ISSN 1688-2806

Universidad de la República
Facultad de Ingeniería

Improving heterogeneous resolution in cryo-EM volumes reconstructions with pose refinement

Tesis presentada a la Facultad de Ingeniería de la Universidad de la República por

Diego Silvera Coeff

en cumplimiento parcial de los requerimientos para la obtención del título de Magister en Ingeniería Eléctrica.

Directores de Tesis

Federico Lecumberry ..... Universidad de la República
Alberto Bartesaghi ..... Duke University

Tribunal

Álvaro Gómez ..... Universidad de la República
Lara Raad ..... Universidad de la República
Pablo Arias ..... Universitat Pompeu Fabra

Director Académico

Federico Lecumberry ..... Universidad de la República

Montevideo
Friday 17th October, 2025

Improving heterogeneous resolution in cryo-EM volumes reconstructions with pose refinement, Diego Silvera Coeff.

ISSN 1688-2806

Esta tesis fue preparada en LaTeX usando la clase iietesis (v1.1).
Contiene un total de 130 páginas.
Compilada el Friday 17th October, 2025.
http://iie.fing.edu.uy/

Acknowledgments/Agradecimientos

This is the only section that is not written in English, since it is intended to be read by people who speak Spanish.

Quiero empezar por agradecer a mi director académico y director de tesis Federico Lecumberry, quien además fue mi tutor en mi proyecto de fin de carrera y ha sido responsable de gran parte de mis tareas como ayudante en el IIE. Gracias por haber confiado en mí y darme oportunidades que nunca pensé tener.

Gracias a mi otro director de tesis Alberto Bartesaghi, por guiarme en este proceso y recibirme con gran calidez en su laboratorio y hogar. La posibilidad de hacer una pasantía en una universidad como Duke University era algo inimaginable para mí, y por haberlo hecho posible voy a estar eternamente agradecido.

También quiero agradecer a los miembros de CICADA e IMAGINA, por permitirme formar parte de su comunidad y enriquecer enormemente mi formación. Utilizar los conocimientos adquiridos en la carrera en el procesamiento de señales biomédicas fue una meta que me propuse a mediados de la carrera, y por momentos pareció poco probable que pudiera suceder en el ámbito laboral. Esta comunidad me abrió las puertas, permitiéndome enfrentar problemas muy interesantes y conocer una comunidad aún más grande y amplia que busca avanzar en el campo de las imágenes biomédicas.

A mis compañeros del IIE, por acompañarme en el día a día promoviendo un excelente clima de trabajo. En especial quiero agradecerle a Ignacio Irigaray. Gracias querido, por todas las enseñanzas, por acompañarme desde mis inicios en el instituto, por ser paciente ante mi falta de paciencia y por el apoyo que le diste a mi formación.

Finalmente, quiero agradecer profundamente a mi familia, amigos y pareja. Gracias por acompañarme, aguantarme, guiarme y ayudarme cada vez que lo necesité. Gracias viejita por estar en todas. Sé que en ocasiones no te la hice fácil, pero ahí estuviste. Gracias Melina por ser mi compañera y el amor de mi vida. No te haces una idea de lo feliz que me hace estar a tu lado. Te amo.

Esta página ha sido intencionalmente dejada en blanco.

A mis hermanas de corazón, Aitana y Catalina Iglesias.

Esta página ha sido intencionalmente dejada en blanco.

Abstract

Single-particle cryo-electron microscopy (cryo-EM) has emerged as a transformative technique for determining the three-dimensional structures of macromolecular complexes at near-atomic resolution. Its ability to visualize biomolecules in multiple functional states without the need for crystallization has provided unprecedented insights into their structure, dynamics, and mechanisms, making it a cornerstone in structural biology and drug discovery.

Despite its success, cryo-EM faces several challenges that limit the achievable resolution and accuracy of reconstructions. Chief among these are the inherently low signal-to-noise ratio (SNR) of raw micrographs, the difficulty in accurately estimating particle orientations (pose estimation), and the presence of conformational and compositional heterogeneity in the sample. In recent years, deep learning has emerged as a leading approach for addressing these limitations, offering powerful methods for denoising, pose refinement, and disentangling structural variability.

In this work, a method designed to exploit particle heterogeneity for iterative pose refinement is presented. The approach integrates two state-of-the-art tools: cryoDRGN, which models structural variability using deep generative networks, and Frealign, which performs high-resolution 3D refinement. These tools were combined into a unified pipeline and tested on real cryo-EM datasets, demonstrating the potential of the method to improve both the accuracy of pose estimation and the quality of heterogeneous reconstructions.

Esta página ha sido intencionalmente dejada en blanco.

Contents

Esta página ha sido intencionalmente dejada en blanco.

Chapter 1

Introduction

1.1 Motivation

Inside cells, macromolecules perform several biological functions within the context of complex molecular networks, often transitioning between different assembly states to regulate activity [6]. The structure and dynamics of these networks can be uncovered through structural analysis of the macromolecules involved. Such analyses are essential not only for understanding fundamental functional mechanisms but also for the design of new drugs aimed at modulating these mechanisms in pathological conditions [14].

Traditionally, the understanding of macromolecular function has been derived from static three-dimensional structures. However, a more comprehensive view can be achieved by analyzing the full distribution of conformational and compositional states that are functionally relevant [65]. Unlike other structural biology imaging techniques such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, cryo-electron microscopy (cryo-EM) offers the unique ability to capture structural dynamics, including a wide range of conformations (continuous motions and flexible states) and compositional (discrete differences in assembly states or subunit presence) variability. Thanks to substantial advances in both hardware [3, 35, 55] and software [5, 9, 35, 53, 55, 63], cryo-EM has rapidly emerged as a leading technique for resolving the high-resolution structures of large macromolecular complexes [1, 62]. An example of cryo-EM particle images and a 3D volume reconstruction can be seen in Figure 1.1, which shows some of the particles used for the volume reconstruction.

The importance of cryo-EM in structural biology was recognized with the 2017 Nobel Prize in Chemistry, awarded to Jacques Dubochet, Joachim Frank, and Richard Henderson for the development of cryo-EM for the high-resolution structure determination of biomolecules in solution. This recognition underscored cryo-EM's potential to study biological structures in near-native states, without the need for crystallization, and to reveal previously inaccessible dynamics and heterogeneity at the molecular level.

Despite its potential, cryo-EM data present significant challenges. The ac-

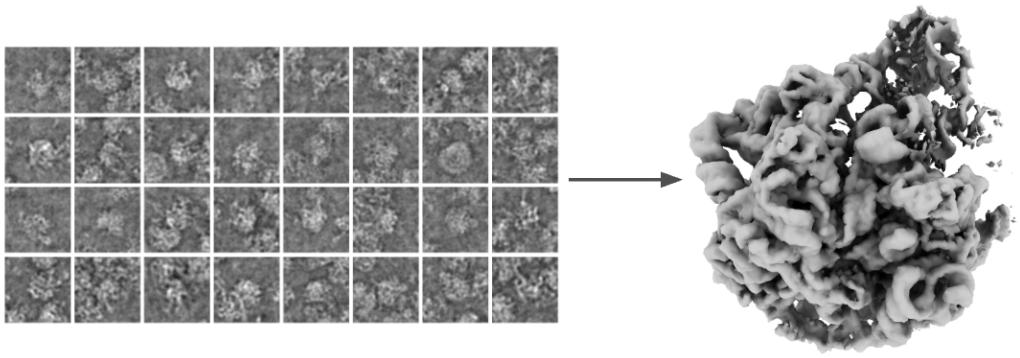


Figure 1.1: Particle images acquired with cryo-EM and their 3D reconstruction.

quired images typically exhibit extremely low signal-to-noise ratios (SNR), and the 3D orientations and 2D translations (poses) of individual particle projections are unknown. Furthermore, most macromolecular samples display heterogeneity, both conformational and compositional [26]. Despite these difficulties, cryo-EM stands out for its ability to capture structural variability of biological importance—variability that is generally inaccessible to predictive tools such as AlphaFold [27], whose main developers Demis Hassabis and John Jumper were also recognized with the 2024 Nobel Prize in Chemistry for their work on protein structure prediction and computational protein design.

A critical limitation in most cryo-EM processing pipelines is that particle poses are generally estimated relative to a single consensus volume. This assumption is problematic in the presence of structural heterogeneity, as it may lead to incorrect pose estimation for particles that significantly deviate from the consensus structure. Consequently, this may become a major obstacle to achieving high-resolution reconstructions for such subpopulations. An example of structural heterogeneity can be seen in Figure 1.2, where different reconstructions of the same macromolecule are shown. It can be seen that there are clear differences in the volumes obtained, which are due to the different compositions the protein can take due to the interaction of certain subunits.

1.2 Objetives

This work aims to improve the resolution of cryo-EM reconstructions in datasets exhibiting structural heterogeneity, particularly in cases of discrete variability. Rather than treating heterogeneity as a source of noise or an obstacle, this thesis explores how it can be exploited to enhance pose estimation and, in turn, produce higher-resolution density maps for distinct molecular states. To achieve this, an iterative framework is proposed in which particle poses are refined based on structure-aware clustering in a low-dimensional space. Specifically, the method involves identifying clusters that correspond to distinct conformational or compositional states of the macromolecule, reconstructing a 3D density map for each

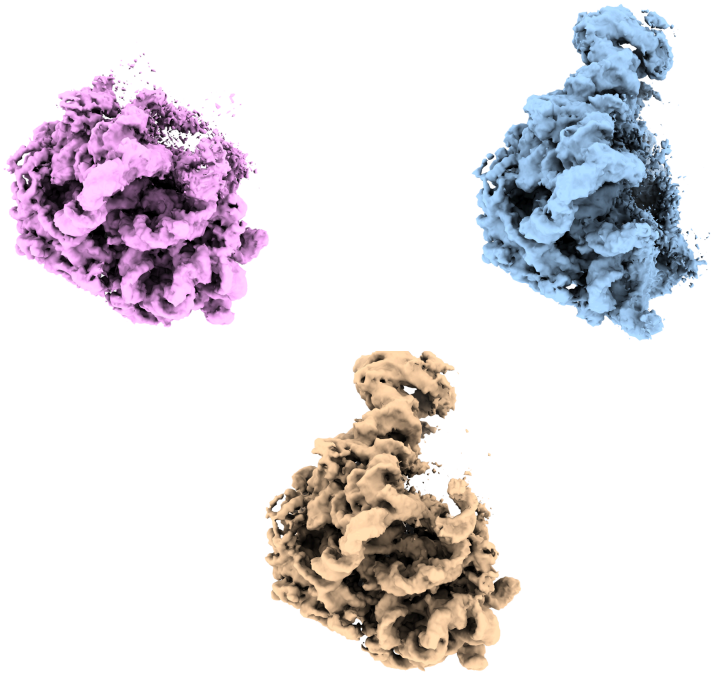


Figure 1.2: Example of structural heterogeneity. The three volumes are reconstructions of different particles from the same dataset, but with different compositions due to the presence or absence of some subunits of the macromolecule.

cluster, and using these reconstructions as references for targeted pose refinement. By integrating clustering and refinement in a feedback loop, this approach leverages the underlying heterogeneity of the sample to inform more accurate alignment of particles.

1.3 Contributions

The main contribution of this work is a fully automated pipeline that combines the feature-learning capabilities of a state-of-the-art deep generative model called cryoDRGN [65] [66], with the high-resolution refinement and reconstruction capabilities of a processing tool called Frealign [20]. This integration enables high-resolution reconstructions, in some cases surpassing the resolutions achieved in previous work [16].

Building upon this pipeline, a second contribution is the development of a set of Python functions designed to facilitate the parsing, manipulation, and conversion of common file formats in cryo-EM workflows. These include:

- `.par` files used by Frealign for alignment and refinement metadata
- `.star` files from RELION, a widely used format for particle metadata

- `.pkl` files produced by Python's `pickle` module
- `.mrc` files for storing particle images and 3D density maps

While cryoDRGN and some other libraries already support integration of `.mrc`, `.star`, and `.pkl`, no existing solution was found that also supports the integration of Frealign's `.par` files.

Although the pipeline can be automated, it has the advantage of showing how the clusters are evolving after each refinement. Clusters can be inspected in low-dimensional spaces (e.g., PCA, UMAP), enabling dynamic feedback on the quality and separability of clusters. This also helps guide hyperparameter tuning in clustering algorithms such as K-Means, Gaussian Mixture Models (GMMs), Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [18], or Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [10], and supports informed decisions regarding class selection in downstream refinement. As such, the pipeline not only supports automated reconstruction, but also provides tools for guided exploration and hyperparameter fine-tuning.

1.4 Document Structure

Chapter 2 provides a brief overview of the biological and scientific context of the problem, along with key concepts needed to understand the following chapters. This includes a biological context and how the levels of organization of proteins are defined, an introduction to the cryo-EM acquisition technique, the available data repositories, the general workflow for cryo-EM image processing, the image formation model used in this work, the Fourier Shell Correlation and the use of it in this context, a dimensionality reduction technique called UMAP, concepts related to the heterogeneity of macromolecules, and how Frealign refines and reconstructs 3D structures.

Chapter 3 presents works that relate deep learning and cryo-EM data processing for volume reconstruction. The selected works present novel approaches to cryo-EM problems such as pose estimation or particle heterogeneity.

Chapter 4 introduces the proposed method, providing a detailed explanation of each component of the pipeline. It also presents preliminary results that guided the decisions made during its design. The experimental setup is described, including the datasets employed and the supporting tools utilized. Finally, the potential advantages and limitations of the pipeline are discussed.

Chapter 5 presents the results of the experiments conducted and provides a detailed analysis of the findings.

Finally, Chapter 6 summarizes the conclusions, outlines directions for future work, and personal reflections about this work.

Chapter 2

Background

Cryo-EM is a biological imaging technique that has recently enabled the reconstruction of molecular structures at near-atomic resolution [67]. These reconstructions are fundamental in structural biology, as the three-dimensional structure of macromolecules is directly linked to their biological function. Accurate maps are therefore critical to progress in biology, chemistry, and pharmacology. Despite its remarkable success, cryo-EM presents significant computational and methodological challenges. Raw images suffer from extremely low signal-to-noise ratio, particle poses are initially unknown, and molecules often exhibit substantial structural heterogeneity. Addressing these issues requires advanced image processing tools.

This chapter introduces the essential concepts needed to understand the remainder of this document. It first provides a biological context, reviewing protein organization levels and the principles of cryo-EM data acquisition. It then describes the image formation model, available data repositories, and key validation metrics such as the Fourier Shell Correlation. Finally, it outlines computational approaches to deal with heterogeneity and reconstruction, motivating the use of methods like UMAP and Frealign, which will be discussed in detail later.

2.1 Structural Biology

Structural biology is a branch of biology that studies the three-dimensional structure of biomolecules, primarily proteins and nucleic acids (DNA and RNA). Unlike other areas that focus on genetic or biochemical aspects, structural biology seeks to understand how the shape of a molecule affects its function, which is fundamental to understanding biological processes at the molecular level.

Among biomolecules, proteins are especially significant due to their diverse and essential functions in the cell. At the most basic level, a protein consists of a sequence of basic components called amino acids, arranged in a specific linear order, called a polypeptide chain. The amino acids in this chain are linked by peptide bonds in a specific order defined by the genetic code. Although the chain itself is linear, it folds into complex three-dimensional structures due to various chemical

interactions between its components. These folded structures are essential for the protein's biological function.

One of the central ideas in this field is that structure determines function. For example, an enzyme (a protein that accelerates chemical reactions) can only fulfill its function if its shape allows it to interact correctly with other molecules. If that shape is altered, either by a genetic mutation or a change in the environment, its activity can be affected, which can lead to disease.

This relationship between form and function can be better understood if we consider the different structural levels at which a protein is organized:

- **Primary structures**

  The primary structure of a protein is its linear sequence of amino acids, which are the fundamental building blocks of proteins. This sequence is encoded by DNA and determines the basic chemical properties of the molecule. Changes in this sequence (for example, a mutation) can affect higher levels of protein structure.

- **Secondary structures**

  The secondary structure refers to simple, repetitive patterns that form when parts of the amino acid chain fold locally. The most common patterns are spiral shapes (called alpha-helices) and flat, sheet-like arrangements (called beta-strands). These structures are held together by internal forces and help give the protein its initial stability and shape.

- **Tertiary structures**

  This is the complete three-dimensional conformation adopted by a polypeptide chain. This shape results from interactions between different parts of the chain, depending on their chemical characteristics. The final folded structure determines how the protein interacts with other molecules and performs its specific role.

- **Quaternary structures**

  The quaternary structure appears when a protein is made up of more than one chain, commonly called subunits. These subunits come together to form a functional unit. The way they fit and work together is essential for the protein's activity.

Figure 2.1 shows an example of each structure level, where the letters in the primary structure represent different amino acids, and the letters from the tertiary and quaternary structures represent the amino (N) and carboxyl (C) groups.

Traditionally, the aim has been to produce a single, high-resolution 3D model that represents the molecule structure. This model can provide important insights into how the molecule works, how it interacts with other molecules, or how mutations might affect its behavior. However, many molecules are not static (normally called homogeneous). They change shape as part of their natural function. For example, ribosomes adopt different conformations during the translation of genetic

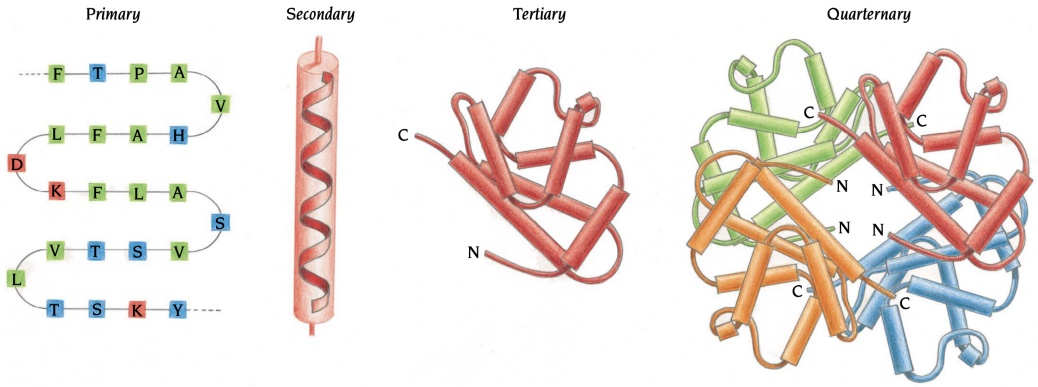


Figure 2.1: Representation of the different structural levels of proteins extracted from [8]. The letters in the primary structure represent differents amino acids, and the letters from the tertiary and quaternary structures represent the amino (N) and carboxyl (C) groups.

information. In those cases where biomolecules are not homogeneous, a single averaged structure may be inaccurate. If structural data from many particles in different conformations/compositions are averaged together, flexible and dynamic regions would become blurred, and important information about molecular motion would be lost. The result would be a model that might look well-defined in some areas, but unclear or misleading in others.

To address these challenges, modern cryo-EM workflows aim not only to reconstruct a single average model, but also to resolve and classify multiple distinct structures from a heterogeneous dataset.

2.2 Cryo-EM

Several experimental techniques are available for determining the 3D structure of biomolecules. The most established methods include X-ray crystallography, NMR spectroscopy, and cryo-EM. Each technique has its strengths and limitations, depending on the nature of the sample and the biological questions being addressed.

In X-ray crystallography, many identical copies of a molecule are packed into a solid crystal. When this crystal is exposed to a beam of X-rays, the atoms in the molecules scatter the radiation, producing a pattern of spots. By analyzing this pattern, it is possible to reconstruct the average 3D structure of the molecule at high resolution. However, this process forces all molecules into a single, repeated arrangement, effectively removing any structural variability from the measurement. While crystallography yields atomic-resolution models, it captures only a single, averaged conformation—typically the most stable one. Transient or flexible states are lost in the process.

NMR spectroscopy studies molecules in solution by placing them in a strong magnetic field and measuring how their atomic nuclei respond to radiofrequency pulses. These responses depend on the chemical environment of the atoms and can be used to infer distances and motions within the molecule. This technique is

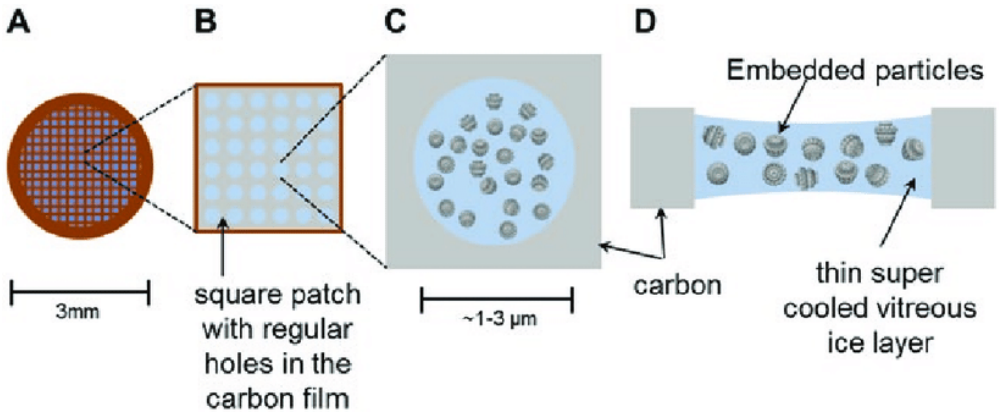


Figure 2.2: Cryo-EM sample preparation. (a) Copper mesh grid, 3 mm in diameter, coated with a thin layer of perforated carbon film. (b) Closer view of one of the grid squares, revealing the microscopic holes within the carbon layer. (c) Detailed view of a single hole, showing a thin layer of vitrified ice in which protein molecules are suspended. (d) Schematic cross-section of a hole illustrating particles embedded in the ice. The figure was extracted from [15].

able to capture molecular dynamics in solution and is well-suited for small, flexible proteins. However, it works best for relatively small biomolecules, such as short proteins or peptides, and becomes increasingly difficult to apply as the molecules get larger. The main reason has to do with the way molecules move in solution. Small molecules rotate and tumble very quickly, and this rapid motion helps NMR produce clear and sharp signals. Large molecules, by contrast, rotate much more slowly. When this happens, the NMR signals die out more quickly, which makes the peaks in the spectrum appear broad and faint. This loss of clarity makes it hard to extract useful information.

Cryo-EM, and particularly single-particle cryo-EM, overcomes many of these limitations. In this technique, thousands to millions of individual particles are imaged in a near-native, frozen-hydrated state. Each image represents a 2D projection of a particle in an unknown 3D orientation and potentially in a different structural state. This makes it well-suited to capture structural variability and explore structural heterogeneity, both conformational and compositional.

In cryo-EM, cryogenic temperatures are used to convert the water into vitreous ice. Vitreous ice is characterized by the absence of an ordered molecular structure, unlike ordinary crystalline ice. In practice, it is formed by rapidly cooling the thin aqueous layer during sample vitrification, which prevents crystallization and preserves biomolecules in a near-native state. This lack of structure is what allows macromolecules to be visualized. Otherwise, the noise and distortions generated by the ice pattern would be so severe that no information would be visible.

The particles are deposited onto a grid, typically made of copper, and subsequently frozen. As illustrated in Figure 2.2, the grid consists of square regions covered by a perforated carbon film, where each hole contains particles embedded within a thin layer of vitreous ice.

Once the sample has been prepared and vitrified, the grid is transferred to

the cryo-electron microscope using a cryo-holder or an automated loading system, ensuring that the specimen remains at liquid nitrogen temperature throughout the process to prevent devitrification. Inside the microscope, the grid is placed in the path of an electron beam, typically generated by a field emission gun operating at 200–300 kV. The electrons interact with the sample, and those transmitted through the thin layer of vitreous ice are projected onto a direct electron detector.

To minimize radiation damage, a low electron dose is used. Otherwise, the sample can become dehydrated and damaged. This is the main cause of the low signal-to-noise ratio in the acquired images.

Images are recorded across different areas of the grid, each containing particles in random orientations. These two-dimensional projections form the raw data from which three-dimensional reconstructions will later be computed.

2.3 Data

Given the importance of characterizing the structures of macromolecules, several specialized public databases have been established to store structural data. Among the best-known are the Electron Microscopy Public Image Archive (EMPIAR) [25], the Electron Microscopy Data Bank (EMDB) [33] and the Protein Data Bank (PDB) [7].

EMPIAR provides access to images collected during cryo-EM experiments. These images may be raw data, which are unprocessed images of thousands of particles called micrographs, or particles that are already selected and cropped from the micrographs.

Particle images are especially useful when the focus is solely on the particles themselves, as there is no need to perform particle picking. Moreover, in some cases, certain artifacts caused by particle motion or the contrast transfer function (see Subsection 2.4) are already partially corrected. Conversely, if the focus is on preprocessing steps such as particle picking, working with raw micrographs is necessary.

Figures 2.3 and 2.4 show examples of a micrograph and some selected particles, respectively. Both figures illustrate the challenges associated with the low SNR and highlight the complexity of the particle picking task. There are deep learning approaches to this problem, such as DeepEM [69], TOPAZ [5] and crYOLO [59], but these will not be addressed in this work because pre-picked particles will be used.

EMPIAR also stores detailed metadata associated with the images. These metadata include crucial information of the microscope like sample conditions, electron beam energy, and defocus angles. Acquisition parameters include the number of micrographs captured, the number of selected particles, sizes of the images, pixel size, and data encoding formats.

Finally, if some processing was made, it is common practice to specify the processing techniques used and the results obtained.

The EMDB stores three-dimensional density maps obtained primarily through electron microscopy techniques, especially cryo-EM. A density map is a volumetric

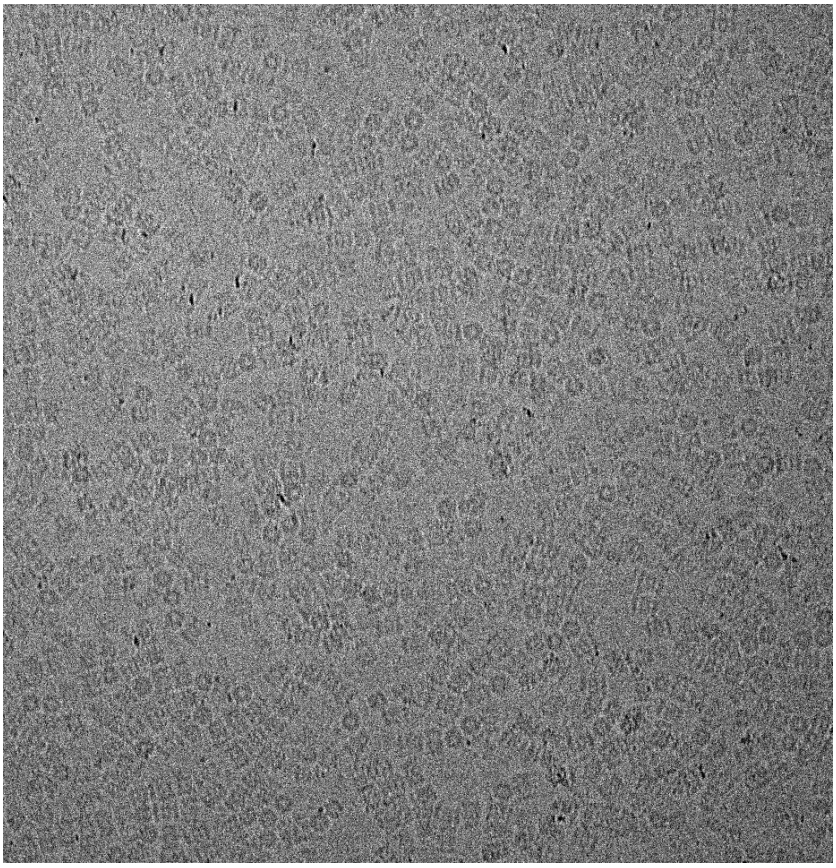


Figure 2.3: Micrograph from EMPIAR-10025.

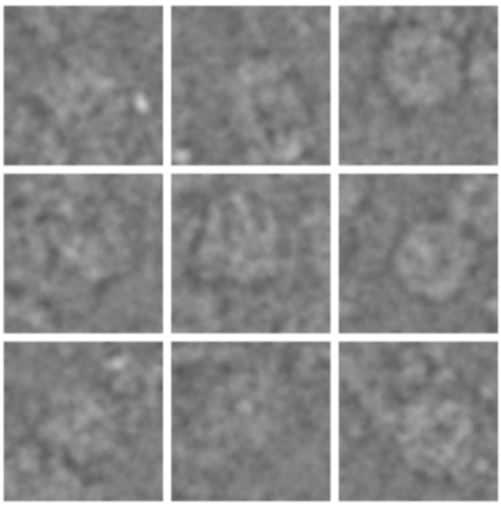


Figure 2.4: Selected particles from Figure 2.3.

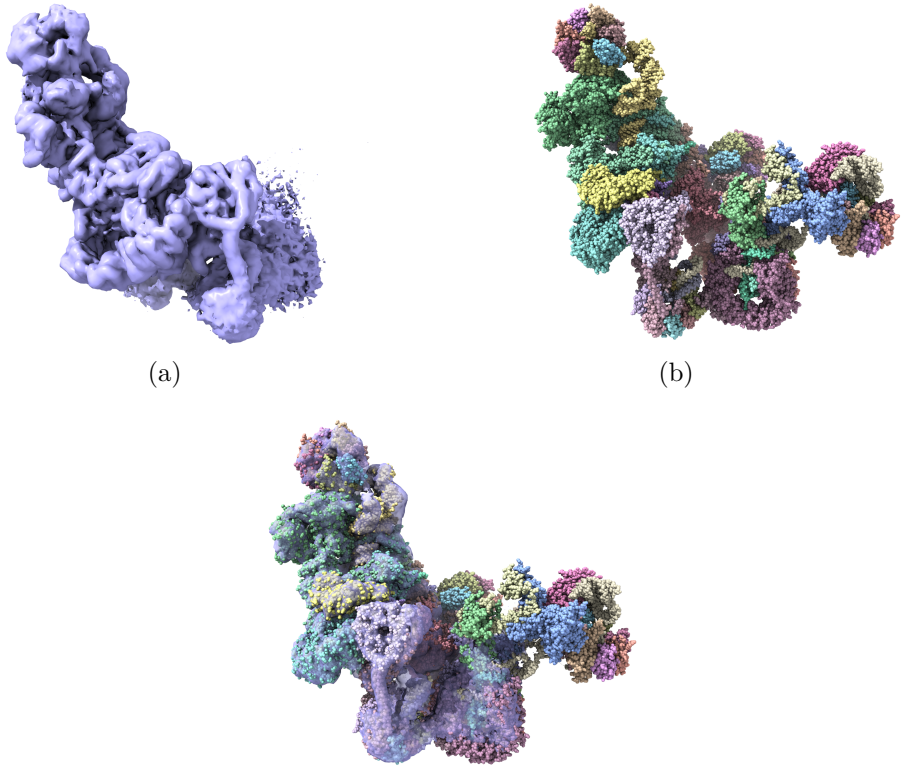


Figure 2.5: (a) Three-dimensional density map of a spliceosome (EMPIAR-10180). (b) Atomic model of pre-catalytic spliceosome from PDB (ID: 5nrl). Atoms are representes as spheres. The colors represent different subunits of the the atomic model. (c) Atomic model fitted into the consensus volume generated from EMPIAR-10180 particles. Atoms are shown in sphere style.

reconstruction that reflects the probability of finding electrons in different regions of space, and therefore gives an estimate of the molecule's structure at a given resolution.

Each EMDB entry also includes associated experimental metadata, such as information about the sample, imaging conditions, achieved resolution, and the software and image processing methods used during reconstruction.

An example of a density map is shown in Figure 2.5a, which was generated using data from the EMPIAR-10180 dataset and represents a pre-catalytic spliceosome.

Because all available particles were used to compute this map, it is commonly referred to as a consensus volume. A consensus volume is a single, averaged reconstruction that captures the dominant structure present in the dataset.

PDB is the central archive for atomic models of biomolecules, including proteins, nucleic acids and large molecular complexes, derived from X-ray crystallography, NMR spectroscopy, and cryo-EM. An atomic model specifies the spatial position of each atom in the molecule, along with other information such as atom

types, connectivity, secondary structure assignments, and experimental metadata. Atomic models are the final result that can be achieved and the one with biological information.

Figure 2.5b shows the atomic model of a pre-catalytic spliceosome.

Together, these databases form an ecosystem for storing, sharing, and analyzing structural data. The images processed in this work are from EMPIAR; the density maps obtained and published from these images are used for the selection of hyperparameters and to compare results, and the atomic models serve as ground-truth.

