nature structural & molecular biology

Technical Report

https://doi.org/10.1038/s41594-025-01713-3

CryoAtom improves model building for cryo-EM

Received: 21 April 2025

Accepted: 16 October 2025

Baoquan Su[1], Kun Huang[2], Zhenling Peng[1]✉, Alexey Amunts[3,4]✉ & Jianyi Yang[1]✉

Constructing atomic models from cryogenic electron microscopy (cryo-EM) density maps is essential for interpreting molecular mechanisms. Here we present CryoAtom, an approach for de novo model building for cryo-EM maps, leveraging recent advancements in AlphaFold2 to improve the state-of-the-art method ModelAngelo. To accommodate the cryo-EM map information, CryoAtom replaces the global attention mechanism in AlphaFold2 with local attention, which is further enhanced by a novel three-dimensional rotary position embedding. CryoAtom produces more complete models, reduces the resolution requirement and accelerates the modeling. The application of CryoAtom to three large maps demonstrates its ability to detect previously uncharacterized proteins with unknown functions, improve the modeling of conformational changes and compartmentalize the map to isolate nonprotein components. A particular case includes a 104-protein complex that was modeled within a few hours and a minor conformational change of a single protein domain was detected at the periphery when models from two different maps were compared. CryoAtom stands as an accurate method currently available for model building of proteins in cryo-EM structure determination. The source code and model parameters are available from GitHub (https://github.com/YangLab-SDU/CryoAtom).

With the advent of direct electron detectors, single-particle cryogenic electron microscopy (cryo-EM) has become the primary method for resolving biomolecular structures[1–3]. A critical step in cryo-EM structure determination is the construction of atomic models from three-dimensional (3D) density maps. The pace of model deposition derived from cryo-EM maps in the Protein Data Bank (PDB) has increased by two orders of magnitude, from 67 in 2012 to 5,788 in 2024 (ref. 4). The exponential growth has been driven by advances in automation of data collection and processing[5], accessible software and web platforms empowering broader adoption[6] and community efforts to further democratize and improve cryo-EM methodologies[7], including the use of artificial intelligence.

Previously, model building relied heavily on manual efforts, involving fitting homology models generated by tools such as I-TASSER[8] into the density maps using graphic software, such as Coot[9] or Isolde[10]. The process of generating a 3D reconstruction involves averaging thousands of nonidentical particle projections, resulting in cryo-EM maps that consist of regions with variable resolutions. Consequently, areas with higher flexibility or partial occupancy are more challenging to interpret with precision.

Several methods have been developed to derive atomic structures from cryo-EM density maps. These can be categorized into two groups: traditional and deep learning. Traditional model-building methods include Pathwalking[11,12], Buccaneer[13], EM-Fold[14], Rosetta[15], MAINMAST[16]

---

[1]MOE Frontiers Science Center for Nonlinear Expectations, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China. [2]School of Life Science, Westlake University, Hangzhou, China. [3]University of Münster, Münster, Germany. [4]Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Dortmund, Germany. ✉e-mail: zhenling@email.sdu.edu.cn; alexey.amunts@gmail.com; yangjy@sdu.edu.cn

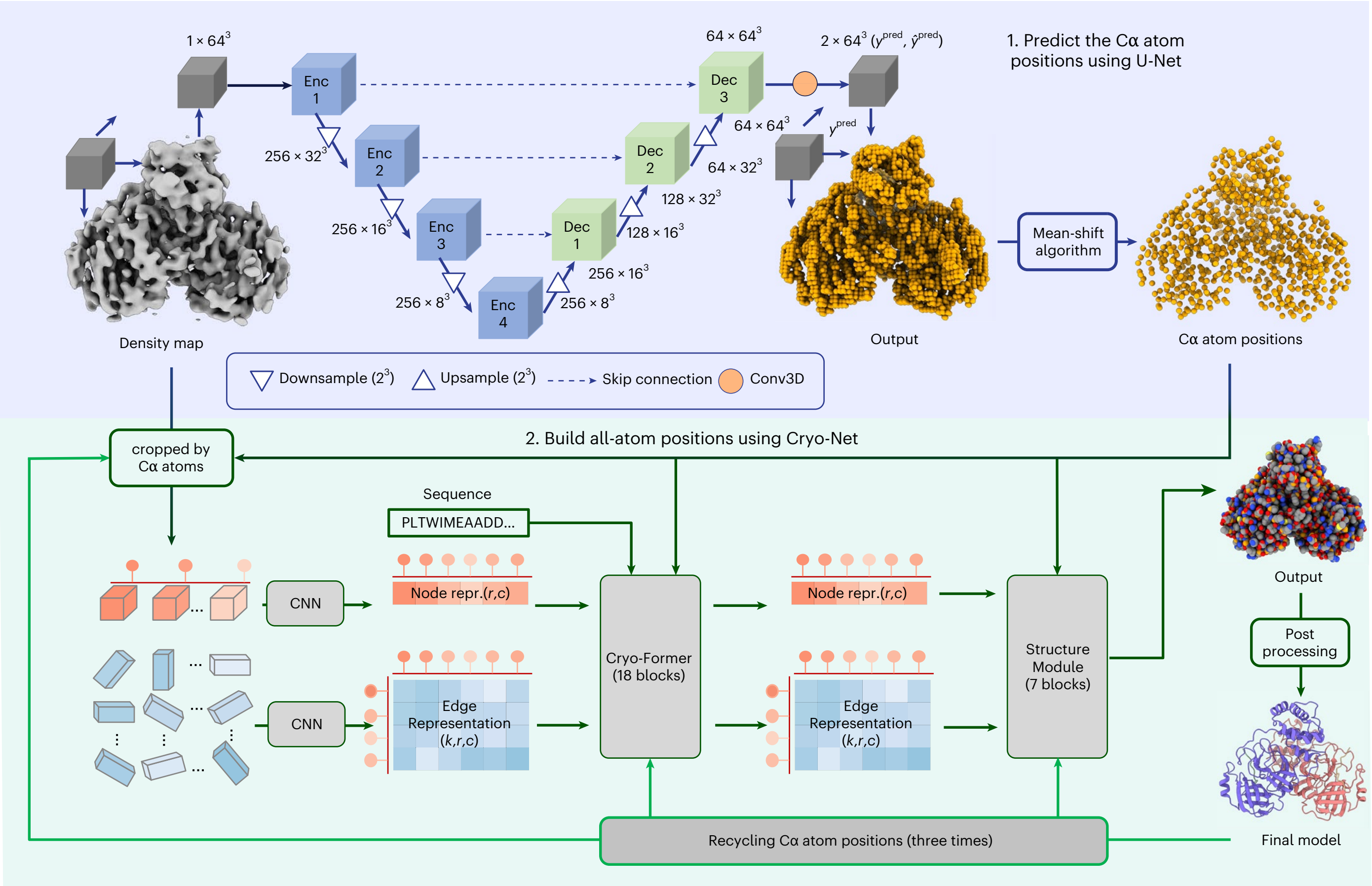


Fig. 1 | Overall architecture of CryoAtom. Two major steps are used to build all-atom structure from a cryo-EM density map and the amino acid sequences of the target proteins. In the first step (purple background), the input density map is fed into a 3D convolution-based network U-Net to predict the Cα atom positions. The map is first cropped continuously into boxes of size 64 × 64 × 64, which are then processed by U-Net and combined at the end. The final U-Net output y^{pred} is of the same size to the input map and the value in each voxel is the probability of containing a Cα atom, which is refined by the mean-shift algorithm. In the second step (green background), all-atom positions are predicted using a local attention network (Cryo-Net) from the amino acid sequence and the predicted Cα atom. The initial node and edge representations are first fetched from density map guided by the predicted Cα atom positions. They are then fed into the Cryo-Former together with the sequence embedding from ESM-2. The updated representations are then decoded into all-atom structure using the Structure Module, followed by a postprocessing step. The parameters *r*, *k* and *c* are the numbers of predicted Cα atoms, neighbors considered and channels, respectively.

and PHENIX[17]. In Pathwalking and MAINMAST, constructing all-atom structures from density maps is treated as a minimization problem, solved using optimization algorithms. Rosetta and EM-Fold first deduce fragment structures from the density map and assemble them using Monte Carlo sampling.

Deep-learning-based methods encompass DeepMainmast[18], EMBuild[19], CR-I-TASSER[20], DeepTracer[21], ModelAngelo[22] and Cryo-2Struct[23]. These approaches typically start by applying a deep network to detect the backbone structure from the density map. They diverge in their subsequent strategies for constructing all-atom models. By incorporating components of protein structure prediction framework, current methods have achieved more accurate results in interpreting density maps[24].

Recently, a breakthrough came from deep-learning-based structure predictors, such as AlphaFold2 (AF2)[25], RoseTTAFold[26] and trRosetta[27], which also contributed to the awarding of the 2024 Nobel Prize in Chemistry. We reasoned that accuracy and the resolution limit in protein model building, especially for the state-of-the-art method ModelAngelo, could be further extended by taking advantage of the recent advances. We, thus, developed CryoAtom. Tests on cryo-EM maps of multiprotein complexes with unknown components demonstrate the robustness of CryoAtom for automated model building with high speed and accuracy.

Overview of the CryoAtom approach

CryoAtom takes as input a cryo-EM density map and the amino acid sequences of the target proteins, following a two-step process for building atomic structures (Fig. 1). In the first step, a 3D convolution-based network, specifically a U-Net[28], is used to predict the positions of Cα atoms from the density map. Because of GPU memory limitations, the map is cropped into smaller patches (64 × 64 × 64), which are later combined to reconstruct the original size. The second step generates all-atom structures using transformer network adapted from AF2 (ref. 25). The all-atom positions are post-processed using a method similar to that used in ModelAngelo[22]. CryoAtom incorporates several key concepts that are inspired by ModelAngelo (Box 1) and we introduced key modifications in the encoder and decoder to fully exploit the structural information present in the density map. The encoder, which we call Cryo-Former, transforms the density map into a node-and-edge representation. The decoder, named Structure Module, generates all-atom positions from these improved representations.

Two key differences distinguish the CryoAtom network (denoted by Cryo-Net) from the AF2 network: an enhanced transformer and the local attention. Cryo-Net features an enhanced transformer that incorporates 3D rotary position embedding (3D-RoPE).

