Machine learning for determining continuous conformational transitions of biomolecular complexes from single-particle cryo-electron microscopy images

Ilyes Hamitouche

DOCTORAL THESIS OF SORBONNE UNIVERSITY

Doctoral school of Computer science, Communications, and Electronics (ED-130)

Presented by

ILYES HAMITOUCHE

To obtain

Doctoral degree of Sorbonne University

In the field of Bioimage informatics

Thesis subject:

Machine learning for determining continuous conformational transitions of biomolecular complexes from single-particle cryo-electron microscopy images

Presented and defended publicly on March 29th, 2023

In front of a jury composed of:

Dr. Charles Kervrann (INRIA, Rennes)	Reporter
Pr. José-Maria Carazo (CSIC, Madrid Espagne)	Reporter
Dr. Elodie Laine (Sorbonne Université, Paris)	Examiner
Dr. Denis Fortun (CNRS, Strasbourg)	Examiner
Dr. Irina Gutsche (CNRS, Grenoble)	Examiner
Dr. Slavica Jonic (CNRS, Paris)	Thesis director

Thesis prepared at IMPMC-UMR 7590 (Sorbonne University).

This work is dedicated to my family,
and friends

Acknowledgment

I would like to express my heartfelt gratitude to all who have supported me throughout my thesis journey. First and foremost, I would like to thank my family. In particular, my parents, my mother, Fattoum Saci, my father, Amar Hamitouche, my brother Abdelmounaim, and my lovely sisters Selma and Rayane.

I would like to express my sincere gratitude to my advisor Slavica Jonic for the continuous support of my Ph.D. study and research and for her patience, motivation, and enthusiasm. Her guidance helped me in all the research and writing this thesis.

I would like to acknowledge the funding provided by the ANR (project EMBioMolMov) for my salary, local computing resources, open-access publications, and participation in conferences; the EMBioMolMov financial management provided by the CNRS; and the fruitful discussions with the collaborators on this project, namely Dr. Bruno Klaholz (IGBMC, Illkirch), Dr. Isabelle Rouiller (The University of Melbourne, Australia), and Florence Tama (Nagoya University, Japan). Also, I would like to acknowledge the CPU/GPU computing hours on French supercomputers granted by GENCI for the duration of this thesis, which allowed the production of the majority of results that are presented in this thesis manuscript.

I thank my fellow labmates, Mohamad Harastani and Rémi Vuillemot, for the stimulating discussions, the collaboration, and all the fun we have had in the last three years.

Last but not least, to my friends who went through hard times together, cheered me on, and celebrated each other accomplishments. Ali Hani Abid, Malik Louail, Mehdi Azil, Tarek Berrahal, Nassim Guergouri and Yasser Hamidullah.

Thank you again for being a part of this journey and for your continued support.

Truly, Ilyes

Abstract

Structural biology aims at determining and analyzing the three-dimensional (3D) architecture of macromolecular assemblies to elucidate and study their working mechanisms. Throughout the past decades, nuclear magnetic resonance (NMR) and X-ray crystallography have been standard approaches in structural biology for macromolecular structure determination. More recently, electron microscopy of cryogenically-cooled purified samples, each sample containing multiple copies of the molecular complex under study (referred to as particles), emerged as the mainstream technique for the 3D structure determination, and it is more commonly known as single-particle cryo electron microscopy (cryo-EM). Single-particle cryo-EM allows solving the 3D structure of macromolecular complexes in their native states from noisy two-dimensional parallel-beam projections of the particles imaged at cryogenic temperatures in unknown random orientations and positions in a thin layer of vitreous ice. Ideally, all the complexes in the sample have the same structure, which leads to a single high-resolution 3D reconstruction from the images. However, macromolecular complexes have flexible structures, and the motion of the complexes (conformational changes) are directly linked to their biological functions (they need to change the conformation to accomplish various biological functions).

In the past decade, the study of conformational variability of biomolecular complexes has gained attention, and researchers used several approaches to address this challenge, which can generally be classified into two groups, namely discrete and continuous conformational heterogeneity methods. Before the start of this thesis, the majority of methods were better suitable for analyzing discrete conformational heterogeneity (a few distinct conformations) than continuous conformational heterogeneity (gradual conformational changes with multiple intermediate conformational states). With the emergence and advances in artificial intelligence, the development of methods for conformational variability analysis has shifted towards data-driven deep learning approaches, which can be better suited for analyzing continuous conformational heterogeneity than the classical cryo-EM approaches.

During my thesis work, I have developed three methods based on deep learning (two supervised and one unsupervised method) to address continuous conformational variability from single-particle cryo-EM images. The following three methods are described in this thesis manuscript, together with their results on test data: DeepHEMNMA[24,25] (supervised), supervised Cryo-VIT, and unsupervised Cryo-VIT. DeepHEMNMA is a fast conformational-space determination method that uses a convolutional neural network to speed up a previously developed method for continuous conformational analysis, HEMNMA[9,26], which combines a simulation of motion computed by Normal Mode Analysis (NMA) with an image processing approach. The convolutional neural network of DeepHEMNMA uses the rigid-body and elastic parameters obtained by HEMNMA (three Euler angles, two shifts, and a small number of normal-mode amplitudes) and learns in a supervised fashion how to map the images to these parameters. In contrast to DeepHEMNMA, the supervised and unsupervised Cryo-ViT approaches learn how to map each single particle image to a large number of atomic coordinates using a variational autoencoder.

Using synthetic and experimental datasets, I demonstrate the effectiveness of DeepHEMNMA in analyzing a strong continuous conformational heterogeneity. Also, I show the encouraging results obtained with both supervised and unsupervised cryo-ViT approaches. Based on these experiments, DeepHEMNMA is expected to be useful for conformational studies of various biomolecular complexes, whereas cryo-ViT approaches need more in-depth tests and, potentially, improvements. The software of DeepHEMNMA is publicly available as part of the cryo-EM data processing pipeline of the open-source software package ContinuousFlex, which is also available as a plugin of Scipion software, extensively used in the cryo-EM field, and uses Scipion's backend software Xmipp. The software of cryo-ViT methods will be available via ContinuousFlex upon the validation with more data.

Keywords: Cryo-EM, Continuous Conformational Variability Analysis, Deep learning, HEMNMA, DeepHEMNMA, Cryo-ViT, ContinuousFlex, Scipion and Xmipp.

Table of contents

Preface

The core focus of this thesis is developing deep learning approaches for continuous conformational variability analysis from single particle cryo-electron microscopy (cryo-EM). The seeds of this thesis were first planted when I was introduced through a seminar to electron microscopy technology and the challenges of its data in developing different analysis approaches. After two master's programs in image processing and machine learning, it was about time to enter an interdisciplinary field, particularly computational biology, where deep learning could play a massive role in pushing the growth of cryo-EM, which led me to express my interest in Dr. Jonic Ph.D. proposal.

My thesis began in December 2019 with the overall goal to develop supervised and unsupervised deep learning approaches to investigate continuous conformational variability of a variety of biomolecular. During the first part of my thesis, my main work was to accelerate a method developed in Dr. Jonic's group called HEMNMA using a supervised deep neural network. During the second part, the focus shifted to developing an unsupervised, fully neural network-based approach for the same purpose, in which minimum human interference is required.

This thesis provides an introduction to cryo-EM technology, a detailed state of the art of conformational variability analysis approaches, and a full description of the methods for the conformational variability analysis developed during this thesis as well as their performance analysis with synthetic and experimental data.

Introduction

The study of macromolecular assemblies' biological functions, such as DNA replication and protein synthesis, is key to novel drug discovery. Structural biology aims to determine and analyze the three-dimensional architecture of macromolecular assemblies to elucidate and study their working mechanism. Throughout the past decades, nuclear magnetic resonance (NMR) and X-ray crystallography have been the two reigning approaches for macromolecular structure determination. Recently, electron microscopy of vitrified cryogenic samples of biomolecular complexes, known as single-particle cryo-electron microscopy, emerged as a major 3D structure determination technique. During my thesis, I worked on developing machine-learning approaches for analyzing the continuous conformational variability of macromolecular complexes using single-particle cryo-EM.

Single particle cryo-EM allows imaging of biomolecular complexes using an electron microscope. During imaging using an electron microscope, the electrons are directed into a biological sample containing copies of the same biomolecule encapsulated in a thin layer of vitreous ice, which helps to maintain the sample in its native state. When the electrons interact with the specimen, they are scattered and then captured with a detector to create an image of the biological sample. The collected images, also called micrographs, are then processed through a set of sophisticated computational techniques, commonly known as Single Particle Analysis (SPA), to provide insight into the structure of protein complexes.

However, when a biomolecular complex undergoes continuous conformational flexibility (gradual changes of the conformation with many intermediate conformational states), the classical SPA workflow will likely not capture this variability, especially in the case of large amplitudes of the conformational change, and the 3D reconstructed structure will be less resolved in the parts where the amplitude of the conformational change is the largest. The conformational heterogeneity problem is currently widely recognized as the main computational challenge in cryo-EM, and sophisticated computational and statistical approaches are needed to disentangle the mixed population of images into a set of reconstructions representing individual unique conformations.

Throughout the last decade, deep learning pushed state-of-the-art research in different interdisciplinary fields and was extensively used in automating different steps in the SPA workflow,

such as automatic particle picking, 2D classification, and image denoising. However, deep learning for conformational variability analysis was proposed only recently, in 2019[27].

In this thesis, three cryo-EM data processing methods have been developed to tackle continuous conformational variability challenge, two supervised methods, DeepHEMNMA and supervised cryo-ViT (supervised), and one unsupervised method, unsupervised cryo-ViT. DeepHEMNMA is a hybrid method that combines a convolutional neural network with HEMNMA [9,26] . HEMNMA uses Normal Mode Analysis (NMA) and an image registration method to extract the hidden motion from single-particle images. DeepHEMNMA is a supervised deep learning method that uses the output of HEMNMA to learn how to simultaneously map an image to three Euler angles, two shifts, and M normal mode amplitudes, where M is the number of normal modes used by HEMNMA. DeepHEMNMA depends on the output of HEMNMA, which, in turn, depends on the simulated motion by NMA. Supervised cryo-ViT is an autoencoder neural network that learns the relationship between single particle images and their corresponding atomic coordinates provided by a newly developed method called MDSPACE[28].

The unsupervised cryoViT is a variational autoencoder that estimates the corresponding displaced 3D atomic coordinates for every single-particle image in an unsupervised fashion. In this approach, the conformational variability analysis is data-driven but requires an initial atomic 3D model (the atomic displacements with respect to this initial model are estimated for each single particle image). Unsupervised cryo-ViT refines the given rigid-body parameters while learning how to map the images to the corresponding 3D atomic coordinates

DeepHEMNMA was validated on simulated data and furtherly tested on a publicly available experimental dataset of the yeast ribosome 80S. Supervised cryo-ViT showed encouraging results on a synthetic dataset, and more extensive experiments are needed to demonstrate its effectiveness. Unsupervised cryo-ViT was validated on synthetic and experimental datasets of the yeast ribosome 80S and a publicly available dataset of transmembrane protein TMEM16F, and it is currently undergoing in-depth tests to endorse its effectiveness on more diverse datasets. DeepHEMNMA software is publicly available as part of the cryo-EM data processing pipeline of the open-source software package ContinuousFlex, which is also available as a plugin of Scipion software, and its backend is based on Xmipp software.

This thesis manuscript is organized as follows:

Chapter 1 reviews cryo electron microscopy principles and data collection challenges. Also, it presents the principles of the SPA workflow and its steps for biomolecular structure determination.

Chapter 2 offers a detailed review of the literature on biomolecular conformational variability analysis methods in single particle cryo-EM other than those developed in this thesis.

Chapter 3 describes the building bricks of the proposed DeepHEMNMA and supervised and unsupervised cryo-ViT methods, namely, NMA, convolutional neural networks, variational autoencoders, vision transformers, and quaternion representation.

Chapter 4 describes in detail the proposed supervised DeepHEMNMA approach and its results obtained with synthetic and experimental datasets. This chapter ends with a description of supervised cryo-ViT and presents preliminary results on synthetic dataset.

Chapter 5 presents the methodologies of the unsupervised neural network, unsupervised cryo-ViT, and its results obtained with synthetic and experimental datasets.

Chapter 6 highlights the software contributions made during this thesis.

This thesis concludes with discussions of the work done and possible future works drawn in Chapter 7.

Chapter 1: Principles of cryogenic electron microscope

At an early stage of biological sciences, the morphology of animals, plants, and tissues has been studied using minimal visual faculties, e.g., the naked eye. With the development of light microscopes, dating back to the seventeenth century [29], researchers could examine considerably smaller objects or microorganisms, such as cells and bacteria.

In a light microscope, shown in Figure 1 (left), a beam of photons is passed through a sample, which will be absorbed and remitted by the sample. Some of these remitted photons are then captured by the microscope's objective lens and brought to a focus at a particular distance, known as the focal length, to create a magnified image. However, the light microscope is limited by the wavelength of light of 400-700 nanometers, limiting the resolution to about 200 nanometers, which restrains to the observation of smaller structures within cells and tissues.

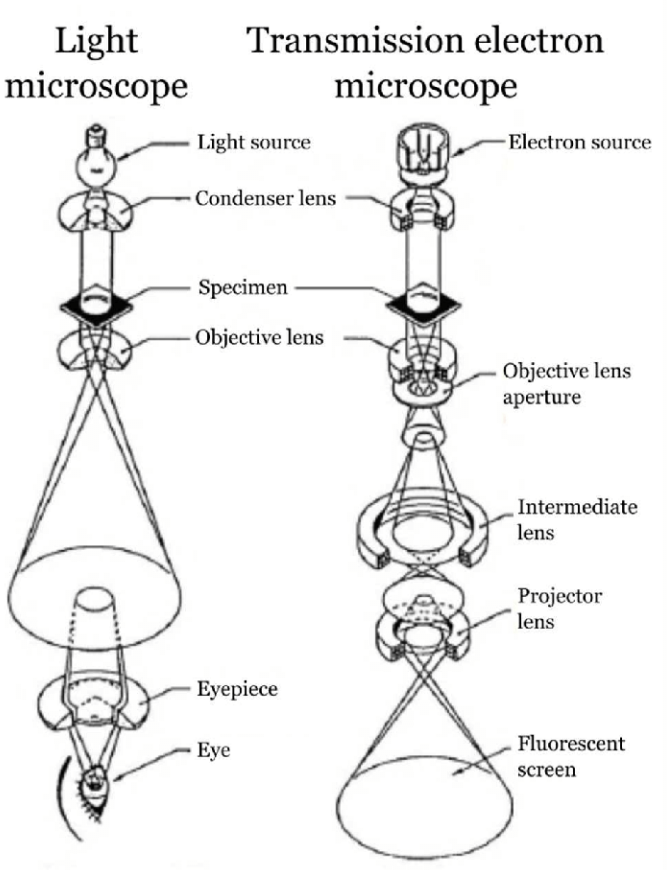


Figure 1 Schematic comparison of imaging instruments of light microscope(left) and electron microscope (right). Adapted from https://www.jeol.com.

To overcome the resolution limit, highlighted in Figure 2, the development of the electron microscope, shown in Figure 1 (right), was a significant advancement in the field of microscopy [30].

Instead of photons, the electron microscope uses a beam of accelerated electrons as a source of illumination to create an image of a sample, in which the wavelength of the electron beam is 10^{-5} shorter than the wavelength of the photons beam which allows to see structures at the atomic level[31].

In 1925, Louis de Broglie first proposed that electrons had wave-like properties with a wavelength much smaller than visible light. This theory was confirmed in 1927 by two independent experiments by Davisson and Germer and Thomson and Reid teams that showed the wave nature of electrons[3]. Soon after, Max Knoll and Ernst Ruska at the Berlin Technische Hochschule introduced the electron microscope in 1932[30], overcoming the limitations of visible light for higher resolution. The ultimate goal was to see atoms, which was accomplished gradually over time. The first electron microscopes only showed that electron beams could produce visible images of matter, and by the late 1930s, 10 nm resolution was achieved. This was further reduced to 2 nm by 1944, compared to the 200 nm resolution of optical light microscopes.

One of the main challenges in early electron microscopy was to increase the contrast of the imaged samples, and one idea was to stain and stabilizes the biological sample with heavy metals since they strongly scatter electrons and produce high amplitude contrast. However, sample staining involves the dehydration of biological specimens, which fundamentally removes them from their native state. In 1981, Dubochet and colleagues introduced cryogenic preservation[32] of the sample to capture its inner structure without damaging the biological sample's structural information, leading to cryo electron microscopy.

Cryo electron microscopy[33] is an electron microscope technique used to study biological samples at very low temperatures, typically around -190°C, where a biological sample is rapidly frozen in a thin layer of vitreous ice to preserve it in a near-native state to analyse the molecular structure and conformational changes without the need for staining or dehydration that can alter the sample's structure, preventing interpretation of the results of 3D reconstruction from the collected images with new instrumental and sample preparation advances. Nowadays, cryo-EM is a powerful technique for studying the structure and function of proteins and other biomolecular assemblies and has contributed significantly to our understanding of cellular processes.

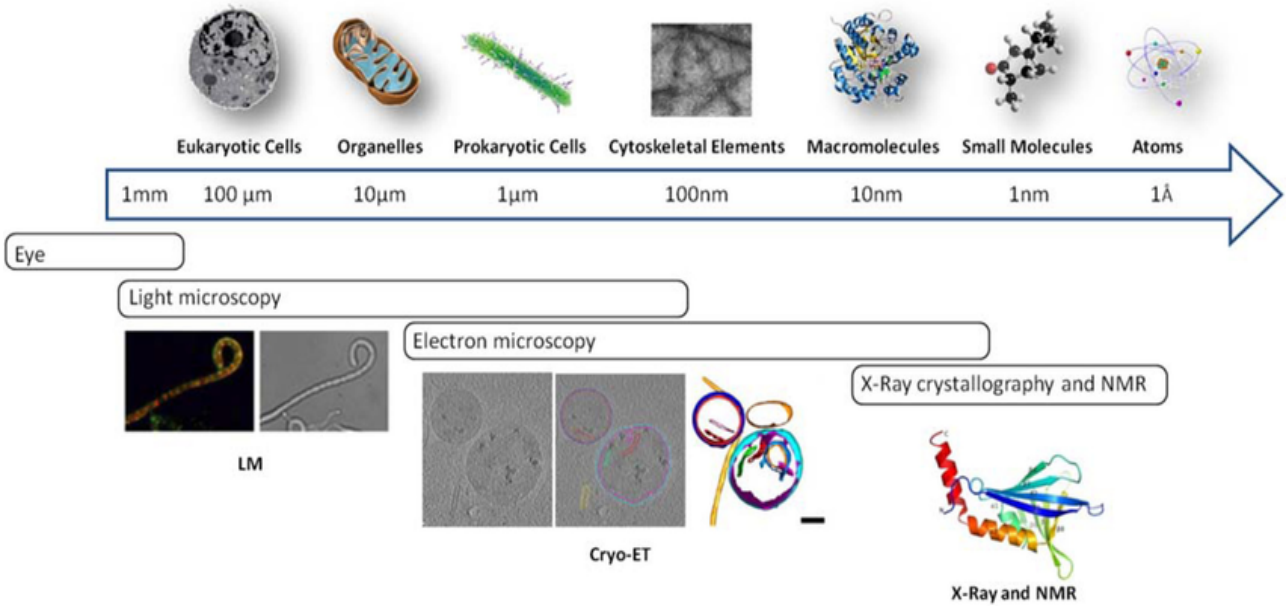


Figure 2 Resolution spectrum demonstrating imaging techniques which can be used at different scales, including light and electron microscopies. Adapted from [10].

Cryo electron microscope is composed of the following components:

- Electron gun: High-energy source of electrons emits electrons that travel down through the vacuum toward the sample and generates contrasts in the plane image.
- Electron lenses: A set of lenses are used to focus and shape the beam of electrons as it travels through the microscope. The first lens, the condenser lens, is located between the electron gun and the objective lens and is responsible for focusing the beam of electrons onto the sample. The second lens is the objective lens which is the closest lens to the specimen; it is used to focus the beam and produce a highly magnified and detailed image of the sample. Another essential lens is the projector lens which is the lens located furthest from the sample and is responsible for projecting an image of the sample onto the detector
- Specimen holder: It holds the sample in place during imaging.
- Detector: It captures the image of the sample, and its recent improvement [34] is considered the largest contributing factor of resolution.

1. Image formation and contrast enhancement

In cryo-EM, image formation involves the interaction of a beam of electrons with a sample, typically frozen in a thin layer of vitreous ice. The sample is placed in the path of the electron beam,

which is focused and aligned using a set of lenses. When the electrons interact with the sample, they are transmitted through the sample, scattered (elastically or inelastically), or absorbed by the sample [3]. The transmitted and scattered electrons, as shown in Figure 3, are then detected by a detector, such as a charge-coupled device camera (CCD) where electrons are first converted into photons before detection [35], or a Direct electron detector (DDD) [36], and the resulting image is recorded. The contrast of the image pixels are related to the thickness of the ice and composition of the sample. Thick films or opaque materials generally have higher contrast in cryo-EM images compared to thin films or transparent materials, as well as the biomolecular samples that are mainly composed of light atoms.

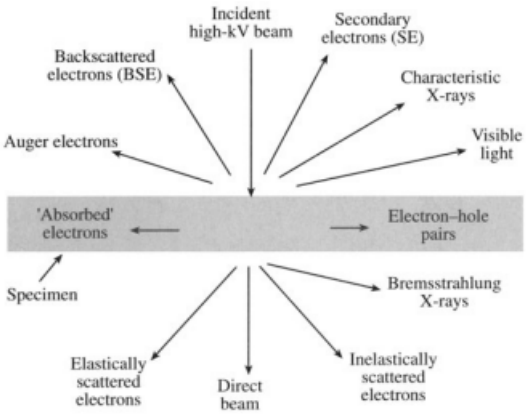


Figure 3 Schematic representation of electrons scattering phenomenon during interaction with the sample. Adapted from [3]

In TEM, two different contrast types arise, amplitude contrast and phase contrast [33,37]. One can think of the electrons as either a single particle with a condensed charge (inducing amplitude contrast) or a plane wave (inducing phase contrast). The amplitude contrast is related to the loss of electrons as they pass through the sample and the lenses. During the interaction of electrons with the matter, the low-angle scattered electrons are focused with the objective lens to produce a contrast. However, the high-angle scattered electrons are blocked by the objective lens aperture, leading to low contrast due to the small amount of electrons arriving at the image plane. Therefore, small aperture leads to a lower the Signal to Noise Ratio (SNR) in the final image. In the microscope, the electron dose is kept to a minimum (usually, 10 electrons per $Å^2$ for single particle analysis data collection) in order to avoid the radiation damage of the sample.

The phase contrast comes from a phase shift between different parts of the beam, producing the interference between different waves. The phase shift of the electrons is a function of the electron's

wavelength and the electron density of the specimen. When the electrons pass through regions of the specimen with different electron densities, they will experience different phase shifts, which can be used to generate contrast in the image. There are several ways to implement phase contrast in cryo-EM. One common method is using a phase plate, a thin film placed in the electron beam that introduces a controlled phase shift to the electrons. Another method is to use the microscope's contrast transfer function (CTF), which describes the phase shift of the electrons as a function of their frequency. It is worth mentioning that the CTF model takes into account the amplitude contrast; however, its contribution to the CTF model depends on the imaging conditions [38].

The image formation in an electron microscope can be described using the model[39] shown in Eq. (1), which relates the image pixel intensity to the properties of the sample and the imaging system. Mathematically, the image formation process in the microscope can be represented by a point spread function (PSF) h (real-space version of the CTF) that convolves the ideal projection φ of the electron density potential function of an object, yielding the experimental TEM image i:

$$i(\mathbf{r}) = h(\boldsymbol{r}) * \big(\varphi(\boldsymbol{r}) + n_b(\boldsymbol{r})\big) + n_a(\boldsymbol{r}) \quad (1)$$

where $\boldsymbol{r}$ is a vector in $\mathbb{R}^2$ representing a real space point, n_b denotes the noise present in the projection image affected by the point spread function, and n_a denotes the noise added by the imaging system.

2. Contrast transfer function (CTF)

A visual representation of the CTF is shown in **(**Figure 4**)**, where the pattern in the frequency domain composed of bright and dark circles, also called Thon rings or diffraction rings, are produced by the interference of electron waves passing through different parts of the specimen. The CTF describes how the amplitude of the electron density at different spatial frequencies is affected by the microscope's properties, such as defocus and astigmatism [40].

To determine the CTF (*H*), the power spectrum density (PSD) of the image is first calculated [41] and then fitted using the following theoretical PSD model (assuming the image formation in Eq.(1)):

$$PSD(\boldsymbol{R}) = K^2|H(\boldsymbol{R})|^2 + PSD_{na}(\boldsymbol{R}) \quad (2)$$

Where $\boldsymbol{R} \in \mathbb{R}^{\mathbf{2}}$ is a given spatial frequency, H is the CTF (i.e., Fourier transform of the PSF), and K^2 is the PSD of the noise before the CTF. In Eq.(2), it is assumed that the noise before the CTF is white and that the PSD of the ideal projection is negligible (the noise power is much more important than the signal power in a typical electron micrograph) [39].

The values of the CTF parameters can be estimated by fitting the PSD model in Eq.(2), and they are then used to perform the CTF correction. The following model of the CTF is often used:

$$H(\boldsymbol{R}) = sin\left(\pi|\Delta f(\boldsymbol{R})|\lambda|\boldsymbol{R}|^2 - \frac{\pi}{2}C_s\lambda^3|\boldsymbol{R}|^4\right) - \cos\left(\pi|\Delta f(\boldsymbol{R})|\lambda|\boldsymbol{R}|^2 - \frac{\pi}{2}C_s\lambda^3|\boldsymbol{R}|^4\right) \quad (3)$$

where $\Delta f(\boldsymbol{R})$ is defocus vector, λ is the electron wavelength, C_s is spherical aberration. It is worth mentioning that the second term is multiplied by a constant Q_0, which represents the fraction of scattered electrons at each frequency.

Once the CTF is estimated, several ways are available to correct for it. One method for CTF correction is called phase flipping[42], which consists of multiplying the phases of the Fourier coefficients by -1 at the spatial frequencies where the CTF is negative. This technique effectively reverses the phase of the image at those frequencies, resulting in an increase in contrast. Another method for CTF correction is called Wiener filtering [43,44]. It involves summing the multiplication of each image Fourier transform by its own CTF, and then dividing by the sum of all CTF's multiplied by the Wiener filter that is designed to minimize the noise in the image while preserving the high-frequency content [45]. The CTF-corrected image is given by:

$$A_\Omega(\boldsymbol{r}) = \frac{\sum_{i=1}^{M} CTF_i(\boldsymbol{r})I_i(\boldsymbol{r})}{\sum_{i=1}^{M} CTF_i^2(\boldsymbol{r})}\Omega(\boldsymbol{r}) \quad (4)$$

Where $\Omega(\boldsymbol{r})$ is the Wiener filter, I_i the i_{th} image, and M is the number of images in the dataset.

