

Fitting Low-resolution Protein Structures into Cryo-EM Density Maps by Multi-objective Optimization of Global and Local Correlations

Biao Zhang^{1,2}, Wenyi Zhang³, Robin Pearce², Yang Zhang^{*,2}, Hong-Bin Shen^{*,1}

¹Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University,
and Key Laboratory of System Control and Information Processing, Ministry of
Education of China, Shanghai, China

²Department of Computational Medicine and Bioinformatics,
University of Michigan, Ann Arbor, MI 48109 USA

³School of life Sciences, Westlake University,
Hangzhou, Zhejiang, 310024, China

Supporting Information

Table of Content

Supporting Tables

- **Table S1.** Summary of the resolutions for the 292 test proteins.
- **Table S2.** Summary of modeling results by I-TASSER structure prediction and the follow-up EM density map fitting methods on 7 proteins with experimentally determined density maps.
- **Table S3.** Summary of comparison results between each single objective fitting method and MOFIT on 292 test proteins.

Supporting Figures

- **Figure S1.** Histogram distribution of the simulated density map resolutions for the 292 benchmark test proteins.
- **Figure S2.** Performance of fitting native structures into density maps. The x-axis represents the range of cRMSD values, while the y-axis represents the corresponding counts numbers.
- **Figure S3.** The correlation coefficients (*CC*) to the EM density map versus TM-score of the initial models for the 292 proteins after fitting. The dashed-dotted line divides the models into two regions with TM-scores below and above 0.5, where the solid lines represent the fitting results by linear regression for the samples in the two regions. The Pearson correlation

coefficients between *CC* and TM-score were 0.18 and 0.74, respectively, for the two regions.

- **Figure S4.** An illustrative example to show the energy distribution of the nondominated particles of the Pareto set in energy space for target T0880.
- **Figure S5.** Illustrative head-to-head comparison between each single objective function. Each block represents the cRMSD value on the corresponding axis.
- **Figure S6.** Objective function values for one case (1efdN) during the PSO searching iterations: the E1, E2 and the E3 values at each iteration are shown in (a), (b) and (c), respectively.
- **Figure S7.** Illustration of the correlation between cRMSD and TM-score of each fitted model at different density map resolutions. (a) the Pearson correlation coefficient between cRMSD and TM-score is -0.61 for 5-6 Å maps, (b) -0.52 for 7-8 Å maps, and (c) -0.50 for 9-10 Å maps.

Supporting Tables

Table S1. Summary of the resolutions for the 292 test proteins.

Protein name	Resolution(Å)	Protein name	Resolution(Å)	Protein name	Resolution(Å)
1efdN_dom1	7	3aliA_dom1	7	3wkuA_dom2	8
1efdN_dom2	7	3aliA_dom2	10	3zvmA_dom1	7
1g87B_dom1	9	3a45A_dom1	6	3zvmA_dom2	7
1g87B_dom2	6	3a45A_dom2	8	4ap5A_dom1	5
1gttA_dom1	9	3ajeA_dom2	6	4ap5A_dom2	6
1iwaA_dom1	9	3aqkA_dom1	7	4axdA_dom1	7
1iwaA_dom2	7	3aqkA_dom2	7	4axdA_dom2	9
1mkmB_dom1	10	3arbA_dom1	5	4bfiB_dom1	7
1mkmB_dom2	7	3arbA_dom2	5	4bfiB_dom2	10
1pprM_dom1	9	3b2zF_dom1	10	4c9zA_dom2	8
1prrA_dom1	5	3b2zF_dom2	7	4cczA_dom1	5
1q19A_dom1	6	3b7wA_dom1	6	4cczA_dom2	7
1q19A_dom2	6	3b7wA_dom2	6	4d0nB_dom1	8
1qwrA_dom1	5	3bt3A_dom1	6	4d0nB_dom2	8
1qwrA_dom2	6	3bt3A_dom2	6	4dj3A_dom1	8
1rh1A_dom1	5	3bu2A_dom2	6	4dj3A_dom2	7
1rh1A_dom2	7	3c4tA_dom1	9	4dqaA_dom1	5
1rktA_dom1	9	3c4tA_dom2	7	4dqaA_dom2	6
1rktA_dom2	8	3craA_dom1	7	4eo3A_dom1	6
1s6lA_dom1	9	3craA_dom2	10	4eo3A_dom2	6
1s6lA_dom2	7	3d30A_dom1	6	4etxA_dom1	5
1st0A_dom2	9	3d30A_dom2	9	4etxA_dom2	10
1upsA_dom1	7	3e00D_dom1	6	4fkcA_dom1	8
1v8cA_dom2	7	3errA_dom1	7	4fkcA_dom2	5
1w3aA_dom1	9	3errA_dom2	8	4fxkC_dom1	8
1w3aA_dom2	8	3g79A_dom1	5	4fxkC_dom2	8
1wfxA_dom1	8	3g79A_dom2	9	4g86A_dom1	9
1wv3A_dom1	5	3hcsA_dom1	5	4g86A_dom2	7
1wv3A_dom2	5	3hcsA_dom2	7	4gbyA_dom1	10
1x7pA_dom1	7	3hyiA_dom1	10	4gbyA_dom2	9
1x7pA_dom2	9	3hyiA_dom2	10	4ggmX_dom1	7
1yiqA_dom1	10	3i2dA_dom1	9	4ggmX_dom2	6
1yiqA_dom2	5	3i2dA_dom2	10	4gslA_dom1	5
1zbuB_dom1	9	3iam2_dom1	7	4gslA_dom2	10
1zbuB_dom2	9	3iam2_dom2	8	4gyjA_dom1	7
1ze1A_dom1	5	3ifrA_dom1	5	4gyjA_dom2	6
1ze1A_dom2	10	3ifrA_dom2	9	4h3tA_dom1	9
2ablA_dom1	8	3isqA_dom1	8	4h3tA_dom2	10

