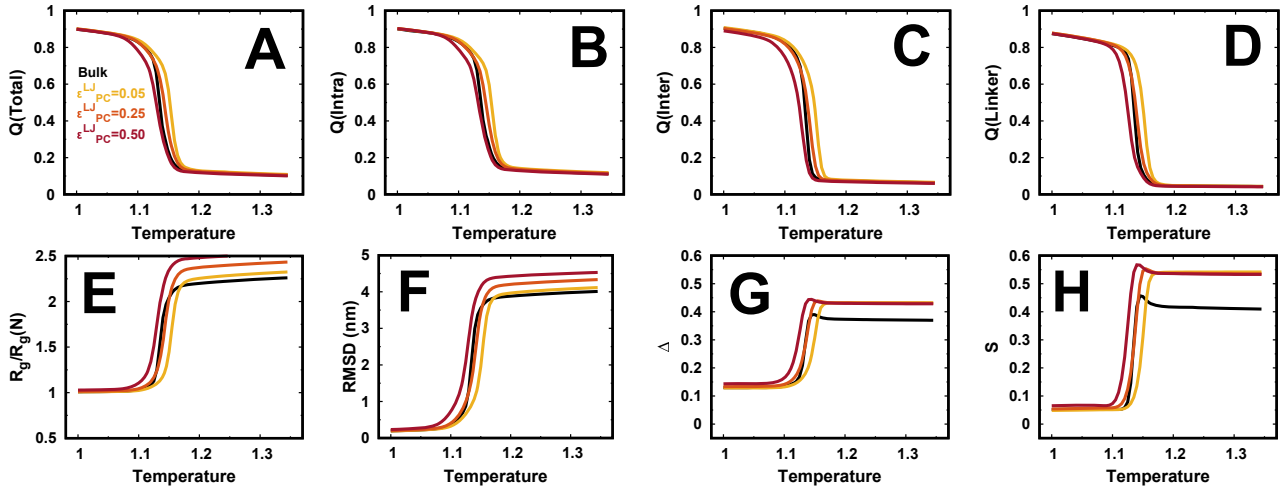

Figure S34: Melting curves of DPO4 for different order and shape parameters with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

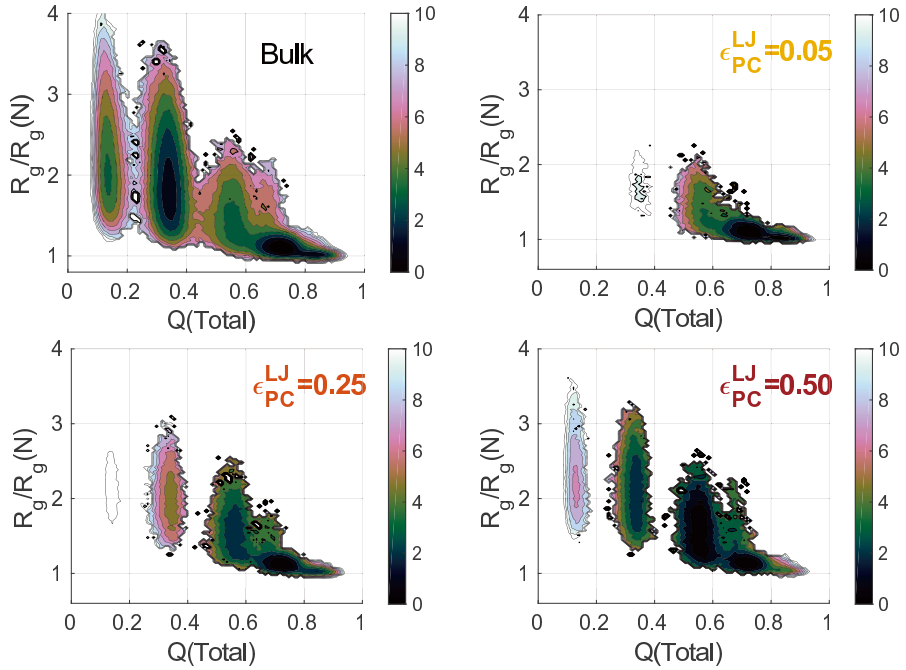

Figure S35: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g with different strengths of attractive protein-crowder interaction at the bulk folding temperature T_f^{bulk} with concentration $\Phi_C = 0.20$. Free energy is in the unit of kT_f^{bulk}.

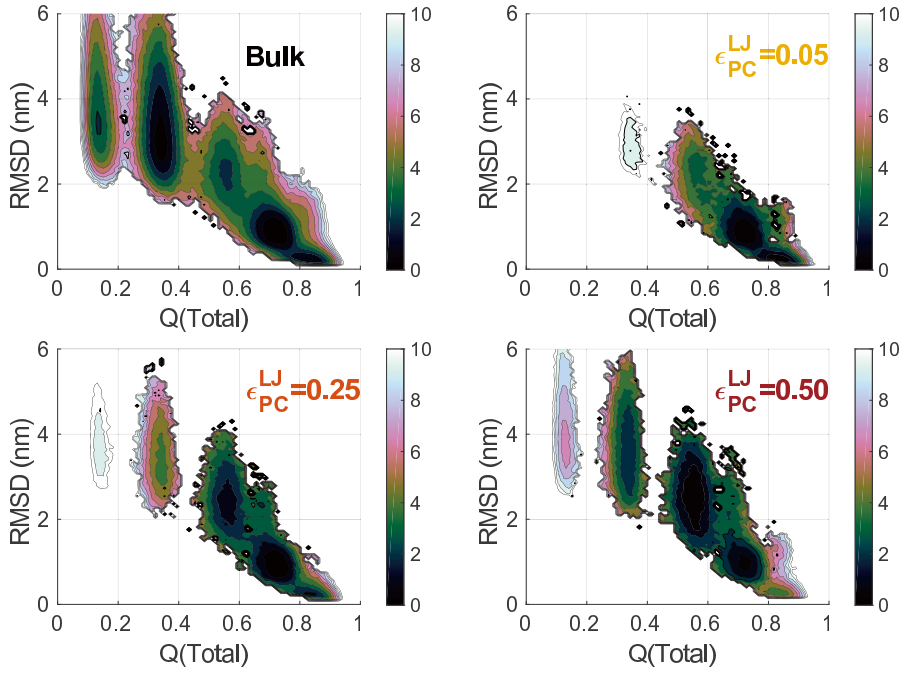

Figure S36: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ different strengths of attractive protein-crowder interaction at the bulk folding temperature T_f^{bulk} with concentration $\Phi_C = 0.20$. Free energy is in the unit of kT_f^{bulk}.

