Electronic Supplementary Information (ESI)

Qing-Miao Nie[1], Li-Zhen Sun[1], Hai-Bin Li[1], Xiakun Chu[2*], and Jin Wang[2*]

[1] Department of Applied Physics, Zhejiang University of Technology, Hangzhou 310023, P.R. China
[2] Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

[*] Corresponding Authors:
xkchu2008@gmail.com (X.C.)
jin.wang.1@stonybrook.edu (J.W.)

Computational details

Structure-based models (SBM) and their variants have been widely used to investigate the mechanisms of protein folding [1] and binding [2, 3], and the validity of SBM has been verified by extensive experiments [1]. In this work, we applied a residue-level C_α coarse-grained SBM to study Dpo4 folding. The interactions among C_α beads are derived from the potential E_{total}:

$$E_{total}=E_{SBM}+E_{ele},$$

where E_{SBM} and E_{ele} are the SBM potential and the electrostatic interaction potential, respectively. The reference structure for building E_{SBM} is the crystal structure of the apo Dpo4 (PDB:2RDI) [4]. E_{SBM} is given by the following expression [1]:

$$\begin{aligned} E_{SBM}= & \sum_{\text{bonds}} K_r(r-r_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 \\ & + \sum_{\text{dihedral}} \left[K_\phi^1\left[1-\cos\left(\phi-\phi_0\right)\right]+K_\phi^3\left[1-\cos\left(3\left(\phi-\phi_0\right)\right)\right]\right] \\ & + \sum_{i<j-3}^{\text{native}} \varepsilon_{ij}\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] \\ & + \sum_{i<j-3}^{\text{non-native}} \varepsilon_{PP}\left(\frac{\sigma_{PP}}{r_{ij}}\right)^{12} \end{aligned}$$

E_{SBM} is made up of bond stretching, angle bending, dihedral rotation, and non-bonded interactions. Non-bonded interactions consist of native interactions and non-native interactions. The parameters K_r, K_θ, K_ϕ, ε_{ij}, and ε_{PP} set the strength of each type of interaction. r, θ, and ϕ are the bond lengths, bending angles, and dihedral angles, with a subscript zero representing the values adopted in the native structure. σ_{ij} is the distance between beads i and j in the native structure. σ_{PP} defines the diameter of the C_α bead, and the associated interaction provides the excluded volume repulsions (non-native interaction) among the beads in Dpo4. The native contact map is generated by the Contacts of Structural Unit (CSU) software [5]. Length is in the unit of nm, and the others are in reduced units. We chose the parameters suggested by Clementi et al. [1], so K_r = 10000.0, K_θ =20.0, K^1_ϕ= 1.0, K^3_ϕ= 0.5, ε_{ij} = ε_{PP} = 1.0, and σ_{PP} = 0.4 nm.

The electrostatic interaction is described by the Debye-Hückel potential [6]:

$$E_{ele}=\varepsilon_{ele}\varepsilon_{Coulomb}\sum_{i,j}\frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon_r r_{ij}},$$

where q_i or q_j is the charge of the bead i or j, and r_{ij} is the distance between the two charged beads (with either the same or opposite charges). In our model, each LYS and ARG amino acid was assigned with one positive charge, each GLU and ASP amino acid was assigned with one negative charge based on their electrostatic charge at neutral PH. The protonation state of HIS is highly dependent on the local environment. For simplicity, we assigned zero charge to each HIS amino acid and other types of amino acids in our model. We note that this charge assignment was widely used by others [7-9]. The Debye screening length κ^{-1} defines the range of electrostatic influences

of an ion, and for monovalent salt at room temperature, we use $\kappa = 0.32\sqrt{C_s}$ Å^{-1}. C_s is the salt concentration in the molar units. We explored Dpo4 folding at the salt concentrations in the range of 0.01–0.30 M. $\varepsilon_{Coulomb} = 138.94$, $\varepsilon_r = 80.0$ and ε_{ele} is set to 0.49 in order to make the magnitude of the electrostatic interaction between two charged residues located at the averaged native contact distance of salt bridges in the native structure of the apo Dpo4 (0.85 nm) be equal to that of native contact interaction ($\varepsilon_{ij} = 1.0$). To avoid the double-counting, the native contact interactions formed by the oppositely charged residues were scaled down to 0.1, similar to the previous studies [8, 10, 11]. Thus, the salt-bridge native contact has a similar energetic contribution compared with the plain native contact and at the same time, retains the attractive interaction at the long-range distance.

