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: deane@stats.ox.ac.uk, michael.saur@astx.com

Abstract:

Cryo-Electron Microscopy (cryo-EM) is a pivotal tool for determining the 3D structures of biological macromolecules. Current cryo-EM workflows, while effective, are computationally demanding and require manual intervention, creating bottlenecks for use in high-throughput scenarios such as structure-based drug discovery. Often in structure-based drug discovery, one can assume that all instances of a protein are equivalent at the resolutions needed for alignment and it therefore should be possible to harness information about particle poses from previous refinements. Current methods, however, do not leverage this form of prior knowledge, instead 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. These features, together with its single-pass operation, can enable real-time reconstructions that provide feedback during data acquisition. We demonstrate cryoPARES's effectiveness through the 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 (Nogales, 2015) with applications in structure-based drug discovery (Evenseth et al., 2020; Saur et al., 2020; Zhang et al., 2024). At the heart of this technique lies Single-Particle Analysis (SPA), which aims to reconstruct the 3D structure of a macromolecule from hundreds of thousands of 2D projection images, termed particles. These particles adopt random orientations (potentially not uniformly distributed) in the ice layer (Passmore & Russo, 2016), with their orientation initially unknown. The main challenge in SPA lies in estimating these orientations amidst large levels of noise and multiple sources of error.

Typically, the orientation estimation is performed via angular refinement (De la Rosa-Trevín et al., 2013; Grant et al., 2018; Ludtke et al., 1999; Penczek et al., 1994; Punjani et al., 2017; Scheres, 2012; Van Heel et al., 1996) that begin with a low-resolution reference volume against which each experimental particle is compared in all possible orientations. The best-matching orientation(s) are then assigned to each particle, potentially with different weights. A new volume is then reconstructed based on these assignments, 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 (Levy, Poitevin, et al., 2022).

For novel samples, the initial reference volume must be estimated *ab initio* (Gomez-Blanco et al., 2019; Punjani et al., 2017; Sorzano et al., 2018; Vargas et al., 2014). However, when prior knowledge exists, this information can be leveraged to generate an alternative 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 (Henderson, 2013; Sigworth, 1998), using a higher resolution initial volume that is closer to the final solution can improve the convergence of angular refinement algorithms, helping with the processing of challenging samples and potentially accelerating the refinement process, although this gain can 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 liganded protein structure determination (Lucas et al., 2023). This is because pose estimation is mostly driven by low- to medium-resolution frequencies (Scheres & Chen, 2012), at which similar proteins are nearly identical. An extreme application of this idea is employed in the baited reconstruction by template-matching approach (Lucas et al., 2023), where a high-resolution reference obtained from an unbound atomic model was used in a single round of template-matching alignment to identify bound ligands.

For ligand determination, template bias presents fewer concerns since ligands are absent from the reference apo structure (Lucas et al., 2023). While false positives may still occur, (e.g., endogenous lipids occupying binding sites being misidentified as ligand hits), these misidentifications can be resolved by improving the map resolution or simply excluding the endogenous ligand from the structure during local refinement. Similarly, false negatives caused by significant conformational changes, flexibility, etc. are less problematic as either poor quality of the results or the presence of model bias should be easily detectable (S. Chen et al., 2013; Lucas et al., 2023; Sorzano et al., 2022; Vargas et al., 2017), leaving the option for reprocessing using classical methods. Thus, the idea of reusing the knowledge from previous refinements offers the potential of more efficient structural determination of protein-ligand complexes.

Despite these examples, the computational burden of processing multiple datasets of similar proteins or of the same protein under different conditions still contributes to the limited adoption of cryo-EM for high-throughput analysis. 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 (Chung, 2024; Gupta et al., 2021; Jiménez-Moreno et al., 2021; Levy, Poitevin, et al., 2022; Levy, Wetzstein, et al., 2022; Lian et al., 2022). 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 (Levy, Poitevin, et al., 2022; Levy, Wetzstein, et al., 2022), others can leverage previous structural knowledge through supervised learning (Chung, 2024; Jiménez-Moreno et al., 2021; Lian et al., 2022; Sanchez-Garcia et al., 2024).

Supervised approaches train surrogate models of the alignment algorithm using pre-computed poses from previous refinements. While requiring pre-aligned particles may be regarded as a potentially insurmountable limitation for *ab initio* refinement, this is less problematic in many practical applications, particularly in 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 (Sanchez-Garcia et al., 2024) which demonstrated the feasibility of supervised deep learning models for pose estimation across multiple specimens, we present here cryoPARES (cryo Pose Assignment for Related Experiments via Supervision), a fully automatic pose estimation tool optimized for ligand screening applications where a single model trained on an apo structure can be used to efficiently process multiple ligand-bound datasets without manual intervention. Our implementation not only surpasses its predecessor's performance on the CESPED benchmark (Sanchez-Garcia et al., 2024) (see Supplementary Table 1) but also demonstrates robust generalization across diverse samples. Furthermore, our tool introduces novel capabilities for automated particle pruning, eliminating the need for manual intervention in particle selection, and 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.

2. Methods

2.1. 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 (Saur et al., 2020): EMPIAR-10644 (β-galactosidase bound to phenylethyl β-D-thiogalactopyranoside/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.

2.2. Datasets image processing

The particle extraction process was identical for both training and testing datasets. CTF estimation was performed using GCtf 1.18 (Zhang, 2016), 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 (Zivanov et al., 2018), 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 Table 1 for additional details.

Table 1 Dataset and image processing details.

Target	Ligand	Microscope and detector	Symmetry	Particle size (pixels)	# extracted particles	# particles after 2D	Low pass 3D (Å)
BGAL	Apo	Krios F3	D2	476	369,690	177,824	20
BGAL	Compound B1	Krios F3	D2	476	134,171	56,764	20
BGAL	Compound B2	Krios F3	D2	476	124,687	42,125	20
PKM2	Apo	Krios F3	D2	334	212,244	125,324	20
PKM2	Compound P1	Glacios F3	D2	334	353,678	254,070	20
PKM2	Compound P2	Krios F3	D2	334	267,921	127,035	20
GDH	Apo	Krios F3	D3	356	156,232	111,645	15
GDH	Compound G1	Krios F3	D3	356	371,299	277,213	15
GDH	Compound G2	Krios F3	D3	356	182,538	120,465	15
GDH	Compound G3	Krios F3	D3	356	409,693	350,312	15

2.3. 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 (Liu et al., 2020) 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 contrastive loss (T. Chen et al., 2020) was also implemented to favour 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 2.8 for details).

2.4. 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 2.8). The remaining particles then undergo local refinement to improve their orientations and/or estimate in-plane shifts (see Methods Section 2.7).

2.5. 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. (Levy, Poitevin, et al., 2022), 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:

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

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.

2.6. Deep learning model

We employ an updated version of our previously published deep learning model (Sanchez-Garcia et al., 2024). The model builds upon the Image2Sphere architecture (Klee et al., 2023). 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 (Gorski et al., 2004)).

In our present implementation, we replace the original ResNet encoder (He et al., 2016) with a U-Net (Ronneberger et al., 2015) 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:

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

where, $G = \{R_j\}$ is the set of rotation matrices in the grid that sample SO(3), and $\mathrm{P}_{\mathrm{nnet}}(R_j|x_i)$ 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:

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

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:

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

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:

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

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 2.5). In our current implementation, we have enhanced the loss function by incorporating label smoothing ($\epsilon = 0.05$) (Szegedy et al., 2015) 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.