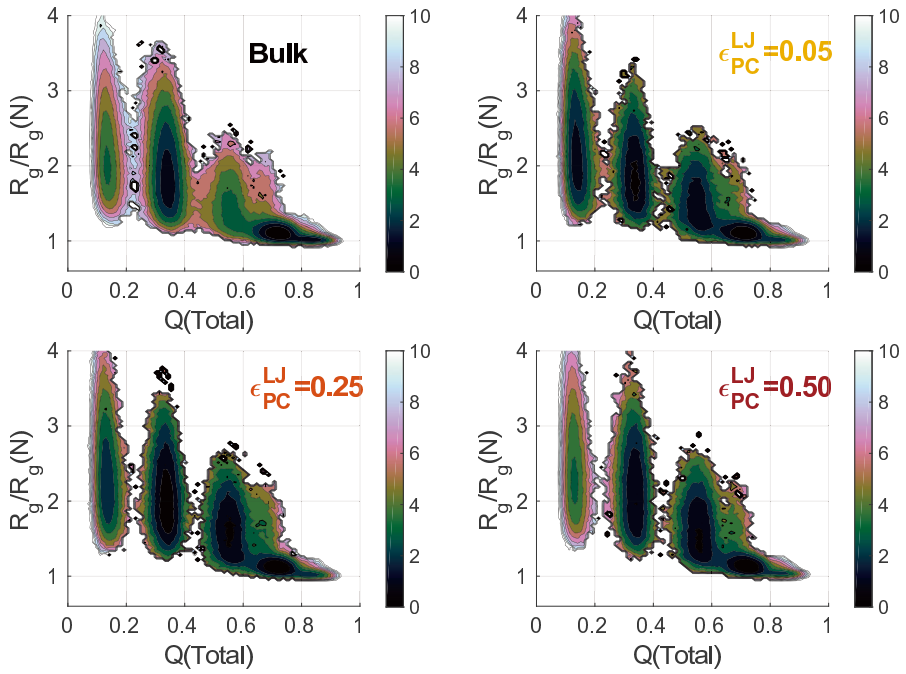

Figure S37: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and R_g different strengths of attractive protein-crowder interaction at the corresponding folding temperatures with concentration $\Phi_C = 0.20$. Free energy is in the unit of kT_f.

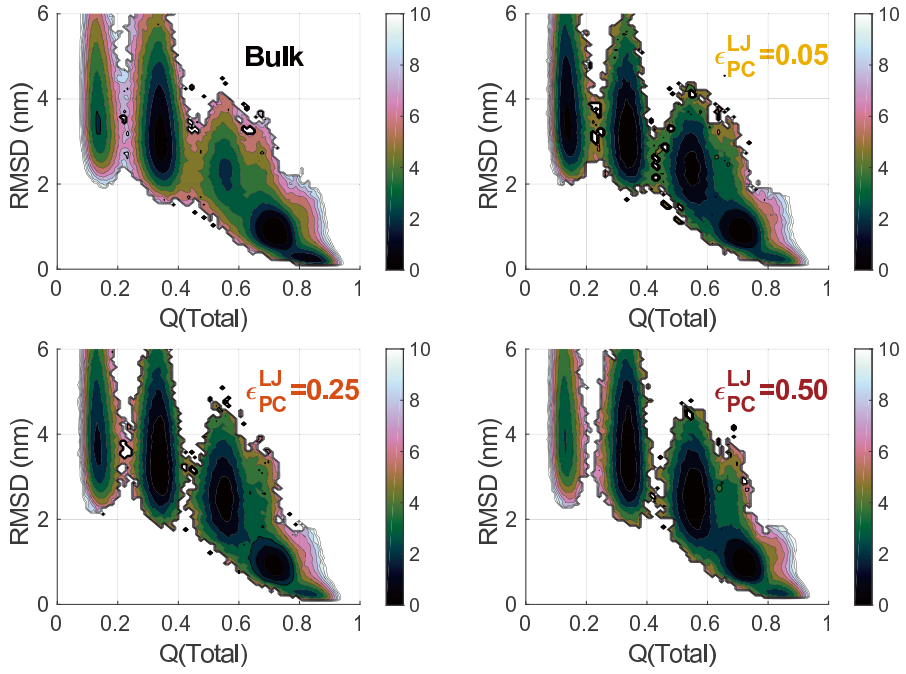

Figure S38: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $RMSD$ different strengths of attractive protein-crowder interaction at the corresponding folding temperatures with concentration $\Phi_C = 0.20$. Free energy is in the unit of kT_f.