Gromacs (version 4.5.7) was used in the simulations [12] and the SBM input files were generated by the standard instructions from the SMOG tool [13]. Further modifications on the strengths of salt-bridge native contacts were made. The Langevin dynamics integrator was used, and the stochastic temperature coupling was set to be 1.0, which corresponds to a friction coefficient of 1.0. The time step was set to be 0.0005, and non-bonded potentials were cut-off at 3.0 nm. Replica-exchange molecular dynamics (REMD) simulations were applied [14]. 40 replicas for each simulation were used with the temperature ranging from 1.00 to 1.35, concentrating around the corresponding melting temperatures. Exchange frequency between neighbor replicas was set to be every 1000 steps. We found that there are reasonable overall exchange probabilities, which are all higher than 0.2, guaranteeing the sampling efficiency of REMD. Finally, all the trajectories were collected, and the Weighted Histogram Analysis Method (WHAM) program was used to obtain the thermodynamic results [15], such as heat capacity curves, free energy landscapes, and melting curves. The melting curve of individual domain/interface can be used to calculate the probability of folded state *p(T)* along with the temperature by fitting the melting curve to a sigmoidal function. Then the thermodynamic coupling index (TCI) was obtained by using *p(T)* with the following expression [16]:

$$TCI(I,J) = -ln\langle\left|\left(p^I(T) - p^J(T)\right)\right|\rangle,$$

where the index *I* and *J* stand for the individual domain/interface. Therefore, a large (small) value of *TCI* means a high (low) folding cooperativity, i.e., a high (low) degree of synchronous folding between the domains/interfaces.

The constant temperature molecular dynamics simulations at the low salt concentration (C_s = 0.01M) and in the absence of electrostatic interactions (corresponding to the infinite salt concentration) were performed to study the conformational fluctuations of Dpo4. The temperature was set to be room temperature, at which Dpo4 remains folded. The room temperature in simulations was determined by an approximate linear energy-dependence behavior between the simulation and experimental temperatures through mapping the folding temperatures at first [17].

The fluctuation matrix *M(i, j)* was calculated through the following expression:

$$M(i,j) = \frac{1}{N_{frames}} \sum_{n}^{N_{frames}} \sum_{\alpha}^{x,y,z} (r_{i,\alpha}^n - \langle r_{i,\alpha}\rangle)(r_{j,\alpha}^n - \langle r_{j,\alpha}\rangle),$$

where $r^{n}_{i,\alpha}$ is the coordinate of residue i at the α axis (α = x, y, z), $\langle r_{i,\alpha} \rangle$ is the corresponding average of all frames in the trajectory, and N_{frames} is the number of frames in the trajectory.

The localized frustration calculation and analyses were obtained from the frustratometer server with and without considering the electrostatic interactions [18, 19]. These were done by using the plain associative-memory, water-mediated, structure and energy model (AWSEM) force field [20] and the AWSEM force field with electrostatic interactions [21], respectively. The PDB of apo Dpo4 was used to upload (2RDI) [4]. The algorithm running behind the frustratometer server compares the energetic contribution to the extra stabilization energy attributed to a given pair of amino acids to the statistics of the energies that would be found by placing different residues in the same native location or by creating a different environment for the interacting pair [22]. To be precise, the algorithm calculates the frustration index for the contacts in the native structure, defined as a Z-score of the energy of the native pair compared with the decoy energies. Based on the frustration index, contacts can be classified into the minimally frustrated contacts (frustration index higher than 0.78), highly frustrated contacts (frustration index lower than -1), and neutral contacts (frustration index is in between 0.78 and -1) [22].