Figure 2.5c shows the density map from Figure 2.5a fitted with the atomic model from Figure 2.5b.

Fitting an atomic model into a density map allows the identification of specific amino acids, nucleotides, or ligands, which is essential for understanding mechanisms of action, binding sites, compositional and conformational changes, or mutations.

In this work, atomic models will be used as ground truth. The obtained reconstructions will be compared against these models by fitting the atomic coordinates into the corresponding density maps, using the flexible fitting procedures from ChimeraX [40]. If the atomic model does not align well with the reconstructed density, it may indicate problems such as poor map quality, incorrect orientation assignments, or misaligned particles. Performing and validating this fitting is also a common requirement when preparing data for deposition in the EMDB.

2.4 General Workflow

The traditional cryo-EM single-particle analysis workflow aims to reconstruct a high-resolution three-dimensional model from a large collection of two-dimensional images. This process involves multiple stages, many of which rely on signal processing and computational techniques to enhance the SNR and extract relevant structural information. Figure 2.6 shows a traditional workflow for single-particle cryo-EM image processing.

2.4.1 Motion Correction

In cryo-EM, motion correction refers to the computational process of compensating for sample movement that occurs during image acquisition. Instead of producing a single still image, modern cryo-EM cameras record a movie consisting of many short-exposure frames (often 20–60 frames over a total exposure of 1–3 seconds). Motion correction aligns and sums these frames to produce a sharp, high-contrast final image of the particles.

Sample motion arises primarily from beam-induced effects. When the high-energy electron beam interacts with the vitreous ice and the supporting film, it can cause electrostatic charging, localized heating, and physical deformation of the specimen. These effects are often anisotropic, meaning different parts of the image can shift in different directions or by different amounts.

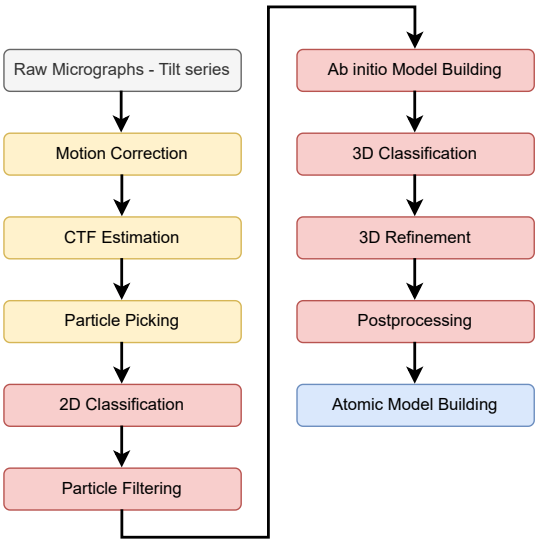


Figure 2.6: Traditional workflow diagram of cryo-EM single particle analysis for 3D reconstruction. Yellow blocks correspond to pre-processing steps, red blocks correspond to the 3D reconstruction process, and the lightgray and blue blocks correspond to the initial input and final result, respectivily. This diagram is based on the diagram from [6].

Motion correction can be performed either globally or locally. In global motion correction, the entire image is treated as if it shifts uniformly between frames, which is effective for correcting mechanical drift but insufficient to fully compensate for beam-induced deformations. Local motion correction addresses this limitation by dividing the image into smaller patches, aligning each patch independently, and then recombining them, allowing the recovery of fine structural detail that would otherwise be lost. In addition to alignment, modern workflows incorporate dose weighting, in which early frames are given higher weight for high-resolution information, while later frames contribute primarily to low-resolution contrast due to the progressive effects of radiation damage.

There are highly effective softwares such as MotionCor2 [64] and Warp [56] that can align movie frames with sub-pixel precision and apply dose weighting, producing near-optimal results for many datasets.

2.4.2 Contrast Transfer Function

The Contrast Transfer Function (CTF) is a fundamental concept in cryo-EM and other imaging techniques that use wave-based signals like electrons or light. It describes how the microscope modifies the contrast of different spatial frequencies in the sample image.

The images obtained are not direct representations of the sample's electrostatic potential distribution. The microscope introduces a series of aberrations to

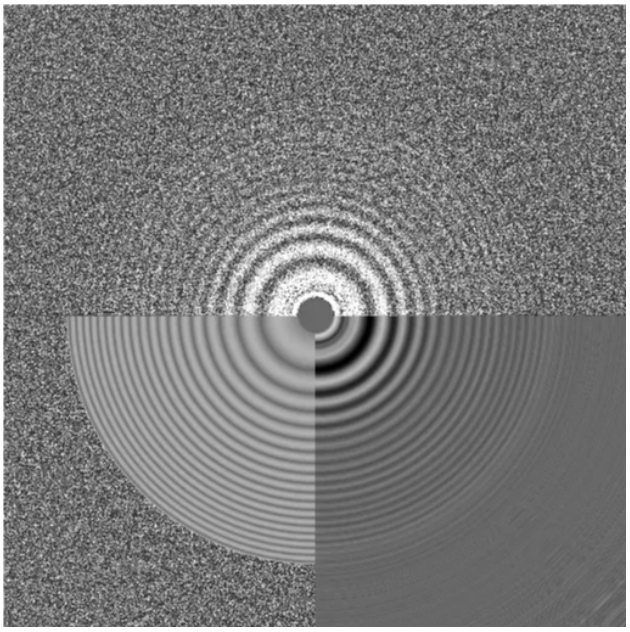


Figure 2.7: Output from CTF estimation showing the experimental and theoretical contrast transfer functions in amplitude and phase. The top panel displays the raw experimental power spectrum from the Fourier transform of the micrograph, the bottom-left panel presents the theoretical CTF amplitude calculated from the estimated microscope parameters, and the bottom-right panel shows the corresponding theoretical CTF phase, with alternating bright and dark bands indicating phase reversals.

the signal, affecting both the amplitude and phase of the various spatial frequencies present. One of these aberrations is caused by the defocus used to enhance contrast, and imperfections in the lensystem.

Figure 2.7 shows a CTF estimation, where the top half shows the raw experimental power spectrum obtained from the Fourier transform of the micrograph. It contains alternating light and dark rings (called Thon rings) along with the original noise background, and the contrast of the rings is exactly as measured, without any normalization or enhancement. The bottom-left quadrant is a noise-free simulation of the CTF amplitude calculated from the estimated microscope parameters such as defocus, astigmatism, spherical aberration, and amplitude contrast. It shows ideal Thon rings without any experimental imperfections and is used to check whether the experimental rings align with the theoretical prediction. The bottom-right quadrant represents the theoretical CTF in phase, displaying alternating positive and negative regions across spatial frequencies. The dark and bright bands correspond to opposite signs of the phase shift introduced by the microscope optics, and like the amplitude representation, it is noise-free, making the zero-crossings and phase reversals easy to identify. The middle grey disc is a masked area ignored during adjustment because the very low frequencies are dominated by variations in illumination, ice thickness, and other non-idealities that do not follow the CTF. It does not contain useful information for adjusting defocus.

Currently, CTFFIND4 [50] is used for CTF estimation and correction, although

there are variants that also work.

2.4.3 Particle Picking

Particle selection is a fundamental step in reconstructing high-resolution density maps. Due to the low SNR present in micrographs, a sufficiently large number of particles are required for reconstruction.

Historically, the approach used for this step has been template matching. Template matching uses cross-correlation between patches of a micrograph and particle templates. This method is highly dependent on image quality and has significant difficulties with heterogeneous datasets, or even with particles that lack symmetry. Other approaches based on edge detection or feature extraction also struggle with heterogeneity and asymmetry.

Methods based on deep neural networks have gained traction in recent years. One of the most widely used is TOPAZ [5], a method based on convolutional networks and trained under the positive-unlabeled data paradigm to avoid having to label too many images. Another method used is WARP [56], which uses a residual architecture, enabling the training of deeper networks. Finally, crYOLO [59], which uses the popular YOLO object detection algorithm, is worth mentioning. This has the advantage of needing to go through the entire image only once.

2.4.4 2D Classification and Particle Filtering

Raw datasets often contain a significant fraction of unsuitable particles, including contaminants, damaged molecules, ice artifacts, carbon edges, or poorly aligned particles. Two-dimensional classification is a fundamental step for identifying and removing such low-quality particles before proceeding to high-resolution refinement. Modern 2D classification algorithms typically rely on maximum-likelihood approaches to assign particles to classes based on their similarity in projection space. In RELION [13], this is implemented through a Bayesian framework that iteratively refines class averages while marginalizing over alignment parameters, allowing for robust classification even at low SNR. CryoSPARC [47] uses a stochastic gradient descent-based optimization of a probabilistic model, which achieves rapid convergence and can handle very large datasets efficiently. Other packages offer alternative implementations, including multivariate statistical analysis followed by multi-reference alignment.

Strategies for effective 2D classification often involve an initial round with a relatively large number of classes (e.g., 100–200) to capture the diversity of particle projections, followed by manual inspection to discard classes lacking discernible structural features. Iterative rounds of classification and filtering can progressively enrich the dataset with high-quality particles while removing spurious or misaligned images.

2.4.5 Postprocessing

Significant advances in equipment and image processing algorithms have enabled cryo-EM to achieve the ability to generate high-resolution three-dimensional maps, from which detailed structural information of biomolecular complexes can be extracted. This progress has made the assessment and verification of resolution in 3D reconstructions an increasingly important aspect to ensure the reliability of the results. However, in the field of electron microscopy, the concept of "resolution" still lacks a unified and universally accepted definition, and the methodologies currently used differ significantly in both their approach and the results obtained. In practice, the most widely used approach to estimate the overall resolution of a 3D map is Fourier Shell Correlation (FSC). This method calculates the correlation, in Fourier space and for different spatial frequency ranges, between two independent reconstructions of the same molecule. Resolution is defined as the point at which the FSC curve drops below a predefined threshold value (e.g., 0.143 or 0.5). Although it is a de facto standard, this procedure has limitations: it requires setting an arbitrary threshold and does not adequately reflect local quality variations within the density map.

To overcome this last limitation, local resolution estimation methods have been developed that assess the quality of different regions of the three-dimensional map. One of the first was BlocRes [11], which applies the FSC principle to cubic windows offset across the entire volume. In each window, the local resolution is calculated independently, allowing for a resolution map to be obtained. However, this method inherits the limitations of the global FSC and adds the need to define the size of the analysis window, which influences the spatial resolution of the resulting map and can introduce bias.

Another method, ResMap [12], takes a different approach: it analyzes the presence of detectable three-dimensional sinusoidal waves above the noise level at each point on the map. The rationale is that the ability to detect a periodic signal at a given frequency is directly related to local resolution. This method avoids the use of moving windows and works point-by-point, although its reliability depends on an accurate noise estimate.

Subsequently, MonoRes [58] was developed. It uses a similar principle to ResMap but uses the monogenic amplitude at different frequencies to quantify resolution. The method compares the monogenic signal in the analyzed region with the monogenic amplitude of the estimated noise, within a predefined resolution range. This approach aims for a more robust detection of the frequencies present, but also requires an accurate estimate of noise variance and intensive computational processing.

Despite the usefulness of these approaches, they all share certain disadvantages: they require long computation times, depend on user-defined parameters (such as window size or frequency ranges), and can produce significantly different results for the same map.

In recent years, the incorporation of deep learning techniques into the field of cryo-EM has led to the development of new local resolution estimation methods that attempt to overcome these limitations. Among them, DeepRes [48] stands out,

a method based on convolutional neural networks (CNNs) trained to predict local resolution directly from density maps. This approach is capable of detecting local quality variations induced by different post-processing processes, such as isotropic filtering, model-based and non-model-based sharpening, and noise suppression, all of which are common in modeling workflows. By not relying on fixed thresholds or window analysis, and by learning patterns directly from the data, DeepRes offers greater flexibility and potential for generalization to different types of maps.

However, given that there is still no general consensus on the optimal method for determining local resolution, and that estimates can vary significantly between approaches, research in this area remains active. The development of methods based on neural networks and other artificial intelligence techniques is emerging as a promising avenue for improving the accuracy, reproducibility, and efficiency of cryo-EM map resolution estimation.

2.4.6 Atomic Model Building

Atomic models play a fundamental role in structural biology, as they provide precise, interpretable representations of the three-dimensional arrangement of atoms in macromolecules. While cryo-EM maps reveal the overall molecular shape and density distribution, atomic models are essential for understanding the structural basis of function. These models serve as the bridge between experimental density maps and biochemical insight, and are critical for guiding hypotheses in mechanistic studies, drug design, and protein engineering.

To address the challenges of model building in intermediate-resolution cryo-EM maps, where manual interpretation is often difficult, several deep learning–based methods have recently been proposed to automate this process. One of these methods is Emap2sec [61], which uses a convolutional neural network to perform local structure detection across the entire 3D map. Benchmarking on maps with resolutions between 5 Å and 10 Å demonstrated that Emap2sec achieves more accurate detection of secondary structure elements and improved validation scores compared to traditional methods. However, the method shows limitations in accurately modeling specific structural elements such as alpha-helices and beta-strands.

A more recent development, EMBuild [22], applies a nested U-Net architecture (UNet++) [68], a fully convolutional network originally designed for image segmentation, to the task of atomic model building from intermediate-resolution maps. EMBuild has been evaluated on single-particle cryo-EM datasets with resolutions between 4 Å and 8 Å, demonstrating excellent performance in fitting reliable atomic structures into the density maps. While further improvements are still necessary, these deep learning–based approaches represent a promising direction for facilitating and accelerating model building in cases where the resolution is insufficient for conventional, fully manual interpretation.

2.5 Image Formation Model

The imaging process can be mathematically modeled as:

$$I_i(\mathbf{x}) = P_{\theta_i}(V_i) * \mathrm{PSF_i} + n(\mathbf{x}), \tag{2.1}$$

where V is the volume representation of the specimen, $P_\theta(V)$ denotes the projection of V with orientation θ, PSF is the point spread function, which is the result of applying the inverse Fourier transform to the CTF, i is the index of a given particle, and $n(\mathbf{x})$ represents noise, typically modeled as a combination of Poisson and Gaussian components.

In the frequency domain, Equation 2.1 can be expressed as

$$\hat{I}_i(\mathbf{k}) = \hat{P}_{\theta_i}(V_i)(\mathbf{k}) \cdot \mathrm{CTF}_i(\mathbf{k}) + \hat{n}(\mathbf{k}), \tag{2.2}$$

where $\hat{I}_i(\mathbf{k})$ is the Fourier transform of the observed image, $\hat{P}_{\theta_i}(V_i)(\mathbf{k})$ is the Fourier transform of the projection of the i-th particle along orientation θ_i, $\mathrm{CTF}_i(\mathbf{k})$ is the CTF, and $\hat{n}(\mathbf{k})$ represents noise in Fourier space. This formulation highlights that the CTF acts as a multiplicative modulation of the projected signal in the frequency domain.

Working in Fourier space is advantageous for cryo-EM image processing because the convolution with the microscope point spread function, which is computationally expensive in real space, becomes a simple element-wise multiplication with the CTF in frequency space. This greatly reduces the computational cost when simulating or correcting images. Moreover, the Fourier Slice Theorem allows a practical reconstruction process in this domain.

2.6 Fourier-Slice Theorem

Also called Central-Slice Theorem or Projection-Slice Theorem, the Fourier-Slice Theorem states that the 2D Fourier Transform of a projection of a 3D volume along a given direction is equal to a central slice of the 3D Fourier Transform of that volume, taken in the corresponding direction. Therefore, collecting Fourier Transforms of different projections is equivalent to sampling the 3D Fourier Transform of the volume over different planes.

For example, given the projection $P(x, y)$ of a volume $V(x, y, z)$ along axis z as follows

$$P(x, y) = \int_z V(x, y, z)\, dz, \tag{2.3}$$

and the Fourier Transformations for 2D $\mathcal{F}_2\{P(x, y)\}$ and for 3D $\mathcal{F}_3\{V(x, y, z)\}$, then the Fourier 2D Transform of $P(x, y)$ is

$$\mathcal{F}_2\{P(x, y)\} = \mathcal{F}_3\{V(x, y, z)\}\big|_{k_z=0}, \tag{2.4}$$

where $k_z = 0$ represents the projection along axis z.

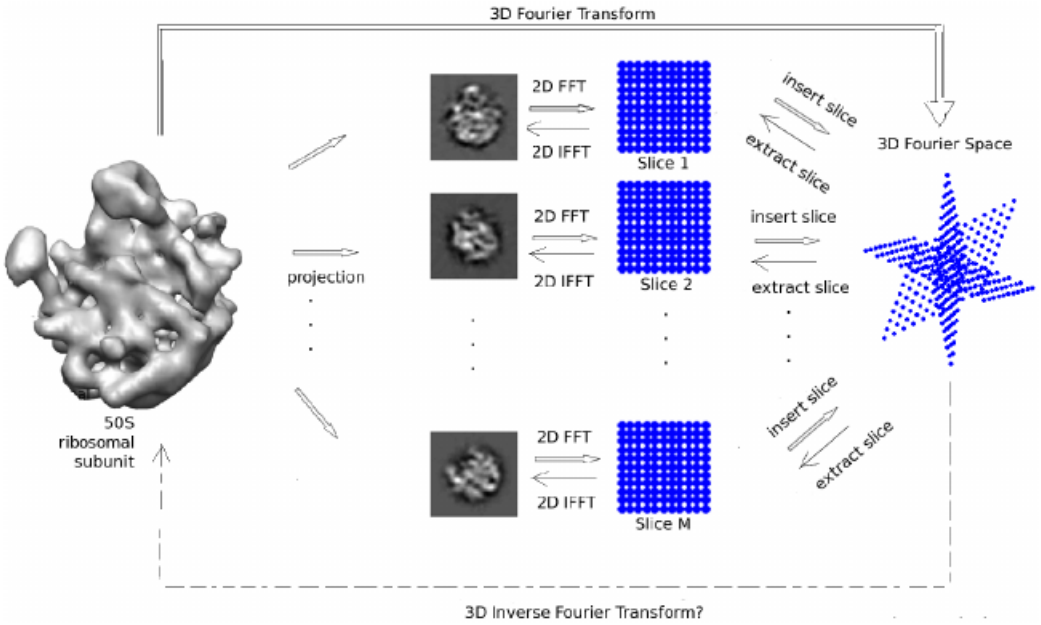


Figure 2.8: Illustration of the Fourier Slice Theorem applied to single-particle cryo-EM. The 3D structure of a particle is projected along different orientations to generate 2D images. Each 2D image undergoes a 2D Fourier transform, producing slices that correspond to central sections of the 3D Fourier transform of the particle. These slices can be inserted into or extracted from the 3D Fourier space, enabling the reconstruction of the full 3D volume via a 3D inverse Fourier transform. The figure was extracted from [60].

The Figure 2.8 illustrates the Fourier Slice Theorem in the context of single-particle cryo-EM reconstruction. According to this principle, each two-dimensional projection image corresponds to a central slice through the three-dimensional Fourier transform of the original particle density. By acquiring projections from multiple orientations, it becomes possible to fill the 3D Fourier space with these slices. The complete 3D structure can then be recovered by performing an inverse 3D Fourier transform on this assembled Fourier volume. This relationship between 2D projections and the 3D structure forms the mathematical foundation of single-particle reconstruction methods in cryo-EM.

2.7 Fourier Shell Correlation

The Fourier Shell Correlation [57] (FSC) is the three-dimensional extension of the two-dimensional Fourier Ring Correlation, a commonly used resolution criterion in electron and fluorescence microscopy [4, 29]. The FSC measures the cross-correlation between two volumes by comparing their respective frequency shells in three-dimensional Fourier space. Mathematically, given two volumes and their 3D Fourier Transformations, also called structure factors, $F_1(r)$ and $F_2(r)$, where r is the 3D spatial frequency vector, then

$$\mathrm{FSC}(r) = \frac{\sum_{r \in r_i} F_1(r) \cdot F_2^*(r)}{\sqrt{\sum_{r \in r_i} F_1^2(r) \cdot \sum_{r \in r_i} F_2^2(r)}}, \tag{2.5}$$

where F_2^* denotes complex conjugation and the summations are taken over all Fourier-space voxels r contained in the shell r_i.

In practice, the FSC is computed by dividing the dataset into two independent halves and performing separate reconstructions (commonly referred to as half-maps). Since both reconstructions are expected to contain the same underlying structural signal but different realizations of noise, their correlation provides an estimate of the SNR at different spatial frequencies. As spatial frequency increases, the amount of structural signal generally decreases relative to noise, and the FSC curve decays accordingly.

To convert the FSC curve into a resolution estimate, it is compared against a cutoff threshold. A commonly adopted criterion is the so-called *gold-standard* threshold of 0.143 [51], which is derived from a variance-based statistical model. This model shows that when two independent half-maps contain the same signal and uncorrelated Gaussian noise, an FSC value of 0.143 corresponds to the frequency where the SNR is equal to 0.5. In other words, at this frequency the structural signal is half as strong as the noise contribution, marking a practical resolution limit beyond which the map is dominated by noise rather than reliable structural features.

In this work, the reported FSC values correspond to the resolution at which the FSC curve intersects the 0.143 threshold. This provides a standardized and reproducible way to summarize the effective resolution of each reconstruction. The units of every FSC intersection result will be represented in Angstroms (Å).

2.8 UMAP

Uniform Manifold Approximation and Projection (UMAP) [39] is a non-linear dimensionality reduction technique designed to preserve local and global features of the data. This technique belongs to the class of k-nearest neighbor (k-NN) based graph learning algorithms, alongside methods like Laplacian Eigenmaps and t-SNE. Like these algorithms, UMAP operates in two main stages. The first stage, graph construction, involves computing a weighted graph that captures the local relationships between points based on their k-NN. Although additional transformations may be applied, the fundamental structure is a weighted k-NN graph that reflects the data's local geometry. The second stage, often called graph layout, defines an objective function that encodes the desired properties of the graph—typically the preservation of local connectivity—and then finds a low-dimensional representation that optimizes this objective.

UMAP relies on three key mathematical assumptions:

- The data is uniformly distributed on a Riemannian manifold. This assumption implies that the high-dimensional data points lie on, or close to, a

smooth and continuous manifold embedded in the ambient space.

- The manifold is locally connected. Local connectivity assumes that small neighborhoods within the manifold form connected regions.
- The Riemannian metric is locally constant (or can be approximated as such). The Riemannian metric defines how distances and angles are measured on the manifold. Assuming it is locally constant means that within small neighborhoods, the geometry of the manifold can be approximated as Euclidean.

For the graph construction stage, given a dataset $X = \{x_1, x_2, ..., x_n\}$, a hyperparameter k and a distance metric d, the k-nearest neighbors (k-NN) are computed for each data point x_i, yielding the set $\{x_{i_1}, x_{i_2}, ..., x_{i_k}\}$. The weights of the resulting weighted graph are then defined as:

$$w_{ij} = \exp\left(-\frac{d\left(x_i, x_{i_j}\right) - \rho_i}{\sigma_i}\right), \tag{2.6}$$

where ρ_i is the distance to the closest neighbor of x_i and σ_i is a normalization factor. The resulting graph (or set of graphs) can be interpreted as encoding local neighborhood probabilities.

In the dimensionality reduction stage, UMAP uses what the original paper refers to as attractive and repulsive forces. Intuitively, this corresponds to bringing closer the points that are neighbors in the original graph, and pushing apart those that are not, but in a lower-dimensional space (typically 2D or 3D). The corresponding similarity function in the low-dimensional space is given by:

$$\tilde{w}_{ij} = \frac{1}{1 + a \times ||y_i - y_j||^{2b}}, \tag{2.7}$$

where y_i and y_j are points in the low-dimensional embedding, and a and b are hyperparameters.

To ensure that the high and low-dimensional graphs are aligned, UMAP minimizes the cross-entropy between them as follows

$$\mathcal{L} = \sum_{(i,j)} w_{ij} \log\left(\frac{w_{ij}}{\tilde{w}_{ij}}\right) + (1 - w_{ij}) \log\left(\frac{1 - w_{ij}}{1 - \tilde{w}_{ij}}\right). \tag{2.8}$$

This objective is optimized using stochastic gradient descent.

Compared to t-SNE, UMAP offers several practical and theoretical advantages. While both methods aim to preserve local structure, UMAP has been shown to better retain global relationships in the data, producing embeddings that are more faithful to the overall topology of the original high-dimensional space. In addition to its improved preservation of global structure, UMAP also demonstrates significantly better computational performance. Its scalability with large datasets and faster runtime make it a more efficient choice, especially in modern workflows involving thousands of data points [39].

For a stronger mathematical foundation, the reader is invited to read the original work [39].

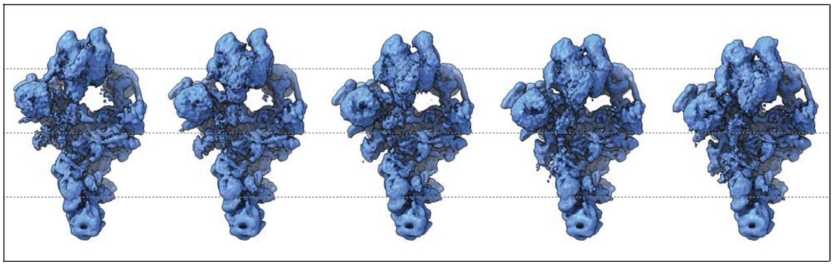


Figure 2.9: 3D reconstructions showing conformational heterogeneity in particles from the EMPIAR-10180 dataset, obtained with CryoDRGN. The figure was extracted from [65].

2.9 Heterogeneity

The objective of cryo-EM processing algorithms is to obtain a 3D reconstruction of a given macromolecule from a dataset of images, where each image corresponds to a 2D projection of the macromolecule with different poses. In order to obtain high-resolution reconstructions, these algorithms must accurately estimate both the orientations of the particles and the underlying structural variability present in the dataset. Without accounting for such variability, the averaging process inherent to reconstruction may blur important features, ultimately limiting resolution and obscuring biologically meaningful information.

Structural heterogeneity is a fundamental characteristic of many macromolecular complexes, particularly proteins, which can adopt multiple distinct structural states under physiological conditions. Structural heterogeneity plays a critical role in regulating protein function, facilitating molecular assembly, and enabling dynamic interactions with other biomolecules. In the context of cryo-EM, addressing structural heterogeneity is essential for accurate reconstruction and interpretation of macromolecular structures, as failure to do so can obscure biologically relevant states or lead to artifacts in the final maps.

The structural heterogeneity of macromolecules can be classified as conformational heterogeneity and compositional heterogeneity. The difference between them has already been emphasized in this document, but a more exhaustive analysis is required. Conformational heterogeneity refers to differences in the shape of a single molecular composition. That is, all particles have the same components (proteins, subunits, ligands, etc.), but these molecules adopt different spatial arrangements. An example of conformational heterogeneity can be seen in Figure 2.9, where each volume represents a different conformation of the same biomolecule, demonstrating the structure's ability to adopt multiple shapes.

Compositional heterogeneity refers to variations in the molecular composition among particles, meaning that certain subunits are present in some particles but absent in others. An example is shown in Figure 2.10, where three main classes can be distinguished: C, D, and E, each with its own subclasses. Class C lacks the central protuberance (CP), class D exhibits reduced density at the ribosomal base, and class E contains both the CP and the base density. Even within subclasses, distinct compositional differences are still apparent.

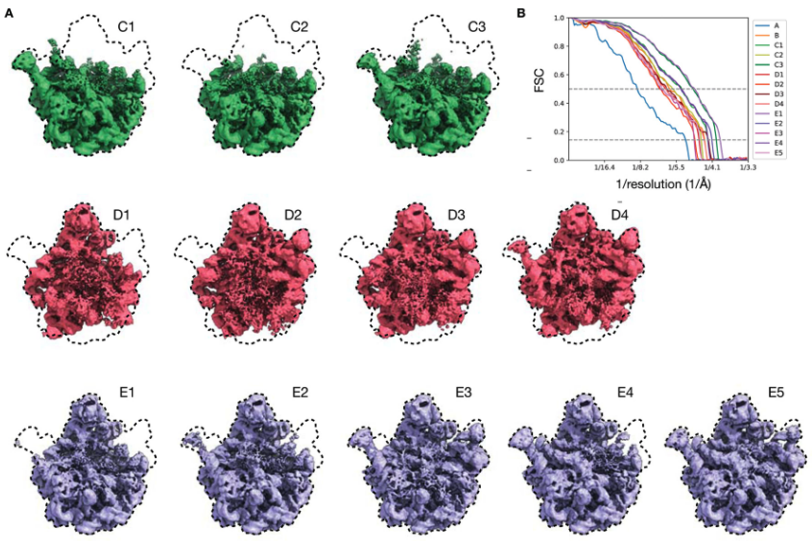


Figure 2.10: (A) Compositional heterogeneity in EMPIAR-10076. The classes differ in the presence of the central protuberance (CP) and the density at the particle's base. (B) shows the FSC curves for each reconstruction. The figure was extracted from [65].

In contrast, conformational heterogeneity involves subtle and often continuous changes in density maps, reflecting smooth structural transitions between states. Compositional heterogeneity is generally discrete, arising from the presence or absence of specific subunits or domains. From a structural perspective, conformational heterogeneity represents rearrangements in the tertiary structure, while compositional heterogeneity reflects differences at the quaternary structure level, affecting the actual assembly of the macromolecular complex.

2.10 Frealign

Frealign is a program for high-resolution refinement of 3D reconstructions from cryo-EM images of single particles. Briefly, Frealign performs projection matching to determine more accurate alignment parameters. Projections are calculated using the reference map provided on input, and alignment parameters for each particle projection are updated according to the projection that generates the highest correlation coefficient. Frealign is capable of performing global refinement through a systematic grid search. However, since this type of search is computationally expensive and initial particle poses are already available for the dataset in use, it will not be necessary in this case.

During the refinement stage, Frealign takes as input a reference density map, the initial poses of the particles, and imaging parameters such as magnification, defocus, and astigmatism. The particle images are transformed into the Fourier domain, where they are multiplied by the estimated CTF. Pose refinement is then carried out directly in Fourier space, which consists of maximizing a weighted correlation coefficient between the Fourier transform of the particle image and a

corresponding projection of the reference structure.

This optimization is carried out using a Powell optimization algorithm [45], which is a derivative-free method particularly suited for problems where gradients are difficult or impractical to compute. Mathematically, the goal of these algorithms is to minimize a scalar function $f : \mathbb{R}^n \to \mathbb{R}$, which can be expressed as

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}). \tag{2.9}$$

In Powell's directional method, the procedure begins from an initial point $\mathbf{x}_0 \in \mathbb{R}^n$ and a set of initial search directions $\mathbf{u}_1, \ldots, \mathbf{u}_n$ (e.g., the canonical basis). The algorithm proceeds as follows:

1. **Unidimensional minimization:** For each direction $\mathbf{u}_i$, a one dimensional minimization is performed:

$$\alpha_i^* = \arg\min_{\alpha} f(\mathbf{x}_{i-1} + \alpha\mathbf{u}_i) \tag{2.10}$$

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \alpha_i^*\mathbf{u}_i \tag{2.11}$$

   This is repeated for $i = 1, \ldots, n$.

2. **Construction of a new direction:** A new direction is defined based on the total displacement:

$$\mathbf{d} = \mathbf{x}_n - \mathbf{x}_0 \tag{2.12}$$

3. **Minimization along the new direction:**

$$\alpha^* = \arg\min_{\alpha} f(\mathbf{x}_n + \alpha\mathbf{d}) \tag{2.13}$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha^*\mathbf{d} \tag{2.14}$$

4. **Direction update:** The first direction $\mathbf{u}_1$ is replaced by the new direction $\mathbf{d}$:

$$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}, \mathbf{d} \tag{2.15}$$

5. **Repeat:** The process is repeated from $\mathbf{x}_{n+1}$ using the updated set of directions.

For reconstruction, the Fourier transforms of each projection are translated and oriented according to the poses calculated during refinement. These transforms are then accumulated following the Fourier Slice Theorem. Since, in general, the sample points of the discrete 2D Fourier transform of the image will not coincide with sample points of the Fourier transform of the reconstruction, an interpolating function has to be used. Such a function is derived from the Fourier transform of a box with dimensions equal to the dimensions of the reconstructed volume. Thus,

in the interpolation step, a box transformation is centered on each sample of the 2D Fourier transform of the image and evaluated at the nearest neighbor of the 3D Fourier transform of the reconstruction. In the process of the reconstruction from images of all particles, sums are accumulated at each sample point of the 3D Fourier transform:

$$R_i = \frac{\sum_j w_j^2 b^2 c_j P_{ij}}{f + \sum_j (w_j b c_j)^2} \tag{2.16}$$

where R_i is the final volume reconstruction in sample i, P_{ij} is the contribution of the particle j in sample i, c_j is the CTF correction of the particle j, b is the box transformation, w_j is a weighting factor and f is a value used to avoid numerical instabilities during reconstruction. The weighting factor allows weighting of contributions from individual images according to their correlation with the reference.

