Supporting Information

Table of Content

Supporting Tables

Table S1 Mean and standard deviation of four H-bond features in regular secondary structure elements as defined in Figure S5.

Secondary structure for residue pair (*i, j*)	Features (mean/standard deviation)			
	$D(O_i,H_j)$	$A(C_i,O_i,H_j)$	$A(O_i,H_j,N_j)$	$T(C_i,O_i,H_j,N_j)$
Helix, *j=i+4*	2.00/0.53	147/10.58	159/11.25	160/25.36
Helix, *j=i+3*	2.85/0.32	89/7.70	111/8.98	-160/7.93
Parallel strands	2.00/0.30	155/11.77	164/11.29	180/68.96
Antiparallel strands	2.00/0.26	151/12.38	163/11.02	-168/69.17

Table S2. Summary of modeling results by I-TASSER structure prediction and the follow-up EM density map refinement methods on 5 proteins with experimentally determined experiment maps.

Protein Name	Methods	Resolution (Å)	Predicted structure		Refined structure	
			TM-score	RMSD	TM-score	RMSD
EMD20239_dom1	Flex-EM	3.1	0.734	14.09 Å	0.735	14.01 Å
	Rosetta				0.815	13.94 Å
	EM-Refiner				0.818	13.92 Å
EMD21923_dom2	Flex-EM	3.3	0.756	9.60 Å	0.761	9.58 Å
	Rosetta				0.789	10.95 Å
	EM-Refiner				0.802	9.53 Å
EMD21040_dom1	Flex-EM	3.8	0.719	5.35 Å	0.726	5.31 Å
	Rosetta				0.855	4.09 Å
	EM-Refiner				0.896	3.09 Å
EMD6847_dom1	Flex-EM	4.4	0.811	7.43 Å	0.796	7.19 Å
	Rosetta				0.828	7.88 Å
	EM-Refiner				0.841	7.28 Å
EMD6708_dom2	Flex-EM	3.9	0.540	7.02 Å	0.538	7.09 Å
	Rosetta				0.713	5.91 Å
	EM-Refiner				0.716	5.10 Å

Table S3. Summary of the resolution of each single domain protein.

