

# Supplement Materials: Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential.

Chi Zhang<sup>1</sup>, Song Liu<sup>1</sup>, and Yaoqi Zhou\*

Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology & Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, NY 14214

\* Corresponding Author: Dr. Yaoqi Zhou, Howard Hughes Medical Institute Center for Single Molecule Biophysics and Department of Physiology & Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, NY 14214, Phone: (716) 829-2985, Fax: (716) 829-2344, Email: yqzhou@buffalo.edu

<sup>1</sup> These two authors contribute equally to this work

Table S1: Number of loops for a given loop length in RAPPER and Jacobson decoy sets.

Loop Length		2	3	4	5	6	7	8	9	10	11	12
RAPPER	Full	34	34	35	35	36	38	32	37	37	33	34
	Filtered <sup>a</sup>	n/a	n/a	31	30	24	27	21	24	23	n/a	n/a
Jacobson	Full	n/a	n/a	40	160	147	114	101	97	49 <sup>b</sup>	37	21
	Filtered <sup>c</sup>	n/a	n/a	35	117	100	82	66	57	29 <sup>b</sup>	18	10

<sup>a</sup>Some loops are removed by Jacobson et al. (2003) based on 1) the pH at which the protein was crystallized, 2) whether the loop interacts with heteroatom species (ligands or ions) and 3) the quality of the crystal structure in the prediction region.

<sup>b</sup>The original set in Jacobson’s work has 71/40(full/filtered) decoy sets, but the current work using Jacobson’s on-line decoy set which is incomplete (<http://francisco.compchem.ucsf.edu/jacobson/decoy.htm>).

<sup>c</sup> Some loops are removed by Jacobson et al. for the same reasons listed above.

Table S2: The average global rmsd(Å) and standard deviations of the lowest energy decoys in RAPPER decoy set given by three energy functions.

Loop length	rmsd (Å)					
	Direct <sup>a</sup>			Minimized <sup>b</sup>		Filtered <sup>c</sup>
	RAPDF <sup>d</sup>	AMBER/GBSA <sup>e</sup>	DFIRE <sup>f</sup>	AMBER/GBSA <sup>e</sup>	DFIRE <sup>f</sup>	DFIRE <sup>c</sup>
2	0.53 ± 0.42	0.49	0.41 ± 0.17	0.35 ± 0.19	0.39 ± 0.15	n/a
3	0.78 ± 0.50	0.49	0.64 ± 0.46	0.37 ± 0.19	0.50 ± 0.21	n/a
4	1.19 ± 0.66	0.70	0.86 ± 0.48	0.47 ± 0.28	0.81 ± 0.44	0.79 ± 0.44
5	1.40 ± 0.97	1.02	1.00 ± 0.54	0.90 ± 0.98	0.97 ± 0.46	0.93 ± 0.44
6	2.30 ± 1.74	1.19	1.85 ± 1.67	0.95 ± 0.86	1.94 ± 1.67	1.26 ± 0.75
7	2.60 ± 1.82	1.57	1.51 ± 0.87	1.37 ± 1.24	1.54 ± 0.89	1.37 ± 0.83
8	2.90 ± 2.10	2.38	2.11 ± 1.29	2.28 ± 1.90	2.28 ± 1.33	1.80 ± 0.96
9	3.92 ± 2.32	2.44	2.58 ± 1.48	2.41 ± 1.65	2.39 ± 1.08	2.39 ± 1.10
10	4.75 ± 2.41	3.54	3.60 ± 1.76	3.48 ± 1.98	3.52 ± 1.78	3.25 ± 1.59
11	5.97 ± 2.60	4.93	4.25 ± 2.23	4.94 ± 2.76	4.11 ± 2.21	n/a
12	6.22 ± 2.57	5.00	4.32 ± 1.96	4.99 ± 2.58	4.31 ± 2.14	n/a

<sup>a</sup> Decoy selection without energy minimization.

<sup>b</sup> Decoy selection with energy minimization.

<sup>c</sup> DFIRE potential with sequential sidechain minimization for the filtered dataset.

<sup>d</sup> The result from RAPDF<sub>1</sub> (de Bakker et al., 2003). (RAPDF<sub>2</sub> in the paper is not shown since the difference from RAPDF<sub>1</sub> is small.)

<sup>e</sup> Using AMBER/GBSA force field for scoring.

<sup>f</sup> Using DFIRE for scoring. This work.

Table S3: Correlation coefficients (and standard deviations) between the global rmsd values ( $\text{\AA}$ ) and DFIRE energy scores with and without energy minimization of sidechain conformations for the full RAPPER decoy set. The results of RAPDF and AMBER/GBSA with minimization are also listed for comparison.

