



CR-I-TASSER: assemble protein structures from cryo-EM density maps using deep convolutional neural networks

Xi Zhang¹, Biao Zhang¹, Peter L. Freddolino^{1,2} and Yang Zhang^{1,2}

Cryo-electron microscopy (cryo-EM) has become a leading approach for protein structure determination, but it remains challenging to accurately model atomic structures with cryo-EM density maps. We propose a hybrid method, CR-I-TASSER (cryo-EM iterative threading assembly refinement), which integrates deep neural-network learning with I-TASSER assembly simulations for automated cryo-EM structure determination. The method is benchmarked on 778 proteins with simulated and experimental density maps, where CR-I-TASSER constructs models with a correct fold (template modeling (TM) score >0.5) for 643 targets that is 64% higher than the best of some other de novo and refinement-based approaches on high-resolution data samples. Detailed data analyses showed that the main advantage of CR-I-TASSER lies in the deep learning-based C α position prediction, which significantly improves the threading template quality and therefore boosts the accuracy of final models through optimized fragment assembly simulations. These results demonstrate a new avenue to determine cryo-EM protein structures with high accuracy and robustness covering various target types and density map resolutions.

Knowledge of three-dimensional (3D) structures of proteins is crucial for understanding their biological functions. Over the past decades, nuclear magnetic resonance (NMR) spectroscopy¹, X-ray crystallography² and electron microscopy (EM)³ have been widely used to obtain protein structures. However, NMR can only be used for relatively small proteins, whereas X-ray crystallography is often constrained by the difficulty of protein crystallization⁴. Although EM can overcome some of these limitations, it suffers from sample damage due to high-energy radiation, or low signal-to-noise ratio when very low electron doses are used⁵. The idea of cryogenic-EM (cryo-EM) was first proposed in the 1980s to reduce sample damage through frozen specimens⁶. Over the last decade, various technological innovations, such as single particle analysis and direct electron detection cameras^{5,7,8}, have made cryo-EM a practical means for probing protein structures without crystallization (X-ray) or size limitations (NMR). However, the success rate of cryo-EM is low with low-resolution density map data and more than half of cryo-EM samples in the EMDDataResource have no atomic structure determined⁹.

To help cryo-EM structure determination, a variety of computational structure modeling methods have been proposed, which can be generally categorized into two groups. The first group of approaches, such as Rosetta-Ref¹⁰, Flex-EM¹¹, iMODFIT¹², MDFF (molecular dynamics flexible fitting)¹³, Situs¹⁴ and EM-Refiner¹⁵, are built on structure refinement guided by correlations between the atomic model and cryo-EM maps. Despite the relative simplicity, most of the refinement programs require predefined model and map superposition, and the success rate critically depends on the quality of initial models and the superposition. The second group is referred to as de novo modeling, which constructs models from sequence and density map alone. One such example is Rosetta de novo (Rosetta-dn)^{16,17} that creates the initial model from a density map followed by RosettaES¹⁷ beam growing and Rosetta folding refinement. Another example is MAINMAST¹⁸ that constructs

initial backbone models from local dense points and then refines the models with the MDFF program¹³. Although these de novo approaches are capable of creating models from density maps alone, their success is highly sensitive to the resolution level of density maps. Additionally, methods such as MAINMAST require manual tuning and combination of multiple parameter-sets, rendering the programs less convenient for automated implementation.

We present a hybrid pipeline, CR-I-TASSER (cryo-EM iterative threading assembly refinement), for fully automated protein structure determination. While it is a de novo type approach in terms of creating models from sequence and density maps alone, CR-I-TASSER does use multithreading algorithms to identify homologous and analogous templates from the Protein Data Bank (PDB) to facilitate structural assembly. Technically, most existing de novo and refinement-based approaches rely on model-to-map correlations to guide the structural modeling simulations, but such correlation information is not precise and specific when the map resolution is low. In CR-I-TASSER, we extend deep residual convolutional neural networks (CNN)¹⁹ to create high-accuracy C α atom trace models from experimental density maps, providing a specific set of target atom positions that can be used to significantly improve threading template quality. In addition, the deep-learning boosted threading models are further assembled with cutting-edge I-TASSER folding simulations under the guidance of specific CNN models and the highly optimized I-TASSER knowledge-based force field²⁰. Our large-scale benchmark tests show a significant advantage of CR-I-TASSER over the traditional de novo and refinement-based approaches in assembling atomic cryo-EM protein structures. The online server and standalone package of CR-I-TASSER have been made publicly available at <https://zhanggroup.org/CR-I-TASSER/>.

Results

CR-I-TASSER is a hybrid method for determining atomic-level protein structures from cryo-EM density maps. As outlined in

¹Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA. ²Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI, USA. ✉e-mail: petefred@umich.edu; zhng@umich.edu

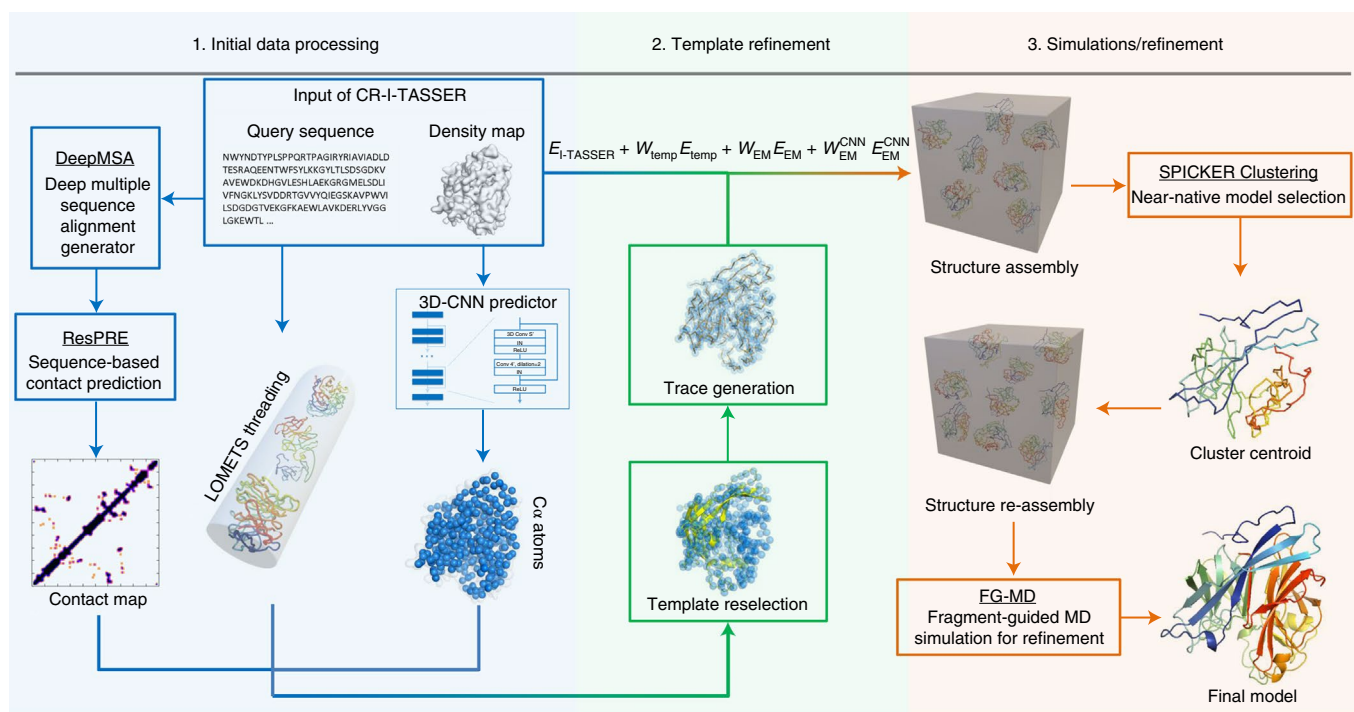


Fig. 1 | CR-I-TASSER pipeline. Starting with a query sequence and cryo-EM density map, CR-I-TASSER constructs atomic models through three consecutive steps: (1) initial data processing to generate 3D-CNN C α conformation, LOMETS threading and ResPRE contact-map prediction; (2) density map-based template reselection and trace generation and (3) density map-guided fragment reassembly simulations and model refinements.

Fig. 1, CR-I-TASSER starts with the creation of a sequence-order-independent C α conformation by deep convolutional neural-network (3D-CNN) training from density maps. The C α conformation is then used to improve the threading templates created by a local meta-threading server (LOMETS)²¹, using multiple heuristic iteration algorithms designed to match the query and template sequences with the C α conformation for template reselection and C α trace regeneration. Finally, the iterative threading assembly refinement method (I-TASSER²⁰) is extended to assemble atomic structure models under the guidance of both cryo-EM density map correlation and deep-learning boosted template restraints. Here, although CR-I-TASSER is built on I-TASSER and LOMETS²¹, the development of a deep-learning approach for cryo-EM-based C α atom prediction and the integration of sequence-order-independent C α models with advanced structure assembly methods represent the main novelty of the pipeline. Although there were previous efforts in applying deep-learning techniques to extract structural information from cryo-EM density maps^{22,23}, CR-I-TASSER marks the first pipeline using sequence-order-independent C α positions to improve threading alignments and regenerate order-dependent C α trace models, so that the deep-learning derived cryo-EM models can be directly used for guiding atomic-level structural assembly simulations. See Supplementary Text 1 for details of CR-I-TASSER datasets.

Density map-based C α significantly improve template quality. A key component of CR-I-TASSER is the deep neural network-based C α atom prediction from cryo-EM density maps, which is used to guide both template regeneration and structure folding simulations. Since the predicted C α atoms from 3D-CNN do not have indexes, we define CRscore to estimate the similarity between the predicted C α atoms and the native structure by

$$\text{CRscore} = \frac{1}{L} \sum_i \frac{1}{1 + \left(\frac{d_{ij}}{d_0}\right)^2} \quad (1)$$

where L is the target length, d_{ij} is the distance between i th atom in the 3D-CNN model and j th atom in the native structure, where the ij correspondence is established by a greedy method selecting the nonredundant ij pairs of the shortest distance (Supplementary Text 2). $d_0 = 1.24\sqrt[3]{N-15} - 1.8$ is a distance scale taken from the template modeling (TM) score to rule out length dependence²⁴. Here, the index information (and index connectivity) of both structures is completely ignored when computing the CRscore since we establish the ij correspondence by using their coordinate information only (Supplementary Text 2).

