Supporting Information

Table of Content

Supporting Tables

Table S1. Summary of the resolution of each membrane protein.

Protein name	Resolution (Å)	Protein name	Resolution (Å)	Protein name	Resolution (Å)
1GZMA	7.13	2B6PC	9.91	2N02A	4.96
2VPZC	6.07	3FI1B	8.43	3RKON	6.16
3VR8D	8.24	4BUOA	4.04	4HUQS	7.6
4M58A	9.47	4QTNC	4.89	4WZ7B	6.6
5A63C	3.49	5EDLA	5.87	6EZNB	3.04
5L2AA	6.27	5NV9A	9.86	5V7PA	6.78
5XLSB	5.63	6C1RB	8.29	6F0KC	8.73
1KF6D	3.74	2BL2A	4.36	2N4XA	9.09
2W2EC	7.75	3GIAA	3.9	3S0XA	4.56
3VVKA	6.73	4C7RA	9.05	4HYGB	4.12
4MLBC	9.58	5EIKA	9.04	4WZ7D	5.13
5AX0A	3.55	5O0TA	8.55	5JKIA	8.56
5XPDA	9.41	6C263	7.8	5VBLB	9.46
1KPLB	9.22	2BS2F	5.74	6F0KF	7.62
2WSWC	8.53	3HD6E	5.85	2NQ2A	6.16
3WMMM	7.24	4CADC	3.3	3TDPA	8.14
4O6YB	8.35	6BQOB	9.23	4K1CA	6.78
5AZBA	4.63	5F15A	9.39	4WZ7E	7.7
5LNKA	9.6	5O5EA	7.7	5JNQB	5.89
5XPRA	8.12	6CAAA	9.48	5VKVA	3.92
1KQFF	4.18	2C3EA	4.21	6F36M	3.48
2WWAA	5.99	3M73A	8.81	2NR9A	5.39
3WO7A	4.55	4CZ9B	7.24	3UKMA	9.06
4OD5A	7.57	4RFSS	5.33	4KNFC	3.34
5B58B	6.4	5FN2B	7.25	4WZ7F	8.51
5LNKN	4.99	5OGEE	8.61	5JWYA	6.07
5XRAA	8.09	6CFWH	9.86	5VOUF	7.93
1LDFB	8.43	2CFPA	5.42	2ONKC	7.37
2X79A	3.88	3MKTA	3.75	3UX4A	9.46
3X3BA	8.55	4DVEA	3.83	4KPPA	4.76
4OR2B	6.7	4RP9B	6.63	4Z3NA	7.81
5BUTJ	4.19	5GJWE	7.75	5KBWA	5.61
5LWYA	5.44	5OONA	8.2	5WUEC	6.19
1NEKD	7.66	6CQ8B	5.45	2Q7RC	8.27
2XQ2A	9.28	2IQRA	5.31	3V2YA	5.28
3ZUYA	4.98	4FI3A	6.17	4L6RA	6.53
4P6VB	8.47	4RYMA	4.68	6BTMF	3.26
5BZ2B	7.21	5GUFA	8.05	5KKZE	7.72

