nature machine intelligence

Article

https://doi.org/10.1038/s42256-025-01056-0

End-to-end cryo-EM complex structure determination with high accuracy and ultra-fast speed

Received: 6 January 2025

Accepted: 9 May 2025

Published online: 24 June 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

While cryogenic-electron microscopy yields high-resolution density maps for complex structures, accurate determination of the corresponding atomic structures still necessitates significant expertise and labour-intensive manual interpretation. Recently, artificial intelligence-based methods have emerged to streamline this process; however, several challenges persist. First, existing methods typically require multi-stage training and inference, causing inefficiencies and inconsistency. Second, these approaches often encounter bias and incur substantial computational costs in aligning predicted atomic coordinates with sequence. Last, due to the limitations of available datasets, previous studies struggle to generalize effectively to complicated and unseen test data. Here, in response to these challenges, we introduce end-to-end and efficient CryoFold (E3-CryoFold), a deep learning method that enables end-to-end training and one-shot inference. E3-CryoFold uses three-dimensional and sequence transformers to extract features from density maps and sequences, using cross-attention modules to integrate the two modalities. Additionally, it uses an SE(3) graph neural network to construct atomic structures based on extracted features. E3-CryoFold incorporates a pretraining stage, during which models are trained on simulated density maps derived from Protein Data Bank structures. Empirical results demonstrate that E3-CryoFold improves the average template modelling score of the generated structures by 400% as compared to Cryo2Struct and significantly outperforms ModelAngelo, while achieving this huge improvement using merely one-thousandth of the inference time required by these methods. Thus, E3-CryoFold represents a robust, streamlined and cohesive framework for cryogenic-electron microscopy structure determination.

Since the invention of the microscope, scientists have sought to observe protein complexes with greater clarity to elucidate their structures and functions and how they affect biological processes[1]. Over centuries of technological advancements within the structural biology community, cryogenic-electron microscopy (cryo-EM)[2], which was awarded the Nobel Prize in 2017, has emerged as a pivotal technique. Cryo-EM is capable of producing nearly atomic-resolution density maps that reveal the shapes and interactions of macromolecules[3–5] without the requirement for crystallization and the damage to samples. Interpreting these three-dimensional (3D) density maps to atomic structural models is a critical step for researchers aiming to understand macromolecular behaviour, however, this process is inherently challenging[6]. It necessitates high levels of expertise to guide interpretation and incurs significant computational costs associated with computer graphics programs[7,8], primarily due to the high dimensionality of the density maps. Furthermore, the absence of accurate templates can

e-mail: chengtan9907@gmail.com; Stan.ZQ.Li@westlake.edu.cn

severely compromise both the accuracy and efficiency of the structural determination[9,10].

In recent years, the integration of artificial intelligence (AI) methods has emerged as a promising approach to address challenges in various fields, including the determination of cryo-EM structures[11–13]. One notable example is DeepTracer[14], a classical deep learning method that uses multiple U-Net architectures[15] to predict the structure of protein complexes. It uses the travelling salesman problem[16] algorithm to connect the predicted alpha carbon (Cα) atoms. Another approach, ModelAngelo[17], enhances structure prediction by incorporating a sequence module and aligning predicted structures with sequences using a hidden Markov model (HMM)[18]. DeepMainmast[19] integrates AlphaFold2 (ref. 20) with a density tracing protocol, significantly improving the quality of atomic models derived from cryo-EM maps. The latest advance, Cryo2Struct[21], uses a 3D Transformer-U-Net architecture[22] for protein structure determination, also using HMM for mapping predicted Cα atoms to sequences. Despite the advances made by these methods in facilitating accurate and expertise-free cryo-EM structure determination, several challenges remain: (1) multiple training and inference stages introduce inefficiency and inconsistency. For example, DeepTracer requires the training of four U-Net models and involves five inference steps. ModelAngelo necessitates the training of convolutional and graph neural networks (GNNs), comprising three inference stages. Similarly, Cryo2Struct trains models for residue and atom predictions and introduces three inference stages. The multi-stage processes used in these existing methods can introduce bias due to inconsistency, and result in run-time inefficiency. (2) There is alignment bias of predicted atom coordinates and sequence. Existing methods, including DeepTracer, Cryo2Struct and ModelAngelo, use convolutional neural networks to obtain atomic coordinates, subsequently using non-parametric algorithms such as the travelling salesman problem or HMMs for alignment with sequence data. This alignment approach often leads to inaccuracies and incurs significant computational costs due to the extensive search space, which generally requires tens of minutes or even hours to predict one sample. (3) There is insufficient ability to generalize. Although the number of new cryo-EM structures in the Electron Microscopy Data Bank[23] is increasing exponentially, fewer than 13,000 cryo-EM structures with resolutions better than 4 Å have been determined so far, many of which are redundant. Consequently, the limited scale of available cryo-EM density maps constrains the ability of deep learning methods to generalize effectively to a broader range of real-world samples.

Here we present our method, end-to-end and efficient CryoFold (E3-CryoFold), which effectively addresses these challenges. E3-CryoFold is an end-to-end training and one-shot inference model that eliminates redundant multi-stage processes, resulting in significantly improved efficiency and accuracy, achieving inference times and template modelling (TM) scores[24] that are one-thousandth and 400% those of existing multi-stage methods, such as Cryo2Struct, and significantly outperforms ModelAngelo. E3-CryoFold concurrently uses 3D and sequence transformers to extract features from both the density map and the sequence. It uses cross-attention[25] modules to integrate spatial information into the sequence features. These spatial-sequence features are then input into an equivariant GNN[26,27] to construct 3D atomic models. This approach circumvents alignment loss between structure and sequence by directly infusing spatial features into the sequence representation. Notably, we have established a training dataset of simulated cryo-EM density maps derived from 163,284 Protein Data Bank (PDB) structures, which enhances model generalization through pretraining. We validate the generalization capability of E3-CryoFold across two test datasets, encompassing different resolutions and lengths, and compare its performance against other robust baselines. Our results demonstrate that E3-CryoFold offers an approach for simpler, more efficient and more robust cryo-EM structure determination.

Results

Overall framework

Figure 1 illustrates the overall framework of E3-CryoFold. Initially, we preprocess the density maps and sequences to align the data and expedite the training process. The density maps and sequences are then input into 3D and sequence transformers, respectively, while a cross-attention module is used to integrate spatial and sequential information from both modalities. Subsequently, an equivariant GNN is constructed to generate atomic structures based on the combined spatial-sequential features. Unlike previous approaches, E3-CryoFold facilitates end-to-end training, allowing users to input the complete cryo-EM density map and sequence (or use the model without sequence information) to directly obtain the atomic structure through a single model. Further details are provided in Methods.

Comparison on a standard cryo-EM structure determination dataset

We initially assessed the modelling performance of E3-CryoFold in comparison to other baseline methods using a standard test dataset from Cryo2StructData[28], where the detailed introduction and filtering process of this dataset is presented in Supplementary Information Section 1.

For our comparisons, we selected Cryo2Struct and Phenix[29] as reference methods. Both E3-CryoFold and Cryo2Struct use the Cryo2StructData for training, and Cryo2Struct represents the most recent related work, being the only openly available source. Structures generated by Phenix were downloaded from its official website for comparison. We evaluated the structural models constructed for 128 test cryo-EM maps against the corresponding true structures in the PDB to assess their quality. The evaluation results, based on four metrics, are presented in Fig. 2.

Consistency of E3-CryoFold's modelling improves accuracy. Figure 3b illustrates the relationship between sequence matching precision and length for E3-CryoFold and Cryo2Struct. E3-CryoFold directly models atomic structures in a sequential manner, thereby ensuring high accuracy in sequence matching. In contrast, Cryo2Struct uses an algorithm that aligns the predicted Cα atom coordinates with the sequence. This approach introduces substantial bias in sequence matching that can strictly impair the quality of the generated structures.

Figure 2a,e shows the global normalized TM scores of the atomic models constructed by the three methods. The TM score is a standard metric for measuring the similarity between a model and its corresponding known structure, ranging from 0 to 1, where higher scores indicate better alignment. We calculated the TM scores using the TM-align[24] program, a widely adopted tool for structural comparison. To ensure a fair comparison of the models built by Cryo2Struct and Phenix, which often differ in residue count, the global TM scores were normalized to the length of the experimental structure. The average global normalized TM score for E3-CryoFold is 0.815, significantly exceeding that of Cryo2Struct (average TM score of 0.2) and Phenix (average TM score of 0.087). Furthermore, across all test density maps, E3-CryoFold consistently achieves higher TM scores than both methods, demonstrating the superior accuracy and stability of the predicted structures.