Protein name	Resolution(Å)	Protein name	Resolution(Å)	Protein name	Resolution(Å)
1efdN_dom1	7	2x8kC_dom1	10	3ub1D_dom2	8
1efdN_dom2	7	2x8kC_dom2	7	3uitD_dom1	7
1g87B_dom1	9	2yilA_dom1	9	3uitD_dom2	10
1g87B_dom2	6	2yilA_dom2	10	3uo3A_dom1	6
1gttA_dom1	9	3a1iA_dom1	7	3uo3A_dom2	8
1iwaA_dom1	9	3a1iA_dom2	10	3vr8B_dom1	6
1iwaA_dom2	7	3a45A_dom1	6	3vr8B_dom2	10
1mkmB_dom1	10	3a45A_dom2	8	3wkuA_dom1	8
1mkmB_dom2	7	3ajeA_dom2	6	3wkuA_dom2	8
1pprM_dom1	9	3aqkA_dom1	7	3zvmA_dom1	7
1prrA_dom1	5	3aqkA_dom2	7	3zvmA_dom2	7
1q19A_dom1	6	3arbA_dom1	5	4ap5A_dom1	5
1q19A_dom2	6	3arbA_dom2	5	4ap5A_dom2	6
1qwrA_dom1	5	3b2zF_dom1	10	4axdA_dom1	7
1qwrA_dom2	6	3b2zF_dom2	7	4axdA_dom2	9
1rh1A_dom1	5	3b7wA_dom1	6	4bfiB_dom1	7
1rh1A_dom2	7	3b7wA_dom2	6	4bfiB_dom2	10
1rktA_dom1	9	3bt3A_dom1	6	4c9zA_dom2	8
1rktA_dom2	8	3bt3A_dom2	6	4cczA_dom1	5
1s6lA_dom1	9	3bu2A_dom2	6	4cczA_dom2	7
1s6lA_dom2	7	3c4tA_dom1	9	4d0nB_dom1	8
1st0A_dom2	9	3c4tA_dom2	7	4d0nB_dom2	8
1upsA_dom1	7	3craA_dom1	7	4dj3A_dom1	8
1v8cA_dom2	7	3craA_dom2	10	4dj3A_dom2	7
1w3aA_dom1	9	3d30A_dom1	6	4dqaA_dom1	5
1w3aA_dom2	8	3d30A_dom2	9	4dqaA_dom2	6
1wfxA_dom1	8	3e00D_dom1	6	4eo3A_dom1	6
1wv3A_dom1	5	3errA_dom1	7	4eo3A_dom2	6
1wv3A_dom2	5	3errA_dom2	8	4etxA_dom1	5
1x7pA_dom1	7	3g79A_dom1	5	4etxA_dom2	10
1x7pA_dom2	9	3g79A_dom2	9	4fkcA_dom1	8
1yiqA_dom1	10	3hcsA_dom1	5	4fkcA_dom2	5
1yiqA_dom2	5	3hcsA_dom2	7	4fxkC_dom1	8
1zbuB_dom1	9	3hyiA_dom1	10	4fxkC_dom2	8
1zbuB_dom2	9	3hyiA_dom2	10	4g86A_dom1	9
1ze1A_dom1	5	3i2dA_dom1	9	4g86A_dom2	7
1ze1A_dom2	10	3i2dA_dom2	10	4gbyA_dom1	10
2ablA_dom1	8	3iam2_dom1	7	4gbyA_dom2	9
2ablA_dom2	10	3iam2_dom2	8	4ggmX_dom1	7
2ahvA_dom1	7	3ifrA_dom1	5	4ggmX_dom2	6
2ahvA_dom2	9	3ifrA_dom2	9	4gslA_dom1	5
2bkpA_dom1	8	3isqA_dom1	8	4gslA_dom2	10

