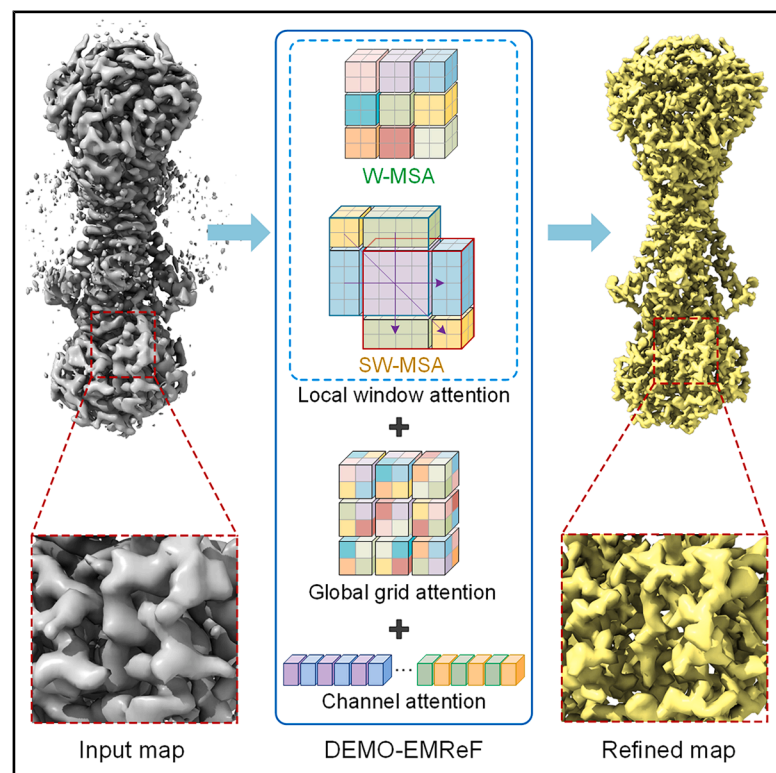


Enhancing interpretability of cryo-EM maps using hybrid attention Transformers

Graphical abstract



Authors

Jie Lin, Ziyang Zhang, Yu Zhang, Feng Cheng, Chengyuan Wang, Guijun Zhang, Xiaogen Zhou

Correspondence

cywang@sim.ac.cn (C.W.),
zgj@zjut.edu.cn (G.Z.),
zxg@zjut.edu.cn (X.Z.)

In brief

Lin et al. present DEMO-EMReF, a hybrid attention Transformer that captures local features and long-range spatial dependencies to refine cryo-EM density maps. DEMO-EMReF improves map quality and interpretability, enabling more accurate automated atomic model building.

Highlights

- DEMO-EMReF combines local and grid-based attention for cryo-EM map refinement
- Hybrid attention captures fine local detail and long-range spatial dependencies
- DEMO-EMReF improves map resolution, isotropy, and atomic resolvability
- Refined maps enable more accurate automated atomic model building



Article

Enhancing interpretability of cryo-EM maps using hybrid attention Transformers

Jie Lin,^{1,5} Ziyi Zhang,^{1,5} Yu Zhang,² Feng Cheng,³ Chengyuan Wang,^{2,*} Guijun Zhang,^{1,4,*} and Xiaogen Zhou^{1,4,5,6,*}

¹College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

²Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 200032, China

³Key Laboratory of Bioorganic Synthesis of Zhejiang Province, College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou 310014, China

⁴Zhejiang Key Laboratory of Intelligent Perception and Control for Complex Systems, Hangzhou, China

⁵These authors contributed equally

⁶Lead contact

*Correspondence: cywang@sim.ac.cn (C.W.), zgj@zjut.edu.cn (G.Z.), zxcg@zjut.edu.cn (X.Z.)

<https://doi.org/10.1016/j.xcrp.2026.103596>

SUMMARY

Cryo-electron microscopy (cryo-EM) is a central technique for determining structures of macromolecular complexes, but noise and heterogeneous, anisotropic resolution often limit map interpretability and complicate atomic model construction. Here, we present DEMO-EMReF, a Swin-Transformer-based framework for cryo-EM density map refinement that integrates channel attention with hybrid spatial attention. By combining local-window-based attention with grid-based sparse global attention, DEMO-EMReF captures fine-grained local density features while modeling long-range, cross-region spatial dependencies, enabling robust refinement of heterogeneous and anisotropic map regions. Across benchmark sets of primary maps and half-maps spanning resolutions of 3.0–8.0 Å, DEMO-EMReF consistently improves resolution-related metrics, map-model correlations, and local atomic resolvability compared with multiple state-of-the-art enhancement methods, with particularly robust gains in challenging heterogeneous or anisotropic cases. The refined maps also facilitate more accurate and efficient automated atomic model building. Together, DEMO-EMReF provides a robust approach for enhancing cryo-EM density maps and enabling more reliable downstream studies.

INTRODUCTION

Recent breakthroughs in single-particle cryo-electron microscopy (cryo-EM) have enabled near-atomic resolution reconstructions of large biomolecular complexes, profoundly transforming the field of structural biology.^{1,2} A central objective of cryo-EM is the construction of high-precision atomic models of macromolecular assemblies, a process that critically depends on the quality of the reconstructed three-dimensional density maps.^{3–6} However, several intrinsic and technical limitations, including electron dose restrictions, radiation damage, and the inherent flexibility and conformational heterogeneity of macromolecules often lead to reconstructed cryo-EM density maps that suffer from low contrast, high noise, and blurred structural regions.⁷ These deficiencies significantly hinder map interpretability and compromise the accuracy of atomic model building.

To improve the interpretability of cryo-EM density maps, a variety of post-processing methods have been developed. Conventional B-factor-based sharpening approaches, implemented in tools such as phenix.auto_sharpen,⁸ RELION,⁹ and CryoSPARC,¹⁰ enhance high-frequency signals globally but do not account for spatial variations in signal-to-noise ratio, often

leading to over-sharpening or under-enhancement in different regions. Localized refinement methods, including LocScale,¹¹ LocalDeblur,¹² LocSpiral,¹³ and OPUS-SSRI,¹⁴ address this limitation by introducing region-specific processing strategies but typically rely on prior information such as atomic models, masks, or local resolution estimates, which limits their general applicability.

More recently, deep-learning-based approaches have been introduced to overcome these constraints. DeepEMhancer¹⁵ employs a U-Net architecture trained on LocScale-enhanced maps to mimic sharpening effects, whereas EM-GAN¹⁶ adopts a 3D generative adversarial network trained on simulated density maps derived from atomic models to reduce noise and preserve structural features. EMReady¹⁷ further introduced a Swin-Conv-UNet (SCUNet)¹⁸ architecture that integrates residual convolutional layers with Swin Transformer modules and achieves robust improvements in map interpretability, while CryoTEN¹⁹ employs a Transformer-based UNETR++ architecture²⁰ for faster inference. Nevertheless, effectively capturing both local structural features and long-range dependencies remains challenging in large and heterogeneous maps. EMReady2²¹ further incorporates a Mamba-based dual-branch U-Net with local-resolution-aware



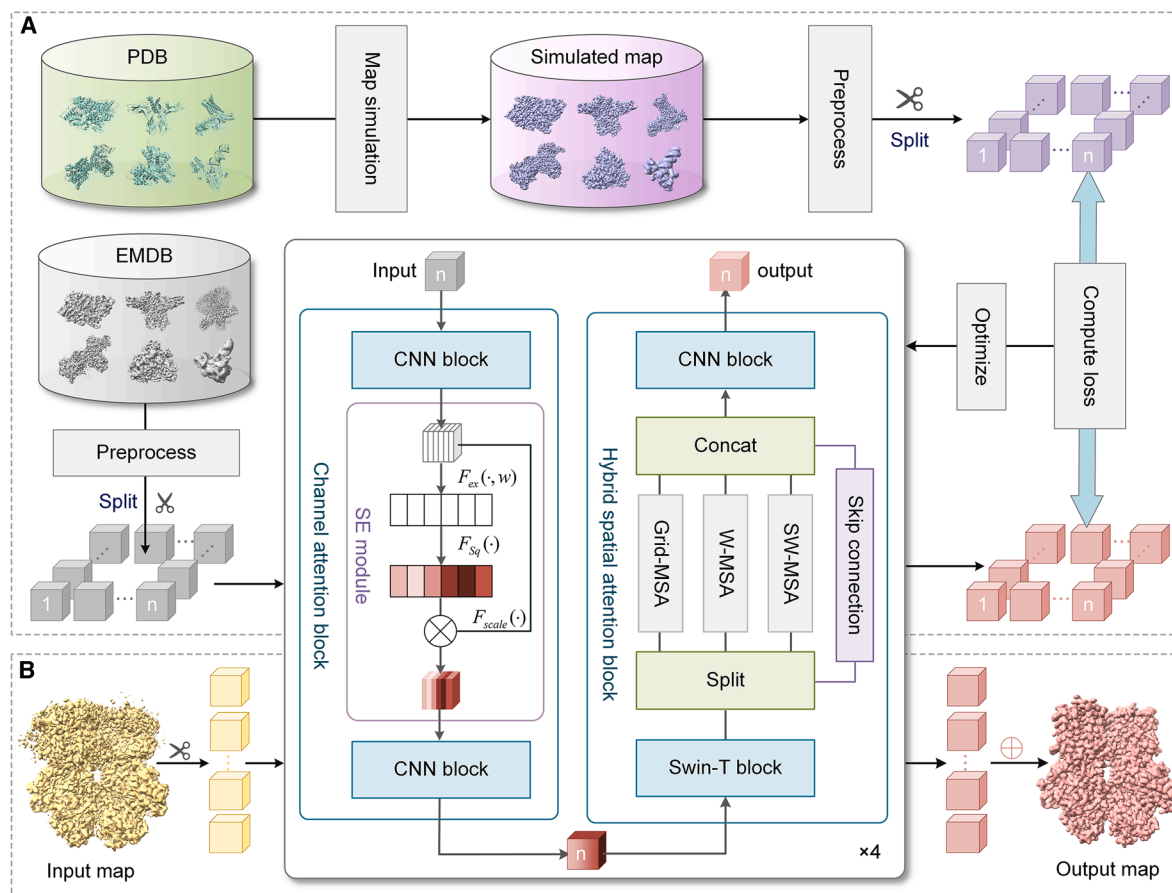


Figure 1. Flowchart of DEMO-EMReF

(A) Training pipeline. Experimental density maps and their corresponding atomic structures were collected from EMDB and PDB, respectively. Noise-free simulated density maps were generated from atomic structures and paired with deposited maps. Both types of maps were preprocessed and partitioned into matched 3D density chunks for supervised training. Each chunk is individually processed by the DEMO-EMReF network (central panel), which integrates convolutional SE blocks for channel-wise attention with a hybrid spatial attention module that combines W-MSA, SW-MSA, and Grid-MSA. The output chunk is compared with the corresponding simulated chunk to compute the training loss, which is used to update the model parameters via backpropagation.

(B) Inference pipeline. A target map is divided into chunks, which are processed independently by the trained model. The enhanced chunks are reassembled into the final improved map.

learning to jointly model local and global features in maps of heterogeneous quality, whereas CryoFM²² employs a flow-based generative foundation model to learn high-quality density priors for signal restoration and map enhancement.

We present DEMO-EMReF, a 3D hybrid attention Transformer architecture that enables long-range information exchange and global context modeling in cryo-EM maps. DEMO-EMReF integrates a channel attention mechanism for global feature recalibration with a hybrid spatial attention module that combines window-based self-attention and a grid-based sparse attention mechanism to overcome the locality constraints of existing methods. This architecture enables selective enhancement of structurally meaningful regions, effective noise suppression, and full-resolution training without information loss induced by downsampling, while also modeling distant structural similarities commonly observed in symmetric or repetitive macromolecular assemblies. Evaluation on diverse test sets of raw experimental primary maps and unprocessed half-maps demonstrates that

DEMO-EMReF consistently improves map quality over state-of-the-art methods. Furthermore, atomic model building confirms that structures reconstructed from DEMO-EMReF-enhanced maps exhibit higher accuracy and structural completeness than those derived from deposited maps, highlighting the practical utility of the method for high-precision cryo-EM structural analysis workflows.