BOX 1

Key concepts of CryoAtom inspired by ModelAngelo

Method	ModelAngelo	CryoAtom
Architecture	Cryo-EM module, sequence module and structure module with spatial IPA	Cryo-Former (integrating cryo-EM module and sequence module using an evoformer-like architecture) and structure module (with local attention and 3D-RoPE)
Representations to update	Node representation	Node representation and edge representation
Attention characteristics	Local, cross-attention	Local, gated cross-attention and gated self-attention
Positional encoding	Sinusoidal	3D-RoPE
Pretrained language model	ESM-1b	ESM-2
Confidence score	Per-residue backbone r.m.s.d.	Per-residue FAPE
Recycling mechanism	Recycling backbone frame	Following ModelAngelo
Sequence alignment	Based on HMM	Following ModelAngelo

Originally, RoPE was used to process one-dimensional (1D) sequences[29]. Here, we extend it to a node position encoding in 3D space (Methods and Supplementary Text 2). 3D-RoPE effectively encodes the positional information of each node into all attention calculations, allowing the attention score to decay with increasing distance between nodes. This design takes advantage of the spatial constraints inherent in the density map. The local attention in Cryo-Net is realized through the improved edge representation (Methods). Instead of an all-against-all approach, Cryo-Net uses an all-against-*k* strategy, where *k* represents the number of spatial neighbors for each node. This adjustment once again capitalizes on the spatial restraints provided by the density map, which are usually unavailable in traditional structure prediction. These features enable efficient training of a deeper network.

The postprocessing procedure, adapted from ModelAngelo, is outlined as follows. The initial all-atom structure is obtained from Cryo-Net and any segments with fewer than four residues are trimmed (referred to as model_net). The amino acid type for each residue in model_net is assigned as the class with the highest probability output from Cryo-Net. The hidden Markov model (HMM) profile is derived from the predicted probability distribution by Cryo-Net and aligned with the input amino acid sequence. The amino acid types of residues in the Cryo-Net structure (before trimming) are then corrected according to this alignment, producing another structure denoted as model_fix. Lastly, residues that do not match the input sequence are pruned, producing model_prune, the default model used for assessment unless otherwise specified.

Quality of the model building for 1–4-Å-resolution maps

To benchmark CryoAtom, we first tested it against 177 maps of resolution better than 4 Å from ModelAngelo. The running time of CryoAtom is presented in Supplementary Fig. 1 and Supplementary Text 1. The specific metrics for performance evaluation are introduced in Supplementary Text 1. For objective comparison, the ModelAngelo models for these maps were downloaded from the website given in the original study[22]. For new maps, atomic models were generated locally using version 0.2.3, which had higher accuracy than other versions in our tests.

While both CryoAtom and ModelAngelo performed well when the map resolution was better than 3 Å (Fig. 2a, top right), CryoAtom's advantages became evident as the map resolution decreased below 3 Å. Notably, with a higher completeness rate, CryoAtom also generated more accurate structures, as indicated by the Cα root-mean-square deviation (r.m.s.d.; 0.31 Å versus 0.33 Å) and backbone r.m.s.d. (0.34 Å versus 0.37 Å) (Fig. 2e).

Completeness is closely linked to backbone recall and amino acid accuracy. A decline in either of these metrics results in reduced completeness. To further investigate this relationship, we calculated backbone recall and amino acid accuracy for the intermediate models (model_net and model_fix). CryoAtom achieved a backbone recall in model_net of 97.2% versus 94.4% of ModelAngelo (Fig. 2g). If segments fewer than four residues were retained, this parameter increased by 0.8% and 2.0% for CryoAtom and ModelAngelo, respectively. The smaller increase for CryoAtom might reflect that the structures generated by our network were less fragmented.

Additionally, CryoAtom exhibited a higher amino acid accuracy before pruning (in both model_net and model_fix; Fig. 2g) and the two methods showed comparable accuracy after pruning (in model_prune; Fig. 2d). The amino acid accuracy for the model generated by Cryo-Net (model_net) was 73.1% compared to 58.5% for ModelAngelo. This indicates that our network possesses a stronger ability to assign amino acid types, likely because of the enhanced transformer architecture. The HMM-based correction from ModelAngelo performs well, improving amino acid accuracy in model_fix to 87.1% for CryoAtom and 84.7% for ModelAngelo. In summary, CryoAtom showed high backbone recall and amino acid accuracy, contributing to the overall completeness.

We also analyzed the accuracy of predicted structures on the basis of the *Q* scores[30] of residues in the deposited structures. The *Q* score reflects the consistency between the map density and the atomic structure, with higher values typically indicating better local resolution. Figure 2c plots the backbone r.m.s.d. and the completeness of models built by CryoAtom and ModelAngelo against the deposited *Q* scores. When the *Q* score was less than 0 or greater than 0.6, CryoAtom models exhibited lower backbone r.m.s.d. For other *Q*-score ranges, the r.m.s.d. of both methods was comparable. In terms of completeness, CryoAtom outperformed ModelAngelo when the *Q* score was below 0.6.

In addition to *Q* scores, we calculated the MolProbity scores[31] and EMRinger[32] metrics. The former reflects the geometric validity of the protein models, whereas the latter indicates the degree of agreement between the protein side chains and the density map. CryoAtom achieved better model quality than ModelAngelo while having a higher level of completeness (MolProbity scores, 3.56 versus 3.69; EMRinger, 3.02 versus 2.90). Detailed results can be found in Supplementary Table 1.

We compared the number of true positives (TPs) in the structures generated by both methods. A residue was classified as a TP if

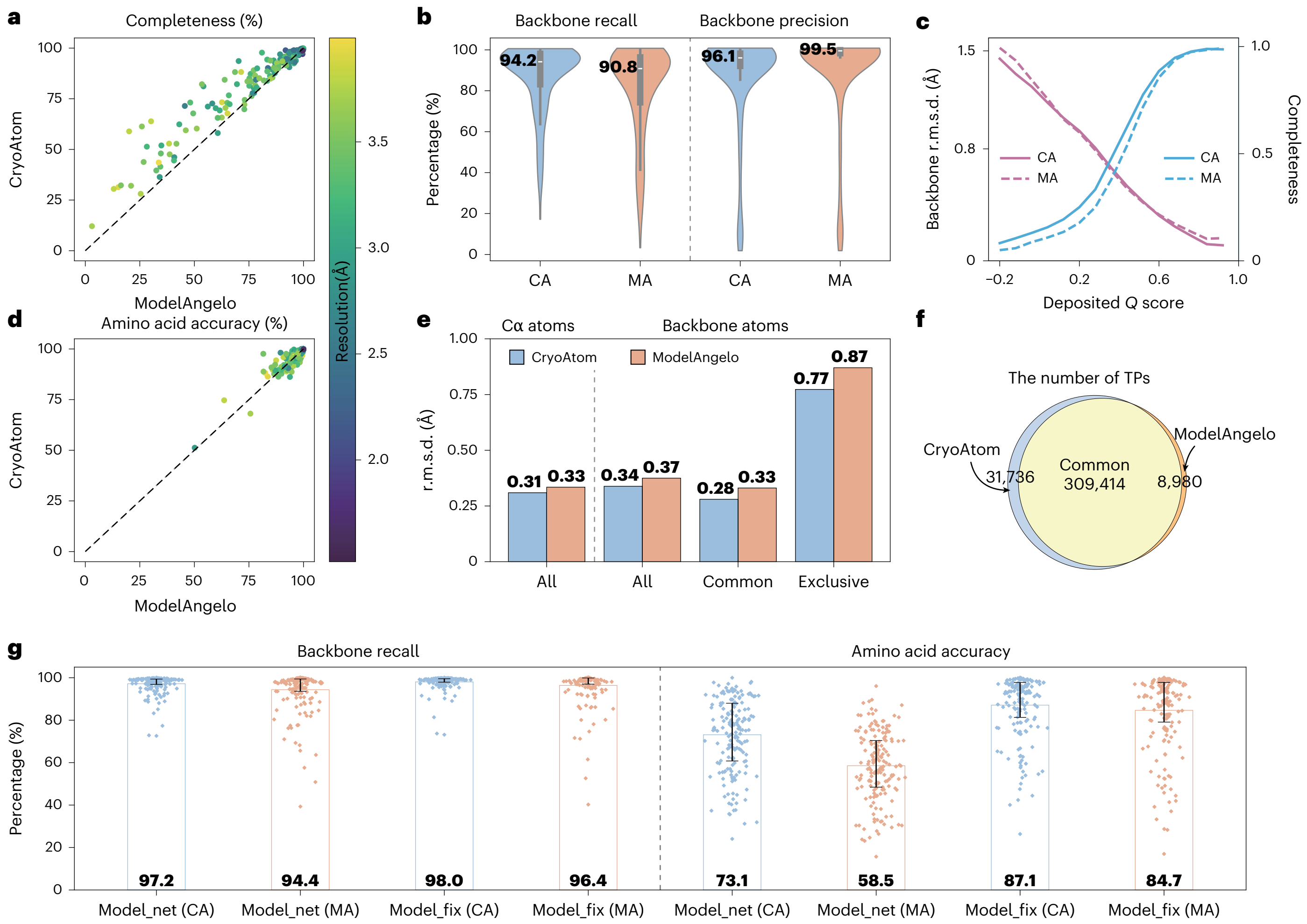


Fig. 2 | Comparison between CryoAtom and ModelAngelo on the test set of 177 high-resolution maps. **a**, Head-to-head comparison based on the completeness. **b**, Violin plots of the backbone recall and the backbone precision ($n = 177$). The shape of the violin plot indicates the distribution. The center, lower and upper lines in each box indicate the median, first quartile and third quartile, respectively. The whiskers extend to the most extreme data points that are within 1.5 times the interquartile range (IQR) from the first and third quartiles. Data points beyond this range are considered outliers. **c**, Backbone r.m.s.d. and completeness plotted as a function of the Q scores of the deposited structures. **d**, Head-to-head comparison of the amino acid accuracy. **e**, Bar plots comparing the average r.m.s.d. of Cα atoms and backbone atoms. The sizes of the sets of TP residues (all, common, exclusive) are shown in **f**. **f**, Venn diagram showing the numbers of TP residues predicted by both methods. **g**, Backbone recall and amino acid accuracy for the intermediate models ($n = 177$). Error bars indicate ±1.0 s.d.

a corresponding deposited residue existed within 3 Å. Both methods had 309,414 common TPs, with backbone r.m.s.d. values of 0.28 Å and 0.33 Å for CryoAtom and ModelAngelo, respectively (Fig. 2e,f). Additionally, CryoAtom identified 31,736 TPs that ModelAngelo missed, contributing to its higher completeness. These residues generally corresponded to lower-resolution areas, as indicated by their lower deposited Q scores (0.50; Supplementary Fig. 2a) and higher backbone r.m.s.d. (0.77 Å; Fig. 2e). In contrast, ModelAngelo identified 8,980 TPs (with a higher deposited Q score of 0.56; Supplementary Fig. 2a) that were missed by CryoAtom. Although the TPs constructed by ModelAngelo but missed by CryoAtom were located in high-resolution regions (average deposited Q score of 0.56, corresponding to a resolution of 3.1 Å)[30], the average backbone r.m.s.d. of these regions in the ModelAngelo models was higher than the overall backbone r.m.s.d. (0.87 Å versus 0.37 Å; Fig. 2e). This reflects the challenges of building atomic structure for these regions.