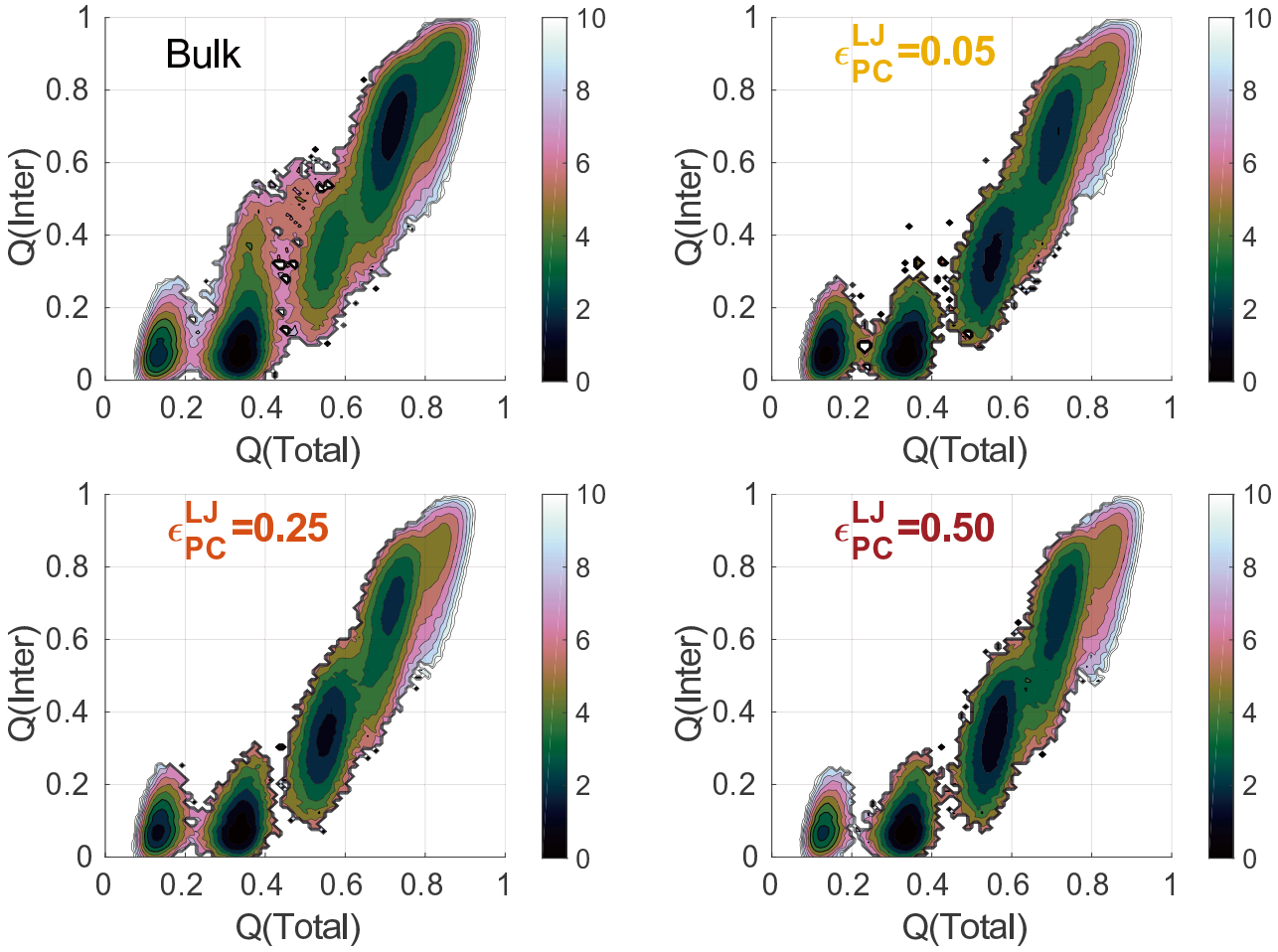

Figure S39: 2D free energy landscapes of DPO4 folding projected onto $Q(Total)$ and $Q(Inter)$ different strengths of attractive protein-crowder interaction at the corresponding folding temperatures with concentration $\Phi_C = 0.20$. Free energy is in the unit of kT_f.

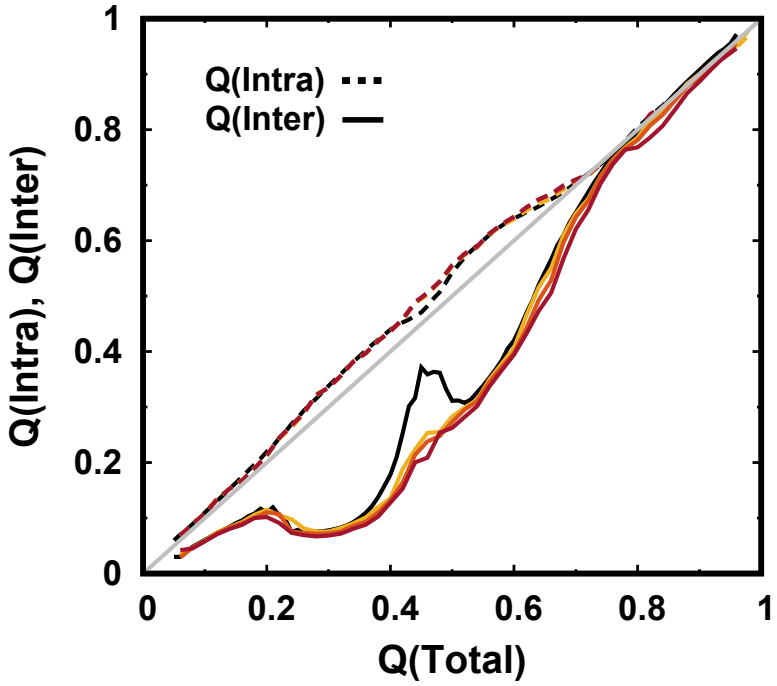

Figure S40: Native contact formation during DPO4 folding with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

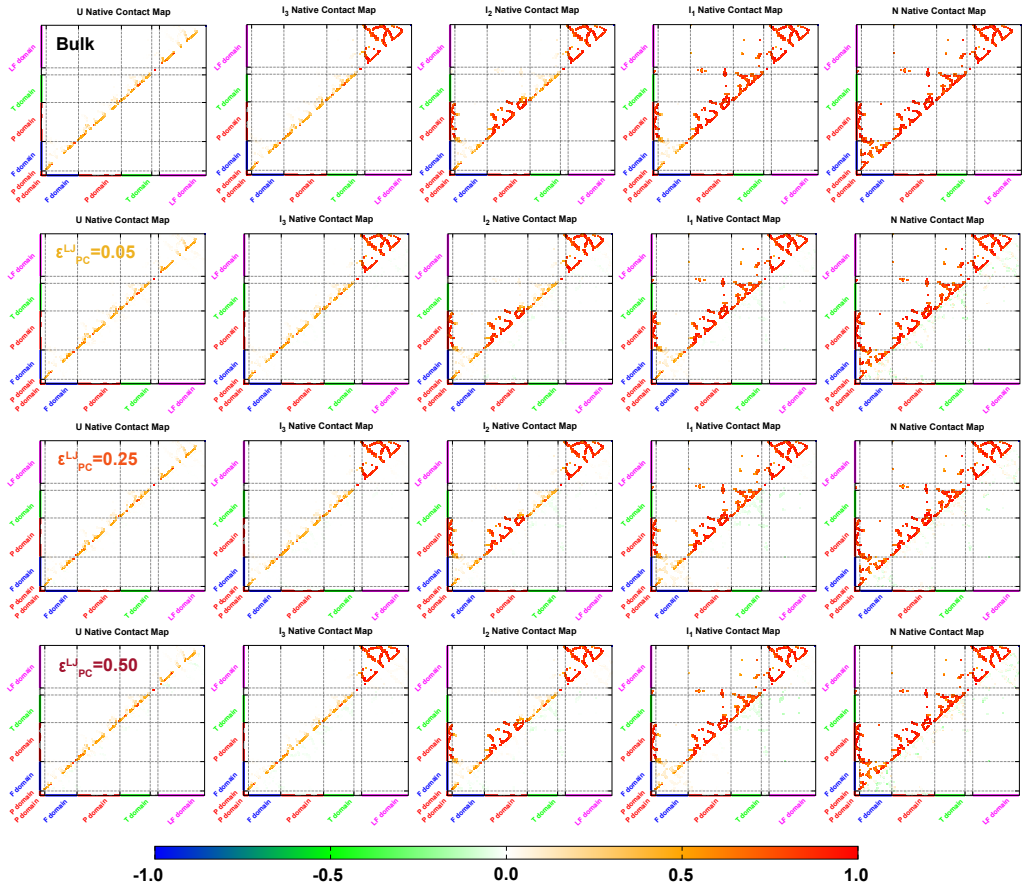

Figure S41: Native contact maps of DPO4 in each state identified from 1D free energy landscape shown in Figure 5 in the main text with different strengths of attractive crowders at concentration $\Phi_C = 0.20$. In each sub-figure, top left shows the native contact probability and bottom right shows the differences to that in bulk. For each state, DPO4 preserves quite similar structural characteristics in respect of native contact formation for different strengths of attractive protein-crowder interaction.

