Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information (ESI)

Non-Specific DNA-Driven Quinary Interactions Promote Structural Transitions in Proteins

*Soundhararajan Gopi & Athi N. Naganathan**

Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India.

e-mail: athi@iitm.ac.in

Supplementary Methods

Total Partition Function Calculation The total partition function of the WSME model[1, 2] is calculated using the transfer-matrix formalism of Wako and Saitô[1] as described below,

$$Z(T) = v_l \left[\sum_{i=1}^{N} X_i \right] v_r^{tr}$$

where,

$$v_l = (1,1,1, \dots, 1)$$

$$v_r = (1,0,0, \dots ,0)$$

$$X_i = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ H_{DNA}^{(i)} z^{(i)} & 0 & 0 & \dots & & 0 \\ 0 & H_1^{(i)} H_{DNA}^{(i)} z^{(i)} & 0 & & & \vdots \\ & & H_2^{(i)} H_{DNA}^{(i)} z^{(i)} & & 0 & 0 \\ & 0 & & & H_{(n-2)}^{(i)} H_{DNA}^{(i)} z^{(i)} & H_{(n-1)}^{(i)} H_{DNA}^{(i)} z^{(i)} \end{bmatrix}$$

and

$$H_k^{(i)} = exp\left(-\beta \sum_{j=i}^{k} \Delta G_{stab} \right) \qquad (k \leq N - i)$$

$$H_k^{(i)} = 0 \qquad (k > N - i)$$

$$H_{DNA}^{(i)} = exp\left(-\beta \Delta G_{stab}^{i,DNA}\right)$$

where $\beta = 1/RT$, $z^{(i)} = \exp\left(\Delta S_i^{conf}/R\right)$ and $R = 8.314\, J\, mol^{-1}K^{-1}$. The stabilization free energy contribution (ΔG_{stab}) arises from the interaction of residue *i* with *k* consecutive protein residues while $\Delta G_{stab}^{i,DNA}$ is the stabilization energy due the interactions of residue *i* with DNA. Both the stabilization terms (ΔG_{stab} and $\Delta G_{stab}^{i,DNA}$) include contributions from van der Waals interactions modeled using a Gō-like criteria (uniform 5 Å distance cut-off to identify intra- and inter-molecular interactions), charge-charge interactions (between charged residues of protein

and negatively charged phosphate group on DNA, in case of protein-DNA interactions) defined by the Debye-Hückel formulation and a solvation term proportional to the number of contacts.[3, 4]

$$\Delta G_{stab} = E_{vdW} + E_{Elec} + \Delta G_{solv}$$

A residue specific entropic term (ΔS_i^{conf}) accounts for the cost of fixing a residue in native conformation. An excess entropic penalty of -6.1 J mol^{-1} K^{-1} per residue is also introduced to account for the larger degree of freedom associated with coiled residues[5] (as identified using STRIDE[6]) and for glycine residues. An entropic penalty of 0 J mol^{-1} K^{-1} per residue is assumed for proline, owing to its limited flexibility.

