Structural biology technologies

https://doi.org/10.1038/s41594-026-01744-4

When cryo-EM modeling meets structure prediction

Xiaogen Zhou, Xiangyu Xu & Guijun Zhang

Accurately interpreting density maps into atomic models is a central yet challenging goal of cryo-EM. Two studies now reveal distinct ways in which protein structure prediction can be incorporated into cryo-EM model building to enable more accurate and robust automated construction of protein atomic models from density maps.

Determining protein structures is fundamental to molecular biology. In recent years, single-particle cryo-electron microscopy (cryo-EM) has become a leading experimental approach for macromolecular protein structure determination[1,2], especially for large and complex assemblies, an advance recognized by the 2017 Nobel Prize in Chemistry. In parallel, advances in artificial intelligence (AI) have revolutionized protein structure prediction[3,4], achieving unprecedented accuracy directly from amino acid sequences, and earning the 2024 Nobel Prize in Chemistry. How these two Nobel Prize-level technologies can be effectively combined for accurate protein model building has been increasingly explored[5–7]. Two recent approaches published in *Nature Structural & Molecular Biology*, CryoAtom by Su et al.[8] and EMProt by Li et al.[9], provide new perspectives on this question by demonstrating how experimental cryo-EM density maps can be tightly integrated with computational structure prediction techniques to build more complete and accurate atomic protein models.

Interpreting cryo-EM maps into accurate atomic models has long been a challenging and labor-intensive task that often requires substantial expert intuition. Current modeling methods can be broadly divided into two main categories: de novo modeling and structure-based fitting approaches. De novo methods typically aim to trace protein backbones directly from density maps, whereas structure-based fitting approaches fit predicted structures or homologous templates into density maps followed by refinement. However, de novo modeling can break down in regions of weak or fragmented density, and fitting-based approaches may fail when predicted models deviate substantially from the true structure or when suitable templates are unavailable. As a result, neither paradigm alone provides a general solution for automated cryo-EM model building across diverse map qualities and structural complexities.

Recent advances in AI-based structure prediction have begun to blur the boundary between de novo modeling and structure-based fitting, offering new ways to address these limitations[10–13]. Representative efforts in this direction, such as ModelAngelo[12], demonstrated that de novo model building can be substantially improved by embedding sequence and structural information directly into density interpretation within a unified learning framework. In parallel, hybrid pipelines exemplified by DeepMainmast[13] showed that incorporating predicted structures with de novo modeling can help to alleviate ambiguities in density interpretation, particularly for large or multi-chain assemblies. Together, these approaches established key building blocks for integrating density maps with structural knowledge, while leaving open how this integration should be formulated as a general and systematic strategy. It is in this context that CryoAtom and EMProt provide a clearer answer, by introducing two distinct yet complementary methods of combining prediction with cryo-EM data: one internalizing structure prediction principles within a de novo framework, and the other explicitly integrating predicted models into density-guided assembly.

CryoAtom[8] represents a natural extension of de novo cryo-EM model building in which principles of protein structure prediction are internalized into the interpretation of density maps. Rather than explicitly incorporating externally predicted models, CryoAtom infers atomic structures directly from density maps by leveraging recent advances in AlphaFold2[3] to extend the de novo framework established by ModelAngelo. In the CryoAtom pipeline (Fig. 1), predicted Cα positions provide a geometric scaffold, upon which an AlphaFold2-inspired structure-aware network decodes density- and sequence-informed representations into all-atom coordinates by predicting backbone frames and side-chain torsion angles. All-atom models are then generated through post-processing based on the predicted geometric information. This internalization strengthens de novo modeling in regions of weak or fragmented density, improving model continuity and completeness and extending automated model building toward intermediate-resolution maps. Conceptually, CryoAtom builds on state-of-the-art density-to-structure de novo approaches by advancing them into the prediction era through embedded structural reasoning.

EMProt[9] adopts a complementary strategy that explicitly integrates structure prediction into de novo model building, while treating predicted structures as flexible hypotheses rather than rigid templates. Following initial prediction of Cα positions and amino acid types from cryo-EM maps (Fig. 1), EMProt constructs an initial de novo model using a three-track attention network that couples density-derived features with three-dimensional backbone representations to infer backbone geometry. When the sequence input is available, predicted structures are decomposed into fragments that are independently aligned to the de novo model and fitted into the density map. Importantly, this fragmentation, together with density-guided selection of non-clashing components, makes the approach tolerant to inaccuracies in global structure prediction by allowing local agreement with the density map to drive model assembly. As a result, EMProt emphasizes completeness through prediction-guided assembly from high-quality maps, while treating predicted models as flexible structural hypotheses rather than rigid templates to be trusted wholesale.