Protein name	Resolution (Å)	Protein name	Resolution (Å)	Protein name	Resolution (Å)
5M87A	8.69	5SV9B	8.29	5XJ9A	5.49
1NEKK	9.34	6CSNB	3.19	2UUHA	5.31
2Z73B	5.07	2K74A	9.28	3VR8C	4.02
4A01A	5.17	3PJZA	9.76	4L6VF	5.84
4P6VD	9.57	4GD3A	5.96	5A43B	8.08
5C8JI	7.71	4TKRA	8.35	5L25B	3.93
5MG3Y	7.52	5GUWB	3.14	5XJJA	9.63
5Y79B	7.9	5TJ5A	8.82	1C17M	6.25
1Q16C	9.12	6D0JB	8.59	1FFTC	7.73
3B4RB	6.06	3PUWG	7.34	2A65B	4.43
4A2NB	4.79	4HEAA	7.82	2KSFA	4.02
4P79A	4.54	4TQ4A	5.01	2LCKA	9.67
5CKRA	4.56	5H36A	8.43	3ORGD	5.54
5NDDA	3.48	5TR1A	9.11	3Q7KC	5.17
5YQ7L	6.46	6DRZA	8.95	3TUIA	7.57
1Q90D	5.54	2KDCB	4.23	4D0AB	9.09
3DDLA	6.78	3QNQB	3.64	4O9PB	3.38
4AINA	4.03	4HEAJ	5.6	4P6VE	5.05
4PHZC	8.88	4TQUM	6.93	4PGRA	6.39
5CKRB	7.92	5I20E	8.18	4QUVA	3.6
5NJGA	6.06	5TSAA	5.83	4RI2A	7.56
5ZKPA	7.95	6EBMD	6.38	4U9NB	9.44
1QLEC	3.54	2KSRA	9.21	4V6MC	4.38
3EFFN	3.27	3RKOJ	6.51	4XU4C	9.86
4B4AA	7.18	4HEAM	8.22	4YMUD	6.37
4PXKA	8.3	4TQUN	7.08	5AEXA	6.84
5CTGB	6.27	5I6CA	8.48	5J4IA	9.56
5NUPB	8.43	5UNGB	4.48	5MRWA	9.8
6BD4A	9.2	6EYUB	8.94	5ULDA	5.34
1YQ3C	8.72	2M3GA	6.85	5UZ7R	5.9
3FH6F	9.75	3RKOM	3.97	5VHYF	6.44
4BPMB	6.27	4HKRC	3.09	5YS3A	9.03
4Q2GB	4.83	4US3A	5.88	5ZBQA	8.18
5DA0A	5.97	5I6XA	3.68	6B2ZM	5.77
5NUPF	3.65	5V6PB	3.49		

Table S2. Summary of the TM-score and RMSD of each refined model by EMCMR for the membrane protein.