Figure 2b,f presents the root-mean-square deviations (r.m.s.d.) of atomic models constructed by the three methods. After aligning the structures, r.m.s.d. measures the Euclidean distance between corresponding atoms in the two models. The average r.m.s.d. for E3-CryoFold is 1.888 Å, which is considerably lower than Cryo2Struct's average of 9.093 Å and Phenix's average of 8.372 Å. Notably, E3-CryoFold consistently demonstrates a lower r.m.s.d. than Cryo2Struct across all evaluated density maps, and it outperforms Phenix in 149 out of 150 maps.

https://doi.org/10.1038/s42256-025-01056-0

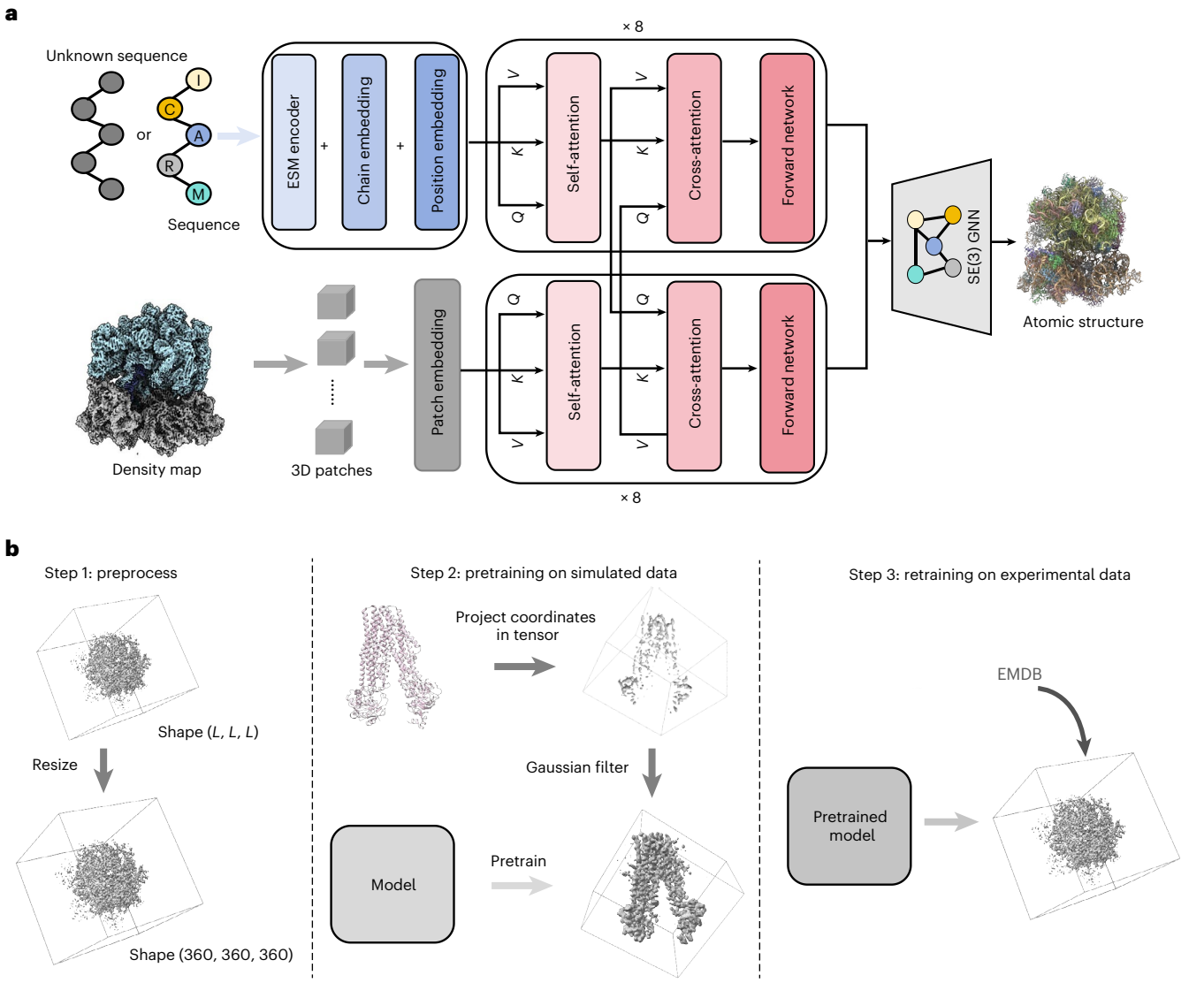

Fig. 1 | The architecture and pipeline of E3-CryoFold. a, The overall framework of E3-CryoFold. 'Unknown sequence' means that this sequence consists of [unk] token without residues prompt. A density map and sequence are input into 3D and sequence transformers, respectively, and a cross-attention module is used for the two modal representations to communicate. Subsequently, conditioning on the extracted features of sequence and density maps, a SE(3) GNN progressively refines the atomic structures. **b**, The complete workflow. *L* is the original dimension of the input density map.

Figure 2c,g illustrates the angle loss of atomic models constructed by the three methods. Angle loss quantifies the average difference in angles among all residues within a complex structure, measuring the disparity between corresponding residues in two different structures: a lower angle loss indicates better alignment. The angle loss can measure the quality of generated structures at the angle level. E3-CryoFold exhibits an average angle loss of 0.2386, significantly lower than that of Cryo2Struct (1.2186) and Phenix (1.7627).

Furthermore, Fig. 2d,h visualizes the bond loss of the atomic models generated by the three methods. Bond loss calculates the average difference in bond lengths among all residues, with lower values indicating improved accuracy. E3-CryoFold's average bond loss is 0.3091, considerably lower than those of Cryo2Struct (1.4525) and Phenix (1.8285). Beyond this, Fig. 3d shows the LDDT (local distance difference test)[30] metric for these atomic models. LDDT is used to compare the differences between predicted protein structures and experimentally obtained structures. It evaluates the accuracy of the model by calculating the difference in distances between localized atoms, which is similar to the bond loss metrics. E3-CryoFold's average LDDT is 0.8746, also better than those of Cryo2Struct (0.2553) and Phenix (0.2300). Across all test density maps, E3-CryoFold demonstrates lower angle, bond losses and higher LDDT metrics compared to the other two methods. These results indicate that E3-CryoFold not only produces superior overall predicted structures but also refines each residue with reduced bias.

We present a comparative analysis of DockQ[31] metrics for three methods in Fig. 3c. DockQ serves as a scoring system for evaluating molecular docking results, providing a quantitative assessment of the accuracy of predicted protein-ligand binding patterns. It integrates contact density and geometric considerations into its evaluation. A DockQ score exceeding 0.8 indicates that the generated atomic models are of high quality. Notably, E3-CryoFold achieves an average DockQ score of 0.900, significantly surpassing the scores of Cryo2Struct (0.027) and Phenix (0.023). These findings illustrate that E3-CryoFold is significantly more effective at capturing the geometric relationships among the complex chains and poses better modelling for the molecules' interaction. Additionally, we introduce the Cα displacement metric, which is calculated by Phenix's chain_comparison tool[29],

https://doi.org/10.1038/s42256-025-01056-0

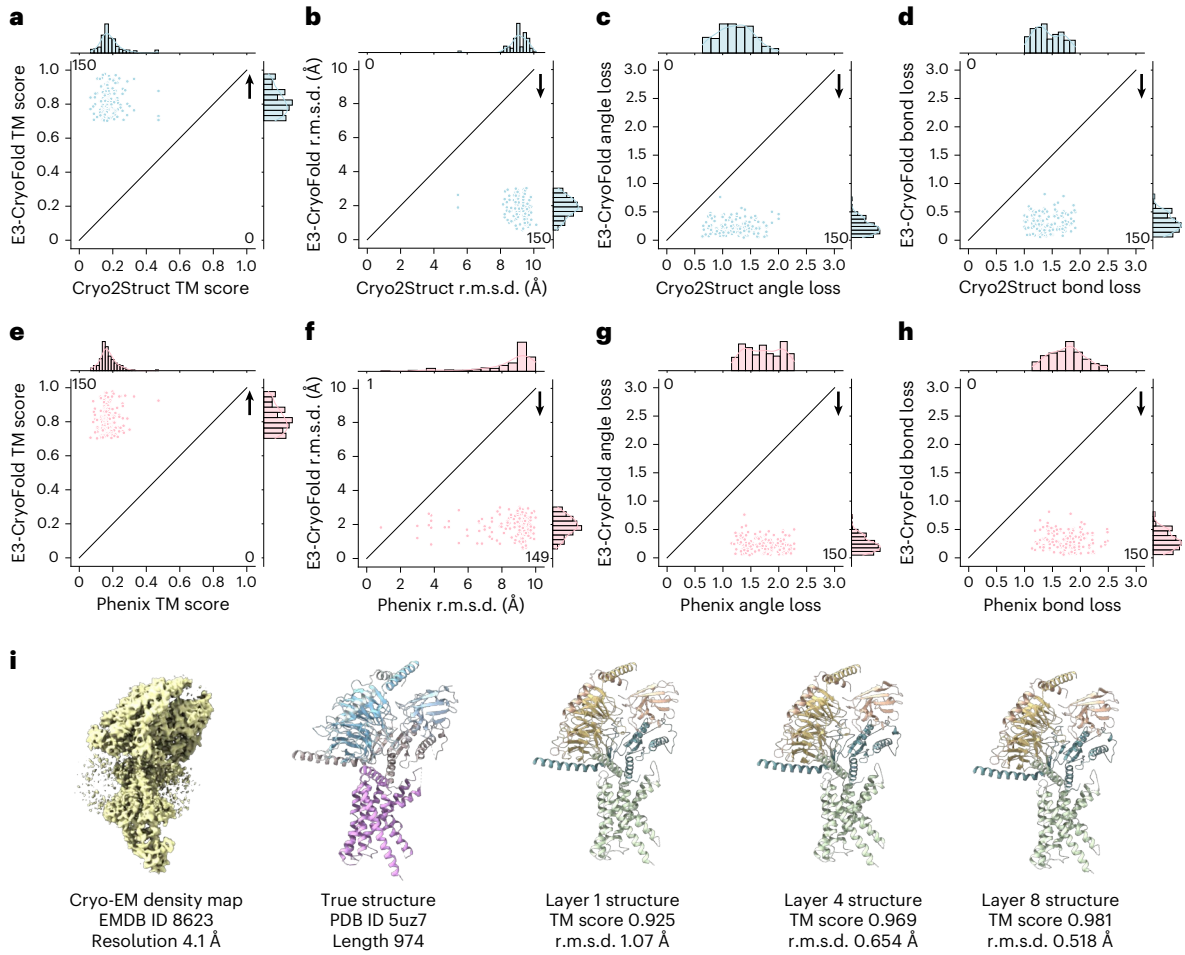

Fig. 2 | The analysis results of atomic structures on 150 test experimental density maps for E3-CryoFold against Cryo2Struct and Phenix in four metrics. In each panel of an evaluation metric, the score of the model built by E3-CryoFold for each map is plotted against that by contrasting the baseline for the same map. A dot above the 45° line indicates that E3-CryoFold has a higher score than the baseline for the map. The number at the top-left and bottom-right corners of a plot are the number of targets plotted above and below the diagonal line, respectively. The up or down arrow in the top-right corner indicates whether this higher or lower metric is better. **a**, The TM score of the atomic structures of E3-CryoFold against Cryo2Struct. The TM score of the atomic structures normalized by the length of the known structure. The normalized TM score is calculated by using TM-align to align the atomic models: the higher the TM score, the better. **b**, The r.m.s.d.[49] of the atomic structures of E3-CryoFold against Cryo2Struct. The r.m.s.d. is also calculated by the TM-align program: the lower the r.m.s.d., the better. **c**, The angle loss of the atomic structures of E3-CryoFold against Cryo2Struct. The angle loss represents the angle difference between the two structures: the lower the angle loss, the better. **d**, The bond loss of the atomic structures of E3-CryoFold against Cryo2Struct. The bond loss represents the atom distance difference between the two structures: the lower the bond loss, the better. **e**, The TM score of the atomic structures of E3-CryoFold against Phenix. **f**, The r.m.s.d. of the atomic structures of E3-CryoFold against Phenix. **g**, The angle loss of the atomic structures of E3-CryoFold against Phenix. **h**, The bond loss of the atomic structures of E3-CryoFold against Cryo2Struct. **i**, The visualization examples of E3-CryoFold's SE(3) GNN multi-layer predictions. The cryo-EM density map 8623 was released on 3 May 2017.

to assess the performance of three models presented in Fig. 3g. The average Cα displacements for E3-CryoFold, Cryo2Struct and Phenix are measured at 0.3163, 1.8237 and 1.2708, respectively. These results indicate a substantial improvement in the predictive accuracy of Cα atom positions. This further corroborates the accuracy of predictions made by E3-CryoFold.