2bkpA_dom2	7	3isqA_dom2	10	4gyjA_dom1	7
2c1lA_dom1	10	3j7aK_dom1	6	4gyjA_dom2	6
2c1yA_dom1	8	3j7aK_dom2	10	4h3tA_dom1	9
2c1yA_dom2	9	3k1rA_dom1	10	4h3tA_dom2	10
2cxcA_dom1	10	3k1rA_dom2	8	4hmoA_dom1	9
2cxcA_dom2	10	3kh5A_dom1	8	4hmoA_dom2	6
2d1cA_dom1	5	3kh5A_dom2	10	4hsaC_dom1	7
2d7iA_dom1	6	3kjpA_dom1	9	4hwiB_dom2	10
2d7iA_dom2	5	3kjpA_dom2	10	4l5gA_dom1	10
2e9xB_dom1	8	3kt1A_dom1	7	4l5gA_dom2	10
2e9xB_dom2	6	3kt1A_dom2	10	4lpqA_dom1	8
2evrA_dom1	9	3kzwA_dom1	9	4lpqA_dom2	9
2evrA_dom2	5	3kzwA_dom2	8	4mspA_dom2	7
2ew9A_dom1	8	3lsgA_dom1	8	4n06B_dom1	8
2fd5A_dom1	7	3lsgA_dom2	9	4n06B_dom2	8
2fd5A_dom2	8	3me4A_dom1	9	4nj5A_dom1	10
2gh8A_dom1	9	3me4A_dom2	7	4nj5A_dom2	10
2gh8A_dom2	8	3ml4C_dom1	6	4up9A_dom1	5
2gt1A_dom1	10	3ml4C_dom2	10	4up9A_dom2	7
2gt1A_dom2	8	3mzfA_dom1	6	4w7sA_dom1	10
2gzaC_dom1	5	3mzfA_dom2	10	4w7sA_dom2	7
2gzaC_dom2	6	3njaB_dom1	7	T0863_dom1	8
2hyxA_dom1	7	3njaB_dom2	7	T0880_dom1	5
2ijd1_dom1	5	3nt8A_dom1	7	T0880_dom2	6
2ijd1_dom2	8	3nt8A_dom2	8	T0886_dom1	5
2iw2A_dom1	9	3o2gA_dom2	10	T0886_dom2	8
2iw2A_dom2	9	3og5A_dom1	5	T0890_dom1	6
2jz4A_dom2	6	3og5A_dom2	6	T0890_dom2	6
2mbgA_dom1	8	3oh0A_dom1	5	T0892_dom1	8
2mbgA_dom2	9	3oh0A_dom2	8	T0892_dom2	5
2nsfA_dom1	7	3qavA_dom1	10	T0893_dom1	7
2nsfA_dom2	7	3qavA_dom2	5	T0893_dom2	10
2nykA_dom1	7	3qf4B_dom2	5	T0894_dom1	10
2nykA_dom2	7	3qjjA_dom1	10	T0894_dom2	5
2o6yA_dom1	10	3qjjA_dom2	8	T0896_dom1	7
2o6yA_dom2	10	3qtdA_dom1	10	T0896_dom2	10
2outA_dom1	6	3qtdA_dom2	7	T0897_dom1	5
2owbA_dom1	10	3r6bA_dom1	10	T0897_dom2	8
2owbA_dom2	6	3rh7A_dom1	6	T0898_dom1	8
2qp2A_dom1	6	3rh7A_dom2	8	T0898_dom2	10
2qp2A_dom2	10	3sb4A_dom1	8	T0912_dom1	9
2qygA_dom1	10	3sb4A_dom2	5	T0914_dom1	5
2qygA_dom2	7	3ssoB_dom2	8	T0914_dom2	7
2r5wB_dom1	6	3swjA_dom1	7	T0918_dom1	9

2r5wB_dom2	8	3swjA_dom2	7	T0918_dom2	8
2r8vA_dom2	7	3t7jA_dom1	9	T0920_dom1	8
2v42A_dom1	9	3t7jA_dom2	6	T0920_dom2	9
2w2gA_dom1	8	3u07C_dom1	8	T0942_dom1	6
2w5fB_dom1	5	3u07C_dom2	10		
2wkxA_dom2	8	3u0oB_dom1	8		
2x7iA_dom1	5	3u0oB_dom2	9		
2x7iA_dom2	9	3ub1D_dom1	9		