2.7. Local refinement implementation

A custom PyTorch-based (Paszke et al., 2019) implementation of the template matching algorithm (Penczek et al., 1994) 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 $R_i^0 = \left(\alpha_i^0, \beta_i^0, \gamma_i^0\right)$

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

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

To optimize computations and reduce the need for projecting very similar orientations, the resulting angles in the grid are discretized into a fixed number of valid orientations using HEALPix (Gorski et al., 2004) through the healpy Python package (Zonca et al., 2019). The HEALPix order is chosen such that the resolution of the discretization is equal to or smaller than δ, with a default of 5 (~1.8°). This discretization allows us to group together particles that need to be compared against the same reference, reducing the number of projection operation required.

Then, for each orientation in the global pool of orientations, the reference volume is projected using the `torch-fourier-slice` Python package (Burt & Chaillet, 2024). 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. (Sorzano et al., 2015):

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

where $\mathrm{P}_{\mathrm{nnet}}\left(R_i^0 \middle| x_i\right)$ 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.

2.8. 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

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

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

where

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

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

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

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

In this study, for robust-z score calculations, we define the particle score $\text{P}_{\text{nnet}}\left(R_j^0 \middle| x_j\right)$ 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.

3. Results and discussion

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

CryoPARES has been designed to perform efficient SPA for one particular use case: the structural determination of specimens for which a pre-existing cryo-EM structure is already available. Although not limited to, this use case is of great relevance in structure-based drug discovery, where it is common to have several structures of a target protein before starting a drug discovery campaign. In such scenarios, where many different ligands are screened against a given protein, the computational cost of traditional refinement algorithms is approximately the same for each sample, as these algorithms start from scratch in each refinement. However, since the samples are highly similar, it is reasonable to believe that knowledge about particle alignment could be reused from experiment to experiment.

CryoPARES's deep learning model can produce an estimate of the probability distribution of the pose of a given particle. After being trained on a set of pre-aligned particles, the cryoPARES model can then be used to infer the poses of new particles from different samples of the same (or similar) specimens at high speed, leading to predicted orientations of high enough quality so that a high-resolution density can be attained after a quick baited local refinement step. Contrary to current methods, cryoPARES's primary computational investment is in the initial training phase. Subsequent analyses of similar specimens can be performed very rapidly, significantly reducing the overall computational burden of large-scale structural studies (see Figure 1).

Given the probabilistic nature of its pose predictions, cryoPARES can be used directly after the particle extraction step, as it automatically filters out junk particles (see Results Section 3.5). As a consequence, cryoPARES enables fully automatic, reproducible reconstruction workflows, eliminating the subjectivity problems inherent in some classical image processing steps such as 2D classification. Unlike traditional iterative approaches, and similar to baited reconstruction, cryoPARES's one-shot workflow determines orientations in a single pass. This feature, combined with its high processing speed (see Results Section 3.4), can enable real-time image processing, potentially providing valuable feedback during data acquisition. Such automation and standardization are of capital importance for high-throughput structural studies where speed, consistency and reproducibility are paramount.

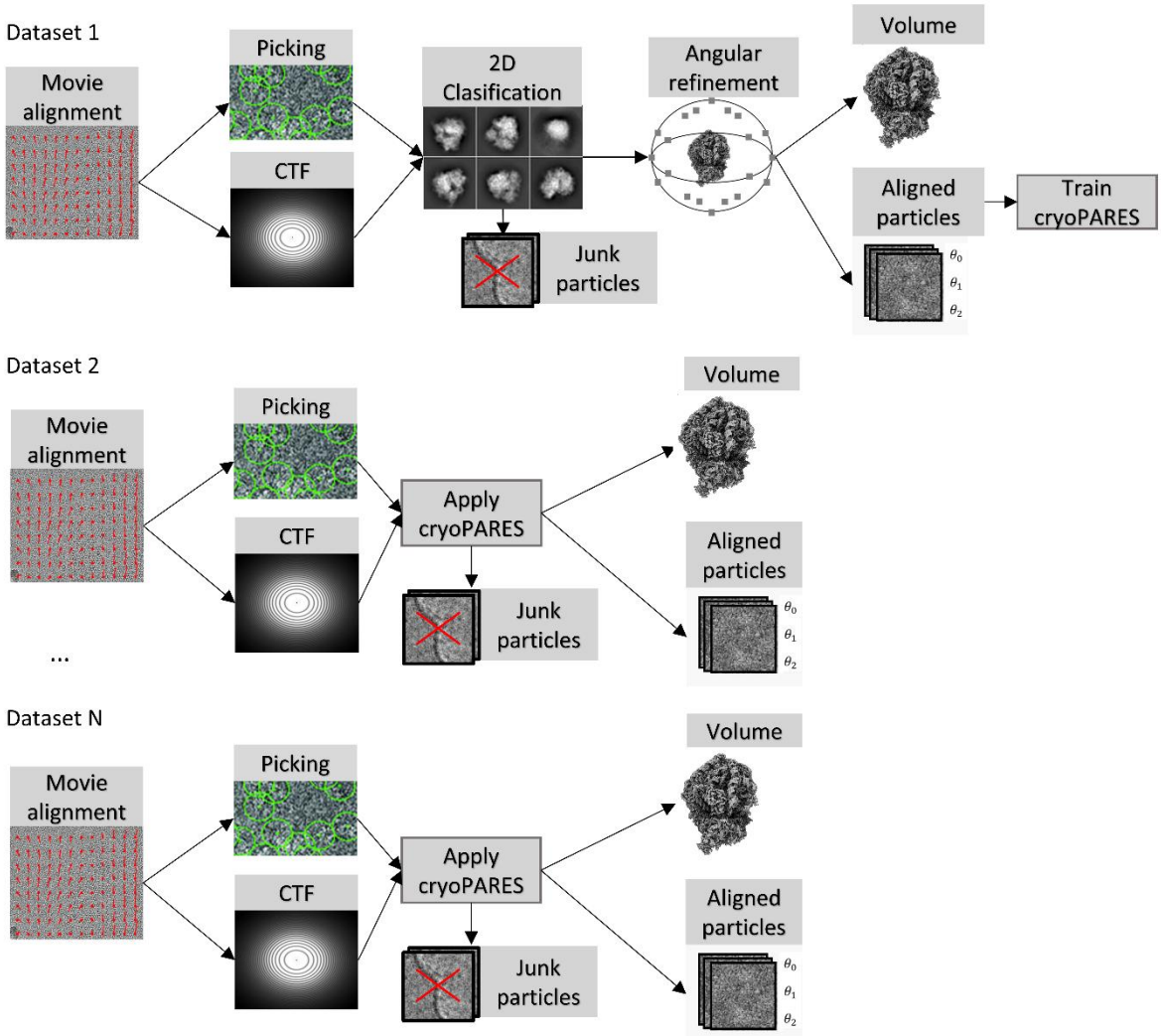

Figure 1. CryoPARES models are trained from a pre-aligned "good" particles dataset that is obtained via classical image processing pipelines. The model is trained in a supervised manner to predict the probability distributions of the pose of each particle. The junk particles in the training dataset can be used to calibrate the particle goodness score (See Methods Section 2.8). 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.

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

In order to illustrate the utility of our tool, we employed cryoPARES to resolve 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 the local refinement step was run using as a reference the reconstruction of the apoprotein low-pass filtered to 6 Å.

