SUPPORTING INFORMATION

Table of Contents

Supporting Texts

Supporting Tables:

Supporting Figures:

Supplementary References

Supporting Texts

Supplementary Text 1. Dataset of CR-I-TASSER training and benchmarking

Since CR-I-TASSER simulations involve both threading template and density map restraints, to have a comprehensive test on the pipeline we first collected a large set of 530 non-redundant proteins from the PDB; these include 229 Easy targets for which LOMETS could identify at least one template with a significant score (Z-score >1), and 301 Hard targets for which LOMETS could not detect any significant templates. These test proteins are non-homologous (with a sequence identity <30%) to the 6,688 training proteins (3,088 with simulated and 3,600 with experimental maps, see Supplementary Table 1 in SI) used for training CR-I-TASSER (including the 3D-CNN pipeline, see Methods). For each test protein, we create simulated density maps using Eq. (2) in Methods with two types of resolution. For high-resolution samples, the resolution parameter R is randomly selected from [2, 5] Å, while for low-resolution samples R is from [5, 15] Å. Additionally, we have collected 248 non-redundant proteins with experimental density maps, with resolution range from 2 to 10 Å. These targets are collected from the EMDataResource, which have (1) a resolution in 2-10 Å, (2) a length >40 residues, (3) a solved structure in the PDB that covers >85% of the query sequence, (4) a pairwise sequence identity <30%, and (5) a sequence identity <30% to the training proteins. As a result, this dataset contains 178 targets with high resolution (2-5 Å) and 70 targets with low resolution (5-10 Å) maps. To reduce the uncertainty of segmentation, we segmented the density maps manually with the known native structures, and re-sample them to a grid space of 1 Å.

To eliminate homology contamination, PDB templates which have a sequence identity >30% to the query have been excluded from the LOMETS library.

Supplementary Text 2. Algorithm for determining Cα correspondences between the native structure and the 3D-CNN models

1. Prepare the coordinates set ($\mathbf{X_1}$, $\mathbf{X_2}$, … $\mathbf{X_N}$) for $C\alpha$ atoms of the native structure and ($\mathbf{Y_1}$, $\mathbf{Y_2}$, … $\mathbf{Y_M}$) for the predicted $C\alpha$ atoms by 3D-CNN.
2. Calculate the $M \times N$ distance matrix $\begin{bmatrix} d_{11} & \cdots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{M1} & \cdots & d_{MN} \end{bmatrix}$.
3. Starting with the shortest distance d_{ij} from the matrix and mark these two Cα atoms as "correspondence" and remove all elements from ith row and jth column.
4. Repeat Step-3 on the remaining matrix till all the Cα atoms in one set are chosen.

Supplementary Text 3. Algorithm of a naïve predictor for Cα atoms prediction

1. Input the original density map, locate the grid position with the largest density value and mark it as the first Cα atoms.
2. Locate the next position with largest density value and with the distance larger than 3 Å to any marked Cα atoms.
3. Repeat step 2 until the number of Cα atoms equals to the length of query sequence.

Supplementary Text 4. Explanation of TM-score

TM-score[1] is a metric for evaluating the topological similarity between protein structures, which can be calculated by

$$TM\text{-}score = max\left[\frac{1}{L_{target}}\sum_{i=1}^{L_{aligned}}\frac{1}{1+\left(\frac{d_i}{d_0(L_{target})}\right)^2}\right] \qquad (S1)$$

where L_{target} is the amino acid sequence length of the target protein, $L_{aligned}$ is the length of the aligned residues to the reference (native) structure, d_i is the distance between the i-th pair of aligned residues, $d_0(L_{target}) = 1.24\sqrt[3]{L_{target} - 15} - 1.8$ is a scale to normalize the match difference, and 'max' refers to the optimized value selected from various rotation and translation matrices for structure superposition. The value of TM-score ranges in [0,1], where 1 indicates that the two structures are identical. Stringent statistics showed that TM-score >0.5 corresponds to a similarity with two structures having the same fold defined in SCOP/CATH[2]. In our case, we need to request at least 1.7 million random structural matches to ensure a TM-score ≥0.487 as obtained by LOMETS threading, while we need at least 10 billion random structural matches to acquire a TM-score≥ 0.707 as obtained by 3D-CNN.

It should be noted that TM-score can be discrepant with the widely used root-mean-square deviation (RMSD) for some protein structure pairs. This is mainly because by definition, RMSD ($=\sqrt{\frac{1}{N}\sum_{i=1}^{N}d_i^2}$) is calculated as an average of distance error (d_i) with equal weight over all residue pairs. Therefore, a large local error on a few residue pairs can result in a quite large RMSD. On the other hand, by putting d_i in the denominator of Eq. (S1), TM-score naturally weights smaller distance errors more strongly than larger distance errors. Therefore, TM-score value is more sensitive to the global structural similarity rather than to the local structural errors, compared to RMSD. Another advantage of TM-score is the introduction of the scale $d_0(L_{target}) = 1.24\sqrt[3]{L_{target} - 15} - 1.8$ which makes the magnitude of TM-score length-independent for random structure pairs, while RMSD is a length-dependent metric[1]. Due to these reasons, our discussion of modeling results is mainly based on TM-score. Since RMSD is intuitively more familiar to most readers, however, we also list RMSD values when necessary in the manuscript.

Supplementary Text 5. Procedure for running MAINMAST

MAINMAST is a *de novo* modeling protocol by directly tracing main-chain connections and C-alpha positions from cryo-EM density map using Tree-Graph model. It was run following the tutorial from http://kiharalab.org/mainmast/Tutorials.html:

1. Generate density map in situs format and predict secondary structure from SPIDER2:
   echo 1|/bin/map2map input.mrc input.situs
   /bin/SPIDER2_local/misc/run_local.sh input.seq
2. Generate main chain traces from density map:
   MAINMAST -m input.situs -t 0.4 -filter 0.3 -Dkeep 0. -Ntb 50 -Rlocal 10 -Nlocal 100 - Nround 5000 > path.pdb

   In this step, multiple parameters combinations are tested.
3. Thread the amino acid sequence on the traces:
   ThreadCA -i path.pdb -a ./20AA.param -spd input.spd3 -fw 1.3 -Ab 3.3 -Wb 0.9 >CA.pdb
   ThreadCA -i path.pdb -a ./20AA.param -spd input.spd3   -fw 1.3   -Ab 3.3 -Wb 0.9 -r >CA_r.pdb

For experimental dataset, 64 different combinations of parameters are used to generate MAINMAST models. These models are then sent to MDFF for a short refinement. The model with lowest energy is chosen as the final model.

Supplementary Text 6. Procedure for running Rosetta de novo

Rosetta de novo is a *de novo* modeling protocol based on Rosetta (rosetta_bin_linux_2019.31.60840_bundle) with cryo-EM density map. It was run following the part 4 and part 5 from the tutorial on https://dimaiolab.ipd.uw.edu/software/:

1. Run Rosetta fragment_picker to create a "fragment file" that predicts local backbone conformations given the amino-acid sequence.
2. *De novo* model-building guided by experimental density data
    Step 4A. Local fragment search
    Step 4B. Placed fragment scoring
    Step 4C. Monte Carlo fragment assembly
    Step 4D. Consensus assignment
    Step 4E. Iterative assembly to increase model coverage
    In this step, Steps 4B-4E are iteratively run multiple times to generate initial models. Specifically, we run at least 10 iterations and also keep the initial models at least 20% completed following the RosettaES paper.
3. Completing partial models guided by experimental density data
    Step 5A. Fragment Picking
    Step 5B. Generate Possible Conformations For Each Segment
    Step 5C. Find a set of consistent conformations
    Step 5D. Interpreting results
    In this step, Steps 5B-5D are iteratively run up to 10 iterations to obtain the final models.

Supplementary Text 7. Procedure for running Phenix

Phenix is a software suite for the automated determination of molecular structures using different methods including cryo-EM. We use Phenix (1.18.2-3874) to benchmark our test dataset following the tutorial from https://www.phenix-online.org/documentation/reference/map_to_model.html and http://www.phenix-online.org/documentation/tutorials/cryo_em_structure_solution_model_building.html:

1. Generate initial model using phenix.map_to_model:
    phenix.map_to_model inputmap.mrc seq.fasta resolution=3 asymmetric_map=true include_trace_and_build=true fix_insertions_deletions=true
2. Fill the gaps of the generated model by using phenix.sequence_from_map:
    phenix.sequence_from_map run_assign_sequence=True model_file=map_to_model.pdb seq_file=seq.fasta map_file=inputmap.mrc resolution=3
3. Refine the model:
    phenix.real_space_refine assign_sequence.pdb inputmap.mrc resolution=3 ignore_symmetry_conflicts=True

In the script, seq.fasta is the input sequence file with fasta format, and inputmap.mrc is the density map in mrc format. When benchmarking the medium-low cases, we set resolution=8.

Supplementary Text 8. Procedure for running MDFF

MDFF (Molecular dynamics flexible fitting) is a molecular-dynamics-based method for flexibly fitting atomic structure into density map.

1. Prepare initial structures (from I-TASSER) and the corresponding density maps.
2. Prepare superposed structures. Although MDFF has its own superposition process, we still choose Situs to superpose the initial structures into the corresponding density maps for a fair comparison with other refinement methods.
3. Set "-gscale 1.0 -minsteps 2000 -numsteps 500000 " in step 1, "-gscale 0.3 -minsteps 2000 -numsteps 500000 -step 2" in step 2 and "-gscale 10.0 -minsteps 2000 -numsteps 0 -step 3" in step 3, and run MDFF.

Supplementary Text 9. Procedure for running EM-Refiner

EM-Refiner is a Monte-Carlo-based method for fully automatic protein structure refinement and determination with cryo-EM density map.

1. Prepare initial structures (from I-TASSER) and the corresponding density maps.
2. Prepare superposed structures. Although EM-Refiner has its own superposition process, we still choose Situs to superpose the initial structures into the corresponding density maps for a fair comparison with other refinement methods.
3. Run EM-Refiner through "UseDensityx4_not_align ./input.pdb input.mrc 3.0 0.0 /bin/version_09_24_2019_EM-Refiner", where input.pdb is the initial structure with pdb format and input.mrc is the density map with mrc format.

When benchmarking the medium-low cases, we set resolution=8.

Supplementary Text 10. Procedure for running Rosetta Refinement

Rosetta Refinement is a Monte-Carlo-based method for protein structure refinement with cryo-EM density map based on Rosetta (rosetta_bin_linux_2019.31.60840_bundle).

1. Prepare initial structures (from I-TASSER) and the corresponding density maps.
2. Prepare superposed structures. Rosetta does not include process to superpose initial structure into density map. Therefore, we first use Situs to superpose the initial structures into the corresponding density maps.
3. Run Rosetta using A_asymm_refine.sh and A_asymm_refine.xml. The A_asymm_refine.xml is as default, and the script of A_asymm_refine.sh is as follow:

```
#!/bin/bash
$ROSETTA3/source/bin/rosetta_scripts.static.linuxgccrelease \
  -database $ROSETTA3/database/ \
  -in::file::s aligned.pdb \
  -parser::protocol A_asymm_refine.xml \
  -parser::script_vars denswt=35 rms=1.5 reso=3. map=input.mrc testmap=input.mrc \
  -ignore_unrecognized_res \
  -edensity::mapreso 5. \
  -default_max_cycles 200 \
  -edensity::cryoem_scatterers \
  -beta \
  -out::suffix _asymm \
  -crystal_refine
```

In the script, aligned.pdb is the initial superposed model in pdb format, and input.mrc is the density map in mrc format. When benchmarking the medium-low cases, we set reso=8.

Supplementary Text 11. Procedure for running Flex-EM

Flex-EM is a protein fitting and refinement program guided by cryo-EM density map. We employed the Flex-EM from Modeller-9.20 in our benchmarking.

1. Prepare initial structures (from I-TASSER) and the corresponding density maps.
2. Prepare superposed structures. Although Flex-EM has its own superposition process, we still choose Situs to superpose the initial structures into the corresponding density maps for a fair comparison with other refinement methods.
3. Run Flex-EM with flex-em.py downloaded from http://topf-group.ismb.lon.ac.uk/flex-em/flex-em.py through "/modeller-9.20/bin/modpy.sh flex-em.py ./aligned.pdb ./inputmap.mrc 3. 0. 0. 0.".

In the script, aligned.pdb is the initial superposed model in pdb format, and inputmap.mrc is the density map in mrc format. When benchmarking the medium-low cases, we set resolution=8.

Supplementary Text 12. Procedure for running iMODFIT
iMODFIT is a Normal-Mode-Analysis-based method for atomic structure flexible fitting into cryo-EM density maps. We used iMODFIT (v1.51_Linux_20190228) in our benchmarking.
1. Prepare initial structures (from I-TASSER) and the corresponding density maps.
2. Prepare superposed structures. Although iMODFIT has its own superposition process, we still choose Situs to superpose the initial structures into the corresponding density maps for a fair comparison with other refinement methods.
3. iMODFIT requires input density map in Situs format (.sit), which can be generate through Situs map2map "echo 1|/bin/map2map inputmap.mrc inputmap.sit".
4. Run iMODFIT through "iMODFIT_v1.51_Linux_20190228/bin/imodfit_mkl aligned.pdb inputmap.sit 3 0".

In the script, aligned.pdb is the initial superposed model in pdb format, and inputmap.mrc is the density map in mrc format when preparing .sit file. When benchmarking the medium-low cases, we set resolution=8.

Supplementary Text 13. Benchmark results on 229 Easy targets with high-resolution simulation maps
In Table 1 (Rows 14-23), we also list the modeling results on the 229 Easy targets in our benchmark set. Due to the improved quality of LOMETS threading templates, I-TASSER generated models of a significantly higher TM-score (0.762) compared to the Hard targets (0.345). Accordingly, Situs could match the I-TASSER models with the density maps given the high-resolution samples, which resulted in a significantly improved model quality by all five refinement-based programs, with TM-score=0.824, 0.799, 0.857, 0.846 and 0.851 for Flex-EM, iMODFIT, MDFF, EM-Refiner and Rosetta-Ref, respectively. However, with the assistance of 3D-CNN models and optimized I-TASSER force field, CR-I-TASSER generates models with a significantly higher TM-score compared with all other tested methods (0.949; $p<10^{-20}$ in all cases, Student's t-test). The average RMSD (1.39 Å) and the number of cases with TM-score >0.9 (198) by CR-I-TASSER are also consistently better than the best competing refinement programs, EM-Refiner (4.00 Å) and MDFF (104). As expected, since the de novo programs do not use template and I-TASSER model information, their performance is comparable to the Hard targets with average TM-score=0.439, 0.474 and 0.493 by MAINMAST, Rosetta-dn and Phenix, respectively, which are worse than both CR-I-TASSER and the refinement-based methods.

Supplementary Text 14. Critical importance of template improvement by 3D-CNN predictor
One of the major driving forces for CR-I-TASSER structure assembly is the utilization of deep-learning $C\alpha$ conformation prediction. However, it is not clear if the improvement was due to template quality improvement, or the better model and density map superposition brought by the 3D-CNN models. In Supplementary Figure 2, we present the TM-score results of CR-I-TASSER (named CR-I-TASSER$_{LOMETS}$) which uses the original LOMETS templates but have them optimally superposed to the density map by TM-score superposition matrix. Although CR-I-TASSER$_{LOMETS}$ still significantly outperforms I-TASSER, the average TM-score (0.524) is significantly lower than that by CR-I-TASSER (0.772), demonstrating the critical importance of template quality improvement by the 3D-CNN models.

Supplementary Text 15. Benchmark results of adding Gaussian noise on simulated maps
To examine the robustness of CR-I-TASSER under noise, we add Gaussian noises to the simulated density maps for both Hard and Easy targets by using Xmipp (see Supplementary Text 16 for details). As

listed in Supplementary Table 4 (Rows 4 and 7), the average TM-scores of CR-I-TASSER drops considerably with the noise samples, from 0.772 to 0.683 for Hard targets and from 0.950 to 0.937 for Easy targets. Nevertheless, these values are still much higher than the control programs with noise-free maps. The TM-score dropped in Hard targets (0.089) is higher than that in Easy targets (0.013), showing that CR-I-TASSER is more sensitive to noise for Hard targets. This is probably due to the fact that the noises reduce the accuracy of 3D-CNN models, which has a higher impact on the Hard targets since the starting models for Hard targets are mainly from the 3D-CNN $C\alpha$ trace regenerations, while for Easy targets, the initial models are also from template reselection and the $C\alpha$ regeneration plays a less dominant role.

Supplementary Text 16. Procedure for adding noise to simulated density maps

We use Xmipp to add zero-mean Gaussian noise to the high-/low-resolution simulated density maps on the 530 test proteins. Here, the noise was added until the Pearson Correlation Coefficient (PCC) between noise map and noise-free map is equal or lower than PCC_{cut}, where PCC_{cut} is set = 0.9 for high-resolution density maps and = 0.8 for low-resolution density maps, respectively[1]. The corresponding average standard deviation is ~0.089 and ~0.074 for high- and low-resolution dataset, respectively.

