nature communications

Article

https://doi.org/10.1038/s41467-026-75327-8

Confidence-guided cryo-EM map optimisation with LocScale-2.0

Received: 5 October 2025

Accepted: 26 June 2026

Published online: 17 July 2026

Check for updates

Alok Bharadwaj, Reinier de Bruin & Arjen J. Jakobi

Cryogenic sample electron microscopy (cryo-EM) maps often display uneven quality, with high-resolution features coexisting alongside weak or poorly ordered regions. Such variation complicates structural interpretation, especially for heterogeneous macromolecular assemblies. Here, we present LocScale-2.0, a context-aware map optimisation framework that operates without prior knowledge of molecular structure or composition. By leveraging general expectations of electron scattering by biological macromolecules, it enhances local detail and connectivity while preserving weak but biologically relevant structural context. We further introduce LocScale-FEM, a Bayesian approximate deep-learning approach that emulates this optimisation to generate feature-enhanced maps. LocScaleFEM provides voxel-wise confidence scores that give a statistically grounded measure of reliability, are sensitive to local phase error in the feature-enhanced map, and highlight regions where interpretation warrants caution. Selected examples illustrate how confidence-guided map optimisation can aid biological interpretation and increase objectivity in cryo-EM density analysis.

Over the past decade, cryogenic electron microscopy (cryo-EM) has transformed structural biology by enabling high-resolution structure determination of macromolecular assemblies that were previously difficult or impossible to study by other structural methods. Cryo-EM now routinely delivers near-atomic resolution reconstructions across a wide range of systems and has become a cornerstone technique for exploring the structure of molecular machines in their functional states and native environments[1,2]. Current applications increasingly target systems that are structurally and compositionally complex: membrane proteins in native or near-native lipid environments[3–5], large assemblies with flexible or low-occupancy subunits[6–8], samples purified from endogenous sources[9–11] and *in-cell* structures obtained by subtomogram averaging from native cellular samples[8,12,13]. Such specimens often contain extensive structural context, including lipid/detergent shells or embedding membranes, heterogeneous modifications and peripheral scaffolds, and exhibit pronounced conformational and compositional heterogeneity. These features are frequently weak or poorly ordered, yet often biologically important. Interpreting such reconstructions, therefore, requires methods that enhance high-resolution detail without suppressing weak but informative low-resolution context, and ideally provide an estimate of confidence in regions where the map is ambiguous. Automating these processes is essential for reproducibility and for making reliable interpretation accessible beyond expert users.

In cryo-EM single-particle analysis and subtomogram averaging, three-dimensional electrostatic potential maps are assembled by averaging information from many molecules in different orientations, using low-signal-to-noise-ratio projection images or tomographic subvolumes. Because the ensemble of imaged molecules inevitably contains molecules in multiple states, the final reconstructions represent averages over conformers and compositions, even after classification into structurally relatively uniform subsets. Regions of structural variability, therefore, exhibit reduced signal-to-noise ratio (SNR) and lower effective resolution[14–16]. The result is sometimes strong spatial variation in map quality across the reconstructed volume. To compensate for the variation in resolution-dependent contrast loss, and to facilitate structural interpretation, maps are typically post-processed after reconstruction by local spectral reweighting of Fourier terms[17–21] and real-space filtering[22–25]. More recently, density-modification approaches have extended these

Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands. e-mail: a.jakobi@tudelft.nl

strategies by incorporating physical and statistical priors about expected map properties into post-processing[21,26,27] and during map reconstruction[28,29]. We refer to this spectrum of procedures collectively as map optimisation.

Deep learning methods offer alternatives that mimic the effects of local sharpening or density modification and produce optimised maps with little or no manual intervention[30–32]. While these methods are widely adopted, their lack of explicit reliability estimates introduces important caveats for interpretation. Deterministic neural networks risk introducing bias by suppressing or hallucinating signals and by enhancing artefacts without signalling their uncertainty, leaving practitioners to rely on subjective judgement when deciding which map features to interpret in ambiguous, low-contrast regions. With the ongoing exponential growth in cryo-EM structures, this poses a risk for misinterpretation, especially by new and inexperienced users[33]. A framework that integrates map optimisation with robust confidence measures would therefore represent a critical step toward more objective and reproducible cryo-EM map analysis.

Here we introduce LocScale-2.0, an automated framework for context-aware and confidence-guided map optimisation. The method extends the original LocScale approach[18] with a procedure for locally adaptive map optimisation that requires no prior model information. In addition, we develop a machine learning strategy that emulates this optimisation and provides a voxel-wise confidence metric to assess the reliability of individual features. Together, these components yield maps that preserve weak but biologically informative context, improve contrast for high-resolution detail, and support more objective structural interpretation.

Results

Multimodal cryo-EM map optimisation in LocScale-2.0

LocScale-2.0 integrates physics-informed and deep learning-based map optimisation into complementary workflows (Fig. 1). The first workflow couples model-independent local scale estimation with windowed Fourier amplitude scaling, which is conceptually related to local map sharpening in the original LocScale approach[18]. Instead of using explicit model information as a reference, it employs physical priors to estimate and locally compensate for resolution-dependent contrast loss. These priors draw on established expectations for the averaged squared structure factor of biological macromolecules[21,34,35]. The second workflow implements a deep convolutional neural network, MC-EMmerNet, that learns from pairs of unmodified input maps and maps optimised with our physics-informed procedure to predict *feature-enhanced maps* directly in real space. Unlike the physics-based method, which adjusts only the amplitudes of Fourier coefficients and leaves their phases locally unchanged, this procedure modifies both amplitudes and phases in a manner similar to density modification. To quantify predictive uncertainty, we adopt a Bayesian approximation via variational inference with Monte Carlo dropout[36]. In addition, the workflow produces a baseline map derived from local amplitude scaling of the unmodified half maps against the mean predicted map, which serves as a reference for computing a voxel-wise confidence metric sensitive to local phase discrepancies and, hence, putative bias in the feature-enhanced map (see Methods).

Physics-informed map optimisation

A limitation of most existing map-optimisation methods is their susceptibility to bias. This limitation affects reference-based sharpening approaches implemented in LocScale-1.0[18] or phenix.autosharpen[19], as well as density modification[27] and deep learning-based methods[30–32]. In reference-based methods, inaccuracies or incompleteness in the input structural model may cause suppression of density in regions not supported by the reference and obscure relevant contextual features for biological interpretation[18,21]. Deep learning-based methods[30–33] face analogous challenges, often compounded by limitations in representativeness of training data and generalisability. An additional concern with these latter methods is positive bias, in which plausible but

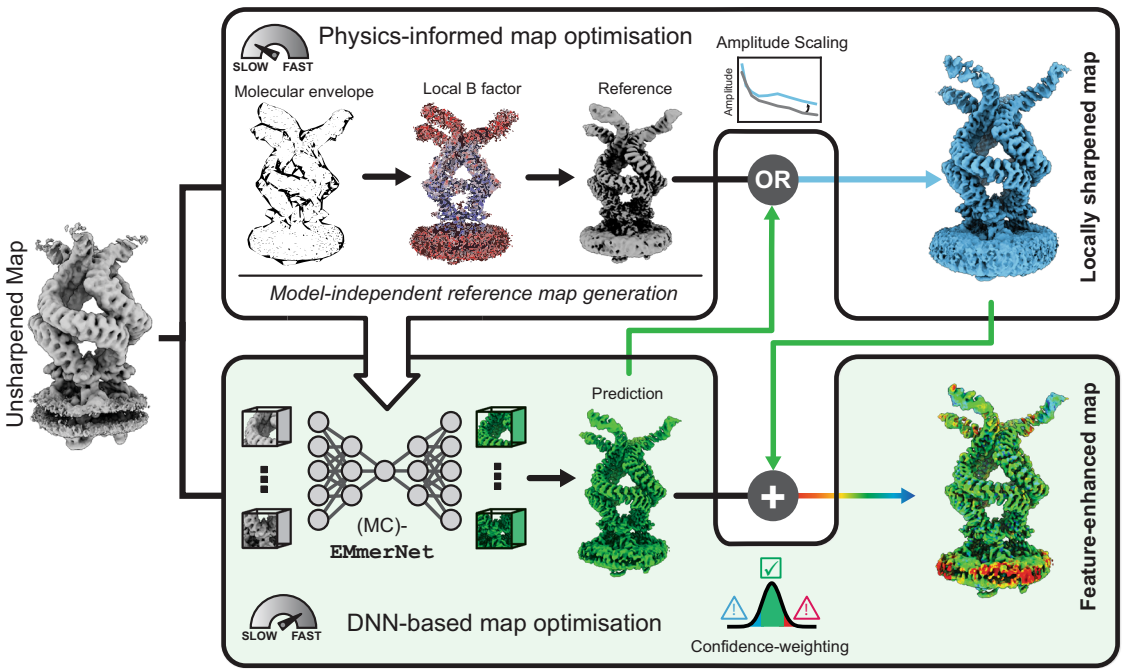

Fig. 1 | Overview of LocScale-2.0. LocScale-2.0 provides two complementary workflows for map optimisation. The physics-informed workflow leverages priors on electron-scattering properties to generate reference maps, which are then used to locally scale radially averaged Fourier amplitudes. The deep neural network (DNN)-based workflow employs EMmerNet, a network trained to emulate the physics-informed process, to directly predict an optimised map. This predicted map can either serve as a scaling reference or, when generated with a variational inference-enhanced network (MC-EMmerNet), as a directly interpretable *feature-enhanced map*. Confidence scores are calibrated against amplitude-modified baseline maps, enabling systematic identification of high- versus low-confidence regions for structural interpretation (see LocScale-FEM).

nonexistent structural details are hallucinated, potentially leading to misinterpretation.

To mitigate these issues, we developed a method for local map optimisation that considers the unbiased molecular signal of the 3D reconstruction and requires no prior structural information. The approach constrains the radially averaged Fourier amplitudes using first-principles expectations and data-driven estimates of the average squared structure factor of biological macromolecules. The spectral profile of this radial structure factor depends on molecular mass distribution, local *B* factors, and secondary structure[21,34,35]. To approximate these parameters, we first estimate the molecular volume using statistical hypothesis testing with false discovery rate (FDR) control[37] and define density consistent with a genuine molecular signal at the 1% FDR threshold (Supplementary Fig. 1a, b). Molecular mass estimates from the volume enclosed by the binarised FDR mask correlate strongly with masses computed from primary sequences and support our chosen threshold as a suitable proxy for the molecular envelope (Supplementary Fig. 1c, d). Importantly, this method reliably generates envelopes that capture density from molecular components absent from the primary sequence, enabling its application to complexes with heterogeneous or unknown composition (Supplementary Fig. 1c, d).

Within this envelope, pseudoatoms are placed randomly to initialise a pseudomodel, which is then iteratively refined against the experimental density (Supplementary Fig. 2a–c). In many practical cases, partial structural information is available from prior experimental models or structure prediction and can be exploited in a hybrid workflow that uses explicit model coordinates where available and a pseudoatomic representation in unmodelled regions (Fig. 2a–c and Supplementary Fig. 1a). This strategy ensures that all density within the molecular boundary is represented and prevents signal suppression in regions lacking atomic models.