Additional figures

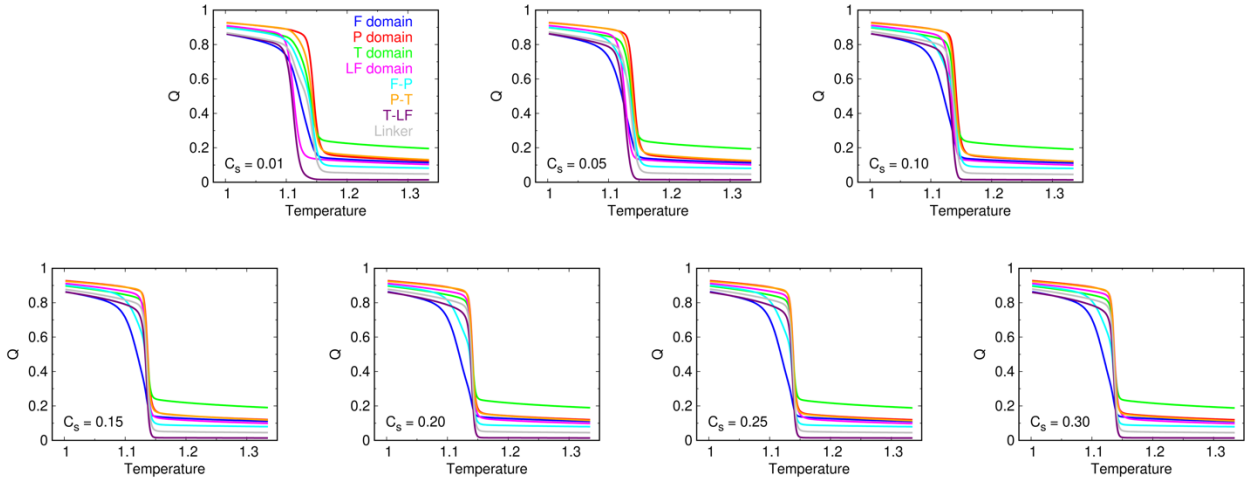

Fig. S1. Melting curves of Q for each intra-, inter-domain, and linker of Dpo4 for different salt concentrations C_s = 0.01-0.30 M. C_s is in the molar units.

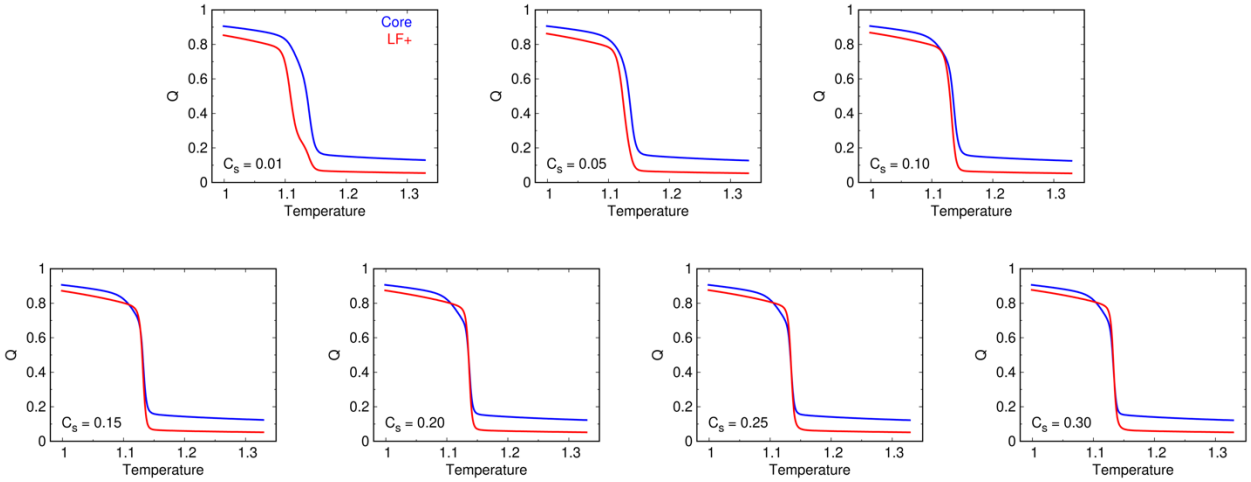

Fig. S2. The melting curves of Q the polymerase core (Core, blue) and the LF domain with the linker (LF+, red) of Dpo4 for different salt concentrations C_s = 0.01-0.30 M. C_s is in the molar units.

