CryoAtom2: inverse folding-inspired deep learning enables accurate model building of protein-nucleic acid complexes in cryo-EM

Baoquan Su[1], Lingfan Zhu[2], Wenfei Li[2], Wenkai Wang[1], Jiahui Liang[1], Kailong Zhao[1], Zhenling Peng[1,*], Jianyi Yang[1,*]

[1]MOE Frontiers Science Center for Nonlinear Expectations, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China. [2]School of Basic Medical Sciences, Cheeloo College of Medicine, Shandong University, Jinan, China.

*Corresponding authors: zhenling@email.sdu.edu.cn; yangjy@sdu.edu.cn

Abstract

Automated model building of protein–nucleic acid complexes in cryo-electron microscopy (cryo-EM) maps is hindered by the highly similar density signatures of chemically related bases at moderate resolutions. Here, we present CryoAtom2, an inverse folding-inspired multimodal deep learning framework for *de novo* model building of proteins, nucleic acids, and their complexes within a single pipeline. CryoAtom2 jointly integrates three complementary modalities: density, sequence, and—crucially—a diverse set of plausible 3D structural conformations. For each residue, rather than generating a single coordinate set, it produces 28 candidate coordinate sets (20 amino acids, 4 RNA nucleotides, and 4 DNA nucleotides) and employs a dedicated Structure Encoder to capture the rich structural context necessary for accurate sequence inference, following ProteinMPNN. By reasoning over this expanded structural space, the model effectively discriminates between similar bases, yielding *globally optimal sequences that best fit the density map*. This approach is analogous to the principle of inverse folding, where structural context guides accurate sequence assignment. Benchmarking on 175 non-redundant cryo-EM maps demonstrates that CryoAtom2 substantially outperforms existing state-of-the-art tools. It delivers more complete models with a superior fit to the density and improved geometric quality across proteins, nucleic acids, and their complexes. For example, under the most stringent criterion—*completeness*—which considers both atomic position and sequence identity—the CryoAtom2 model achieves an average of 60.0% for nucleic acids, significantly outperforming other methods: EM2NA (40.6%), ModelAngelo (32.6%), CryoREAD (32.4%), and Phenix (11.7%). We successfully applied CryoAtom2 to three challenging, unseen experimental structures: an *in situ* human 80S ribosome, an archaeal 70S ribosome, and a large RNA-only nanocage. In

each case, CryoAtom2 achieved near-expert levels of model completeness, exceptional sequence accuracy, and stereochemically sound geometry. Together, these results position CryoAtom2 as a transformative tool for interpreting complex biological machinery in cryo-EM.

Introduction

Determining the structures of nucleic acids (NAs) and their interactions with proteins is inherently more challenging than for proteins alone, due to greater flexibility, conformational heterogeneity, and difficulties in crystallization or sample preparation. As of early 2026, the Protein Data Bank [1] (PDB) contains approximately 248,000 released structures, of which more than 210,000 are protein-only entries. In contrast, structures involving NAs are far scarcer: only about 4,900 are NA-only, and roughly 15,700 are protein–NA complexes. This severe data imbalance largely explains why the structures of NAs and protein–NA complexes remain difficult to predict accurately, with current methods showing reduced accuracy or failing on novel systems [2, 3]. Consequently, a major leap forward in biomolecular structure prediction will likely require a substantial expansion of high-quality experimental structures of NAs and protein–NA complexes to empower robust machine-learning approaches [4, 5].

Cryo-electron microscopy (cryo-EM) has emerged as the leading technique for determining structures of large, dynamic macromolecular assemblies, including many protein–NA complexes that resist crystallization. Over 31,000 cryo-EM structures are now deposited in the PDB, accounting for a rapidly growing fraction of annual entries. However, the critical final step—atomic model building—remains a major bottleneck, historically reliant on manual, expert-driven interpretation.

For proteins, this challenge has been substantially alleviated by the advent of AI-based model-building tools. Methods such as ModelAngelo [6], EmodelX [7], Cryo2Struct [8], and our previous work, CryoAtom [9], use deep learning to identify amino acids and build atomic models *de novo* from cryo-EM maps. For lower-resolution maps that are challenging for *de novo* approaches, methods such as EMProt [10] and DiffModeler [11] build models by fitting predicted structures into density maps. Such methods, however, rely on highly accurate structure predictions as inputs, which are not always available.

In contrast, specialized AI tools for NA model building from cryo-EM maps have only recently emerged and remain considerably less mature. Emerging methods, including

CryoREAD [12], EM2NA [13], and EMRNA [14], signal promising progress. Typically, they employ convolutional neural networks to detect key nucleotide atoms (phosphates, sugars, and/or bases), trace backbones via graph-based or clustering approaches, and assign bases. Despite their utility, these tools face persistent limitations rooted in the intrinsic properties of NAs. First, the limited chemical diversity among the 4 canonical bases (versus 20 diverse amino acids) creates severe ambiguity in base identification from density. Second, most methods require relatively high-resolution maps to achieve reliability, struggling with the noise and conformational variability prevalent at lower resolutions. Furthermore, apart from ModelAngelo, no other AI-based method currently supports automated atomic model building for protein–NA complexes. However, ModelAngelo's performance remains limited for NAs.

To address these challenges, we introduce CryoAtom2, a significantly enhanced and extended successor of our prior CryoAtom. CryoAtom2 is a unified deep-learning-based system for automated, *de novo* atomic model building that handles pure proteins, pure NAs, and protein–NA complexes within a single pipeline. To address the notorious challenge in NA modeling, rather than merely fitting a presumed sequence, CryoAtom2 adopts an inverse folding-inspired strategy: building 28 (one for each residue type) candidate sets of atomic structures for each position and then inferring the most probable sequences that are compatible with the density map, analogous to the principle of inverse folding. This innovation is coupled with a Bayesian framework for backbone tracking and a refined scoring system for sequence assignment.

These innovations enable CryoAtom2 to achieve a breakthrough in the automated interpretation of cryo-EM maps. Benchmarks and applications demonstrate that CryoAtom2 significantly outperforms existing methods for pure proteins, pure NAs, and protein–NA complexes. For the first time, it generates near-complete and highly accurate protein–NA models for high-resolution maps, approaching the quality of human experts. Thus, CryoAtom2 represents a powerful advance for streamlining cryo-EM map interpretation, accelerating comprehensive elucidation of the integrated protein–NA machinery fundamental to life.

Overview of the CryoAtom2 approach

CryoAtom2 extends the capabilities of CryoAtom to enable *de novo* model building for proteins, NAs, and their complexes. The central conceptual advance is the introduction of an inverse folding-inspired Structure Encoder for solving the ambiguity problem in density-based sequence assignment. We term this an "inverse folding-inspired"

framework not because CryoAtom2 performs inverse folding per se, but because it adopts the core principle of inverse folding—reasoning from structure to sequence—as an auxiliary learning signal that guides sequence assignment. This trimodal integration—density, sequence, and structure—is distinct from prior approaches that derive sequence assignments solely from density maps: by explicitly re-encoding the evolving structures as an independent input modality, CryoAtom2 evaluates sequence plausibility within structural space rather than inferring it from density alone. In the following, we provide a brief overview of the CryoAtom2 methodology; full details are given in the Methods.

Identify key atoms with RU-Net. As outlined in Fig. 1a, the pipeline begins by processing the input cryo-EM density map using a recurrent U-Net [15] (RU-Net; Supplementary Fig. 1). By integrating convolutional and attention layers, the RU-Net captures both local structural details and global context, allowing it to identify candidate locations for protein Cα atoms and NA phosphorus (P) atoms.

Generate node and edge representations with Cryo-EM Encoder. From the input density map and the key atoms predicted by RU-Net, the Cryo-EM Encoder, a convolutional network adapted from CryoAtom, is employed to produce node representations (1D features) and edge representations (2D features). Note that each node is connected to its k nearest spatial neighbors ($k = 30$) rather than all other nodes. This locally restricted connectivity ensures computational efficiency, enabling effective local attention mechanisms in subsequent stages of the pipeline.

Update node and edge representations with Cryo-Former. Similar to CryoAtom, a 16-layer Cryo-Former module—analogous to AlphaFold2's Evoformer [16] but with untied weights—is used to iteratively refine the node and edge representations. Input sequences are first encoded into embeddings using dedicated language models (ESM-2 [17] for proteins and RNA-FM [18] for NAs), and these embeddings are integrated via cross-attention within the Cryo-Former to enhance sequence assignment. Beyond updating the node and edge representations, this module also performs two critical prediction tasks: classifying each node into one of 28 possible residue types (20 amino acids, 4 RNA nucleotides, and 4 DNA nucleotides; Supplementary Fig. 2a) and predicting the sequence adjacency for each edge.

Decode structure and sequence using inverse folding–inspired deep learning. The key conceptual innovation of CryoAtom2 is the inverse folding–inspired strategy, which

fundamentally redefines how sequences are inferred from cryo-EM density. As shown in Fig. 1c, CryoAtom and other methods such as ModelAngelo [6] and EM2NA [13] directly predict the sequence from the density map (dashed arrow). In contrast, CryoAtom2 first builds all possible atomic structures for each node (identity unknown) and then infers the globally optimal sequences that are compatible with the density map (solid arrows), analogous to the principle of inverse folding. This strategy filters out sequences that locally fit the density but are structurally inconsistent (Fig. 1d).

Specifically, the updated representations from Cryo-Former are decoded by a 7-layer weight-tied Structure Module to predict 28 candidate sets of atomic coordinates per node—one for each possible residue type. This results in 28^r candidate 3D structures (r is the total number of residues predicted in the density map), which are then transformed into 1D and 2D features using a Structure Encoder (Fig. 1b; Supplementary Fig. 3). Following the representational strategy of ProteinMPNN [19], the encoder calculates 15 $\times$15 atomic distances for each pair of nodes and projects the distances into a 2D edge representation via a linear layer. This 2D representation—along with the 1D node representation from the Structure Module—is then fused back into the node and edge embeddings from the previous iteration cycle. After three cycles, CryoAtom2 assigns a definitive sequence identity to each node by integrating the predicted residue type probabilities with a sequence-alignment algorithm. The final all-atom model is then constructed by selecting the atomic coordinate set that matches the assigned identity for every node.

We acknowledge that CryoAtom2 is inspired by ModelAngelo (Box 1), which introduced a graph neural network capable of processing multimodal inputs—density maps and sequence information—together with a structure module inspired by AlphaFold2, enabling de novo modeling of protein–NA complexes. Building on these ideas, CryoAtom reformulated the ModelAngelo pipeline within the AlphaFold2 network framework, demonstrating that structure prediction architectures can effectively reduce the dependence on map resolution while improving overall model completeness. CryoAtom2 takes a step further by incorporating an inverse folding objective (Fig. 1c) through the Structure Encoder (Fig. 1b). Although ModelAngelo and CryoAtom also employ structure recycling, they primarily treat backbone atomic coordinates as positional encodings for graph nodes. By contrast, CryoAtom2 explicitly decomposes 3D structural information into 1D and 2D features following ProteinMPNN, re-encoding the evolving structure as an independent modality. To our knowledge, CryoAtom2 is the first automated model-building approach to integrate

trimodal encoding—density, sequence, and structure—and ablation studies presented below confirm that each modality is indispensable.

Benchmark design

To evaluate the performance of CryoAtom2, we compiled a benchmark set of 175 non-redundant cryo-EM density maps containing NAs, ensuring that there is no overlap with training data in terms of release date and sequence similarity (see Methods). This benchmark consists of 151 protein–NA complex maps and 24 NA-only maps, allowing separate assessment for each category. As described in the Methods section, we further evaluated a subset of 50 non-redundant protein–NA complexes and a subset of 20 pure protein complexes to assess whether redundancy affects the performance of CryoAtom2. All evaluated methods, including CryoAtom2, CryoAtom, EM2NA, ModelAngelo, CryoREAD [12], and Phenix [20], were run locally using default parameters on identical inputs (density maps and corresponding sequences).

We conducted a comprehensive evaluation of CryoAtom2 across three key aspects: model accuracy against deposited structures, model geometric quality such as steric clashes, and model–map consistency. The results are presented separately for NAs, proteins, and protein–NA complexes. Beyond benchmarking, we showcase CryoAtom2's practical utility by applying it to three novel density maps, demonstrating its capacity to build near-complete atomic models for large, complex systems. Finally, we include ablation studies to assess component contributions and an analysis of computational efficiency.

Nucleic acid model accuracy against deposited structures

Model accuracy for NAs is assessed using the metrics adapted from the original CryoAtom study [9]: completeness, sequence accuracy (termed amino acid accuracy in CryoAtom), backbone recall, backbone precision, and backbone RMSD (see Methods for full definitions).