2ablA_dom2	10	3isqA_dom2	10	4hmoA_dom1	9
2ahvA_dom1	7	3j7aK_dom1	6	4hmoA_dom2	6
2ahvA_dom2	9	3j7aK_dom2	10	4hsaC_dom1	7
2bkpA_dom1	8	3k1rA_dom1	10	4hwiB_dom2	10
2bkpA_dom2	7	3k1rA_dom2	8	4l5gA_dom1	10
2c1lA_dom1	10	3kh5A_dom1	8	4l5gA_dom2	10
2c1yA_dom1	8	3kh5A_dom2	10	4lpqA_dom1	8
2c1yA_dom2	9	3kjpA_dom1	9	4lpqA_dom2	9
2cxcA_dom1	10	3kjpA_dom2	10	4mspA_dom2	7
2cxcA_dom2	10	3kt1A_dom1	7	4n06B_dom1	8
2d1cA_dom1	5	3kt1A_dom2	10	4n06B_dom2	8
2d7iA_dom1	6	3kzwA_dom1	9	4nj5A_dom1	10
2d7iA_dom2	5	3kzwA_dom2	8	4nj5A_dom2	10
2e9xB_dom1	8	3lsgA_dom1	8	4up9A_dom1	5
2e9xB_dom2	6	3lsgA_dom2	9	4up9A_dom2	7
2evrA_dom1	9	3me4A_dom1	9	4w7sA_dom1	10
2evrA_dom2	5	3me4A_dom2	7	4w7sA_dom2	7
2ew9A_dom1	8	3ml4C_dom1	6	T0863_dom1	8
2fd5A_dom1	7	3ml4C_dom2	10	T0880_dom1	5
2fd5A_dom2	8	3mzfA_dom1	6	T0880_dom2	6
2gh8A_dom1	9	3mzfA_dom2	10	T0886_dom1	5
2gh8A_dom2	8	3njaB_dom1	7	T0886_dom2	8
2gt1A_dom1	10	3njaB_dom2	7	T0890_dom1	6
2gt1A_dom2	8	3nt8A_dom1	7	T0890_dom2	6
2gzaC_dom1	5	3nt8A_dom2	8	T0892_dom1	8
2gzaC_dom2	6	3o2gA_dom2	10	T0892_dom2	5
2hyxA_dom1	7	3og5A_dom1	5	T0893_dom1	7
2ijd1_dom1	5	3og5A_dom2	6	T0893_dom2	10
2ijd1_dom2	8	3oh0A_dom1	5	T0894_dom1	10
2iw2A_dom1	9	3oh0A_dom2	8	T0894_dom2	5
2iw2A_dom2	9	3qavA_dom1	10	T0896_dom1	7
2jz4A_dom2	6	3qavA_dom2	5	T0896_dom2	10
2mbgA_dom1	8	3qf4B_dom2	5	T0897_dom1	5
2mbgA_dom2	9	3qjjA_dom1	10	T0897_dom2	8
2nsfA_dom1	7	3qjjA_dom2	8	T0898_dom1	8
2nsfA_dom2	7	3qtdA_dom1	10	T0898_dom2	10
2nykA_dom1	7	3qtdA_dom2	7	T0912_dom1	9
2nykA_dom2	7	3r6bA_dom1	10	T0914_dom1	5
2o6yA_dom1	10	3rh7A_dom1	6	T0914_dom2	7
2o6yA_dom2	10	3rh7A_dom2	8	T0918_dom1	9
2outA_dom1	6	3sb4A_dom1	8	T0918_dom2	8
2owbA_dom1	10	3sb4A_dom2	5	T0920_dom1	8
2owbA_dom2	6	3ssoB_dom2	8	T0920_dom2	9

