Supporting Information

for

Conformational Sampling of Peptides in the Presence of Protein Crowders from AA/CG-Multiscale Simulations

Alexander V. Predeus[1], Seref Gul[2], Srinivasa M. Gopal[1], and Michael Feig[1,2]*

[1]Department of Biochemistry & Molecular Biology and [2]Department of Chemistry, Michigan State University, East Lansing, MI, 48824, United States

Figure S1: Crystals of protein-protein complexes composed of two proteins (chains A and B) were obtained from the PDB data base (see PDB ID on individual figures). Four types of structures were generated: all-atom chains A and B (red curves), all-atom A and PRIMO B (green curves), PRIMO A and all-atom B (blue curves), PRIMO chains A and B (purple curves). Chain B was then translated along the axis connecting centers of mass of chains A and B to distances ranging from -1 to +15 Å, and system total energies with GBMV implicit solvent were calculated.

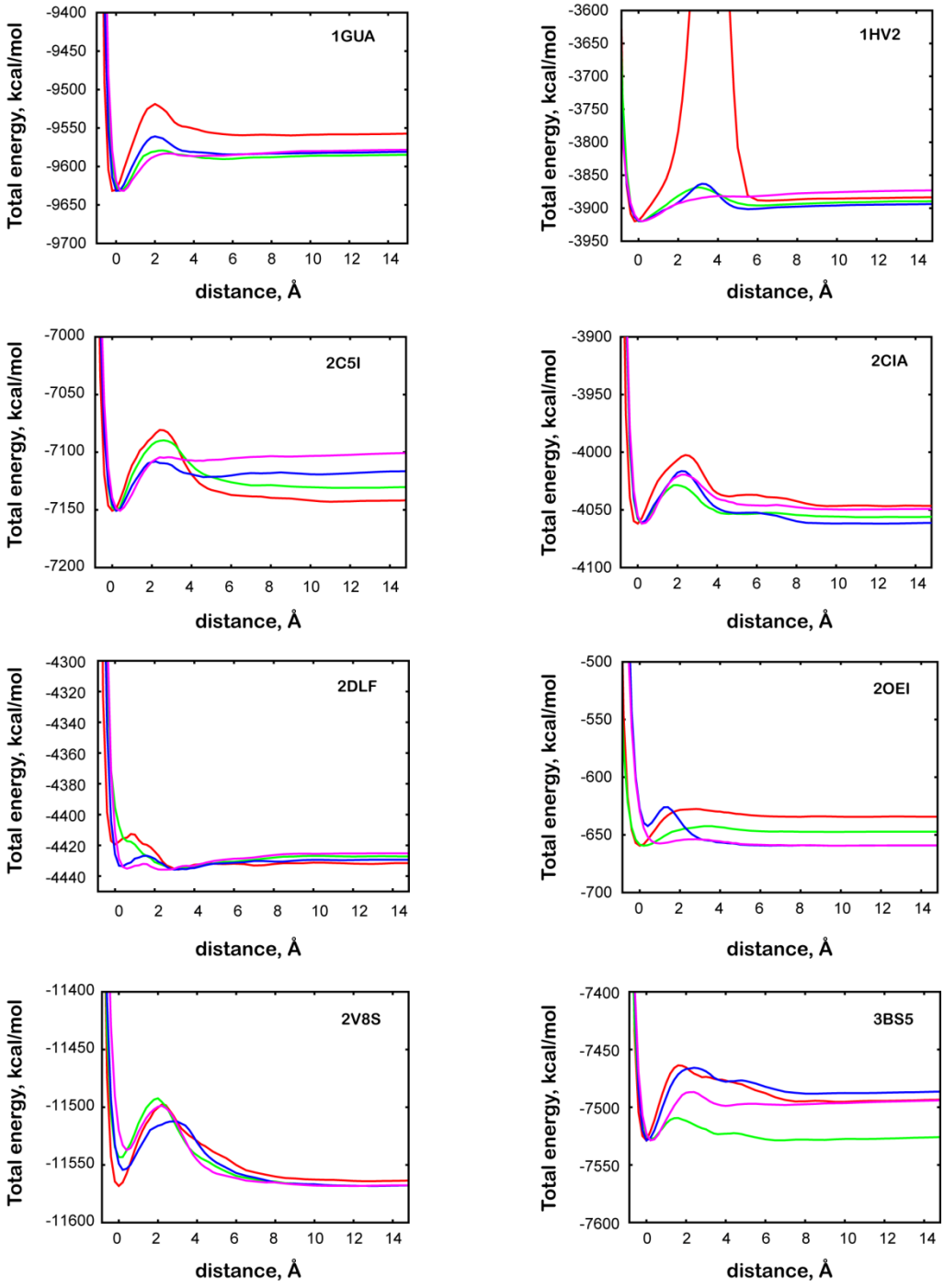

Figure S2: The graph of $state_i$ function, as well as its first and second derivatives, depending on the distance between a pair of residues d_i. Cutoff of 8.0 Å is used in this illustration.