Supporting Figures

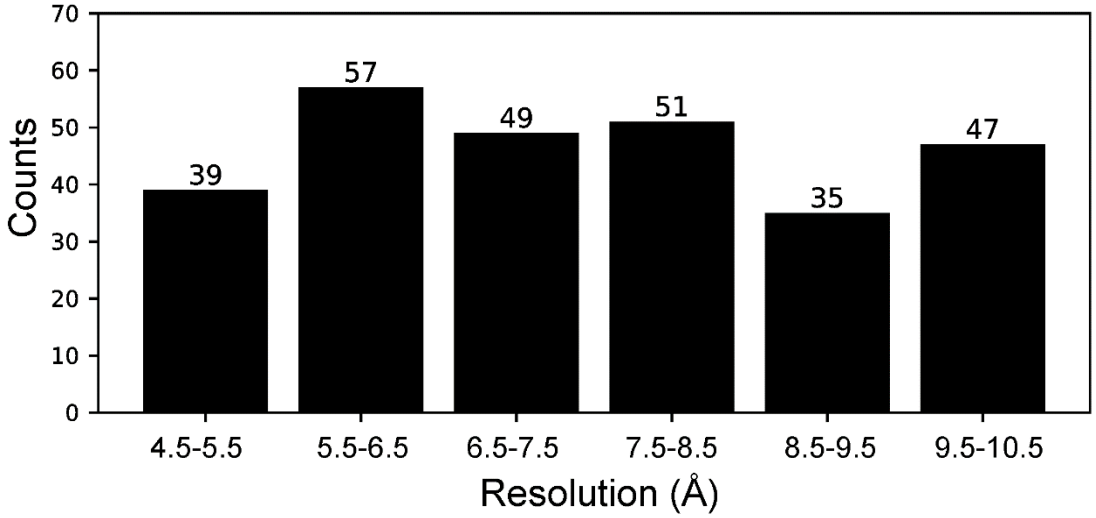

Figure S1. Histogram distribution of the simulated density map resolutions for the 278 benchmark test proteins.

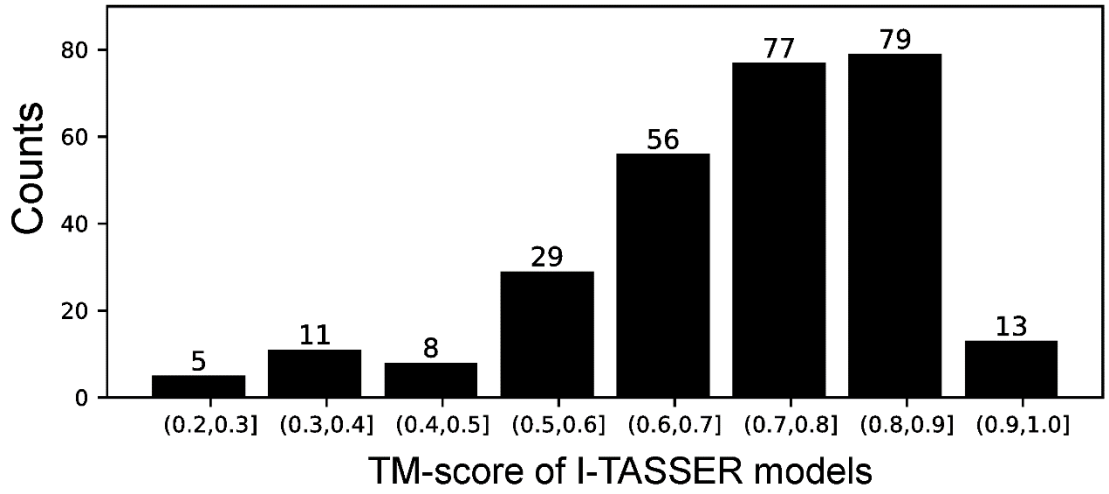

Figure S2. Histogram distribution of the TM-scores of the I-TASSER models for the 278 benchmark test proteins.