Frealign Fortran code is available. The pipeline in this work uses version 10, also known as FrealignX.

Esta página ha sido intencionalmente dejada en blanco.

Chapter 3

Related Work

This chapter presents a selection of works that address the reconstruction of one or more density maps using deep learning techniques. These works were chosen based on their strategies for tackling key challenges in this stage, such as low SNR, pose estimation, and the representation of structural heterogeneity. Their advantages and limitations are discussed, along with relevant considerations and decisions related to the reconstruction problem itself.

3.1 Problems to be addressed

Given the biological importance of accurately capturing the structural information of macromolecules, and the complexity of this task, several deep learning-based approaches have recently been proposed to address this challenge. Particularly for modeling structural heterogeneity in macromolecular complexes, traditional reconstruction pipelines often struggle to represent variability, especially when it lies along continuous manifolds or when discrete states are not well separated. Deep learning approaches have shown great promise in overcoming these limitations by leveraging data-driven priors and flexible generative architectures.

A key difficulty in this context involves the estimation of nuisance variables, which are the poses and the CTF parameters. While these variables do not carry biological significance, they are essential for producing high-resolution reconstructions.

Another major challenge tackled by deep learning techniques is the inherently low SNR of cryo-EM images, a consequence of the need to minimize electron dose to preserve the integrity of the sample. Neural networks must therefore learn to extract meaningful structural features from highly noisy data.

The works presented in this chapter were selected for their relevance in addressing the central issues in cryo-EM reconstruction. For readers interested in a more comprehensive overview of deep learning methods applied to cryo-EM reconstruction, it is recommended to read the following review on the subject [17], which was very useful for this work.

3.2 Homogeneous Methods

The homogeneous methods presented in this section were selected because, although they follow a homogeneous reconstruction strategy, they propose alternative approaches to pose estimation. Traditional homogeneous methods typically rely on aligning particles to a consensus volume to estimate their poses. In macromolecules with structural heterogeneity, this can lead to poor results, since a single volume is not representative of all particles, and consequently, the alignment of some particles with the volume is poor.

In contrast, the methods discussed here avoid relying on a single-volume alignment for pose estimation. While they still generate only one density map, they explore alternatives such as bypassing pose estimation entirely or inferring poses through neural networks, thus mitigating the limitations imposed by structural variability.

3.2.1 CryoGAN

CryoGAN [21] is a reconstruction technique based on Generative Adversarial Networks (GANs) [19] that aims to reconstruct a consensus density map without the need for pose estimation. To achieve this, it adopts the traditional GAN framework, but replaces the generator with a cryo-EM physics simulator, which generates projections of an estimated volume $\hat{X}$ according to the image formation model described in Equation 2.1. The CryoGAN framework is illustrated in Figure 3.1, where $\mathbf{H}_\varphi$ represents the convolution with the CTF, the projection of the volume, and the translation and rotation of the particle. In this approach, a physics-based cryo-EM simulator generates 2D projections from the current estimate of the 3D density map, which is the quantity that the network learns to reconstruct. The parameters of CryoGAN are directly the voxels of the volume, which are passed through the cryo-EM physics simulator. In contrast, the pose and CTF parameters are sampled from a prior probability distribution.

One of the main advantages of CryoGAN is that it avoids the computationally expensive and highly non-convex problem of pose estimation. Additionally, CryoGAN is a reference-free method, meaning that it does not require an initial volume to begin the reconstruction (it is initialized with a zero-valued volume).

Despite these advantages, the method has several limitations. First, it assumes structural homogeneity and is therefore not suited for analyzing conformational or compositional variability of macromolecules. This limitation is closely linked to the relatively low resolution achieved by the method, with reported results reaching only 10.8Å. Another commonly cited drawback in the context of these algorithms is the lack of validation on experimental data. CryoGAN reports results exclusively on synthetic datasets, which may not accurately reflect the inherent challenges of real cryo-EM data. Moreover, the field lacks a standardized synthetic benchmark dataset for evaluating such methods, making comparisons across different approaches difficult.

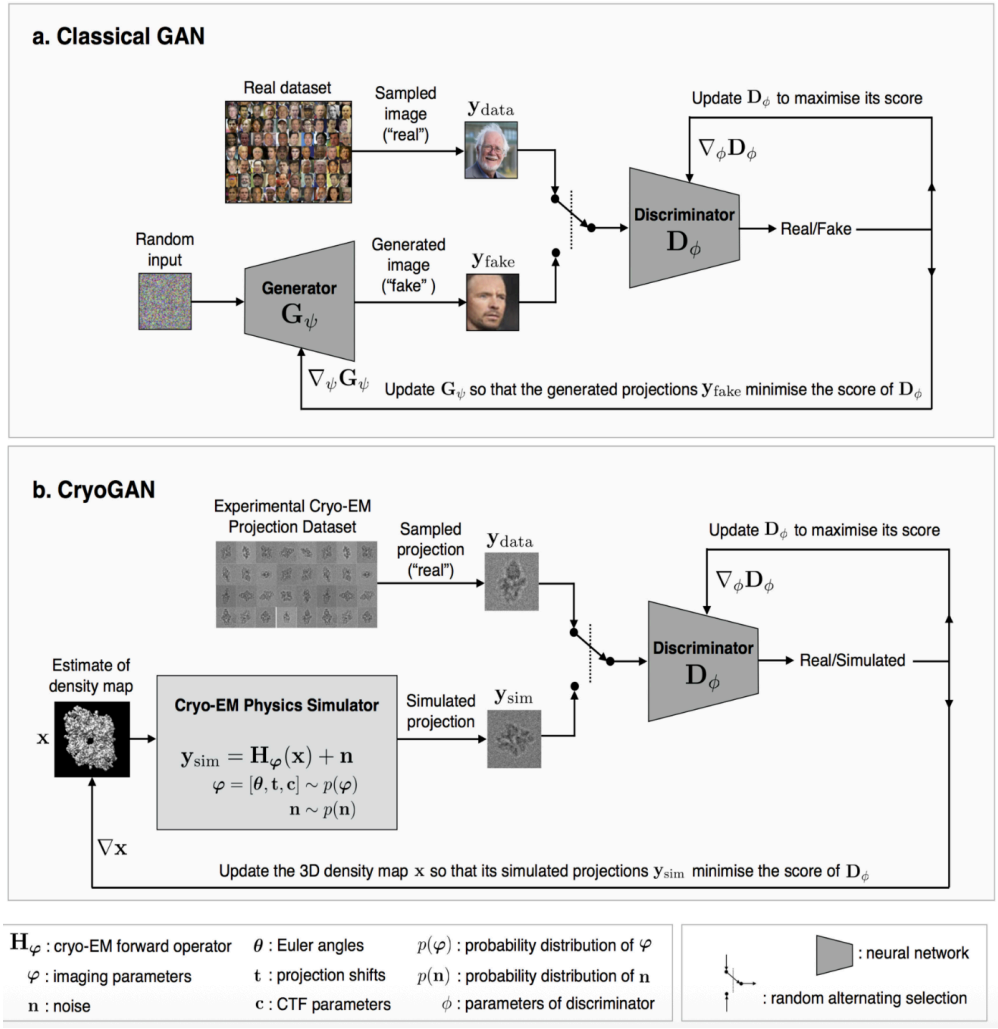


Figure 3.1: Comparison of a classical GAN and CryoGAN extracted from [21]. (a) In a classical GAN, a generator network produces synthetic images from random inputs, which are then evaluated by a discriminator that distinguishes real images from generated ones. Both networks are updated adversarially. (b) In CryoGAN, instead of a neural network generator, a physics-based cryo-EM simulator generates projections from an estimated 3D density map using imaging parameters and noise models. The discriminator distinguishes real experimental cryo-EM projections from simulated ones. The density map is updated iteratively so that its simulated projections increasingly resemble experimental data.

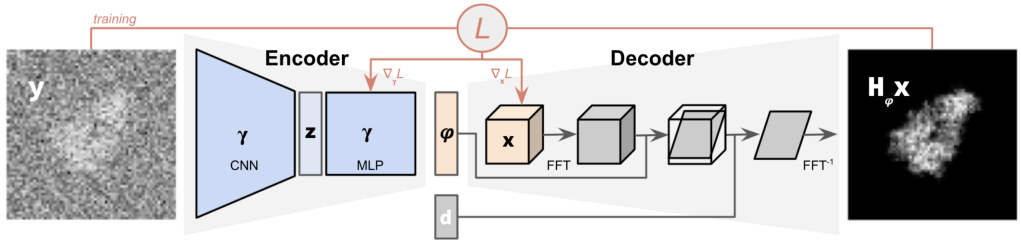


Figure 3.2: In CryoPoseNet, each noisy cryo-EM particle image y is passed through an encoder that estimates the particle's orientation φ. This information is then used by a decoder to generate a simulated projection of the 3D structure in that specific orientation. The projection process is modeled using the Fourier slice theorem: the current estimate of the structure x is transformed into Fourier space, a plane orthogonal to φ is extracted, and the result is modified by the known CTF with parameter d. During training, both the encoder's convolutional and multilayer perceptron weights γ and the 3D structure x itself are refined. The optimization minimizes a loss function L that measures the discrepancy between the experimental particle image y and its corresponding reconstruction $H_\varphi x$. This figure was extracted from the original CryoPoseNet work [43].

3.2.2 CryoPoseNet

Another homogeneous framework worth mentioning is CryoPoseNet [43], which introduces a fully unsupervised method for reconstructing a density map along with the rotations of individual particles projections, without requiring any reference volume (it is also reference-free). The implementation consists of a convolutional neural network that learns shift-invariant features, followed by a multilayer perceptron (MLP) that maps these features to Euler angles. A cryo-EM physics model, similar to that used in CryoGAN, is then applied to generate a projection that resembles the input image. The decoder operates in Fourier space in order to leverage the Fourier Slice Theorem. Figure 3.2 illustrates the basic architecture, where γ denotes the network parameters, φ the predicted poses for each projection, and $\mathbf{d}$ the CTF parameters (which are assumed to be pre-estimated).

A major strength of CryoPoseNet is the fact that it is capable of estimating particle orientations, where the computational cost is independent of the number of input images. This is achieved by learning a mapping function φ from particles to rotations, parameterized by γ (the trainable parameters of the encoder). Crucially, the number of parameters does not scale with the number of orientations to be estimated. It is important to note, however, that the method assumes known CTF parameters and in-plane translations. Also, CryoPoseNet is a global method (using a shared encoder function across all particles), which allows it to integrate information from the entire dataset when estimating individual poses.

Nonetheless, the method also has several limitations. Like CryoGAN, it operates under the assumption of structural homogeneity and has only been evaluated on synthetic data. The authors report challenges in handling noise, which is a major characteristic of real cryo-EM datasets. Additionally, the cryo-EM physics model used in the framework is subject to limitations due to interpolation artifacts when extracting 2D projections from the 3D volume in Fourier space.

3.3 Heterogeneous Methods

The focus of the present work is to exploit protein heterogeneity to improve pose estimation. This section focuses on deep learning methods for heterogeneous reconstruction, where neural networks are used to model conformational or compositional variability in macromolecular structures.

It is important to note that this is not intended as an exhaustive review of all deep learning applications in cryo-EM. For example, methods such as Topaz and DeepEM apply neural networks during the particle picking stage, performing classification directly from micrographs and obtaining labels that are used to generate multiple reconstructions. These approaches operate at different stages of the processing pipeline and are not included in the focus of this section.

Additionally, traditional methods for modeling structural heterogeneity, such as Multibody Refinement or 3D classification, are not included either, as they do not involve deep learning techniques.

3.3.1 3DFlex

3DFlex [46] is a deep learning framework designed to model continuous structural heterogeneity in cryo-EM data. It is based on the hypothesis that conformational variability arises from physical processes that tend to preserve mass and local geometry. To capture this, the method leverages a neural network architecture that simultaneously learns a canonical 3D density map (consensus volume) and a deformation field. The deformation field is conditioned on the location of each particle projection in a latent space, effectively allowing each particle to be represented as a deformed version of the shared canonical volume.

A distinguishing feature of 3DFlex is its ability to model smooth, physically plausible motions while maintaining high-resolution features in the parts of the structure that remain rigid. The deformation is applied in real space rather than Fourier space, which enables the incorporation of spatial constraints but increases the computational cost. This approach avoids the Fourier Slice Theorem and instead performs real-space convolutions with the PSF rather than frequency-domain multiplications with the CTF. Although a reference volume is used, it is not provided externally but is instead learned jointly during training.

Importantly, 3DFlex is trained on real cryo-EM datasets, and its authors report that it captures characteristic motions associated with conformational heterogeneity. However, the method is not primarily designed for high-resolution refinement. The resulting reconstructions can often be further refined using conventional techniques. This is aligned with the common observation that conformational heterogeneity tends to manifest as local blurring in the density map—regions undergoing variability appear diffuse due to averaging over different states. In contrast, compositional heterogeneity typically results in partially visible features, as components may be present in some particles but not in others.

The core components of the model include the canonical map, the latent vectors that represent particle-specific conformations, and a flow generator that maps each latent vector to a deformation field applied to the canonical volume. The use of a

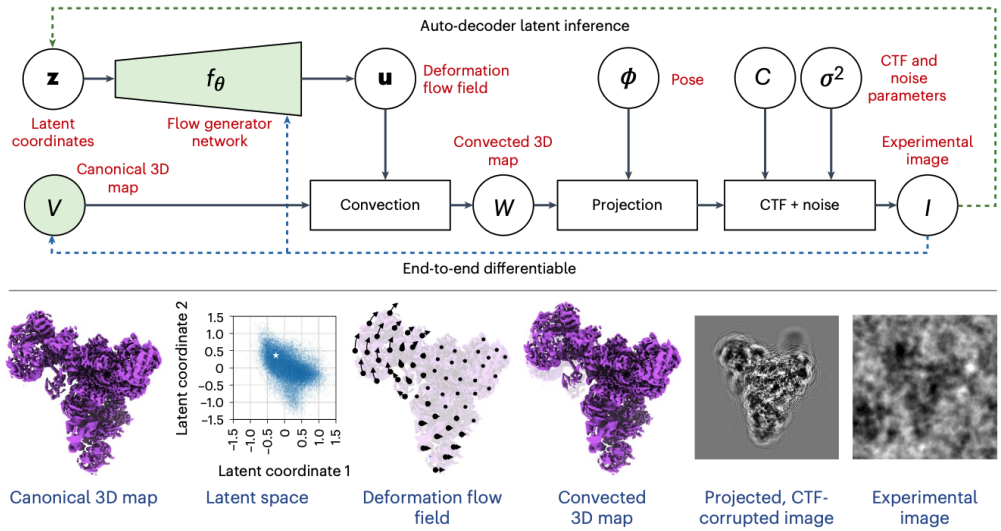


Figure 3.3: Overview of the 3DFlex model for capturing continuous conformational variability in cryo-EM. The model starts from a canonical 3D map V and latent coordinates $\mathbf{z}$ that describe the conformational state of each particle. A flow generator network f_θ predicts a deformation flow field $\mathbf{u}$ from the latent coordinates, which is then applied to the canonical map through a convection operator to produce a convected 3D map. This map is projected into 2D according to the particle's pose ϕ, modulated by the CTF and corrupted with noise to generate a simulated image. The model is trained end-to-end by comparing these simulated images to the experimental images I, simultaneously inferring latent coordinates, deformation fields, poses, and imaging parameters. The bottom panel illustrates the processing pipeline: starting from the canonical 3D map, through the latent space, deformation flow field, and convected map, to the projected CTF-corrupted image and its corresponding experimental image. The green blocks indicate components with learnable parameters. This figure was extracted from the original 3DFlex work [46].

global latent space ensures that information is shared across the dataset, and the learned deformation fields offer interpretable insights into the structural flexibility of the molecule.

Compared to other deep learning-based methods, 3DFlex focuses on physically constrained modeling of continuous motions. Alongside cryoDRGN, it is considered one of the most promising approaches for characterizing structural variability in cryo-EM [17], particularly in the context of continuous heterogeneity. Unlike traditional autoencoders, the inference mechanism in 3DFlex is tightly coupled to its deformation model, enabling the system to disentangle structural dynamics from static features.

Similarly to CryoGAN, 3DFlex performs backpropagation directly through the image formation process to update the voxel values of the 3D density map. This means that the gradients are propagated from the comparison between experimental and simulated projections all the way back to the underlying volume, allowing the network to iteratively refine the density itself.

The main limitations of 3DFlex come from its design assumptions and computational requirements. First, the method is not intended to model compositional

heterogeneity, which involves the presence or absence of distinct molecular components, rather than smooth deformations of a common structure. As such, 3DFlex may not adequately capture discrete or large-scale structural differences across particles. Second, its reliance on real-space operations introduces significant computational overhead compared to Fourier-based methods, which can exploit the efficiency of the Fourier Slice Theorem. Additionally, 3DFlex assumes that nuisance variables such as particle poses, in-plane translations, and CTF parameters are known a priori. These must therefore be estimated through separate preprocessing steps using other tools or algorithms, which can add further complexity to the workflow.

3.3.2 DynaMight

DynaMight [54] is another deep learning framework designed to address continuous structural heterogeneity in cryo-EM data. Unlike methods that aim to explore the full conformational landscape, DynaMight focuses primarily on identifying and modeling conformational variability with the specific goal of improving the resolution of the consensus density map. Its objective is not to recover the spectrum of conformations per se, but rather to correct for structural flexibility in order to refine the averaged structure.

The method is based on a variational autoencoder (VAE) architecture, where the decoder outputs 3D deformations of Gaussian pseudo-atoms. In this representation, each particle is modeled as a conformational variation of a shared set of pseudo-atoms—simple Gaussian density components that together approximate the macromolecule. This formulation assumes that all particles belong to a continuum of conformational states and does not support compositional heterogeneity. In fact, the authors explicitly state that datasets containing compositional variability must be pre-filtered before applying DynaMight.

Figure 3.4 illustrates the architecture, including the use of two encoders to compute the FSC between resulting density maps, which serves as a proxy for resolution. This dual-encoder setup is used to ensure that the model produces consistent and reproducible reconstructions, and to monitor structural variability across reconstructions.

DynaMight has been applied to real data, including the well-known EMPIAR-10180 dataset, and demonstrates its ability to improve map resolution in regions affected by conformational flexibility. By accounting for continuous variability during the reconstruction process, it aims to reduce local blurring and sharpen flexible regions of the molecule, which are typically problematic in traditional consensus averaging.

One of the strengths of DynaMight lies in its targeted approach: rather than attempting to model the full deformation field like 3DFlex or cryoDRGN, it focuses on a constrained, interpretable representation via pseudo-atoms. This can be advantageous for resolution-driven applications but limits its use for conformational landscape analysis.

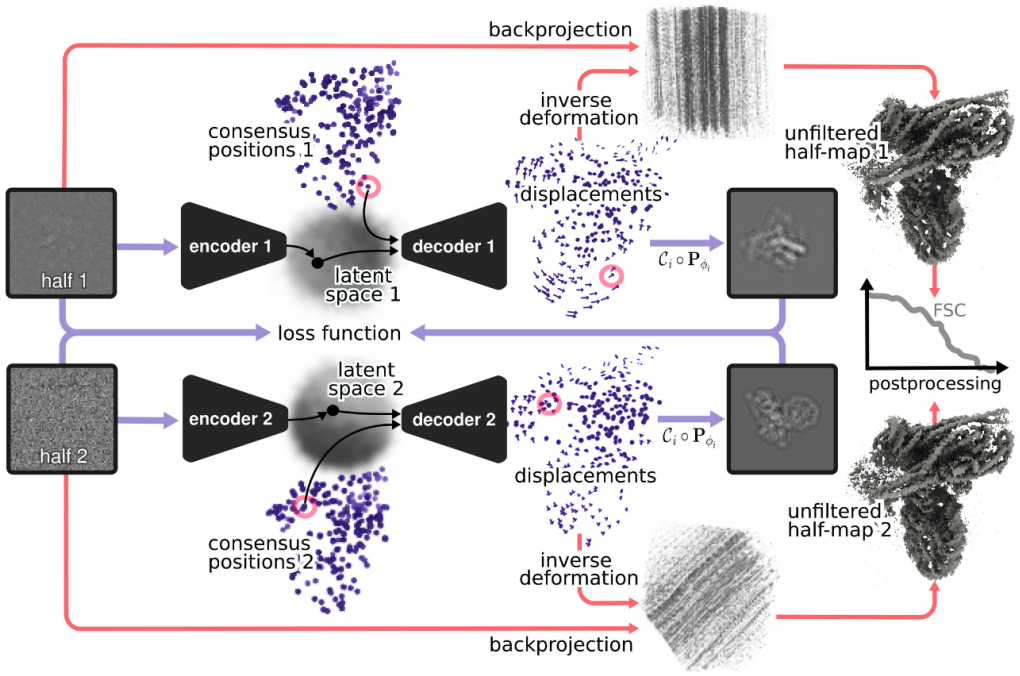


Figure 3.4: Schematic representation of the DynaMight architecture for modeling continuous heterogeneity in cryo-EM data. The particle images are split into two independent halves to enable cross-validation. Each half-dataset is processed through its own encoder–decoder, which maps the particle images into a latent space representation. The latent vectors are aligned to consensus positions, and particle-specific displacements are predicted. These displacements are inverted and applied to the corresponding particle images before 3D reconstruction through backprojection. The resulting unfiltered half-maps are then used to compute the FSC for resolution estimation and to generate the final postprocessed maps. The loss function enforces consistency between the latent representations learned from each half-dataset, ensuring that the model captures genuine structural variability rather than overfitting to noise. This figure was extracted from the original DynaMight work [54].

3.3.3 CryoDRGN

CryoDRGN (Cryo-EM Deep Reconstructing Generative Network) is a method designed to address the problem of structural heterogeneity. It is based on a VAE. The input consists of 2D particle images already selected and cropped, which are passed through an encoder that maps each image to a Gaussian distribution in a continuous, low-dimensional latent space. The idea is that each Gaussian in latent space represents the conformational and/or compositional state of the particle.

The decoder samples a latent vector from this distribution using the reparameterization trick and, together with the particle's pose vector, generates a 2D projection of an implicit density map. The training process minimizes the difference between the real particle image and the simulated image generated from the volume, using the estimated poses and CTF parameters. As a result, the model simultaneously learns a latent space that captures the heterogeneity of the particles

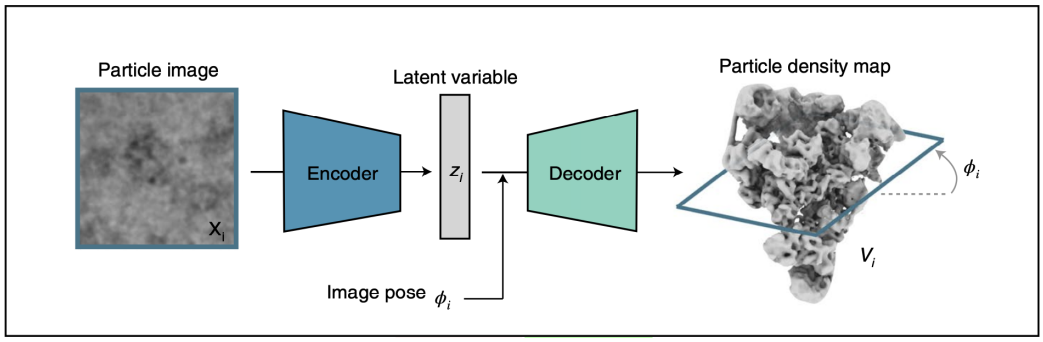


Figure 3.5: Block diagram of the cryoDRGN framework for heterogeneous cryo-EM reconstruction using deep generative models. A particle image x_i is passed through an encoder network that infers a low-dimensional latent variable z_i, representing the particle's conformation. Along with the known or estimated image pose ϕ_i, the latent variable is provided to a decoder network that reconstructs the corresponding 3D density map V_i. By jointly learning the encoder, decoder, and latent space, cryoDRGN models continuous structural variability across the dataset, enabling the generation of a range of conformations consistent with the experimental images.This figure was extracted from the original cryoDRGN work [65].

and a continuous 3D generative model.

To represent the density maps, cryoDRGN uses what is referred to as a coordinate-based neural network. This network learns a representation of the density map through a function that takes 3D coordinates and the latent vector as input and returns a density value. Formally, this can be described as $V : \mathbb{R}^{3+n} \rightarrow \mathbb{R}$, where V is the learned function and n is the dimensionality of the latent space.

It is important to note that the pose vector in this case is 3D, as the decoder does not model translations. Instead, it only applies a rotation to the coordinate system in Fourier space and compares the simulated projection with the translation-corrected and CTF-corrected particle image.

CryoDRGN also includes a suite of interactive tools that are highly valuable for exploring and interpreting results. Some of these tools enable visualization of the latent space by applying dimensionality reduction techniques such as PCA, t-SNE, or UMAP to the learned latent vectors. This allows for the inspection of how particles are distributed according to their conformational or compositional states. Clustering techniques such as K-Means or Gaussian Mixture Models (GMMs) can also be applied, either directly on the latent vectors or on their low-dimensional representation. This facilitates the identification of discrete structural states or compositional classes within the dataset.

Furthermore, it is possible to reconstruct density maps for arbitrary points in latent space, enabling the generation of as many 3D reconstructions as there are particles used during training. In fact, cryoDRGN can even be used as a synthetic data generator by sampling new points from the latent space (e.g., from a standard normal distribution) and generating corresponding 2D projections using the decoder.

Another useful functionality is the ability to filter out particles, either by removing entire clusters or using an interactive selection tool to manually exclude

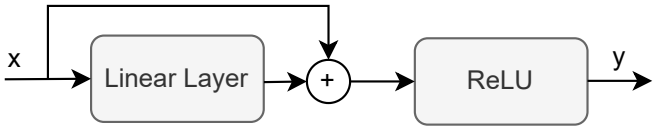


Figure 3.6: Residual block from cryoDRGN architecture

undesired particles. This is particularly helpful for cleaning up the dataset before further analysis or refinement.

One of the main advantages of cryoDRGN is its ability to represent both conformational and compositional heterogeneity in a continuous and data-driven manner. In various studies, trajectories in PCA or UMAP space have been used to investigate conformational changes, while clustering has been employed to uncover discrete compositional variability.

Its interactive tools are especially practical for exploring large and complex datasets, providing an intuitive way to interpret the latent space and generate hypotheses about structural variability.

However, a notable limitation of cryoDRGN is that it does not currently achieve high-resolution reconstructions. The reconstructions are typically limited to low or intermediate resolution, which may not be sufficient for atomic model building or fine structural interpretation.

The encoder from cryoDRGN's VAE maps 2D particle images to a distribution in a low-dimensional latent space, and the decoder maps coordinates in 3D space together with a latent variable back to a density value. The model learns a continuous function

$$f_\theta : (\mathbf{x}, \mathbf{z}) \mapsto \rho \in \mathbb{R},$$

where $\mathbf{x} \in \mathbb{R}^3$ are spatial coordinates in the volume, $\mathbf{z} \in \mathbb{R}^{d_z}$ is a latent vector describing the conformation, and ρ is the predicted electron density at that point.

In the residual decoder architecture, the function f_θ is parameterized by a multilayer perceptron (MLP) composed of residual linear blocks. Each block computes

$$y = ReLU\left(Wx + b + x\right), \tag{3.1}$$

where x is the block input, W and b are the learnable weights and biases, and $ReLU$ is the nonlinear activation function. The skip connection ensures that the block can learn residual corrections to the identity mapping, which improves gradient flow and stabilizes training in deeper networks. An example of a residual block from cryoDRGN can be seen in Figure 3.6

During training, the encoder network $q_\phi(\mathbf{z} \mid I, R)$ processes each 2D particle image I together with its known or estimated orientation R, producing the mean $\boldsymbol{\mu}$ and log-variance $\log \boldsymbol{\sigma}^2$ of the approximate posterior distribution in latent space. A latent vector $\mathbf{z}$ is then sampled using the reparameterization trick,

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, Id),$$

and concatenated to the 3D coordinates $\mathbf{x}$ when evaluating the decoder. The decoder output is compared against the Fourier transform of the experimental images, projected to the same orientation R, yielding a reconstruction loss in Fourier space. The total loss is the sum of the reconstruction term and the Kullback–Leibler divergence between the approximate posterior $q_\phi(\mathbf{z} \mid I, R)$ and the prior $\mathcal{N}(0, I)$, as in a standard VAE.

In this setup, the residual architecture improves the ability of the decoder to represent high-frequency details and subtle conformational changes in the protein structure, while maintaining stability during optimization. The network can be evaluated on any $\mathbf{z}$ in latent space to generate a corresponding 3D density, enabling both continuous conformational analysis and discrete clustering in the learned manifold.

3.4 Summary

The methods discussed above address, through various deep learning approaches, the challenges of pose estimation and structural heterogeneity in cryo-EM data. Some, such as CryoPoseNet, propose explicit solutions for estimating individual particle poses. Others, like cryoDRGN, 3DFlex, and DynaMight, are focused on modeling structural variability through continuous latent spaces or deformation fields. CryoGAN, in contrast, avoids direct pose estimation and learns a 3D volume via an adversarial learning framework. In general, these methods tend to focus on either pose estimation or heterogeneity modeling, but not both simultaneously, which limits their ability to accurately reconstruct structurally diverse systems where these two aspects are strongly interdependent.

Another common limitation among these approaches is the difficulty in achieving high-resolution density maps. Some of these methods have only been validated on synthetic datasets, where particle orientations and ground-truth structures are known. This is related to the low SNR problem.

Notably, cryoDRGN and DynaMight can generate one density map per particle, which is particularly valuable for studying continuous conformational changes or discrete subpopulations. However, these reconstructions pose significant challenges for validation: since they are not based on well-defined subsets of particles, traditional resolution metrics such as FSC cannot be applied, making it difficult to objectively assess the quality of the resulting maps.

The following chapter presents the method proposed in this work, which aims to jointly address structural heterogeneity and pose estimation. By leveraging structural variability as a source of information, the method improves pose estimation and captures structural diversity. It combines the noise robustness and feature learning capabilities of VAEs with a reconstruction and refinement strategy grounded in maximum likelihood estimation.

Esta página ha sido intencionalmente dejada en blanco.

Chapter 4

Proposed Method and Experimental Setup

This chapter presents the proposed method, detailing each step of the processing pipeline. Preliminary results are provided to justify specific design choices. Additionally, the expected advantages and potential limitations of the method are outlined based on theoretical considerations.

4.1 Workflow

This work aims to integrate cryoDRGN's particle feature learning capability with Frealign's high-resolution reconstruction and refinement techniques. In addition, the structural heterogeneity of particles will be leveraged to improve pose estimation by applying various clustering and dimensionality reduction techniques to identify distinct conformations and/or compositions. These clusters are then preprocessed to enable refinement and density map reconstruction for each group. Finally, the resulting poses from Frealign's refinement are processed to make them compatible for reuse in cryoDRGN.

This workflow can be seen on Figure 4.1. The colors are intended to differentiate the Frealign, cryoDRGN, and other blocks implemented in this work. The yellow block represents Frealign reconstructions and refinements, the red represents the cryoDRGN VAE training, and the blue blocks are blocks implemented in this work to achieve compatibility between both tools. The purple block is intended to be a combination of red and blue because the analysis function used was based on the one proposed by cryoDRGN, but many modifications were made to it. The light gray blocks outside the iterative cycle represent initial preprocessing, and various cryo-EM image processing tools can be used.

The iterative block was automated using a main bash script that iteratively calls the scripts corresponding to each block in the diagram in Figure 4.1.