RESULTS

Overview of DEMO-EMReF

Figure 1 illustrates the overall architecture and workflow of DEMO-EMReF. The central panel shows a hybrid attention Transformer network specifically designed for cryo-EM density map enhancement. The model consists of four stacked Transformer layers, each containing two key modules. The first is a convolutional Squeeze-and-Excitation²³ (SE)-based channel attention block that adaptively adjusts channel-wise feature

responses to capture global contextual dependencies. The second is a hybrid spatial attention module that integrates local window self-attention including window-based multi-head self-attention (W-MSA) and shifted-window self-attention (SW-MSA) with a grid-based multi-head self-attention (Grid-MSA)²⁴ to facilitate long-range interactions across spatially distant regions. This architectural design enables simultaneous modeling of fine-grained local features and long-range correlations in cryo-EM density maps. For model training (Figure 1A), experimental density maps with reported resolutions ranging from 3.0 to 8.0 Å were collected from the Electron Microscopy DataBank (EMDB),²⁵ and their corresponding atomic models were obtained from the Protein DataBank (PDB).²⁶ All maps were preprocessed by normalizing density values and resampling voxel spacing to 1.0 Å. Each map was then partitioned into overlapping density chunks of $64 \times 64 \times 64$ voxels, from which $48 \times 48 \times 48$ voxel sub-volumes were randomly cropped. The corresponding noise-free simulated density maps generated from the atomic models were processed in the same manner and used as ground-truth targets. The model was optimized using a combined loss function consisting of mean squared error (MSE) and structural similarity index (SSIM), and data augmentation strategies, including random rotations and cropping, were applied to improve the generalization capability of the model. During inference (Figure 1B), the target map was divided into multiple overlapping density chunks that were independently processed by the trained model. The enhanced chunks were subsequently stitched together through weighted averaging in the overlapping regions to reconstruct the final enhanced full map. Further details are provided in the [methods](#) section.

Improvement of primary maps

The performance of DEMO-EMReF was first evaluated on a benchmark dataset comprising 111 deposited experimental density maps (primary maps), with resolutions ranging from 3.0 to 8.0 Å. For each map, the unmasked map-to-model FSC-0.5 resolution was computed using phenix.mtriage.²⁷ Detailed parameter settings and commands for all comparison methods are provided in [Notes S1–S6](#). Figure 2 and Table S1 summarize the results of the deposited maps before and after refinement by different methods. DeepEMhancer successfully processed 101 out of 111 maps, whereas CryoTEN failed to produce valid unmasked FSC-0.5 values for two cases, and one deposited map did not yield a valid unmasked FSC-0.5 value. To ensure a fair comparison, average values were calculated only over test cases where all methods produced valid results. Unless otherwise specified, all reported resolutions correspond to the standard unmasked FSC-0.5 values, and the phenix baseline refers specifically to phenix.auto_sharpen. Overall, DEMO-EMReF significantly improved the average unmasked FSC-0.5 resolution from 5.03 Å for deposited maps to 3.52 Å. As shown in Figures 2A and 2B, DEMO-EMReF enhanced FSC-0.5 resolutions of deposited maps in 97.3% of cases, with 83% of maps exhibiting improvements greater than 0.5 Å. Compared with state-of-the-art methods, DEMO-EMReF (3.52 Å) outperformed EMReady2 (3.66 Å), CryoFM (3.80 Å), EMReady (3.88 Å), CryoTEN (4.45 Å), DeepEMhancer (4.67 Å), and

phenix (5.03 Å). Moreover, compared with the strongest baseline method, EMReady2, DEMO-EMReF achieved superior FSC-0.5 resolutions in 76.6% of cases (Figure S1A), with the difference being statistically significant as assessed by a two-sided Wilcoxon signed-rank test. Detailed FSC-0.5 resolutions for all test cases are provided in Table S2.

We further evaluated the interpretability of the enhanced maps using the Q-score²⁸ metric implemented in the MapQ plugin of UCSF Chimera.^{29,30} The Q-score quantifies the correlation between the density at atomic positions and a reference Gaussian density function and is therefore commonly used to assess local map interpretability at the atomic level. The average Q-score over the entire model has been shown to correlate strongly with map resolution and overall interpretability. The average Q-scores of different methods are summarized in Table S1. As shown in Figures 2C and S1B, the maps processed by DEMO-EMReF exhibited higher Q-scores than those by CryoTEN, DeepEMhancer, phenix, EMReady, and CryoFM in most cases. Specifically, DEMO-EMReF improved the Q-score in 99 out of 111 deposited maps (Figure S1C), yielding an average Q-score of 0.549, which is higher than those of the deposited maps (0.498), CryoTEN (0.501), DeepEMhancer (0.442), phenix (0.486), EMReady (0.537), and CryoFM (0.544). EMReady2 achieved a marginally higher average Q-score of 0.552 than DEMO-EMReF, but the difference was not statistically significant. Detailed Q-score values for each test case are provided in Table S4.

To comprehensively evaluate the performance of DEMO-EMReF, we further analyzed the correlation coefficients (CC) between the model-derived and deposited maps, including CC_box, CC_mask, and CC_peaks metrics (Table S1). As shown in Figures 2D–2F and S1D–S1F, DEMO-EMReF significantly improved all CC metrics for the deposited maps and outperformed most of the compared methods. Specifically, the average CC_box, CC_mask, and CC_peaks of the DEMO-EMReF-enhanced maps were 0.882, 0.822, and 0.781, respectively, corresponding to improvements of 12.2%, 1.2%, and 14.9% relative to the deposited maps (0.786, 0.810, and 0.680). DeepEMhancer and phenix failed to improve the CC values of the deposited maps on average. CryoTEN achieved average CC_box, CC_mask, and CC_peaks values of 0.868, 0.794, and 0.735, respectively, while EMReady maps obtained corresponding values of 0.872, 0.806, and 0.761, all of which were lower than those of DEMO-EMReF. Compared with CryoFM and EMReady2, DEMO-EMReF achieved higher CC_mask and CC_peaks values, while ranking second only to CryoFM in CC_box. Although the average differences were marginal, correlation-based metrics primarily quantify voxel-wise similarity between maps and can therefore be influenced by shared signal and noise components. Nevertheless, DEMO-EMReF enhanced the CC_box, CC_mask, and CC_peaks in 100, 80, and 98 out of 111 deposited maps (Figures S1G–S1I), corresponding to 3.8% fewer, 31.1% more, and 27.3% more cases than CryoFM and 8.7%, 29.0%, and 11.4% more cases than EMReady2. Detailed CC values for each test case are provided in Tables S5–S7.

Figure 3A presents the 3.1 Å cryo-EM density map of the Desulfovibrio thermocuniculi IsrB (DtIsrB) complexed with omega

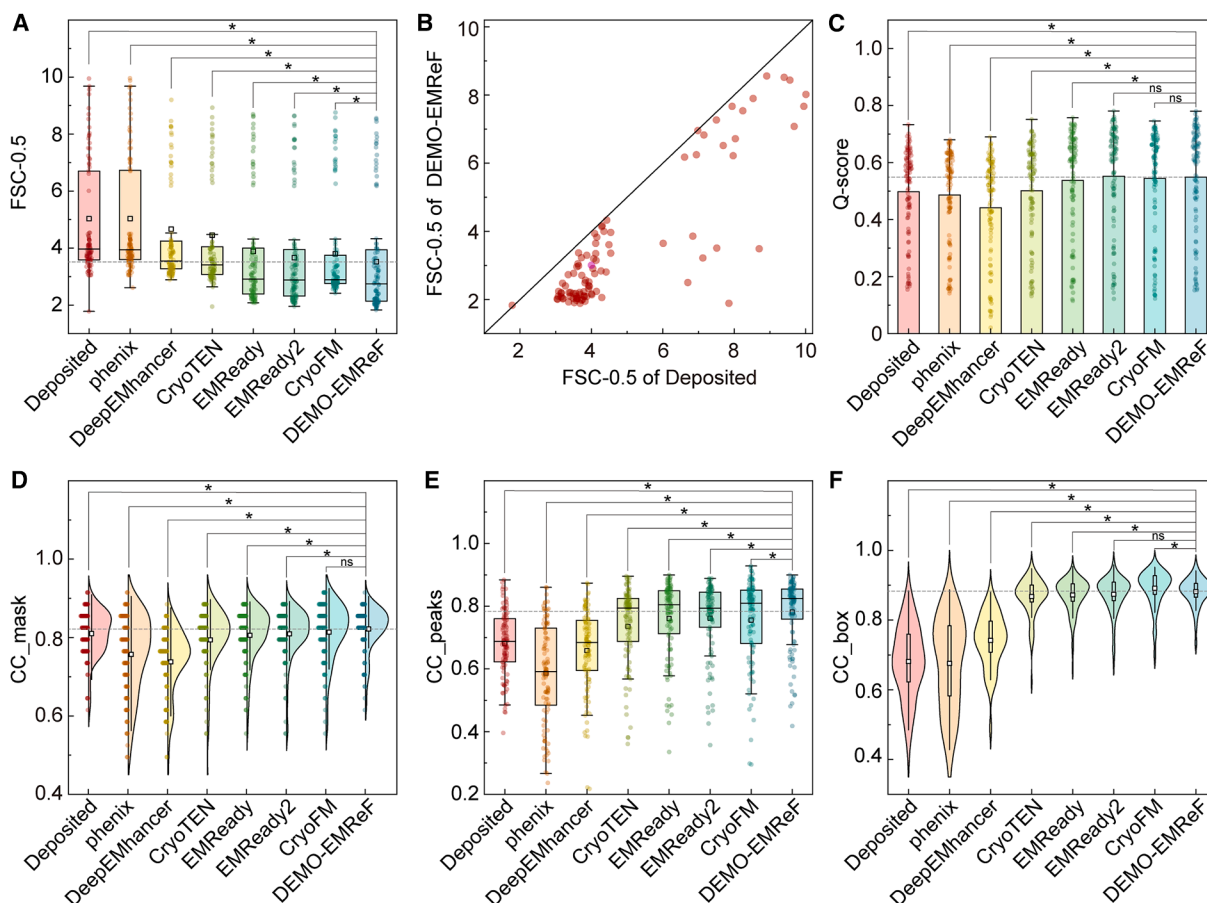


Figure 2. Comprehensive comparison of map refinement methods on the 111 deposited primary maps

(A) Boxplots showing the distribution of unmasked map-to-model FSC-0.5 resolutions for each method. The boxes represent the 25th to 75th percentiles, horizontal line within each box denotes the median, square markers indicate the mean, whiskers extend to the most extreme data points within $1 \times \text{IQR}$ beyond the quartiles, and circles represent individual data points. Asterisks (*) indicate statistically significant differences between DEMO-EMReF and the corresponding method, as determined by a two-sided Wilcoxon signed-rank test ($p < 0.0001$), whereas “ns” indicates no statistically significant difference. The corresponding effect sizes, 95% confidence intervals, and multiple-comparison-adjusted p values are reported in Table S3.

(B) Scatterplot comparing FSC-0.5 resolutions of the deposited maps before refinement and after refinement by DEMO-EMReF.

(C) Bar chart summarizing average Q-scores for different methods, with individual data points shown as circles.

(D) Violin plots of CC_mask for each method, overlaid with individual data points and boxplots; square markers indicate the mean, and error bars denote the standard deviation.

(E) Boxplots showing CC_peaks values for different refinement methods, with the same statistical annotations as in (A).

(F) Violin plots showing the distribution of CC_box values across different methods.

RNA and target DNA³¹ (EMD-27533). Although the deposited map is of good quality (FSC-0.5 = 3.79 Å; Q-score = 0.624), portions of the nucleic acid regions display missing or weak density and substantial background noise. DEMO-EMReF markedly improved the map, yielding an FSC-0.5 of 2.23 Å and a Q-score of 0.687. As highlighted in the enlarged view of Figure 3A, DEMO-EMReF enhanced the continuity of the DNA phosphate-backbone density and substantially reduced background noise. Accordingly, the processed map obtained significantly higher CC_box/CC_peaks values (0.911/0.859/0.855) than the deposited map (0.688/0.670/0.812). The unmasked map-model FSC curves plotted against inverse resolution (Figure 3C) further demonstrate the improvement of the DEMO-EMReF-processed map over the deposited map.