Protein name	TM-score/RMSD	Protein name	TM-score/RMSD	Protein name	TM-score/RMSD
1GZMA	0.8747/4.208	6BTMF	0.948/3.479	6EZNB	0.7726/5.145
2VPZC	0.9083/2.2	2B6PC	0.7932/5.35	2N02A	0.7345/3.259
3VR8D	0.7189/7.106	3FI1B	0.7072/9.138	3RKON	0.961/1.729
4M58A	0.8418/2.906	4BUOA	0.9235/3.741	4HUQS	0.9022/1.84
5A63C	0.8268/2.965	4QTNC	0.8243/3.749	4WZ7B	0.876/4.807
5L2AA	0.8146/5.785	5EDLA	0.9424/1.434	5V7PA	0.8355/3.642
5XLSB	0.9507/1.846	5NV9A	0.8145/5.362	6F0KC	0.7601/6.67
1KF6D	0.8647/1.993	6C1RB	0.7305/8.921	2N4XA	0.2726/14.392
2W2EC	0.9018/2.317	2BL2A	0.9466/1.325	3S0XA	0.2115/17.045
3VVKA	0.8653/5.879	3GIAA	0.8929/3.614	4HYGB	0.8946/3.335
4MLBC	0.9273/3.304	4C7RA	0.8985/6.038	4WZ7D	0.9076/3.618
5AX0A	0.9222/2.997	5EIKA	0.8711/2.665	5JKIA	0.8724/2.28
5XPDA	0.6619/13.491	5O0TA	0.9237/1.694	5VBLB	0.9298/1.934
1KPLB	0.9189/3.936	6C263	0.8055/3.279	6F0KF	0.8088/6.596
2WSWC	0.9491/2.025	2BS2F	0.8733/6.537	2NQ2A	0.9593/1.616
3WMMM	0.817/7.316	3HD6E	0.7349/15.202	3TDPA	0.9415/1.575
4O6YB	0.8483/2.739	4CADC	0.8747/3.013	4K1CA	0.8712/3.426
5AZBA	0.6993/9.268	5F15A	0.8796/3.402	4WZ7E	0.852/4.874
5LNKA	0.5272/9.539	5O5EA	0.7679/8.217	5JNQB	0.8987/3.388
5XPRA	0.8913/2.735	6CAAA	0.8606/4.851	5VKVA	0.2646/15.256
1KQFF	0.8112/3.662	2C3EA	0.8074/3.994	6F36M	0.7078/9.673
2WWAA	0.8742/4.406	3M73A	0.9175/2.209	2NR9A	0.8466/2.886
3WO7A	0.7236/8.567	4CZ9B	0.8985/3.367	3UKMA	0.618/9.263
4OD5A	0.8875/2.784	4RFSS	0.7807/3.77	4KNFC	0.9219/2.164
5B58B	0.9565/1.6	5FN2B	0.7872/5.747	4WZ7F	0.6459/14.451
5LNKN	0.9368/2.218	5OGEE	0.8826/3.109	5JWYA	0.6517/9.995
5XRAA	0.8905/3.173	6CFWH	0.9475/2.029	5VOUF	0.8392/3.101
1LDFB	0.8714/2.967	2CFPA	0.9256/2.397	2ONKC	0.9501/1.389
2X79A	0.8738/4.542	3MKTA	0.8891/2.989	3UX4A	0.785/3.763
3X3BA	0.8917/2.572	4DVEA	0.8045/3.249	4KPPA	0.8315/4.916
4OR2B	0.9232/1.998	4RP9B	0.6925/7.56	4Z3NA	0.9345/2.495
5BUTJ	0.8685/6.07	5GJWE	0.8701/2.421	5KBWA	0.96/1.052
5LWYA	0.85/4.901	5OONA	0.8911/2.431	5WUEC	0.9539/1.275
1NEKD	0.8803/1.953	6CQ8B	0.662/14.11	2Q7RC	0.8682/2.198
2XQ2A	0.8035/7.3	2IQRA	0.8651/5.499	3V2YA	0.9322/2.188
3ZUYA	0.9584/1.445	4FI3A	0.9429/1.792	4L6RA	0.9138/2.177
4P6VB	0.2451/22.838	4RYMA	0.9626/0.972	5KKZE	0.8289/5.65
5BZ2B	0.9452/1.813	5GUFA	0.7348/5.209	5XJ9A	0.902/3.273
5M87A	0.8308/4.413	5SV9B	0.9254/2.744	2UUHA	0.9193/1.648
1NEKK	0.9129/1.58	6CSNB	0.771/6.602	3VR8C	0.8548/2.597
2Z73B	0.8884/6.368	2K74A	0.6087/7.66	4L6VF	0.5908/11.005