To approximate the physical decay of the overall structure factor, atomic displacement factors (ADPs or *B* factors) are then refined using restrained refinement based on local neighbourhood averaging across an empirically determined radius (Fig. 2d, e and Supplementary Fig. 2d, e; Methods). These restraints ensure that *B* factors of pseudoatoms, which are independent and nonbonded, vary smoothly, thereby effectively reducing the number of free refinement parameters. Refined local *B* factors from pseudomodels agree closely with those from restrained ADP refinement of atomic models (Fig. 2f, g and Supplementary Fig. 3d, e), without evidence of overfitting (Fig. 2e and Supplementary Fig. 2e). The resulting *B* factor distributions follow the shifted inverse-gamma behaviour characteristic for macromolecular structures[38,39]. In several cases, we observed multimodal *B* factor distributions consistent with structural heterogeneity such as flexible loops, peripheral or low-occupancy subunits, detergent micelles, and with prior reports that local mixtures of shifted inverse-gamma distributions reflect conformational or compositional variability[39] (Supplementary Fig. 3h, i). Together, these steps provide model-independent estimates of local structure factors and their resolution-dependent falloff, from which radially averaged structure factors can be computed. Because the pseudomodel representations only approximate secondary structure, we apply empirical priors to imprint expected spectral texture of secondary structure elements from proteins and nucleic acids into the local radial structure factor over the relevant frequencies[21,29]. The spectral profiles of the locally averaged Fourier amplitudes are then used to for local map sharpening, as described previously[18]. A detailed account of the implementation of the above procedure is given in the Methods.

Comparison of model-to-map Fourier Shell Correlation (FSC) curves for 83 model-map pairs illustrates the respective strengths and limitations of atomic, pseudoatomic, and hybrid reference models for local amplitude scaling (Fig. 2h). Pseudomodels, which lack explicit stereochemical information, are well suited to capturing low-resolution scattering behaviour by representing signal containing regions not typically covered by atomic models, but they cannot fully reproduce high-resolution features. Atomic models, in contrast, provide an accurate description of high-resolution signal but tend to suppress density in weak, unmodelled or heterogeneous regions. Hybrid models combine these complementary advantages and yield the most balanced representation of the experimental density. In the following, we will therefore focus our discussion on applications of this hybrid approach for reference generation and subsequent local sharpening in LocScale-2.0.

Preserving structural context with hybrid map optimisation

To benchmark our procedure, we applied LocScale-2.0 to generate optimised maps for a set of representative single-particle reconstructions and subtomogram averages from the Electron Microscopy Data Bank (EMDB), and compared them with corresponding maps produced by two established local sharpening approaches: LocScale-1.0[18] and phenix.autosharpen[19] (see Methods for details). Across test cases, LocScale-2.0 produced maps whose high-resolution detail was generally comparable to, and in some cases improved relative to, the alternative methods, while additionally preserving and enhancing weak but informative density in regions that are flexible, disordered, or generally unmodelled such as detergent shells, lipid belts, N-linked glycans, and low-occupancy subunits (Fig. 3 and Supplementary Fig. 4). A notable feature of the optimised maps is that both high-resolution detail and weak, low-resolution contextual density remain visible at a single density threshold, allowing co-visualisation of these densities during map interpretation and facilitating direct comparison between structurally heterogeneous regions.

For example, in the *Klebsiella pneumoniae* type II secretin (EMD-0193), the LocScale-2.0 map co-visualises the plug density that occludes the lumen, a feature likely unresolved at high resolution due to its mismatch with the *C*15 symmetry of the secretin complex[40]. In the deposited half maps, this feature is only detectable after strong low-pass filtering (Fig. 3a and Supplementary Fig. 4e). In subtomogram averages of *Mycoplasma pneumoniae* 70S ribosomes (EMD-13234) from *in-cell* tomograms, the LocScale-2.0 map co-visualises heterogeneous density at the mRNA entry and A-site tRNA without loss of high-resolution detail in well-resolved regions (Supplementary Fig. 4a–c). Similarly, in Tweety homologue 2 (TTYH2; EMD-51108), LocScale-2.0 captures the largely continuous lipid/detergent belt surrounding the transmembrane region as well as the N-linked glycans, while retaining fine detail of the transmembrane helices, including individually resolved, ordered lipids (Fig. 3b and Supplementary Fig. 4d). Overall, these observations show that hybrid reference-based sharpening simultaneously enhances weak contextual features and high-resolution detail, and yields maps that represent heterogeneous complexes and assemblies in a manner more consistent with their structural context. This extends to highly heterogeneous assemblies: In apolipoprotein B100 (apoB-100)-assembled low-density lipoprotein (LDL) particles (EMD-44469), for instance, the method emphasises well-resolved density for the structured apoB-100 N-terminal domain including morphological features of its β-belt formed by amphipathic β-sheets wrapped around the surface of the core lipids of LDL, while retaining the weak density consistent with cholesteryl ester plates within the LDL core (Fig. 4a,b and Supplementary Fig. 4f). By contrast, these weak and low complexity densities are entirely suppressed by the alternative methods.

LocScale-2.0 optimised maps aid automated model building

We next asked whether the improved balance in representing high- and low-resolution features in LocScale-2.0-optimised maps can aid automated methods for atomic model building. To test this, we used the cryo-EM model-building software ModelAngelo[41] on a benchmark set of 50 map-model pairs from the EMDB and PDB, each determined at better than 4.0 Å FSC resolution. For each case, an optimised map was generated using the hybrid LocScale-2.0 procedure, and

https://doi.org/10.1038/s41467-026-75327-8

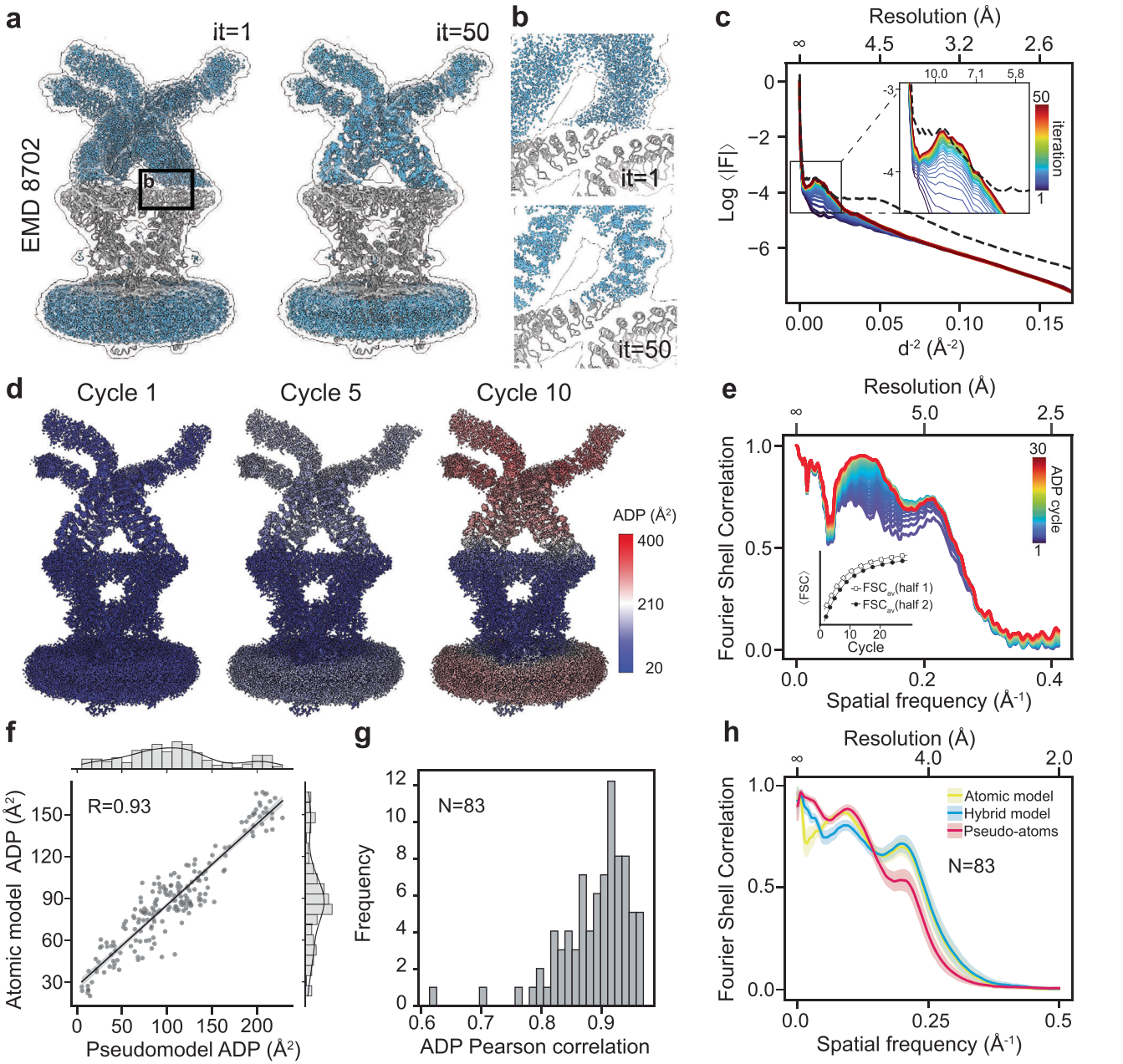

Fig. 2 | Pseudomodel generation and ADP refinement. a Example for pseudoatom placement and refinement for *Drosophila melanogaster* NOMPC (EMDB: EMD-8702) with partial prior model information (hybrid model). Pseudoatoms (blue) are uniformly placed within the molecular boundary defined by the FDR confidence mask except for regions for which prior model information is available (grey). Models are shown for the start (iteration 1) and end (iteration 50) of the positional refinement. **b** Close-up of a region at the interface between explicit and pseudomodel. **c** Radially averaged Fourier amplitudes of model maps computed throughout the iterations of pseudomodel refinement. Inset: close-up of low-to-medium resolution region defining the overall molecular envelope. **d** Atomic displacement factors (ADP) mapped onto atomic coordinates shown for different snapshots across the ADP refinement cycles. **e** The evolution of model-to-map Fourier shell correlation (FSC) is shown for a series of ADP refinement cycles. The insets shows the half-map validation displaying the average integrated FSC for work and test maps as a function of refinement cycle. **f** Correlation of atomic model and pseudomodel ADPs averaged in 25 Å windows. *R* denotes the Pearson correlation coefficient; the shaded band indicates the 95% confidence interval. **g** Distribution of Pearson correlation coefficients for window-averaged ADPs between atomic and pseudomodel representations across 83 EMDB entries. **h** Average model-to-map Fourier Shell Correlation computed for 83 EMDB entries and their corresponding PDB-deposited models, pseudomodels and hybrid models. The error band indicates the 95% confidence intervals. Maps were resampled to uniform pixel size for comparison. Source data are provided as a Source Data file.

