Uncovering Hidden Functional States in Cryo-EM Datasets with EMPROVE

Supplementary information

This PDF file includes:
- Sensitivity of scoring functions to subtle image variations
- Software Architecture and Computational Efficiency
- Supplementary Results
- Supplementary References
- Supplementary Figures 1 to 10

Sensitivity of scoring functions to subtle image variations

The Structural Cross-correlation Index (SCI) was compared to three widely used scoring methods—Cross Correlation (CC), Normalised Mutual Information (NMI), and Structural Similarity Index (SSIM)—to evaluate their sensitivity in distinguishing subtle structural variations within images. This assessment was conducted using classical image-processing features to illustrate the capability of each method to discern minor but biologically meaningful differences (Fig. S1).

Cross-correlation (CC) measures the similarity between two images by computing the normalised sum of products of mean-centred pixel intensities, and it is a widely applied measure in image analysis[1]. Normalised Mutual Information (NMI) assesses image similarity based on the statistical dependence between discretized intensity histograms of two images[2]. The Structural Similarity Index (SSIM) quantifies perceptual image similarity by comparing luminance, contrast, and structural information[3], making it particularly effective for images with subtle differences perceivable by the human visual system.

The illustrative example in Fig. S1 suggests that SCI better discriminates against localised structural differences compared to CC, NMI, and SSIM. CC and NMI show limited sensitivity, particularly for subtle distortions affecting specific regions, while SSIM captures global intensity and contrast differences but is less responsive to fine local variations. SCI shows greater sensitivity to small, localised structural changes—such as the altered eye in Image 2—making it well suited for tasks such as particle selection and 3D class reassignment in cryo-EM data processing.

Software Architecture and Computational Efficiency

EMPROVE is implemented as a hybrid architecture combining computationally efficient C++ core routines with a high-level Python interface designed for workflow orchestration and parallel execution. The core algorithms, including SCI, directional derivatives, Fourier amplitude equalisation, and per-particle scoring functions, are written in C++ for optimal numerical performance and compiled into shared libraries. These are accessed via Python bindings, enabling flexible integration into user workflows without sacrificing execution speed.

To accelerate processing of large datasets, EMPROVE supports parallelisation through a native multiprocessing approach implemented in Python. Particle scoring and reconstruction tasks are automatically distributed across multiple cores using MPI-enabled routines, allowing efficient execution on multi-core workstations or high-performance computing (HPC) environments. This is particularly critical for large cryo-EM datasets where scoring and selection need to be applied across hundreds of thousands to millions of particles. Performance is further optimised by detecting previously computed reconstructions and avoiding redundant processing.

The reconstruction and local resolution estimation steps are currently delegated to RELION, which introduces a performance bottleneck due to external I/O and serial processing constraints. Nevertheless, EMPROVE's internal pipeline efficiently handles batch execution of reconstruction subsets, enabling iterative resolution-driven thresholding strategies without excessive overhead. Future developments aim to address this dependency by incorporating native reconstruction tools directly within the EMPROVE framework.

The modular design also facilitates extensibility, with session managers and bash wrappers enabling creation of selection and classification workflows using simple command-line interfaces. This structured, script-driven approach ensures reproducibility while simplifying job submission on remote servers or clusters.

Supplementary Results

EMPIAR-10308: a comparative evaluation of scoring-based particle ranking strategy for optimised selection in Cryo-EM

For benchmarking scoring functions, we used the serotonin receptor dataset EMPIAR-10308. This dataset was selected due to its minimal overall structural flexibility, except in the ligand-binding region, and its large number of particles. These characteristics make EMPIAR-10308 particularly suitable for assessing particle selection methods. Furthermore, EMPIAR-10308 includes a nanodisc, added by the original investigators to mimic the native phospholipid bilayer. The nanodisc exhibits flexibility, variable size, and a heterogeneous density distribution, which complicates particle selection and makes it an ideal test case for evaluating scoring functions under variable particle conditions.

We evaluated our Structural Cross-correlation Index (SCI; see Methods) against established scoring functions and compared EMPROVE's selection method to the CryoSieve algorithm[4]. To isolate the impact of selection strategies, we used identical particle counts rather than optimising particle numbers. Specifically, we selected 10,000 particles (0.7% of the deposited stack), matching the count used in the original EMPIAR-10308 study for their *ab initio* reference map. This approach ensured that observed differences were attributed to selection strategies rather than dataset size variability.

Comparison of particle selection strategies in Cryo-EM: EMPROVE versus CryoSieve

We further assessed EMPROVE's performance by comparing its selected particle set to that produced by CryoSieve. Weevaluated both the scoring function and the Euler Equalisation-ranking architecture (Fig. 1A, step 3). As before, 10,000 particles were selected to ensure consistency with the original dataset and isolate the impact of particle quality.