Figure 3B shows a comparison between the deposited map and the DEMO-EMReF-enhanced map of EMD-60471, corresponding to the structure of the zinc-activated channel (ZAC) in the Cys-loop receptor superfamily³² (PDB: 8ZTS), at different contour levels. At the lower contour threshold (left panel), the transmembrane domain of OIZAC is embedded within the surrounding lipid density and is nearly indistinguishable. The DEMO-EMReF enhancement increases the contrast between lipid and protein while suppressing background noise, thereby making the transmembrane domain clearly visible. The right panel shows maps at a higher contour level, highlighting DEMO-EMReF’s ability to enhance high-density features. In the deposited map, one of the four transmembrane helices within the OIZAC subunit appears weakened or even absent,

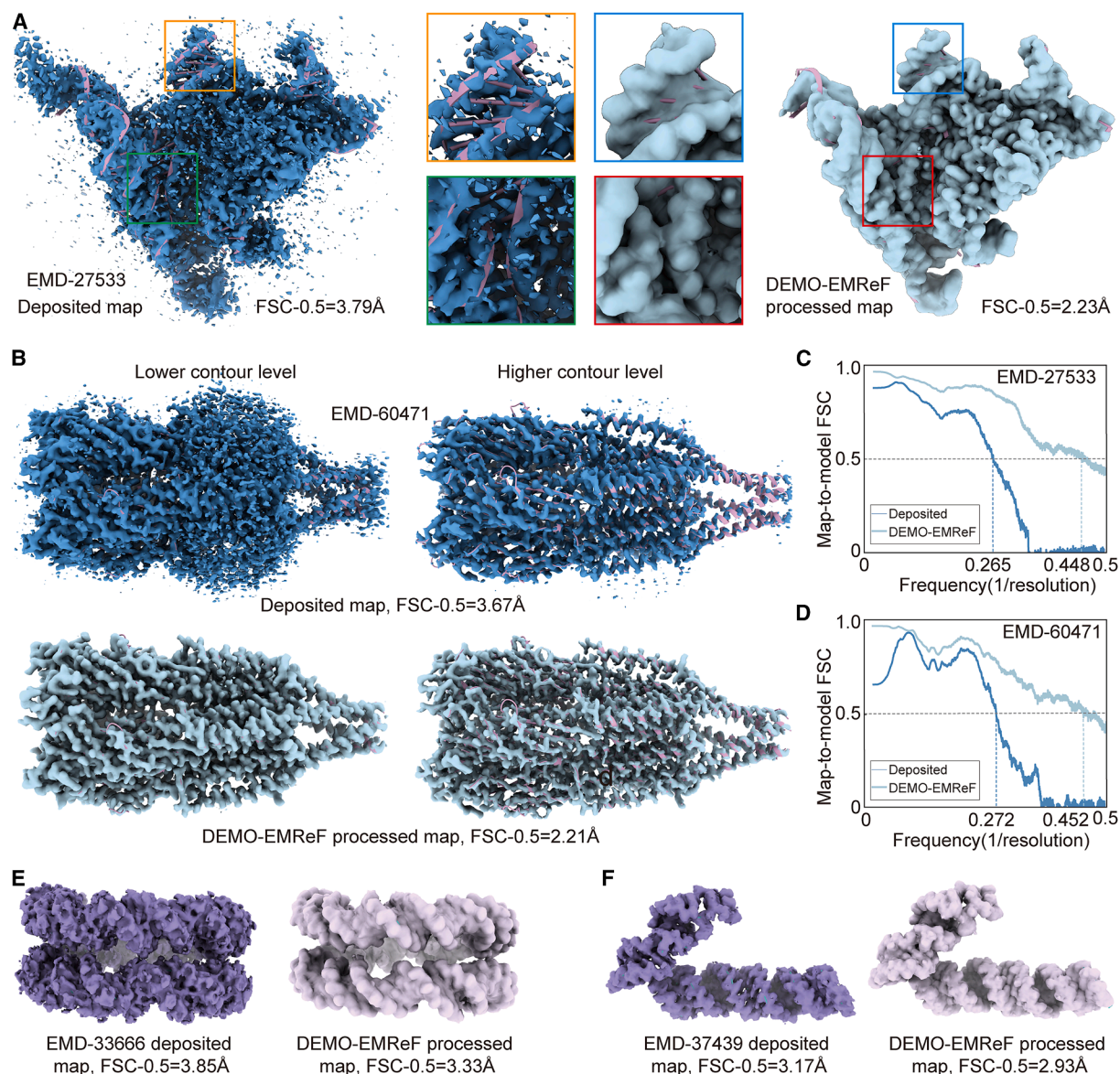


Figure 3. Examples of density maps enhanced by DEMO-EMReF

(A) Cryo-EM density maps of the DtlrB-omega RNA-target DNA complex (EMD-27533, PDB: 8DMB) with a resolution of 3.1 Å before and after DEMO-EMReF enhancement.

(B) Comparison of the deposited map (EMD-60471, PDB: 8ZTS) with a resolution of 3.6 Å and the DEMO-EMReF-enhanced map of the ZAC zinc-activated channel at different contour levels.

(C and D) Unmasked map-to-model FSC curves for EMD-27533 and EMD-60471, respectively.

(E) DNA density in the CENP-A nucleosome-ggKNL2 complex (PDB: 7Y7I, chains I and J) from the deposited map (EMD-33666) with a resolution of 3.42 Å and the corresponding enhanced map.

(F) RNA duplex density of the Cas13h1-crRNA-target RNA complex (PDB: 8WCE, chains G and T) in the deposited map (EMD-37439) with a resolution of 3.09 Å and enhanced maps.

whereas DEMO-EMReF restores these densities, bringing the processed map into close agreement with the PDB model. Quantitatively, DEMO-EMReF improves the unmasked FSC-0.5 resolution from 3.67 to 2.21 Å (Figure 3D) and raises the Q-score from 0.603 to 0.699. In addition, CC_mask, CC_box, and CC_peaks reach 0.861, 0.901, and 0.857, respectively, which

are substantially higher than the corresponding values of 0.791, 0.631, and 0.523 for the deposited map.

Figure 3E focuses on the DNA structure (chains I and J) from the CENP-A nucleosome in complex with ggKNL2³³ (PDB: 7Y7I, EMD-33666). Although the deposited map exhibits recognizable DNA double-helix features, the overall density

boundaries are relatively blurred. DEMO-EMReF substantially reduced background noise, revealing a clearer DNA backbone and more continuous helical features. DEMO-EMReF improved the FSC-0.5 from 3.68 to 2.15 Å and enhanced Q-score from 0.442 to 0.498. [Figure 3F](#) focuses on the RNA duplex region (chains G and T) of Cas13h1 bound to guide and target RNA³⁴ (PDB: 8WCE, EMD-37439). While the deposited map reveals the overall RNA folding, local density discontinuities are apparent, especially near secondary structure boundaries. DEMO-EMReF markedly enhanced the RNA helical features and density continuity in connecting regions, yielding higher contrast. Accordingly, the DEMO-EMReF-processed map achieved an unmasked FSC-0.5 of 2.09 Å and an average Q-score of 0.674, compared to 3.8 Å and 0.622 for the deposited map. These examples indicate that DEMO-EMReF can effectively enhance key structural features of nucleic acid (e.g., DNA base-stacking interactions and RNA secondary structures), facilitating subsequent atomic structure modeling.

[Figure S2](#) presents the refined density map of the highly symmetric human ferritin complex (EMD-6714). This hollow protein nanocage contains repeating subunits and equivalent ligand-binding environments, resulting in highly similar density patterns across symmetric regions. Compared with other methods, both EMReady and DEMO-EMReF improved side-chain density reconstruction, while DEMO-EMReF showed superior preservation of water molecule densities at ligand-binding sites. DEMO-EMReF achieved the highest Q-score of 0.780 and the best FSC-0.5 resolution of 2.02 Å among all compared methods, indicating enhanced local map resolvability and overall reconstruction quality. These results demonstrate that the hybrid attention architecture effectively captures long-range structural correlations in symmetric macromolecular assemblies, enabling the selective enhancement of structurally meaningful regions while suppressing noise.

[Figure S3](#) presents a representative case (EMD-24991) in which DEMO-EMReF underperforms the deposited map according to certain evaluation metrics. Although DEMO-EMReF improves the overall interpretability of the macromolecular density, as reflected by an increase in Q-score from 0.682 to 0.752, noticeable degradation is observed in local ligand-binding regions (green arrows). Specifically, the reconstructed density is excessively smoothed relative to the deposited map, resulting in the partial loss of weak ligand features and a slight decrease in the mask-based correlation coefficient from 0.907 to 0.891. This performance degradation is likely attributable to the current training strategy, which is primarily optimized for the refinement of protein and nucleic acid densities and therefore provides limited supervision for sparse ligand densities with low signal-to-noise ratios.

Evaluations on half-maps

Most post-processing methods require half-maps with nominally independent errors in addition to the primary cryo-EM maps. To comprehensively assess the performance of DEMO-EMReF, we further evaluated it on a test set of 27 pairs of half-maps. The average of the two half-maps (averaged map) was used as the input for DEMO-EMReF and the other methods. [Figure 4](#) and [Table S8](#) report the results of improved maps by different

methods. As shown in [Figure 4A](#), DEMO-EMReF significantly improved the unmasked map-to-model FSC-0.5 resolution across all 27 half-map cases, outperforming all other methods tested ([Figure S4A](#)), with the average FSC-0.5 improving from 5.02 Å in the averaged map to 3.38 Å after enhancement.

DEMO-EMReF also increased the Q-scores of the averaged maps ([Table S8](#)). As shown in [Figure S4B](#), DEMO-EMReF obtained an average Q-score of 0.571, surpassing the original averaged maps (0.503) and other post-processing methods (DeepEMhancer: 0.468; phenix: 0.531; CryoTEN: 0.527; EMReady: 0.569). DEMO-EMReF raised the Q-score in 25 out of 27 cases ([Figure 4B](#)). Further model-to-map correlation analysis ([Table S8](#); [Figure 4C](#)) also revealed significant improvement of DEMO-EMReF for the deposited half-map in terms of CC. As shown in [Figure S4C](#), on average, DEMO-EMReF improved CC_box, CC_mask, and CC_peaks of averaged maps from 0.701 to 0.886, 0.785 to 0.819, and 0.657 to 0.798, respectively. As shown in [Figure 4C](#), DEMO-EMReF improved CC_box, CC_mask, and CC_peaks in 26, 25, and 26 out of 27 cases, respectively. Detailed results for each case are provided in [Tables S9–S13](#).

[Figure 4D](#) presents an example of the DEMO-EMReF-processed averaged map for EMD-18345 (PDB: 8QE8). In the original averaged map, only the overall morphology of the complex is visible, with α -helical regions largely buried in background noise (top left). After DEMO-EMReF processing, the map displays well-defined and continuous α -helices. Enlarged views highlight sharply resolved local features with reduced noise and improved contrast relative to the original maps (top right). When both maps are displayed at the same high contour level for the complete map, the processed DEMO-EMReF density retains more density signal along the main-chain tracing path ([Figures S4D](#) and [S4E](#)). Quantitative analysis shows an improvement in FSC-0.5 resolution from 8.16 to 2.60 Å and an increase in Q-score from 0.621 to 0.678. Consistently, the CC_box, CC_mask, and CC_peaks of the map were improved to 0.897, 0.858, and 0.853, respectively, compared with 0.622, 0.792, and 0.622 for the original averaged map. Additionally, [Figure 4E](#) shows the three-dimensional Fourier Shell Correlation³⁵ (3DFSC) of both the original and processed half-maps of EMD-18345. The original half-maps exhibit strong directional anisotropy with a sphericity of 0.60, whereas DEMO-EMReF-processed half-maps increased the sphericity to 0.92, indicating a substantially more isotropic map. Consistently, anisotropy curves based on Fourier Shell Occupancy (FSO)³⁶ and Bingham test *p* values ([Figure 4F](#)) show a marked reduction in directional variation, where the resolution gap between the 0.9 and 0.1 cutoff frequencies narrowed, the FSO = 1 frequency range broadened, and the FSC resolution at FSO = 0.5 improved from 9.29 to 3.48 Å. These results demonstrate that DEMO-EMReF effectively suppresses anisotropic artifacts and greatly enhances the overall isotropy and quality of the density map.