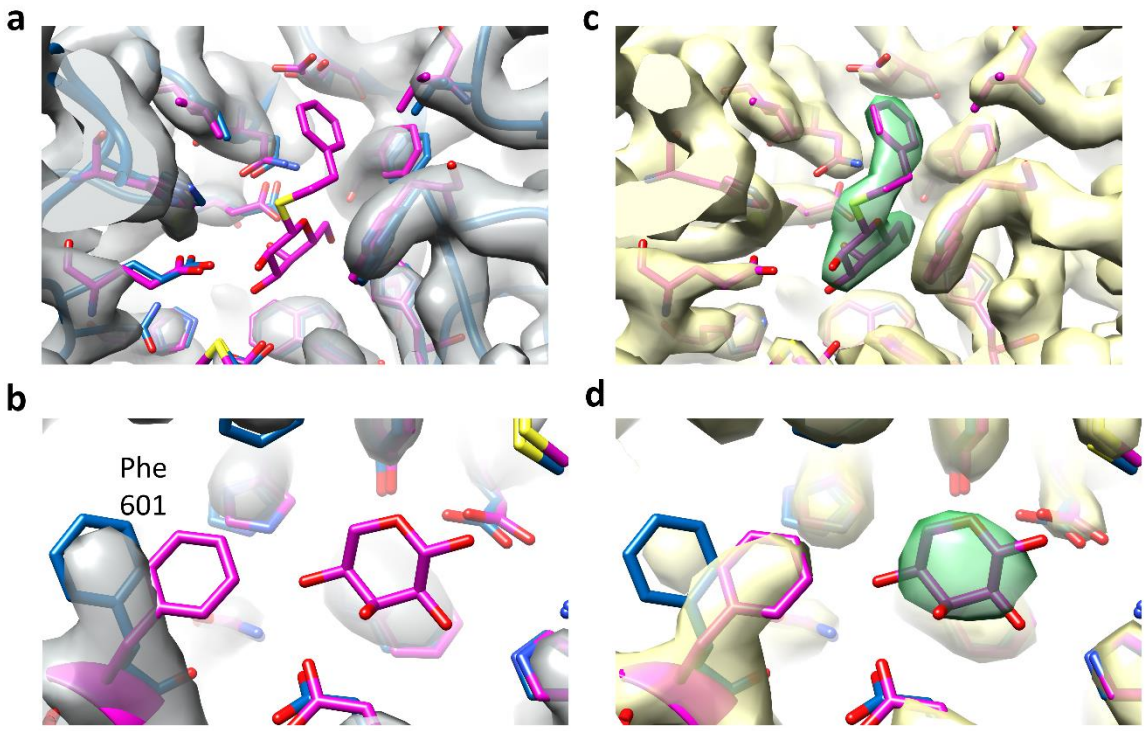

Figure 2. Reconstructed maps of β-galactosidase (BGAL) at the ligand binding site using orientations inferred with cryoPARES. a) Map of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound B1 (pink). b) Same as (a) but with the atomic 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 phenylalanine 601 in the Compound B2 case (b vs d).

We first tested the method on the previously published datasets of the *E. coli* β-galactosidase (BGAL) bound to Compounds B1 and B2 (Saur et al., 2020). Figure 2 shows the reconstructed densities for the Compound B1 (c) and Compound B2 (d), compared to the structure of the apoprotein used as the template for local refinement and whose particles were used to train the deep learning model (a,b). The map features of the ligands using the orientations estimated with cryoPARES fit well the ground-truth atomic model of the protein-ligand complexes built in our previous publication. 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 (see 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 3.4). Similarly, the map-to-model FSC resolutions (threshold 0.5) were 2.9 Å and 3.2 Å, confirming the overall similarity of the reconstructed map and the ground truth atomic model.

Our second test was conducted on our previously published datasets for the human pyruvate kinase M2 (PKM2) bound to Compounds P1 and P2 (Saur et al., 2020). Similar to our observations on the first test case, the densities for the ligand-bound dataset reconstructed with cryoPARES poses matched the ground-truth atomic models (see Figure 3). In this case, the FSC resolution obtained with cryoPARES poses was closer to the resolution obtained with the RELION workflow: 3.5 Å and 3.3 Å for cryoPARES (see Figure 5 c-d), and 3.4 Å and 2.8 Å using the standard RELION workflow. 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, suggesting the method is robust under different imaging conditions.

The last test case consists of two previously unpublished datasets of glutamate dehydrogenase (GDH) bound to Compound G1 and Compound 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 on the poses inferred with cryoPARES on the ligand-bound datasets. 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 3.0 Å and 3.1 Å.

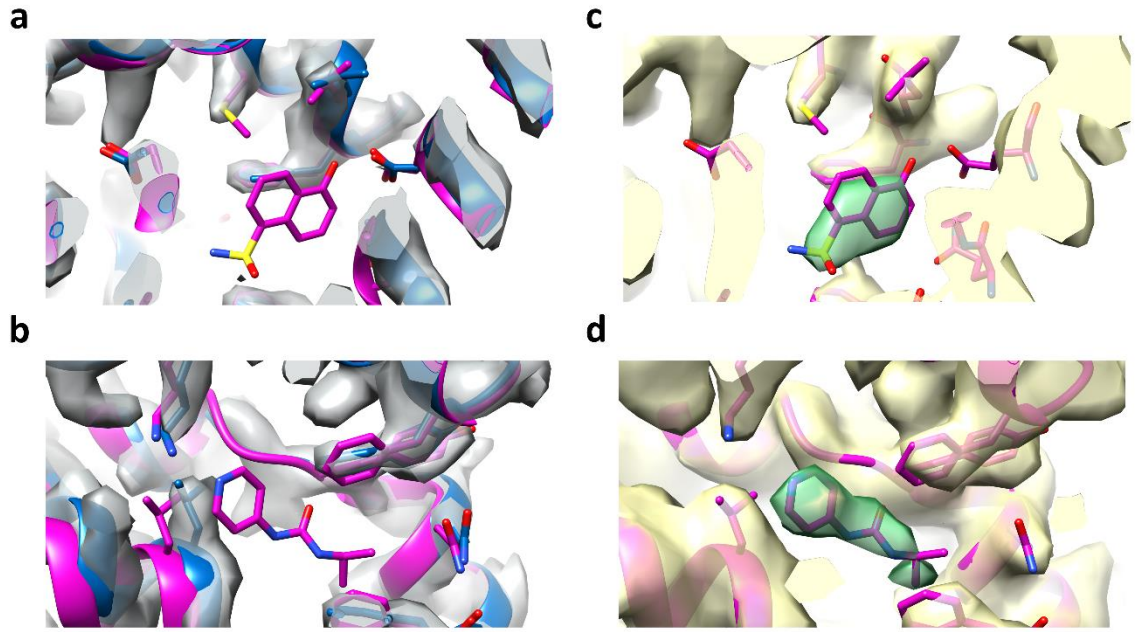

Figure 3. Reconstructed maps of pyruvate kinase M2 (PKM2) at the ligand binding site using orientations inferred with cryoPARES. a) Density of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound P1 (pink). b) Same as (a) but with 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).

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, the several observations strongly indicate that the effect of this, if present, is small on the structures presented above: 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 about template bias can be found in Results Section 3.7.