Supplementary Text 17. Case study on polyomavirus VP1 pentamer protein

In Supplementary Figure 3, we present an illustrative example from polyomavirus VP1 pentamer protein (PDB ID: 1vps-A) which is 285-residue long with density map resolution of 3.42 Å. As shown in Supplementary Figure 3a, I-TASSER creates a model of incorrect fold (TM-score=0.173) due to the lack of good templates (best template with a TM-score=0.308); this resulted in incorrect superposition with density map by Situs (Supplementary Figure 3b) which is far away from the ideal superposition in Supplementary Figure 3a obtained by TM-score. Consequently, all refinement programs failed to refine the models due to the local conformational traps. As shown in Supplementary Figure 3c, although Rosetta-Ref seemingly fits the structure into the density map, the topology and fold are completely incorrect with a TM-score 0.163 as shown in the zoom-in figures of Supplementary Figure 3c. The TM-scores from Flex-EM, iMODFIT, EM-Refiner and MDFF are also poor (0.170, 0.186, 0.178 and 0.154, respectively). Among three *de novo* modeling methods, only Phenix built a better model with TM-score of 0.445. The other two programs, MAINMAST and Rosetta-dn, also failed in this case and generated models with TM-score of 0.161 and 0.157, respectively, probably due to the complicated topology of the target structure which can trap the conformational search to local minima. As shown in Supplementary Figure 3f, Phenix manages to generate the model fitting into the density map. However, there are various local misconnections shown in the zoom-in figures of Supplementary Figure 3f, which leads to a partially incorrect sequence mapping despite the general correct topology.

In CR-I-TASSER, although no good templates were generated from LOMETS, 3D-CNN creates an excellent Cα model with CRscore=0.990, where the subsequent Cα trace connection algorithm (outlined in Supplementary Figure 11) results in a Cα trace model with a great TM-score of 0.832, as shown in Supplementary Figure 3d. Guided by the improved templates, CR-I-TASSER eventually constructed the first model with TM-score=0.835 (Supplementary Figure 3e), significantly outperforming the control programs. This example highlights again the critical role of 3D-CNN based template refinements.

Supplementary Text 18. Benchmark results on 229 Easy targets with low-resolution simulation maps

A similar trend is seen on the modeling results on the 229 Easy targets as summarized in Table 1 (Rows 36-47). While the performance of the de novo programs is close to that observed on Hard targets, the TM-score of the refinement-based models is much higher than that of the Hard targets, due to the improved quality of the starting models by I-TASSER on the Easy targets. Accordingly, CR-I-TASSER achieves a much-improved TM-score (0.898), which is significantly higher than all the comparison programs with a p-value below 10^{-23} in Student's t-test compared with each of the other methods. Finally, in Rows 10 and 13 of Supplementary Table 4, we list the performance of CR-I-TASSER on the low-resolution datasets with

Gaussian noise added. It is shown that the noise-induced TM-score reduction (0.090 for Easy and 0.176 for Hard targets) is larger than that on high-resolution data, mainly due to the enhanced impact of noises on the 3D-CNN model predictions for the low-resolution density maps. Like what was observed for the high-resolution data, the TM-score of CR-I-TASSER models even with noise added is still significantly higher than that of the other tested programs with noise-free density maps on the same datasets.

Supplementary Text 19. Case study on Q6MIM9 from *Bdellovibrio bacteriovorus*

As an illustration of protein modeling from low-resolution maps, we present in Supplementary Figures 4c-h the modeling results on Q6MIM9 from *Bdellovibrio bacteriovorus* which contains 139 residues with a density map of very low resolution (11.5 Å). This protein was FM target T1001-D1 in CASP13[25], where I-TASSER constructed a poor model with TM-score 0.228 (Supplementary Figure 4c). Accordingly, Situs generates a nearly randomly initial structure superposition (Supplementary Figure 4d), which resulted in completely incorrect models by the refinement programs with TM-score=0.238, 0.210, 0.231, 0.221 and 0.196 for Flex-EM, iMODFIT, MDFF, EM-Refiner and Rosetta-Ref, respectively. Similarly, the *de novo* programs also failed to create correct models due to the extreme low-resolution density map which have TM-scores=0.203, 0.176 and 0.149 for MAINMAST, Rosetta-dn and Phenix, respectively. On CR-I-TASSER, however, 3D-CNN creates a reasonable Cα conformation with CRscore=0.975 (Supplementary Figure 4g). Thus, a structure matching with the 3D-CNN predicted conformation resulted in a good threading template (TM-score=0.621), which was originally ranked low in LOMETS but pulled up to rank-1 by the CRscore between template and predicted Cα conformation. In addition, the Cα trace regeneration creates a new template model of TM-score=0.702 from the 3D-CNN conformation. Finally, after the density-map guided structure reassembly simulations, CR-I-TASSER creates a significantly improved model with TM-score=0.874 as shown in Supplementary Figure 4h. This example highlights the hybrid effects of both template reselection and regeneration processes, as well as the optimized structure assembly simulations, which contributed to the successful modeling of a Hard target with very low-resolution density maps.

Supplementary Text 20. Further benchmarking with experimental maps

To examine the impact of template homology cutoff (i.e., maximum homology to the target for any structures available in the template library), we tested CR-I-TASSER with a much more stringent sequence identity cutoff=10%. As expected, this cutoff resulted in a significant TM-score reduction of the LOMETS threading templates (from 0.518 to 0.162), which subsequently led to a much worse I-TASSER model quality with average TM-score reduced from 0.637 to 0.300. Accordingly, the model quality of all refinement-based methods decreases dramatically, where none of the methods have a TM-score above 0.3 with either low- or high-resolution maps (Supplementary Table 4, Column 15-20). Nevertheless, the TM-score drop-off of CR-I-TASSER is limited (by 0.185 TM-score units) and the average TM-score (=0.598) is still significantly higher than the best of the de novo programs (=0.303 by Rosetta-dn) for the 248 experimental targets. To further examine the sensitivity of our results to neural network settings and the datasets used in training, we tested several 3D-CNN networks with similar architecture and hyperparameters as the original one (Supplementary Figure 10) but using different block counts (4, 6, 8, 10) which were trained on a purely experimental training set with more than 950 epochs. Here, as shown in Supplementary Figure 6a, we calculated the average CRscore loss from the experimental test set every 48 hours and stop the training if: (1) training epochs > 950 and max average CRscore > 0.8 and the latest average CRscore is 0.02 less than the max average CRscore or (2) training epochs > 1000. In Supplementary Figure 6b, the ascending (but close) average TM-scores against networks with increasing number of blocks (4, 6, 8 vs 10) indicates that a sufficient complexity of networks is preferred for deep-learning algorithms

to train models on the complicated systems such as cryo-EM density-maps. We also attempted to train a bigger network with 15 blocks, which takes much more computational resources but does not show better performance than that with 10 blocks, suggesting that 10 blocks might be sufficient for 3D-CNN in our case. Interestingly, the overall performances in Supplementary Figure 6b shows the average TM-score of the 10-block model trained without simulated data ('N10') is lower than that of the original 10-block model trained on a mixture of simulated/experimental dataset ('origin10'). Additionally, we also present the average TM-score of 10-block model trained with pure simulated data ('sim10'), which is lower than that of the other two 10-block models. This probably suggests that although the model trained in noise-free data is not as good as that trained in experimental data when applied to practical cases with experimental maps, the inclusion of simulated map data might help the neural network to extract the essential patterns more efficiently, as physical signals from the noise-free maps are easier and clearer for neural network to learn. Nevertheless, although the current network achieves the overall best performance, there are still many targets where other networks obtain higher TM-score as shown in Supplementary Figures 6c-h, indicating a possible optimization by further integrating different networks with reasonable combination rules.

Although CR-I-TASSER is designed for single-chain protein modeling, we tested it on the experimental dataset with full density maps instead of segmented maps. Since the current version of 3D-CNN predictor is not suited for DNA/RNA, we excluded those maps with DNA/RNA inside. As shown in Supplementary Figure 7a, the performance of CR-I-TASSER dropped in the full-map modeling with average TM-score=0.670, compared to the TM-score=0.752 with segmented maps. Nevertheless, the TM-score is higher than that of I-TASSER (0.620), showing that the maps have still some effects on the modeling. To further explore the reason for the performance dropping, we calculated the relative size of the proteins to the density maps by comparing their radii of gyration (Supplementary Text 21). As shown in Supplementary Figure 7b, when the target protein is much smaller than the full map size (with size ratio<0.3), the average TM-score (0.595) is much smaller than that using segmented maps (~0.753). With the increasing ratio of protein/map sizes, however, the performance of CR-I-TASSER in full-map is enhanced and close to that in segmented map (e.g., 0.804 vs 0.819 for ratio>0.6). This is understandable because there is a smaller model-map matching space to search through when the ratio increases and CR-I-TASSER could therefore perform reasonably well even with full density maps. However, with reduced size ratio, the initial model-map match becomes much more difficult and the impact of density map on the folding simulations will become negligible when an incorrect model-map match is taken. In this case, either an efficient segmentation algorithm or an extended version of CR-I-TASSER-complex that models multi-chain structures simultaneously will be needed, and both are currently under development. With the current program, CR-I-TASSER provides an option to allow users to include their (superposed) partial models (if applicable) as initial templates, which should help enhance the accuracy of full map modeling.

Supplementary Text 21. Compare the sizes of a full map and a target protein

The sizes of full maps and target proteins are measured by their radii of gyration, $Rg = \sqrt{\sum_i W_i(|x_i - \bar{x}|)^2 / \sum_i W_i}$. For target proteins, W_i and x_i are the mass and coordinate for ith atom; for full maps, W_i and x_i are the density value and grid coordinate for ith grid. After obtaining the radii of gyration, we further compare their sizes by calculating the ratio of their Rg, ratio $= Rg_{protein}/Rg_{full-map}$, where $Rg_{protein}$ and $Rg_{full-map}$ are Rg for protein and for full-map, respectively.

Supplementary Text 22. Map segmentation using Phenix segment_and_split_map

In our application of CR-I-TASSER to the beta galactosidase enzyme complex, we employ Phenix segment_and_split_map for making the map segmentation by following the instruction of http://www.phenix-online.org/documentation/reference/segment_and_split_map.html. Since the map of EMD-10564 contains four identical chains with resolution 2.6 Å, we choose n_au_box=1 as there is only 1 asymmetrical in the map, and regions_to_keep=4 as we would like to keep 4 segmented maps for four chains. The final commend is

"phenix.segment_and_split_map   inputmap.map   seq.fasta   resolution=2.6   regions_to_keep=4 n_au_box=1 density_threshold=0.3"

where inputmap.map and seq.fasta are the input density map and the query sequence. We also implement this protocol into the CR-I-TASSER web server as an option, in case the users need a tool for possible map segmentation.

Supplementary Text 23. "Keep-tracing mode" in template regeneration

As describe in the Methods, a Cα trace will stop growing if it is an endpoint, which could happen at the true end of the protein, or if there is no available atom in the probing radius (5.5 Å), or if there are other atoms but are already fully connected in an unused fragment. However, if there are still lots of un-connected Cα atoms and the users would like to try to generate a near-full-length template, using the "keep-tracing mode" described as follows.

In this mode, after encountering the normal endpoint, CR-I-TASSER will evaluate the length of the current trace. If the length of trace is lower than 90% of the query sequence, it will keep building the trace from the endpoint instead of ending it by two strategies: 1), breaking the connection patterns around the endpoint, see Supplementary Figure 12 a-b; 2), doing a gap jumping by searching the next atom with a large radius (15 Å), as shown in Supplementary Figure 12 c-d.

For strategy 1, CR-I-TASSER will search all the available Cα atoms less than 5.5 Å to the current endpoint (shown in yellow in Supplementary Figure 12) and break their connections (shown in red sticks in Supplementary Figure 12). After breaking the connection, it is possible that the current endpoint is able to connect to the next Cα atoms, which are shown in green in Supplementary Figure 12 b.

For strategy 2, if CR-I-TASSER cannot locate the next Cα atoms after using strategy 1, it will try to enlarge the search radius to 15 Å in case there is gap in the model or density map. As shown in Supplementary Figure 12 c-d, CR-I-TASSER connects through the gap which cannot be connected with the original strategy. After gap jumping, CR-I-TASSER will also do a gap jumping in the corresponding sequence assignment by $N_{gap}=L_{gap}/3.8 - 1$, where L_{gap} is the length of the gap calculated from the Euclidian distance of two atoms on the gap, which are shown in green in Supplementary Figure 12 c-d.

If CR-I-TASSER is still unable to locate new available Cα atoms for the next connection after using strategies 1 and 2, it will end the current tracing and record the trace for the later sequence assignment/energy evaluation. This mode has been implemented in the web server as an optional argument. Additionally, since CR-I-TASSER can generate near-full-length templates in a short time, users could also choose if they would like to run a full version of CR-I-TASSER (with simulations and refinement), or just wait for these templates to finish being generated.

Supplementary Text 24. CR-I-TASSER extension for superposing templates from two chains into a density map

In the case study of low-resolution (13.5 Å) map EMD-30703, we successfully build the three large receptor-binding domains (RBD) but fail to build the 2-chain 2H2 Fab model because these two chains

share very similar structures (TM-score=0.730 using TM-align and normalized by the longer chain). To solve this, we modify the template superposition and reselection part as follows:

1. Since the resolution is extremely low, we first employ BFGS algorithm (details see Supplementary Text 27) to separately superpose templates from both chains to the map; for each template, we record the top-100 poses with correlation-coefficient larger than 0.25.
2. With these two conformation lists, we enumerate the combinations of the poses from chain 1 and chain 2, with energy function $E_{two-chain} = Ecc_1 + Ecc_2 + E_{exc}$, where Ecc is the negative of PCC (see Supplementary Text 26). E_{exc} is the energy due to excluded volume. For two Cα atoms from different chains, if the distance between them is less than 4 Å, an extra energy of $(1 - f^{12})/(2N_1N_2)$ is added to the E_{exc}, where $f = distance/4$Å, N_1 and N_2 are the number of residues of chain 1 and chain 2.
3. Re-rank the templates (poses) pair with the energy function and select top-40 to the next simulations.

Supplementary Text 25. Residual neural network architecture in 3D-CNN predictor

The purpose of a residual neural network is to prevent gradient vanishing/exploding by allowing the network to skip any layers using a shortcut architecture. As shown in Supplementary Figure 7a, a basic block (shown in orange) contains two 3D convolutional layers, two instance normalization layers and two ReLU activation functions. Notably, a shortcut connects the beginning input and the output before the final ReLU directly, which potentially allows the network to skip the middle layers if gradient vanishing/exploding happened to them. This would allow us to build a deeper network without concern about the gradient vanishing/exploding problem.

Supplementary Text 26. Procedure for calculating PCC between structure and target density map

The PCC between a structure (e.g., template or structural decoy) and a target density map ρ_0 can be calculated by first converting the structure to a simulated density map ρ by Eq. 2 in the main text. The PCC is then calculated by $PCC = \dfrac{E[\rho,\rho_0] - E[\rho]E[\rho_0]}{\sqrt{E[\rho^2] - (E[\rho]^2)}\sqrt{E[\rho_0^2] - (E[\rho_0]^2)}}$, where E[…] is the expectation value over all the grids on simulated/target density map.

Supplementary Text 27. BFGS algorithm for structure rigid-body docking into density map

Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is a method of solving unconstrained optimization problems by iteratively approximating the Hessian matrix of the loss function and moving to the local maximum/minimum along the gradient. In the rigid-body docking problem, we choose the PCC between the structure and target density map defined in Supplementary Text 26 as the loss function for maximization. The dimension of this optimization problem is 6, including structure center coordinate (x,y,z) and Euler angle (θ, φ, ψ). The procedure is as follows,

1, Initialize vector $\boldsymbol{X}^{(0)} = (x, y, z, \theta, \varphi, \psi)$ and matrix $B_0^{-1} = I$, t=0;

2, Compute gradient $\boldsymbol{g}_t$ for every dimension, then calculate the searching direction $\boldsymbol{d}^{(t)} = -B_0^{-1} \cdot \boldsymbol{g}_t$;

3, Search from $\boldsymbol{X}^{(t)}$ along different dimension along $\boldsymbol{d}^{(t)}$:

$$\lambda_t = argmax\, f\left(\boldsymbol{X}^{(t)} + \lambda \cdot \boldsymbol{d}^{(t)}\right)$$
$$\boldsymbol{X}^{(t+1)} = \boldsymbol{X}^{(t)} + \lambda_t \cdot \boldsymbol{d}^{(t)}$$

where f is the PCC in our system;

4, If $|g_{t+1}| < \epsilon$, finish docking; otherwise, move to step 5;

5, $\Delta g = g_{t+1} - g_t$, $\Delta X = X^{(t+1)} - X^{(t)}$, update B_t^{-1} to

$$B_{t+1}^{-1} = \left(I - \frac{\Delta X \Delta g^T}{\Delta X^T \Delta g}\right) B_t^{-1} \left(I - \frac{\Delta g \Delta X^T}{\Delta X^T \Delta g}\right) + \frac{\Delta X \Delta X^T}{\Delta X^T \Delta g};$$

6, t = t+1; back to step 2.

Supplementary Text 28. Knowledge-based energy terms of inherent I-TASSER force field

The inherent I-TASSER knowledge-based energy force field contains 17 energy terms which can be categorized into six classes.