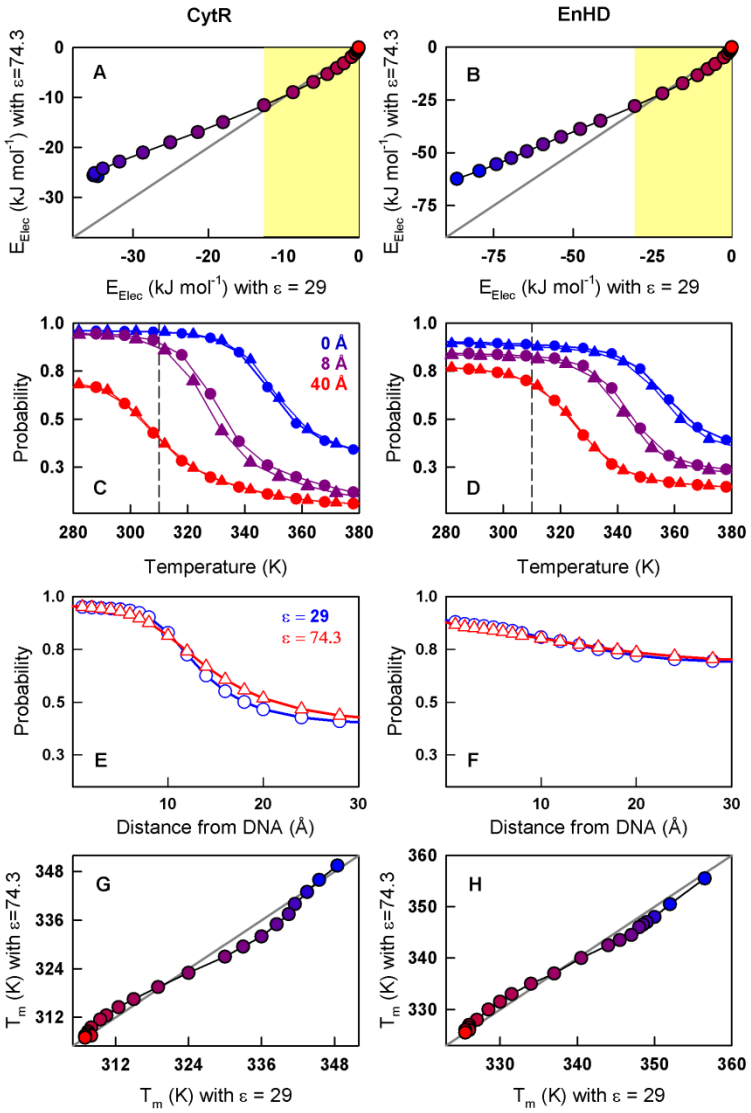

Figure S1 (A-B) Comparison of the electrostatic interaction energy between protein and DNA (kJ mol^{-1}) using inter-molecular dielectric constants ($\varepsilon_{DNA,prot}$) of 29 and 74.3. The color gradient from blue to red represents increasing distances between protein and DNA and the gray line indicates the 1:1 correlation line. Note that with the increasing distance between protein and DNA, the electrostatic interaction energy becomes increasingly linear and comparable (circles in the yellow shaded region). (C-D) Comparing the thermal melting profiles of CytR and EnHD using $\varepsilon_{DNA,prot}$ of 29 (circles) and 74.3 (triangles). The large differences in the electrostatic interaction energy close to the DNA are masked by the increasing contribution from the short-range contacts between protein and DNA. (E-F) The folding probability as a function of distance from DNA, at 310 K and 100 mM ionic strength condition, follows the same trend with very small differences in the predicted probabilities. (G-H) Comparison of the apparent melting temperatures (T_m) estimated using different inter-molecular dielectric constants. The color gradient is the same as in the panels A and B and the 1:1 correlation line is shown in gray.