[Figure S5](#) presents a comparative analysis of DEMO-EMReF and the anisotropy-correction method splsoNet³⁷ on the representative anisotropic case EMD-45551, which was previously evaluated in the original splsoNet study. The comparison was performed using 3DFSC and FSC-based resolution metrics. As expected, splsoNet substantially improved directional isotropy,

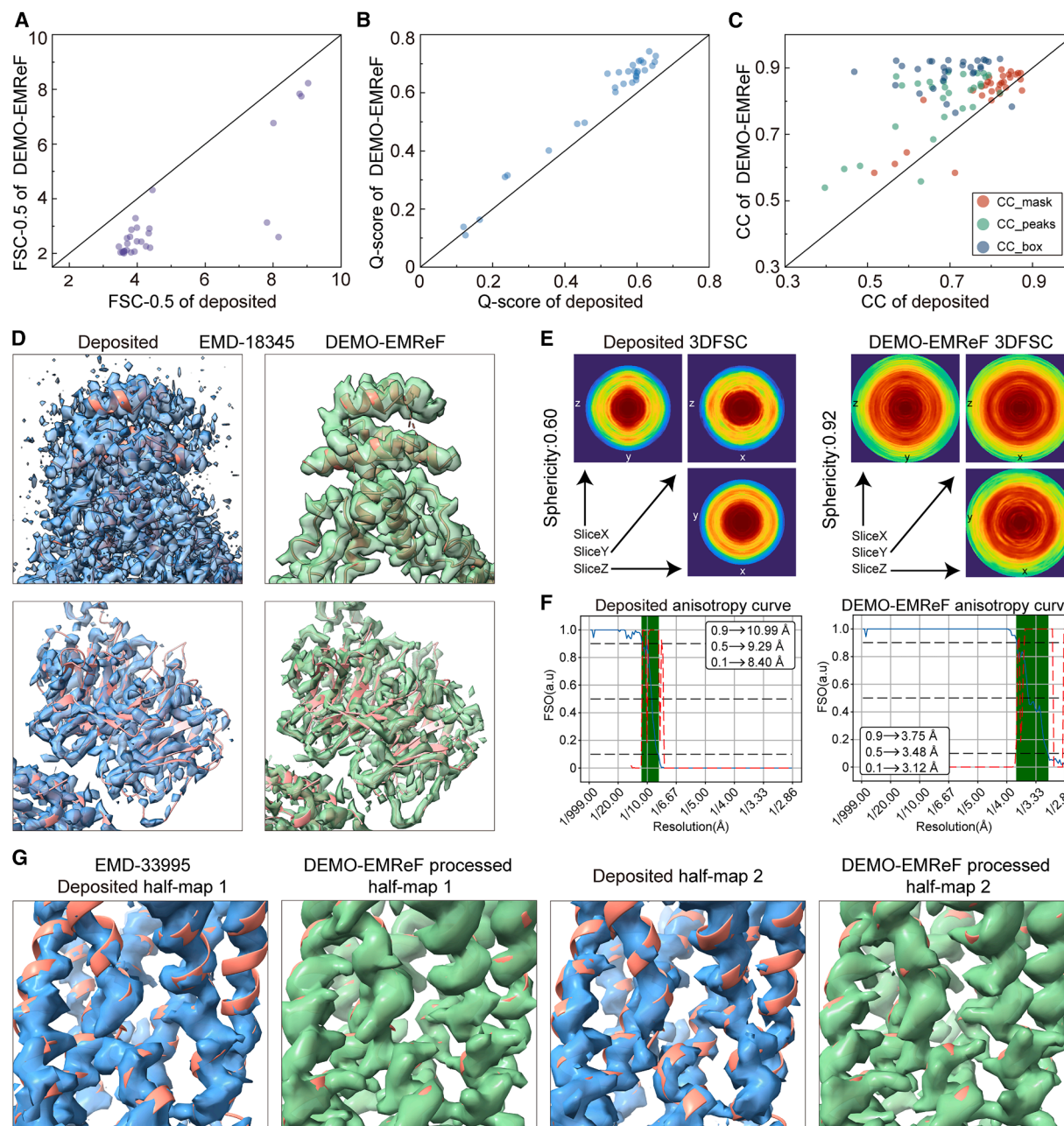


Figure 4. Results of half-map evaluations

(A–C) Scatterplots comparing FSC-0.5, Q-scores, and CC values between the original averaged maps and maps refined by DEMO-EMReF.

(D) Comparison between the original averaged map of EMD-18345 and the DEMO-EMReF-enhanced map at different contour levels.

(E) Anisotropy analysis of the original half-maps of EMD-18345 and the corresponding DEMO-EMReF-enhanced half-maps. Central slices along the X, Y, and Z directions of the 3DFSC are shown for both the original and enhanced half-maps.

(F) Anisotropy curves of the original half-maps and DEMO-EMReF-enhanced half-maps, where the FSO (green line) and the *p* value for the Bingham test (red dashed line) were computed using Scipion. a.u., arbitrary units.

(G) Pairwise comparison of half-maps of EMD-33995, in which one half-map remains unprocessed while the other is enhanced by DEMO-EMReF.

increasing the 3DFSC sphericity from 0.652 to 0.884 (Figure S5C). Although DEMO-EMReF is not specifically designed for anisotropy correction and produced a smaller increase in sphericity, from 0.652 to 0.775, it markedly improved

density continuity, structural interpretability, and fine structural details. Accordingly, the FSC-based resolution improved from 4.30 to 3.08 Å (Figure S5B). Notably, the splsoNet-processed map (EMD-45555) alone did not improve this FSC-based

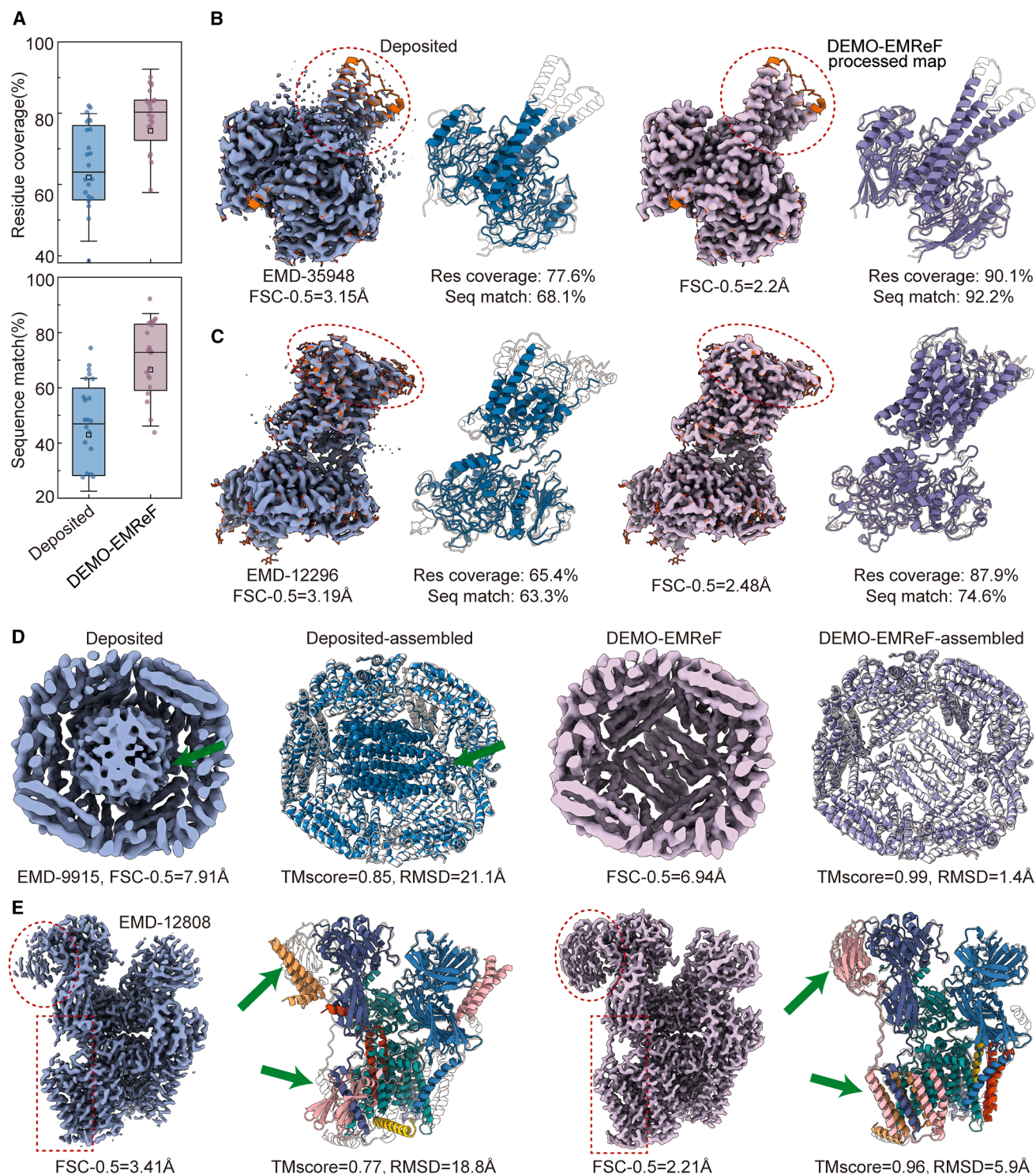


Figure 5. Improved atomic model building enabled by DEMO-EMReF-enhanced maps

(A) Bar plots with overlaid data points comparing residue coverage and sequence match percentage of models built from deposited maps and DEMO-EMReF-enhanced maps. Boxes represent the 25th to 75th percentiles, horizontal center lines indicate medians, square markers indicate means, and whiskers extend to the most extreme data points within $1.5 \times \text{IQR}$ beyond the quartiles. Metrics were calculated against the corresponding deposited PDB structures using phenix.chain_comparison.

(legend continued on next page)

resolution metric. However, subsequent application of DEMO-EMReF further improved the resolution to 1.31 Å while maintaining a high 3DFSC sphericity of 0.855 (Figure S5D). These results indicate that the two methods address complementary aspects of anisotropic map enhancement. splsoNet is more effective in correcting directional anisotropy, whereas DEMO-EMReF provides greater improvements in overall density quality and structural detail.

To rigorously verify DEMO-EMReF's ability to extract genuine structural signals from maps, we performed a cross-validation experiment using the 27 pairs of cryo-EM half-maps from the above experiment and 16 additional pairs of cryo-electron tomography (cryo-ET) half-maps. Each half-map was processed independently, and the unmasked FSC-0.5 resolution between the processed half-map and its corresponding unprocessed complementary half-map was computed. As shown in Table S14 and Figure S6A, DEMO-EMReF improved the average cross FSC-0.5 resolution to 6.18 Å and 6.15 Å for half-map 1 and half-map 2, respectively, compared with 6.86 Å for the unprocessed pairs. A case study on EMD-33995 further confirms that DEMO-EMReF enhances key structural features such as α -helices and β -sheets. The processed half-map shows improved continuity and bridging of previously fragmented densities relative to the deposited half-map (Figures 4G, S6B, and S6D). Consistently, the FSC curve between a DEMO-EMReF-processed half-map and the unprocessed complementary half-map exceeds that between the two unprocessed half-maps across the entire resolution range (Figure S6C), resulting in improved cross FSC-0.5 resolutions of 5.88 and 4.55 Å, compared with 7.50 Å for the unprocessed pair. We also provide a representative β -sheet-rich case (EMD-12309; PDB: 7NG8) in Figure S7, in which improved continuity of β -sheet densities and clearer side-chain features are observed after DEMO-EMReF processing. These results further confirm DEMO-EMReF's ability to enhance true structural signals within half-maps.

Improved atomic model building

The primary objective of map improvement is to enhance interpretability and thereby improve the accuracy of atomic model building. To evaluate the practical impact of DEMO-EMReF on atomic structure construction, we performed *de novo* model building on 25 cryo-EM primary maps. Fully automated model building was independently performed on both the deposited maps and the corresponding DEMO-EMReF-enhanced maps using phenix.map_to_model³⁸ with default parameters and auto-refinement disabled. The resulting models were compared with deposited PDB structures using phenix.chain_comparison³⁹ with a maximum distance threshold of 2 Å. Model quality was evaluated using two metrics: residue coverage and sequence match percentage. As shown in Figure 5A and Table S15, models built from DEMO-EMReF-enhanced maps achieved substantially better downstream modeling performance than those built from the deposited maps. The average

residue coverage increased from 59.4% to 73.6%, while the sequence match increased from 38.9% to 63.3%. Consistent improvements were also observed in local map-model validation, with DAQ increasing from −0.14 to 0.11, EMRinger increasing from 1.14 to 1.71, and MEDIC error decreasing from 0.85 to 0.64. Moreover, DEMO-EMReF improved the FSC-0.5 resolution from 4.49 to 2.82 Å and increased the Q-score and map-model correlation metrics. These results demonstrate that DEMO-EMReF improves map interpretability, map-model agreement, and the completeness and accuracy of automated atomic model construction. Detailed case-level results are provided in Tables S16 and S17.