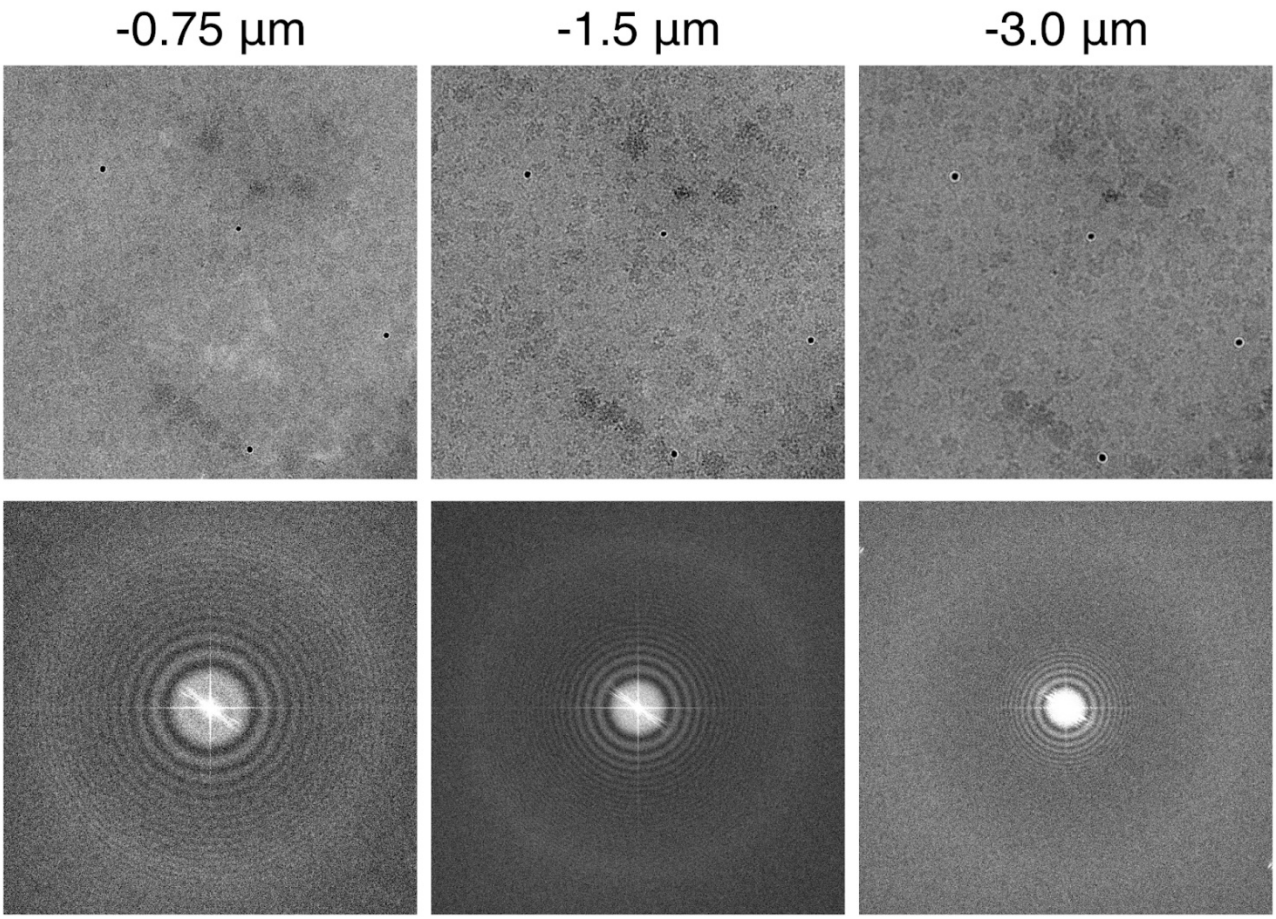


Figure 4 Effect of different defocus values on the resulting micrographs during data collection step in cryo-EM. Adapted from https://cryoem101.org/

The Wiener filter $\Omega(\boldsymbol{r})$ is defined as follows:

$$\Omega(\boldsymbol{r}) = \frac{1}{1 + \frac{1}{SNR_F(\boldsymbol{r})} \sum_{i=1}^{M} CTF_i^2(\boldsymbol{r})} \tag{5}$$

Where $SNR_F(\boldsymbol{r})$ is the ratio of particle signal power and the average noise power and is given by:

$$\mathrm{SNR_F(r)} = \frac{\mathrm{M|\varphi(r)|^2}}{\sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{|n_a(r)|^2}} \tag{6}$$

It is worth mentioning that the $SNR_F(\boldsymbol{r})$ was approximated by a constant in early works, and more advanced statistical approaches determine it as a function of the data[46].

Other approaches like iterative data refinement [47] and maximum entropy [48] can also be used to correct for the CTF.

3. Radiation damage and electron dose

The radiation damage phenomenon indicates the damage caused to the specimen by the beam of high-energy electrons emitted[49], as shown in Figure 5. When the beam of electrons hits the specimen, it can either break the chemical bonds or knock out electrons from the atoms. The first phenomenon, also called radiolysis, breaks the chemical bonds due to exposure to ionizing radiation, which can affect their structure and function (the loss of functional groups, changes in the conformation of proteins "formation of new bonds", and the formation of free radicals)[50]. The knock-out damage [51] is caused by the electrons interacting with the atoms, knocking out some of them, which can damage the sample's structural integrity, including the displacement of atoms from their native positions. Both radiolysis and knock-out damage can negatively impact the quality of the final cryo-EM image. Other pheromones can occur due to radiation damage that affect the quality of the acquired image.

The extent of the radiation damage caused by the beam of electrons depends on the electron dose, or the total number of electrons, used to produce the micrographs. The electron dose is expressed in electrons per square angstrom (e/Å2)[1], and the optimal electron dose for a particular sample will depend on emperical parameters, including the nature and size of the sample, the ice thickness, and the resolution at which the image is being taken. Higher electron doses are needed

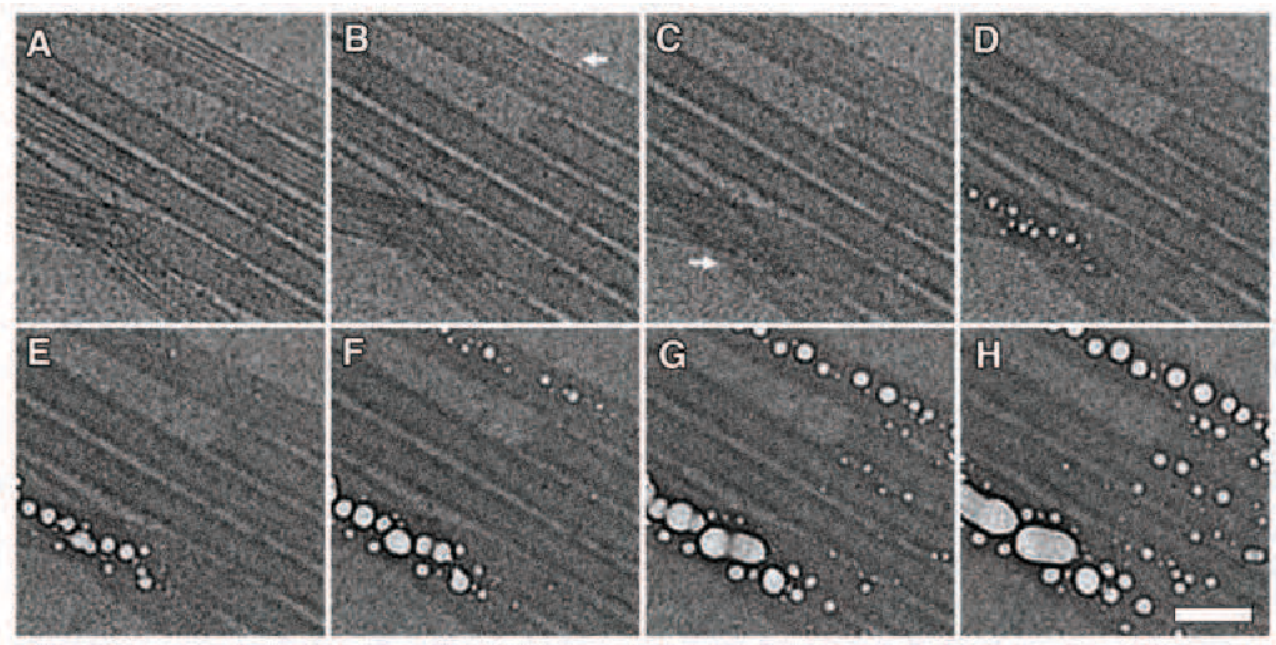


Figure 5 Radiation damage effect. The electron dose accumulation effect on the sample. Adapted from [1].

when dealing with thicker ice to produce good-quality images. Therefore, it is essential to find a good trade-off between the electron dose and the extent of the radiation damage.

Limiting electron exposure and using appropriate cryoprotection can effectively reduce radiation damage. However, low exposures result in under-sampled, noisy images and complicate obtaining 3D models of biological specimens from cryo-EM images.

4. Single particle cryo-EM for structure determination

SPA has emerged over the last decades as the mainstream processing technique that provides insight into the structure of protein complexes. In particular, the recent advances in sample preparation[52,53], computation[36,54-56], and instrumentation for SPA allowed for solving near-atomic-resolution structures of macromolecular complexes that resist crystallization. The principle of SPA is computational averaging of thousands of identical particles. Therefore, the collected single-particle dataset should be as homogenous as possible to obtain a near-atomic-resolution structure. The main steps of the SPA workflow are shown in Figure 6 and will be detailed in the coming sections.

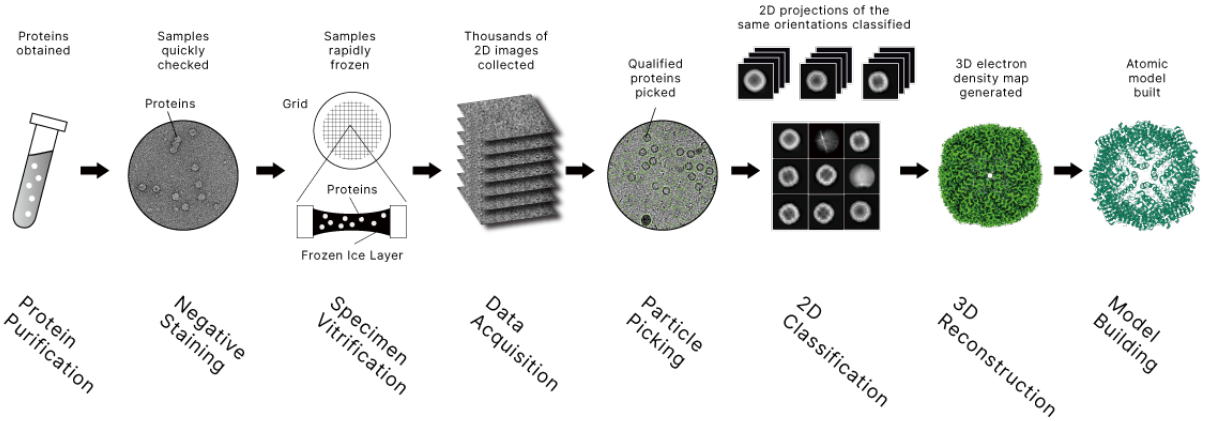


Figure 6 Main steps of the SPA workflow for structure determination. Adapted from https://shuimubio.com/services/cryo-em-spa

Specimen manipulation

For SPA data collection and analysis, the given specimen is purified to remove contaminants and impurities that can negatively affect the imaging of the specimen. In particular, the goal of the sample purification is to obtain a highly homogeneous and monodisperse sample[52]. Once the specimen is purified, it is applied to an EM grid to be imaged.

The grid with the sample on it is plunged into liquid ethane cooled by liquid nitrogen[53], a process called plunge-freezing, to cool the sample and vitrify it rapidly. The rapid cooling preserves the native structure of the sample without allowing ice crystals to form, which can damage the sample. Indeed, the sample needs to be fixed in a native state before imaging because the microscope is maintained in a high vacuum environment, which can cause hydrated biological specimens to dehydrate rapidly. The preservation in a vitrified state allows for good sample preservation. The high-resolution images obtained from the vitrified samples are then used to perform the rest of the computational SPA steps to produce the 3D structure of the biomolecules in question.

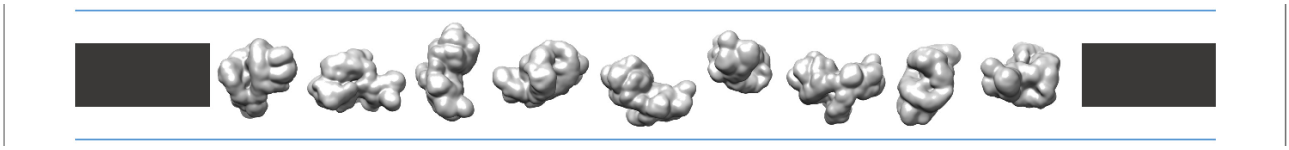


Figure 7 Schematic diagrams of copies of the same biomolecular complexes trapped in vitreous ice. Blue stripes represents the ice and the dark areas represents the carbon film Adapted from [21]

5. Single particle analysis for high-resolution structure determination

The data acquisition step in SPA has several shortcomings that make data processing challenging. The micrographs have low contrast and SNR due to the low electron dose used for data collection. Another challenge is that the individual particles in the collected micrographs are often heterogeneous, meaning they are in different orientations, locations, and conformations, which restrains to obtain high-resolution structural information, as the data from the different particles may not be directly comparable. To address this challenge, researchers typically use computational

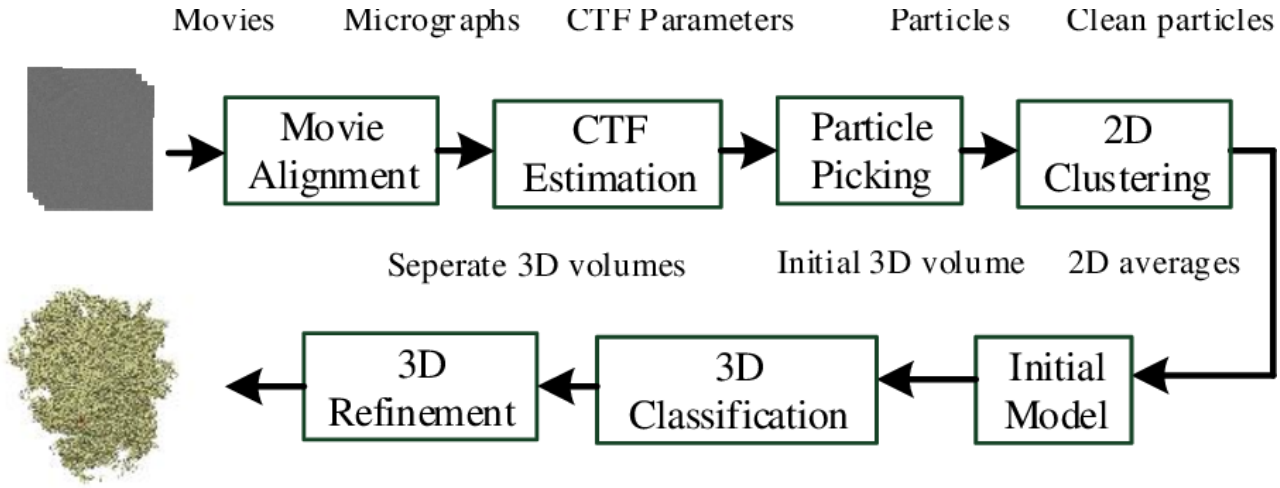


Figure 8 Overview of Single Particle Analysis (SPA) computational steps for structure determination. Adapted from [4]

techniques to align and average the data from multiple particles, ignoring slight conformational heterogeneity to obtain a 3D representation of the overall structure of the molecule. SPA is a computationally intensive process and requires specialized expertise; however, it has become increasingly feasible with the development of more advanced computational approaches and computing technologies. In a typical SPA workflow, several steps are needed to obtain the final high-resolution structure as illustrated in Figure 8.

Motion correction

The first step of a typical SPA workflow is to perform motion correction on the acquired micrographs. This step consists of correcting for the movement of the sample induced either by stage drift, where the specimen sample as a whole may move, or anisotropic deformation, where the sample may shift due to the electron beam intensity [57]. Motion correction [58-62] involves using computational algorithms to align the images of the sample in order to correct for any global, full frame motion correction, or local, per particle motion correction, movements that may have occurred during the data collection step. This can help improve the resolution and signal of the sample's structural information, which improves all the latter steps of SPA.

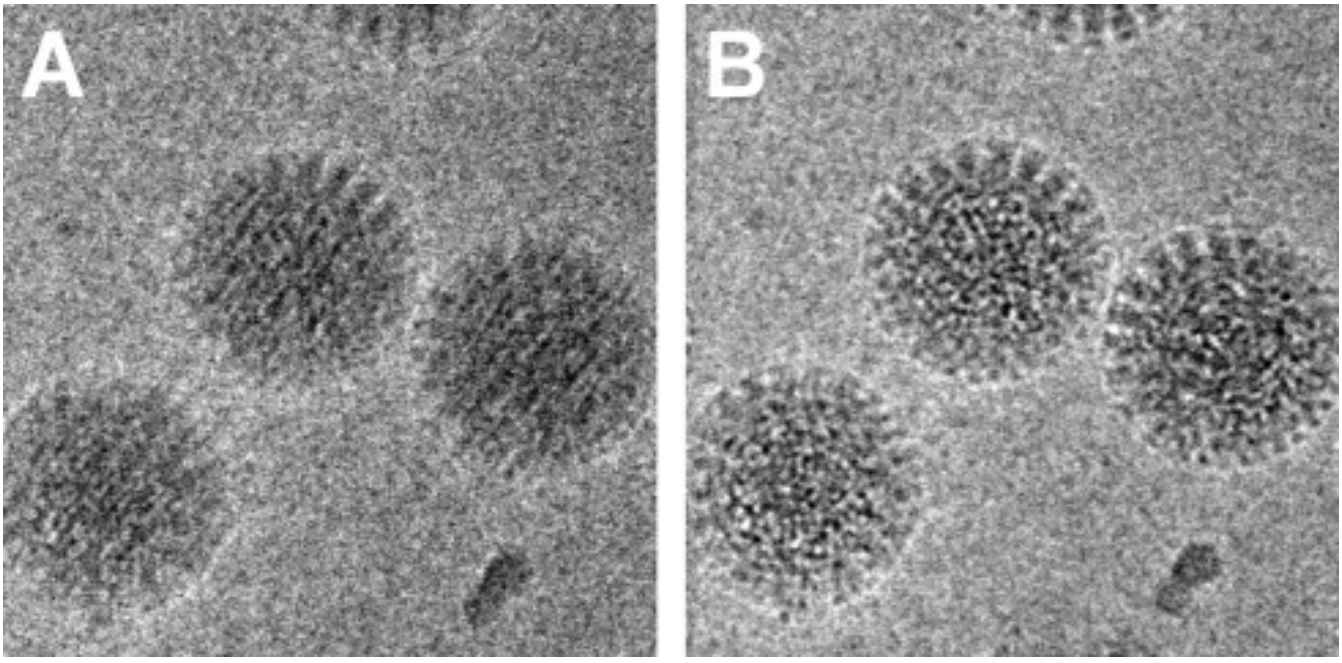


Figure 9 Motion correction and recovery of high-resolution information. Average of frames of rotavirus particles before (A) and after (B) motion correction. Features are blurred before the motion correction. Adapted from[22]

CTF estimation and correction

The next step of the SPA workflow is the CTF estimation. Accurate estimation of the CTF is critical for a near-atomic resolution cryo-EM reconstruction. The CTF affects the visibility and resolution of small details in the sample and should be corrected. The CTF estimation and correction methods were discussed in the beginning of this chapter.

Particle picking

In order to determine the 3D structure at high resolution, many projection images are needed. Furthermore, particles must be extracted from the collected micrographs into individual images for further processing [57]. In the beginning of single particle cryo-EM, extracting the particles was done manually, by an expert, exploring few micrographs one by one, and particle windowing (extracting image patches) from the micrograph regions that contain the particle as opposed to the regions that contain noise or contaminants. Throughout the last two decades, different techniques have been proposed to automate this time-consuming step.

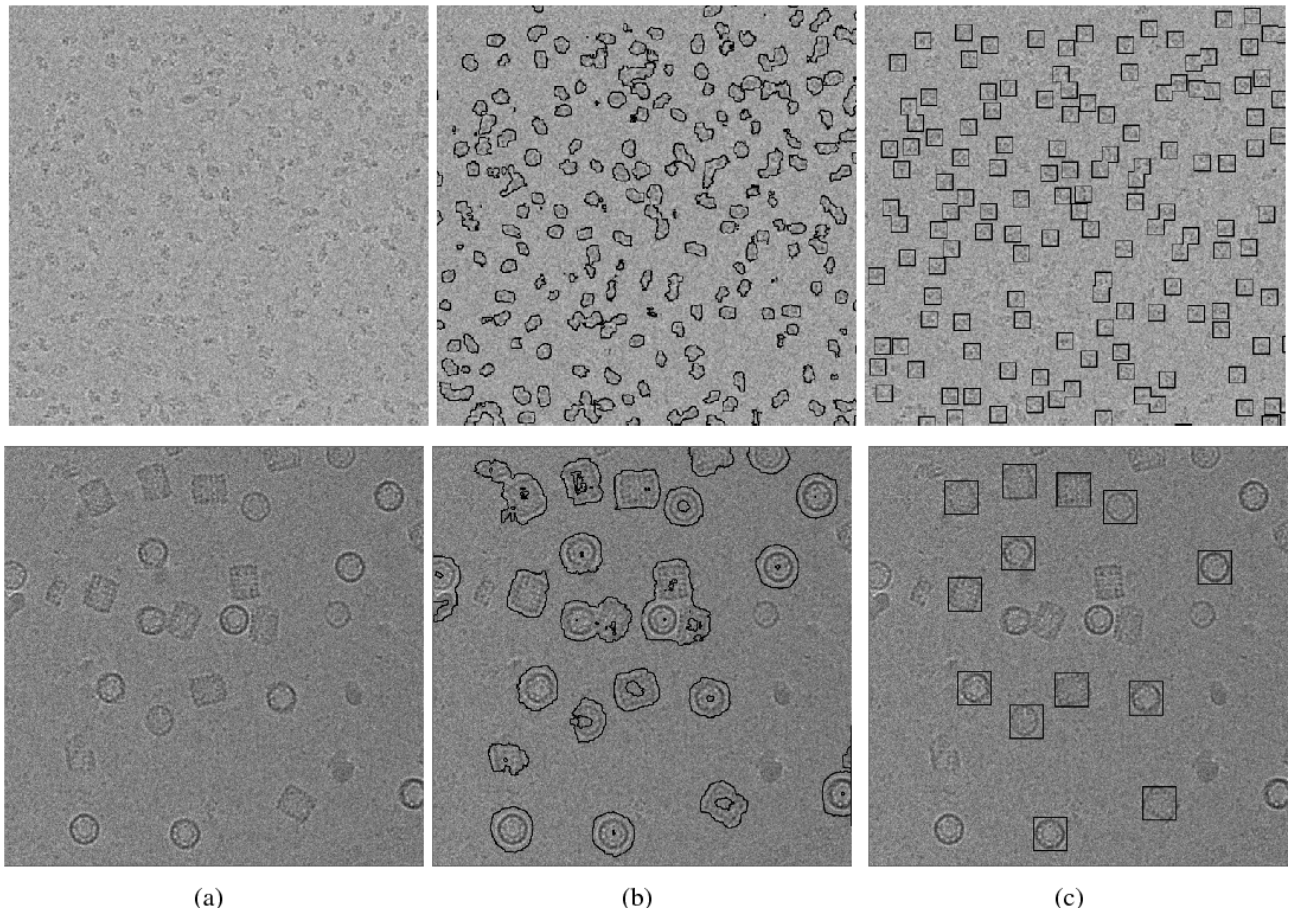


Figure 10 An example of an automatic particle picker (APPLE), Top row contains a β-Galactosidase micrograph. Bottom row contains a KLH micrograph. The left column contains plain micrographs without any selected region. The middle column contains the micrographs with the output of the object detection. The right column contains the micrographs with the picked particles. Adapter from[13]

There are several approaches to particle picking, including template-based matching [63], feature-based matching [64], and machine learning-based matching [13,65-69]. Template-based matching involves using a template (model of the particle) to identify similar particles in the images based on a metric score, usually the cross-correlation between the template and the micrograph. Feature-based matching involves extracting specific features, such as pixel intensity, edge detection, or texture features, from the images and using them to identify the particles. Deep learning-based methods involve training a neural network on a set of images to recognize the particles of interest.

2D classification

2D classification is often used in the later stages of single particle analysis, specifically in the alignment and averaging steps and in the model building and refinement steps. In 2D classification, the images of individual particles are divided into classes based on similarity, and a separate average image is generated for each class [70]. This can help improve the structural model's resolution and accuracy by separating out variations in the structure that may be present within the sample [71]. Specifically, it is often the step where orientation bias and structural heterogeneity are first identified, and if only a limited number of distinct 2D classes are present, it may indicate that the dataset lacks a wide variety of unique views. The clustering of similar particle images was first introduced by van Heel and Frank[72], when the correspondent analysis was used to classify images. Another famous approach is the maximum-likelihood 2D classification implemented in RELION [73]. These 2D class

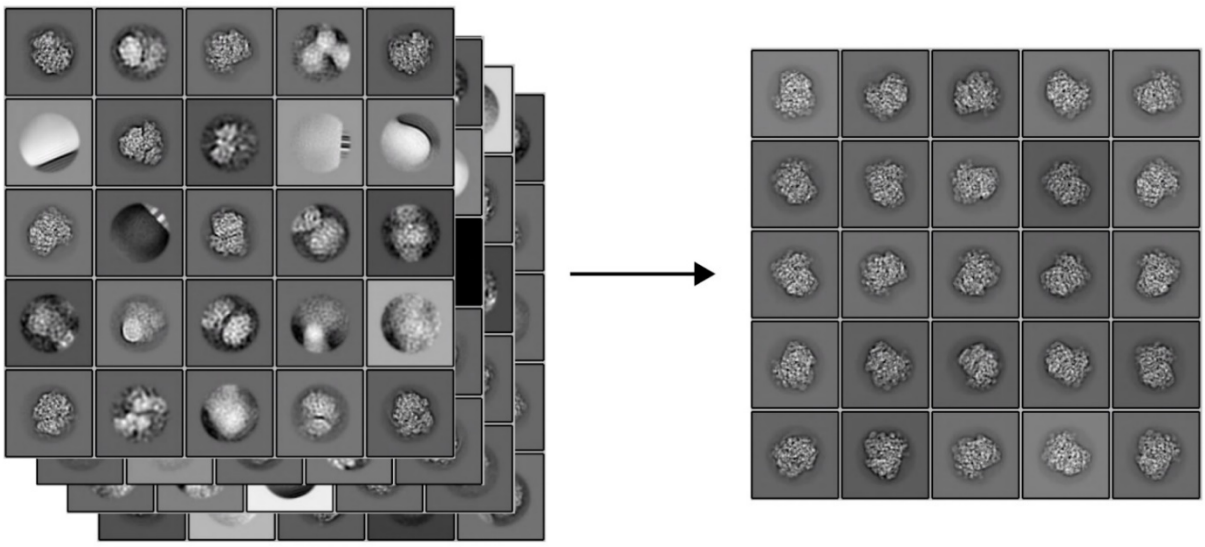


Figure 11 An example of 2D classification step. The output of 2D classification with a mixture of junk and well-resolved classes (left). The best classes extracted from the output of the 2D classification, keeping only well-resolved classes and discarding junk classes (right). Adapted from https://cryoem101.org/

averages can then be used to generate an initial structural (ab-initio) model of the protein or biomolecule. The initial structural model is refined using additional images and information in the model building and refinement step. 2D classification can be used to identify and correct any errors or inconsistencies in the model by comparing the model to the class averages and adjusting the model as needed.

In addition to evaluating the quality of the particle images, 2D classification also curates the stack of extracted particles. Well-resolved classes are separated, and a refined stack is created, while poorly resolved or "junk" classes are discarded. It is common at this stage to discard a significant number of particles, depending on the accuracy of the particle picking process. Multiple rounds of 2D classification can be performed to refine the class averages further and increase confidence in the downstream steps. The refined stack is then used as input for 3D classification and reconstruction.