We investigated whether there is a systematic bias in the construction of these regions. We defined the interchain interface as Cα atoms from different chains that were within 8 Å of each other. Terminal segments were defined as the first and last 10% segments of the entire chain that did not belong to interface regions. The results, shown in Supplementary Fig. 2b,c, indicate that, among the TPs (8,980) missing by CryoAtom but constructed by ModelAngelo, approximately 21% were located in interface regions, 9% belonged to terminal segments and 58% of the secondary structure constituted coils. The high flexibility of coil segments may have contributed to the difficulties in modeling these regions. Correspondingly, among the CryoAtom TPs (31,736) missed by ModelAngelo, about 14% were located in interface regions, 9% belonged to terminal segments and 50% of the secondary structure constituted coils. The systematic bias in both methods was similar. These data reinforce CryoAtom's advantage in constructing atomic structures for lower-resolution maps or regions.

Lastly, we analyzed the sensitivity of CryoAtom to hyperparameters and evaluated its performance across different biological systems. The sensitivity analysis of CryoAtom to hyperparameters can be found in Supplementary Fig. 3 and Supplementary Text 1. Details of CryoAtom's performance across different biological systems are shown in Supplementary Fig. 4. CryoAtom achieved a completeness of 83.8% across the entire high-resolution test set (average resolution of 3.1 Å), 85.9% for membrane proteins (average resolution of 3.0 Å), 88.7% for monomers (average resolution of 2.9 Å), 84.9% for homooligomer complexes (average resolution of 3.1 Å), 82.4% for heterooligomer

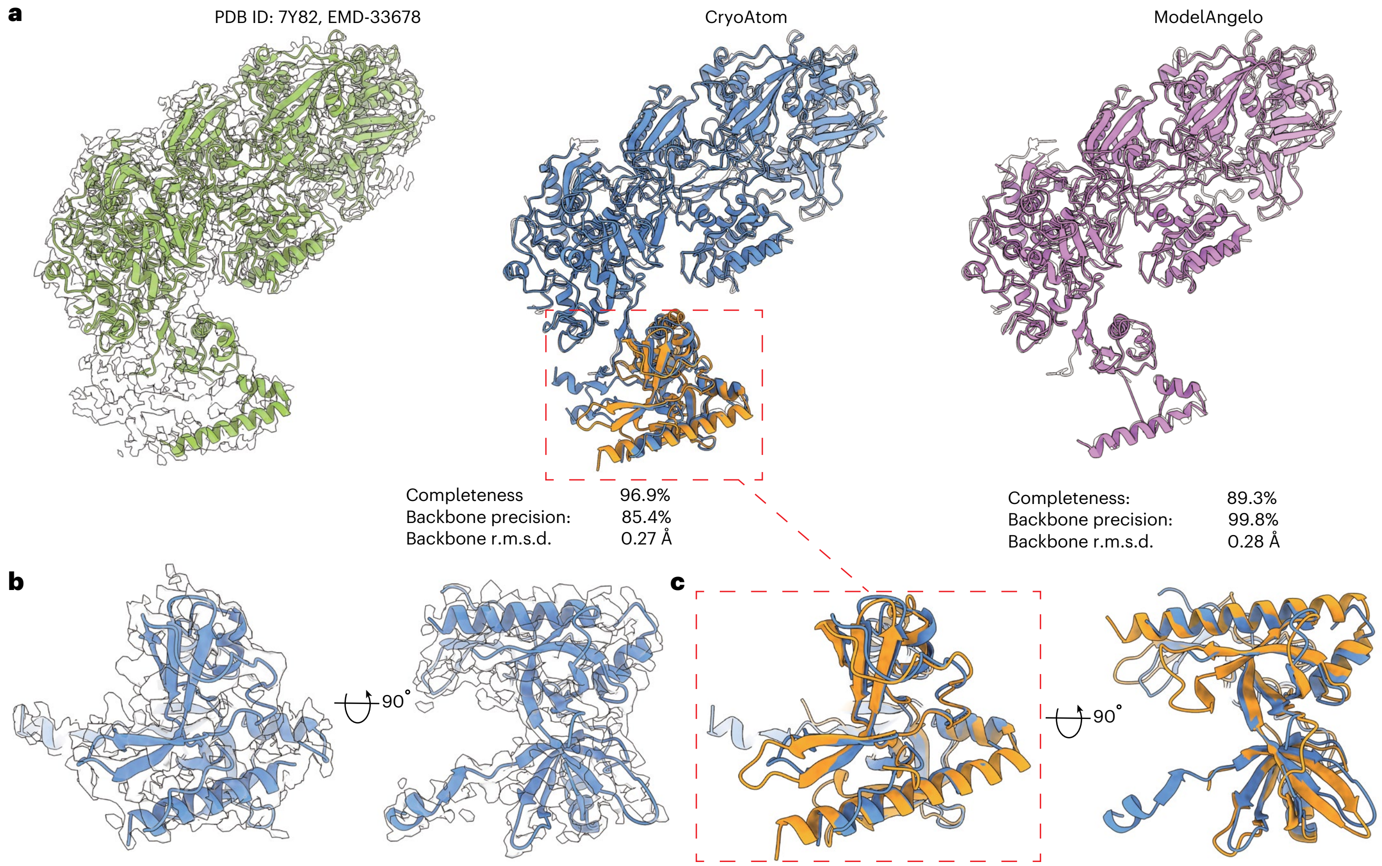


Fig. 3 | An example of how CryoAtom has a more complete model compared to the deposited structure. **a**, The gray surface is the density map EMD-33678 (reported resolution: 2.83 Å). The green/gray cartoon is the deposited structure (PDB 7Y82). The blue cartoon is the CryoAtom model, the magenta cartoon is the ModelAngelo model and the orange cartoon is the AF3 model for the unmodeled regions in the deposited structure. **b**, The blue cartoon is the modeling of the absent residues 1041–1380 in the deposited structure by CryoAtom and the gray surface is the corresponding density map. **c**, An enlarged view of the red box in **a**.

complexes (average resolution of 3.1 Å) and 85.1% for protein complexes bound to nucleic acids (average resolution of 3.2 Å). These data suggest that CryoAtom demonstrates good robustness across various biological systems.

CryoAtom extends the map resolution limit for structure construction

Our tests indicate that CryoAtom exhibited slightly lower backbone precision compared to ModelAngelo (Fig. 2b). There are two possible reasons for this. The first is that map masking is not used in CryoAtom (Supplementary Fig. 5 and Supplementary Text 1). The second may stem from CryoAtom's ability to construct structures in lower-resolution regions, as discussed below, which are often absent in deposited structures and classified as false positives during evaluation. For instance, in the cryo-EM structure of the type III-E CRISPR Craspase gRAMP–crRNA complex (PDB 7Y82, EMD-33678; green cartoon and gray surface in Fig. 3a), residues 1041–1380 are largely absent in both the deposited structure (green cartoon in Fig. 3a) and the ModelAngelo model (magenta cartoon in Fig. 3a). The structures for these regions are modeled by CryoAtom (blue cartoon in Fig. 3a,b) but classified as false positives, resulting in a lower precision than ModelAngelo (85.4% versus 99.8%). However, these residues are modeled with high confidence (>70). The definition of CryoAtom's confidence can be found in Supplementary Fig. 6 and Supplementary Text 1. We also submitted the domain sequence to AF3 (ref. 33) for structure prediction, yielding a high confidence score (AF3 confidence score > 0.8). The structure superimposition indicates that the AF3 model closely resembled the CryoAtom model (template modeling (TM)-score, 0.86; r.m.s.d., 2.07 Å; Fig. 3c). This cross-validation supports the correctness of the CryoAtom model for the missing residues.

The above phenomenon is not isolated and we observed it in numerous cases, potentially impacting overall backbone precision. On average, the CryoAtom models for the 177 maps included approximately 75,000 new residues predicted with high confidence (>60; Supplementary Fig. 7a) absent in deposited structures. All structures generated by CryoAtom are available for download (https://yanglab.qd.sdu.edu.cn/CryoAtom). On the other hand, we investigated the *Q* scores of these false positives (FPs), as shown in Supplementary Fig. 7b. For CryoAtom, only 18.6% of these FPs had a *Q* score below 0.4, indicating that most (that is, 81.4%) of the newly constructed regions, which are not present in the deposited structures, are likely reasonable. If we also classified residues with a *Q* score > 0.4 as TPs, then the backbone precision of CryoAtom would increase from 96.1% (Fig. 2b) to 98.2%; correspondingly, the backbone precision of ModelAngelo would increase from 99.5% (Fig. 2b) to 99.9%.