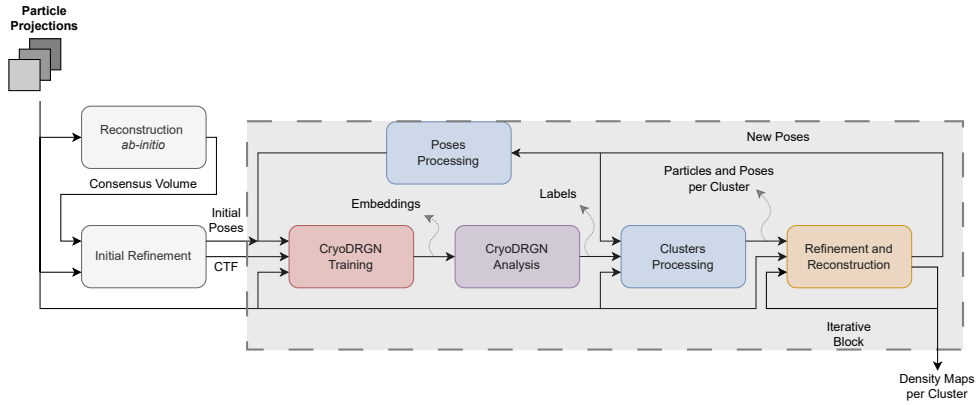


Figure 4.1: Proposed pipeline for discrete heterogeneous reconstruction and refinement. The yellow block represents Frealign code, the red block represents cryoDRGN code and the blue blocks represent code implemented in this work. The purple block intends to be a mix between red and blue, due to cryoDRGN analysis function was heavily modified. The light gray blocks are preprocessing steps that can be than with any cryo-EM software.

4.1.1 Reconstruction *ab-initio* and Initial Refinement

There are no strict constraints for the *ab initio* reconstruction and initial refinement steps. The user may use CryoSPARC for both stages, or use cryoDRGN decoder to generate density maps to be used as reference, and the refinement may be done with Frealign. The important step here is to obtain initial poses and the CTF estimation, which will not be refined in this work.

The pipeline also requires at least one reference density map, since Frealign needs it for pose refinement. There are several possible approaches to obtain such references. One approach is to use the consensus map employed in each class refinement; however, this forces all clusters to start from the same initial condition, which may obscure the true structural heterogeneity of the macromolecule. A second option is to use the maps generated by cryoDRGN for each cluster. Importantly, these are not reconstructions of the clusters themselves, but rather particle-level reconstructions. A practical strategy to obtain representative density maps in this case is to select the particle closest to the centroid of each cluster and use its reconstruction. Finally, another option is to perform a dedicated clustering and reconstruction step with Frealign prior to iteration, thereby generating high-resolution reference volumes for each cluster. While this last approach is the most computationally demanding, it provides the most accurate reference density maps for the initial iteration without compromising the heterogeneous representation. The latter option was used in the pipeline to achieve the best possible resolution results.

4.1.2 CryoDRGN Training

For cryoDRGN training, some hyperparameters had to be defined, such as the architecture size, minibatch size, number of epochs, and latent space size, among others. For this purpose, some experiments were conducted using only the training and analysis blocks, evaluating the network's performance using the reconstructions generated by the decoder instead of Frealign. Due to the training computational time, a smaller model with subsampled images was also used, and the model's capacity was evaluated for this case.

The hyperparameters used were as follows:

Hyperparameter	**Value**
Number of initial epochs	40
Number of iterative epochs	5
Batch size	8
Learning rate	0.0001
Number of hidden layers (enc)	3
Number of nodes in hidden layers (enc)	256/1024
Number of hidden layers (dec)	3
Number of nodes in hidden layers (dec)	256/1024
Activation function	ReLU
Dimension of latent space	8

Table 4.1: Hyperparameters used in the cryoDRGN training block. The number of nodes in the hidden layers depends on the size of the images; downsampled images use 256 nodes in each layer, while full-size images use 1024. This decision is due to the fact that smaller images do not require as much network capacity as full-size images.

The iterative training of the cryoDRGN neural network does not start from zeros or a random initial condition, as this would entail very high computational costs. For this stage, an initial training session was performed, which was cut off before convergence. The weights from the last epoch of this training session were used as the initial condition for the iterative cycle. This logic is repeated in each iteration, starting from the weights from the last epoch of the previous iteration and only training a few epochs. The weights do not change substantially if the initial pose estimate is relatively good, so it is not necessary to train for many epochs.

4.1.3 CryoDRGN Analysis

The cryoDRGN analysis block was adapted to emphasize the low-dimensional representation of latent space vectors and clustering. Although cryoDRGN supports 3D UMAP projections, their use is not commonly reported in the literature. In this work, 3D UMAP is incorporated to better capture both local and global structures in the latent space. Compared to 2D UMAP, the additional degree of freedom in

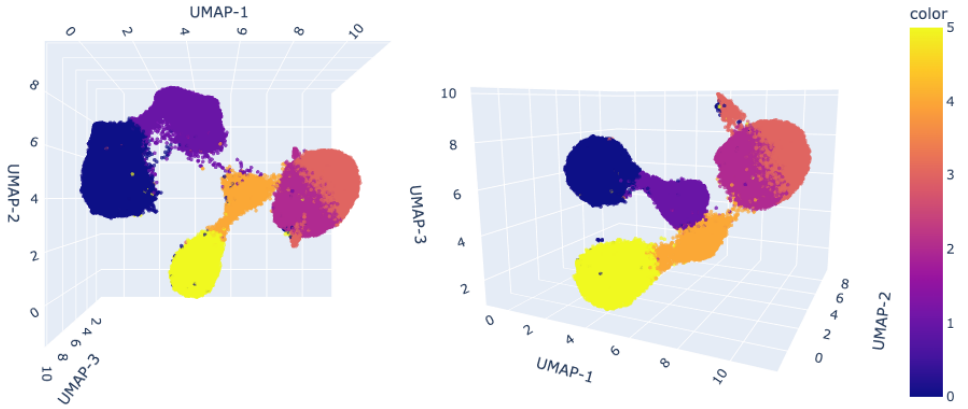


Figure 4.2: Different views from 3D UMAP plot of EMPIAR-10076. The colors correspond to labels that represent the clustering results of applying K-Means in 2D UMAPs.

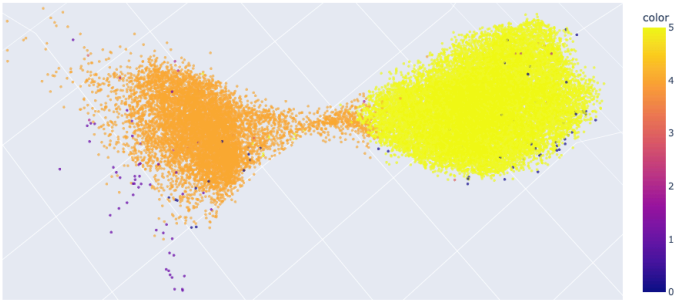
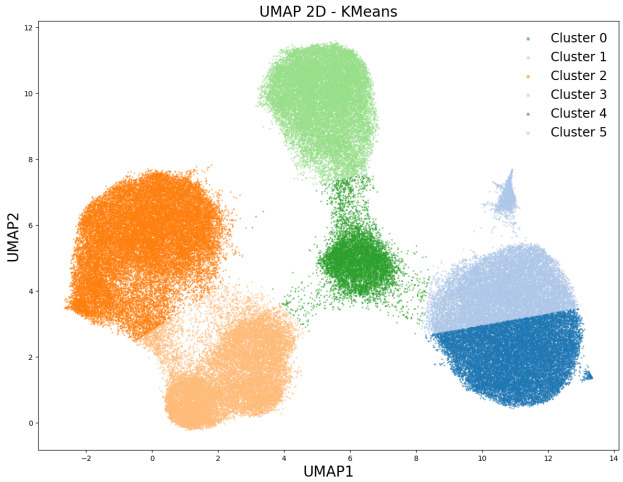


Figure 4.3: Figure 4.2 with zoomed clusters. It can be seen that some of the points were classified to diferent clusters acording to 2D UMAP, but expanding to 3D it looks like miss-clasification (blue points around both yellow and orange clusters).

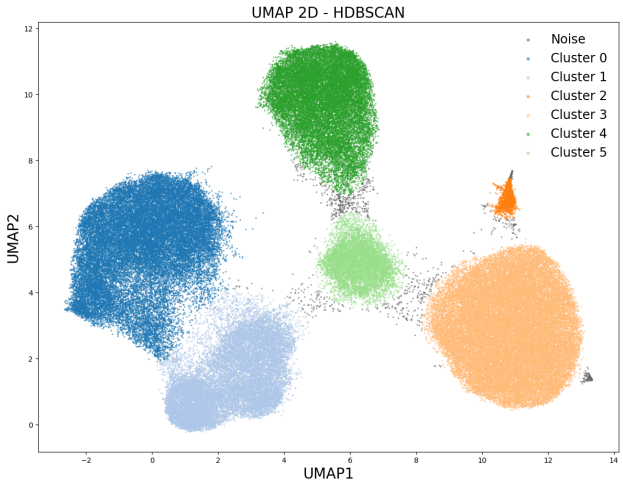
3D allows for a more faithful separation of points, reducing cases where distant vectors in the latent space are projected closely together or even overlap. Figures 4.2 and 4.3 illustrate these benefits. Figure 4.2 presents two different 3D visualizations colored by K-Means labels derived from the 2D UMAP, while Figure 4.3 zooms in on a region with potentially misclassified points.

Another key component of the analysis is the incorporation of clustering techniques. Previous work has shown that clustering techniques can be improved with dimensionality reduction algorithms. More specifically, K-Means has shown improvements in its accuracy with UMAP [2]. But for clusters with arbitrary shapes, K-Means does not work that well. For this reason, density-based algorithms were added to this analysis, more specifically DBSCAN [18] (Density-Based Spatial Clustering of Applications with Noise) and HDBSCAN [10] (Hierarchical DBSCAN).

A major advantage of these methods over K-Means is the fact that they do not necessarily associate each particle projection with a cluster; some projections can be considered noise. Figure 4.4 shows an example of 2D UMAP clustering with



(a) Clustering on K-Means with $K = 6$.



(b) Clustering on HDBSCAN.

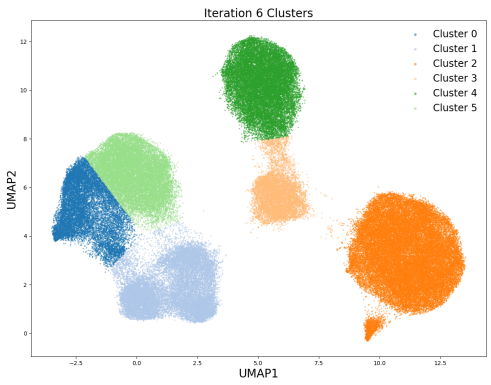
Figure 4.4: 2D UMAP representation of EMPIAR-10076 with different clustering techniques. Gray dots in Figure 4.4b represent particle projections that were not assigned to any cluster.

K-Means and HDBSCAN, where the gray dots were not assigned to any cluster.

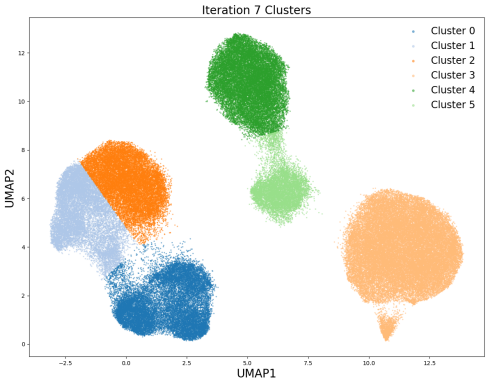
The fact that some projections are not assigned to any cluster implies that their poses are not refined. The effect this has on cryoDRGN's feature learning will be studied in the next chapter.

4.1.4 Clusters Processing

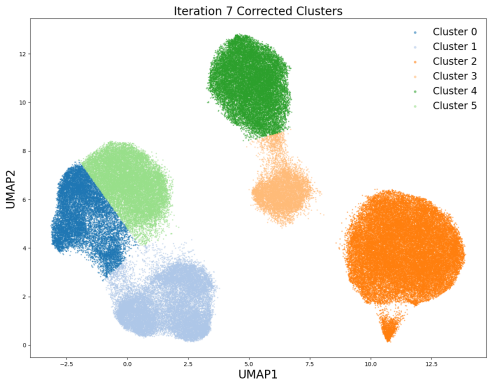
To achieve iterative refinement by re-estimating clusters on each iteration, it is essential to ensure that the clusters are consistent between iterations. Otherwise, a reference density map from another cluster will be used, resulting in particles being aligned with an unrepresentative volume.



(a) 2D UMAP with clusters from iteration 6.



(b) 2D UMAP with clusters from iteration 7, before the alignment.



(c) 2D UMAP with clusters from iteration 7, after the alignment.

Figure 4.5: Example of clustering results before 4.5b and after 4.5c labels correction using clusters assignment with the previous iteration 4.5a.

An example of this can be seen in Figure 4.5, where Figure 4.5a shows the 2D UMAP clustering results from a previous iteration, Figure 4.5b shows the 2D UMAP and calculated clusters on the current iteration, and Figure 4.5c shows the results after the cluster assignment between iterations. If consistency in clusters between iterations is not ensured, it will be difficult to obtain improvements in the pose estimation. This is why it is necessary to correctly assign the clusters to each other.

In order to do this, a distance matrix where the number of rows is defined by the number of clusters in iteration $i-1$ and the number of columns is defined by the number of clusters in iteration i is calculated. The optimal assignment between clusters is sought by minimizing the total sum of label differences, based on the Hungarian Algorithm [2]. Readers interested in the details of the algorithm can refer to the original publication or to standard tutorials that clearly illustrate its procedure.

In the case of clustering techniques where the number of clusters is not fixed (DBSCAN, HDBSCAN), the problem arises of how to handle this. If the number of clusters in iteration i-1 is greater than the number of clusters in i, there is no problem; some of the previous clusters are simply ignored. Otherwise, the clusters that can be assigned are assigned, and the clusters that do not have a counterpart in the previous iteration are aligned to the consensus volume.

In addition to cluster assignment, this block handles the preprocessing required to run Frealign. This involves splitting the dataset and generating Frealign parameter files (.par) by cluster.

4.1.5 Refinement and Reconstruction

Once the clustering and preprocessing are complete, pose refinement is performed for each cluster in Frealign [20]. To do this, each of the .par files generated in the previous block is iterated over. The main inputs are the particles in the corresponding cluster, along with their .par file and a reference density map. Both the .mrc file with the projections and the .par file with the parameters were generated in the cluster processing block. The reference volume is the density map corresponding to this cluster generated in the previous iteration. In this way, the projections of each cluster are aligned with a more representative volume. This processing pipeline can be seen in Figure 4.6, where the .mrc with projections and the .par with parameters from the particles are the ones generated in the cluster processing block. The other parameters are microscope parameters (e.g. beam energy, spherical aberration, amplitude contrast), macromolecule information (e.g. molecular mass, symmetry) and preferences of the algorithm (e.g. masking, frequency thresholds, normalization).

4.1.6 Poses Processing

Once the refinement is complete, there are as many .par files with the new poses as there are clusters. This block groups all the Frealign output files, recovering the previously saved indices, and generates a pickle file for cryoDRGN to read. This can be seen in Figure 4.7, where the number of clusters is K (therefore the number of .par files is also K). The .par files are unified and converted to a pickle file to be used as input by cryoDRGN.

For the conversion from a parameter file into a pickle file, it is necessary to extract the Euler angles and translations from the .par file, then convert the angles to a rotation matrix and change the translation units from Angstroms to pixels. In the case of methods that do not necessarily assign a cluster to each projection (DBSCAN and HDBSCAN), the final pose file will not match the number of projections used. To avoid this, the particles from the previous iteration are taken and inserted into the current pose file. Note that this means that the poses of those projections were not refined in this iteration.

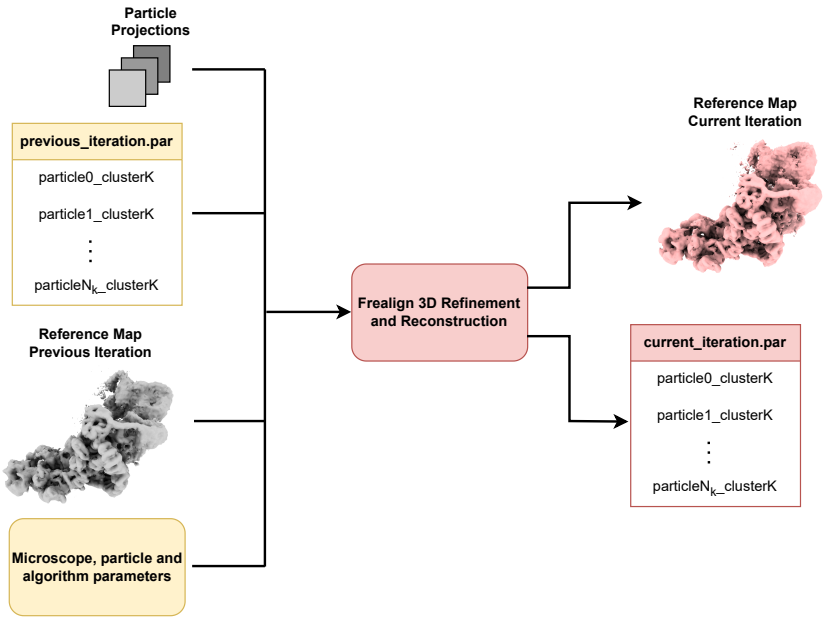


Figure 4.6: Workflow of the 3D refinement and reconstruction block for a single cluster. The inputs are the particle projections of a particular cluster (K for example), a parameter file of this cluster, a reference map and parameters associated with the microscope, the macromolecule, and Frealign's options.

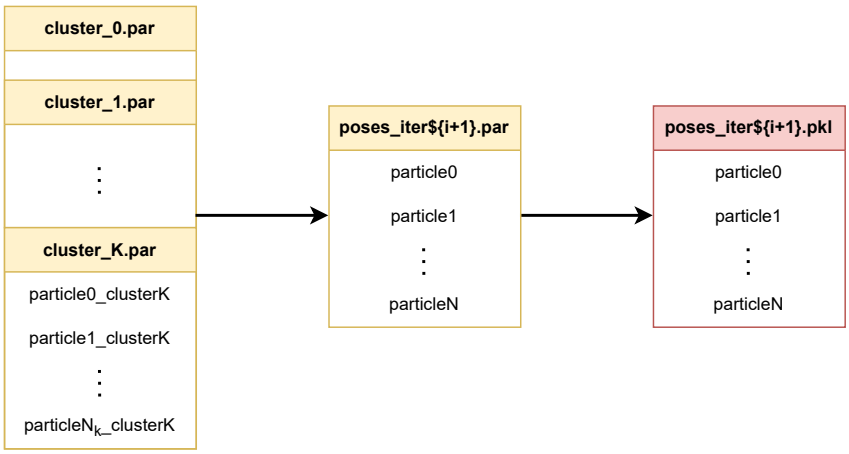


Figure 4.7: Workflow of the poses processing block. The parameter files from the refine block are unified in one file and converted into a pickle file with all N particles for cryoDRGN processing.

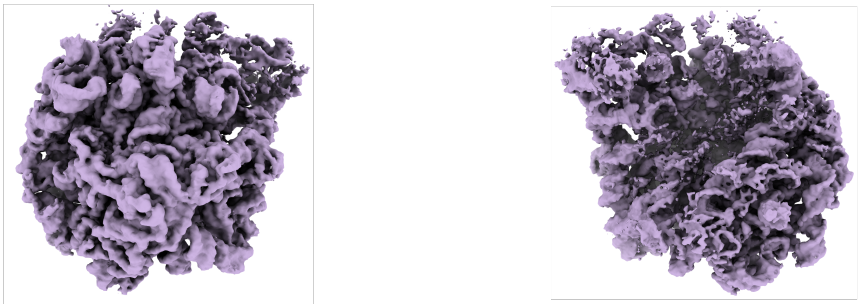


Figure 4.8: Consensus volume generated with the EMPIAR-10076 dataset from different views.

4.2 Datasets

4.2.1 EMPIAR-10076

One of the datasets used in this work is EMPIAR-10076, titled *CryoEM Dataset of L17-Depleted 50S Ribosomal Intermediates*, which contains cryo-EM images of ribosomal particles from *Plasmodium falciparum*, the parasite responsible for the most severe form of malaria [16]. Although this organism is eukaryotic, the dataset focuses on mitochondrial or apicoplast ribosomes, which are structurally similar to prokaryotic ribosomes. These consist of a small (30S) and a large (50S) subunit that together form the functional 70S ribosome.

The ribosome can be viewed as a modular molecular machine whose function depends on the correct assembly of its components. In this dataset, particular attention is given to intermediates of the 50S subunit lacking the L17 protein, which plays a role in the maturation and assembly process. The observed heterogeneity is not only due to continuous conformational changes, but also to discrete compositional variability arising from the stepwise assembly of the 50S subunit and its potential association with the 30S subunit. Understanding this variability is essential for characterizing different structural states and interpreting how assembly defects may affect ribosomal function. Figure 4.8 shows a 3D density map reconstruction of the consensus volume obtained from this dataset from two different views.

4.2.2 EMPIAR-10180

The second dataset, EMPIAR-10180, corresponds to the pre-catalytic spliceosome [44], a large and transient ribonucleoprotein complex involved in the removal of introns from precursor messenger RNA. The spliceosome undergoes a complex assembly and activation process involving large-scale rearrangements of its RNA and protein components. Capturing these intermediate states experimentally is extremely challenging due to their short lifetimes and structural diversity.

This dataset is particularly valuable for testing methods aimed at modeling structural heterogeneity, as it contains particles representing a continuum of conformations rather than a single, stable structure.

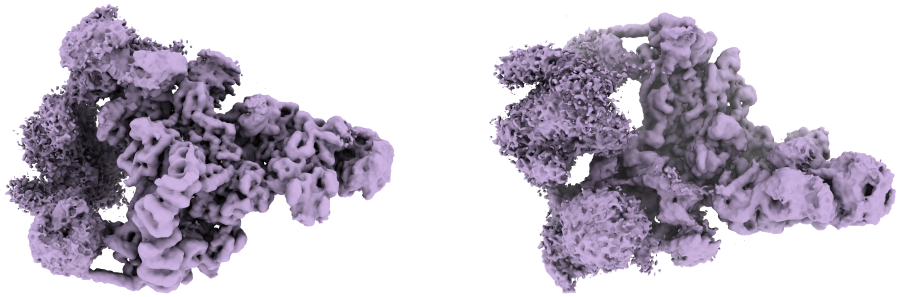


Figure 4.9: Consensus volume generated with the EMPIAR-10180 dataset from different views.

4.3 Complementary Tools

All density map visualizations in this document were generated using ChimeraX. ChimeraX [40] is a molecular visualization program developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI). The version used in this document is 1.6.1 and 1.10.1 for fitting atomic models in density maps.

Another tool used to a lesser extent was CryoSPARC [47]. CryoSPARC is a commercial software widely used by researchers for cryo-EM image analysis and processing. CryoSPARC was initially part of the pipeline, but due to its incompatibility with the proposed framework (it did not allow for a fully automatization), it was replaced by Frealign for density map refinement and reconstruction.

Also worth mentioning is IMOD [31], an open-source software package developed at the University of Colorado Boulder for the processing, analysis, and visualization of electron microscopy data. IMOD was useful for efficiently managing the metadata of large .mrc and .mrcs files, and data visualization using 3dmod.

Finally, much of the work was done in the nextPYP [37] container. nextPYP is an end-to-end cryo-EM and cryo-ET image processing platform. This platform integrates, among other things, the functionality of cryoDRGN and Frealign, so the dependencies of both codes are in the container.

4.4 Advantages and Limitations

As mentioned above, three of the main difficulties in macromolecule reconstruction are low SNR, pose calculation, and their structural heterogeneity.

In the low SNR regime, the processing pipeline does not include an explicit denoising step. Instead, reconstruction is performed directly from the noisy particle images, while clustering is carried out in the latent space of a VAE. VAEs are known to learn latent representations that can be relatively robust to noise, as the encoder-decoder architecture tends to focus on features that are consistent across samples and contribute meaningfully to reconstruction. Since random noise does not help minimize the reconstruction loss, it is often not represented in the latent space. This property has been leveraged in several applications for denoising and

dimensionality reduction, even though the model is not explicitly trained for denoising in this case. This property makes the identification of heterogeneity more robust to image noise.

Pose estimation is refined iteratively as distinct compositions or conformations are revealed within the dataset. This iterative process benefits reconstruction, since improved identification of heterogeneity leads to better alignment during refinement with Frealign, ultimately increasing pose accuracy.

The central goal of the proposed method is the identification of structural heterogeneity. Specifically, it is tailored to detect compositional heterogeneity through a discrete clustering strategy. Nonetheless, the method will also be applied to datasets exhibiting significant conformational variability, with the aim of identifying dominant conformations and evaluating whether this alone can lead to improvements in the resolution of the reconstructed volumes.

In addition to this, the pipeline offers several advantages. One of them is that, by updating the poses every certain number of epochs, it helps prevent the network from getting stuck in local minima, potentially improving cluster identification.

Another advantage is the ability to visualize how clusters evolve throughout the iterations. This enables better tuning of hyperparameters in clustering techniques such as K-Means or GMM, where the number of clusters is often chosen only once at the beginning of the process. This approach aligns with the idea behind Multibody Refinement [42], which involves selecting a number of clusters and classifying the data (either softly or rigidly). The pipeline thus provides support for determining this critical hyperparameter—a nontrivial benefit.

Additionally, by using Frealign instead of cryoDRGN's decoder, the pipeline produces higher-resolution structures, which is inherent to the reconstruction methods—cryoDRGN does not generate high-resolution density maps. Furthermore, cryoDRGN reconstructs density maps per particle, while Frealign generates one map per group of particle projections (in this case, one per cluster).

The cryoDRGN analysis function was modified to incorporate clustering techniques such as DBSCAN and HDBSCAN, and to be able to operate on three-dimensional UMAP representations.

Although the conformational changes of many macromolecules are, in principle, continuous, the use of discrete clustering methods to group particles into different states can be justified from both practical and theoretical perspectives. First, in complex biological systems, certain conformational states are often more stable and therefore more populated. This means that, within a continuous conformational space, there may exist local energy minima that act as "attractors" where the molecule tends to remain for longer periods. These dominant states can appear experimentally as frequent discrete modes, and discrete clustering can effectively capture them, enabling more accurate reconstruction of these representative conformations.

Moreover, even if transitions between conformations are smooth, current 3D reconstruction and pose estimation techniques face limitations in resolution and sensitivity to noise. In this context, discrete clustering can serve as a form of regularization, segmenting the latent space into relatively homogeneous regions

and thereby improving reconstruction quality by averaging similar particles.

Finally, discrete clustering can be useful as an initial exploratory step for analyzing structural diversity in a dataset. By dividing the particles into a limited number of groups, it provides a simplified representation of the conformational space, which can aid both in the interpretation and visualization of structural heterogeneity.

The code of this work is available at `https://github.com/DiegoSC13/master`

Chapter 5

Results and Discusion

This chapter provides a comprehensive analysis and comparison of the results obtained from the various experiments performed. It begins by studying the relevance of poses and showing how they affect the features of the reconstructions. It then discusses different clustering techniques, and finally presents the experimental setup.

5.1 Poses Relevance

As intuitive as it may seem, it is important to verify that particle poses play a significant role in the reconstruction of density maps and the resolution achieved. To assess this, an initial experiment was conducted in which the original poses were randomly altered, and cryoDRGN was trained using these modified orientations. This was accomplished by adding random uniform noise to the initial poses, varying the distribution to analyze its impact on the cryoDRGN's training and resulting volumes. The result can be seen in Figure 5.1, which illustrates the reconstruction of a single particle using cryoDRGN's decoder. In this case, the input translations were deliberately perturbed by adding random noise sampled from a uniform distribution, $\mathcal{U}(-n, n)$, with $n \in 3, 5, 10, 20$.The reason for using the cryoDRGN decoder instead of Frealign is to make the results easily comparable. Since they are different models, the representation of each particle can be different, and this can affect clustering. The cryoDRGN decoder allows for the reconstruction of each particle, so only one was selected to display the results.

It can be seen that particle features are lost as the additive noise increases. One way to assess whether features are being lost is by observing the visibility of elements such as alpha helices. The presence of well-defined alpha helices in a density map is an indicator of a high-quality reconstruction. Figure 5.2 shows a zoomed area of each particle reconstruction, where some alpha helices can be seen. The degradation of the helices is more evident in those on the right, where the characteristic helical shape progressively fades and the structures start looking more cylindrical. The reconstruction with noise sampled from $\mathcal{U}(-20, 20)$ completely lost these features.

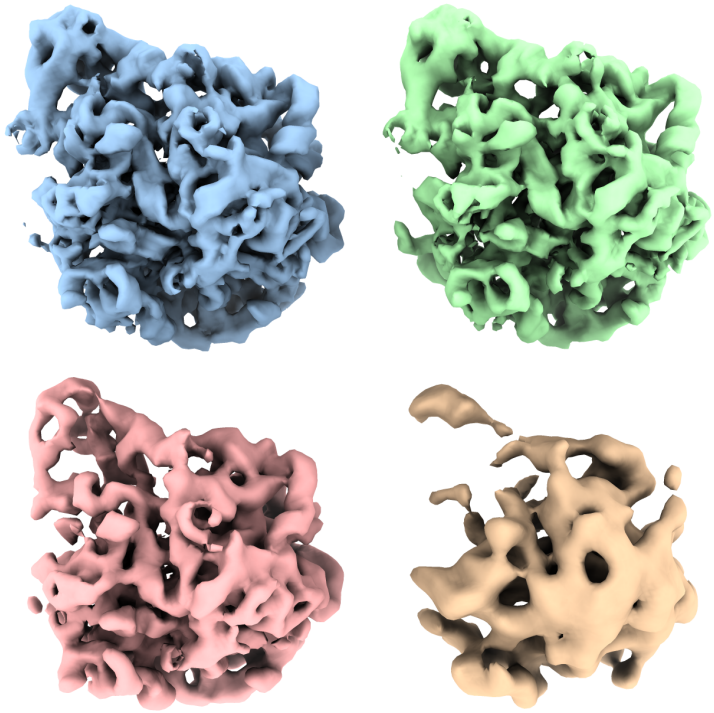


Figure 5.1: CryoDRGN reconstruction of a single particle with corrupted translations. The translations were corrupted by adding random noise sampled from an uniform distribution $\mathcal{U}(-n, n)$. The four maps show the results for $n \in \{3, 5, 10, 20\}$, arranged from left to right and top to bottom.

The same experiments were performed for rotations. In this case, uniform noise is added to the angles before being converted to the rotation matrix used by cryoDRGN. Instead of representing pixels, the noise in this case represents errors in degrees.

The results of adding uniform noise to the Euler angles can be seen in Figures 5.3 and 5.4.

The results of both experiments show that incorrect pose estimation leads to feature loss in the computed reconstructions. This indicates that pose estimation is indeed a sensitive parameter in reconstruction, and its correct estimation is key to achieving good resolutions.

5.2 Low-Dimensional Representation

A fundamental part of the proposed pipeline is the accurate identification of distinct structural states. Without this, it would not be possible to exploit the heterogeneity of the macromolecule to refine poses. To achieve this, various clustering techniques were employed in combination with different low-dimensional representations. The spaces explored include the original latent space, its representation in the first two principal components, and UMAP.

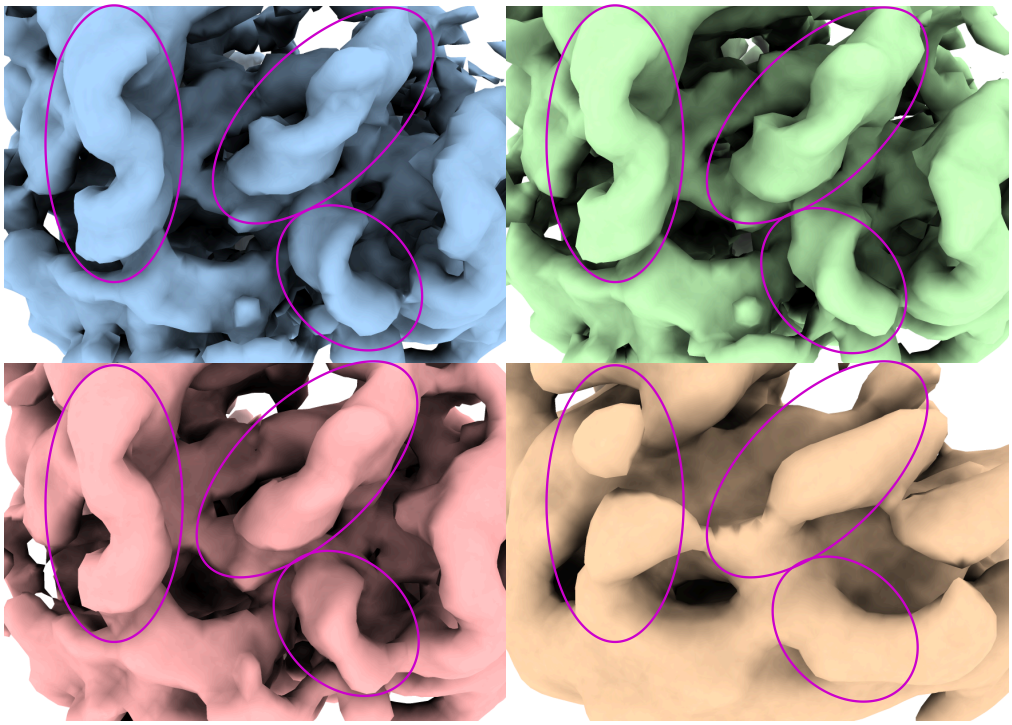


Figure 5.2: Zoomed area with some alpha helices for the reconstructions from Figure 5.1. The ellipses show the location of alpha helices in each of the reconstructions.