Loop length	Corr. coeff.			
	Direct <sup>a</sup>	Minimization <sup>b</sup>	RAPDF <sup>c</sup>	AMBER/GBSA <sup>d</sup>
2	0.71 ± 0.28	0.77 ± 0.24	0.63 ± 0.37	0.80 ± 0.29
3	0.64 ± 0.20	0.75 ± 0.13	0.58 ± 0.26	0.55 ± 0.46
4	0.66 ± 0.19	0.70 ± 0.17	0.52 ± 0.27	0.56 ± 0.43
5	0.56 ± 0.25	0.63 ± 0.23	0.47 ± 0.37	0.59 ± 0.31
6	0.39 ± 0.25	0.45 ± 0.25	0.34 ± 0.33	0.49 ± 0.34
7	0.46 ± 0.21	0.53 ± 0.22	0.34 ± 0.36	0.45 ± 0.35
8	0.46 ± 0.22	0.50 ± 0.21	0.38 ± 0.31	0.31 ± 0.27
9	0.37 ± 0.22	0.42 ± 0.22	0.21 ± 0.31	0.30 ± 0.26
10	0.28 ± 0.23	0.31 ± 0.25	0.14 ± 0.28	0.16 ± 0.30
11	0.34 ± 0.20	0.39 ± 0.20	0.12 ± 0.31	0.24 ± 0.24
12	0.28 ± 0.21	0.35 ± 0.21	0.14 ± 0.26	0.12 ± 0.25

<sup>a</sup> DFIRE without rotamer library re-optimization.

<sup>b</sup> DFIRE with rotamer library re-optimization.

<sup>c</sup>RAPDF (de Bakker et al., 2003).

<sup>d</sup> AMBER/GBSA with minimization(de Bakker et al., 2003).

Table S4: The average global rmsd ( $\text{\AA}$ ) and standard deviation of the lowest energy decoys in Jacobson decoy set.

Loop length	rmsd ( $\text{\AA}$ )			
	Full <sup>a</sup>		Filtered <sup>b</sup>	
	OPLS/SGB-NP <sup>c</sup>	DFIRE <sup>d</sup>	OPLS/SGB-NP <sup>c</sup>	DFIRE <sup>d</sup>
4	0.24 ± 0.15	0.40 ± 0.24	0.24 ± 0.15	0.38 ± 0.22
5	0.44 ± 0.62	0.76 ± 0.72	0.43 ± 0.69	0.73 ± 0.71
6	0.59 ± 0.82	0.80 ± 0.76	0.52 ± 0.69	0.72 ± 0.69
7	0.77 ± 1.36	1.15 ± 1.27	0.61 ± 1.03	0.97 ± 0.99
8	0.98 ± 1.09	1.47 ± 1.41	0.84 ± 0.88	1.24 ± 1.14
9	1.37 ± 1.59	1.97 ± 2.00	1.28 ± 1.56	1.54 ± 1.54
10	1.70 ± 1.75	1.52 ± 1.43 <sup>e</sup>	1.22 ± 1.17	1.28 ± 1.28 <sup>e</sup>
11	2.66 ± 2.29	2.48 ± 1.81	1.63 ± 2.02	1.52 ± 1.20
12	2.67 ± 2.02	2.18 ± 1.66	2.28 ± 1.84	1.92 ± 1.16

<sup>a</sup> The results for the full decoy set.

<sup>b</sup> The results for the filtered set.

<sup>c</sup> Scoring function by OPLS/SGB-NP.

<sup>d</sup> This work.

<sup>e</sup> Incomplete online decoy set (See note under Table S1).

Table S5: Correlation coefficients between the rmsd values ( $\text{\AA}$ ) of decoys and energy scores for the Forrest-Woolf set using various force fields.

Method	Rhodopsin	Ca <sup>2</sup> -ATPse
VAC <sup>a</sup>	-0.26	0.37
VAC_UN <sup>a</sup>	0.57	0.77
VAC+ASP <sup>a</sup>	-0.20	0.44
VAC+ASP_UN <sup>a</sup>	0.61	0.80
ACS <sup>a</sup>	0.63	0.81
ACS_UN <sup>a</sup>	0.63	0.81
DDD <sup>a</sup>	0.11	0.59
DDD_UN <sup>a</sup>	0.57	0.77
DDD+ASP <sup>a</sup>	0.19	0.64
DDD+ASP_UN <sup>a</sup>	0.61	0.80
DDD+EEF1_UN <sup>a</sup>	0.62	0.77
FDPB <sup>a</sup>	0.62	0.77
FDPB_UN <sup>a</sup>	0.60	0.73
DFIRE <sup>b</sup>	0.89	0.96

<sup>a</sup> The results from Forrest and Woolf (2003). These physical based energy functions are the combination of CHARMM with various solvation models denoted by FDPB, ACS\_UN, DDD+EEF\_UN, DDD+ASP\_UN, DDD\_UN, and VAC\_UN. (Here, FDPB denotes finite-difference Poisson Boltzmann, ACS, Analytical continuum solvent model, UN, uncharged neutral ionic residues, DDD, the distance-dependent dielectric constant, ASP, the atomic solvation parameters, EEF, effective energy function, VAC, CHARMM vacuum energy function).

<sup>b</sup> This work.