Benchmarking performance on 4–7-Å-resolution maps

Next, we assessed the performance of CryoAtom in a lower resolution range. We tested 104 maps with resolutions of 4–7 Å (Fig. 4a–c). As expected, the accuracy was lower compared to high-resolution maps. Nevertheless, CryoAtom built relatively complete models (>50% completeness) for approximately 40% of the tested maps. CryoAtom generally constructed more complete models than ModelAngelo

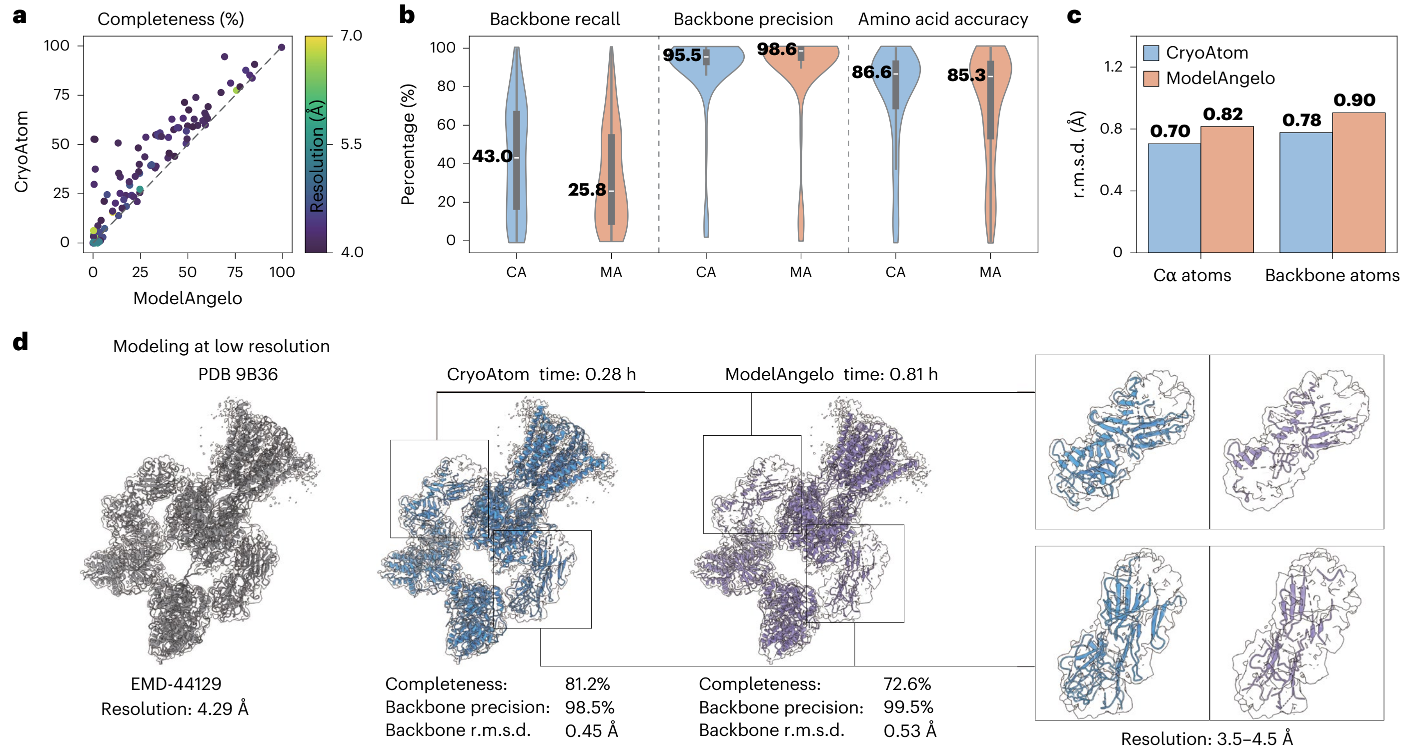


Fig. 4 | Comparison of CryoAtom and ModelAngelo on 104 test maps with resolutions of 4–7 Å. a, Head-to-head comparison of the completeness. **b**, Violin plots of backbone recall, backbone precision and amino acid accuracy (*n* = 104). The shape of the violin plot indicates the distribution. The center, lower and upper lines in each box indicate the median, first quartile and third quartile, respectively. The whiskers extend to the most extreme data points that are within 1.5 times the IQR from the first and third quartiles. Data points beyond this range are considered outliers. **c**, Bar plot of the main-chain atom r.m.s.d. **d**, Modeling at low-resolution regions showing comparison between the deposited model (PDB 9B36; gray cartoon, left), the CryoAtom model (blue cartoon, middle) and the ModelAngelo model (purple cartoon, right) with corresponding close-up views. The density map (EMD-44129; reported resolution of 4.29 Å) is shown in transparent representation.

(backbone recall, 43.0% versus 25.8%; completeness, 36.8% versus 23.6%) while maintaining similar accuracy (amino acid accuracy, 86.6% versus 85.3%; backbone r.m.s.d., 0.78 Å versus 0.90 Å). An illustrative low-resolution example is provided in Fig. 4d, comparing models generated by CryoAtom and ModelAngelo. The case involves an open state of kainate receptor GluK2 (PDB 9B36) with 4.3-Å resolution. CryoAtom successfully constructed structures for 81.2% of the residues, achieving backbone r.m.s.d. of 0.45 Å. In contrast, ModelAngelo constructed structures for 72.6% of the residues, achieving backbone r.m.s.d. of 0.53 Å. Because of the lower resolution in the peripheral regions of the density map, ModelAngelo encountered difficulties during modeling. However, CryoAtom constructed relatively complete atomic models for these regions (Fig. 4d). In addition, CryoAtom took less time than ModelAngelo for the modeling (0.28 h versus 0.81 h). This example clearly demonstrates CryoAtom's superior modeling capability and efficiency for handling low-resolution regions within maps. The assessment on a nonredundant set of maps indicated that the key factor impacting CryoAtom's performance was the map resolution rather than the sequence similarity or structural similarity to the training data (Supplementary Fig. 8 and Supplementary Text 1).

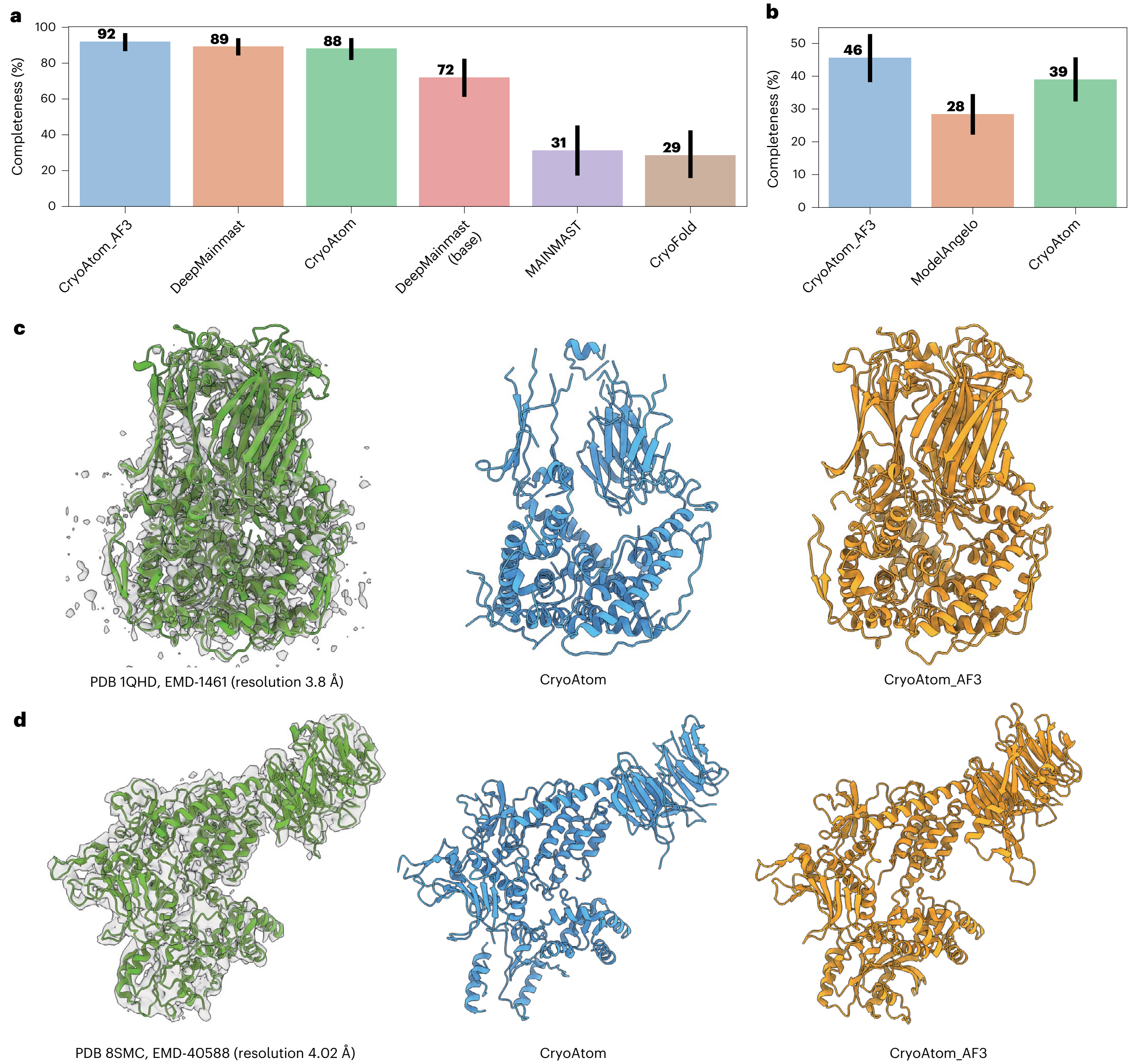
Comparison to other methods that use additional constraints

We further compared CryoAtom to DeepMainmast[18] that incorporates additional constraints. DeepMainmast constructs complex structures from density maps using deep learning and predicted structures from AF2. The DeepMainmast protocol comprises two components. The first, DeepMainmast(base), constructs models using de novo main-chain tracing guided by deep learning. The second integrates the DeepMainmast(base) model with structure predictions from AF2.

Modeling a target with DeepMainmast typically requires several hours. For convenience, we referenced the data from the original DeepMainmast paper[18], which included results for 20 protein complexes. In addition, the DeepMainmast paper also included the results of two classical methods, CryoFold[34] and MAINMAST[16], which perform de novo modeling on the basis of cryo-EM density maps. Here, we cite their results on single-chain datasets and reweight the completeness (referred to as AA match in the DeepMainMast paper) according to chain length to obtain the results for the complexes. We found that CryoFold failed to run in certain cases; as MAINMAST is a step within the CryoFold pipeline, the results from MAINMAST were used directly for these cases of failure. Additionally, we found that using the original density map (EMD-5495) directly as input for CryoAtom yielded very poor results. This may be related to the fact that EMD-5495 was released early (in 2012) with high background noise. Thus, we preprocessed EMD-5495 in the same way as ModelAngelo and then used it as input for CryoAtom.