ModelAngelo was run both on these maps and on the corresponding EMDB-deposited maps (Supplementary Fig. 5). Overall, ModelAngelo achieved higher sequence coverage from LocScale-2.0 maps than from deposited maps, both before (82%) and after (72%) pruning of short peptide fragments (Supplementary Fig. 5a–c). On average, these models also exhibited smaller predicted backbone root-mean-squared deviation (r.m.s.d.) for both protein and nucleic acid components. Interestingly, improvements were observed for targets both represented and not represented in the ModelAngelo training dataset. In several cases, the improvements were substantial. For example, in the *Azospirillum brasilense* glutamate synthase (EMD-10106), the LocScale-2.0 map enabled building of 1306 additional residues compared to the deposited map (Supplementary Fig. 5d, e). Given that this $\alpha_6\beta_4$ complex comprises 11,108 residues in total, this represents an increase of

https://doi.org/10.1038/s41467-026-75327-8

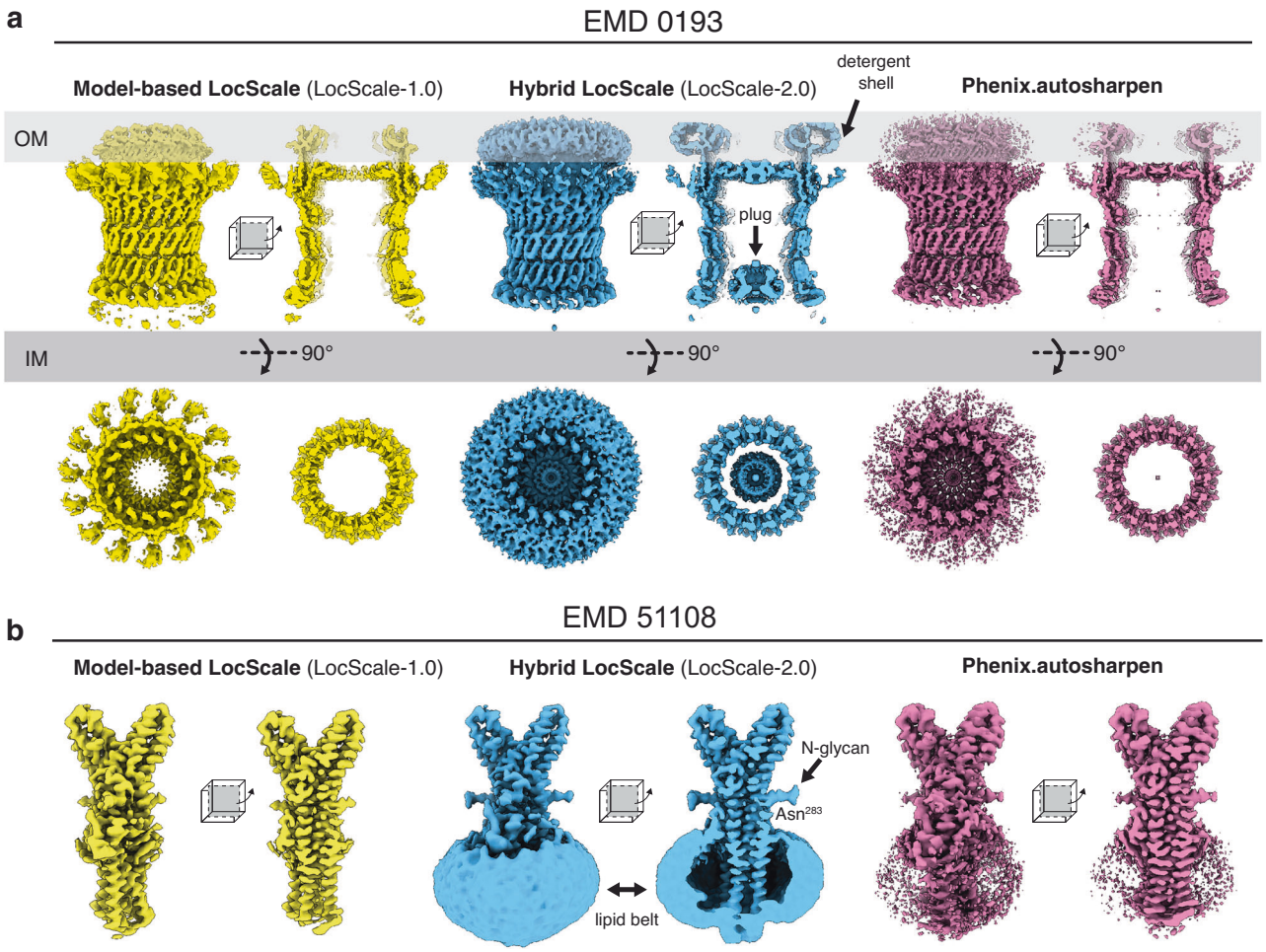

Fig. 3 | Comparison of LocScale-2.0 hybrid mode map optimisation with other local sharpening methods. a Map optimisation of *Klebsiella pneumoniae* type II secretion system outer membrane complex (EMDB: EMD-0193) using model-based sharpening with LocScale-1.0 (left), hybrid reference sharpening in LocScale-2.0 (middle) and phenix.autosharpen (right). Approximate locations of the outer (OM) and inner (IM) membrane are indicated. Arrows mark the visible detergent shell of the outer membrane-spanning segment and the section views reveal the central disordered plug occluding the lumen. **b** Optimised maps for human Tweety homologue 2 (EMDB: EMD-51108). The arrows mark density for N-linked glycans and the lipid belt around the the hydrophobic cavity.

~12% in modelled sequence. In 17% of the benchmark cases, however, the EMDB-deposited maps produced higher sequence coverage than LocScale-2.0 maps and in 11% the sequence coverage was equal, indicating that improvements are not uniform across the dataset. The sequence recall (i.e. the fraction of residues correctly identified compared to the PDB-deposited model) computed from the 50 test map-model pairs was comparable for EMDB-deposited and LocScale-2.0 maps (Supplementary Fig. 5f–j). Since ModelAngelo was trained on generally *B*-factor–augmented maps, these results suggest that additional gains might be realised by retraining ModelAngelo or related deep learning-based model building tools directly on LocScale-2.0 maps, thereby extending their reach to lower resolution. Complementary to retraining, LocScale-2.0 optimisation could also be applied at the inference step, so that model building benefits from context-aware maps during prediction. In practice, these strategies may be combined to maximise performance.

Feature-enhanced maps from uncertainty-aware deep learning

We next investigated whether the physics-informed map optimisation procedure can be replicated and accelerated using machine learning. To this end, we trained a 3D U-Net on pairs of unmodified half maps and hybrid LocScale-2.0 maps to predict optimised maps directly from unmodified input densities. Predictions from this network, which we term MC-EMmerNet, emulate the sharpening achieved by the physics-informed approach (Supplementary Fig. 6a–d) at substantially reduced computational cost. For example, for a large and context-rich structure such as the apolipoprotein B100 assemblies (EMD-44469; Fig. 4), network prediction speeds up map optimisation about 16-fold compared to the physics-informed procedure (14 min vs. 4 h; Supplementary Table 1). However, deterministic networks provide point estimates without a measure of predictive uncertainty, which would allow identification of regions where the network's output may be unreliable. To overcome this limitation, we implemented variational inference in the 3D U-Net using Monte Carlo dropout, which retains dropout layers at inference time and samples multiple stochastic forward passes[36]. This produces an ensemble of predictions from which voxel-wise means and variances can be computed, yielding both the average predicted map (which we call Feature-Enhanced Map, or LocScale-FEM) and a spatially resolved estimate of predictive uncertainty (Fig. 5a, Supplementary Fig. 6e–g). Rather than using the raw voxel-wise variance from these predictions as the uncertainty metric, we use the ensemble mean as a reference to generate a phase-conservative baseline map. This baseline map is obtained by locally scaling the amplitudes of the experimental half maps to match the predicted amplitude distribution while preserving the experimental phases. Analysis of the predictive variance during training shows rapid

https://doi.org/10.1038/s41467-026-75327-8

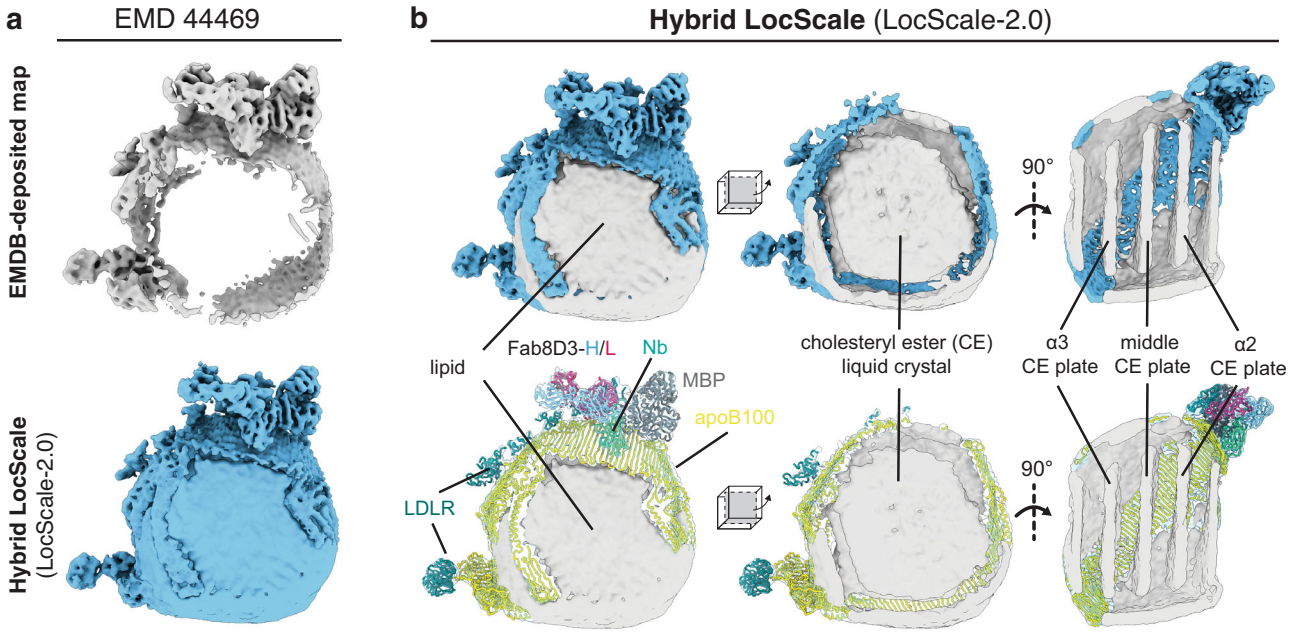

Fig. 4 | Preservation of contextual structure by LocScale-2.0. a Cryo-EM density maps of Apolipoprotein B 100 bound to LDL receptor and legobody (EMDB: EMD-44469 – top; deposited[6]) and after hybrid reference sharpening in LocScale-2.0 (bottom). **b** LocScale-2.0 optimised map EMD-44469. Top: Density assigned to protein is shown in blue; unassigned contextual lipid density is shown in grey. Bottom: The deposited atomic model (PDB: 9bdt) fitted into the optimised map with colours as labelled. Cross-sections show the cholesteryl ester plates in the interior of the low-density lipoprotein (LDL) particle.

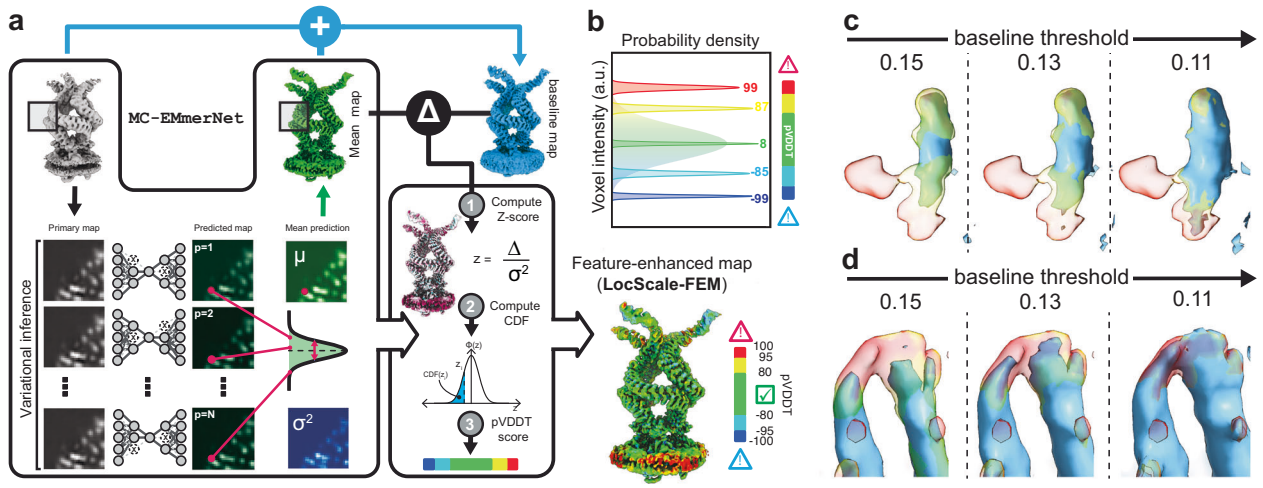

Fig. 5 | Confidence-guided map optimisation and feature-enhanced maps (LocScale-FEM). a Schematic overview of Bayesian-approximate variational inference in MC-EMmerNet (left), and computation of voxel-wise pVDDT confidence scores by statistical comparison to a baseline reference (right). The far right shows the resulting feature-enhanced map (FEM), weighted by local confidence. **b** Illustration of predicted and scaled central voxel intensity distributions for five voxels with varying pVDDT scores. For clarity, only a single predicted distribution is shown; in practice, a unique distribution is generated per voxel. **c**,**d** Comparison of baseline and LocScale-FEM maps for plant phytochrome A (EMDB: EMD-35691). Shown are two regions with positive pVDDT scores (threshold: 0.15), overlaid with baseline maps displayed at varying thresholds, with (**c**) showing an example where the baseline map never recovers the positively flagged density. Source data are provided as a Source Data file.

stabilisation, indicating reliable convergence of the posterior predictive distribution and supporting its use as a calibration target (Supplementary Fig. 6n–p). The statistical difference of voxel intensities between the predicted ensemble distribution and the baseline reference then allows computing a probability-derived confidence measure in the form of a rescaled z-score that we term the predicted voxel-wise difference test (pVDDT) score. Since Fourier amplitudes have been spectrally matched in baseline and feature-enhanced maps, the intensity residuals are dominated by phase errors in the predicted map (see Methods), and the pVDDT metric captures these errors in an interpretable way (Fig. 5a and Supplementary Fig. 6l–k).

