Supporting Information - Effects of pH on proteins: Predictions for ensemble and single molecule pulling experiments

Edward P. O'Brien[1,2*], Bernard R. Brooks[2], and D. Thirumalai[1†]

[1] *Biophysics Program,*

Institute for Physical Science and Technology

and Department of Chemistry and Biochemistry

University of Maryland,

College Park, MD 20742

[2] *Laboratory of Computational Biology*

National Heart Lung and Blood Institute

National Institutes of Health,

Bethesda, MD 20892

(Dated: September 19, 2011)

* Current address: Dept. of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.

† Corresponding author: Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, phone: 301-405-4803; fax: 301-314-9404; e-mail: thirum@umd.edu

TABLE I: pK_a values of titratable side chains in the native and denatured states

CI2[a]				protein G[b]			
Number	Residue	pK_a^N	pK_a^D	Number	Residue	pK_a^N	pK_a^D
4	Glu	2.9	4.0	15	Glu	4.4	4.0
7	Glu	2.9	4.0	19	Glu	3.7	4.0
14	Glu	3.5	4.0	22	Asp	2.9	3.6
15	Glu	2.8	4.0	27	Glu	4.5	4.0
23	Asp	2.4	3.6	36	Asp	3.8	3.6
26	Glu	3.65	4.0	40	Asp	4.0	3.6
41	Glu	3.14	4.0	42	Glu	4.4	4.0
45	Asp	3.6	3.6	46	Asp	3.6	3.6
52	Asp	2.5	3.6	47	Asp	3.4	3.6
55	Asp	4.95	3.6	56	Glu	4.0	4.0

[a]Values taken from[1].

[b]Values taken from[2].

TABLE II: Parameters used in the C_α-SCM force-field[3] (see Eqs. 4-6 in the main text).

Parameter	value[a]
K_A	30
K_{D_1}	0.70 (0.550)[b]
K_{D_2}	0.00
K_{D_3}	0.35 (0.275)
n_1[c]	1
n_2	2
n_3	3
K_{Ch}	18.98 (21.87)[d]
ϵ_{HB}	0.75 (1.5)[e]
ϵ_i^N	NB[f]
ϵ_i^{NN}	10^{-12}

[a]The unit of energy is kcal/mol.

[b]Values in parentheses are for protein G.

[c]n_j is the dimensionless period of the cosine function of Eq. 3 in reference[3].

[d]For CI2 $K_{Ch} = 18.98$ kcal mol^{-1} degree^{-2}. For protein G $K_{Ch} = 21.87$ kcal mol^{-1} degree^{-2}

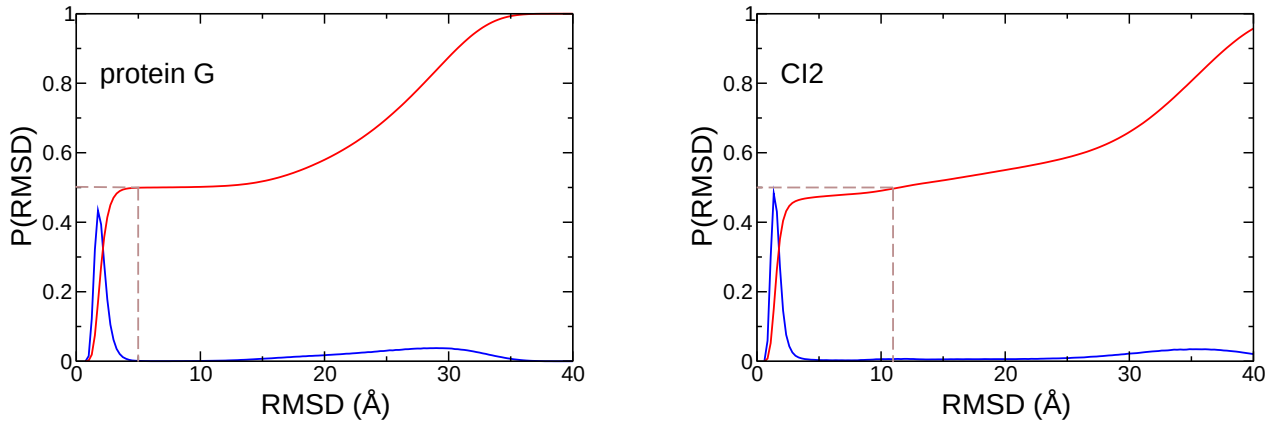
[e]Residue pairs that make just one backbone hydrogen bond are assigned $\epsilon_{HB} = 0.75$. For pairs that make two hydrogen bonds $\epsilon_{HB} = 1.5$.