3D classification

As with 2D classification, particles can also be classified in 3D to sort them among different structural states. This 3D classification is beneficial for distinguishing particles that exhibit compositional and conformational heterogeneity. A popular method for 3D classification is based on the maximum likelihood[71], where each particle image is assigned a probability to yield a particular 3D model. After multiple iterations of refinement, the assignment of each particle image to a particular 3D model is usually unambiguous, and separate reconstructions can be made from the separately assigned particle stacks. 3D classification can be useful for in-silico purification of the particles (removing the particles that are not the object of the study). However, it is not always accurate and often requires multiple classification rounds.

Once all steps of the SPA workflow are done, one or several final average 3D density maps are obtained. However, when a biomolecular complex is underlying some conformational flexibility, the classical SPA workflow will not capture this variability, and the final structure will be less resolved in the parts where the changes happen most. Chapter 2 reviews the conformational heterogeneity problem in-depth and provides a detailed state of the art of the existing techniques.

Chapter 2: State of the art

Studying biological functions of macromolecular complexes, such as DNA replication, transducing chemical and electrical signals, and protein synthesis, via structural studies of the complexes, is the key to novel drug discovery. The structure of the complexes is directly linked to their biological functions. Therefore, the determination of the structure and its flexibility allows the elucidation of the working mechanisms of the complexes. Indeed, biomolecular complexes are dynamic entities, and their conformational changes are linked to the multiple functions that they perform in cells. The conformational variability of the complexes generally results from continuous conformational changes, where the complex gradually transitions from one state to another while passing through many intermediate conformational states. Another form of conformational changes arises from the binding or unbinding of substrates and is usually referred to as discrete conformational changes, where the complex can take a finite set of conformational states. Continuous and discrete

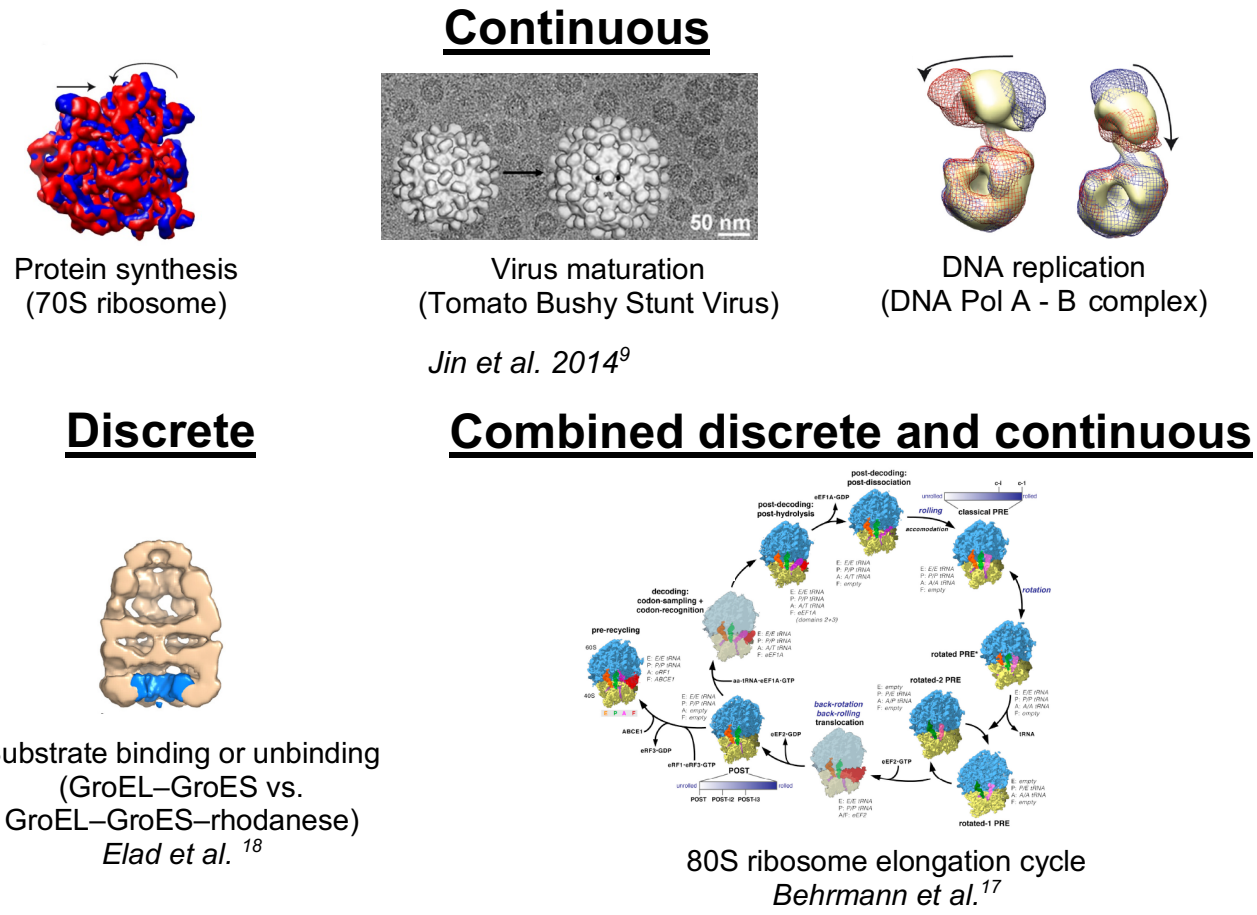


Figure 12 Types of conformational heterogeneity, Continuous conformational heterogeneity (top), Discrete conformational variability (bottom left), Combined discrete and continuous conformational heterogeneity (bottom right).

conformational variabilities may manifest simultaneously, where the main complex gradually changes its shape while a substrate is binding or unbinding.

In this chapter, we review the methods specialized in conformational heterogeneity analysis, which were available before the start of this thesis work (in 2019), and the more recently proposed methods. In the last decade, much progress has been made in addressing conformational variability and some of the most commonly used approaches for this are multivariate and subspace analysis, maximum likelihood classification, multi-reference classification, and deep learning approaches.

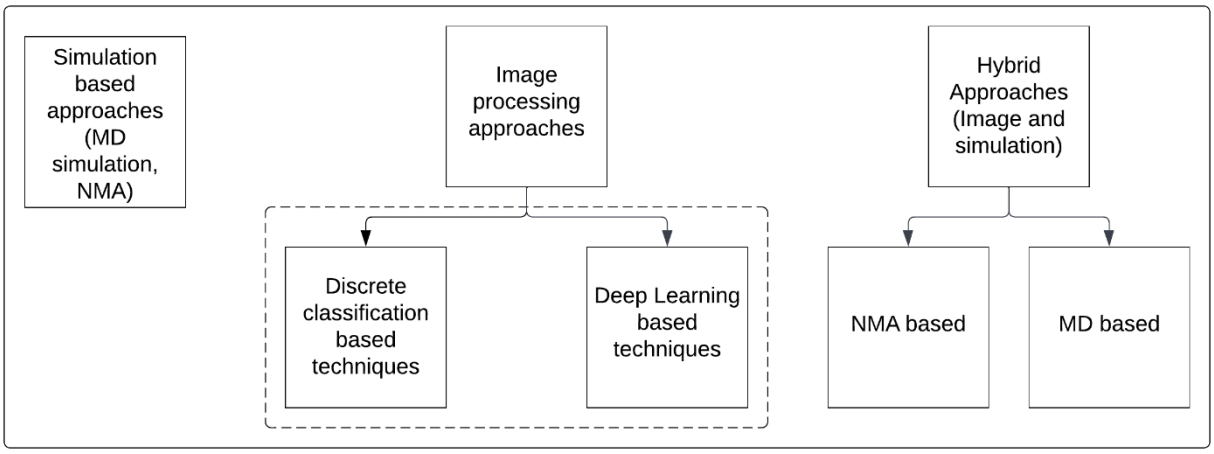


Figure 13 Approaches for conformational heterogeneity analysis available in the literature.

Multivariate analysis approaches use statistical techniques, such as principal component analysis (PCA) [74] and correspondent analysis, to identify patterns in the data and assign the particles to distinct classes accordingly. Maximum likelihood classification approaches align the particles, compare their structures to identify similarities and differences and assign the particles to distinct classes based on the likelihood that they belong to a particular class. Multi-reference classification approaches align the particles to a set of reference structures and assign the particles to distinct classes based on the similarity of their structures to the references.

Deep learning-based approaches use deep learning neural networks solely or in combination with other approaches for continuous conformational heterogeneity analysis.

Conformational heterogeneity analysis methods can be categorized in several manners, such as based on the type of heterogeneity for which the methods are better suited (continuous or discrete heterogeneity) or based on the dimension of the space used for the conformational variability analysis

(2D or 3D variability analysis). As the focus of this thesis was the development of deep learning approaches for continuous conformational heterogeneity analysis, we below review the conformational heterogeneity analysis methods categorized according to the most suitable type of heterogeneity (discrete or continuous), and detail some of the most famous approaches in the literature.

1. Discrete heterogeneity analysis methods

Discrete heterogeneity methods are a straightforward solution to the conformational heterogeneity problem. They group images into discrete clusters and minimize the intraclass variance within the groups, where each group contains images sharing structural similarities [75]. Discrete heterogeneity methods provide readily interpretable results, where each particle is assigned to a small number (K) of 3D models reflecting K states of the molecular complex. In the cases where the free-energy landscape has distinct local minima, discrete heterogeneity methods are suitable. However, the free-energy landscape is not known in advance, and the fundamental assumption when using discrete heterogeneity methods is that the selected number K is representative of the number of local energy minima on the free-energy landscape. In practice, K has an arbitrary value that is decided by the user based on some prior information about the complex and its potential conformational variability[76]. Often, such methods are run several times, using the same data with different values of K, and the results of the runs are compared.

Multi-Reference classification

In multi-reference classification, a set of reference structures is used to classify particles into different classes based on their similarity to the reference structures. Once a set of reference structures is defined, the similarity between each particle and the reference structures is usually quantified using a metric such as cross-correlation[77]. More precisely, the image of each particle is compared to the image of the reference structures projected in a predetermined projection direction. The reference structure with the highest cross-correlation value determines the class of the particle[78].

Multireference classification can be supervised[77,79], where the process is guided by a set of predefined references provided before the heterogeneity analysis. Here, the references structures are first projected along predefined projection directions estimated for each single particle image and

compared to all single particle to obtain the cross-correlation. Once the images are assigned to the reference structures, 3D reconstructions are produced for each class.

Similarly, unsupervised multireference classification[80-82] uses a first set of reference structures and iteratively assigns single particles to one of the cryo-EM maps; then, new references are generated. This process is repeated to improve the classes and the final cryo-EM maps.

Maximum likelihood approaches

Alternatively, the particles can be assigned to the different classes based on their similarity to the reference structures using Maximum likelihood (ML) methods. Maximum likelihood estimation (MLE) is a powerful statistical approach used to estimate a set of parameters of a probability distribution, e.g., mean and standard deviation values of a Gaussian distribution, that best describes a given observed dataset. In the context of conformational heterogeneity analysis in cryo-EM, MLE is a widely used approach to estimate the probability distribution of conformations in a population of molecules[83].

The idea behind MLE is to find a set of parameters that maximizes the likelihood of the data, given the model. In the case of conformational heterogeneity analysis, the data is the set of 2D noisy single-particle images, and the model is a probability distribution that describes the conformations of the molecules. The goal is to find the set of parameters that best describes the conformations of the molecules in the population, as determined by images [83,84].

MLE is implemented using the Expectation-Maximization (EM) [85] algorithm. EM is an iterative algorithm that estimates the model's parameters starting with an initial guess for them; it alternates between two steps: the expectation step (E-step) and the maximization step (M-step) until convergence. The modeling of the 3D structure of the biomolecular complex typically involves defining a probability density function (PDF) that describes the 3D structure of the particle, such as a Gaussian function or a more complex function like a Gaussian mixture model (GMM) or a hidden Markov model (HMM).

In cryo-EM, ML approaches were first introduced by Sigworth [86] for aligning a set of noisy simulated homogeneous single-particle images. Several contributions followed this work to analyze conformational variability. Pascual-Montano et. al [87] proposed an ML variant of the conventional K-means classifier. In this work, a kernel density self-organizing map was applied to a set of prealigned

single-particle images, and each image was assigned to a predefined multireference K 2D class. This approach was used to study large T-antigen and hexameric helicase G40P. To overcome the bias of aligning the 2D single-particle images prior to the conformational variability analysis, Scheres et al. [88] proposed a multireference 2D ML variant called ML2D. The approach models both the in-plane rigid-body displacements and the class assignments as hidden variables. The effectiveness of this approach was demonstrated on simulated datasets and led to the discovery of an overhanging dsDNA in the large T-antigen complex. With the limitations of addressing conformational heterogeneity at the 2D level, Scheres et al. [83] proposed an extension of ML2D to perform the analysis at the 3D level called ML3D. A 3D multireference refinement scheme was proposed as an EM-ML approach for single-particle analysis. This method simultaneously refines K 3D reference maps against a structurally diverse set of images. The parameters of this problem include class assignments and 3D orientations, as well as in-plane translations, for each image. Optimizing the log-likelihood function in this 3D refinement is more complex than in 2D[89]. The ML3D classification protocol separates images based on the projections of different 3D structures by starting from random variations of a single, low-pass filtered initial model, making it unsupervised and robust to noise. This method was successfully applied to two challenging cryo-EM data sets, separating projections of Escherichia coli ribosomes in different conformations and of large-T antigen in various states of bending. Other variants of ML, involving replacing EM step with stochastic gradient descent (SGD) or branch-and-bound were introduced [54,76] to speed up the processing.

ML methods have certain limitations that can impact their effectiveness in certain scenarios. One limitation is that the number of classes, K, needs to be provided beforehand. When this number is underestimated, some classes with merged features are formed, while when it is overestimated, unnecessary small classes are split off at significant computational expense. Additionally, as an iterative method without guaranteed convergence, it depends on the quality of the initial references. This can be challenging in specific applications and may require multiple rounds of processing in a stepwise, hierarchical scheme[84]. Assuming that the heterogeneity is discrete is a serious limitation of ML approaches as they do not address changes in structures that are by their nature continuous. Attempts to capture fine subdivisions in a continuum of states by specifying a large number of classes to account for the continuous changes, sampling a continuum with a finite number of classes will

produce arbitrary boundaries of meaningless classes, leading to a significant increase in computational expense, which becomes prohibitive when the number of classes exceeds 10-15.

2. Continuous conformational heterogeneity methods

The limitations encountered in discrete based approaches pushed towards more convenient approaches that consider the gradual changes that a biomolecular complex may underly during its work. Therefore, several approaches were proposed to address continuous conformational variability, and they are presented in the coming sections.

Manifold Learning

Manifold learning is a technique used to organize observations, in this case, particle images, based on a norm of the difference between the observations. The idea is to embed high-dimensional data into a lower-dimensional space while preserving specific properties of the data, such as the similarity between observations. This idea was first proposed to characterize continuous heterogeneity, as the space of particle images forms a manifold of images. However, this process presents several challenges, including the introduction of additional dimensions to the problem due to the variables such as viewing direction, in-plane rotations, and in-plane translations, as well as CTF and the noise in each image. The straightforward idea was to reduce the complexity of the problem by assuming the images are prealigned to address these challenges. Dashti et al. [90] demonstrate such an approach's effectiveness on the yeast ribosome. Several works have recently explored manifold embedding (ManifoldEM) [19,91-94]. Here, the conformational heterogeneity is analyzed by finding the conformational manifold along a projection direction. The resultant manifold is mapped to another coordinate system to construct the free energy landscape sampled by the system. At this level, the trajectories can be built to explore 2D movies. Then, the process is repeated along orthogonal projection directions to obtain the intrinsic dimensions that best depict the conformational manifold. Additionally, the manifolds are aligned and sidestepped to study the 3D conformational heterogeneity by allowing 3D reconstructions from different projection directions. One can use either a nonlinear

Laplacian spectral analysis or embedded subspace partitioning and eigenfunction realignment (ESPER).

Zernike polynomials have been exploited to produce a deformation field [95,96] that describes the conformational variability analysis. A given reference cryo-EM density map is formally deformed using a deformation field approximated by Zernike polynomials to obtain a confirmation. Then, this process is repeated, and the Zernike polynomials are computed for all possible pair combinations in a set of N cryo-EM density maps to form a distance matrix in this case, the cross-correlation. Then the cross-correlation matrix is embedded in a low-dimension space using multidimensional scaling (MDS)[97] to visualize the conformational space and produce 3D heterogeneous reconstructions and animation trajectories.

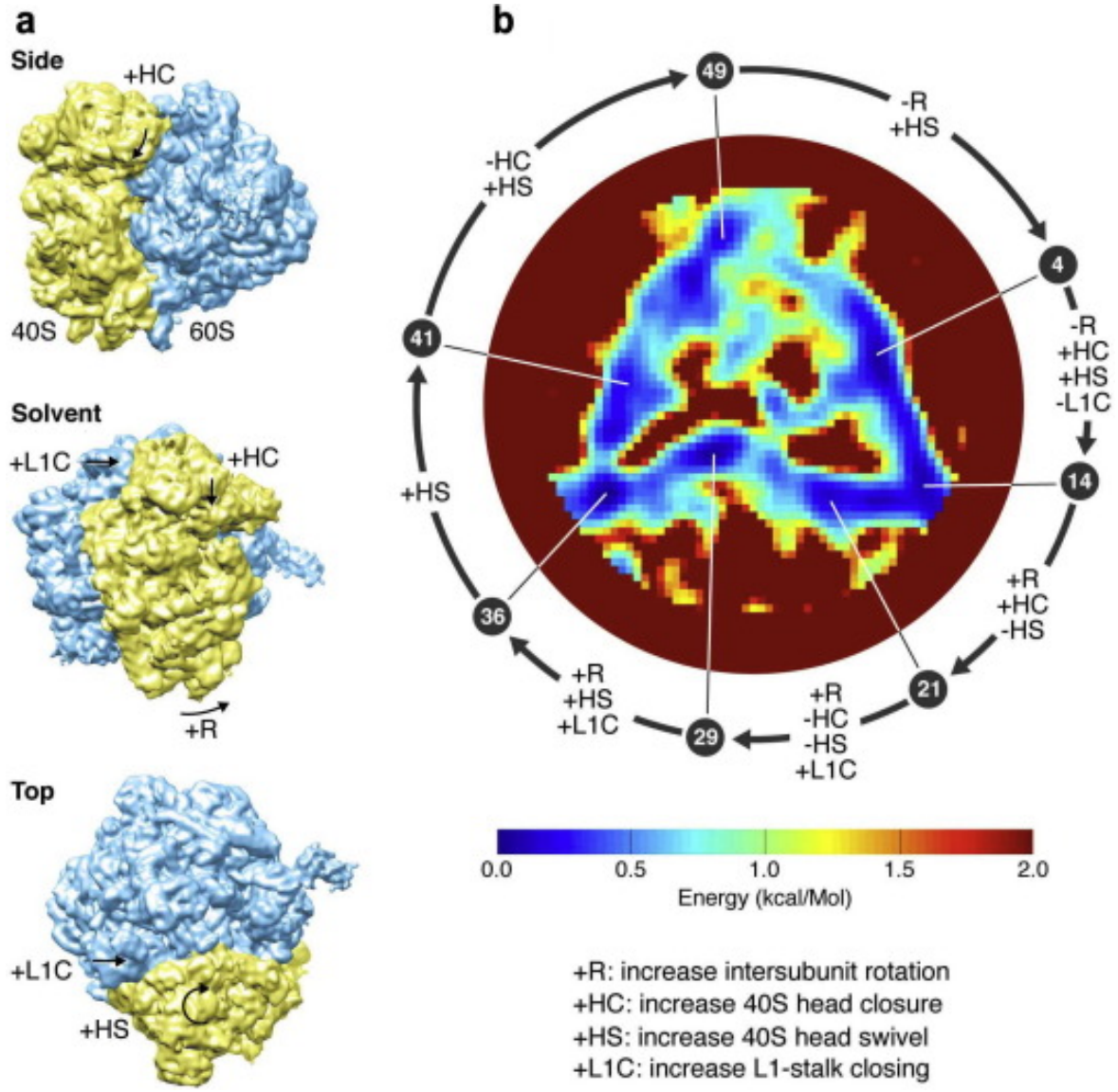


Figure 14 ***a*** *3D reconstructions representing the conformational variability of the ribosome 80S obtained using ManifoldEM.* ***b*** *The 2D Conformational variability and energy landscape of the ribosome 80S obtained. Adapted from* [19]

A fundamental limitation of ManifoldEM-based approaches is the prerequisite of known and accurate viewing directions for each single-particle image. For instance, significant conformational changes may result in imprecise rigid-body alignment parameters estimation, which will affect the variability analysis results.

Hybrid methods combining simulation and image processing

In this class of methods, NMA or molecular dynamics simulations are used to model the conformations through iterative fitting of models with the images [9,26,28]. The fitted conformational models (each corresponding to a single particle image) are then projected onto a low-dimensional space, which determines the conformational space. NMA is one of the first methods [98] that treated continuous conformational heterogeneity in cryo-EM. It is based on a linearized model for small changes around the equilibrium position of atoms in a reference structure and the principle that a macromolecule can be displaced from one conformation to another by applying a linear combination of normal modes vectors representing the motion directions. The mathematical principles of NMA will be detailed in Chapter 3. HEMNMA [9,26] uses a reference structure, either an atomic structure or an EM density map, as input and employs NMA to simulate the motion directions.

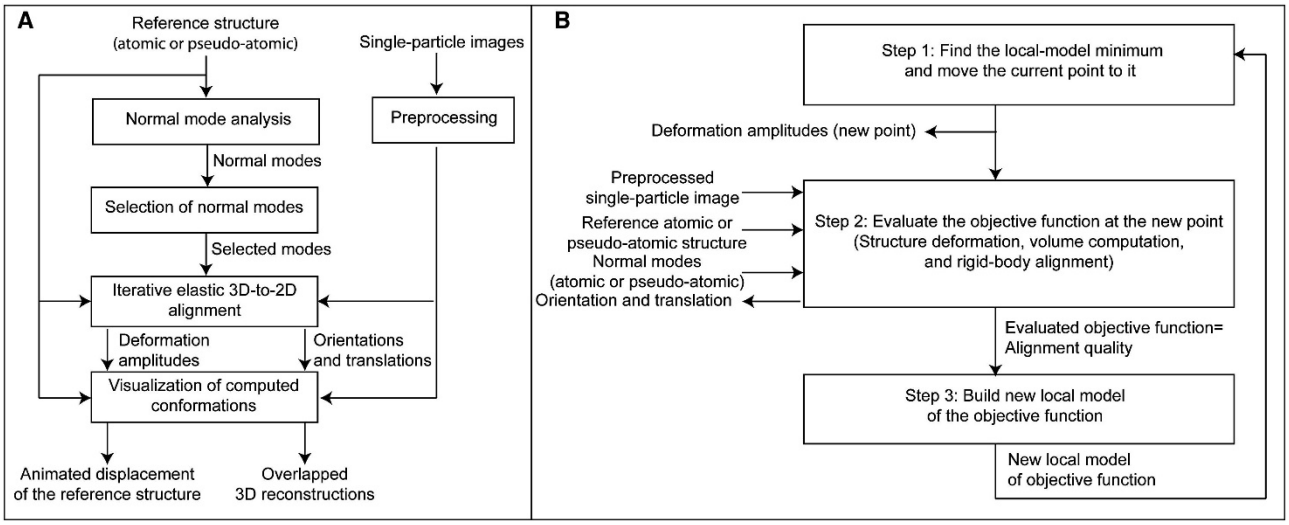


*Figure 15 Overview of the pipeline of HEMNMA for continuous conformational analysis. **A** Steps of HEMNMA, including NMA and 3D-to-2D iterative alignment. **B** Steps of the numerical optimizer used for 3D-to-2D iterative alignment. Adapted from [9].*

HEMNMA aims to determine a set of normal mode amplitudes producing the conformation that fits the given particle image. The method uses an iterative process independent for each particle image, which deforms the input structure using normal modes and a set of test normal-mode

amplitudes, converts the deformed structure into a density map, and compares 2D projections of that map with the given image, until the best matching conformation (normal-mode amplitudes) and the best matching 2D projection orientation and shift are found for this image.

After the conformational parameters are found for the image ensemble, they are fed to a dimensionality reduction method, such as PCA, to map images onto a low-dimensional space. This helps identify the most densely populated regions in the low-dimensional space, which correspond to the trajectories of conformational changes. By exploring these regions, it becomes possible to visualize the changes in 3D by displacing the reference structure along the trajectories. Additionally, density maps can be reconstructed along the trajectories from images with similar identified conformations.

Recently, the exploitation of the free energy paths to study conformational variability has gained attention. These computational methods provide a detailed quantitative description of the different energy minima a biomolecular complex can adopt, in which each minimum is considered a unique conformation, as well as the barriers that must be overcome for the molecule to transition between these conformations. By building a free energy landscape from the cryo-EM data, researchers can identify the different conformations that the particles adopt and quantify the relative population of each conformation. Giraldo-Barreto et al.[99] proposed a Bayesian approach for continuous conformational variability called Cryo-BIFE. This approach extracts the free energy profiles and their uncertainties from single-particle images and assign each particle to a single minima along the free energy path. Cryo-BIFE consists of three mains steps (1) a set of conformations is selected before the analysis; (2) a likelihood of the image given the conformations is calculated for each particle; (3) an energy profile is constructued by sampling from the likelihood computed in te previous step usin Markov chain Monte Carlo. A limitation of this technique is its dependency on the main conformations provided before the analysis and the accuracy of precomputed orientations and shifts.

Multivariate statistical analysis approaches

Multivariate statistical analysis approaches[100], also known as Subspace analysis approaches, exploit the idea of analyzing the data in a lower dimensional subspace to simplify the analysis while still preserving the information of the data. There are several different subspace analysis techniques, such as PCA, Correspondent analysis (CA) [101], and Non-negative Matrix Factorization (NMF)[102].

They are considered the first class of methods used for conformational heterogeneity, and they allow researchers to extract information about the different conformations a biomolecule can adopt and the transitions between them.

First approaches date back to the 1980s[72,103,104], where a set of single-particle images is compressed to a reduced size matrix using correspondent analysis[101]; this helps to accelerate the analysis procedure. Then, a distance matrix is used, such as χ^2 distance, to compute the distances between two rows of a symmetric matrix where the eigenvalues and eigenvectors are determined to investigate the variation in the data. Typically, the eigenvalues and eigenvectors reflect variations in the distribution between the different projections, and the analysis of eigenvectors patterns is used to cluster images into groups with characteristic features. MSA was used to study different biomolecular complexes' experimental datasets. Van heel et al. [105,106] analyzed the conformational heterogeneity of the 30S ribosomal subunits of Bacillus stearothermophilus. MSA approaches can be combined with clustering algorithms such as K-means [107] to facilitate the exploration of the low dimensional representation of the data.

Liu et al. [108] introduce 3D variance analysis, in which the studied volumes are obtained through homogeneous reconstruction using subsets of the particle images[109]. Covariance matrix estimation [110-112] have demonstrated that it is possible to estimate the 3D covariance matrix and principal volumes directly from 2D particle images. Probabilistic PCA approaches [113,114] has followed-up and improved covariance matrix estimation approaches for estimating the principal volumes, where the principal volumes are estimated directly from the data without computing the covariance matrix first, which allows for reconstruction of volumes at higher resolution.

This category of methods does not require any prior knowledge of conformational variability; it relies on the compressed eigenvectors-eigenvalues representation of the data to extract it. Although they can be used solely for conformational heterogeneity analysis, combining them with unsupervised clustering techniques may help explore the compressed representation more adequately to produce more stable conformations. However, they suffer from the aspect of known viewing directions for each single-particle image before conformational heterogeneity analysis.