In Supplementary Fig. 1a, we list the average CRscore of 3D-CNN models on the 530 test proteins in different resolution ranges. The average CRscore is >0.95 when the resolution is high ($<5 \text{ \AA}$), but slightly decreases when the resolution becomes lower ($>10 \text{ \AA}$). This is consistent with the trend of root mean squared (r.m.s.d.) shown in Supplementary Fig. 1b, which is around $2\text{--}3 \text{ \AA}$ for high-resolution density maps but rises to $3\text{--}5 \text{ \AA}$ for low-resolution maps. As a comparison, we use an established algorithm, MAINMAST, which can generate C α locations from the density map. In addition, we create C α atom models by a naive greedy procedure that picks C α atom positions of the highest density values not in an excluded volume (Supplementary Text 3). As shown in Supplementary Fig. 1, the average CRscore and r.m.s.d. from our 3D-CNN C α models are considerably better than MAINMAST and the naive greedy procedure when resolution is high to medium ($1\text{--}8 \text{ \AA}$), and the scores become much better as the resolution drops, demonstrating the efficiency of the deep-learning training process for C α position prediction.

Using the 3D-CNN models, CR-I-TASSER creates two types of template by either density map-based template reselection or C α trace regeneration, followed by score reranking. In Supplementary Table 2, we compare TM scores of the templates from LOMETS with those after 3D-CNN-based refinement, where TM score is a metric defined to assess structural similarity of two structures, which has values ranging (0,1) with a higher value indicating closer similarity²⁴ (see Supplementary Text 4 for a more detailed

Table 1 | Modeling results by CR-I-TASSER and other methods on 778 benchmark test proteins involving different density map types and resolutions

Methods	TM score	N (TM > TM ₀) ^a	r.m.s.d. (Å)	P value
301 Hard targets with high-resolution density map (resolution in 2–5 Å) (TM ₀ = 0.5)				
I-TASSER ^b	0.345	27	12.0	8.0×10^{-91}
Flex-EM ^c	0.318	22	12.4	3.8×10^{-96}
iMODFIT ^c	0.340	25	11.9	6.6×10^{-91}
MDFF ^c	0.331	26	12.1	3.4×10^{-91}
EM-Refiner ^c	0.315	18	12.2	6.9×10^{-96}
Rosetta-Ref ^c	0.297	30	14.0	1.2×10^{-99}
MAINMAST ^d	0.438	121	10.2	9.8×10^{-47}
Rosetta-dn ^d	0.419	94	12.2	8.7×10^{-52}
Phenix ^d	0.466	134	8.6	8.7×10^{-42}
CR-I-TASSER ^d	0.772	251	4.4	–
229 Easy targets with high-resolution density map (resolution in 2–5 Å) (TM ₀ = 0.9)				
I-TASSER ^b	0.762	16	5.1	8.4×10^{-75}
Flex-EM ^c	0.824	66	4.4	4.6×10^{-35}
iMODFIT ^c	0.799	43	4.7	5.6×10^{-48}
MDFF ^c	0.857	104	4.1	4.8×10^{-21}
EM-Refiner ^c	0.846	76	4.0	3.5×10^{-37}
Rosetta-Ref ^c	0.851	103	4.0	6.9×10^{-21}
MAINMAST ^d	0.439	9	11.8	5.7×10^{-78}
Rosetta-dn ^d	0.474	17	12.0	8.0×10^{-77}
Phenix ^d	0.493	8	8.4	1.4×10^{-76}
CR-I-TASSER ^d	0.950	198	1.4	–
301 Hard targets with low-resolution density map (resolution in 5–15 Å) (TM ₀ = 0.5)				
I-TASSER ^b	0.345	27	12.0	2.0×10^{-48}
Flex-EM ^c	0.303	13	12.3	1.2×10^{-61}
iMODFIT ^c	0.316	23	12.0	2.0×10^{-56}
MDFF ^c	0.319	29	11.8	6.8×10^{-55}
EM-Refiner ^c	0.305	19	12.1	2.3×10^{-60}
Rosetta-Ref ^c	0.268	18	13.9	1.6×10^{-70}
MAINMAST ^d	0.204	3	14.3	2.1×10^{-86}
Rosetta-dn ^d	0.201	7	14.6	6.7×10^{-91}
Phenix ^d	0.180	0	12.5	5.5×10^{-95}
CR-I-TASSER ^d	0.597	191	6.3	–
229 Easy targets with low-resolution density map (resolution in 5–15 Å) (TM ₀ = 0.9)				
I-TASSER ^b	0.762	16	5.1	8.4×10^{-75}
Flex-EM ^c	0.666	0	5.3	3.5×10^{-90}
iMODFIT ^c	0.767	34	4.4	4.0×10^{-29}
MDFF ^c	0.788	46	4.3	5.5×10^{-23}
EM-Refiner ^c	0.739	21	4.7	5.3×10^{-42}
Rosetta-Ref ^c	0.714	14	4.9	7.5×10^{-49}
MAINMAST ^d	0.202	0	15.6	5.7×10^{-311}
Rosetta-dn ^d	0.225	1	9.2	1.5×10^{-238}
Phenix ^d	0.174	0	13.8	3.2×10^{-309}
CR-I-TASSER ^d	0.898	137	2.1	–

Continued

Table 1 | Modeling results by CR-I-TASSER and other methods on 778 benchmark test proteins involving different density map types and resolutions (continued)

Methods	TM score	N (TM > TM ₀) ^a	r.m.s.d. (Å)	P value
178 targets with experimental density map (resolution in 2–5 Å) (TM ₀ = 0.9)				
I-TASSER ^b	0.647	6	8.3	4.0×10^{-15}
Flex-EM ^c	0.681	24	8.5	3.6×10^{-9}
iMODFIT ^c	0.695	19	7.8	6.8×10^{-8}
MDFF ^c	0.709	37	7.3	4.9×10^{-6}
EM-Refiner ^c	0.690	32	8.3	2.5×10^{-7}
Rosetta-Ref ^c	0.688	40	8.5	7.1×10^{-7}
MAINMAST ^d	0.323	2	15.2	7.4×10^{-72}
Rosetta-dn ^d	0.353	5	15.7	1.4×10^{-60}
Phenix ^d	0.349	1	13.3	2.7×10^{-63}
CR-I-TASSER ^d	0.810	75	4.9	–
70 targets with experimental density map (resolution in 5–10 Å) (TM ₀ = 0.5)				
I-TASSER ^b	0.612	49	9.2	2.7×10^{-3}
Flex-EM ^c	0.546	45	9.3	4.3×10^{-7}
iMODFIT ^c	0.603	48	8.9	1.7×10^{-3}
MDFF ^c	0.573	46	8.7	5.9×10^{-5}
EM-Refiner ^c	0.576	45	8.8	9.7×10^{-5}
Rosetta-Ref ^c	0.554	43	9.3	9.7×10^{-6}
MAINMAST ^{d,e}	0.221	0	16.1	2.0×10^{-31}
Rosetta-dn ^c	0.176	1	15.6	5.4×10^{-41}
Phenix ^c	0.118	0	18.3	1.5×10^{-43}
CR-I-TASSER ^c	0.714	63	6.2	–

P values are calculated using two-tailed Student's t -tests between the TM scores produced by CR-I-TASSER and the other methods. Numbers in bold font highlight the performer that obtained the best average result in each category. ^aTM₀ is 0.5 for simulated Hard targets and low-resolution experimental targets, and 0.9 for simulated Easy targets or high-resolution experimental targets. ^bProtein structure prediction methods. ^cCryo-EM-based structure refinement methods. ^dCryo-EM-based de novo structure modeling methods. ^eOnly 61 targets are solved with MAINMAST, probably due to the low resolution and experimental noise.

description of TM score). In general, 3D-CNN makes the largest improvement for Hard targets in which C α traces deduced from 3D-CNN models have a substantially higher TM score (0.690 and 0.527 with high- and low-resolution density maps, respectively) than that of the original LOMETS (0.283). Combining both Easy and Hard targets, the TM score of the first models by 3D-CNN (0.707) is 45% higher than that of the original LOMETS (0.487), which corresponds to $P = 1.3 \times 10^{-174}$ using Student's t -test, showing that the template quality improvement brought by 3D-CNN is statistically highly significant.

CR-I-TASSER on high-resolution simulated density maps. To examine the efficiency of the CR-I-TASSER pipeline, we first apply it to the 301 Hard targets from our benchmark set that lack homologous templates in the PDB. Overall, CR-I-TASSER creates models with average TM score of 0.772 and r.m.s.d. of 4.4 Å. If we count the targets with TM score >0.5, which corresponds to a model with the correct fold²⁵, CR-I-TASSER creates correct folds for 251 targets, which is 9.3 times that obtained by I-TASSER (27, Table 1), showing a major impact of cryo-EM density maps on I-TASSER-based structure modeling.