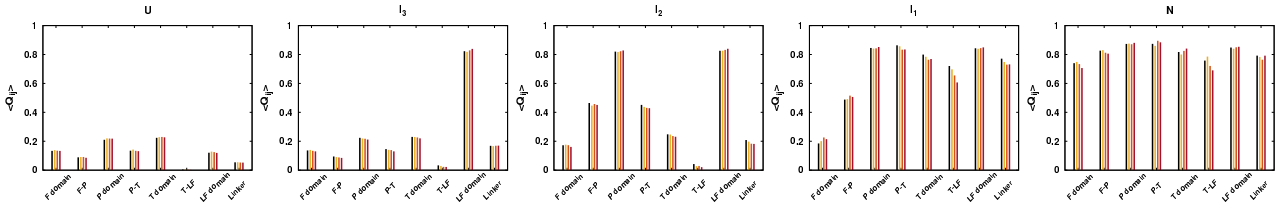

Figure S42: Native contact formations for each domain and interface of DPO4 during folding with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

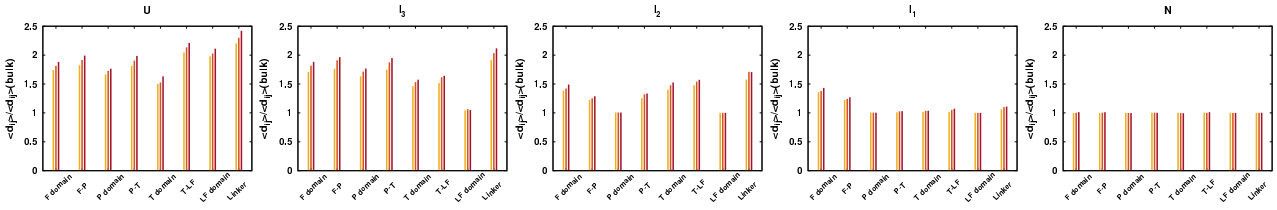

Figure S43: Spatial distance formations for the pairs in each domain and interface of DPO4 during folding with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

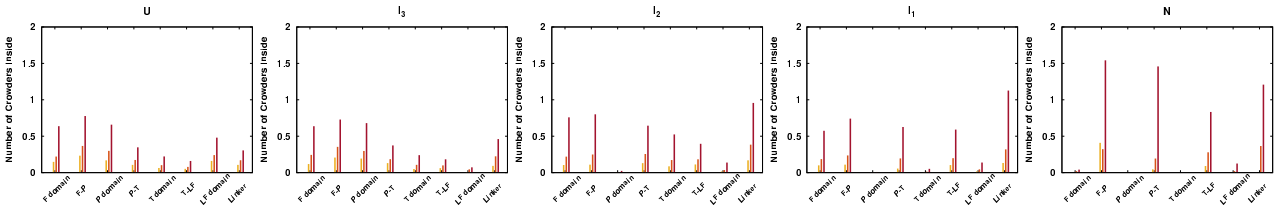

Figure S44: Number of crowders inside each domain/interface of DPO4 during folding with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

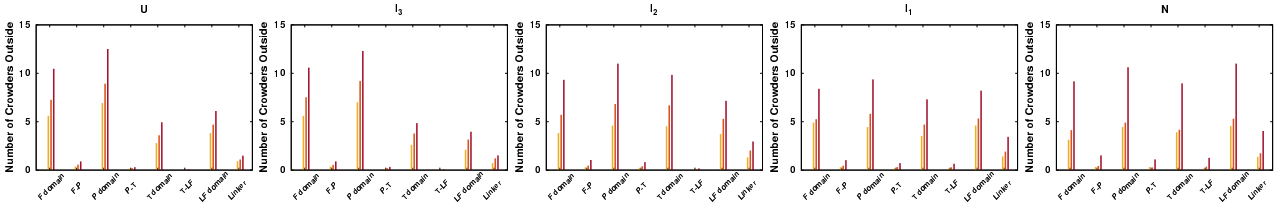

Figure S45: Number of crowders outside each domain/interface of DPO4 during folding with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

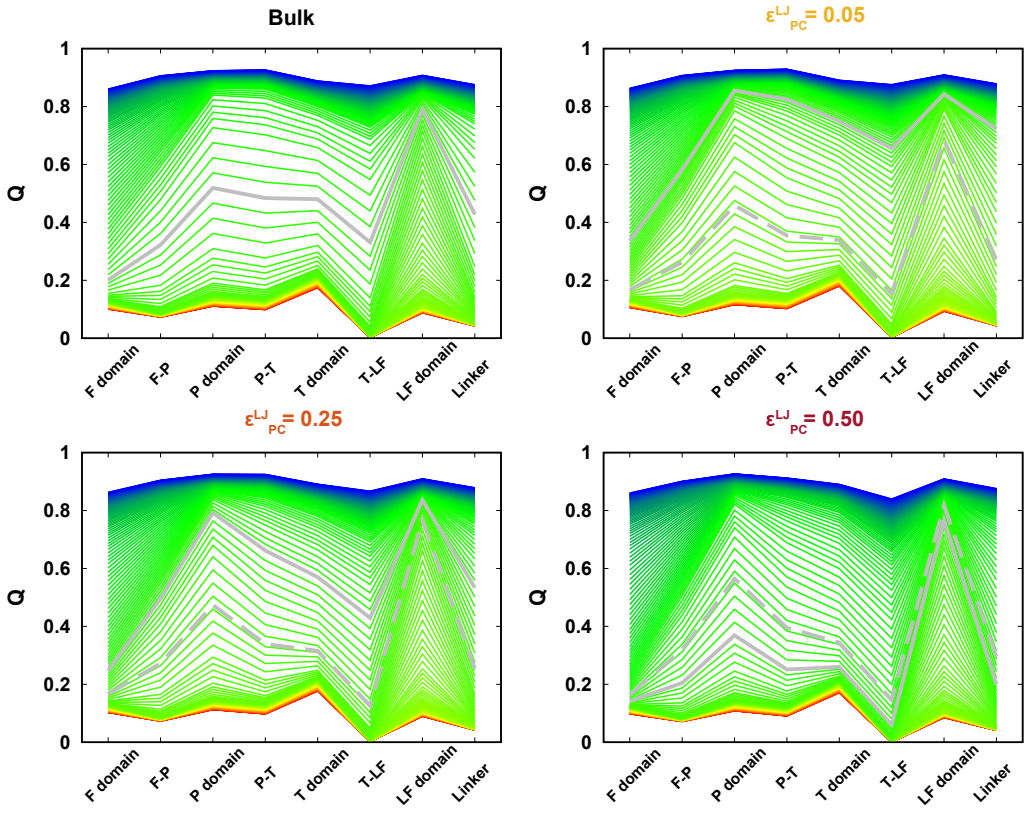

Figure S46: Native contact evolution for each domain and interface along with temperature under different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

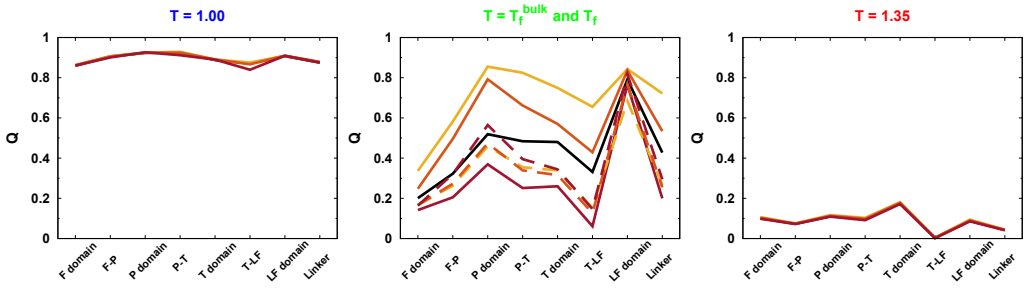

Figure S47: The mean Q for each domain and interface in DPO4 at the low, high and folding temperatures with different strengths of attractive protein-crowder interaction at concentration $\Phi_C = 0.20$.