First, under a fair comparison without considering additional inputs, the result is shown in Fig. 5a. The average completeness of CryoAtom was 0.88, demonstrating a notable advantage over DeepMainmast(base) (0.72), MAINMAST (0.31) and CryoFold (0.29). When compared to the full DeepMainmast protocol, CryoAtom remained competitive (completeness, 0.88 versus 0.89). In the DeepMainmast paper, the density map with the shortest runtime (14.85 h) in the dataset was EMD-5185, while the density map with the longest runtime (613.66 h) was EMD-6374. In contrast, CryoAtom took only 4.44 h and 82.77 h for these two density maps, respectively.

The comparisons above demonstrate that CryoAtom remained competitive with DeepMainmast, despite the additional constraints used by the method. We intentionally avoided incorporating extra



PDB 1QHD, EMD-1461 (resolution 3.8 Å) CryoAtom CryoAtom_AF3

PDB 8SMC, EMD-40588 (resolution 4.02 Å) CryoAtom CryoAtom_AF3

Fig. 5 | Enhancement of CryoAtom through the introduction of AF3-predicted structure models. CryoAtom and CryoAtom_AF3 represent the default version of CryoAtom and the version enhanced with the AF3 model, respectively. **a**, Bar plots comparing completeness using DeepMainmast, CryoFold and MAINMAST for *n* = 20 protein complexes. Data are the mean and the 95% confidence interval of the mean, estimated from 1,000 bootstrap samples. **b**, Bar plots comparing completeness using CryoAtom_AF3, CryoAtom and ModelAngelo on *n* = 69 proteins from low-resolution test sets. The height of each bar graph represents the average value, while the error bars were calculated on the basis of the 95% confidence interval. **c**,**d**, Example models built with CryoAtom and CryoAtom_AF3 for two maps (EMD-1461 in **c** and EMD-40588 in **d**). The green structure represents the deposited model, the gray surface indicates the density map and the blue and orange structures represent the models constructed by the default CryoAtom and the enhanced CryoAtom, respectively.

constraints into CryoAtom to prioritize ease of use and efficiency. As a result, CryoAtom is several times faster than DeepMainmast. In fact, CryoAtom can also enhance its modeling performance by integrating additional inputs. We conducted a preliminary test to explore the possibility of enhancing CryoAtom for density maps with poor resolution by integrating predicted structure models. To this end, we introduce a variant of CryoAtom that includes an additional input from the AF3-predicted structure model (Supplementary Text 2) and we refer to this AF3-enhanced version as CryoAtom_AF3.

The incorporation of predicted structure models improved the models built by CryoAtom (Supplementary Table 7). On these 20 targets, the completeness of CryoAtom models was very close to that of the DeepMainmast models (0.88 versus 0.89). However, after integrating the AF3 models, the completeness of the CryoAtom_AF3 models improved to 0.92, outperforming DeepMainmast (Fig. 5a). An example from this dataset is illustrated in Fig. 5c (PDB 1QHD). The density map had medium resolution (3.8 Å), with the top regions at even lower resolution, making it challenging to build a complete model. CryoAtom

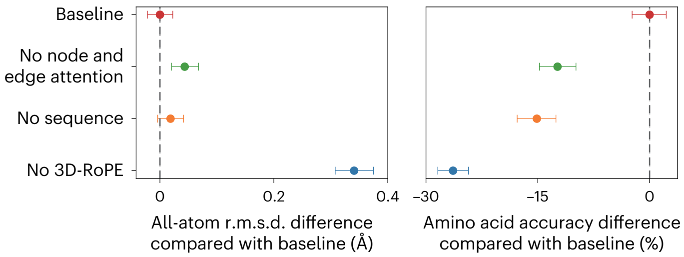


Fig. 6 | Results of ablation study. The experiments were conducted on 177 high-resolution maps. Left: all-atom r.m.s.d. Right: amino acid accuracy. The mean is indicated by the center point. The error bar was calculated on the basis of the 95% confidence interval.

only constructed the structures with a completeness of 0.66. This issue is not unique to our method; it is a common phenomenon for all methods that rely solely on the density map, including ModelAngelo and DeepMainmast(base). After incorporating the predicted models from AF3, we successfully built a model very close to the deposited version, achieving over 0.95 completeness.

Furthermore, to validate the robustness of the CryoAtom_AF3 version, we selected a larger low-resolution dataset for testing. To avoid the long time required for structure prediction, we chose density maps with fewer than 5,000 residues from the original low-resolution test set (selecting 69 of 104). When incorporating the predicted AF3 models, the average completeness of CryoAtom_AF3 was 0.46 (Fig. 5b), demonstrating a notable advantage over CryoAtom (0.39) and ModelAngelo (0.28). An example from this dataset is illustrated in Fig. 5d, where the structures in the low-resolution region at the periphery were successfully built by CryoAtom_AF3.

The data presented above indicate potential avenues for enhancing CryoAtom in the context of low-resolution density maps. However, low-resolution density maps still pose challenges. In regions with excessively low resolution, the default models provided by CryoAtom may be a collection of fragments, missing some core structures. For such fragmented results, simply using US-align[35] may lead to incorrect alignments because these fragments do not specify the correct chain order.

Ablating CryoAtom components to assess drivers of performance

To investigate the contributions of different components to CryoAtom, we conducted ablation experiments on the 177 high-resolution maps. Four configurations of its network, Cryo-Net, were assessed: the default version (18-layer Cryo-Former), a version without node and edge attention, a version without sequence input and a version without 3D-RoPE. These experiments evaluated the impact of network depth, sequence information and 3D-RoPE on performance. Metrics included amino acid accuracy and all-atom r.m.s.d., calculated against the deposited structures (Fig. 6). The default model achieved 0.51 Å all-atom r.m.s.d. and 73.1% amino acid accuracy. The r.m.s.d. increased slightly for shallower networks and the version without sequence input, while amino acid accuracy dropped substantially, especially for the version lacking sequence input. The configuration without 3D-RoPE exhibited the most substantial differences, with an increased all-atom r.m.s.d. of 0.34 Å and a 26.4% decrease in amino acid accuracy. Notably, the loss curve for the ablation experiment without 3D-RoPE displayed oscillations during training (Supplementary Fig. 9), suggesting that this encoding is crucial for the effective training of Cryo-Net.

Case studies to assess protein identification and conformational changes

We showcase the advanced capabilities of CryoAtom with three complementary examples, each addressing a distinct and challenging aspect of modern macromolecular modeling (Fig. 7). They are introduced below in detail.

Detecting new proteins in multicomponent complexes

A challenge in protein modeling for complex organisms is determining the identity of previously uncharacterized proteins directly from the cryo-EM map. Inspired by ModelAngelo, CryoAtom implements the function to identify unknown proteins by performing sequence searches using HMM profiles generated by deep learning. Specifically, an initial set of sequences, which can be completely random or part of the sequences related to the density map, can be fed into CryoAtom. The HMM profiles generated by CryoAtom are then compared to the sequences in the sequence database related to the density map (for example, downloaded from the UniProt database by organism information or simply using the whole UniProt database if nothing is known). Successfully matched sequences are added into the initial sequence set. CryoAtom is run again with the updated set of sequences. This iterative process continues until no new sequences are identified. The final set of sequences represents all sequences related to the density map. Once the sequence database is prepared, this process becomes fully automated and requires no manual intervention. This enables the automatic search of sequences within the density map from a vast sequence database, thereby constructing atomic models. The prerequisite for this process to work is that the provided sequence database must encompass all protein sequences in the density map. Given the rapid development of the genome sequencing projects, this assumption is not difficult to meet. We can use this function to identify unknown or missing subunits that human experts failed to build.

We tested CryoAtom's ability to tackle this using a reconstruction of a photosynthetic complex with multiple chains from a chromerid alga (EMD-66850)[36]. In this example, CryoAtom identified and modeled 22 proteins, six of which were not previously characterized (Fig. 7a). Additionally, two universally conserved chains exhibited fully resolved extensions of 71 and 99 residues, which effectively doubled the original size of each protein in the core complex, indicating potential functional relevance (Fig. 7b). The newly modeled sections were not previously identified and the new chains were not predicted to be part of the complex. CryoAtom's ability to detect new proteins and structural extensions within intricate cryo-EM data demonstrates its efficiency, with calculations taking ~1 h on an A800 GPU.

Capturing minor conformational changes in a megadalton complex

To explore CryoAtom's sensitivity to subtle conformational changes, we used a mitochondrial supercomplex functioning as a respirasome, with a mass of 1.8 MDa, in which only one protein displayed a slight rotation[37]. First, CryoAtom generated an atomic model from the map of the native source, constructing the 104-protein complex in 5.6 h, with half of the chains being species specific (Fig. 7c). Compared to ModelAngelo, CryoAtom provided higher completeness in lower-resolution regions of the map (Supplementary Fig. 10). To further assess CryoAtom's precision, we classified two maps and found a subtle 15-Å movement in a single peripheral domain of a protein. This minor motion was resolved in two conformations within the 261 × 296-Å supercomplex. The resolution in the area of this conformational shift was 3.0–4.0 Å. CryoAtom's ability to capture such minor structural variations demonstrates its effectiveness for investigating dynamic biological particles and targeting structural regulation.