(1) Burial interaction restraints

The first potential uses ellipsoid representation for protein to consider the general solvation effect of residues:

$$E_{burial}^{SG} = -\sum_{i=1}^{L} E(x_i, y_i, z_i) * P(ASA_i), \tag{S2}$$

$$E(x_i, y_i, z_i) = \min[0, \max(-1, \frac{(x_i - x_c)^2}{x_0^2} + \frac{(y_i - y_c)^2}{y_0^2} + \frac{(z_i - z_c)^2}{z_0^2} - 2.5)]. \tag{S3}$$

Here, $P(ASA_i)$ is the accessible surface of the i-th residue predicted by PSSpred[2]. The value of $P(ASA_i)$ is negative if the i-th residue is predicted as buried. (x_i, y_i, z_i) is the coordinate of the center of side-group heavy atoms for the i-th residue. (x_c, y_c, z_c) and (x_0, y_0, z_0) are the center and the length of principal axes of the protein ellipsoid representation, respectively.

(2) Secondary structure-based restraints

Secondary structure restraints for C_α. The following energy terms are designed to enhance local backbone secondary structures predicted by PSSpred[2]:

$$\begin{aligned}
E_{sec}^{C\alpha} = w_{sec1} \sum_{i=1}^{L-4} E_{sec}^{C\alpha}(d_{i,i+4}) \\
+ w_{sec2} \sum_{i=1}^{L-4} E_{sec}^{C\alpha}(\overrightarrow{H_\iota}, \overrightarrow{H_{\iota+4}}) \\
+ w_{sec3} \sum_{i=1}^{L-2} E_{sec}^{C\alpha}(\overrightarrow{M_\iota}, \overrightarrow{M_{\iota+2}}),
\end{aligned} \tag{S4}$$

$$E_{sec}^{C\alpha}(d_{i,i+4}) = \begin{cases} -2 - \dfrac{DF_i * DF_{i+1} + DF_{i+3} * DF_{i+4}}{2}, & if\ \alpha - helix \\ -2 - (DF_i * DF_{i+1} + DF_{i+3} * DF_{i+4}), & if\ \beta - sheet \\ 0, & otherwise \end{cases} \tag{S5}$$

$$\begin{aligned}
&E_{sec}(\overrightarrow{H_\iota}, \overrightarrow{H_{\iota+4}}) = \\
&\begin{cases} -\dfrac{DF_i * DF_{i+1} + DF_{i+3} * DF_{i+4}}{2}, & if\ S_{i,i+4}\ is\ helix\ and\ \overrightarrow{H_\iota} \cdot \overrightarrow{H_{\iota+4}} > 0.9 \\ -(DF_i * DF_{i+1} + DF_{i+3} * DF_{i+4}), & if\ \overrightarrow{H_\iota} \cdot \overrightarrow{H_{\iota+4}} < -0.3\ or\ \overrightarrow{H_\iota} \cdot \overrightarrow{H_{\iota+4}} > 0.5 \\ 0, & otherwise \end{cases}
\end{aligned} \tag{S6}$$

$$E_{sec}(\overrightarrow{M_\iota}, \overrightarrow{M_{\iota+2}}) = -\frac{DF_i + DF_{i+1} + DF_{i+2}}{2} * \frac{\min(0.71, \overrightarrow{M_\iota} * \overrightarrow{M_{\iota+2}})}{0.71}, \tag{S7}$$

$$DF_i = \min\left(\max\left(\frac{2.2L^{0.38}}{(x_i - x_c)^2 + (x_i - x_c)^2 + (x_i - x_c)^2}, 0.5\right), 1\right). \qquad (S8)$$

In Eq. S4, the conditions for *α-helix* forming are: $d_{i,i+4} < 7.53\text{Å}$, $4\text{Å} < d_{i,i+3} < 8\text{Å}$, $\vec{U_i} \cdot \overrightarrow{U_{i+2}} < 0$, $\overrightarrow{U_{i+1}} \cdot \overrightarrow{U_{i+3}} < 0$, $\vec{U_i} \cdot \overrightarrow{U_{i+3}} > 0$, and the local segment $S_{i+1,i+3}$ is not predicted as a sheet. $\vec{U_i}$ is the unit vector from i-th to $(i+1)$-th Cα atom. The conditions for forming β-sheets include: $d_{i,i+4} > 11\text{Å}$, $arccos\frac{\overrightarrow{H_{i+1}} \cdot \overrightarrow{H_{i+3}}}{|\overrightarrow{H_{i+1}}| * |\overrightarrow{H_{i+3}}|} < 45°$, $arccos\frac{\overrightarrow{H_{i+1}} \cdot \overrightarrow{H_{i+2}}}{|\overrightarrow{H_{i+1}}| * |\overrightarrow{H_{i+2}}|} > 135°$, and the local segment $S_{i+1,i+3}$ is not predicted as a helix. $\overrightarrow{H_{i+1}} = \frac{\vec{U_i} \times \overrightarrow{U_{i+1}}}{|\vec{U_i} \times \overrightarrow{U_{i+1}}|}$ is the hydrogen bond direction of the $(i+1)$-th residue. Eq. S5 is designed to organize the direction of hydrogen bonding direction $\overrightarrow{H_i}$. Eq. S6 is for $\overrightarrow{M_i}$, which is equal to $\frac{\overrightarrow{U_{i-1}} - \vec{U_i}}{|\overrightarrow{U_{i-1}} - \vec{U_i}|}$.

In Eq. S7, (x_c, y_c, z_c) and (x_c, y_c, z_c) are the coordinate for the C_α atom of the i-th residue and the center of of the protein ellipsoid representation, respectively. $2.2L^{0.38}$ is the estimated radius of gyration for a protein with length L.

Penalty for crumpling structures. The following energy term is to panelize irregular crumpled structures:

$$E_{crumpling} = \sum_{i=1}^{L-8} E_{crumpling}(i), \qquad (S9)$$

$$E_{crumpling}(i) = \begin{cases} 1, & if \ \overrightarrow{U_{i,i+4}} \cdot \overrightarrow{U_{i+4,i+8}} < 0, \ \overrightarrow{U_{i+4,i+8}} \cdot \overrightarrow{U_{i+8,i+12}} < 0 \ and \ \overrightarrow{U_{i,i+4}} \cdot \overrightarrow{U_{i+8,i+12}} > 0 \\ 0, & otherwise \end{cases}, \qquad (S10)$$

where $\overrightarrow{U_{i,j}}$ is the unit vector from i-th to j-th C_α atoms.

Alpha/beta fragment restraints. The following energy term encourages the alpha/beta fragments continuity:

$$E_{sec}^{frag} = \sum_{i=1}^{L} E_{sec}^{frag}(i), \qquad (S11)$$

$$E_{sec}^{frag}(i) = \begin{cases} |d_{i,i+7} - 10.5|, & if \ S_{i,i+7} \ is \ helix \\ |d_{i,i+6} - 19.1| * 2, & if \ S_{i,i+6} \ is \ sheet \\ 0, & otherwise \end{cases}. \qquad (S12)$$

(3) Pairwise potentials

Cα-SG pairwise potential. The following energy term is designed for atomic packing and solvation between Cα atoms and the center of side-group heavy atoms:

$$E_{pair}^{C\alpha-SG} = \sum_{i}^{L} \sum_{j \neq i}^{L} E_{pair}^{C\alpha-SG}(d_{ij}^{C\alpha-SG}), \qquad (S13)$$

$$E_{pair}^{C\alpha-SG} = \begin{cases} \left(\frac{r_1}{d_{ij}^{C\alpha-SG}}\right)^2 - \frac{1}{2}, & if \ r_1 \leq d_{ij}^{C\alpha-SG} < r_2 \\ \frac{1}{2}, & if \ d_{ij}^{C\alpha-SG} < r_1 \\ 0, & otherwise \end{cases}. \qquad (S14)$$

where $d_{ij}^{C\alpha-SG}$ is the distance between the C_α atoms of the i-th residue and the center of side-group heavy atoms of the j-th residue. $r_1 = 3.14$ Å and $r_2 = 5.22$ Å.

SG-SG pairwise potential. The following energy term is for generic atomic packing between the centers of side-group heavy atoms:

$$E_{pair}^{SG} = \sum_{i}^{L} \sum_{j \neq i}^{L} E_{pair}^{SG}(d_{ij}^{SG}), \tag{S15}$$

$$E_{pair}^{SG}(d_{ij}^{SG}) = \begin{cases} U_{i,j}^{ori}, & if \ d_{ij}^{SG} < d_{cut}^{SG}(AA_i, AA_j) \\ 0, & otherwise \end{cases}, \tag{S16}$$

where d_{ij}^{SG} is the distance between i-th and j-th centers of the side group heavy atoms in the structure model. $d_{cut}^{SG}(AA_i, AA_j)$ is a residue type-dependent cutoff value for d_{ij}^{SG}. $U_{i,j}^{ori}$ is the generic orientation-dependent contact potential derived from 6,500 non-redundant high-resolution PDB structures[3]. These contacts are weighted by the summation of the BLOSUM mutation score between the residue pairs of the query and the PDB structures over a window of ±5 neighboring residues.

Parallel C_α-C_α pairwise potential. The following energy term is designed for atomic packing and solvation between parallel C_α atoms:

$$E_P^{C\alpha} = \sum_{i}^{L-i} \sum_{j>i}^{L} E_P^{C\alpha}(d_{ij}), \tag{S17}$$

$$E_P^{C\alpha}(d_{ij}) = \begin{cases} \min\left(0, -\dfrac{r_1^2}{max(r_1^2, d_{ij}^2)} + \dfrac{1}{2}\right), & if \ \overrightarrow{M_i} \cdot \overrightarrow{M_j} > 0.5 \\ 0, & otherwise \end{cases}, \tag{S18}$$

where $\overrightarrow{M_i} = \dfrac{\overrightarrow{U_{i-1}} - \overrightarrow{U_i}}{|\overrightarrow{U_{i-1}} - \overrightarrow{U_i}|}$ and r_1=4.77Å. $\overrightarrow{M_i} \cdot \overrightarrow{M_j} > 0.5$ indicates $\overrightarrow{M_i}$ and $\overrightarrow{M_j}$ are parallel.

Non-parallel C_α-C_α pairwise potential. The following energy term is designed for atomic packing and solvation between non-parallel C_α atoms:

$$E_{NP}^{C\alpha} = \sum_{i}^{L-i} \sum_{j>i}^{L} E_{NP}^{C\alpha}(d_{ij}), \tag{S19}$$

$$E_{NP}^{C\alpha}(d_{ij}) = \begin{cases} \dfrac{r_1^2}{d_{ij}^2} - \dfrac{1}{2}, & if \ \overrightarrow{M_i} \cdot \overrightarrow{M_j} \leq 0.5, \ d_{ij} < 5\text{Å} \\ 0, & otherwise \end{cases}. \tag{S20}$$

where r_1=3.48Å. $\overrightarrow{M_i} \cdot \overrightarrow{M_j} < 0.5$ indicates $\overrightarrow{M_i}$ and $\overrightarrow{M_j}$ are non-parallel.

(4) Hydrogen bond restraints

The hydrogen bonds are specified by the backbone geometry following the STRIDE[4] secondary structure assignments:

$$E_{HB} = \sum_{i=1}^{L-1} \sum_{j>i}^{L} E_{HB}(d_{ij}), \tag{S21}$$

$$E_{HB}(d_{ij}) =$$

$$\begin{cases} -w_{HB}(1 - |cc - cc_0|)(1 - |bb - bb_0|)\left[\dfrac{1}{(1 + |bri - br_0|)} + \dfrac{1}{(1 + |brj - br_0|)}\right], \\ \qquad if \ helix \ and \ |i - j| = 3 \\ -w_{HB}(|bb| * cc)\left[\dfrac{1}{1 + bri/2} + \dfrac{1}{1 + brj/2}\right], \\ \qquad if \ sheet \ and \ |i - j| < 4 \ for \ parallel \ or \ |i - j| > 20 \ for \ antiparallel \end{cases} \qquad (S22)$$

where $cc = \overrightarrow{M_i} \cdot \overrightarrow{M_j}$, $bb = \overrightarrow{H_i} \cdot \overrightarrow{H_j}$, $bri = |\varepsilon\overrightarrow{H_i} - \vec{r}|$ and $brj = |\varepsilon\overrightarrow{H_j} - \vec{r}|$. Here, ε=5.0Å or 4.6Å if both donor and receptor residues are predicted as *α-helices* or *β-sheets*. Similarly, $w_{HB} = 1$ if both donor and receptor residues are predicted as *α-helices* and *β-sheets*; otherwise $w_{HB} = 0.5$. The cutoff parameters for standard hydrogen bonds (cc_0, bb_0, br_0) were calculated from an average of 500 high resolution PDB structures with their secondary structure elements assigned by STRIDE[4].

(5) Statistical restraints

Short-range correlation restraints. This type of potential includes three energy terms counting for short-range C_α distance correlation.

$$E_{corr}^{C\alpha} = w_{corr1} \sum_{i=1}^{L-2} corr\left(AA_i, AA_{i+2}, bin(d_{i,i+2})\right)$$
$$+w_{corr2} \sum_{i=1}^{L-3} corr\left(AA_{i+1}, AA_{i+2}, bin(d_{i,i+3}), \varepsilon_i, S_{i+1,i+3}\right)$$
$$+w_{corr3} \sum_{i=1}^{L-4} corr\left(AA_{i+1}, AA_{i+2}, bin(d_{i,i+4}), S_{i+1,i+3}\right). \qquad (S23)$$

The first term $corr\left(AA_i, AA_{i+2}, bin(d_{i,i+2})\right)$ is the short-range C_α distance correlation between the *i*-th and the (*i*+2)-th residues, which comes from a look-up table. $d_{i,i+2}$ is the C_α distance between the *i*-th and (*i*+2)-th residues of the model. $bin(d_{i,i+2})$ indicates that $d_{i,i+2} < 6.03$ or that $d_{i,i+2} \geq 6.03$.

The second term $corr\left(AA_{i+1}, AA_{i+2}, bin(d_{i,i+3}), \varepsilon_i, S_{i+1,i+3}\right)$ is from a look-up table for short-range C_α distance correlation between the *i*-th and the (*i*+3)-th residues. $d_{i,i+3}$ is the C_α distance between *i*-th and (*i*+3)-th residues of the model. $bin(d_{i,i+3})$ indicates that $d_{i,i+3} \in$ (0, 1Å], (1Å, 2Å], …, or (11Å, ∞]. ε_i denotes the local structure chirality of three consecutive C_α-C_α vectors from the *i*-th to (*i*+3)-th residue. $S_{i+1,i+3}$ denotes that the local segment from the *i*-th to (*i*+3)-th residue is an alpha-helix, beta-sheet or coil.

The third term $corr\left(AA_{i+1}, AA_{i+3}, bin(d_{i,i+4}), S_{i+1,i+3}\right)$ also comes from a look-up table for correlation between the *i*-th and the (*i*+4)-th residues. $d_{i,i+4}$ is the C_α distance between the *i*-th and (*i*+4)-th residues of the model. $bin(d_{i,i+4})$ indicates that $d_{i,i+4} \in$ (0, 1Å], (1Å, 2Å], …, or (15Å, ∞]. $w_{corr1}, w_{corr2}, w_{corr3}$ are the weights used to balance each energy term.

Excluded volume restraints. The following energy mimics a smaller hard-core exclusion plus a bigger 1/r type of soft-core exclusion:

$$E_{vol}^{SG} = \sum_{i}^{L-i} \sum_{j>i}^{L} E_{vol}^{SG}(d_{ij}^{SG}), \qquad (S24)$$

$$E_{vol}^{SG}(d_{ij}^{SG}) =$$

$$
\begin{cases}
1, \; if \begin{cases} \overrightarrow{M_i} \cdot \overrightarrow{M_j} > 0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{pa}(AA_i, AA_j), d_{max}^{pa}(AA_i, AA_j) \right) \\
or \; \overrightarrow{M_i} \cdot \overrightarrow{M_j} < -0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{an}(AA_i, AA_j), d_{max}^{an}(AA_i, AA_j) \right) \\
or \; -0.5 \le \overrightarrow{M_i} \cdot \overrightarrow{M_j} \le 0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{pe}(AA_i, AA_j), d_{max}^{pe}(AA_i, AA_j) \right)
\end{cases} \\
0, \; otherwise
\end{cases} \quad . (S25)
$$

Here, $\left(d_{min}^{pa}(AA_i, AA_j), d_{max}^{pa}(AA_i, AA_j) \right)$, $\left(d_{min}^{an}(AA_i, AA_j), d_{max}^{an}(AA_i, AA_j) \right)$ and $\left(d_{min}^{pe}(AA_i, AA_j), d_{max}^{pe}(AA_i, AA_j) \right)$, which correspond to parallel/antiparallel/perpendicular C_α vectors, are residue type-dependent statistical values that were extracted from the PDB. d_{ij}^{SG} is the distance between i-th and j-th centers of the side group heavy atoms in the model.