Figure 3e,f shows the performance of models predicted by E3-CryoFold in comparison to deposited models (target structures), assessed using the *Q*-score[32] and Fourier shell correlation (FSC) metrics[33]. The *Q*-score quantifies the resolvability of individual atoms within cryo-EM maps, thereby reflecting the overall quality of the constructed model. A well-constructed model typically exhibits *Q*-scores that correlate with local resolution, which may vary across cryo-EM maps. FSC is a quantitative measure used to assess the resolution of 3D reconstructions from cryo-EM data. It evaluates the similarity between two independent reconstructions of the same specimen by comparing their Fourier transforms within specific spatial frequency shells. Additionally, we calculate the *Q*-scores using the MapQ tool[32] and determine the FSC using Servalcat[34] after refining both models, focusing only on the residues present in E3-CryoFold and the deposited models. The average *Q*-scores for models generated by E3-CryoFold and the deposited models are 0.4470 and 0.5017, respectively. Similarly, the average FSC scores for the same models are 0.5127 and 0.5345. Although the models produced by E3-CryoFold do not outperform the deposited models in these metrics, it is important to consider that E3-CryoFold was trained using the deposited models as targets. These findings indicate that E3-CryoFold achieves performance comparable to that of the target structures, particularly in terms of the FSC metric. This suggests a strong alignment between the models generated by E3-CryoFold and the corresponding cryo-EM density maps.

Figure 3h,i,j presents a detailed analysis of LDDT metrics, angle and bond loss for each residue type. Bulky amino acids, particularly tryptophan (Trp), tyrosine (Tyr), arginine (Arg) and phenylalanine (Phe), exhibited minimal angle and bond loss, along with high LDDT

https://doi.org/10.1038/s42256-025-01056-0

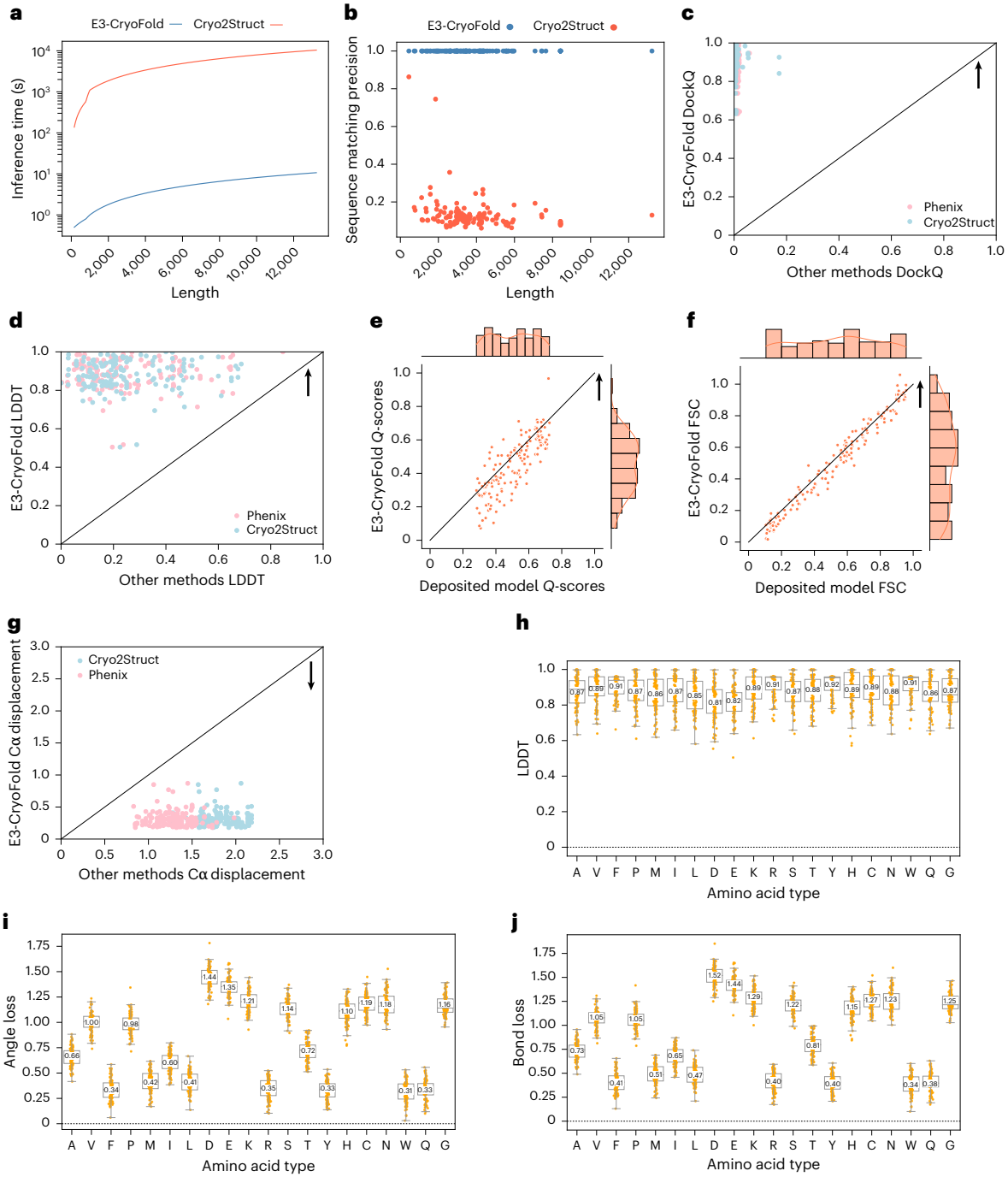

Fig. 3 | The analysis results in various metrics for E3-CryoFold, Cryo2Struct and Phenix on 150 test experimental density maps. a, The inference time of E3-CryoFold and Cryo2Struct against different lengths. **b**, The sequence matching precision of E3-CryoFold and Cryo2Struct against different lengths. **c**, The DockQ of the atomic structures of E3-CryoFold against other methods: the higher the DockQ, the better. **d**, The LDDT of the atomic structures of E3-CryoFold against other methods: the higher the LDDT, the better. **e**, The *Q*-score of structures of E3-CryoFold against the deposited models: the higher the *Q*-score, the better. **f**, The FSC of structures of E3-CryoFold against the deposited models: the higher the FSC, the better. **g**, The Cα displacement of atomic structures of E3-CryoFold against other methods: the lower the Cα displacement, the better. **h**, The LDDT analysis of each amino acid type for 300,000 amino acids in 150 target structures. **i**, The angle loss. **j**, The bond loss. In the three box plots in **h**, **i** and **j**, the centre line, along with the bottom and top edges of the box, represents the median, first quartile and third quartile values, respectively. The boundaries of the whiskers extend to 1.5 × the interquartile range from the upper and lower quartiles.

scores. In contrast, negatively charged residues, such as glutamic acid (Glu) and aspartic acid (Asp), which are more susceptible to radiation damage[35–37], demonstrated higher levels of angle and bond loss and lower LDDT scores. Notably, hydrophobic amino acids, which are less prone to radiation damage[37], also exhibited lower angle and bond loss alongside higher LDDT values.

We observed that E3-CryoFold demonstrates significant advantages in all metrics. All the improved metrics indicate that the atomic models produced by E3-CryoFold exhibit greater rationality and similarity. These advantages stem from two key perspectives: (1) E3-CryoFold integrates spatial features from density maps into the sequence modality, thereby eliminating bias associated with the

alignment of sequences and predicted Cα atoms; and (2) E3-CryoFold predicts atomic models in an end-to-end manner, which mitigates the inconsistency typically encountered across multiple training and inference stages through gradient descent. Thus, E3-CryoFold is capable of generating more reasonable and high-quality structures. Furthermore, we believe that enhancing the alignment of E3-CryoFold's predicted structures with the coordinates of Cα atoms from density maps could yield additional improvements. Consequently, this positions E3-CryoFold as a valuable tool for structural biologists aiming to elucidate protein functions and interactions at an atomic level.

Figure 2i illustrates a high-quality visualization generated by E3-CryoFold alongside predictions from various SE(3) GNN layers. Our analysis reveals that higher layers generally yield more accurate and complete predictions compared to lower layers, demonstrating the progressive refinement of atomic structures by the SE(3) GNN. Additionally, we offer further example analyses of E3-CryoFold in comparison to Cryo2Struct and Phenix in Supplementary Information Section 2.