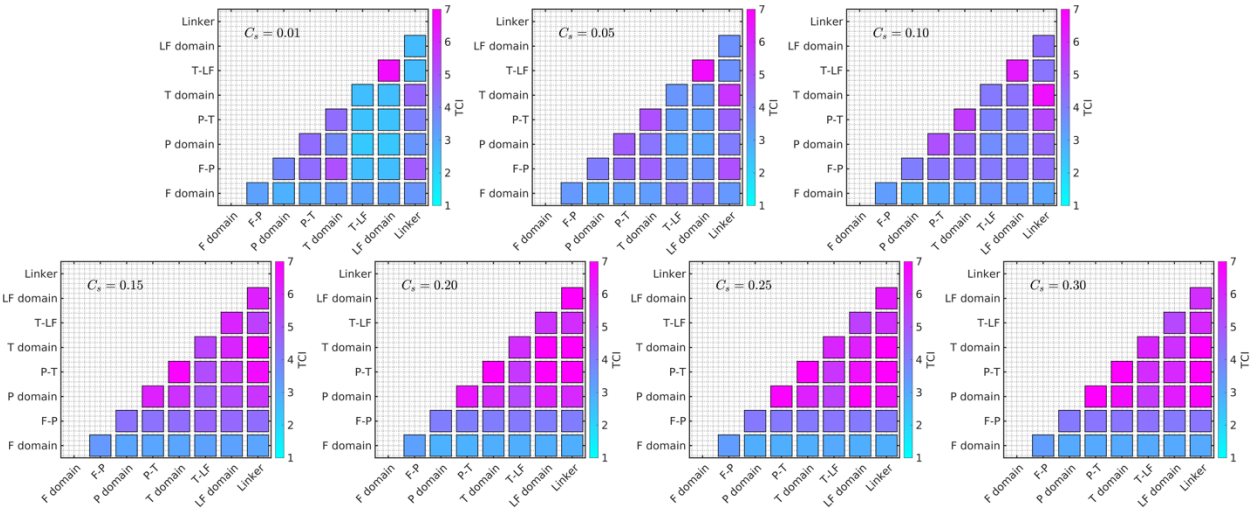

Fig. S3. Thermodynamic coupling index *TCI(I, J)* between different domains/interfaces of Dpo4 for different salt concentrations C_s = 0.01-0.30 M. The C_s is in the molar units.