Protein name	TM-score/RMSD	Protein name	TM-score/RMSD	Protein name	TM-score/RMSD
4A01A	0.8529/5.169	3PJZA	0.8111/8.408	5A43B	0.8252/2.213
4P6VD	0.7741/4.29	4GD3A	0.7687/5.548	5L25B	0.8007/5.537
5C8JI	0.4541/13.249	4TKRA	0.8485/2.484	5XJJA	0.9509/1.792
5MG3Y	0.6864/8.613	5GUWB	0.9645/1.795	1C17M	0.5777/6.766
5Y79B	0.9021/2.405	5TJ5A	0.7587/5.685	1FFTC	0.8264/2.956
1Q16C	0.8216/3.982	6D0JB	0.9328/2.14	2A65B	0.9076/3.096
3B4RB	0.8576/3.035	3PUWG	0.8898/3.146	2KSFA	0.3395/9.89
4A2NB	0.8555/3.414	4HEAA	0.5457/16.755	2LCKA	0.7442/4.636
4P79A	0.7945/3.11	4TQ4A	0.8316/4.692	3ORGD	0.83/6.861
5CKRA	0.9047/3.321	5H36A	0.926/1.543	3Q7KC	0.9127/2.477
5NDDA	0.9064/2.772	5TR1A	0.8194/5.988	3TUIA	0.9197/1.945
5YQ7L	0.7384/23.333	6DRZA	0.9115/2.092	4D0AB	0.8881/3.121
1Q90D	0.5073/15.52	2KDCB	0.2666/17.435	4O9PB	0.7277/16.015
3DDLA	0.8501/4.541	3QNQB	0.5344/9.89	4P6VE	0.9144/2.072
4AINA	0.9035/6.948	4HEAJ	0.8002/2.955	4PGRA	0.6391/8.644
4PHZC	0.7112/4.566	4TQUM	0.9383/1.779	4QUVA	0.8437/4.123
5CKRB	0.8779/3.386	5I20E	0.925/1.888	4RI2A	0.5179/11.392
5NJGA	0.9193/4.148	5TSAA	0.8496/2.841	4U9NB	0.7476/8.892
5ZKPA	0.9014/2.509	6EBMD	0.7224/8.332	4V6MC	0.2371/24.622
1QLEC	0.9216/2.226	2KSRA	0.7643/3.719	4XU4C	0.7679/4.847
3EFFN	0.864/2.26	3RKOJ	0.8818/2.04	4YMUD	0.9402/1.611
4B4AA	0.7563/4.287	4HEAM	0.9379/2.114	5AEXA	0.8295/9.46
4PXKA	0.8937/2.576	4TQUN	0.8943/3.073	5J4IA	0.8893/2.999
5CTGB	0.9341/1.67	5I6CA	0.8339/5.439	5MRWA	0.7041/11.66
5NUPB	0.8538/3.388	5UNGB	0.9109/4.118	5ULDA	0.7187/5.865
6BD4A	0.8533/3.858	6EYUB	0.8995/2.295	5UZ7R	0.8802/2.804
1YQ3C	0.8788/1.859	2M3GA	0.9134/2.004	5VHYF	0.8359/2.735
3FH6F	0.782/6.201	3RKOM	0.9599/2.351	5YS3A	0.829/2.727
4BPMB	0.8874/2.184	4HKRC	0.7287/12.448	5ZBQA	0.8715/3.977
4Q2GB	0.8993/2.394	4US3A	0.9052/3.296	6B2ZM	0.9045/3.123
5DA0A	0.9542/1.731	5I6XA	0.8534/5.603	6BQOB	0.798/2.697
5NUPF	0.4563/8.6	5V6PB	0.3718/16.832		

Table S3 Mean and standard deviation of four H-bond features in regular secondary structure elements.

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

Supporting Figures

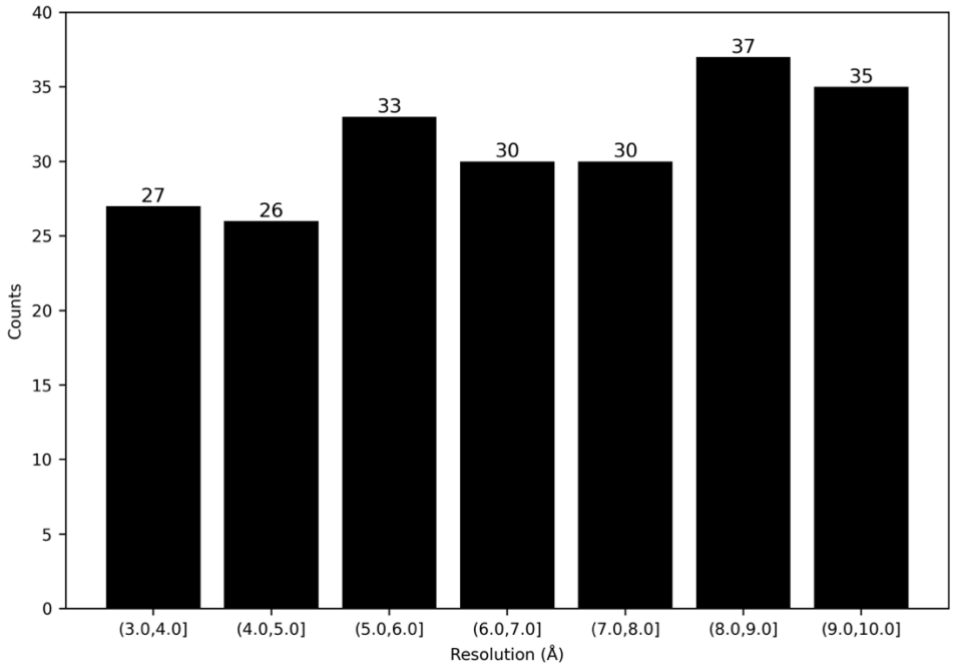

Figure S1. Histogram distribution of the simulated density map resolutions for the 218 benchmark test proteins.