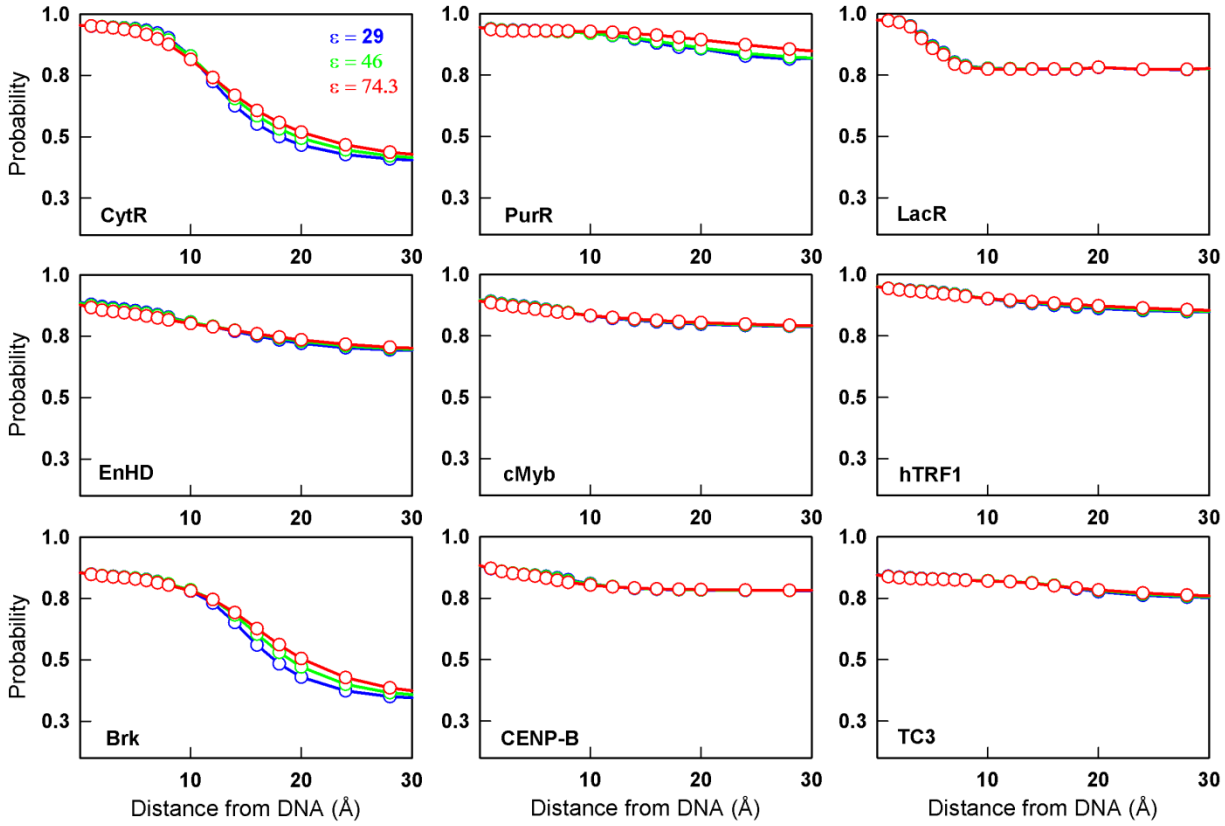

Figure S2 Comparison of folding probability as the function of distance from DNA at 310 K and 100 mM ionic strength conditions calculated using varying inter-molecular dielectric constants. The color coding is same across all the panels.

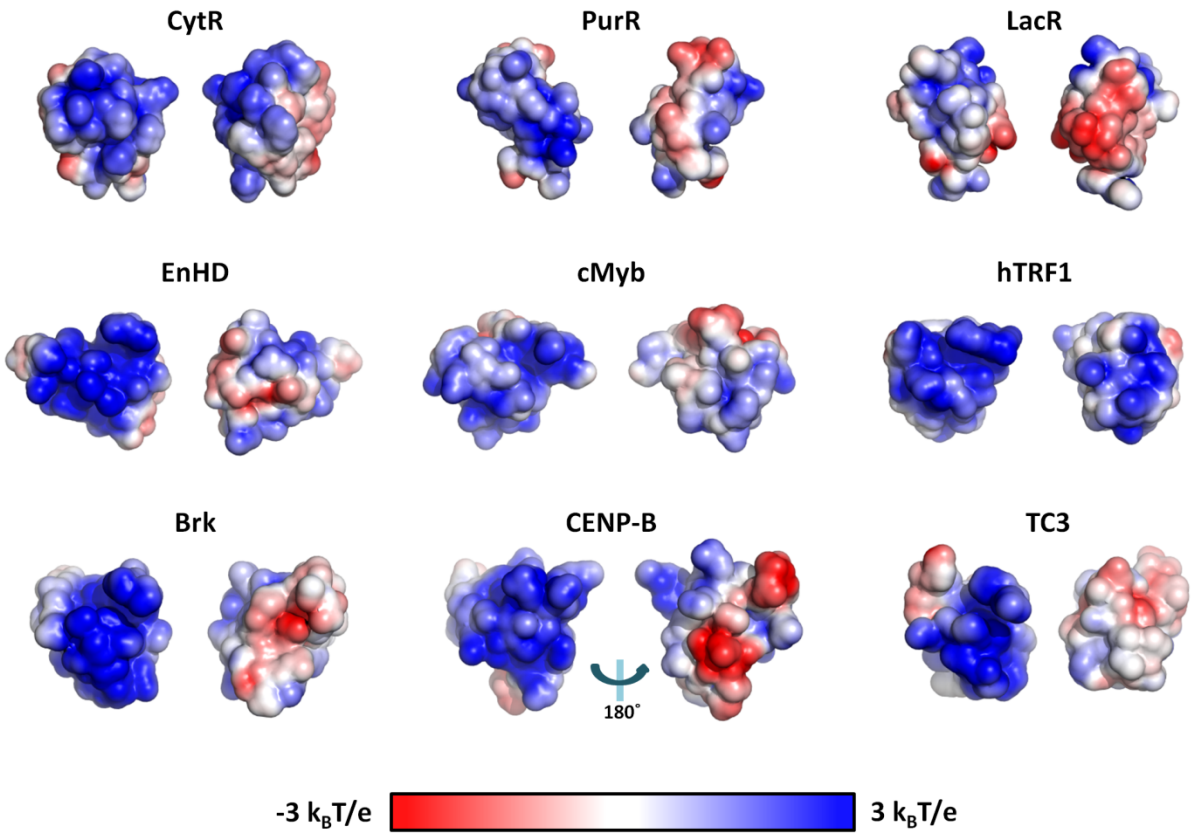

Figure S3 Electrostatic potential mapped onto the surfaces of all nine proteins calculated by solving the non-linear Poisson-Boltzmann equation at 310 K, 100 mM ionic strength condition. Both the DNA-binding face (left) and the opposite face (right; by rotating 180˚ along the vertical axes) are shown for each protein. Note the preferentially segregated negative surface potential (red) on the non-binding face of the proteins, and that is quite prominent for LacR, Brk and CENP-B.