The end-to-end model benefits efficiency considerably. Figure 3a depicts the relation of inference time and the length for E3-CryoFold and Cryo2Struct. On a single A100 graphical processing unit (GPU), the minimum inference time for E3-CryoFold is 0.5 s (Electron Microscopy Data Bank (EMDB) ID 6555; sequence length is 190), the maximum inference time is 10.7 s (EMDB ID 6270; sequence length is 13,224) and the average time is 3.1 s. Cryo2Struct's minimum inference time is 137.85 s, maximum inference time is 10,398.8 s and the average time is 3,001.2 s. On average, E3-CryoFold is up to nearly 1,000 times faster than Cryo2Struct, demonstrating the efficiency of the E3-CryoFold modelling method. We believe this prominent merit positions E3-CryoFold to be able to serve as a simpler, more efficient and more accurate tool to revolutionize the research community of cryo-EM structure determination.

Evaluation of E3-CryoFold on a new established test dataset

To evaluate the generalization of E3-CryoFold, we assess the performance of E3-CryoFold on a large independent test dataset comprising 500 new density maps. We remove the test samples in which the PDB structure only contains the asymmetric units and this results in 428 density maps. The details of this new test dataset can be found in Supplementary Information Section 1.

Comparing E3-CryoFold with ModelAngelo on 109 test samples. To evaluate the accuracy of E3-CryoFold predictions and thoroughly investigate the advantages and limitations of the proposed model, we conduct a comprehensive comparison with ModelAngelo, a prominent method in cryo-EM structure determination. Given that the density maps in the standard test set referenced in the section 'Comparison on a standard cryo-EM structure determination dataset' were normalized before download, and considering ModelAngelo's limitations with normalized density maps, we selected a dataset for comparing E3-CryoFold with ModelAngelo. In accordance with ModelAngelo's approach, we removed structures containing insertion codes and other irregularities to minimize computational costs and structural biases, resulting in 109 test samples.

Figure 4a presents the TM scores of the predicted atomic structures from both methods. The average TM score for E3-CryoFold is 0.863, whereas ModelAngelo achieves an average of 0.329. Notably, E3-CryoFold outperforms ModelAngelo in 97 out of 109 samples, indicating a significant advancement in TM-score performance.

Figure 4b illustrates the r.m.s.d. values for the predicted atomic structures. The average r.m.s.d. for E3-CryoFold is 1.508 Å, compared to 1.849 Å for ModelAngelo. While E3-CryoFold's predictions are superior in 44 out of 109 samples, its average r.m.s.d. is comparable to that of ModelAngelo, albeit with fewer instances of superior performance.

Figure 4c,d depicts the angle and bond loss metrics for the predicted atomic structures. The average angle loss for E3-CryoFold is 0.1722, while ModelAngelo's is 0.2199. The average bond loss for the two models is 0.2207 for E3-CryoFold and 1.4966 for ModelAngelo. E3-CryoFold outperforms ModelAngelo in 79 and 105 samples for angle and bond loss, respectively. These metrics reveal that E3-CryoFold exhibits superior average performance over ModelAngelo, although it demonstrates greater variance in angle loss, while ModelAngelo shows more significant variance in bond loss. This highlights the differing strengths of the two methods in predicting atomic angles and bond lengths.

Figure 4e displays the Cα displacements of the predicted atomic structures. E3-CryoFold's average Cα displacement is 0.2769, significantly better than ModelAngelo's average of 0.7035. Moreover, E3-CryoFold surpasses ModelAngelo in 99 out of 109 samples. Unlike the comparable r.m.s.d. values, the Cα displacement metric clearly indicates that E3-CryoFold generates models with superior local quality.

Figure 4f,g presents the *Q*-scores and FSC for the predicted atomic structures from both E3-CryoFold and ModelAngelo. The average *Q*-score for E3-CryoFold is 0.4413, while ModelAngelo exhibits a score of 0.5037. In terms of FSC, E3-CryoFold achieves an average of 0.5278 compared to ModelAngelo's 0.5699. Additionally, E3-CryoFold outperforms ModelAngelo in 20 and 28 predictions, respectively. These findings indicate that ModelAngelo demonstrates superior alignment between predicted structures and density maps, as its atomic predictions are directly derived from the density data.

Figure 4h illustrates the inference time comparison between the two methods. E3-CryoFold significantly reduces average inference time to 4.1 s, in stark contrast to ModelAngelo's 4,536.2 s.

A comprehensive analysis of these results reveals key differences between the two methods.

Regarding efficiency, E3-CryoFold's average inference time is substantially lower than that of ModelAngelo, demonstrating an efficiency increase of nearly 1,000 times.

Regarding accuracy, various metrics, including TM score, angle loss and bond loss, indicate a significant advancement in E3-CryoFold's capabilities for sequential and global structure modelling. Furthermore, both r.m.s.d. and Cα displacement metrics show that E3-CryoFold matches or exceeds ModelAngelo's performance in atomic position prediction. We observe that in some cases ModelAngelo shows poor TM scores despite having low r.m.s.d. values. We attribute this phenomenon to the bias in Cα atom coordinate-sequence alignment in ModelAngelo, as well as the inherent differences between TM score and r.m.s.d. The r.m.s.d. focuses solely on the absolute distances between atoms in two structures without considering sequence similarity, whereas TM score emphasizes the overall folded structure, the relative positions of proteins and sequence similarity[24]. Consequently, even if some predicted structures from ModelAngelo show very similar atomic coordinates to the ground truth, they may still yield poor TM scores if ModelAngelo fails to accurately align the predicted coordinates and sequences. Instead, E3-CryoFold maintains consistency in TM score and r.m.s.d. due to the perfect Cα atom-sequence alignment of E3-CryoFold. We also provide two visualization examples of E3-CryoFold and ModelAngelo in Fig. 4i,j. In these two visualization examples, we observe that while ModelAngelo can generate overall structures similar to the ground truth, it produces many disconnected or additional chains that differ from the ground truth, therefore showing more different colours of chains in the visualization. Additionally, the structure generated by ModelAngelo in Fig. 4j shows a poor TM score but a good r.m.s.d. These results further support our earlier conclusion that ModelAngelo struggles to accurately align the predicted Cα atoms with their corresponding sequences. In contrast, E3-CryoFold demonstrates high homogeneity with ground truth in chains and accurately generates structures, highlighting the superior performance of E3-CryoFold.

https://doi.org/10.1038/s42256-025-01056-0

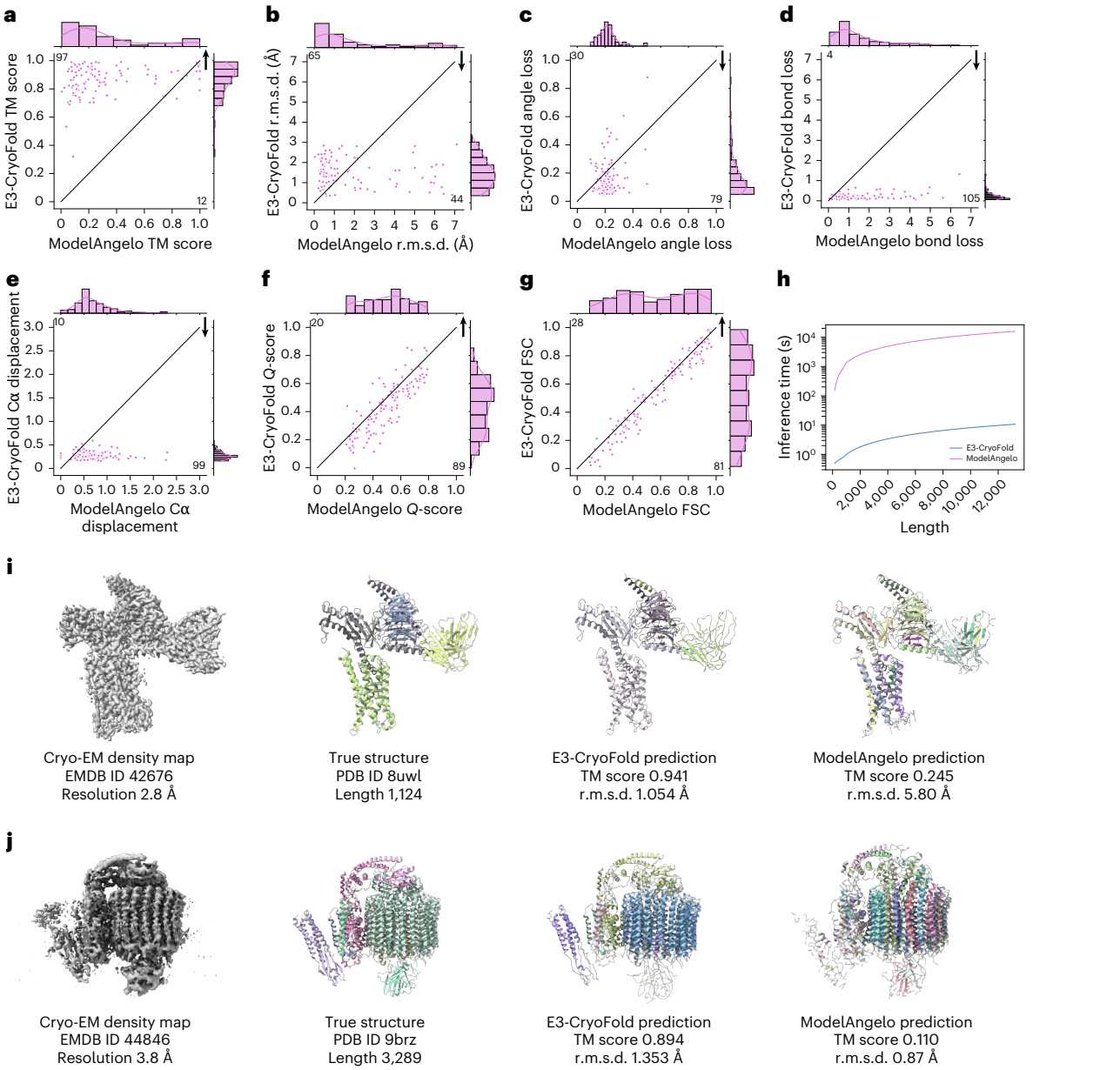

Fig. 4 | The analysis results of atomic models built on 109 test experimental cryo-EM maps for E3-CryoFold against ModelAngelo in eight metrics. **a**, The TM score of the atomic structures of E3-CryoFold against ModelAngelo. **b**, The r.m.s.d. **c**, The angle loss. **d**, The bond loss. **e**, The Cα displacement. **f**, The *Q*-scores. **g**, The FSC. **h**, The inference time of two models versus the length of atomic structures. **i**,**j**, The visualization examples (EMDB ID 42676 (**i**), EMDB ID 44846 (**j**)) of E3-CryoFold and ModelAngelo.