Performance on a large protein–nucleic acid complex

Lastly, we tested whether CryoAtom works for large-scale biological macromolecules. As a representative example, we used the human mitoribosome, which has over 80 protein chains and a large overall dimensions of 307 × 287 × 286 Å (ref. 38). Unlike ModelAngelo, CryoAtom does not currently model nucleic acids; nevertheless, it

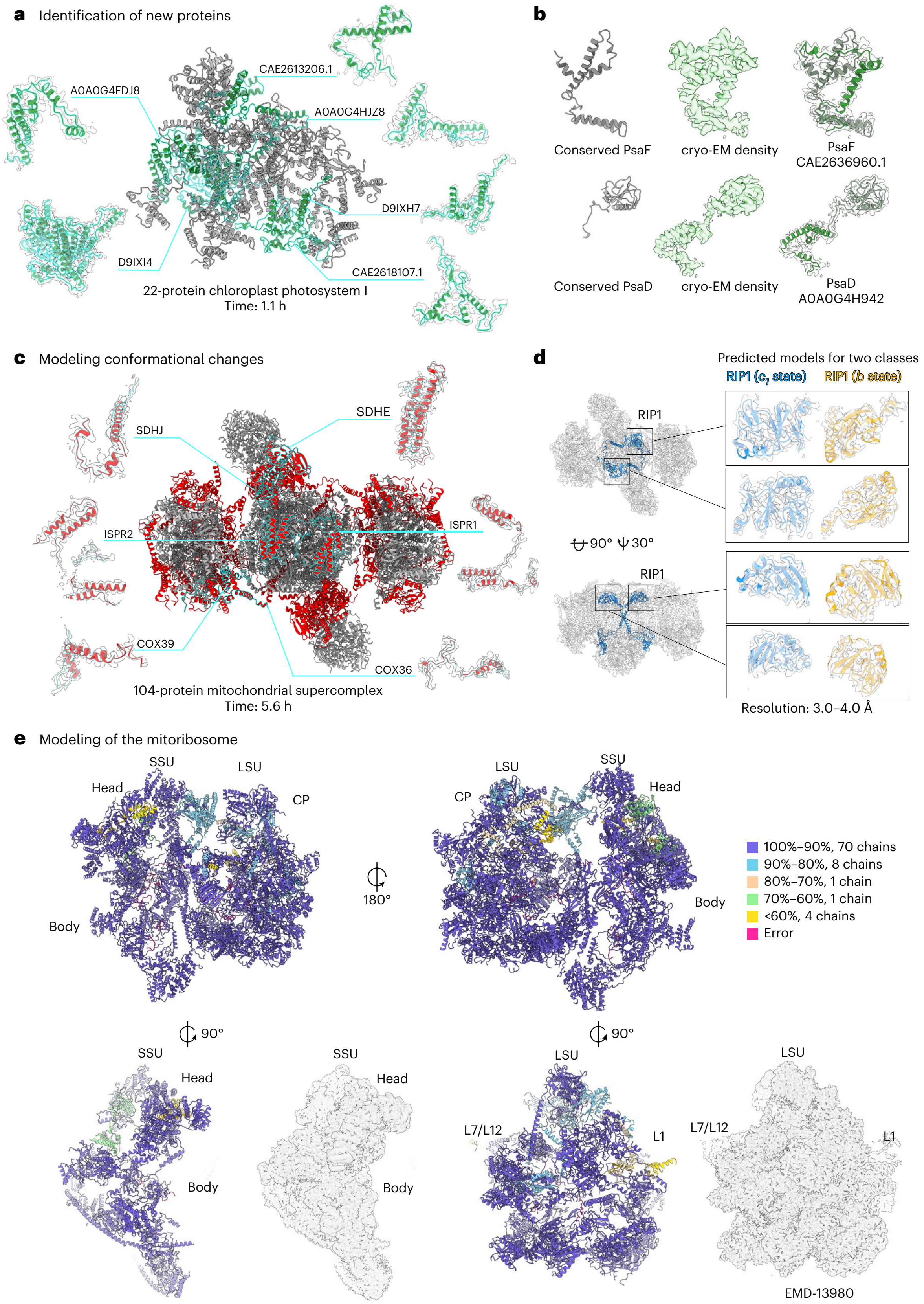


Fig. 7 | Capabilities of CryoAtom with three complementary examples. **a**, Identification of six new proteins (green) in a 22-protein photosystem complex, shown with their corresponding map densities. **b**, Two substantially extended proteins identified by CryoAtom in the same complex (right) compared to the canonical models (left). **c**, The CryoAtom model of the mitochondrial respirasome II$_2$–III$_2$–IV$_2$ (gray)[37] with 52 of 104 proteins that were not reported in other complexes (red). Six of these proteins are shown with their densities and compared to the ModelAngelo in Supplementary Fig. 10. **d**, Two distinct conformations (blue and yellow in close-up views) of a single protein (RIP1) in the 104-protein supercomplex modeled by CryoAtom in two classes. **e**, Overall view of the CryoAtom model of the human mitoribosome and its individual subunits. Proteins are colored according to model completeness. The corresponding density map (EMD-13980) is shown for individual subunits.

successfully segmented the heterogeneous density, where one third corresponded to nucleic acids[39]. Notably, 78 of 87 protein chains were modeled with >80% completeness (Fig. 7e). This example demonstrates that CryoAtom effectively handles large-scale structures containing a notable number of nucleic acids.

Discussion

In this study, we demonstrated the advanced capabilities of CryoAtom in addressing complex challenges in macromolecular modeling using a local attention network and 3D-RoPE. CryoAtom increases the accuracy and completeness of automatically built protein models while reducing the execution time and requirement for map resolution. Tests on large complexes show identification of previously unknown proteins that could not be revealed by other automated tools or predicted. For most tested maps, it constructed models with over 80% completeness, including for protein complexes with nucleic acids, and could also model previously missing residues. Through specific examples, we illustrated CryoAtom's applications in addressing challenging tasks of modern macromolecular modeling. They include successfully detecting novel proteins in multicomponent complexes even without a well-established protein database, capturing subtle conformational changes within large protein assemblies and precisely compartmentalizing the map to isolate nonprotein components. CryoAtom also goes beyond simple structural modeling by segmenting maps with multiple cofactors. Together, these results underscore CryoAtom's ability to tackle a variety of complex challenges in macromolecular modeling in a more time-efficient manner. The versatility, cost efficiency and precision mark CryoAtom as a valuable tool for high-throughput macromolecular modeling and functional analysis of dynamic biological assemblies.

Although CryoAtom leverages deep learning to reduce resolution requirements, modeling at low resolution remains challenging. This stems from the difficulty of identifying protein side chains in low-resolution regions. The challenge arises from the heterogeneous nature of cryo-EM data; stable residues are well-resolved (reflected in high-resolution areas), whereas dynamic regions appear blurred (reflected in low-resolution areas). One potential solution for low-resolution regions is to represent them with an ensemble of conformations rather than a single fixed structure. CryoFold[34] serves as a good example of using molecular dynamics simulations to address this issue. Additionally, CryoAtom can incorporate MDFF[40] during the postprocessing stage, which is a method for flexible fitting of density maps using molecular dynamics. These enhancements will be explored in future work.

Online content

Any methods, additional references, Nature Portfolio 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/s41594-025-01713-3.

References

1. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
2. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
3. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
4. Burley, S. K. et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **51**, D488–D508 (2023).
5. Chua, E. Y. D. et al. Better, faster, cheaper: recent advances in cryo-electron microscopy. *Annu. Rev. Biochem.* **91**, 1–32 (2022).
6. de la Rosa-Trevin, J. M. et al. EMhub: a web platform for data management and on-the-fly processing in scientific facilities. *Acta Crystallogr. D* **80**, 780–790 (2024).
7. Stuart, D. I., Subramaniam, S. & Abrescia, N. G. The democratization of cryo-EM. *Nat. Methods* **13**, 607–608 (2016).
8. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
9. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
10. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).
11. Chen, M., Baldwin, P. R., Ludtke, S. J. & Baker, M. L. De novo modeling in cryo-EM density maps with Pathwalking. *J. Struct. Biol.* **196**, 289–298 (2016).
12. Baker, M. R., Rees, I., Ludtke, S. J., Chiu, W. & Baker, M. L. Constructing and validating initial Cα models from subnanometer resolution density maps with pathwalking. *Structure* **20**, 450–463 (2012).
13. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
14. Lindert, S. et al. EM-fold: de novo folding of α-helical proteins guided by intermediate-resolution electron microscopy density maps. *Structure* **17**, 990–1003 (2009).
15. Wang, R. Y. et al. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **12**, 335–338 (2015).
16. Terashi, G. & Kihara, D. De novo main-chain modeling for EM maps using MAINMAST. *Nat. Commun.* **9**, 1618 (2018).
17. 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).
18. 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).
19. He, J., Lin, P., Chen, J., Cao, H. & Huang, S. Y. Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nat. Commun.* **13**, 4066 (2022).
20. Zhang, X., Zhang, B., Freddolino, P. L. & Zhang, Y. CR-I-TASSER: assemble protein structures from cryo-EM density maps using deep convolutional neural networks. *Nat. Methods* **19**, 195–204 (2022).
21. 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).
22. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
23. Giri, N. & Cheng, J. De novo atomic protein structure modeling for cryoEM density maps using 3D transformer and HMM. *Nat. Commun.* **15**, 5511 (2024).
24. Wang, X., Zhu, H., Terashi, G., Taluja, M. & Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nat. Methods* **21**, 2307–2317 (2024).
25. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
26. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
27. Du, Z. et al. The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* **16**, 5634–5651 (2021).

28. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention 2015* (eds Navab, N. et al.) 234–241 (Springer, 2015).
29. Su, J. et al. RoFormer: enhanced transformer with rotary position embedding. *Neurocomputing* **568**, 127063 (2024).
30. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
31. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
32. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).
33. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
34. Shekhar, M. et al. CryoFold: determining protein structures and data-guided ensembles from cryo-EM density maps. *Matter* **4**, 3195–3216 (2021).
35. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).
36. Feng, Y., Huang, K., Li, X. & Amunts, A. Structure of Photosystem I with SOD reveals coupling between energy conversion and oxidative stress protection. Preprint at *bioRxiv* https://doi.org/10.1101/2025.07.05.663305 (2025).
37. Wú, F. et al. Structure of the II_2–III_2–IV_2 mitochondrial supercomplex from the parasite *Perkinsus marinus*. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.25.595893 (2024).
38. Singh, V. et al. Structural basis of LRPPRC–SLIRP-dependent translation by the mitoribosome. *Nat. Struct. Mol. Biol.* **31**, 1838–1847 (2024).
39. Singh, V. et al. Mitoribosome structure with cofactors and modifications reveals mechanism of ligand binding and interactions with L1 stalk. *Nat. Commun.* **15**, 4272 (2024).
40. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).