Interpreting pVDDT confidence scores

While the variances from Monte Carlo dropout provide a first estimate of predictive uncertainty, we found that they were systematically smaller than the actual prediction errors, measured as the absolute difference between the predicted means and the corresponding baseline values (Supplementary Fig. 6h). This mismatch reflects the common tendency of Monte Carlo dropout to underestimate uncertainty and produce overconfident predictions[42,43]. We therefore used isotonic regression calibration to adjust the predicted standard deviations so that they matched the empirically observed residuals (Supplementary Fig. 6i). The resulting calibrated uncertainties

establish a statistically robust baseline for voxel-wise pVDDT scores, which quantify both the probability and direction of signal deviations in the predicted maps (Fig. 5b, c and Supplementary Fig. 6l–m). The magnitude of the score reflects the statistical strength of the deviation. For interpretability, we define three ranges of pVDDT values. Scores between −80 and +80 fall within the reliable range, corresponding approximately to the central 80% confidence interval of the standard normal distribution. This threshold provides a practical compromise between the metric's sensitivity and robustness. Scores below −80 or above +80 indicate statistically significant deviations, while values beyond ± 95 mark voxels where deviations are highly unlikely given the calibrated uncertainty. As feature-enhanced maps tend to accentuate secondary structure, negative pVDDT confidence scores often appear along the grooves of αhelices. High positive pVDDT scores (greater than +95) may represent network hallucination of density features not present in the phase-conservative baseline map (Fig. 5c and Supplementary Fig.7a), but in other cases may correspond to real features (Fig. 5d and Supplementary Fig. 7b). Thus, pVDDT confidence scores help draw attention to regions that merit closer inspection, and joint examination of feature-enhanced and baseline maps provides an efficient strategy for guiding interpretation. One limitation of our procedure is that pVDDT values are referenced to the feature-enhanced maps for visualisation; as a result, strongly negative scores corresponding to genuine densities that are completely suppressed in the prediction will not be reported. Feature-enhanced maps should therefore always be evaluated alongside the baseline map to ensure that such features are not overlooked.

Applications of LocScale-FEM in structure analysis

We examined how pVDDT confidence scores can guide the interpretation of feature-enhanced maps from LocScaleFEM. As a first example, we applied LocScale-FEM to the unmodified half maps of EMD-13234 to re-examine the *in-cell* subtomogram averaged reconstruction of M. pneumoniae 70S ribosomes[8]. For comparison, we also generated the corresponding map using DeepEMhancer, which has been trained on LocScale-1.0 (model-based) maps[30]. The LocScale-FEM map effectively recovered weak density at levels comparable to the equivalent hybrid LocScale-2.0 map (cf. Supplementary Fig. 4a–c), including density at the mRNA entry site, A-site tRNA and the nascent peptide (Fig. 6a). pVDDT confidence scores supported the reliability of most of these features, with the exception of the nascent peptide density[8]. While density near the mRNA entry site remained heterogeneous and uninterpretable, the A-site density in the LocScale-FEM map supported confident tRNA placement (Fig. 6b). In contrast, DeepEMhancer maps showed pronounced downscaling of the same regions, leaving them essentially uninterpretable (Fig. 6a, b).

A similar trend was observed for subtomogram averages of tubular lattices formed by the capsid (CA) and nucleocapsid (NC) domains of Human T-cell leukaemia virus type I (EMD-17929)[44]. LocScale-FEM produced interpretable density for dimers within the irregular hexamers formed by the CA C-terminal domain (CA-CTD), which in the EMDB-deposited map is visible only after strong low-pass filtering (Fig. 6c; shown at 25 Å low-pass) and entirely suppressed in the corresponding EMReady map. It has been hypothesised that CA-CTD interactions buttress the NTD scaffold but leave the lattice free to adopt diverse curvatures[44], highlighting that co-visualisation of ordered elements (here, the NTD lattice) and less-ordered components (here, CTD-mediated contacts) in such reconstructions may help inform mechanistic models.

Another class of structures for which co-visualisation of contextual density is highly informative is transmembrane or membrane-associated complexes. An illustrative case is the *in-cell* structure of mitochondrial ATP synthase from *Polytomella*[45]. The deposited consensus map (EMD-19999) captures ATP synthase dimers on cristae ridges, with the best-resolved region (4 Å local resolution) at the central-peripheral stalk. The F_1 and F_O subcomplexes are less well resolved due to averaging over different rotational states[45], but remain visible after low-pass filtering (Fig. 6d). Map optimisation with DeepEMhancer and EMReady entirely suppressed density outside the central peripheral stalk, whereas the LocScale-FEM map co-visualised the peripheral stalks alongside both F_1 and F_O subcomplexes and the strongly curved cristae membrane, without loss of detail in the peripheral stalks compared to the deposited or otherwise density-modified maps. The LocScale-FEM map also distinguished individual α-helical densities for several of the ten outer c-ring subunits embedded in the membrane (Fig. 6e, f), with pVDDT scores suggesting high confidence for these features. This case exemplifies how LocScale-FEM can preserve the broader architectural context of complex membrane-bound assemblies. Similarly, LocScale-FEM maps of the yeast COPII inner coat assembled on membrane tubules (EMD-15949)[5,46] allowed confident placement of the Sar1 amphipathic helix inserting into the outer leaflet of the membrane (Supplementary Fig. 7d). The map clearly visualises the parallel insertion of the amphipathic helices, which has been posited to reinforce curvature through cooperative interactions across the lattice[47]. Additional examples are provided in Supplementary Fig. 7e, f.

The above examples suggest that LocScale-FEM is less prone to inadvertent signal suppression, preserving contextual density features that are often obliterated by other deep learning-based map optimisation methods. To quantify this effect, we measured volumetric recall relative to an unbiased signal envelope derived from the original half-maps using FDR thresholding[48]. For each case, LocScale-FEM, DeepEMhancer, and EMReady maps were normalised by rescaling their positive voxels to the unit interval [0,1]. We then calculated volumetric recall across varying thresholds to generate recall curves for each method. Figure 6g, h illustrates an example for this procedure. Applied to 26 randomly selected EMDB reconstructions, LocScale-FEM showed significantly higher integrated volume recall (AURC) compared to DeepEMhancer and EMReady, confirmed by permutation testing ($p < 0.0004$ and $p < 0.0002$, respectively; Fig. 6i). To further assess whether the recovered density corresponds to reproducible experimental signal, we additionally evaluated cross-FSCs between feature-enhanced maps and independent half maps. In this analysis, feature-enhanced maps were generated from one-half of the dataset and correlated with the independent half of the dataset that was not used for their generation. Thus, any increase in cross-FSC can only arise from improved contrast for genuine features supported by both datasets. Across the analysed examples, we generally observe improved cross-FSC values for LocScale-FEM maps (Supplementary Fig. 7g, h), consistent with the interpretation that the recovered density reflects genuine signal rather than prediction-induced artefacts. This analysis establishes that the advantages observed in individual examples generalise across diverse datasets.

Interpretation of small-molecule ligands

We next examined the utility of feature-enhanced maps and pVDDT scores for interpreting small-molecule ligands and other non-protein components. As a representative case, we analysed the GPCR–Gi complex of human relaxin family peptide receptor 4 (RXFP4) bound to compound 4, an amidrazone scaffold (EMD-33888)[49,50]. LocScale-FEM reproduced the ligand density observed in the deposited (DeepEMhancer-processed) map, with visible features corresponding to both the hydroxyindole and chlorophenyl substituents (Fig. 7a). Analysis of pVDDT confidence scores, however, revealed important distinctions between these moieties. While the scores supported a reliable interpretation of the hydroxyindole group, strongly positive scores around the chlorophenyl substituent indicated possible overemphasis of density at this site relative to an amplitude-modified baseline. This differential confidence scoring demonstrates how our