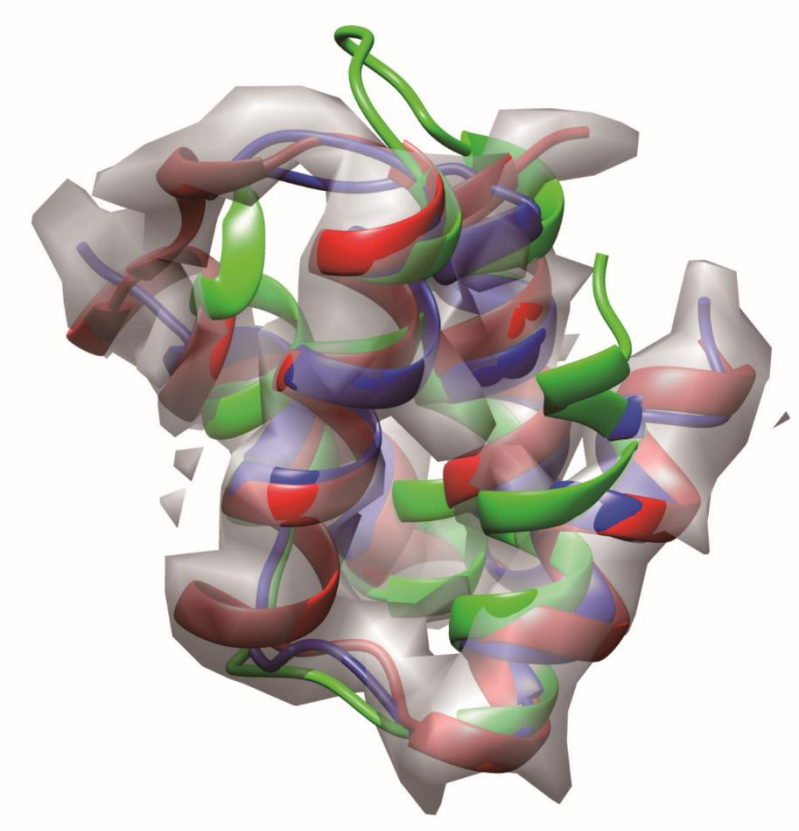

Figure S3. An illustrative example of refinement using simulated experimental data for a single domain of 5NV4. Compared with the native model (red), the EM-Refiner model (blue) had a TM-score=0.53 and RMSD=11.522, while the model from I-TASSER (green) had a TM-score=0.37 and an RMSD=12.06.