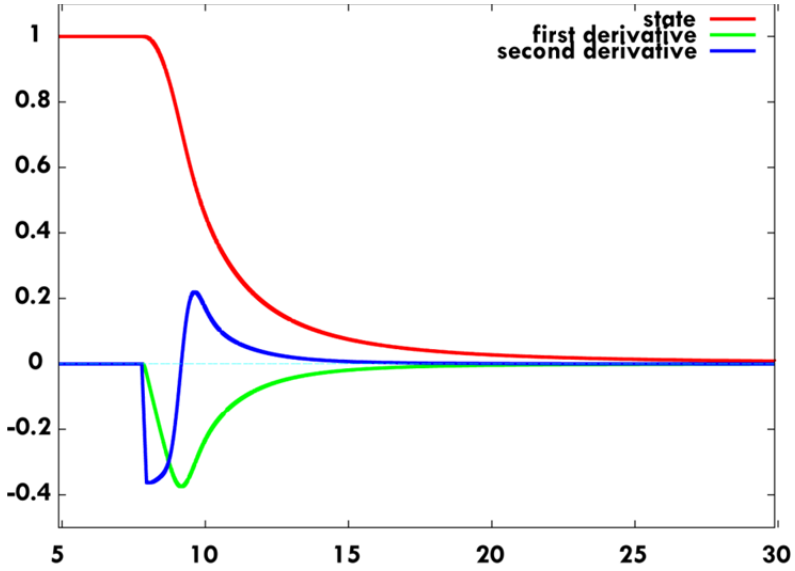

Table S1: Block averaged cluster populations of trp-cage during three simulation intervals

ε	time	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
5	10-40 ns	0.0	30.0	0.1	22.8	47.1	0.0	0.0	0.0
	40-70 ns	0.0	81.0	3.4	8.2	7.4	0.0	0.0	0.0
	70-100 ns	0.0	78.1	0.4	10.7	10.8	0.0	0.0	0.0
20	10-40 ns	16.8	57.0	6.4	15.7	3.4	0.0	0.0	0.0
	40-70 ns	15.6	14.1	10.8	55.1	3.9	0.0	0.0	0.0
	70-100 ns	6.7	50.9	5.4	32.7	4.2	0.0	0.0	0.0
40	10-40 ns	57.1	3.5	35.9	1.1	0.1	0.0	0.0	0.0
	40-70 ns	71.8	1.4	18.6	1.5	0.2	0.0	0.0	0.0
	70-100 ns	47.0	7.6	37.1	3.4	0.4	0.0	0.1	0.0
80	10-40 ns	47.7	2.9	39.8	1.1	0.0	0.0	0.7	0.0
	40-70 ns	58.8	1.0	32.5	2.7	0.0	0.0	0.4	0.0
	70-100 ns	55.2	4.2	24.5	1.5	0.0	0.0	0.0	0.0
multi	5-10 ns	33.0	14.3	28.4	4.5	0.0	9.6	5.6	2.7
	10-15 ns	21.2	14.5	42.4	11.7	0.0	3.7	3.7	2.8
	15-20 ns	11.4	19.3	36.6	16.7	1.3	4.8	5.8	4.0

Table S2: Block averaged cluster populations of melittin during three simulation intervals

ε	time	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
5	7.5-30 ns	90.1	0.0	9.9	0.0	0.0	0.0	0.0
	30-52.5 ns	89.8	0.0	9.2	0.0	0.0	0.0	0.0
	52.5-75 ns	85.6	0.0	10.8	0.0	0.0	0.0	0.0
20	7.5-30 ns	58.9	17.3	11.3	3.5	7.7	0.0	0.0
	30-52.5 ns	28.6	27.6	6.6	7.3	29.7	0.1	0.0
	52.5-75 ns	29.4	34.0	7.5	12.3	16.5	0.0	0.0
40	7.5-30 ns	37.9	32.0	10.8	7.4	0.2	0.3	0.0
	30-52.5 ns	32.6	28.1	9.6	12.4	0.1	0.0	0.0
	52.5-75 ns	23.6	51.4	8.1	11.3	0.0	0.4	0.0
80	7.5-30 ns	15.4	55.7	4.5	9.9	0.3	0.1	0.0
	30-52.5 ns	0.5	68.2	0.1	13.3	0.0	0.0	0.0
	52.5-75 ns	0.0	85.0	0.0	9.5	4.7	0.0	0.0
multi	5-10 ns	57.7	6.4	4.4	0.0	0.0	16.8	9.6
	10-15 ns	52.7	11.4	13.6	0.0	0.0	13.2	2.2
	15-20 ns	37.4	11.8	22.0	0.0	0.0	13.3	2.0

Table S3. Contacts found in melittin's native conformation obtain through CHARMM's DMAT facility. Left to right, residue 1 number, residue 1 name, residue 2 number, residue 2 name, and distance between centers of geometry of side chain's heavy atoms in Å. For GLY, C_α was used as a side chain.