https://doi.org/10.1038/s41467-026-75327-8

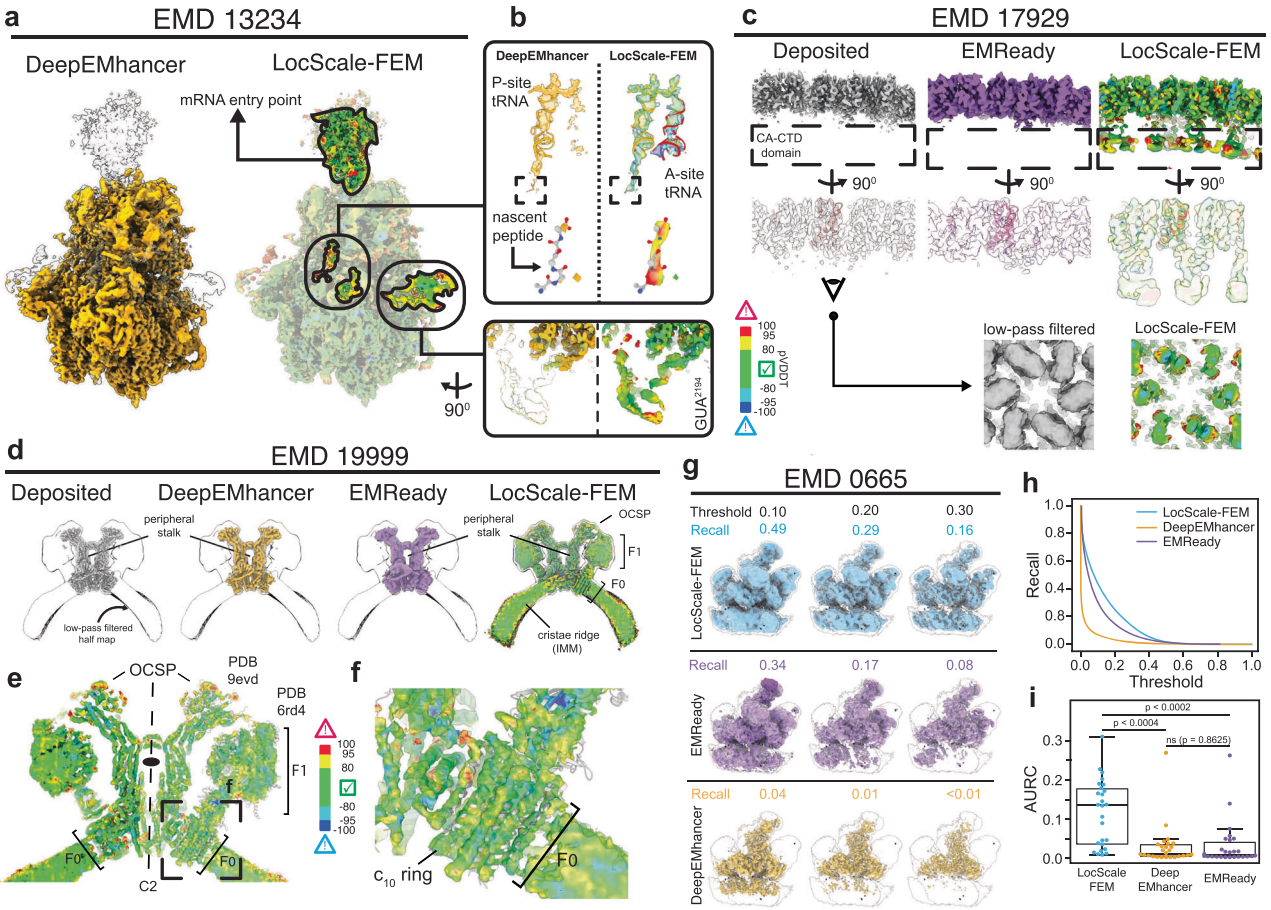

Fig. 6 | Comparison of LocScale-FEM and alternative deep learning approaches for contextual structure recovery. a LocScale-FEM and DeepEMhancer maps of *Mycoplasma pneumoniae* 70S ribosome (EMDB: EMD-13234). Selected contextual densities are highlighted in the LocScale-FEM map. The 1% FDR contour computed from the unmodified half maps is outlined for both maps. **b** Close-up views from (**a**), showing improved interpretability of densities for the A-site tRNA, nascent peptide, and a peripheral region. **c** Comparison of author-deposited, EMReady-optimised, and LocScale-FEM maps for subtomogram averages of the Human T-cell leukaemia virus type I immature capsid (CA; EMDB: EMD-17929). The capsid protein C-terminal domain (CA-CTD) is highlighted. A close-up shows the CA-CTD viewed from below, comparing the 25 Å low-pass filtered EMDB map with the LocScale-FEM map. **d** Author-deposited and optimised maps for in situ subtomogram averages of the peripheral stalks of *Polytomella* ATP synthase (EMDB: EMD-19999). All optimised maps are superimposed on the low-pass filtered contour of the EMDB-deposited map. Contextual densities for the F_1 and F_O ATP synthase subunits, oligomycin sensitivity-conferral protein (OCSP), and the crista membrane are indicated. **e** Close-up of the C_2-symmetric ATP synthase dimer in the LocScale-FEM map. The right half of the map is shown transparent and superposed on PDB 6rd4 (F_1 and F_O) and PDB 9evd (peripheral stalk). The pVDDT colour bar is shown for reference. **f** Zoom of the F_O subunit showing resolved helical densities for the C_{10} ring. **g** Volume recall for EMD-0665 computed for different thresholds for LocScale-FEM, EMReady and DeepEMhancer optimised maps. **h** Volume recall curves for LocScale-FEM, DeepEMhancer and EMReady maps generated for EMD-0665. **i** Area under the recall curve (AURC) for LocScale-FEM, DeepEMhancer, and EMReady, plotted for randomly selected EMDB maps ($N = 26$). Statistical significance of mean differences was evaluated by two-sided permutation testing with 10,000 resamples. Pairwise p-values were: LocScale-FEM vs DeepEMhancer, $p = 0.00039$; LocScale-FEM vs EMReady, $p = 0.00019$; and DeepEMhancer vs EMReady, $p = 0.8659$. The box plot centre line indicates the median, the box bounds indicate the first and third quartiles (Q1 and Q3), and the whiskers extend to the most extreme data points within 1.5 × the interquartile range (IQR = Q3–Q1). Points outside this range are shown as outliers. Source data are provided as a Source Data file.

approach can provide additional guidance for ligand interpretation by flagging regions where structural features may be less reliable or potentially artifactual. To validate this approach across different ligand chemistries, we examined β-galactosidase bound to the inhibitor 2-phenylethyl β-D-thiogalactopyranoside (PETG; EMD-7770)[51], where ligand pose can be challenging[52]. The LocScale-FEM analysis showed high confidence density for the well-ordered pyranose moiety while appropriately downscaling density for the somewhat more flexible phenylethyl substituent (Supplementary Fig. 8b). This pattern suggests that our confidence scoring effectively distinguishes between well-ordered structural elements and regions with inherently higher mobility or weaker density. LocScale-FEM demonstrated density preservation across diverse ligand types and retained ligand density also in cases where other deep-learning methods tend to downscale density excessively (Fig. 7b, c and Supplementary Fig. 8c–e). Despite this apparent robustness, fundamental limitations remain. Ligands are strongly underrepresented in the EMDB, and their structural diversity spans an enormous chemical space, which places inevitable constraints on any deep-learning model's ability to generalise across all molecular scaffolds. This limitation is exemplified in our analysis of ordered lipid densities in the Connexin-43 gap junction channel (EMD-33394)[53] (Supplementary Fig. 8a). While such densities are clearly visible in the baseline LocScale maps, they were substantially down-weighted in LocScale-FEM as well as in DeepEMhancer and EMReady predictions. This case shows that all three deep learning models may systematically suppress the genuine signal. Taken together, these examples illustrate both the potential and the limitations of LocScale-FEM for non-protein components. Confidence scores can provide useful discrimination in some cases, while in others, suppression of genuine features persists. In practice, joint inspection of feature-enhanced maps, confidence scores, and baseline references will therefore be essential for interpretation.

https://doi.org/10.1038/s41467-026-75327-8

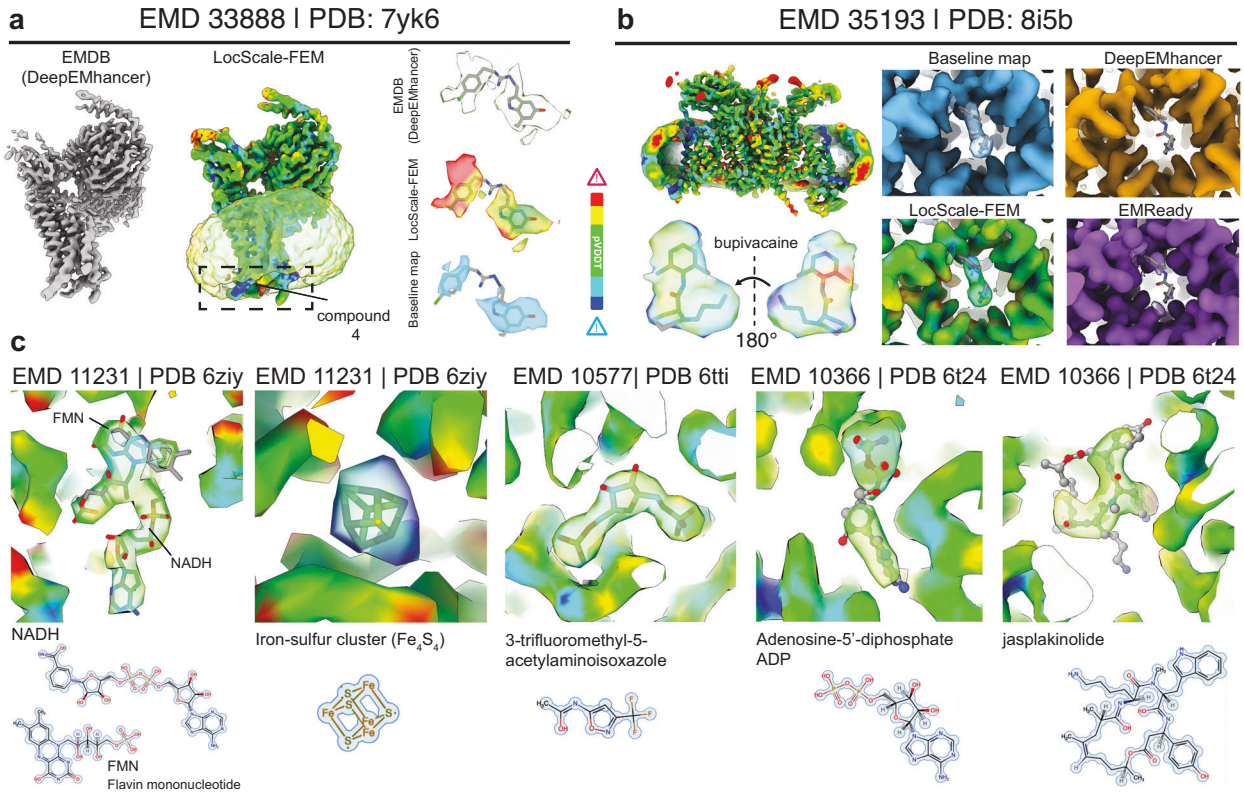

Fig. 7 | Small-molecule ligand interpretation with LocScale-FEM. a Author-deposited (DeepEMhancer) and LocScale-FEM map of (RXFP4)-Gi bound to 1-[2-(4-chlorophenyl)ethyl]-3-[(7-ethyl-5-oxidanyl-1H-indol-3-yl)methylideneamino]guanidine (compound 4; EMDB: EMD-33888). Ligands superposed on their respective densities are shown on the right. **b** LocScale-FEM map for human voltage-gated sodium channel (Nav1.7; EMDB: EMD-35193) in complex with bupivacaine[75]. The close-ups show ligand-superposed densities for LocScale-FEM, DeepEMhancer and EMReady-optimised maps. The LocScale-FEM baseline map is also shown for comparison. **c** LocScale-FEM maps for *Thermus thermophilus* complex I (EMDB: EMD-11231)[76], pyruvate kinase PKM (EMDB: EMD-10577)[77], and stabilised F-actin filaments (EMDB: EMD-10366)[78] containing a chemically diverse array of ligands. 2D chemical representations of the respective ligands are drawn below.

Discussion

In this work, we demonstrate new capabilities of LocScale2.0 map optimisation for recovering weak and contextual structure in reconstructions affected by conformational or compositional heterogeneity, both from cryo-EM single-particle analysis and cryo-ET subtomogram averaging. LocScale-2.0 integrates physics-informed map sharpening with uncertainty-aware deep learning to address limitations of current map post-processing methods. Both hybrid sharpening and feature-enhanced maps generated by LocScale-FEM consistently mitigate negative bias in regions that remain poorly resolved in consensus maps or retain residual heterogeneity after extensive classification. These improvements are most pronounced for low-resolution regions and contextual density, while performance in the highest-resolution areas varies due to limitations of current physical priors and residual trade-offs between preserving low-resolution context and maximally accentuating high-resolution detail. Importantly, LocScaleFEM introduces calibrated voxel-wise confidence scores that enable systematic identification of ambiguous regions, thereby reducing reliance on experience and subjective judgement. Although not all retained low-resolution density will be chemically interpretable, the ability of LocScale-2.0 maps to co-visualise high-resolution details and weaker contextual features at a single density threshold represents an important practical advantage for structural analysis and the visual communication of complex macromolecular assemblies.

Several remaining limitations warrant consideration. pVDDT confidence estimates are referenced to the feature-enhanced maps, and strongly negative deviations corresponding to genuine features that are completely suppressed during prediction may go undetected. This is most relevant for small-molecule ligands and other components underrepresented in the training set and underscores the need to inspect feature-enhanced maps alongside their baseline references. Likewise, further improvements may be possible with extensions to the physical priors beyond general scattering behaviour. Finally, although throughout we refer to significant differences between predicted and baseline maps as phase errors in the context of confidence estimates, it should be emphasised that the experimental Fourier phases themselves are subject to error. Since the true, error-free structure is unknown, the metric should therefore be seen as a conservative measure erring on the side of caution. Put differently, it should be recognised that some apparent errors may reflect genuine improvements over noisy observations but cannot be distinguished as such.