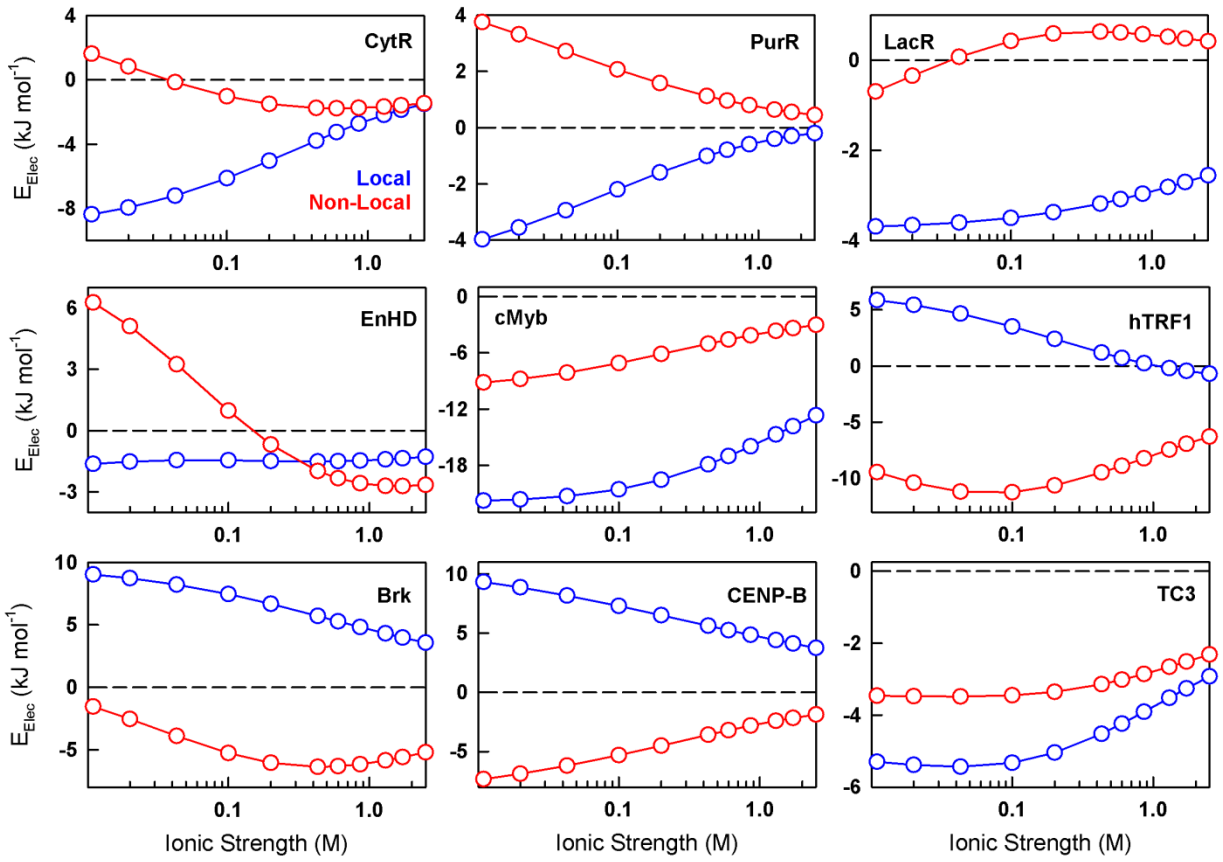

Figure S4 Electrostatic interaction energy as a function of ionic strength (11 mM to 2.5 M) at 310 K calculated using the Tanford-Kirkwood algorithm.[7] Interactions between residues with sequence separation ≤ 4 are considered as local interactions.