Generative deep learning approaches

Throughout the last decade, deep learning pushed state-of-the-art research in different interdisciplinary fields and was extensively used in automating different steps in SPA workflow, as shown in Chapter 2, such as automatic particle picking, 2D classification, and image denoising [4,55]. However, conformational heterogeneity analysis waited until 2019 [27] for the first fully deep learning-based approach that addresses conformational variability, where an encoder-decoder neural network was used, leading to CryoDRGN [8]. CryoDRGN is an end-to-end unsupervised variational autoencoder (VAE) that consists of (1) an encoder that maps the input single particle images to a latent space representation by approximating the true posterior distribution given the input single particle images; and (2) a decoder that takes the representation and maps it back to a single particle image. VAEs will be explained in Chapter 3.

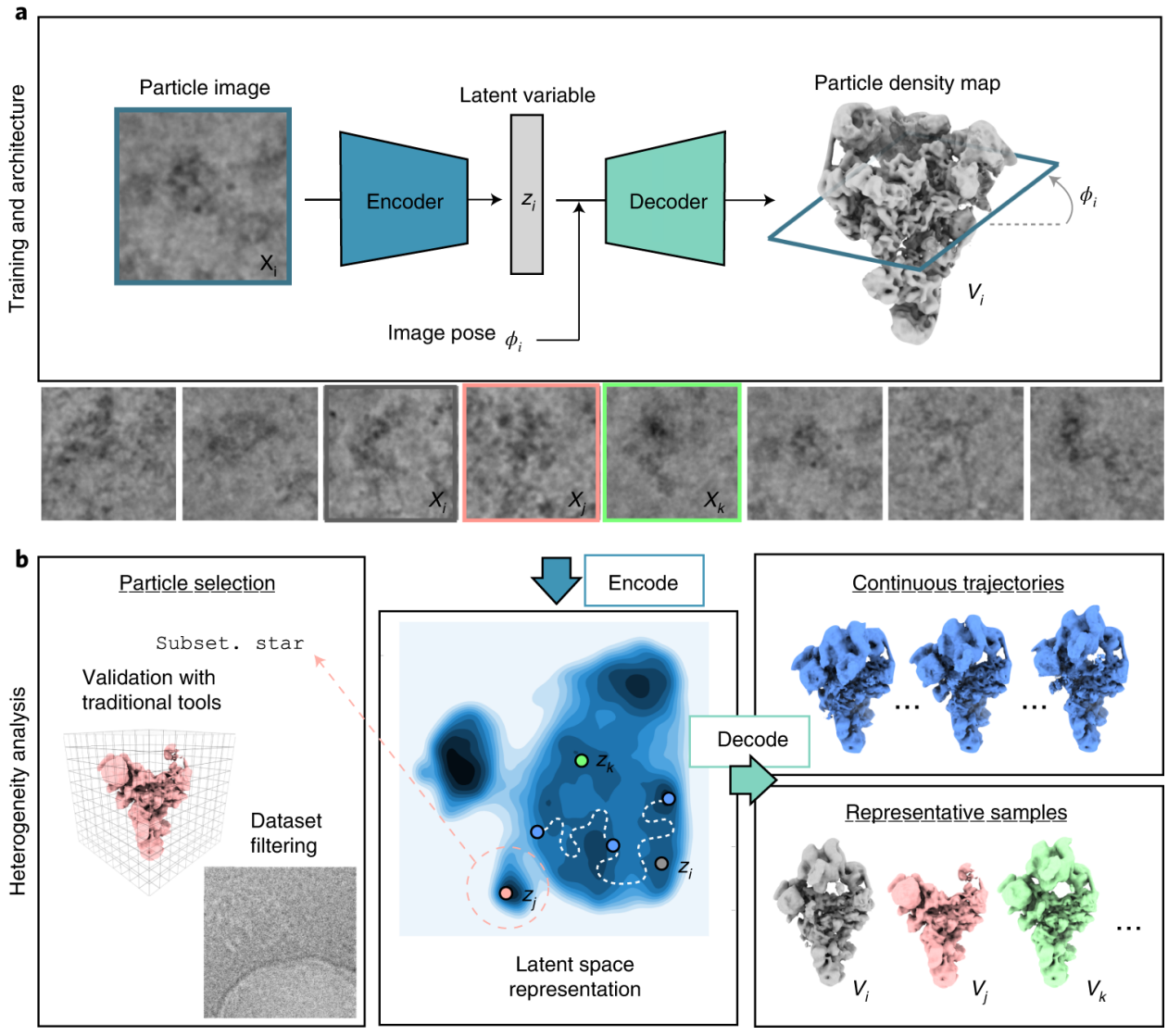


Figure 16 Pipeline of CryoDRNG neural network for conformational variability analysis. ***a*** *CryoDRGN architecture.* ***b*** *Results interpretation and heterogeneity analysis. Adapted from* [8]

To extract conformational heterogeneity, CryoDRGN treats the problem as an inverse problem of the image formation model where each image corresponds to an oriented central slice of the 3D volume in the Fourier domain. The encoder learns a pose-invariant representation of the protein's structural heterogeneity for a given image and encodes it in a latent representation. Then, given a sample from the encoder distribution and the positional encoder of the 3D coordinates, the decoder reconstructs the image pixel-by-pixel. Finally, uniform manifold approximation and projection (UMAP) [115] is applied to the learned latent representation to visualize the conformational variability and perform 3D heterogeneous reconstructions. CryoDRGN uses an estimation of alignments parameters for training and refining them using a branch-and-bound (BNB) searching algorithm, and an acceleration of the BNB alignment was implemented using a feedforward neural network [116].

In recent work, amortized inference was used to improve variational autoencoder latent representation. CryoFIRE [117] follows the same approach as CryoDRGN and simultaneously uses the encoder to learn the orientations and conformational variability from scratch; however, Instead of re-estimating the parameters of the approximate posterior, the amortized inference reuses the parameters of the distribution, in this case, Gaussian distribution, for multiple points which improves its efficiency and expressiveness of more complex distributions.

Furthermore, Generative adversarial network (GAN) approaches were used for the heterogeneity analysis. Multi-CryoGAN [118] is an unsupervised approach based on CryoGAN, which was initially developed to solve homogenous 3D volumetric structures from single-particle images. CryoGAN replaces the generative network with a physics simulator. Under CryoGAN, the physics simulator projects an input EM map using a set of pre-estimated imaging parameters (Euler angles, shifts, CTF parameters). The simulator must then add realistic noise to the clean projections. The physics simulator is paired with a convolutional discriminator whose role is to distinguish between the synthetic samples generated by the physics simulator and real experimental single-particle images. Like CryoGAN, MultiCryo-GAN uses the same approach with a twist for conformational variability analysis, combining CryoGAN with a CNN. The CNN is first fed a decided a priori latent representation of the conformational variability to generate an EM map (conformation). This output conformation is then passed to CryoGAN, as explained earlier, and the whole approach is trained simultaneously to minimize an optimal transport objective function, the Wasserstein distance.

With the success of CryoDRGN in extracting the protein's structural heterogeneity, VAEs were extensively used, and a second work followed CryoDRGN, namely E2GMM[119] was implemented. First, the decoder is trained with a 0-valued latent variable to obtain a Gaussian model of 5N parameters by minimizing the Fourier Ring Correlation (FRC) between the 2D projection of the predicted Gaussian model and the 2D projection of the initial model. Then, the encoder is trained to learn an invariant low-dimensional representation of the conformational heterogeneity from each input single particle image. This latent representation is then used to continue the training of the decoder. During training, the predicted 3D Gaussian model is projected to a 2D image, and the FRC is minimized between the 2D projection of the 3D Gaussian model and the input single-particle. Then PCA is applied to the latent space to extract and visualize the conformational space and perform heterogenous 3D reconstruction. E2GMM also uses pre-estimated alignment parameters, and a separate network can be trained to refine them.

Instead of an encoder-decoder approach, CryoSPARC proposed a variational auto-decoder neural network for 3D flexible refinement (3DFlex) [6]. Under this model, a single particle image is associated with a low-dimensional latent coordinate **z** that encodes the conformational heterogeneity information. Formally, each z_i can be seen as a point estimate that maximizes the posterior distribution over conformational states. Then a neural flow generator network converts the latent representation into a flow field, and a convection operator then deforms a high-resolution canonical density using a weighted pre-segmented 3D mesh to generate a convected map. This map is projected along a predetermined pose direction, CTF-corrupted, and compared against the input experimental image. The latent representation is zero-initialized and optimized during the neural flow generator training.

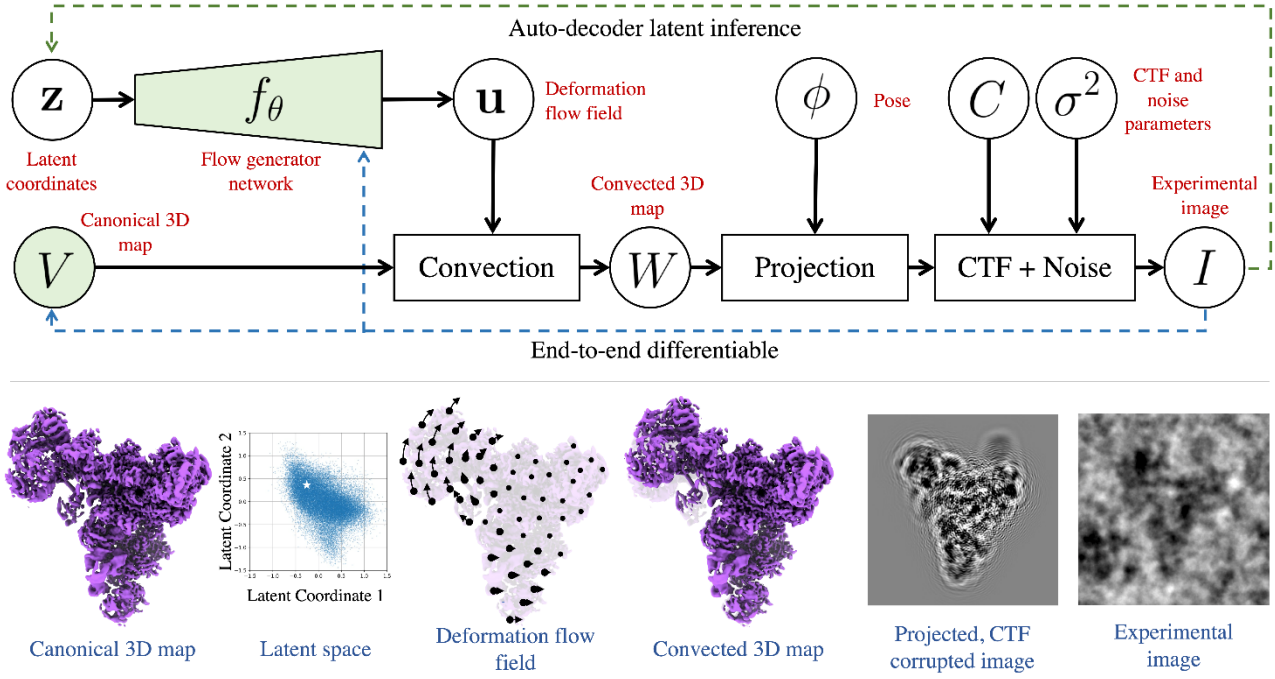


Figure 17 Pipeline of 3DFlex network for conformational variability analysis. Adapted from [6]

3. Summary of conformational heterogeneity analysis approaches

Discrete variability analysis approaches are considered a good representation of continuous heterogeneity because continuous functions can be approximated by a large number of discrete samples. However, continuous conformational analysis approaches have several advantages over discrete models. First, they do not require the user to specify the number of classes, unlike discrete models, where the choice of the number of classes can be challenging, as shown in ML approaches. Second, continuous models use all available data to produce high-resolution volumes, while discrete models only use a subset of the data. Third, continuous models have an advantage in analyzing rare conformations due to the continuity between conformations. Finally, it is worth mentioning that while traditional 3D classification and multi-body analysis are well-established, the use of continuous heterogeneity models is still a challenge for practitioners, as they face the dilemma of choosing the right software, model assumptions, and parameters, as summarized in **Table 1**.

Approach	**Nature of heterogeneity**	**Space**	**Deep Learning-based**	**Hidden variables**	**Known variables**	**Reference volume**
CryoDRGN[8]	Continuous	Fourier	Yes	Conformational representation z Rotation	CTF	Free

				Translation		
Multi-Cryo-GAN[118]	Continuous	Real	Yes	Conformational representation z	CTF Rotation Translation	Yes
3DFlex [6]	Continuous	Real	Yes	Conformational representation z	CTF Rotation Translation	Yes
E2GMM[119]	Continuous	Real	Yes	Conformational representation z	CTF Rotation Translation	Free
CryoFIRE [117]	Continuous	Fourier	Yes	Conformational representation z Rotation Translation	-	Free
3DVA[114]	Continuous	Real	NO	-	CTF Rotation Translation	Free
ManifoldEM [19]	Continuous	Real	No	-	CTF Rotation Translation	Free
HEMNMA [9]	Continuous	Real and Fourier	No	Rotation Translation Normal-mode amplitudes (conformational parameters)	-	Yes
Maximum Likelihood	Discrete	Real	No	Rotation Translation	CTF K classes of conformation	Yes
Cryo-BIFE[99]	Continuous	Real	No	-	Rotation Translation	Yes
MDSPACE[28]	Continuous	Real	No	-	Rotation Translation	Yes

Table 1 . Classification of conformational heterogeneity analysis method based on several factors, the processing is done in real or Fourier space, the known and hidden parameters, the nature of conformational heterogeneity analysis. Table is adapted from [120] and further extended.

	Biomolecular complex	Data type
CryoDRGN[8]	Ribosome 80S, Spliceosome on pre-cursor mRNA (pre-mRNA)	Synthetic and Experimental
Multi-Cryo-GAN[118]	Sars-Cov2	Synthetic
3DFlex [6]	Ribosome 80 S, Tri-snRNP spliceo-some	Synthetic and Experimental
E2GMM[119]	Ribosome 50S, spliceosome on pre-cursor mRNA (pre-mRNA)	Synthetic and Experimental
CryoFIRE [117]	Ribosome 80S, spliceosome on pre-cursor mRNA (pre-mRNA)	Synthetic and Experimental
3DVA[114]	Ribosome 80 S, Tri-snRNP spliceo-some	Synthetic and Experimental
ManifoldEM [19]	Ribosome 80S	Synthetic and Experimental
HEMNMA [9]	E. coli 70S ribosome (70S DNA polymerase Pol α and B subu-nit complex Tomato bushy stunt virus (TBSV)	Synthetic and Experimental
MDSPACE [28]	Ribosome 80S, Tmr ABC exporter	Synthetic and Experimental
Maximum Likelihood	Multitude of complexes	Synthetic and Experimental

Table 2 Summary of synthetic and experimental cryo-EM datasets on which the conformational heterogeneity analysis methods have been tested. Table is adapted from [120] and further extended.

4. Conclusions and remarks

This chapter provided a review and critical comparison of conventional approaches and deep generative neural networks, focusing on explaining their advantages and drawbacks. This review highlights the differences between the approaches by categorizing them according to the nature of the heterogeneity for which each approach is the best suited. It also highlights the new wave of methods, based on deep generative learning. While the use of maximum likelihood approaches may result in consistent results, its dependence on the estimation of the number K of the classes before the analysis affects the convergence, besides the assumption that the heterogeneity of the studied biomolecular

complex is discrete. With the introduction of approaches such as manifold embedding and subspace analysis, the assumption of the nature of conformational heterogeneity vanished. However, the separation of alignment parameters estimation from the heterogeneity analysis is a crucial limitation of these approaches. The larger the conformational changes, the less accurate the alignment parameters estimation, which leads to an iterative process to determine and refine conformation and rigid-body parameters, implemented in hybrid methods, combining simulation (NMA or MD) and image processing based on a reference model. One can notice that these methods may suffer from simulated conformations that might not exist in the real data, potentially leading to misinterpretation of the image analysis results. A new wave of approaches based on variational deep learning could overcome several problems of the previous approaches. However, choosing a set of hyperparameters is subjective and may affect the results. Table 1. summarizes the most used methods in the literature and provides a comparative analysis of known parameters and hidden parameters needed before the heterogeneity analysis for each method. From Table 2, one can notice the absence of universal benchmarks, which interferes with comparative studies of the old and newly developed methods and prevents users from selecting methods adequate for their use case.

Chapter 3: Background

During this thesis, three approaches were developed to address the continuous conformational variability of biomolecular complexes, mainly Deep-HMEMNA, supervised cryo-ViT, and unsupervised cryo-ViT.

Deep-HEMNMA is a deep neural network that accelerates HEMNMA using Normal mode analysis (NMA), image registration, and Convolutional Neural Network (CNN).

Supervised cryo-ViT and Unsupervised cryo-ViT are deep learning approaches based on a variational autoencoder (VAE) that uses a vision transformer (ViT) to study conformational variability from single-particle images.

This chapter highlights the core approaches, NMA, CNN, VAE, and ViT, used to develop the methods presented in this thesis.

1. Molecular motion analysis

Biomolecular complexes motion analysis is a vital process that helps to elucidate their biological functions, and their interactions with the environment can be challenging. Moreover, molecular motion was studied at the atomic level, where a biomolecular complex is represented as a collection of atoms. Each atom is represented by a point in space in the Cartesian coordinate system, where atoms are connected with chemical bonds. When a biomolecular complex changes its shape, the analysis of how the atoms may potentially move was studied mainly using two techniques, NMA, and Molecular dynamics (MD) simulation.

Molecular dynamics simulation

Molecular dynamics dates back to the 1950s and is the reigning method for molecular motion analysis. It involves simulating the motion of atoms over time using classical mechanics. In an MD simulation, the motion of each atom is calculated using Newton's laws of motion, and the interactions between atoms are described using interatomic potentials (force fields) [121]. MD simulations can be used to study the behavior of a molecule under different conditions, such as different temperatures and pressures, and they can provide detailed information about the dynamics of the molecule. Although MD is a powerful tool for biomolecular motion analysis, it is computationally expensive.

It can be limited by time and length scales, where it is not well-suited for studying processes that occur over highly long periods or at microscopic lengths scales.

Normal mode analysis

Normal mode analysis (NMA) is a fast-molecular computational method used to study the large-scale conformational changes that can occur in proteins. It uses a mathematical model to identify the most likely ways a protein can move and change shape based on its atomic-level structure. This information can help understand the function of a protein, as well as for predicting how it might

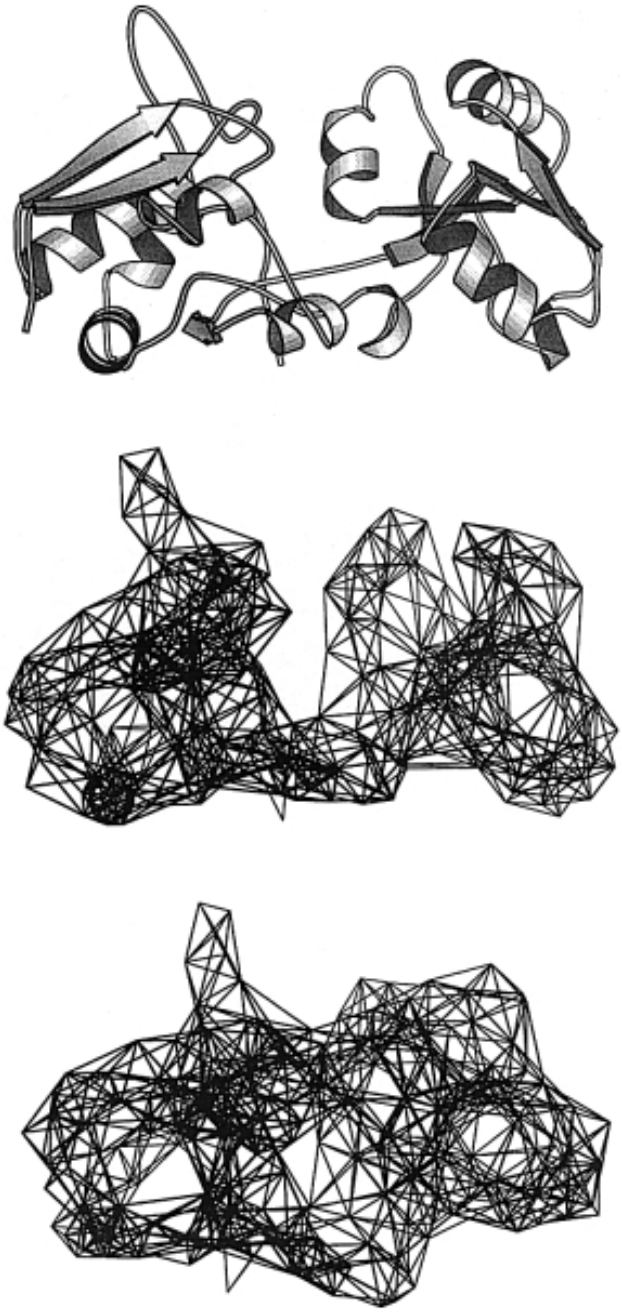


Figure 18 An example of elastic network model (ENM) modeling of the lysine-arginine-ornithine (LAO) binding protein. The open state of the LAO protein (top), the open state of the LAO protein modeled with an ENM (middle), the closed state of the LAO protein modeled with an ENM. Adapted from [12]

behave in different environments. This section explains in depth the idea of NMA and presents its mathematics. [122]

When a biomolecular complex changes its shape, the movement of atoms is generally more complex and requires sophisticated mathematical approaches to study it. NMA is based on the Elastic Network Model (ENM)[123], which assumes that a biomolecular complex can be represented as a network of interconnected atoms, where atoms closer than a given cutoff distance are coupled by a harmonic spring, and the force exerted by a harmonic spring is given by Hooke's Law, which states that the force is proportional to the displacement of the atoms from their equilibrium (minimum energy) position[124]. Consider a biomolecular complex consisting of N atoms, each of position $q_i = (x_{i1}, x_{i2}, x_{i3})$ and the equilibrium position is $q_i^{(0)}$, and each atom can vibrate about its equilibrium position. The displacement from the equilibrium position following the Hookean single potential energy is given by:

$$q_{i\alpha} = q_{i\alpha} - q_{i\alpha}^{(0)} \tag{7}$$

Where $\alpha = 1, 2, or\ 3$ represent the position of the atom in the Cartesian coordinates.

In Tirion's[123] ENM that follows the Anisotropic Network Model (ANM) [124], the Hookean pairwise potential energy of a harmonic spring between two atoms is simplified as follows:

$$E(q_i, q_j) = \frac{\gamma_{ij}}{2}\left(|q_{ij}| - |q_{ij}^0|\right)^2 \tag{8}$$

Where $|q_{i,j}|$ is the distance of the vector connecting atom i and atom j, and similarly $q_{i,j}^0$ is the distance of the correspondent atoms i and j in the initial conformation. This energy function links all atoms with a spring of elastic constant $\boldsymbol{\gamma_{ij}} = C$ [123]. Therefore, the potential energy within a molecule can be expressed:

$$E(q) = \sum_{\substack{q_{i,j}^0 < R \\ i \neq j}} E(q_i, q_j) \tag{9}$$

Here R is the radius of interaction between the atoms.

One may study the change in energy with respect to one reference conformation, q^0 and since the displacements are small, the potential energy can be expanded in a Taylor series as follows:

$$E = E(q_0) + \sum_i \left(\frac{\partial E}{\partial q_i}\right)(\mathrm{q}_i - {q_i}^0) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 E}{\partial q_i \partial q_j}(q_i - q^0)\left(q_j - {q_j}^0\right) + \ldots \quad (10)$$

The first term $E(q_0)$, is the potential energy when the atoms are in their equilibrium configuration and can be neglected and $E(q_0)$ is assumed to be zero (near the global minimum of the energy surface), and since $E(q_0)$ is zeroed, the first derivative vanishes because it is evaluated at the equilibrium conformation where there is no force on any atom. Since the displacements of interest are small [125], the Taylor series is truncated and approximated by the quadratic terms; thus, Eq. (9) can be rewritten as:

$$\begin{aligned} E &= \frac{1}{2}\sum_{i,j}(q_i - {q_i}^0)\frac{\partial^2 E}{\partial q_i \partial q_j}\left(q_j - {q_j}^0\right) \quad (11)\\ &= \frac{1}{2}\sum_{i,j}(q_i - {q_i}^0)\, H_{i,j}\left(q_j - {q_j}^0\right) \\ &= \frac{1}{2}\,\Delta q^T H\, \Delta q \end{aligned}$$

Where H is the Hessian matrix obtained from the second derivatives of the potential energy with respect to the components of r. The matrix H is an N x N block matrix of 3 x 3 matrices:

$$H = \begin{bmatrix} H_{12} & H_{12} & \ldots & H_{1N} \\ H_{21} & H_{22} & \ldots & H_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ H_{N1} & H_{N2} & \ldots & H_{NN} \end{bmatrix} \quad (12)$$

where each H_{ij} is given by:

$$H_{ij} = \frac{\partial^2 E}{\partial r_i \partial r_j} = \begin{bmatrix} \frac{\partial^2 E}{\partial x_i \partial x_j} & \frac{\partial^2 E}{\partial x_i \partial y_j} & \frac{\partial^2 E}{\partial x_i \partial z_j} \\ \frac{\partial^2 E}{\partial y_i \partial x_j} & \frac{\partial^2 E}{\partial y_i \partial y_j} & \frac{\partial^2 E}{\partial y_i \partial z_j} \\ \frac{\partial^2 E}{\partial z_i \partial x_j} & \frac{\partial^2 E}{\partial z_i \partial y_j} & \frac{\partial^2 E}{\partial \square\square_i \partial z_j} \end{bmatrix} \quad (13)$$

Using the ANM described in Eq.(8) it is possible to readily write a closed form of the Hessian matrix H using Eq.(8)[125] in Eq.(11), and in this case the second derivatives of the potential are given by:

$$\frac{\partial^2 E}{\partial x \partial y_j} = -\frac{C(x_j - x_i)(y_j - y_i)}{q_{ij}^2} \tag{14}$$

$$\frac{\partial^2 E}{\partial x_i \partial z_j} = -\frac{C(x_j - x_i)(z_j - z_i)}{q_{ij}^2} \qquad \frac{\partial^2 E}{\partial y_i \partial z_j} = -\frac{C(y_j - y_i)(z_j - z_i)}{q_{ij}^2}$$

Following the notation $x_{ij}^0 = (x_j^0 - x_i^0)$, $y_{ij}^0 = (y_j^0 - y_i^0)$, and $z_{ij}^0 = (z_j^0 - z_i^0)$ we can rewrite Eq.(13) as follows:

$$H_{ij} = -\frac{\gamma}{q_{ij}^2}\begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\ x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix} \tag{15}$$

In the context of conformational changes, it is necessary to account for kinetic energy as well as potential energy, and this gives us the following equation:

$$Hu_k = \omega_k^2 M u_k \tag{16}$$

Where M is the matrix of mass of the atoms, and u_k and ω_k^2 are the eigenvectors are the eigenvalues of the Hessian matrix and represent the normal mode vectors and the squared normal mode frequencies, respectively. However, for simplicity, Tirion's ENM assumes that each atom is assigned a mass of 1[126]. Therefore, Eq.(16) can be rewritten as:

$$Hu_k = \omega_k^2 u_k \tag{17}$$

Since H is real and symmetrical, Eq 3.11 can be solved using diagonalization to obtain the eigenvalues and eigenvectors. This leads to the following:

$$\mathrm{H} = \mathrm{UAU}^{\mathrm{T}} \tag{18}$$

The matrix U is a unitary matrix whose columns are the eigenvectors of the matrix H, and A is a diagonal matrix with the corresponding eigenvalues. The eigenvalues, ω_k (where k ranges from 1 to $3N$), are related to the frequency of the movement and the corresponding eigenvector, u_k, is referred to as the k_{th} normal mode. The normal modes with the smallest eigenvalues represent more collective movements, while those with larger eigenvalues represent localized movements. The first six normal modes correspond to a combination of the six rigid body degrees of freedom (three global rotations and three global translations), which do not change the potential energy of the conformation.