Modified excluded volume restraints. The following term is a modified version of the excluded volume restraints:

$$
E_{mvol}^{SG} = \sum_{i}^{L-i} \sum_{j>i}^{L} E_{mvol}^{SG}(d_{ij}^{SG}), \quad\quad (S26)
$$

$$
E_{mvol}^{SG}(d_{ij}^{SG}) =
\begin{cases}
U^{pa}(AA_i, AA_j), if \; \overrightarrow{M_i} \cdot \overrightarrow{M_j} > 0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{pa}(AA_i, AA_j), d_{max}^{pa}(AA_i, AA_j) \right) \\
U^{an}(AA_i, AA_j), if \; \overrightarrow{M_i} \cdot \overrightarrow{M_j} < -0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{an}(AA_i, AA_j), d_{max}^{an}(AA_i, AA_j) \right) \\
U^{pe}(AA_i, AA_j), if \; -0.5 \le \overrightarrow{M_i} \cdot \overrightarrow{M_j} \le 0.5 \; and \; d_{ij}^{SG} \in \left(d_{min}^{pe}(AA_i, AA_j), d_{max}^{pe}(AA_i, AA_j) \right) \\
0, \quad\quad\quad otherwise
\end{cases} \quad . (S27)
$$

where $U^{pa}(AA_i, AA_j)$, $U^{pa}(AA_i, AA_j)$, and $U^{pe}(AA_i, AA_j)$, which correspond to parallel/antiparallel/perpendicular, are residue type-dependent statistical values that were extracted from the PDB.

Separated C_α-C_α pairwise potential. The following energy considers the distance between C_α atoms from separated residues:

$$
E_{Spair1-5}^{C\alpha} = \sum_{i=3}^{L-3} \sum_{j>i}^{L-1} E_{Spair1-5}^{C\alpha}(d_{ij}), \quad\quad (S28)
$$

$$
E_{Spair1-5}^{C\alpha}(d_{ij}) =
\begin{cases}
-corr\left(AA_{i-1}, AA_{i+1}, bin(d_{i-2,i+2}), S_{i-1,i+1}\right) \\
* corr\left(AA_{j-1}, AA_{j+1}, bin(d_{j-2,j+2}), S_{j-1,j+1}\right), \\
\quad if \begin{cases} \overrightarrow{M_i} \cdot \overrightarrow{M_j} > 0.5 \; and \; d_{ij}^{SG} \in \left(0, d_{max}^{pa}(AA_i, AA_j) \right) \\
or \; \overrightarrow{M_i} \cdot \overrightarrow{M_j} < -0.5 \; and \; d_{ij}^{SG} \in \left(0, d_{max}^{an}(AA_i, AA_j) \right) \\
or \; -0.5 \le \overrightarrow{M_i} \cdot \overrightarrow{M_j} \le 0.5 \; and \; d_{ij}^{SG} \in \left(0, d_{max}^{pe}(AA_i, AA_j) \right)
\end{cases} \\
0, \quad otherwise
\end{cases} \quad . (S29)
$$

where d_{ij} is the C_α distance between the i-th and j-th residues of the model; d_{ij}^{SG} is the distance between the i-th and j-th centers of the side-group heavy atoms in the model. $corr\left(AA_{i-1}, AA_{i+1}, bin(d_{i-2,i+2}), S_{i-1,i+1}\right)$ is similar to the definition in Eq. S22.

Contact profile constraints. The following energy is designed for counting the contact environment for each residue:

$$E_{cprof} = \sum_{i=1}^{L} E_{cprof}\left(N_i^{pa}, N_i^{an}, N_i^{pe}, AA_i\right). \tag{S30}$$

where $N_i^{pa}, N_i^{an}, N_i^{pe}$ are the number of residues that are in parallel/antiparallel/perpendicular contact with the i-th residue. $E_{cprof}(N_i^{pa}, N_i^{an}, N_i^{pe}, AA_i)$ is the statistic value from the PDB calculated through the negative logarithm of the relative frequency histogram.

Contact number constraints. The following energy considers the bias to the expected contact number and contact order:

$$E_{Ncon} = \left|N^{Con} - N_0^{Con}\right| + \left|\overline{S^{Con}} - S_0^{Con}\right|, \tag{S31}$$

where N^{Con} is the number of contacts of model and $\overline{S^{Con}}$ is the average sequence separation of the contacts. N_0^{Con} and S_0^{Con} are statistical values obtained from the PDB.

(6) Sequence-based contact restraints

The contact-map predictions by ResPRE[5] on both C_α and C_β atoms are considered in CR-I-TASSER by

$$E_{Scon}^{C\alpha/C\beta} = \sum_{i=1}^{L-1} \sum_{j>i}^{L} E_{Scon}^{C\alpha/C\beta}(d_{ij}), \tag{S32}$$

$$E_{Scon}^{C\alpha/C\beta}(d_{ij}) = \begin{cases} -U_{ij}, & d_{ij} < d_{cut} \\ -\dfrac{1}{2}U_{ij}\left[1 - sin\left(\dfrac{d_{ij} - \left(\frac{d_{cut} + D}{2}\right)}{D - d_{cut}}\pi\right)\right], & d_{cut} \leq d_{ij} < D \\ \dfrac{1}{2}U_{ij}\left[1 + sin\left(\dfrac{d_{ij} - \left(\frac{D + 80}{2}\right)}{(80 - D)}\pi\right)\right], & D \leq d_{ij} < 80\text{Å} \\ U_{ij}, & d_{ij} \geq 80\text{Å} \end{cases} \tag{S33}$$

where d_{ij} is the C_α or C_β distance between the i-th and j-th residues of the model. $d_{cut} = 8$ Å and D is length-dependent constant. $U_{ij} = 2.5F_w[1 + conf(i,j) - conf_{cut}(i,j)]$, where $conf(i,j)$ is the confidence score of predicted contacts between residues i and j, $conf_{cut}(i,j) = (0.607, 0.581, 0.654)$ is the confidence score cutoff for range type $|i - j| \in ([0,11], [12,23], [24, \infty])$. $F_w = 0.62, 1.25, 6.25$ and 5 for Trivial, Easy, Hard and Very Hard targets, respectively, where the target type is specified by LOMETS threading programs.

Supplementary Text 29. Template based restraints in CR-I-TASSER

The template-based restraints are extended from I-TASSER and contain four aspects:

Template-based short-range distance restraints. The following energy term considers a short-range restraint ($|i\text{-}j|\leq6$ for the i-th and j-th residues of the model) from the templates:

$$E_{dist}^{Short} = \sum_{i=1}^{L-1} \sum_{j>i}^{i+6} E_{dist}^{Short}(d_{ij}), \qquad (S34)$$

$$E_{dist}^{Short}(d_{ij}) = \begin{cases} 1, & if \; |d_{ij} - d_{ij}^T| > \sigma_{ij}^T \\ 0, & otherwise \end{cases}. \qquad (S35)$$

where d_{ij} is the distance between C_α atoms from the i-th and j-th residues of the model. d_{ij}^T and σ_{ij}^T are the mean value and standard deviation of the C_α distances for the i-th and j-th residues collected from the templates.

Template-based long-range distance restraints. The following energy term considers a long-range restraint ($|i\text{-}j|>6$ for the i-th and j-th residues of the model) from the templates.

$$E_{dist}^{Long} = \sum_{i=1}^{L-7} \sum_{j>i+6}^{L} E_{dist}^{Long}(d_{ij}), \qquad (S36)$$

$$E_{dist}^{Long}(d_{ij}) = -\frac{1}{\max(1, |d_{ij} - d_{ij}^T|)}. \qquad (S37)$$

where d_{ij} is the distance between C_α atoms from the i-th and j-th residues of the model. d_{ij}^T is the mean value of the C_α distances for the i-th and j-th residues collected from the templates.

Template-based contact restraints for C_a. The following energy term takes the C_α atoms contact information into account, which is extracted from the templates:

$$E_{Tcon}^{C\alpha} = \sum_{i=1}^{L-1} \sum_{j>i}^{L} E_{Tcon}^{C\alpha}(d_{ij}), \qquad (S38)$$

$$E_{Tcon}^{C\alpha}(d_{ij}) = \begin{cases} -U_{ij}, & if \; d_{ij} < 6.5\text{Å} \\ 0, & otherwise \end{cases}, \qquad (S39)$$

$$U_{ij} = \begin{cases} 1 + 4 * |conf_{ij}^{C\alpha} - conf_{cut}^{C\alpha}|, & if \; conf_{ij}^{C\alpha} > conf_{cut}^{C\alpha} \\ 1 - 2 * |conf_{ij}^{C\alpha} - conf_{cut}^{C\alpha}|, & otherwise \end{cases}. \qquad (S40)$$

where d_{ij} is the distance between C_α atoms from the i-th and j-th residues of the model. $conf_{ij}^{C\alpha}$ is the contact confidence score for the i-th and j-th atoms of the model, where the confidence scores are obtained based on the threading results. $conf_{cut}^{C\alpha}$ is a pre-tuned cutoff value which is query type-dependent.

Template-based contact restraints for the center of side-group heavy atoms (SG). The following energy term takes the contact information from the center of side-group heavy atoms into account, which is extracted from the templates:

$$E_{Tcon}^{SG} = \sum_{i=1}^{L-1} \sum_{j>i}^{L} E_{Tcon}^{SG}(d_{ij}^{SG}), \qquad (S41)$$

$$E_{Tcon}^{SG}(d_{ij}^{SG}) =$$

$$E_{Tcon}^{SG}\left(d_{ij}^{SG}\right) = \begin{cases} -U_{ij}^{SG}, & d_{ij}^{SG} < d_{cut}^{SG}\left(AA_i, AA_j\right) \\ -\dfrac{1}{2}U_{ij}^{SG}\left[1 - sin\left(\dfrac{d_{ij}^{SG} - \left(\dfrac{d_{cut}^{SG}\left(AA_i, AA_j\right) + D}{2}\right)}{D - d_{cut}^{SG}\left(AA_i, AA_j\right)}\pi\right)\right], & d_{cut}^{SG}\left(AA_i, AA_j\right) \leq d_{ij}^{SG} < D \\ \dfrac{1}{2}U_{ij}^{SG}\left[1 + sin\left(\dfrac{d_{ij}^{SG} - \left(\dfrac{D + 80}{2}\right)}{(80 - D)}\pi\right)\right], & D \leq d_{ij}^{SG} < 80\text{Å} \\ U_{ij}^{SG}, & d_{ij}^{SG} \geq 80\text{Å} \end{cases} \quad \text{(S42)}$$

$$U_{ij}^{SG} = \begin{cases} 1 + 4 * \left|conf_{ij}^{SG} - conf_{cut}^{SG}\right|, & if\ conf_{ij}^{SG} > conf_{cut}^{SG} \\ 1 - 2 * \left|conf_{ij}^{SG} - conf_{cut}^{SG}\right|, & otherwise \end{cases}. \quad \text{(S43)}$$

where d_{ij}^{SG} is the distance between i-th and j-th centers of the side group heavy atoms in the model. $conf_{ij}^{SG}$ is the contact confidence score for the i-th and j-th pseudo side-group heavy atoms of the model, where the confidence scores are obtained based on the threading results. $conf_{cut}^{SG}$ is a pre-tuned cutoff value which is query type-dependent. $D = 2 + d_{cut}^{SG}\left(AA_i, AA_j\right)$, in which $d_{cut}^{SG}\left(AA_i, AA_j\right)$ is a residue type-dependent cutoff value.

Supporting Tables

Supplementary Table 1. Distribution of training and testing datasets collected for CR-I-TASSER.

	Datasets	**# maps**	**Sources**	**Density maps**	**Classes ratio[1]**
Training	Dataset 1	3,088	PISCES[2]	Simulated	134:1
	Dataset 2	3,600	EMDataResource	Experimental[3]	46:1
Validation	Dataset 3	382	PISCES[2]	Simulated	128:1
Testing	Dataset 4	530	PDB	Simulated high[4]	104:1
		530	PDB	Simulated low[5]	167:1
	Dataset 5	248	EMDataResource	Experimental	214:1[6]

[1]Classes ratio is defined as $\frac{\#class\ 0}{\#class\ 1}$, where class 1 stands for Cα atom around a grid while class 0 for not having Cα atom.

[2]These two datasets come from the same non-redundant set but randomly separated with ratio ~9:1 as training set and validation set.

[3]Experimental training set was obtained by segmenting maps with size $50 \times 50 \times 50$ Å^3 from 36 very large (>2G) experimental maps. To make the training process focus more on central parts (more Cα atoms) instead of marginal parts (only a few Cα atoms), we set a filter of these segmented maps by containing at least 250 Cα atoms.

[4]Simulated maps are calculated with resolution range from [2, 5] Å.

[5]Simulated maps are calculated with resolution range from [5, 15] Å.

[6]The classes ratio in experimental testing set is much higher than others because some experimental maps contain most of the space with no Cα atom.

Supplementary Table 2. Average TM-scores for the templates generated by LOMETS and those refined by 3D-CNN $C\alpha$ models.

Test dataset	Counts	LOMETS	3D-CNN refinement
301 Hard targets (resolution<5Å)	First template	0.283	0.690
	Best in top-5 templates	0.299	0.709
	Average on top-10 templates	0.237	0.658
	Average on top-40 templates	0.219	0.639
301 Hard targets (resolution>5Å)	First template	0.283	0.527
	Best in top-5 templates	0.299	0.540
	Average on top-10 templates	0.237	0.423
	Average on top-40 templates	0.219	0.368
229 Easy targets (resolution<5Å)	First template	0.756	0.872
	Best in top-5 templates	0.765	0.874
	Average on top-10 templates	0.703	0.753
	Average on top-40 templates	0.650	0.733
229 Easy targets (resolution>5Å)	First template	0.756	0.812
	Best in top-5 templates	0.765	0.821
	Average on top-10 templates	0.703	0.725
	Average on top-40 templates	0.650	0.702
All 530 targets	First template	0.487	0.707
	Best in top-5 templates	0.500	0.720
	Average on top-10 templates	0.438	0.626
	Average on top-40 templates	0.405	0.596

Supplementary Table 3. Clash and Molprobity Scores of the models by CR-I-TASSER and other methods on 530 test protein (301 Hard and 229 Easy) with high (2-5 Å) and low (5-15 Å) resolution density maps and 248 experimental targets with 2-9 Å density maps. These scores are calculated by Phenix[6] for final predicted models with all heavy atoms (without H atoms). The lower the clash score and Molprobity score are, the better the local structures are. Bold fonts highlight the best scores in each category.

Methods	Clash Score	Molprobity Score
301 Hard targets with high-resolution density map (resolution in 2-5 Å)		
I-TASSER[1]	69.834	3.832
Flex-EM[2]	268.803	4.672
iMODFIT[2]	465.781	5.277
MDFF[2]	122.197	4.502
EM-Refiner[2]	45.570	3.788
Rosetta-Ref[2]	**13.815**	**2.273**
MAINMAST[3]	451.274	5.276
Rosetta-dn[3]	399.407	4.695
Phenix[3]	131.286	3.878
CR-I-TASSER[3]	37.342	2.553
CR-I-TASSER[3] (noise)[5]	59.713	3.052
229 Easy targets with high-resolution density map (resolution in 2-5 Å)		
I-TASSER[1]	11.035	2.826
Flex-EM[2]	162.151	4.109
iMODFIT[2]	215.777	4.126
MDFF[2]	24.078	3.402
EM-Refiner[2]	**5.691**	2.162
Rosetta-Ref[2]	7.013	**1.743**
MAINMAST	468.513	5.303
Rosetta-dn[3]	358.396	4.487
Phenix[3]	122.412	3.829
CR-I-TASSER[3]	8.885	1.921
CR-I-TASSER[3] (noise)[5]	9.759	2.006
301 Hard targets with low-resolution density map (resolution in 5-15 Å)		
I-TASSER[1]	69.834	3.832
Flex-EM[2]	408.607	4.873
iMODFIT[2]	478.211	5.327
MDFF[2]	135.976	4.561
EM-Refiner[2]	89.236	4.217
Rosetta-Ref[2]	**16.457**	**2.420**
MAINMAST[3]	440.690	5.452
Rosetta-dn[3]	570.660	4.824
Phenix[3]	177.937	4.221
CR-I-TASSER[3]	120.964	3.920
CR-I-TASSER[3] (noise)[5]	79.992	3.966
229 Easy targets with low-resolution density map (resolution in 5-15 Å)		
I-TASSER[1]	11.035	2.826
Flex-EM[2]	344.531	4.546
iMODFIT[2]	104.477	4.058
MDFF[2]	71.799	4.067
EM-Refiner[2]	27.739	3.253
Rosetta-Ref[2]	12.893	**2.154**
MAINMAST[3]	671.014	5.551
Rosetta-dn[3]	142.284	3.270
Phenix[3]	165.281	4.176
CR-I-TASSER[3]	14.531	2.422
CR-I-TASSER[3] (noise)[5]	**10.980**	2.283

248 targets with experimental density map (resolution in 2-10 Å)		
I-TASSER[1]	21.985	2.425
Flex-EM[2]	207.835	4.337
iMODFIT[2]	261.163	4.513
MDFF[2]	45.305	3.775
EM-Refiner[2]	20.177	2.724
Rosetta-Ref[2]	**6.233**	**1.691**
MAINMAST[3,4]	501.738	5.385
Rosetta-dn[3]	96.885	3.015
Phenix[3]	117.184	3.848
CR-I-TASSER[3]	30.058	2.615

[1]Protein structure prediction methods

[2]Cryo-EM based structure refinement methods

[3]Cryo-EM based de novo structure modeling methods

[4]Nine low-resolution targets cannot be solved by MAINMAST, probably due to the low resolution and experimental noise

Supplementary Table 4. Comparison of CR-I-TASSER modeling results in different options. Bold fonts highlight the performer which obtains the best average result in each category.

