SUPPORTING MATERIALS

Improving Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization

Dong Xu and Yang Zhang

Table S1. Training set of 262 "hard" proteins. Each protein name is represented by 4-letter PDB ID and 1-letter chain ID.

2gpiA	1tifA	1em8A	1vbvA	1pm4A	1h9dB	3eipA	3doeB	1kshB	3feaA
3di2A	1b3tA	2vlqB	2qv8A	2ia7A	1vlsA	1bjaA	2ehpA	3bt4A	1ifrA
2pk8A	1bm8A	1h05A	4ubpB	1ygtA	1v74B	1yozA	1o6dA	2gj2A	1xvhA
2sicI	3cniA	1szhA	2v75A	1h2cA	1ozjA	2nmmA	2gtsA	3bcyA	3fxhA
1x8yA	1wddS	1ztdA	2rhkA	2rh3A	3eziA	1k2xB	1jiwI	2p58C	2ehwA
2vqgA	1bcpD	2qffA	1w1wE	2hjmA	2bcmA	2oyyA	2cu6A	2o6kA	3dt5A
3d1mA	3epvA	2r19A	3d33A	1hxrA	1p5vB	2a7bA	2gr8A	2nrrA	2endA
2z1cA	1yz1A	1tuwA	3dqgA	2ai4A	3bl2A	2inwA	1tulA	3evyA	3dtdA
3fapB	2o5hA	1pucA	2va0A	2uubR	2gnrA	1fc3A	3f6zB	2qklB	2gu3A
1a1xA	2iiaA	1zvaA	1z0pA	3fyfA	2vxgA	3proC	1mr1C	1egwA	2v1lA
2gmfA	2fupA	1l8dA	1j27A	2b3gA	1zhsA	1dszA	3e0jB	2pi2E	2efvA
2vt8A	2z0tA	3ejjA	2qzbA	2zkoA	2huhA	3c5oA	2p9xA	1z21A	2z5bA
2jbyA	1r0uA	2v1tA	2gmqA	1l7lA	3by7A	2bzvA	1vjlA	3db3A	2od5A
3c5xC	1nz0A	2qq4A	2vh3A	3encA	1f32A	2rdcA	2hw2A	1zbxB	2rbgA
2ga1A	3dcxA	2rjiA	1z8uA	1hekA	2v89A	3dnxA	2v3sA	1ahsA	3c1qA
3ebqA	1zt3A	1usuB	2bnlA	2yskA	2dm9A	3f59A	1zpsA	2iihA	1qqrA
2ahmA	3bxuA	2qzqA	1s4cA	3fgxA	3f62A	2fztA	2gujA	1g2rA	1nigA
1gvpA	3c8iA	1vsrA	2ebeA	3be3A	3eoiA	3facA	3fxtA	1jg5A	2w0gA
1uj8A	3cqxC	3fkeA	1wnaA	2dstA	1y9lA	2dpfA	2o9uX	1g4yB	2qziA
2r2cA	3c9pA	3bypA	1wjxA	1f46A	2zfdB	3esmA	1t3uA	2b8iA	2ccqA
1osyA	2oy9A	1regX	3fbiB	1cmcA	3e0zA	1ujwB	3e9vA	1lj2A	1l0sA
1thqA	3fb9A	1gd8A	1y0nA	2hc8A	1oz9A	1sedA	1a62A	1i8nA	2o0qA
3d0wA	1eteA	2oszA	1q8dA	1m5iA	1u2mA	3dnjA	1zldA	1ed1A	2mcmA
2dbsA	1efnB	3e11A	1whiA	3g73A	1hcnB	2hu9A	3bqaA	1xg8A	1lm8V
1q5zA	3bp3A	2opeA	2fomB	2es9A	2fufA	1wv8A	2qlcA	1v2zA	2qudA
1k8rB	1lktA	2gagD	3d4uB	3dfdA	2f6mB	1zs4A	1e44B	1vqo3	2vlqA
2hd0A	1ezjA								

Table S2. Testing set of 148 “hard” proteins. Each protein name is represented by 4-letter PDB ID and 1-letter chain ID.