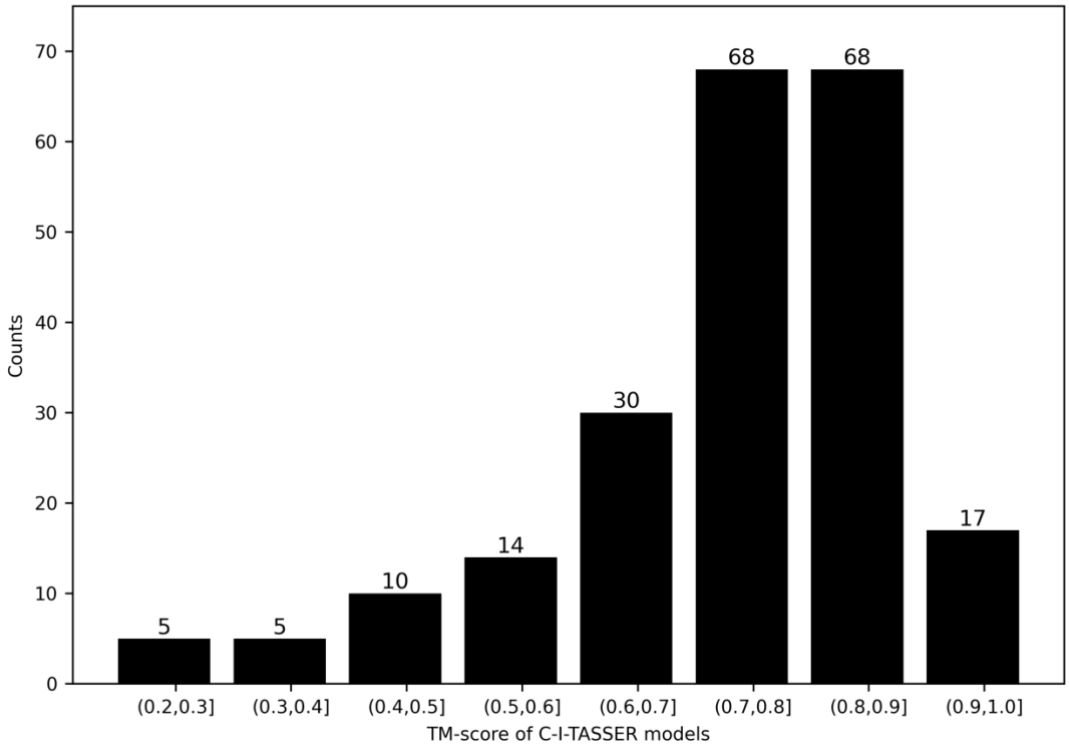

Figure S2. Histogram distribution of the TM-scores of the C-I-TASSER models for the 218 benchmark test proteins.

```
Res_id Res_id space probability average standard
1 2 0.0 0.43953 5.41582 0.41805
1 3 0.0 0.23098 6.67030 1.35847
1 4 0.0 0.15621 5.41712 0.87072
1 5 0.0 0.18136 6.55145 0.55000
1 6 0.0 0.23998 9.45999 0.51862
…
```

Figure S3. Illustration of the format of contact/distance map file which include residue id, probability, predicted the average of distance value and standard value.