As a comparison, in Table 1 (rows 9–11) we list the results from three de novo programs, MAINMAST¹⁸, Rosetta-dn^{16,17}

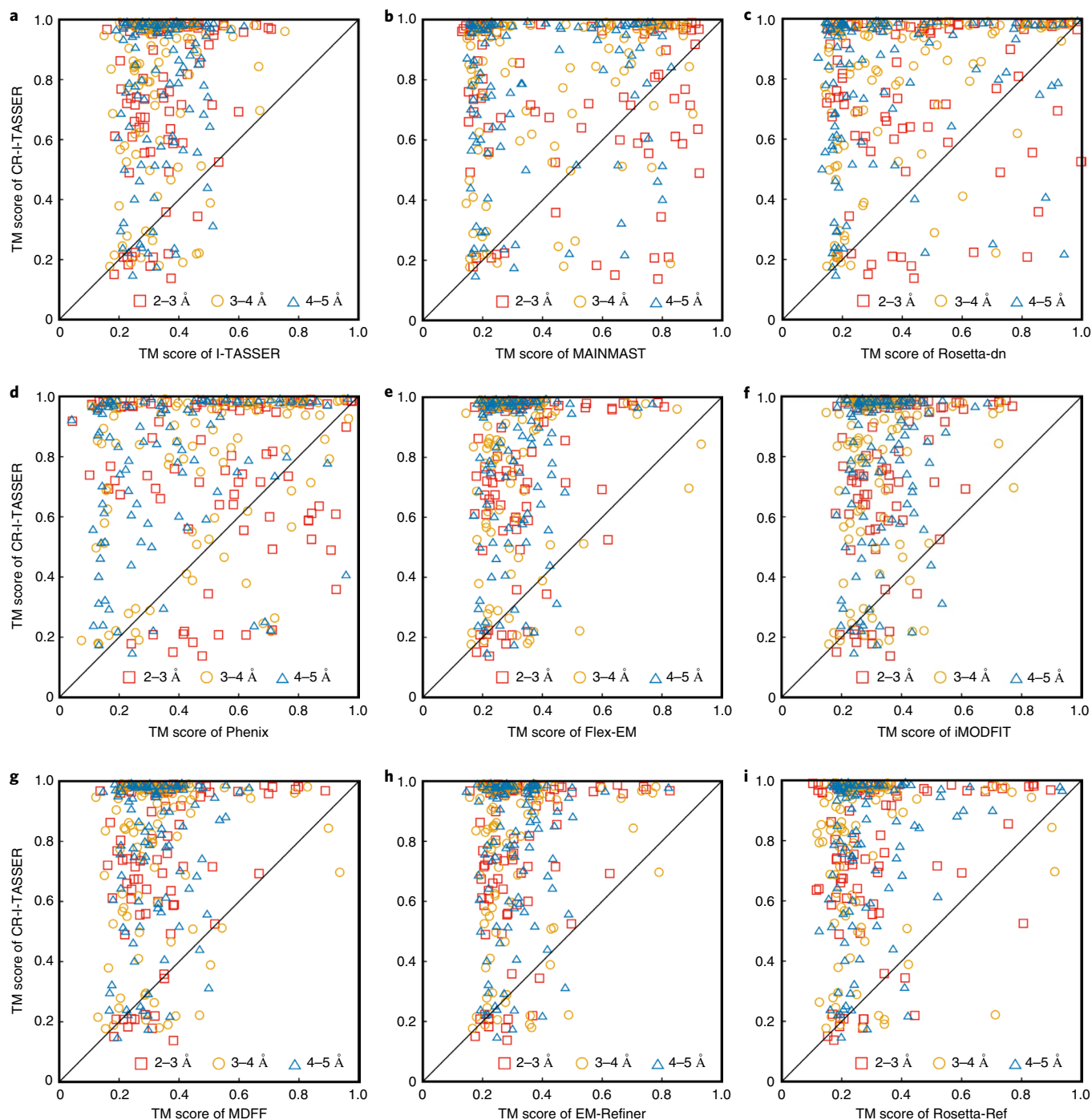


Fig. 2 | TM score comparisons of CR-I-TASSER with I-TASSER and eight other control methods on 301 Hard targets with 2–5 Å resolution simulated density maps. **a–i**, CR-I-TASSER versus I-TASSER (**a**); MAINMAST (**b**); Rosetta-dn (**c**); Phenix (**d**); Flex-EM (**e**); iMODFIT (**f**); MDFF (**g**); EM-Refiner (**h**) and Rosetta-Ref (**i**). The symbols with different colors and shapes denote different ranges of resolution: red squares, 2–3 Å; yellow circles, 3–4 Å and blue triangles, 4–5 Å.

and Phenix²⁶, which create models from the same set of density map data (see Supplementary Texts 5–7 for setting). It shows that CR-I-TASSER outperforms these programs substantially, with the average TM score 76% higher than MAINMAST (0.438), 84% higher than Rosetta-dn (0.419) and 66% higher than Phenix (0.466). In Fig. 2b–d, we present a head-to-head TM score comparison of CR-I-TASSER with the three control programs, where CR-I-TASSER has a higher TM score in 259, 270 and 252 cases than MAINMAST, Rosetta-dn and Phenix, and the competing

programs outperform CR-I-TASSER do so only in 42, 31 and 49 cases, respectively. In Fig. 2e–i, we also list the modeling results by five state-of-the-art cryo-EM refinement programs from Flex-EM¹¹, iMODFIT¹², MDFF¹³, EM-Refiner¹⁵ and Rosetta-Ref¹⁰, which start with the I-TASSER models after superposition of the density maps using Situs¹⁴ (Supplementary Texts 8–12). Overall, the refinement programs do not work well for the Hard targets, where their TM scores are even lower than those of the initial I-TASSER models for

the Hard proteins that have an average TM score of 0.345. This result is consistent with a previous observation¹⁵, which showed that the correlation between model quality and model-to-map correlation coefficient vanishes when the TM score of the initial models <0.5 , and therefore there is no sufficient correlation coefficient gradient to guide the programs for refining structures. We also benchmarked CR-I-TASSER on 229 Easy targets, where it outperforms other control groups with a significantly higher TM score (0.949; $P < 10^{-20}$ in all cases, Student's *t*-test). Details can be found in Supplementary Text 13.

In addition to the global structure quality listed in Table 1, we also calculate the local structure scores, including clashes and Molprobity²⁷, in Supplementary Table 3. CR-I-TASSER achieves the second-best clash and Molprobity scores following Rosetta-Ref, indicating that the CR-I-TASSER models have a reasonable local structure quality. Moreover, we demonstrate that improvement of template quality plays a critically important role in CR-I-TASSER structure assembly (Supplementary Text 14), and benchmarked CR-I-TASSER under Gaussian noises added by Xmipp²⁸ (see Supplementary Texts 15 and 16 for details). Furthermore, in Supplementary Fig. 3, we present an illustrative example from polyomavirus VP1 pentamer protein (PDB ID 1vps-A), which demonstrates that the template regeneration process can create high-quality templates from the 3D-CNN C α traces and result in much-improved full-length structure models, even though the initial threading templates are completely incorrect (see Supplementary Text 17 for details).

CR-I-TASSER on low-resolution simulated density maps. While cryo-EM experiments are now achieving increasingly good resolutions, it is still of importance to model structures from medium- and low-resolution density maps, especially for molecules with high flexibility or conformational/compositional heterogeneity⁵. In Table 1 (rows 25–34), we examine the performance of CR-I-TASSER on the 301 Hard proteins with resolution ranging from 5 to 15 Å. Compared to the models with high-resolution density maps (2–5 Å), the overall performance of CR-I-TASSER is reduced in the low-resolution set with an average TM score of 0.597; this is mainly due to the reduction of the 3D-CNN C α model quality with lower map resolution, as shown in Supplementary Fig. 1. Nevertheless, the TM score of CR-I-TASSER is substantially higher than the de novo programs by MAINMAST (0.204), Rosetta-dn (0.201) and Phenix (0.180), as well as the refinement programs by Flex-EM (0.303), iMODFIT (0.316), MDFF (0.319), EM-Refiner (0.305) and Rosetta-Ref (0.268). A similar trend can be found on the 229 Easy targets as summarized in Table 1 (rows 36–45); see Supplementary Text 18 for details.

In Supplementary Fig. 4a,b, we list a head-to-head TM score comparison of CR-I-TASSER with the best de novo and refinement programs, where CR-I-TASSER outperforms MAINMAST and MDFF in 296 and 265 cases, respectively, while the competing programs outperform CR-I-TASSER in only five out of 36 cases, respectively. If we count the number of cases with TM score >0.5 , CR-I-TASSER constructs the correct fold for 191 out of the 301 targets, which is 63 times that by MAINMAST (3) and 7.3 times that by MDFF (26). As an illustration, we present in Supplementary Fig. 4c–h the modeling results on Q6MIM9 from *Bdellovibrio bacteriovorus*, which highlights that the hybrid effects of both template reselection and regeneration processes, as well as the optimized structure assembly simulations, make a major contribution to the modeling of a difficult target with very low-resolution density maps (Supplementary Text 19).

Overall, although the average TM score of CR-I-TASSER drops for low-resolution maps in 530 Hard or Easy targets, the magnitude of the TM score reduction for CR-I-TASSER (by 17% from 0.849 to 0.727) is much smaller than that of the other de novo

methods, including MAINMAST (54%), Rosetta-dn (53%) and Phenix (73%). Even with the low-resolution maps, the average TM score of CR-I-TASSER is 87% higher than that of the second-best method (MDFF) for Hard targets, and 14% (299%) higher than other refinement-based (de novo) methods for Easy targets. This advantage on low-resolution data modeling is mainly attributed to the integration of multithreading alignments and the deep C α trace learning with the Broyden–Fletcher–Goldfarb–Shanno and Monte Carlo assembly simulations, which makes CR-I-TASSER a robust pipeline for a wide range of map densities.

Structure modeling on experimental density maps. To examine our pipeline in a realistic setting, we further tested CR-I-TASSER on 248 nonredundant proteins with experimental density maps (see Supplementary Text 1 for details of the dataset). On average, CR-I-TASSER achieves an average TM score of 0.783 for the 248 EMDDataResource targets, which is 158% higher than the best de novo program Rosetta-dn (0.303) and 17% higher than the best refinement program MDFF (0.671). In Fig. 3, we present a head-to-head comparison of CR-I-TASSER with I-TASSER and other control programs, where CR-I-TASSER outperforms the control methods (including I-TASSER) in most of the cases. Especially, CR-I-TASSER outperforms the sequence-based I-TASSER method in 228 out of 248 cases (92%). The average TM score of CR-I-TASSER (0.783) is 23% higher than that of I-TASSER (0.637), which corresponds to a $P = 3.8 \times 10^{-6}$ in Student's *t*-test, showing significant impact of the introduction of cryo-EM data in the cutting-edge structure assembly simulations. If we count the number of cases with TM score >0.5 or 0.9 for low- or high-resolution targets, CR-I-TASSER achieves good predictions in 138 cases, which is 23 and 1.7 times that by the best de novo program (Rosetta-dn, 6) and the best refinement program (MDFF, 83), respectively. In the bottom of Table 1 (rows 46–67), we split the data samples into high- and low-resolution, where a similar trend of the superiority of CR-I-TASSER over other methods is seen. The gap between CR-I-TASSER and the comparison methods, as assessed by $\Delta\text{TM} = \text{TM score}_{\text{CR-I-TASSER}} - \text{TM score}_{\text{other}}$ is slightly larger for the low-resolution (0.543 / 0.141 for Rosetta-dn / MDFF) than the high-resolution samples (0.457 / 0.101), despite the fact that all methods perform better for high- rather than low-resolution samples. This is probably due to the fact that TM scores of the control methods for low-resolution samples are lower and therefore have more room for improvement. Furthermore, we specifically checked whether any particular secondary structure components would affect the performance of CR-I-TASSER. As shown in Supplementary Fig. 5, although CR-I-TASSER performs better in high- than low-resolution maps, there is no obvious correlation between the average TM score and the ratio of secondary components for both high- and low-resolution cases. More benchmark results (for example, template homology cutoff, different network training sessions, full maps and so on) can be found in Supplementary Text 20.