Methods

Training and test data

Training set and validation set. To train our method and compare to ModelAngelo[22], we constructed the training dataset with the same settings as ModelAngelo. Specifically, the training set consists of 6,422 EM Data Bank[41] cryo-EM density maps released before 1 April 2022 with resolution better than 4 Å and paired structures in PDB[4]. The full training set was used to train an initial model. The model was used to filter the training dataset on the basis of the precision (>0.5) of the predicted Cα atoms. After filtering, there were 5,731 remaining map–model pairs, of which 10% were randomly selected for validation and 90% for training.

High-resolution test set. The test dataset we used consisted of 177 high-resolution (better than 4 Å) maps from ModelAngelo[22] (released between 1 April 2022 and 9 February 2023).

Low-resolution test set. To test the performance of CryoAtom on low-resolution maps, we collected another set of maps at lower resolution. We first obtained 135 maps with resolutions of 4–7 Å, which were released between 1 January 2024 and 1 June 2024 (about 2 years after the training data). Among these maps, ModelAngelo failed to build atomic structures for 31 maps (the program did not respond for a long time or encountered errors during the process). Therefore, the final low-resolution test dataset consisted of 104 maps with resolutions of 4–7 Å. The number of residues in these maps ranges from 250 to 17,520 (3,461 residues on average).

Nonredundant test set. To investigate whether there was any issue of data leakage, we constructed a nonredundant set. Specifically, if the test density map and the training density map had at least one pair of protein chains with over 40% sequence identity or 0.5 TM-score, we removed it from the test dataset. Additionally, the test dataset was clustered internally on the basis of 40% sequence identity. After applying the above steps, of 281 map–model pairs, 54 map–model pairs were retained.

CryoAtom algorithm

CryoAtom builds the atomic structure from the cryo-EM density map and amino acid sequence information following two major steps (Fig. 1). The first step predicts the Cα atom positions using U-Net[28]. The second step builds all-atom positions using an enhanced transformer network Cryo-Net. They are described in detail below.

Predict the Cα atom positions using U-Net in step 1

We train a 3D convolutional neural network (CNN) to predict the Cα atom positions of all residues in the density map, which are represented by the Cα atoms coordinates. The CNN adopts the U-Net architecture[28]. A detailed architecture can be found in Supplementary Fig. 11. The input to the network is the cropped density map, with each box size being 64 × 64 × 64. The shapes of the network outputs (that is, y^{pred} and $\hat{y}^{\mathrm{pred}}$ in Fig. 1) are the same as the original map but the value in each voxel (that is, each cube 1 × 1 × 1) is the probability of containing a Cα atom (obtained through a sigmoid layer). The difference between y^{pred} and $\hat{y}^{\mathrm{pred}}$ is that their corresponding labels are slightly different; the label for the former is a hard label, while the label for the latter is a soft label. Here, y^{pred} is used as the final output of the U-Net.

The loss function for U-Net[28] consists of the focal loss[42] (FC) and L_2 distance loss (L_2) as follows:

$$L_1 = \mathrm{FC}(y^{\mathrm{pred}}, y^{\mathrm{true}}) + L_2(\hat{y}^{\mathrm{pred}}, \hat{y}^{\mathrm{true}}) \tag{1}$$

where y^{pred} and $\hat{y}^{\mathrm{pred}}$ are both predicted probability density maps for Cα atoms; y^{true} represents the map labeled by the positions of the real Cα atoms, in which a value of 1 or 0 indicates the presence or absence of a Cα atom, whereas $\hat{y}^{\mathrm{true}}$ represents the simulated Cα probability density map, which is calculated from the deposited Cα atom coordinates as follows[43]:

$$\hat{y}^{\mathrm{true}}(x) = \max\left\{ e^{-\frac{\|x-t\|^2}{d_0^2}}, \forall t \in J \right\} \tag{2}$$

where x is a grid point, t represents the coordinates of the Cα atoms, J represents the set of all Cα atoms in the deposited structure and d_0 represents a normalization factor (3 Å here). If the distance from the grid point x to the nearest Cα atom is d_0, the probability value for the grid point will be $1/e$. We combine both losses, aiming for the network to learn rich information from the density map.

The Cα atom coordinates predicted above are refined through the mean-shift algorithm[16,44]. The initial Cα atom coordinates with a probability value above 0.6 are selected first. Then, the coordinates for each of the selected Cα atoms are updated as the weighted linear sum of its neighboring Cα atoms within 1.7 Å (including itself). After this update, the neighboring Cα atoms are removed because the van der Waals radius of a carbon atom is 1.7 Å. Such refinement is necessary because the real coordinates are continuous while the predicted ones from the density maps are discrete. After the refinement, we obtain a set of predicted Cα atoms (the total number of Cα atoms is denoted by r).

Build all-atom positions using Cryo-Net in step 2

Inspired by AF2, we designed Cryo-Net, a transformer-based network that predicts all-atom positions from the input density map, amino acid sequences and the predicted Cα atoms. Cryo-Net consists of two transformer-based modules. The first is called Cryo-Former (Fig. 1, bottom, and Supplementary Fig. 12a) and the second is called Structure Module (Fig. 1, bottom, and Supplementary Fig. 13). Cryo-Net iteratively updates two key representations of the protein structure: node representation and edge representation (corresponding to the single representation and the pair representation in AF2, respectively), which are introduced below.

Node representation. The node representation ($r \times c$, where r is the number of predicted Cα atoms and c is the number of channels) is constructed from the predicted Cα atoms and the density map. A cubic box with size 17 × 17 × 17 Å^3 centered at each of the r predicted Cα atoms is cropped from the density map first. Then, to extract information from the density map and reduce the dimensionality, a CNN is used to convert the density map in each cubic box into c channels. The architecture of this CNN is the same as the encoder module in U-Net[28] (that is, Supplementary Fig. 11a).

Edge representation. The edge representation ($k \times r \times c$) is constructed similarly, measuring the relationship between neighboring nodes (k is the number of nearest neighbors considered). For each edge, a rectangular box with size 3 × 3 × 12 Å^3 is cropped from the density map. The Cα atom is located at the center of the bottom box (that is, the 3 × 3 square) and the direction of the box is the same as the edge vector. Lastly, another CNN (with kernel size 3 × 3 × 1) is used to convert the density map in each rectangular box into c channels.

Note that the above representations are different to those used in AF2. The representations here are constructed from the density map guided by the predicted Cα atoms (inspired by ModelAngelo), rather than from multiple-sequence alignments and homologous templates in AF2. In addition, the edge representation (called pair representation in AF2) is a square matrix ($r \times r$) in AF2, whereas a nonsquare matrix ($k \times r$) is used here. This is because the density map already provides rich structure information and we only need to consider the k-nearest neighbors rather than all nodes.

The input and output of Cryo-Net in the nth iteration are defined by the following equation:

$$g\left(T^{(n)}, V, S\right) = \left(T^{(n+1)}, \alpha^{(n+1)}, A^{(n+1)}, P^{(n+1)}, E^{(n+1)}, M^{(n+1)}\right) \tag{3}$$

where $T^{(n)} \in \mathbb{R}^{r\times3\times4}$ is the backbone frame (N, Cα, C) at the nth step, V is the density map, $S \in \mathbb{R}^{r\times1280}$ is the sequence embedding from the protein language model ESM-2 (ref. 45), $\alpha \in \mathbb{R}^{r\times7\times2}$ is the sine and cosine encoding of the three backbone and up to four side-chain torsion angles, $A \in \mathbb{R}^{r\times20}$ is the probability vector for the 20 types of amino acids, $P \in \mathbb{R}^{r}$ is the confidence score of each node, $E \in \mathbb{R}^{r\times k}$ is a binary vector indicating whether each node is sequentially adjacent to its k-nearest neighbors and $M \in \mathbb{R}^{r}$ is the probability of a node being a true residue.

Cryo-Former. The Cryo-Former module consists of a total of 18 blocks with nonshared weights, as shown in Supplementary Fig. 12a. The key difference between Cryo-Former and the Evoformer in AF2 is the concentration on local attention rather than global attention. This is reflected by the number of neighbors considered (that is, the parameter k). Evoformer is equivalent to Cryo-Former when k is equal to the total number of nodes (that is, $k = r$). The key components of Cryo-Former are described below.

Sequence attention. The use of pretrained protein language models is inspired by ModelAngelo (Box 1). This layer incorporates the protein sequence information into the node features with a gated cross-attention mechanism (Supplementary Text 2 and Supplementary Algorithm 1). The node representations are queries and the sequence embeddings from a pretrained protein language model (that is, ESM-2)[45] are keys and values.

Node attention. As shown in Supplementary Fig. 12b, the node attention updates the representation of a given node on the basis of self-attention on the node and its k-nearest neighbors, with the information of edge representations serving as bias. As the depth of this network increases, it can collect information from more nodes, resulting in a larger receptive field. It is worth noting that this node attention is equivalent to the row-wise gated self-attention in AF2 when $k = r$.

Transition and outer product. Inspired by AF3 (ref. 33), the transition layer uses the SwiGLU[46] instead of ReLU[47] to update the node and edge features. It uses the nonlinear activation function to enrich the representation of the two types of features. The outer product layer updates the edge features on the basis of the node features. It incorporates the node features of two adjacent nodes into the edge features.

Edge attention. As shown in Supplementary Fig. 12c, the edge attention updates the edge representation through a self-attention mechanism. The information of edge ij is aggregated from the information of all edges connected to node i and all edges connected to node j. The edge attention can be seen as a special type of the node attention, according to the Whitney graph isomorphism theorem (Supplementary Fig. 14). In addition, it is worth mentioning that, when $k = r$, the edge attention is equivalent to applying AF2's row and column attentions simultaneously.

3D-RoPE. 3D-RoPE is not an independent component. Originally, RoPE[29] was used to process 1D sequences. We extend it to a node position encoding in 3D space. It is inserted into the node attention as a multiplicative positional encoding. It elegantly encodes the position information of each node into the attention calculations. The attention score decays when the distance between two nodes increases. More details can be found in Supplementary Text 2.

Amino acid probability for each node. To generate all-atom positions, we try to assign each node to one of the 20 possible amino acids. This is achieved by using the information from the density map of each node (that is, the original node representation), the information from neighboring nodes (that is, the updated node representation) and the sequence information (obtained from the sequence attention in Cryo-Former). A separate MLP is trained to generate the amino acid probability vectors for all nodes (that is, A in equation (3); Supplementary Fig. 15). Ablation experiments indicated that removing the information of neighboring nodes (that is, no node attention and edge attention) or removing the sequence information would affect the subsequent determination of amino acid identity.