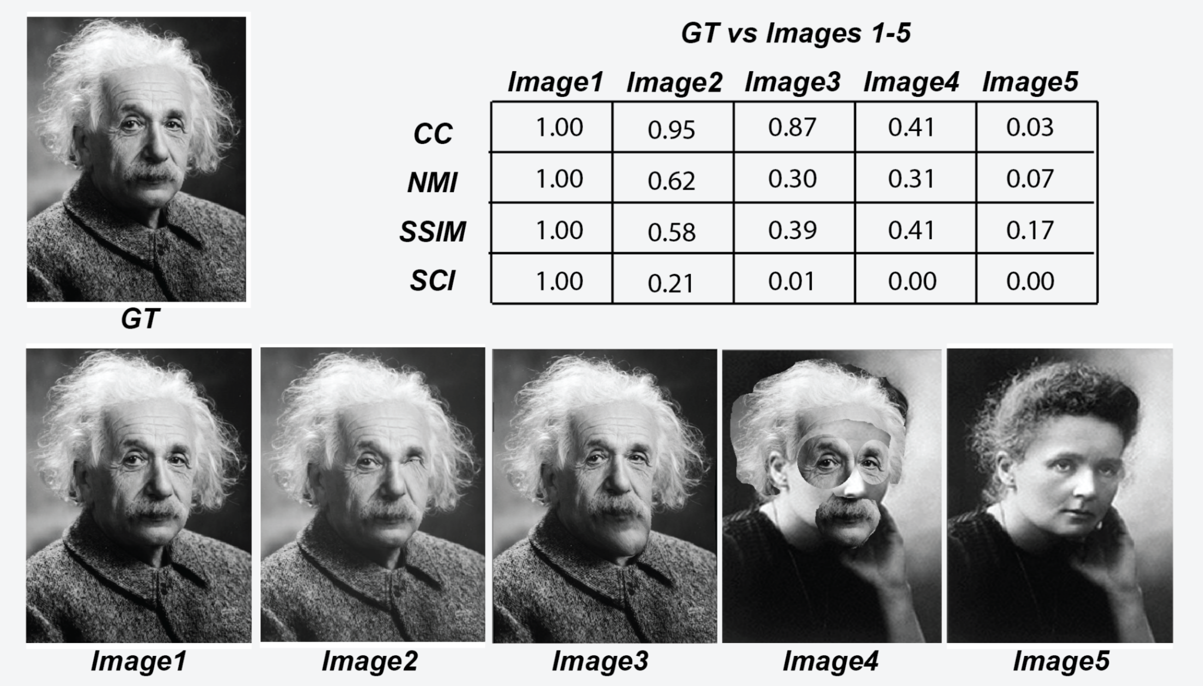
Local resolution maps and Fourier Shell Correlation (FSC) curves demonstrated that EMPROVE outperformed CryoSieve in both global and local resolution (Fig. S2.B). EMPROVE consistently yielded higher-resolution maps, with regions achieving ~4 Å compared to ~6 Å resolution for CryoSieve. Moreover, EMPROVE reconstructions revealed greater structural detail, with more clearly defined secondary structure features than those seen in CryoSieve maps.

The FSC curves confirmed EMPROVE's superior overall resolution (Fig. S2.A), demonstrating that integrating SCI-based scoring with the Euler Equalisation-ranking approach effectively enhances particle selection and improves final reconstruction quality.

Supplementary References

1. Bracewell, R. N. Strip Integration in Radio Astronomy. Aust. J. Phys. 9, 198–217 (1956).
2. Pluim, J. P. W., Maintz, J. B. A. & Viergever, M. A. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22, 986–1004 (2003).
3. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13, 600–612 (2004).
4. Zhu, J. et al. A minority of final stacks yields superior amplitude in single-particle cryo-EM. Nat. Commun. 14, 7822 (2023).

Supplemental Figures



GT vs Images 1-5

	Image1	Image2	Image3	Image4	Image5
CC	1.00	0.95	0.87	0.41	0.03
NMI	1.00	0.62	0.30	0.31	0.07
SSIM	1.00	0.58	0.39	0.41	0.17
SCI	1.00	0.21	0.01	0.00	0.00

Fig. S1. Illustrative comparison of general-purpose image similarity metrics for detecting subtle structural variations. A set of five altered portraits (Image1–Image5) is compared against a ground truth (GT) reference using four image similarity metrics: Cross-Correlation (CC), Normalised Mutual Information (NMI), Structural Similarity Index (SSIM), and Structural Cross-correlation Index (SCI). While CC, NMI, and SSIM display relatively high scores even for significantly altered images, SCI shows a sharper decline, capturing local minute structural inconsistencies. This highlights SCI's higher sensitivity to local structural changes, supporting its use in cryo-EM particle selection and 3D class reassignment where fine differences are critical.