[f]The statistical potential of Miyazawa-Jernigan[4] formed the basis for choosing ϵ_i^N values for $SC - SC$ interactions. Values reported in Table 5 of[4] were subtracted by 1.2 so that all pair energies would be negative. To obtain protein melting temperatures above 300 K we scaled the resultant values by multiplying them by 0.65, for CI2, and by 0.85 for protein G. The resulting values were assigned to ϵ_i^N based on the amino acids forming the native contact. For native backbone-side chain interactions $\epsilon_i^N = 0.37$ kcal mol^{-1}

Figure Captions

Figure S1: The probability density (blue line) as a function of the RMSD (P(RMSD)) from the crystal structure at the simulated melting temperature for protein G (A), and CI2 (B). The integral over P(RMSD), starting from a lower limit of 0 Å, as a function of the RMSD is shown as a red line. The upper limit at which this integral equals 0.5, indicated by dashed lines, is defined as the RMSD threshold value separating the native and denatured ensembles. The simulation reference pH is 2.3 and 3.5 for protein G and CI2, respectively, and $f = 0$ pN.

Figure S2: Fraction of native contacts for structural elements (Q_{SE}) in CI2 as a function of the distance between the N-terminus and C-terminus projected on to the x-axis (the pulling direction) at pH = 2.5. 'β-12' corresponds to the fraction of native contacts formed between β-strands 1 and 2, 'Helix-β_3' to the Q_{SE} between the α-helix of CI2 and β-strand 3, and so on.


protein G
P(RMSD)
RMSD (Å)
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
CI2
P(RMSD)
RMSD (Å)

(a) (b)

FIG. 1:

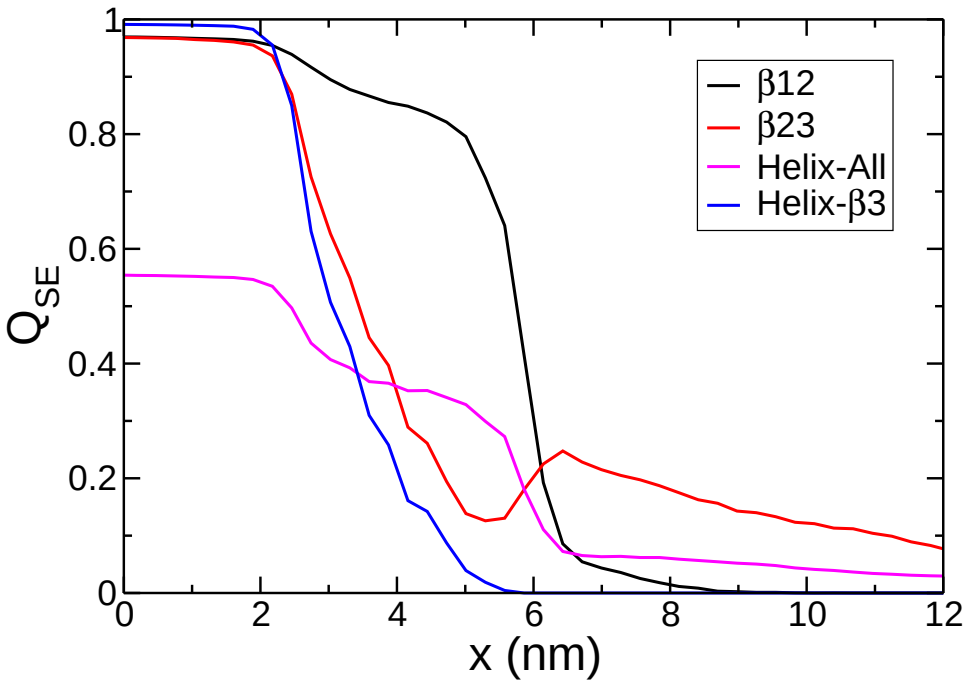


FIG. 2:

References

[1] Tan, Y. J.; Oliveberg, M.; Davis, B.; Fersht, A. R. *J. Molec. Biol.* **1995**, *254*, 980–992.

[2] Khare, D.; Alexander, P.; Antosiewicz, J.; Bryan, P.; Gilson, M.; Orban, J. *Biochemistry* **1997**, *36*, 3580–3589.

[3] O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13403–13408.

[4] Miyazawa, S.; Jernigan, R. L. *Prot. Struc. Func. Gene.* **1999**, *34*, 49–68.