2qp2A_dom1	6	3swjA_dom1	7	T0942_dom1	6
2qp2A_dom2	10	3swjA_dom2	7	T1008	5
2qygA_dom1	10	3t7jA_dom1	9	T0998	7
2qygA_dom2	7	3t7jA_dom2	6	T0991	8
2r5wB_dom1	6	3u07C_dom1	8	T0980s1	6
2r5wB_dom2	8	3u07C_dom2	10	T0953s1	7
2r8vA_dom2	7	3u0oB_dom1	8	T0953s2	9
2v42A_dom1	9	3u0oB_dom2	9	T0986s2	5
2w2gA_dom1	8	3ub1D_dom1	9	T1015s1	8
2w5fB_dom1	5	3ub1D_dom2	8	T1015s2	9
2wkxA_dom2	8	3uitD_dom1	7	T0975	6
2x7iA_dom1	5	3uitD_dom2	10	T0968s2	8
2x7iA_dom2	9	3uo3A_dom1	6	T1019S1	5
2x8kC_dom1	10	3uo3A_dom2	8	T0955	9
2x8kC_dom2	7	3vr8B_dom1	6	T0992	7
2yilA_dom1	9	3vr8B_dom2	10		
2yilA_dom2	10	3wkuA_dom1	8		

Note: Txxx represents the target from CASP experiment case.

Table S2. Summary of modeling results by I-TASSER structure prediction and the follow-up EM density map fitting methods on 7 proteins with experimentally determined density maps.

Protein Name	Methods	Resolution (Å)	Predicted structure		Fitted structure
			TM-score	RMSD	cRMSD
EMD20239_dom1	ADP_EM	3.1	0.734	14.09 Å	1.57 Å
	Situs				1.81 Å
	PowerFit				2.74 Å
	MOFIT				0.59 Å
EMD21923_dom2	ADP_EM	3.3	0.756	9.60 Å	1.93 Å
	Situs				1.92 Å
	PowerFit				3.78 Å
	MOFIT				1.82 Å
EMD21040_dom1	ADP_EM	3.8	0.719	5.35 Å	6.46 Å
	Situs				2.84 Å
	PowerFit				4.96 Å
	MOFIT				2.58 Å
EMD6847_dom1	ADP_EM	4.4	0.811	7.43 Å	0.96 Å
	Situs				0.86 Å
	PowerFit				1.36 Å
	MOFIT				0.86 Å
EMD6708_dom2	ADP_EM	3.9	0.540	7.02 Å	13.62 Å
	Situs				18.43 Å
	PowerFit				11.46 Å
	MOFIT				2.20 Å
EMD7454_dom2	ADP_EM	6.72	0.666	4.65 Å	1.60 Å
	Situs				1.74 Å
	PowerFit				1.88 Å
	MOFIT				0.86 Å
EMD2221_dom1	ADP_EM	8.4	0.709	4.04 Å	0.84 Å
	Situs				0.70 Å
	PowerFit				1.75 Å
	MOFIT				1.58 Å

Table S3. Summary of comparison results between each single objective fitting method and MOFIT on 292 test proteins.