Figures 5B and 5C compare atomic models built from the deposited maps with those built from DEMO-EMReF-processed maps for two representative cases. As highlighted by the dashed outlines, the deposited maps exhibit locally weak and discontinuous density, leading to backbone-tracing breaks and ambiguous side-chain placements. In contrast, DEMO-EMReF markedly enhances the signal in these regions, reduces background noise, and improves density continuity, thereby improving backbone tracing and side-chain assignment. Figure 5B shows the first example (EMD-35948, PDB: 8J2F). The model built from the deposited map achieved 77.6% residue coverage and 68.1% sequence match, whereas the model built from the DEMO-EMReF-processed map reached 90.1% residue coverage and 92.2% sequence match. Figure 5C displays another example (EMD-12296, PDB: 7NF6). The residue coverage and sequence match of the model generated from the deposited map were 65.4% and 63.3%, respectively, while these metrics were improved to 87.9% and 74.6% when using the DEMO-EMReF-processed map. These results demonstrate that DEMO-EMReF effectively enhances weak density features and overall map continuity, leading to higher-accuracy atomic models in automated *de novo* model building. Figure S8 presents an additional case (EMD-18638) containing substantial regions with local resolutions worse than 4.5 Å, beyond the recommended resolution range of phenix.map_to_model. Nevertheless, both the residue coverage and sequence match of the automatically built model were substantially improved when the DEMO-EMReF-processed map was used.

We further assessed the benefits of DEMO-EMReF-enhanced maps for structure-based model assembly using two representative cases. Monomeric structures were predicted by AlphaFold3⁴⁰ and assembled into complex models using our previously developed DEMO-EMol.^{41–43} The first case is *Bacterioferritin from Streptomyces coelicolor* (EMD-9915, PDB: 6K4M), whose deposited map contains low-resolution density around the iron ion (dashed circles in Figure 5D). Assembly based on the deposited map resulted in several chains misfitting around the iron-binding region, yielding a complex model with a TM-score⁴⁴ of 0.85 and a root-mean-square deviation (RMSD) of 21.1 Å. In contrast, DEMO-EMReF effectively filters out spurious density signals related to the iron ions, resulting in

(B–E) Comparisons of atomic models built from deposited maps and DEMO-EMReF-enhanced maps for EMD-35948 (B), EMD-12296 (C), EMD-9915 (D), and EMD-12808 (E). Deposited maps and their corresponding model are shown in blue, DEMO-EMReF-enhanced maps and corresponding model in purple, and the reference-deposited PDB structure in white. Dashed outlines indicate regions with enhanced density signals after DEMO-EMReF processing. Arrows highlight regions where model assembly was corrected when using DEMO-EMReF-enhanced maps.

a near-native model with TM-score = 0.99 and RMSD = 1.4 Å. The second case is the yeast *Ost6p-containing oligosaccharyl-transferase complex* (EMD-12808, PDB: 7OCI), whose deposited map exhibits substantial missing structural density (dashed box in Figure 5E). Assembly using this raw map produced a complex model with a TM-score of 0.77 and an RMSD of 18.8 Å due to four misfitting chains. When using the DEMO-EMReF-enhanced map, the complex was reconstructed with markedly improved accuracy (TM-score = 0.96, RMSD = 5.9 Å), owing to the strengthening of previously ambiguous structural features. In addition, modeling efficiency was markedly improved after DEMO-EMReF processing, as the removal of irrelevant density signals substantially reduced ambiguity during chain placement.

To further assess the downstream structure-modeling performance of DEMO-EMReF, we expanded the comparative analysis to four independent automated model-building pipelines: phenix.map_to_model, ModelAngelo,⁴ CryoAtom,⁴⁵ and EM-Build.⁴⁶ Detailed parameter settings and commands for these methods are provided in Notes S7–S11. The evaluation was stratified by map resolution, including maps at resolutions of 4.5 Å or better and worse than 4.5 Å, as well as by molecular composition, including protein-only and nucleic-acid-containing cases. The benchmark comprised 30 cryo-EM density maps, including five lower-resolution cases selected to broaden the resolution coverage. Model quality was assessed using coordinate accuracy, sequence assignment, map-model agreement, and stereochemical validation metrics. The results are summarized in Table S18. Overall, DEMO-EMReF demonstrated broad compatibility with different automated model-building pipelines and achieved robust and competitive downstream modeling performance across diverse map resolutions and molecular compositions. Its most pronounced benefits were observed for phenix.map_to_model on maps at resolutions of 4.5 Å or better, with improvements in sequence assignment, coordinate accuracy, map-model agreement, and several geometry-related metrics. DEMO-EMReF also maintained competitive performance when combined with EM-Build on intermediate-resolution protein maps, further supporting its practical utility in automated cryo-EM structure-modeling workflows.

Assessing density-value optimization

DEMO-EMReF improves map quality by directly optimizing density values in real space. It is therefore important to evaluate the validity of this density optimization. However, direct evaluation is difficult, as absolute ground-truth maps or models are typically unavailable in practical applications. Here, we first assessed DEMO-EMReF's ability to suppress artificially added noise using a benchmark set of primary maps. Specifically, noise-free simulated density maps were generated from the corresponding PDB structures with a grid spacing of 1.0 Å. After normalizing density values to the range of 0.0–1.0, Gaussian white noise with a standard deviation of 0.05 was added using Xnipp⁴⁷ to generate noisy maps. These noisy maps were then processed by DEMO-EMReF. The results are summarized in Table S19, where values closer to those of the simulated maps indicate better density optimization. As shown in the table, all quality metrics of the simulated maps declined substantially after noise addition: CC_box decreased from 0.885 to 0.749, CC_mask from 0.828

to 0.781, and CC_peaks from 0.834 to 0.689. After DEMO-EMReF processing, map quality was largely restored, with CC_box, CC_mask, and CC_peaks improving to 0.890, 0.811, and 0.814, respectively. These values closely approach those of the original noise-free simulated maps (Figure S9A), indicating effective noise suppression through direct density optimization.

To further evaluate the performance under extreme noise conditions, we introduced Gaussian noise with a standard deviation of 0.2 into a simulated map generated from a PDB structure (PDB: 8QJX). As shown in Figure S9B, DEMO-EMReF effectively distinguished noise from authentic structural signals and successfully recovers the original density distribution. Unmasked FSC analysis (Figure S9C) shows that the global FSC-0.5 resolution improved markedly from 7.46 Å (between the noisy and simulated maps) to 2.49 Å (between the DEMO-EMReF-processed and simulated maps), with consistent enhancement across a broad resolution range. Furthermore, as observed in previous examples (Figures 3A and 3B), DEMO-EMReF can robustly handle real-world noise by directly optimizing density values.

We then evaluated whether DEMO-EMReF overfits the low-resolution regions caused by heterogeneity or flexibility through a representative example. Figure 6A shows the deposited structure solved from EMD-22879 at 3.54 Å resolution, where the B-factor values of the RIFIN domains were notably higher than those in the core region of the complex, indicating high structural flexibility in these regions.³⁹ As shown in Figure 6B, these flexible structural regions exhibited relatively lower local map resolution as evaluated by CryoRes.⁴⁸ To assess the reliability of density refinement in DEMO-EMReF, we report the local map-to-model correlation in Figure 6C, with the left panel showing the deposited map and the right panel showing the DEMO-EMReF-processed map. Several key observations can be made. First, regions with the lowest local correlation (i.e., highest discrepancy) in the DEMO-EMReF-processed map predominantly localize to the two RIFIN domains, which are characterized by lower resolution. In contrast, regions with higher local correlation corresponded to the more rigid parts of MDB1 Fab, where the local resolution was better. This trend is consistent with the local resolution variations observed in the deposited map. Second, compared with the simulated map, DEMO-EMReF enhanced local correlation in structurally reliable regions of MDB1 Fab at approximately 3–5 Å resolution. However, in flexible regions of the RIFIN domains, where the resolution exceeded 5 Å, the density correlation remained comparable to that of the deposited map. This indicates that DEMO-EMReF effectively refines density in well-resolved regions without overfitting poorly resolved or flexible parts, thereby preserving map interpretability and reliability. Overall, the DEMO-EMReF-processed map demonstrated improvements across multiple quality metrics compared with the deposited map: the FSC-0.5 resolution improved from 4.41 to 2.81 Å (Figure 6F); the Q-score increased from 0.598 to 0.658; and CC_box, CC_mask, and CC_peaks increased from 0.777, 0.833, and 0.754 to 0.929, 0.879, and 0.886, respectively. These results collectively support the rationality and effectiveness of the density refinement in DEMO-EMReF.

Finally, we assessed the reasonableness of density optimization for a map with relatively uniform resolution and minimal

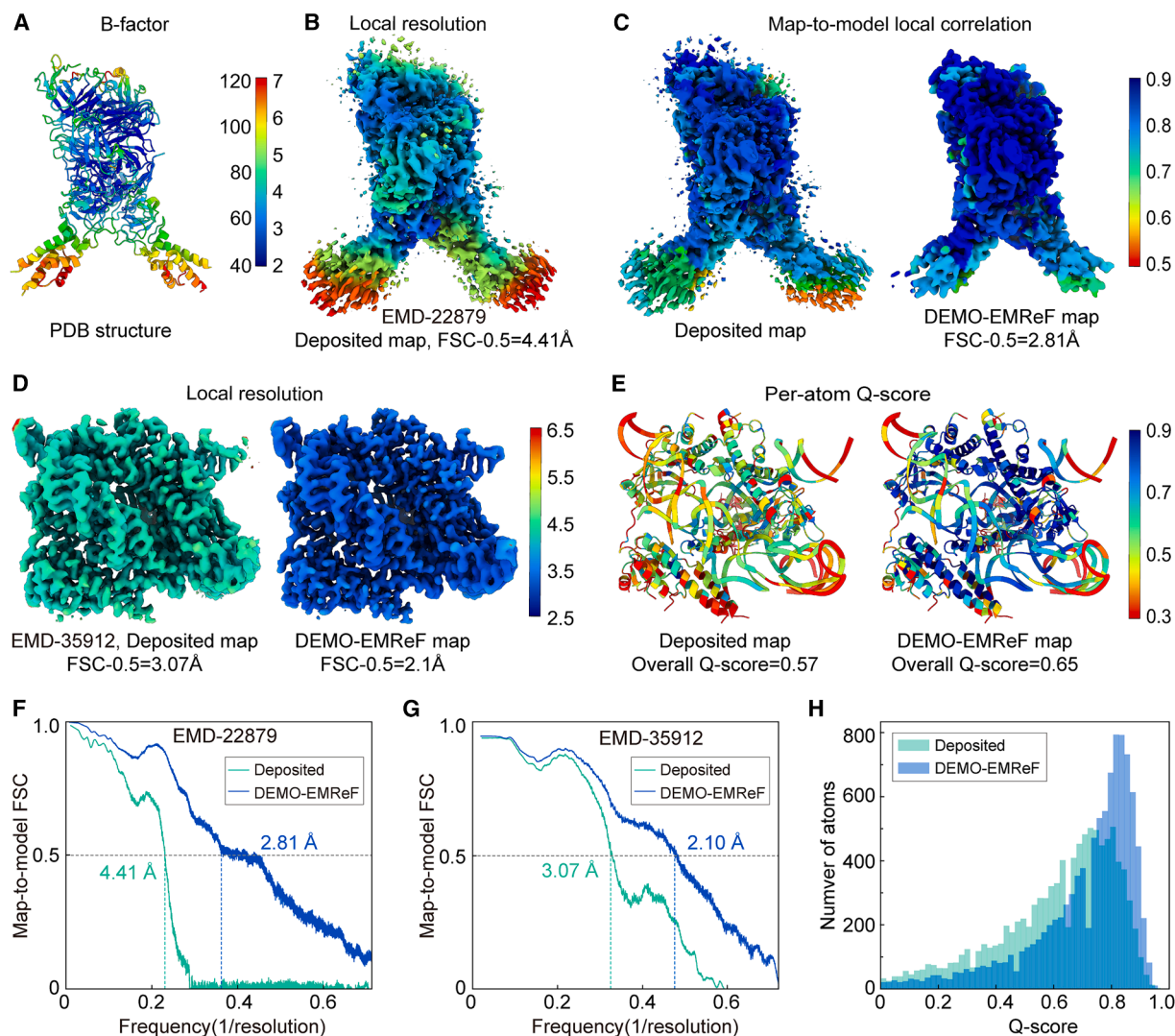


Figure 6. Representative examples illustrating the rationality of density value optimization by DEMO-EMReF

(A) Deposited structure (PDB: 7KHF) colored by B-factor.
(B) Deposited map of EMD-22879 colored by the local resolution.
(C) Local map-model correlation comparison between the deposited map of EMD-22879 and the corresponding DEMO-EMReF-processed map.
(D) Local resolution comparison between the deposited map of EMD-35912 and the corresponding DEMO-EMReF-processed map.
(E) Atomic structures (PDB: 8J12) colored by Q-score computed using the deposited map of EMD-35912 and the DEMO-EMReF-processed map.
(F and G) Unmasked map-model FSC curves for EMD-22879 and EMD-35912, respectively.
(H) Q-score distributions of the deposited map (EMD-35912) and the corresponding DEMO-EMReF-processed map.

noise using a representative case. Figure 6D (left panel) shows the deposited map of EMD-35912 at 3.08 Å resolution, colored by local resolution. The local resolution of the map is mostly around 3.5 Å, with little apparent noise. The right panel of Figure 6D shows the map after DEMO-EMReF processing, where the local resolution is uniformly improved to around 2.5 Å with clearer structural details. The FSC-0.5 resolution of the deposited map was improved from 3.07 to 2.10 Å (Figure 6G) after DEMO-EMReF processing. Figure 6E shows the deposited structures colored by the Q-score calculated from the deposited and DEMO-EMReF-processed maps,

respectively, and the corresponding Q-score distributions are shown in Figure 6H. The improvement in Q-score is consistent with the enhancement in local resolution. As a result, DEMO-EMReF increased the average Q-score of the deposited map from 0.570 to 0.653. In addition, compared with the CC_box (0.843), CC_mask (0.838), and CC_peaks (0.824) of the deposited map, DEMO-EMReF achieved higher CC_box (0.907), CC_mask (0.863), and CC_peaks (0.866). These results further highlight the reasonableness of density optimization by DEMO-EMReF, demonstrating its ability to improve local map quality without inducing overfitting.