Regarding structure-map alignment, in terms of *Q*-score and FSC metrics, E3-CryoFold shows a clear disadvantage compared to ModelAngelo. This discrepancy is expected, as E3-CryoFold relies on spatial-sequential features where sequence information is paramount, whereas ModelAngelo primarily uses density map data for atomic structure predictions, using sequence information as an auxiliary tool. These results underscore the potential for integrating both methods as a promising direction for future cryo-EM structure determination.

Evaluating E3-CryoFold on the whole test dataset. As Fig. 5 shows, on this newly established test dataset, E3-CryoFold demonstrates an average TM score of 0.854, r.m.s.d. of 1.632 Å, angle loss of 0.184 and bond loss of 0.2360. These figures represent improvements of 4.5, 13.2, 21.6 and 22.4%, respectively, compared to the standard test dataset. In contrast, Cryo2Struct yields an average TM score of 0.213, r.m.s.d. of 8.804 Å, angle loss of 1.107 and bond loss of 1.325. While E3-CryoFold maintains a leading TM score, angle loss and bond loss relative to Cryo2Struct, its results exhibit greater bias from the target structure due to a lack of constraints. This further underscores E3-CryoFold's capacity to generate accurate atomic structures while highlighting the challenges associated with bias in the absence of stringent constraints.

Figure 5a illustrates the relationship between TM score and resolution. While the TM scores of Cryo2Struct's predicted structures decline as resolution decreases, E3-CryoFold shows enhanced quality of generated structures with increasing resolution. This suggests that E3-CryoFold is more robust to low resolution and less dependent on the quality of the density map. Additionally, the lower Pearson correlation for E3-CryoFold compared to Cryo2Struct supports this conclusion. Figure 5b,c,d depicts r.m.s.d., angle and bond loss versus resolution,

https://doi.org/10.1038/s42256-025-01056-0

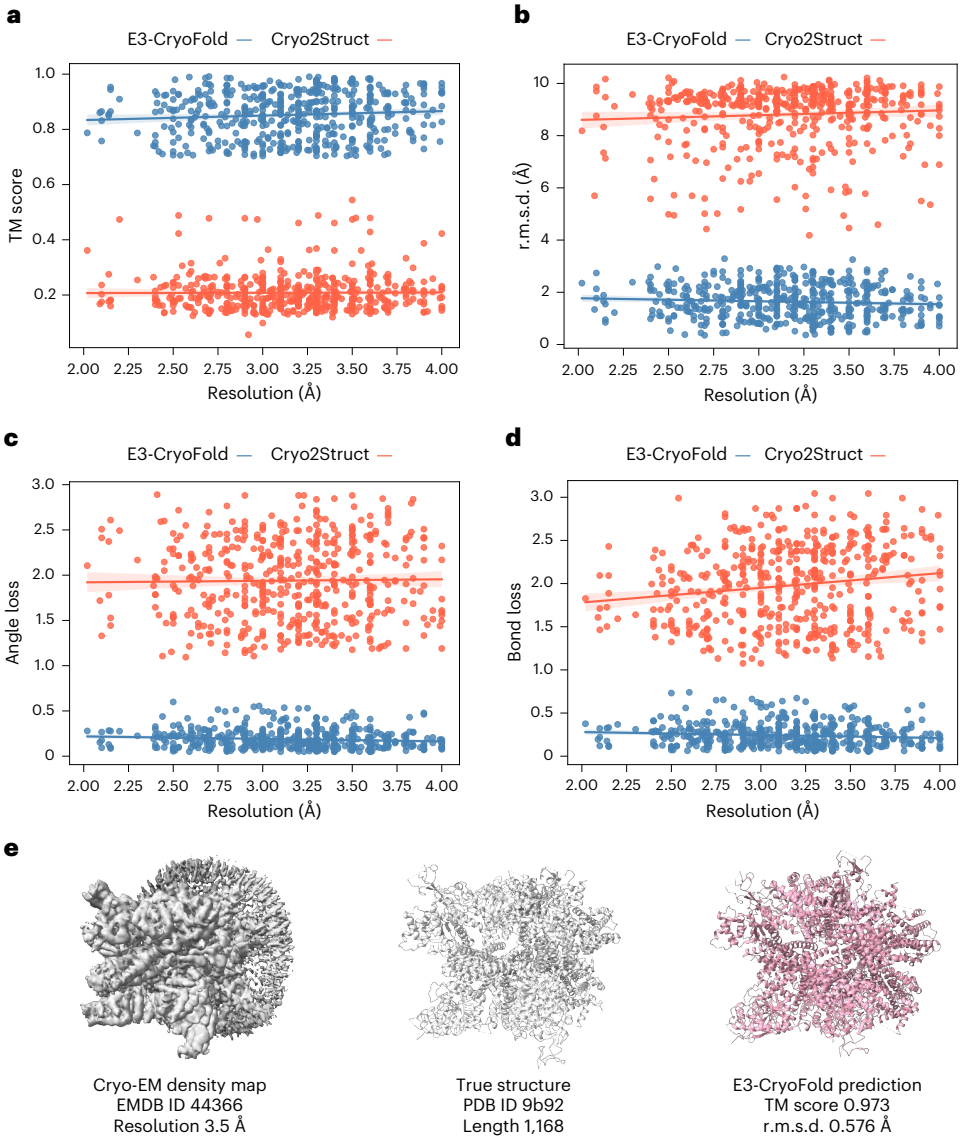

Fig. 5 | The analysis results of atomic models built for 428 test experimental cryo-EM maps. The solid lines depict linear regression lines and the coloured area represents a 95% confidence interval. **a**, The TM score versus resolution; the E3-CryoFold regression equation, $0.0153x + 0.8054$; Pearson's correlation, 0.079; the Cryo2Struct regression equation, $-0.0011x + 0.184$; correlation, −0.010. **b**, The r.m.s.d. versus resolution; the E3-CryoFold regression equation, $-0.1034x + 1.958$; correlation, −0.069; the Cryo2Struct regression equation, $0.1833x + 8.234$; correlation, 0.067. **c**, The angle loss versus resolution; the E3-CryoFold regression equation, $-0.0266x + 0.2679$; correlation, −0.107; the Cryo2Struct regression equation, $0.0316x + 1.793$; correlation, 0.0304. **d**, The bond loss versus resolution; the E3-CryoFold regression equation, $-0.0350x + 0.3462$; correlation, −0.113; the Cryo2Struct regression equation, $0.0375x + 1.895$; correlation, 0.035. **e**, The cryo-EM density map 44366 was released on 15 May 2024.

respectively. These results further reinforce the earlier conclusion that E3-CryoFold demonstrates improved structural quality with increased resolution. Collectively, these findings indicate that E3-CryoFold performs competitively well with relatively low-resolution data.

Figure 6a–f presents a visual analysis of six metrics in relation to target structure length. All figures demonstrate that the atomic structures generated by E3-CryoFold experience a minor decline in performance as the length of target structures increases. However, structures of considerable length still maintain good TM scores above 0.7 and r.m.s.d. values below 3.0 Å. These findings further affirm E3-CryoFold's capability to handle test data comprising very long sequences, demonstrating the significance of E3-CryoFold in biological structural analysis, particularly in addressing the complexities associated with extended sequence lengths.

Figure 6g–l presents an analysis of six metrics in relation to the maximum dimension of the target density map (for example, for a target density map with dimensions of (540, 500, 500), the maximum dimension is 540). Theoretically, resizing during preprocessing can lead to greater information loss as the dimensions of the target density map increase. Indeed, the figures indicate that the predicted structures generated by E3-CryoFold exhibit a slight decline in performance as the maximum dimension increases. However, the results remain consistent with those observed for length, showing that the degradation is not pronounced, with TM scores consistently above 0.7 and r.m.s.d. values below 3.0 Å. This demonstrates that E3-CryoFold is robust against information loss resulting from resizing. Nevertheless, the issue of information loss is significant and warrants further investigation in future work.

Supplementary Fig. 13 illustrates several high-quality examples generated by E3-CryoFold on the test dataset, further demonstrating its capability to model diverse complexes with intricate structures and combinations.