3chbD	1wn2A	1ykuA	1wxcB	1ufyA	2ofcA	2covD	2j9wA	1t6uA	2fcwA
2nmlA	1kafA	2ip6A	1j3wA	1nlqA	2chhA	1mn8A	2jekA	1n8vA	1u84A
2v33A	1w53A	1ugiA	1o82A	1ok0A	1ka8A	1xzqB	1prtF	1prtD	1oczE
2r41D	1xpxA	1jc7A	2bseA	2fioB	1h4lD	1cf7B	2o26A	1hl6D	2d5rB
2j4bB	1jsgA	2hdiB	2bhmA	2pt7G	2yx8A	1cqaA	2gjvA	2e79A	1qynD
2gwfB	1smpI	1e2aA	2qvtA	2g38C	1qc7A	1b4bA	1v1jA	2bl7A	1im3D
1bg8A	2qyaA	1ipiA	1k3bA	2rcfB	2fu2A	1j1vA	1xeqA	2qm2B	3bjoA
2h5nD	1pzwA	1r7lB	1y96B	2czvD	2b91A	1lddA	2gy9K	2gybT	2gy9Q
1nzeA	1y71B	2o4tA	2v94B	1stmA	1r75A	1mwpA	1v54H	2igpA	2fygA
2oq8A	1kvdB	1kp6A	1r77B	2j97A	1kptA	1tsfA	2e12A	2f5gA	2hl7A
1t1jB	2oqbB	1yo3A	1dj7A	1vccA	2d68B	1zkeA	2ciuA	3bs1A	1g8qA
1rkiA	2aydA	1t92A	2q3tA	2czsA	1jr8A	1lmiA	1dp7P	1vyiA	1wpuA
1xawA	1bgfA	2hewF	2ohwA	1n13B	1mc2A	2b97A	2vb1A	1jo0A	1wckA
1ijyA	1vmhA	2qt7A	2j6bA	2qsbA	1nqjB	2bt9A	2h8eA	1vr7A	2fhzA
2fhzB	1sauA	1uz3A	2axwA	1lwbA	1i27A	2aibA	1c5eA		

Table S3. Testing set of 113 “easy” proteins. Each protein name is represented by 4-letter PDB ID and 1-letter chain ID.

1b0xA	3f75P	1gyyA	1sfuA	2vb2X	1obxA	1scjB	2pneA	2p2rA	2q5wD
3cjsB	1nlwB	2eboA	2utgA	3bs9A	3b5nC	1dokA	2z4pA	1iqzA	2ec9T
1l0hA	2qbvA	1wm2A	1jq0A	1h75A	2fq3A	1ucrB	2guzA	2pmrA	1ljoA
1n1jB	1l9lA	3bs7A	2oo2A	1wz3A	2o30A	1v5iB	1xmkA	3bktA	2o2vB
1fipA	1c75A	2evbA	1eptB	3b7hA	1pufA	3htsB	1dgwX	3e19C	1ybzA
2uwjE	2hueB	2bkfA	1uw4A	1d3bE	1bt0A	1tafB	2z3rB	1zpvC	1pk3C
1guaB	2ek1D	3bs5B	2bv2A	1f60B	1vmgA	1g33A	1wlzA	1nfjA	1krnA
2zeqA	2qvoA	2vc8A	1r8oB	1na3A	2fazA	1wmhA	1y96A	1c53A	2nzcB
3f6qB	2hzcA	2ckxA	2vs0B	1x6jA	3cecA	2ieqB	1i81A	2o49A	2o1rA
2icpA	3cu4A	1a68A	2ca3B	2zboA	2eaqA	1ynrA	1kwaA	3eusA	2i6vA
3d1mD	1cxyA	1luzA	2obpA	3bqqA	1wveC	2r2zA	2pliA	2bf3A	3cjkB
3fojA	1oeyA	2h61B							

Table S4. Comparison of structural relaxation results by different versions of Rosetta programs. Each bold number is the best performance at each category.