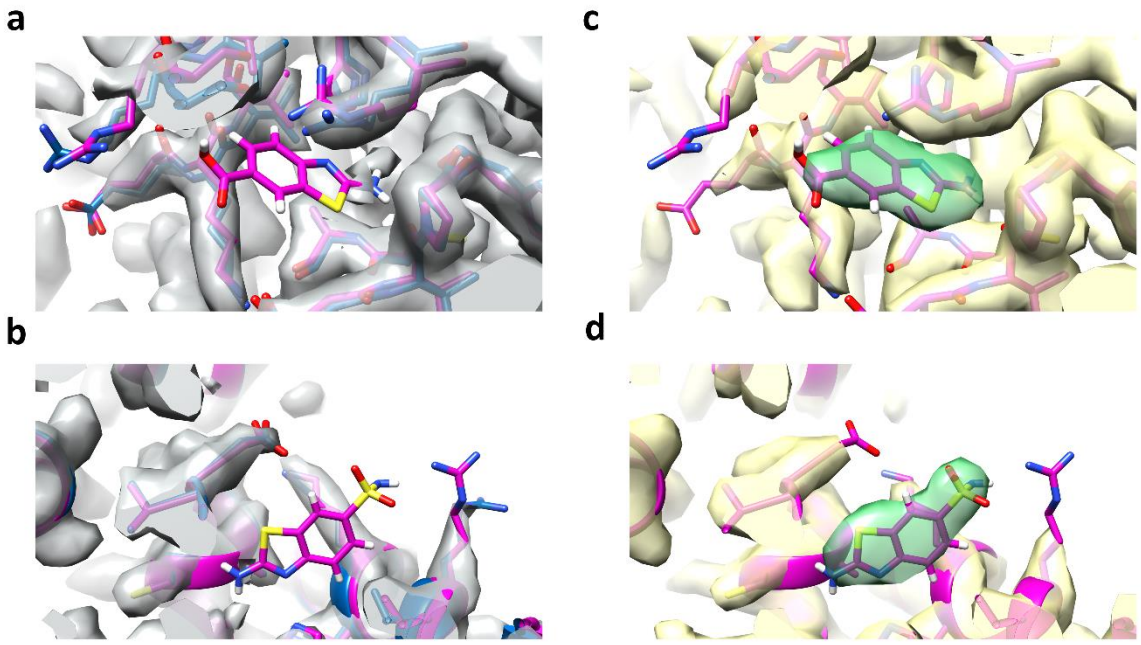

Figure 4. Reconstructed maps of glutamate dehydrogenase (GDH) at the ligand binding site using orientations inferred with cryoPARES. a) Density of the apoprotein (grey) with its associated atomic model (blue) and the aligned atomic model of the protein bound to Compound G1 (pink). b) Same as (a) but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

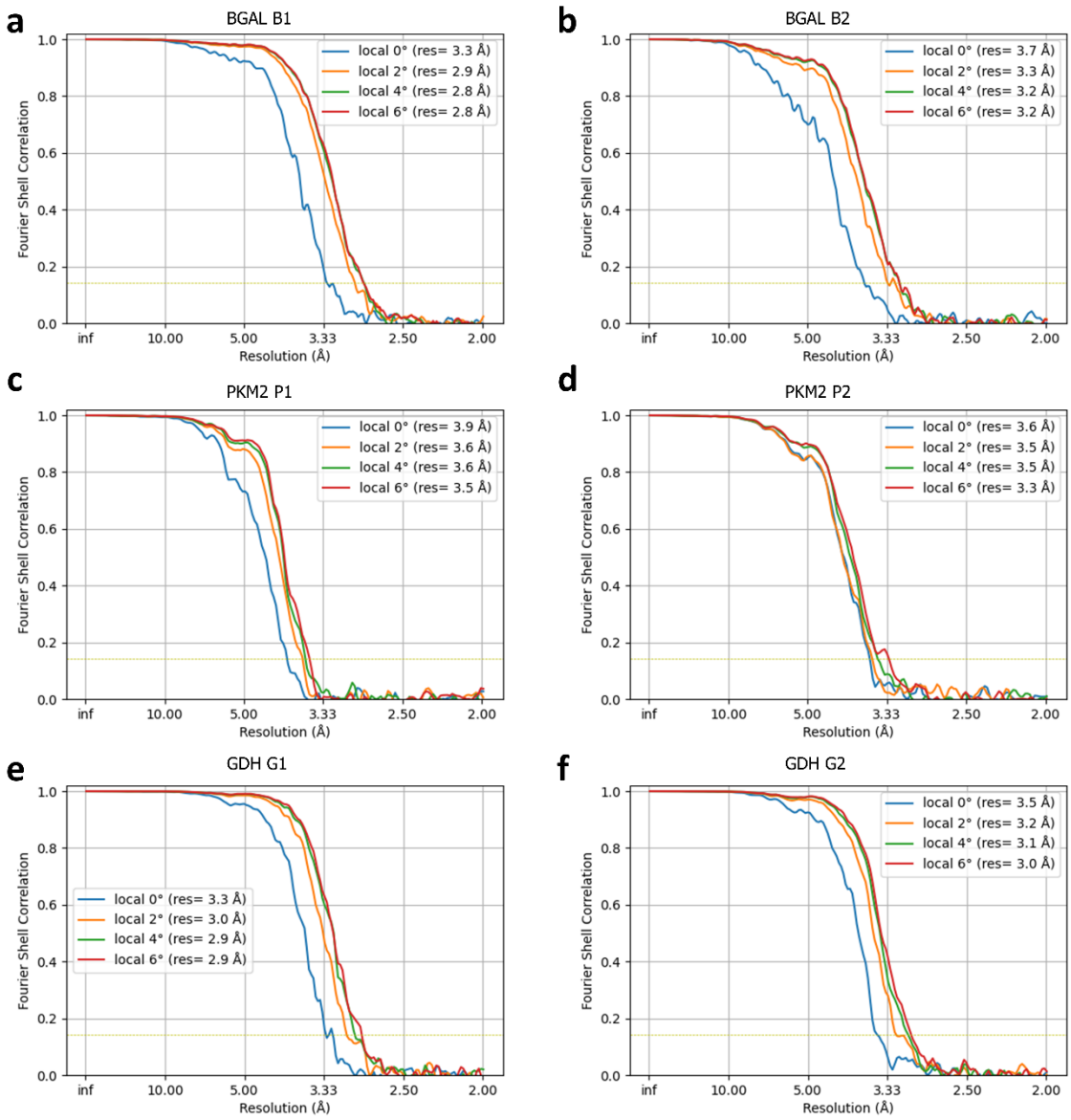

Figure 5. FSC curves (half maps) for the maps obtained by cryoPARES at different levels of local refinement angular search range (0° blue, ±2° orange, ±4° green, ±6° red) for the protein-ligand complexes BGAL bound to Compound B1 (a) and Compound B2 (b), PKM2 bound to Compound P1 (c) and Compound P2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

3.3. CryoPARES predicts accurate global alignment parameters

We first evaluated the accuracy of cryoPARES using simulated data, where ground truth particle orientations are known. On these datasets, the neural network achieved validation average 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 presents a unique challenge due to the lack of ground truth poses.

For real 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% (Jiménez-Moreno et al., 2021). This substantial variability in pose estimation between identical RELION runs underscores that high levels of correspondence between different methods may actually be undesirable, as it could indicate shared systematic biases rather than accurate pose determination. 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. As can be seen in all cases, the FSC curve improves with a local refinement with a search range as small as ± 2°, while extending to ± 4° yields only marginal gains, and almost no extra improvement when extended from ± 4° to ± 6°. This suggests that many of the orientations inferred by the neural network are as close to the "correct" orientation as 2° to 4°.

3.4. CryoPARES performs real time particle alignment

Supplementary Table 4 summarizes the measured running times for pose estimation across different particle datasets. The cryoPARES neural network achieves throughputs exceeding 30,000 particles min^{-1} GPU^{-1} during neural network inference. This speed means that the neural network can process a 250,000-particle stack in less than 8 minutes. Assuming that 50% of the picked particles are junk particles that do not require local refinement, only 7 additional minutes are needed to run a local refinement to improve the alignment, resulting in a total runtime of approximately 15 minutes on a single GPU. 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.