For this analysis, a cryoDRGN model was trained, and the different representations of the results were evaluated. Figure 5.5 shows the two-dimensional PCA and UMAP representation of the EMPIAR-10076 dataset, and Figure 5.6 shows the same representations of EMPIAR-10180.

Figures 5.5a and 5.6a show that it is very difficult to identify clusters in PCA. This seems to coincide with the variance represented in each of the principal components, which represents 39% and 40%, respectively. To further this analysis, the explained variance ratio was calculated and plotted for both datasets, which can be seen in Figure 5.7. Explained variance is a statistical measure of how much variation in a dataset can be attributed to each of the principal components generated with PCA. In other words, it tells us how much of the total variance is "explained" by each component.

Both plots show that variance is spread across many components, so much of the data's diversity is lost when using this technique for low-dimensional representation. Given this, it is very difficult for PCA to capture the structural heterogeneity of the proteins under study. Neither the explained variance ratios nor the difficulty in identifying clusters using PCA on the latent space is unexpected, and may in fact generalize to other VAEs under certain conditions. When the input data contain complex structures and the latent space is low-dimensional,

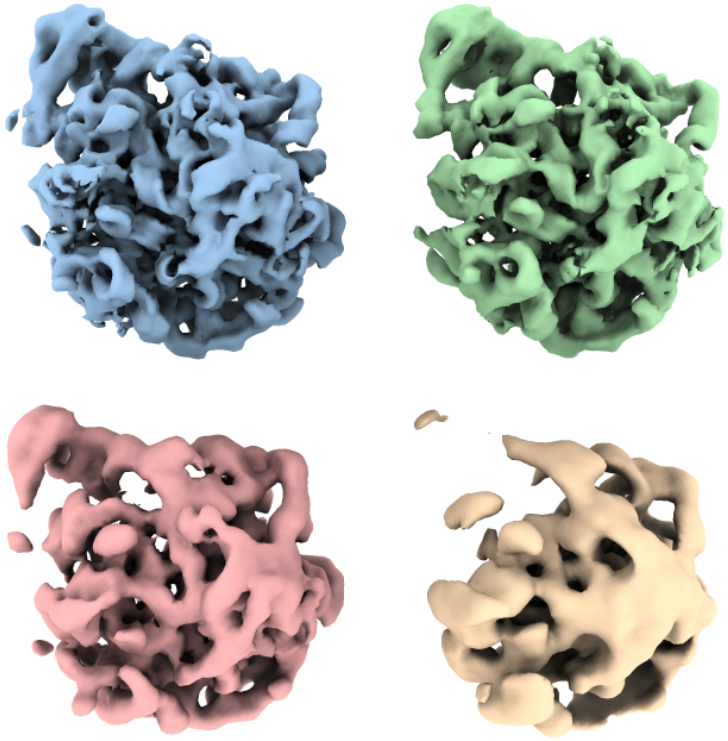


Figure 5.3: CryoDRGN reconstruction of a single particle with corrupted rotations. The rotations were corrupted by adding random noise sampled from an uniform distribution $\mathcal{U}(-n, n)$. The four maps show the results for $n \in \{3, 5, 10, 20\}$, arranged from left to right and top to bottom.

it is reasonable to assume that the model avoids allocating capacity to represent redundant information.

This is similar to results already discussed regarding VAEs being good denoisers; this is because noise does not contribute to minimizing the VAEs' loss function. Likewise, if two features are highly redundant (for example, a linear combination), then it makes no sense for the model to use two different dimensions to represent them. The goal of VAE (minimizing a reconstruction loss plus a KL regularization term, as can be seen on A.2) pushes it toward a compact, non-redundant representation.

This may also explain why no sharp drops are seen in the explained variance ratios in Figure 5.7. The latent space is a dense informative representation, with little redundancy between components, and therefore applying PCA does not reveal structures dominated by a few principal components.

In contrast, Figure 5.5b shows the presence of very well-defined clusters, and others that are not so well-defined. The existence of such discrete clusters demonstrates high compositional heterogeneity in the ribosome and is closely aligned with the working hypothesis of the proposed pipeline. This is due to UMAP's ability to model nonlinear relationships, which allows it to preserve both the global and local structure of the data.

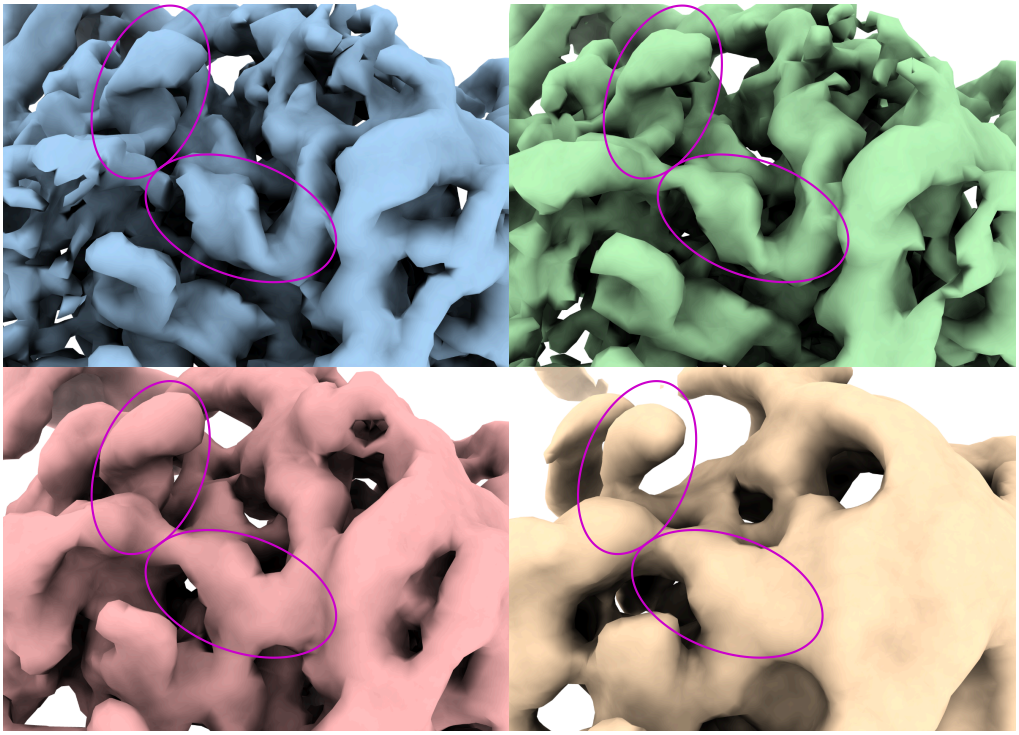
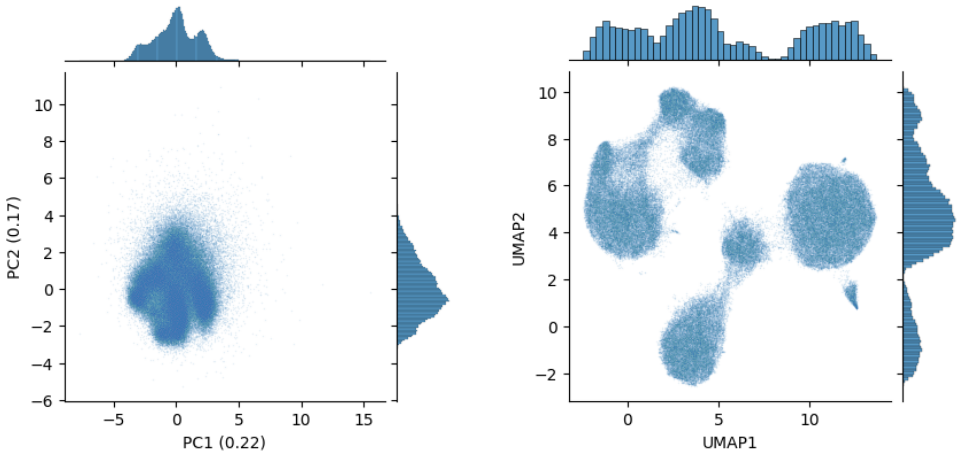


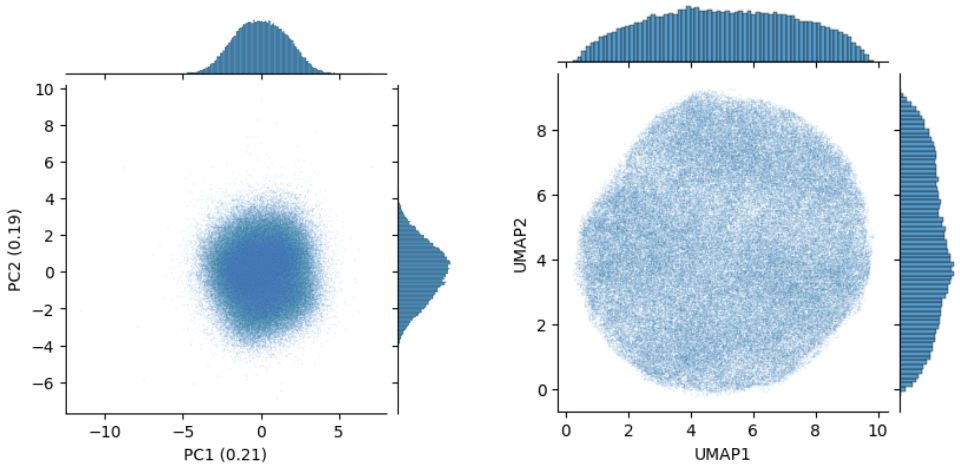
Figure 5.4: Zoomed area with some alpha helices for the reconstructions from Figure 5.3. The ellipses show the location of alpha helices in each of the reconstructions.



(a) PCA embedding of the latent space for EMPIAR-10076

(b) UMAP embedding of the latent space for EMPIAR-10076

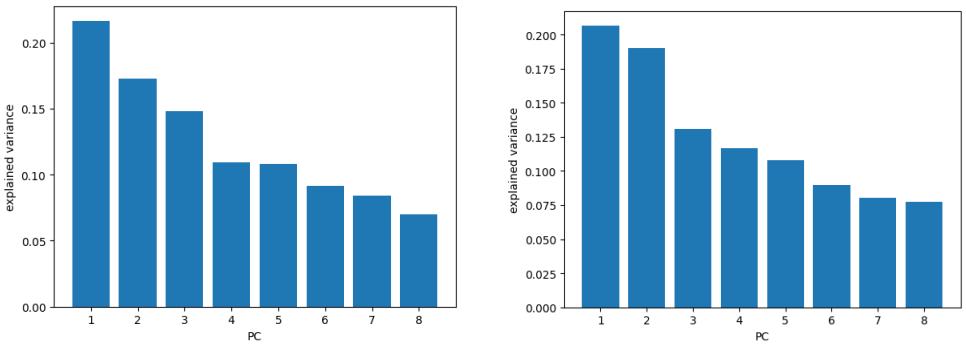
Figure 5.5: Two-dimensional representations of the latent space learned by cryoDRGN for the EMPIAR-10076 dataset, using PCA (a) and UMAP (b). These embeddings are used for clustering and identifying distinct structural states.



(a) PCA embedding of the latent space for EMPIAR-10180

(b) UMAP embedding of the latent space for EMPIAR-10180

Figure 5.6: Two-dimensional representations of the latent space learned by cryoDRGN for the EMPIAR-10180 dataset, using PCA (a) and UMAP (b). These visualizations support the identification of structural variability in the dataset.



(a) Explained variance ratio for EMPIAR-10076

(b) Explained variance ratio for EMPIAR-10180

Figure 5.7: Explained variance ratios for the principal components of EMPIAR-10076 (Figure 5.7a) and EMPIAR-10180 (Figure 5.7b). The plots show that the variance is spread across many components, with the first few accounting for less than half of the total variance. This supports the observation that PCA may not yield easily separable low-dimensional embeddings for clustering in these datasets.

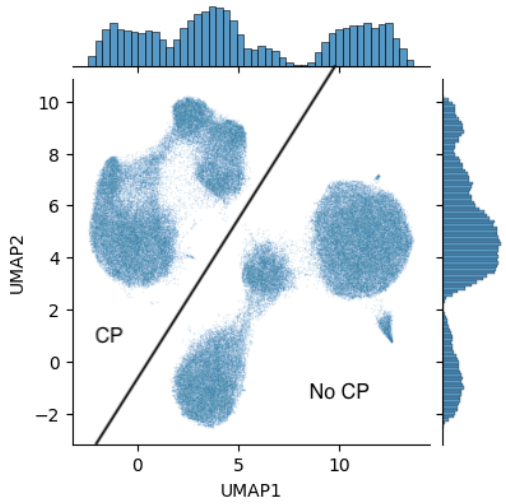


Figure 5.8: UMAP representation and classification of structures with and without the central protuberance (CP) subunit from EMPIAR-10076.

In fact, the subunit called central protuberance (CP) can easily be identified, and data can easily be classified between particles with this subunit and particles without it. This separation can be seen in Figure 5.8, where a linear boundary was drawn to differentiate the two classes. The clarity of this separation suggests that UMAP preserves relevant structural information present in the latent space, likely by exploiting nonlinear relationships among the latent variables that PCA fails to capture.

Unfortunately, Figure 5.6b shows that probably there will be problems with the performance of the pipeline in EMPIAR-10180. These results seem to break with the initial hypothesis that there could be predominant conformations, since there are no obvious clusters.

For these experiments, the particle filtering proposed by the cryoDRGN example repository was used, which retains 139722 of the 327490 particles. A cryoDRGN model was trained with all the particles, and some clusters were observed. Initially, running the pipeline with all particles instead of using the cryoDRGN indices was considered, but this was computationally very expensive; not in training, but in the pose refinement stage.

Despite this, it was decided to run the pipeline on the filtered EMPIAR-10180 dataset to experimentally evaluate its performance and analyze the results.

As a final analysis for low-dimensional representation, the points closest to the K-Means centroids with K = 5 in EMPIAR-10076 were reconstructed to view the volumes found, which in principle should be representative of different ribosomal compositions. The results can be seen in Figures 5.9, 5.10 and 5.11, which are the closest points to the centroids for K-Means with K=5 applied to the latent vectors, their first two components of PCA embeddings, and UMAP embeddings, respectively.

Both the original latent space and its UMAP representation achieved good

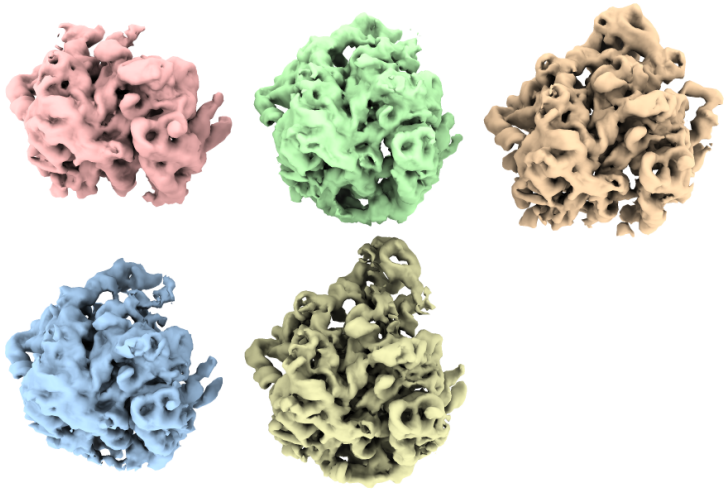


Figure 5.9: Reconstruction of nearest particles from K-Means centers with $K = 5$ applied to z. Each reconstruction should represent a different composition or conformation.

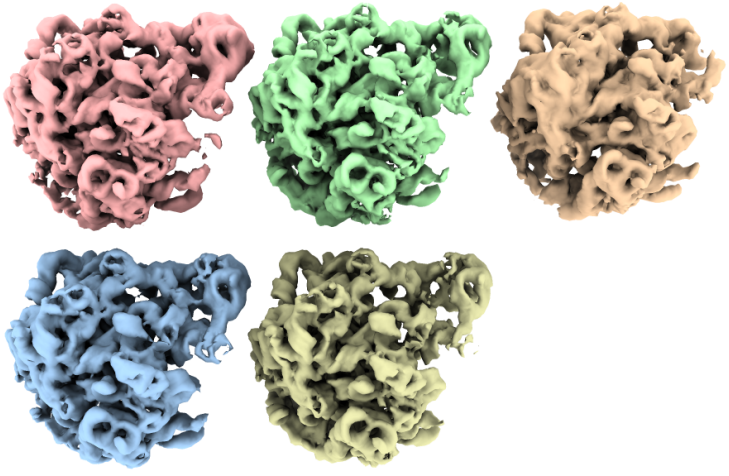


Figure 5.10: Reconstruction of nearest particles from K-Means centroids with $K = 5$ applied to PCA. Each reconstruction should represent a different composition or conformation.

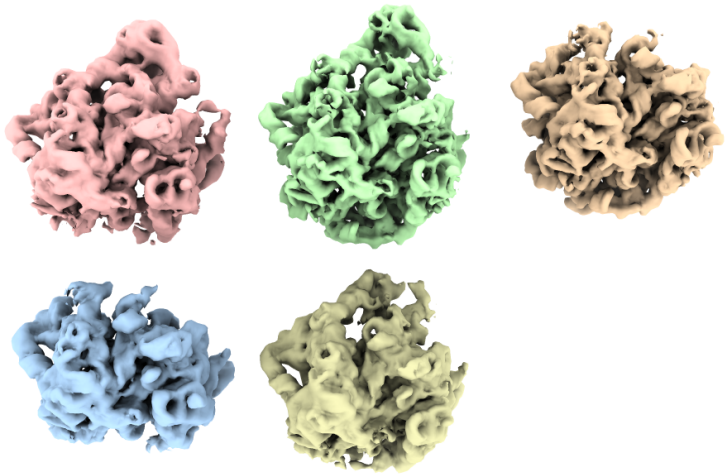


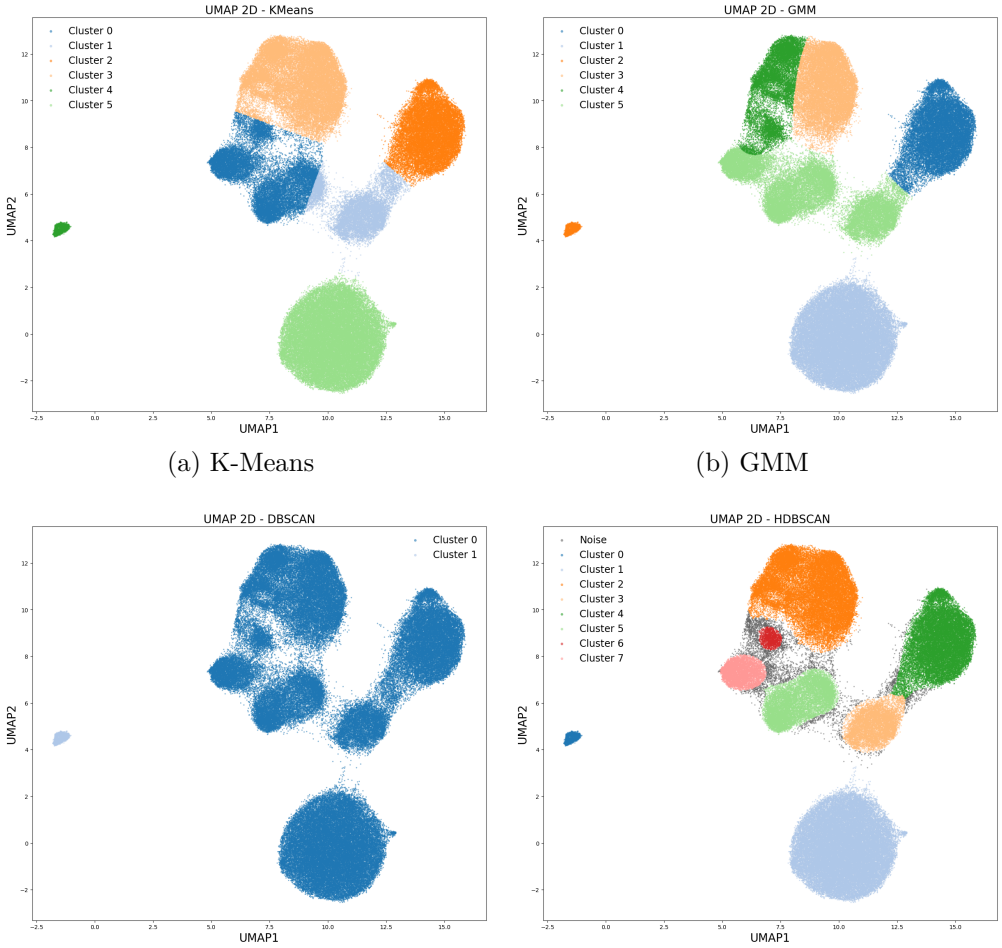
Figure 5.11: Reconstruction of nearest particles from K-Means centers with $K = 5$ applied to UMAP. Each reconstruction should represent a different composition or conformation.

results representing heterogeneity with K-Means. Given the potential challenges associated with the curse of dimensionality, as well as the need for more interpretable visualizations, clustering will be performed in the UMAP space. However, it is known that dimensionality reduction inevitably entails some degree of information loss, which may imply loss of subtle variations present in the original latent space. To mitigate this problem, three-dimensional UMAPs will also be used. This way, there is less loss of information, but the results remain interpretable.

5.3 Clustering

The proposed pipeline is capable of applying K-Means, GMM, DBSCAN, and HDBSCAN clustering techniques. It can even switch between methods across iterations, although this feature is not explored in the present work. Figure 5.12 illustrates the results of each technique applied to the 2D UMAP representation of EMPIAR-10076.

Considering the computational cost of each experiment, and the fact that not only the clustering technique but also the dataset, the input image size, and the UMAP embedding dimension vary between experiments, we decided to focus on only two methods for the reported results. The first method chosen was K-Means, due to its straightforward interpretation and its comparability with previous works [65]. The second method was HDBSCAN, selected because of its ability to identify dense clusters regardless of their geometric shape. In addition, HDBSCAN offers a natural way of handling minority clusters and varying densities, which stands in contrast with K-Means. Although DBSCAN shares some of these advantages, HDBSCAN extends the method by removing the need to predefine an absolute distance threshold and by providing a hierarchical structure of



(a) K-Means (b) GMM

(c) DBSCAN (d) HDBSCAN

Figure 5.12: Result of applying different clustering techniques to 2D UMAP representation of EMPIAR-10076

clusters from which the most stable ones can be extracted. This makes HDBSCAN more robust in high-dimensional embeddings such as those obtained from UMAP, where density can vary substantially across regions of the latent space. For these reasons, HDBSCAN is preferred over DBSCAN in this work, as it combines the flexibility of density-based clustering with improved stability and adaptability to heterogeneous datasets. This can be seen by comparing the results in Figures 5.12c and 5.12d, where the need to find another absolute distance threshold for DBSCAN is clearly visible. This hyperparameter can vary between iterations, making the method highly unstable.

5.4 Workflow

To run the pipeline on the datasets, it was first necessary to ensure that suitable initial conditions were available. Two key requirements are the initial pose estimations of the particles and the estimation of the CTF. Pose estimation requires at least one reference density map, which allows the particles to be aligned. This density map can either be generated with CryoSPARC, Frealign, or any cryo-EM reconstruction software, or downloaded from EMDB. CTF parameters, on the other hand, can be obtained using tools such as CTFFIND [41, 50], which is the standard choice in cryo-EM. A practical advantage of working with EMPIAR datasets is that they often already include these estimates. In this work, both the initial poses and CTF parameters were taken directly from EMPIAR. For EMPIAR-10076, however, the poses were further refined, since the consensus volume generated from the provided poses initially had very low resolution. These refined poses, together with the CTF estimates, serve as the input for the iterative block shown in Figure 4.1.

The other thing the pipeline needs outside of the iterative block is one or more volumes for the initial pose refinement. Remember that refinement is performed first, followed by reconstruction with the refined poses. Refinement requires both a pose file (in local refinement, which is the case) and a volume to align the poses. For this step, there are several alternatives.

One is to align all particles with the consensus volume generated for the initial pose estimation (if a pose file is not used or available). This option has the advantage that the initial volume likely has better overall resolution because it was generated with all the data. This is because, with such noisy images, more particles generally mean better SNR. In fact, this may be one of the reasons why DynaMight and 3DFlex start from a consensus volume even when they want to represent data heterogeneity.

On the other hand, if the goal is to emphasize heterogeneity, refining each cluster on the same density map doesn't seem like the best approach. This can lead to heterogeneity being poorly represented for some clusters and pose refinement being suboptimal, since part of the motivation for this work is that aligning under-represented particles to the consensus volume can lead to poor results. For this reason, instead of using a single volume, a clustering and density map reconstruction stage was performed for each cluster. The clusters and volumes obtained are those used for the first iteration, aligning the clusters as appropriate and using the logic outlined in Subsection 4.1.4.

The selected approach also has a disadvantage, which is essentially obtaining a structure with lower local resolution. It is important to understand that although the resolution measure with half FSC maps will be reported in this chapter, a significant part of the project is based on local resolution, which will be evaluated by visualizing regions not well represented in the consensus volume or early iterations of the pipeline, and seeing how these regions better fit the atomic model as iterations progress.

For the experiments reported below, several modifications were introduced to

Dataset	**Architecture and Image Size**	**Clustering Technique**	**UMAP**	**ID**
EMPIAR-10180	256x3 layers, D=128	K-Means (K=7)	2D	1
			3D	2
	1024x3 layers, D=320		2D	3
			3D	4
EMPIAR-10076	256x3 layers, D=128	K-Means (K=6)	2D	5
			3D	6
		K-Means (K=15)	2D	7
			3D	8
		HDBSCAN	2D	9
			3D	10
	1024x3 layers, D=320	K-Means (K=6)	2D	11
			3D	12
		HDBSCAN	2D	13
			3D	14

Table 5.1: Overview of the experimental configurations evaluated, varying the dataset (EMPIAR-10180 and EMPIAR-10076), the cryoDRGN architecture (number of neurons per layer and number of hidden layers in both encoder and decoder) and input image size (images have $D \times D$ pixels), the clustering method used (K-Means or HDBSCAN), and the dimensionality of the UMAP projection (2D or 3D).

the pipeline. First, the input images were downsampled to evaluate whether meaningful features can still be learned at a lower resolution; accordingly, the cryoDRGN architecture was also reduced, under the assumption that a smaller network may be sufficient to capture the relevant variability in smaller images. In addition, different clustering techniques were tested in order to compare the performance of approaches based on partitions versus those based on density estimation, thus assessing the robustness of the results to the choice of method. Finally, the dimensionality of the UMAP embeddings was varied to explore whether increasing the number of latent dimensions preserves additional structural variability that could be lost in a strict two-dimensional representation. The experiments performed can be seen in Table 5.1.

5.4.1 EMPIAR-10180

To appreciate the conformational heterogeneity of EMPIAR-10180, one of the first steps was to visualize the reconstructions of volumes sampled along the first principal component obtained with PCA. PCA was chosen in this context not as a superior alternative to UMAP for clustering, but simply as a straightforward way to explore the main axis of variability in the latent space. By reconstructing volumes along this principal direction, one can be confident that structural differences

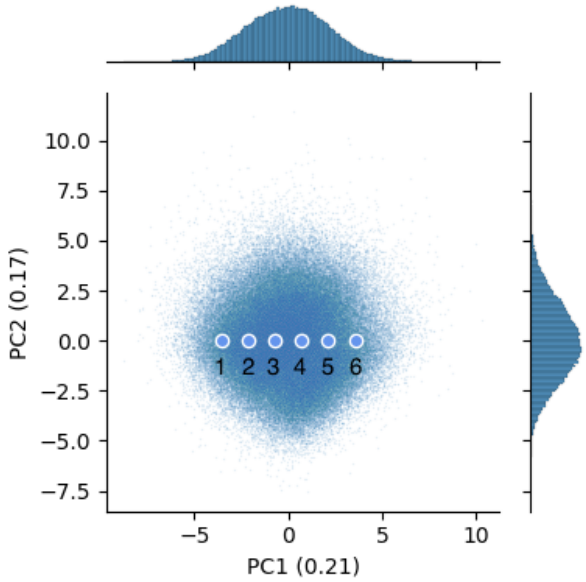


Figure 5.13: Conformational heterogeneity of EMPIAR-10180 represented along the first principal component (PC1) of the latent space obtained with PCA. The scatter plot shows the distribution of particles, with density indicated by the color intensity. The marginal histograms display the distribution of values for each component separately. To explore conformational variability, six points (1–6) were sampled along the first principal component, which captures the largest source of variance in the data.

present in the dataset are being captured, even if the method is limited to a linear projection. UMAP remains the preferred technique for nonlinear dimensionality reduction and clustering, while PCA is used here exclusively to provide an intuitive validation that the learned latent space encodes meaningful conformational variability. Figure 5.13 shows the PCA results in a hexbin plot along equidistant points taken along the principal axis, choosing the 5th percentile as the starting point and the 95th percentile as the ending point.

To visualize conformational changes, the nearest embeddings to those selected on the principal axis were reconstructed, using the cryoDRGN decoder to reconstruct the desired particles. The resulting density maps can be seen in Figure 5.14, where it is possible to observe, for example, how the reconstructions get curved. This result is consistent with the density maps in Figure 2.9.

Another way to visualize these results is by aligning the density maps, as can be seen in Figure 5.15a and 5.15b, where the first shows all the maps in Figure 5.14 aligned, and the second shows only the maps corresponding to the start and end points.

These figures provide a better comparison of the relative motion of the particle, where it can be seen that not the entire macromolecule is moving, but only the so-called SF3b subcomplex [65]. This rotation occurs continuously, which is consistent with the idea of conformational heterogeneity. Since the motion is continuous, if there are no predominant conformations, it is very difficult for UMAP to identify

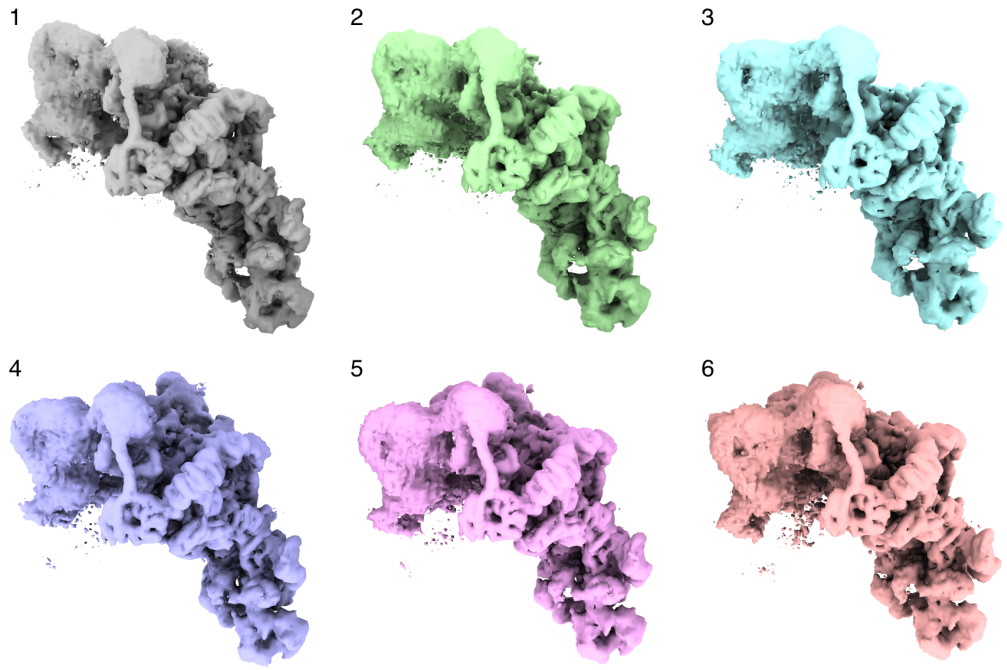
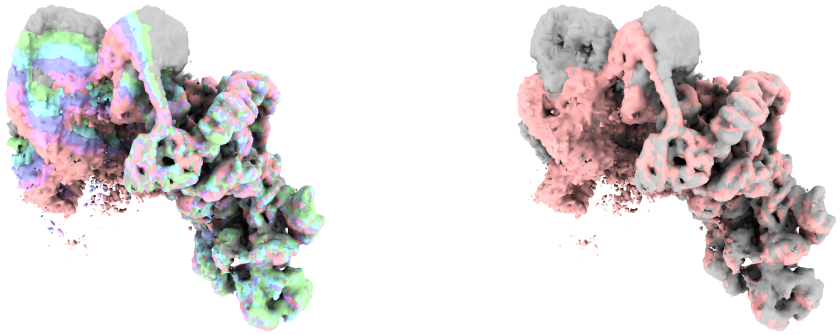


Figure 5.14: Reconstructed density maps of EMPIAR-10180 corresponding to the six points sampled along the first principal component in Figure 5.13. Each map is labeled (1–6) according to its position in the PCA plot, ordered from left to right along the principal axis. The gradual structural changes observed across these reconstructions (e.g., how maps become curved) illustrate the conformational heterogeneity encoded in the latent space.