Comparison EMPROVE – CryoSieve, 10k particle selection, 1 iteration

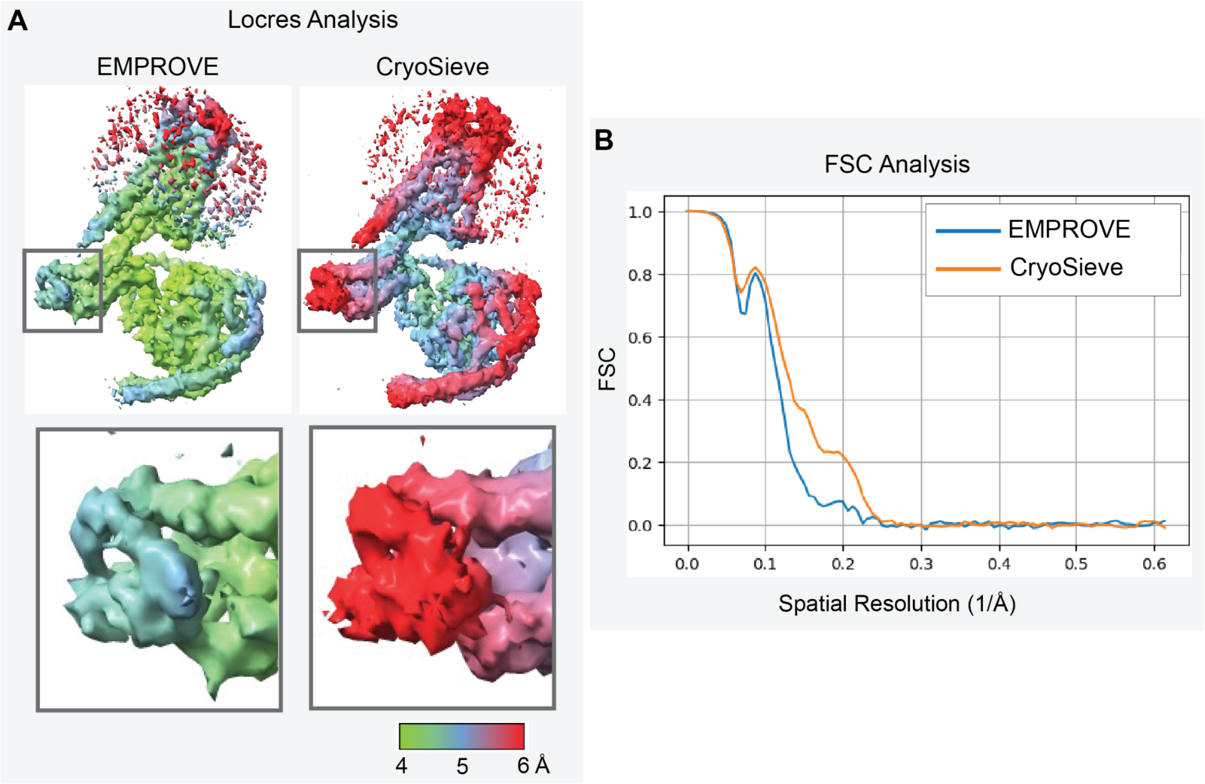


Fig. S2. EMPROVE vs. CryoSieve: Comparative Local Resolution and FSC Analysis. Comparison using 10k particles (0.7% of the full dataset) for reconstruction using automatically selected particles by EMPROVE and CryoSieve v1.2. The left panels illustrate 3D reconstructions where local resolution is color-coded: green for 4 Å, blue for 5 Å, and red for 6 Å. Notably, no post-processing or B-factor sharpening has been applied to these volumes. EMPROVE demonstrates superior performance, yielding resolutions primarily in the 4 Å range, while CryoSieve shows a trend towards the 6 Å range. This improvement in detail is visually apparent and endorses the quantitative local resolution estimation. The right panel's FSC curves further confirm EMPROVE's enhanced resolution capability over CryoSieve, indicating a more effective differentiation of high-quality ('good') versus low-quality ('bad') particles in the selected subset.

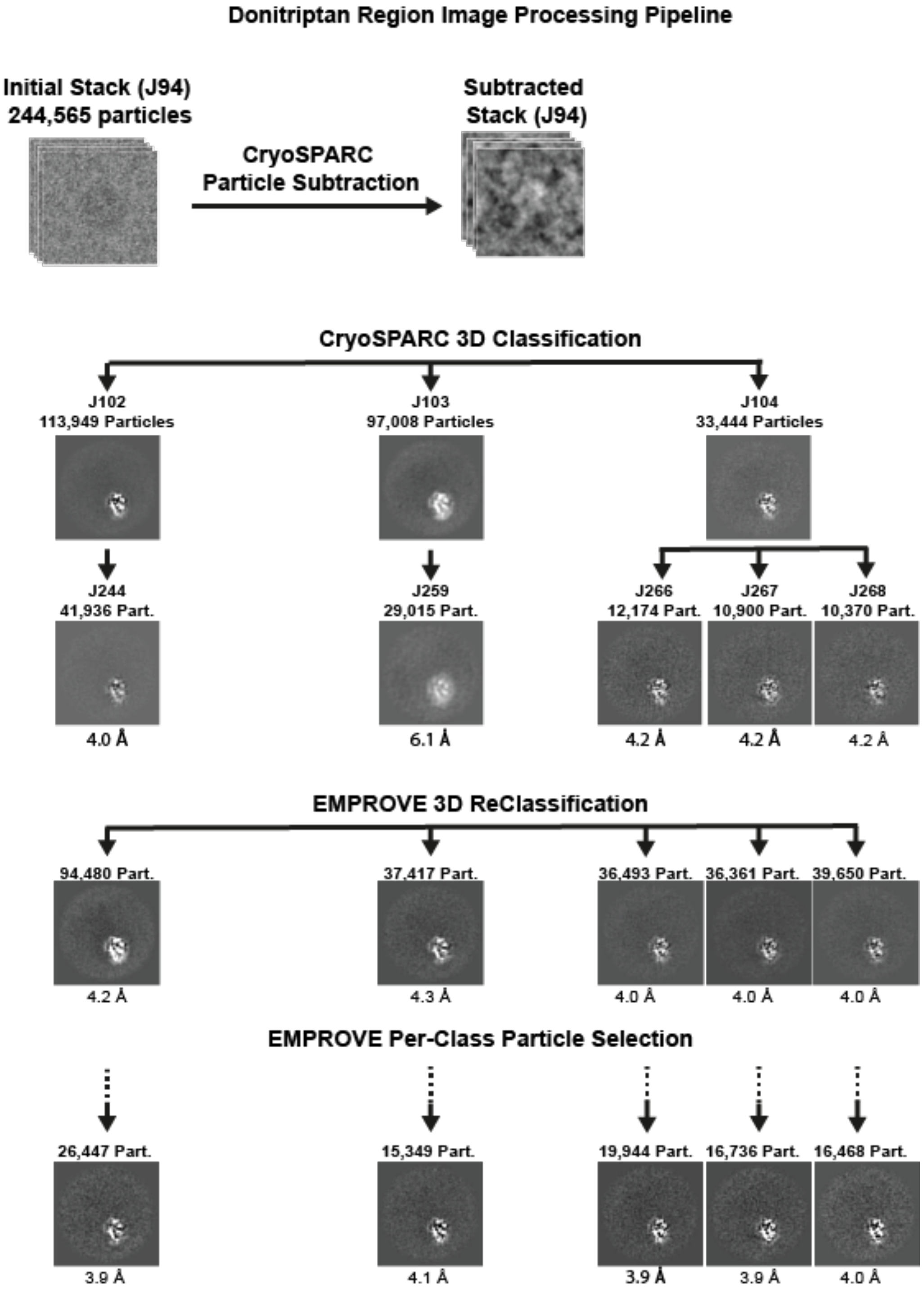


Fig. S3. 5-HT1B receptor–Go complex (EMPIAR-10308) – Donitriptan Region Image Processing Pipeline. The workflow illustrates the use of EMPROVE in combination with cryoSPARC to improve the structural resolution of the 5-HT1B receptor's ligand-binding pocket. Starting from an initial stack of 244,565 particles, cryoSPARC particle subtraction was employed to isolate signal from the ligand-binding region, resulting in a subtracted stack. Subsequent cryoSPARC 3D classification generated five distinct classes with particle counts ranging from 10,370 to 41,936 and local resolutions between 4.0Å and 6.1Å. EMPROVE's 3D reassignment and particle selection were then applied, reducing the resolution gap among classes to 4.0Å–4.3Å and refining particle distributions within each class. Final EMPROVE-selected particles further enhanced resolution to a range of 3.9Å–4.1Å across the classes. The class with 16,736 particles and a resolution of 3.9Å was selected for detailed investigation, demonstrating improved density in the ligand-binding region.