As a further case study focusing on difficult targets, we examine in detail a hard example from the anthrax toxin antigen pore protein (PDB ID 3j9c-A) in Fig. 4. This target consists of 423 residues and the cryo-EM density map has a resolution of 2.6 Å. In this case, LOMET's failed to locate good templates (the best template has a TM score of 0.257), which resulted in an incorrect fold of the final I-TASSER model with a TM score of 0.132. Therefore, the superposition between Situs is nearly random. Consequently, all refinement-based methods failed to model the target and have a final model with TM scores of 0.144, 0.132, 0.136, 0.143 and 0.153 for Flex-EM, iMODFIT, MDFF, EM-Refiner and Rosetta-Ref, respectively. As illustrated in Fig. 4a,d, the Rosetta-Ref model does not match the native structure both globally and locally. On the other hand, Phenix built a model from density map alone that fits the global conformation with the density map. However, there are multiple misconnections and disordered local structures in the

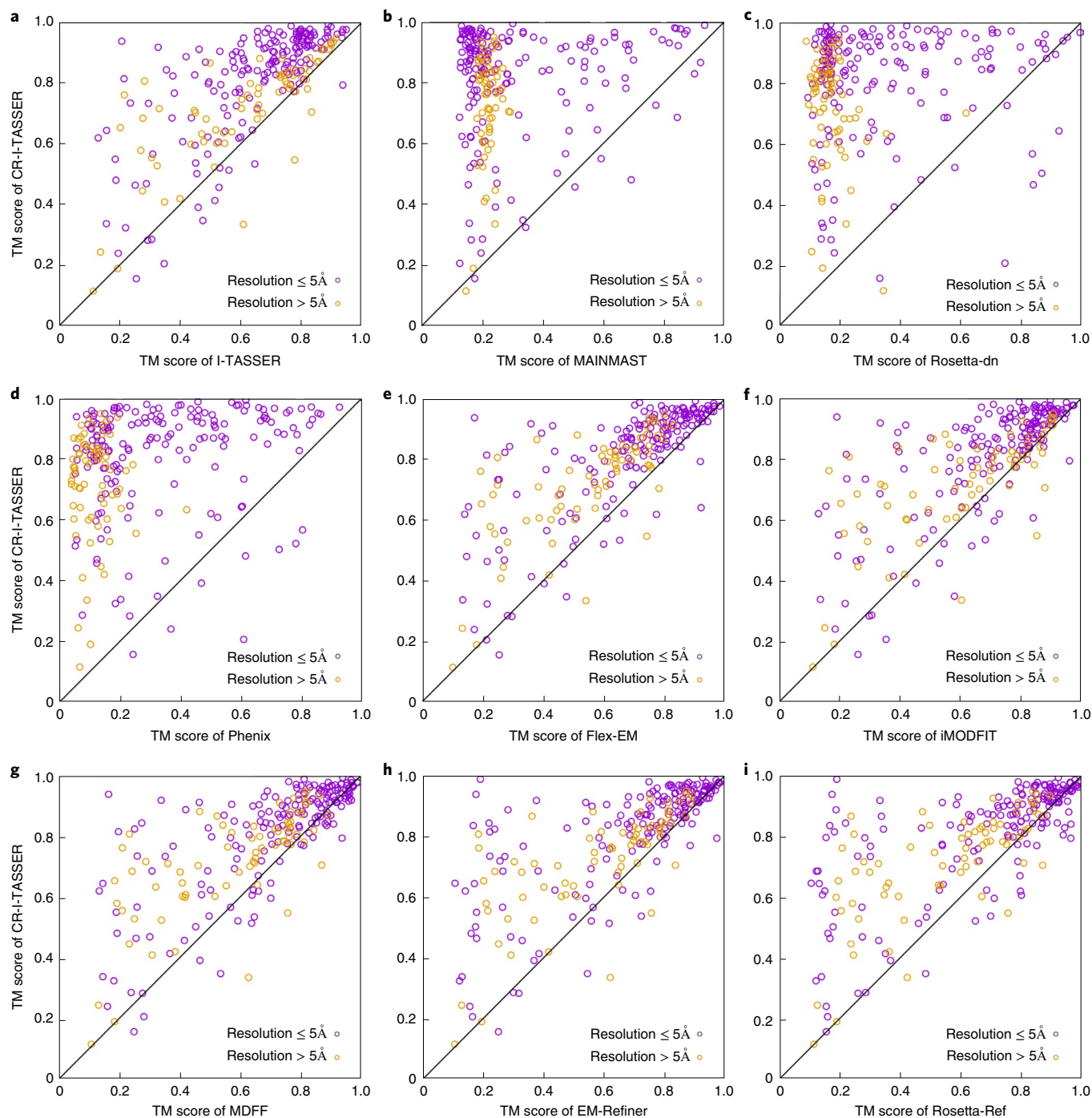


Fig. 3 | Modeling results on 248 targets with experimental density maps by different methods. CR-I-TASSER versus I-TASSER (a), MAINMAST (b), Rosetta-dn (c), Phenix (d), Flex-EM (e), iMODFIT (f), MDFF (g), EM-Refiner (h) and Rosetta-Ref (i). The symbols with different colors denote different ranges of resolution: purple, 2–5 Å and yellow, 5–10 Å.

model, resulting in an incorrect topology and sequence mapping with a TM score of 0.274 (Fig. 4b,e). Similar results were obtained by MAINMAST and Rosetta-dn with TM scores of 0.165 and 0.245, respectively.

Given the high resolution of the density map, 3D-CNN generated a well-predicted $\text{C}\alpha$ conformation with CRscore of 0.947. Benefiting from this high-quality prediction, the template regeneration algorithm created a reasonable $\text{C}\alpha$ trace model with TM score of 0.534. Following the CR-I-TASSER reassembly, the final model achieves a TM score of 0.725 for the head globular domain

(Fig. 4c) and TM score of 0.620 for the overall chain (Fig. 4f), which are both higher than that by all template and cryo-EM-based modeling programs.

It is notable that the TM score of the sequence-ordered $\text{C}\alpha$ trace model in CR-I-TASSER is considerably lower than the CRscore calculated from the order-independent $\text{C}\alpha$ conformation in the anthrax toxin antigen pore protein case. This is mainly due to the extreme complexity of target structure consisting of a three-domain globular head flanked with a long beta-hairpin stem that form an antigen pore with other homo-chains; such structural complexity

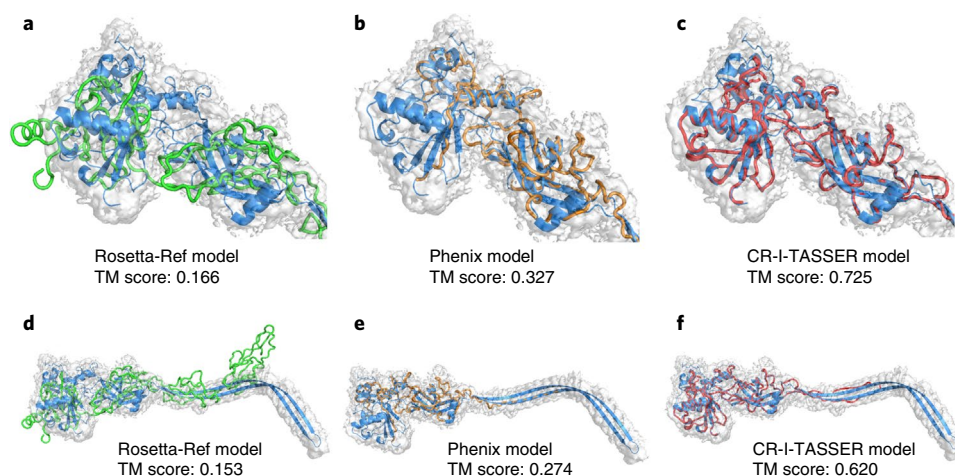


Fig. 4 | Structure modeling results on a protective antigen pore protein (PDB ID 3j9c-A) with high-resolution (2.9 Å) density map. a–c, Predicted models by Rosetta-Ref (green) (**a**), Phenix (orange) (**b**) and CR-I-TASSER (red) (**c**) are shown along with the native structure on the head globular domain (residues 1–98; 185–423, blue). **d–f,** The corresponding full-length models including the stem region: Rosetta-Ref (**d**), Phenix (**e**) and CR-I-TASSER (**f**). The predicted C α conformations and connection pattern can be found in Supplementary Fig. 8.

not only introduces noise to C α position predictions due to the high flexibility of the long stem, but also results in a huge conformational space of fragment connection patterns, which makes the true backbone difficult to trace. As shown in Supplementary Fig. 8, there are many mis-predicted C α atoms around the long stem. Additionally, the connection conformational space is huge because the two long beta strands are close to each other, making it hard for the fragment-tracing program to interpret the correct connection patterns. Given the specific local structures, however, this issue could be amended by using the density map-based secondary structure prediction models because the backbone conformational space could be substantially reduced by excluding the zigzag connection patterns in the predicted beta zone. A separate computational pipeline implementing real-space secondary structure prediction powered with deep learning is currently under development, which may in the future highly benefit modeling for targets with extremely low-resolution maps as well.

End-to-end studies on protein complexes EMD-10564/EMD-30703. As end-to-end case studies from raw density map to final structure, we first present an illustrative example in Fig. 5a–f and Supplementary Figs. 9a–c for a large-size homo-tetramer complex Beta-galactosidase (PDB ID 6tsk), with each chain consisting of 1,040 residues. The corresponding density map EMD-10564 has a resolution of 2.3 Å and is segmented by Phenix segment_and_split_map that has been integrated in the CR-I-TASSER pipeline (Supplementary Text 22), resulting in a reasonable segmentation model as shown in Supplementary Fig. 9a. Here, we construct four models from the four segmented density maps separately and look specifically into chain A. As shown in Supplementary Fig. 9b, 3D-CNN creates a high-quality C α model with CRscore of 0.946, which is subsequently used for template reranking and selection from the LOMETS alignment pool (outlined in Supplementary Fig. 12) and for C α trace generation with the C α trace connection algorithm (outlined in Supplementary Fig. 14). In this case, the best template with a TM score of 0.666 was identified by both LOMETS and the predicted C α trace conformation, as shown in Supplementary Fig. 9c. However, the rest of the threading templates are not as good as the best one, resulting in an average TM score of 0.446 for the top 40 LOMETS templates. By combining the template reranking and C α trace generation processes, CR-I-TASSER improved the TM score from 0.446 to 0.513 for the top 40 templates.