Methods	⟨TM-score⟩	N (TM>TM$_0$)[1]	⟨RMSD⟩ (Å)
301 Hard targets with high-resolution density map (resolution in 2-5 Å) (TM$_0$=0.5)			
CR-I-TASSER	**0.772**	**251**	**4.4**
CR-I-TASSER (noise)[2]	0.683	217	6.1
229 Easy targets with high-resolution density map (resolution in 2-5 Å) (TM$_0$=0.9)			
CR-I-TASSER	**0.950**	**198**	**1.4**
CR-I-TASSER (noise)[2]	0.937	192	1.6
301 Hard targets with low-resolution density map (resolution in 5-15 Å) (TM$_0$=0.5)			
CR-I-TASSER	**0.597**	**191**	**6.3**
CR-I-TASSER (noise)[2]	0.422	96	9.9
229 Easy targets with low-resolution density map (resolution in 5-15 Å) (TM$_0$=0.9)			
CR-I-TASSER	**0.898**	**137**	**2.1**
CR-I-TASSER (noise)[2]	0.818	54	3.6
178 targets with experimental density map (resolution in 2-5 Å) (TM$_0$=0.9)			
I-TASSER (cutoff=0.1)[3]	0.300	0	17.4
Flex-EM (cutoff=0.1)[3]	0.281	0	17.5
iMODFIT (cutoff=0.1)[3]	0.294	0	17.4
MDFF (cutoff=0.1)[3]	0.287	0	17.3
EM-Refiner (cutoff=0.1)[3]	0.274	0	17.0
Rosetta-Ref (cutoff=0.1)[3]	0.250	0	19.5
CR-I-TASSER	**0.810**	**75**	**4.9**
CR-I-TASSER (cutoff=0.1)[3]	0.632	33	8.8
70 targets with experimental density map (resolution in 5-10 Å) (TM$_0$=0.5)			
I-TASSER (cutoff=0.1)[3]	0.300	4	17.1
Flex-EM (cutoff=0.1)[3]	0.269	2	17.4
iMODFIT (cutoff=0.1)[3]	0.290	4	16.9
MDFF (cutoff=0.1)[3]	0.275	1	16.7
EM-Refiner (cutoff=0.1)[3]	0.264	2	17.0
Rosetta-Ref (cutoff=0.1)[3]	0.246	2	18.3
CR-I-TASSER	**0.714**	**63**	**6.2**
CR-I-TASSER (cutoff=0.1)[3]	0.513	39	12.1

[1]TM$_0$=0.5 for simulated Hard targets or low-resolution experimental targets, =0.9 for simulated Easy targets or high-resolution experimental targets
[2]Benchmark results in simulated dataset with Gaussian noise
[3]Sequence identity cutoff=0.1 was employed in threading process

Supplementary Table 5. Pearson correlation coefficient calculated between local PCC/confidence and actual modeling errors on 248 test proteins with experimental density maps. The statistics results are obtained by calculating the local PCC/confidence as well as actual modeling errors for each residue of each predicted structures in experimental dataset, where 183,033 residues from 757 CR-I-TASSER models are used in the benchmarking.

	Local PCC[1]	Local confidence[2]
Distance error from native[3]	-0.341	-0.561
Local CRscore from native[4]	0.444	0.684

[1]"Local PCC" is calculated by Eq. (10) in the manuscript.
[2]"Local confidence" is calculated by Eq. (12) in the manuscript.
[3]"Distance error from native" is the distance error between Cα atoms of the same residue from the predicted structure and the native structure.
[4]"Local CRscore from native" is calculated by Eq. (1) in the manuscript with $d_{i,j}$ from 3 for only one residue.