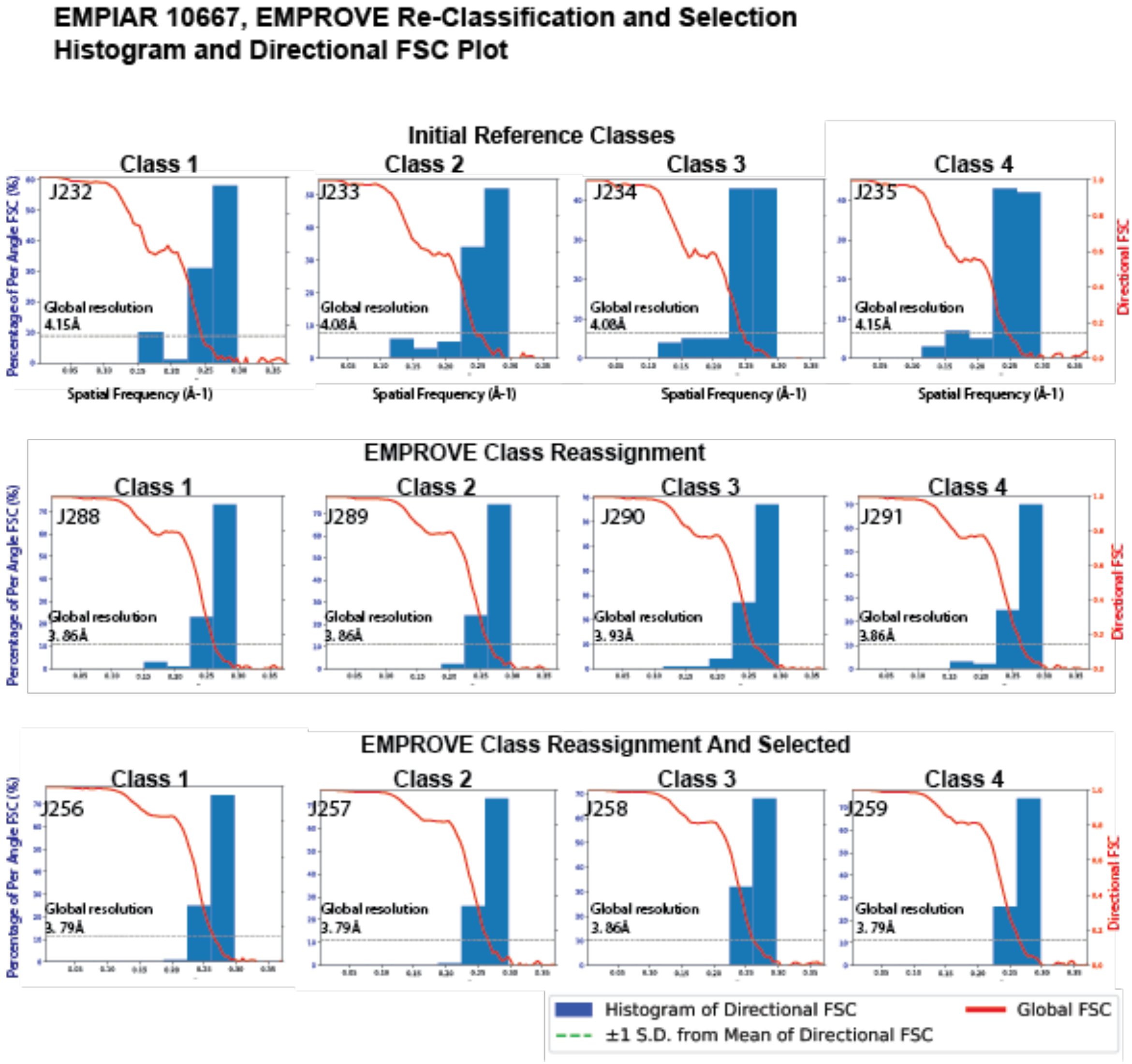


Fig. S4. 46Q HTT–HAP40 complex (EMPIAR-10667) – Histogram and Directional FSC Analysis of Initial, Reassigned, and Selected Classes. Histograms of directional FSC (blue bars) and global FSC plots (red lines) for the four classes derived from EMPIAR-10667, a dataset characterised by a strong preferred orientation (efficiency = 0.48). The top row represents the initial cryoSPARC reference classes (Class 1 to Class 4), showing global resolutions of 4.15 Å, 4.08 Å, 4.08 Å, and 4.15 Å, respectively. The second row shows the results after EMPROVE 3D class reassignment, with improved global resolutions of 3.86 Å for Classes 1, 2, and 4, and 3.93 Å for Class 3. The third row represents the final results following four rounds of EMPROVE particle selection, further refining each class. Global resolutions improved to 3.79 Å for Classes 1, 2, and 4, and remained at 3.86 Å for Class 3. Directional FSC histograms illustrate the distribution of resolution across orientations, highlighting the impact of strong preferred orientation and the reduction of striations and artefacts, which indicate enhanced isotropy and map quality following EMPROVE reassignment and selection. Improvements in global and local resolution across all classes highlight the efficacy of the EMPROVE pipeline in addressing challenges related to misassignment and orientation bias, critical for analysing flexible regions such as the HAP40 N-terminal BΦ motif.