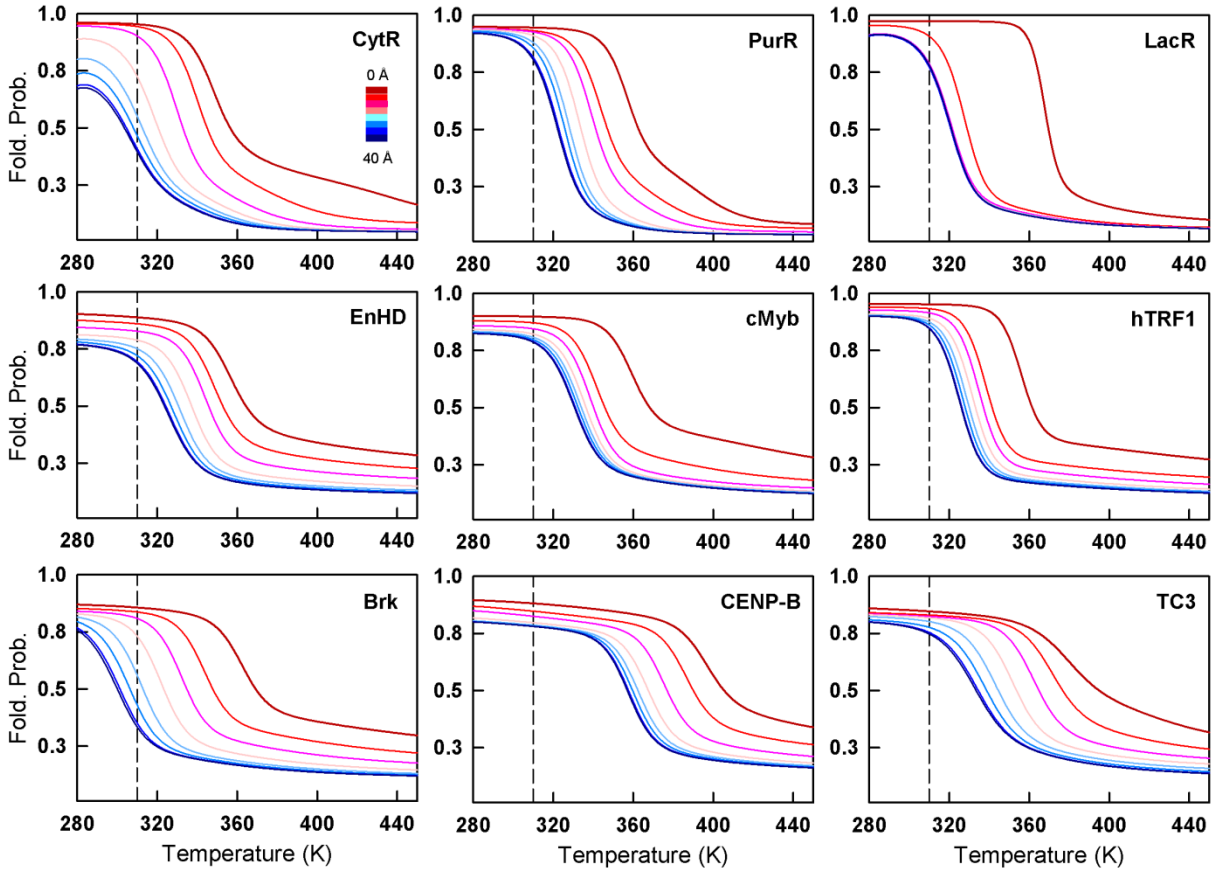

Figure S5 Thermal unfolding curves as a function of distance from DNA (as defined by the color scale in first panel) maintaining the native orientation and at 310 K, 100 mM ionic strength conditions. The vertical dashed line signals 310 K.

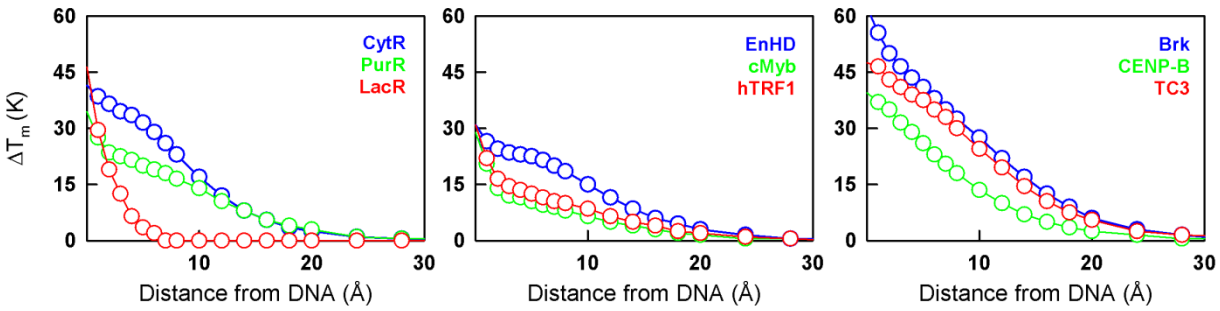

Figure S6 Relative change in the thermal stability calculated as $\Delta T_m = T_m(X) - T_m(\text{unbound})$ where $T_m(X)$ is the thermal stability of protein at distance *X* from DNA maintaining the orientation of the bound conformation.