The run times of cryoPARES are more than one order of magnitude faster than a RELION auto-refinement job and comparable to the non-open source cryoSPARC homogeneous refinement job (see Supplementary Table 5). More importantly, considering that a high-end microscope can collect up to 600 micrographs per hour, and assuming 400 particles per micrograph, a minimum processing throughput of 4,000 particles min^{-1} would allow for real-time analysis. CryoPARES exceeds this threshold, delivering >10,000 particles min^{-1} GPU^{-1} including the local refinement step. This means that a single GPU card can process particles far faster than even the most demanding data acquisition rates.

Unlike current methods, which 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.

3.5. CryoPARES can perform automatic particle pruning

CryoPARES provides probability estimates for the inferred poses of the particles. Those estimates can be used to detect putative junk particles or particles that are likely to have their poses wrongly estimated. While it is possible to directly use the probability of the predicted pose as a pruning score, we have observed that the distribution of predicted probabilities can be heavily dependent on

the orientation, so that particle pruning based on the probability alone could lead to reconstructions affected by artefactual preferred orientations (see Supplementary Figure 3). Instead, we have designed a directional normalization scheme that allows us to compute a more robust score for particle pruning (see Methods Section 2.8).

Figure 6 shows the histograms of direction-normalized scores of "good" particles (particles that survived a 2D classification round), compared to the score for particles that were ruled out by 2D classification. Due to the imperfections of existing cleaning algorithms, there are always false positives and false negatives in the sets of "good" and "bad" particles, respectively. Thus, a perfect separation of the two categories according the 2D classification criteria is not feasible nor desirable, as it will imply that our method would suffer from the same errors. Figure 6 shows that, in all cases, the cryoPARES scores for the "good" particles tend to be larger than the scores for "bad" particles, with very few "bad" particles showing positive values. This confirms that particle pruning can be conducted by setting a simple threshold based on this score. The results presented in the previous two sections have employed this method.

To further assess the pruning capabilities of our algorithm, we ran RELION auto-refine jobs of 1) datasets after 2-D classification, 2) datasets pruned with cryoPARES and 3) datasets in which a subset of particles was randomly selected (control). In the last two cases, to maintain comparability, the number of selected particles was set to match the number of particles surviving 2D classification. Note that, for this reason, the thresholds used for cryoPARES predictions in this section are different from the automatically determined threshold (see Supplementary Material Section 9) used in the previous sections.

Table 2 shows that the resolution obtained with particles pruned by cryoPARES is comparable to that achieved through 2D classification pruning. Since for some of our datasets the effect of clean-up is less prominent (random particle selection yields similar resolution), we also analysed two more challenging datasets: a membrane protein bound to two different ligands (MPL1 and MPL2). In these cases, comparing resolutions between 2D-classified and randomly selected particles clearly demonstrates the significant impact of particle clean-up procedures and the effectiveness of cryoPARES doing that task.

Table 2. 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

3.6. CryoPARES is robust against preferred orientations

CryoPARES, similar to baited reconstruction approaches (Lucas et al., 2023), is not affected by the attractor problem (Gomez-Blanco et al., 2022; Sorzano et al., 2010), as the reference image in all views has the same Signal-to-Noise Ratio. The attractor problem occurs during conventional refinement when views with higher particle counts (thus higher SNR) artificially "attract" particles that would normally align to other orientations, creating a self-reinforcing bias in the reconstruction (Sorzano et al., 2021). This advantage makes cryoPARES particularly robust when analysing specimens with preferred orientations. Their reconstruction should, therefore, show a diminished impact of preferred orientation when compared to traditional algorithms providing unbiased pose distributions were used for training.

To illustrate this situation, we analysed a 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 7 a). However, employing a higher-resolution density (15 Å) as reference yields substantially improved reconstruction quality, enabling ligand visualization (Figure 7 b,d). This scenario resembles the baited reconstruction strategy, where a high-resolution reference is used as template. In the same manner, reconstruction using cryoPARES-derived orientations generates a density map with good overall shape and with better resolved local features. For instance, the ligand density in the cryoPARES reconstruction seems more continuous than in the RELION reconstruction and the lysine next to the ligand is also better resolved in cryoPARES (Figure 7 c,e).

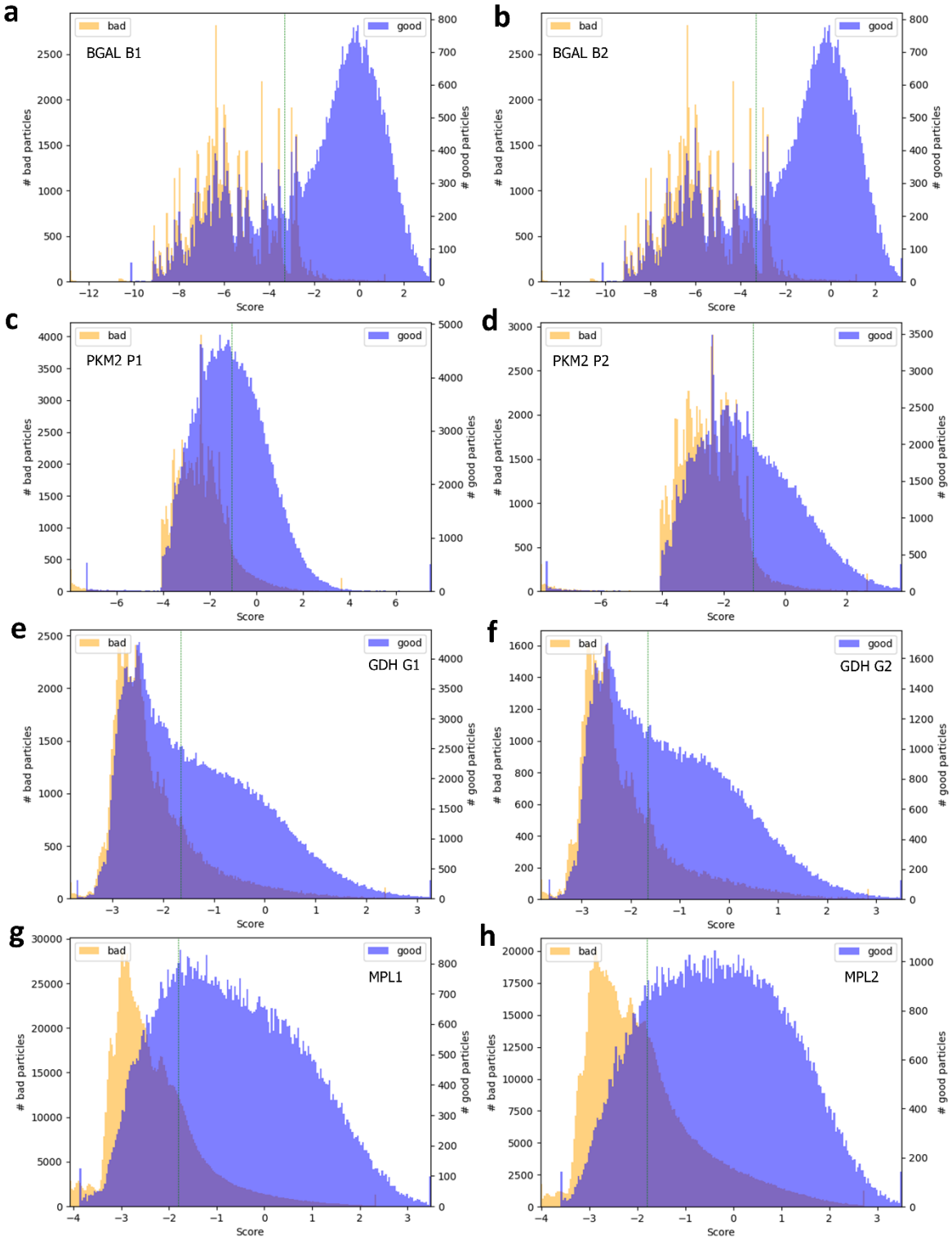

Figure 6. Histograms showing the distribution of direction-normalized pruning scores for particles that were discarded after 2D classification (bad, orange), and particles that survived the 2D classification step (good, blue). Note the different scales for number of "good" and "bad" particles. (a) BGAL bound to Compound B1, (b) and Compound B2, (c) PKM2 bound Compound P1, (d) and Compound P2, (e) GDH bound to Compound G1, (f) and Compound G2, (g) membrane protein bound to ligand MPL1, (h) and ligand MPL2. The vertical green line corresponds to the threshold automatically estimated for each target (see Methods Section 2.8 and Supplementary Material Section 9).