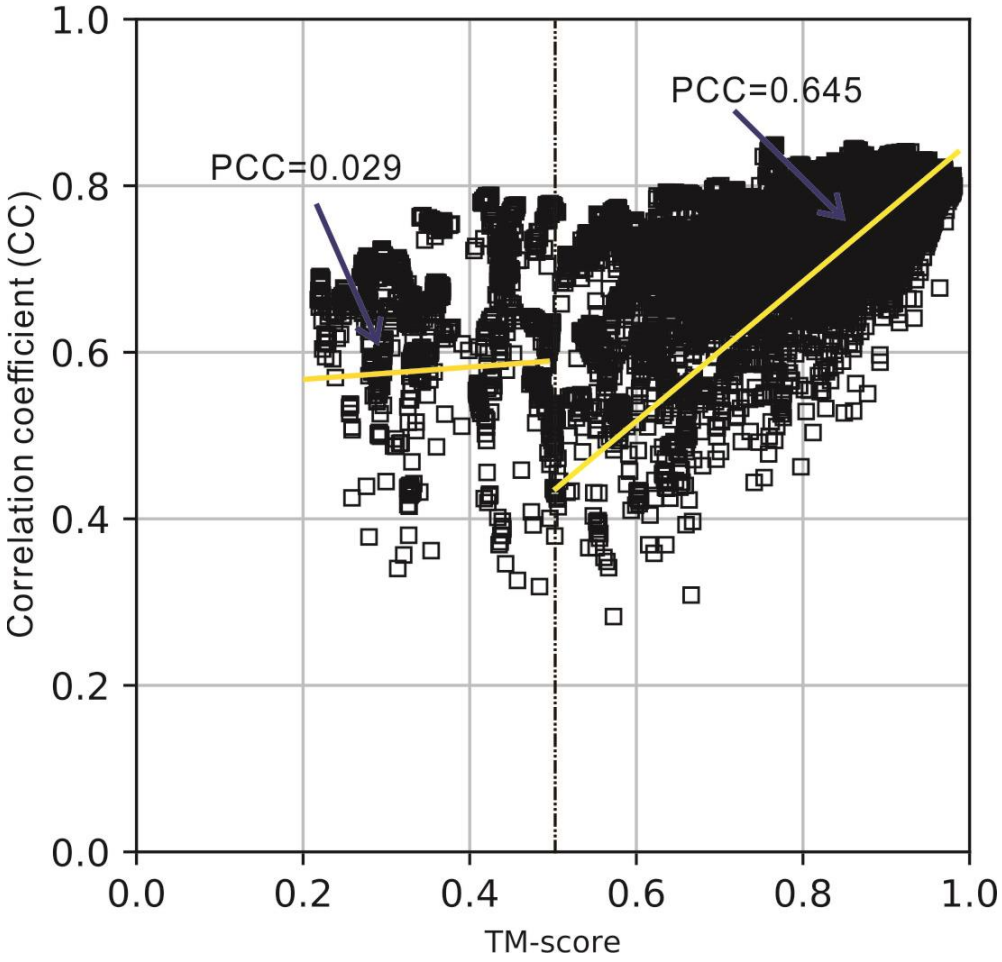

Figure S4. Correlation coefficient (CC) to the EM density map versus TM-score of the structural decoy conformations generated by EM-Refiner simulations for 278 test proteins. The dashed-dotted line divides the decoys into two regions (TM-score below and above 0.5), where the yellow lines represent the fitting results by linear regression for the samples in the two regions. Pearson correlation coefficients between CC and TM-score are 0.029 and 0.645, respectively, in the two regions.

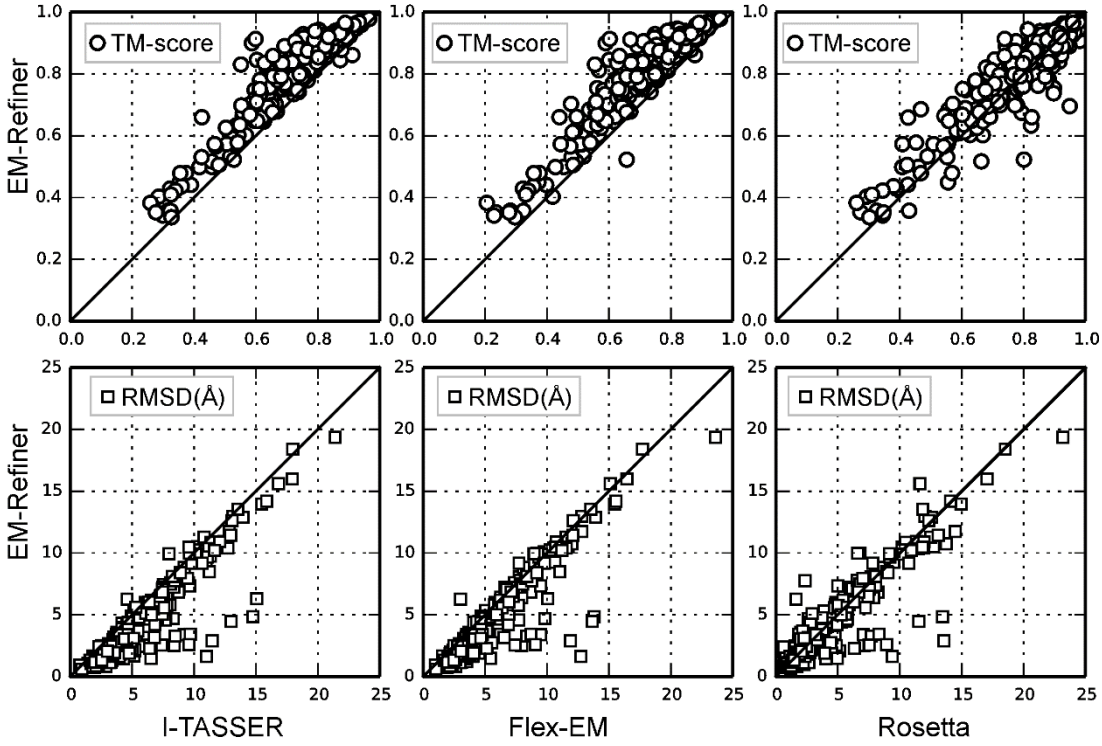

Figure S5. Head-to-head comparison of the TM-score (upper row) and RMSD (lower row) of the EM-Refiner models versus the initial predicted models and the two control programs. The model refinement started from the optimal model superposition created by the TM-score program that directly overlays the initial model with the target structure.