Future extensions could refine physical priors through explicit anisotropy corrections to account for direction-dependent blurring, and incorporate expectations on map behaviour in real-space, such as solvent flattening[54–56], histogram matching[57,58], and the use of symmetry-related correlations. Several of these ideas have been investigated in previous methods, but their combination remains unexplored[26,27]. For LocScale-FEM, while Monte Carlo dropout offers a practical means to estimate uncertainty, Bayesian alternatives that are currently computationally too demanding may in the future yield better calibrated or more accurate uncertainty estimates[59–61]. By introducing a statistical measure of reliability, LocScale2.0 represents a step toward automated cryo-EM map-interpretation frameworks that transparently report predictive uncertainty. More broadly, integrating confidence-aware optimisation into automated model building and refinement could increase the accuracy and sequence recall of these methods for lower-resolution maps. As cryoEM expands to

increasingly complex systems, such approaches will be important for ensuring rigorous and objective structural analysis.

Methods

Programme outputs

LocScale-2.0 retains all intermediate and final results generated during map optimisation. The main output directory contains the optimised maps and auxiliary files required for interpretation, while intermediate files are stored in a dedicated processing directory. For the physics-informed workflow (Hybrid LocScale), the principal output is the optimised LocScale map. For LocScale-FEM, the outputs include the feature-enhanced map, the baseline map, a voxel-wise pVDDT confidence map, and a ChimeraX script for visualising confidence scores mapped onto the density. An overview of the most relevant output files produced by both workflows is provided in Supplementary Fig. 9.

Estimation of molecular envelope

The regions in a map that correspond to macromolecule and associated components (such as lipid, glycans or ligands) as opposed to solvent (i.e., the molecular envelope or boundary) are identified using FDR-controlled statistical thresholding[37]. Sampling windows located near the box edges are used to estimate the background noise distribution; we note that the procedure will fail if the noise sampling box overlaps with part of the signal, as can happen for filaments or subtomogram averages in which the signal extends to (some of) the box edges. In these cases, noise centre coordinates of sampling windows can be individually supplied. The size of the noise box is set to 20 pixels or 10% the unit cell dimension, whichever is higher. FDR confidence maps are binarised at 1% FDR to yield confidence masks.

Estimation of total scattering mass

Using the molecular volume enclosed by the confidence mask, the number of required pseudoatoms is calculated assuming constant values for the mean protein density and the average atomic mass of the pseudoatoms:

$$N_{atoms} = \frac{N_A \cdot \rho_{avg} \cdot V_{mask}}{\overline{m}_{pseudo}}$$

where N_A denotes Avogadro's constant, ρ_{avg} (1.35 g/cm^3 = 0.813 Da/Å^3) is the mean protein density determined from partial specific volumes using hydrodynamic compressibility[62,63] and V_{mask} is the volume of the FDR mask (for pure pseudomodels) or the difference mask (for hybrid models). The average atomic mass for pseudoatoms $\overline{m}_{pseudo}$ was estimated by computing the ratio of molecular mass to the number of non-hydrogen atoms over 119 atomic models from a reference dataset, which is derived from the training data used for DeepEMhancer. The median ratio was 13.2 Da. For hybrid model construction, we estimate the unmodelled volume by computing the difference between the FDR-based confidence mask and a binary mask derived from the available atomic model. The atomic model mask is obtained by placing a spherical mask of radius 3 Å at each atomic coordinate, excluding hydrogen atoms. To minimise artefacts arising from subtraction, the difference mask is smoothed with a uniform filter across three voxels and subsequently binarised at a threshold of 0.5. The calculated number of pseudoatoms is initially placed at random within the boundaries of the FDR confidence or difference mask. They are further randomly displaced by a magnitude of 0.5 pixels to prevent two atoms from overlapping.

Positional refinement of pseudoatoms

The positions of pseudoatoms are then refined using a regularised gradient-based molecular dynamics scheme, with the cryo-EM density map acting as an external potential. This refinement integrates density-gradient forces and interatomic Lennard–Jones (LJ) interactions using a simple Verlet integration with friction damping[64]. For simplicity, pseudoatom mass is set to unity, so that forces directly translate to accelerations. Note that the atoms corresponding to the input atomic models remain unchanged during this operation. The total force on a pseudoatom is computed as:

$$\mathbf{F} = \mathbf{F}_{\text{grad}} + \mathbf{F}_{\text{LJ}} + \mathbf{F}_{\text{fric}}$$

The attraction to high-density regions is modelled by the local gradient of the cryo-EM map:

$$\mathbf{F}_{\text{grad}} = g \cdot \nabla_{\rho},$$

where ∇_{ρ} is the local gradient of the density map, and g is a scaling factor to scale the gradient force magnitude relative to that of the Lennard-Jones term (default $g = 10$). This approach is conceptually similar to other density-biased refinement methods[65,66].

To avoid pseudoatom clustering and promote spatial regularity, a pairwise Lennard–Jones potential is applied between pseudoatoms within a 3 Å neighbourhood. To stabilise the refinement and assist convergence, we apply linear friction:

$$\mathbf{F}_{\text{fric}} = -\mu\mathbf{v},$$

where $\boldsymbol{v}$ is the instantaneous velocity of the pseudoatom and $\mu = 10$ is the friction coefficient. We implemented velocity Verlet integration to propagate the positions and velocities of pseudoatoms. By default, the refinement runs for 50 iterations with a timestep of $\Delta t = 0.05$ (in arbitrary units), which we found to reach convergence in most cases.

Once the positions of all pseudoatoms are refined, pseudoatoms are merged with the input atomic model to form a single structure. The pseudoatoms are introduced as a separate chain, with the sequence number of the added chain recorded for downstream purposes. For consistency, identifiers of the original atomic model chains always begin with "A". If no input model is supplied, LocScale-2.0 assumes the entire map is unmodelled. Only the pseudoatomic model is retained for further processing. The overall workflow for constructing the pseudomodel and hybrid model is illustrated in Supplementary Fig. 1a.

ADP refinement of pseudoatoms and hybrid models

LocScale 2.0 performs ADP (isotropic *B*-factor) refinement with REFMAC (v5.8)[67] via the Servalcat wrapper (v0.4.100)[68], and supports explicit atomic models, pure pseudomodels, and hybrid models. In all cases, positional refinement is disabled (refi bonly), all atomic *B* factors are initialised to 40 Å, and refinement is carried out against the unmodified map up to the Nyquist frequency.

REFMAC restrains ADPs using a weighted Kullback–Leibler divergence between atom pairs with the isotropic *B* factors as the optimised parameters, where the weights depend on distance and bonding connectivity[67]. For hybrid models and pure pseudoatomic models, the absence of chemical bonds for the pseudomodel component necessitates a different strategy. Refinement is performed one cycle at a time, and each cycle is followed by spherical averaging of *B* factors within a 3 Å radius to enforce local smoothness and reduce the effective number of free parameters. In hybrid models, this averaging is restricted to the pseudoatomic chain(s) identified during model construction; coordinates belonging to explicit atomic chains are left unchanged except for their refined *B* factors. Pure pseudoatomic models lack valid chemical topology and are therefore refined in REFMAC with unconstrained refinement (refi type unre) while still applying the post-cycle spherical averaging step described above. This additional keyword is not required for hybrid models. By default, the sequence of a single refinement cycle followed by local spherical averaging is repeated ten times; the number of iterations can be

adjusted by the user. Independent half-map refinement is performed in all cases and automatically monitored for overfitting.

Symmetry averaging

If the input map has non-*C*1 point-group symmetry, reference maps simulated from pseudomodels or hybrid models do not necessarily preserve the original map's point-group symmetry. To correct for this, symmetry averaging on the simulated density using a user-defined point group symmetry. The desired point group symmetry can be specified using the -sym command-line argument.

Amplitude scaling with generalised radial profiles

For each voxel position within the molecular envelope mask, cubic subvolumes are extracted both from the unmodified experimental map and from the corresponding reference map. The spherically averaged power spectra of these cubes are then compared, and the radial profile of the experimental cube is rescaled to match that of the reference cube. This procedure follows the principle of LocScale-1.0 but is extended to accommodate pseudomodels and hybrid models.

For voxels that lie within the FDR or difference mask corresponding to pseudoatomic regions, additional corrections are introduced to incorporate average scattering profiles. First, the frequency limits for applying these corrections are defined: the low-resolution limit corresponds to the transition from a continuous-volume description to a discrete i.i.d. distribution, estimated from the local molecular volume[35], while the high-resolution limit is set by the half-map FSC. Within these limits, scattering amplitude deviations are computed by rescaling the scattering profile to the *B* factor of the reference profile, constructing an exponential fall-off function with the same *B* factor, and subtracting the scaled scattering amplitudes from this function. Finally, the computed deviations are added back to the reference radial profile within the relevant frequency range. This ensures that the local scaling not only matches the reference map but also reflects expected scattering behaviour in regions represented by pseudomodels.

Monte Carlo (MC-) EMmerNet

Network architecture. We implemented a convolutional 3D U-Net, following a previous design[30,69] with several modifications (Supplementary Fig. 6a; Table 7). The architecture consists of an encoder–decoder with skip connections, using strided convolutions for downsampling and transposed convolutions for upsampling. PReLU was chosen as the activation function throughout the network.

Dataset. The network was trained on cubic subvolumes extracted from the average of unmodified EMDB-deposited half maps and their corresponding LocScale-2.0 reference maps. Cubes were 32 Å on each side and sampled from the training maps on a 24 Å grid. The dataset was derived from the DeepEMhancer map–model pairs[30], excluding cases where reliable confidence masks could not be generated. EMDB entries were grouped according to 30% sequence identity clusters, and the training, validation, and test sets were constructed such that they are mutually exclusive under this criterion. To avoid bias during evaluation, only one representative per sequence cluster was included in the validation and test sets. Hybrid model maps were available for 114 entries, of which 89 were used for training and 14 for validation (86:14 split). A separate set of 11 maps was reserved for testing and uncertainty calibration. In addition, 24 EMDB additional maps were specifically selected from the EMDB to assess the ability of LocScale-2.0 to enhance contextual density and retain non-protein components. These examples were also used to validate pVDDT scores (Supplementary Fig. 7a–c, g–h).

Data augmentation. Training cubes were classified as signal- or noise-containing based on the resampled confidence mask. A 4:1 signal-to-noise ratio was enforced. Data augmentation included random rotations (0–2π), sharpening or blurring corresponding to random *B* factors between 0 and 400 Å^2, and Gaussian low-pass filtering.

Training. Input intensities were standardised to zero mean and a standard deviation of 0.1, and maps were resampled to 1 Å per pixel. The Adam optimiser was used with a constant learning rate of 10^{-4}, and mean absolute error (MAE) served as the loss function. Networks were trained for 15 epochs.

Prediction. At the inference, maps were preprocessed identically to the training data. Subvolumes of size 32 Å were extracted every 16 Å and restricted to the confidence mask region. A default batch size of 8 per GPU was used, with multi-GPU data parallelism handled via `MirroredStrategy`.

Uncertainty estimates were obtained by enabling dropout during inference (training=True), yielding multiple stochastic forward passes. Voxel-wise means and standard deviations were computed across the ensemble. Overlapping cube predictions were merged using a weight matrix to reduce boundary artefacts. Final mean and standard deviation volumes were resampled to the original pixel size and saved as `.mrc` files.

Uncertainty prediction

Multiple stochastic forward passes (MC samples) are performed to generate a distribution of predictions. From these, the voxel-wise mean and variance are computed as follows:

$$\hat{Y} = \frac{1}{T}\sum_{n=1}^{T} Y_n$$

$$\sigma^2 = \frac{1}{T}\sum_{n=1}^{T}\left(\hat{Y} - Y_n\right)^2$$

where z is the predictive mean, σ^2 is the predictive variance, and *T* is the number of MC samples.