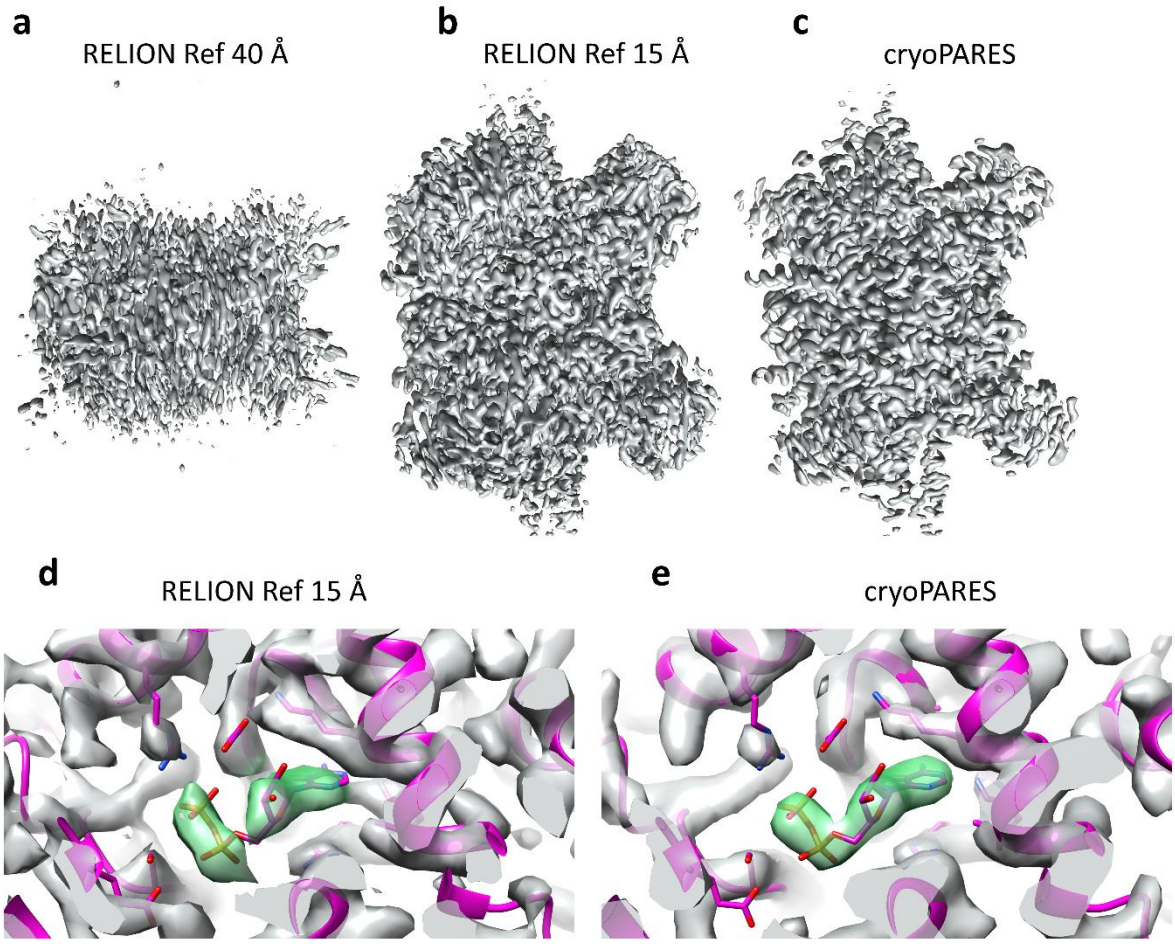

Figure 7. Reconstructed densities from a dataset of GDH bound to ADP that contains preferred orientations. a) Overview of the map obtained with a RELION workflow of 2D classification followed by auto-refine with a reference low-pass filtered at 40 Å b) Same as (a) with a reference low-pass filtered at 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 (in 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.

3.7. Dealing with model bias

The core idea behind cryoPARES is that the structural features that guide particle alignment in an apoprotein are largely preserved when small molecules bind to it. This hypothesis is based on the observation that ligand binding often induces localized conformational changes while maintaining the protein's overall shape, which is the crucial part for alignment. However, this assumption may not hold in cases where ligand binding triggers substantial conformational changes. In such scenarios, two potential sources of bias need careful consideration: model bias from the neural network predictions could lead to incorrect density interpretations, and template bias could arise when using high-resolution apo structures as references during local refinement, as shown by Lucas et al. (Lucas et al., 2023).

As discussed in the introduction, when used in ligand screening, the false positive case (classifying a ligand miss structure as bound), is less likely, since the ligands are not present in the apo dataset used for training or can be left out of the template for local refinement. However, the false negative

case, in which a ligand-bound structure is not resolved either due to model bias or insufficient reconstruction quality, could occur and should be detected to ensure that structural screenings are reliable.

Our approach to detect model bias is based on applying the gold standard for the local refinement step while using a medium-resolution reference (by default, low-pass filtered at 6 Å). If the half-to-half FSC resolution does not improve substantially beyond the filtering resolution, we consider the density to be affected by model bias or to be of low quality and cryoPARES recommends users to run a classical reconstruction program to rule out a false negative case.

Figure 8 illustrates the detection of two extreme cases. In the first case, we calculate a local refinement (4°) on a dataset with randomised orientations, simulating the complete failure of the neural network to infer poses (Figure 8 a,b). In the second case, we simulate the effect of high model bias by creating an Einstein from noise map that matches the training protein and running local refinement after (Figure 8 c,d). As shown in Figure 8 g, using a mid-resolution, 6Å template for local refinement (instead of a high-resolution template) easily identifies reference bias in both cases by the FSC curves quickly dropping to zero at the template's resolution. Conversely, if a high-resolution template is used instead, we obtain well-defined high-resolution structures with high FSC resolution that would lead to incorrect structural interpretation. When orientations correctly predicted by cyoPARES are used instead, the usage of a reference low-pass filtered at 6 Å does not prevent the reconstruction from reaching a resolution significantly better than the filtering cut-off (Figure 8 e,f).

We computed the Ω metric proposed by Lucas et al. (Lucas et al., 2023) (see Supplementary Material Section 13). Unsurprisingly, the computed metrics show that the amount of template bias tends to be smaller when we use the low-pass filtered reference than when the unfiltered reference map is used. It also shows that the template bias is much larger when random orientations are provided to the local refinement steps. These results are in accordance with what is observed in the FSC curves and show that the usage of high-resolution templates could present a significant risk. We were able to show, however, that simple, FSC-based model bias detection mechanisms can detect such situations. Additionally, we have implemented several preventive measures to help avoid model bias (see Supplementary Section 13).