Method	Average cRMSD	Median	P value
E1	4.03	1.29	1.65E-06
E2	5.25	1.85	6.36E-08
E3	6.41	1.88	6.44E-11
MOFIT	2.46	0.75	--

Supporting Figures

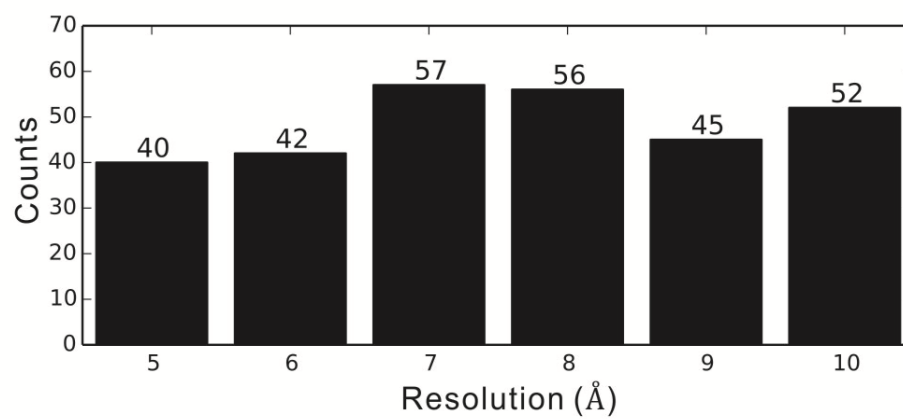


Figure S1. Histogram distribution of the simulated density map resolutions for the 292 benchmark test proteins.

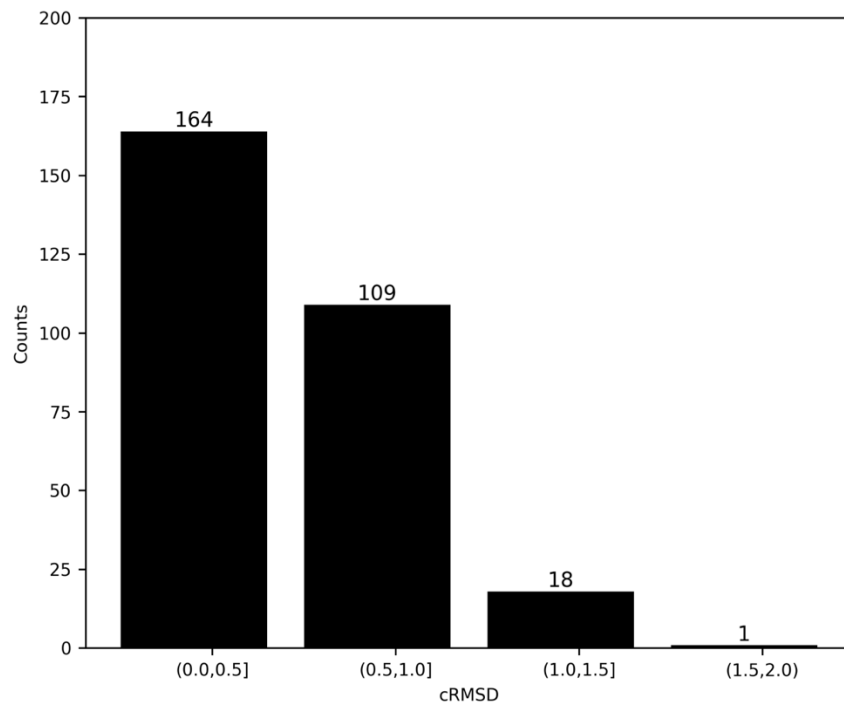


Figure S2. Performance of fitting native structures into density maps. The x-axis represents the range of cRMSD values, while the y-axis represents the corresponding counts numbers.