These templates are submitted to the structural assembly simulations that are guided by the restraint-enhanced I-TASSER force field and the density map correlations. Eventually, CR-I-TASSER constructed the final model with TM score of 0.705 (Fig. 5c), which is 41% higher than that of the original I-TASSER prediction (0.500). Due to the size and complexity of the model, Situs does not correctly superpose the I-TASSER model into the density map, resulting in the general low quality from the refinement-based programs with TM scores of 0.476, 0.474, 0.343, 0.359 and 0.353 for Flex-EM, iMODFIT, MDFF, EM-Refiner and Rosetta-Ref, respectively. Meanwhile, the de novo programs that we tested are also unsuccessful in creating correct folds because of the complexity of tracing/building such a large protein, resulting in final TM scores of 0.194, 0.105 and 0.251, for MAINMAST, Rosetta-dn and Phenix, respectively.

Although CR-I-TASSER successfully built a model with the highest TM score among the state-of-the-art programs, there is still room for improvement. In fact, the final model in Fig. 5c shows that the structure of the three domains in the left side of the picture is very close to the native, but that for the remaining two domains in the right side is poor. This is partly because the correct LOMETS alignments are mostly located in the left domains. However, the connection patterns of the C α trace model shown in Fig. 5a overlaps well with the target structure, indicating the connections are mostly correct. A closer view shows that there are several small flaws of misconnections in beta sheets of the right part, where these misconnections can terminate the growth of the long traces as the target atoms may be out of the probing radius of the last C α atom, as shown in the zoom-in figure in Fig. 5b. The probing radius request is used as the default in CR-I-TASSER to ensure the reasonability of the C α tracing models for general sequences. Nevertheless, if we use the option of ‘keep-tracing mode’ provided in the CR-I-TASSER pipeline, which allows for the end point of current trace to break the connection patterns (Supplementary Text 23), the created C α trace models are greatly improved with the average TM score increased from 0.446 to 0.708 for this case, where the TM score of the first template is improved from 0.666 to 0.749. These high-quality C α trace templates lead to a much-improved full-length model with TM score of 0.857 (Fig. 5e). Despite the improved performance for this case, the keep-tracing mode is not used as the default setting in CR-I-TASSER as the drop off of the probing radius could increase connection uncertainty and reduce the average performance for

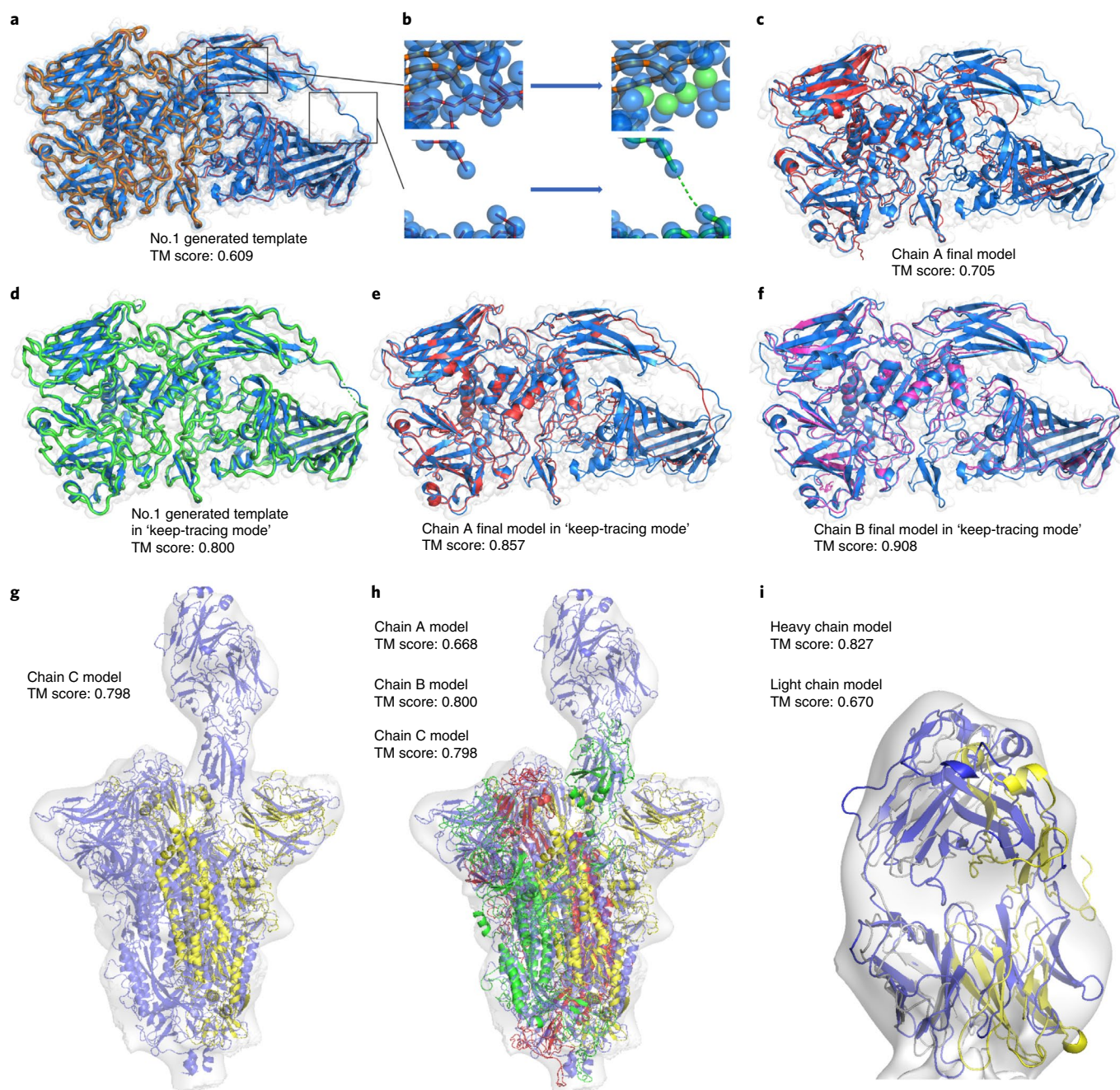


Fig. 5 | Illustrative examples of end-to-end structural modeling by CR-I-TASSER from unsegment maps. Through all pictures, native structures are shown in blue overlaid on density map in gray. **a–f**, Beta-galactosidase in complex with L-ribose (PDB ID 6tsk) from density map (EMD-10564, resolution 2.3 Å). **a**, Best C α trace model (orange) superposed with the native. **b**, Zoom-in pictures of breaking connections can be remedied by the 'keep-tracing mode' (see Supplementary Fig. 15 for details). **c**, Full-length model by CR-I-TASSER with default setting (red). **d**, C α trace model generated with 'keep-tracing mode' (green). **e**, Full-length model by CR-I-TASSER in 'keep-tracing mode' (red). **f**, Full-length model with the highest eTM score among four chains (magenta). **g–i**, the SARS-CoV-2 spike protein with RBDs bound with a 2H2 Fab (PDB ID 7dk5) from a density map (EMD-30703, resolution 13.5 Å). **g**, First CR-I-TASSER model (yellow) built on the map as in the chain C location. **h**, Models of chains A (green), B (red) and C (yellow) built on the map. **i**, Final CR-I-TASSER models of heavy and light chains of 2H2 Fab (gold and silver) using the complex-based superposition process described in Supplementary Text 24.

normal proteins. Additionally, since we have separately modeled four segmented chains, we could choose a possibly better model by examining the estimated TM scores (see equation (8) in Methods), which are 0.777, 0.912, 0.834 and 0.856 for chains A, B, C and D, respectively. By selecting the model for chain B, we obtained the final full-length model with a TM score of 0.908 as shown in Fig. 5f.

Overall, this example demonstrates the practicality of CR-I-TASSER for generating high-quality models from unsegmented raw

density map data, but also exposes the potential weaknesses of the default CR-I-TASSER pipeline that is sometimes too conservative when generating C α traces for targets involving long loops/tails and disorder regions, where the keep-tracing mode may help provide an alternative solution for better C α tracing and final model constructions for these cases when the first try fails.

In Fig. 5g–i, we present another example of models built from a raw low-resolution density map (13.5 Å), which is for the

complex of the SARS-CoV-2 spike protein with a 2H2 Fab (PDB ID 7dk5). In this complex, three large homo-chains (each with 1,261 residues) are bound with the two heavy and light chains of a 2H2 Fab with 214 and 218 residues. Due to the low resolution, it is not feasible to automatically segment using only density map information. Thus, we attempted to build models on the whole map. Given that CR-I-TASSER performs better for the cases with a higher protein-map size ratio as shown in Supplementary Fig. 7b, we first tried to build a long spike protein chain in the map. In this case, LOMETS recognize the top-1 template with TM score of 0.562, where the CR-I-TASSER reranked the alignments and chose a better first-rank template with TM score of 0.671. As shown in Fig. 5g and Supplementary Fig. 9d, CR-I-TASSER superposed the first-rank template into the low-resolution density map correctly and built a final model with TM score of 0.798 to the deposited structure in the chain C position, where the model built by I-TASSER has only a TM score of 0.682. After that, the density map was masked by deleting the part that overlaps with the model just built. The remaining density map was then used by CR-I-TASSER to build the second and third spike chains subsequently by repeating this process. As shown in Fig. 5h and Supplementary Fig. 9e, CR-I-TASSER eventually built three spike protein models on the low-resolution map with TM scores of 0.668, 0.800 and 0.798 for the chain A (with an up receptor-binding domain, RBD) and chains B/C with down RBDs, respectively (compared to 0.599, 0.677 and 0.682 by I-TASSER). Although the resolution is low, CR-I-TASSER still assembles spikes with up and down RBD conformations in the correct position.