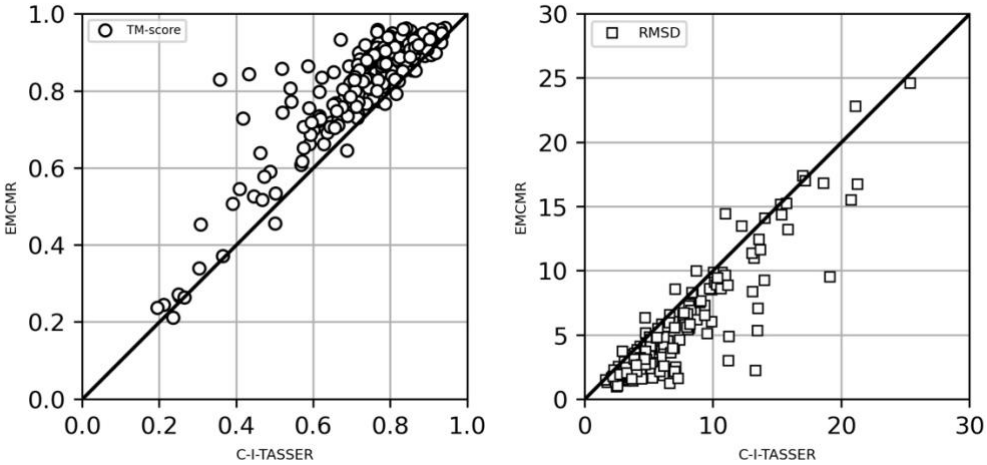

Figure S4. Comparison of structure models obtained by EMCMR and C-I-TASSER on 218 test proteins. Left and right figures are the TM-score and RMSD to the native structure, respectively.

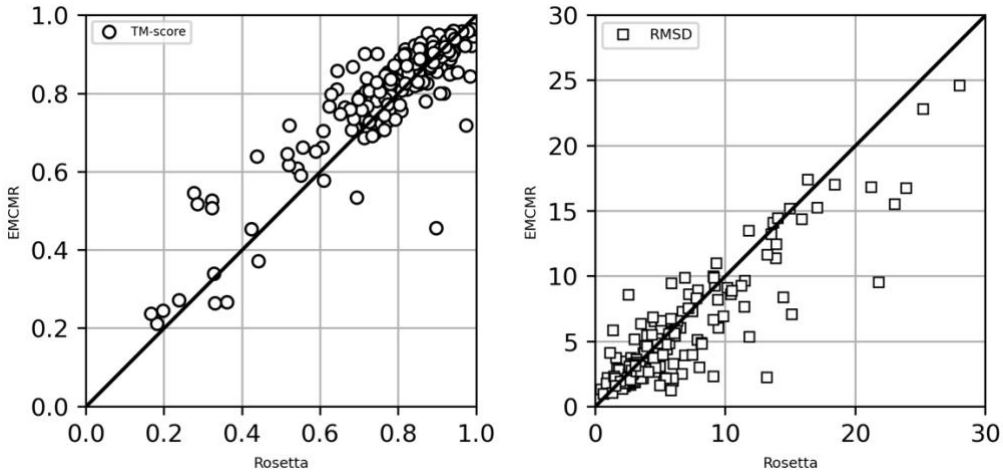

Figure S5. Comparison of structure models obtained by EMCMR and Rosetta on 218 test proteins. Left and right figures are the TM-score and RMSD to the native structure, respectively.