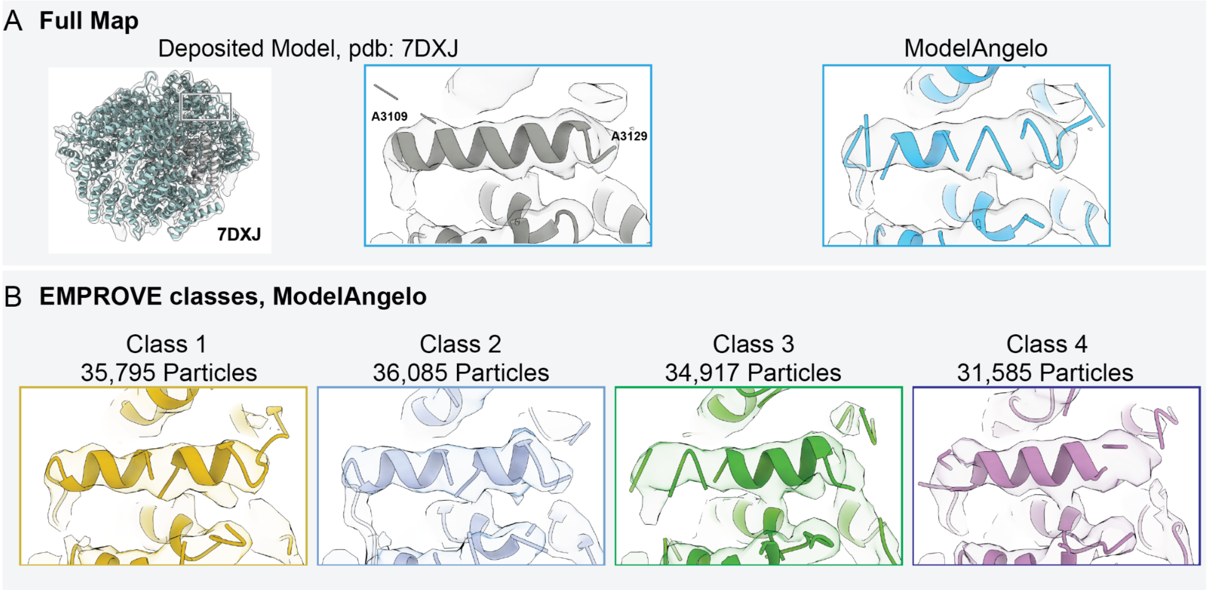


Fig. S5. 46Q HTT–HAP40 complex (EMPIAR-10667) – EMPROVE improves resolution and conformational interpretability of an α-helical segment near the C-terminus. (A) Overview of the full cryo-EM map of the HTT-HAP40 complex (PDB: 7DXJ), with the boxed region indicating an α-helix spanning residues A3109–A3129, located adjacent to the C-terminal domain. Left: deposited cryo-EM map and atomic model, showing incomplete density in this region. Middle: detail of the deposited atomic model, reflecting a single conformation. Right: ModelAngelo-derived model from the full EMPROVE-refined map, showing discontinuities in the α-helix. (B) EMPROVE-assisted reconstructions for each of four 3D classes, followed by automated model building with ModelAngelo. Despite reduced particle numbers, each class shows improved density and more continuous modelling of the α-helix, with distinct conformational features. By selectively reassigning particles based on local structural agreement, EMPROVE enables the recovery of flexible or previously under-resolved elements, even in regions affected by low signal or strong orientation bias.