The mean and variance volumes are stored for each voxel in the map. During post-processing, these volumes are reintegrated into their original positions on the global grid and corrected for overlaps.

Uncertainty calibration

The uncertainty estimates from Monte Carlo dropout are typically uncalibrated and therefore not directly suitable for statistical inference without further adjustment[42,70]. In a perfectly calibrated regression model, the predicted uncertainties should match the expected residuals, defined as the absolute difference between the prediction and the true value. As shown in Supplementary Fig. 6h, the raw uncertainties predicted via Monte Carlo dropout tend to be significantly smaller than the actual residuals.

For a given voxel *i*, the prediction interval is defined as:

$$\hat{Y}_i \pm z_\alpha \cdot \sigma_i$$

where $\hat{Y}_i$ is the predicted mean, and σ_i is the standard deviation estimated via Monte Carlo dropout. The constant z_α is the z-score corresponding to a desired confidence level α.

A well-calibrated model produces uncertainty estimates that correspond to empirical probabilities. Specifically, for any confidence level $\alpha \in [0,1]$, the probability that the true value Y_i lies within the interval defined by the predictive mean Y_i and scaled uncertainty $\pm z_\alpha \sigma_i$ should be approximately α:

$$P(\hat{Y}_i - z_\alpha \cdot \sigma_i \le \hat{Y}_i \le \hat{Y}_i + z_\alpha \cdot \sigma_i) \approx \alpha, \forall \alpha \in [0,1]$$

In this study, we employ isotonic regression to calibrate the predicted uncertainties[43,71] by learning a monotonically increasing function I that maps predicted standard deviations to expected residuals.

The calibration procedure uses a set of N voxel-level predictions and their corresponding ground-truth values:

$$S = \left\{ \left(Y_1, \hat{Y}_1\right), \left(Y_2, \hat{Y}_2\right), \ldots, \left(Y_N, \hat{Y}_N\right) \right\}$$

Residuals are computed for each voxel as:

$$R_i = \left| Y_i - \hat{Y}_i \right|$$

For the set $U = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ of predicted standard deviations from Monte Carlo dropout, the isotonic regression function I is learned by minimising the squared error between residuals and calibrated uncertainties:

$$\text{Minimise} : \sum_{i=1}^{N} \left(R_i - I(\sigma_i)\right)^2$$

$$\text{subject to} : I(\sigma_i) \le I(\sigma_{i+1}) \text{for all } i \in \{1, 2, \ldots, N-1\}.$$

The resulting function I can then be applied to all predicted uncertainties to obtain calibrated standard deviations for statistical inference.

For the calibration procedure, 11 maps from the test dataset were selected. For each map, a mean and variance map was generated via Monte Carlo dropout. The mean map was used as input to generate the corresponding LocScale map, serving as the ground-truth reference. Calibration was performed voxel-wise by extracting the predicted mean and variance from the MC-EMmerNet predictions and the baseline values from the LocScale maps.

Since most voxels in a map correspond to background (i.e., regions with little or no signal), including them in the calibration process would bias the results and lead to underestimation of uncertainties. These background voxels typically exhibit very low predicted variance. However, simply discarding low-variance voxels is not trivial, as some signal-bearing voxels, particularly those with low intensity, also fall in this range. To address this, we apply a variance cutoff of 0.005: voxels with predicted variance below this threshold are excluded from the calibration process. While this introduces some miscalibration for low-intensity signal voxels, it significantly improves overall reliability. In practice, miscalibration tends to appear most prominently when low-threshold contours are used to visualise the feature-enhanced map. For consistent interpretation, we recommend using a contour threshold of 0.1 or higher when displaying predictions alongside pVDDT scores, as the variance is well calibrated in this range. Additionally, predicted variances are capped at a maximum value observed within the dataset. The isotonic regression model is also bounded accordingly: any input variance exceeding this cap is mapped to a fixed upper value. The effective calibration range is visualised in Supplementary Fig. 6i.

Within this valid range, the calibrated variance serves as a reliable estimate of prediction error. This calibration is essential for the statistical interpretation of voxel-wise deviations and facilitates the detection of regions where the predicted distribution significantly diverges from the experimental reference. Such regions can then be analysed using standard hypothesis testing techniques.

Relationship of pVDDT score and phase differences between feature-enhanced and baseline maps

To evaluate voxel-wise significance of differences between predicted and baseline maps, we consider the data in Fourier space where density maps are represented by their series of Fourier coefficients. The structure factors of observed and predicted maps are:

$$F_{\text{obs}}(\boldsymbol{s}) = \left|F_{obs}(\boldsymbol{s})\right| e^{i\phi_{obs}(\boldsymbol{s})}$$

$$F_{\text{pred}}(\boldsymbol{s}) = \left|F_{pred}(\boldsymbol{s})\right| e^{i\phi_{pred}(\boldsymbol{s})}$$

LocScale modifies the original observation by adjusting the amplitudes of its Fourier coefficients to match the radially averaged spectrum of the prediction while, within the scaling window, preserving their phases:

$$F_{scaled}(\boldsymbol{s}) = k(\|\boldsymbol{s}\|) \cdot |F_{obs}(\boldsymbol{s})| e^{i\phi_{obs}(\boldsymbol{s})}$$

with:

$$k(\|\boldsymbol{s}\|) = \sqrt{\frac{\sum_{s' \in \Delta(\boldsymbol{s})} |F_{pred}(\boldsymbol{s}')|^2}{\sum_{s' \in \Delta(\boldsymbol{s})} |F_{obs}(\boldsymbol{s}')|^2}}$$

The local Fourier-space difference between the prediction and amplitude-scaled observation is:

$$F_{\text{diff}}(\boldsymbol{s}) = F_{pred}(\boldsymbol{s}) - F_{scaled}(\boldsymbol{s}) = |F_{pred}(\boldsymbol{s})| e^{i\phi_{pred}(\boldsymbol{s})} - k(\|\boldsymbol{s}\|) \cdot |F_{obs}(\boldsymbol{s})| e^{i\phi_{obs}(\boldsymbol{s})}$$

This can be rearranged to highlight the role of phase and amplitude differences:

$$F_{\text{diff}}(\boldsymbol{s}) = [|F_{pred}(\boldsymbol{s})| - k(\|\boldsymbol{s}\|) \cdot |F_{obs}(\boldsymbol{s})|] e^{i\phi_{obs}(\boldsymbol{s})} + |F_{pred}(\boldsymbol{s})| [e^{i\phi_{pred}(\boldsymbol{s})} - e^{i\phi_{obs}(\boldsymbol{s})}]$$

Here, the first term reflects residual amplitude differences between Fourier coefficients and the second term captures phase differences. When amplitude matching is effective (i.e., $\left|F_{pred}(\boldsymbol{s})\right| \approx k(\|\boldsymbol{s}\|) \cdot \left|F_{obs}(\boldsymbol{s})\right|$), the first term becomes small, and the difference density depends primarily on the phase difference (Supplementary Fig. 6l,m):

$$F_{\text{diff}}(\boldsymbol{s}) \approx \left|F_{\text{pred}}(\boldsymbol{s})\right| \left[e^{i\phi_{pred}(\boldsymbol{s})} - e^{i\phi_{obs}(\boldsymbol{s})}\right]$$

The inverse discrete Fourier transform over the local window grid G gives the real-space difference density:

$$\rho_{\text{diff}}(\boldsymbol{r}) = \frac{1}{N^3} \sum_{s \in G} F_{\text{diff}}(\boldsymbol{s}) e^{2\pi i \boldsymbol{s} \cdot \boldsymbol{r}}$$

At the central voxel ($\mathbf{r} = 0$):

$$\rho_{\text{diff}}(\boldsymbol{r} = 0) = \frac{1}{N^3} \sum_{s \in G} F_{diff}(\boldsymbol{s}) \approx \frac{1}{N^3} \sum_{s \in G} \left|F_{pred}(\boldsymbol{s})\right| \left[e^{i\phi_{pred}(\boldsymbol{s})} - e^{i\phi_{obs}(\boldsymbol{s})}\right]$$

This demonstrates that, under the condition that amplitude differences are small, the central voxel intensity of the local difference map is primarily determined by the accumulated phase differences between Fourier terms of predicted and unmodified maps across spatial frequencies. These differences manifest as measurable voxel intensity variations in real space that can be statistically evaluated, and justify the use of the intensity difference of the central voxel $\rho_{\text{diff}}(0)$ as a phase-sensitive metric.

We use $\rho_{\text{diff}}(0)$ to compute a confidence metric based on a voxel-wise difference density test (*pVDDT*) score. A calibrated standard deviation σ_{cal} is first assigned to each voxel of the predicted map using a learned uncertainty from the Monte Carlo dropout predictions. The voxel-wise z-score for the difference between predicted and baseline

map is then computed as:

$$z_0 = \frac{\rho_{\text{diff}}(\boldsymbol{r}=0)}{\sigma_{\text{cal}}}$$

Here, σ_{cal} represents the expected difference between predicted and baseline densities. The z-score, therefore, measures how strongly the observed deviation exceeds this expectation.

The probability of observing a z-score less than or equal to z_0 under the null hypothesis H_0: $z_0 = 0$ is given by:

$$\Phi(z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{t^2}{2}} dt$$

where $\Phi(z)$ is the cumulative distribution function of a standard normal distribution. To reflect both the magnitude and sign of the deviation in a bounded and interpretable scale, the result is linearly rescaled to the range [−100,100]:

$$pVDDT = 200\Phi(z_0) - 100$$

Positive pvDDT scores indicate voxels where predicted voxel intensity exceeds the baseline reference (possible overestimation), negative values indicate underestimation, and larger absolute values reflect greater statistical significance.

Automated model building with ModelAngelo

All ModelAngelo model building was done using ModelAngelo v.1.0.1 with default parameters. For each of the 50 map model pairs in our ModelAngelo test dataset, FASTA files containing primary sequences of all complex components were pulled from the PDB. If applicable, protein and nucleic acid sequences were subsequently split over separate FASTA sequence files. For each model-map pair, ModelAngelo was run using both the EMDB-deposited map and the LocScale-2.0 optimised map. To test the accuracy of sequence prediction, residue identities for each protein residue in the predicted structure were compared to those of the closest residue in the deposited PDB reference structure based on the coordinates of their respective C-α atoms. Recall statistics were then computed for all residues from each map-model pair in our dataset.

Reporting summary

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

Data availability

All published data sets used for training, validation, testing and analysis were accessed and dowloaded from the EMDB and PDBe with accession codes specified in the Figure legends and Supplementary Tables. Data generated in this study are available from the 4TU.ResearchData repository (https://doi.org/10.4121/7fac93cc-d8f3-4f14-85a5-c7f1daa5c455)[72]. Pretrained MC-EMmerNet models (v0.3) can be accessed via Zenodo (https://zenodo.org/records/8211668)[73]. Scripts used to generate figures shown in this study can be found at [https://github.com/cryotud/publications]. Source data are provided with this paper.

Code availability

LocScale-2.0 is open-source software released under a BSD license and is available at [https://github.com/cryotud/locscale][74], where detailed documentation is also provided.