(a) Reconstructions of the particles nearest to each point from Figure 5.13.

(b) Reconstruction of the first and the last point from Figure 5.13.

Figure 5.15: Superimposed density maps of EMPIAR-10180 along the first principal component, corresponding to the reconstructions shown in Figure 5.14. Figure 5.15a overlap of all six reconstructions, highlighting the gradual structural variability observed across the principal axis. Figure 5.15b is a direct comparison between the first and the last reconstruction, which makes the conformational change more apparent. The region that exhibits the most pronounced movement corresponds to a subunit called SF3b, providing visual evidence of conformational heterogeneity.

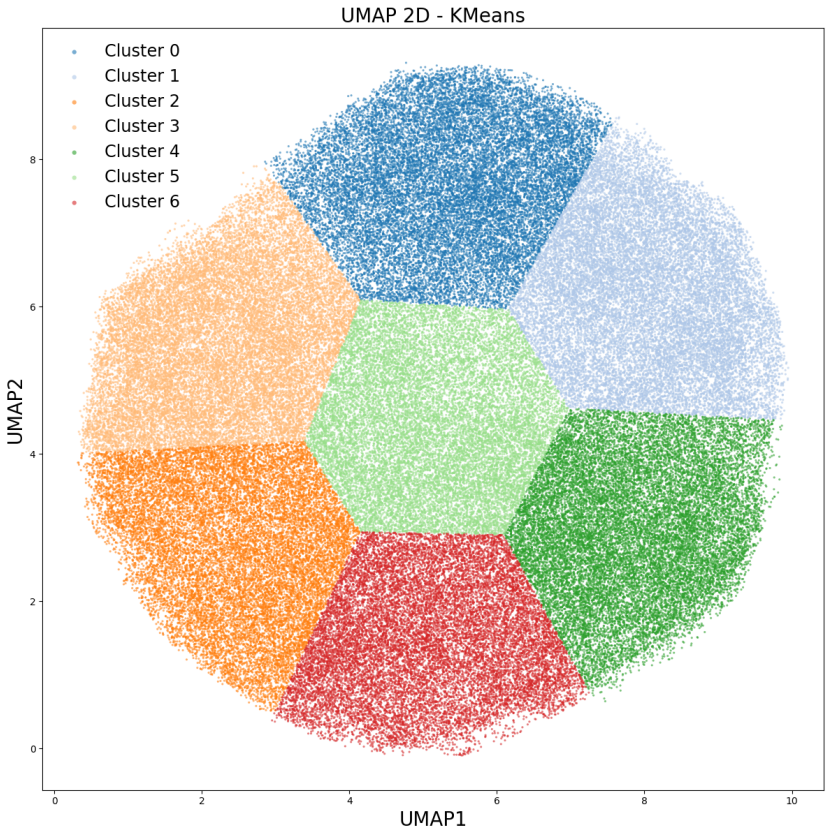


Figure 5.16: Clustering results from applying K-Means with $K = 7$ to 2D UMAP from EMPIAR-10180.

clusters. This is partly because there are none, and the proposed method will not perform well a priori, but it was decided to verify this experimentally anyway.

To evaluate the pipeline, both large and small network architectures were tested, and clustering was performed on 2D and 3D UMAP embeddings. HDBSCAN was not employed, as Figure 5.6 does not reveal well-separated, high-density clusters. Under these conditions, the algorithm would be unable to identify meaningful clusters. A value of K = 7 was selected based on the EMPIAR-10180 dataset and the density maps deposited in the EMDB from that study. However, this number should not be considered a strict reference, since a substantial fraction of particles was filtered out during their processing.

The results of applying K-Means clustering with K=7 to the 2D UMAP embedding obtained from the small architecture are shown in Figure 5.16.

These clusters do not appear to have any major significance, but rather appear to be forced limitations to obtain the requested groupings. This result is consistent with what was expected, since no well-defined groupings were initially visible.

It is important to note that for datasets dominated by continuous conformational changes, this approach is not expected to perform particularly well. Methods such as the original cryoDRGN pipeline or 3DFlex are specifically designed to ad-

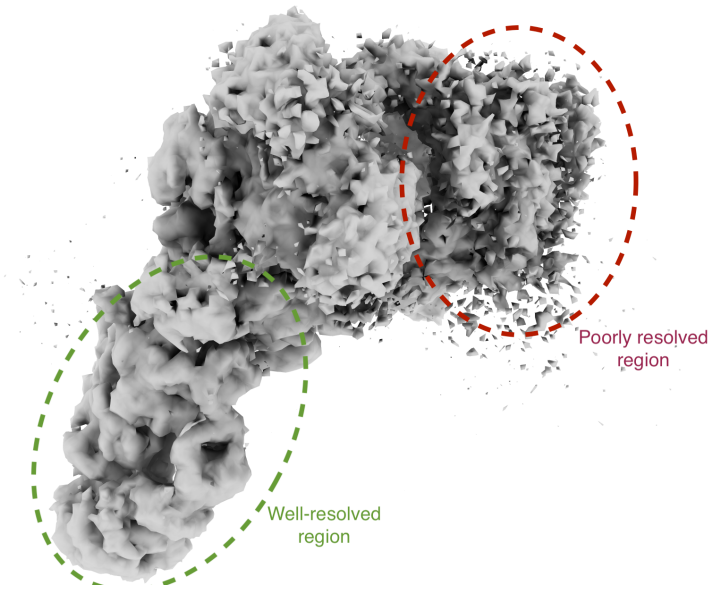


Figure 5.17: Example of reconstruction obtained after clustering. The SF3b subunit noticeably loses its local resolution.

dress this type of variability. At best, clustering on UMAP embeddings might capture discrete differences associated with the most prevalent conformations, or subtle variations within each group. However, in the case of EMPIAR-10180, UMAP does not produce well-defined clusters, and consequently, clustering algorithms applied to these embeddings are unlikely to reveal clear conformational heterogeneity.

When reconstructing the different clusters, it was observed that most were very similar. One of the reconstructions can be seen in Figure 5.17, where a significant loss of resolution can be seen in the subunit called SF3b subcomplex, while the rest of the particle appears to retain it. This is because different conformations are not being captured by clustering; rather, this conformational heterogeneity is still present in each cluster, causing averaging to return poor results for the mobile part.

Because most of the reconstructions are similar to the one in Figure 5.17, it can be concluded that the method fails to detect conformational heterogeneity, at least in cases where the transition between states is smooth and there are no predominant states that can be visualized as clusters in UMAP. This was expected, given the nature of both UMAP and the clustering used.

UMAP is a dimensionality reduction method that seeks to preserve the neighborhoods of the data, making it naturally useful for maintaining clusters in high-dimensional data. In the case of conformational heterogeneity, it does not seek to identify clusters, but rather trajectories, which is why analyses such as reconstructing density maps along the principal component of the data are more revealing. In fact, due to the nature of particle motions, it is possible that some particles have similar features, but others do not. This prevents UMAP from correctly

interpreting this heterogeneity.

On the other hand, the clustering in both cases is rigid. Techniques that allow the data to partially belong to more than one cluster were not used, which could be useful for showing transitions between clusters and relating them to conformational variations. For this purpose, it may be an option to study clustering techniques such as GMMs or their variants, where each particle has a probability of belonging to each cluster.

5.4.2 EMPIAR-10076

For K-Means clustering, two different K values were tested, based on the results obtained in the original work of this dataset [16]. In this work, the first 3D classification produced six structural groups (A–F) with average resolutions ranging from 3.7 Å to 7.9 Å.

- Group A: incomplete structure with several missing regions.
- Group F: a complete small subunit, expected from the sample preparation.

Groups A and F were not analyzed further. Groups B–E correspond to a larger subunit at different formation stages, mainly differing in the presence or absence of two key regions: the top (CP) and the base.

- B: missing both regions, the least complete.
- C: includes the base but still missing the top.
- D: includes the top but missing both the base.
- E: includes both the top and the base.

A second classification of C, D, and E produced twelve subgroups at 4–5 Å resolution, without revealing additional clear differences. Overall, this process described about 80% of the particles, ranging from very incomplete structures to ones almost identical to the final form. Figure 5.18 shows these results, along with the number of particles and resolution achieved in each class.

Given these results, $K = 6$ and $K = 15$ were selected for K-Means clustering.

The results obtained for the small architecture with downsampled images, K-Means with $K = 6$ and 2D UMAP can be seen in Table 5.2.

The results show notable improvements in all clusters. This is partly due to a low initial resolution, possibly due to poor clustering performance. Nevertheless, the improved resolution is consistent, and even better resolutions than those obtained in the original work can be seen.

In this case, we must also be careful with the performance of K-Means, as it can group the data differently between iterations, making them truly representative of the same composition. This can be seen in Figure 5.19, where clusters 2 and 3 are now considered a single cluster, and cluster 0 is now considered two distinct clusters.

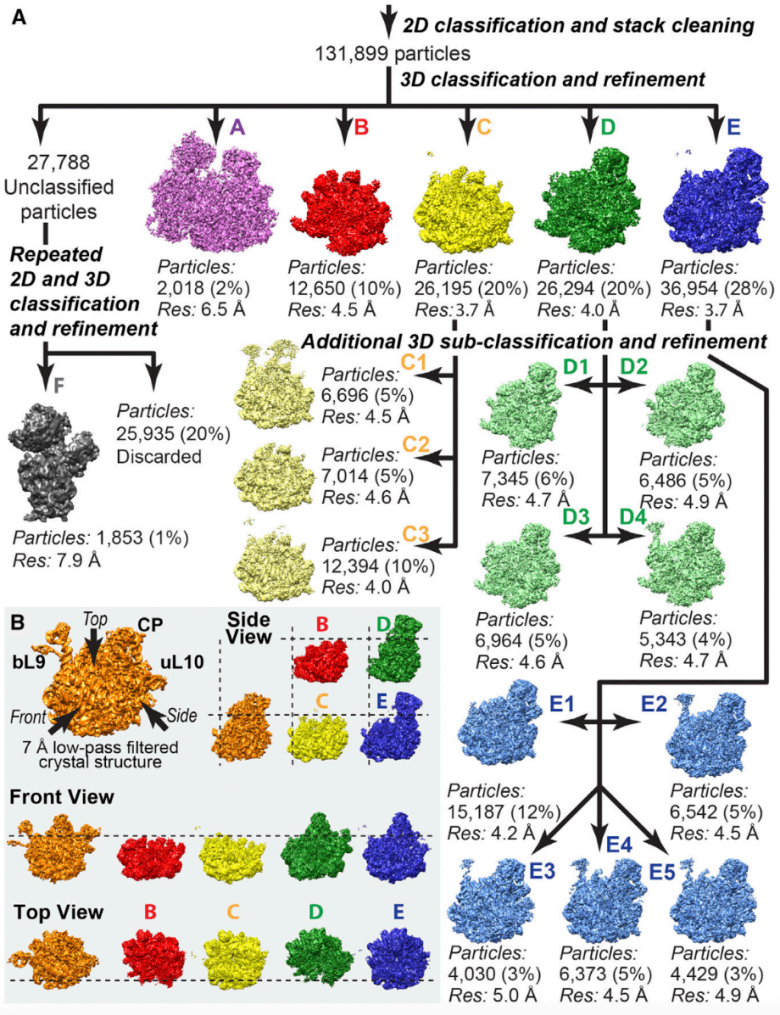
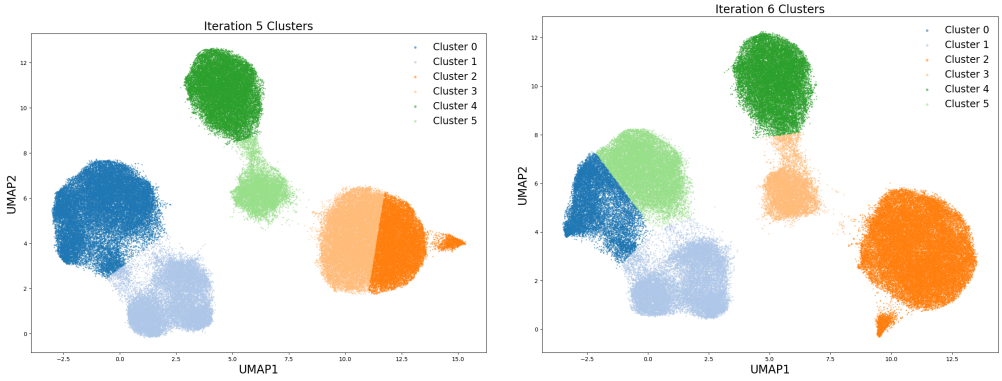


Figure 5.18: Hierarchical scheme of EMPIAR-10076 heterogeneity. (A) Density maps reconstructions of each class. Labels show the number of particles included at each refinement stage and the global resolution (Res) of the resulting maps. The super-classes (A–F) and the sub-classes (C1–E5) obtained from a second classification/refinement round are shown. (B) Side, front, and top views illustrate regions of EM density missing in each super-class compared to a native 50S subunit. The reference subunit is displayed as EM density derived from the mature ribosome LSU model (PDB: 4YBB). All structures are aligned for direct comparison, with dotted lines indicating key reference points. The figure was extracted from [16].

Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	6.62	3.63	3.47	3.57	3.48	**3.46**	3.74	3.8	3.92	3.65
cluster1	6.85	4.14	3.97	3.96	4.00	3.98	3.98	**3.95**	3.99	4.03
cluster2	7.22	7.46	6.69	6.09	5.87	5.00	**3.99**	4.83	8.24	9.90
cluster3	7.23	7.66	6.33	5.90	5.79	**5.08**	5.61	5.45	5.38	5.50
cluster4	6.82	3.96	3.78	3.84	3.83	3.83	3.81	**3.74**	3.77	3.85
cluster5	7.49	5.49	5.09	5.04	5.14	5.11	4.09	3.78	**3.75**	6.82

Table 5.2: FSC values for EMPIAR-10076 with K-Means, 128x128 images and 2D UMAP. The numbers in bold represent the best resolution for a given cluster.

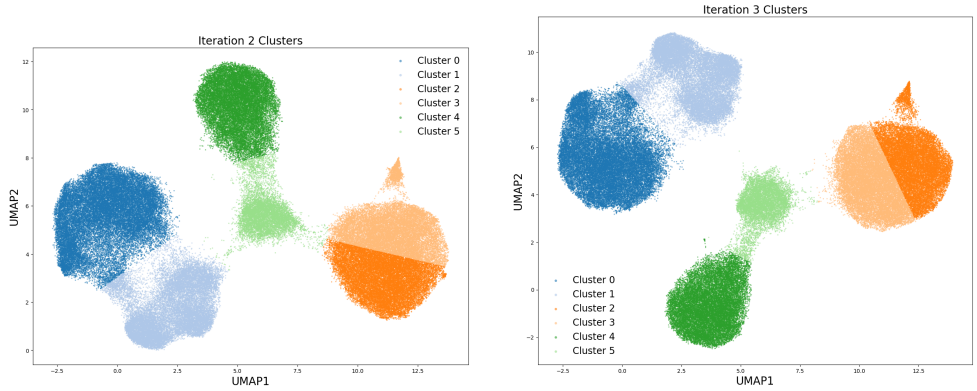


(a) Iteration 5 (b) Iteration 6

Figure 5.19: UMAP plots with the clustering results of K-Means in different iterations. Figure 5.19a represents the clustering at Iteration 5, while Figure 5.19b shows the clustering at Iteration 6. The figure demonstrates how K-Means can adjust the number and boundaries of clusters between iterations, sometimes merging clusters that were previously separate or vice versa.

This case is not corrected by `align_clusters.py` because it is not really an alignment problem between clusters, but rather a performance issue of the clustering technique by taking random centroids in each iteration. This alters the performance of the algorithm because the affected clusters will be aligned with density maps that correspond to them, in this case clusters 3 and 5. One way to solve this is by fixing the initial centroids. This can be counterproductive as a standard approach, because you are biasing the algorithm based on the clusters obtained in a certain iteration. Furthermore, although UMAP maintains global and local features, phenomena like those in Figure 5.20 can occur. These are due to the fact that UMAP does not preserve the orientation or scale of the original space, which implies that the embeddings can be rotated, mirrored, or even scaled even if the data are very similar to each other. What is preserved in UMAP are actually the k nearest neighbors and, consequently, the closeness between points.

Note that in this case, `align_clusters.py` is able to correctly align the clus-



(a) Iteration 2 (b) Iteration 3

Figure 5.20: UMAP representation of the data in iterations 2 5.20a and 3 5.20b, illustrating how projections of very similar data can appear mirrored, while still preserving both the global and local structure of the clusters.

Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	7.93	7.38	6.54	6.26	6.07	6.07	**5.89**	6.36	7.81	8.68
cluster1	7.88	5.86	5.07	**5.05**	5.09	5.09	5.07	5.24	5.34	5.47
cluster2	3.92	3.78	**3.72**	3.81	3.82	3.83	3.77	3.73	3.79	3.79
cluster3	5.31	3.91	**3.67**	3.74	3.74	3.69	3.71	3.73	3.80	3.79
cluster4	4.15	3.95	**3.89**	3.94	3.94	3.96	3.94	3.94	3.97	4.02
cluster5	**3.64**	3.87	3.86	4.04	4.00	3.98	3.94	3.95	3.99	4.03

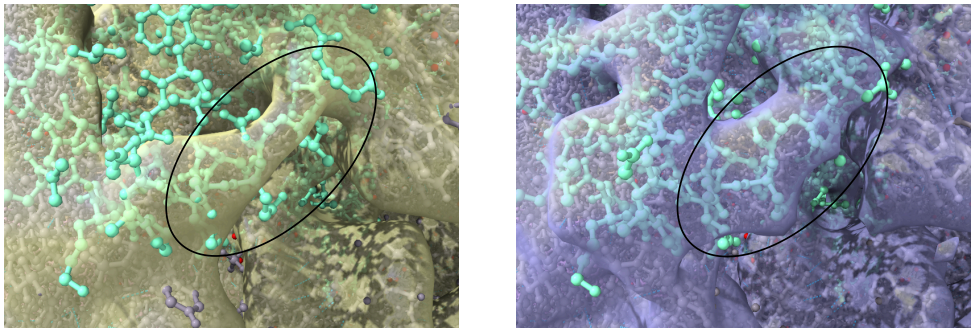
Table 5.3: FSC values for EMPIAR-10076 with K-Means, 128x128 images and 3D UMAP. The numbers in bold represent the best resolution for a given cluster.

ters. This is because the general structure of the data is maintained, and the performance of the clustering algorithm is similar in both cases.

Similar to the K-Means problem, in UMAP, one could also start from a non-random initial arrangement (i.e., the results of a previous iteration). In fact, by setting the initial conditions for K-Means and UMAP, more stable results could be achieved. This approach seems reasonable after a certain number of iterations, where the pipeline does not show large variations in the results obtained between iterations. Although it seems like a promising idea, it will not be explored in this work.

The results for the three-component UMAP case can be seen in Table 5.3. In this case, the improved resolutions occur at earlier iterations. This may be partly due to the fact that the initial iterations in this case are much better than those in the two-component UMAP case, which may be due precisely to the additional degree of freedom in representing the data.

Another way to analyze the obtained density maps, in addition to calculating



(a) Region of initial density map (b) Region of iterative refined density map

Figure 5.21: Comparison between the initial density map (5.21a) and the refined density map after iterative processing (5.21b), overlaid with the corresponding atomic model. The black ellipse highlights a region where the refined map more closely follows the atomic structure: the density better encloses the backbone and side chains of the model, aligning with bonds and atoms. Such correspondence between density and the atomic model is an indicator of improved local resolution.

the FSC and looking for features such as well-defined alpha helices, is to compare both results with the atomic model.

To do this, the atomic model is fitted to the obtained density maps, in the same way as shown in Figure 2.5c, and their similarity to each other is studied.

Remembering that an atomic model shows the location of each atom in the protein and its peptide bonds, the goal is for the model to accurately reflect these geometries. When an atomic model is aligned with a density map, each atom or group of atoms should correspond to a density peak in the correct location. In high-resolution maps, these peaks are not diffuse, but rather well-defined and centered on the atom's position. Figure 5.21 illustrates this principle. In the initial map (5.21a), the density around the highlighted region is diffuse and does not fully capture the features of the atomic model. After iterative refinement (5.21b), the same region shows a much better match. The density follows the contour of the polypeptide more closely, with bridges and bulges that align with the positions of bonds and side chains. This improvement demonstrates the effect of refinement in enhancing the quality of the reconstruction, providing maps that are more consistent with the atomic model and therefore more useful for structural interpretation.

Tables 5.2 and 5.3 show that the last iterations can become unstable. This is due to a Frealign parameter that defines a resolution threshold. This parameter acts as a low-pass filter when aligning the images, filtering out high frequencies to avoid overfitting with the noise. There is a trade-off in this component because by not using high-frequency information, image details may be lost when aligning, and a very low threshold can make it difficult to obtain high-resolution reconstructions. The decision was to lower this threshold depending on the iteration number, using the values in Table 5.4.

Iteration	EMPIAR-10076	EMPIAR-10180
0	6	6
1	6	6
2	5	5
3	4	4
4	3	4
5	3	3.5
6	3	3.5
7	2.7	3.5
8	2.7	3
9	2	–

Table 5.4: High-resolution limit values by iteration in Angstroms, for EMPIAR-10076 and EMPIAR-10180.

It is important to note that this parameter does not directly apply a filter to the images, nor does it cause information loss. Instead, it limits the information used for image alignment to frequencies below the specified threshold. High-resolution reconstructions can still be obtained even with a relatively high value for this threshold, if the alignment is accurate enough using only low-frequency information. The strategy of iteratively decreasing the resolution limit is standard practice in cryo-EM, as it helps the alignment algorithm by eliminating high-frequency noise components. In fact, it is common in global refinement approaches to start with a resolution limit of 10Å or even higher [32].

It is also important to ensure that this threshold does not exceed the Nyquist frequency, which is defined as:

$$f_{Nyquist} = \frac{1}{2 \times pixel\ size}. \tag{5.1}$$

If the frequency threshold is higher, there will be no information and therefore the only thing added to the alignment is noise. The final iteration intentionally exceeds $f_{Nyquist}$ to show how this affects the FSC.

The pipeline results using HDBSCAN for two- and three-component UMAP can be seen in Tables 5.5 and 5.6, respectively.

In both cases, it can be observed that there are incomplete rows. This is because HDBSCAN does not have a fixed number of clusters, so they do not necessarily match between iterations. An example of this can be seen in Figure 5.22, where an additional cluster was identified in iteration 3. In fact, in both experiments, the incomplete row corresponds to that specific cluster.

One limitation of the method is that if no reconstruction of that cluster is found in the previous iteration, the particles are aligned with the consensus volume. In this case, this can be particularly detrimental, as it is a very small cluster, which might not be well represented by the consensus volume due to the weight of all the other particles in the reconstruction. This, coupled with the few particles used for refinement, results in a very low resolution for that class.

Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	3.94	3.78	**3.72**	3.80	3.81	3.77	3.77	3.74	3.77	3.83
cluster1	5.35	5.20	**5.00**	5.01	5.09	5.09	5.03	5.15	5.36	5.37
cluster2	4.16	3.95	**3.91**	3.94	3.96	3.94	3.93	3.93	4.00	4.03
cluster3	3.66	3.52	3.47	3.54	3.49	3.43	3.44	**3.37**	3.43	3.45
cluster4	10.18	**9.59**								
cluster5	7.75	7.46	7.15	6.09	5.72	4.84	**4.79**	7.21	7.82	8.84
cluster6				9.85				10.52		10.75

Table 5.5: FSC values for EMPIAR-10076 with HDBSCAN, 128x128 images and 2D UMAP. The numbers in bold represent the best resolution for a given cluster. Incomplete rows correspond to clusters that were not identified in all iterations.

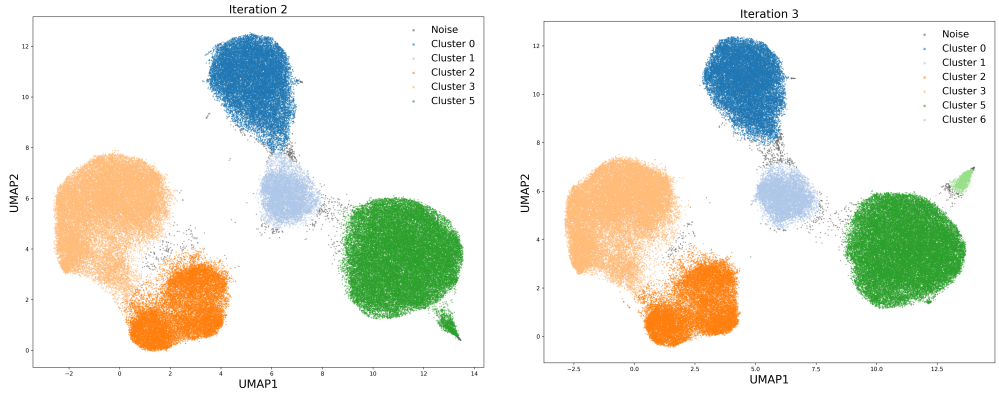
Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	3.65	3.51	3.48	3.54	3.46	3.43	3.41	**3.34**	**3.34**	4.65
cluster1	7.74	7.42	6.70	6.16	6.00	5.89	**5.77**	6.57	7.85	9.27
cluster2	4.14	3.95	**3.91**	3.94	3.95	3.95	3.96	4.78	4.90	9.53
cluster3	5.33	5.21	5.04	**4.99**	5.09	5.04	10.37	14.37	16.73	10.56
cluster4	3.94	3.76	**3.72**	3.81	3.83	3.76	3.70	3.68	3.74	3.88
cluster6		10.05			10.59					

Table 5.6: FSC values for EMPIAR-10076 with HDBSCAN, 128x128 images and 3D UMAP. The numbers in bold represent the best resolution for a given cluster. Incomplete rows correspond to clusters that were not identified in all iterations.

Regarding the overall workflow, it is worth asking whether updating the clusters actually has any effect. An important experiment to validate the workflow is to use only refinement and reconstruction, without updating the clusters in the iterative block. This would involve modifying the block diagram to function like the one in Figure 5.23, where the training of cryoDRGN, the analysis of the results obtained, and the processing of the clusters with the various techniques mentioned above are all performed once. In this case, the `.par` file generated by the refinement function becomes the reference parameter file for the next iteration, and reconstructions are generated at each iteration to be used as the reference volume for the next iteration.

These experiments were performed using the high-resolution limits in Table 5.4 so that the results are comparable. The FSC results for each cluster and iteration can be seen in Tables 5.7, 5.8, 5.9 and 5.10.

The results appear promising; however, an unexpected observation emerges. The resolution continues to improve even when operating below the Nyquist frequency, a behavior that should not occur. After inspecting the volumes, it was found that they were identical beyond a certain point, and that what was aligning



(a) Iteration 2: This plot shows the clustering results from the second iteration of HDBSCAN, where six clusters are identified.

(b) Iteration 3: In the third iteration, HDBSCAN adjusts the clustering, detecting seven clusters.

Figure 5.22: UMAP plots showing the clustering results of HDBSCAN at different iterations. Figure 5.22a represents the clustering at Iteration 2, while Figure 5.22b shows the clustering at Iteration 3. The figures show that the number of clusters detected by HDBSCAN can change between iterations.

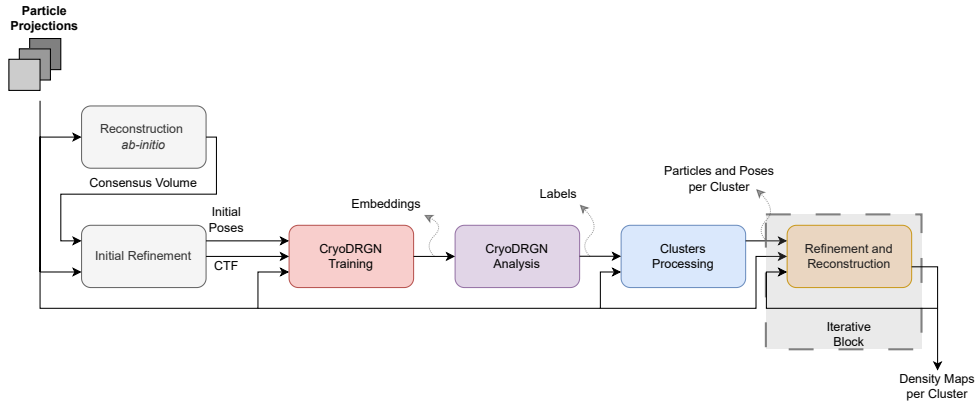


Figure 5.23: Workflow without iterative clustering.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	3.67	3.53	3.58	3.48	3.41	3.38	3.35	3.34	**3.28**
cluster1	4.21	4.02	4.00	3.96	3.91	3.87	3.84	3.82	**3.75**
cluster2	7.53	6.57	5.96	5.80	5.73	**5.69**	**5.69**	**5.69**	7.52
cluster3	7.66	6.53	6.12	6.04	5.99	**5.96**	5.98	5.98	7.67
cluster4	4.00	3.85	3.87	3.82	3.74	3.69	3.63	3.60	**3.53**
cluster5	5.70	5.19	5.13	5.21	5.18	5.13	5.05	**4.99**	5.09

Table 5.7: FSC values for EMPIAR-10076 with K-Means (K=6), 128x128 images and 2D UMAP iterating only the refinement. The numbers in bold represent the best resolution for a given cluster.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	7.63	7.40	7.26	7.24	7.21	**7.20**	7.23	7.24	7.79
cluster1	7.61	7.37	7.16	7.06	5.93	5.90	5.90	**5.89**	7.59
cluster2	3.81	3.77	3.84	3.78	3.70	3.66	3.62	3.60	**3.54**
cluster3	5.19	5.04	5.03	5.08	5.07	5.05	5.01	**4.99**	5.05
cluster4	3.99	3.95	3.96	3.93	3.88	3.86	3.83	3.81	**3.76**
cluster5	3.56	3.52	3.56	3.46	3.39	3.37	3.35	3.34	**3.28**

Table 5.8: FSC values for EMPIAR-10076 with K-Means (K=6), 128x128 images and 3D UMAP iterating only the refinement. The numbers in bold represent the best resolution for a given cluster.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	3.83	3.79	3.84	3.79	3.72	3.68	3.64	3.62	**3.54**
cluster1	5.21	5.03	5.00	5.04	5.03	5.02	4.99	**4.97**	5.05
cluster2	4.00	3.95	3.96	3.92	3.88	3.85	3.83	3.81	**3.79**
cluster3	3.58	3.52	3.59	3.48	3.39	3.37	3.35	3.34	**3.28**
cluster4	9.86	9.77	9.65	**9.62**	9.65	9.70	9.98	10.18	14.85
cluster5	7.43	7.11	5.98	5.75	4.94	4.89	4.86	**4.84**	6.97

Table 5.9: FSC values for EMPIAR-10076 with HDBSCAN, 128x128 images and 2D UMAP iterating only the refinement. The numbers in bold represent the best resolution for a given cluster.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8
cluster0	3.57	3.53	3.58	3.48	3.41	3.38	3.36	**3.35**
cluster1	7.43	7.06	6.15	6.01	5.93	5.88	5.90	**5.88**
cluster2	4.01	3.97	3.96	3.94	3.90	3.88	3.85	**3.84**
cluster3	5.21	5.05	5.02	5.06	5.04	5.03	5.00	**4.98**
cluster4	3.82	3.78	3.82	3.81	3.72	3.68	3.62	**3.59**

Table 5.10: FSC values for EMPIAR-10076 with HDBSCAN, 128x128 images and 3D UMAP iterating only the refinement. The numbers in bold represent the best resolution for a given cluster.

was particle noise due to overfitting.