		RMSD	TM-score	GDT-TS	GDT-HA	GDT-SC	HMA	HMC	HSA	HSC
Rosetta decoys	Rosetta2.3.0	10.48Å	**0.386**	40.60	27.54	10.25	0.585	**0.523**	**0.171**	**0.087**
	Rosetta3.2.1a	**10.42Å**	**0.386**	40.63	27.54	10.14	0.590	0.518	0.137	0.082
	Rosetta3.2.1b	10.44Å	**0.386**	40.63	27.60	10.18	0.594	0.520	0.141	0.079
	Rosetta3.2.1c	10.47Å	**0.386**	**40.64**	**27.73**	**10.36**	**0.595**	0.516	0.144	0.081
I-Tasser decoys	Rosetta2.3.0	7.85Å	**0.526**	**53.44**	**38.06**	**17.99**	0.670	**0.546**	**0.205**	**0.099**
	Rosetta3.2.1a	**7.75Å**	0.523	53.16	37.65	17.74	0.668	0.539	0.151	0.090
	Rosetta3.2.1b	7.80Å	0.525	53.37	38.02	17.73	**0.680**	0.542	0.152	0.082
	Rosetta3.2.1c	7.91Å	0.524	53.22	37.89	17.80	0.678	0.540	0.159	0.087

Table S5. Command lines for the four different Rosetta relaxation programs.

Rosetta programs	Command lines
Rosetta2.3.0	./rosetta.release aa 1xxx A -relax -farlx -s initfull -fa_input -fa_output -ex1 -ex2
Rosetta3.2.1a	./relax.linuxgccrelease -database rosetta_database -in:file:s initfull.pdb -in:file:fullatom -relax:thorough
Rosetta3.2.1b	./relax.linuxgccrelease -database rosetta_database -in:file:s initfull.pdb -in:file:fullatom -relax:fast
Rosetta3.2.1c	./relax.linuxgccrelease -database rosetta_database -in:file:s initfull.pdb -in:file:fullatom -relax:quick

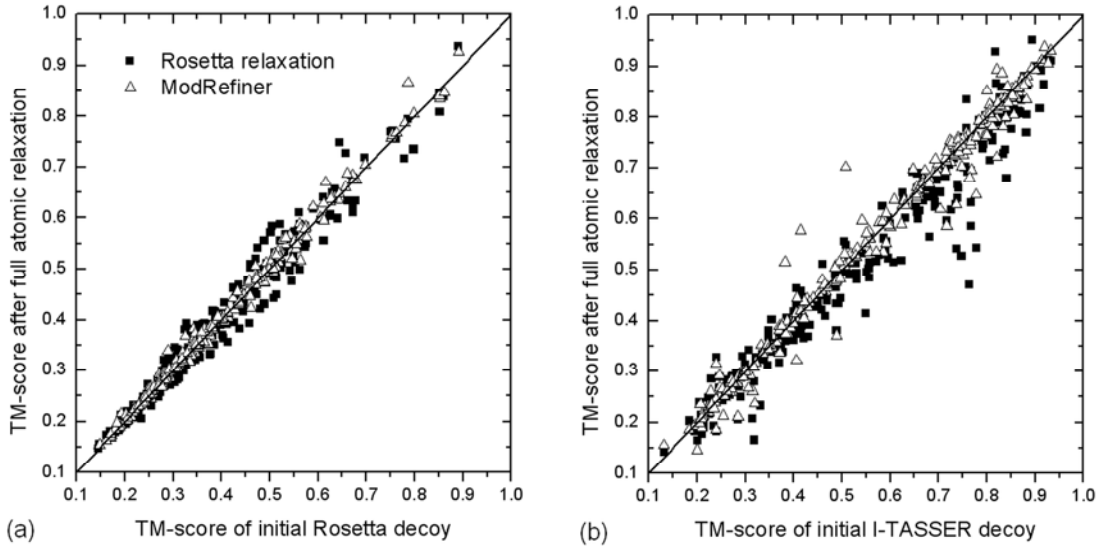

Figure S1. TM-score comparison of ModRefiner and Rosetta relaxation models with initial models generated by (a) Rosetta and (b) I-TASSER.