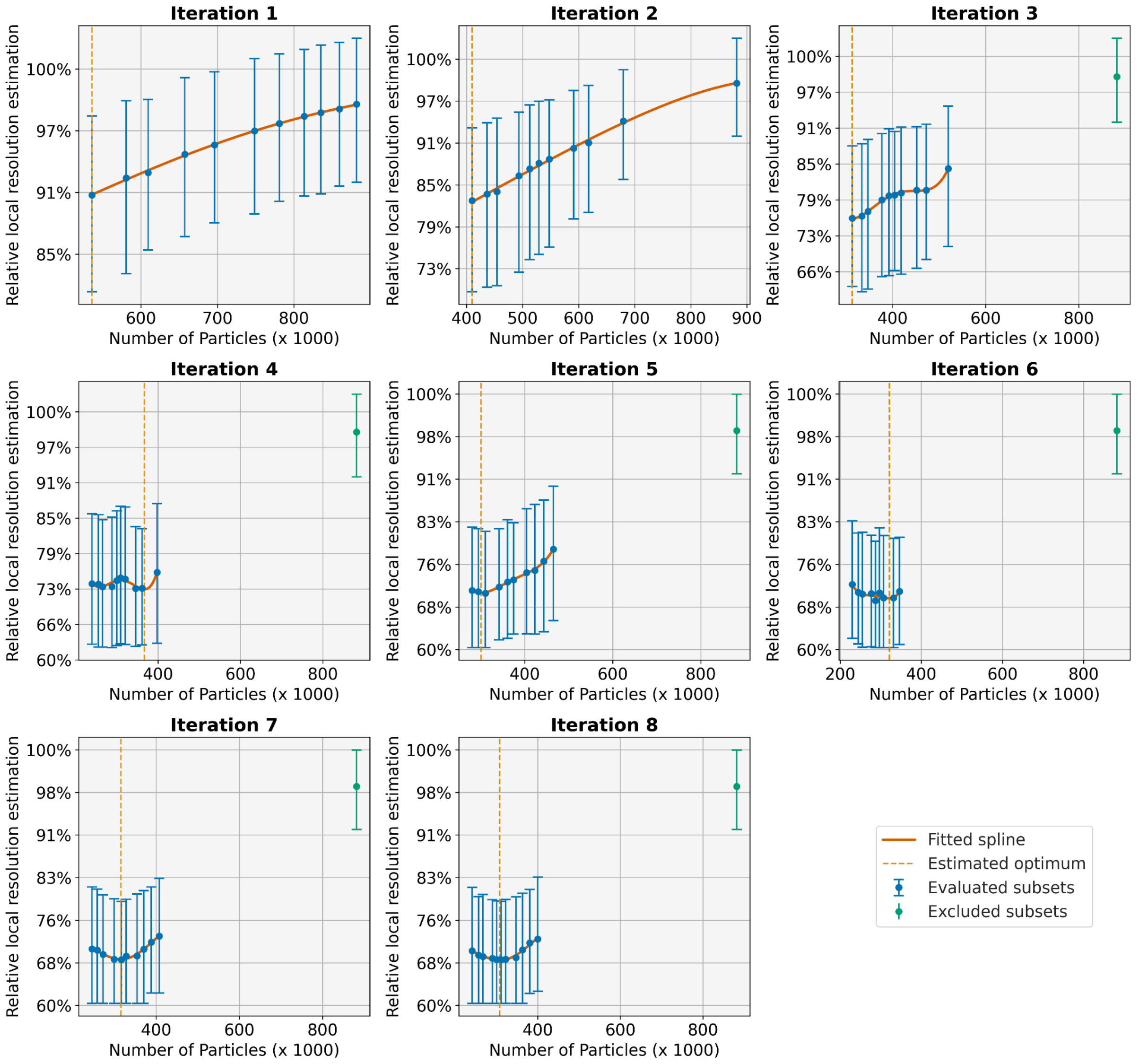


Fig. S6. Group II intron dataset – EMPROVE-guided optimization of particle count based on relative local map quality. EMPROVE iteratively identifies the optimal number of particles for cryo-EM reconstruction through adaptive subset sampling, starting from the largest dataset. At each iteration, subsets of particles are randomly selected using a negatively skewed Gaussian distribution centred on the previous estimate and biased towards smaller particle counts. For each subset, relative local resolution is computed to evaluate map quality. A univariate spline interpolation (orange line) is fitted to the evaluated subsets (blue), and the estimated optimum (dashed vertical line) is updated accordingly. Iteration 7 (315,214 particles) identifies the optimal balance between data quantity and structural resolution, beyond which local quality declines due to the inclusion of low-contributing particles or undersampling. Iteration 8 marks the termination of the optimisation, as no further improvement in subset separation was observed.

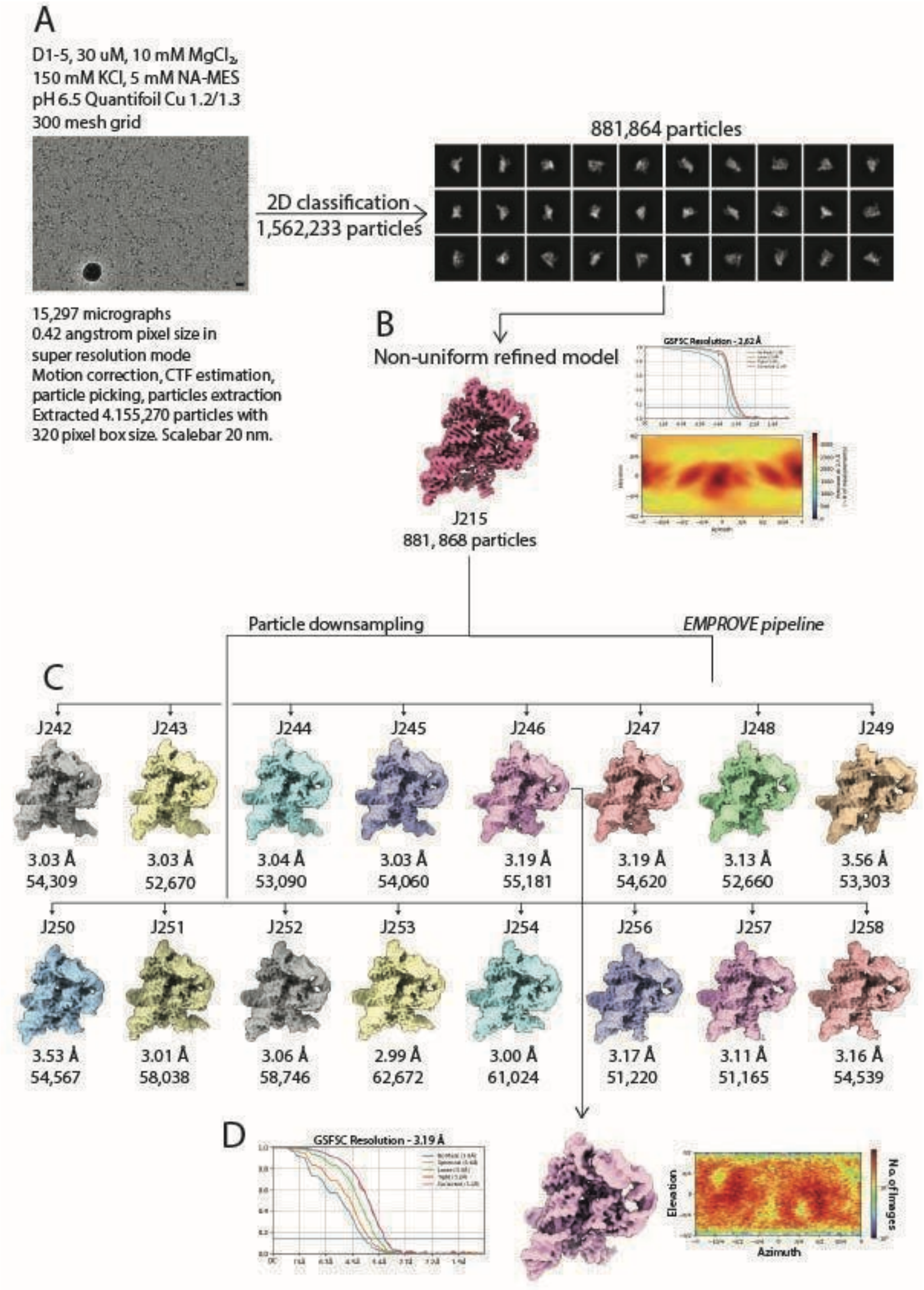


Fig. S7. Group II intron dataset – cryo-EM data processing pipeline. (A) Representative micrograph of the Oceanobacillus iheyensis Group II intron and outlines the initial data processing steps, from motion correction and CTF estimation to particle extraction and 2D classification, which resulted in 881,864 selected particles. In (B), these particles were used to generate a non-uniform refinement in cryoSPARC, yielding a 2.62 Å reconstruction with the corresponding FSC curve and angular distribution. To investigate potential conformational variability, the particle set was downsampled and split into 16 subsets refined independently, as shown in (C), producing maps with resolutions between 2.99 Å and 3.56 Å. One of these, map J246, was selected for downstream analysis; its reconstruction, resolution estimate, and particle orientation distribution are shown in (D).