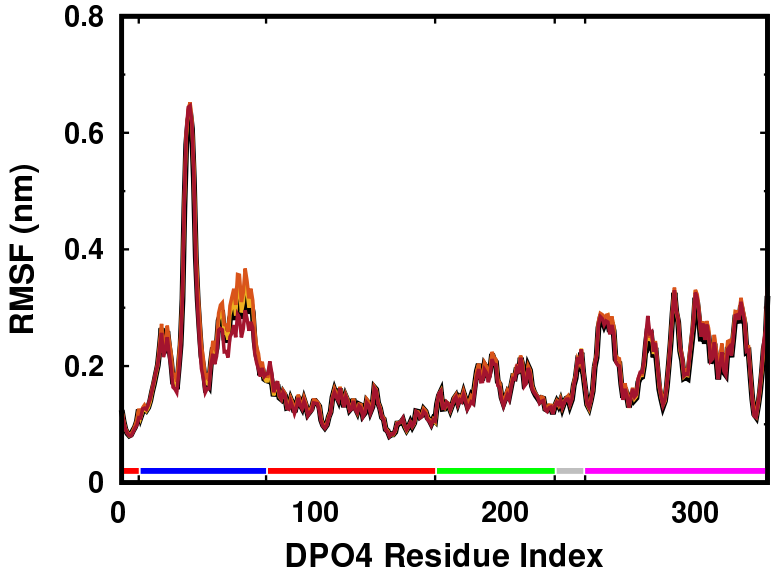

Figure S48: RMSF of DPO4 with different strengths of attractive protein-crowder interaction at room temperature with concentration $\Phi_C = 0.20$.

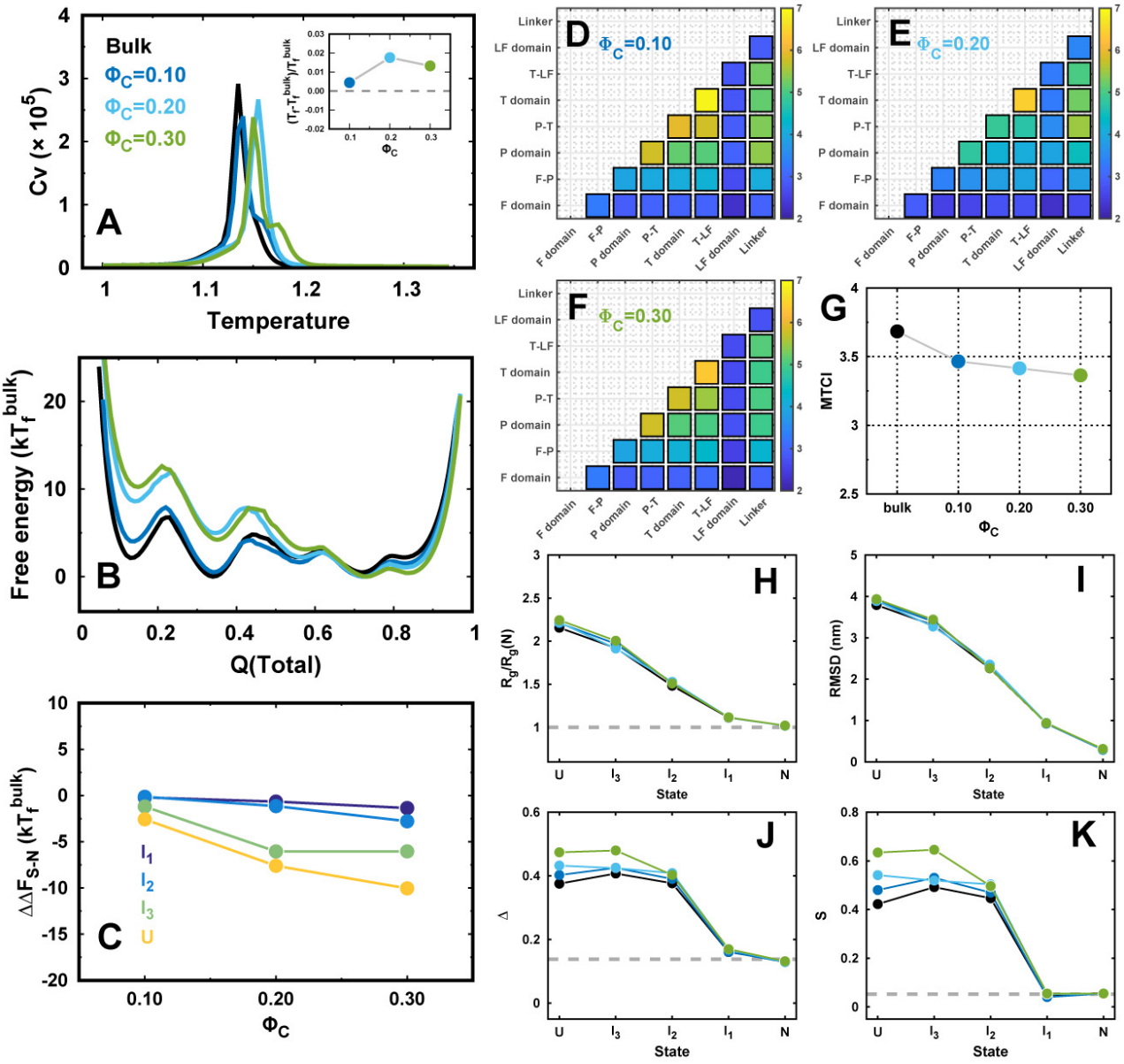

Figure S49: DPO4 folding under different concentrations of attractive crowders. The strength of the attractive protein-crowder interaction is $\varepsilon_{PC}^{LJ} = 0.05$.