Ablation experiments

To systematically assess the contributions of key components in DEMO-EMReF, we performed a series of ablation studies. Unlike other deep learning approaches that employ U-Net architectures, DEMO-EMReF adopts an enhanced Swin Transformer that integrates convolution-enhanced SE blocks and sparse grid-window attention to capture complex local and non-local features across both channel and spatial domains. To examine how specific components affect the performance, we trained three ablated variants on the same dataset: (1) a reduced-depth model to quantify the impact of network depth (DEMO-EMReF-3); (2) a model employing SCUNet, which incorporates Swin Transformer blocks into a U-Net backbone; and (3) an SCUNet++-based model⁴⁹ augmented with dense skip connections to strengthen feature propagation.

We evaluated the performance of all ablation models on the benchmark set of primary maps and using FSC-0.5, Q-score, and CC as evaluation metrics. The averaged results of all cases are shown in Table S20. It should be noted that the SCUNet baseline uses the same architecture as EMReady but was re-trained on the DEMO-EMReF dataset under the same training configuration as the other ablation models. In addition, Table S20 reports results for the full benchmark dataset, whereas Table S1 includes only cases with valid outputs from all compared methods. Therefore, the results in the two tables are not directly comparable. Although larger input density chunks can enhance a model's ability to capture non-local features, excessively large chunks drastically increase computational cost. Following recommendations in the EMReady paper, DEMO-EMReF adopts a density chunk of $48 \times 48 \times 48$ voxels. To verify the superiority of DEMO-EMReF in capturing non-local features, we compared it against the two SCUNet-based models using the same chunk size. As shown in Figure S10, the baseline DEMO-EMReF model achieved consistently better performance in terms of average FSC-0.5 resolution and Q-score compared with the SCUNet-based models. This demonstrates that integrating convolution-enhanced SE blocks with sparse grid-window attention enables more stable and effective learning of non-local region features than SCUNet-based architectures.

In general, network depth plays a critical role in determining feature extraction capability. Blindly reducing the number of layers may degrade model accuracy, whereas overly deep networks often require substantially more training data and computational resources. Figure S10 shows that the baseline DEMO-EMReF model, which employs a four-layer network, outperforms the reduced-depth variant DEMO-EMReF-3. Accordingly, we set the network depth of DEMO-EMReF to four layers to balance performance and computational efficiency.

Furthermore, as illustrated in Figure S11 for EMD-27533 (PDB: 8DMB), the baseline DEMO-EMReF model effectively reduces the risk of misidentifying genuine density signals near structural boundaries as noise. In contrast, the SCUNet-based model exhibits partial density loss in edge regions (highlighted by dashed outlines), whereas DEMO-EMReF successfully preserves these critical structural features, generating a more continuous and well-defined density distribution. These results further validate the rationality and robustness of the DEMO-EMReF network architecture.

To further evaluate the contribution of individual network blocks in DEMO-EMReF, we conducted leave-one-out ablation studies on the primary-map dataset. Specifically, we evaluated two variants: (1) w/o channel attention block, where the channel attention module was removed while retaining the full spatial attention design, and (2) w/o Grid-MSA, where the grid-based sparse attention module was removed while preserving the remaining hybrid attention components. The quantitative results are summarized in Table S21. For a fair comparison, metrics were computed only on cases where all variants produced valid outputs. Removing either component resulted in consistent decreases across all evaluation metrics, demonstrating that the channel attention and Grid-MSA modules provide complementary contributions. The best overall performance was achieved only when both components were jointly incorporated.

Computational efficiency analysis

To assess the computational efficiency of DEMO-EMReF, we compared its inference cost with those of existing cryo-EM map enhancement methods, including DeepEMhancer, CryoTEN, EMReady, and CryoFM. As summarized in Table S22, the reported efficiency metrics were averaged over the same set of 10 representative cryo-EM maps evaluated under identical benchmarking conditions. DEMO-EMReF achieved a favorable balance between computational cost and map enhancement performance. Specifically, it required substantially fewer FLOPs and MACs than DeepEMhancer and CryoFM, resulting in markedly shorter inference times. Although its peak GPU memory consumption was higher than that of lightweight methods such as EMReady and CryoTEN because of its attention-based architecture, DEMO-EMReF maintained substantially higher throughput than DeepEMhancer and CryoFM. Table S23 provides the map-specific inference results for these 10 representative cryo-EM maps with different voxel dimensions, demonstrating relatively consistent per-chunk processing efficiency across maps of varying sizes. Overall, these results show that DEMO-EMReF maintains practical computational efficiency while preserving the advantages of attention-based structural feature recovery.

DISCUSSION

In this study, DEMO-EMReF demonstrates that explicitly facilitating information exchange across distant regions is critical for effective cryo-EM density map refinement. Unlike conventional Swin Transformer architectures that primarily aggregate contextual information through fixed sliding-window attention, the hybrid attention design of DEMO-EMReF enables a more balanced integration of fine local features and long-range structural context. This capability is particularly relevant for cryo-EM map refinement, where accurate modeling of global structural context is critical for mitigating anisotropic resolution artifacts and preserving overall density coherence. In addition, channel-wise feature recalibration allows DEMO-EMReF to adaptively emphasize informative density channels while suppressing noise-dominated responses, thereby improving discrimination between true density signals and background noise. Together, the complementary roles of spatial and channel attention provide

an interpretable framework for understanding the improved interpretability of cryo-EM maps observed in our experiments.

Systematic benchmarking on both primary maps and half-maps indicates that DEMO-EMReF consistently improves map quality relative to all compared methods. Beyond enhancing overall resolution and local structural details, the refined maps exhibit reduced anisotropic resolution artifacts, leading to more reliable density interpretation. These improvements translate directly into increased accuracy and efficiency in both *de novo* atomic model building and fitting-based model assembly. Importantly, ablation analyses suggest that these gains primarily stem from the hybrid attention strategy, in which channel-wise feature recalibration and cross-region spatial interactions jointly contribute to effective modeling of both local and global density features.

Despite these promising results, the performance of DEMO-EMReF could be further improved in several aspects. First, the current training strategy relies on simulated density maps derived from atomic structures as supervision, which may not fully capture the complexity of experimental noise and reconstruction artifacts. More physically informed target-generation strategies, such as electron-scattering-factor-based calculations, multi-Gaussian representations, and resolution- or B-factor-dependent simulations, may provide more faithful supervision, particularly in chemically heterogeneous regions containing charged atoms, ions, ligands, nucleic acids, or solvent molecules.^{50–53} Incorporating these representations together with more chemically diverse training data may further improve model generalization and local reconstruction accuracy. Second, DEMO-EMReF operates on deposited maps that may have already lost information during reconstruction. Integrating experimental single-particle images, together with simulated images from reference models such as homologous structures or AlphaFold3 predictions, represents a potential avenue for further improvement. Third, the current chunk-based processing and grid-based sparse attention mechanisms provide an efficient approximation of global context but its computational complexity still increases with the number of grid tokens, which may limit scalability for increasingly large density maps. Future work will explore more adaptive and scalable global-context modeling strategies. Finally, commonly used evaluation metrics such as CC are inherently resolution dependent, and discrepancies in resolution distributions between target and reference maps can bias quantitative assessment. Developing resolution-aware or task-oriented metrics therefore remains an important direction for future work. Overall, given its accuracy and robustness in enhancing both resolution and interpretability, DEMO-EMReF provides a robust and effective tool for improving cryo-EM map quality to support downstream structural and functional studies in the era of cryo-EM structural biology.

METHODS

Dataset construction

To effectively train and evaluate the DEMO-EMReF framework, we curated a high-quality dataset containing density maps and their corresponding atomic models. Specifically, single-particle cryo-EM and cryo-ET entries with reported resolutions between 3.0 and 8.0 Å were collected from the EMDB. For maps associ-

ated with multiple PDB models, only one representative structure was selected to avoid sample redundancy. Entries were excluded if they exhibited non-orthogonal map axes, had mismatches between the structure and the density map, or lacked reported resolution values determined using the FSC-0.143 criterion. To ensure spatial consistency between structures and maps, the `phenix.map_model_cc`²⁷ tool was employed to compute two quality metrics: CC_mask and CC_box. Only samples with CC_mask > 0.75 and CC_box > 0.6 were retained. For maps with reported resolutions of 6.0–8.0 Å, the CC_mask threshold was moderately relaxed to 0.6 to expand the dataset.

In addition, to eliminate redundancy, the FASTA sequences of all samples were clustered using MMseqs2⁵⁴ with a sequence identity threshold of 30% and a minimum sequence coverage of 80%. The detailed clustering procedure and the corresponding MMseqs2 command are provided in [Note S12](#). For each cluster, the first sequence listed by MMseqs2 was selected as the representative entry and retained for subsequent dataset construction. The resulting dataset comprised 672 protein-only maps, 242 protein-nucleic acid complex maps, and 9 nucleic acid-only maps, with resolution ranges of 3.0–8.0 Å, 3.0–8.0 Å, and 3.02–7.29 Å, respectively. The resolution distribution of the resulting dataset is summarized in [Figure S12](#). Among these samples, 111 cryo-EM maps were randomly assigned to the test set using a fixed random seed of 200 to ensure reproducibility. In addition, 16 cryo-ET cases for which half-map files could be successfully retrieved from the EMDB and showed good agreement with their corresponding atomic models were included in the test set. The remaining samples were divided into training and validation sets at a ratio of 4:1 using the same random seed of 200. Detailed information on all dataset entries is provided in [Table S24](#), and a comparison of the datasets used by representative map enhancement methods is presented in [Table S25](#).

Map data preprocessing

To ensure consistent preprocessing across all maps and improve training efficiency, a unified preprocessing protocol was applied to both experimental maps and their corresponding simulated maps. All maps were resampled to a uniform voxel spacing of 1.0 Å using cubic interpolation. Negative density values were clipped to zero to remove non-physical noise. To normalize density amplitudes across different maps, all density values were linearly scaled to the range [0, 1] using the 99.999th percentile of each map's density distribution. In the training set, the input consisted of deposited experimental maps, while the supervision was provided by simulated density maps generated from the corresponding PDB structures. These simulated maps were generated using a Gaussian kernel-based method, in which each atom contributes to the density at a grid point according to its atomic number and spatial distance. Specifically, the density value at a grid point x is computed as follows:

$$\rho(x) = \sum_{i=1}^N \theta s_i \exp(-k \|x - p_i\|^2)$$

where s_i and p_i denote the atomic number and Cartesian coordinates of the i -th atom, respectively, and N is the total number of atoms. The attenuation factor k was determined from the

reported FSC resolution R_0 at the 0.143 threshold⁵⁵ as $k = (\pi/0.9R_0)^2$, and the scaling factor was defined as $\theta = (k/\pi)^{3/2}$. In addition, given the large size of cryo-EM maps, all maps were partitioned into overlapping 3D density chunks to enable efficient training and inference. Only chunks containing meaningful density signals were retained, while empty regions were discarded. During validation and testing, maps were split into overlapping chunks of size $48 \times 48 \times 48$ with a stride of 24 voxels. The final enhanced maps were reconstructed by weighted averaging in the overlapping regions.