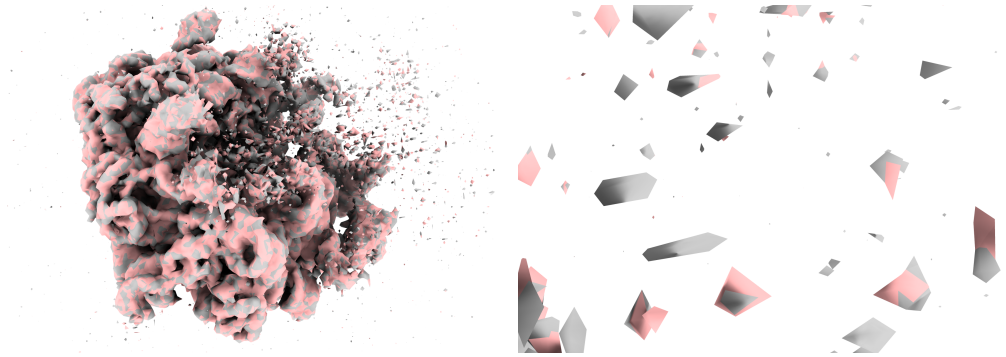
Overfitting in the context of cryo-EM occurs when the refinement algorithm begins to fit not only the true structural signal of the particle images but also the random noise inherent to the dataset. This happens because, with repeated exposure to the same particles in the same grouping, the algorithm progressively adapts its model to reproduce small, noise-driven variations that are not representative of genuine structural features. As a result, the refined map may appear to display high-resolution details that are not supported by reproducible signals, leading to artificially inflated resolution estimates and potentially misleading structural interpretations.

The mechanism behind this overfitting is straightforward. Each refinement iteration updates particle orientations, shifts, and weighting parameters to maximize the agreement between the 3D reconstruction and the experimental data. When the same set of particles and their corresponding noise patterns are repeatedly presented to the algorithm, these adjustments become increasingly specialized to the dataset, rather than to features common to all possible instances of the underlying structure. This phenomenon is especially problematic when working with low SNR data, where noise patterns can be mistaken for structural information, and when the dataset contains conformational or compositional heterogeneity that is not explicitly modeled.

Figure 5.24 shows an example of overfitting, where even the noise features are matched between maps, a clear sign that the reconstruction is fitting to noise rather than the true signal.

The dynamic clustering proposed in the pipeline helps to mitigate this effect. By reassigning particles to different clusters between refinement cycles, the refinement process is exposed to slightly different noise realizations in each iteration. This approach reduces the likelihood of the algorithm overfitting to the noise associated with a fixed subset of particles, as it effectively acts as a form of regularization similar in spirit to stochastic optimization techniques in machine learning. In this way, the refinement is forced to adapt to structural features that are consistent across varying subsets, thereby improving the robustness and generalizability of the final reconstruction.

To avoid confusing good reconstructions with overfitting, it is necessary to visualize the reconstructions, compare them with each other, and fit them with



(a) Density maps from different iterations. (b) Noise captured in both reconstructions.

Figure 5.24: Example of overfitting during excessive Frealign iterations. Figure 5.24a shows nearly identical density maps despite the intersection of the FSC with the threshold increasing, indicating overfitting. Figure 5.24b shows that even the noise features are matched between maps.

the corresponding atomic model.

Experiments were then performed using a larger architecture and with the original image size ($D = 320$). The results of applying UMAP with two and three components for this architecture can be seen in Figures 5.25 and 5.26.

In both results, the appearance of a new outlier-like cluster can be observed. This may be more consistent with the results reported in the original work on the dataset, where there are two classes with around two thousand particles.

Figure 5.27 shows a reconstruction of this new cluster, which appears to have a new subunit that none of the previous particles had. This ribosomal subunit is called the 30S.

Figure 5.28 shows an example of the atomic model of the ribosome, which is made up of two main parts: the large subunit, called 50S, and the small subunit, called 30S. The 50S subunit can be thought of as the part where the assembly of proteins takes place, while the 30S subunit is mainly responsible for reading the genetic instructions. When both come together, they form the complete 70S ribosome.

The results for the larger architecture can be seen in Tables 5.11 and 5.12

Regarding clustering with HDBSCAN, the results can be seen in figures 5.29 and 5.30. These show that the method can be very unstable, with the number of clusters changing abruptly between iterations.

Another observation that can be made about clustering with this architecture, and which could even be seen in unclassified UMAPs, is that by working with a deeper architecture and the original size of the images, the network may be able to identify more clusters.

This raises the question of whether, in this dataset and with this architecture, it is possible to identify predominant conformations. This was reviewed, but the reality is that even in HDBSCAN iterations with more clusters identified, all reconstructions have compositional differences.

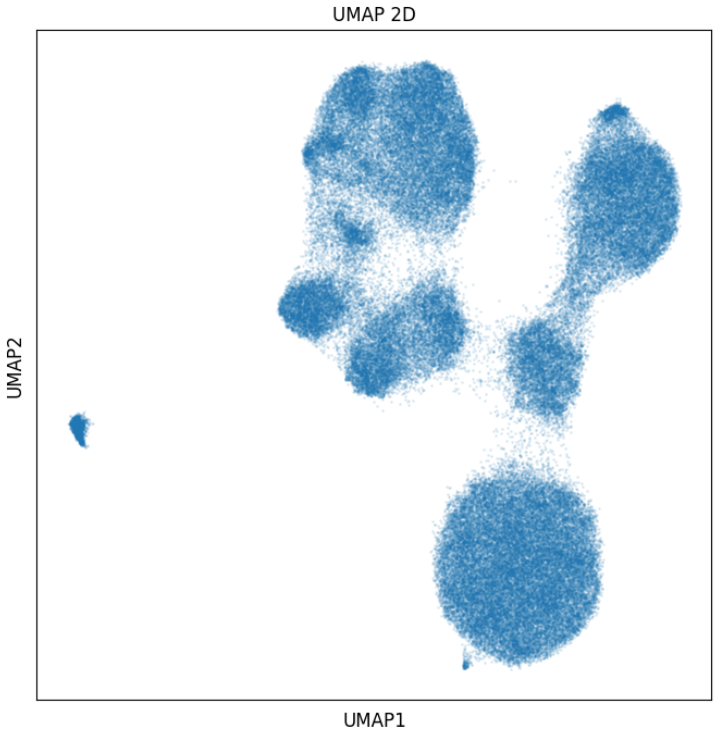


Figure 5.25: 2D UMAP of EMPIAR-10076 for the large cryoDRGN architecture and the original image size.

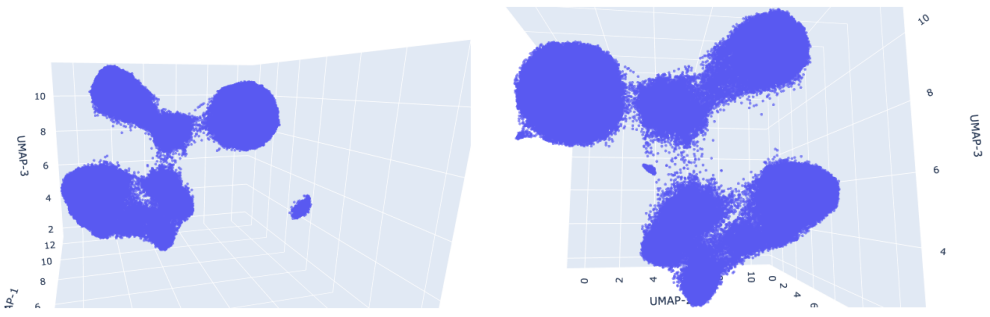


Figure 5.26: 3D UMAP of EMPIAR-10076 for the large cryoDRGN architecture and the original image size.

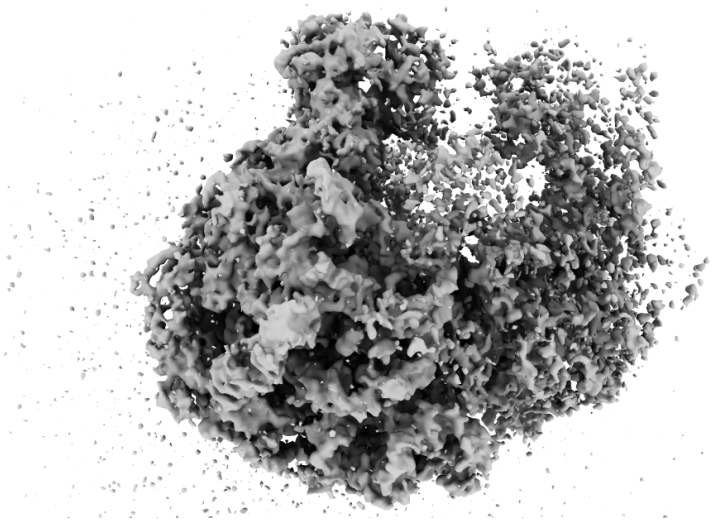
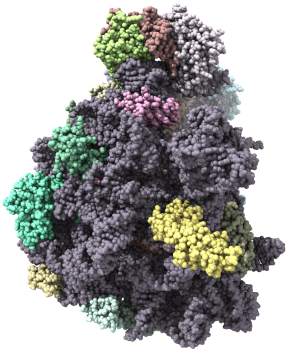
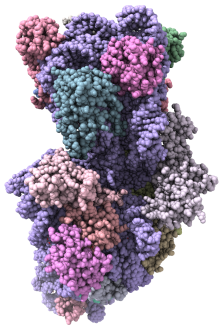


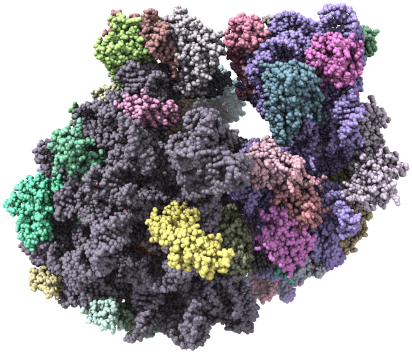
Figure 5.27: Density map corresponding to 70S subunit of EMPIAR-10076.



(a) Atomic model of 50S subunit



(b) Atomic model of 30S subunit



(c) Atomic model of 70S subunit

Figure 5.28: Atomic models of 50S subunit, 30S subunit and 70S subunit. The 70S subsunit is composed by subunits 50S and 30S.

Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	5.31	5.23	5.91	5.85	5.24	5.49	5.67	**5.21**	5.46	6.20
cluster1	7.77	7.47	7.23	7.06	5.63	4.79	4.39	4.06	**3.88**	7.35
cluster2	4.10	3.89	3.88	3.94	3.95	3.93	3.95	3.88	**3.83**	3.89
cluster3	3.63	3.51	3.46	3.56	3.53	3.50	3.48	3.44	3.44	**3.42**
cluster4	3.89	3.77	**3.69**	3.80	3.82	3.81	3.77	3.72	3.72	3.75
cluster5	5.18	5.03	4.91	**4.53**	4.69	4.60	4.55	4.64	4.75	4.94

Table 5.11: FSC values for EMPIAR-10076 with K-Means (K=6), 320x320 images and 2D UMAP. The numbers in bold represent the best resolution for a given cluster.

Cluster	iter0	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	5.34	**5.23**	5.92	5.83	**5.23**	5.59	5.54	5.36	5.34	6.18
cluster1	7.75	7.46	7.21	5.86	5.71	5.67	4.73	4.68	**4.66**	7.27
cluster2	5.23	5.08	4.95	4.89	5.03	4.99	4.98	**4.79**	4.81	5.03
cluster3	3.93	3.77	3.71	3.81	3.82	3.79	3.77	3.71	**3.69**	3.73
cluster4	3.66	3.52	3.47	3.56	3.52	3.47	3.47	3.46	3.43	**3.41**
cluster5	4.14	3.93	3.87	3.94	3.95	3.91	3.91	3.87	**3.84**	3.87

Table 5.12: FSC values for EMPIAR-10076 with K-Means (K=6), 320x320 images and 3D UMAP. The numbers in bold represent the best resolution for a given cluster.

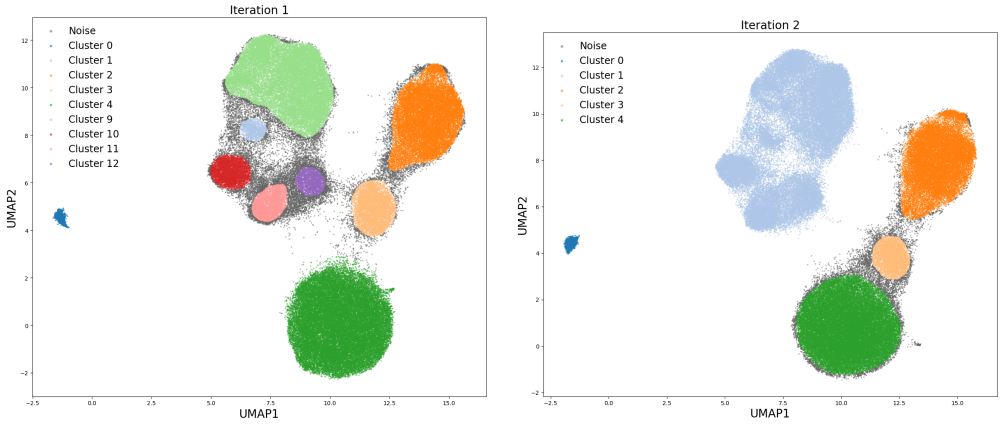


Figure 5.29: 3D UMAP clustering with HDBSCAN between consecutive iterations.

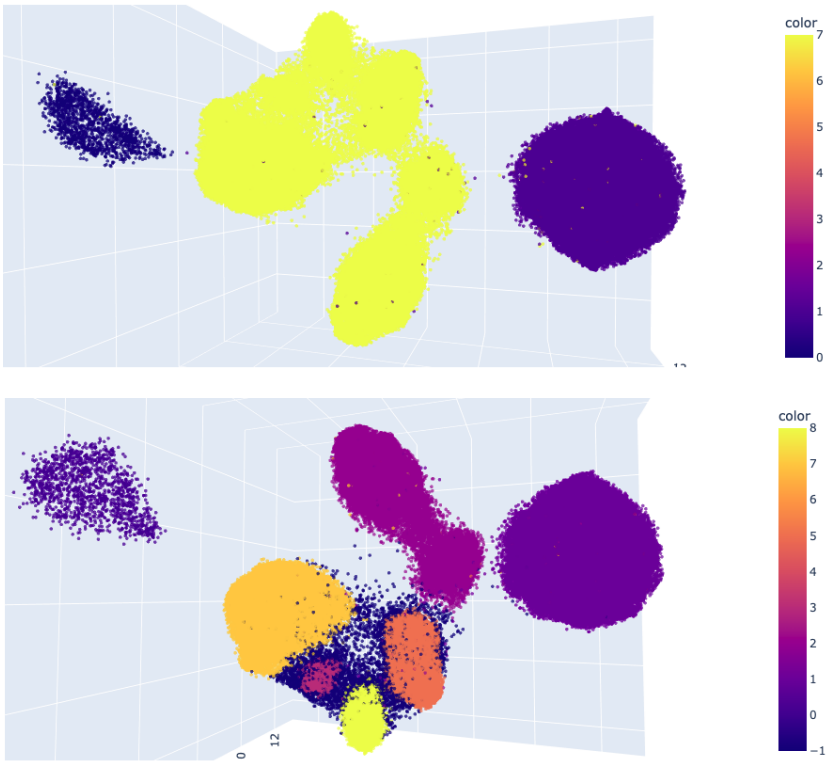


Figure 5.30: 3D UMAP clustering with HDBSCAN between consecutive iterations.

The FSCs for HDBSCAN can be seen in Tables 5.13 and 5.14. In these experiments, it can be seen that the number of clusters identified increases in both cases, and the number of clusters identified in each iteration is considerably more irregular than in other experiments with HDBSCAN. This becomes a problem during refinement, because both the number of clusters and the size of each cluster vary significantly, and this can lead to instability between iterations. In fact, the results obtained show poor resolution in most of the clusters.

To conclude the experiments, and continuing along the line of conformational heterogeneity, an experiment with $K = 15$ was performed. Figure 5.31 shows the clustering results obtained with the smaller architecture using downsampled images at 128×128 pixels.

The clustering performance remains poor, once again suggesting that the data are being forced into an inappropriate number of clusters. It is worth noting that, even when predominant conformations are present and correctly captured by the network, the expected outcome in UMAP visualizations is to observe some degree of clustering, even if the clusters appear less distinct or more diffuse than in the case of compositional heterogeneity.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	5.19	5.99	5.90	5.36	5.63	5.31	5.14	**5.10**	5.89
cluster1	5.25	3.76	5.05	3.66	3.50	5.16	3.44	**3.41**	5.02
cluster10	4.30		3.53			4.28		**4.18**	
cluster11	**4.90**								
cluster12	**5.31**								
cluster2	3.80	3.75	3.82	3.84	3.79	3.78	3.74	**3.70**	3.82
cluster3	7.76	6.39	5.38	5.99	5.91	5.35	5.22	5.04	**5.02**
cluster4	7.45	7.18	5.84	4.85	4.77	**4.63**	5.94	5.79	5.63
cluster8			4.14			4.32		**4.18**	
cluster9	3.54		4.19			**3.49**		3.51	

Table 5.13: FSC values for EMPIAR-10076 with HDBSCAN, 320x320 images and 2D UMAP. The numbers in bold represent the best resolution for a given cluster. Incomplete rows correspond to clusters that were not identified in all iterations.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	5.18	5.94	5.93	5.46	5.50	5.49	**5.15**	5.18	5.82
cluster1	7.46	7.20	6.98	6.84	6.06	**4.74**	5.95	5.81	5.65
cluster10						4.30	4.31	4.28	**4.19**
cluster11						5.14	**4.98**	4.98	
cluster12						5.19		**4.90**	
cluster2	3.78	3.73	3.79	**3.75**					
cluster3	3.49	**3.44**	5.07	5.50					
cluster5		4.96	**4.36**	4.38					
cluster7			3.51	3.49	**3.33**	3.41	3.48	3.45	3.56
cluster8			4.19	4.35		4.22	4.24	**4.18**	4.27
cluster9						3.72	3.72	**3.71**	3.80

Table 5.14: FSC values for EMPIAR-10076 with HDBSCAN, 320x320 images and 3D UMAP. The numbers in bold represent the best resolution for a given cluster. Incomplete rows correspond to clusters that were not identified in all iterations.

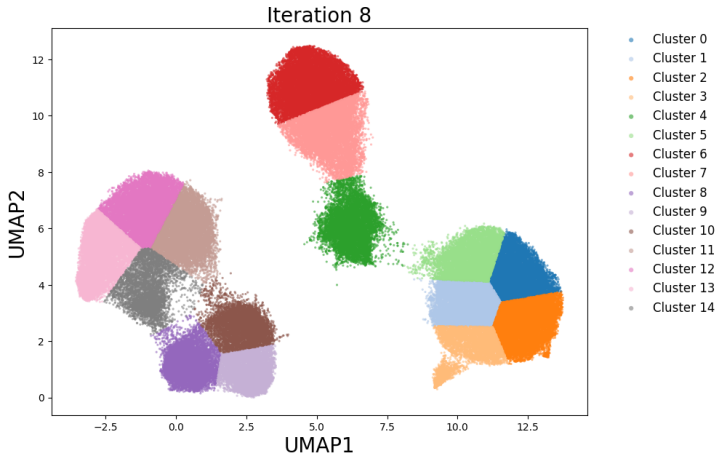


Figure 5.31: 2D UMAP K-Means clustering with $K = 15$.

Cluster	iter1	iter2	iter3	iter4	iter5	iter6	iter7	iter8	iter9
cluster0	8.60	7.74	7.56	8.07	8.31	8.27	9.90	14.74	**5.06**
cluster1	8.06	7.91	7.64	7.91	7.83	**4.89**	8.18	9.79	6.15
cluster2	7.96	7.93	7.62	8.16	7.77	8.21	9.64	11.02	**5.78**
cluster3	7.97	7.96	7.84	7.88	7.85	8.39	9.53	10.43	**4.62**
cluster4	8.00	5.61	5.32	5.03	8.00	8.61	6.00	5.63	**4.27**
cluster5	**4.21**	5.26	4.74	7.78	5.23	5.21	4.64	8.93	5.06
cluster6	4.42	4.60	4.25	4.40	4.32	4.58	4.37	4.06	**3.93**
cluster7	4.12	4.03	4.40	4.08	**4.01**	4.27	4.35	4.25	8.48
cluster8	**4.06**	4.14	4.90	4.29	4.23	4.26	4.42	4.54	8.39
cluster9	4.30	4.95	4.83	**4.82**	4.77	4.84	4.85	5.13	8.72
cluster10	**4.14**	4.32	4.64	4.86	4.75	4.86	4.81	5.09	9.24
cluster11	5.15	4.40	4.04	**3.97**	4.04	4.12	4.13	4.02	8.17
cluster12	4.61	4.22	4.02	**3.94**	4.09	4.15	4.15	3.95	4.88
cluster13	4.22	4.30	4.31	**4.08**	**4.08**	4.17	4.26	4.19	4.16
cluster14	4.31	4.81	**4.05**	4.55	4.67	4.34	4.79	4.98	4.06

Table 5.15: FSC values for EMPIAR-10076 with K-Means (K=15), 128x128 images and 2D UMAP. The numbers in bold represent the best resolution for a given cluster.

Class	Method	Iteration	Resolution	Reference
A	11	7	5.21	6.5
B	11	3	4.53	4.5
C	11,12	2,8	3.69	3.7
D	11	8	3.83	4.0
E	10	7,8	3.34	3.7

Table 5.16: Comparison of results obtained with different methods. Class labels were assigned according to their definitions in the original study [16]. Clusters were aligned to each class to determine their correspondence.

5.5 Concluding Remarks

The density maps with the highest resolutions were compared against the corresponding entries available in the EMDB. Reconstructions whose reported resolution limits exceeding the Nyquist frequency were excluded from further consideration. Additionally, each map was inspected to ensure that the observed resolution genuinely reflected structural detail rather than artifacts caused by overfitting, thus confirming that the resolution values were supported by meaningful signal.

Table 5.16 shows the resolution obtained for each class, the method used (using the IDs from Table 5.1), the iteration in which this resolution was obtained, and the resolution of the EMDB density maps used as a reference.

The results obtained show that the K-Means method with $K = 6$ on the large architecture and 2D UMAP performed better, achieving the best resolution in four of the five density maps. This is due in part to the stable performance of K-Means across the iterations and the unstable behavior of HDBSCAN, but also to the architecture's ability to better capture the image features. This is particularly noticeable in class A, which could be identified as a class, allowing for a substantial improvement in its reconstruction. It is also worth mentioning that better resolutions are obtained for classes D and E in most experiments.

Figure 5.32 shows in the top row the EMDB density maps, and in the bottom row the best density maps obtained with the proposed method.

At first glance, it is very difficult to distinguish improvements between the different maps. The most notable case is in class A, which also shows the greatest improvement in resolution. Figures 5.33 and 5.34 show examples where the improvement can be seen in this class. It is worth noting that the best resolution can be observed in the 50S subunit. Although the network managed to capture this class, the representation of the 30S subunit is still poor.

In the local comparisons between the atomic model and the density maps, clear differences can be observed in the degree of fit depending on the region analyzed.

In Figure 5.35, which corresponds to an area containing helices and loops, the map obtained with the proposed pipeline (5.35a) shows a density that is more centered and confined around the polypeptide chain. As a result, both the backbone and the side-chain carbons of the model remain enclosed within the density envelope. In contrast, in the reference EMDB map (5.35b), the density appears shifted

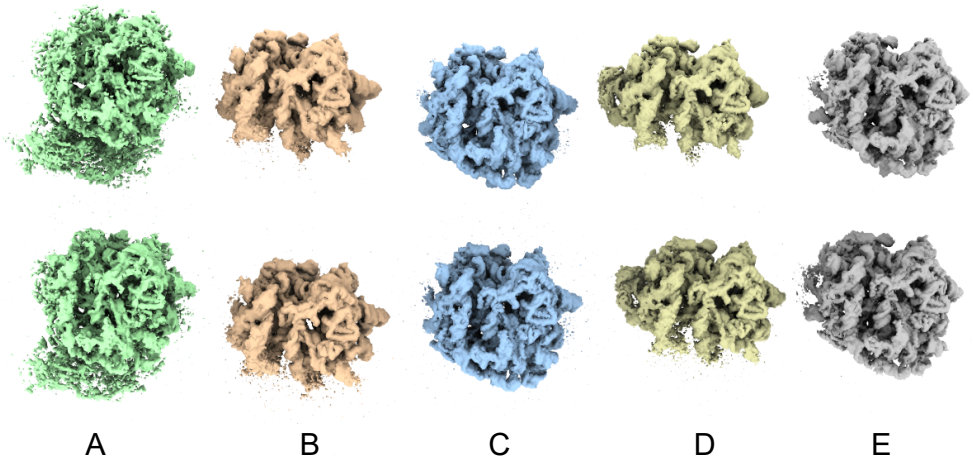


Figure 5.32: Comparison between EMDB density maps (top row) and the best density maps achieved (bottom row) for each class (A, B, C, D and E).

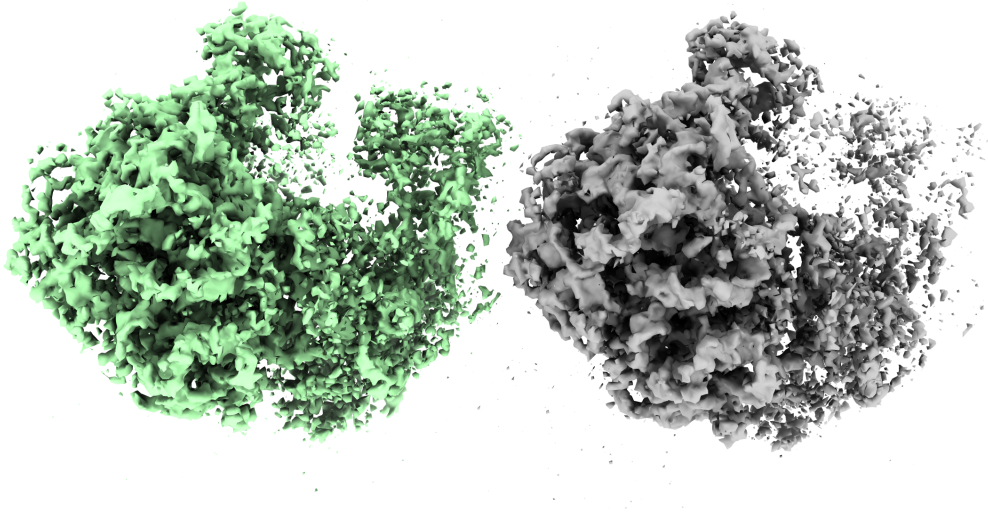
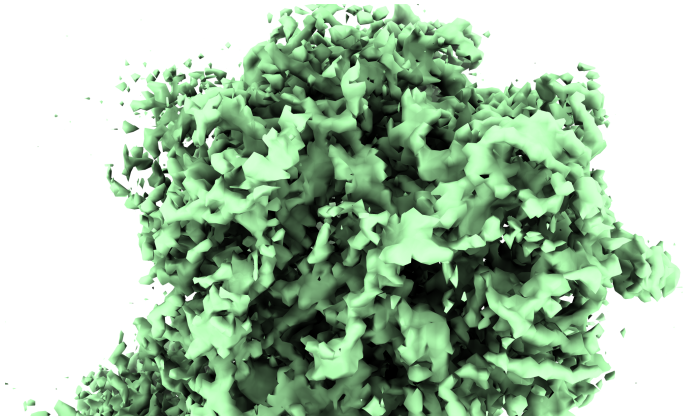
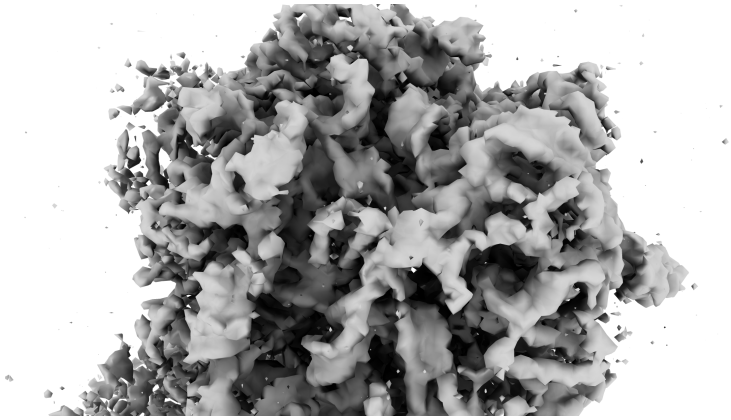


Figure 5.33: Comparison between the EMDB density map corresponding to class A (left) and the density map achieved with the proposed method corresponding to this class (right).



(a) Reconstruction from EMDB



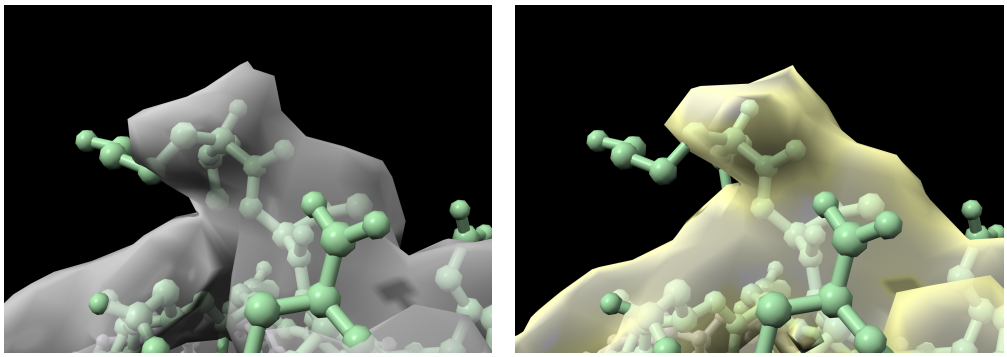
(b) Result from the proposed pipeline

Figure 5.34: Density maps from class A. 5.34a appears to be degraded, losing features such as the alpha helices present in the protein. Figure 5.34b also shows some degradation, but to a lesser extent.

relative to the atomic coordinates, leaving parts of the chain outside the highest density regions. This misalignment suggests that the reconstruction obtained with the proposed pipeline provides a more accurate description of the molecular structure in this region.

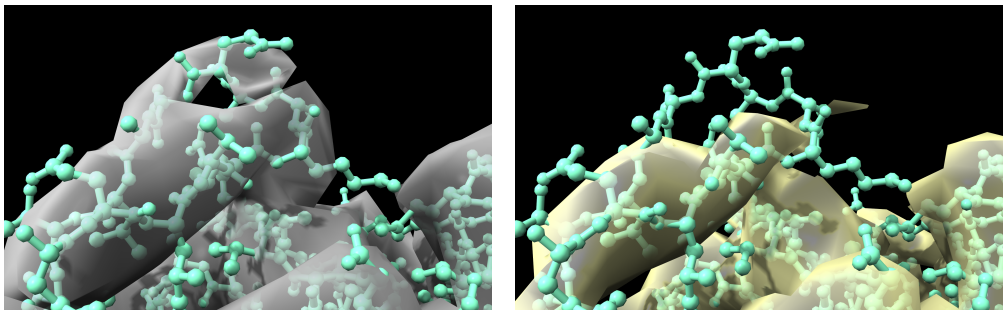
In Figure 5.36, which corresponds to a different area of the same class, a similar trend is observed. The map reconstructed by the pipeline (5.36a) follows the trajectory of the backbone more consistently and provides clear density around the side chains. In the reference map (5.36b), however, the density is more diffuse and in some cases does not fully enclose the model, with atoms extending beyond the density contours. These differences indicate that the proposed pipeline is able to capture local structural features more faithfully, resulting in a closer match between the atomic model and the experimental density.