As summarized in Fig. 2, CryoAtom2 demonstrates comprehensive superiority over all other methods. A key challenge in NA modeling—unlike in proteins—lies in the accurate placement of bases, as the key hydrogen-bonding atoms critical for base pairing reside in the nucleotide bases themselves. We therefore introduce a composite metric, *completeness*, defined as the fraction of nucleotides in the deposited structure

that are both correctly positioned (within 3 Å RMSD to the reference for the P atom) and assigned the correct base identity. By this stringent criterion, CryoAtom2 achieves an average completeness of 60.0%, substantially outperforming EM2NA (40.6%), ModelAngelo (32.6%), CryoREAD (32.4%), and Phenix (11.7%) (Fig. 2a).

When interpreted together with the backbone recall rates (Supplementary Table 1), these results reveal a critical limitation of prior approaches: although EM2NA and ModelAngelo successfully reconstruct 79.4% and 78.7% of the NA backbones, respectively, they misassign base identities in roughly half of cases, resulting in their substantially lower completeness scores. This observation is further reinforced by the sequence accuracy metric (fraction of correctly identified bases among modeled residues; Fig. 2b), where CryoAtom2 attains an average of 70.2%, far surpassing EM2NA (49.9%), ModelAngelo (39.1%), CryoREAD (48.6%), and Phenix (36.5%).

These results highlight a longstanding bottleneck: reliable identification of nucleotide types directly from cryo-EM density remains challenging. A major contributing factor is the high similarity in local density signatures in purines (A vs. G) and pyrimidines (C vs. T/U), particularly at moderate resolutions (typically 3–5 Å), where base-specific features are often poorly resolved.

CryoAtom2 addresses this challenge through its inverse folding-inspired deep learning framework, which jointly integrates three complementary modalities: density, sequence, and—crucially—plausible structural information. The structural modality is explicitly incorporated via the Structure Encoder, enabling nucleotide base identity to be inferred not solely from local density features (which are often ambiguous between purines or between pyrimidines at moderate resolutions), but also from global geometric and stereochemical consistency with realistic 3D biomolecular structures. By reasoning over plausible conformations, CryoAtom2 substantially enhances discrimination of chemically similar bases, resulting in markedly improved base-type recognition and overall sequence accuracy.

When relaxing the evaluation to assess backbone accuracy alone—i.e., considering a prediction correct if the backbone atom P is within 3 Å of the deposited structure, irrespective of base identity—CryoAtom2 maintains a clear superiority (Supplementary Table 1). CryoAtom2 achieves the highest balanced performance, with precision of 88.3%, recall of 83.6%, and the lowest average RMSD of 0.82 Å for aligned backbone atoms. In comparison, EM2NA and ModelAngelo exhibit competent backbone

reconstruction but lag several percentage points behind in both precision and recall. CryoREAD performs notably weaker across all backbone metrics. The traditional Phenix pipeline recovers less than 30% of the backbone on average, underscoring the conservative nature of classical approaches and the substantial gains offered by modern deep learning methods in both accuracy and coverage.

We further evaluated the performance of various automated methods for NA modeling across different resolution ranges. To obtain sufficient data for statistical analysis, we considered 175 density maps containing NAs. Because the reported global resolution of these maps is dominated by the protein component, we computed the mean deposited Q-score [21] of the NA portion alone in each map and converted it to an effective resolution, thereby providing a resolution estimate specific to the NA regions. As shown in Fig. 2c, for high-resolution maps (better than 2.5 Å), only CryoAtom2 achieved a NA modeling completeness approaching that of human experts (~80%), substantially outperforming ModelAngelo (53%), the best-performing alternative at this resolution. Notably, the completeness achieved by CryoAtom2 at moderate resolution (3.5–4.0 Å; 52%) already matched that of other automated methods operating on high-resolution maps.

To examine how modeling accuracy depends on local map quality, we analyzed all 40,281 nucleotides across the 175 protein–NA complexes in our test set. Local map quality is quantified using the deposited Q-score [21], a per-residue metric that reflects local density–model agreement; higher Q-scores indicate better local resolution.

The completeness and backbone RMSD are plotted against the deposited Q-score in Supplementary Fig. 4. Across the full spectrum of local resolutions, CryoAtom2 consistently achieves higher completeness and lower backbone RMSD than all competing methods. Notably, in the ultra-high-resolution regime (deposited Q-score > 0.8), CryoAtom2 attains exceptional performance: 95% completeness and 0.5 Å RMSD, outperforming EM2NA (56%, 0.8 Å), ModelAngelo (66%, 0.6 Å), CryoREAD (50%, 1.9 Å), and Phenix (35%, 0.5 Å). These results position CryoAtom2 as the first automated method capable of reconstructing NA models from high-resolution cryo-EM density maps with accuracy approaching that of expert manual curation. With continued advances in cryo-EM technology, an increasing number of high-resolution NA cryo-EM density maps are expected to become available. CryoAtom2 therefore has strong potential to accelerate the workflow of NA structure determination from cryo-EM data.

Here, we present two representative examples to illustrate the accuracy of CryoAtom2 in nucleotide type recognition. In the first example (Fig. 2d), CryoAtom2 reconstructs the structure for *raiA RNA motif from Clostridium acetobutylicum* (2.9 Å resolution; PDB 9ELY, EMD-48162). It recovers 95.5% of the nucleotides with 99.5% sequence accuracy and a backbone RMSD of 0.60 Å. In contrast, EM2NA achieves only 53.0% completeness (59.4% sequence accuracy, 0.85 Å RMSD), and ModelAngelo recovers 51.5% of the structure with 55.0% sequence accuracy and 0.75 Å RMSD. This example demonstrates that CryoAtom2 exhibits substantially higher accuracy in recognizing the four RNA nucleotide types.

The second example (Fig. 2e) features the *elongating DRT2 reverse transcriptase* in complex with its non-coding RNA and dNTPs at 3.17 Å resolution (PDB 9C0J; EMD-45086), presenting a more challenging scenario with intertwined protein–RNA–DNA interactions. CryoAtom2 again delivers outstanding performance, achieving 89.0% sequence accuracy, far surpassing EM2NA (1.4%) and ModelAngelo (6.8%). Notably, EM2NA misclassified all RNA as DNA, while ModelAngelo correctly identified RNA but erroneously assigned all RNA nucleotides as pyrimidines, failing to distinguish purines from pyrimidines. In contrast, CryoAtom2 accurately captures the functional distinction between RNA and DNA and robustly discriminates all four nucleotide types in both RNA and DNA.

Nucleic acid model validation: geometric correctness and model–map agreement

We validate the model quality using two standard criteria in the cryo-EM community: geometric correctness and model–map agreement. Geometric correctness measures the geometric quality of the model with three metrics: MolProbity score [22], RMS(bonds), and RMS(angles). These metrics are computed using the program phenix.molprobity [23]. Model–map agreement is assessed using four correlation coefficients (CC): CC_{mask}, CC_{volume}, CC_{peaks}, and CC_{box}, which are computed with the program phenix.map_model_cc [24].

Note that CryoREAD generates non-standard NA structures, leading to errors in running the above programs. Moreover, models produced by Phenix are often too incomplete relative to deposited references (see Fig. 2a). Therefore, the following analyses focus on CryoAtom2, EM2NA, and ModelAngelo.

We first evaluate the geometric correctness. Fig. 3a shows that across the 175 test maps, CryoAtom2 achieves the best average MolProbity score (3.31), outperforming EM2NA (3.95) and ModelAngelo (3.92). This indicates that, in addition to achieving higher model accuracy, CryoAtom2 also generates models that are intrinsically more geometrically sound. We further assess bond lengths and bond angles separately. The results show that CryoAtom2 produces the most reasonable stereochemistry, with the lowest RMS(bonds) (0.03 Å) and RMS(angles) values (3.15°), substantially better than EM2NA (0.22 Å and 13.00°) and ModelAngelo (0.11 Å and 9.41°).

An illustrative example is shown in Fig. 3d. For the *Rous sarcoma virus frameshifting pseudoknot RNA EM straight dimer* structure (PDB 9DII, EMD-46906, reported resolution of 4 Å), CryoAtom2 yields the best stereochemistry (RMS(bonds) 0.02 Å and RMS(angles) 2.47°), compared with EM2NA (0.26 Å and 14.90°) and ModelAngelo (0.13 Å and 9.80°). This results in geometrically continuous chains in CryoAtom2 models, whereas EM2NA and ModelAngelo frequently exhibit chain-break artifacts due to incorrect backbone conformations and/or missing connecting nucleotides. CryoAtom2 also achieves the highest backbone recall (94.6%) in this case, versus 91.8% (EM2NA) and 85.9% (ModelAngelo) (Supplementary Table 2). These results indicate that CryoAtom2 effectively learns NA topology and generates geometrically plausible models even from low-resolution density.

We next assess the model–map consistency. Because phenix.map_model_cc calculations on protein-containing maps produced errors for EM2NA models (designed specifically for NAs), we restricted this analysis to the 24 NA-only density maps in the test set. As shown in Fig. 3b, CryoAtom2 again achieves the highest correlations across all four metrics (CC_{mask}: 0.65, CC_{volume}: 0.65, CC_{peaks}: 0.55, CC_{box}: 0.70), surpassing both EM2NA (0.55, 0.60, 0.51, 0.63) and ModelAngelo (0.45, 0.46, 0.40, 0.60). These results indicate that CryoAtom2 achieves a superior fit to the experimental density for NA modeling without compromising geometric quality.

An illustrative example is presented in Fig. 3c for the structure of the *Rail* RNA motif. The corresponding cryo-EM density map (EMD-45988) has a reported resolution of 3 Å. The model generated by CryoAtom2 demonstrates a strong correlation with the density map, achieving CC_{mask}, CC_{volume}, CC_{peaks}, and CC_{box} values of 0.73, 0.73, 0.70, and 0.87, respectively. These values outperform those obtained by EM2NA (0.61, 0.69, 0.60, 0.71) and ModelAngelo (0.46, 0.46, 0.48, 0.74). Visually, the CryoAtom2 model (blue) aligns well with both the density map (gray) and the deposited structure (PDB

9CXF, in green), whereas EM2NA (yellow) places false-positive bases in regions lacking density and misassigns base identities, and ModelAngelo (magenta) produces implausible backbone conformations with poor density fit. This case confirms that CryoAtom2 generates more accurate, geometrically sound, and density-consistent atomic models.

CryoAtom2 further improves protein model building

To enable a direct comparison with CryoAtom and ModelAngelo, we assessed CryoAtom2 on the 54 protein test cases reported in the CryoAtom study. As shown in Fig. 4, CryoAtom2 exhibits consistent and substantial improvements across multiple metrics, including global completeness, geometric correctness, and model–map agreement. In terms of completeness (Fig. 4a), CryoAtom2 achieves an average of 85.2%, compared with 73.6% for CryoAtom and 65.5% for ModelAngelo. Geometric quality is further compared in Fig. 4b. CryoAtom2 achieves the lowest average MolProbity score (2.99) across the 54 proteins, outperforming CryoAtom (3.65) and ModelAngelo (3.74). This indicates that the protein models generated by CryoAtom2 not only exhibit higher overall completeness but also possess more physically plausible local geometry with fewer steric clashes. Importantly, this improvement in geometric quality is not obtained at the expense of low model–map consistency. Compared with NA models, two additional model–map correlation coefficients (CC_{mc} and CC_{sc}) are further calculated for protein models. As shown in the radar plot in Fig. 4c, CryoAtom2 achieves the best performance across all CC metrics (CC_{mask}, CC_{volume}, CC_{peaks}, CC_{box}, CC_{mc}, and CC_{sc}) with values of 0.63, 0.64, 0.48, 0.58, 0.62, and 0.60, surpassing CryoAtom (0.61, 0.61, 0.44, 0.54, 0.60, 0.57) and ModelAngelo (0.63, 0.63, 0.40, 0.50, 0.61, 0.58).

A representative example is the *Isobutyryl-CoA mutase Q341A fusion in complex with GTP* (4.6 Å resolution; PDB 8SSL, EMD-40751; Fig. 4d). CryoAtom2 reconstructs a markedly more complete and continuous model (66.1% completeness), whereas CryoAtom and ModelAngelo recover only 24.5% and 7.4%, respectively. The CryoAtom2 model appears more coherent, clearly defined, and biologically interpretable within the density envelope, while other methods produce fragmented or sparsely traced structures. Importantly, despite modeling substantially more residues and building a more complex all-atom representation, CryoAtom2 maintains a favorable MolProbity score (3.66) compared with CryoAtom (4.04) and ModelAngelo (3.95). These results demonstrate that CryoAtom2 improves the overall completeness without compromising local geometric quality.