Neural network architecture

DEMO-EMReF adopts a Transformer-based neural network architecture with a hybrid attention mechanism, designed to enhance both the resolution and resolvability of density maps. As illustrated in Figure 1B, the model takes as input a 3D density chunk of size $48 \times 48 \times 48$ voxels (voxel spacing of 1.0 Å) and outputs an enhanced density chunk of the identical dimensions. The overall architecture (Figure 1, center panel) consists of four sequentially stacked residual hybrid attention modules, each comprising two core components: a channel attention module and a hybrid spatial attention module.

The channel attention module is built upon the SE mechanism, which dynamically modulates feature responses along the channel dimension. Specifically, the input density chunks are first processed by a $3 \times 3 \times 3$ convolutional layer with an expansion factor of six to increase channel dimensionality. A global squeeze operation with a reduction ratio of 0.5 is then applied to aggregate global context, followed by a $1 \times 1 \times 1$ convolution (excitation) to restore the original channel dimension. Through global context compression and channel-wise reweighting, this module enhances informative features while suppressing less relevant ones, thereby improving feature discrimination and signal-to-noise characteristics.

To capture both local and long-range spatial dependencies in density maps, DEMO-EMReF employs a hybrid spatial attention module built upon the Swin Transformer framework. An initial Swin Transformer block is applied to extract local and non-local spatial representations. To further expand the effective receptive field and strengthen cross-region interactions, feature maps are partitioned and processed using a Grid-MSA, together with W-MSA and SW-MSA. In this process, feature maps are divided into non-overlapping grid regions with small spatial offsets, enabling sparse attention computation across grids and facilitating efficient long-range information exchange. The resulting features are subsequently integrated through a multi-layer perceptron and a convolutional fusion module to combine local and global representations. A detailed schematic of the grid-based sparse attention module and its integration into the three-dimensional hybrid attention framework of DEMO-EMReF is provided in Figure S13, while its implementation details are summarized in Table S26.

Overall, the network synergistically integrates channel-wise feature recalibration with an enhanced spatial attention, enabling simultaneous learning of fine-grained local details and long-range structural dependencies. This design allows DEMO-EMReF to robustly refine cryo-EM density maps across a wide range of resolutions and structural complexities.

Network training

The network of DEMO-EMReF was trained for 100 epochs on an NVIDIA A800 GPU with 80 GB of memory. To balance convergence stability and memory efficiency, a batch size of 8 and a 4-layer Transformer encoder-decoder architecture were adopted. All maps were split into overlapping density blocks of size $48 \times 48 \times 48$ voxels after preprocessing, consistent with the block size used during inference. To reduce overfitting and improve generalization, several data augmentation strategies were applied, including random cropping from $64 \times 64 \times 64$ to $48 \times 48 \times 48$ voxel chunks, random 90° rotations, and axis flipping. In addition, Gaussian white noise with a variance of 0.1 was added to each density block with a probability of 0.1 to enhance the denoising capability of the model. Dropout layers with a rate of 0.1 were applied in both the Transformer encoder and decoder modules.

The training loss function combined the MSE⁵⁶ loss and the SSIM⁵⁷ loss. The MSE loss quantifies voxel-wise local differences and is defined as follows:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (V_i - v_i)^2 \quad (\text{Equation 1})$$

where N denotes the total number of nonzero voxels (i.e., voxels that are nonzero in either the predicted or target blocks), V_i is the predicted voxel value, and v_i is the corresponding ground-truth value.

To capture global structural similarity, the SSIM loss was incorporated and computed as:

$$L_{SSIM} = 1 - \frac{2C_{PT} + \delta}{S_P^2 + S_T^2 + \delta} \quad (\text{Equation 2})$$

where S_P and S_T denote the standard deviations of the predicted and target density chunks, respectively; C_{PT} is their covariance; and $\delta = 10^{-6}$ is a small constant to prevent division by zero. The total loss function is defined as a weighted sum of the two terms:

$$L = \lambda_1 L_{MSE} + \lambda_2 L_{SSIM} \quad (\text{Equation 3})$$

where $\lambda_1 = 1$ and $\lambda_2 = 4$ are weighting factors used to balance their contributions.

Model optimization was performed using the Adam optimizer with an initial learning rate of 5×10^{-4} , which was decayed to a minimum of 1×10^{-5} using a cosine annealing schedule over the course of training. Model performance was evaluated on the validation set after each epoch, and the checkpoint with the lowest validation loss was selected for final evaluation. The training process was implemented using PyTorch 2.3.1 and CUDA 12.1. Where applicable, the hyperparameters were based on the recommended settings of EMReady, and any necessary adjustments were made exclusively based on performance on the validation set. Detailed system configurations are provided in Note S13. The average training time was approximately 2.5 h per epoch, resulting in a total training time of about 250 h.

Evaluation metrics

First, frequency-domain resolution and anisotropy were evaluated. FSC curves between maps (before and after enhancement) and their corresponding atomic models were calculated using phenix.mtriage,²⁷ with the resolution at FSC = 0.5 adopted as

the primary metric for map resolution. To assess directional resolution anisotropy, 3DFSC was adopted. 3DFSC provides voxel-wise visualization of directional FSC values across Fourier space, enabling detection of anisotropic regions arising from preferred particle orientation or data-collection geometry. The sphericity derived from the 3DFSC distribution was used as a quantitative descriptor of isotropy and directional resolution variation in the enhanced maps.

In addition, FSO was also employed to evaluate frequency-dependent anisotropy and completeness of the enhanced maps. FSO measures the fraction of directions in each Fourier shell for which the directional FSC exceeds 0.143. Values close to 1 indicate an isotropic Fourier signal distribution, whereas lower values reflect increasing anisotropy. The frequency over which FSO declines from 0.9 to 0.1 defines the anisotropy transition zone, and the shell with FSO = 0.5 approximately corresponds to the global resolution. Thus, FSO complements FSC and 3DFSC by providing a compact descriptor of both resolution and directional uniformity.

Second, real-space map-model correlation was calculated. Real-space correlation coefficients (CC_mask, CC_box, and CC_peaks) were computed using phenix.map_model_cc to quantify the agreement between density maps and atomic models. CC_mask measures correlation within a molecular mask, CC_box assesses global correlation across the entire map, and CC_peaks evaluates fit at the highest-density regions by comparing prominent density peaks.

Third, local atomic resolvability was assessed. Local model interpretability was evaluated using Q-score, computed with the MapQ plugin in UCSF Chimera. Q-score quantifies the similarity between the observed local density profile around an atom and an idealized Gaussian-shaped peak, with higher values indicating clearer atomic features and higher local resolvability. In addition, local resolution was estimated using CryoRes, a deep-learning method that predicts voxel-wise resolution directly from a map. For each voxel, CryoRes outputs a resolution value learned from FSC-based local-resolution training data, producing a 3D map that reflects regional variations in density quality, where lower values indicate better-resolved features.

Finally, the accuracy of reconstructed atomic models was computed. The accuracy of *de novo* reconstructed atomic models from maps was assessed using residue coverage and sequence matching rate calculated by phenix.chain_comparison. Residue coverage quantifies the percentage of residues whose representative atoms (C α for proteins; P for nucleic acids) lie within 2.0 Å of corresponding residues in the reference structure, regardless of residue identity. Sequence matching rate measures the proportion of residues that are both spatially aligned and correctly assigned to the reference sequence. Higher values for both metrics indicate better map interpretability, more accurate backbone placement, and more reliable residue typing.

For atomic models assembled using computationally predicted chain structures, structural accuracy was evaluated by TM-score and RMSD calculated with US-align.⁵⁸ TM-score evaluates the global topological similarity between two protein structures on a scale of (0, 1), with higher values indicating closer

overall agreement. RMSD measures the average distance between corresponding atoms after optimal alignment, with lower RMSD values denoting greater structural similarity.

A comprehensive classification and descriptions of all evaluation metrics used in this study are provided in Table S27.

Performance evaluation of DEMO-EMReF

The performance of DEMO-EMReF was evaluated on a benchmark set of 111 primary maps and compared with DeepEMhancer, EMReady, CryoTEN, phenix, EMReady2, and CryoFM, under identical hardware conditions using default parameters for all methods. Average performance metrics were computed only for cases that yielded valid outputs across all methods. DeepEMhancer provides three model configurations: tightTarget, wideTarget, and highRes. For maps with resolutions between 3 and 6 Å, the best-performing result among the three models was selected, whereas for maps in the 6–8 Å range, only the tightTarget and wideTarget models were considered.

To examine the robustness of DEMO-EMReF under low signal-to-noise conditions and its ability to recover true structural signals, an independent evaluation was performed on 43 half-map pairs (27 EM pairs and 16 ET pairs). DEMO-EMReF was first applied to the averaged map from each half-map pair. In addition, a cross-half-map validation was performed, in which one half-map was processed by DEMO-EMReF and FSC was computed between the processed half-map and its unprocessed complementary half-map.

To evaluate the benefits of DEMO-EMReF for atomic model building, *de novo* model building was performed on 25 enhanced maps using phenix.map_to_model. The resulting models were assessed using phenix.chain_comparison and compared with models built from the corresponding deposited maps. The effect of DEMO-EMReF on structure-based assembly was further evaluated by assembling complex structures from AlphaFold3-predicted monomers using our previously developed DEMO-EMol pipeline. Models assembled using DEMO-EMReF-enhanced maps were compared with those generated from deposited maps.

To assess the reasonableness of the density optimization performed by DEMO-EMReF, artificial Gaussian noise was added to simulated maps generated from the PDB structures corresponding to the 111 primary maps using Xmipp. DEMO-EMReF was then applied to the noisy maps to evaluate its ability to recover the simulated maps by identifying the noise. Furthermore, DEMO-EMReF was tested on representative cases exhibiting low-resolution regions caused by structural heterogeneity or flexibility, as well as on cases with relatively uniform resolution and minimal noise, to assess its ability to improve local density quality.

Finally, to quantify the contributions of individual architectural components in DEMO-EMReF, ablation studies were conducted under identical experimental settings, using the same datasets and hyperparameters as the baseline DEMO-EMReF model. Five ablated variants were examined: (1) DEMO-EMReF-3, a reduced-depth version of the baseline model; (2) a SCUNet-based model incorporating Swin Transformer blocks into a U-Net backbone; (3) a SCUNet++-based model enhanced with dense skip connections; (4) a variant excluding the channel attention module; and (5) a variant excluding the Grid-based MSA module.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xiaogen Zhou (zgx@zjut.edu.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

- All data supporting the findings and conclusions of this study are available within the paper and its [supplemental information](#). The experimental density maps and the deposited structures used in this study were downloaded from EMDB and PDB.
- All original code, the trained DEMO-EMReF model weights, and the density maps used for training and testing have been deposited in Zenodo at <https://doi.org/10.5281/zenodo.22111435> and GitHub at <https://github.com/zhoulab/DEMO-EMReF> and are publicly available as of the date of publication. A freely accessible online server is available at <http://zhouxglab.com/DEMO-EMReF/>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2022ZD0115103 to G.Z.), the National Natural Science Foundation of China (62203389 to X.Z. and 62573386 to G.Z.), Fundamental Research Funds for the Provincial Universities of Zhejiang (RF-C2024006 to X.Z.), Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2023R01006 to X.Z.), Zhejiang Provincial Special Support Program for High-Level Talents (2023R5248 to G.Z.), and the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (2025C01121 to X.Z. and 2025C01190 to G.Z.).

AUTHOR CONTRIBUTIONS

X.Z. conceived and designed the project. J.L. developed and implemented the computational pipeline and performed the test. G.Z. and C.W. jointly supervised the study. Z.Z. performed data analysis and prepared the figures. Y.Z. and F.C. assisted with data analysis. J.L. and X.Z. wrote the manuscript. All authors revised and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrp.2026.103596>.