CryoAtom and EMProt use different modeling strategies that make them well suited to different cryo-EM modeling scenarios. As schematically illustrated in Fig. 1, CryoAtom and EMProt share a

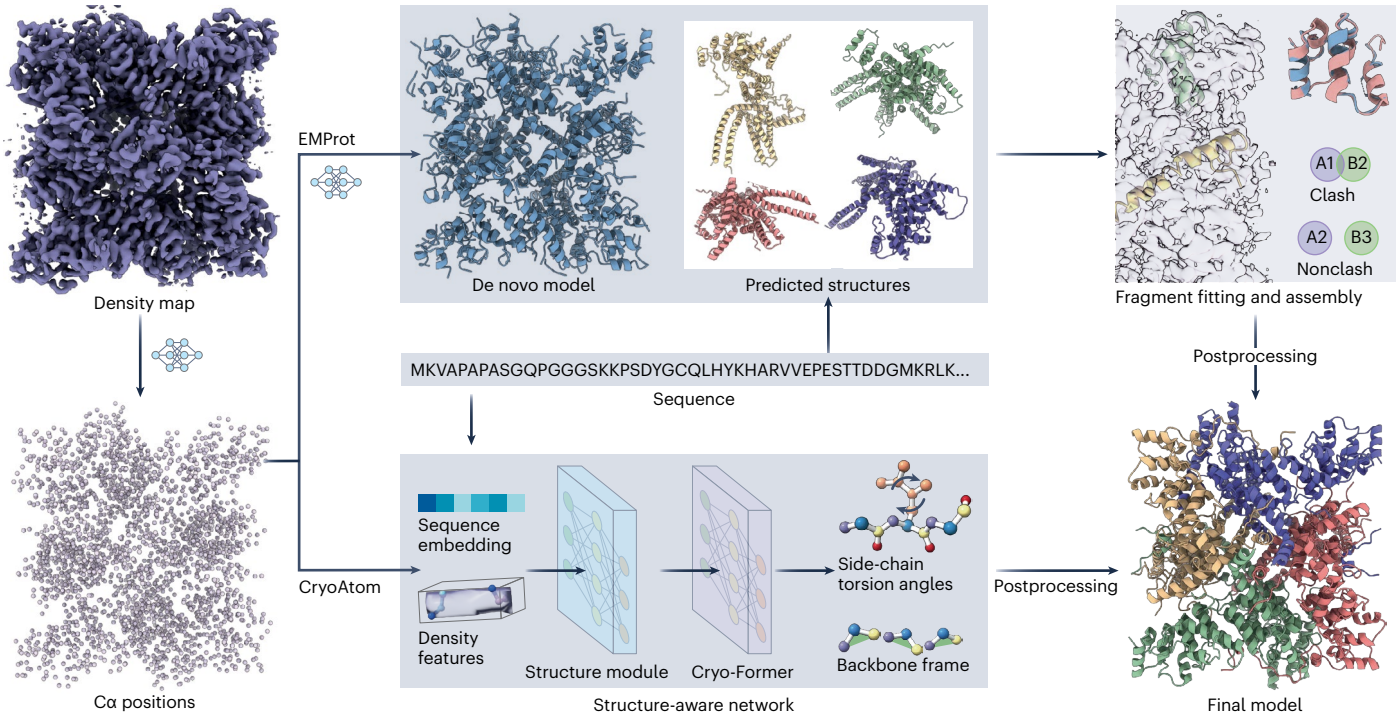

Fig. 1 | Conceptual comparison of CryoAtom and EMProt. Both CryoAtom and EMProt start from experimental cryo-EM density maps and protein sequences and share a common density-driven Cα localization step. CryoAtom internalizes structure-prediction principles within a single, map-driven structure inference pipeline, directly embedding predictive reasoning into density interpretation (bottom). By contrast, EMProt explicitly integrates structure prediction into model building by assembling density-consistent fragments derived from predicted structures, which are treated as flexible hypotheses rather than rigid templates (top). Molecular graphics and analyses performed with UCSF ChimeraX[14], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

common density-driven backbone for model construction but differ in how and where structure prediction approaches are incorporated into the modeling pipeline. CryoAtom is best viewed as an advanced de novo model builder and is particularly attractive when reliable prior structural information is limited or when minimizing dependence on predicted models is desirable. By internalizing structure-prediction principles, CryoAtom can extract more coherent structural information from intermediate-resolution density maps than earlier de novo approaches, while remaining constrained by the intrinsic ambiguity of low-resolution data. EMProt, by contrast, is most effective in scenarios in which model completeness is a primary goal and where sequence information and predicted structures can be leveraged. For large, multi-chain assemblies or maps with clear global features but uneven local detail, EMProt's hybrid strategy can often recover substantially more of the underlying atomic structure. Seen in this light, CryoAtom and EMProt should not be regarded as competing solutions, but as complementary tools that together define a practical spectrum for automated cryo-EM model building, spanning internalized and explicitly integrated uses of structure prediction.