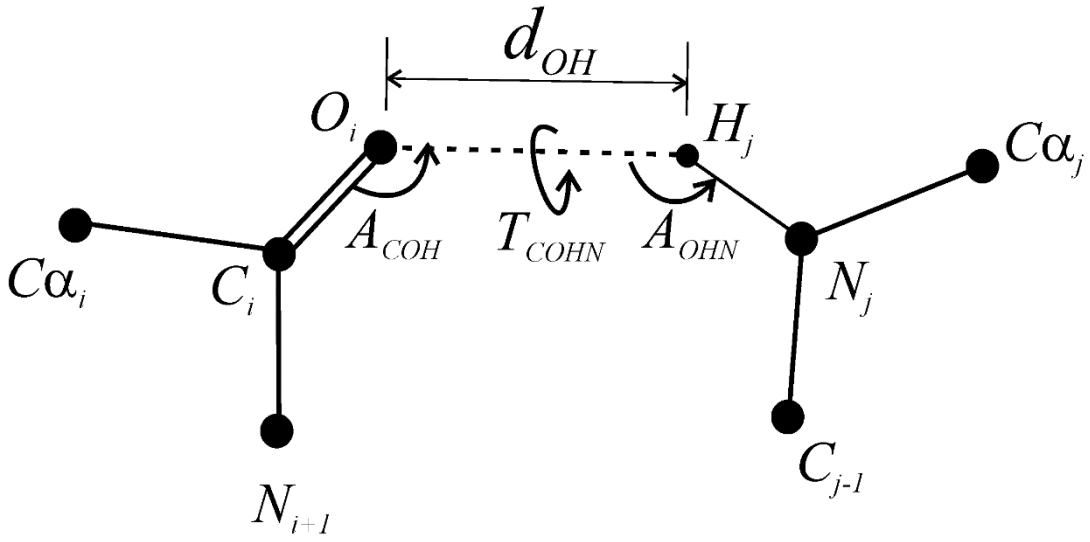

Figure S6. Illustration of four distance and angle features used by EM-Refiner to calculate the H-bonding potential.

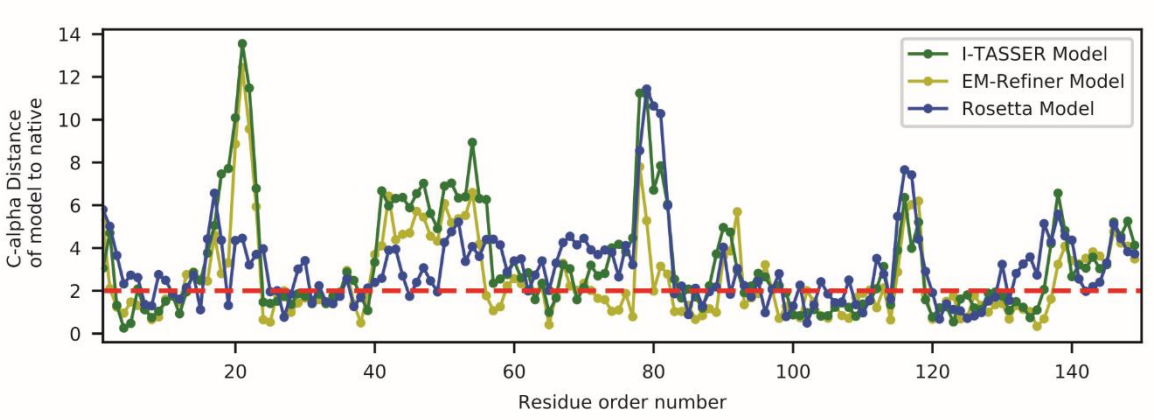

Figure S7. Residue-level distance of C-alpha atom pairs between model and native after TM-score superposition. Different curves refer to the models generated by I-TASSER, Rosetta and EM-Refiner respectively.