Following the long-chain structure modeling for the spike proteins, we further attempted to build models of the heavy and light chains of 2H2 Fab. Since these two chains are of similar lengths but not identical, it is hard to tell which one should be built first. By randomly selecting the heavy chain to start, CR-I-TASSER created models with TM scores of 0.702 and 0.518 for the heavy and light chains, respectively, which are marginally better than I-TASSER (TM scores of 0.524 and 0.571), where the positions of the two chains on the map are apparently incorrect (Supplementary Fig. 9f,g). The failure in improvement is partly because the native structures of these two chains share similar folds (TM score of 0.730 by TM-align²⁹), and hence they have very similar density maps, which make it harder to locate the correct position in such a low-resolution map. Instead of one-by-one modeling, a better strategy may be to introduce complex modeling. Here, we slightly extended the current pipeline to simultaneously superpose the templates from two chains and choose the best combination poses (see details in Supplementary Text 24). With this, good templates for both chains were correctly ranked and superposed in the density map as shown in Supplementary Fig. 9h. These templates were then submitted to CR-I-TASSER simulations separately, which resulted in the final models with higher TM scores (0.827 and 0.670 for heavy and light chains, Fig. 5i and Supplementary Fig. 9i). Despite the simplicity of the strategy, this result demonstrates the feasibility to extend CR-I-TASSER for complex-based structural modeling on full density maps.

Conclusion

We present a hybrid pipeline, CR-I-TASSER, for automated protein structure modeling from cryo-EM density maps. The core component of the pipeline is the density map-based C α trace predictions from 3D CNNs, which are used for threading template selection and initial model generations through fragment tracing. The advanced I-TASSER folding simulation platform is then extended to reassemble the template and C α trace models, under the guidance of an optimized force field combining 3D-CNN density map and template restraints with the advanced knowledge-based energy potentials.

CR-I-TASSER was benchmarked on a large-scale dataset containing 778 proteins with both computer-simulated and

experimental density maps, compared to three state-of-the-art de novo (Rosetta-dn^{16,17}, MAINMAST¹⁸ and Phenix²⁶) and five refinement-based (Flex-EM¹¹, iMODFIT¹², MDFF¹³, EM-Refiner¹⁵ and Rosetta-Ref¹⁰) methods. Overall, CR-I-TASSER generates models with an average TM score of 0.839 when high-resolution (2–5 Å) density maps are used, which is 88% higher than the best de novo modeling program (Phenix) and 41% higher than the best refinement program (MDFF), with $P < 10^{-66}$ using Student's *t*-test for both comparisons. When the medium-to-low resolution (5–15 Å) maps are used, although the average TM score of CR-I-TASSER is slightly reduced (0.726), it still generates the correct fold with a TM score > 0.5 for 482 cases, which is 66% higher than the best of other methods (289 by the MDFF program). Detailed data analyses showed that the density map-based deep-learning C α trace models from 3D-CNN play a critical role in the structure quality improvement. Since deep learning can derive specific and precise information on C α atoms from density map, the 3D-CNN C α trace models can therefore be used to more efficiently constrain both initial template regeneration and CR-I-TASSER model assembly simulations, compared to traditional de novo and refinement-based approaches that are guided solely by model-to-map correlations. Thus, CR-I-TASSER provides currently best-in-class performance for automated structure prediction from cryo-EM density maps.

Despite the encouraging results, it is important to note that the current CR-I-TASSER pipeline relies on the success of 3D-CNN on C α trace prediction, and we observe that the accuracy can decrease on low-resolution data. There are also issues in converting C α positions into ordered tracing models when the target structure involves long loops/tails or disordered regions. Given the exciting progress witnessed in hybrid deep-learning and evolution-based protein structure prediction^{30–32}, the combination of 3D-CNN with deep multiple sequence alignments (MSAs) collected from metagenome databases should help further improve the 3D-CNN C α trace and CR-I-TASSER model accuracy. Additionally, a new module of CR-I-TASSER aimed to further enhance its performance on low-resolution data is in development, in which we use density map-based real-space secondary structure modeling powered by deep neural-network learning to assist cryo-EM model construction. The preliminary result is encouraging and shows that since secondary structure is 'coarser' than C α positions, the models are easier to learn and can provide more relevant information to improve the modeling accuracy for the targets with poorer resolution maps. Meanwhile, CR-I-TASSER mainly focuses on monomer proteins, for which the density maps need to be segmented manually in the first place. We expect that it will be possible to combine CR-I-TASSER in a modular fashion with improved upstream or downstream tools for other modeling tasks (for example, segmentation or refinement) to further enhance future performance. Given that a chief advantage of cryo-EM is on large-size protein complex structure determination, however, an important next step is to extend the deep learning-based structure assembly simulations for protein–protein/protein–nucleic acid complex structure modeling and determination. While one of the current state-of-the-art segmentation programs has been integrated into CR-I-TASSER, new algorithms built on I-TASSER homology modeling and heuristic structure-map alignment iterations²⁹ can be a meaningful solution; investigations along these lines are in progress.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-021-01389-9>.

Received: 31 January 2021; Accepted: 17 December 2021;
Published online: 7 February 2022

References

1. Wuthrich, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**, 45–50 (1989).
2. Drenth, J. *Principles of Protein X-Ray Crystallography* (Springer Science & Business Media, 2007).
3. Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32 (1975).
4. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).
5. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2016).
6. Knapek, E. & Dubochet, J. Beam damage to organic material is considerably reduced in cryo-electron microscopy. *J. Mol. Biol.* **141**, 147–161 (1980).
7. Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
8. Glaeser, R. M. How good can cryo-EM become? *Nat. Methods* **13**, 28–32 (2015).
9. Lawson, C. L. et al. EMDDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403 (2016).
10. Wang, R. Y.-R. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).
11. Joseph, A. P. et al. Refinement of atomic models in high resolution EM reconstructions using Flex-EM and local assessment. *Methods* **100**, 42–49 (2016).
12. Lopez-Blanco, J. R. & Chacon, P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).
13. Singharoy, A. et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *eLife* **5**, e16105 (2016).
14. Kovacs, J. A., Galkin, V. E. & Wriggers, W. Accurate flexible refinement of atomic models against medium-resolution cryo-EM maps using damped dynamics. *BMC Struct. Biol.* **18**, 12 (2018).
15. Zhang, B., Zhang, X., Pearce, R., Shen, H.-B. & Zhang, Y. A new protocol for atomic-level protein structure modeling and refinement using low-to-medium resolution cryo-EM density maps. *J. Mol. Biol.* **432**, 5365–5377 (2020).
16. Wang, R. Y.-R. et al. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **12**, 335–338 (2015).
17. Frenz, B., Walls, A. C., Egelman, E. H., Veesler, D. & DiMaio, F. RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat. Methods* **14**, 797–800 (2017).
18. Terashi, G. & Kihara, D. De novo main-chain modeling for EM maps using MAINMAST. *Nat. Commun.* **9**, 1618 (2018).
19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
20. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
21. Zheng, W. et al. LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res.* **47**, W429–W436 (2019).
22. Si, D. et al. Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps. *Sci. Rep.* **10**, 4282 (2020).
23. Maddhuri Venkata Subramaniya, S. R., Terashi, G. & Kihara, D. Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning. *Nat. Methods* **16**, 911–917 (2019).
24. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
25. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
26. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 531–544 (2018).
27. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 12–21 (2010).
28. De la Rosa-Trevín, J. et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
29. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
30. Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).
31. Senior, A. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
32. Zheng, W. et al. Deep-learning contact-map guided protein structure prediction in CASP13. *Proteins Struct. Funct. Bioinf.* **87**, 1149–1164 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

CR-I-TASSER is a hierarchical method integrating I-TASSER with cryo-EM density maps for high-accuracy protein structure determination. As outlined in Fig. 1, the pipeline consists of three consecutive steps: (1) initial data processing; (2) deep learning-based template refinement and regeneration and (3) density map-guided structural reassembly simulations.

Initial data processing. Starting from query sequence and cryo-EM density map, CR-I-TASSER extracts three parts of information.

Predicting C α locations using deep neural-network learning. A 3D-CNN with a residual network architecture¹⁹ (see Supplementary Text 25 for details) is used to predict C α atom locations in a grid system, where the input of the 3D-CNN is the cryo-EM density map, and the output is the grid values ranging from 0 to 1 representing the possibility of C α atoms at the grids. The overall 3D-CNN architecture is shown in Supplementary Fig. 10a, where the density map in 3D grid space is taken as input signal to send through a 3D convolutional layer followed by instance normalization and the rectified linear unit, and extended to 32 channels. Next, ten basic blocks with residual network architecture are used to enhance the network capability of learning essential information of density maps. Eventually, the signal goes through the last layer that contains a 3D convolutional layer with two output channels and a SoftMax layer. The final outputs of 3D-CNN contain two complementary probability maps with the same size of the input density map, in which one map represents the probability of class 1 ('having a C α atom') while the other one stands for class 0 ('not having a C α atom'). Since only a few grids are with C α atoms around them, these two classes are highly imbalanced. Specifically, if we look at the central part (instead of marginal part) of the density maps where proteins are located, the ratio of the numbers of class 0 to 1 in experimental training set is 440,462,749 to 9,537,251, which is approximately 50 to 1 (Supplementary Table 1). Therefore, to make the training process more balanced, we set the weights as 1.0 and 50.0 for class 0 and 1, respectively, when computing the loss function, for which the cross-entropy loss is used. Although the weights are important in imbalance training and can affect the training process, the slightly different weights (for example, 1/25 or 1/75 for class 0/1) will have negligible effect on the final training result and hence we only used the weights that are most naturally derived from statistics result. During the training, the Adam optimizer is used to enhance learning efficiency with a learning rate of 0.0005. To reduce overfitting, random dropout is also used with a drop_rate of 0.2, and the batch sizes are set to 1.