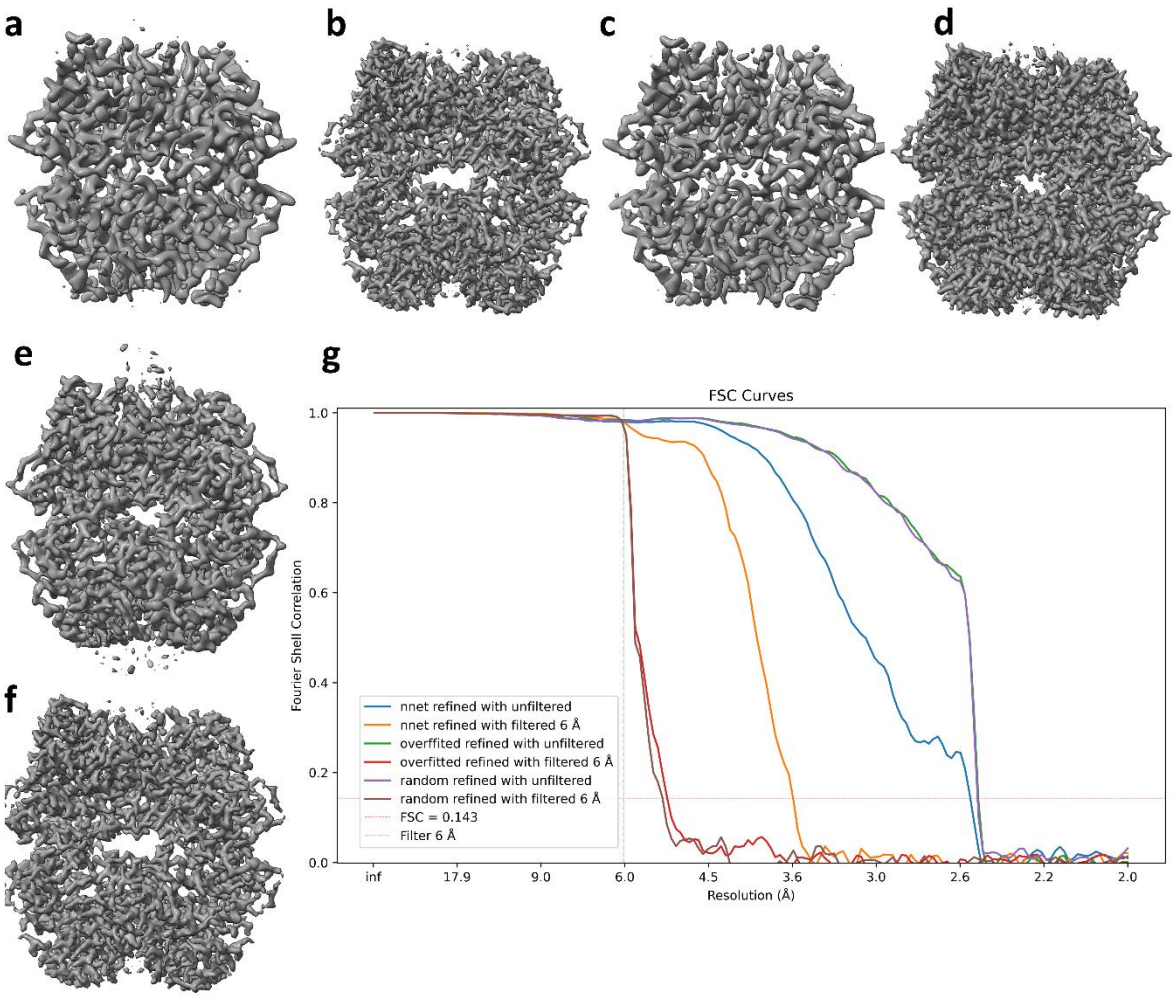

Figure 8. Reconstructed PKM2 Compound P1 under different model bias situations. a) Map obtained after applying a local refinement step to a set of particles with randomised orientations when the template reference is low-pass filtered at 6 Å, and b) when an unfiltered template is used instead. c) Map obtained when a local refinement step is applied to the already overfitted angular assignments of (b) if the reference map is filtered at 6 Å or d) left unfiltered. e) Map obtained after applying a local refinement step to a set of particles with cryoPARES orientations when the template reference is low-pass filtered at 6 Å, and f) when an unfiltered template is used instead. g) FSC curves for the experiments a-f. The vertical hashed, green line shows the low-pass cut-off of the filtered references.

4. Conclusions

In this work we introduced cryoPARES, a supervised deep learning framework that performs efficient cryo-EM pose determination for use in cases where a pre-existing dataset of aligned particles is available. CryoPARES rapidly predicts particle orientations, while performing automatic particle pruning, significantly reducing the computational burden of single-particle analysis. Our experiments with β-galactosidase, pyruvate kinase M2, and glutamate dehydrogenase demonstrate that cryoPARES effectively reconstructs high-resolution structures of ligand-bound complexes, revealing ligand densities absent in the apo structures used for training.

CryoPARES's robustness is evidenced by its ability to generalize across different datasets and imaging conditions, maintaining high performance even when trained on data from various microscopes. Incorporating a local refinement step enhances reconstruction quality, while automatic particle

pruning streamlines preprocessing by eliminating subjective steps like 2D class selection. Furthermore, cryoPARES could potentially address some limitations of current methods caused by the attractor problem such as their high sensitivity to preferred orientation, thus producing reliable results in challenging datasets where current approaches may struggle.

CryoPARES high-speed and one-shot nature can enable real-time data analysis during cryo-EM acquisition, offering immediate feedback on data quality and experimental parameters. This capability, together with automatic particle pruning, is particularly valuable in high-throughput structural studies and drug discovery campaigns, where rapid iteration, automation and reproducibility are essential.

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 will be available at https://github.com/rsanchezgarc/cryoPARES

8. References

Burt, A., & Chaillet, M. (2024). *teamtomo/torch-fourier-slice*.

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., & Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, *135*, 24–35. https://doi.org/10.1016/j.ultramic.2013.06.004

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. *37th International Conference on Machine Learning, ICML 2020*, *PartF168147-3*, 1575–1585. https://arxiv.org/abs/2002.05709v3

Chung, S. C. (2024). Cryo-forum: A framework for orientation recovery with uncertainty measure with the application in cryo-EM image analysis. *Journal of Structural Biology*, *216*(1), 108058. https://doi.org/10.1016/J.JSB.2023.108058

De la Rosa-Trevín, J. M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M., & Sorzano, C. O. S. (2013). Xmipp 3.0: An improved software suite for image processing in electron microscopy. *Journal of Structural Biology*, *184*(2), 321–328. https://doi.org/10.1016/j.jsb.2013.09.015

Gomez-Blanco, J., Kaur, S., Ortega, J., & Vargas, J. (2019). A robust approach to ab initio cryo-electron microscopy initial volume determination. *Journal of Structural Biology*, *208*(3), 107397. https://doi.org/10.1016/J.JSB.2019.09.014

Gomez-Blanco, J., Kaur, S., Strauss, M., & Vargas, J. (2022). Hierarchical autoclassification of cryo-EM samples and macromolecular energy landscape determination. *Computer Methods and Programs in Biomedicine*, *216*, 106673. https://doi.org/10.1016/J.CMPB.2022.106673

Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelman, M. (2004). HEALPix -- a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, *622*(2), 759–771. https://doi.org/10.1086/427976