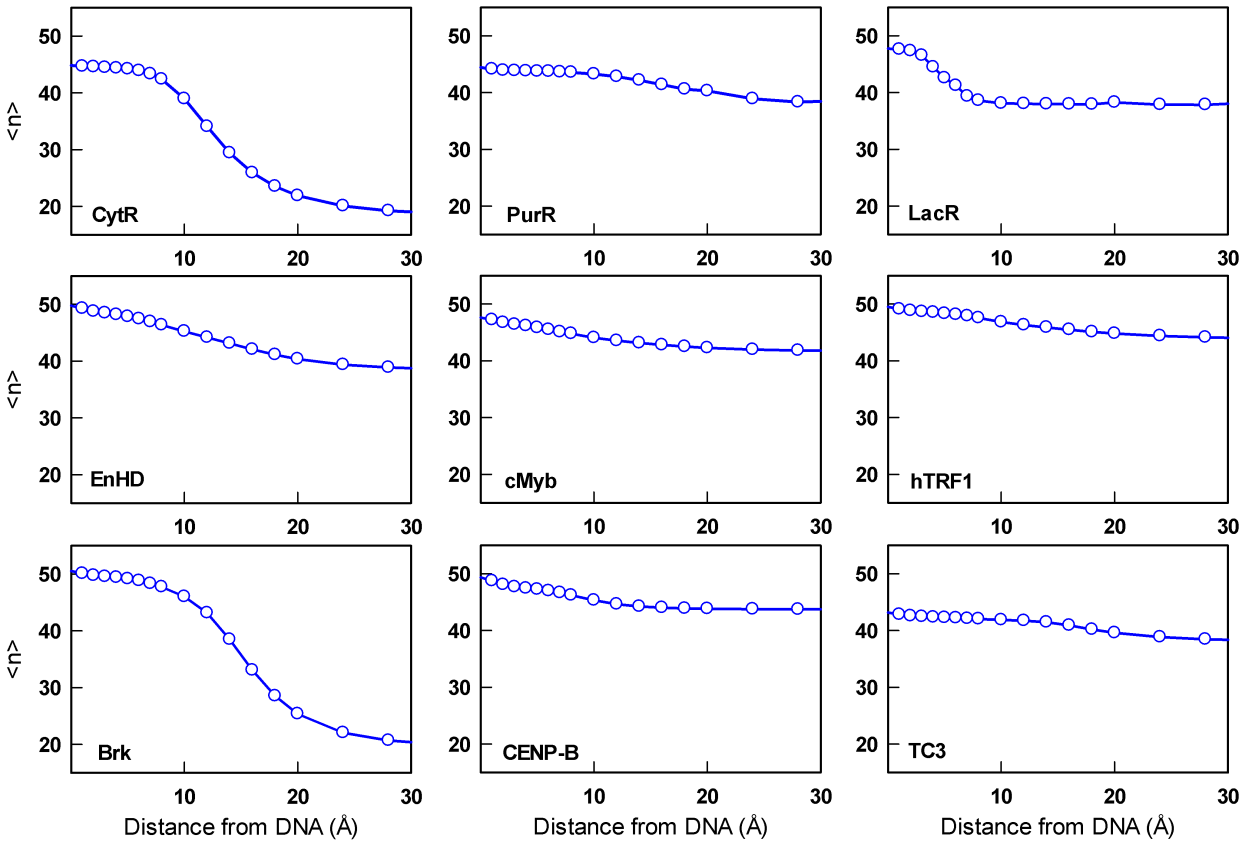

Figure S7 Probability weighted reaction coordinate (<n>) as a function of distance from DNA calculated as a sum over the product of probability density (from the free energy profiles shown in Figure 2) and the corresponding reaction coordinate value (n). Higher the value the more folded is the protein. Clear folding transitions are observed for CytR (top left) and Brk (bottom left) that fold in the presence of DNA.