Performance on protein–NA complexes

In the previous sections, we demonstrated substantial advances of CryoAtom2 in both NA and protein modeling through detailed and separate evaluations. In this section, we shift our focus to protein–NA complexes as integrated macromolecular systems. We evaluate CryoAtom2 on a benchmark of 151 protein–NA complexes (derived from 175 test maps after excluding 24 NA-only cases). ModelAngelo, currently the only available AI-based method for protein–NA complex model building, is selected for comparison. Unlike the previous analyses, where protein or NA components were evaluated separately, all evaluation metrics in this section jointly consider both protein and NA components.

CryoAtom2 outperforms ModelAngelo on nearly all complexes (Fig. 5a), achieving an average completeness of 85.6% (vs. 64.1% for ModelAngelo) and sequence accuracy of 93.2% (vs. 83.3%). Despite reconstructing substantially more structural regions, CryoAtom2 maintains a better overall backbone RMSD than ModelAngelo (0.64 Å vs. 0.72 Å). The performance advantage of CryoAtom2 over ModelAngelo becomes more pronounced at lower resolutions, suggesting that its inverse folding-inspired multimodal framework reduces reliance on high-resolution density, resulting in enhanced robustness across varying resolutions.

Furthermore, we evaluate the intrinsic geometric quality and model–map correlation of the models generated by CryoAtom2 and ModelAngelo. First, as shown in Fig. 5b, CryoAtom2 achieves a lower average MolProbity score (2.79) than ModelAngelo (3.60), indicating that CryoAtom2 generates geometrically more plausible structures. Despite imposing stricter geometric restraints, CryoAtom2 simultaneously exhibits superior agreement with the experimental density maps (Fig. 5c). CryoAtom2 outperforms ModelAngelo across all six correlation coefficients, achieving CC_{mask}, CC_{volume}, CC_{peaks}, CC_{box}, CC_{mc}, and CC_{sc} values of 0.66, 0.66, 0.56, 0.68, 0.67, and 0.65, respectively, compared with 0.57, 0.58, 0.44, 0.59, 0.67, and 0.56 for ModelAngelo. These results demonstrate that CryoAtom2 achieves a remarkable balance by preserving local geometric quality without sacrificing it for enhanced density fitting. In conclusion, CryoAtom2 stands at the forefront of density-map-based modeling for protein–NA complexes, and we expect it to become an indispensable tool for atomic modeling within the cryo-EM community.

We illustrate the modeling capabilities of CryoAtom2 on a challenging *147-bp 5S rDNA nucleosome–closed* complex (PDB ID: 9D3R; EMD-46545; reported resolution: 3.3 Å). Notably, this example is annotated in the EMDB [25] with a deposited Q-score of 0.393, indicating an effective resolution worse than 4 Å, which poses significant challenges for accurate *de novo* modeling. As shown in Fig. 5d, CryoAtom2 achieves a modeling completeness of 76.1%, substantially outperforming ModelAngelo (24.3%). ModelAngelo struggles to reliably reconstruct the backbones for both protein and NA (backbone recall: 49.1%; Supplementary Table 3). In contrast, CryoAtom2 excels, successfully constructing a continuous backbone structure from the noisy density map, achieving a remarkable backbone recall of 97.0% (Supplementary Table 3). Despite the limited map quality, CryoAtom2 attains higher sequence accuracy and lower backbone RMSD than ModelAngelo (78.4% and 0.64 Å vs. 49.5% and 0.94 Å, respectively). This example highlights the advantage of CryoAtom2 in modeling protein–NA complexes from medium- to low-resolution density maps using its inverse folding-inspired multimodal framework.

Performance on non-redundant sets

For the 54 protein test cases from the CryoAtom study, we did not take their similarity to the newly expanded CryoAtom2 training set into account. Similarly, to obtain a sufficient number of NA maps for benchmarking, protein similarity was not considered when constructing the 175 NA-containing maps. Here, we applied a 30% sequence identity cutoff to remove redundancy among proteins, resulting in non-redundant test sets comprising 20 pure proteins (Supplementary Table 4) and 50 protein–NA complexes (Supplementary Table 5) for further evaluation.

We first analyzed the 20 non-redundant pure protein test cases, with the overall results shown in Supplementary Fig. 5. Notably, after redundancy removal, the completeness of CryoAtom2 decreased from 85.2% to 81.6%. This performance drop may be attributed to the lower average resolution of the test set after redundancy removal, which decreased from 3.44 Å to 3.74 Å. Nevertheless, CryoAtom2 still substantially outperformed CryoAtom (65.3%) and ModelAngelo (56.4%) in terms of completeness. While constructing more regions, CryoAtom2 also maintained superior model quality, achieving a MolProbity score of 3.17 compared to 3.74 for CryoAtom and 3.82 for ModelAngelo. In terms of map interpretability, CryoAtom2 achieved the best performance across all correlation coefficient (CC) metrics—CC_{mask}, CC_{volume}, CC_{peaks}, CC_{box}, CC_{mc}, and CC_{sc}—with values of 0.62, 0.62, 0.46, 0.58, 0.60, and 0.59,

respectively, surpassing CryoAtom (0.60, 0.60, 0.41, 0.50, 0.58, 0.55) and ModelAngelo (0.60, 0.60, 0.38, 0.45, 0.58, 0.55).

Next, we analyzed the 50 non-redundant protein–NA test cases, with the overall results shown in Supplementary Fig. 6. We observed that after redundancy removal, the completeness of CryoAtom2 remained largely unchanged (85.6% vs. 85.8%). This stability may be due to the similar average resolution of the test set before and after redundancy removal (3.13 Å vs. 3.13 Å). Nevertheless, CryoAtom2 still significantly outperformed ModelAngelo, which achieved 61.5% completeness. While modeling more regions, CryoAtom2 also maintained superior model quality, achieving a MolProbity score of 2.85 compared to 3.69 for ModelAngelo. Regarding map interpretability, CryoAtom2 achieved the best performance across all correlation coefficient (CC) metrics—CC_{mask}, CC_{volume}, CC_{peaks}, CC_{box}, CC_{mc}, and CC_{sc}—with values of 0.64, 0.64, 0.53, 0.65, 0.66, and 0.62, respectively, surpassing ModelAngelo (0.52, 0.53, 0.41, 0.55, 0.65, 0.52).

These results indicate that CryoAtom2 significantly outperforms other methods even on non-redundant protein and protein–NA test sets. The completeness of CryoAtom2 is primarily determined by the resolution of the density maps (Fig. 2c) rather than the sequence similarity with the training data, demonstrating the strong generalizability of CryoAtom2.

Applications of CryoAtom2 to three novel cryo-EM maps

To assess the practical utility of CryoAtom2 in modeling complex biomolecular systems, we applied it to three recently solved cryo-EM maps outside our training and benchmark datasets: two large protein–NA complexes (an 80S ribosome and a 70S ribosome) and one large RNA nanocage.

In situ human hibernating 80S ribosome in the rotated-3 state with Z-site tRNA [26]. This *in situ* structure constitutes a highly complex macromolecular assembly, comprising 13,520 amino acids and 5,743 nucleotides (PDB 9P79; EMD-71335; reported resolution: 3.1 Å). As shown in Fig. 6a, CryoAtom2 successfully modeled 77.4% of the amino acids and 60.1% of the nucleotides in this ribosome. In contrast, ModelAngelo reconstructed only 54.1% of the amino acids and 37.0% of the nucleotides, resulting in a substantial gap in overall completeness (CryoAtom2: 72.3% vs. ModelAngelo: 49.0%). A clear qualitative difference can be observed at the peripheral regions of the structure (highlighted by the black box at the top of Fig. 6a), where CryoAtom2 builds substantially more atomic structures than ModelAngelo. This example shows that

CryoAtom2 delivers superior performance on large, *in situ* ribosomal maps and holds promise for atomic-level modeling of other *in situ* datasets.

Archaeal 70S ribosome [27]. This is a newly determined structure with a reported resolution of 2.3 Å, which has been deposited in the PDB (PDB 9XHK, EMD-66877). We compare the models generated by CryoAtom2 and ModelAngelo against the reference structure. The results demonstrate that CryoAtom2 successfully reconstructs all three rRNAs and all 56 ribosomal proteins, achieving an overall completeness of 95.9% (Fig. 6b)—significantly surpassing the 89.3% attained by ModelAngelo. Unmodeled segments are primarily localized in highly flexible regions where the local resolution dropped to approximately 8 Å (Fig. 6c). Notably, even in moderately poorly resolved areas (~5 Å local resolution), the predicted model exhibits high accuracy with a backbone RMSD of 0.58 Å relative to the reference, underscoring the robustness of CryoAtom2 in poorly resolved density (Fig. 6d).

Regarding sequence assignment, both CryoAtom2 and ModelAngelo show near-perfect accuracy (99%) for protein components. However, CryoAtom2 demonstrates a substantial advantage in NA modeling, achieving a sequence accuracy of 96.2% compared to 81.5% for ModelAngelo. Furthermore, the atomic models generated by CryoAtom2—both for individual bases and the overall NA chains—show superior stereochemical plausibility. Specifically, CryoAtom2 exhibits remarkable conformational accuracy for unpaired bases (Fig. 6e), which are typically the most difficult to model *de novo.*

GOLLD RNA-only nanocage. The structure of the RNA-only nanocage has recently been solved by at least four independent research groups using cryo-EM [28-31], reflecting the growing interest in elucidating RNA-only assemblies. We sought to test if CryoAtom2 could successfully reconstruct this structure using one of the cryo-EM maps (PDB ID: 9MEE; EMD-48214; reported resolution: 3 Å). This structure comprises a total of 11,032 nucleotides from 14 homomeric subunits (Fig. 6f).

Given the sufficiently high resolution of this density map, both CryoAtom2 and ModelAngelo are able to reconstruct the majority of the NA regions, achieving a backbone recall of over 95% (Fig. 6f). The primary distinction between the two models lies in their base type assignment (Fig. 6g). CryoAtom2 achieves an impressive overall sequence accuracy of 97.8%, significantly surpassing ModelAngelo's 84.3%. This difference contributes to a higher completeness in CryoAtom2's model compared to ModelAngelo, with values of 93.5% versus 80.8%. Notably, CryoAtom2's base type

assignment approaches the accuracy of human experts, benefiting from its inverse folding-inspired multimodal framework.

In summary, these applications demonstrate that for large macromolecular assemblies, CryoAtom2 not only excels in structural completeness but also maintains high accuracy in low-resolution regions. Its ability to significantly outperform existing tools in both accuracy and stereochemical quality of NA modeling highlights its comprehensive capabilities for interpreting complex biological systems in cryo-EM.

Ablation studies

To evaluate the contribution of CryoAtom2's key components while controlling computational cost, we conducted a series of ablation experiments. Because a full training run of CryoAtom2 is resource-intensive (approximately 10 days for 50 epochs on 8 NVIDIA H200 GPUs), we first established a lightweight baseline model for efficient experimentation. This baseline retains the core architecture of CryoAtom2 but incorporates the following hyperparameter reductions to lower computational cost: node and edge representation dimensions in CryoAtom2 were reduced from 384 to 256 and from 192 to 128, respectively. The number of nearest neighbors (k) for graph edge construction was decreased from 30 to 20. The number of Cryo-Former layers was reduced from 16 to 14. The total training epochs were shortened to 30. This streamlined baseline enables rapid iteration while preserving the fundamental architectural innovations of CryoAtom2.

Using this baseline, we created two additional variants of the baseline to evaluate the contributions of individual components. The first removes the input of language model-based sequence embedding. The second eliminates the dedicated Structure Encoder and reverts to the strategy employed in the original CryoAtom, where predicted Cα (protein) and P (NA) atoms are iteratively recycled as input to the network.

Ablation results for nucleic acids and proteins.

The impact of each component is shown for NAs in Fig. 7a and for proteins in Fig. 7b, with consistent conclusions across both biomolecule types. For NAs at 30 epochs, the baseline achieves a sequence accuracy of 56.8%. This drops to 54.6% without sequence embeddings and falls to 51.3% without the Structure Encoder, confirming that both the sequence embedding and the inverse folding–inspired strategy are critical for nucleotide type recognition.

We also observe that early in training (around epoch 4), the baseline model shows lower sequence accuracy than the variant without the Structure Encoder. We hypothesize this occurs because the network has not yet learned sufficient NA topology to allow the Structure Encoder to provide useful context to complement sequence prediction. Once geometric accuracy improves (all-atom RMSD < 3 Å around epoch 7), the Structure Encoder in the baseline begins to contribute effectively, boosting sequence accuracy. At 30 epochs, the all-atom RMSD is 1.36 Å for the baseline, 1.38 Å without sequence embedding, and 1.64 Å without the Structure Encoder. This demonstrates that the Structure Encoder module helps the network refine and correct the predicted structures.