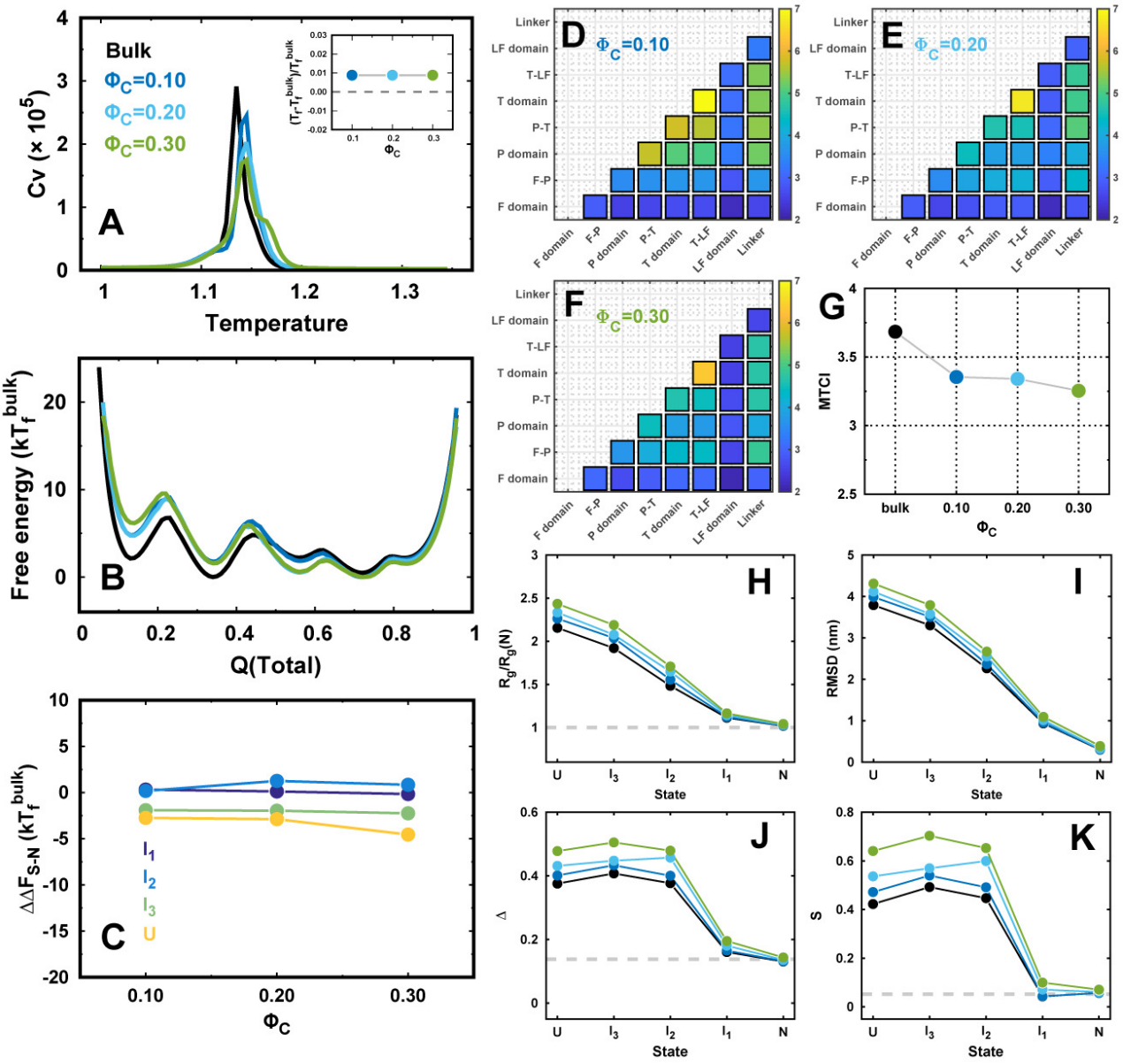

Figure S50: DPO4 folding under different concentrations of attractive crowders. The strength of the attractive protein-crowder interaction is $\varepsilon_{PC}^{LJ} = 0.25$.

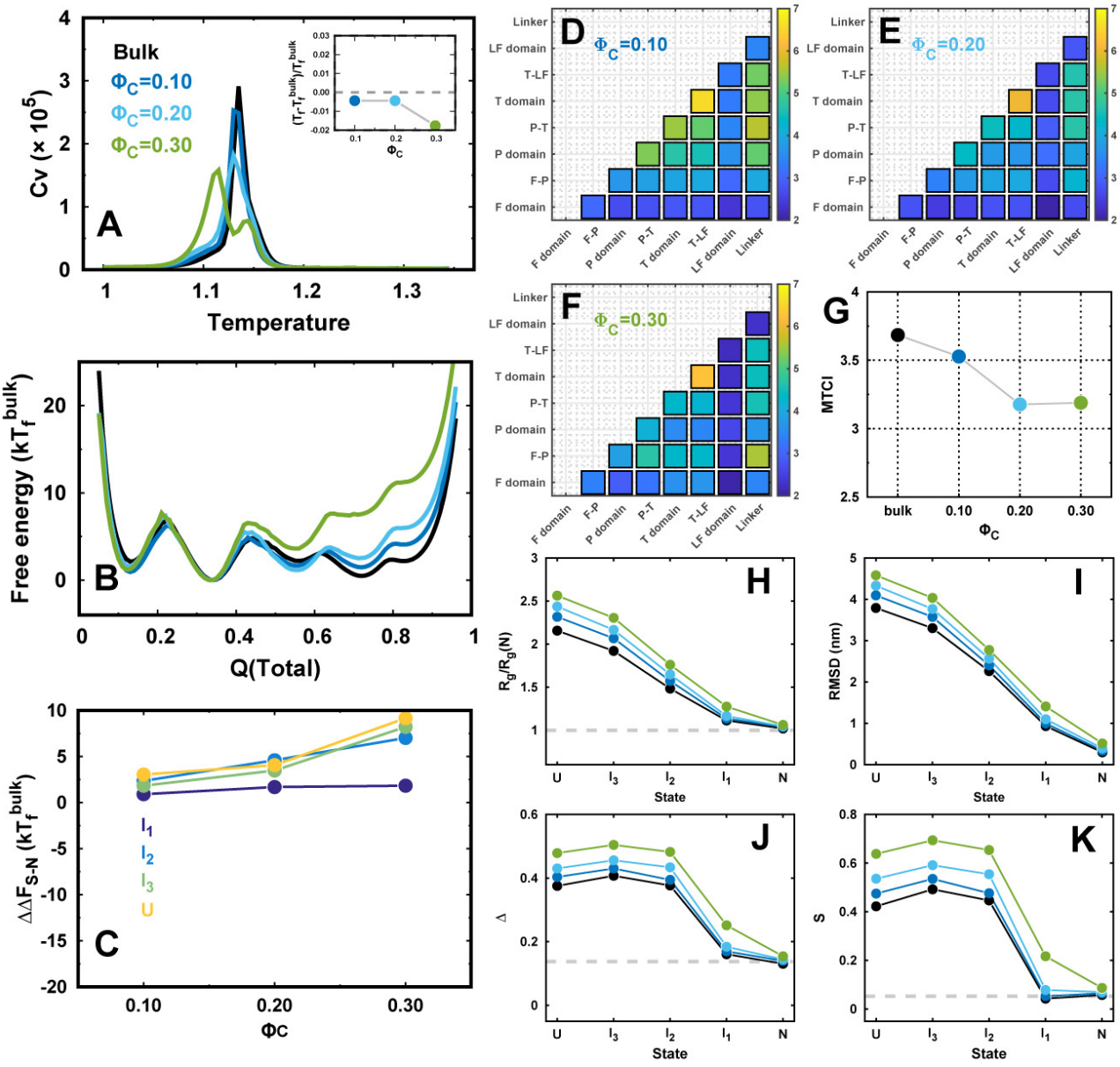

Figure S51: DPO4 folding under different concentrations of attractive crowders. The strength of the attractive protein-crowder interaction is $\varepsilon_{PC}^{LJ} = 0.50$.