Despite these advances, a closely related challenge concerns the dependence of current approaches on map resolution. Both CryoAtom and EMProt are primarily developed for high-resolution density maps, typically better than 4 Å, where the map provides sufficient local structural constraints for atomic model construction. Although CryoAtom extends modeling into lower-resolution maps, reaching approximately 7 Å, model accuracy inevitably degrades as the density map becomes less informative. This limitation reflects a broader, unresolved issue in the field: how to construct reliable atomic models when density maps alone no longer support unambiguous backbone prediction. In this context, tighter coupling with structure prediction and potentially reducing reliance on explicit density-derived backbone information, may offer a promising route toward extending automated modeling to intermediate- and lower-resolution cryo-EM maps.

Beyond resolution, CryoAtom and EMProt remain primarily focused on protein modeling, and systematic benchmarking of these methods on nucleic acid structures is still lacking. Although ModelAngelo has demonstrated the ability to trace nucleotide backbones, accurate base assignment remains challenging across current approaches. Given the growing importance of ribonucleoprotein complexes, viral genomes and other mixed macromolecular assemblies, extending prediction-integrated cryo-EM modeling frameworks beyond proteins represents a crucial and timely direction for future development.

In summary, the emergence of CryoAtom and EMProt underscores a broader shift in cryo-EM atomic structure modeling: the integration of density-driven modeling with structure prediction approaches. Rather than asking whether density-driven modeling or prediction-driven modeling should dominate model building, these methods illustrate how different balances between the two can be tailored to yield more reliable atomic models. As cryo-EM continues to push toward larger, more heterogeneous and more dynamic molecular systems, the field

is likely to benefit from diverse techniques rather than a single universal solution. In this sense, CryoAtom and EMProt together mark an important step toward more flexible, robust and data-regime-aware automated modeling, helping experimentalists extract deeper structural insight from increasingly complex cryo-EM data.

Xiaogen Zhou [1,2] ✉, **Xiangyu Xu** [1] **& Guijun Zhang** [1,2] ✉
[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. [2]Zhejiang Key Laboratory of Intelligent Perception and Control for Complex Systems, Hangzhou, China.
✉e-mail: zxg@zjut.edu.cn; zgj@zjut.edu.cn

Published online: 23 January 2026

References

1. Yip, K. M. et al. *Nature* **587**, 157–161 (2020).
2. Strack, R. *Nat. Methods* **17**, 1175–1175 (2020).
3. Jumper, J. et al. *Nature* **596**, 583–589 (2021).
4. Abramson, J. et al. *Nature* **630**, 493–500 (2024).
5. He, J. et al. *Nat. Commun.* **13**, 4066 (2022).
6. Zhang, Z. et al. *Nucleic Acids Res.* **53**, W228–W237 (2025).
7. Chen, S. et al. *Nat. Commun.* **15**, 8808 (2024).
8. Su, B. et al. *Nat. Struct. Mol. Biol.* https://doi.org/10.1038/s41594-025-01713-3 (2025).
9. Li, T. et al. *Nat. Struct. Mol. Biol.* https://doi.org/10.1038/s41594-025-01723-1 (2025).
10. Zhang, X. et al. *Nat. Methods* **19**, 195–204 (2022).
11. Zhou, X. et al. *Nat. Computat. Sci.* **2**, 265–275 (2022).
12. Jamali, K. et al. *Nature* **628**, 450–457 (2024).
13. Terashi, G. et al. *Nat. Methods* **21**, 122–131 (2024).
14. Pettersen, E. F. et al. *Protein Sci.* **30**, 70–82 (2021).

Acknowledgements

The authors acknowledge support from the National Key R&D Program of China (2022ZD0115103), the National Nature Science Foundation of China (62203389, 62173304), Fundamental Research Funds for the Provincial Universities of Zhejiang (RF-C2024006), Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2023R01006), Zhejiang Provincial Special Support Program for High-Level Talents (2023R5248), and the 'Pioneer' and 'Leading Goose' R&D Program of Zhejiang (2025C01121, 2025C01190).

Competing interests

The authors declare no competing interests.