Structure Module. The Structure Module here (Supplementary Fig. 13) is largely similar to that in AF2 but with a few key differences (detailed below). It predicts the backbone frame (that is, T in equation (3)) and the backbone and side-chain torsion angles (that is, α in equation (3)), using the updated node and edge representations from Cryo-Former. The Structure Module consists of seven blocks with shared weights. Two types of attention mechanisms are used in this module (Supplementary Text 2 and Supplementary Algorithm 2).

The first one is a self-attention mechanism applied to the node features (similar to the node attention in Cryo-Former). Given the constraints from the density map, the attention is restricted to each node and its k-nearest neighbors (rather than all other nodes in AF2). In addition, the position information between nodes is introduced into the attention mechanism using 3D-RoPE, ensuring that the attention score incorporates position information.

The second type of attention is the invariant point attention (IPA) from AF2. It is conducted on the backbone frames and is invariant under global Euclidean transformations. Like the first type of attention, the attention is restricted to a node and its k neighboring nodes (rather than all others in AF2).

Cryo-Net will be run iteratively to improve its output. During training, 0–2 rounds of iterations are performed. Only the backbone frame is iterated. During inference, two rounds of iterations are performed. The reason for choosing $n = 2$ here is to keep a balance between computational cost and performance. No notable improvement in the network output was observed when more rounds of iterations are performed.

Lastly, all-atom positions for each node are then generated using the predicted backbone frame, the backbone and side-chain torsion angles and the predicted amino acid type (the amino acid type with the maximum probability). This calculation is performed using the function frames_and_literature_positions_to_atom14_pos in the AF2 script all_atom.py.

The predicted all-atom positions are postprocessed to connect the unordered nodes into chains of amino acid sequences. This is achieved using the heuristic algorithm in ModelAngelo[22]. Intuitively, it connects nodes by minimizing the total length of peptide bond on the basis of the fact that the C–N distance in a peptide bond is less than 1.4 Å. Profile-sequence-based alignments between the predicted sequence profiles (that is, A in equation (3)) and the input sequences are further used to correct the sequence assignment. Inspired by AF2 and ModelAngelo, recycling is also used here by replacing the predicted Cα atoms in the first step with those from the postprocessed structure. Three rounds of recycling are used in this work.

The loss function for Cryo-Net. The loss function here can be written as follows:

$$L_2 = L_{\text{multi}} + L_{\text{single}} \tag{4}$$

$$L_{\text{multi}} = L_{C\alpha} + 1.3L_{\text{FAPE}} + 0.025L_{\text{pFAPE}} + L_{\text{existence}} \tag{5}$$

$$L_{\text{single}} = L_{\text{atom}} + L_{\text{torsion}} + L_{\text{attention}} + 2L_{\text{violation}} + 0.3L_{\text{aa}} + L_{\text{edge}} \tag{6}$$

where the overall L_2 can be divided into L_{single} (single-step loss) and L_{multi} (multistep loss). The single-step loss is the final loss output after passing through seven structure modules, while the multistep loss is calculated seven times within the seven structure modules and averaged. They are described in detail below.

L_{atom} is the all-atom r.m.s.d. It calculates the r.m.s.d. loss of all atoms on the main chain and side chain, which can be written as follows:

$$L_{\text{atom}} = \frac{1}{r}\sum_{i=1}^{r}\sqrt{\frac{1}{3n_i}\sum_{j=1}^{n_i}\sum_{d=1}^{3}\left(\left[x_{ij}^{\text{pred}}\right]_d - \left[x_{ij}^{\text{true}}\right]_d\right)^2} \tag{7}$$

where r represents the total number of residues, n_i represents the number of atoms in the ith residue, x_{ij} represents the atomic coordinates of the jth atom in the ith residue and d represents the dth dimension of the 3D coordinates.

L_{torsion} is the L_2 distance between the torsion angles of the network output (that is, α in equation (3)) and the torsion angles of the deposited model. It is the same as the torsion angle loss defined in AF2.

$L_{\text{attention}}$ is the classification loss when the nodes attend to the amino acids in the sequence. The calculation of this loss is the focal loss[42]. The purpose of this loss is to ensure that each node can correctly match each amino acid in the sequence.

$L_{\text{violation}}$ penalizes unreasonable bond angles and bond lengths in the model. It imposes constraints to ensure that the output structure maintains ideal bond angles and bond lengths. The definition of this loss is the same as AF2.

L_{aa} represents the amino acid classification loss. This loss calculates the focal loss by comparing the 20-dimensional amino acid probability vector (that is, A in equation 3)) and the one-hot vector corresponding to the amino acid identity in the deposited structure. The main purpose of this loss is to assign an amino acid identity to each node.

L_{edge} represents the edge classification loss. The edge classification is based on whether there is a connection between two nodes in the sequence. If there is a connection, we want them to be classified as neighbors. Specifically, this loss calculates the focal loss between E in equation (3) and the classification corresponding to the edge.

$L_{\text{C}\alpha}$ is only the r.m.s.d. loss of Cα atoms. Its definition is similar to the all-atom r.m.s.d. (that is, $n_i = 1$ in equation (7)).

L_{FAPE} (frame-aligned point error) comes from AF2, which has a role in updating the backbone frame. It can be written as follows:

$$d_{ij} = \sqrt{\left\|T_i^{-1} \circ \vec{t}_j - \left(T_i^{\text{true}}\right)^{-1} \circ \vec{t}_j^{\,\text{true}}\right\|^2 + \varepsilon} \tag{8}$$

$$L_{\text{FAPE}} = \frac{1}{Z} \underset{\substack{i \in \{\text{frames}\} \\ j \in \{\text{atoms}\}}}{\text{mean}} \left(\min(d_{\text{clamp}}, d_{ij})\right) \tag{9}$$

where T_i represents the ith backbone frame (that is, T in equation (3)). t_j represents the coordinates of the jth Cα atom. Both d_{clamp} and Z are taken as 10 Å here.

L_{pFAPE} is a simple regression loss. It calculates the mean squared error loss between the predicted FAPE loss for each residue (that is, P in equation (3)) and the actual FAPE loss (that is, equation (9)). It is worth noting that, in the postprocessing, P will be normalized to a confidence score between 0 and 100 as the final confidence score.

$L_{\text{existence}}$ represents the residue existence loss. Each residue can be classified into two classes: artificially added (Supplementary Text 2) and naturally occurring. This loss calculates the focal loss between M in equation (3) and the corresponding one-hot vector of the residue class.

Reporting summary

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

Data availability

All models constructed by CryoAtom in the paper can be downloaded from figshare (https://doi.org/10.6084/m9.figshare.28919303.v3)[48]. The training and test datasets can be obtained online (http://yanglab.qd.sdu.edu.cn/CryoAtom/benchmark). The cryo-EM map of photosynthetic complex was deposited in Electron Microscopy Data Bank (EMD-66850).

Code availability

The source code and model parameters are available online (https://github.com/YangLab-SDU/CryoAtom and https://yanglab.qd.sdu.edu.cn/CryoAtom/).

References

41. Turner, J. et al. EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Res.* **52**, D456–D465 (2024).
42. Lin, T. Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 318–327 (2020).
43. Li, T. et al. All-atom RNA structure determination from cryo-EM maps. *Nat. Biotechnol.* **43**, 97–105 (2025).
44. Wang, X., Terashi, G. & Kihara, D. CryoREAD: de novo structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nat. Methods* **20**, 1739–1747 (2023).
45. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
46. Shazeer, N. GLU variants improve Transformer. Preprint at https://arxiv.org/abs/2002.05202 (2020).
47. Glorot, X., Bordes, A. & Bengio, Y. Deep sparse rectifier neural networks. In *Proc.14th International Conference on Artificial Intelligence and Statistics* (eds Geoffrey, G. et al.) 315–323 (PMLR, 2011).
48. Su, B. et al. CryoAtom improves model building for cryo-EM. *figshare* https://doi.org/10.6084/m9.figshare.28919303.v3 (2025).

Acknowledgements

We would like to thank K. Jamali and S. H. W. Scheres for their kind help in the benchmarking of ModelAngelo and S. Caffarri, A. Naschberger, R. Sobotka and F. Wu for help with the structural studies. This work is supported by the National Key Research and Development Program of China (2024YFA0916901 to J.Y.), National Natural Science Foundation of China (T2225007 and 32430063 to J.Y.; T2222012 to Z.P.), the Fundamental Research Funds for the Central Universities, European Research Council (ERC-805230) and Westlake Visiting Professor Fellowship to A.A.

Author contributions

J.Y. conceptualized and administered the study. B.S. designed and implemented the network. A.A. and Z.P. cosupervised the study. K.H. contributed to the application of CryoAtom on new maps. All authors revised and approved the final draft of 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/s41594-025-01713-3.

Correspondence and requests for materials should be addressed to Zhenling Peng, Alexey Amunts or Jianyi Yang.

Peer review information *Nature Structural & Molecular Biology* thanks Sheng-You Huang, Dari Kimanius and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Sara Osman, in collaboration with the *Nature Structural & Molecular Biology* team.

nature portfolio

Corresponding author(s): Jianyi Yang, Alexey Amunts, Zhenling Peng

Last updated by author(s): Aug 7, 2025

Reporting Summary

Nature Portfolio 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 Portfolio 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	Only publicly available data from the EMDB and PDB were used in this study.
Data analysis	CryoAtom is distributed freely under an MIT licence. We used version 1.0 for protein modeling on the benchmark. Version 1.3.2 was used for analyzing new proteins.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All models constructed by CryoAtom in the paper can be downloaded from figshare https://figshare.com/s/746b50330fa837a47477. The training and test datasets can be downloaded from http://yanglab.qd.sdu.edu.cn/CryoAtom/benchmark/train_datasets.zip, http://yanglab.qd.sdu.edu.cn/CryoAtom/benchmark/test_datasets.zip, respectively.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Life sciences study design

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

Sample size	We used all available structures from EMDB/PDB and collected a sufficient number of samples for a fair benchmark.
Data exclusions	In the non-redundant test set, we removed data with over 40% sequence identity and a TM-score of 0.5 to the training set.
Replication	N/A
Randomization	N/A
Blinding	N/A

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

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern
☒	☐	Plants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*