References

1. Nogales, E. & Mahamid, J. Bridging structural and cell biology with cryo-electron microscopy. *Nature* **628**, 47–56 (2024).
2. Hutchings, J. & Villa, E. Expanding insights from in situ cryo-EM. *Curr. Opin. Struct. Biol.* **88**, 102885 (2024).
3. Waltz, F. et al. In-cell architecture of the mitochondrial respiratory chain. *Science* **387**, 1296–1301 (2025).
4. Coupland, C. E. et al. High-resolution electron cryomicroscopy of V-ATPase in native synaptic vesicles. *Science* **385**, 168–174 (2024).
5. Hutchings, J. et al. Structure of the complete, membrane-assembled COPII coat reveals a complex interaction network. *Nat. Commun.* **12**, 2034 (2021).
6. Reimund, M. et al. Structure of apolipoprotein B100 bound to the low-density lipoprotein receptor. *Nature* **638**, 829–835 (2025).
7. Leung, M. R. et al. Structural specializations of the sperm tail. *Cell* **186**, 2880–2896.e17 (2023).
8. Xue, L. et al. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610**, 205–211 (2022).
9. Verbeke, E. J., Mallam, A. L., Drew, K., Marcotte, E. M. & Taylor, D. W. Classification of single particles from human cell extract reveals distinct structures. *Cell Rep.* **24**, 259–268.e3 (2018).
10. Ho, C.-M. et al. Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods* **17**, 79–85 (2020).
11. Su, C.-C. et al. and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. *Nat. Methods* **18**, 69–75 (2021).
12. Zila, V. et al. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021).
13. Zhang, X. et al. Molecular mechanisms of stress-induced reactivation in mumps virus condensates. *Cell* **186**, 1877–1894.e27 (2023).
14. Palmer, C. M. & Aylett, C. H. S. Real space in cryo-EM: the future is local. *Acta Crystallogr. Sect. D. Struct. Biol.* **78**, 136–143 (2022).
15. Read, R. J., Millán, C., McCoy, A. J. & Terwilliger, T. C. Likelihood-based signal and noise analysis for docking of models into cryo-EM maps. *Acta Crystallogr. Sect. D. Struct. Biol.* **79**, 271–280 (2023).
16. Forsberg, B. O., Shah, P. N. M. & Burt, A. A robust normalized local filter to estimate compositional heterogeneity directly from cryo-EM maps. *Nat. Commun.* **14**, 5802 (2023).
17. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
18. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *eLife* **6**, e27131 (2017).
19. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. D: Struct. Biol.* **74**, 545–559 (2018).
20. Ramírez-Aportela, E. et al. Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics* **36**, 765–772 (2020).
21. Bharadwaj, A. & Jakobi, A. J. Electron scattering properties of biological macromolecules and their use for cryo-EM map sharpening. *Faraday Discuss.* **240**, 168–183 (2022).
22. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
23. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
24. Vilas, J. L. et al. MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* **26**, 337–344.e4 (2018).
25. Ramlaul, K., Palmer, C. M. & Aylett, C. H. S. A local agreement filtering algorithm for transmission EM reconstructions. *J. Struct. Biol.* **205**, 30–40 (2019).
26. Spiegel, M., Duraisamy, A. K. & Schröder, G. F. Improving the visualization of cryo-EM density reconstructions. *J. Struct. Biol.* **191**, 207–213 (2015).

27. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V., Adams, P. D. & Read, R. J. Density modification of cryo-EM maps. *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 912–925 (2020).
28. Kimanius, D. et al. Exploiting prior knowledge about biological macromolecules in cryo-EM structure determination. *IUCrJ* **8**, 60–75 (2021).
29. Gilles, M. A. & Singer, A. A molecular prior distribution for Bayesian inference based on Wilson statistics. *Computer Methods Prog. Biomed.* **221**, 106830 (2022).
30. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 1–8 (2021).
31. He, J., Li, T. & Huang, S.-Y. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nat. Commun.* **14**, 3217 (2023).
32. Maddhuri Venkata Subramaniya, S. R., Terashi, G. & Kihara, D. Enhancing cryo-EM maps with 3D deep generative networks for assisting protein structure modeling. *Bioinformatics* **39**, btad494 (2023).
33. Berkeley, R. F., Cook, B. D. & Herzik, M. A. Machine learning approaches to cryoEM density modification differentially affect biomacromolecule and ligand density quality. *Front. Mol. Biosci.* **11**, 04885 (2024).
34. Morris, R. J., Blanc, E. & Bricogne, G. On the interpretation and use of ⟨|E|2⟩(d*) profiles. *Acta Crystallogr. Sect. D Struct. Biol.* **60**, 227–240 (2004).
35. Singer, A. Wilson statistics: derivation, generalization and applications to electron cryomicroscopy. *Acta Crystallogr. Sect. A Found. Adv.* **77**, 472–479 (2021).
36. Gal, Y. & Ghahramani, Z. Dropout as a bayesian approximation: representing model uncertainty in deep learning. In *Proc. 33rd International Conference on Machine Learning*. 1050–1059 (PMLR, 2016).
37. Beckers, M., Jakobi, A. J. & Sachse, C. Thresholding of cryo-EM density maps by false discovery rate control. *IUCrJ* **6**, 18–33 (2019).
38. Masmaliyeva, R. C. & Murshudov, G. N. Analysis and validation of macromolecular B values. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 505–518 (2019).
39. Masmaliyeva, R. C., Babai, K. H. & Murshudov, G. N. Local and global analysis of macromolecular atomic displacement parameters. *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 926–937 (2020).
40. Chernyatina, A. A. & Low, H. H. Core architecture of a bacterial type II secretion system. *Nat. Commun.* **10**, 5437 (2019).
41. Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
42. Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. In *Proc. 34th International Conference on Machine Learning*. Vol. 70, ICML'17, 1321–1330 (Sydney, NSW, Australia, ICML, 2017).
43. Kuleshov V., Fenner N. & Ermon S. Accurate uncertainties for deep learning using calibrated regression. In *Proc. of the 35th International Conference on Machine Learning*. PMLR 80 (2018).
44. Obr, M. et al. Distinct stabilization of the human T cell leukemia virus type 1 immature Gag lattice. *Nat. Struct. Mol. Biol.* **32**, 268–276 (2025).
45. Dietrich, L., Agip, A.-N. A., Kunz, C., Schwarz, A. & Kühlbrandt, W. In situ structure and rotary states of mitochondrial ATP synthase in whole Polytomella cells. *Science* **385**, 1086–1090 (2024).
46. Zivanov, J. et al. A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife* **11**, e83724 (2022).
47. Hutchings, J., Stancheva, V., Miller, E. A. & Zanetti, G. Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nat. Commun.* **9**, 4154 (2018).
48. Beckers, M., Palmer, C. M. & Sachse, C. Confidence maps: statistical inference of cryo-EM maps. *Acta Crystallogr. Sect. D. Struct. Biol.* **76**, 332–339 (2020).
49. DeChristopher, B. et al. Discovery of a small molecule RXFP3/4 agonist that increases food intake in rats upon acute central administration. *Bioorg. Med. Chem. Lett.* **29**, 991–994 (2019).
50. Chen, Y. et al. Ligand recognition mechanism of the human relaxin family peptide receptor 4 (RXFP4). *Nat. Commun.* **14**, 492 (2023).
51. Bartesaghi, A. et al. Atomic resolution Cryo-EM structure of βGalactosidase. *Structure* **26**, 848–856.e3 (2018).
52. Lawson, C. L. et al. Outcomes of the EMDataResource cryo-EM ligand modeling challenge. *Nat. Methods* **21**, 1340–1348 (2024).
53. Lee, H.-J. et al. Conformational changes in the human Cx43/GJA1 gap junction channel visualized using cryo-EM. *Nat. Commun.* **14**, 931 (2023).
54. Wang, B.C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112 (1985).
55. Cowtan, K. D. & Main, P. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr. Sec. D Struct. Biol.* **49**, 148–157 (1993).
56. Terwilliger, T. C. Reciprocal-space solvent flattening. *Acta Crystallogr. Sect. D Struct. Biol.* **55**, 1863–1871 (1999).
57. Zhang, K. Y. J. & Main, P. Histogram matching as a new density modification technique for phase refinement and extension of protein molecules Acta Crystallogr. *Sect. A Foundations of Crystallogr.* **46**, 41–46 (1990).
58. Cowtan, K. Recent developments in classical density modification. *Acta. Crystallogr. Sect. D Biol. Crystallogr.* **66**, 470–478 (2010).
59. Minderer, M. et al. Revisiting the calibration of modern neural networks. *Advances in Neural Information Processing Systems.* **34**, 15682–15694 (2021).
60. B. Lakshminarayanan, A. Pritzel and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proc. 31st International Conference on Neural Information Processing Systems, NIPS'17*, pp. 6405–6416, Red Hook, NY, USA, (NIPS, 2017).
61. C. Blundell, J. Cornebise, K. Kavukcuoglu and D. Wierstra. Weight Uncertainty in Neural Network. In *Proc. 32nd International Conference on Machine Learning*, pages1613–1622. (PMLR, 2015).
62. Kuntz, I. & Kauzmann, W. Hydration of proteins and polypeptides. *Adv. Protein Chem.* **ume 28**, 239–345 (1974).
63. Squire, P. G. & Himmel, M. E. Hydrodynamics and protein hydration. *Arch. Biochem. Biophys.* **196**, 165–177 (1979).
64. Verlet, L. Computer "Experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103 (1967).
65. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
66. Topf, M. et al. Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16**, 295–307 (2008).
67. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
68. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 1282–1291 (2021).
69. O. Ronneberger, P. Fischer and T. Brox. U-Net: convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi, editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, (MICCAI, 2015).
70. Laves, M.-H., Ihler, S., Kortmann, K.-P. & Ortmaier, T. Calibration of Model Uncertainty for Dropout Variational Inference, [2020]. arXiv:2006.11584 [cs].7
71. Chakravarti, N. Isotonic median regression: a linear programming approach. *Math. Oper. Res.* **14**, 303–308 (1989).

72. Bharadwaj, A. et al. Data underlying the publication: Confidence-guided cryo-EM map optimisation with LocScale-2.0. Version 1. 4TU.ResearchData dataset. https://doi.org/10.4121/7fac93cc-d8f3-4f14-85a5-c7f1daa5c455.v1 (2025).
73. Bharadwaj, A., de Bruin, R. M. M. & Jakobi, A. J. LocScale-EMmerNet deep learning models for contrast optimisation of cryo-EM maps. *Zenodo*, 0.3, https://doi.org/10.5281/zenodo.8211668 (2022).
74. Bharadwaj, A. & Jakobi, A. J. LocScale 2.0 – confidence-weighted cryoEM map optimisation. *Zenodo*, 2.3.1, https://doi.org/10.5281/zenodo.15488220 (2025).
75. Wu, Q. et al. Structural mapping of Nav1.7 antagonists. *Nat. Commun.* **14**, 3224 (2023).
76. Gutiérrez-Fernández, J. et al. Key role of quinone in the mechanism of respiratory complex I. *Nat. Commun.* **11**, 4135 (2020).
77. Saur, M. et al. Fragment-based drug discovery using cryo-EM. *Drug Discov. Today* **25**, 485–490 (2020).
78. Pospich, S., Merino, F. & Raunser, S. Structural effects and functional implications of phalloidin and jasplakinolide binding to actin filaments. *Structure* **28**, 437–449.e5 (2020).

Acknowledgements

We thank Keitaro Yamashita and Tom Burnley for valuable discussions, and Agnel Joseph for help with integrating LocScale-2.0 in CCP-EM Doppio. We are grateful to Ambroise Desfosses, Benedikt Junglas and the participants of the CCP-EM Icknield model-building workshops for testing early versions of the method and for useful feedback. We acknowledge the use of computational resources of the DelftBlue supercomputer provided by the Delft High Performance Computing Centre.

Author contributions

A.B. and A.J.J. designed the method. A.B. implemented the algorithm with contributions from R.B. A.B. and A.J.J. analysed the data and wrote the manuscript. A.J.J. conceived and supervised the project.

Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC StG 852880 to AJ), the TU Delft AI initiative and the Collaborative Computational Project for Electron Cryo-Microscopy (CCP-EM).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-75327-8.

Correspondence and requests for materials should be addressed to Arjen J. Jakobi.

Peer review information *Nature Communications* thanks Mark Herzik, Jr., José-Maria Carazo, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.