https://doi.org/10.1038/s42256-025-01056-0

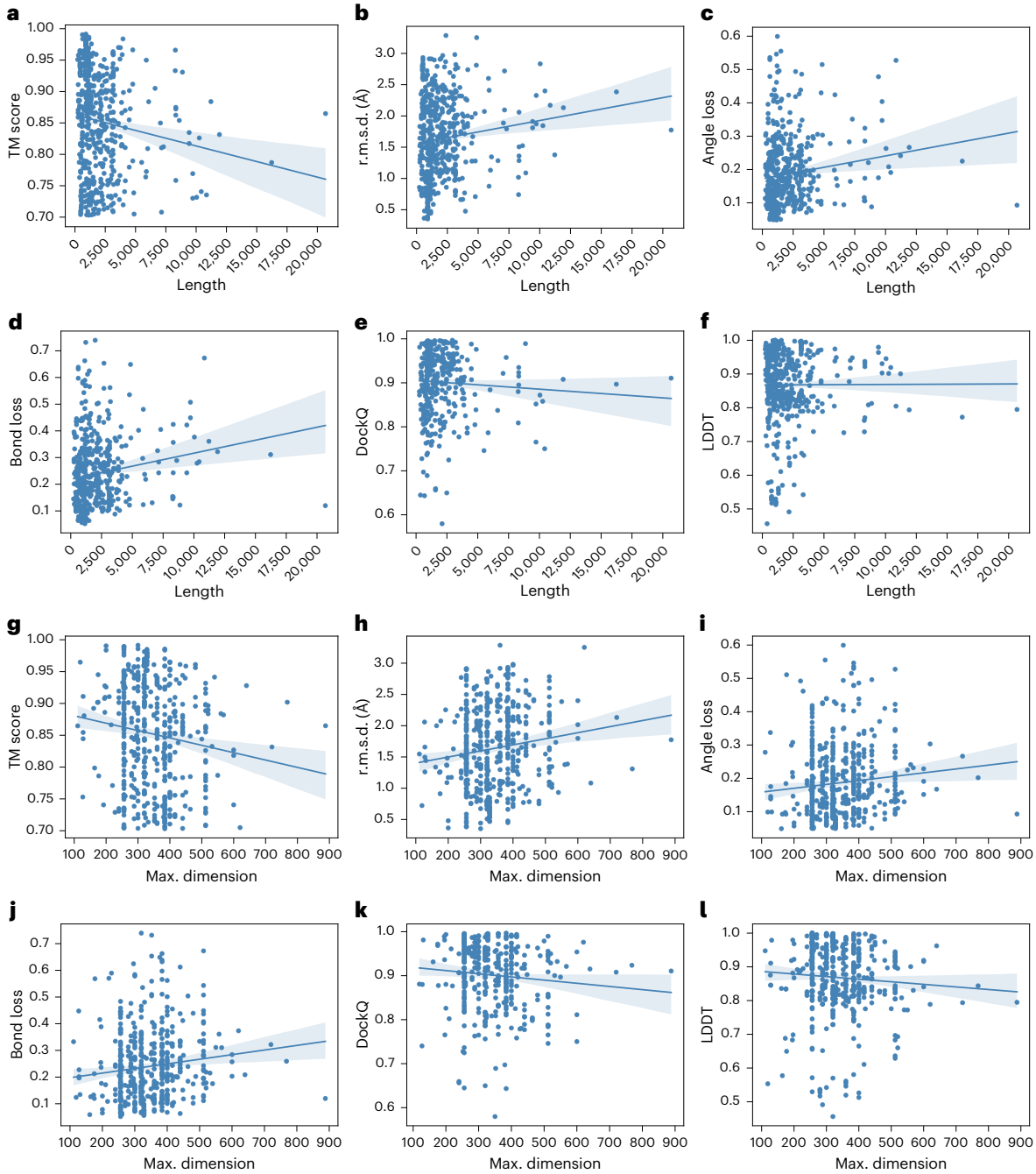

Fig. 6 | The analysis results of atomic models built for 428 test cryo-EM maps. The coloured area represents a 95% confidence interval. **a**, The TM score versus length; the regression equation, $-4.91 \times 10^{-6}x + 0.8596$; Pearson's correlation, −0.128. **b**, The r.m.s.d. versus length; the regression equation, $3.59 \times 10^{-5}x + 1.5613$; correlation, 0.123. **c**, The angle loss versus length; the regression equation, $6.89 \times 10^{-6}x + 0.1713$; correlation, 0.145. **d**, The bond loss versus length; the regression equation, $9.78 \times 10^{-6}x + 0.2199$; correlation, 0.164. **e**, The DockQ versus length; the regression equation, $-1.89 \times 10^{-6}x + 0.9112$; correlation, −0.052. **f**, The LDDT versus length; the regression equation, $1.53 \times 10^{-7}x + 0.8654$; correlation, 0.0034. **g**, The TM score versus maximum (max.) dimension of density maps; the regression equation, $-1.26 \times 10^{-4}x + 0.8911$; Pearson's correlation, −0.129. **h**, The r.m.s.d. versus max. dimension; the regression equation, $9.82 \times 10^{-4}x + 1.3059$; correlation, 0.152. **i**, The angle loss versus max. dimension; the regression equation, $1.32 \times 10^{-4}x + 0.1390$; correlation, 0.111. **j**, The bond loss versus max. dimension; the regression equation, $1.31 \times 10^{-4}x + 0.1853$; correlation, −0.133. **k**, The DockQ versus max. dimension; the regression equation, $-7.44 \times 10^{-5}x + 0.9301$; correlation, −0.112. **l**, The LDDT versus max. dimension; the regression equation, $-7.82 \times 10^{-5}x + 0.8921$, correlation, −0.077.

Discussion

E3-CryoFold introduces a model for efficient, robust and generalizable cryo-EM structure determination. It leverages a 3D and sequence transformer to extract information from cryo-EM density maps and sequences, using a cross-attention module to integrate these two modalities. Furthermore, an efficient SE(3) GNN is proposed to construct the complete atomic structure, facilitating end-to-end training and inference.

We evaluated the performance of E3-CryoFold models on both standard and newly established test datasets. Our results demonstrate a

significant improvement in accuracy and efficiency, achieving accurate predictions with only one-thousandth of the inference time compared to previous methods, including ModelAngelo, Cryo2Struct and Phenix.

Despite these advancements, E3-CryoFold faces certain challenges. First, the irregular shapes of density maps necessitate resizing them to a uniform shape, which may introduce bias and lead to a loss of spatial information, particularly in larger maps. Second, because E3-CryoFold generates atom coordinates without constraints, the r.m.s.d. between predicted and target structures can be relatively volatile. We believe that combining E3-CryoFold predictions with the atom coordinates derived from density maps could effectively mitigate these issues. Last, E3-CryoFold currently supports only the modelling of the residue backbone, while the side chains, which are also critical, have not been considered. The modelling of side chains represents a significant extension and an important area for future development of E3-CryoFold.

Methods

Problem definition

E3-CryoFold is designed to predict the 3D atomic structure of a protein complex by leveraging both a cryo-EM density map and the corresponding protein sequence. The cryo-EM density map is represented as a 3D voxel grid $M \in \mathbb{R}^{L\times L\times L}$, where L denotes the dimension of the density map. In practice, the density map is divided into cubic patches of size $L' \times L' \times L'$, resulting in $(\frac{L}{L'})^3$ smaller units that facilitate efficient processing of the spatial data. The protein sequence consisting of N residues is denoted as $S = (s_1, s_2, \ldots, s_N)$, where each residue s_i belongs to the set of amino acids $\mathbb{A}$. It is important to highlight that, in our model, the specific types of amino acid in the sequence are not required; the primary role of the sequence input is to convey the number of residues. This allows E3-CryoFold to focus on reconstructing the protein's structure based on spatial data from the density map while using the sequence primarily as a guide for residue count.

The goal of E3-CryoFold is to predict the 3D atomic structure of the protein backbone, which includes the spatial coordinates of four backbone atoms for each residue: nitrogen (N), alpha carbon (Cα), carbonyl carbon (C) and oxygen (O). The output is represented as a tensor $X \in \mathbb{R}^{N\times 4\times 3}$, where N is the number of residues, 4 corresponds to the four backbone atoms and 3 refers to the x, y, z spatial coordinates. The overarching objective is to learn a mapping function:

$$f(M, S) \rightarrow X, \tag{1}$$

which takes as input the cryo-EM density map M and the sequence S, and outputs the 3D atomic coordinates of the protein's backbone. This function must effectively integrate the spatial information embedded in the cryo-EM density map with the structural constraints implied by the protein sequence, ultimately producing an accurate and physically plausible atomic reconstruction of the protein complex.

Background

Self-attention. The self-attention mechanism[22] was initially introduced to capture long-range dependencies. Given a d-dimensional embedding $H \in \mathbb{N} \times \mathbb{D}$, the self-attention operation computes attention scores between all pairs of elements using query (Q), key (K) and value (V) matrices. These matrices are linear transformations of the input embeddings:

$$Q = HW_Q, K = HW_K, V = XW_V. \tag{2}$$

W_Q, W_K and W_V are the matrices that project the embedding into the hidden dimension. The attention weights are calculated as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D}}\right)V. \tag{3}$$

T is the transpose operation; D is the hidden dimension size. This mechanism allows each residue or voxel to aggregate information from the entire sequence or density map, making it particularly useful for modelling long-range spatial dependencies in cryo-EM density maps and protein sequences.

Cross-attention. While self-attention captures intra-modality relationships (within the sequence or within the density map), the cross-attention mechanism[25] extends this concept to interactions between different modalities. Cross-attention allows one set of embeddings (queries) to attend to another set (keys and values), facilitating the integration of information from multiple sources. In E3-CryoFold, cross-attention is used to merge features from the density map and the protein sequence. The sequence embeddings are updated by attending to the spatial embeddings derived from the density map, effectively allowing the sequence representation to integrate spatial information.

Mathematically, if $Q_{\text{seq}} \in \mathbb{R}^{N\times D}$ represents the sequence embedding and $K_{\text{des}}, V_{\text{des}} \in \mathbb{R}^{L'\times D}$ represent the embedding from the density map, the cross-attention operation is formulated as:

$$\text{Cross-attention}(Q_{\text{seq}}, K_{\text{des}}, V_{\text{des}}) = \text{softmax}\left(\frac{Q_{\text{seq}}K_{\text{des}}^T}{\sqrt{D}}\right)V_{\text{des}}. \tag{4}$$

This allows the sequence to integrate spatially contextualized information from the density map, ensuring a coherent representation of both modalities.

Overall workflow

Preprocess. Normalizing the density values of cryo-EM density maps involves applying scaling and clipping techniques. Positive density values within these maps indicate regions where the protein is predicted to be located. However, the distribution of these positive density values varies significantly across different maps, with some ranging from −2.32 to 3.91 and others from −0.553 to 0.762. To ensure comparability across diverse datasets, a percentile-based normalization approach is used. This involves computing the 95th percentile of positive density values within each map and subsequently normalizing all values relative to this threshold.

Given the diverse dimensions of density maps, handling all maps with a single configuration presents significant challenges. Therefore, we standardize all density maps to a uniform shape of 360 × 360 × 360 using a cubic interpolation algorithm. Although this resizing operation may compromise some content, we demonstrate that E3-CryoFold remains robust to such information loss in 2.3. For sequence encoding, we use the Evolutionary Scale Modelling (ESM)[38] alphabet to represent all complex sequences. Additionally, we use the [unk] token to encode residues not included in the alphabet. In constructing the target atomic structures, we predict the coordinates of backbone atoms (including C, Cα, N and O atoms) for each amino acid during training.