Standard diagonalization methods are time consuming; thus, Tama et al. [98] proposed a method called Rotation translation block (RTB), which speeds up the Hessian matrix diagonalization by reducing the size of the Hessian matrix. The RTB method involves dividing the system into smaller blocks, or subunits, each of which is treated as a rigid-body and can adopt only rotations and translations.

Chapter 4 presents the exploitation of NMA in continuous conformational variability analysis from single-particle cryo-EM.

2. Flexible fitting

Flexible fitting[127] allows determining the most likely 3D structure that would produce the observed density map, given some assumptions about the possible flexibility and conformational changes for the biomolecule. Therefore, it aims to obtain a new conformation of an already-known structure that fits a given electron density map. Given an atomic model obtained using X-ray crystallography, the process is to adjust a set of parameters that control the flexibility of the atomic model within an electron density map to obtain the best fit.

In flexible fitting, the fitness of the electron density map and the atomic model is defined by the cross-correlation between the electron density map and the simulated electron density map generated from the atomic model. Flexible fitting is subject to various constraints, such as ensuring that the atomic model remains physically reasonable, and NMA[106] and MD simulations[128,129] are the two reigning methods, in which the normal mode amplitudes in NMA or the potential energy in MD simulations are adjusted iteratively to drive the fitting process [130].

Flexible fitting can be performed in 3D [128,131,132], and the process consists of adjusting the position and orientation of the atomic model to fit into (1) a set of low-resolution cryo-EM density maps[127] produced by a discrete classification approach, or (2) a set of subtomogram averages[131,132] to study the conformational variability. This process is done iteratively through a series of steps to optimize an objective function, where the optimization algorithm seeks to minimize the cross-correlation between the electron density map and a model synthesized from the atomic model. The

optimization process is repeated until the fit reaches convergence, meaning that the adjustments in the orientations and positions of the atomic model are no longer improving the fit.

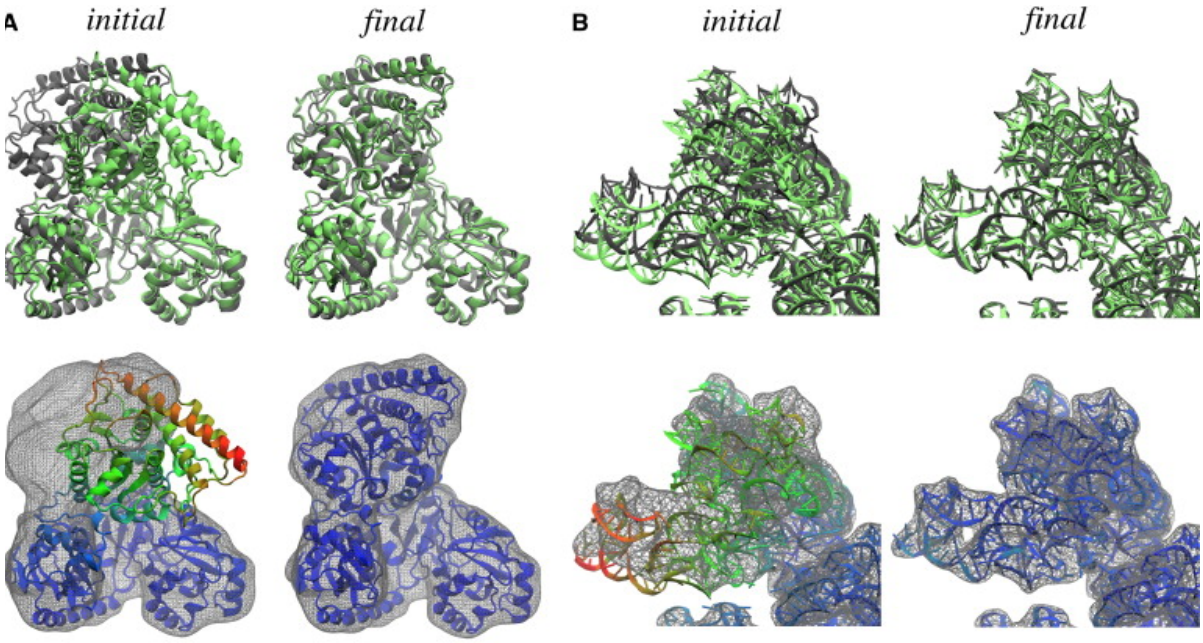


*Figure 19 An example of flexible fitting of two different complexes. **A** the initial and final Acethyl-CoA synthase/carbon monoxide dehydrogenase atomic structures fitted into an electron density map. **B** the initial and final 16S rRNA atomic structures fitted into an electron density. Adapted from*[20]

On the other hand, flexible fitting can be performed in 2D[9,28]. For SPA, it has been an essential tool for elucidating dynamical aspects of biomolecular complexes and it consists of adjusting the flexibility parameters and the orientations of the atomic model in a single plane to study conformational variability from single particles. Here, the atomic model is projected onto a single plane and optimized in that plane, in which the optimization algorithm is similar to that used in 3D flexible fitting and seeks to minimize the cross-correlation between the 2D projection of the atomic model and a single-particle image, this projection eliminates the need to perform computationally intensive calculations in three dimensions.

3. Convolutional neural network

Convolutional neural networks (CNNs) are a type of artificial neural network (ANN) that are particularly well-suited to processing grid-like (2D and 3D) data, such as images, video frames, etc...

ANN is a subset of machine learning algorithms based on research on the nervous system and the brain and they mimic the biological neural networks. They consist of input, hidden, and output layers with connected neurons (nodes) to simulate the human brain

CNNs are composed of multiple layers of artificial neurons, simply neurons [133]. In an artificial neural network, a neuron is a crucial unit, which is a mathematical function that is connected to the input data. In CNNs the neuron is arranged in a grid-like data and considered to be filters that will be applied to the input data, also called a receptive field, and are trained to recognize patterns within this region. The building block of CNNs is the convolutional layers, which are responsible for learning features and patterns within the data. Convolutional layers consist of a set of filters learned during training, and each convolutional layer processes the input data by applying the set of filters to produce a set of feature maps. Each filter is a small matrix of weights that is trained to recognize a specific pattern in the input data. As the filter is moved across the input data, it produces a feature map by taking the dot product between the filter weights and the input data at each position. The convolution operation can be written as:

$$(f * g)(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m, n) g(x - m, y - n) \tag{19}$$

Here, f and g are the image and the filter, and $*$ denotes the convolution operation. A visual representation of the convolution operation is illustrated in **Error! Reference source not found.**.

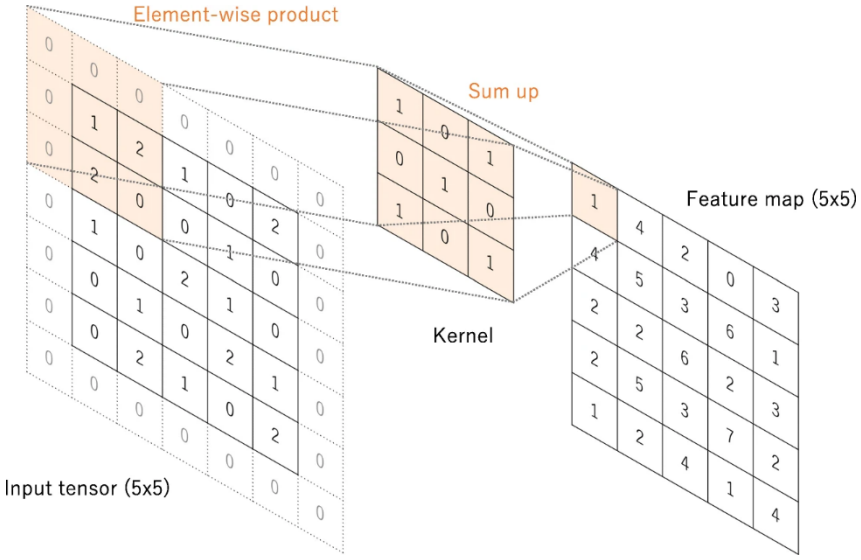


Figure 20 Schematic representation of the convolution operation. Adapted from[11]

CNNs were first introduced in the 1980s by Kunihiko Fukushima [134], however, due to the limitations in advanced computing power and the availability of large datasets they waited until the 2010s to become widely used. However, The first successful application of CNNs dates back to 1998 and was in digit recognition, where Yann Lecun and colleagues implemented the first successful CNN architecture called LeNet [135]. In 2012, a breakthrough in CNNs was achieved by a team led by Alex Krizhevsky[136], where they developed a deep CNN called AlexNet to win the ImageNet Challenge[137].

In 2014, Simonyan and Zisserman 2014[5] followed AlexNet with a new network called VGGNet (Visual Geometry Group), which achieved state-of-the-art results and outperformed AlexNet on the ImageNet challenge. The architecture of VGGNet consists of a deep network (16-19 convolutional layers), and it has been influential in the development of new deep learning architectures. It has been widely used as a feature extractor in some use cases, such as object detection and classification pipelines, and as a benchmark for evaluating new architectures and techniques in the field, Despite VGG's effectiveness, it suffered from the vanishing gradient problem, which limited the number of the layers one can use.

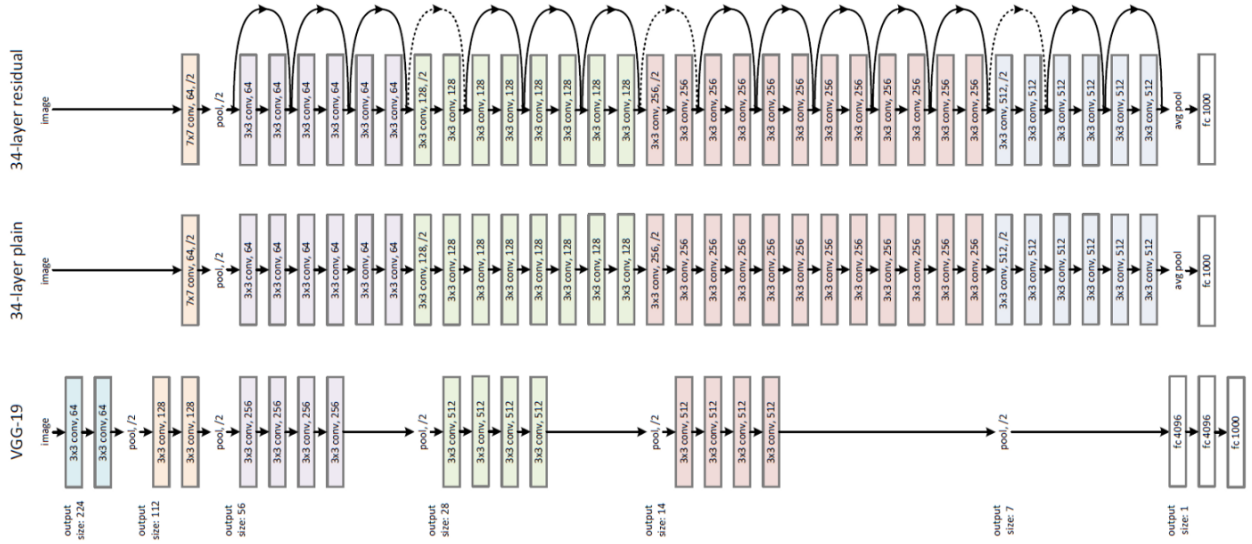


Figure 21 Schematic representation of a comparison between 34-layer residual network architecture (top), 34-layer plain neural networks (middle) and 19-layer VGG architecture (bottom) [5]. Adapted from [23]

In recent years, the main focus in the field has been improving the efficiency and performance of CNNs, including developing new architectures such as ResNets (Residual network)[138]. A comparative schematic representation between a plain deep neural network, VGG architecture, and residual neural network is shown in **Error! Reference source not found.**The critical innovation of

ResNet is the use of residual connections, which allow the network to learn much deeper architectures (i.e., with more layers) than was previously possible without suffering from the vanishing gradient problem, in which the gradient values in the first layers are minimal for updating the filters weights during backpropagation. The introduction of residual layers allows for training more accurate models for image classification and other tasks. The impact of ResNet on CNNs has been significant, as it has led to a renewed interest in profound neural networks and has been the foundation for many state-of-the-art models in computer vision and other fields.

4. Vision-transformer

A new wave of deep learning approaches has surged based on the Transformer architecture[139] that was initially developed for natural language processing tasks, called Vision Transformer (ViT)[16]. ViT, shown in (Figure 22), is another type of neural network architecture designed specifically for image processing tasks.

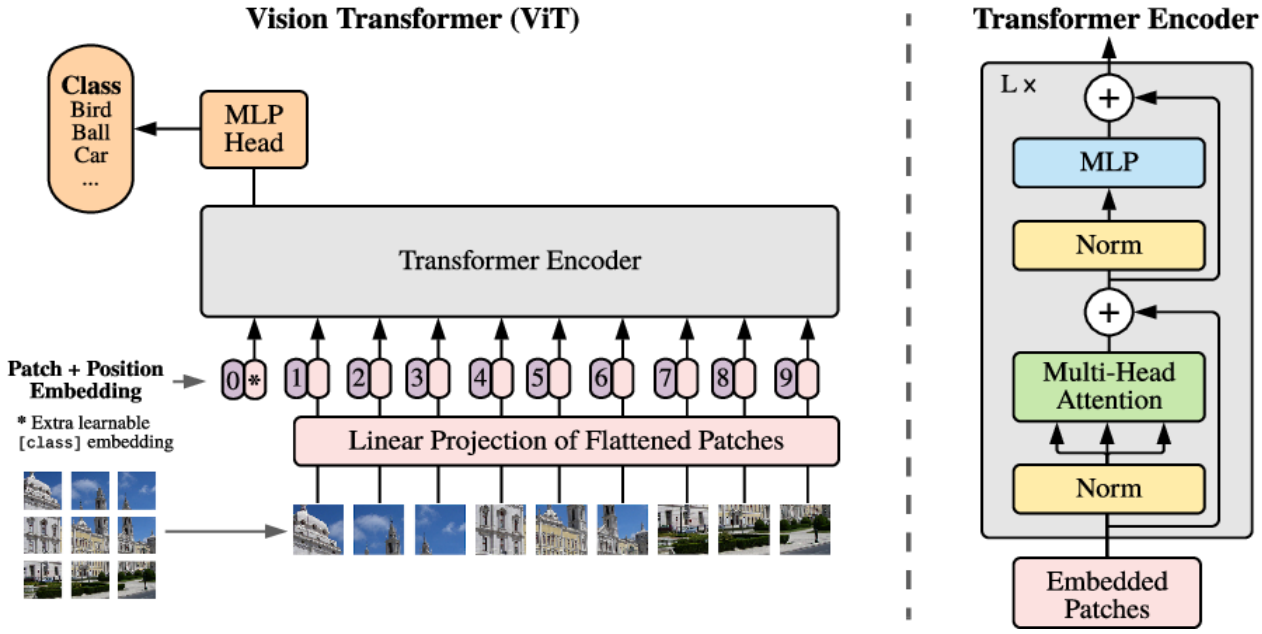


Figure 22 Overview of the pipeline of the original implementation of ViT, all the component of the vision transformer (left), the transformer encoder used in ViT that consists of a Norm Layer, a multihead attention mechanism and an MLP(right). This Adapted from [16].

The idea behind the Transformer architecture is to use self-attention mechanisms [140] to process input sequences, a sequence of words in the case of NLP tasks, in a parallel and efficient manner. In the case of ViT [16], the image is split to a fixed-size non-overlapping patches, which helps the network to see the whole image at once[141]. First, each patch is flattened, and a position embedding is computed for each. The position embedding incorporates the spatial relationships between patches and preserves the position of the patches in the image. The computed position embedding is added to the flattened

patch and passed to a set (L) of Transformer encoders. The transformer encoder consists of a series of layers that include (1) self-attention mechanisms, which weigh the importance of different parts of an input image, allowing each to focus on different aspects of the input; typically, multiple attention mechanisms are used (multi-head attention); (2) feedforward neural networks (MLP head), which is a set of fully connected layers to compute the desired output.

Layer normalization[5] is used in ViT, referred to as *Norm* in Figure 22, and it stabilizes the training process and ensures that the inputs to each layer have a similar distribution by reducing the internal covariate shift, which occurs when the distribution of the inputs to a layer changes during training.

In the original implementation of ViT, the output of the final layer is used to predict the class of an input image. However, the pipeline can be tweaked for any specific output depending on the study case (e.g., regression, classification, object detection).

Self-attention mechanism

The key component in Transformer-based models is the Self-attention mechanism [140], commonly known as the "attention mechanism". It was first used in NLP tasks as a way to model the dependencies between words in a sentence. The self-attention mechanism was first introduced in[15], enabling the model to attend to different parts of the input as it processes it rather than processing the input in a fixed order; this allows the model to capture long-range dependencies and complex relationships between input elements which reduce inductive bias. The self-attention mechanism has been widely used in various fields, including machine translation (BeRT [142]), computer vision (ViT EffiecentNet).

Instead of a sequence of words, in ViT the self-attention mechanism operates on a sequence of flattened image patches and computes a weighted sum of the patches based on their relative importance which allows the network to see the global features of th input image.

Mathematically, the attention mechanism can be expressed as follows:

$$\text{Attention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \text{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\right)\mathrm{V} \tag{20}$$

Here, Q, K, and V are matrices representing the "query", "key", and "value" tokens, respectively. They are typically derived from a learned linear transformation of the input sequence. The query serves as a question that the model asks about the input sequence when computing attention

scores. The key represents spatial information (position) in the input sequence that is used to compare with the query to determine which parts of the input sequence are relevant to the task at hand. It acts as an index or reference for each position in the input data. The value indicates how much each position in the value sequence should contribute to the final representation of the sequence. The attention mechanism computes the dot product of the query and key values, and scales it by the inverse square root of the key dimensionality (d_k), and applies a softmax function to produce a positive weight for each patch and multiply it by the value tokens. The intuition here is to keep only relevant features with high scores (the softmax score) and discard the irrelevant features with low scores [143].

The attention mechanism is used multiple times in parallel in the ViT architectures to form the multihead attention module, shown in Figure 23. Multihead attention runs through an attention mechanism several times in parallel and allows the network to build up a hierarchical representation of the input image, with each layer focusing on different levels of abstraction. The independent attention outputs are then concatenated and linearly transformed into the expected dimension to be used in the following layers.

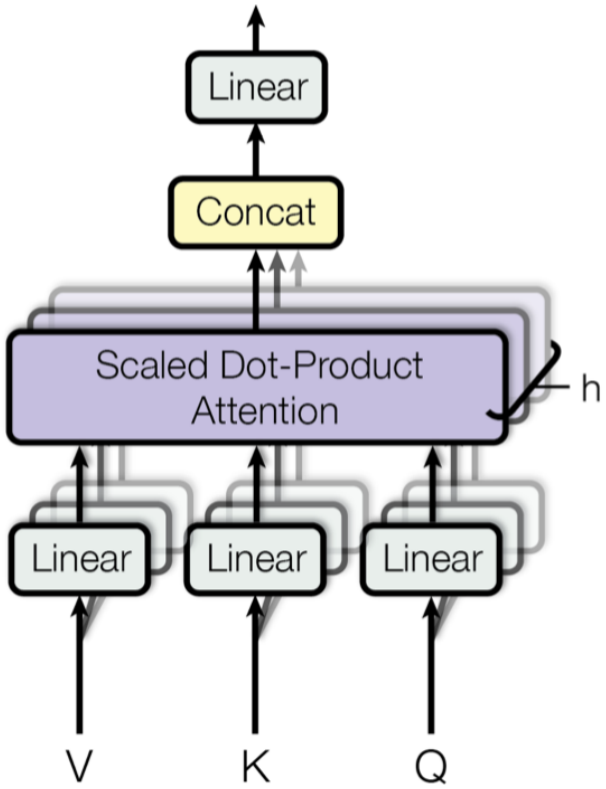


Figure 23 Schematic representation of the multihead attention module used in the transofrmer architecutre. Q,K,V represent the query, key value connected to a scale dot product node. h represnts the number of attention mechanisms concatenated and attached to a linear layer. Adapted from [15]

5. Variational autoencoder

Variational autoencoder (VAE)[144] is an unsupervised deep learning approach, which is a type of neural network models called autoencoders. Autoencoders, are a set of neural network models that learn how to compress the input data to a discrete representation of a small dimension while preserving the ability to reconstruct it back to its original dimension.

Instead of learning a discrete representation, VAEs learn the dataset's probability distribution to generate new samples of the input dataset. It provides a probabilistic description of the input data in a latent representation by combining autoencoders with variational inference. A graphical representation of VAE architecture is shown in Figure 24.

Because VAEs are a type of autoencoder, they consist of an encoder, $q_\theta(z|x)$, a neural network that outputs the parameters to $q_\theta(z|x)$, which maps the input data to a probability distribution parameters, the Gaussian distribution is commonly used and in this case the encoder predicts the parameters of the Gaussian distribution mean and standard deviation, then the latent variable is sampled using the predicted parameters of the distribution using a technique called reparameterization trick, explained in the next section, and a decoder, $p_\phi(x|z)$ a neural network that reconstructs back the input images given the latent variable[145]. VAEs are widely used in different fields, audio codecs, text generation

and image generation, and the architecture of the encoder and decoder can be an MLP, a CNN, or any other neural network architecture depending on the type of input data.

The objective of VAE is to learn the parameters θ, ϕ of the encoder and decoder networks that maximize the likelihood of the input data x. However, directly maximizing the likelihood is challenging; therefore, VAE uses the concept of variational inference and maximizes a lower bound on the likelihood. The lower bound is referred to as the evidence lower bound (ELBO) and is given by:

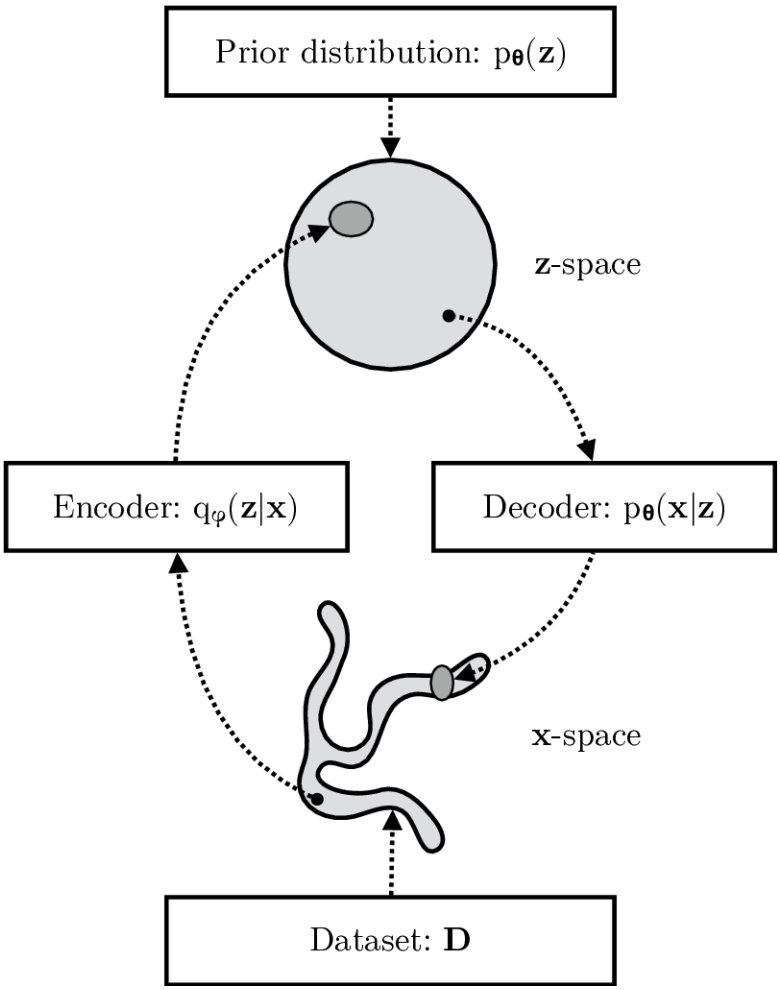


Figure 24 Overview of the architecture of the original variational autoencoder. Adapted from [2]

$$ELBO(\theta, \phi) = Eq_\theta(z|x)[log\ p_\phi(x|z)] - KL[q_\theta(z|x)\ ||\ p(z)] \quad (21)$$

where $q_\theta(z|x)$ is the approximate posterior over z given the input x, and $p(z)$ is the prior distribution over the latent representation z. In the ELBO equation (21), the first term is the reconstruction error, which reflects how well the model has reconstructed the input data, and is

calculated between the input and the output of the decoder, and the second term is the KL divergence between the approximate posterior extracted by the encoder and the predetermined prior distribution, the prior can be a simple distribution as mentioned before, which encourages the approximate posterior distribution to be as close as possible to the prior distribution. It is worth to mention that in VAE, instead of maximizing the ELBO, we minimize it to optimize the parameters of the encoder and decoder networks.

Reparameterization trick

During training, VAE estimates the parameters of a specific probability distribution, and the marginal likelihood is computed to evaluate the distribution's parameters. However, the marginal likelihood can sometimes be intractable, meaning it is difficult or impossible to calculate analytically. This happens when the integral of the likelihood function over the prior distribution cannot be computed in closed form or when the integral is too difficult to evaluate using numerical methods.

The reparameterization trick [144,146] is a technique used in VAE to make the marginal likelihood computationally tractable and perform efficient gradient-based optimization of the parameters. In VAE, the encoding is modeled as a random variable with a learned distribution (e.g., a Gaussian distribution); however, to compute gradients during training, the encoding must be a differentiable function of the input and the model parameters. The reparameterization trick achieves this by reparametrizing the random variable as a deterministic function of the input, the parameters of the distribution, and a sample noise ε from a fixed standard distribution (e.g., a standard Gaussian distribution). Introducing a non-deterministic noise is crucial in the reparameterization trick as it allows overcoming the challenge of optimizing a sampling process involving a non-differentiable cumulative distribution function (CDF). The gradient with respect to the parameters can then be computed using the chain rule, enabling efficient training of the VAE using gradient-based optimization methods[7]. The left panel of Figure 25 demonstrates the original form of the sampling process of z using the given input data x and the variational parameters ϕ, in the context of VAE, x is typically an embedded vector produced by the encoder. Right panel of Figure 25 demonstrates the sampling process using the given input x and the variational parameters ϕ and the noise variable ε (randomness) to allow the gradient flow and solve the marginal likelihood intractability problem.