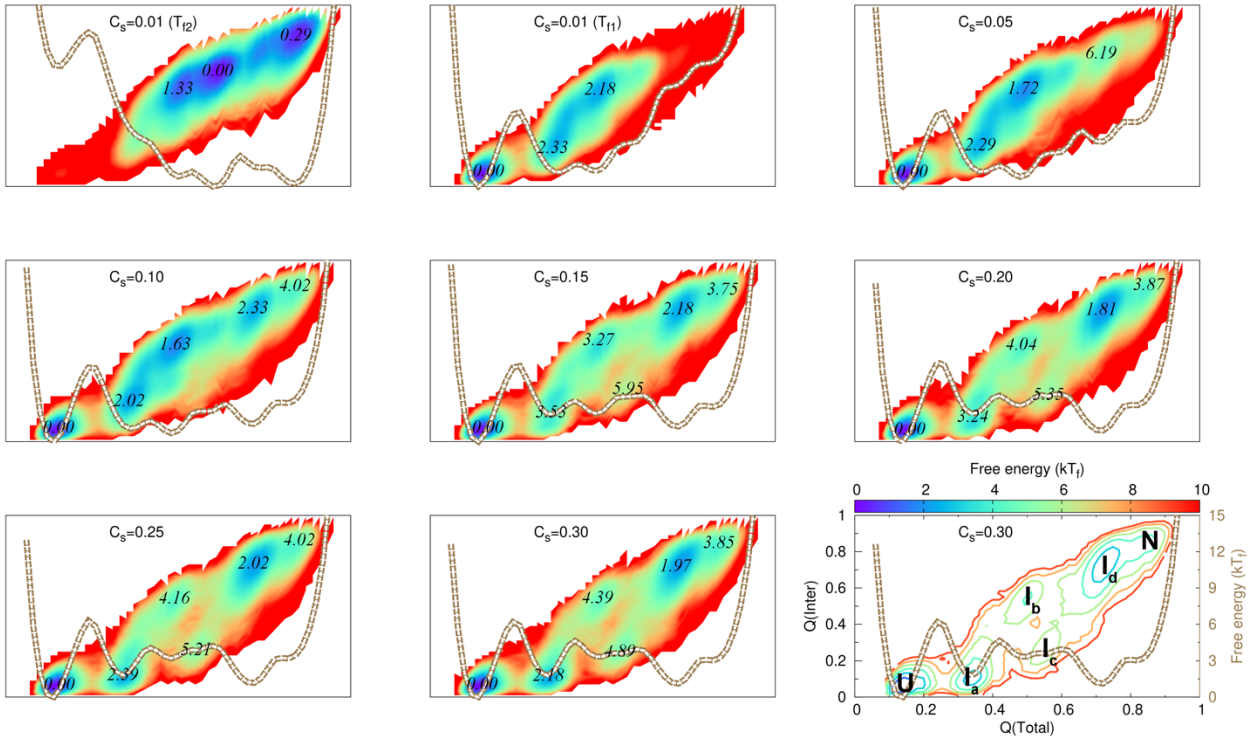

Fig. S4. Salt-dependent Dpo4 folding energy landscapes. The figures are the same with Fig. 3 in the main text but for different salt concentrations C_s = 0.01-0.30 M with the free energy values of the (meta)stable states shown on each free energy landscape plot.

References

1. Clementi, C., H. Nymeyer, and J.N. Onuchic, *Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins.* J Mol Biol, 2000. **298**(5): p. 937-53.
2. Levy, Y., P.G. Wolynes, and J.N. Onuchic, *Protein topology determines binding mechanism.* Proc Natl Acad Sci U S A, 2004. **101**(2): p. 511-6.
3. Levy, Y., et al., *A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes.* J Mol Biol, 2005. **346**(4): p. 1121-45.
4. Wong, J.H., et al., *Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4.* J Mol Biol, 2008. **379**(2): p. 317-30.
5. Sobolev, V., et al., *Automated analysis of interatomic contacts in proteins.* Bioinformatics, 1999. **15**(4): p. 327-32.
6. Givaty, O. and Y. Levy, *Protein sliding along DNA: dynamics and structural characterization.* J Mol Biol, 2009. **385**(4): p. 1087-97.
7. Azia, A. and Y. Levy, *Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt.* J Mol Biol, 2009. **393**(2): p. 527-542.
8. Ganguly, D., et al., *Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins.* J Mol Biol, 2012. **422**(5): p. 674-684.
9. Tzul, F.O., K.L. Schweiker, and G.I. Makhatadze, *Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling.* Proc Natl Acad Sci U S A, 2015. **112**(3): p. E259-E266.
10. Chu, X., et al., *Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone Chz1 and histone H2A. Z-H2B.* PLoS Comput Biol, 2012. **8**(7): p. e1002608.
11. Levy, Y., J.N. Onuchic, and P.G. Wolynes, *Fly-casting in protein− DNA binding: frustration between protein folding and electrostatics facilitates target recognition.* J Am Chem Soc, 2007. **129**(4): p. 738-739.
12. Hess, B., et al., *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation.* J Chem Theory Comput, 2008. **4**(3): p. 435-447.
13. Noel, J.K., et al., *SMOG@ctbp: simplified deployment of structure-based models in GROMACS.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W657-61.
14. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding.* Chem Phys Lett, 1999. **314**(1): p. 141-151.
15. Kumar, S., et al., *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* J Comput Chem, 1992. **13**(8): p. 1011-1021.
16. Sadqi, M., D. Fushman, and V. Muñoz, *Atom-by-atom analysis of global downhill protein folding.* Nature, 2006. **442**(7100): p. 317-21.

17. Chu, X. and V. Muñoz, *Roles of conformational disorder and downhill folding in modulating protein–DNA recognition.* Phys Chem Chem Phys, 2017. **19**(42): p. 28527-28539.
18. Parra, R.G., et al., *Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics.* Nucleic acids research, 2016. **44**(W1): p. W356-W360.
19. Jenik, M., et al., *Protein frustratometer: a tool to localize energetic frustration in protein molecules.* Nucleic Acids Res, 2012. **40**(W1): p. W348-W351.
20. Davtyan, A., et al., *AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing.* J Phys Chem B, 2012. **116**(29): p. 8494-8503.
21. Tsai, M.Y., et al., *Electrostatics, structure prediction, and the energy landscapes for protein folding and binding.* Protein Sci, 2016. **25**(1): p. 255-269.
22. Ferreiro, D.U., et al., *Localizing frustration in native proteins and protein assemblies.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 19819-19824.