Supporting Figures
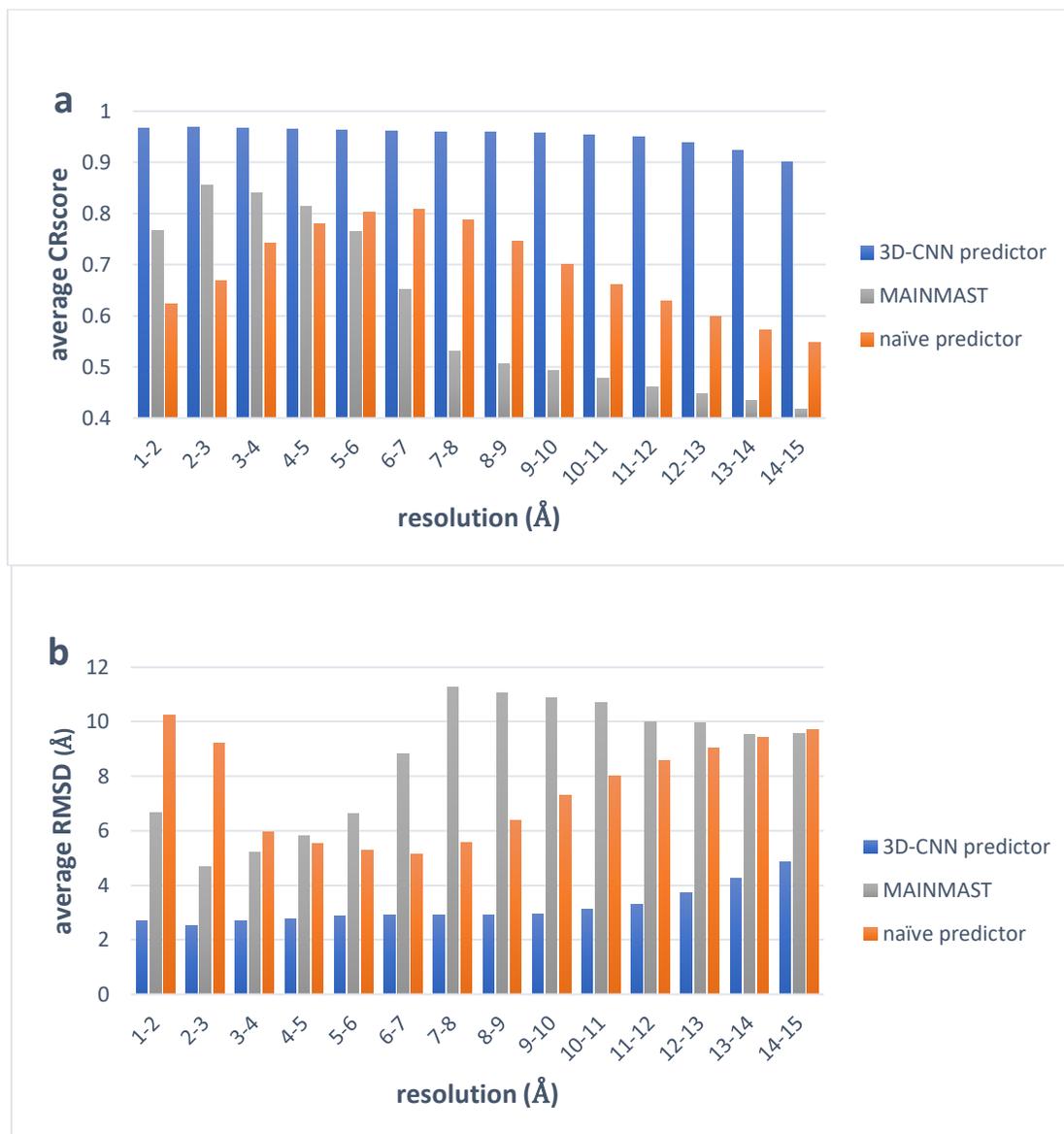
Supplementary Figure 1. Quality of Cα models by 3D-CNN, MAINMAST and a naïve method described in Supplementary Text 2 for 530 test proteins. Benchmark results of the 3D-CNN predictor are shown in blue columns where those of MAINMAST and naïve predictor shown in grey and orange columns, respectively. (a) Average CRscores calculated by Eq. 1. (b) RMSD calculated between paired atoms following Supplementary Text 1.
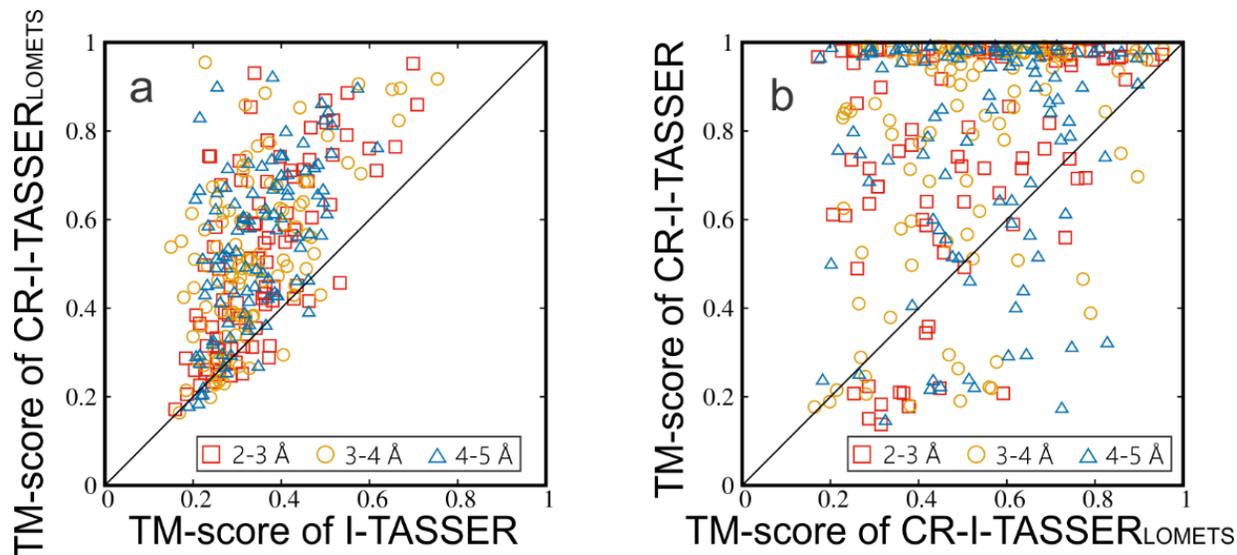
Supplementary Figure 2. Comparisons of TM-scores on 301 hard target proteins with 2-5 Å resolution density maps between I-TASSER, CR-I-TASSER and CR-I-TASSER_{LOMETS}: (a) CR-I-TASSER_{LOMETS} versus I-TASSER; (b) CR-I-TASSER versus CR-I-TASSER_{LOMETS}. The symbols with different colors and shapes denote different ranges of resolution: red square, 2-3 Å; yellow circle, 3-4 Å, blue triangle: 4-5 Å.
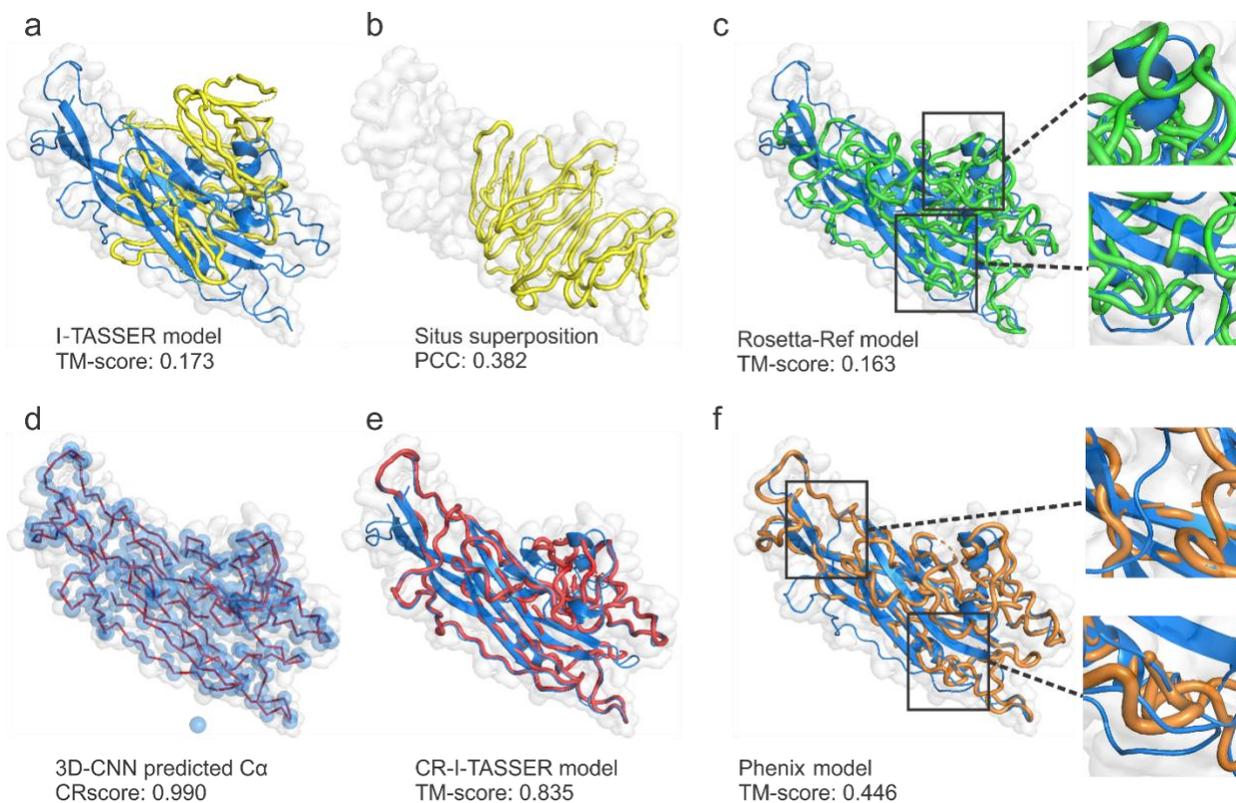
Supplementary Figure 3. Illustrative example of structure models by different methods on a Hard target from polyomavirus VP1 pentamer protein (PDB ID: 1vps-A) with high-resolution (3.42 Å) density map. (a) I-TASSER model (yellow) superposed to the native structure (blue) by TM-score; (b) I-TASSER model overlaid on density map by Situs; (c) model built by Rosetta-Ref (green) compared to the native (blue); (d) 3D-CNN $C\alpha$ trace (light blue) overlaid on density map; (e) first CR-I-TASSER model (red) compared to the native (blue); (f) model built by Phenix (orange) compared to the native (blue).
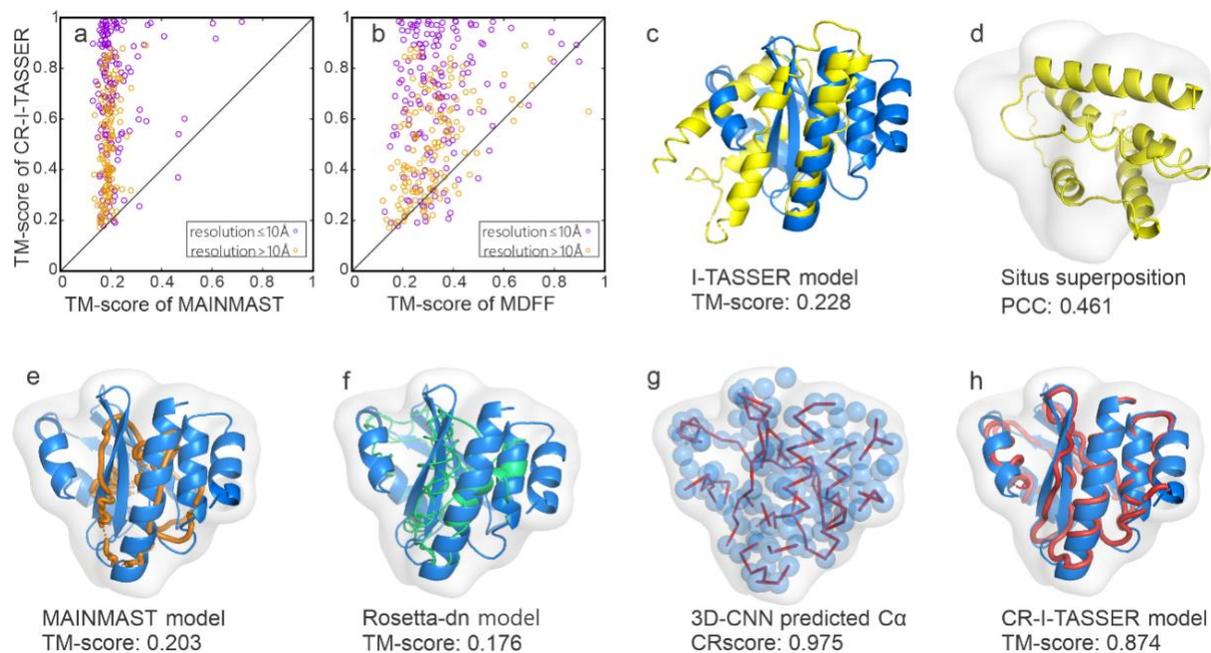
Supplementary Figure 4. Modeling results of CR-I-TASSER and the comparison methods on 301 Hard targets with 5-15 Å resolutions density maps. (a, b) TM-score comparison of CR-I-TASSER versus MAINMAST and MDFF. (c-h) Case study on a CASP13 FM target (T1001-D1). (c) I-TASSER model (yellow) superposed with the native (blue); (d) I-TASSER model overlaid on density map by Situs; (e) MAINMAST model (orange) over native structure (blue) and density map; (f) Rosetta-dn model (green) over native structure (blue) and density map; (g) predicted $C\alpha$ trace from 3D-CNN (light blue) overlaid on density map; (h) first CR-I-TASSER model (red) over native structure (blue).
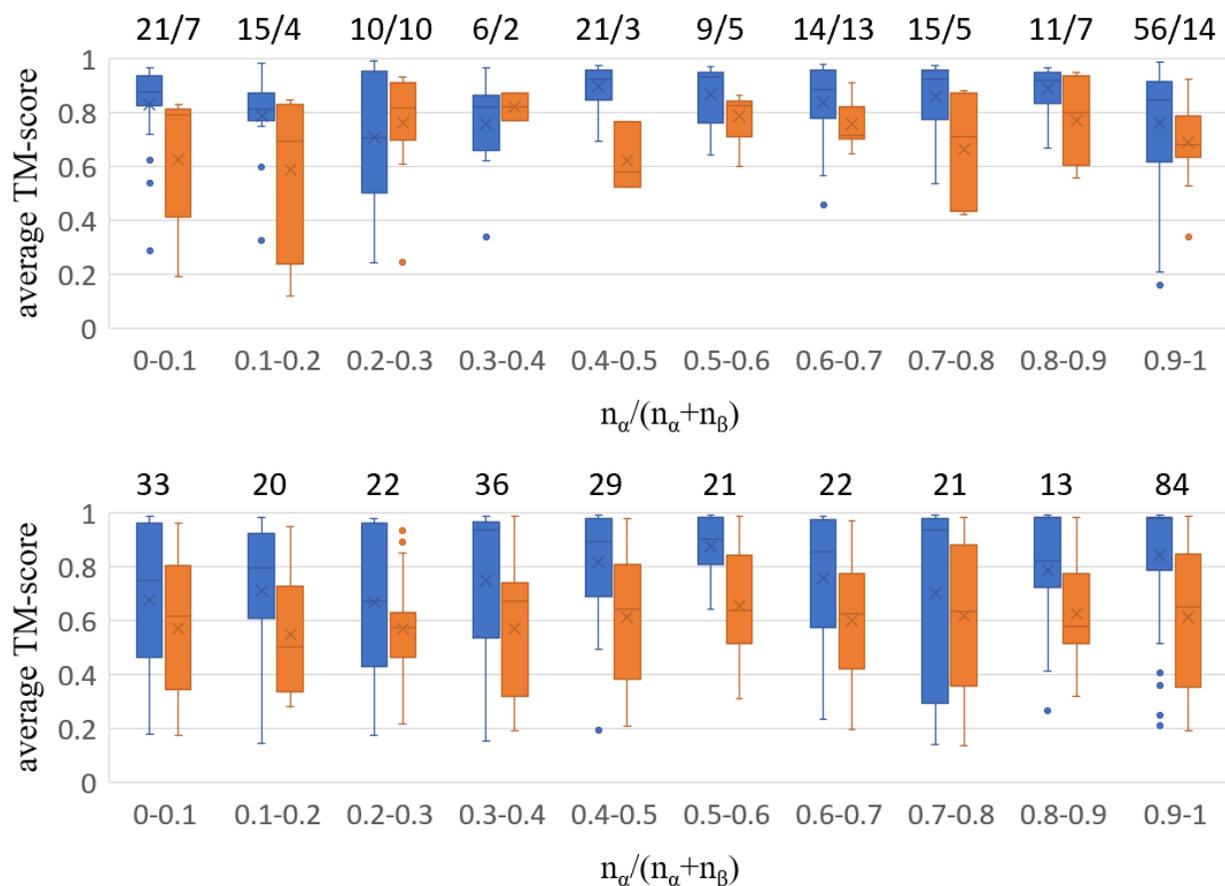
Supplementary Figure 5. Box-and-whisker plot for average TM-score versus $n_\alpha/(n_\alpha+n_\beta)$, where n_α and n_β stand for the number of residues that have alpha helix and beta strand, respectively. (a) Statistics results for high-resolution (blue) and low-resolution (orange) in experimental set. (b) Statistics results for high-resolution (blue) and low-resolution (orange) in 301 Hard targets set. The numbers of samples are shown above the corresponding category: (21, 15, 10, 6, 21, 9, 14, 15, 11, 56) and (7, 4, 10, 2, 3, 5, 13, 5, 7, 14) for high-/low-resolution experimental set in (a), and (33, 20, 22, 36, 29, 21, 22, 21, 13, 84) for 301 Hard targets set in (b). The "x" in each box is the corresponding mean value.
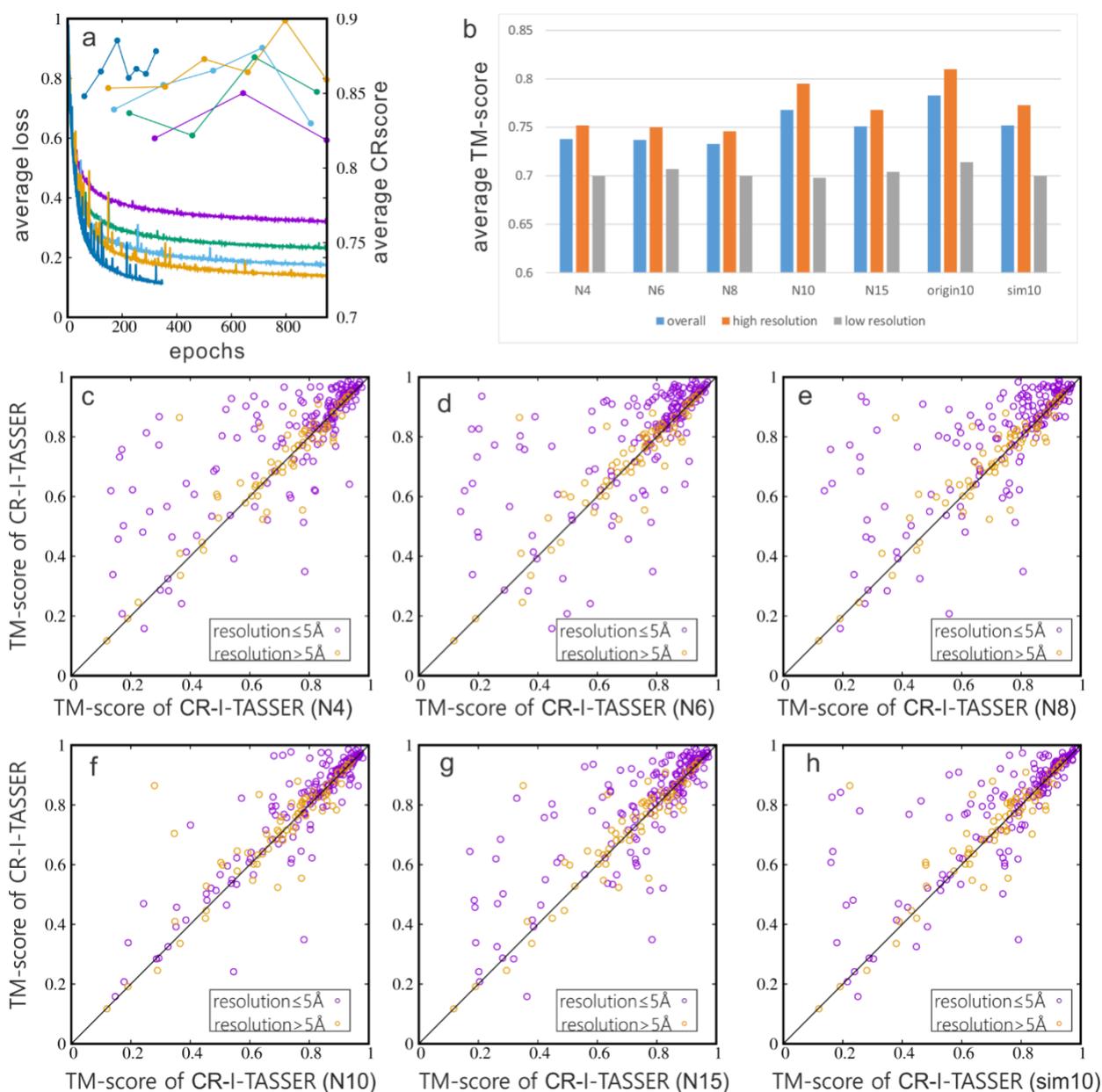
Supplementary Figure 6. (a) Average loss of neural networks with different numbers of blocks trained in pure experimental set (down lines) and the corresponding average CRscores for model choosing (upper lines): 4 blocks (purple), 6 blocks (light blue), 8 blocks (yellow), 10 blocks (yellow) and 15 blocks (deep blue). (b) Average TM-scores benchmarked on 248 experimental targets by CR-I-TASSER with different network models. (c-h) TM-score comparisons of CR-I-TASSER with original network model versus those with different network models: 4 blocks (c), 6 blocks (d), 8 blocks (e), 10 blocks (f), 15 blocks (g) and 10 blocks but trained on simulated data only (h). Purple circles are high-resolution targets while yellow circles are low-resolution targets.
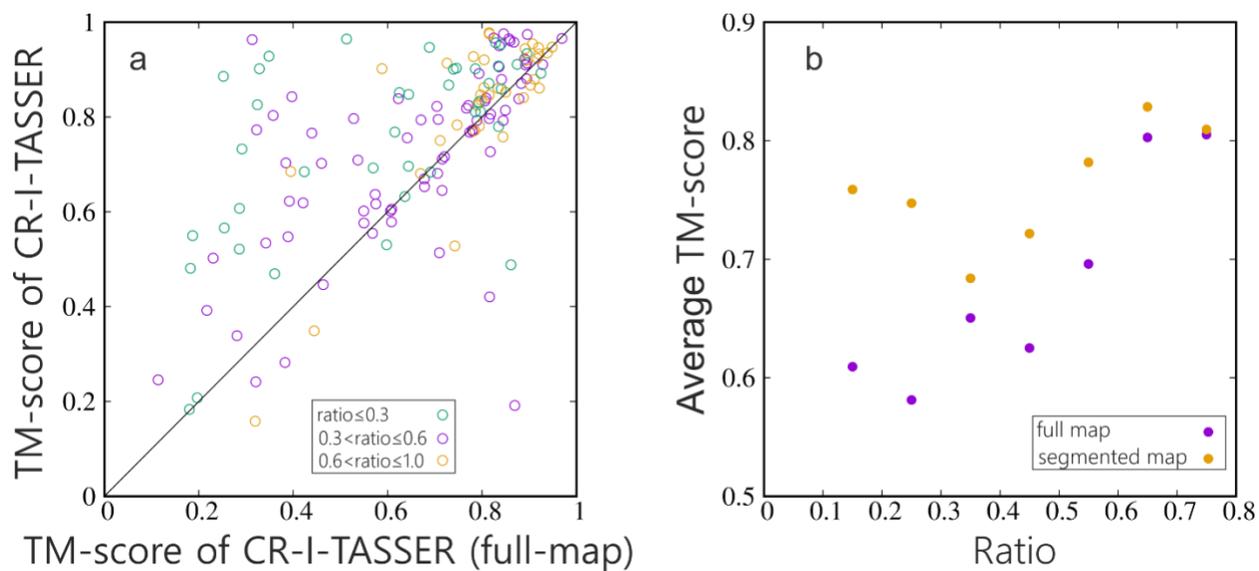
Supplementary Figure 7. Modeling results on 153 experimental targets with full density maps (without DNA/RNA). (a) CR-I-TASSER benchmarked on full-maps versus benchmarked on segmented maps. (b) Average TM-score against the relative size of proteins and the density maps by comparing their radii of gyration (see Supplementary Text 20).
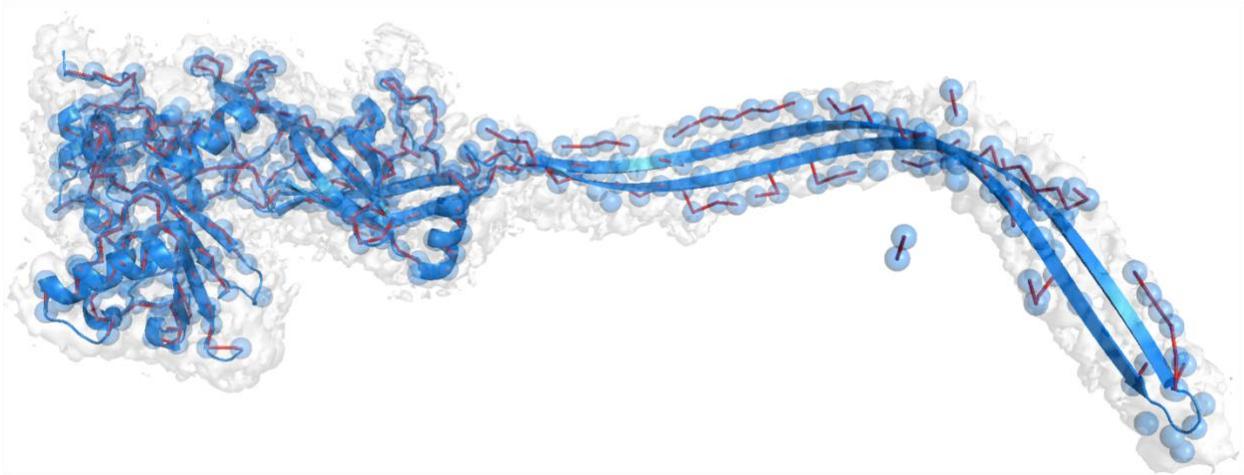
Supplementary Figure 8. Predicted Cα conformations (light blue) and connection pattern (red) calculated by CR-I-TASSER on the anthrax toxin protective antigen pore protein (PDB ID: 3j9c-A, shown in blue).
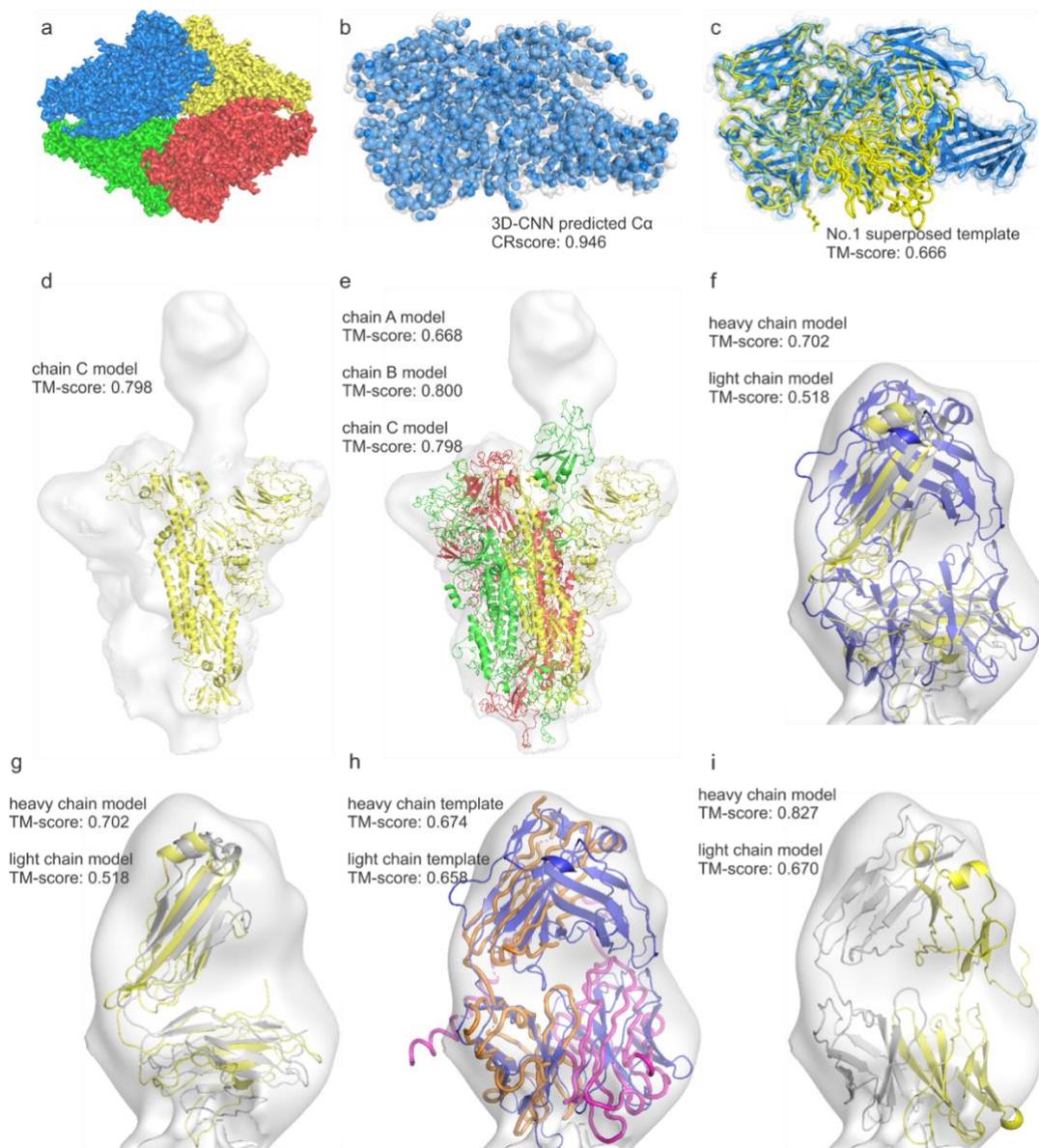
Supplementary Figure 9. Illustrative examples of end-to-end structural modeling by CR-I-TASSER from unsegment maps. Native structures are shown in blue in (c), (f) and (h). (a-c) Beta-galactosidase in complex with L-ribose (PDB ID: 6tsk) from density map (EMD-10564, resolution 2.3 Å). (a) The raw density map is segmented into four parts by Phenix segment_and_split_map. (b) Predicted Cα positions (blue) from the segmented map. (c) Best threading template from LOMETS (yellow) superposed with the native. (d-i) the SARS-CoV-2 spike protein with receptor-binding domains (RBD) bound with a 2H2 Fab (PDB ID: 7dk5) from density map (EMD-30703, resolution 13.5 Å). Native structures are shown in blue (right/left for heavy/light chains of 2H2 Fab) in (f) and (h), the rest are shown in Figure 5 in the manuscript. (d) First CR-I-TASSER model (yellow) built on the map as in the chain C location. (e) Models of chains A (green), B (red) and C (yellow) built on the map. (f-g) Models of heavy/light chains of 2H2 Fab (gold/silver) which are incorrectly built on the map due to the low resolution. (h) The best templates couple of heavy/light chain (magenta/orange) using superposition process described in Supplementary Text 23. (i) Final CR-I-TASSER models of heavy/light chains of 2H2 Fab (gold/silver).
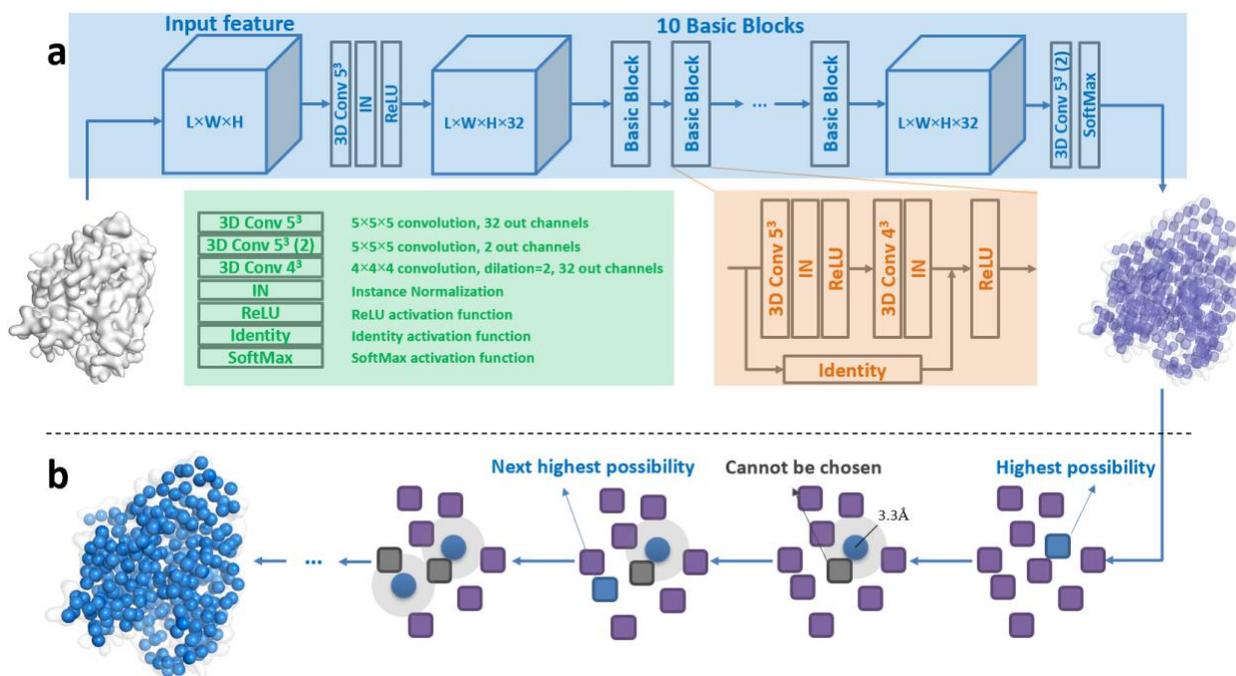
Supplementary Figure 10. 3D-CNN pipeline for sequence-independent $C\alpha$ conformation prediction. (a) Architecture of 3D-CNN predictor which uses cryo-EM density map as input and predict Cα atom possibility maps. (b) The predicted Cα atom possibility maps are converted to Cα atoms by searching for the maximum possibility grids under the consideration of excluded volume.
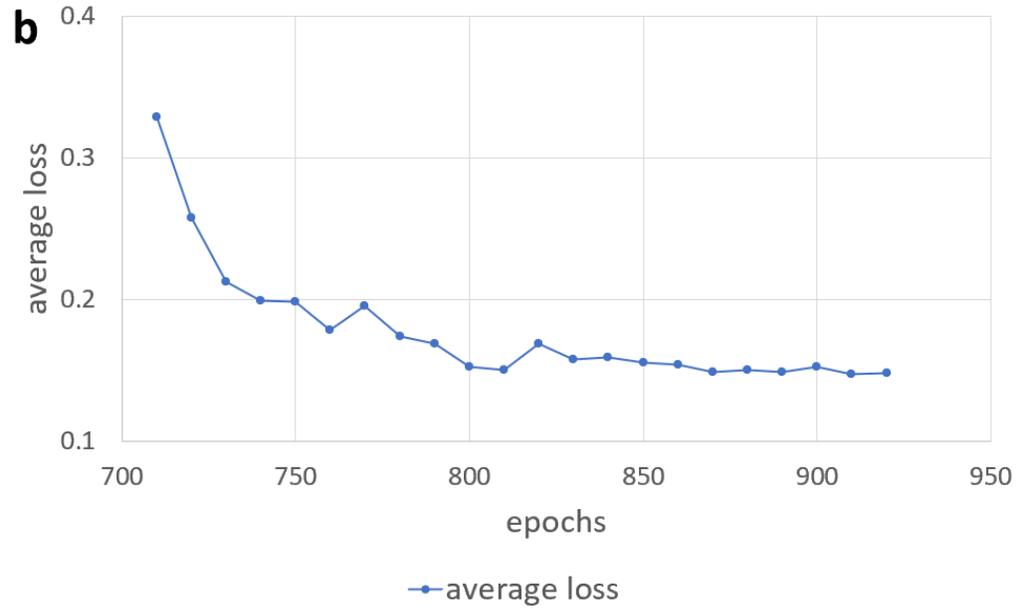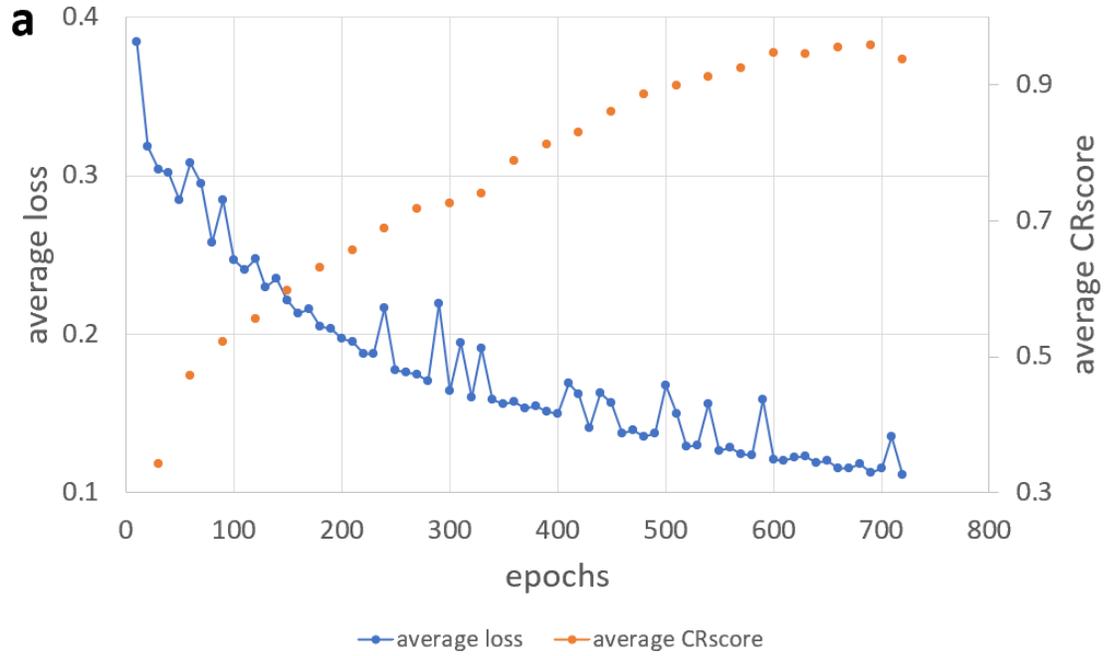
Supplementary Figure 11. (a) Average loss (blue) of the simulated training set versus average CRscore (orange) of the validation set along with training epochs. (b) Average loss (blue) of the experimental training set started from the model trained in (a).
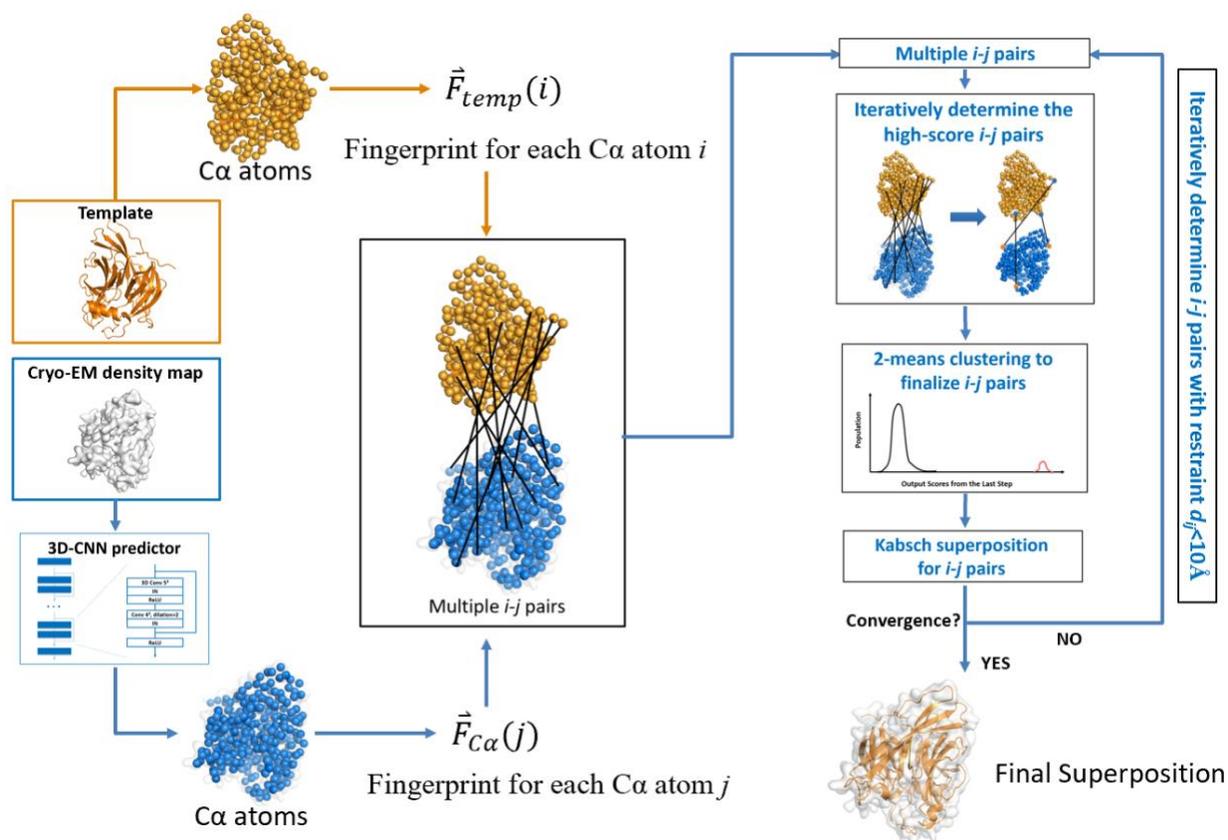
Supplementary Figure 12. Pipeline of template reselection, in which the templates will be superpose into the cryo-EM density map and re-ranked automatically. The procedure for iterative i-j atom pair determination between template and 3D-CNN conformation is further explained in Supplementary Figure 10.