Pretraining on simulated density maps. Despite consistent progress in the field of cryo-EM, the total number of available cryo-EM density maps remains limited to approximately 30,000. Additionally, a significant portion of these maps has resolutions below 4 Å, complicating structural determination. To address the limitations of dataset scale and improve the generalization of our models, we simulate high-resolution density maps using a comprehensive collection of PDB structures. To avoid the sequence overlapping between the test and pretraining dataset, we select PDB structures with less than 20% sequence identity to any of the proteins in the test set. We have curated a dataset comprising 163,284 high-quality PDB structures, from which we map atomic coordinates into density maps of dimensions (360, 360, 360). For constructing simulated density maps, we initialize a zero tensor of dimensions (360, 360, 360). Based on the scale of the PDB structure size, we determine which pixel each atom corresponds to.

Each atom is then mapped into its respective pixel within the initialized tensor. Following this mapping process, we smooth the density values by applying a Gaussian filter[39]. We set the width of Gaussian distribution as 2, and the amplitude is 1 to simulate a high-quality density map. Furthermore, to enhance data diversity, we introduce random Gaussian noise (mean 0, scale 0.05) as part of our data augmentation strategy.

For model training, we set the learning rate to 1×10^{-4}, the batch size to 4 per GPU and the number of training epochs to 50. We use the AdamW[40] optimizer in conjunction with the Onecycle[41] learning rate scheduler for pretraining our models. The training is conducted on eight A100 GPUs over a period of approximately 5 days, with validation performed on the Cryo2StructData[28] test set.

Fine-tuning on experimental density maps. Despite our models being pretrained on a large-scale dataset of high-quality simulated density maps, it remains essential to refine these models using experimental data for real-world applications. Following the pretraining phase, we further fine-tune the models on the Cryo2StructData. This dataset comprises 7,389 experimental cryo-EM density maps with resolutions ranging from 1 to 4 Å. These density maps were released by the EMDB on 27 March 2023 (ref. 23). We split the whole dataset into 7,000 density maps for training and 389 for validation. We maintain the same training hyperparameters as in the pretraining stage, with the exception of increasing the training epoch to 500.

Model architecture

3D and sequence transformers extract features. To integrate spatial information from cryo-EM density maps into sequence data, we should embed both modalities into a shared hidden space. To achieve this, we use 3D and sequence encoders to model the input data within the same dimensional framework. The density map is divided into 1,000 cubes, each with dimensions of 36 × 36 × 36. Convolutional kernels of the same dimensions are used to encode these cubes into spatial embeddings. E3-CryoFold comprises eight blocks, with each block implementing spatial self-attention modules to facilitate communication among different spatial positions for each embedding. For enhanced generalization of the sequence module, we use the ESM-2 pretrained model to derive sequence embeddings, supplemented by sequential self-attention modules to connect individual residues. Besides, we use two embedding modules—'Chain Embedding' and 'Position Embedding'—to capture the sequence position and chain information for each residue. Crucially, cross-attention modules are integrated to enable interaction between spatial and sequential features. At the end of each block, multi-layer perceptrons update the features from both modalities. The output sequence features from the sequence encoder are subsequently used as node features within the equivariant GNN.

Equivariant GNN constructs atomic structures. Conditioned on these extracted node features, we use an equivariant GNN to construct the final structure of complexes. Initially, virtual atoms are obtained by projecting the node features into Euclidean space. These virtual atoms, with the shape of (N,4,3) (where N represents the number of nodes), along with the node features, are then fed into eight layers of the equivariant GNN[42] to refine the virtual atomic representations progressively. Following the approach outlined in Chroma[43], we use multiple loss functions to train E3-CryoFold, specifically, Global Loss, Fragment Loss, Pair Loss and Neighbour Loss. Further details are provided in the next section 'Protein backbone reconstruction based on SE(3) GNN'.

Protein backbone reconstruction based on SE(3) GNN

Backbone reconstruction with a protein graph. The backbone reconstruction process in E3-CryoFold begins by initializing random coordinates $\hat{X} \in \mathbb{R}^{N\times4\times3}$. Based on these initial coordinates, a k-nearest neighbours (kNN) graph is constructed, which defines the local spatial relationships between residues in the protein structure. Each residue serves as a node in this graph, and its neighbours are determined based on proximity in the backbone structure. The initial node embeddings $\mathcal{V}^{(0)}$ are derived from the integrated feature embeddings $H \in \mathbb{R}^{N\times D}$, which are generated by combining spatial information from the cryo-EM density map with sequence information. These embeddings encapsulate both local and global structural features of the protein, allowing the model to leverage the inherent relationships between the protein's sequence and its spatial configuration.

For each residue, a local frame $T = (R, \mathbf{t})$ is defined, where $R \in \mathbb{R}^{3\times3}$ is a rotation matrix encoding the residue's orientation, and $\mathbf{t} \in \mathbb{R}^3$ is a translation vector specifying the residue's position in 3D space. These local frames are updated iteratively using a SE(3) GNN[42], which respects the symmetries of 3D space by ensuring that the operations on rotations and translations remain equivariant to transformations such as rotations and translations. At each iteration, the SE(3) GNN aggregates relative rotation and relative translation information from neighbouring residues to update the frame of a given residue. For a residue s, the update rules for the rotation matrix R_s and translation vector $\mathbf{t}_s$ are as follows:

$$\begin{cases} \mathrm{vec}(R_s) = \sum\limits_{j\in\mathcal{N}_s} a^{\mathrm{r}}_{sj}\mathrm{vec}(R_{sj}) \\ R_s \leftarrow \mathrm{Quat2Rot} \circ \mathrm{Norm} \circ \mathrm{MLP}^{9\to4} \odot \mathrm{vec}(R_s) \\ \mathbf{t}_s = \sum\limits_{j\in\mathcal{N}_s} a^{\mathrm{t}}_{sj}\mathbf{t}_j \end{cases}$$

where $\mathcal{N}_s$ denotes the set of neighbouring residues for the residue s in the kNN graph, and a^{r}_{sj} and a^{t}_{sj} are learnable attention weights that determine the influence of neighbour j's rotation and translation on the residue s. The rotation matrix R_{sj} is flattened into a nine-dimensional vector using the vec(·) operation before being aggregated. After aggregation, the vector is transformed back into a valid rotation matrix using a quaternion-to-rotation function, Quat2Rot(·), which ensures numerical stability and smooth rotations. Similarly, the translation vector $\mathbf{t}_s$ is updated based on the weighted contributions from the translations of neighbouring residues. Through this frame-level message passing, the SE(3) GNN allows each residue's position and orientation to be iteratively refined based on the local structural context provided by its neighbours. This ensures that the backbone reconstruction respects the geometric relationships within the protein while remaining equivariant to spatial transformations.

Once the local frames $T_s = (R_s, \mathbf{t}_s)$ for each residue have been refined over multiple layers of the SE(3) GNN, the 3D coordinates of the backbone atoms can be recovered. The updated coordinates x_s of each residue s are computed by applying the predicted local transformation:

$$x_s = T_s^{(l)} \odot \mathcal{V}^{(l)}. \tag{5}$$

where $\mathcal{V}^{(l)}$ represents the node embedding at the final iteration.

Reconstruction loss. Inspired by Chroma[43], E3-CryoFold uses multiple loss functions to train the model effectively. The total loss $\mathcal{L}$ is the sum of several components designed to enforce global structural accuracy, local fragment fidelity, pairwise consistency and accurate modelling of neighbourhood relationships:

$$\mathcal{L} = \mathcal{L}_{\mathrm{global}} + \mathcal{L}_{\mathrm{fragment}} + \mathcal{L}_{\mathrm{pair}} + \mathcal{L}_{\mathrm{neighbour}} + \mathcal{L}_{\mathrm{distance}} \tag{6}$$

The loss terms are defined as follows:

- Global loss ($\mathcal{L}_{\mathrm{global}}$): this term evaluates the r.m.s.d. between the ground truth 3D coordinates $X \in \mathbb{R}^{N\times4\times3}$ and the reconstructed coordinates $\hat{X} \in \mathbb{R}^{N\times4\times3}$:

$$\mathcal{L}_{\mathrm{global}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{4}\sqrt{\sum_{k=1}^{3}\left(X_{i,j,k} - \hat{X}_{i,j,k}\right)^2} \tag{7}$$

- Fragment loss ($\mathcal{L}_{\text{fragment}}$): this loss computes the r.m.s.d. between local fragments of residues. For each residue, the loss is evaluated over its c nearest neighbouring residues:

$$\mathcal{L}_{\text{fragment}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{c}\sum_{k=1}^{4}\sqrt{\sum_{q=1}^{3}\left(X_{i,j,k,q}-\hat{X}_{i,j,k,q}\right)^2} \tag{8}$$

- Pair loss ($\mathcal{L}_{\text{pair}}$): this loss enforces consistency between pairs of residues, evaluating the r.m.s.d. over k pairs for each kNN[44] pair:

$$\mathcal{L}_{\text{pair}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\sum_{k=1}^{c^2}\sum_{q=1}^{4}\sqrt{\sum_{m=1}^{3}\left(X_{i,j,k,q,m}-\hat{X}_{i,j,k,q,m}\right)^2} \tag{9}$$

- Neighbour loss ($\mathcal{L}_{\text{neighbour}}$): this loss enforces consistency between each residue and its k-nearest neighbours:

$$\mathcal{L}_{\text{neighbour}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\sum_{k=1}^{4}\sqrt{\sum_{q=1}^{3}\left(X_{i,j,k,q}-\hat{X}_{i,j,k,q}\right)^2} \tag{10}$$

- Distance loss ($\mathcal{L}_{\text{distance}}$): the distance loss directly evaluates the mean squared error between the predicted and ground truth pairwise distance matrices. Given the pairwise distance matrix $\mathcal{D} \in \mathbb{R}^{N\times N}$ and its reconstructed counterpart $\hat{\mathcal{D}}$, the loss is computed as follows:

$$\begin{cases} \mathcal{D}_{i,j} = \sqrt{\frac{1}{4}\sum_{k=1}^{4}\sum_{q=1}^{3}\left(X_{i,k,q}-X_{j,k,q}\right)^2} \\ \hat{\mathcal{D}}_{i,j} = \sqrt{\frac{1}{4}\sum_{k=1}^{4}\sum_{q=1}^{3}\left(\hat{X}_{i,k,q}-\hat{X}_{j,k,q}\right)^2} \\ \mathcal{L}_{\text{distance}} = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathcal{D}_{i,j}-\hat{\mathcal{D}}_{i,j}\right)^2 \end{cases} \tag{11}$$

By default, $c = 7$ represents the number of fragments, and $k = 30$ indicates the number of kNNs[44]. These loss functions are applied at each layer of the decoder, and the final loss is computed as the average across all layers. This multi-layer loss application improves model performance by ensuring that intermediate representations are consistent with the final prediction, leading to more robust and accurate reconstructions of the protein backbone.