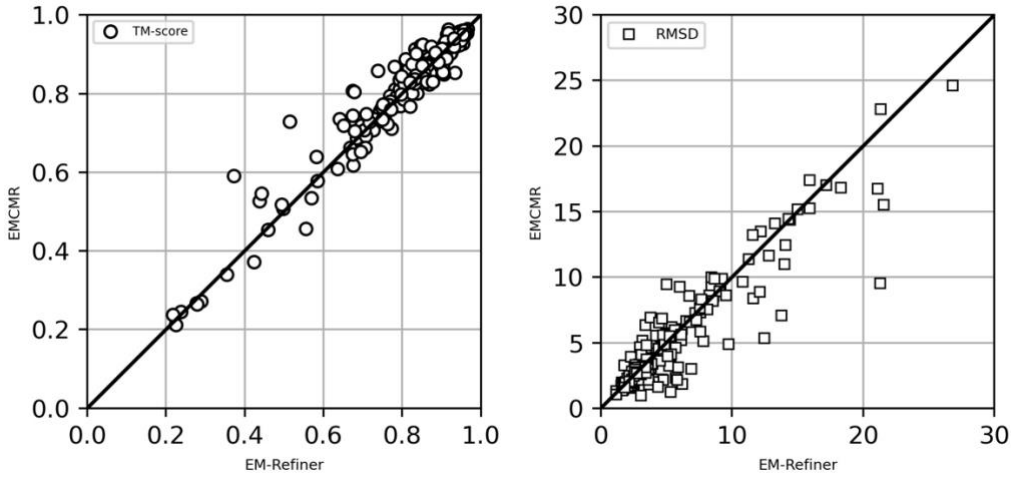

Figure S6. Comparison of structure models obtained by EMCMR and EM-Refiner on 218 test proteins. Left and right figures are the TM-score and RMSD to the native structure, respectively.

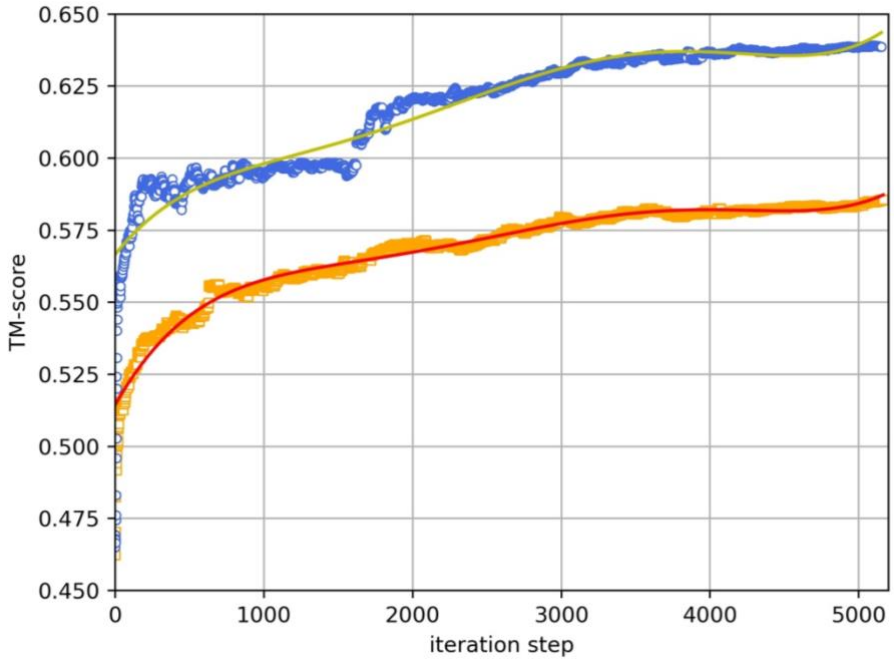

Figure S7. The TM-score of the structural decoy conformations changes with iteration steps in EMCMR simulations (blue circle) and EM-Refiner simulations (orange square).

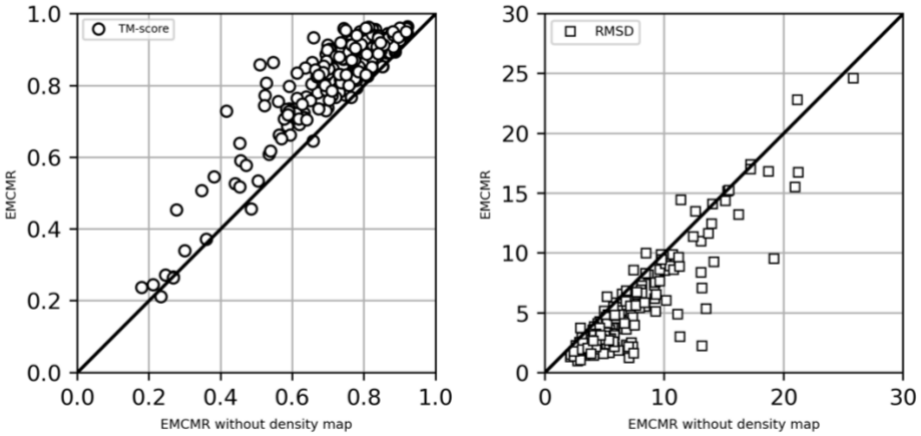

Figure S8. Comparison of structure models obtained by EMCMR and EMCMR without density map on 218 test proteins. Left and right figures are the TM-score and RMSD to the native structure, respectively.