a

Initialize $(w(i)=1, \forall i \in template)$ for each Ca atom in template; input the pairing mapping between ith Ca atom in template and jth predicted Ca atom $pair(i)=j$ by calculating $Fscore_{ij}$

↓

For atom i in template, calculate Euclidean distances $d(i, ii)$ between Ca atoms i and ii in template

↓

Calculate the Euclidean distances $d(j, jj)$ between the Ca atoms j and jj in predicted Ca atoms file corresponding to Ca atoms i and ii in template

↓

$$S(i,j) = \sum_{\substack{ii \neq i \\ jj \neq j}} \begin{cases} W(i,ii) & if\,|d(i,ii)-d(j,jj)| \leq 1 \\ \dfrac{W(i,ii)}{(d(i,ii)-d(j,jj))^2} & if\,|d(i,ii)-d(j,jj)| > 1 \end{cases}$$

where $j = pair(i)$; $jj = pair(ii)$;
and $W(i,ii) = w(i) \cdot w(ii)$

$$SumW(i) = \sum_{ii \neq i} W(i,ii)$$

$$\overline{SumW} = \frac{1}{N_{temp}} \sum_{i} SumW(i)$$

↓

Update $w(i) = S(i)/\overline{SumW}$

↓

If $Max\left(\Delta \dfrac{S(i,j)}{SumW(i)}, \forall i \in template\right) \leq 10^{-5}$?

No → (loop back)