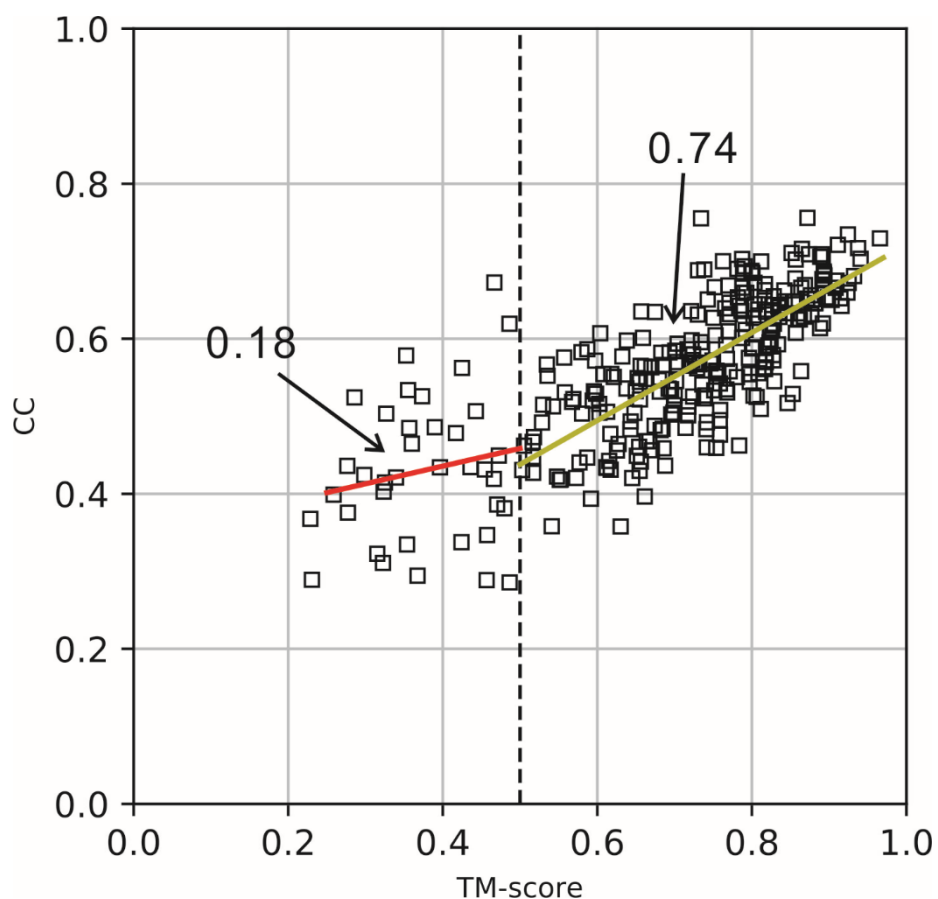


Figure S3. The correlation coefficients (CC) to the EM density map versus TM-score of the initial models for the 292 proteins after fitting. The dashed-dotted line divides the models into two regions with TM-scores below and above 0.5, where the solid lines represent the fitting results by linear regression for the samples in the two regions. The Pearson correlation coefficients between CC and TM-score were 0.18 and 0.74, respectively, for the two regions.