For proteins, the contributions are analogous. The baseline achieves a sequence accuracy of 88.4%, which drops to 67.6% without sequence embeddings and 81.6% without the Structure Encoder. Structural accuracy follows the same trend (baseline: 0.89 Å, w/o sequence embedding: 0.91 Å, w/o Structure Encoder: 0.97 Å). These results highlight that protein language models are particularly powerful for amino acid identification. The comparatively smaller gain from NA language models reflects their current lower informational power relative to protein models [5]. Together with the ablation study in CryoAtom using different protein language models (ESM-2 vs. ESM-1b [32]; Supplementary Fig. 7), CryoAtom2's performance could be further improved by integrating more advanced NA language models as they become available.

Analysis of nucleic acid sequence alignment strategy.

CryoAtom2 employs an advanced strategy to resolve ambiguity in NA sequence alignment (see Methods). To illustrate its effectiveness in distinguishing between DNA and RNA, we examine the recently engineered *IscB*, which facilitates persistent epigenome editing (PDB 9NVU, Fig. 7c). This structure features an RNA (in blue/green nucleotides) alongside a double-helical DNA (in red/yellow nucleotides).

In scenarios where no reference sequence is provided (denoted as "w/o sequence align"), nucleotide assignments rely solely on the network-predicted most probable types, resulting in a sequence accuracy of 55.3%. While the DNA nucleotides are accurately identified, some RNA nucleotides are incorrectly classified as DNA. This misassignment highlights the similarities in density features between DNA and RNA in cryo-EM maps.

Conversely, when a reference sequence is available, accuracy improves significantly. Utilizing a simplified, two-letter (purine/pyrimidine) alignment strategy—similar to

ModelAngelo (denoted as "MA align")—one DNA strand is misclassified as RNA, achieving an accuracy of 82.3%. In contrast, by integrating network-predicted probabilities for all eight nucleotide types (four RNA nucleotides and four DNA nucleotides), CryoAtom2 successfully distinguished the double-helical DNA and correctly assigned most RNA components, achieving an accuracy of 90.0%. This demonstrates the importance of a nuanced alignment strategy in resolving ambiguities within hybrid NA complexes.

Computational efficiency

In this section, we investigate the computational efficiency of *de novo* atomic modeling methods for NA and protein complexes. Notably, methods that rely solely on CPUs for modeling, such as Phenix, were excluded from this analysis. Supplementary Fig. 8 shows the inference time as a function of the total length of the constructed model, based on data from a single A100 GPU and 175 test density maps. The results reveal a clear linear relationship between inference time and total model length across all methods. CryoAtom2 and ModelAngelo are comparable and are approximately 10× and 35× faster than EM2NA and CryoREAD, respectively. For example, CryoAtom2 can build 5,000 nucleotides (or amino acids) in ~30 minutes on a single A100. In contrast, EM2NA and CryoREAD require ~400 minutes and ~1,000 minutes, respectively. The high accuracy and efficiency of CryoAtom2 position it well to meet the demands of high-throughput automated modeling.

Discussion

Building on the recent breakthrough in inverse folding and structure prediction, we have developed CryoAtom2, an inverse folding-inspired multimodal framework for *de novo* model building of NAs and protein–NA complexes in cryo-EM. Comprehensive benchmarks demonstrate that CryoAtom2 substantially outperforms existing tools across NAs, proteins, and protein–NA complexes. In addition to higher completeness, CryoAtom2 consistently improves the geometric quality and model–map correlation, indicating not only more accurate atomic placement but also better structural interpretability and stereochemical correctness. These results highlight that CryoAtom2 advances beyond simply “building more atoms”—it produces models that are both physically plausible and better explained by the density, addressing a long-standing bottleneck in cryo-EM modeling of NAs and protein–NA complexes. Beyond accuracy, CryoAtom2 belongs to the top tier in computational efficiency, achieving near-linear scaling with model length and enabling high-throughput automated modeling.

Application to three recently solved maps of large complex biomolecular systems involving NAs further demonstrates the robustness of CryoAtom2. Taken together, these results underscore CryoAtom2's unique combination of accuracy, robustness, and efficiency, establishing it as a practical and scalable solution for interpreting complex biological machinery in cryo-EM.

Nevertheless, we acknowledge that CryoAtom2 still has room for further improvement in NA model building. On one hand, ablation studies indicate that the sequence accuracy of NAs predicted directly by the neural network is substantially lower than that of proteins, with a gap of approximately 30%. On the other hand, when employing HMMER for sequence alignment, NAs are represented by an alphabet of only four letters, which conveys considerably less information than the 20-letter alphabet for proteins. These two factors together limit the accuracy of nucleotide identity assignment, which has not yet reached the level achieved for proteins. To address the first limitation, we anticipate progress from the development of more powerful NA language models that can compete with current protein language models. Regarding the second limitation, incorporating secondary structure constraints presents a promising approach to enhance the information content of sequence alignments. We are committed to continuously improving CryoAtom2 in future studies by integrating these new features.

Methods

CryoAtom2 algorithm

We present CryoAtom2, a substantially more powerful version that extends protein model building in CryoAtom to protein–NA complexes (Fig. 1a). Given the challenges in building accurate all-atom models for NAs, CryoAtom2 incorporates concepts from protein and NA sequence design. It encodes structural information to predict plausible sequences (Fig. 1b), thereby helping low-resolution density maps recover fine-grained atomic features of complex assemblies.

CryoAtom2 is an inverse folding-inspired deep-learning framework that tightly couples three modalities—density maps (via the Cryo-EM Encoder), sequences (via the Cryo-Former), and structures (via the Structure Encoder). CryoAtom2 inherits the overall architecture of the previous generation, with several key adjustments introduced to address the challenges in NA modeling (see below).

Determining atomic positions in Stage 1

Compared with CryoAtom, we incorporate both convolution and attention mechanisms to simultaneously capture both local and global features of the density map (Supplementary Fig. 1 and Supplementary Algorithm 1). Moreover, we adopt a recurrent U-Net architecture (RU-Net), which increases the effective inference depth without adding parameters (Supplementary Algorithm 1.1), thereby improving network performance.

The network takes as input a cropped density map (size 1 × 64 × 64 × 64) and outputs a four-channel probability density map (size 4 × 64 × 64 × 64) corresponding to Cα atoms for proteins, P and C4′ atoms for NAs, and others. RU-Net is trained using a weighted Focal Loss [33] defined as follows:

$$\mathcal{L}_1 = \sum_{c=1}^{4} \sum_{x} -\frac{1}{N_c} y_c(x)\left(1 - p_c(x)\right)^{\gamma} \log\left(p_c(x)\right) \quad (1)$$

Here, N_c represents the number of samples belonging to class c. $y_c(x)$ represents the one-hot label corresponding to voxel x. $p_c(x)$ represents the predicted probability that voxel x belongs to class c. The hyperparameter γ is set to 2.

If the predicted probability of a channel exceeds the threshold (>0.5), the corresponding voxel is assigned to the class represented by that channel. For Cα atoms, we retain the same post-processing strategy as in CryoAtom. For P atoms, we follow the carefully designed mean-shift algorithm from EM2NA to cluster the discrete voxel coordinates and produce continuous coordinates (local maxima of the probability map). The resulting P and Cα atoms are then used to initialize the residues, which are subsequently refined in Stage 2.

Building all-atom model in Stage 2

In this stage, we first construct node and edge representations for proteins and NAs using a convolutional network (referred to as the Cryo-EM Encoder) based on the Stage 1 output (P and Cα atoms). Next, the 16-layer Cryo-Former module with non-shared weights integrates sequence information into both the node (1D) and edge (2D) representations, producing updated 1D and 2D features. Finally, the 7-layer Structure Module with shared weights decodes these representations to generate the all-atom structure of the protein–NA complex.

During the recycling stage, inter-atomic distances within the complex are encoded by a

Structure Encoder, injecting structural information back into the 1D and 2D representations. This step is inspired by inverse folding tasks for proteins and NAs, where structural context is used to infer plausible amino-acid or nucleotide identities. Such structure-guided inference plays an important role in accurately determining residue and base identities from density maps (see Ablation studies).

The main extensions in this stage compared with CryoAtom include the construction of sequence embeddings, the introduction of the Structure Encoder module, the Bayesian-estimated tracing algorithm, sequence assignment, and additional geometric losses. They are introduced in detail below.

Construction of sequence embeddings. To enable the model to jointly process protein and NA sequence embeddings, we use protein and RNA language models to generate their respective representations. For DNA sequences, the letter T is replaced with U so that they can be fed into the RNA language model. More specifically, the composite sequence embedding can be defined by the following formulation.

$$S(\text{seq}) = \begin{cases} \left[S^{prot}; \vec{0}_{640}; 0 \right], & \text{if seq is protein,} \\ \left[\vec{0}_{1280}; S^{RNA}; 0 \right], & \text{if seq is RNA,} \\ \left[\vec{0}_{1280}; S^{RNA}; 1 \right], & \text{if seq is DNA} \end{cases} \tag{2}$$

where $S(\text{seq}) \in \mathbb{R}^{L\times 1921}$ represents the composite sequence embedding for sequences of length L. $S^{prot} \in \mathbb{R}^{L\times 1280}$ represents the sequence embedding of length L generated by the protein language model (ESM-2). $S^{RNA} \in \mathbb{R}^{L\times 640}$ represents the sequence embedding of length L generated by the RNA language model (RNA-FM).

With this design, the sequence embedding can uniformly handle representations of both proteins and NAs. Moreover, it can be easily shown that when such embeddings pass through a linear layer in the neural network, the RNA embedding and the DNA embedding differ only by a bias (a translation vector). This fully leverages the capability of the pretrained RNA language model as well as the inherent similarity between DNA and RNA.

Structure Encoder. Inspired by protein inverse folding tasks (e.g., ProteinMPNN [19]), incorporating structural information as additional inputs to CryoAtom2 can help improve modeling performance. On the one hand, we integrate the node representation output by the Structure Module—containing rich structural information—into the initial node representation. On the other hand, CryoAtom2 incorporates inter-atomic distance information from the predicted structures into the initial edge representation. As shown in Fig. 1b and Supplementary Fig. 3, CryoAtom2 predicts 28 sets of all-atom

coordinates for each node (20 standard amino acids, 4 RNA nucleotides, and 4 DNA nucleotides). In this module, we construct a protein–NA graph by labeling Cα atom nodes as alanine (ALA) and P atom nodes as guanine (G) and cytosine (C). This design facilitates CryoAtom2 to consider representative atoms in proteins (N, Cα, C, O, Cβ) and NAs (P, O3′, O2, N3, N4 for cytosine, N7, O6, N1, N2, N3 for guanine), which define the backbone geometry and base chemistry that underpin secondary-structure organization. Moreover, for NAs, representing each nucleotide using a purine–pyrimidine categorical representation enables the network to evaluate different combinatorial interaction patterns, which in turn helps resolve whether a given nucleotide is more consistent with a purine or a pyrimidine identity. This module fully exploits the structural information embedded in cryo-EM density maps and automatically captures the dominant interactions between nodes, such as peptide bonds, phosphodiester bonds, and base pairing.

Bayesian-estimated tracing algorithm. After localizing all-atom coordinates from the cryo-EM density map, a critical subsequent step is to connect isolated residues into one or more continuous chains. In ModelAngelo and related algorithms [6, 12], the chain-linking problem is formulated as a graph optimization task that minimizes the total distance between connected nodes. Based on the empirical observation that peptide bonds and phosphodiester bonds have characteristic lengths of approximately 1.3 Å and 1.6 Å, respectively, heuristic algorithms are employed to identify a connectivity pattern that minimizes the total peptide-bond or phosphodiester-bond distance. However, this strategy is susceptible to false-positive residues predicted by the model, which may lead to incorrect connections between unrelated chains, as illustrated in Supplementary Fig. 2b.