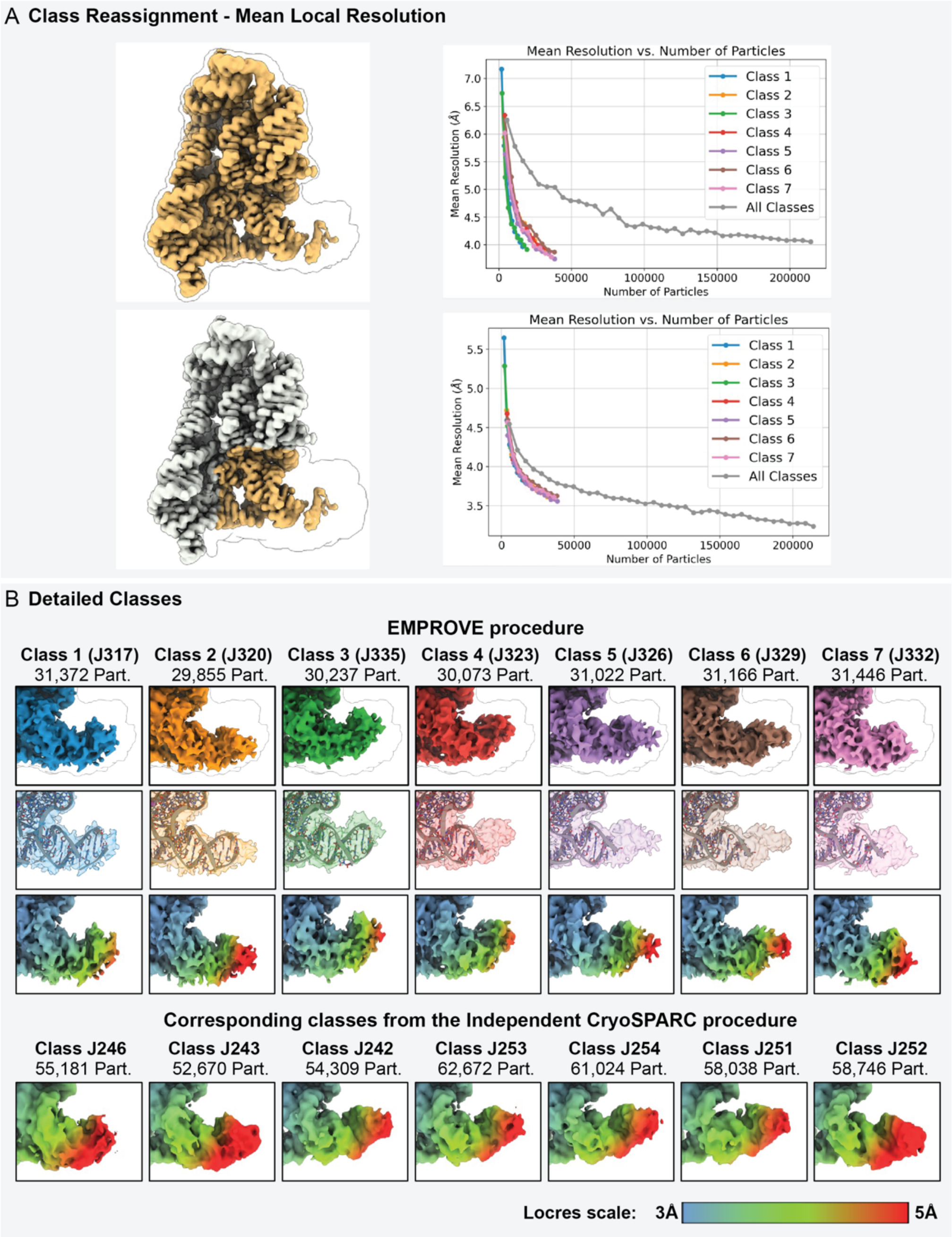


Fig S8. Group II intron dataset – Local resolution analysis across EMPROVE-selected classes. Approximately 200,000 particles were divided into 7 classes, with each class and the collective dataset ('All Classes') undergoing individual reconstructions. Reconstructions were conducted at regular intervals using randomly selected subsets of particles from each class. In **A**, local resolution computations for these reconstructions were performed within the confines of both the 'Overall Mask' and a 'Local Mask' focused on a flexible region with masks shown in pink. The two accompanying graphs synthesise the comparative outcomes. From the overall maps, local flexibility made a small impact on the graph, with the whole set of particles reconstructed largely exceeding the maximum resolution obtained for each individual class. When resolution was assessed in a specific region, all the individual reconstructions had better resolution than the whole set of particles reconstructed. This demonstrates the variability in the ideal particle number when considering entire sets of classes compared to individual ones. The significance of this finding is particularly relevant when datasets contain particles exhibiting multiple conformations central to the research question, precludes the use of a global quality estimator metric as a selection criterion. In **B,** a closeup of a region which was better resolved in the seven classes than in the ‘All Classes’ reconstruction. The grey silhouette corresponds to the density occupied by all seven classes as is shown to help illustrate the conformational variability in this region.