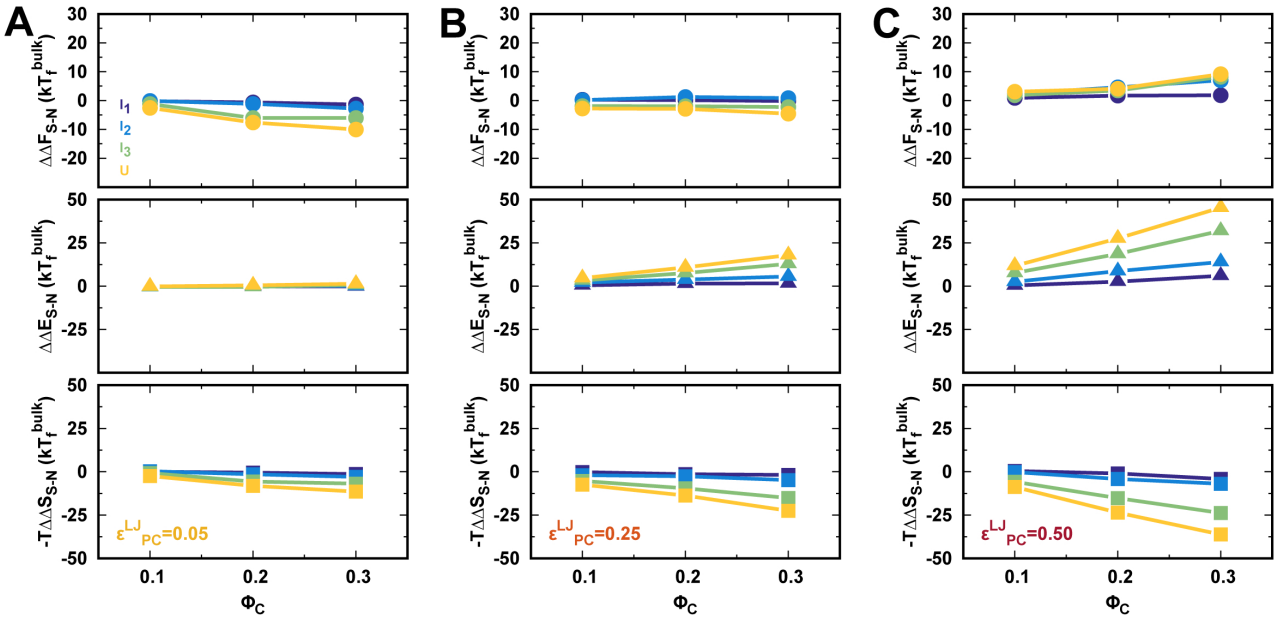

Figure S52: The relative change of the differences in free energy (top), energy (middle) and entropy (bottom) for different folding states of DPO4 to those in bulk with different attractive crowder concentrations for strengths (A) $\varepsilon_{PC}^{LJ} = 0.05$, (B) $\varepsilon_{PC}^{LJ} = 0.25$ and (C) $\varepsilon_{PC}^{LJ} = 0.50$.

6 Mapping of the effects of confinement and repulsive crowder on DPO4 folding

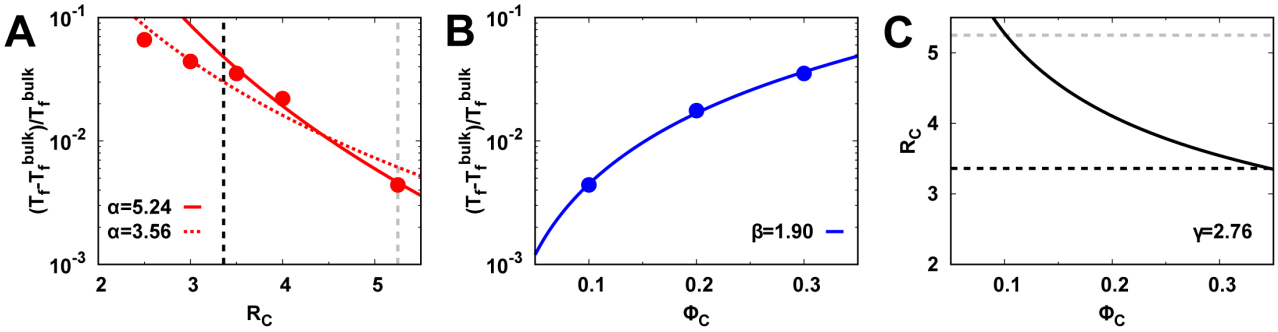

Figure S53: (A) The change of folding temperature led by confinements. The change of folding temperature is fitted to the power-law relation of R_C: $\sim R_C^{-\alpha}$. Two dashed vertical lines dictate the half length of DPO4 native structure at longest axis (black, 3.36 nm) and radius of gyration R_g of DPO4 at unfolded states obtained from high temperature simulations in bulk (grey, 5.25 nm). The solid red line is fitted to power-law function with the data between 3.36 and 5.25 nm, while the dotted red line is fitted to power-law function with the data of whole range. We found the α for fitting our entire data is 3.56, in reasonable agreement with previous studies[5–8], where a range of α from 2 to 4 was found. (B) The change of folding temperature led by repulsive crowder. The change of folding temperature is fitted to the power-law relation of Φ_C: $\sim {\Phi_C}^{-\beta}$. We found $\beta = 1.90$, close to that found in[6]. (C) The relation between R_C and Φ_C established by the fitting parameters obtained in (A) and (B) with $R_C \sim \Phi_C^{-1/\gamma}$. We found $\gamma = 2.76$, in reasonable agreement of 3.00, predicted by theory[9] and verified by simulation[6].