To address this issue, CryoAtom2 introduces a *prior* probability $p(e_{ij})$ for each potential edge, representing the likelihood that two residues form a peptide bond (or a phosphodiester bond). These probabilities are predicted by a supervised neural network, as shown in Supplementary Fig. 2a. Specifically, we reformulate the original minimum-total-distance chain-linking problem as a *posterior probability maximization* problem under a Bayesian framework:

$$\omega^{*} = \underset{\omega\in\Omega,\omega_{ij}\in\{0,1\}}{\arg\max} \sum_{i\neq j} \omega_{ij} \log p\left(e_{ij}\left|structure\right.\right) + \sum_{i\neq j}\left(1-\omega_{ij}\right)\log\left(1-p\left(e_{ij}\left|structure\right.\right)\right)$$

$$p\left(e_{ij}\left|structure\right.\right) \propto p\left(e_{ij}\right) p\left(structure\left|e_{ij}\right.\right) \tag{3}$$

$$p\left(structure\left|e_{ij}\right.\right) = 1 - e^{-\left(\sigma/d_{ij}\right)^{6}}$$

where ω_{ij} is a binary indicator variable denoting whether an edge between nodes i and

j is selected. Ω denotes the set of valid connectivity configurations, for example, enforcing that each node has at most one successor and at most one predecessor, and prohibiting self-loops or conflicting connections. $p(e_{ij}|\text{structure})$ represents the *posterior* probability of adjacency between nodes i and j along the chain under a given atomic structure. $p(e_{ij})$ denotes the predicted *prior* probability of adjacency between nodes i and j along the chain, $p(\text{structure}|e_{ij})$ characterizes the structural plausibility after connecting nodes i and j, which can be approximated by mapping the attractive term of the Lennard–Jones potential [34] into probability space (the corresponding functional form is illustrated in Supplementary Fig. 2c). d_{ij} represents the distance between the C atom of the i-th residue and the N atom of the j-th residue for proteins, or between the O3′ atom of the i-th residue and the P atom of the j-th residue for NAs. Here, σ is set to 1.7 for proteins and 1.9 for NAs. Finally, the problem can be solved using the heuristic flood-fill algorithm adopted in CryoAtom.

Bayesian estimation refines the neural network–predicted prior $p(e_{ij})$ using the inferred all-atom coordinates to obtain the posterior probability. If the prior term predicted by the neural network is ignored (i.e., $p(e_{ij})\equiv 1$) in Eq. 3, the problem degenerates into a graph optimization task that minimizes the total distance between connected nodes, leading to the failure mode illustrated in Supplementary Fig. 2b. In contrast, the incorporation of the prior probability term prevents the algorithm from favoring locally shortest edges and promotes globally consistent chain connectivity (Supplementary Fig. 2d).

Sequence assignment. After tracing and connecting isolated residues into multiple backbone chains, the next step is sequence assignment. In the absence of reference sequences, each residue is assigned the type with the highest probability predicted by the neural network (Supplementary Fig. 2a). When reference sequences are available, all chains generated by CryoAtom2 are treated as query sequences for sequence alignment.

For each protein chain, following the same strategy as CryoAtom, the predicted probabilities of the 20 amino acid types are converted into an HMM profile, which is used in HMMER [35] to perform multiple sequence alignment between the query and reference sequences. The alignment with the largest number of matched positions is selected, and residues at matched positions are assigned the amino acid identities of the corresponding reference sequence positions.

For each NA chain, directly performing sequence alignment using the predicted

probabilities of the eight nucleotide types is challenging, because purines are highly similar to each other (G vs. A), and pyrimidines are likewise highly similar to each other (C vs. T/U) in cryo-EM density maps. Inspired by ModelAngelo, to address this issue, CryoAtom2 maps both query and reference nucleotide sequences from eight nucleotide types into two classes—purines (encoded as the letter G) and pyrimidines (encoded as the letter C)—and performs multiple sequence alignment in HMMER using this reduced alphabet. By reducing the effective alphabet size, this representation alleviates information sparsity caused by prediction noise and uncertainty, thereby improving alignment hit rates. However, this strategy may also introduce false-positive alignments, for example, aligning an RNA query sequence to a DNA reference sequence by removing the distinction between RNA and DNA (see Ablation studies).

To mitigate this issue, we redefine the matching score for each alignment by ignoring gap penalties as follows:

$$
\begin{aligned}
T(s) &= \begin{cases} G & s = A, G, DA, DG, \\ C & s = C, U, DC, DT. \end{cases} \\
\left(p \circ T^{-1}\right)(y) &\triangleq \begin{cases} \sum_{T(s)=G} p(s) & y = G, \\ \sum_{T(s)=C} p(s) & y = C. \end{cases} \\
\pi^{*} &= \arg\max_{i} \sum_{(k,m)\in\pi_i} \log p_k\left(s_m\right)
\end{aligned}
\tag{4}
$$

where *T(s)* denotes a mapping that collapses the eight nucleotide symbols (with DA representing adenine in DNA) into two categories: purines (encoded as G) and pyrimidines (encoded as C). The $p \circ T^{-1}$ maps the probability values in the nucleic-acid HMM profile file to the probabilities of the two-category purine/pyrimidine scheme. The π_i denotes the set of matched position pairs for the *i*-th alignment. The p_k represents the nucleotide probability at position *k* of the query sequence as predicted by CryoAtom2. The s_m denotes the nucleotide symbol at position *m* of the *i*-th reference sequence. Under this redefined matching score, we identify the optimal alignment π^* and mutate each nucleotide in the query sequence to the nucleotide at the corresponding matched position in the alignment.

Geometric losses. The model employs a loss function similar to that used in CryoAtom. To overcome the geometric clashes observed in protein structures generated by CryoAtom, CryoAtom2 introduces additional geometric loss terms to produce physically plausible protein–NA complex structures.

$$\mathcal{L}_2 = \mathcal{L}_{\text{multi}} + \mathcal{L}_{\text{single}} + \underbrace{\mathcal{L}_{\text{bond}} + \mathcal{L}_{\text{clash}}}_{\text{fine-tune}}$$
$$\mathcal{L}_{\text{multi}} = \mathcal{L}_{\text{bb}} + 1.3\mathcal{L}_{\text{FAPE}} + 0.025\mathcal{L}_{\text{pbb}} + \mathcal{L}_{\text{existence}} \quad (5)$$
$$\mathcal{L}_{\text{single}} = \mathcal{L}_{\text{atom}} + \mathcal{L}_{\text{torsion}} + \mathcal{L}_{\text{attention}} + 2\mathcal{L}_{\text{violation}} + 0.3\mathcal{L}_{\text{aa}} + \mathcal{L}_{\text{edge}}$$

A detailed description of the above loss terms for proteins can be found in the CryoAtom paper [9]. These terms can be directly extended from proteins to NAs in a straightforward manner. For example, for proteins, the backbone loss $\mathcal{L}_{\text{bb}}$ is defined as the RMSD of the main-chain framework (N–Cα-C), whereas for NAs, the backbone framework is defined by the phosphate group (OP1–P–OP2). During the later stage of training, CryoAtom2, similar to trRosettaRNA2 [36], enables geometric loss terms for fine-tuning. Specifically, for NAs, the bond loss $\mathcal{L}_{\text{bond}}$ enforces the O3′–P bond length and bond angles such as O3′–P–O5′ to remain within ideal ranges, which are derived from statistics of the training data; for proteins, the same loss constrains the C–N bond length and bond angles such as Cα-C-N. The clash loss $\mathcal{L}_{\text{clash}}$ penalizes inter-atomic distances that are smaller than min$(d - 0.1\ \text{Å},\ 2\ \text{Å})$, where d denotes the inter-atomic distance observed in the corresponding experimental structure. In the early stage of training, these two geometric loss terms are disabled to prevent excessive structural regularization from hindering the learning of global protein–NA topology.

Training and test data

Training set and validation set. CryoAtom2 expands the training dataset of CryoAtom, which originally contained 5,731 protein map-model pairs, by incorporating NA–containing samples. Specifically, we selected cryo-EM density maps from the Electron Microscopy Data Bank (EMDB) [25] that were released before July 2024 and had resolutions better than 4.0 Å, provided that the corresponding atomic models were available in the Protein Data Bank (PDB) [1]. These map–model pairs were then integrated into the training dataset, resulting in a total of 8,815 map–model pairs. Among them, 10% were randomly held out for validation, while the remaining 90% were used for training.

Non-redundant test set. To construct a strictly non-redundant test set, we collected 1,100 newly released NA–containing map–model pairs from the EMDB and PDB that became available after the release of the training data (between July 2024 and September 2025), with resolutions ranging from 2-5 Å. We then defined redundancy in the most stringent manner: if there existed at least one pair of NA chains between a test density map and any training (or validation) density map whose sequence identity

exceeded a predefined threshold (80%) [37], the corresponding map–model pair was removed from the test set. After this procedure, 246 map–model pairs remained. Finally, density maps in which the NA components had a deposited Q-score < 0.3 (corresponding to resolutions worse than 5 Å) were excluded. In addition, to ensure a fair and consistent comparison, we removed any density maps for which any of the control methods (EM2NA, ModelAngelo, CryoREAD, and Phenix) failed to run successfully or produced empty structural models. This resulted in a final benchmark test set consisting of 175 map–model pairs dedicated to NA evaluation. It contains 151 protein–NA complexes and 24 pure NA complexes. For the 151 protein–NA complexes, we applied the same procedure as for NAs, performing redundancy reduction on the protein sequences at 30% sequence identity. This resulted in 50 non-redundant protein–NA complexes for additional testing. To evaluate pure protein complexes, we used a set of 54 non-redundant protein test cases from the CryoAtom paper. Notably, because CryoAtom2 incorporated additional protein–NA complexes during training set expansion, we applied the same redundancy removal procedure with a more stringent sequence identity threshold of 30%—compared to that used in the original CryoAtom paper—resulting in 20 non-redundant protein test maps for additional evaluation, thereby ruling out any performance bias arising from training–test set redundancy.

Evaluation metrics

We adopted the evaluation metrics defined in the CryoAtom paper [9] to assess model-building accuracy, including completeness, sequence accuracy, backbone recall, backbone precision, and backbone RMSD.

- Completeness refers to the fraction of residues in the deposited structure that are successfully reconstructed in the predicted model with the correct residue identity, where the corresponding backbone atom lies within 3 Å of the deposited structure (Cα atoms for proteins and P atoms for NAs).
- Sequence accuracy is defined as the proportion of predicted residues that are assigned the correct residue identity.
- Backbone recall/precision are defined in the same way as in CryoAtom: for proteins they are based on Cα atoms, and for NAs they are based on P atoms.
- Backbone RMSD is defined as the root-mean-square deviation (RMSD) between the backbone atoms of matched residue pairs in the predicted and deposited structures. Backbone atoms are defined as (N, Cα, C, O) for amino acids and (P, O3′, C1′, C4′) for nucleotides.

In addition to the above evaluation metrics that rely on deposited reference structures, we further assessed model quality using the Phenix program with two complementary

categories of criteria: geometric quality metrics and model–map correlation coefficients (CCs). The geometric quality metrics include the MolProbity score, RMS(bonds), and RMS(angles), while the model–map agreement is quantified using multiple correlation coefficients computed under different spatial constraints.

- **MolProbity score** evaluates the geometric and stereochemical quality of the atomic model, including covalent geometry, steric clashes, and side-chain conformations.
- **RMS(bonds)** (root-mean-square deviation of bond lengths) measures the deviation of all covalent bond lengths in the model from their ideal target values defined by standard chemical restraints. Lower RMS(bonds) values indicate better adherence to chemically reasonable bond lengths and reflect the internal geometric consistency of the model.
- **RMS(angles)** (root-mean-square deviation of bond angles) quantifies the deviation of bond angles from their ideal values specified by stereochemical restraints. Smaller RMS(angles) values suggest improved angular geometry and reduced local strain within the molecular structure.
- $\mathbf{CC_{mask}}$ represents the correlation between the model-derived density and the experimental map calculated within a mask generated from the atomic model.
- $\mathbf{CC_{volume}}$ denotes the correlation computed within a molecular envelope defined by the model, reflecting the overall agreement between the modeled structure and the density inside the molecular region.
- $\mathbf{CC_{peaks}}$ evaluates the correlation at high-density peak regions of the map, emphasizing the agreement at well-resolved and information-rich density features.
- $\mathbf{CC_{box}}$ measures the correlation over the entire map box, providing a global but less localized assessment of model–map agreement.
- $\mathbf{CC_{mc}}$ quantifies the correlation between the density map and the protein main-chain atoms.
- $\mathbf{CC_{sc}}$ quantifies the correlation between the density map and the protein side-chain atoms.

Data availability

All training and benchmark datasets are freely available from public sources, including cryo-EM maps from EMDB (https://www.ebi.ac.uk/emdb/) and atomic structure from PDB (https://www.rcsb.org/). The benchmark dataset and the list of IDs for the training dataset are available at https://yanglab.qd.sdu.edu.cn/CryoAtom/benchmark/.

Code availability

The source code is publicly available at https://github.com/YangLab-SDU/CryoAtom.

Author contributions

J.Y. conceived and supervised the project; B.S. designed and implemented the experiments; Z.P. co-supervised the study; L.Z. and W.L. conducted the applications; W.W. contributed to the geometric loss; J.L. and K.Z. participated to results analysis. All authors reviewed and approved the final version of the paper.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC T2225007, T25B2009, 32430063, 62501364, 62503282), and Fundamental Research Funds for the Central Universities.