In Figure 5.37, the comparison highlights complementary strengths between the two reconstructions. In the map generated with the proposed pipeline (5.37a),



(a) Result from the proposed pipeline (b) Reconstruction from EMDB

Figure 5.35: Comparison of the density map obtained from the proposed pipeline (5.35a) and the density map from EMDB (5.35b) for class E. It can be seen that the polypeptide chain is more centered in the first figure, indicating a more accurate reconstruction in that region of the map.

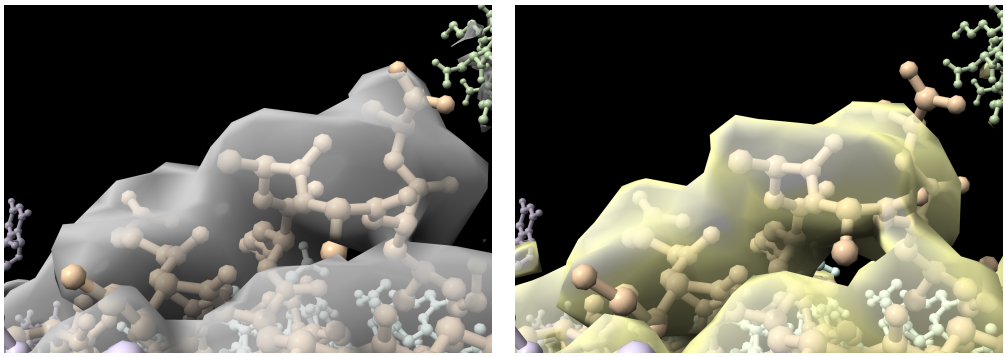


(a) Result from the proposed pipeline (b) Reconstruction from EMDB

Figure 5.36: Comparison of the density map obtained from the proposed pipeline (5.36a) and the density map from EMDB (5.36b) for class E. The pipeline results better capture information in this region of the density map.

one end of the polypeptide chain is more clearly captured, with density features that better represent this local region. However, in the reference EMDB map (fig:compa32), the density appears more consistently aligned with the overall atomic model, with the backbone fitting more centrally within the envelope. These observations indicate that the two maps emphasize different aspects of the structure: while the pipeline result seems to better resolve specific local features, the EMDB map provides a more globally consistent alignment with the model. This suggests that improvements in the reconstruction may come from combining the strengths of both approaches—accurate local detail and robust global alignment.

It's important to clarify that even if one density map has better resolution than another, not all features are necessarily better. In the search to identify improvements between EMDB density maps and the results obtained, differences that favor the EMDB results were also seen.



(a) Result from the proposed pipeline (b) Reconstruction from EMDB

Figure 5.37: Comparison of the density map obtained from the proposed pipeline (5.37a) and the density map from EMDB (5.37b) for class E. In this case, the pipeline result better captures one end of the polypeptide chain, but the EMDB result appears to be better aligned with the atomic model.

This clarifies that it is absolutely necessary to find a more reliable metric of resolution, especially in cases where heterogeneity is involved. In these cases, there is an intrinsic trade-off between accurately capturing structural heterogeneity and achieving high global resolution. A reconstruction that better preserves subtle conformational or compositional differences may present a lower nominal resolution, even though it provides more biologically relevant information. On the other hand, maximizing global resolution can lead to an averaging effect that suppresses heterogeneous features.

These could be addressed with metrics that integrate both spatial and structural variability, rather than relying on a single global value. This would provide a more meaningful and accurate measure of reconstruction quality in the presence of heterogeneity.

Chapter 6

Conclusions

This work presents an iterative framework for pose refinement and heterogeneous macromolecule reconstruction from cryo-EM projections. The work successfully exploits the inherent heterogeneity of the particles studied to refine poses and achieve improved resolutions. However, this method has limitations.

The first conclusion drawn from the results is that the current pipeline is not well-suited for exploring conformational heterogeneity. The continuous nature of conformational changes is inherently incompatible with the UMAP algorithm, since one favors continuous trajectories while the other prioritizes local neighborhood preservation. In contrast, compositional changes appear to be accurately captured by this dimensionality reduction approach.

At high resolution, pipeline stability becomes critical. Achieving good resolutions without overfitting is challenging. Iterative clustering is essential, as overfitting in Frealign mainly occurs when pose and parameter refinements begin to fit the noise of individual particles after excessive iterations on the same dataset. Nevertheless, allowing excessively large variations between clusters destabilizes the pipeline and does not lead to improved resolutions. One possible solution is to initialize UMAP or K-Means centroids from the previous iteration, or to perform a hyperparameter grid search for HDBSCAN to maintain a consistent number of clusters between iterations.

The ability of cryoDRGN to represent features in a lower-dimensional space is particularly valuable, as it enables the use of clustering techniques while avoiding the curse of dimensionality. This approach is notably more effective than applying algorithms directly to noisy particle images.

Furthermore, the use of nonlinear dimensionality reduction methods such as UMAP facilitates the visualization and interpretation of results, thanks to their ability to preserve both local and global data structure. UMAP also captures compositional heterogeneity effectively, aiding the clustering process—something that linear techniques like PCA fail to achieve. This limitation arises because variational autoencoders tend to distribute information across all latent dimensions. In datasets with high variability and a low-dimensional latent space, the different latent dimensions will not be redundant nor simple linear combinations of each other, provided that the network has sufficient capacity to learn the data's

features.

Another important conclusion is the need for a faster decoder. While Frealign's pose refinement yields excellent resolution, it is a CPU-based tool whose computational cost scales steeply with the number and size of input images. Recent advances in this area, driven by the boom in generative deep learning and the evident need for efficient heterogeneous refinement methods, offer promising alternatives [34] [36]. However, developing a decoder capable of achieving high-resolutions in heterogeneous settings remains an open challenge.

Pipeline automation is valuable, but at high resolution, careful inspection and hyperparameter tuning between iterations are essential. Precision in parameter selection becomes increasingly important as resolution increases.

Refining poses after a set number of epochs can be viewed as a form of network regularization, helping to prevent convergence to local minima. This strategy may lead to clusters that favor certain structural classes over others, even if not all clusters benefit equally. While this does not guarantee improvements across all clusters, it may enable higher resolution for specific ones.

Finally, the postprocessing block from Figure 2.6 is gaining more relevance and is becoming increasingly necessary when working in high resolution. The development and adoption of more reliable resolution assessment metrics is essential, especially those capable of quantifying local resolution variations and evaluating the fidelity of heterogeneous features.

6.1 Future Directions

Reconstruction of heterogeneity density maps with cryo-EM images still has many unsolved problems. One of the main challenges is incorporating a decoder that effectively generates high-resolution reconstructions. All the works presented in Chapter 3 have issues with the decoder generating high-resolution structures. This could be explored by incorporating more modern architectures in the image processing area, such as diffusion models or transformers. To date, there are already works with these approaches [30] [38], but no good results have been obtained so far.

Another option is to dedicate efforts to cryo-electron tomography (cryo-ET) processing. Cryo-ET is an acquisition technique that, instead of taking a single projection of each particle, acquires a series of images, called a tilt series, of the same particle as it is tilted in the microscope. This provides multiple projections of the same particle, opening the possibility of obtaining complementary views of complex regions of the protein and facilitating reconstruction when the particle adopts a minority conformation or composition. However, the problem of pose estimation is not completely resolved, since cryo-ET only provides the relative tilt geometry of the series, while the absolute orientations of the particles within the tomogram remain unknown. The main disadvantage of these methods today is the limitation of the electron beam dose, which must be lower than that of cryo-EM. This is because the sample is irradiated multiple times, so the power must be reduced to avoid damaging the sample.

Another promising direction would be the development of a foundation model for protein density map reconstruction. Unlike AlphaFold, which operates at the folding level by taking amino acid sequences as input, such a model would work directly with experimental density data from cryo-EM. Current approaches are usually trained on a single protein, but structural motifs such as -helices and -sheets are shared across all proteins. Leveraging datasets from multiple proteins could therefore improve generalization and lead to more robust reconstructions. This approach, however, introduces challenges in data preprocessing, network design, and the computational demands of training on large-scale datasets. As a proof of concept, one could start with a more contained problem, such as training on the ribosome family, where many datasets are publicly available.

6.2 Personal Reflections

My master's degree in electrical engineering gave me the opportunity to study in depth such an interesting, important, and difficult topic as macromolecule reconstruction using cryo-EM images. This allowed me to learn a lot about research, an area I've been delving into since completing my undergraduate degree. One of the most important things I think I'm beginning to understand is the difficulty involved in delving deeply into a particular topic. Despite having studied, used, and partially implemented image processing and volume reconstruction techniques in this area, I feel there's still a lot of ground to cover and a huge amount to learn.

This master's degree also inspired me to pursue a PhD, something that was unthinkable for me before starting this path. I'm aware that I still have a lot to learn, but I'm excited to continue advancing in my career, and I think a PhD is a great way to do that. I hope and trust that everything I've learned here can be useful in this new step of my career.

Esta página ha sido intencionalmente dejada en blanco.

Appendix A

Generative Models

A.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) [19] are a class of generative models that frame training as a two-player minimax game between a *generator* and a *discriminator*. The generator learns to map samples from a simple prior distribution (e.g., Gaussian noise) into realistic data, while the discriminator learns to distinguish between real data and generator-produced samples.

A.1.1 Adversarial Objective

Let $p_{\text{data}}(x)$ denote the true data distribution and $p_z(z)$ a prior (e.g., $\mathcal{N}(0, I)$). The generator $G_\theta(z)$ transforms latent noise z into a synthetic sample $\tilde{x}$. The discriminator $D_\phi(x) \in (0, 1)$ outputs the probability that x is real. The value function is

$$\min_\theta \max_\phi V(D_\phi, G_\theta) \;=\; \mathbb{E}_{x\sim p_{\text{data}}}[\log D_\phi(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D_\phi(G_\theta(z)))]. \quad \text{(A.1)}$$

At equilibrium, G_θ produces samples indistinguishable from real data, and D_ϕ outputs $1/2$ for both real and fake inputs.

A.1.2 Training Dynamics

The training of GANs follows a minimax game between the generator G_θ and the discriminator D_ϕ. The two networks are trained alternately in an iterative process:

1. **Discriminator step:** Update D_ϕ to maximize its ability to distinguish between real and generated samples. This typically involves maximizing

$$\mathcal{L}_D = \mathbb{E}_{x\sim p_{\text{data}}}[\log D_\phi(x)] + \mathbb{E}_{z\sim p(z)}[\log(1 - D_\phi(G_\theta(z)))].$$

2. **Generator step:** Update G_θ to minimize the probability of the discriminator correctly identifying its outputs as fake. A naive formulation minimizes

$$\mathcal{L}_G = \mathbb{E}_{z\sim p(z)}[\log(1 - D_\phi(G_\theta(z)))].$$

However, this objective often leads to vanishing gradients when the discriminator is too strong. In practice, one instead maximizes

$$\mathcal{L}'_G = \mathbb{E}_{z\sim p(z)}[\log D_\phi(G_\theta(z))],$$

which provides stronger gradients for learning.

This adversarial dynamic resembles a two-player zero-sum game, where the generator improves by "fooling" the discriminator, and the discriminator improves by better distinguishing real from synthetic data. In theory, the equilibrium is reached when $p_g = p_{\text{data}}$, meaning the generator has learned to replicate the true data distribution, and the discriminator cannot distinguish real from fake samples better than random guessing.

In practice, the dynamics are far more complex. If the discriminator learns too quickly, it may perfectly classify real and fake samples, providing no useful gradient to the generator. Conversely, if the generator improves faster, the discriminator may fail to adapt, leading to unstable training and oscillatory behavior. It may even be able to learn to fool the discriminator with a small sample of nearly identical images, which is called mode collapse.

A.1.3 Evaluation

Evaluating GANs is non-trivial because they do not provide an explicit likelihood, making direct statistical comparison with the true data distribution difficult. As a result, evaluation typically relies on quantitative metrics and qualitative assessments.

A straightforward approach is manual inspection of generated samples. While subjective, this remains informative for perceptual quality, especially in image synthesis.

Another alternative is the Inception Score [52], which uses a pretrained classifier (commonly the Inception-v3 network) to evaluate both the quality and diversity of generated samples. For a set of generated images $\{\tilde{x}_i\}$, the conditional label distribution is $p(y|\tilde{x})$, and the marginal class distribution is $p(y) = \frac{1}{N}\sum_i p(y|\tilde{x}_i)$. The score is defined as

$$\text{IS} = \exp\left(\mathbb{E}_{\tilde{x}}\left[\mathrm{D}_{\mathrm{KL}}\big(p(y|\tilde{x}) \,\|\, p(y)\big)\right]\right). \tag{A.2}$$

A high IS indicates (i) low-entropy conditional distributions (the classifier is confident, suggesting high-quality samples), and (ii) high-entropy marginal distribution (samples are diverse across classes). Limitations include sensitivity to the pretrained classifier and lack of direct comparison with the real data distribution.

The Fréchet Inception Distance (FID) [23] compares the feature statistics of real and generated images using embeddings from a pretrained network (again, typically Inception-v3). If $\mathcal{N}(\mu_r, \Sigma_r)$ and $\mathcal{N}(\mu_g, \Sigma_g)$ are Gaussian approximations to the real and generated feature distributions, then

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}\Big(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}\Big). \tag{A.3}$$

Lower FID values indicate closer alignment between the real and generated distributions. FID is widely used as it accounts for both sample quality and diversity, though it is still dependent on the choice of embedding network and may not correlate perfectly with human perception.

Additional methods include Kernel Inception Distance (KID), precision-recall tradeoffs for generative models, and task-specific evaluations (e.g., downstream classification accuracy when augmenting training data). These highlight the importance of choosing evaluation metrics aligned with the intended application.

A.1.4 Limitations

Despite their successes, GANs face several fundamental challenges stemming from the adversarial training dynamics.

GAN optimization is formulated as a minimax game between the generator and the discriminator. In practice, this dynamic can fail to converge: the generator may oscillate between strategies, or the discriminator may saturate and provide vanishing gradients. Carefully balanced learning rates, architecture choices, and regularization are required to maintain stability.

Another common issue is mode collapse, where the generator learns to produce a limited variety of outputs that successfully fool the discriminator. This occurs because the generator is rewarded for exploiting weaknesses of the current discriminator rather than learning the full data distribution. Techniques such as minibatch discrimination, unrolled GANs, or Wasserstein-based objectives partially mitigate this phenomenon. Otherwise, if the discriminator learns too quickly, it can perfectly separate real from fake samples, providing no useful gradient signal to the generator.

Unlike VAEs or autoregressive models, GANs do not assign explicit likelihoods to data. This complicates tasks such as density estimation, anomaly detection, or model comparison. Evaluation thus depends on proxy metrics (IS, FID), which may not always reflect real-world utility.

Overall, the strengths of GANs—sharp, realistic generations—come at the cost of complex training dynamics and evaluation challenges. Ongoing research addresses these limitations through alternative objectives (e.g., Wasserstein distances), architectural innovations, and hybrid models combining the advantages of GANs with those of likelihood-based approaches.

A.2 Variational Autoencoders (VAEs)

Variational Autoencoders (VAEs) [28, 49] are latent-variable generative models that combine deep neural networks with approximate Bayesian inference. A VAE specifies a prior over latent variables $z \in \mathbb{R}^d$, typically $p(z) = \mathcal{N}(0, I)$, and a likelihood (decoder) $p_\theta(x \mid z)$ parameterized by θ. An inference network (encoder) $q_\phi(z \mid x)$ with parameters ϕ amortizes variational inference across the dataset.

A.2.1 Objective: the Evidence Lower Bound (ELBO)

For data $x \sim p_{\text{data}}(x)$, maximizing the marginal likelihood $\log p_\theta(x)$ is intractable in general. Introducing a variational posterior $q_\phi(z \mid x)$ yields the standard decomposition:

$$\log p_\theta(x) \;=\; \underbrace{\mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x \mid z)\big] - \mathrm{D}_{\mathrm{KL}}\big(q_\phi(z \mid x) \,\|\, p(z)\big)}_{\mathcal{L}(\theta,\phi;x)\ \text{(ELBO)}} + \mathrm{D}_{\mathrm{KL}}\big(q_\phi(z \mid x) \,\|\, p_\theta(z \mid x)\big), \tag{A.4}$$

where the last term is always nonnegative. Thus, maximizing the *evidence lower bound* (ELBO)

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x \mid z)\big] - \mathrm{D}_{\mathrm{KL}}\big(q_\phi(z \mid x) \,\|\, p(z)\big) \tag{A.5}$$

provides a tractable surrogate. The reconstruction term encourages faithful reconstructions, while the KL divergence ensures that the approximate posterior remains close to the prior.

A.2.2 Reparameterization Trick

Directly sampling $z \sim q_\phi(z \mid x)$ would prevent gradients from flowing through the encoder parameters. The reparameterization trick addresses this issue. For a Gaussian posterior $q_\phi(z \mid x) = \mathcal{N}\big(z;\, \mu_\phi(x), \mathrm{diag}(\sigma^2_\phi(x))\big)$, one writes

$$z \;=\; \mu_\phi(x) \;+\; \sigma_\phi(x) \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{A.6}$$

This isolates randomness in ϵ, which is independent of ϕ, enabling low-variance gradient estimates via backpropagation. Without this trick, training VAEs with stochastic latents would require high-variance estimators such as score-function gradients.

A.2.3 Common Likelihoods and Losses

The reconstruction term depends on the choice of likelihood. For real-valued data (e.g., images scaled to $[0, 1]$), a Gaussian likelihood yields a mean-squared-error loss. For binary or bounded intensities, a Bernoulli likelihood leads to a cross-entropy loss. The minibatch objective for $\{x_i\}_{i=1}^B$ is:

$$\mathcal{J}(\theta, \phi) \;=\; -\frac{1}{B} \sum_{i=1}^{B} \Big(\mathbb{E}_{q_\phi(z|x_i)}\big[\log p_\theta(x_i \mid z)\big] - \mathrm{D}_{\mathrm{KL}}\big(q_\phi(z \mid x_i) \,\|\, p(z)\big)\Big). \tag{A.7}$$

A.2.4 β-VAE and Variants

A common extension introduces a scaling factor $\beta > 0$ on the KL term [24]:

$$\mathcal{L}_\beta(\theta, \phi; x) \;=\; \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x \mid z)\big] - \beta\, \mathrm{D}_{\mathrm{KL}}\big(q_\phi(z \mid x) \,\|\, p(z)\big). \tag{A.8}$$

When $\beta > 1$, the model enforces stronger regularization, often yielding more *disentangled* latent representations, where independent factors of variation in the data are captured by distinct dimensions. This comes at the expense of reconstruction fidelity. Conversely, $\beta < 1$ prioritizes reconstruction but weakens regularization. Other extensions include richer priors (e.g., VampPrior), more flexible posteriors via normalizing flows, and hierarchical latent-variable structures.

A.2.5 Practical Considerations

- **Architecture:** Encoders and decoders often mirror each other (e.g., convolutional networks for images). Latent dimensionality is task-dependent: too small limits expressivity, too large weakens regularization.
- **Training stability:** VAEs are generally stable to train, but *KL collapse* can occur when the decoder is overly expressive, causing $q_\phi(z|x) \approx p(z)$. Remedies include KL annealing, "free bits," or restricting decoder capacity.
- **Evaluation:** Metrics include the ELBO or tighter bounds such as the importance-weighted ELBO, reconstruction quality (e.g., PSNR/SSIM for images), and downstream utility of the learned latent space (e.g., clustering, interpolation).
- **Uncertainty and interpretability:** The probabilistic formulation naturally provides uncertainty estimates and enables meaningful interpolation in latent space, useful for anomaly detection and scientific modeling.

A.2.6 Limitations

Despite their probabilistic appeal, VAEs face several limitations:

- **Blurry outputs:** With simple Gaussian decoders, reconstructions and samples tend to be overly smooth compared to GANs. Richer decoders (e.g., autoregressive, diffusion) improve sharpness but increase computational cost.
- **KL collapse:** When the KL term vanishes, the encoder ignores latent variables, leading to poor generative performance.
- **Posterior mismatch:** The Gaussian assumption for $q_\phi(z|x)$ may not capture complex true posteriors. Techniques like normalizing flows expand posterior flexibility.
- **Tradeoff between disentanglement and fidelity:** Methods like β-VAE reveal a fundamental tension between interpretable latent factors and accurate reconstructions.

Esta página ha sido intencionalmente dejada en blanco.

Bibliography

[1] T. Ahmed, Z. Yin, and S. Bhushan. Cryo-em structure of the large subunit of the spinach chloroplast ribosome. *Scientific Reports*, 6:35793, 2016.

[2] Mebarka Allaoui, Mohammed Lamine Kherfi, and Abdelhakim Cheriet. Considerably improving clustering algorithms using umap dimensionality reduction technique: A comparative study. In Abderrahim El Moataz, Driss Mammass, Alamin Mansouri, and Fathallah Nouboud, editors, *Image and Signal Processing*, pages 317–325, Cham, 2020. Springer International Publishing.

[3] B. E. Bammes, R. H. Rochat, J. Jakana, D.-H. Chen, and W. Chiu. Direct electron detection yields cryo-em reconstructions at resolutions beyond 3/4 nyquist frequency. *Journal of Structural Biology*, 177:589–601, 2012.

[4] Niccolò Banterle, Khanh Huy Bui, Edward A. Lemke, and Martin Beck. Fourier ring correlation as a resolution criterion for super-resolution microscopy. *Journal of Structural Biology*, 183(3):363–367, 2013.

[5] T. Bepler, K. Kelley, A. Noble, and B. Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16:1153–1160, 2019.

[6] H. J. Berendsen and S. Hayward. Collective protein dynamics in relation to function. *Current Opinion in Structural Biology*, 10:165–169, 2000.

[7] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, TN Bhat, Helge Weissig, Innokentiy N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[8] Carl I. Branden and John Tooze. *Introduction to Protein Structure*. Garland Science, 2nd edition, 1999.

[9] M. A. Brubaker, A. Punjani, and D. J. Fleet. Building proteins in a day: efficient 3d molecular reconstruction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3099–3108, 2015.

[10] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.

[11] Giovanni Cardone, J. Bernard Heymann, and Alasdair C. Steven. One number does not fit all: mapping local variations in resolution in cryo-em reconstructions. *Journal of Structural Biology*, 184(2):226–236, November 2013.

[12] Ning Chen, Shuai Zhang, Sheng Zhang, Yuting Yan, Yu Chen, and Sanglu Lu. Resmap: Exploiting sparse residual feature map for accelerating cross-edge video analytics. In *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, pages 1–10, 2023.

[13] Xiao chen Bai, Thomas G. Martin, Sjors H. W. Scheres, and Hendrik Dietz. Cryo-em structure of a 3d dna-origami object. *Proceedings of the National Academy of Sciences*, 109(49):20012–20017, 2012.

[14] J. M. Chung and H. S. Jung. Cryo-electron tomography: A tool for in situ structural analysis of macromolecular complexes. *Applied Spectroscopy Reviews*, 53:195–202, 2017.

[15] Tiago R. D. Costa, Athanasios Ignatiou, and Elena V. Orlova. *Structural Analysis of Protein Complexes by Cryo Electron Microscopy*, pages 377–413. Springer New York, New York, NY, 2017.

[16] Joseph H. Davis, Yong Zi Tan, Bridget Carragher, Clinton S. Potter, Dmitry Lyumkis, and James R. Williamson. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167(6):1610–1622.e15, 2016.

[17] Claire Donnat, Axel Levy, Frédéric Poitevin, Ellen D. Zhong, and Nina Miolane. Deep generative modeling for volume reconstruction in cryo-electron microscopy. *Journal of Structural Biology*, 214(4):107920, December 2022. Review article.

[18] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231, 1996.

[19] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[20] Nikolaus Grigorieff. Frealign: High-resolution refinement of single particle structures. *Journal of Structural Biology*, 157(1):117–125, 2007. Software tools for macromolecular microscopy.

[21] Harshit Gupta, Michael T. McCann, Laurène Donati, and Michael Unser. Cryogan: A new reconstruction paradigm for single-particle cryo-em via deep adversarial learning. *IEEE Transactions on Computational Imaging*, 7:759–774, 2021.

[22] Jiahua He, Peicong Lin, Ji Chen, Hong Cao, and Sheng-You Huang. Model building of protein complexes from intermediate-resolution cryo-em maps with deep learning-guided automatic assembly. *Nature Communications*, 13(1):4066, July 2022.

[23] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, pages 6626–6637, 2017.

[24] Irina Higgins, Loïc Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017.

[25] Andrei Iudin, Paul K Korir, Josep Salavert-Torres, Gerard J Kleywegt, and Ardan Patwardhan. Empiar: A public archive for raw electron microscopy image data. *Nature Methods*, 13(5):387–388, 2016.

[26] Minkyu Jeon, Rishwanth Raghu, Miro Astore, Geoffrey Woollard, Ryan Feathers, Alkin Kaz, Sonya M Hanson, Pilar Cossio, and Ellen D Zhong. Cryobench: Diverse and challenging datasets for the heterogeneity problem in cryo-em. *arXiv:2408.05526*, 2024.

[27] J. Jumper, R. Evans, A. Pritzel, and et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021.

[28] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014.

[29] Sami Koho, Giorgio Tortarolo, Marco Castello, Takahiro Deguchi, Alberto Diaspro, and Giuseppe Vicidomini. Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nature Communications*, 10(1):3103, 2019.

[30] Karsten Kreis, Tim Dockhorn, Zihao Li, and Ellen Zhong. Latent space diffusion models of cryo-em structures, 2022.

[31] JR Kremer, DN Mastronarde, and JR McIntosh. Computer visualization of three-dimensional image data using imod. *Journal of Structural Biology*, 116(1):71–76, 1996.

[32] Grigorieff Lab. Frequently asked questions about frealign. `https://grigoriefflab.umassmed.edu/frequently_asked_questions_0`, 2016. Accessed: 2025-09-07.

[33] Catherine L Lawson, Ardan Patwardhan, Matthew L Baker, Corey Hryc, Brian P Hudson, Ingemar Lagerstedt, Steven J Ludtke, Helen M Berman,

Wah Chiu, and Gerard J Kleywegt. Emdatabank unified data resource for 3dem. *Nucleic Acids Research*, 44(D1):D396–D403, 2016.

[34] Axel Levy, Rishwanth Raghu, J. Ryan Feathers, Michal Grzadkowski, Frédéric Poitevin, Jake D. Johnston, Francesca Vallese, Oliver Biggs Clarke, Gordon Wetzstein, and Ellen D. Zhong. Cryodrgn-ai: neural ab initio reconstruction of challenging cryo-em and cryo-et datasets. *Nature Methods*, 22(7):1486–1494, 2025.

[35] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. Agard, and Y. Cheng. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em. *Nature Methods*, 10:584–590, 2013.

[36] Hsuan-Fu Liu, Ye Zhou, Qinwen Huang, Jeffrey Martin, and Alberto Bartesaghi. In situ structure determination of conformationally flexible targets using nextpyp. *Nature Protocols*, August 2025.

[37] Hsuan-Fu Liu, Ye Zhou, Qinwen Huang, Jonathan Piland, Weisheng Jin, Justin Mandel, Xiaochen Du, Jeffrey Martin, and Alberto Bartesaghi. nextpyp: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nature Methods*, 20(12):1909–1919, December 2023.

[38] Xinhang Liu, Yan Zeng, Yifan Qin, Hao Li, Jiakai Zhang, Lan Xu, and Jingyi Yu. Cryoformer: Continuous heterogeneous cryo-em reconstruction using transformer-based neural representations, 2023.

[39] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.

[40] Elaine C. Meng, Thomas D. Goddard, Eric F. Pettersen, Greg S. Couch, Zach J. Pearson, John H. Morris, and Thomas E. Ferrin. Ucsf chimerax: Tools for structure building and analysis. *Protein Science*, 32(11):e4792, 2023.

[41] J. A. Mindell and N. Grigorieff. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3):334–347, June 2003.

[42] Takanori Nakane and Sjors H. W. Scheres. Characterisation of molecular motions in cryo-em single-particle data by multi-body refinement. *eLife*, 7:e36861, 2018.

[43] Youssef S.G. Nashed, Frédéric Poitevin, Harshit Gupta, Geoffrey Woollard, Michael Kagan, Chun Hong Yoon, and Daniel Ratner. Cryoposenet: End-to-end simultaneous learning of single-particle orientation and 3d map reconstruction from cryo-electron microscopy data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 4066–4076, 2021. Fully unsupervised encoder-decoder for pose volume estimation, simulated data only.

[44] Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546(7660):617–621, June 2017.

[45] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2):155–162, January 1964.

[46] Ali Punjani and David Fleet. 3dflex: determining structure and motion of flexible proteins from cryo-em. *Nature Methods*, 20:1–11, 05 2023.

[47] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14(3):290–296, March 2017.

[48] Erney Ramírez-Aportela, Javier Mota, Pablo Conesa, Jose Maria Carazo, and Carlos Oscar S. Sorzano. *DeepRes*: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*, 6(6):1054–1063, Nov 2019.

[49] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 1278–1286. PMLR, 2014.

[50] Alexis Rohou and Nikolaus Grigorieff. Ctffind4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, 2015.

[51] Peter B. Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745, 2003.

[52] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, volume 29, pages 2234–2242, 2016.

[53] S. H. W. Scheres. A bayesian view on cryo-em structure determination. *Journal of Molecular Biology*, 415:406–418, 2012.

[54] Johannes Schwab, Dari Kimanius, Alister Burt, Tom Dendooven, and Sjors H. W. Scheres. Dynamight: estimating molecular motions with improved reconstruction from cryo-em images. *Nature Methods*, 21(10):1855–1862, October 2024.

[55] C. Suloway, J. Shi, A. Cheng, J. S. Pulokas, D. M. Carragher, B. R. Potter, C. S. Zheng, N. Fellmann, F. M. Voorhees, and S. H. C. De Carlo. Automated molecular microscopy: the new leginon system. *Journal of Structural Biology*, 151:41–60, 2005.

[56] Dimitry Tegunov and Patrick Cramer. Real-time cryo–em data pre-processing with warp. *Nature Methods*, 16(11):1146–1152, 2019.

[57] Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005.

[58] Jose Vilas, Josue Gomez-Blanco, Pablo Conesa, Roberto Melero, Jose Rosa-Trevín, Joaquín Otón, Julliana Cuenca, Roberto Marabini, José Carazo, Javier Vargas, and Carlos Sorzano. Monores: Automatic and accurate estimation of local resolution for electron microscopy maps. *Structure*, 26, 01 2018.

[59] Thorsten Wagner, Francisco Merino, Michael Stabrin, Takeshi Moriya, Celeste Antoni, Abigail Apelbaum, Pascal Hagel, Oliver Sitsel, Tobias Raisch, Dominik Prumbaum, Dennis Quentin, Daniela Roderer, Sandra Tacke, Benedikt Siebolds, Elke Schubert, Tanweer R. Shaikh, Peter Lill, and Stephan Raunser. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications Biology*, 2(1):218, 2019.

[60] Lanhui Wang, Yoel Shkolnisky, and Amit Singer. A fourier-based approach for iterative 3d reconstruction from cryo-em images. 07 2013.

[61] Xiao Wang, Eman Alnabati, Tunde W. Aderinwale, Sai Raghavendra Maddhuri Venkata Subramaniya, Genki Terashi, and Daisuke Kihara. Emap2sec+: Detecting protein and dna/rna structures in cryo-em maps of intermediate resolution using deep learning. *bioRxiv*, 2020.

[62] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science*, 367:1260–1263, 2020.

[63] K. Zhang. Gctf: real-time ctf determination and correction. *Journal of Structural Biology*, 193:1–12, 2016.

[64] Shawn Q Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A Verba, Yifan Cheng, and David A Agard. Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332, April 2017.

[65] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature Methods*, 18(2):176–185, February 2021.

[66] Ellen D. Zhong, Tristan Bepler, Joseph H. Davis, and Bonnie Berger. Reconstructing continuous distributions of 3d protein structure from cryo-em images, 2020.

[67] Z. Hong Zhou. Chapter 1 - atomic resolution cryo electron microscopy of macromolecular complexes. In Steven J. Ludtke and B.V. Venkataram

Prasad, editors, *Recent Advances in Electron Cryomicroscopy, Part B*, volume 82 of *Advances in Protein Chemistry and Structural Biology*, pages 1–35. Academic Press, 2011.

[68] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.

[69] Yanan Zhu, Qi Ouyang, and Youdong Mao. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics*, 18:348, 2017.

Esta página ha sido intencionalmente dejada en blanco.

List of Tables

List of Figures

Esta es la última página.
Compilado el Friday 17th October, 2025.
http://iie.fing.edu.uy/