Data availability

The experimental dataset can be downloaded at https://doi.org/10.7910/DVN/FCDG0W (ref. 45), and the standard test dataset can be downloaded at https://doi.org/10.7910/DVN/2GSSC9 (ref. 46). The low-resolution and simulated datasets are accessible at https://zhang-group.org/CR-I-TASSER/. All source data are accessible from ref. 47 (standard_test_data.xlsx for standard test dataset, novel_test_data.xlsx for novel established test dataset, low_resolution_experimental_data.xlsx for low-resolution density maps, simulated_data.xlsx for simulated density maps). Source data are provided with this paper.

Code availability

The source code of E3-CryoFold is available via GitHub at https://github.com/A4Bio/CryoFold/ (ref. 48). This repository also contains the instructions and tutorial for applying E3-CryoFold on an example cryo-EM map to generate a complex structure.

References

1. Boadu, F., Cao, H. & Cheng, J. Combining protein sequences and structures with transformers and equivariant graph neural networks to predict protein function. *Bioinformatics* **39**, i318–i325 (2023).
2. Bai, X.-C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
3. Lawson, C. L. et al. Outcomes of the EMDataResource cryo-EM ligand modeling challenge. *Nat. Methods* **21**, 1340–1348 (2024).
4. Dhakal, A., Gyawali, R., Wang, L. & Cheng, J. A large expert-curated cryo-EM image dataset for machine learning protein particle picking. *Sci. Data* **10**, 392 (2023).
5. Dhakal, A., Gyawali, R., Wang, L. & Cheng, J. Cryotransformer: a transformer model for picking protein particles from cryo-EM micrographs. *Bioinformatics* **40**, btae109 (2024).
6. Giri, N., Roy, R. S. & Cheng, J. Deep learning for reconstructing protein structures from cryo-EM density maps: recent advances and future directions. *Curr. Opin. Struct. Biol.* **79**, 102536 (2023).
7. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
8. Croll, T. I. Isolde: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. Sect. D: Struct. Biol.* **74**, 519–530 (2018).
9. Gao, Y., Thorn, V. & Thorn, A. Errors in structural biology are not the exception. *Acta Crystallogr. Sect. D Struct. Biol.* **79**, 206–211 (2023).
10. Croll, T. I. et al. Making the invisible enemy visible. *Nat. Struct. Mol. Biol.* **28**, 404–408 (2021).
11. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. Cryodrgn: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
12. Rangan, R. et al. CryoDRGN-ET: deep reconstructing generative networks for visualizing dynamic biomolecules inside cells. *Nat. Methods* **21**, 1537–1545 (2024).
13. Levy, A., Wetzstein, G., Martel, J. N., Poitevin, F. & Zhong, E. Amortized inference for heterogeneous reconstruction in cryo-EM. *Adv. Neural Inf. Process. Syst.* **35**, 13038–13049 (2022).
14. Pfab, J., Phan, N. M. & Si, D. Deeptracer for fast de novo cryo-EM protein structure modeling and special studies on cov-related complexes. *Proc. Natl Acad. Sci. USA* **118**, e2017525118 (2021).
15. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Part III* Vol. 18, 234–241 (Springer, 2015).
16. Hoffman, K. L. & Padberg, M. et al. Traveling salesman problem. *Encycl. Oper. Res. Manag. Sci.* **1**, 1573–1578 (2013).
17. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
18. Rabiner, L. & Juang, B. An introduction to hidden Markov models. *IEEE ASSP Mag.* **3**, 4–16 (1986).
19. Terashi, G., Wang, X., Prasad, D., Nakamura, T. & Kihara, D. Deepmainmast: integrated protocol of protein structure modeling for cryo-EM with deep learning and structure prediction. *Nat. Methods* **21**, 122–131 (2024).
20. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
21. Giri, N. & Cheng, J. De novo atomic protein structure modeling for cryoEM density maps using 3D transformer and HMM. *Nat. Commun.* **15**, 5511 (2024).
22. Vaswani, A. Attention is all you need. In *Proc. Advances in Neural Information Processing Systems* Vol. 30 (eds Guyon, I. et al.) 6000–6010 (Curran Associates, 2017).
23. Lawson, C. L. et al. Emdatabank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403 (2016).
24. Zhang, Y. & Skolnick, J. Tm-align: a protein structure alignment algorithm based on the Tm-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
25. Chen, C.-F. R., Fan, Q. & Panda, R. Crossvit: cross-attention multi-scale vision transformer for image classification. In *Proc. IEEE/CVF International Conference on Computer Vision* 357–366 (IEEE, 2021).

26. Satorras, V. G., Hoogeboom, E. & Welling, M. E (n) equivariant graph neural networks. In *Proc. International Conference on Machine Learning* 9323–9332 (PMLR, 2021).
27. Han, J., Rong, Y., Xu, T. & Huang, W. Geometrically equivariant graph neural networks: a survey. Preprint at https://arxiv.org/abs/2202.07230 (2022).
28. Giri, N., Wang, L. & Cheng, J. Cryo2StructData: a large labeled cryo-EM density map dataset for AI-based modeling of protein structures. *Sci. Data* **11**, 458 (2024).
29. Terwilliger, T. C., Adams, P. D., Afonine, P. V. & Sobolev, O. V. A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* **15**, 905–908 (2018).
30. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. LDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
31. Basu, S. & Wallner, B. DockQ: a quality measure for protein-protein docking models. *PLoS ONE* **11**, e0161879 (2016).
32. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
33. Van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *J. Struct. Biol.* **151**, 250–262 (2005).
34. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using servalcat. *Biol. Crystallogr.* **77**, 1282–1291 (2021).
35. Allegretti, M., Mills, D. J., McMullan, G., Kühlbrandt, W. & Vonck, J. Atomic model of the F_{420}-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* **3**, e01963 (2014).
36. Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam, S. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc. Natl Acad. Sci. USA* **111**, 11709–11714 (2014).
37. Hattne, J. et al. Analysis of global and site-specific radiation damage in cryo-EM. *Structure* **26**, 759–766 (2018).
38. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
39. Young, I. T. & Van Vliet, L. J. Recursive implementation of the Gaussian filter. *Signal Process.* **44**, 139–151 (1995).
40. Loshchilov, I. et al. Fixing weight decay regularization in Adam. Preprint at https://arxiv.org/abs/1711.05101 (2017).
41. Smith, L. N. Cyclical learning rates for training neural networks. In *Proc. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* 464–472 (IEEE, 2017).
42. Gao, Z., Tan, C. & Li, S. Z. Foldtoken4: Consistent & hierarchical fold language. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.04.606514 (2024).
43. Ingraham, J. B. et al. Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023).
44. Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K. & Kerdprasop, N. An empirical study of distance metrics for *k*-nearest neighbour algorithm. In *Proc. 3rd International Conference on Industrial Application Engineering* Vol. 2, p. 4 (Institute of Industrial Applications Engineers, 2015).
45. Giri, N., Wang, L. & Cheng, J. Cryo2StructData: full dataset. *Harvard Dataverse* https://doi.org/10.7910/DVN/FCDG0W (2023).
46. Giri, N., Wang, L. & Cheng, J. Cryo2StructData: test dataset. *Harvard Dataverse* https://doi.org/10.7910/DVN/2GSSC9 (2023).
47. Wang, J. cryofold_source_data.zip. *figshare* https://doi.org/10.6084/m9.figshare.28530359.v1 (2025).
48. Wang, J. & Tan, C. End-to-end Cryo-EM complex structure determination with high accuracy and ultra-fast speed. *Zenodo* https://doi.org/10.5281/zenodo.14970359 (2025).
49. Hodson, T. O. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geosci. Model Dev.* **15**, 5481–5487 (2022).

Acknowledgements

This work was supported by National Science and Technology Major Project (grant no. 2022ZD0115101), National Natural Science Foundation of China Project (grant no. 624B2115, and grant no. U21A20427), Project (grant no. WU2022A009) from the Center of Synthetic Biology and Integrated Bioengineering of Westlake University and Project (grant no. WU2023C019) from the Westlake University Industries of the Future Research Funding.

Author contributions

J.W. conceived the idea and developed the framework. Z.G. provided the crucial technology for SE(3) GNN and structure generation, as well as the pretraining dataset. J.W. drafted the paper and C.T. helped in writing Methods. C.T., Z.G., Y.Z., G.Z. helped in editing the paper. J.W. and C.T. prepared codes and released them on GitHub. S.Z.L. supervised the project and helped revise the paper.

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/s42256-025-01056-0.

Correspondence and requests for materials should be addressed to Cheng Tan or Stan Z. Li.

Peer review information *Nature Machine Intelligence* thanks the anonymous reviewers for their contribution to the peer review of this work.

Jue Wang[1,5], Cheng Tan[1,5] ✉, Zhangyang Gao[1,5], Guijun Zhang[2], Yang Zhang[3,4] & Stan Z. Li[1] ✉

[1]AI Laboratory, Research Center for Industries of the Future, Westlake University, Hangzhou, China. [2]College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. [3]Department of Computer Science, School of Computing, National University of Singapore, Singapore, Singapore. [4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [5]These authors contributed equally: Jue Wang, Cheng Tan, Zhangyang Gao. ✉e-mail: chengtan9907@gmail.com; Stan.ZQ.Li@westlake.edu.cn