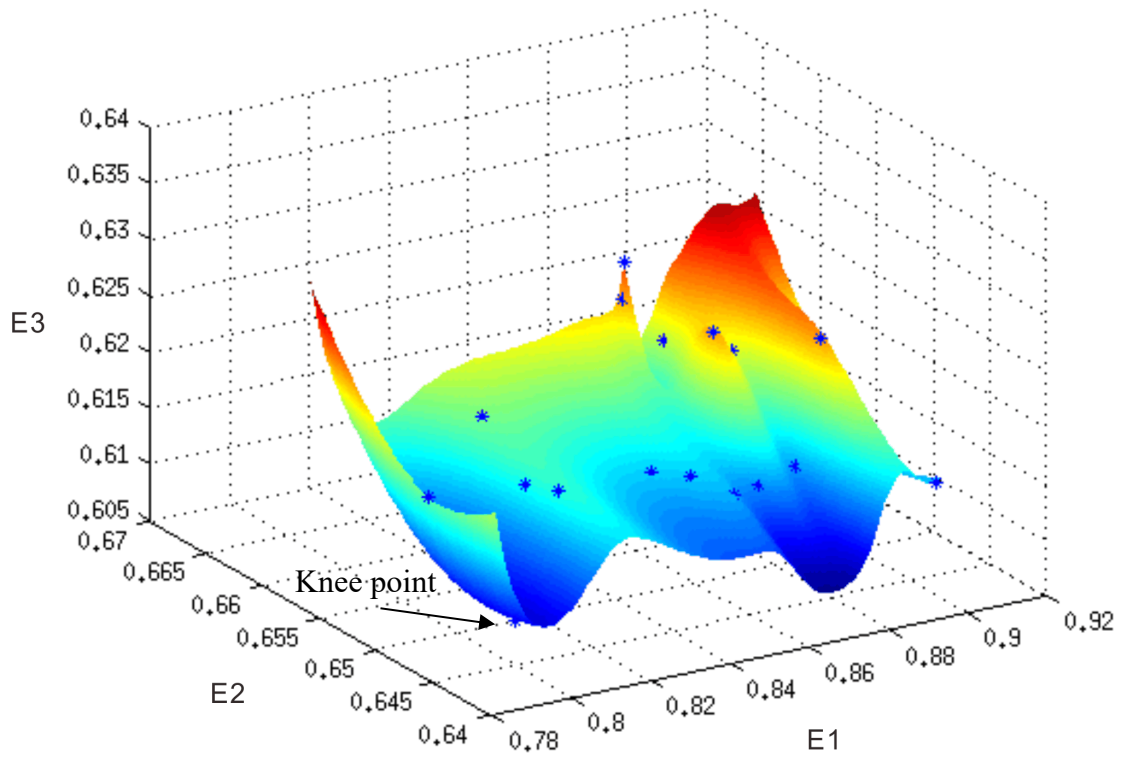


Figure S4. An illustrative example to show the energy distribution of the nondominated particles of the Pareto set in energy space for target T0880.

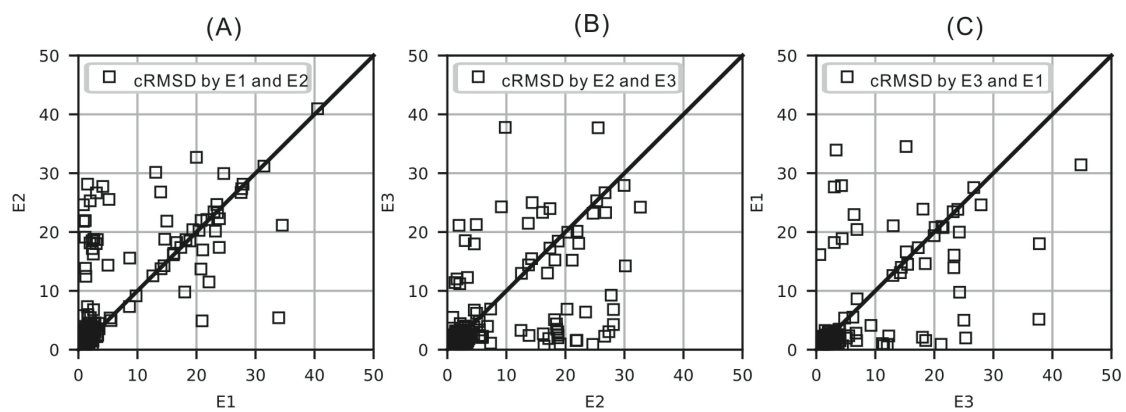


Figure S5. Illustrative head-to-head comparison between each single objective function. Each block represents the cRMSD value on the corresponding axis.