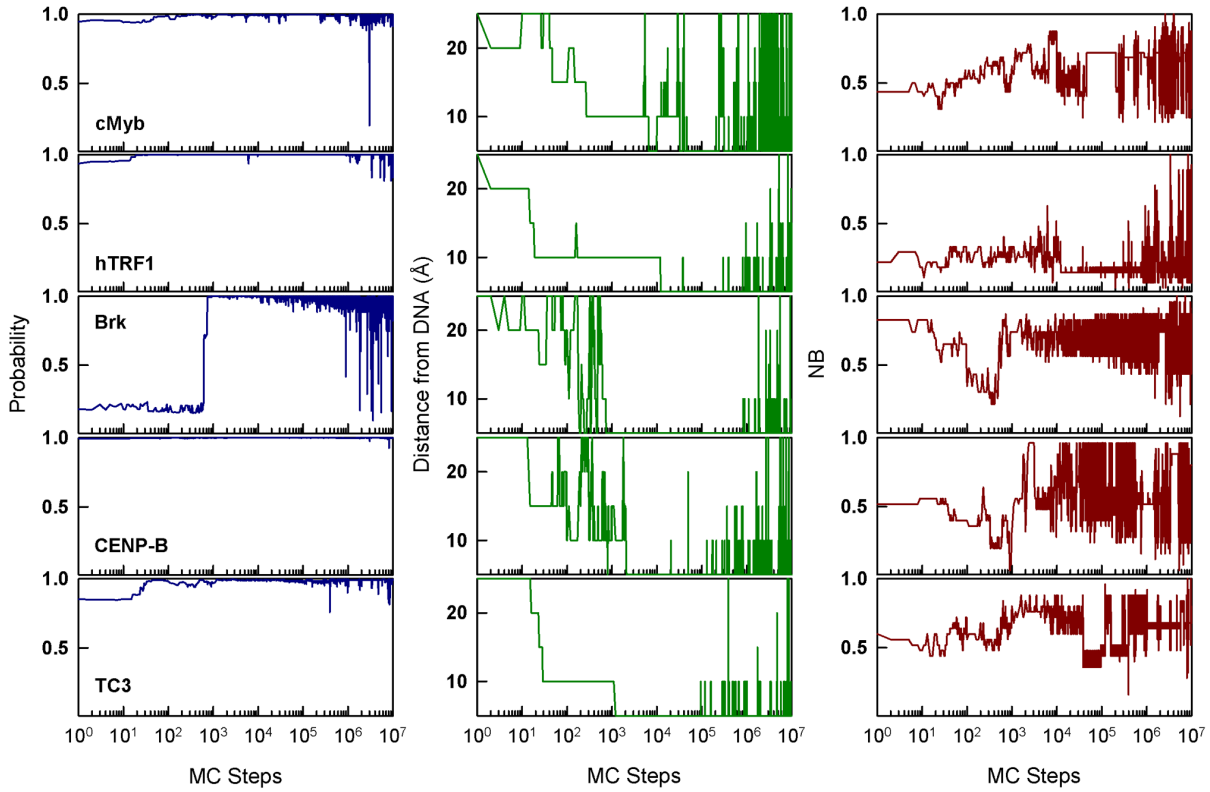

Figure S8 Results of Monte-Carlo simulations on the simulated distance- and orientation-dependent free energy profiles. The left, middle and right columns represent the probability of finding a protein folded, distance from DNA and the fraction of native binding (NB) residues facing DNA.

Protein	Charged residues in the DNA binding face of the protein	Fraction of charged residues in DNA binding face	Charged residues in the opposite face of the protein	Fraction of charged residues in opposite face
CytR	K13, R28, K35 D14, D34	33.3% 50.0%	K18, K20, R41, R43, K46, R49 E45, E50	66.7% 50.0%
PurR	K5, R26 D6	28.6% 25.0%	K9, R10, K24, R33, K41 E30, E31, E42	71.4% 75.0%
LacR	K2, R22 	40% 0%	K33, R35, K37 D8, E11, E36, E39, E44	60% 100%
EnHD	R5, R24, R31, K46, K50, R53, K55, K57, K58 E28	64.3% 16.7%	R15, K17, R18, R29, R30 E11, E19, E22, E37, E42	35.7% 83.3%
cMyb	K144, R165, K171, R176, K182, R190, R191 E151, D152, E168, D178	58.3% 57.1%	K143, R153, K160, R161, K192 E141, E149, E150	41.7% 42.9%
hTRF1	K379, R380, K405,R415, K421, R423, R425, K428, K429 E387, D388, E400, D422	64.3% 80.0%	K389, R392, R396, K397, K411 E386	35.7% 20.0%
Brk	R45, R46, K53, R61, R71, R75, R81, R82, K86 	75.0% 0%	K67, K76, R95 E58, D63, D65, E91	25.0% 100%
CENP-B	K4, R5, R6, K13, R27, K28, R33, K47 E30	61.5% 14.3%	R11, R15, R34, K49, R50 E12, E19, E21, E22, D25, E56	38.5% 85.7%
TC3	R203, R230, K231, R234, R236, R240 E211, E227	66.7% 40.0%	R212, K219, K244 D209, D216, D245	33.3% 60.0%