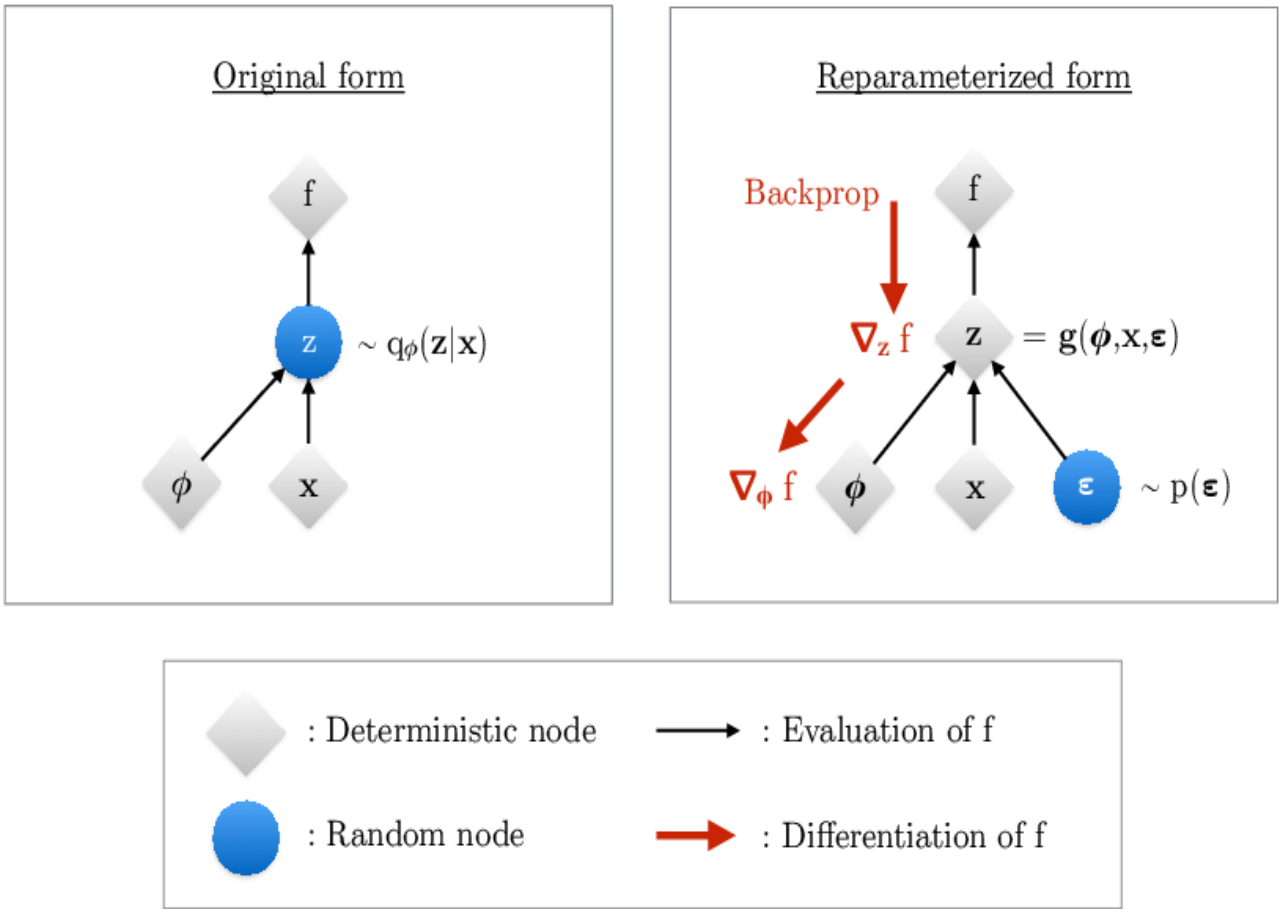


Figure 25 Schematic representation of the reparameterization trick. The original sampling process where the variational parameters ϕ affect the objective f through the random variable z (left). The reparameterization trick process to allow the backpropagates and differentiation by externalizing the randomness variable ε. Adapted from [7]

6. Unit Quaternion

Quaternions are based on complex numbers, and many fundamental geometric and algebraic properties of quaternions originated from complex numbers. In this section, we will define the quaternion and discuss some properties that were used during our experiments.[147]

A quaternion $\mathbf{q}$ is a vector in 4-dimensions that is defined as a hypercomplex number composed of a real part and three imaginary parts $\mathbf{q} = q_0 + q_1\mathbf{i} + q_3\,\mathbf{j} + q_4\mathbf{k}$ (w, x, y, z in some works), where the standard orthonormal basis for R^3 is given by three-unit vectors $\mathbf{i}$ = (1, 0, 0), $\mathbf{j}$ = (0, 1, 0), $\mathbf{k}$ = (0, 0, 1).

Euler angles to unit quaternion conversion

Converting from Euler angles to quaternions is important because quaternions have a number of mathematical properties that make them useful for certain types of computations, such as interpolation and integration. Quaternions also do not suffer from the gimbal lock problem, which can occur with Euler angles when an object's rotation is such that two of the Euler angles become equal[147].

Euler angles can be converted to unit quaternion for, and the common rotations convention used in the Cryo-EM community is the ***ZYZ*** convention, meaning that our rotation by Euler angles is a combination of rotating the biomolecules about the *z*-axis (ϕ), then by rotating about the *y*-axis (θ), and finally rotating about the *z*-axis (ψ), and with this convention, the corresponding unit quaternion rotation is:

$$q_{\phi\theta\psi} = q_{\psi} \otimes q_{\theta} \otimes q_{\phi} \tag{22}$$

Where

$$q_{\phi} = \begin{pmatrix} \cos\frac{\phi}{2} \\ 0 \\ 0 \\ \sin\frac{\phi}{2} \end{pmatrix}, q_{\theta} = \begin{pmatrix} \cos\frac{\theta}{2} \\ 0 \\ \sin\frac{\theta}{2} \\ 0 \end{pmatrix}, q_{\psi} = \begin{pmatrix} \cos\frac{\psi}{2} \\ 0 \\ 0 \\ \sin\frac{\psi}{2} \end{pmatrix} \tag{23}$$

Thus, the corresponding unit quaternion vector is:

$$q_{\phi\theta\psi} = \begin{pmatrix} \cos\frac{\theta}{2}\cos\frac{\psi+\phi}{2} \\ -\sin\frac{\theta}{2}\sin\frac{\psi-\phi}{2} \\ \sin\frac{\theta}{2}\cos\frac{\psi-\phi}{2} \\ \cos\frac{\theta}{2}\sin\frac{\psi+\phi}{2} \end{pmatrix} \tag{24}$$

Chapter 4: Supervised deep learning approaches developed in this thesis

This chapter presents two supervised deep learning approaches developed in this thesis work for analyzing a combination of continuous conformational and rigid-body heterogeneity of biomolecular complexes in cryo-EM single-particle images. The first approach is DeepHEMNMA, which is based on a Resnet deep learning network architecture. We describe DeepHEMNMA and show its performance with synthetic and experimental datasets originally published in one peer-review conference article and one peer-review journal article [24,25]. The second approach (unpublished) is based on a Vision Transformer.

1. DeepHEMNMA

DeepHEMNMA is a supervised convolutional neural network that imitates the conformational variability analysis that can be performed by HEMNMA, in order to speed up the analyses of larger datasets for which HEMNMA is impractical to use as too computationally demanding. HEMNMA combines elastic and rigid-body 3D-to-2D iterative alignments of a flexible 3D reference (atomic structure or EM density map) to match the conformation, orientation, and position of the complex in each single particle image. The elastic matching in HEMNMA combines molecular mechanics simulation (by NMA of the 3D reference) and experimental, single-particle image data analysis. The conformational parameters (amplitudes of normal modes) of the complexes in each single particle image are obtained through the alignment and are processed to visualize the distribution of conformations in the space of lower dimension (typically, 2D or 3D), referred to as the space of conformations. This allows a visually interpretable insight into the dynamics of the complexes, by calculating 3D reconstructions of images with similar structural conformational information from selected regions and by recording movies of the 3D reference's displacement along selected trajectories through the densest regions.

The first article published during this thesis work (a conference article [60]) describes the concepts of DeepHEMNMA, and uses synthetic datasets to show preliminary results. The second article published during this thesis work [119] presents the DeepHEMNMA method and its

performance with synthetic data in detail and demonstrates the effectiveness of DeepHEMNMA on a publicly available experimental dataset EMPIAR-10016.

Methods

DeepHEMNMA workflow is shown in (Figure 26**)** and has three stages. It uses an input set of images split into two subsets (indicated as Image set 1 and Image set 2 in Figure 26) and an input atomic structure or EM map (the reference for HEMNMA rigid-body and elastic alignment). In the first stage, HEMNMA is used to estimate the conformational (normal-mode amplitudes), orientational, and translational parameters for the images in Image set 1, through an iterative normal-mode-based elastic and rigid-body 3D-to-2D alignment of the reference with each single-particle image. In the second stage, the neural network is trained using Image set 1 (referred to as training set from now on) and the parameters estimated by HEMNMA for this set of images; then, the trained network is used to predict the parameters for the images in Image set 2 (referred to as test set from now on). The third stage consists of projecting the predicted conformational parameters onto a low-dimensional space and exploring this space, using a HEMNMA module. In this low-dimensional space, which could be considered as an essential conformational space, close points correspond to images with similar conformations and distant points to images with different conformations. The exploration of this space includes (1) generating animations of the displacement of the reference along the data distribution directions and (2) interactive grouping of images with similar conformations and calculating 3D reconstructions from these groups. In DeepHEMNMA, the parameters predicted by the network and those estimated by HEMNMA can optionally be combined into a single conformational space.

The deep neural network in DeepHEMNMA is a ResNet CNN feature extractor followed by a Multilayer Perceptron (MLP) block. The ResNet feature extractor consists of a ResNet 34 architecture (a 34-layer network) that extracts general relevant features from single-particle images. The MLP block predicts the conformational, orientational, and shift parameters based on the features extracted by ResNet.

In the remaining part of this section, we present the different steps of DeepHEMNMA in more detail.

Stage 1: hemnma estimation of the conformational and rigid-body parameters from the training images (image set 1)

HEMNMA combines cryo-EM image analysis and NMA of the reference. It simultaneously estimates the particle's conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientations and translations) in each particle image. If the reference is an EM map, this EM map must be converted into a collection of 3D Gaussian functions, referred to as pseudoatoms [148], before NMA can be performed.

We next briefly recall the theory of NMA and the iterative elastic and rigid-body 3D-to-2D alignment of HEMNMA, which are mandatory steps at Stage 1 of DeepHEMNMA. The projection of images onto a low-dimensional conformational space and the analysis of this space, which were originally developed for HEMNMA and are now also used in DeepHEMNMA, will be recalled at Stage 3 of DeepHEMNMA.

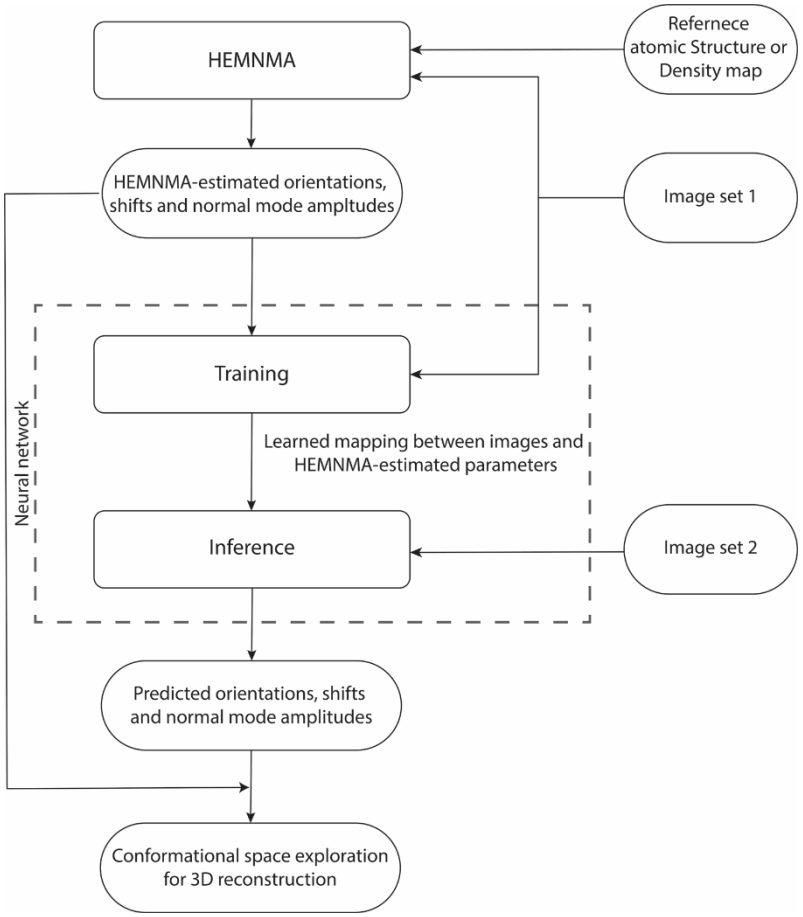


Figure 26 Flowchart of DeepHEMNMA combining HEMNMA and deep neural network methods. It uses an input atomic structure or EM map (reference) and an input set of images split into two subsets indicated as Image set 1 (referred to as training set) and Image set 2 (referred to as test set).

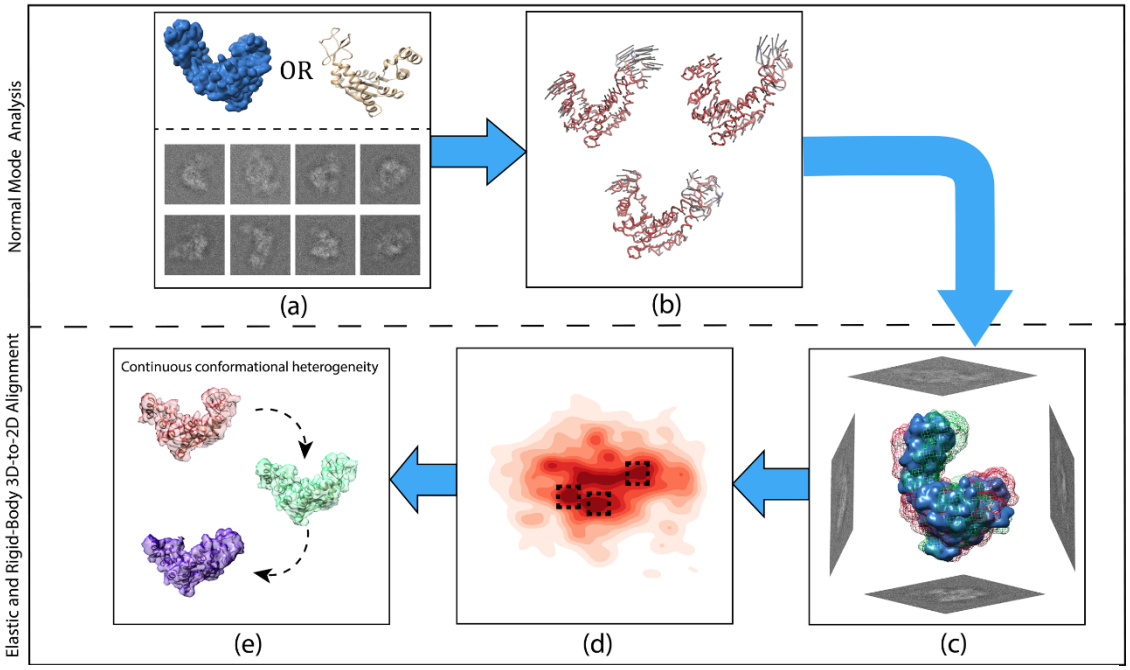


Figure 27 Graphical summary of HEMNMA steps. (a) Input EM map or atomic structure (the reference) and input single particle images. (b) Normal mode analysis of the reference and selection of normal modes (vectors). (c) Elastic and rigid-body alignment of each single particle image with the reference using the selected normal modes. (d) Mapping of single particle images onto a low-dimensional (here, 2D) conformational space in which the reference can be animated (denser regions are marked with a darker red color; close points correspond to images with similar conformations and distant points to images with different conformations). (e) 3D reconstructions from the densest areas in the low-dimensional conformational space shown by squares in (d).

Normal Mode Analysis (NMA)

NMA is based on the so-called elastic network model (ENM) of the molecular system [123], which is a simple and fast method to calculate vibrational modes and has been successfully used to predict biologically relevant motions [149-153]. Normal modes are the vectors along which the system is displaced and are calculated using a harmonic approximation of the potential energy function of the system around a given reference conformation. The reference conformation can be represented with atoms or pseudoatoms (3D Gaussian functions with which an EM map, reconstructed from single-particle images, can be represented [148]). In the ENM, close atoms or pseudoatoms are connected with elastic springs (the interaction radius is a parameter that determines the size of the region beyond which the atom is not connected with other atoms and do not interact with them) [123]. Normal modes are calculated by diagonalizing the Hessian matrix (the matrix of the second derivatives of the potential energy function) [123], which can be made faster in the case of atomic structures by splitting

the structure into blocks of consecutive residues (RTB blocks) that are only allowed to rotate and translate[98]. Normal modes and their squared frequencies are eigenvectors and eigenvalues of the Hessian matrix, respectively. Lower-frequency normal modes describe more collective motions (displacing most of the atoms or pseudoatoms together, synergistically), whereas higher-frequency normal modes describe more localized movements of atoms. Several studies have shown that low-frequency normal modes correspond to functionally relevant biomolecular motions and that conformational transitions can be globally well described using a few low-frequency modes [149-153]. Therefore, only a few low-frequency normal modes are usually selected for further analyses. In general, the six lowest-frequency normal modes are not used as related to rigid-body motions.

The elements of a normal-mode vector provide information on the direction of the displacement of each atom or pseudoatom with this normal mode (in HEMNMA, this displacement is in angstroms, Å, which are the standard atomic-coordinate units). The total number of normal modes and the length of each vector are equal to 3 times the number of atoms or pseudoatoms (the total number of the atomic or pseudoatomic coordinates). Atoms or pseudoatoms are displaced to form a new conformation (model), using a linear combination of normal modes. Normal-mode amplitudes are the coefficients of the linear combination and indicate the contributions of the different normal modes to the global displacement (in HEMNMA, the normal-mode amplitudes have no physical units). NMA allows calculating normal modes (vectors of the displacement), but not the normal-mode amplitudes (amplitudes of the displacement along the vectors). The normal-mode amplitudes can be determined by fitting the conformational model with the experimental data, through numerical optimization of the coefficients of the linear combination of normal modes used for modeling, as described next.

Iterative elastic and rigid-body 3D-to-2D alignment

In this step, HEMNMA iteratively maximizes a measure of similarity between a given particle image and the 2D projection of the reference conformation being elastically modified (using normal modes), rotated, and shifted, until the best elastic and rigid-body alignment is achieved between the image and the projection. It results in a quasi-simultaneous determination of the conformation (the coefficients of the linear combination of normal modes used for the conformational model, i.e., normal-mode amplitudes), orientation (three Euler angles), and position (two in-plane shifts) of the particle in each particle image. The HEMNMA-estimated parameters (normal-mode amplitudes,

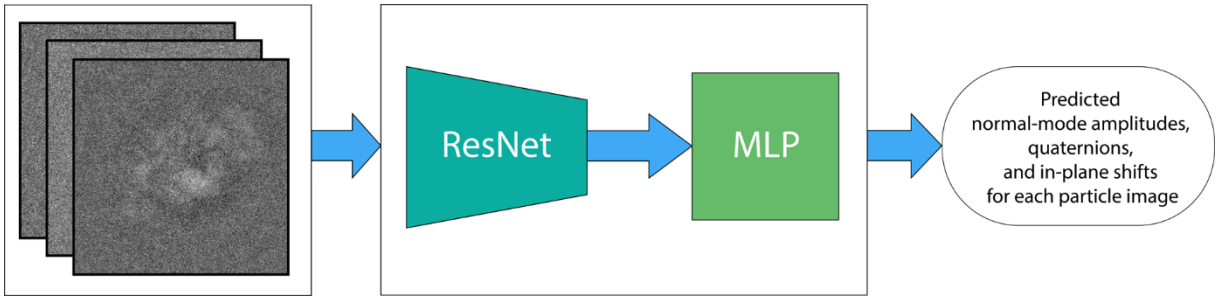
three Euler angles, and two in-plane shifts) are then used to train the neural network at Stage 2 of DeepHEMNMA.

Stage 2: Deep learning of the relationships between the training images and their HEMNMA-estimated parameters (Image set 1) and prediction of the unknown parameters from the test images (Image set 2)

At Stage 2, DeepHEMNMA uses a deep learning neural network, which accelerates the determination of the conformational and rigid-body parameters (normal-mode amplitudes, Euler angles, and in-plane shifts) for large sets of single particle images. This network is trained to learn the complex non-linear relationships between a subset of images (Image set 1) and their conformational and rigid-body parameters estimated at Stage 1 of DeepHEMNMA. The same network architecture is separately trained for each of the three types of parameters (normal-mode amplitudes, angles, and shifts). The three trained network models are then used to predict the three sets of parameters for the remaining subset of images (Image set 2 unseen by the network during the training).

The neural network in DeepHEMNMA is a combination of a ResNet feature extraction block and an MLP estimator (predictor) block (Figure 28). Residual networks allow training of very deep CNNs, by introducing residual blocks (skip connections) in the network architecture [138]. They are very effective as feature extractors and have shown great results in classification tasks [154,155]. DeepHEMNMA uses ResNet 34 CNN architecture, which has 34 layers [138]. In the training phase, ResNet 34 takes a subset of the input particle images (Image set 1) and extracts features that capture the pose (orientations and translations) and the motions of the biomolecular complex in the images. The extracted features are passed onto the MLP that maps them onto each of the three sets of parameters (normal-mode amplitudes, orientations, and translations). The training involves updating the weights of the whole network (ResNet and MLP blocks) to minimize the error of the parameter prediction by the network with respect to the parameters estimated by HEMNMA (mean absolute error type of loss), though Adam backpropagation stochastic optimization method [156]. The MLP takes the input flattened features maps, obtained by ResNet, and captures a multimodal distribution of the particle pose and motion parameters through a stack of 4 fully connected layers. The first 3 layers (1000, 512, 128 nodes, respectively) have a nonlinear function (Rectified Linear Unit) applied to each layer, to model complex nonlinear functions. The last layer has the nodes with linear functions and their number is equal to the number of the output parameters. In the test phase, the trained entire

network model (ResNet and MLP blocks) predicts the pose and motion parameters of the particle from the remaining input particle images (Image set 2). The network is implemented using Python 3.8 and PyTorch 1.8.



.

Figure 28 DeepHEMNMA neural network step. The deep learning neural network is a combination of a ResNet 34 feature extractor (ResNet block) and a 4-layer multilayer perceptron (MLP block). It is trained to map each single-particle image onto the corresponding, HEMNMA-estimated conformational parameters (M normal-mode amplitudes), orientational parameters (3 Euler angles), and positional parameters (2 in-plane shifts) of the particle in the image. DeepHEMNMA converts the Euler-angle representation of the orientation used in HEMNMA into a 4-parameter quaternion representation, which is learned by the neural network internally. The learned quaternion representation of the orientation is then converted back to the Euler-angle representation for the analysis at Stage

DeepHEMNMA uses a unit quaternion representation for the orientation in 3D space, meaning that the three Euler angles estimated with HEMNMA for each single-particle image are converted into the corresponding quaternion and these quaternions are used to train the network. Similarly, the quaternions predicted by the network are converted back to the Euler-angle representation, for use with methods based on the orientation representation with Euler angles (the majority of cryo-EM methods), as the 3D reconstruction method used at Stage 3 of DeepHEMNMA. Quaternions provide an extensive representation of the orientations through a four-tuple system and help overcome the gimbal lock drawback of the representation by Euler angles [147]. For more information, the reader is referred to the recent review [147]. We have compared the performance of our deep learning network using the two representations and found that the network achieves slightly worse results with the Euler-angle representation. Therefore, we decided to use the quaternion representation for our deep learning network.

As the network is trained separately for each of the three types of parameters, the number of outputs in the final MLP layer is different for the three trained models (*M* outputs for *M* normal-mode

amplitudes, 4 outputs for the quaternion representation of 3D orientation, and 2 outputs for the shifts in *x* and *y* directions in the image plane).

We have tested DeepHEMNMA with the ResNet architectures deeper than ResNet 34 (ResNet 50 and ResNet 101 having 50 and 101 layers, respectively) and found that the little improvement of the results obtained with such deeper networks does not to justify the extra time required for their training.

In this work, the neural network training was performed on a 4-GPU computing node (NVIDIA V100, 5120 CUDA cores per GPU card) using a batch size of 2 and 400 epochs of Adam optimization method. The starting learning rate was 10^{-5}. The learning rate was gradually decreased by dividing it by 10 each 80 epochs.

The conformational parameters (*M* normal-mode amplitudes), orientational parameters (3 Euler angles obtained by conversion from 4-parameter quaternions), and translational parameters (2 shifts in *x* and *y* directions in the image plane) predicted at Stage 2 are then analyzed at Stage 3 of DeepHEMNMA, as explained next.

Stage 3: conformational-space dimension reduction and analysis

At Stage 3 of DeepHEMNMA, a dimensionality reduction method is first used to project the set of *M* normal-mode amplitudes predicted by the neural network onto a lower-dimensional space (usually, a 2D or 3D space), which can then be visualized. The dimensionality reduction in DeepHEMNMA is a feature brought by HEMNMA. Several dimensionality reduction methods are available in HEMNMA and we usually use PCA, which is a widely used and intuitively clear dimensionality reduction method. In the lower-dimensional conformational space (Figure 29), each point represents a conformation predicted for a given single-particle image and close points correspond to similar conformations. For each point, the predicted orientation and position of the particle in the image are also available and can be used to calculate 3D reconstructions from groups of images with similar conformations, interactively selected in high-density regions of this space. The

interactive grouping of images with similar conformations in DeepHEMNMA is also a feature brought by HEMNMA.

Beside using 3D reconstructions, the conformations predicted by the neural network can also be inspected by visualizing movies of the motion of the reference along the data distribution directions in this space. Additionally, it can be noted that the dimensionality reduction and further analysis can also be performed for the conformational space that combines the conformations estimated by HEMNMA and those predicted by the network.

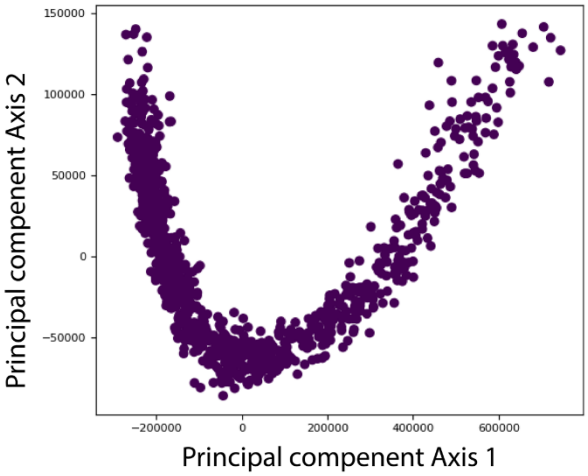


Figure 29 Illustration of a lower-dimensional (here, 2D) conformational space obtained by principal component analysis of the conformational parameters (normal-mode amplitudes) estimated by HEMNMA or predicted by the neural network of DeepHEMNMA. Different points represent different particle images. Each point corresponds to the conformational parameters (normal-mode amplitudes) of the molecular complex in the corresponding sigle-particle image. For each point, the orientation and position of the molecular complex are also available (estimated by HEMNMA or predicted by the neural network) and can be used to calculate 3D reconstructions from interactivelly selected groups of images with similar conformations in the densest regions of this space (not shown in this illustration but in the experiments below).

2. Results

To evaluate the performance of DeepHEMNMA thoroughly, we carefully designed and run several experiments with synthetic datasets of the chain A of adenylate kinase (AK) from the PDB database (PDB:4AKE) [157] and with the experimental cryo-EM dataset of yeast 80S ribosome-tRNA complexes from the EMPIAR database (EMPIAR:10016) [158]. In this section, we describe these experiments and show their results.

Performance of DeepHEMNMA with synthetic data

In this section, we present results obtained with synthetic single particle images affected by noise and contrast transfer function (CTF) of the simulated microscope, to demonstrate the entire DeepHEMNMA protocol and show its accuracy and speed. The dataset was obtained by randomly sampling synthetic continuous conformational transitions, orientations, and positions of AK. The parameters of the synthetic AK conformation, orientation, and position were used as the ground-truth parameters to assess the accuracy of the prediction of these parameters by the neural network. As the network was trained using HEMNMA-estimated parameters, the accuracy of the neural network prediction was also assessed with respect to the HEMNMA-estimated parameters.