Yes ↓

Output score set $\{S\}$

b

Template Cα

Predicted Cα

Initialize weight=1 for each Ca atom

Calculate intra-distance $d(i,ii)$

Iteratively update values of $w(i)$

$$S(i,j) = \sum_{\substack{ii \neq i \\ jj \neq j}} \begin{cases} W(i,ii) & if\,|d(i,ii)-d(j,jj)| \leq 1 \\ \dfrac{W(i,ii)}{(d(i,ii)-d(j,jj))^2} & if\,|d(i,ii)-d(j,jj)| > 1 \end{cases}$$
where $j = pair(i)$; $jj = pair(ii)$;
and $W(i,ii) = w(i) \cdot w(ii)$

$$SumW(i) = \sum_{ii \neq i} W(i,ii)$$
$$\overline{SumW} = \frac{1}{N_{temp}} \sum_{i} SumW(i)$$
update $w(i) = S(i)/\overline{SumW}$

output score set $\{S\}$ when convergent

Initial paring $j=pair(i)$ from Fscore

Same color (number) indicates one pairing

Calculate intra-distance $d(j,jj)$

The grey (#4) pairing is detected with low $S(i,j)$

Supplementary Figure 13. Flowchart of the algorithm to iteratively determine the atom pair between LOMETS template and 3C-CNN predicted conformation. The Σ in the flowchart is the summation from 1 to N_{temp}, which is the number of Cα atoms on the template. (a) Text representation of the pairing scheme. (b) Graphic illustration of the same pairing scheme.
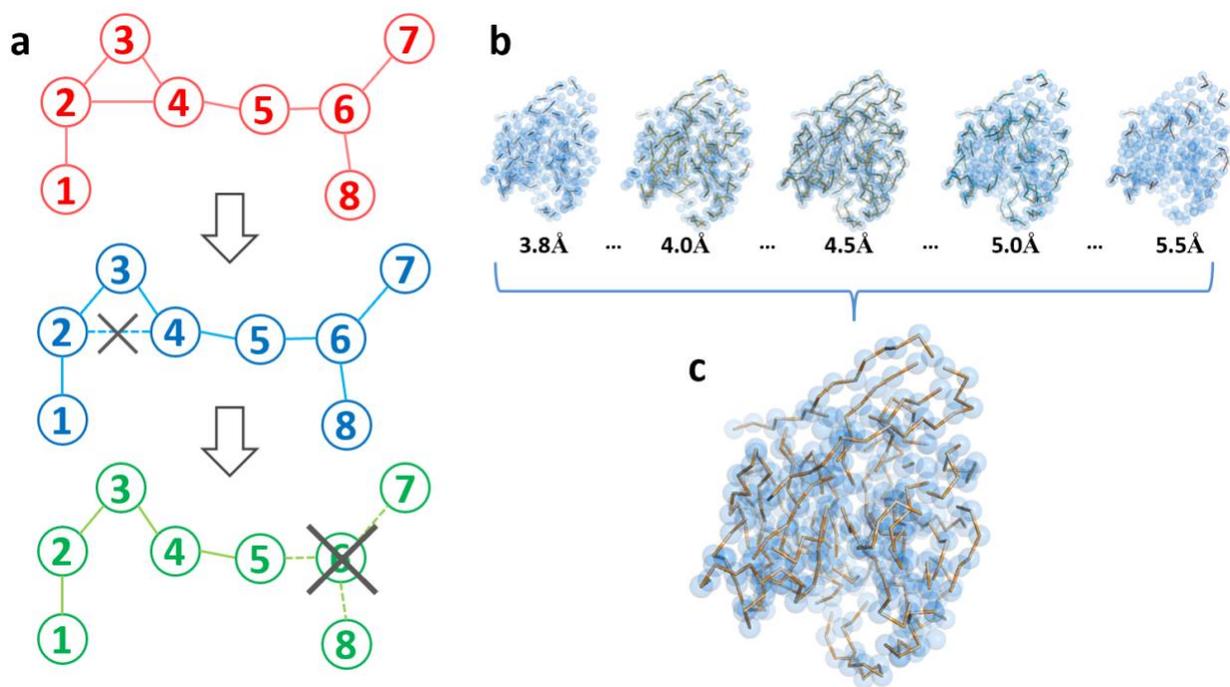
Supplementary Figure 14. Procedure for deducing order-dependent Cα traces from order-independent 3D-CNN $C\alpha$ conformations. (a) an illustrative example to define 'true' connections that have bond-length $<d_b$ and the number of connects to an atom is ≤ 2. Subsequently, the incorrect connections and atoms contradicting to the true connects are eliminated. (b) Examples of resultant connection patterns under different bond-length d_b cutoff. (c) An example of the final $C\alpha$ traces by connecting the continuous fragments from (b).
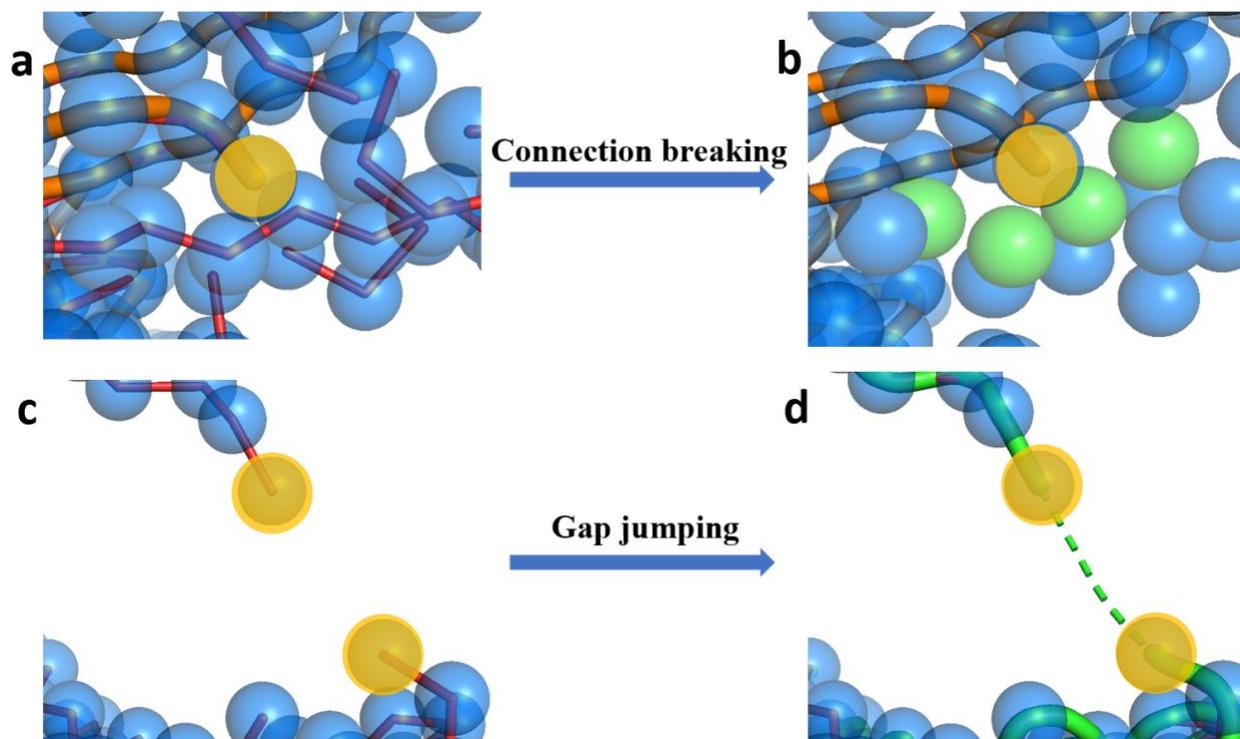
Supplementary Figure 15. Strategy of "keep-tracing mode". (a) Endpoint (orange) cannot find the next connection because the connection patterns around it are already settled down. (b) Breaking the connection around the endpoint (orange) so that atoms become available for connection (green). (c) Endpoint (orange) cannot find the next connection due to a large gap. (d) Gap jumping with a larger probing radius so that the endpoint can connect to the atom from the other side of the gap.
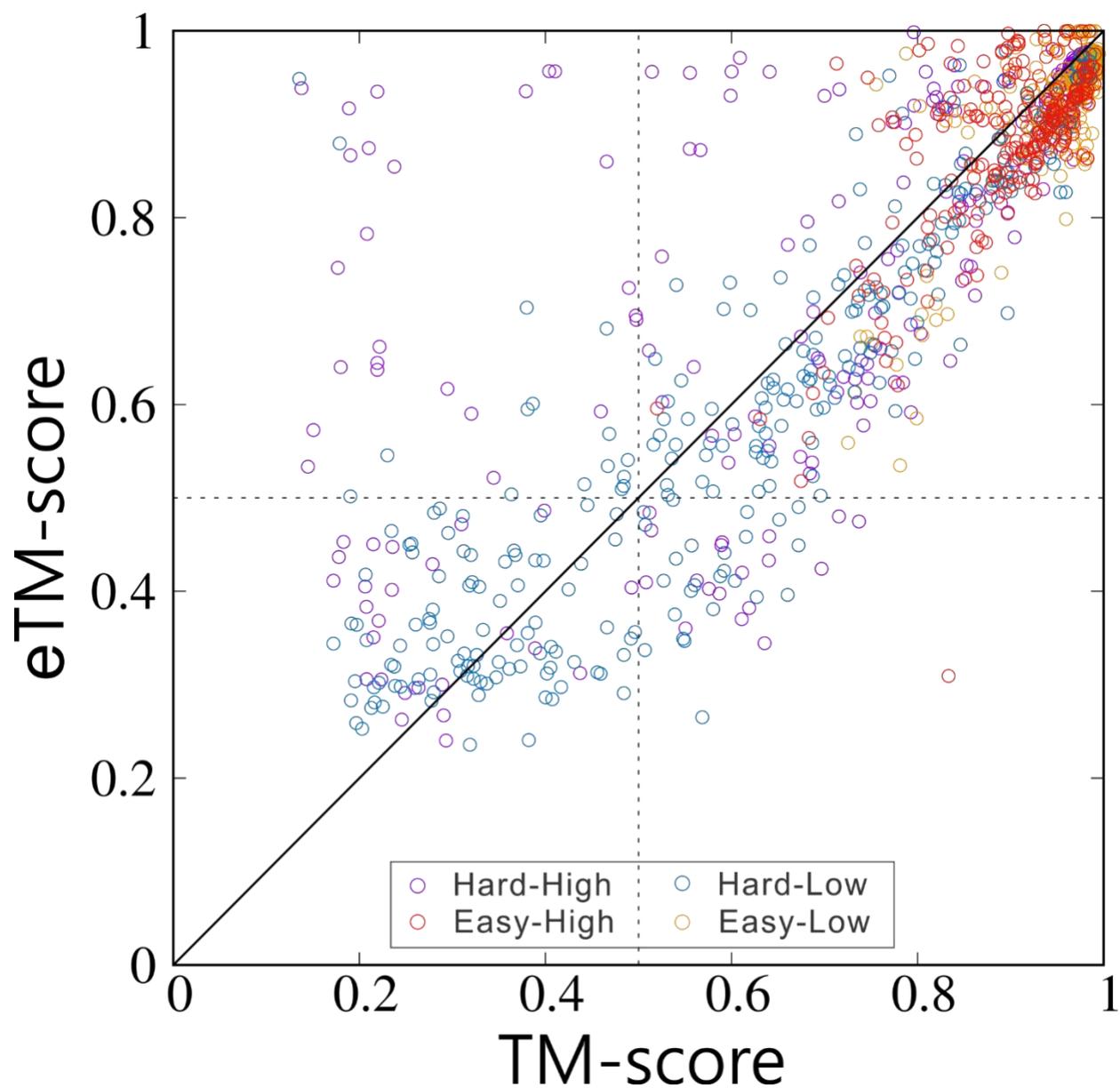
Supplementary Figure 16. Benchmark results of eTM-score versus TM-score calculated with native structure on 1,060 CR-I-TASSER models from different datasets/resolutions, including 301 Hard targets with high-resolution density maps (Hard-High, purple), 301 Hard targets with low-resolution density maps (Hard-Low, cyan), 229 easy targets with high-resolution density maps (Easy-High, red), and 229 easy targets with low-resolution density maps (Easy-Low, orange).
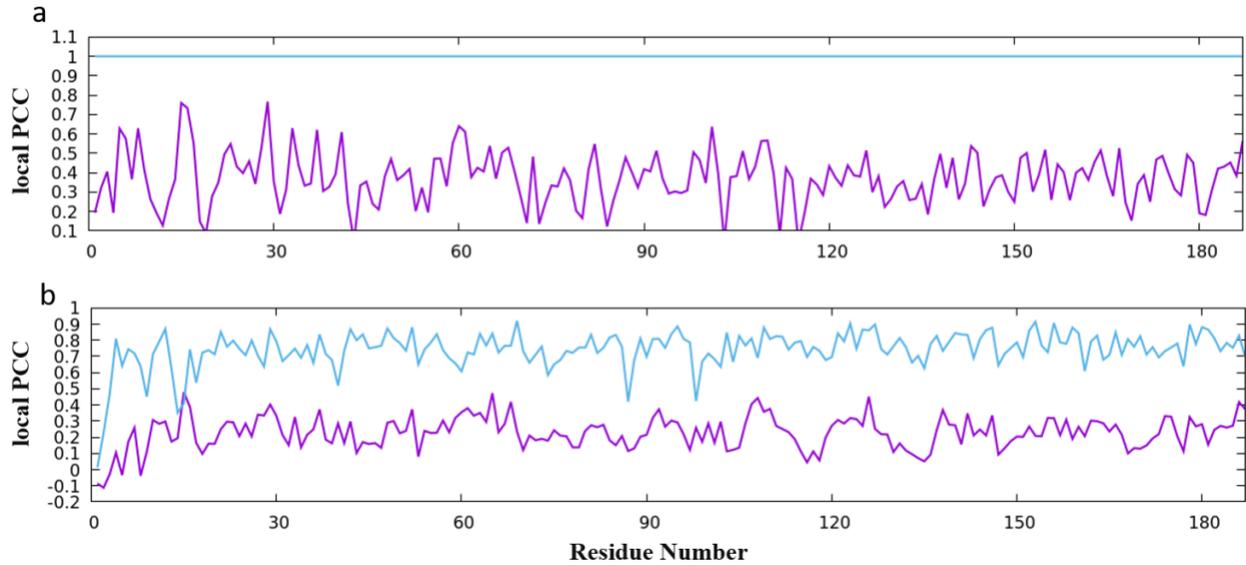
Supplementary Figure 17. Toy models to illustrate local PCCs computed with background density $\rho_0(\mathbf{y})$ (purple) and with $\rho'_m(\mathbf{y}, i)$ (blue) calculated from Eq. (11). (a) Native structure of a simulated map is used to calculate local PCC. Since the simulated background map is identical to the map calculated from the native structure, local PCCs should all equal to 1 as that computed with $\rho'_m(\mathbf{y}, i)$ (blue), while a zigzag line with relatively low PCC is obtained with $\rho_0(\mathbf{y})$ (purple). (b) Structure predicted by CR-I-TASSER (TM-score=0.770) from the background map is employed to calculate local PCC. The overall PCC between the predicted structure and the background map is 0.65, where the average local PCCs are 0.74 (blue, calculated with $\rho'_m(\mathbf{y}, i)$) and 0.23 (purple, calculated with $\rho_0(\mathbf{y})$). The closer average local PCCs calculated with $\rho'_m(\mathbf{y}, i)$) demonstrates that $\rho'_m(\mathbf{y}, i)$) is more reliable than directly using $\rho_0(\mathbf{y})$.
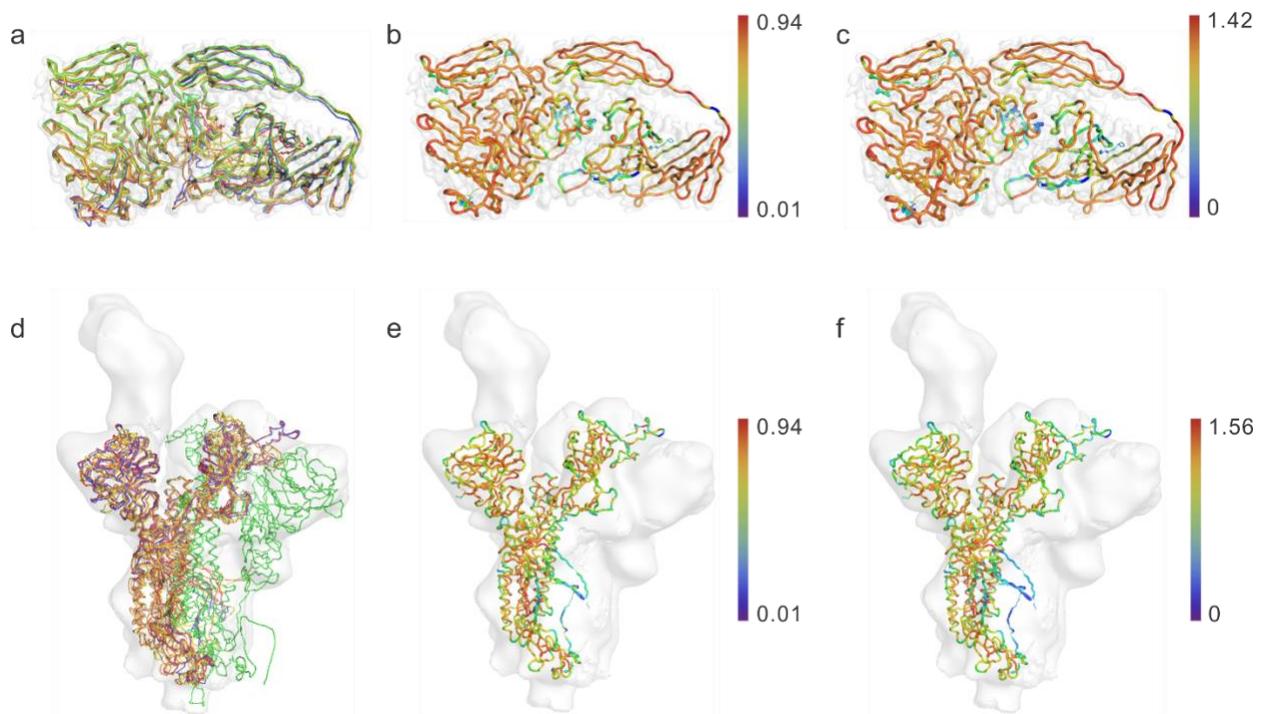
Supplementary Figure 18. Illustration of local PCC (LPCC) and local confidence (LC) scores on the two end-to-end study proteins in the manuscript. (a-c) 6tsk-B; (d-f) 7dk5; (a,d) display of the two-five proteins; (b,e) local PCC; (c,f) local confidence.

Supplementary References

1   Woetzel, N., Lindert, S., Stewart, P. L. & Meiler, J. BCL:: EM-Fit: rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. *Journal of structural biology* **175**, 264-276 (2011).
2   Yan, R., Xu, D., Yang, J., Walker, S. & Zhang, Y. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. *Sci Rep* **3**, 2619, doi:10.1038/srep02619 (2013).
3   Zhang, Y., Kolinski, A. & Skolnick, J. TOUCHSTONE II: a new approach to ab initio protein structure prediction. *Biophys J* **85**, 1145-1164, doi:10.1016/S0006-3495(03)74551-2 (2003).
4   Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579, doi:10.1002/prot.340230412 (1995).
5   Li, Y., Hu, J., Zhang, C., Yu, D. J. & Zhang, Y. ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. *Bioinformatics* **35**, 4647-4655, doi:10.1093/bioinformatics/btz291 (2019).
6   Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861-877 (2019).