Table S1 Distribution of charged residues in the proteins. The positively charged residues and the associated percentage distribution on the DNA binding face and the opposite face are shown in blue and those of negatively charged residues are shown in red.

Protein	PDB	N_{res}	ξ (J mol^{-1}) per native contact	ΔS_{conf} (J mol^{-1} K^{-1}) per residue	ΔC_p^{cont} (J mol^{-1} K^{-1}) per native contact
CytR	2L8N[#]	47	-212.7	-33.1[*]	-2.13
PurR	1PRU[#]	47	-181.4	-28.9	-1.71
LacR	1CJG	49	-159.3	-25.9	-1.60
EnHD	2HOS	56	-92.9	-17.4	-0.67
cMyb	1MSE	53	-90.0	-17.4	-0.67
hTRF1	1W0T	52	-89.0	-17.4	-0.67
Brk	2GLO	59	-90.5	-17.4	-0.67
CENP-B	1BW6[#]	56	-104.9	-17.4	-0.67
TC3	1TC3	51	-94.6	-17.4	-0.67

Table S2 Details of proteins and the cxWSME model parameters employed in the current work. Native contacts are identified with a uniform heavy-atom distance cut-off of 5 Å. An excess entropic cost of -6.1 J mol^{-1} K^{-1} per residue is assigned for both non-helical residues identified using STRIDE and glycine residues, while an entropic cost of 0 J mol^{-1} K^{-1} per residue is used to describe the rigidity of proline residues. Both intra- and inter-molecular electrostatic interactions are scaled using a uniform dielectric constant (ε) of 29.

* As the protein is known to be disordered in the absence of DNA, structure independent but sequence dependent entropic cost is used (-39.2 and 0 J mol^{-1} K^{-1} per residue for glycine and proline residues, respectively and a uniform entropic cost of -33.1 J mol^{-1} K^{-1} per residues for all other residues).

[#] DNA bound structures are not available and hence are modeled in PyMOL using the DNA-bound structures of structural homologues as template (i.e.) DNA-bound structure of CytR and PurR (CENP-B) are modeled using LacR (Brk).

	5 Å from DNA	10 Å from DNA	15 Å from DNA	20 Å from DNA	25 Å from DNA	Total
CytR	5805	20363	22104	22104	22104	92480
PurR	4871	17574	21897	22104	22104	88550
LacR	4800	17932	21656	22104	22104	88596
EnHD	1424	15616	21503	22104	22104	82751
cMyb	228	9181	20250	22104	22104	73867
hTRF1	1918	18522	22104	22104	22104	86752
Brk	1547	15931	21854	22104	22104	83540
CENP-B	1621	14210	21426	22104	22104	81465
TC3	1760	17552	21786	22104	22104	85306

Table S3 Total number of non-redundant orientations (without steric clashes with DNA) considered for each protein as the function of distance from DNA. Only few orientations are allowed at 5 Å from DNA due to steric clashes.

Supplementary References

1. H. Wako and N. Saito, *J. Phys. Soc. Japan*, 1978, **44**, 1939-1945.
2. V. Muñoz and W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 11311-11316.
3. A. N. Naganathan, *J. Chem. Theory Comput.*, 2012, **8**, 4646-4656.
4. S. Munshi, S. Gopi, G. Asampille, S. Subramanian, L. A. Campos, H. S. Atreya and A. N. Naganathan, *Nucleic Acids Res.*, 2018, **46**, 8700-8709.
5. N. Rajasekaran, S. Gopi, A. Narayan and A. N. Naganathan, *J. Phys. Chem. B*, 2016, **120**, 4341-4350.
6. M. Heinig and D. Frishman, *Nuc. Acids Res.*, 2004, **32**, W500-W502.
7. C. Tanford and J. G. Kirkwood, *J. Am. Chem. Soc.*, 1957, **79**, 5333-5339.