Received: April 20, 2026

Revised: July 25, 2026

Accepted: September 17, 2026

REFERENCES

- Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* 13, 24–27. <https://doi.org/10.1038/nmeth.3694>.
- Strack, R. (2020). Cryo-EM goes atomic. *Nat. Methods* 17, 1175. <https://doi.org/10.1038/s41592-020-01014-1>.
- Yip, K.M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature* 587, 157–161. <https://doi.org/10.1038/s41586-020-2833-4>.
- Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S.H.W. (2024). Automated model building and protein identification in cryo-EM maps. *Nature* 628, 450–457. <https://doi.org/10.1038/s41586-024-07215-4>.
- Zhou, X., Xu, X., and Zhang, G. (2026). When cryo-EM modeling meets structure prediction: Structural biology technologies. *Nat. Struct. Mol. Biol.* 33, 200–202. <https://doi.org/10.1038/s41594-026-01744-4>.
- Cai, Y., Zhang, Z., Xu, X., Xu, L., Chen, Y., Zhang, G., and Zhou, X. (2025). Fitting atomic structures into Cryo-EM maps by coupling deep learning-enhanced map processing with global-local optimization. *J. Chem. Inf. Model.* 65, 3800–3811. <https://doi.org/10.1021/acs.jcim.5c00004>.
- Fernández-Giménez, E., Martínez, M., Sánchez-García, R., Marabini, R., Ramírez-Aportela, E., Conesa, P., Carazo, J.M., and Sorzano, C.O.S. (2021). Cryo-EM density maps adjustment for subtraction, consensus and sharpening. *J. Struct. Biol.* 213, 107780. <https://doi.org/10.1016/j.jsb.2021.107780>.
- Terwilliger, T.C., Sobolev, O.V., Afonine, P.V., and Adams, P.D. (2018). Automated map sharpening by maximization of detail and connectivity. *Biological Crystallography* 74, 545–559. <https://doi.org/10.1107/S2059798318004655>.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745. <https://doi.org/10.1016/j.jmb.2003.07.013>.
- Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296. <https://doi.org/10.1038/nmeth.4169>.
- Jakobi, A.J., Wilmanns, M., and Sachse, C. (2017). Model-based local density sharpening of cryo-EM maps. *eLife* 6, e27131. <https://doi.org/10.7554/eLife.27131>.
- Ramírez-Aportela, E., Vilas, J.L., Glukhova, A., Melero, R., Conesa, P., Martínez, M., Maluenda, D., Mota, J., Jiménez, A., Vargas, J., et al. (2020). Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* 36, 765–772. <https://doi.org/10.1093/bioinformatics/btz671>.
- Kaur, S., Gomez-Blanco, J., Khalifa, A.A.Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J.S., Bui, K.H., and Vargas, J. (2021). Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Nat. Commun.* 12, 1240. <https://doi.org/10.1038/s41467-021-21509-5>.
- Luo, Z., Campos-Acevedo, A.A., Lv, L., Wang, Q., and Ma, J. (2021). Sparseness and Smoothness Regularized Imaging for improving the resolution of Cryo-EM single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* 118, e2013756118. <https://doi.org/10.1073/pnas.2013756118>.
- Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* 4, 874. <https://doi.org/10.1038/s42003-021-02399-1>.
- Venkata Subramaniya, S.R.M., Terashi, G., and Kihara, D. (2021). Super resolution Cryo-EM maps with 3D deep generative networks. *Biophys. J.* 120, 283a. <https://doi.org/10.1016/j.bpj.2020.11.1801>.
- He, J., Li, T., and Huang, S.-Y. (2023). Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nat. Commun.* 14, 3217. <https://doi.org/10.1038/s41467-023-39031-1>.
- Guo, X., O'Neill, W.C., Vey, B., Yang, T.C., Kim, T.J., Ghassemi, M., Pan, I., Gichoya, J.W., Trivedi, H., and Banerjee, I. (2021). SCU-Net: A deep learning method for segmentation and quantification of breast arterial calcifications on mammograms. *Med. Phys.* 48, 5851–5861. <https://doi.org/10.1002/mp.15017>.
- Selvaraj, J., Wang, L., and Cheng, J. (2025). CryoTEN: efficiently enhancing cryo-EM density maps using transformers. *Bioinformatics* 41, btaf092. <https://doi.org/10.1093/bioinformatics/btaf092>.
- Shaker, A., Maaz, M., Rasheed, H., Khan, S., Yang, M.H., and Shahbaz Khan, F. (2024). UNETR++: delving into efficient and accurate 3D medical

- p>image segmentation.
- IEEE Trans. Med. Imag.*
- 43, 3377–3390.
- <https://doi.org/10.1109/TMI.2024.3398728>
- .
21. Cao, H., Zhu, Y., Li, T., Chen, J., He, J., Wang, X., and Huang, S.Y. (2026). EMReady2: improvement of cryo-EM and cryo-ET maps by local quality-aware deep learning with Mamba. *Nat. Commun.* 17, 6493. <https://doi.org/10.1038/s41467-026-71794-1>.
 22. Zhou, Y., Li, Y., Yuan, J., and Gu, Q. (2024). CryoFM: A flow-based foundation model for cryo-EM densities. Preprint at arXiv:2410.08631. <https://doi.org/10.48550/arXiv.2410.08631>.
 23. Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141.
 24. Chu, S.-C., Dou, Z.-C., Pan, J.-S., Weng, S., and Li, J.H. (2024). Hybrid multi-axis aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6257–6266. <https://doi.org/10.1109/CVPRW63382.2024.00629>.
 25. Lawson, C.L., Patwardhan, A., Baker, M.L., Hryc, C., Garcia, E.S., Hudson, B.P., Lagerstedt, I., Ludtke, S.J., Pintilie, G., Sala, R., et al. (2016). EMDatabank unified data resource for 3DEM. *Nucleic Acids Res.* 44, D396–D403. <https://doi.org/10.1093/nar/gkv1126>.
 26. Westbrook, J., Feng, Z., Jain, S., Bhat, T.N., Thanki, N., Ravichandran, V., Gilliland, G.L., Bluhm, W., Weissig, H., Greer, D.S., et al. (2002). The protein data bank: unifying the archive. *Nucleic Acids Res.* 30, 245–248. <https://doi.org/10.1093/nar/30.1.245>.
 27. Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Biological Crystallography* 74, 814–840. <https://doi.org/10.1107/S2059798318009324>.
 28. Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M.F., and Chiu, W. (2020). Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* 17, 328–334. <https://doi.org/10.1038/s41592-020-0731-1>.
 29. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
 30. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. <https://doi.org/10.1002/pro.3943>.
 31. Hirano, S., Kappel, K., Altae-Tran, H., Faure, G., Wilkinson, M.E., Kannan, S., Demircioglu, F.E., Yan, R., Shiozaki, M., Yu, Z., et al. (2022). Structure of the OMEGA nickase IrsB in complex with ω RNA and target DNA. *Nature* 610, 575–581. <https://doi.org/10.1038/s41586-022-05324-6>.
 32. Jin, F., Lin, Y.Y., Wang, R.C., Xie, T.X., Zhao, Y., Shen, C., Sheng, D., Ichikawa, M., Yu, Y., Wang, J., and Hattori, M. (2024). Cryo-EM structure of the zinc-activated channel (ZAC) in the Cys-loop receptor superfamily. *Proc. Natl. Acad. Sci. USA* 121, e2405659121. <https://doi.org/10.1073/pnas.2405659121>.
 33. Jiang, H., Ariyoshi, M., Hori, T., Watanabe, R., Makino, F., Namba, K., and Fukagawa, T. (2023). The cryo-EM structure of the CENP-A nucleosome in complex with ggKNL2. *EMBO J.* 42, e111965. <https://doi.org/10.15252/embj.2022111965>.
 34. Chen, F., Zhang, C., Xue, J., Wang, F., and Li, Z. (2024). Molecular mechanism for target RNA recognition and cleavage of Cas13h. *Nucleic Acids Res.* 52, 7279–7291. <https://doi.org/10.1093/nar/gkac324>.
 35. Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carraher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* 14, 793–796. <https://doi.org/10.1038/nmeth.4347>.
 36. Vilas, J.-L., and Tagare, H.D. (2023). New measures of anisotropy of cryo-EM maps. *Nat. Methods* 20, 1021–1024. <https://doi.org/10.1038/s41592-023-01874-3>.
 37. Liu, Y.-T., Fan, H., Hu, J.J., and Zhou, Z.H. (2025). Overcoming the preferred-orientation problem in cryo-EM with self-supervised deep learning. *Nat. Methods* 22, 113–123. <https://doi.org/10.1038/s41592-024-02505-1>.
 38. Terwilliger, T.C., Adams, P.D., Afonine, P.V., and Sobolev, O.V. (2020). Cryo-EM map interpretation and protein model-building using iterative map segmentation. *Protein Sci.* 29, 87–99. <https://doi.org/10.1002/pro.3740>.
 39. Lawson, C.L., Kryshtafovych, A., Adams, P.D., Afonine, P.V., Baker, M.L., Barad, B.A., Bond, P., Burnley, T., Cao, R., Cheng, J., et al. (2021). Cryo-EM model validation recommendations based on outcomes of the 2019 EMDatabank challenge. *Nat. Methods* 18, 156–164. <https://doi.org/10.1038/s41592-020-01051-w>.
 40. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
 41. Zhang, Z., Xu, L., Zhang, S., Peng, C., Zhang, G., and Zhou, X. (2025). DEMO-EMol: modeling protein-nucleic acid complex structures from cryo-EM maps by coupling chain assembly with map segmentation. *Nucleic Acids Res* 53, W228–W237. <https://doi.org/10.1093/nar/gkaf416>.
 42. Zhou, X., Li, Y., Zhang, C., Zheng, W., Zhang, G., and Zhang, Y. (2022). Progressive assembly of multi-domain protein structures from cryo-EM density maps. *Nat. Comput. Sci.* 2, 265–275. <https://doi.org/10.1038/s43588-022-00232-1>.
 43. Zhang, Z., Cai, Y., Zhang, B., Zheng, W., Freddolino, L., Zhang, G., and Zhou, X. (2024). DEMO-EM2: assembling protein complex structures from cryo-EM maps through intertwined chain and domain fitting. *Briefings Bioinf.* 25, bbae113. <https://doi.org/10.1093/bib/bbae113>.
 44. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–2309. <https://doi.org/10.1093/nar/gki524>.
 45. Su, B., Huang, K., Peng, Z., Amunts, A., and Yang, J. (2026). CryoAtom improves model building for cryo-EM. *Nat. Struct. Mol. Biol.* 33, 351–361. <https://doi.org/10.1038/s41594-025-01713-3>.
 46. He, J., Lin, P., Chen, J., Cao, H., and Huang, S.-Y. (2022). Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nat. Commun.* 13, 4066. <https://doi.org/10.1038/s41467-022-31748-9>.
 47. Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., et al. (2021). Advances in xMipp for cryo-electron microscopy: From xMipp to scipion. *Molecules* 26, 6224. <https://doi.org/10.3390/molecules26206224>.
 48. Dai, M., Dong, Z., Xu, K., and Zhang, Q.C. (2023). CryoRes: local resolution estimation of cryo-EM density maps by deep learning. *J. Mol. Biol.* 435, 168059. <https://doi.org/10.1016/j.jmb.2023.168059>.
 49. Chen, Y., Binfeng, Z., Zhaoxin, G., Yiyu, H., Yifan, H., Feiwei, Q., Qinai, L., and Changmiao, W. (2024). Scunet++: Swin-unet and cnn bottleneck hybrid architecture with multi-fusion dense skip connection for pulmonary embolism ct image segmentation. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 7744–7752. <https://doi.org/10.1109/WACV57701.2024.00758>.
 50. Wang, J., and Moore, P.B. (2017). On the interpretation of electron microscopic maps of biological macromolecules. *Protein Sci.* 26, 122–129. <https://doi.org/10.1002/pro.3060>.
 51. Yonekura, K., and Maki-Yonekura, S. (2016). Refinement of cryo-EM structures using scattering factors of charged atoms. *Applied Crystallography* 49, 1517–1523.
 52. Yonekura, K., Matsuoka, R., Yamashita, Y., Yamane, T., Ikeguchi, M., Kidera, A., and Maki-Yonekura, S. (2018). Ionic scattering factors of atoms that compose biological molecules. *IUCr J* 5, 348–353. <https://doi.org/10.1107/S2052252518005237>.

53. Beton, J.G., Mulvaney, T., Cragolini, T., and Topf, M. (2024). Cryo-EM structure and B-factor refinement with ensemble representation. *Nat. Commun.* 15, 444. <https://doi.org/10.1038/s41467-023-44593-1>.
54. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. <https://doi.org/10.1038/nbt.3988>.
55. DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W., and Baker, D. (2009). Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* 392, 181–190. <https://doi.org/10.1016/j.jmb.2009.07.008>.
56. Chicco, D., Warrens, M.J., and Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* 7, e623. <https://doi.org/10.7717/peerj-cs.623>.
57. Hore, A., and Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. *2010 20th international conference on pattern recognition (IEEE)*, pp. 2366–2369.
58. Zhang, C., Freddolino, L., and Zhang, Y. (2026). A graphic and command line protocol for quick and accurate comparisons of protein and nucleic acid structures with US-align. *Nat. Protoc.* 21, 517–541. <https://doi.org/10.1038/s41596-025-01189-x>.