References

1. Burley, Stephen K. *et al.* Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research* **53**, D564-D574 (2025).
2. Kretsch, R.C. *et al.* Assessment of Nucleic Acid Structure Prediction in CASP16. *Proteins: Structure, Function, and Bioinformatics* **94**, 192-217 (2026).
3. Wang, W. *et al.* trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nature Communications* **14**, 7266 (2023).
4. Doerr, A. Predicting RNA structures. *Nature Methods* **22**, 2495-2495 (2025).
5. Wang, W., Su, B., Peng, Z. & Yang, J. Integrated experimental and AI innovations for RNA structure determination. *Nature Biotechnology* (2026).
6. Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024).
7. Chen, S. *et al.* Protein complex structure modeling by cross-modal alignment between cryo-EM maps and protein sequences. *Nature Communications* **15**, 8808 (2024).
8. Giri, N. & Cheng, J. De novo atomic protein structure modeling for cryoEM density maps using 3D transformer and HMM. *Nature Communications* **15**, 5511 (2024).
9. Su, B., Huang, K., Peng, Z., Amunts, A. & Yang, J. CryoAtom improves model building for cryo-EM. *Nature Structural & Molecular Biology* (2025).
10. Li, T., Chen, J., Li, H., Cao, H. & Huang, S.-Y. EMProt improves structure determination from cryo-EM maps. *Nature Structural & Molecular Biology* (2025).
11. Wang, X., Zhu, H., Terashi, G., Taluja, M. & Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nature Methods* **21**, 2307-2317 (2024).
12. Wang, X., Terashi, G. & Kihara, D. CryoREAD: de novo structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nature Methods* **20**, 1739-1747 (2023).
13. Li, T., Cao, H., He, J. & Huang, S.-Y. Automated detection and de novo structure modeling of nucleic acids from cryo-EM maps. *Nature Communications* **15**, 9367 (2024).
14. Li, T. *et al.* All-atom RNA structure determination from cryo-EM maps. *Nature Biotechnology* **43**, 97-105 (2025).
15. Wang, W., Yu, K., Hugonot, J., Fua, P.V. & Salzmann, M. Recurrent U-Net for Resource-Constrained Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2142-2151 (2019).
16. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
17. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023).
18. Chen, J. *et al.* Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. *arXiv preprint arXiv:2204.00300* (2022).
19. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).
20. 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. *Nature*

Methods **15**, 905-908 (2018).

21. Pintilie, G. *et al.* Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nature Methods* **17**, 328-334 (2020).
22. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* **35**, W375-W383 (2007).
23. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D* **75**, 861-877 (2019).
24. Afonine, P.V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D* **74**, 814-840 (2018).
25. The ww, P.D.B.C. EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Research* **52**, D456-D465 (2024).
26. Zheng, W. *et al.* Visualizing the translation landscape in human cells at high resolution. *Nature Communications* **16**, 10757 (2025).
27. Zhu, L. *et al.* A dual-factor complex governs archaeal ribosome hibernation by sensing energy status. *bioRxiv*, 2026.2001.2019.700304 (2026).
28. Wang, L. *et al.* Cryo-EM reveals mechanisms of natural RNA multivalency. *Science* **388**, 545-550 (2025).
29. Kretsch, R.C. *et al.* Naturally ornate RNA-only complexes revealed by cryo-EM. *Nature* **643**, 1135-1142 (2025).
30. Ling, X. *et al.* Cryo-EM structure of a natural RNA nanocage. *Nature* **644**, 1107-1115 (2025).
31. Zhang, S. *et al.* Structural insights into higher-order natural RNA-only multimers. *Nature Structural & Molecular Biology* **32**, 2012-2021 (2025).
32. Rives, A. *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences* **118**, e2016239118 (2021).
33. Lin, T.Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal Loss for Dense Object Detection. *IEEE transactions on pattern analysis and machine intelligence* **42**, 318-327 (2020).
34. Schwerdtfeger, P. & Wales, D.J. 100 Years of the Lennard-Jones Potential. *Journal of Chemical Theory and Computation* **20**, 3379-3405 (2024).
35. Eddy, S.R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **7**, e1002195 (2011).
36. Wang, W., Peng, Z. & Yang, J. Predicting RNA 3D structure and conformers using a pre-trained secondary structure model and structure-aware attention. *bioRxiv*, 2025.2004.2009.647915 (2025).
37. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

Tables

Box 1. Key concepts of CryoAtom2 inspired by ModelAngelo.

Method	ModelAngelo	CryoAtom2
Architecture	Cryo-EM module, sequence module and structure module with spatial IPA	Cryo-Former (integrating cryo-EM and sequence modules via an Evoformer-like architecture) and structure module with local attention and 3D-RoPE
Multimodal encoding	Cryo-EM density and sequence	Cryo-EM density, sequence, and inverse-folding-inspired 3D structural features
Representations to update	Node representation	Node representation and edge representation
Attention mechanism	Local, cross attention	Local, gated cross attention and gated self-attention
Positional encoding	Sinusoidal	3D-RoPE
Pre-trained language model	ESM-1b	ESM-2 and RNA-FM
Recycling mechanism	Recycling backbone frame	Structure Encoder
Sequence alignment	HMM-based	Score-corrected HMM alignment

Figures

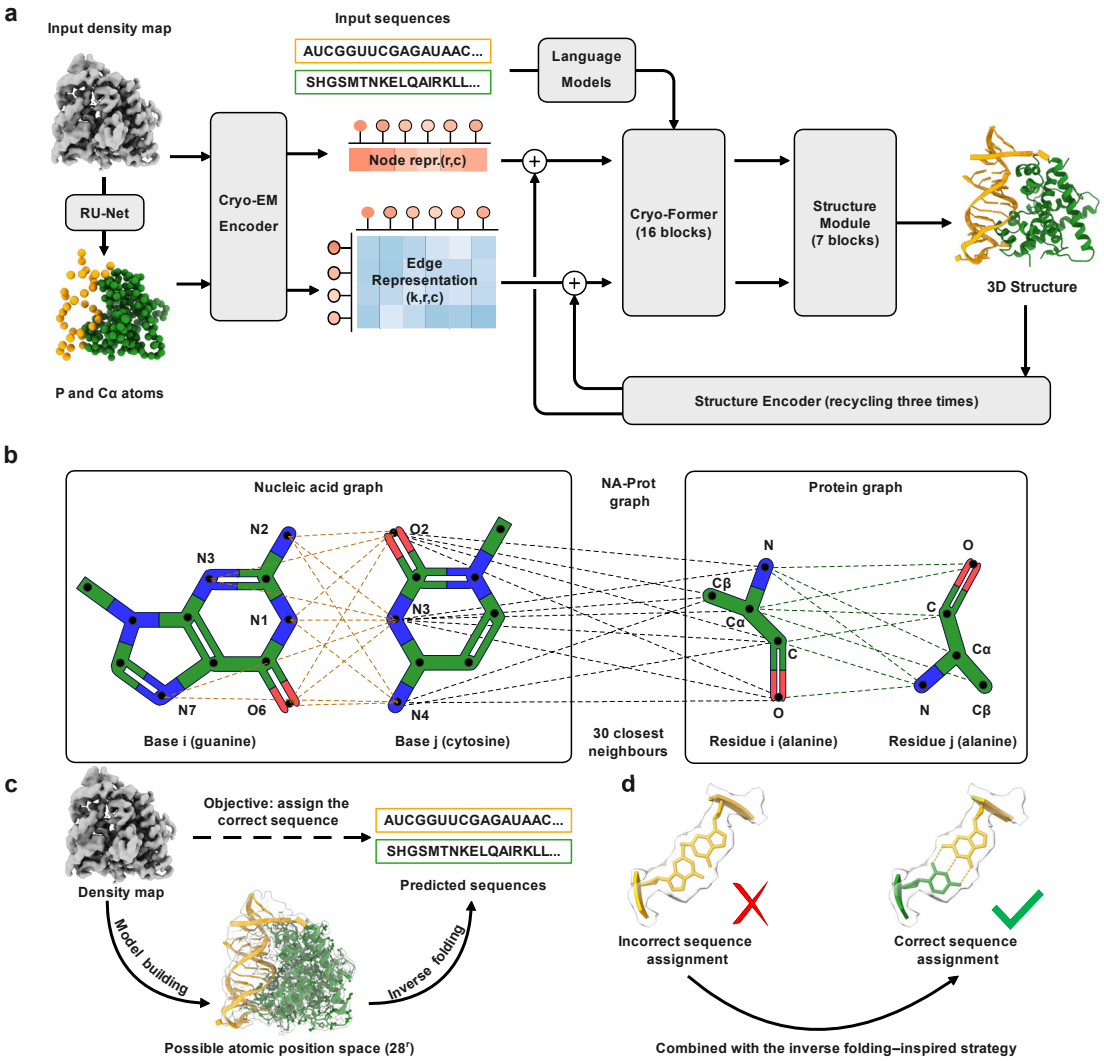

Fig. 1. Overall architecture of CryoAtom2. (a) CryoAtom2 is an extension of CryoAtom that enables cryo-EM model building of nucleic acids (NAs), proteins and protein–NA complexes. Its node and edge representations are expanded to support NA tokens as well as NA–NA and NA–protein pair features. r is the total number of predicted atoms (Cα and P), c is the number of channels and k is 30. In terms of network architecture, CryoAtom2 inherits CryoAtom's components for processing density maps (Cryo-EM Encoder) and sequences (Sequence Attention in Cryo-Former), while additionally incorporating a new modality—3D structural information—through a Structure Encoder as an extra network input. (b) Schematic illustration of the Structure Encoder encoding pairwise representations of NA–NA, NA–protein, and protein–protein interactions. When the amino acid and nucleotide identities are unknown, proteins are uniformly labeled as alanine, and NAs are labeled as either guanine and cytosine, representing purines and pyrimidines, respectively. The Structure Encoder emphasizes capturing biomolecular interactions by focusing on the protein backbone

atoms and the NA base atoms. (c) Inferring sequences from density maps is performed in two stages. In the first stage, a Transformer explores the possible atomic coordinate space under experimental density constraints, and CryoAtom2 predicts 28 candidate atomic coordinate sets for each node (20 standard amino acids, 4 RNA nucleotides, and 4 DNA nucleotides). In the second stage, these candidates are filtered by integrating an inverse folding–inspired strategy, allowing the network to infer the most plausible atomic coordinates with the correct sequence. (d) A supplementary illustration of the inverse-folding idea in (c). On the left, although the assigned sequence fits the density map well, structural conflict exists. On the right, after incorporating the inverse folding–inspired strategy, the new sequence inferred by CryoAtom2 resolves the conflict.