Methods used to assess the neural-network prediction (inference) accuracy

The accuracy of the parameters predicted (inferred) by the neural network from images (normal-mode amplitudes, angles, and shifts) was assessed with respect to the ground-truth and HEMNMA-estimated parameters. The metrics to assess the accuracy of the inferred normal-mode amplitudes and shifts was the mean absolute error. The metrics to assess the accuracy of the inferred Euler angles was the average angular distance between the rotated coordinate-system axes (the inferred Euler angles mean the angles obtained by conversion from the inferred quaternions). As a complementary metrics to assess the accuracy of the inferred parameters, we used the root mean squared deviation (RMSD) between the atomic coordinates of AK displaced with the inferred and ground-truth parameters. More precisely, for each synthetic particle image, we calculated the RMSD between the AK atomic coordinates displaced with the inferred and ground-truth parameters using, for the displacement, one type of parameters at a time (normal-mode amplitude, angles, or shift). Then, we averaged the RMSDs over all images, for each parameter type separately. Additionally, we assessed the inference accuracy using 3D reconstructions from the groups of images with similar inferred conformations (the groups selected from different dense areas of the low-dimensional conformational space obtained by PCA of the inferred normal-mode amplitudes). We assessed the quality of each of these 3D reconstructions using Fourier Shell Correlation (FSC) with respect to the map simulated from the atomic model of conformation corresponding to the centroid of the image group used for 3D reconstruction.

Data synthesis: To synthesize the data for the experiment shown in this section, we followed the steps in the flowchart presented in Figure 30. The synthetic conformations were obtained by modifying the atomic AK structure using a linear combination of modes 7-9 (three lowest-frequency non-rigid normal modes), which is an arbitrary choice of normal modes made for this experiment. The linear combination of modes 7-9 was determined by their amplitudes q_7- q_9, respectively, which were randomly sampled from an arbitrary synthetic continuous conformational transition, as follows:

$$q_7(r) = -200 \cdot r, \;\; q_8(r) = 200 \cdot sin\;(\pi \cdot r)\;, \qquad q_9(r) = 200 \cdot cos\;(\pi \cdot r), \qquad (25)$$

where r is a random variable, uniformly distributed between 0 and 1. It should be noted that the hypothetical ground-truth trajectory here, randomly sampled, has a parabolic shape that facilitates a qualitative (visual) inspection of the inference accuracy in the synthetic-data experiments shown in this article. Indeed, a quick visual inspection of the spread of the inferred points around the hypothetical ground-truth trajectory can be an additional indicator of the inference accuracy, beside the quantitative assessment by evaluating the parameter inference errors and 3D reconstructions. The obtained conformations were then converted into density maps [159] (map size 256 × 256 × 256 voxels; voxel size: 0.325 Å × 0.325 Å × 0.325 Å). These maps were rotated and shifted using random angles and shifts (random uniform distribution) and, then, projected onto the image plane of size of 256 × 256 pixels (pixel size: 0.325 Å × 0.325 Å). The total number of synthesized images was 70, 000. It can be noted that the synthesized data are such that the conformation in each particle image can be unique (a different conformation can be present in each different image). The rotation followed the *ZYZ* angular convention, with the first and third rotation angles (around *z*-axis) between 0° and 360° and the second rotation angle (around *y*-axis) between 0° and 180°. The shifts were between -5 and +5 pixels in *x* and *y* directions. Finally, noise and CTF were applied to each synthesized image. In the experiment shown in this section, the SNR was 0.1 and the CTF was simulated for a 200-kV microscope with a spherical aberration of 2 mm, a magnification of 50,000 and a defocus of -0.5 μm. An example of simulated images following Figure 30, with different defocus and SNR values is shown in Figure 31.

Experiment and results: The synthesized set of images was split into a training set of 20,000 images (Image set 1 in Figure 26) and a test set of 50,000 images (Image set 2 in Figure 26). Before running HEMNMA, the images were CTF-phase corrected (phase flip), as it would be done with experimental cryo-EM images. The CTF-phase flipped images were then downscaled to the size of 128 × 128 pixels (pixel size: 0.65 Å × 0.65 Å). The image size reduction was preceded by an antialiasing low-pass filtering, as usually done before image downscaling (in this case, the low-pass cutoff was 1.3 Å). Image size reduction not only speeds up processing, but also reduces noise in images, which generally yields better results, as also observed in our experiments

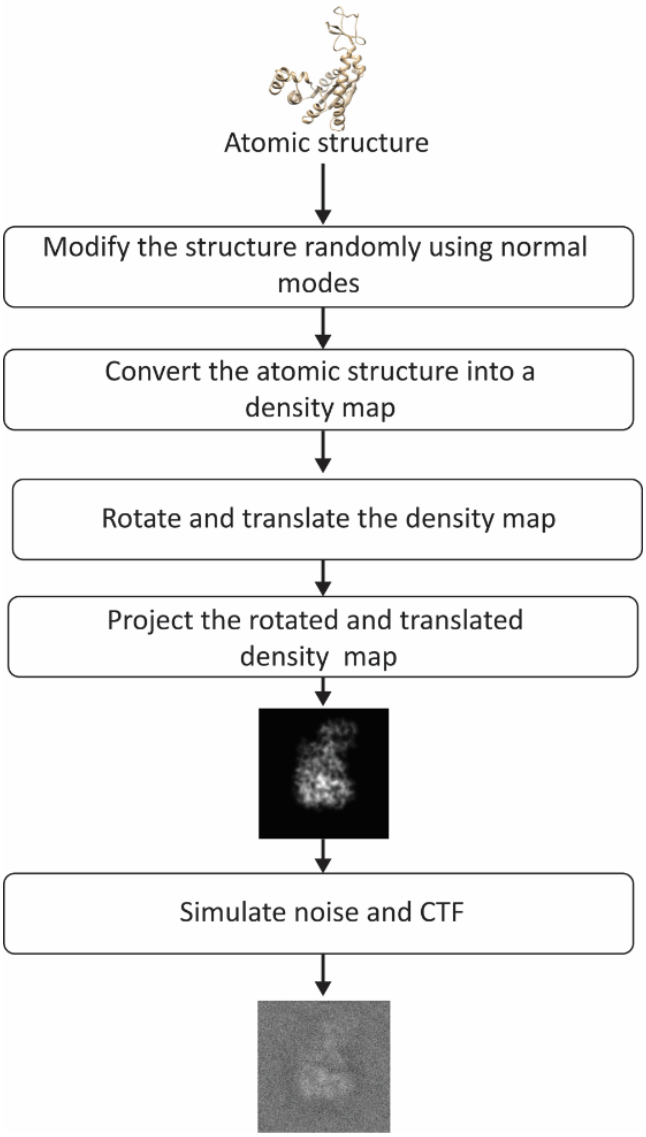


Figure 30 Flowchart of image synthesis for evaluating the performance of DeepHEMNMA. See the text for the details on the synthesis of random normal-mode amplitudes, angles, and shifts.

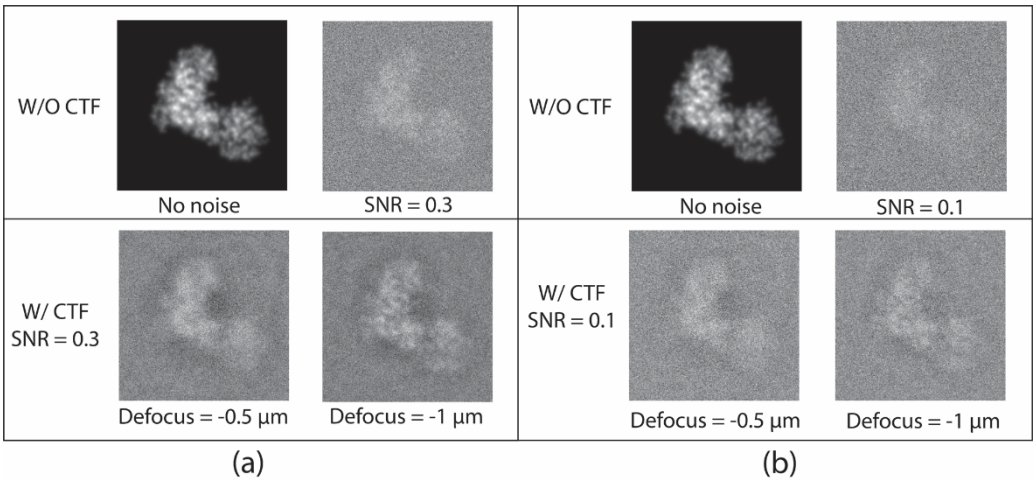


Figure 31 Examples of noisy and CTF-affected images of Adenylate Kinase chain A (same view) synthesized with the SNR of 0.3 (a) and 0.1 (b) and with the CTF defocus of -0.5 µm (bottom left in (a) and (b)) and -1 µm (bottom right in (a) and (b)). Images without noise (top left in (a) and (b)) and without CTF (top right in (a) and (b)) are also shown.

HEMNMA was used to estimate the normal-mode amplitudes, angles, and shifts for the training set of images (20,000 images). The images whose HEMNMA-estimated normal-mode amplitudes were far away from the majority were removed using the Mahalanobis distance measure [160]. The Mahalanobis distance threshold of 3.2 was applied to the normal-mode amplitudes, which resulted in keeping 18,055 images for further processing. The network was trained using 14,055 images (from the kept 18,055 images). From the remaining 4,000 images, we used 2,000 images for tuning the network's hyperparameters (the step referred to as validation in neural network terminology). The remaining 2,000 images were used for quickly testing and comparing the finally trained models and this set of images will here be referred to as small test set. The test set of 50,000 images was used to test the finally selected trained model and will here be referred to as large test set. In this section, we show the results of both tests (with 2,000 and 50,000 images).

Table 3 shows the distance (mean and standard deviation) of each inferred parameter with respect to its ground-truth and HEMNMA-estimated values, obtained using the small test set (2,000 images), and also includes the distance between the HEMNMA-estimated and ground-truth values for the same test set. The distance between the inferred and ground-truth values of each parameter, expressed in RMSD terms, is shown in Table 4. For the metrics used, please recall ***Methods used to assess the neural-network prediction (inference) accuracy*** paragraph in this section. An overlap between the inferred, ground-truth, and HEMNMA-estimated normal-mode amplitudes obtained

using the small test set is provided in **(**Figure 32**)**, which shows that the inferred normal-mode amplitudes follow the ground-truth continuous conformational transition globally well. The distances between the inferred and ground-truth values of parameters obtained using the large test set (50,000 images) and these distances expressed in RMSD terms are shown in Table 3 and Table 4, respectively. These tables show the same range of the parameter inference errors for the small and large test datasets, which indicates that the network model has successfully generalized during the training.

Figure 33 shows a 2D conformational space obtained by PCA of the inferred normal-mode amplitudes. In this space, it is possible to calculate 3D reconstructions either from the reduced-size images (128 × 128 pixels), which were used for training and inference, or from the original-size images (256 × 256 pixels). Here, we demonstrate the reconstructions from the original-size images (using the inferred shifts, after their multiplication by 2, and the inferred angles). Ten 3D reconstructions were calculated from the images in the corresponding ten dense regions of the 2D PCA space. In Figure 33, each reconstructed map is overlapped with the atomic model that corresponds to the centroid of the region used for the reconstruction. Figure 33 also shows the number of images used for the reconstruction and the 0.5-FSC resolution of the reconstructed map with respect to the map simulated from the corresponding centroid atomic model. The resolution is in the range 3-4 Å

Parameter distance	Normal-mode amplitudes							Angles [°]		Shifts X [Å]		Shifts Y [Å]	
	Mean over modes 7-9	Mode 7		Mode 8		Mode 9							
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Inferred *vs.* Ground-truth	7.5	5.4	6.5	8.2	9.2	8.9	10.5	2.5	3.3	0.2	0.1	0.2	0.1
Inferred *vs.* HEMNMA	6.9	5.4	6.7	7.3	9.0	7.9	9.6	1.9	3.4	0.2	0.1	0.2	0.1
HEMNMA *vs.* Ground-truth	6.6	5.7	8.4	6.2	7.2	7.8	7.2	1.0	0.9	0.2	0.2	0.2	0.2

Table 3 Mean and standard deviation (Std) of the distance between inferred, ground-truth, and HEMNMA-estimated values of parameters (normal-mode amplitudes, angles, and shifts) for a small test set of 2,000 synthetic images (the data used for quick tests at the training step

RMSD	Normal-mode amplitudes [Å]		Angles [Å]		Shifts [Å]	
	Mean	Std	Mean	Std	Mean	Std
Inferred *vs.* Ground-truth	0.4	0.2	0.9	1.0	0.3	0.2

Table 4 Mean and standard deviation (Std) of the distance between inferred and ground-truth parameters from ***Table 3*** *(for a small test set of 2,000 synthetic images), but expressed in RMSD terms.*

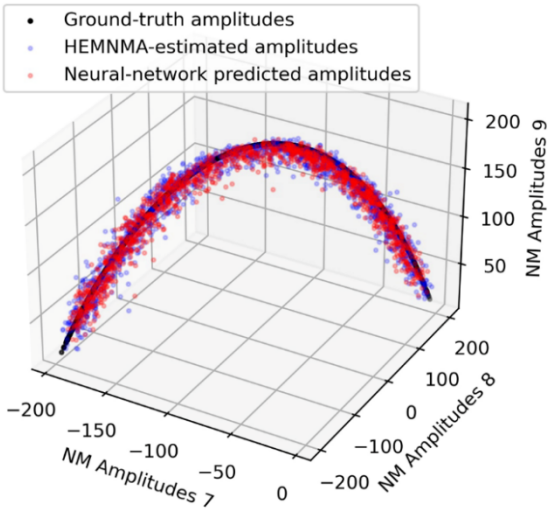


Figure 32 Overlap between inferred, ground-truth, and HEMNMA-estimated values of conformational parameters (normal-mode amplitudes) for a small test set of 2,000 synthetic images. Each point corresponds to an image and a molecular conformation inside it. Close points correspond to similar conformations and vice versa. See also ***Tables 3-4****.*

Parameter distance	Normal-mode amplitudes							Angles [°]		Shifts X [Å]		Shifts Y [Å]	
	Mean over modes 7-9	Mode 7		Mode 8		Mode 9							
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Inferred *vs.* Ground-truth	7.8	6.6	8.7	9.5	10.6	7.3	9.9	2.6	3.4	0.2	0.2	0.2	0.2

Table 5 Mean and standard deviation (Std) of the distance between inferred and ground-truth values of parameters (normal-mode amplitudes, angles, and shifts) for a large test set of 50,000 synthetic images (the data used to test the generalization of the finally trained network on a large set of images).

RMSD	Normal-mode amplitudes [Å]		Angles [Å]		Shifts [Å]	
	Mean	Std	Mean	Std	Mean	Std
Inferred *vs.* Ground-truth	0.4	0.2	0.9	1.2	0.3	0.2

Table 6 Mean and standard deviation (Std) of the distance between inferred and ground-truth parameters from Table 5 (for a large test set of 50,000 synthetic images), but expressed in RMSD terms.

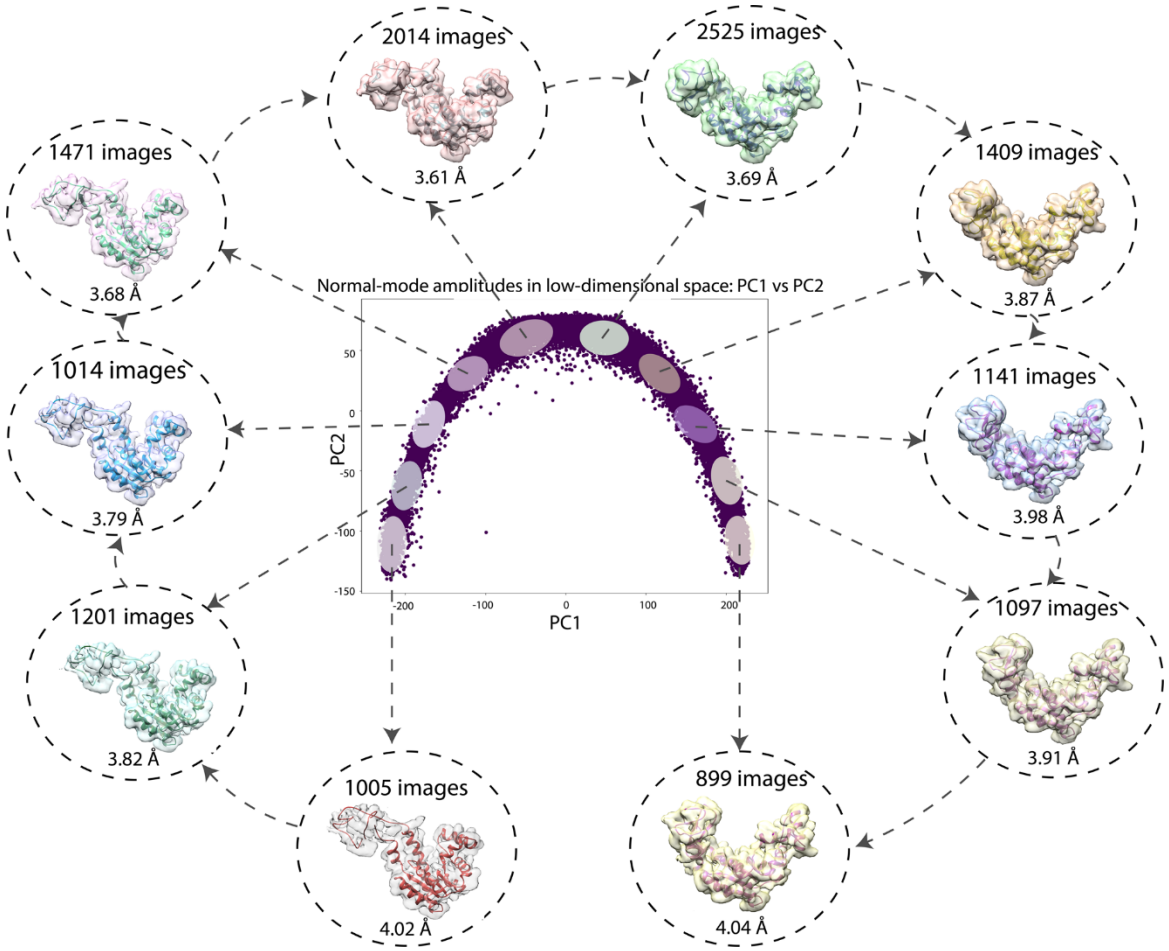


Figure 33 Low-dimensional (here, 2D) conformational space obtained by principal component analysis of the inferred conformational parameters (normal-mode amplitudes) for a large test set of 50,000 synthetic images, together with ten 3D reconstructions from ten different dense regions of this space supperposed with the corresponding atomic models (centroids of the regions). The network training and inference of normal-mode amplitudes, angles, shifts were done using images of size 128 × 128 pixels (for the inferrence accuracy, see Tables 5-6) and the reconstructions were obtained from images of size 256 × 256 pixels. The number of images used for each reconstruction and the 0.5-FSC resolution of the reconstructed map are also shown . Each point in the conformational space corresponds to an image and a molecular conformation inside it. Close points correspond to similar conformations and vice versa.

Speed assessment: DeepHEMNMA is faster than HEMNMA alone and it is even faster for larger datasets. HEMNMA was run on 160 INTEL 2.6 GHz CPU cores. The neural network was run on 4 GPU cards at the training step and on 1 GPU card at the inference step (NVIDIA V100 with 5120 CUDA cores per card). The estimated total number of computing hours needed by DeepHEMNMA for obtaining normal-mode amplitudes, angles, and shifts for 1 million synthetic AK images of size 128 × 128 pixels with 3 normal modes is around 44 times smaller compared to HEMNMA. Indeed, HEMNMA alone would require 64,000 CPU hours, whereas DeepHEMNMA would require 1,232 CPU hours and 233 GPU hours. A detailed comparison of HEMNMA and DeepHEMNMA in terms of processing time is detailed Table 7 Times needed for HEMNMA estimation of all three types of parameters (three normal-mode amplitudes, three angles, and two shifts). The shaded cells of the table show a rough estimation of the expected time.Table 9 Times needed for CNN inference of one of the three types of parameters (normal-mode amplitudes, angles, and shifts). The shaded cells of the table show a rough estimation of the expected time..

HEMNMA 160 INTEL 2.6 GHz CPU cores	1 image	20,000 images	10^6 images
256x256 pixels	8 min	15.6 h	800 h
128x128 pixels	4 min	7.7 h	400 h

Table 7 Times needed for HEMNMA estimation of all three types of parameters (three normal-mode amplitudes, three angles, and two shifts). The shaded cells of the table show a rough estimation of the expected time.

Training 4 NVIDIA V100 / 5120 CUDA cores	6,000 images	14,000 images	50,000 images
256x256 pixels	15 h	28 h	75 h
128x128 pixels	11 h	19 h	55 h

Table 8 Times needed for CNN training to learn one of the three types of parameters (normal-mode amplitudes, angles, and shifts). The shaded cells of the table show a rough estimation of the expected time. Images used for validation (2000 images) are not counted in the number of images displayed in the labels of the table columns.

Inference 1 NVIDIA V100 / 5120 CUDA cores	1 image	2,000 images	50,000 images	10^6 images
256x256 pixels	10 ms	0.3 min	5.6 min	16.7 min
128x128 pixels	5 ms	0.2 min	4.5 min	8.3 min

Table 9 Times needed for CNN inference of one of the three types of parameters (normal-mode amplitudes, angles, and shifts). The shaded cells of the table show a rough estimation of the expected time.

Performance of DeepHEMNMA with experimental data

In this subsection, we show the results of DeepHEMNMA using cryo-EM data of yeast 80S ribosome-tRNA complexes available in EMPIAR database under the accession code EMPIAR-10016 [158].

Dataset: The dataset consists of a stack of single particle images of size 360 × 360 pixels and pixel size of 1.05 Å (normalized so that the average of the image is zero and the standard deviation is 10) and 5 metadata files containing the orientation and translation parameters for 5 image classes obtained in [158] using FREALIGN [161]. Two of these metadata files, with the parameters of 23,726 and 22,369 images, were used in [158] to reconstruct two cryo-EM maps, accessible in EMDB database with the codes EMD-5976 (rotated conformation with 1 tRNA at resolution of 6.2 Å) and EMD-5977 (nonrotated conformation with 2 tRNA at resolution of 6.3 Å), respectively.

Data preprocessing and data splitting for neural network: After inspecting all 5 classes obtained in [158] (quality and number of images in each class as well as 3D reconstruction reproduced for each class), we decided to run DeepHEMNMA only on images used for reconstructing EMD-5976 and EMD-5977 (46,095 images in total). The other 3 classes seemed less "clean" (many images seem to contain different objects than ribosomes) and the number of images in these classes was much smaller. Before running DeepHEMNMA, images were CTF-phase flipped and downscaled to the size of 128 × 128 pixels (pixel size: 2.95 Å). Our preliminary tests with this experimental cryo-EM dataset have shown large angular prediction errors (with respect to HEMNMA estimation) for the network trained using 20,000 images (recall that this is the number of images used to train the network with synthetic data). Therefore, we decided to split the set of 46,095 images as follows: (1) 32,000 images for

training; (2) 2,000 images for validation (adjusting hyperparameters of the network); (3) 12,095 images for testing (large test set), out of which 2,000 images for quickly testing and comparing the trained models (small test set). Images from both FREALIGN classes were uniformly distributed in each of these image subsets.

Reference model and normal mode analysis: The reference model used by HEMNMA to calculate normal modes and to analyze images with these normal modes was a coarse-grain model of the nonrotated conformation, which was made by keeping only Cα and P atoms from the atomic model available in the PDB database under the code PDB:3j78 (the atomic model derived from EMD-5977 map in [158]). The coarse-grain model had 17,082 atoms (Cα and P). Its normal modes were calculated using an RTB block size of 20 residues and an interaction radius of 20 Å.

Selection of normal modes for image analysis with HEMNMA: Regarding the selection of normal modes, an option was to only select the mode that describes the rotation between the large and small subunits of the ribosome, which is often informative enough to separate different ribosome states, as shown in our previous work [9]. However, we decided to include more normal modes to demonstrate, using this experimental dataset, the performance of our deep neural network learning and prediction of a larger number of normal-mode amplitudes. Therefore, in this work, we selected normal modes by analyzing the motion field between the conformations obtained in [158] with FREALIGN. More precisely, we performed flexible fitting of the coarse-grain reference model (obtained from PDB:3j78) into EMD-5976 map, using 7 lowest-frequency non-rigid-body normal modes (modes 7-13), by employing our normal-mode-based 3D-to-3D flexible fitting approach of HEMNMA-3D [132]. The 7 obtained normal-mode amplitudes indicate that all 7 modes contribute to the motion between the two conformations. From this set of modes, we selected 4 modes with the highest contribution (modes 7-9 and 11), among which the mode describing the rotation between the ribosome subunits.

DeepHEMNMA data analysis: HEMNMA was run to analyze images with the four selected normal modes, to obtain the conformations (normal-mode amplitudes), Euler angles, and shifts corresponding to these images, which were then used for the network training. The trained network was used to predict (infer) the normal-mode amplitudes, Euler angles, and shifts for the test images. The inferred normal-mode amplitudes were analyzed by PCA and 3D reconstructions were calculated from groups of images in this space using their inferred Euler angles and shifts.

The 2D PCA space obtained for the set of 12,095 test images was split along the first principal axis into two groups of images, one with 4,741 images and the other with 4,219 images, as illustrated in Figure 34. The two 3D reconstructions obtained from these two groups (Figure 36**A-D**) indicate two different average conformations, with an additional mass in one reconstruction where the additional tRNA is expected (the region indicated by a red ellipse in Figure 36**A**) and without this additional mass in the other reconstruction (Figure 36**B**). The reconstructions obtained using FREALIGN metadata files from EMPIAR-10016 [158] (Figure 36**I-L**) show similarity with those obtained with DeepHEMNMA (Figure 36**A-D** and Figure 36**M-P**). Note however that the two reconstructions from FREALIGN metadata files were obtained using 22,369 and 23,726 images (related to EMD-5977 and EMD-5976 maps, respectively).

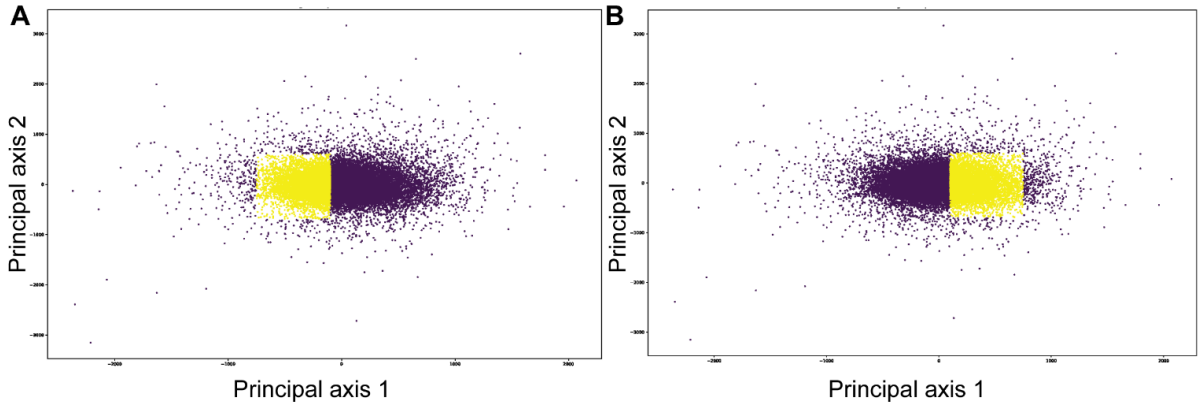


*Figure 34 Two-dimensional conformational space for the EMPIAR-10016 dataset (cryo-EM single particle images of yeast 80S ribosome-tRNA complexes) obtained by principal component analysis of normal-mode amplitudes inferred from 12,095 images, with panels A and B showing two selected groups of images (yellow) used for the 3D reconstructions shown in Figure 36**A** (4,741 images) and Figure 36**B** (4,219 images), respectively. The groups of images were selected automatically using logical operators on the coordinates of the two principal axes (principal axis 1: [-900, -100] in A and [100, 900] in B; principal axis 2: [-900, 900] in A and B).*

Furthermore, we found that the additional mass in the map reconstructed using inferred parameters (Figure 36**A**) could be better resolved if more images were used for this 3D reconstruction. We illustrate this by using a larger set of 22,095 images that were obtained by combining (1) 12,095 images with inferred parameters and (2) 10,000 images with HEMNMA-estimated parameters (from 32,000 images used for network training). The 2D PCA space for this set of 22,095 images, shown in Figure 35, was split along the first principal axis into two groups of images, one with 7,870 images and the other with 6,682 images. The 3D reconstructions from the latter two groups of images (Figure 36**E-H**) are similar to those obtained from the images with inferred parameters (Figure 36**A-D**) but

some details are better resolved in Figure 36**E-H**, such as the additional mass related to tRNA (region marked by red in Figure 36**E**), which is directly linked to the use of more images for the reconstructions in Figure 36**E-H**.