The network was trained on two datasets to obtain two network models separately. To obtain the first training dataset, we download the file 'cullpdb_pc20_res1.6_R0.25_d190404_chains3470.gz' from http://dunbrack.fccc.edu/Guoli/piscis_download.php, which contains 3,470 nonredundant proteins and was then randomly split into a training (3,088 proteins) and a validation (382 proteins) set at a ratio of roughly 9:1 to prevent overfitting. The density map for the first dataset is simulated by

$$\rho(\mathbf{y}) = \sum_i \frac{A_i}{\sqrt{2\pi\sigma^2}} e^{-\frac{|\mathbf{y}-\mathbf{x}_i|^2}{2\sigma^2}} \quad (2)$$

where $\sigma = R/\sqrt{2\pi}$ with R being the resolution parameter randomly taken from [1, 15 Å], \mathbf{y} is the coordinate vector of the density map, \mathbf{x}_i and A_i indicate the coordinate vector and atomic number of i th atom of the protein, respectively. The second training dataset contains 3,600 targets with experimental density maps whose resolutions range from 2.1 to 10.0 Å. These experimental maps were generated from 36 large complexes with well-superposed experimental structures by randomly segmenting them into small maps with a size of $50 \times 50 \times 50 \text{ \AA}^3$. To make the training process focus more on C α atoms, we set a filter of these small maps by containing at least 250 C α atoms. This can avoid the issue of containing too few C α atoms in a map, which could happen in the marginal parts of experimental maps. Through the 3D-CNN networks, the first model was trained on the simulated training set with more than 720 epochs. We calculated the average CRscore loss from the validation set every 30 epochs and stopped the training if: (1) training epochs > 500 and maximum average CRscore > 0.8 and the latest average CRscore was 0.02 less than the maximum average CRscore, or (2) training epochs > 2,000. After stopping training, we selected the model with the max CRscore (708 epochs, Supplementary Fig. 11a). The second model started from the first model and was trained on the experimental training set for 217 more epochs, where the average loss against training epochs is shown in Supplementary Fig. 11b. The loss in the first model started to saturate around 600–700 epochs, while that in the second model did so after 800 epochs, probably because of the relatively higher complexity associated with the experimental maps.

Following the 3D-CNN model, a quick procedure was designed to convert the C α possibility map into C α atom coordinates (Supplementary Fig. 10b). The procedure first locates the grid with the highest possibility and labels it as the first C α atom. It then iteratively searches for the next C α atom with the highest possibility at the grids with a distance no less than 3.3 Å from all the labeled C α atoms. The procedure repeats to ensure at least L (query length) C α atoms are located. It will continue until $1.2 \times L$ C α atoms are located if the next highest possibility is > 0.9.

Initial template identification by LOMETS. We used LOMETS²¹, a meta-threading method containing 11 leading fold-recognition programs, to identify homologous and analogous templates from the PDB. For each query sequence, the top 300 templates were collected based on the normalized Z-score (Z_n), which measures the significance of query-template alignments by each program. Accordingly, a target was defined as 'Easy' if there was on average one or more good templates with $Z_n > 1$ for each program, while others were labeled 'Hard' due to the lack of good templates.

Inter-residue contact-map prediction. ResPRE³³ is used to predict the residue-residue contact maps. From a query sequence, ResPRE first uses DeepMSA³⁴ to collect MSAs from the whole-genome and metagenome sequence databases, where the inter-residue contact maps are then predicted from the inverse covariance matrix derived from the MSAs, based on deep residual convolutional network training¹⁹.

Deep learning-based template selection and regeneration. We designed two procedures using the deep learning-based C α conformations to improve initial template quality of CR-I-TASSER through template reselection and model regeneration, respectively.

Template reselection by C α and density map matching. LOMETS creates multiple threading templates, but the best templates do not always rank at the top by the Z-score. We rerank the top 300 template structures based on their match with the C α conformations predicted by the 3D-CNN from a cryo-EM density map, using a procedure outlined in Supplementary Fig. 12. Because the 3D-CNN C α conformation has no sequence index assigned, the matching procedure starts with the calculation of the 'fingerprint' for each C α atom in a given LOMETS template and C α conformation, where a fingerprint vector of i th C α atom $\vec{\mathbf{F}}(i)$ is defined as a set of 20 ascending-ranking intradistances between i th C α atom and 20 nearest C α atoms in the template (or C α conformation). A pairing score of i th atom at template with j th atom at C α conformation is then calculated by

$$Fscore_{ij} = \left| \frac{\vec{\mathbf{F}}(i) - \vec{\mathbf{F}}(j)}{\mathbf{F}_{temp}} \right|^2 \quad (3)$$

The lower the $Fscore_{ij}$ is, the more similar the environment two atoms (i, j) are in, indicating a higher probability for (i, j) to be correctly paired. Therefore, we initially select the C α atom pairs with the minimum $Fscore_{ij}$ and pair them in the ascending order, where each atom can only be paired once. Generally, if i th and j th C α atoms from the template are correctly paired to j th and j th C α atoms from the C α conformation, the intradistance between i th and i th C α atoms, $d(i, ii)$, should be close to that between j th and j th C α atoms, $d(j, jj)$. Based on this assumption, we further refine the initial pairing using a weighted matching score $S(i, j)$ defined by

$$S(i, j) = \sum_{\substack{ii \neq i \\ jj \neq j}} \begin{cases} W(i, ii) & \text{if } |d(i, ii) - d(j, jj)| \leq 1 \\ \frac{W(i, ii)}{(d(i, ii) - d(j, jj))^2} & \text{if } |d(i, ii) - d(j, jj)| > 1 \end{cases} \quad (4)$$

Here, $W(i, ii) = w(i)w(ii)$, where $w(i)$ is the weight for i th C α atom from the template, which is initially set at 1 and updated iteratively by an algorithm outlined in Supplementary Fig. 12. After the convergence, only the pairs with a matching score $S(i, j) > S_0$ are selected, where the threshold S_0 is defined by the two-mean clustering of the matching scores. Based on the selected C α pairing, the Kabsch r.m.s.d. superposition of template and C α conformation is performed³⁵, where the interchain distance $d_{ij} < 10 \text{ \AA}$ will be used as a new condition to select C α pairing in addition to equations (3–4). This new pairing will be used as the input of pairing refinement and Kabsch superposition to generate a newer pairing. The procedure will repeat until the final pairing and structure superposition converge (Supplementary Fig. 13). Overall, the idea of the superposition process described above is to identify the correct pairs of atoms between C α conformation (index-free) and template alignments (indexed) by comparing their intraenvironments.

Finally, the CRscore is calculated for each template with the 3D-CNN C α conformation based on the selected C α pairing, where the 300 LOMETS templates selected by Z-score are reranked based on the calculated CRscores. A template will be defined as a 'good' template if the CRscore > 0.5. Up to 30 good templates ($N_{\text{rank}} \leq 30$) are selected from this template reselection procedure.

Initial C α trace model generation from 3D-CNN C α conformations. Since CR-I-TASSER uses 40 replicas in the replica-exchange Monte Carlo (REMC) simulations and each replica starts with different templates, we generate $N_{\text{gen}} = 40 - N_{\text{rank}}$ new templates directly from the 3D-CNN C α conformations; this contains two steps of C α -trace connection and sequence-trace mapping (Supplementary Fig. 14).

For C α -trace connection, we first connect all neighboring C α atoms that have a distance below a bond length d_c . All connections to a C α atom that has

the number of connections ($n_{\text{conn}} \leq 2$) are considered as 'true' connections (for example, connections to atoms 1, 3, 5, 7 and 8 in Supplementary Fig. 14a), while all other connections that contradict with the true connections and make $n_{\text{conn}} > 2$ for other atoms are removed (for example, connections 2–4 in Supplementary Fig. 14a). After this scan, if a C α still contains >2 connections, this atom will be removed from the trace (for example, atom 6 in Supplementary Fig. 14a). As shown in Supplementary Fig. 14b, the remaining C α trace pattern will depend on the selection of d_c . In CR-I-TASSER, we implement the procedure under 18 different cutoffs of $d_c = 3.8, 3.9, \dots, 5.5 \text{ \AA}$ separately, and keep only the connections with a frequency of occurrence $>40\%$ in the final C α connection.

This connection procedure creates multiple C α fragments, where up to 1,000,000 C α traces are generated by randomly connecting the fragments, until no atom is available for the next connection. This could happen at the true end of the protein, 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. Although the constraints involved in the connection process can help improve the accuracy of the template generation on average, it cannot always result in a full-length C α trace model because the growth could stop anywhere under the constraints. To address this issue, CR-I-TASSER provides an alternative 'keep-tracing mode' to improve the fragment-tracing success rate for some special cases by partially releasing some of the restraints or additional iterations (see Supplementary Text 23 for details).

Assuming that each fragment is continuous, we map the query sequence to each C α trace by gapless threading and calculate the C α –C α contact map using a distance cutoff $d < 8 \text{ \AA}$. The top 300 C α traces are selected based on the Pearson correlation coefficient (PCC) of the C α –C α contact map with the predicted contact map from ResPRE, as well as the PCC of the template structure with the target density map (Supplementary Text 26). Finally, N_{gen} templates are selected from the 300 traces based on the PCC of the template structure with the target density map. This PCC is also used to rerank all top 40 templates including those from template reselection and regeneration.

It is noted that two 3D-CNN models have been trained on the simulated and experimental density map datasets separately, which generates two sets of C α conformations for each target. If the two conformations are close, that is, with the CRscore between them >0.85 , which usually indicates good quality of the conformations, we will take the average for each C α atom pair to generate the final C α conformation and use it for the template reselection and regeneration as described above. In case the C α conformations are different (CRscore <0.85), which, while rare, happens in some cases with low-resolution experimental cases and usually indicates that the predicted C α conformation is not reliable, we skip the C α conformation-based template reselection and regeneration. Instead, we match each of the LOMETS templates directly with the density maps using the Brodyen–Fletcher–Goldfarb–Shanno algorithm (Supplementary Text 27) followed by a short Metropolis Monte Carlo simulation under the guidance of template-density correlation as defined in Supplementary Text 26, with movements including 2,000 rigid-body translations/rotations. The top 40 templates are then selected based on the correlation coefficients from high to low.