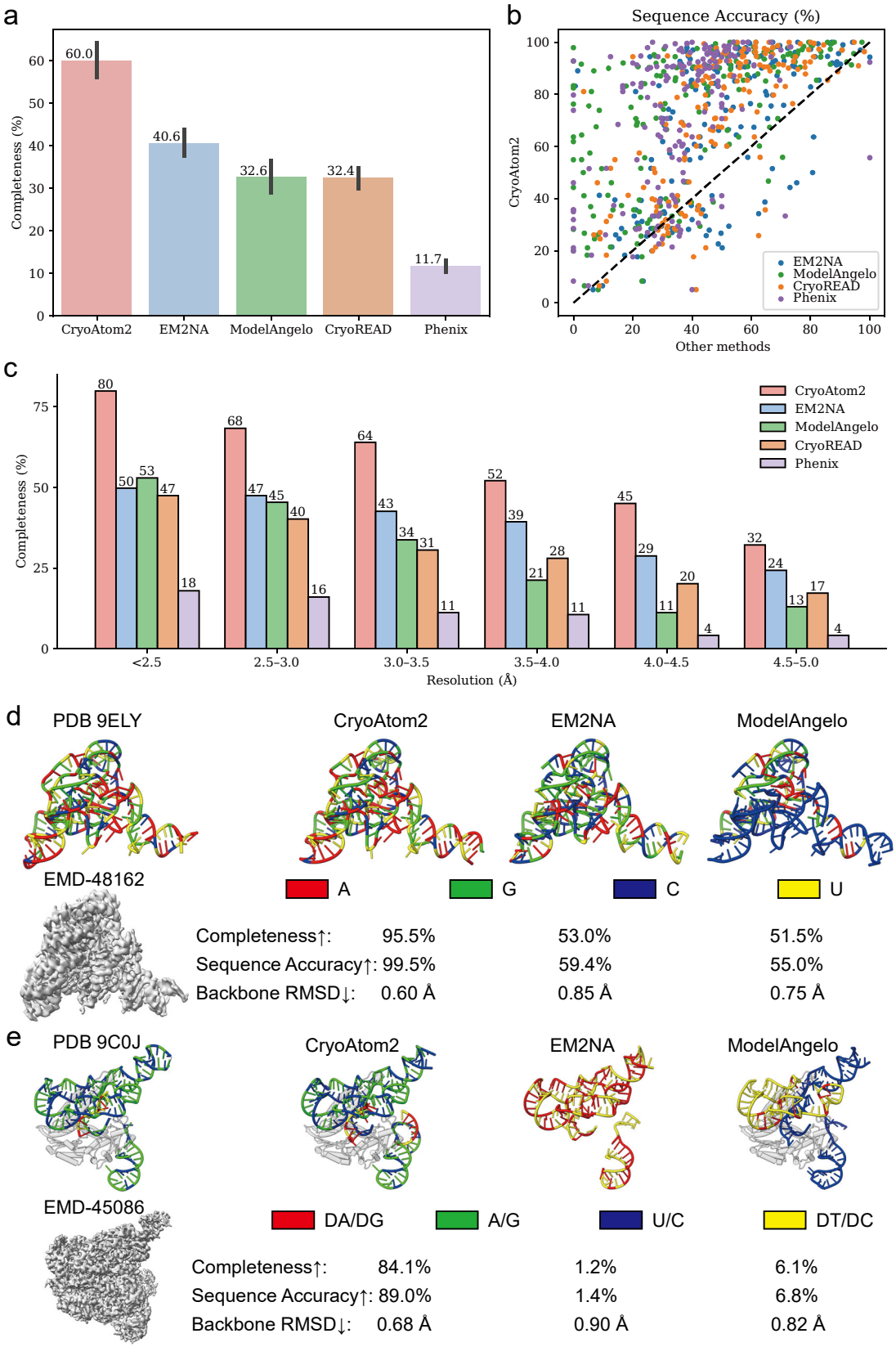

Fig. 2. Nucleic acid model accuracy against the deposited PDB structures. The comparisons among CryoAtom2, EM2NA, ModelAngelo, CryoREAD, and Phenix are based on the independent test set of 175 cryo-EM maps containing NAs. (a) Bar plots

comparing completeness for n = 175 NAs. Data are the mean and the 95% confidence interval of the mean, estimated from 1,000 bootstrap samples. (b) Head-to-head comparison based on the sequence accuracy. (c) Model completeness for various automated model-building software for different resolution ranges in the maps. (d) The gray surface represents the density map EMD-48162 (reported resolution: 2.9 Å). This example shows a comparison of CryoAtom2, EM2NA, and ModelAngelo, with RNA types (A/G/C/U) labeled in different colors (red/green/blue/yellow). (e) This example demonstrates the ability of CryoAtom2, EM2NA, and ModelAngelo to recognize RNA and DNA in a complex system. In this visualization, purines in DNA are colored red, pyrimidines in DNA are colored yellow, purines in RNA are colored green, and pyrimidines in RNA are colored blue. The gray surface represents the density map EMD-45086 (reported resolution: 3.17 Å). The protein component is shown in transparent gray cartoon.

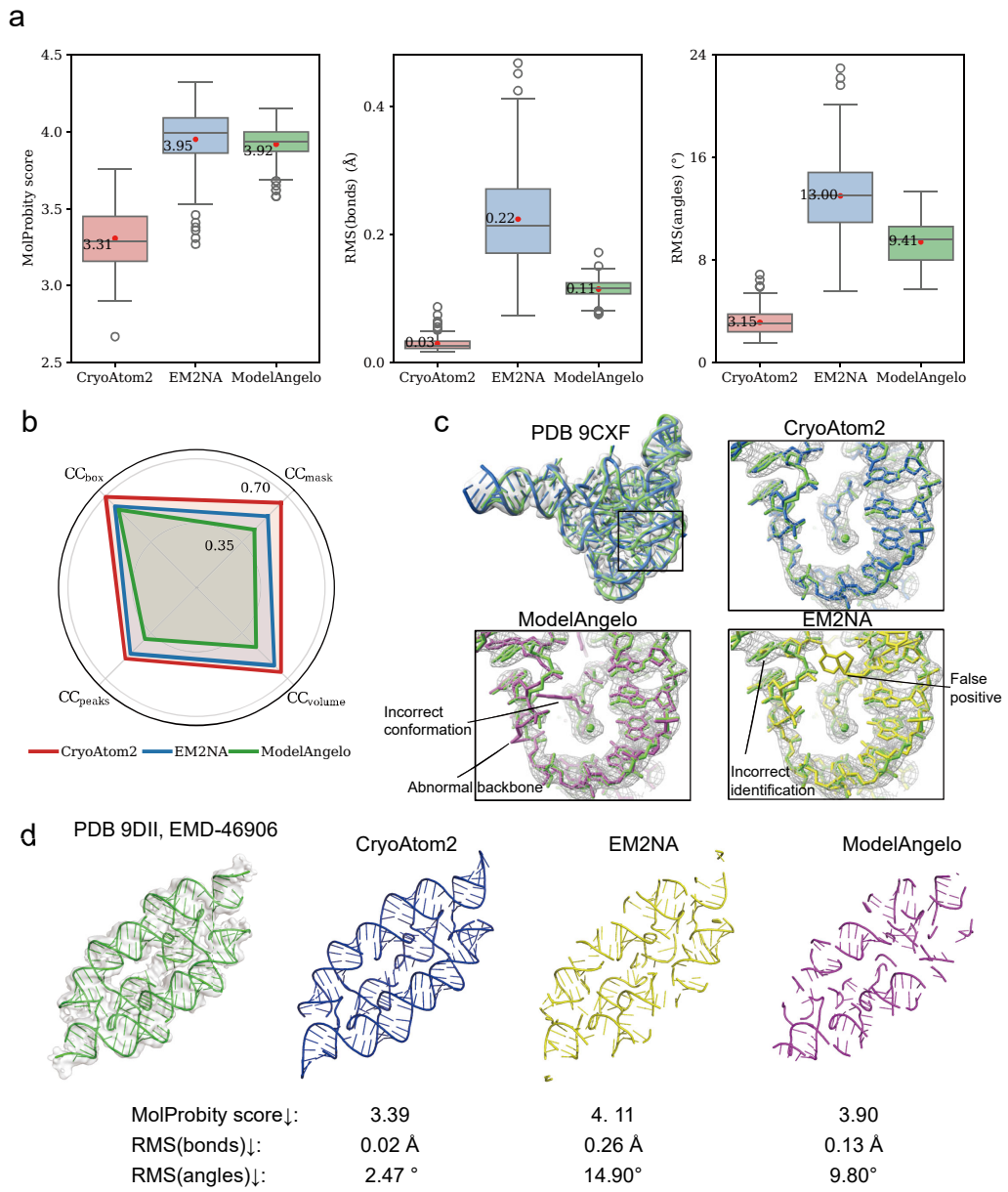

Fig. 3. Nucleic acid model quality validation. The evaluation is based on the 175 NA-containing test maps. (a) Box plots showing the distributions of the MolProbity score, RMS(bonds), and RMS(angles), computed exclusively for NAs (n = 175), with red dots denoting the mean values. The center, lower and upper lines in each box indicate the median, the first quartile and the third quartile, respectively. The whiskers extend to the most extreme data points that are within 1.5 times the interquartile range from the first and third quartiles. Data points beyond this range are considered outliers. (b) Radar plot showing the model–map correlation coefficients (CCs) computed on 24 NA-only complexes among the 175 test density maps, using models generated by CryoAtom2, EM2NA, and ModelAngelo. (c) The gray transparent surface represents the density map EMD-45988 (reported resolution: 3.02 Å). The green model corresponds to the deposited structure (PDB 9CXF). The blue model was built by CryoAtom2, the yellow

model was built by EM2NA, and the magenta model was built by ModelAngelo. (d) The gray transparent surface represents the density map EMD-46906 (reported resolution: 4.00 Å). The green model corresponds to the deposited structure (PDB 9DII). The blue model was built by CryoAtom2, the yellow model was built by EM2NA, and the magenta model was built by ModelAngelo.

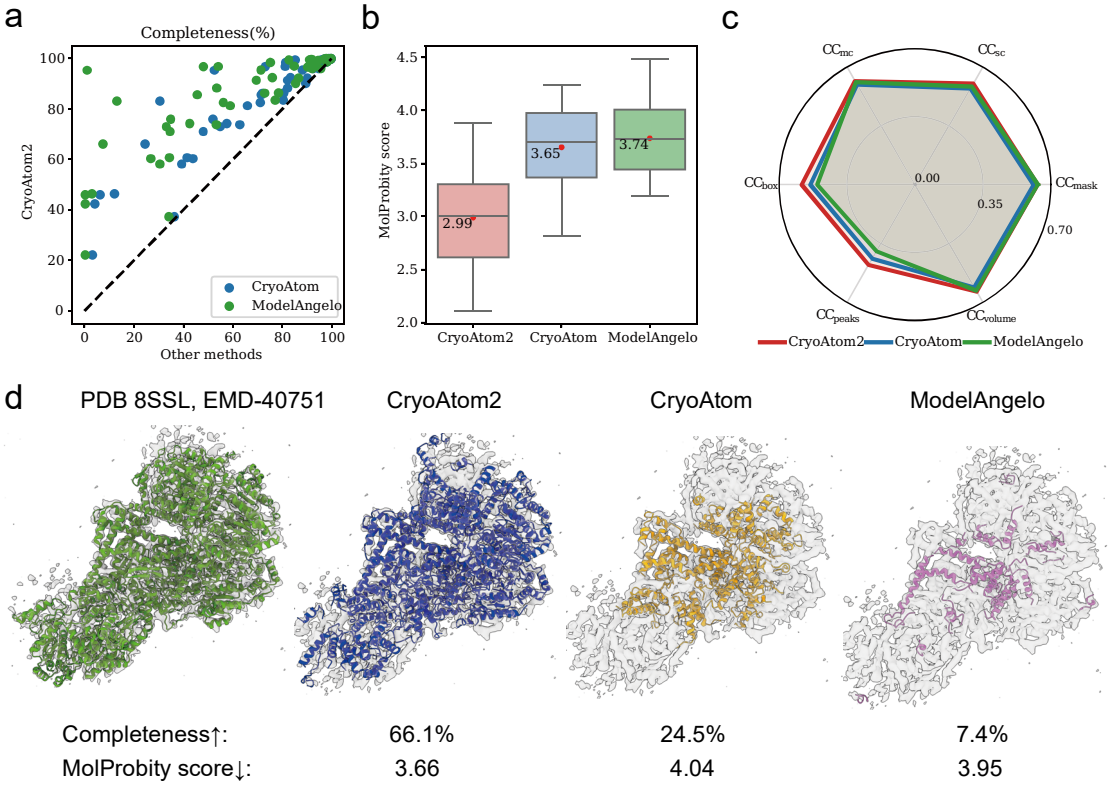

Fig. 4. Comparison among CryoAtom2, CryoAtom, and ModelAngelo on the 54-protein test maps. (a) Head-to-head comparison based on the completeness. (b) Box plots of the MolProbity score (n=54), in which the red dots represent the mean values. The center, lower and upper lines in each box indicate the median, the first quartile and the third quartile, respectively. The whiskers extend to the most extreme data points that are within 1.5 times the interquartile range from the first and third quartiles. Data points beyond this range are considered outliers. (c) Radar plot showing the model–map correlation coefficients (CCs) computed on the 54-protein test set, using models generated by CryoAtom2, CryoAtom, and ModelAngelo. (d) The gray transparent surface represents the density map EMD-40751 (reported resolution: 4.60 Å). The green model corresponds to the deposited structure (PDB 8SSL). The blue model was built by CryoAtom2, the orange model was built by CryoAtom, and the magenta model was built by ModelAngelo.

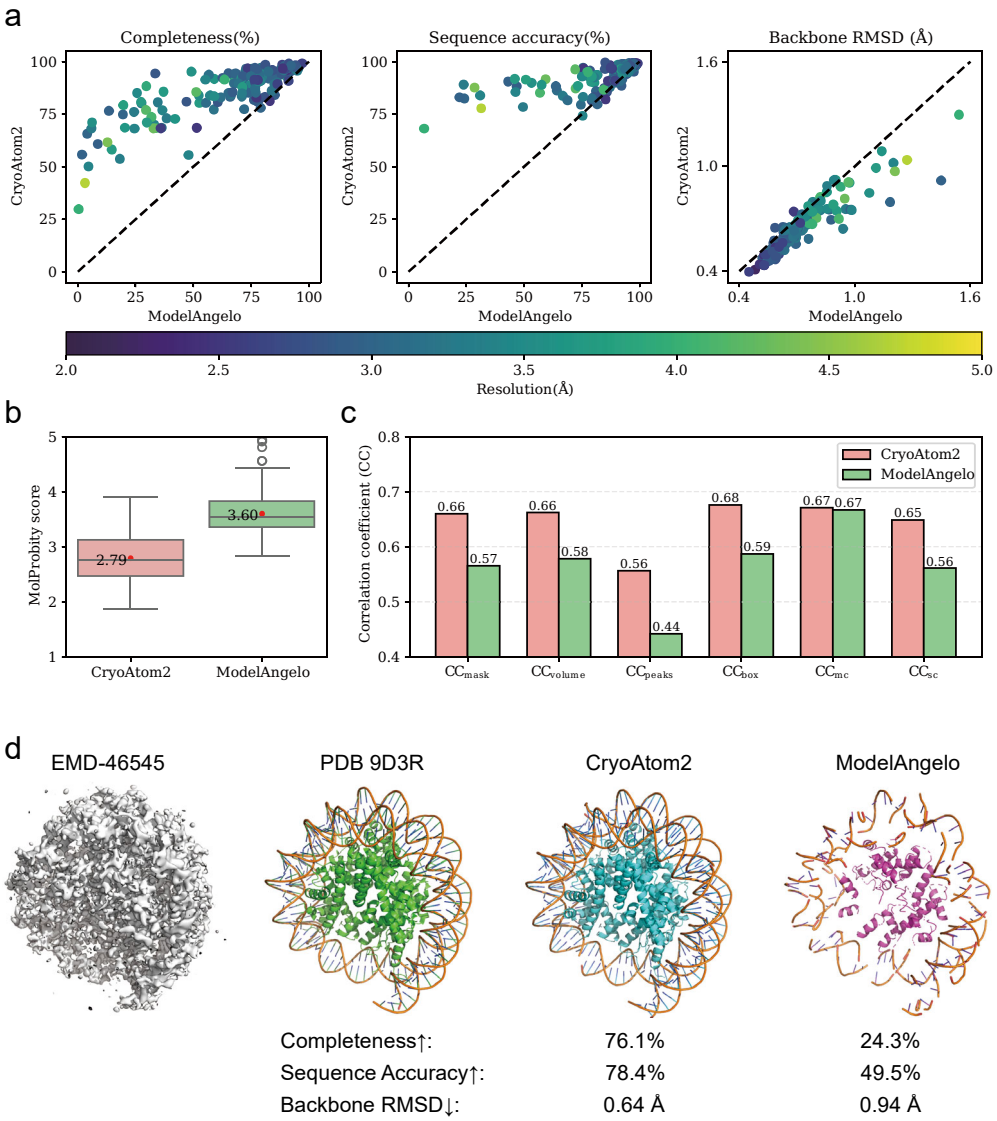

Fig. 5. Comparison among CryoAtom2 and ModelAngelo on the 151 protein–NA complex test set with resolutions of 2–5 Å. (a) Head-to-head comparison of completeness, sequence accuracy, and backbone RMSD. (b) Box plots of the MolProbity score (n = 151), in which the red dots represent the mean values. The center, lower and upper lines in each box indicate the median, the first quartile and the third quartile, respectively. The whiskers extend to the most extreme data points that are within 1.5 times the interquartile range from the first and third quartiles. Data points beyond this range are considered outliers. (c) Bar plots of the model–map correlation coefficients (CCs) using models generated by CryoAtom2 and ModelAngelo. (d) The gray surface represents the density map EMD-46545 (reported resolution: 3.3 Å). The green model corresponds to the deposited structure (PDB 9D3R). The cyan model was built by CryoAtom2, and the magenta model was built by ModelAngelo. The NA backbone is highlighted in orange.

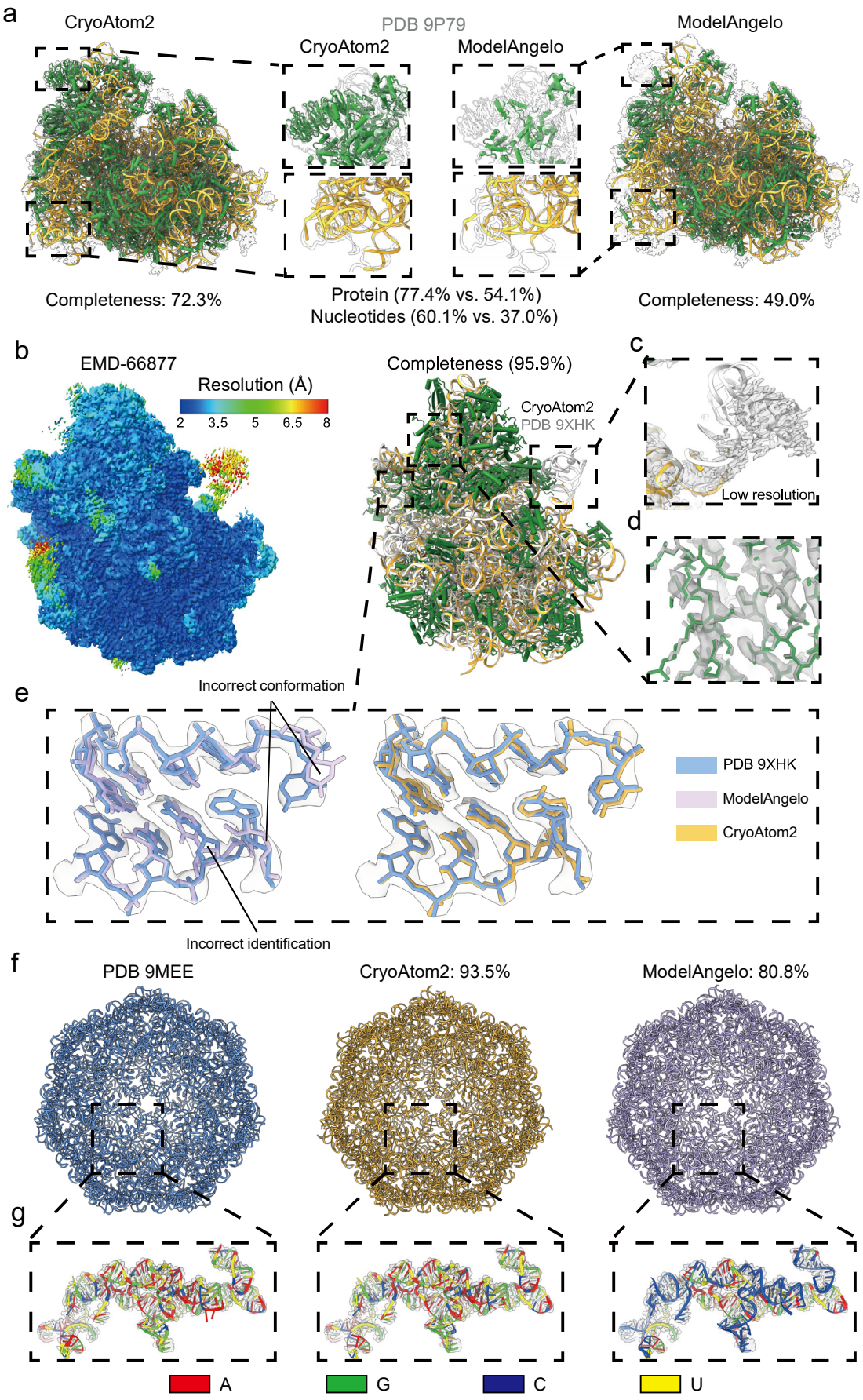

Fig. 6. Applications of CryoAtom2 to three novel maps and comparison with

ModelAngelo. (a) An example of a large 80S ribosome structure (PDB 9P79; EMD-71335; reported resolution: 3.1 Å). The protein and NA structures built by CryoAtom2 (or ModelAngelo) are shown in green and orange, respectively. The deposited structure is outlined in black. (b) The left panel shows the local resolution map of EMD-66877 for an archaeal 70S ribosome, while the right panel presents a structural superposition of the model built by CryoAtom2 with the reference structure (PDB 9XHK). Proteins and NAs are displayed in cartoon representation, with the CryoAtom2 predictions colored green and orange, respectively, and the reference model shown in gray. (c-d) Magnified views of low-resolution regions in the ribosome, with local resolutions of approximately 8 Å (c) and 5 Å (d). NAs are depicted as cartoons and proteins as sticks. (e) Comparison of ModelAngelo and CryoAtom2 predictions superposed on the reference structure (PDB 9XHK) for a local region. Representative nucleotides are shown as sticks with their corresponding densities. Incorrect base identification and conformational errors are highlighted. Color code: blue (PDB 9XHK), pink (ModelAngelo), and orange (CryoAtom2). (f) Comparison of CryoAtom2 and ModelAngelo in modeling an RNA-only nanocage structure (PDB 9MEE, EMD-48214, reported resolution: 3 Å). The numbers in the figure indicate the modeling completeness achieved by each method. (g) A comparison of sequence accuracy between CryoAtom2 and ModelAngelo in a local region, with RNA types (A/G/C/U) labeled in different colors (red/green/blue/yellow). The gray surface represents the density map.

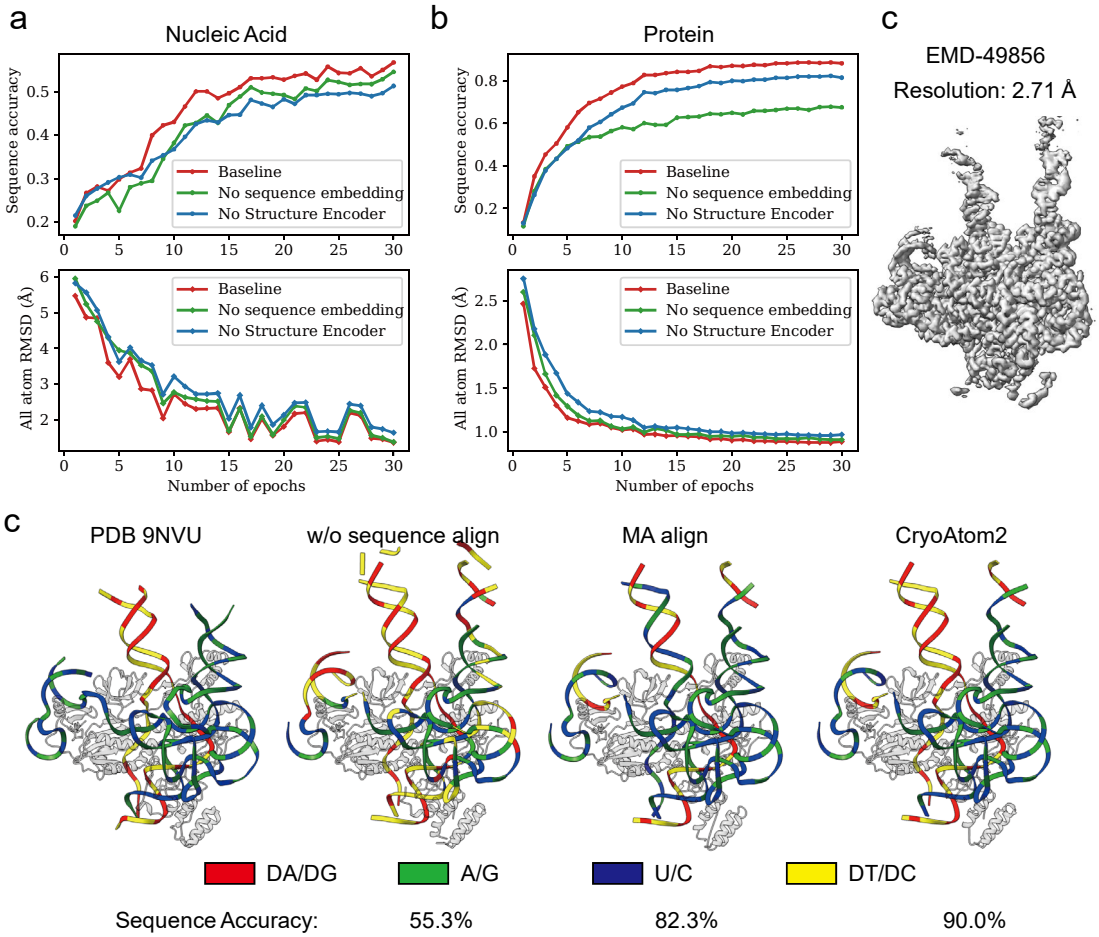

Fig. 7. Ablation studies of CryoAtom2. The experiments were conducted on 880 validation density maps. (a–b) Evolution of sequence accuracy and all-atom RMSD for NAs (a) and proteins (b) under different model variants during training. (c) Ablation study of the sequence alignment strategy. In this visualization, purines/pyrimidines in DNA are colored red/yellow; purines/pyrimidines in RNA are colored green/blue. The gray surface represents the density map EMD-49856 (reported resolution: 2.71 Å). The protein structures in the model are shown in transparent gray.