4	ALA	1	GLY	3.632038
5	VAL	4	ALA	5.849712
7	LYS	4	ALA	5.587463
8	VAL	4	ALA	6.339722
8	VAL	5	VAL	6.054344
9	LEU	5	VAL	5.681451
9	LEU	8	VAL	6.498944
10	THR	6	LEU	4.997338
10	THR	9	LEU	6.265362
11	THR	10	THR	5.467035
13	LEU	8	VAL	4.829834
13	LEU	9	LEU	6.37523
14	PRO	11	THR	6.444817
14	PRO	13	LEU	5.932506
15	ALA	14	PRO	6.053349
16	LEU	13	LEU	6.183247
17	ILE	13	LEU	6.289258
17	ILE	14	PRO	5.878261
17	ILE	16	LEU	6.462913
18	SER	15	ALA	6.000315
19	TRP	16	LEU	5.688582
20	ILE	16	LEU	5.975275
20	ILE	17	ILE	5.547878
20	ILE	19	TRP	6.305514
21	LYS	17	ILE	6.373267
21	LYS	18	SER	5.289639
22	ARG	18	SER	5.114005
23	LYS	19	TRP	4.586513
24	ARG	21	LYS	6.21103
26	GLN	23	LYS	5.966605

Table S4. Contacts found in trp-cage native conformation. Left to right, residue 1 number, residue 1 name, residue 2 number, residue 2 name, and distance between centers of geometry of side chain's heavy atoms in Å. For GLY, C_α was used as a side chain.

3	TYR	2	LEU	5.751879
6	TRP	3	TYR	6.098105
7	LEU	3	TYR	5.940856
7	LEU	6	TRP	5.360793
8	LYS	5	GLN	5.252648
9	ASP	8	LYS	3.798363
12	PRO	6	TRP	5.184951
13	SER	12	PRO	5.626726
14	SER	9	ASP	5.17474
16	ARG	9	ASP	4.362751
16	ARG	14	SER	4.3398
18	PRO	6	TRP	4.918295
18	PRO	12	PRO	6.40614
18	PRO	17	PRO	5.862235
19	PRO	2	LEU	4.732281
19	PRO	3	TYR	5.138424
19	PRO	6	TRP	4.659135
19	PRO	18	PRO	5.669231

Definition of the coordinate based on the fraction of native contacts

The coordinate is based on the definition used in Sheinerman, F.B.; Brooks, C. L. *J. Mol Biol.*, **1998**, *278*, 439-456. The definition was modified to allow for more continuous switching, which in turn makes the resulting population more evenly distributed (the use of original definition resulted in ladder-like energy profiles due to relatively small number of native contacts).

The coordinate (ρ) is defined as follows:

$$\rho = \frac{\sum_{i=1}^{N} w_i(1 - state_i)}{\sum_{i=1}^{N} w_i} \tag{S1}$$

The coordinate is based on a map of contacts between N pairs of residues. In the definition that we use in this work, a pair of residues is in contact if the distance between centers of geometry of residues' side chains (heavy atom only) is less than 6.5 Å. Since the reference state can be generated not only from a single snapshot but from a dynamic trajectory, every contact's abundance is accounted for by its weight w_i that is based on the probability of this contact (e.g. if in a 1000 frame trajectory the contact appears 700 times, it's weight would be 0.7). If we use a single structure, as in this work, w_i = 1.0 for all contacts.

Every contact is then described by a state switch function, that is equal to ~ 1 everywhere within a cutoff distance (in our case, 6.5 Å), and then is gradually switched to zero on longer distances. The exact form of this function is given in Equation S2:

$$state_i = \begin{cases} 1, & if\ d_i - cutoff < -0.1 \\ 1 - \frac{C_1}{3C_2}(d_i - cutoff + 0.1)^3, & if\ -0.1 \leq d_i - cutoff \leq 0.066090 \\ \frac{1}{C_2}\left(\frac{1}{1+C_3(d_i - cutoff+5)^{20}} + \frac{25}{5+(d_i - cutoff)^2}\right), & if\ d_i - cutoff > 0.066090 \end{cases} \tag{S2}$$

The function is of class C2, with a smooth first and a continuous second derivative (see Figure S2). Coefficients are C_1 = 6.5147216196, C_2 = 5.9933363671, and C_3 = 1.0e-16.