Density map-guided structural assembly simulations. CR-I-TASSER performs REMC simulations to assemble full-length structure models, under a composite energy force field of

$$E_{\text{CR-I-TASSER}} = E_{\text{I-TASSER}} + W_{\text{temp}} E_{\text{temp}} + W_{\text{EM}} E_{\text{EM}} + W_{\text{EM}}^{\text{CNN}} E_{\text{EM}}^{\text{CNN}} \quad (5)$$

where $E_{\text{I-TASSER}}$ is the inherent knowledge-based potential extended from I-TASSER²⁰ and described in Supplementary Text 28, E_{temp} contains four aspects of distance and contact restraints collected from the top templates determined by LOMETS and 3D-CNN models (Supplementary Text 29). E_{EM} counts for the global correlation between structure conformation and experimental density map ρ_0 by

$$E_{\text{EM}} = - \sum_{\mathbf{y}} \rho_0(\mathbf{y}) \rho(\mathbf{y}) \quad (6)$$

where $\rho(\mathbf{y})$ is calculated by equation (2). The $E_{\text{EM}}^{\text{CNN}}$ counts for the correlation between structure conformation and the 3D-CNN predicted C α conformation:

$$E_{\text{EM}}^{\text{CNN}} = - \sum_{\mathbf{y}} \rho_0^{\text{CNN}}(\mathbf{y}) \rho(\mathbf{y}) \quad (7)$$

where ρ_0^{CNN} is the density maps calculated by equation (2) for the 3D-CNN C α conformation. This term is performed only when CRscore between the two 3D-CNN conformations is >0.85 , which is designed to enhance the convergence of simulations to the consensus C α conformations. It is noted that the negative cross correlation in equations (6–7) instead of PCC defined in Supplementary Text 26 is implemented because the former is computed as being faster than the latter. Additionally, benefiting from the linear combination form of equations (6–7), energy terms need to be computed only for the local segment involved in each movement, which is much faster than the calculations on the entire chain after each movement. The resolution for $\rho(\mathbf{y})$ and $\rho_0(\mathbf{y})$ calculations is automatically detected and set by a short-trained 3D-CNN predictor for resolution prediction.

Our benchmark results showed that the final model quality is not sensitive to the value of setting resolution. The weight parameters in equation (5), as well as those in the inherent knowledge-based I-TASSER force field, are determined in a separate training protein dataset, which is nonhomologous to the test proteins of this work, by maximizing the average TM score of the final models.

Final model selection and model quality estimation. The structure conformations generated by CR-I-TASSER (referred to as 'decoys') in eight low-temperature replicas are clustered using SPICKER to select the states corresponding to the lowest free energy states³⁶. Specifically, an all-to-all r.m.s.d. matrix is calculated among all decoys where a pair of decoys are considered as neighbors if their r.m.s.d. is within a cutoff. The decoy with the largest number of neighbors is selected as the center of the first cluster and the representative centroid model for the cluster is obtained by averaging all decoys included. The second cluster is obtained in a similar way on the remaining decoys after excluding all decoys from the first cluster, and the procedure repeats until five clusters are obtained. Thus, a decoy cluster captures the inherent statistics of the Monte Carlo process, that is, the larger the size of the decoy cluster, the higher the convergence and, accordingly, the lower the uncertainty of the model sampling. As the cluster centroid models from SPICKER often contain steric clashes, the centroids of the five biggest clusters are reassembled by a second round of REMC simulation to improve the hydrogen-bonding network and local structural geometry. The lowest energy conformations are selected from the second-round simulations and further refined at atomic level by the fragment-guided molecular dynamics³⁷ to create final models.

To evaluate the quality of predicted structures, we calculate the estimated TM score (eTM score) of the m th CR-I-TASSER model relative to the target structure by

$$\text{eTMscore}_m = 0.18 + 0.82 \max \left(C_m, \max_{n \neq m} (\text{TMscore}_{mn} - 0.5(1 - C_n)) \right) \quad (8)$$

where TM score_{mn} is the TM score between m th and n th predicted models. The confidence score C_m is defined as

$$C_m = \frac{\text{CRscore}_m}{1 + 0.05 (M_{\text{tot}} \langle \text{r.m.s.d.} \rangle_m) / M_m} \quad (9)$$

where M_{tot} is the total number of decoy conformations submitted to SPICKER, M_m is the number of decoys at m th cluster, $\langle \text{r.m.s.d.} \rangle_m$ is the average r.m.s.d. of the decoys to the cluster centroid and the CRscore_m is the matching score of the model with the 3D-CNN predicted C α conformation by equation (1).

Supplementary Fig. 16 displays the data of eTM score versus the actual TM scores on the first predicted models of all 530 test proteins with high-/low-resolution density maps, where most of the data points are located near the diagonal line, showing a strong linear correlation. The PCC and cosine similarities between eTM and TM scores are 0.858 and 0.989, respectively. If we use an eTM score of 0.5 as a cutoff to split 'positive' and 'negative' cases, the numbers of cases for true positive, false negative, true negative and false positive are 856, 44, 119 and 41, respectively, which correspond to the true positive, false negative, true negative and false positive rates of 95.1, 4.9, 74.4 and 25.6%, and the overall Matthews correlation coefficient (MCC) of 0.710. The strong correlation indicates that eTM score can be used to reliably estimate the quality of predicted models.

In addition to the eTM score for overall quality estimation, we introduce two metrics, local PCC and local confidence, to estimate the local agreement to the density for the final models. First, the local PCC for i th-residue modeling quality from the m th predicted model is defined as

$$\text{LPCC}(m, i) = \frac{\sum_{\mathbf{y}} [\rho_m(\mathbf{y}, i) - E[\rho_m(i)]] [\rho'_m(\mathbf{y}, i) - E[\rho'_m(i)]]}{\left\{ \sum_{\mathbf{y}} [\rho_m(\mathbf{y}, i) - E[\rho_m(i)]]^2 \sum_{\mathbf{y}} [\rho'_m(\mathbf{y}, i) - E[\rho'_m(i)]]^2 \right\}^{1/2}} \quad (10)$$

where $\rho_m(\mathbf{y}, i)$ is the density on grid \mathbf{y} calculated by equation (2) but only from the i th residue of the m th predicted model. Equation (10) is very similar to the normal PCC (Supplementary Text 26) except that we use a modified density ρ'_m instead of the experimental density ρ_0 :

$$\rho'_m(\mathbf{y}, i) = \rho_0(\mathbf{y}) \cdot \frac{\rho_m(\mathbf{y}, i)}{\sum_j \rho_m(\mathbf{y}, j)} \quad (11)$$

The reason we use the modified density to compute local PCC for the i th residue is because the experimental density $\rho_0(\mathbf{y})$ on grid \mathbf{y} contains contributions from all residues, where equation (11) is designed to decouple the experimental density for i th residue specifically. Toy model results shown in Supplementary Fig. 17 demonstrate that the $\rho'_m(\mathbf{y}, i)$ is more reasonable than $\rho_0(\mathbf{y})$ when computing the local PCC.

Second, the local confidence for i th residue from the m th predicted model is defined by integrating eTM score and local PCC:

$$LC(m, i) = T(m, i) \sum_j \frac{eTMscore_m}{T(m, j)} \quad (12)$$

where $T(m, i)$ is defined as

$$T(m, i) = \frac{(LPCC(m, i) + 1)}{N_{\text{model}}} \sum_{n=1}^{N_{\text{model}}} \frac{eTMscore_n}{1 + (d_i(m, n)/d_0)^2} \quad (13)$$

Here, N_{model} is the distance of i th residue between m th and n th models, and d_0 is a scaling parameter from TM score (Supplementary Text 4). N_{model} is the number of final models predicted by CR-I-TASSER, which is no more than five.

As an illustration, Supplementary Fig. 18 displays the local PCC and local confidence scores on two end-to-end study proteins (PDB 6tsk-B and 7dk5), where Supplementary Table 5 lists the average correlation coefficients between the local quality scores and the local error of predicted models from the experimental structure for all 248 test proteins with experimental density maps. The data show that both scores can be used for local model quality estimation. Although the local confidence shows a slightly higher correlation with the local modeling errors, CR-I-TASSER output both scores for alternative local quality estimations. In addition, CR-I-TASSER produces up to five models that allow the user to estimate the global/local quality using other methods such as ensemble structure comparison.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All training and testing data are available at <https://zhanggroup.org/CR-I-TASSER/> and <https://zenodo.org/record/5774563#.YbWYaNpJeXA> (<https://doi.org/10.5281/zenodo.5774563>). Source data are provided with this paper.

Code availability

The standalone package of the CR-I-TASSER programs, including library and manual documents, are available to download at <https://zhanggroup.org/CR-I-TASSER/download.html> and <https://zenodo.org/record/5774535#.YbWYudPJeXA> (<https://doi.org/10.5281/zenodo.5774535>).

References

33. 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 (2019).

34. Zhang, C., Zheng, W., Mortuza, S., Li, Y. & Zhang, Y. DeepMSA: constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. *Bioinformatics* **36**, 2105–2112 (2020).
35. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A* **32**, 922–923 (1976).
36. Zhang, Y. & Skolnick, J. SPICKER: a clustering approach to identify near-native protein folds. *J. Comput. Chem.* **25**, 865–871 (2004).
37. Zhang, J., Liang, Y. & Zhang, Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure* **19**, 1784–1795 (2011).

Acknowledgements

We thank Y. Li, C. Zhang and X. Zhou for their stimulating discussions. This work is supported in part by the National Institute of General Medical Sciences (grant nos. GM136422 and S10OD026825 to Y.Z.), the National Institute of Allergy and Infectious Diseases (grant no. AI134678 to Y.Z.), and the National Science Foundation (grant nos. IIS1901191, DBI2030790 and MTM2025426 to Y.Z.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This work used computing resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (grant no. ACI1548562).

Author contributions

Y.Z. conceived and designed the project. X.Z. developed the methods and performed the experiments. X.Z. and Y.Z. wrote the manuscript. B.Z. and P.L.F. participated in the discussion and edited the manuscript. All authors proofread and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-021-01389-9>.

Correspondence and requests for materials should be addressed to Peter L. Freddolino or Yang Zhang.

Peer review information *Nature Methods* thanks Georgios Skiniotis and the other, anonymous, reviewers for their contribution to the peer review of this work. Arunima Singh was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Python (2.6) was used in data collection.

Data analysis Python (3.7) and TMscore were used in data analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All training and testing data are available at <https://zhanggroup.org/CR-I-TASSER/>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The training dataset used in this research contained 3,088 simulated density maps generated from non-redundant proteins and 3,600 segmented experimental density maps. The benchmark dataset included 778 targets which is sufficient for statistics analysis.
Data exclusions	Proteins homologous to the test targets were excluded from the threading library to avoid homology contamination on the test results.
Replication	All results could be reproduced by our web server and standalone package, or based on the Online Methods/Supporting Information.
Randomization	The continuum single-domain proteins in the benchmark dataset were collected from CASP12 and CASP13 (simulated), as well as randomly collected from the PDB (simulated) or EMDDataResource (experimental), after the consideration of non-redundancy among the protein entries.
Blinding	There was no blinding group or analysis in this manuscript.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging