Supervised Deep Learning for Efficient Cryo-EM Image Alignment in Drug Discovery with cryoPARES

Ruben Sanchez-Garcia[1,2,3,*], Alex Berndt[2], Amir Apelbaum[2], Judith Reeks[2], Pamela A Williams[2], Carl Poelking[2], Charlotte M Deane[1,+,*], Michael Saur[2,+,*]

[1]Department of Statistics, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, United Kingdom

[2]Astex Pharmaceuticals, 436 Cambridge Science Park, Cambridge, CB4 0QA, United Kingdom

[3] School of Science and Technology, IE University, Paseo de la Castellana 259, 28046 Madrid, Spain

[+] Joint last authors

* Correspondence: ruben.sanchez-garcia@stats.ox.ac.uk, deane@stats.ox.ac.uk, michael.saur@astx.com

Abstract:

Cryo-Electron Microscopy (cryo-EM) is a pivotal tool for determining 3D structures of biological macromolecules. Current workflows are computationally demanding and require manual intervention, creating bottlenecks for high-throughput applications like structure-based drug discovery. In such contexts, where all protein samples can be assumed to be equivalent at resolutions relevant for image alignment, information about particle poses from previous refinements could be reused. Existing methods, however, ignore this prior knowledge, aligning each dataset from scratch. We present cryoPARES, a deep learning pose estimation method trained on pre-aligned datasets. Our method not only provides accurate angular predictions significantly faster than traditional approaches but also introduces automated particle pruning capabilities that eliminate manual intervention. Together with its single-pass operation, these features enable real-time reconstructions that provide feedback during data acquisition. We demonstrate cryoPARES's effectiveness through rapid structural determination of six ligand-bound complexes across three distinct protein targets and release three new fragment-bound cryo-EM datasets.

1. Introduction:

Cryo-Electron Microscopy (cryo-EM) has emerged as a powerful tool in structural biology[1] with applications in structure-based drug discovery (SBDD)[2–4]. At the heart of this technique lies Single-Particle Analysis (SPA), which reconstructs 3D structures of macromolecules from hundreds of thousands of 2D projection images, termed particles. Ideally, these particles adopt random orientations in the ice layer[5], with their orientation initially unknown. The main challenge in SPA lies in estimating these orientations amidst high noise and multiple sources of error.

Typically, the orientation estimation is performed via iterative angular refinement[6–12], which begins with a low-resolution reference volume against which each particle is compared in all possible orientations. The best-matching orientation(s) are then assigned to each particle and a new volume is then reconstructed, serving as the reference for the next iteration. This process repeats until convergence. While powerful, this approach is computationally demanding due to the large number of iterations required and to the large number of comparisons performed in each iteration[13].

For novel samples, the initial reference volume must be estimated *ab initio*[7,14–16]. However, when prior knowledge exists, it can be used to generate an initial volume. For instance, a previously refined map or a map simulated from an atomic model from a homologous protein can be used. While this strategy may introduce template bias[17,18], using a higher resolution initial volume that is

closer to the final solution can improve convergence of refinement algorithms and speed processing, although gains may be limited by the need for denser angular sampling at higher resolutions.

The use of already solved volumes as references has proven particularly successful in several situations. For example, in cases where it can be assumed that the overall protein structure remains largely unchanged upon ligand binding, a high-resolution structure of the apoprotein can serve as an excellent initial model for the determination of ligand-bound structures[19]. This is because pose estimation is mostly driven by low- to medium-resolution frequencies[20] at which similar proteins are nearly identical. An extreme application of this idea is employed in the baited reconstruction by template-matching approach[19], where a high-resolution reference obtained from an unbound atomic model was used to identify bound ligands via template-matching.

Despite these examples, the computational burden of processing multiple datasets of similar samples still limits wider adoption of cryo-EM in high-throughput studies. Moreover, current approaches require substantial manual intervention for parameter optimization and particle selection, creating additional bottlenecks in the pipeline. To address these challenges, several deep learning alternatives have been proposed[13,21–25]. Among these, "amortised pose inference" approaches, where a neural network is trained to predict particle poses directly from images, are particularly promising. While some are unsupervised *ab initio* approaches[26,27], others can leverage previous structural knowledge through supervised learning[23–25,28].

Supervised approaches train surrogate models of the alignment algorithm using pre-computed poses from previous refinements. While requiring pre-aligned particles may seem a limitation for *ab initio* refinement, this is less problematic in many applications, including ligand screening. In this context, the same protein target is repeatedly screened against different small molecule ligands, generating multiple datasets of previously aligned particles that can be used to train a model that is expected to generalize to similar particles bound to different ligands. However, none of these methods has yet demonstrated practical applicability in this scenario.

Building upon our previous work[28] , we present here cryoPARES (cryo-EM Pose Assignment for Related Experiments via Supervision), a fully automatic pose estimation tool optimized for ligand screening. A single model trained on an apo structure can be used to efficiently process multiple ligand-bound datasets without manual intervention, demonstrating robust generalization across diverse samples. Furthermore, our tool introduces novel capabilities for automated particle pruning, eliminating the need for manual intervention in particle selection. It also shows promise in alleviating the preferred orientations problem often observed in cryo-EM samples. We validate these capabilities through the rapid structural determination of six ligand-bound complexes across three distinct protein targets and release three new ligand-bound cryo-EM datasets to the community.

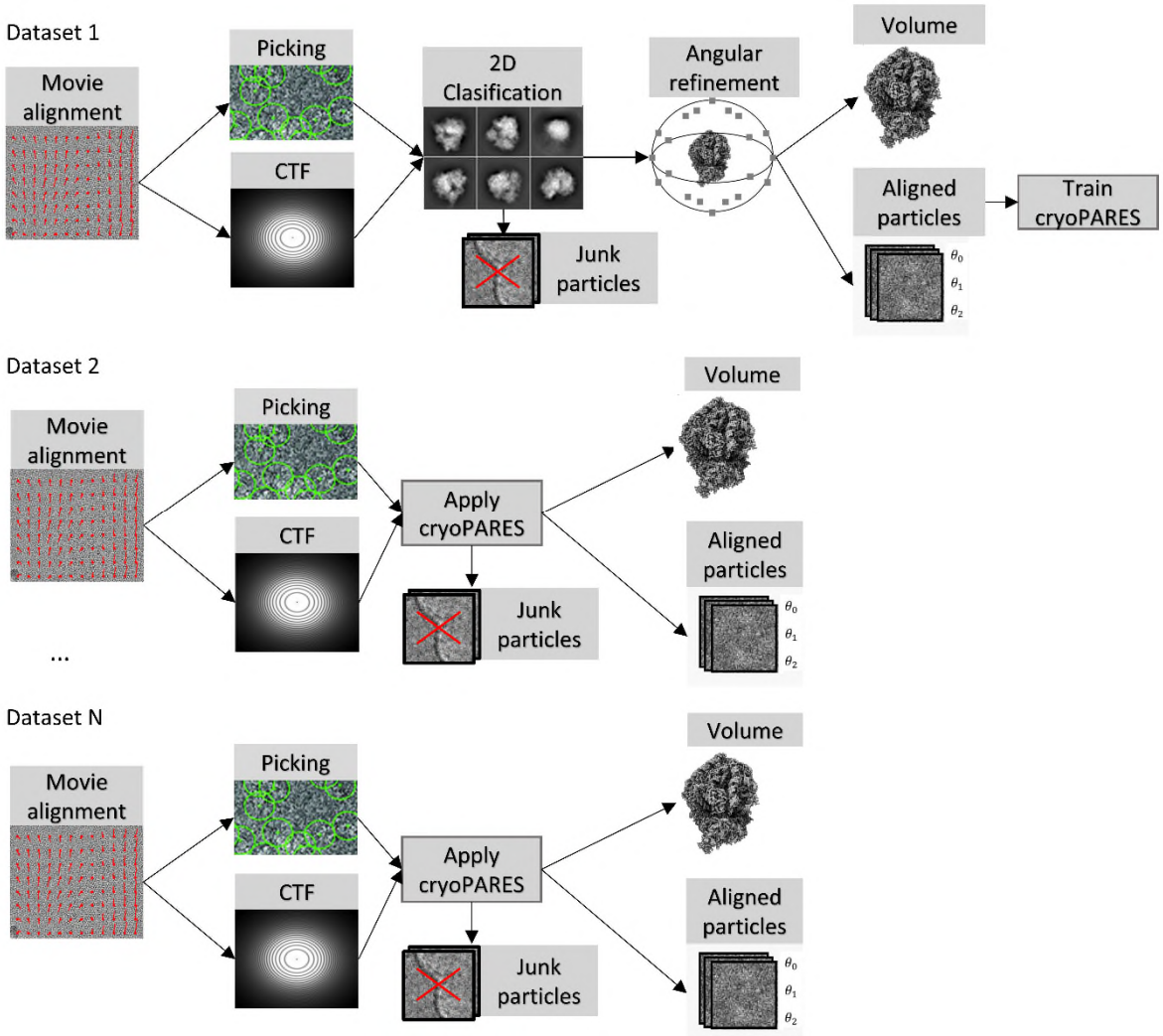

Figure 1. CryoPARES models are trained from a pre-aligned "good" particles dataset that is obtained via classical image processing pipelines. The model is trained in a supervised manner to predict the probability distributions of the pose of each particle. The junk particles in the training dataset can be used to calibrate the particle goodness score (see Methods Section 4.8). Once the models are trained, cryoPARES performs efficient structural determination of subsequent samples by replacing some of the most time-consuming steps by a quick one-shot estimation step.

2. Results and discussion

2.1. CryoPARES is a tool to perform fast structural determination of related samples

CryoPARES addresses an increasingly common challenge in cryo-EM: processing multiple, closely related datasets. It is particularly valuable in SBDD, where work often proceeds only after at least one structure of the target protein has been resolved, and where many similar ligand-bound complexes are subsequently determined in succession. By reusing alignment information across these related experiments, cryoPARES minimizes redundant computation and accelerates structural determination.

To achieve this, cryoPARES uses a supervised deep-learning model trained on pre-aligned particles to predict the probability distribution of particle orientations. Once trained, the network rapidly determines poses for new, related datasets, yielding orientations accurate enough for high-resolution reconstruction. Contrary to conventional refinement methods, most of the computational

cost occurs during the initial training, while subsequent analyses of similar specimens can be completed within minutes, greatly reducing the burden of large-scale structural studies (see Figure 1).

Because cryoPARES predicts pose probabilities, it can automatically identify and exclude junk particles (see Results Section 2.5). This allows its direct use immediately after particle extraction, enabling a fully automated and reproducible reconstruction workflow while removing subjective steps such as manual 2D classification. Its one-shot inference, akin to baited reconstruction, determines orientations in a single pass and, together with its high processing speed (see Results Section 2.4), makes real-time image processing feasible, providing immediate feedback during data acquisition. Such automation and standardization are essential for high-throughput structural studies that demand speed, consistency, and reproducibility.

2.2. CryoPARES can resolve the structure of ligands bound to different protein targets

To illustrate the utility of our tool, we employed cryoPARES to determine the structure of two protein-ligand complexes each for three different protein targets. In all cases, the cryoPARES models were trained using the pre-aligned particles of the apoprotein previously solved and a 6 Å low-pass-filtered apo map was used as the reference for local refinement.

We first tested our method on the previously published datasets of *E. coli* β-galactosidase (BGAL) bound to Compounds B1 and B2[3]. Figure 2 shows the reconstructed densities of the ligand-bound datasets (c,d) compared with the apo reference used for training (a,b). The map features of the ligands using the orientations estimated with cryoPARES agree with the ground-truth atomic models reported previously. As a negative control, it can be observed that the densities of the ligands are not present in the apoprotein maps (Figure 2 a-b). Overall, the Fourier Shell Correlation (FSC) resolution of the reconstructed maps (half-maps at threshold 0.143) was measured to be 2.8 Å for the Compound B1 dataset and 3.2 Å for the Compound B2 dataset (Figure 5 a-b) compared to 2.4 Å and 2.6 Å obtained with a standard workflow of 2D classification and auto-refine in RELION, but at a fraction of the computational cost (see Results Section 2.4). Similarly, the map-to-model FSC resolutions (threshold 0.5) were 2.9 Å and 3.2 Å, confirming good agreement between the reconstructed maps and their corresponding ground-truth atomic models.

Our second test was conducted on our previously published datasets of human pyruvate kinase M2 (PKM2) bound to Compounds P1 and P2[3]. As with BGAL, the densities for the ligand-bound reconstructions obtained using cryoPARES poses matched the ground-truth atomic models (Figure 3). In this case, the FSC resolution obtained with cryoPARES poses was closer to the resolution obtained with the standard RELION workflow: 3.5 Å and 3.3 Å for cryoPARES (Figure 5 c-d), compared to 3.4 Å and 2.8 Å. The map-to-model resolutions were 3.7 Å and 3.5 Å. For this protein target, the apo dataset used for training was collected on a Krios microscope, but the PKM2 Compound P1 dataset was collected on a Glacios microscope, demonstrating the method is robust under different imaging conditions.

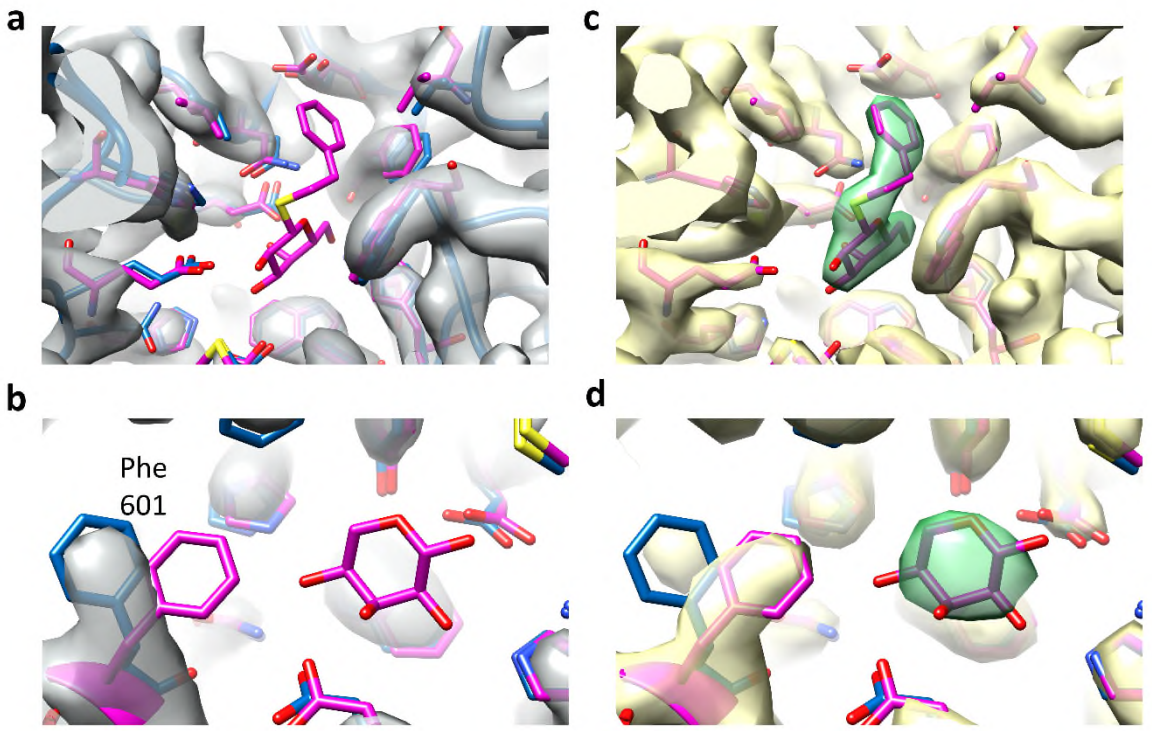

Figure 2. Reconstructed maps of β-galactosidase (BGAL) at the ligand binding site using orientations inferred with cryoPARES. a) Map of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound B1 (pink). b) Same as (a) but with the atomic model of the protein bound to Compound B2 (pink). c) Reconstructed map (yellow) of the protein bound to Compound B1 using orientations estimated with cryoPARES. The ligand and residues within 4 Å of the ligand are displayed in pink. The density closer than 1.8 Å to the ligand is coloured in green. d) Same as (c), but for the protein bound to Compound B2. The features for the ligands are absent in the apo maps (a, b) but clearly visible in the reconstructed maps from the bound datasets (c, d). In this case, the binding of ligands does not induce significant changes in the structure of the protein pocket, with the exception of the Phe 601 in the Compound B2 case (b vs d).

The last test case consists of two in-house datasets of bovine glutamate dehydrogenase (GDH) bound to Compounds G1 and G2. Figure 4 shows the maps for the apo structure used in the training and local refinement steps, as well as the densities of the reconstructed maps reconstructed from cryoPARES-predicted poses. In this case, we measured FSC resolutions of 2.9 Å, and 3.0 Å (see Figure 5 e-f), compared to 2.7 Å and 2.8 Å for the RELION workflow. The map-to-model resolutions were 2.9 Å and 3.1 Å.

These three experiments confirm that cryoPARES can deliver high-resolution structures of ligands, (in the 3.5 to 3.0 Å resolution range). While FSC measurements can be affected by template bias, there are several observations strongly indicating that the effect of this, if present, is small: 1) the ligands reconstructed were neither present in the training data nor the template, 2) the measured resolutions of ~3 Å are significantly better than the frequency threshold used to low-pass filter the local refinement templates (6 Å), and 3) the density of phenylalanine 601 in Figure 2 b,d shows a clear rotamer change upon ligand binding. Further tests and discussion on template bias can be found in Supplementary Section 13.

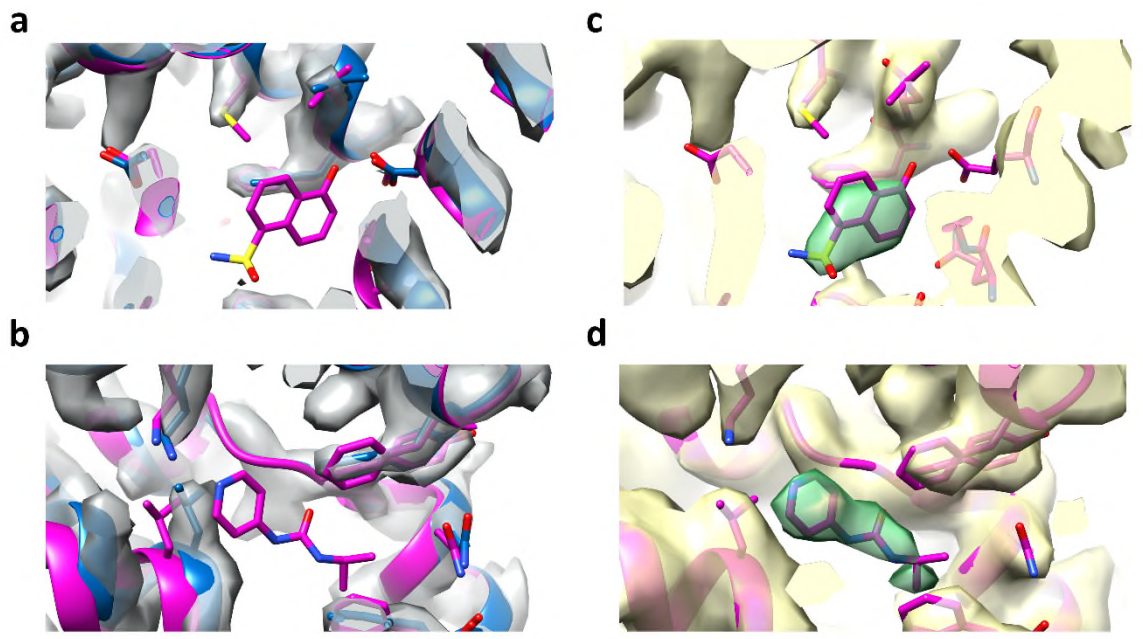

Figure 3. Reconstructed maps of pyruvate kinase M2 (PKM2) at the ligand binding site using orientations inferred with cryoPARES. a) Density of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound P1 (pink). b) Same as (a) but with the atomic model of the protein bound to Compound P2 (pink). c) Reconstructed density (yellow) of the protein bound to Compound P1 using orientations estimated with cryoPARES. The ligand and residues within 4 Å of the ligand are displayed in pink. The density closer than 1.8 Å to the ligand is coloured in green. d) Same as (c), but for the protein bound to Compound P2. The density for the ligands is absent in the apo maps (a, b) but clearly visible in the reconstructed maps from the bound datasets (c, d).

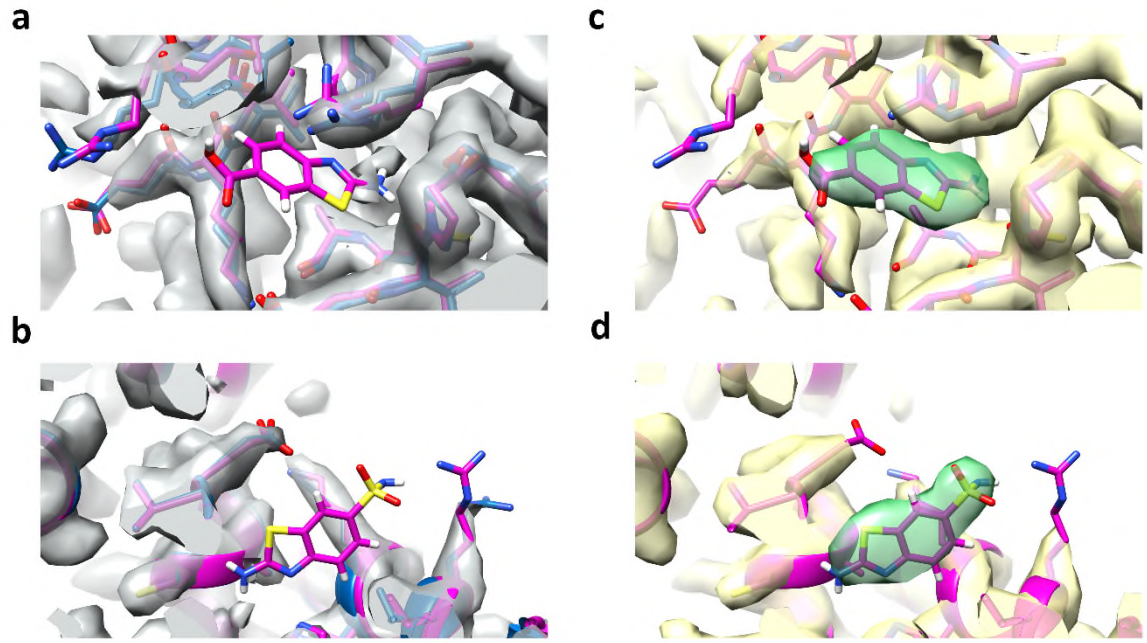

Figure 4. Reconstructed maps of glutamate dehydrogenase (GDH) at the ligand binding site using orientations inferred with cryoPARES. a) Density of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound G1 (pink). b) Same as

(a) but with the atomic model of the protein bound to Compound G2 (pink). c) Reconstructed density (yellow) of the protein bound to Compound G1 using orientations estimated with cryoPARES. The ligand and residues within 4 Å of the ligand are displayed in pink. The density closer than 1.8 Å to the ligand is coloured in green. d) Same as (c), but for the protein bound to Compound G2. The density for the ligands is absent in the apo maps (a, b) but clearly visible in the reconstructed maps from the bound datasets (c, d).

2.3. CryoPARES predicts accurate global alignment parameters

We first evaluated accuracy on simulated data with perfect orientations, where the cryoPARES's neural network achieved validation top-1 median angular errors of 3.9°, 4.6°, and 4.1° for the three protein targets, approaching the SO(3) grid resolution of ~3.7° used within the model. While these results demonstrate the model's capability under ideal conditions, evaluating angular errors on real datasets is more challenging due to the lack of ground truth poses.

For experimental datasets, we compared our predicted poses with those from RELION (see Supplementary Table 3). The neural network's top-1 raw predictions are within 5° of RELION estimates for approximately 30% of particles, increasing to more than 60% when followed by local refinement. Notably, even two consecutive RELION auto-refine jobs on the same dataset can differ by more than 5° for over 20% of particles (see Supplementary Table 3, last column), with studies reporting disagreement levels as high as 50%[23]. This substantial variability between identical RELION runs underscores that perfect agreement between methods is neither expected nor necessarily desirable, since it could reflect shared systematic bias rather than true accuracy. Therefore, our observed level of correspondence with RELION, combined with the quality of our reconstructed maps, supports the reliability of cryoPARES poses.

Another indirect method for assessing the accuracy of the neural network angular assignment is to measure how much local refinement is needed to improve the reconstruction before no gains are observed. Figure 5 shows the FSC curves for all the samples where local refinement with an angular search range of 0°, ±2°, ±4° and ±6° and step size of 2° was applied. In every dataset, most of the gain was achieved with a ± 2° search, with only marginal improvement beyond ± 4°, indicating that the predicted orientations are typically within 2–4° of the optimal alignment.

2.4. CryoPARES performs real time particle alignment

Supplementary Table 4 summarizes the measured runtimes for pose estimation across the different datasets. During inference, the cryoPARES neural network achieves throughputs exceeding 30,000 particles min^{-1} GPU^{-1}, allowing a 250,000-particle stack to be processed in less than eight minutes. If we assume that cryoPARES rejects as junk particles approximately 50% of the dataset, the local refinement of the remaining particles would take about seven additional minutes, resulting in a total runtime of roughly 15 minutes on a single GPU for 250,000 picked particles.

Since our local refinement implementation is a proof-of-concept demonstrator, performance could be further enhanced through better alignment algorithms or by reducing the amount of local refinement required via improvements of the neural network. Even in its present form, cryoPARES is more than an order of magnitude faster than RELION auto-refinement and comparable in speed to the non-open source cryoSPARC homogeneous refinement (see Supplementary Table 5). Given that a high-end microscope can collect up to 700 micrographs h^{-1}, and assuming 400 particles per micrograph, real-time processing requires ~4,700 particles min^{-1}. CryoPARES comfortably exceeds even the most demanding acquisition rates, sustaining > 10,000 particles min^{-1} GPU^{-1} including the local-refinement step.

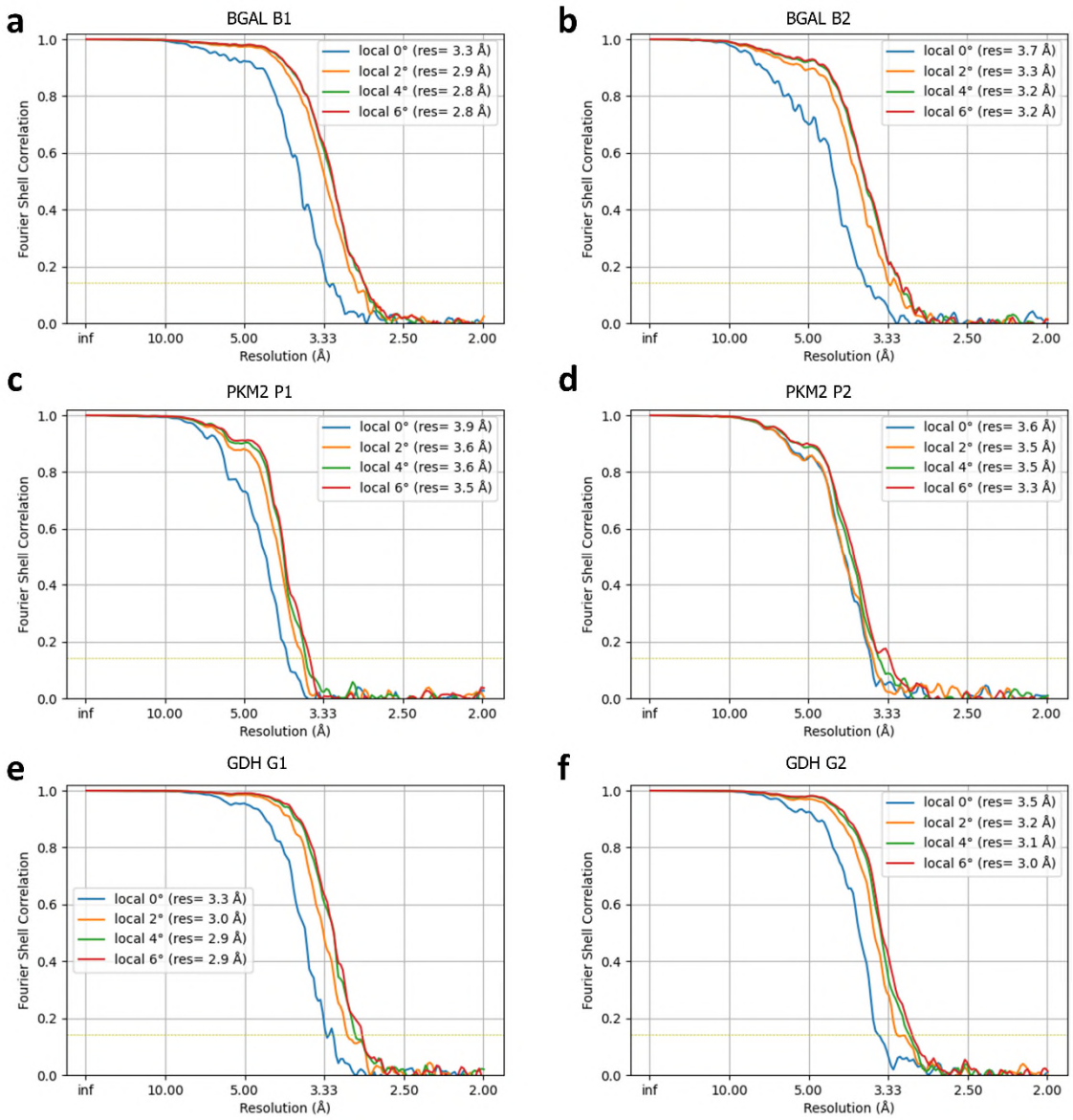

Figure 5. FSC curves (half maps) for the maps obtained by cryoPARES at different levels of local refinement angular search range (0° blue, ±2° orange, ±4° green, ±6° red) for the protein-ligand complexes BGAL bound to Compound B1 (a) and Compound B2 (b), PKM2 bound to Compound P1 (c) and Compound P2 (d), GDH bound to Compound G1 (e) and Compound G2 (f).

Unlike conventional workflows that require accumulating a minimal number of particles to perform statistical estimations and preprocessing steps like 2D classification, cryoPARES processes each particle independently. This fundamental difference not only enables processing to begin immediately after the first micrograph is acquired but also allows for continuous updates of the 3D reconstruction as new particles become available. Together, the combination of high-throughput processing and per-particle analysis can transform cryo-EM data processing from a batch-based, post-acquisition task into a real-time monitoring system that provides immediate feedback on data quality and experimental parameters.

2.5. CryoPARES can perform automatic particle pruning

CryoPARES provides probability distributions over particle orientations, which can be used to identify junk particles and those with unreliable estimated poses. While one could directly use the

probability of the predicted orientation as a quality metric, we observed that the distribution of these raw probabilities depends strongly on viewing direction. Applying a single global threshold would therefore bias reconstructions toward more frequent orientations, potentially producing artefactual preferred orientations (Supplementary Figure 3).

To overcome this, we implemented a pruning score based on direction-normalized robust z-scores of the predicted probability distributions (see Methods Section 4.8). This normalization allows particle and pose quality to be evaluated consistently across all orientations. Supplementary Figure 4 presents histograms of the direction-normalized scores for "good" particles (those retained after 2D classification) and "bad" particles (those discarded). In every case, the distributions for good particles extend toward higher scores, whereas those for bad particles are concentrated at lower values, confirming that a single threshold can separate the two populations. All reconstructions reported in the previous sections were obtained from the datasets pruned using automatically estimated thresholds of this score (see Supplementary Section 9).

To further assess pruning performance, we compared RELION auto-refine results obtained using (i) particles cleaned by 2D classification, (ii) particles selected using cryoPARES pruning scores, and (iii) randomly chosen subsets containing the same number of particles. Note that, for this experiment only, we manually adjusted the cryoPARES thresholds to match the number of retained particles from 2D classification. For this reason, the thresholds used here differ from the automatically determined values employed in the previous sections. Table 1. Resolution obtained with a RELION auto-refine job when the input particles are (1) cleaned with a 2D classification step, (2) cryoPARES directional pruning scores, or (3) randomly selected particles. The same number of particles in the second and third cases was selected to match the 2D classification outcome.Table 1 summarizes the resulting resolutions. Datasets pruned by cryoPARES achieved resolutions comparable to those obtained through manual 2D classification, while random subsets consistently produced lower-resolution reconstructions. Because the datasets analyzed so far were relatively clean (as evidenced by the small effect of random subsetting in several cases), we also evaluated cryoPARES pruning on two in-house membrane-protein datasets bound to distinct ligands (MPL1 and MPL2), which contain a higher proportion of poorly aligned and junk particles. In these more challenging cases, the benefits of pruning were particularly pronounced: cryoPARES-selected particles improved the final resolutions by 0.4–1.2 Å relative to random subsets and reached values nearly identical to those obtained after exhaustive 2D classification.

Table 1. Resolution obtained with a RELION auto-refine job when the input particles are (1) cleaned with a 2D classification step, (2) cryoPARES directional pruning scores, or (3) randomly selected particles. The same number of particles in the second and third cases was selected to match the 2D classification outcome.

Sample	Resolution after 2D classification (Å)	Resolution after cryoPARES (Å)	Resolution after random subset (Å)
BGAL Compound B1	2.51	2.50	2.74
BGAL Compound B2	2.75	2.70	3.10
PKM2 Compound P1	3.32	3.36	3.39
PKM2 Compound P2	2.88	2.82	2.96
GDH Compound G1	2.52	2.50	2.54
GDH Compound G2	2.67	2.63	2.74
MPL1	2.54	2.57	3.02
MPL2	2.79	2.67	3.92

Together, these analyses show that cryoPARES can effectively replace 2D classification as a particle-pruning step, offering two key advantages: on-the-fly operation and full reproducibility, as no manual class selection is required.

2.6. CryoPARES is robust against preferred orientations

CryoPARES, similar to baited reconstruction approaches[19], avoids the attractor effect that arises in iterative refinement. It happens when orientations represented by more particles—and thus higher signal-to-noise ratios—pull neighbouring particles toward themselves, amplifying over-represented directions and biasing the reconstruction[29,30]. This property makes cryoPARES particularly robust for specimens exhibiting preferred orientations, since its training on unbiased pose distributions prevents reinforcement of dominant views.

To illustrate this situation, we analyzed another GDH dataset (bound to ADP, Compound G3) exhibiting significant preferred orientations. When reconstructed using a standard RELION workflow —comprising 2D classification followed by auto-refine with a 40 Å low-pass filtered reference— the resulting volume shows severe anisotropy that precludes ligand identification (Figure 6 a). However, employing a higher-resolution density (15 Å) as reference yields substantially improved reconstruction quality, enabling ligand visualization (Figure 6 b,d). This scenario resembles the baited reconstruction strategy, where a high-resolution reference is used as template.

Reconstruction based on cryoPARES-predicted orientations produced a map with the correct global features and markedly better-resolved local features (Figure 6 c,e). The ligand density in the cryoPARES map was more continuous than in the RELION result, and nearby side-chains—such as LYS488, adjacent to the ADP molecule—were better defined. These results confirm that cryoPARES alleviates orientation bias and remains reliable even for datasets dominated by preferred views.

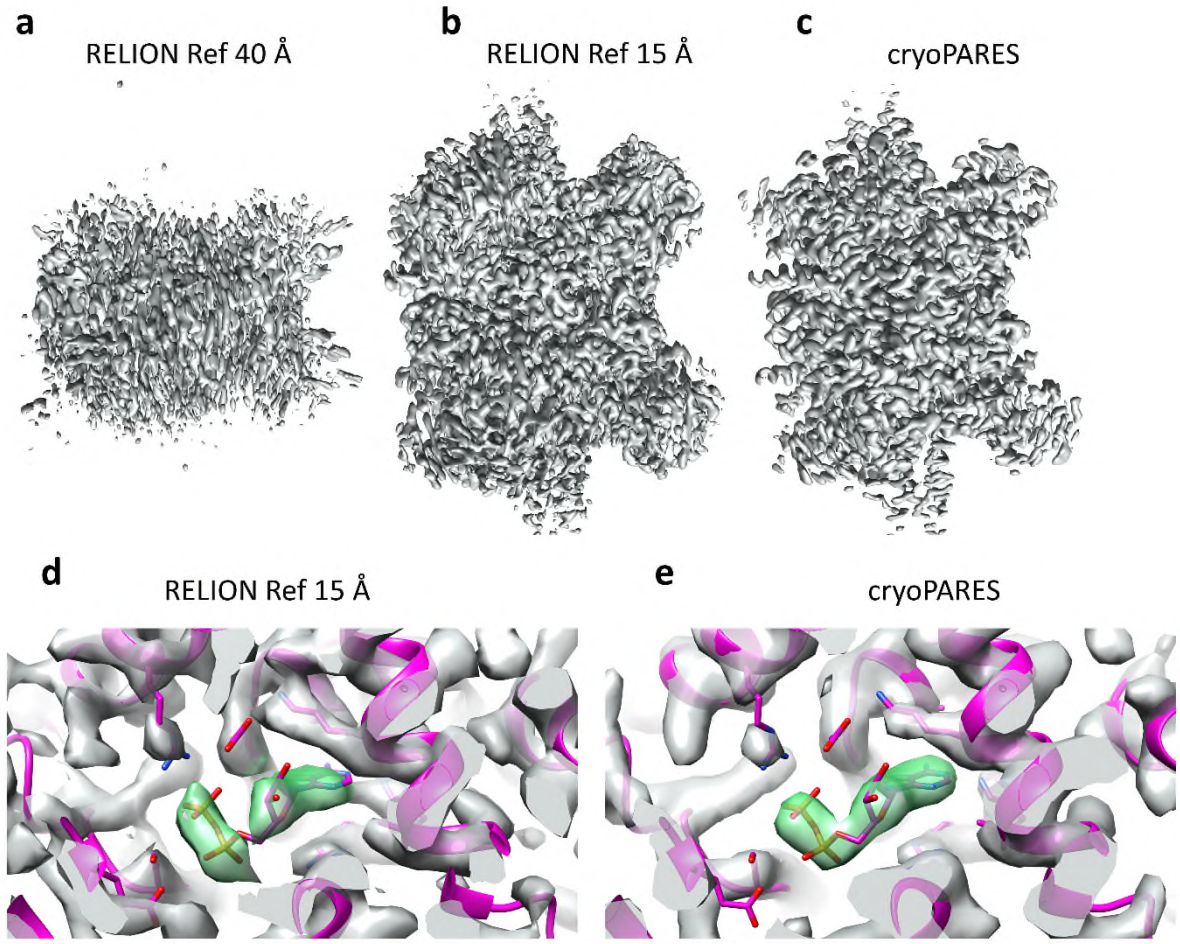

Figure 6. Reconstructed densities from a dataset of GDH bound to ADP that contains preferred orientations. a) Overview of the map obtained with a RELION workflow of 2D classification followed by auto-refine with a reference low-pass filtered at 40 Å b) Same as (a) with a reference low-pass filtered at 15 Å. c) Overview of the map reconstructed using cryoPARES. d) Close-up of (b) showing the ADP binding site. e) Close-up of (c) showing the ADP binding site. The atomic model (pink) corresponds to the ground-truth structure, computed from another stack of ADP-bound GDH particles not affected by preferred orientations. In panels d-e, the density within 1.8 Å of the ADP is coloured in green.

3. Conclusions

In this work, we introduced cryoPARES, a supervised deep-learning framework that performs efficient cryo-EM pose determination in cases where pre-aligned particle datasets are available. By learning from these datasets, cryoPARES rapidly predicts particle orientations while performing automatic particle pruning, thereby significantly reducing the computational burden of single-particle analysis.

Experiments with BGAL, PKM2, and GDH demonstrate that cryoPARES reconstructs high-resolution ligand-bound complexes that accurately reproduce ligand densities absent in the apo structures used for training. The method generalizes robustly across different datasets and imaging conditions, maintaining high performance even when trained on data acquired with multiple microscopes. Furthermore, cryoPARES can alleviate some limitations of current methods associated with the attractor problem, such as their high sensitivity to preferred orientations, thus producing reliable results in challenging datasets where traditional approaches may struggle.

The high-speed, one-shot nature of cryoPARES enables real-time data analysis during cryo-EM acquisition, providing immediate feedback on data quality and experimental parameters. This capability, combined with automatic particle pruning, is particularly valuable for high-throughput structural studies and drug discovery campaigns, where rapid iteration, automation, and reproducibility are essential.

4. Methods

4.1. Datasets

We studied three protein targets, β-galactosidase (BGAL), pyruvate kinase M2 (PKM2) and glutamate dehydrogenase (GDH). For each protein target, we prepared two types of datasets: training data from apo samples and testing data from ligand-bound samples (see Supplementary Table 2 for a list of compounds).

The testing sets for the BGAL and the PKM2 targets were prepared from previous work[3]: EMPIAR-10644 (β-galactosidase bound to PETG, Compound B1), EMPIAR-10646 (β-galactosidase bound to L-ribose, Compound B2), EMPIAR-10648 (PKM2 bound to 5-hydroxynaphthalene-1-sulfonamide, Compound P1), and EMPIAR-10649 (PKM2 bound to 3-(propan-2-yl)-1-(pyridin-4-yl)urea, Compound P2). The training sets for BGAL and PKM2 were prepared following an identical protocol as the ligand-bound datasets.

For a description of the GDH datasets sample preparation, data collection and atomic modelling processes, see Supplementary Material Sections 3-5. The employed ligands were 2-amino-1,3-benzothiazole-6-carboxylic acid (Compound G1), and 2-amino-1,3-benzothiazole-6-sulfonamide (Compound G2). An additional dataset, ADP-bound (Compound G3), was employed for the preferred orientations analysis.

4.2. Datasets image processing

The particle extraction process was identical for both training and testing datasets. CTF estimation was performed using Gctf 1.18[34], using a 1024-pixel box size, 30 Å minimum resolution, and 0.07 amplitude contrast. We then conducted particle picking via template matching using Gautomatch 0.56 (http://www.mrc-lmb.cam.ac.uk/kzhang/), with references low-pass filtered to 20 Å, and extracted all particles at 1 Å/pixel. For testing, these extracted particles were directly fed to cryoPARES, which predicts their poses without requiring any additional processing. However, to train cryoPARES and to evaluate cryoPARES predictions, we needed reference poses and an estimation of good and bad particles for particle pruning. We obtained these by processing both training and testing datasets through a 2D classification job with 25 steps in RELION 3[35], followed by 3D refinement using RELION auto-refine with a reference map low-pass filtered at 20 Å (BGAL and PKM2) or 15 Å (GDH) and a soft mask created from the reference map. See Supplementary Table 1 for details.

4.3. Training strategy

For each of the targets, the apo dataset containing only "good" particles is evenly split into two half-sets of particles. Then, for each of the half-sets, a model is trained using 70% of the data as the training set and the remaining 30% as the validation set. In each of the training runs, we first pre-train our model for 5 epochs using a simulated dataset of particles (see Supplementary Material Section 6 for more details about the simulation). After that, we continue training for at most 100 epochs with the training set of experimental particles. While the pre-training phase using simulated data is not necessary – as models can be directly trained on experimental particles alone – it can lead to better results in some cases (see Supplementary Figure 1).

For both the pre-training and the training steps, we employ RAdam[36] as the optimizer, with an initial learning rate of 1e-3 and a weight decay of 5e-6. The metric used to monitor the training progress was the mean validation geodesic distance,

$$\text{distance}(trueR_i, predR_i) = \arccos\left(\frac{\text{trace}\left(trueR_i\, predR_i^T\right) - 1}{2}\right)$$

where $trueR_i$ and $predR_i$ are, respectively, the ground-truth and predicted rotation matrix for the i-th particle of the validation set. When not stated otherwise, $predR_i$ refers to the highest probability rotation matrix, referred as the top-1 orientation. During training, we halve the learning rate if the validation error did not improve for more than 5 epochs. Early stopping is applied when the validation error did not improve for more than 11 epochs. Data augmentation is applied for the training set (see Supplementary Section 7 for details). We use a batch size of 32 images with gradient accumulation over 20 batches. Each batch includes 4 copies of the same particles subjected to different data augmentation techniques. This is intended to help the network learn augmentation-invariant features, particularly for shifts and in-plane rotations. An optional SimCLR-like contrastive loss[37] was also implemented to favor learning such features. However, no significant performance gains for the datasets presented here were observed.

Once training is completed, we apply the trained model to predict poses for particles in the validation set. When comparing these predicted poses with their ground truth values, we also compute the metrics needed to develop our particle pruning criteria based on direction-normalized robust z-scores, which help us filter out junk and/or unreliably aligned particles in new datasets (see Methods Section 4.8 for details).

4.4. Pose inference strategy

For pose inference, we start with the ligand-bound particles extracted directly after particle picking. We split these particles into two equally-sized subsets and assign each subset to one of our two independently trained apoprotein models for initial pose prediction. Following this initial prediction, we employ our particle pruning strategy based on direction-normalized robust z-scores to filter out junk particles and incorrectly aligned particles (see Methods Section 4.8). The remaining particles then undergo local refinement to improve their orientations and/or estimate in-plane shifts (see Methods Section 4.7).

4.5. Data pre-processing

Before being fed to the neural network, the particle images undergo several preprocessing steps. First, they are downsampled to 1.5 Å/pixel and normalized such that the corner regions, which should contain only noise, exhibit a mean of 0 and standard deviation of 1. The images are then enriched with an additional channel containing their CTF-corrected (phase-flipped) versions. From these processed images, we extract central crops. Then, a circular mask is applied with a diameter equal to the particle diameter. To enhance feature detection across multiple levels of noise, as suggested by Levy et al. [13], we apply a series of Gaussian filters (σ = 0, 0.3, 1, and 2) to both the raw and phase-flipped channels. This preprocessing pipeline results in 8-channel images going into the network

For orientation representation, particle rotations are encoded as rotation matrices and mapped to their nearest neighbours in a predefined SO(3) grid. Formally, given a grid G of rotation matrices over SO(3) and a point symmetry group Ω, the mapping is defined as:

$$k(R_i) = \underset{0<j<|G|}{\text{argmin}}\ \min\left\{\text{distance}\left(\omega R_i, R_j\right) \mid \forall \omega \in \Omega\right\}$$

During training, the orientation labels $L_i \in \mathbb{R}^{|G|}$ are encoded as sparse vectors containing $|\Omega|$ nonzero values, with the remaining elements set to zero.

$$L_i^j = \begin{cases} \frac{1}{|\Omega|} & if\ j \in\ k(R_i) \\ 0 & \text{otherwise} \end{cases}$$

To ensure that the less reliable labels are not used during training, we also exclude particles with the RELION metadata value rlnMaxValueProbDistribution below 0.01 from the training set.

4.6. Deep learning model

We employ an upgraded version of our previously published deep learning model[28] (see Supplementary Table 8 for a performance comparison). The model builds upon the Image2Sphere architecture[38]. It uses first an encoder network to compute N-dimensional 2D features for the input image. These 2D features are then orthographically projected onto a 3D hemisphere. The projected features are interpreted as spherical harmonics and subsequently processed with a group equivariant convolution layer with support in S^2, followed by a group equivariant convolution with SO(3) support. Finally, the output of the SO(3) convolution, which takes the form of Wigner D-matrix coefficients, is projected into a discretized representation of SO(3) with 294,912 rotation matrices (corresponding to a discretization of ~3.7°, HEALPix order 4[39]).

In our present implementation, we replace the original ResNet encoder[40] with a U-Net[41] with 5 encoder blocks and 4 decoder blocks. This allows for larger spatial dimensions in the 2D feature map representing the image, which should lead to better expression upon projection onto the hemisphere.

We also refined the method for estimating the output rotation matrix $predR_i$. The original implementation directly selected the rotation matrix with the highest probability from a discretized SO(3) grid:

$$predR_i = G_K \quad \text{where} \quad k = \underset{0<j<|G|}{\operatorname{argmax}} \mathrm{P_{nnet}}(R_j|x_i)$$

where, $G = \{R_j\}$ is the set of rotation matrices in the grid that sample SO(3), and $\mathrm{P_{nnet}}(R_j|x_i)$ is the model-estimated probability that R_j represents the pose of particle x_i. In our current approach, we compute the rotation matrix as a weighted average over neighbouring grid points. Specifically, the predicted rotation matrix $predR_i$ is calculated as:

$$predR_i = \mathrm{ProjectOntoSO3}(\sum_{j \in N(k)} \mathrm{P_{nnet}}(R_j|x_i)\, R_j)$$

where $N(k)$ represents the neighbour grid indices of the k-th index, $\mathrm{ProjectOntoSO3(M)} = \mathrm{UV^T}$, and U and V are obtained via Singular Value Decomposition $\mathrm{M} = \mathrm{U}\,\Sigma\,\mathrm{V^T}$.

We further improved the method by symmetrizing predictions according to the protein point symmetry:

$$\mathrm{P}(R_j|x_i) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \mathrm{P_{nnet}}(\omega R_j|x_i)$$

where Ω is the set of rotation matrices given a point symmetry group. As in our previous work, we employ a weighted cross-entropy loss function:

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{|G|} -conf_i \cdot L_i^j \cdot \log \mathrm{P_{nnet}}(R_j|x_i)$$

where $conf_i$represents the pose reliability estimate given by rlnMaxValueProbDistribution RELION score and L_i^j is the label value for the i-th particle and j-th rotation matrix (see Methods Section 4.5). In our current implementation, we have enhanced the loss function by incorporating label smoothing ($\epsilon = 0.05$)[42] to mitigate overfitting and account for potential inaccuracies in the data labels. Additionally, we implement gradient clipping (default value of 5, value-based) to ensure stable training dynamics.

For a comprehensive overview of all model hyperparameters, please refer to Supplementary Material Section 8.

4.7. Local refinement implementation

A custom PyTorch-based[43] implementation of the template matching algorithm[10] has been used in this work as a proof of concept to illustrate the capabilities of our neural network. The algorithm begins by generating, for each particle x_i, a grid of orientations (G) around the Euler angles predicted by the neural network $R_i^0 = \left(\alpha_i^0, \beta_i^0, \gamma_i^0\right)$

$$G(x_i) = \left\{ R_i^0 + (\Delta\alpha, \Delta\beta, \Delta\gamma) \middle| \Delta\alpha, \Delta\beta, \Delta\gamma \in \{\mathrm{i}\delta \mid \mathrm{i} \in \mathbb{Z}, -d \leq \mathrm{i}\delta \leq d\}\right\}$$

where $2d$ is the grid width (12° by default), and δ is the angular step size (2° by default).

Then, for each orientation in the global pool of orientations, the reference volume is projected using the `torch-fourier-slice` Python package[45]. Next, for each particle associated to the orientation, the cross-correlation (CC) between the particle and the CTF modulated projection is computed.

$$CC(x_i, R) = \mathcal{F}^{-1}(\mathcal{F}(x_i)^* \, CTF_i \, \Pi_R \, \mathcal{F}(V))$$

where V represents the reference volume, Π_R is the projector operator at orientation R, CTF_i is the contrast transfer function of the i-th particle and $\mathcal{F}$ and $\mathcal{F}^{-1}$ represents the Fourier transform and inverse Fourier transform respectively. For efficiency, the Fourier transform of the volume and the particles, as well as CTFs of the particles are precomputed.

After that, the orientation of a particle is assigned to the orientation that yielded the highest cross-correlation peak:

$$R_i^{CC} = \underset{r \in G(x_i)}{\operatorname{argmax}} \, CC_{peak}(x_i, r)$$

where

$$CC_{peak} = \max CC(x_i, r)$$

and the shifts of the particle are estimated as the location of the highest cross-correlation peak

$$\delta_{w,h} = \underset{w,h}{\operatorname{argmax}} \, CC(x_i, R_i^{CC})$$

where w, h represent the pixel coordinates that are at a distance smaller than 20% of the particle radius from the centre of the image. Finally, a weight for the particle to be used at reconstruction is computed by combining the score given by the neural network and an estimation of the significance of the orientation in a similar fashion as Sorzano et al.[46]:

$$W\left(R_i^{CC}\right) = \mathrm{P}_{\mathrm{nnet}}\left(R_i^0 | x_i\right) \left(1 - \mathrm{P}\left(CC_{peak}\,(x_i, r) > CC_{peak}\left(x_i, R_i^{CC}\right)\right)\right)$$

where $\mathrm{P}_{\mathrm{nnet}}(R_i^0|x_i)$ is the probability given by the neural network to the pose R_i^0 and $\mathrm{P}\left(CC_{peak}\ (x_i, r)\right)$ is the probability of a cross-correlation peak value for the i-th particle at a given pose. For computational reasons, $\mathrm{P}\left(CC_{peak}\ (x_i, r)\right)$ is assumed to follow a normal distribution whose parameters are estimated as

$$\mu_{x_i} = \frac{1}{|G(x_i)|} \sum_{r \in G(x_i)} CC_{peak}\ (x_i, r)$$

$$\sigma_{x_i} = \sqrt{\frac{\sum_{r \in G(x_i)} \left(CC_{peak}\ (x_i, r) - \mu_{x_i}\right)^2}{|G(x_i)|}}$$

where $|G(x_i)|$ is the number of orientations in the grid of orientations for the particle x_i.

4.8. Particle pruning

After the neural network has predicted a probability distribution for the orientations of a particle x_i, we compute its direction-normalized pruning score as a robust z-score based on the median absolute deviation (MAD) of the distribution of predicted scores at a given orientation obtained from the validation set at training time. Particles with a robust z-score below a given threshold are regarded as "bad" particles and ignored in subsequent steps.

The median and MAD calculations required to compute our directional robust z-scores can be regarded as part of the training process and are computed as follows. First, the particles in the validation set are assigned to one of the Kcones in which we discretize the projection sphere (first and second Euler angles according to RELION's convention). By default, we use cones of 15° (HEALPix order of 2) to ensure a sufficient number of particles in each cone. Then, the cone median and MAD are computed as

$$\mathrm{median}^K = \mathrm{median}(Y^K)$$

$$\mathrm{MAD}(K) = \mathrm{median}(Y^K - \mathrm{median}^K)$$

where

$$Y^K = \{\mathrm{P}_{\mathrm{nnet}}(R_i^0|x_i) |\ i \in \mathrm{coneIdxs}(K) \wedge \mathrm{predictedCone}(x_i) = K\}$$

Here, $\mathrm{coneIdxs}(K)$ is the set of particle indices that correspond to the cone Kaccording to the ground truth pose, and $\mathrm{predictedCone}(x_i)$ is the cone index of the highest score orientation of the i-th particle.

The median^K and $\mathrm{MAD}(K)$ estimated from the validation dataset are then used to compute the per-cone robust z-score of new particles following the next equation:

$$rzscore(y_j) = \frac{\mathrm{P}_{\mathrm{nnet}}(R_j^0|x_j) - \mathrm{median}^{\mathrm{predictedCone}(x_j)}}{\mathrm{MAD}(\mathrm{predictedCone}(x_j))}$$

In this study, for robust-z score calculations, we define the particle score $\mathrm{P}_{\mathrm{nnet}}(R_j^0|x_j)$ as the sum of the top-10 highest probabilities predicted by the neural network. This approach makes the score more robust in cases where the neural network is uncertain about particle orientation, as the top-1 probability might not capture enough of the learnt signal.

There are several approaches to estimate a per-cone robust z-score threshold. For instance, it could be computed based on the expected fraction of “bad” picks given historical data. Here we estimate it by comparing the distribution of per-cone robust z-scores of the validation dataset against a dataset of "bad" particles (particles that were removed after 2D classification). See Supplementary Material Section 9 for details about automatic thresholds.

5. Acknowledgments

R.S.-G. is a Sustaining Innovation Postdoctoral Research Associate at Astex Pharmaceuticals and thanks Astex Pharmaceuticals for funding. We acknowledge the Cambridge Pharmaceutical Cryo-EM Consortium and the Astbury BioStructure Laboratory for access to their cryo-EM facilities.

6. Author contributions

RSG. co-designed, implemented, and tested the method, performed image processing and wrote the first draft of the manuscript. AA, AB, JR, and PAW prepared the protein samples and collected electron microscopy data. CP co-designed the method. MS prepared the training and testing data, analysed the results, and co-supervised the project. CD co-supervised the project and secured funding. All authors contributed to the writing of the final version of the manuscript.

7. Code availability

CryoPARES is available at https://github.com/rsanchezgarc/cryoPARES

8. Data availability

The cryo-EM maps generated with cryoPARES have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-55146 (β-galactosidase–B1), EMD-55241 (β-galactosidase–B2), EMD-55304 (PKM2–P1), EMD-55305 (PKM2–P2), EMD-55535 (GDH–G1), EMD-55549 (GDH–G2), and EMD-55532 (GDH–G3). Atomic coordinates for β-galactosidase and PKM2 correspond to the models reported by Saur et al. (2020) and are available in the Protein Data Bank (PDB) under accession codes 6TTE (β-galactosidase–B1), 6TSK (β-galactosidase–B2), 6TTF (PKM2–P1), and 6TTQ (PKM2–P2). The GDH models generated in this study have been deposited in the PDB under accession codes 9T4W (G1), 9T4X (G2), and 9T4U (G3). Raw particle images used to reconstruct the GDH maps are available in the Electron Microscopy Public Image Archive (EMPIAR) under accession numbers EMPIAR-13104 (G1), EMPIAR-13105 (G2), and EMPIAR-13106 (G3).

9. References

1. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat Methods* **13**, 24–27 (2015).
2. Zhang, M. *et al.* G protein-coupled receptors (GPCRs): advances in structures, mechanisms and drug discovery. *Signal Transduction and Targeted Therapy 2024 9:1* **9**, 1–43 (2024).
3. Saur, M. *et al.* Fragment-based drug discovery using cryo-EM. *Drug Discov Today* **25**, 485–490 (2020).
4. Evenseth, L. S. M., Gabrielsen, M. & Sylte, I. The GABAB Receptor—Structure, Ligand Binding and Drug Development. *Molecules 2020, Vol. 25, Page 3093* **25**, 3093 (2020).
5. Passmore, L. A. & Russo, C. J. Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol* **579**, 51–86 (2016).
6. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519–530 (2012).
7. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods 2017 14:3* **14**, 290–296 (2017).

8. Grant, T., Rohou, A. & Grigorieff, N. CisTEM, user-friendly software for single-particle image processing. *Elife* **7**, (2018).
9. De la Rosa-Trevín, J. M. *et al.* Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J Struct Biol* **184**, 321–328 (2013).
10. Penczek, P. A., Grassucci, R. A. & Frank, J. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**, 251–270 (1994).
11. Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17–24 (1996).
12. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**, 82–97 (1999).
13. Levy, A. *et al.* CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **13681 LNCS**, 540–557 (2022).
14. Sorzano, C. O. S. *et al.* Swarm optimization as a consensus technique for Electron Microscopy Initial Volume. *Applied Analysis and Optimization* **2**, 299–313 (2018).
15. Vargas, J., Álvarez-Cabrera, A. L., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* **30**, 2891–2898 (2014).
16. Gomez-Blanco, J., Kaur, S., Ortega, J. & Vargas, J. A robust approach to ab initio cryo-electron microscopy initial volume determination. *J Struct Biol* **208**, 107397 (2019).
17. Sigworth, F. J. A maximum-likelihood approach to single-particle image refinement. *J Struct Biol* **122**, (1998).
18. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc Natl Acad Sci U S A* **110**, 18037–18041 (2013).
19. Lucas, B. A., Himes, B. A. & Grigorieff, N. Baited reconstruction with 2D template matching for high-resolution structure determination in vitro and in vivo without template bias. *Elife* **12**, (2023).
20. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods 2012 9:9* **9**, 853–854 (2012).
21. Levy, A., Wetzstein, G., Martel, J., Poitevin, F. & Zhong, E. D. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Adv Neural Inf Process Syst* **35**, 13038–13049 (2022).
22. Gupta, H., McCann, M. T., Donati, L. & Unser, M. CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning. *IEEE Trans Comput Imaging* **7**, 759–774 (2021).
23. Jiménez-Moreno, A., Střelák, D., Filipovič, J., Carazo, J. M. & Sorzano, C. O. S. DeepAlign, a 3D alignment method based on regionalized deep learning for Cryo-EM. *J Struct Biol* **213**, 107712 (2021).
24. Chung, S. C. Cryo-forum: A framework for orientation recovery with uncertainty measure with the application in cryo-EM image analysis. *J Struct Biol* **216**, 108058 (2024).
25. Lian, R. *et al.* End-to-end orientation estimation from 2D cryo-EM images. *Acta Crystallogr D Struct Biol* **78**, 174–186 (2022).
26. Levy, A., Wetzstein, G., Martel, J., Poitevin, F. F. & Zhong, E. D. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Adv Neural Inf Process Syst* **35**, (2022).
27. Levy, A. *et al. CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images*. *Lecture Notes in Computer Science (including*

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) vol. 13681 LNCS 540–557 (Springer Science and Business Media Deutschland GmbH, 2022).
28. Sanchez-Garcia, R., Saur, M., Vargas, J., Poelking, C. & Deane, C. M. CESPED: A benchmark for supervised particle pose estimation in cryo-EM. *Phys Rev Res* **6**, 023245 (2024).
29. Sorzano, C. O. S. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol* **171**, 197–206 (2010).
30. Gomez-Blanco, J., Kaur, S., Strauss, M. & Vargas, J. Hierarchical autoclassification of cryo-EM samples and macromolecular energy landscape determination. *Comput Methods Programs Biomed* **216**, 106673 (2022).
31. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
32. Sorzano, C. O. S. *et al.* On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Acta Crystallogr D Struct Biol* **78**, 410–423 (2022).
33. Vargas, J., Melero, R., Gómez-Blanco, J., Carazo, J. M. & Sorzano, C. O. S. Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis. *Scientific Reports 2017 7:1* **7**, 1–14 (2017).
34. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1–12 (2016).
35. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
36. Liu, L. *et al.* ON THE VARIANCE OF THE ADAPTIVE LEARNING RATE AND BEYOND. in *8th International Conference on Learning Representations, ICLR 2020* (International Conference on Learning Representations, ICLR, 2020).
37. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *37th International Conference on Machine Learning, ICML 2020* **PartF168147-3**, 1575–1585 (2020).
38. Klee, D. M., Biza, O., Platt, R. & Walters, R. Image to Sphere: Learning Equivariant Features for Efficient Pose Prediction. *International Conference on Learning Representations* http://arxiv.org/abs/2302.13926 (2023).
39. Gorski, K. M. *et al.* HEALPix -- a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere. *Astrophys J* **622**, 759–771 (2004).
40. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vols 2016-Decem 770–778 (IEEE Computer Society, 2016).
41. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention-MICCAI* vol. 9351 234–241 (Springer, Cham, 2015).
42. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016-December**, 2818–2826 (2015).
43. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems* vol. 32 8026–8037 (2019).
44. Zonca, A. *et al.* healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *J Open Source Softw* **4**, 1298 (2019).
45. Burt, A. & Chaillet, M. teamtomo/torch-fourier-slice. Preprint at (2024).
46. Sorzano, C. O. S. *et al.* A statistical approach to the initial volume problem in Single Particle Analysis by Electron Microscopy. *J Struct Biol* **189**, 213–219 (2015).