Grant, T., Rohou, A., & Grigorieff, N. (2018). CisTEM, user-friendly software for single-particle image processing. *ELife*, *7*. https://doi.org/10.7554/eLife.35383

Gupta, H., McCann, M. T., Donati, L., & Unser, M. (2021). CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning. *IEEE Transactions on Computational Imaging*, *7*, 759–774. https://doi.org/10.1109/TCI.2021.3096491

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18037–18041. https://doi.org/10.1073/PNAS.1314449110

Jiménez-Moreno, A., Střelák, D., Filipovič, J., Carazo, J. M., & Sorzano, C. O. S. (2021). DeepAlign, a 3D alignment method based on regionalized deep learning for Cryo-EM. *Journal of Structural Biology*, *213*(2), 107712. https://doi.org/10.1016/j.jsb.2021.107712

Klee, D. M., Biza, O., Platt, R., & Walters, R. (2023). Image to Sphere: Learning Equivariant Features for Efficient Pose Prediction. *International Conference on Learning Representations*. http://arxiv.org/abs/2302.13926

Levy, A., Poitevin, F., Martel, J., Nashed, Y., Peck, A., Miolane, N., Ratner, D., Dunne, M., & Wetzstein, G. (2022). 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. https://doi.org/10.1007/978-3-031-19803-8_32/FIGURES/5

Levy, A., Wetzstein, G., Martel, J., Poitevin, F., & Zhong, E. D. (2022). Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Advances in Neural Information Processing Systems*, *35*, 13038–13049. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10392957/

Lian, R., Huang, B., Wang, L., Liu, Q., Lin, Y., & Ling, H. (2022). End-to-end orientation estimation from 2D cryo-EM images. *Acta Crystallographica Section D: Structural Biology*, *78*(2), 174–186. https://doi.org/10.1107/S2059798321011761

Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2020, August 8). ON THE VARIANCE OF THE ADAPTIVE LEARNING RATE AND BEYOND. *8th International Conference on Learning Representations, ICLR 2020*. https://arxiv.org/abs/1908.03265v4

Lucas, B. A., Himes, B. A., & Grigorieff, N. (2023). Baited reconstruction with 2D template matching for high-resolution structure determination in vitro and in vivo without template bias. *ELife*, *12*. https://doi.org/10.7554/ELIFE.90486

Ludtke, S. J., Baldwin, P. R., & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology*, *128*(1), 82–97. https://doi.org/10.1006/JSBI.1999.4174

Nogales, E. (2015). The development of cryo-EM into a mainstream structural biology technique. *Nature Methods*, *13*(1), 24–27. https://doi.org/10.1038/nmeth.3694

Passmore, L. A., & Russo, C. J. (2016). Specimen Preparation for High-Resolution Cryo-EM. *Methods in Enzymology*, *579*, 51–86. https://doi.org/10.1016/BS.MIE.2016.04.011

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*, 8026–8037. https://doi.org/10.5555/3454287.3455008

Penczek, P. A., Grassucci, R. A., & Frank, J. (1994). The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, *53*(3), 251–270. https://doi.org/10.1016/0304-3991(94)90038-8

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods 2017 14:3*, *14*(3), 290–296. https://doi.org/10.1038/nmeth.4169

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, *9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sanchez-Garcia, R., Saur, M., Vargas, J., Poelking, C., & Deane, C. M. (2024). CESPED: A benchmark for supervised particle pose estimation in cryo-EM. *Physical Review Research*, *6*(2), 023245. https://doi.org/10.1103/PHYSREVRESEARCH.6.023245

Saur, M., Hartshorn, M. J., Dong, J., Reeks, J., Bunkoczi, G., Jhoti, H., & Williams, P. A. (2020). Fragment-based drug discovery using cryo-EM. *Drug Discovery Today*, *25*(3), 485–490. https://doi.org/10.1016/J.DRUDIS.2019.12.006

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Scheres, S. H. W., & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature Methods 2012 9:9*, *9*(9), 853–854. https://doi.org/10.1038/nmeth.2115

Sigworth, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, *122*(3). https://doi.org/10.1006/jsbi.1998.4014

Sorzano, C. O. S., Bilbao-Castro, J. R., Shkolnisky, Y., Alcorlo, M., Melero, R., Caffarena-Fernández, G., Li, M., Xu, G., Marabini, R., & Carazo, J. M. (2010). A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of Structural Biology*, *171*(2), 197–206. https://doi.org/10.1016/J.JSB.2010.03.011

Sorzano, C. O. S., Jimenez-Moreno, A., Maluenda, D., Martinez, M., Ramirez-Aportela, E., Krieger, J., Melero, R., Cuervo, A., Conesa, J., Filipovic, J., Conesa, P., Del Cano, L., Fonseca, Y. C., Jiménez-De La Morena, J., Losana, P., Sanchez-Garcia, R., Strelak, D., Fernandez-Gimenez, E., De Isidro-

Gómez, F. P., … Carazo, J. M. (2022). On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Acta Crystallographica Section D: Structural Biology*, *78*(4), 410–423. https://doi.org/10.1107/S2059798322001978/IC5116SUP1.PDF

Sorzano, C. O. S., Semchonok, D., Lin, S. C., Lo, Y. C., Vilas, J. L., Jiménez-Moreno, A., Gragera, M., Vacca, S., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Melero, R., Cuervo, A., Conesa, J. J., Conesa, P., Losana, P., Caño, L. del, de la Morena, J. J., Fonseca, Y. C., … Carazo, J. M. (2021). Algorithmic robustness to preferred orientations in single particle analysis by CryoEM. *Journal of Structural Biology*, *213*(1), 107695. https://doi.org/10.1016/J.JSB.2020.107695

Sorzano, C. O. S., Vargas, J., de la Rosa-Trevín, J. M., Otón, J., Álvarez-Cabrera, A. L., Abrishami, V., Sesmero, E., Marabini, R., & Carazo, J. M. (2015). A statistical approach to the initial volume problem in Single Particle Analysis by Electron Microscopy. *Journal of Structural Biology*, *189*(3), 213–219. https://doi.org/10.1016/J.JSB.2015.01.009

Sorzano, C. O. S., Vargas, J., Vilas, J., Jiménez-Moreno, A., Mota, J., Majtner, T., Maluenda, D., Martínez, M., Sánchez-García, R., Segura Mora, J., Oton, J., Melero, R., Cano, L., Conesa, P., Gomez-Blanco, J., Rancel, Y., Marabini, R., & Carazo, J. M. (2018). Swarm optimization as a consensus technique for Electron Microscopy Initial Volume. *Applied Analysis and Optimization*, *2*(10), 299–313.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *Journal of Structural Biology*, *116*(1), 17–24. https://doi.org/10.1006/JSBI.1996.0004

Vargas, J., Álvarez-Cabrera, A. L., Marabini, R., Carazo, J. M., & Sorzano, C. O. S. (2014). Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics*, *30*(20), 2891–2898. https://doi.org/10.1093/BIOINFORMATICS/BTU404

Vargas, J., Melero, R., Gómez-Blanco, J., Carazo, J. M., & Sorzano, C. O. S. (2017). Quantitative analysis of 3D alignment quality: its impact on soft-validation, particle pruning and homogeneity analysis. *Scientific Reports 2017 7:1*, *7*(1), 1–14. https://doi.org/10.1038/s41598-017-06526-z

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *ELife*, *7*. https://doi.org/10.7554/eLife.42166

Zonca, A., Singer, L. P., Lenz, D., Reinecke, M., Rosset, C., Hivon, E., & Gorski, K. M. (2019). healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *Journal of Open Source Software*, *4*(35), 1298. https://doi.org/10.21105/JOSS.01298