METHODS

EMCMR force field. The energy function used to guide the EMCMR programs is a linear combination of seven energy terms:

$$E_{main} = w_1 * S_{cc} * \mathrm{E}_{cc} + w_2 \mathrm{E}_{cl} + w_3 \mathrm{E}_{hb} + w_4 \mathrm{E}_{bl} + w_5 \mathrm{E}_{tor} + w_6 \mathrm{E}_{ref} + w_7 * S_{cp} * \mathrm{E}_{cp} \quad (1)$$

where $E_{cc} = 1 - CC(res)$ and $CC(res)$ is the correlation coefficient between the calculated and experimental density maps.

The second term in Eq. (2) is the Lennard-Jones potential, which accounts for the van der Waals interactions:

$$\mathrm{E}_{cl} = \sum_{i>j} \varepsilon_{ij} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^{6} \right] \quad (4)$$

where d_{ij} is the distance between atoms *i* and *j*; ε_{ij} and r_{ij} are the well depth and the sum of the two atoms' radii, respectively, which are taken from CHARMM19[1].

E_{hb} is the main-chain hydrogen bonding energy adopted from QUARK[2]. Here, four geometric features are considered for a residue pair (*i* and *j*, Figure S6), including the distance between O_i and H_j (d_{OH}), the inner angle of $C_i - O_i - H_j$ (A_{COH}), the inner angle of $O_i - H_j - N_j$ (A_{OHN}), and the torsion angle of $C_i - O_i - H_j - N_j$ (T_{COHN}) (the torsion angle term is only used for helical residues). The H-bond energy is calculated by as follows

$$E_{hb} = \sum_{i>j} \sum_{k=1}^{4} \frac{[F_k(i,j) - \overline{F_k}]^2}{2\sigma_k^2} \quad (5)$$

where $F_k(i,j)$ denotes the *k*th feature for the candidate residue pair (*i*, *j*). $\overline{F_k}$ and σ_k are the mean and standard deviation of the *k*th feature, respectively, which were calculated from 3,881 high-resolution X-ray PDB structures with their values listed in Table S3.

E_{bl} is a bond-length potential calculated by

$$E_{bl} = \sum_{i>j} \kappa_{i,i+1} \left(d_{i,i+1} - d_0 \right)^2 \quad (6)$$

where $d_{i,i+1}$ is the distance between neighboring heavy atoms connected by a covalent bond. The parameters for statistical bond-length d_0 and the force constant κ were taken from CHARMM[1].

E_{tr} is a main-chain dihedral torsion potential calculated by

$$E_{tr} = -\sum_{i=2}^{L-1} \log \left(P\left(\phi_i, \psi_i \middle| A(i), S(i) \right) \right) \quad (7)$$

where ϕ_i/ψ_i are the Ramachandran torsion-angle pair for residue i, and $P(\phi,\psi|A,S)$ is the conditional probability of ϕ and ψ given the residue type A and the secondary structure type S, which was taken from the QUARK force field[2].

Finally, E_{res} is designed to constrain the structural differences between the decoys and the initial model by

$$E_{res} = \sum_{i,j\in G, j-i>1} \left| d_{ij}^{d} - d_{ij}^{i} \right| \quad (8)$$

where G represents all residue pairs that have Cα-Cα distances below 8.0 Å in the initial model; d_{ij}^{d} and d_{ij}^{i} are the Cα-Cα distances in the decoy and initial model, respectively.

The weight parameters are set to w_1=520, w_2=1.0, w_3=0.8, w_4=1.0, w_5=0.6, w_6=10, and w_7=1.0, which were tuned using a training set to balance the contributions of the different energy terms in Eq. (1).

References

1. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **2010,** *4* (2), 187-217.

2. Xu; D.; Zhang; Y., Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins Structure Function & Genetics*.