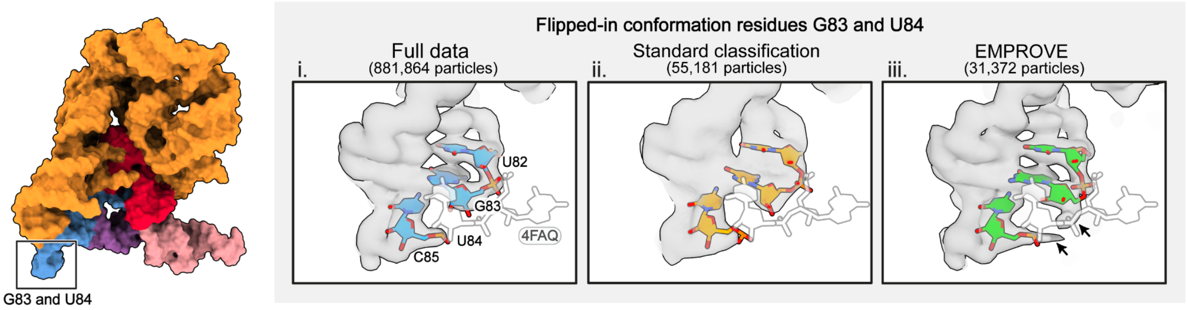


Fig S9. Group II intron dataset – EMPROVE reveals improved density and local structural rearrangements of G83 and U84 residues. The figure highlights key residues involved in the active site architecture: G83, C85, U82, and U84, comparing them against the crystal structure (PDB ID: 4FAQ) shown in white. (i) The full dataset reconstruction (881,864 particles, 2.6 Å resolution) shows G83 in a flipped-in conformation, positioned within the RNA helix, while U84 is flipped out, suggesting increased flexibility. (ii) The standard classification reconstruction (55,181 particles, 3.2 Å resolution) exhibits weaker density in this region, making it difficult to accurately determine the orientation of G83 and U84. (iii) The EMPROVE-selected reconstruction (31,372 particles, 2.8 Å resolution) provides improved density for this region, confirming the flipped-in conformation of G83 and the outward flexibility of U84, providing new insights into the local structural rearrangements of these residues

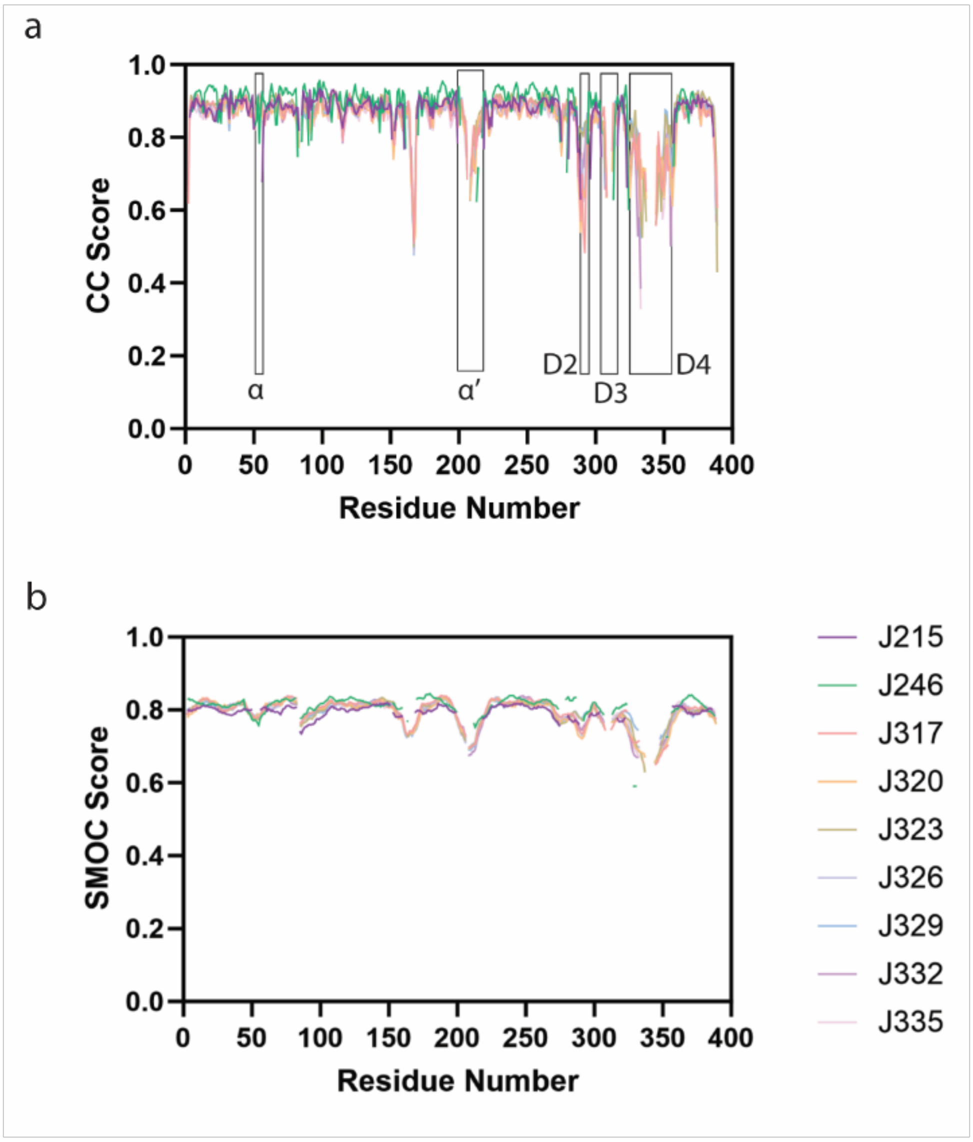


Fig. S10. Cross-correlation and SMOC score plot. (a) The graph displays the cross-correlation (abbreviated as CC) score for CRYOSPARC (J215, J246) and EMPROVE (J317, J320, J323, J326, J329, J332 and J335) models. Rectangles highlight the region of CC score from EMPROVE models corresponding to α-α’ motif, the terminal loop of D2 and D3, and D4. (b) The graph displays the distribution of SMOC score calculated for CRYOSPARC (J215, J246) and EMPROVE (J317, J320, J323, J326, J329, J332 and J335) models