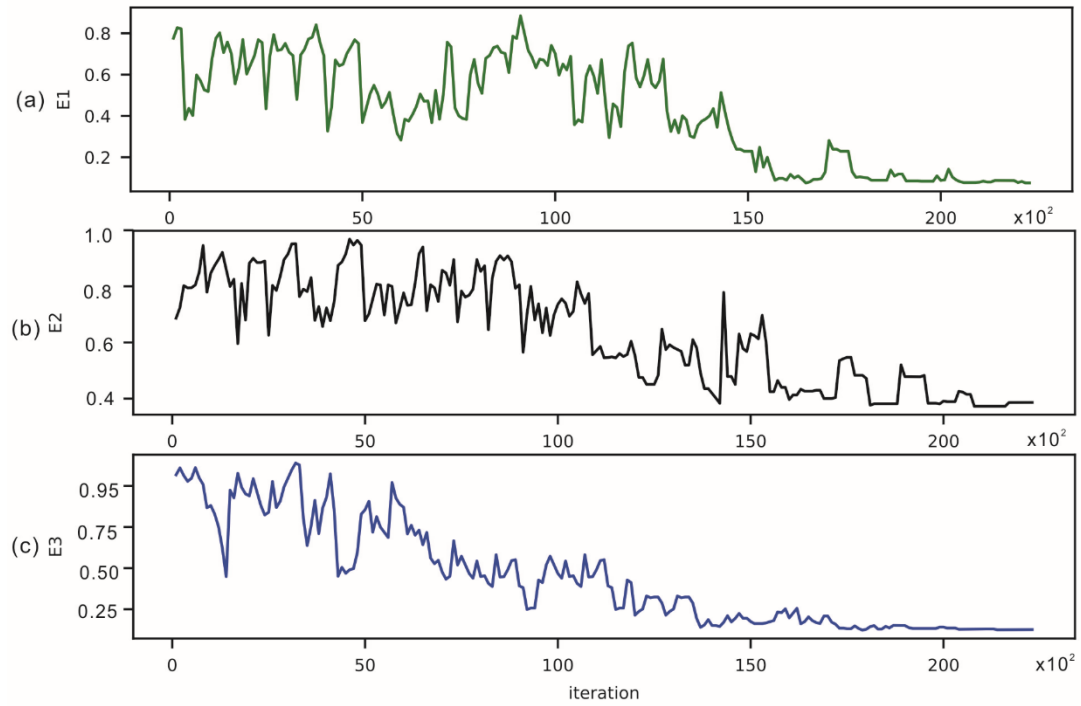


Figure S6. Objective function values for one case (1efdN) during the PSO searching iterations: the E_1 , E_2 and the E_3 values at each iteration are shown in (a), (b) and (c), respectively.

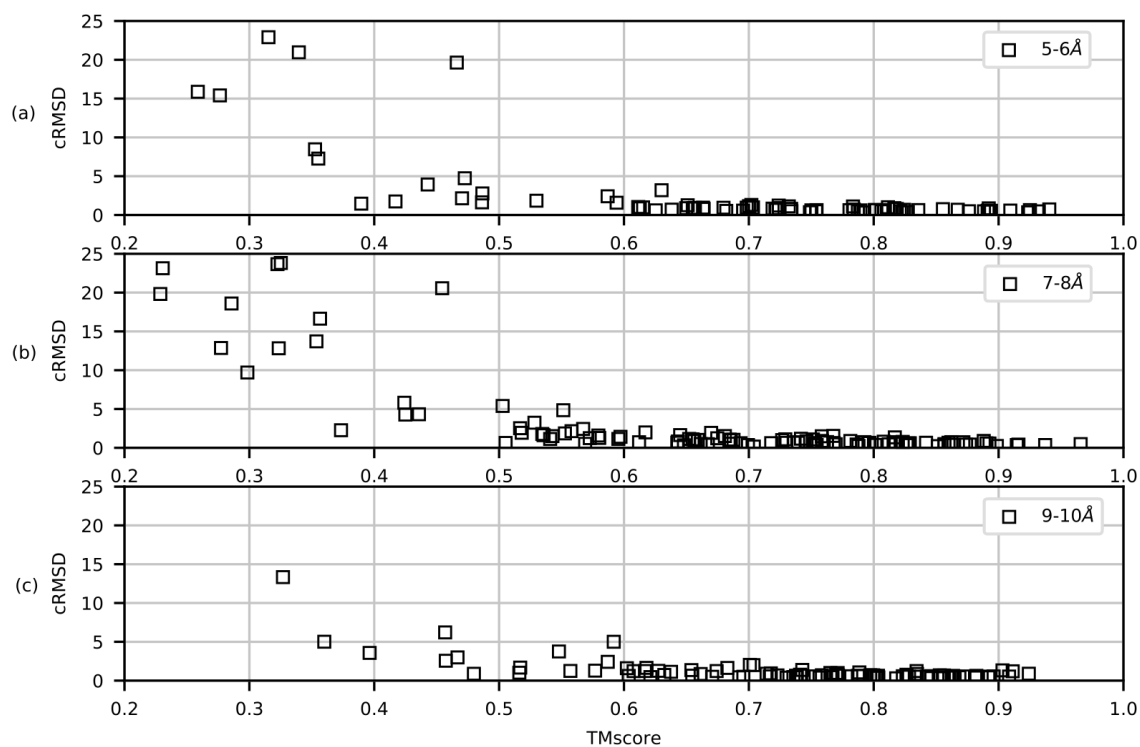


Figure S7. Illustration of the correlation between cRMSD and TM-score of each fitted model at different density map resolutions. (a) the Pearson correlation coefficient between cRMSD and TM-score is -0.61 for 5-6 Å maps, (b) -0.52 for 7-8 Å maps, and (c) -0.50 for 9-10 Å maps.