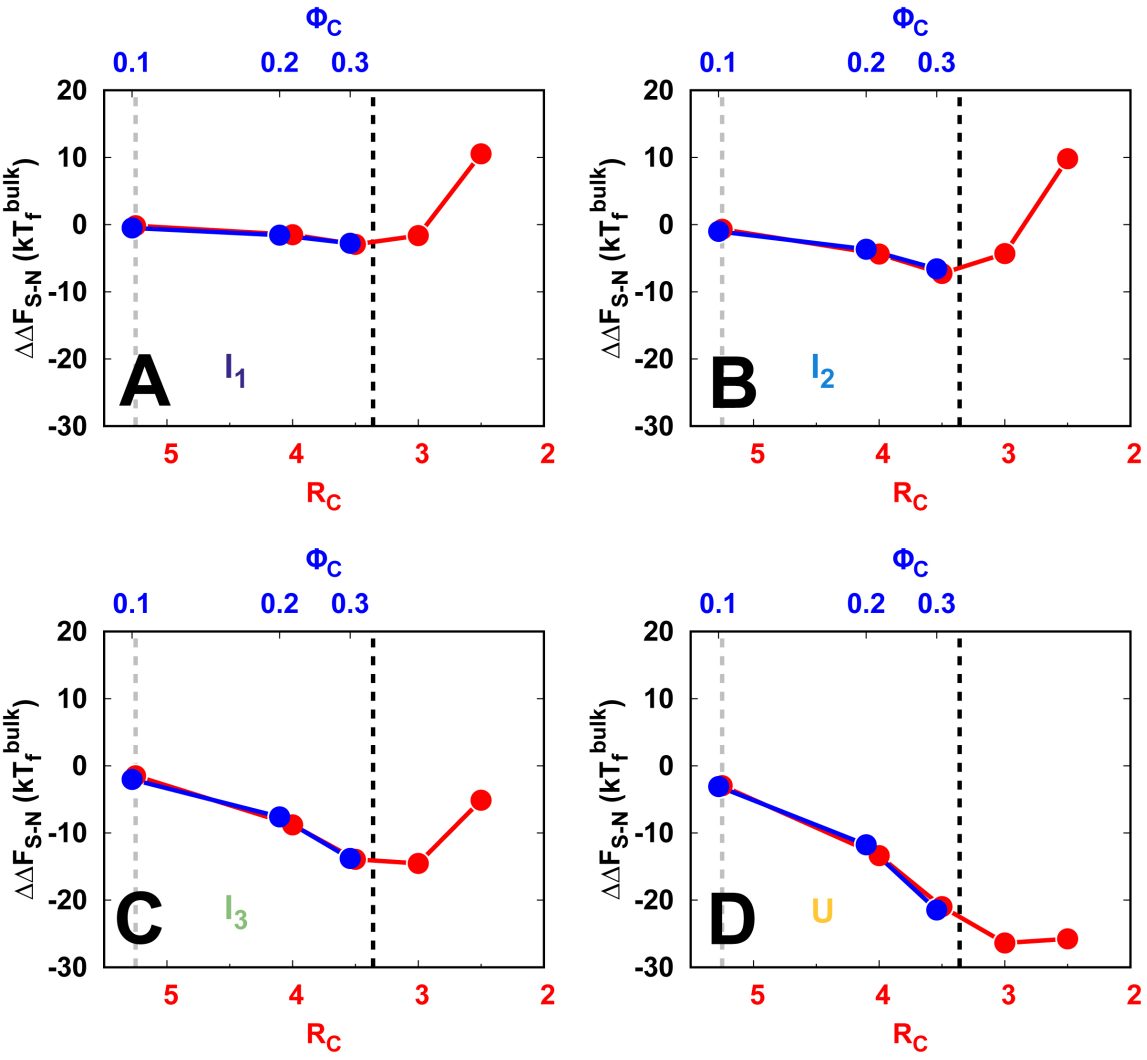

Figure S54: The mapping of free energy change led by confinement and repulsive crowder for different DPO4 folding states.

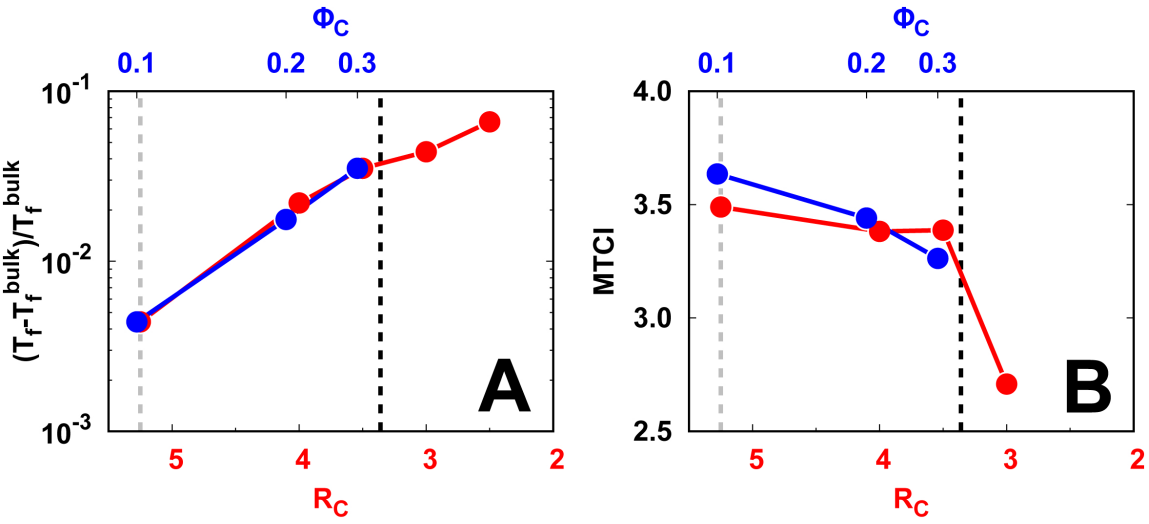

Figure S55: (A) The mapping of folding temperature change led by confinement and repulsive crowder. (B) The change of folding cooperativity in terms of *MTCI* led by confinement and repulsive crowder.

References

[1] Lorenzo Sborgi, Abhinav Verma, Mourad Sadqi, Eva de Alba, and Victor Muñoz. *Protein folding at atomic resolution: analysis of autonomously folding supersecondary structure motifs by nuclear magnetic resonance*, pages 205–218. Humana Press, Totowa, NJ, 2012.

[2] Shanen M Sherrer, Brian A Maxwell, Lindsey R Pack, Kevin A Fiala, Jason D Fowler, Jun Zhang, and Zucai Suo. Identification of an unfolding intermediate for a dna lesion bypass polymerase. *Chem. Res. Toxicol.*, 25(7):1531–1540, 2012.

[3] David De Sancho and Robert B Best. Modulation of an idp binding mechanism and rates by helix propensity and non-native interactions: association of hif1α with cbp. *Mol. BioSyst.*, 8(1):256–267, 2012.

[4] Jimson H Wong, Kevin A Fiala, Zucai Suo, and Hong Ling. Snapshots of a y-family dna polymerase in replication: substrate-induced conformational transitions and implications for fidelity of dpo4. *J. Mol. Biol.*, 379(2):317–330, 2008.

[5] Fumiko Takagi, Nobuyasu Koga, and Shoji Takada. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl. Acad. Sci. U. S. A.*, 100(20):11367–11372, 2003.

[6] Margaret S Cheung, Dmitri Klimov, and D Thirumalai. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 102(13):4753–4758, 2005.

[7] Nitin Rathore, Thomas A Knotts IV, and Juan J de Pablo. Confinement effects on the thermodynamics of protein folding: Monte carlo simulations. *Biophys. J.*, 90(5):1767–1773, 2006.

[8] Jeetain Mittal and Robert B Best. Thermodynamics and kinetics of protein folding under confinement. *Proc. Natl. Acad. Sci. U. S. A.*, 105(51):20233–20238, 2008.

[9] Michael R Shaw and D Thirumalai. Free polymer in a colloidal solution. *Phys. Rev. A*, 44(8):R4797, 1991.