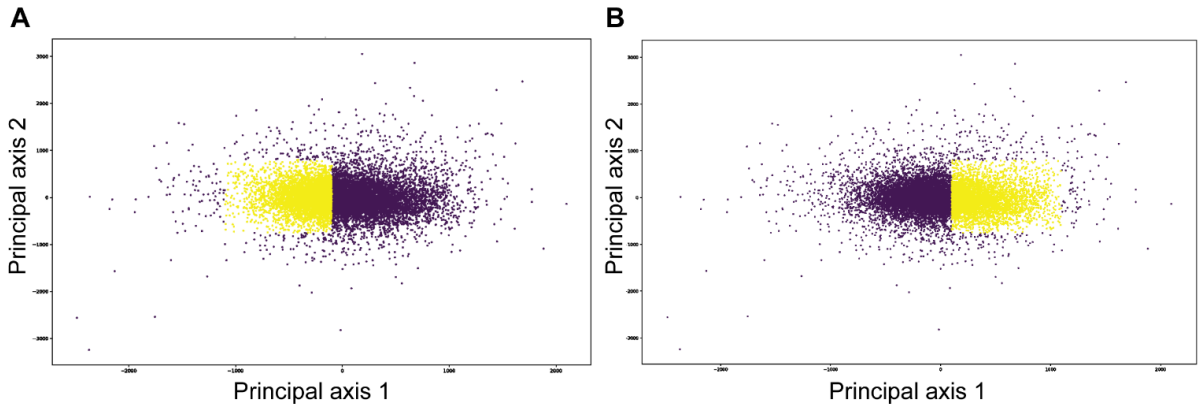


*Figure 35 Two-dimensional conformational space for the EMPIAR-10016 dataset (cryo-EM single particle images of yeast 80S ribosome-tRNA complexes) obtained by principal component analysis of a combination of normal-mode amplitudes inferred from 12,095 images and HEMNMA-estimated from 10,000 images (the total of 22,095 images represented in this space), with panels A and B showing two selected groups of images (yellow) used for the 3D reconstructions shown in Figure 36**E** (7,870 images) and Figure 36**F** (6,682 images), respectively. The groups of images were selected automatically using logical operators on the coordinates of the two principal axes (principal axis 1: [-1100, -100] in A and [100, 1100] in B; principal axis 2: [-900, 900] in A and B).*

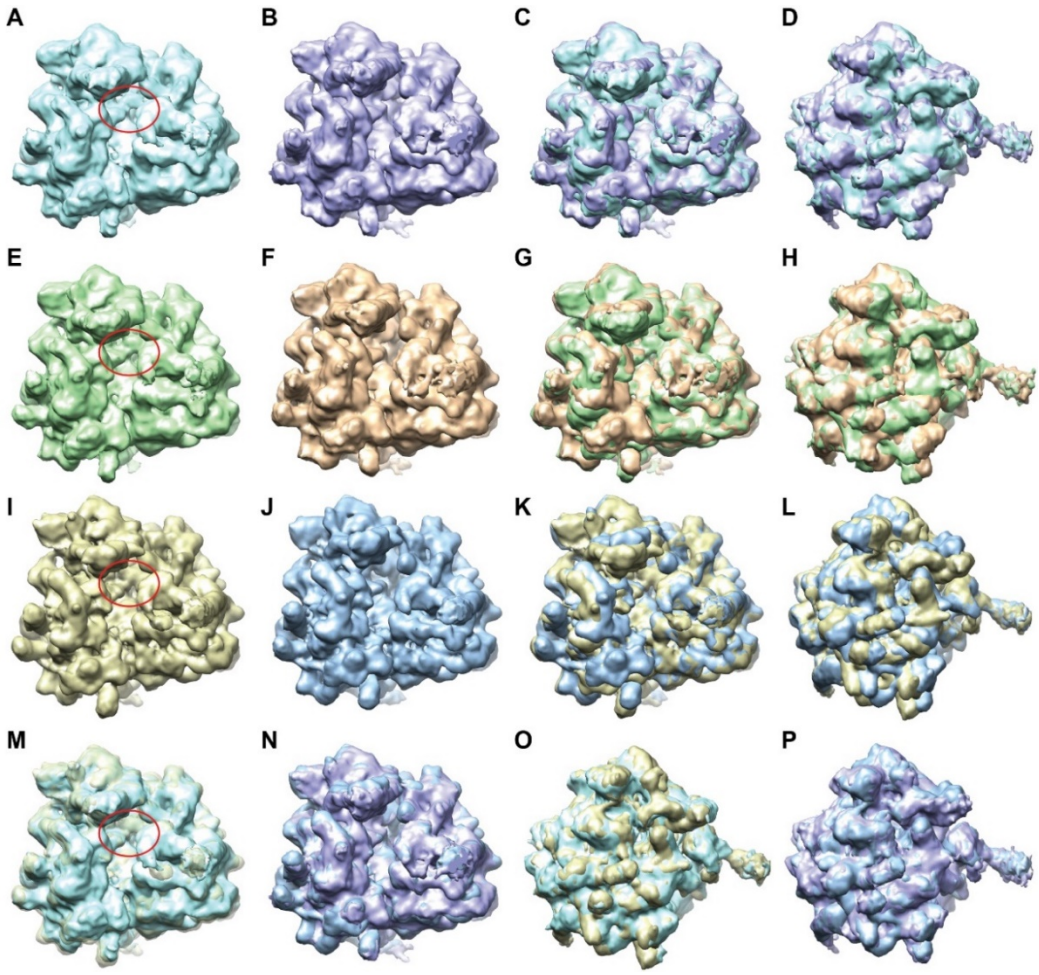


Figure 36 Two average conformations of yeast 80S ribosome-tRNA complexes obtained by 3D reconstruction from EMPIAR-10016 cryo-EM images, with and without additional mass in the region marked with a red ellipse, using DeepHEMNMA and using the original Frealign classification parameters from EMPIAR-10016 dataset.
A-B *Same view of two reconstructions obtained from the conformational space based on 12,095 images with inferred parameters from which groups of 4,741 and 4,219 images were used for the reconstructions.* ***C-D*** *Two views of the superposed reconstructions from A and B.*
E-F *Same view of two reconstructions obtained from the conformational space based on 12,095 images with inferred parameters and 10,000 images with HEMNMA-estimated parameters, from which groups of 7,870 and 6,682 images were used for the reconstructions.* ***G-H*** *Two views of the superposed reconstructions from E and F.*
I-J *Same view of two reconstructions obtained using Frealign parameters for 22,369 and 23,726 images resulting in EMD-5977 and EMD-5976 maps, respectively.* ***K-L*** *Two views of the superposed reconstructions from I and J.*
M-N *Superposition of the reconstructions obtained from images with inferred parameters and those obtained using Frealign parameters (M: overlap between the reconstructions shown in A and I; N: overlap between the reconstructions shown in B and J).* ***O-P*** *Different view of the superposed volumes shown in M-N, respectively.*
The red ellipse shown in panels A, E, I, M indicates the region with the additional mass (corresponding to the additional tRNA), with respect to the same region in panels B, F, J, N, respectively. All surfaces are shown in solid color except for the yellow surface in M that is shown transparent for a better visualization of the additional mass (red ellipse).

The need to use more images for 3D reconstruction in order to better resolve the tRNA could be explained by a larger conformational heterogeneity of the dataset. In Figure 37, we show more extensively the conformational variability using 3D reconstructions from a larger number of groups of images selected along the first principal axis of the 2D PCA space of the 12,095 images used for the inference. The PCA space was split quasi-uniformly in the way to get at least 900 images per group. One can note a variable degree of rotation between the small and large subunits as well as the presence and absence of the additional tRNA over the seven maps reconstructed from 1018, 1148, 1461, 1816, 1771, 975, and 949 images (Figure 37 bottom, from left to right).

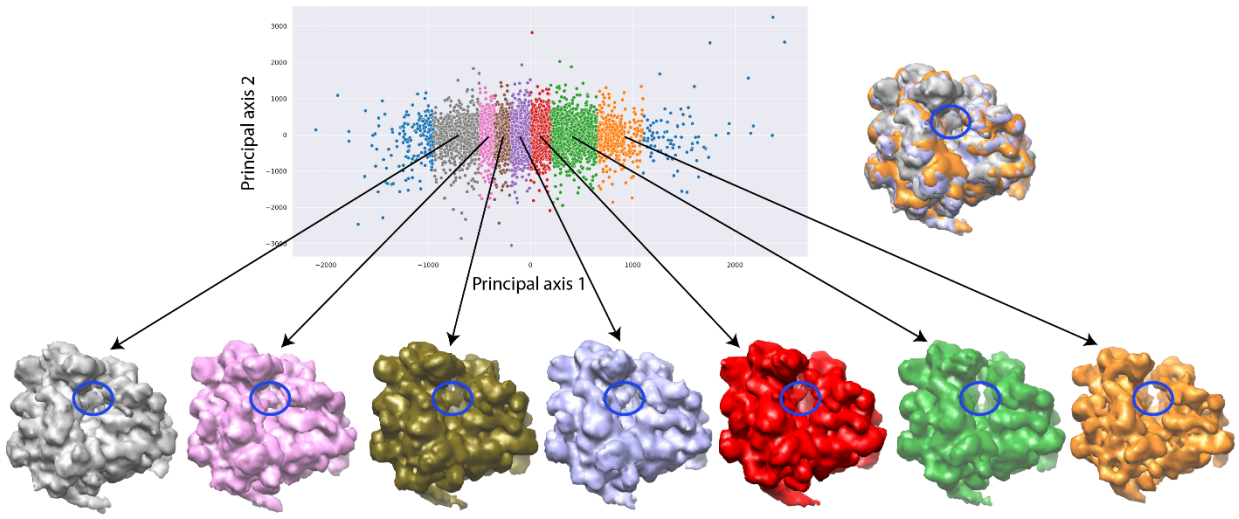


Figure 37 Two-dimensional conformational space of yeast 80S ribosome-tRNA complexes from the EMPIAR-10016 cryo-EM images, obtained by principal component analysis of normal-mode amplitudes inferred from 12,095 images (top, left), with 7 average conformations obtained by 3D reconstruction (bottom) from groups of images selected along the first principal axis, and an overlap of the maps reconstructed from the first (gray), fourth (violet), and seventh (orange) groups (top, right). The far left and far right images (blue) were not used for the reconstructions and the remaining part of the conformational space was split quasi-uniformly in the way to get at least 900 images per group. The reconstructions were obtained from 1018, 1148, 1461, 1816, 1771, 975, and 949 images (bottom, from left to right). The color of the reconstructed map (bottom) corresponds to the color of the group of points in the conformational space (top). The blue ellipse overlapped with the reconstructed maps indicates the region where the additional mass, corresponding to the additional tRNA, is present or absent

This experiment shows that DeepHEMNMA can be useful for extensive analyses of conformational variability of biomolecular complexes, where large sets of experimental single particle images can be obtained. At least 30,000 particle images would be required for the neural network training. To take full advantage of the power of the trained network, one should aim at analyzing millions of single-particle images. The analysis of such large datasets is not practical with conventional methods, whereas it has a low computational cost with trained neural networks.

3. Discussion

This cahpter introduces DeepHEMNMA, a hybrid method using HEMNMA image analysis (based on normal mode analysis) and a deep ResNet-based neural network to study continuous conformational variability of biomolecular complexes from single particle cryo-EM images. The purpose of the neural network is to accelerate HEMNMA-based continuous conformational landscape determination from cryo-EM images. DeepHEMNMA determines the conformational parameters (normal-mode amplitudes) and rigid-body parameters (three Euler angles and two in-plane shifts) of the biomolecular complex in each single particle images. To this goal, HEMNMA is first used to estimate these parameters from a subset of images. Then, the neural network is trained to learn the relationships between this subset of images and its HEMNMA-estimated parameters. The network is a ResNet 34 feature extractor followed by a multilayer layer perceptron. The trained network is then used to predict the parameters from the remaining images (unseen during the training). Finally, the conformational landscape is obtained by mapping the inferred normal-mode amplitudes onto a lower-dimensional space, which allows 3D reconstructions using the inferred angles and shifts. Also, this space allows animations of a model displacement and identification of possible hidden conformations.

We described this new approach and showed its performance with synthetic and experimental data. Using a synthetic dataset and a publicly available experimental dataset, we demonstrated a good generalization capability of the network (no overfitting against the training data), meaning that the trained network is able to accurately predict the conformation, orientation, and position of the molecule in the images that were not used for the training.

DeepHEMNMA has a general purpose and could be useful in analyzing conformational variability of various molecular complexes, as is the case for HEMNMA on which it is based. HEMNMA has been demonstrated on complexes of various sizes and architectures [9]. It is thus expected that DeepHEMNMA performs like HEMNMA on the same complex. However, it should be noted that the network should be trained for each different molecular complex because each different complex will require a separate normal mode analysis, which depends on the shape of the complex.

We trained the network separately for normal-mode amplitudes, orientations, and shifts. This training strategy has the advantage that the number of images used for training can be adjusted for the different types of parameters. Indeed, with experimental data, we observed that learning of orientations requires around twice more images than learning of shifts or normal-mode amplitudes. However, in the future, we will add an option to our open-source DeepHEMNMA software to allow a combined training for all three types of parameters, which is expected to be faster than the separate training for each parameter type, for the same size of the training dataset.

4. Supervised Cryo-ViT

In this subsection, we describe a second supervised deep learning method that was developed in this thesis for continuous conformational variability analysis, which assumes that the given input dataset contains a set of single particle images and the corresponding 3D atomic representation for each particle image. The recently published MDSPACE method [28], based on a combined use of NMA and MD simulation to extract an atomic model from each individual particle image, is able to provide a set of atomic models, which, together with the corresponding set of particle images could be fed to the method described below. The idea would be to make the method below imitate the MDSPACE analysis, so as to be able to predict the atomic models for larger datasets for which MDSPACE is impractical to use as very computing-time demanding.

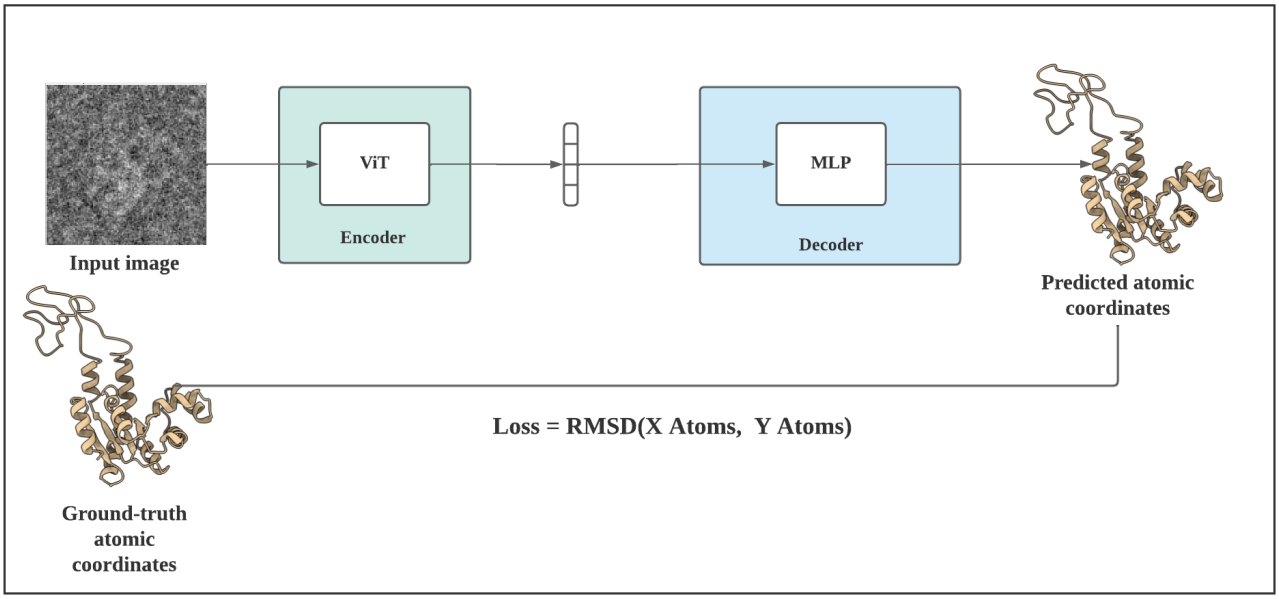


Figure 38 Pipeline of the proposed neural network (supervised Cryo-ViT) for image to atomic coordinates estimation from single-particle images. The network consists of a VAE composed of a ViT encoder, and a 6-layerd MLP decoder.

This thesis implemented a second supervised approach for continuous conformational variability analysis called supervised. Instead of estimating M normal mode amplitudes, in this approach, an encoder-decoder network was developed to estimate the rotated, translated, and displaced 3D atomic coordinates with N atoms given a single particle image. The encoder is a ViT that encodes a given single-particle image represented as patches into a discrete latent representation. The patches are of size 16 × 16 pixels and they are generated internally for each image. The decoder is a feedforward network, with 6 LeakyReLU non-linear layers, and a final linear layer of a size $3N$. The network is trained in a supervised fashion where the encoder is fed a single particle image and encodes it in a latent representation of $256D$. The decoder then estimates the correspondent displaced 3D atomic coordinates. Then the predicted rotated shifted and displaced atomic coordinates are

aligned against a reference atomic structure, and a dimensionality reduction, in this case PCA, on the predicted and aligned atomic coordinates is needed to visualize the conformational space.

Preliminary results on synthetic dataset

This experiment aims to retrieve the ground truth of continuous conformational variability from a synthesized dataset by training the proposed approach in a supervised fashion with various configurations. For instance, our goal is to train the neural network to learn how to project each image back to its corresponding 3D deformed atomic coordinates and retrieve the conformational heterogeneity present in the image set. For this experiment, we used the same image set used in DeepHEMNMA experiment, where we synthesized 10,000 images of the atomic structure of AK (PDB:4AKE) following the same steps shown in Figure 30, used in DeepHEMNMA synthetic data experiment. The synthetic conformations were obtained by modifying the atomic structure using a linear combination of the normal modes (7-9) shown in Eq.(25). Moreover, the deformed rotated and shifted atomic coordinates for each image were generated to permit training of the neural network in a supervised fashion.

This experiment aims to test this approach's ability to retrieve the conformational heterogeneity for each image. Therefore, we trained the neural network on 8,000 images and inferred on the rest 2,000 images. The RMSDs between the inferred and ground-truth atomic coordinates are shown in Table 10**.**

RMSD	Atomic coordinates [Å]	
	Mean	Std
Inferred *vs.* Ground-truth	1.3	0.9

Table 10 Mean and standard deviation (Std) of the distance between inferred and ground-truth atomic coordinates (predicted on 2000 images), expressed in terms of RMSD.

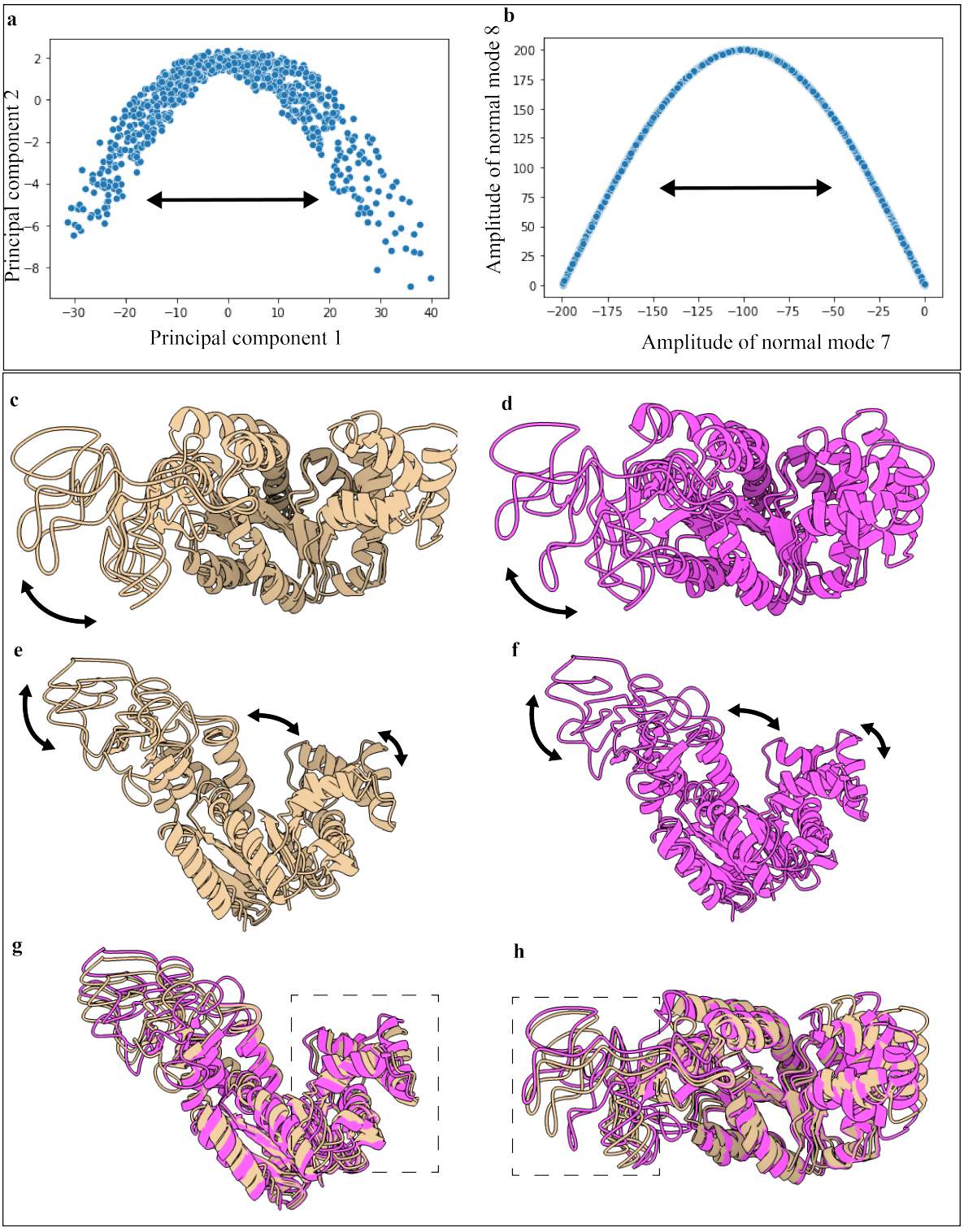


Figure 39 The predicted conformational variability for Chain A of AK synthetic dataset with supervised cryo-ViT. ***a*** *2D Conformational space obtained by PCA of the aligned predicted atomic coordinates.* ***b*** *2D scatter plot of the ground-truth amplitude of normal modes 7-8 used to synthesize the images.* ***c-e*** *Three predicted atomic coordinates selected from the predicted 2D conformational space in two different views* ***a****.*
d-f *Three ground-truth atomic coordinates selected from the ground-truth 2D conformational space* ***b*** *in two different views.*
g-h *superposition of three predicted atomic coordinates and their correspondent ground-truth atomic coordinates in two different views.*

Qualitative results are shown in Figure 39**.** Figure 39.**a-b** shows the 2D conformational space obtained by applying PCA on the predicted atomic coordinates aligned against a reference atomic model and the ground-truth atomic coordinates aligned using the angles and shifts used to synthesize the images. In this space, each point represents an atomic structure, and points close to each other share similar structural information. We selected three points from the predicted conformational space **a** and fetched their correspondent from the ground-truth conformational space **b**, the atomic models are shown in Figure 39.**c-f**. Figure 39.**g-h** show the superposition of the predicted and ground-truth atomic coordinates. Although the neural network captured the motion as illustrated in Figure 39.**b** and the inferred atomic coordinates follow the ground-truth continuous conformational transition globally (Figure 39.**a-b**), the amplitude of the predicted motion is smaller compared to the ground-truth motion as one can notice from the predicted and ground-truth PCA space , which means that the network had some difficulties in estimating the parameters precisely. Further investigations of this problem led to our conclusion that this may come from the difficulty of the network to simultaneously predict the conformational and rigid-body parameters for a given, relatively small set of images.

With the unavailability of such a dataset where each single-particle image is associated with its displaced 3D atomic coordinates, this approach is impractical, and few adjustments of this neural network are needed to exploit it. Hence, the straightforward idea is to benefit from the advantages of unsupervised learning and dispose of the bias of supervised learning. The next chapter extensively reviews the unsupervised version of this neural network and provides all the details of the network and how it is trained with synthetic and experimental datasets. The supervised version of the network is currently referred to as supervised Cryo-ViT, whereas the unsupervised version of it is referred to as unsupervised Cryo-ViT.

Chapter 5: Unsupervised deep learning approaches developed in this thesis

This chapter presents unsupervised cryo-ViT, an unsupervised extension of the supervised deep learning method presented in Chapter 5 for analyzing macromolecular continuous conformational variability in cryo-EM single-particle images. Unsupervised cryo-ViT benefits from the VAE architecture to learn a continuous lower-dimensional conformational space that allows exploring conformational variability, generating motion trajectories and obtaining 3D reconstruction by grouping conformationally similar single-particle images without imposing any prior knowledge on the analysis of the conformational variability. The lower-dimensional conformational space reveals the nature of the heterogeneity, continuous or discrete, that the studied complex may underly.

This chapter presents the unsupervised cryo-ViT method and its performance in analyzing continuous conformational variability. Unsupervised cryo-ViT was tested on simulated datasets and on experimental datasets of the yeast Ribosome 80S and the TMEM16F Calcium-Activated Ion Channel. The results obtained from these experiments are encouraging and coherent with previous findings.

1. Unsupervised cryo-ViT

Unsupervised cryo-ViT is a VAE-based method that learns the motion of flexible protein molecules at the atomic level from single-particle cryo-EM images. To learn how to map between single particle images and the three-dimensional atomic coordinates, unsupervised cryo-ViT uses three inputs (1) a cryo-EM single particle image set; (2) the estimated rigid-body alignment parameters of these images (three Euler angles in ZYZ convention and in-plane shifts, which can be estimated using standard discrete-classification approaches that do not take into account continuous conformational heterogeneity); (3) an initial model of the complex represented by backbone (Cα) atomic coordinates (referred to as “reference atomic coordinates”). The workflow of unsupervised cryo-ViT is illustrated in Figure 40. The network learns the conformational variability by mapping each single particle image to the corresponding displacement of the reference atomic coordinates through a comparison of the projection of the mapped structure with the input image. To interpret the learned conformational heterogeneity, molecular movies, and 3D reconstructions can be computed

(1) in the conformational latent representation of 32 dimensions by reducing its dimension first; and (2) in a low-dimension space of the predicted and aligned atomic structures. The main difference with respect to the method presented in the previous chapter regarding the inputs, is the use of a reference model and initial rigid-body alignment parameters of the images. This method is trained to predict, for each input particle image, the corresponding coordinate displacements with respect to the reference atomic coordinates. This predicted coordinate displacement is a combination of conformational displacement and rigid-body displacement. In the end, the predicted rigid-body displacements are combined with the initial rigid-body alignment parameters to correct for the rigid-body alignment of the images before they can be used for 3D reconstructions.

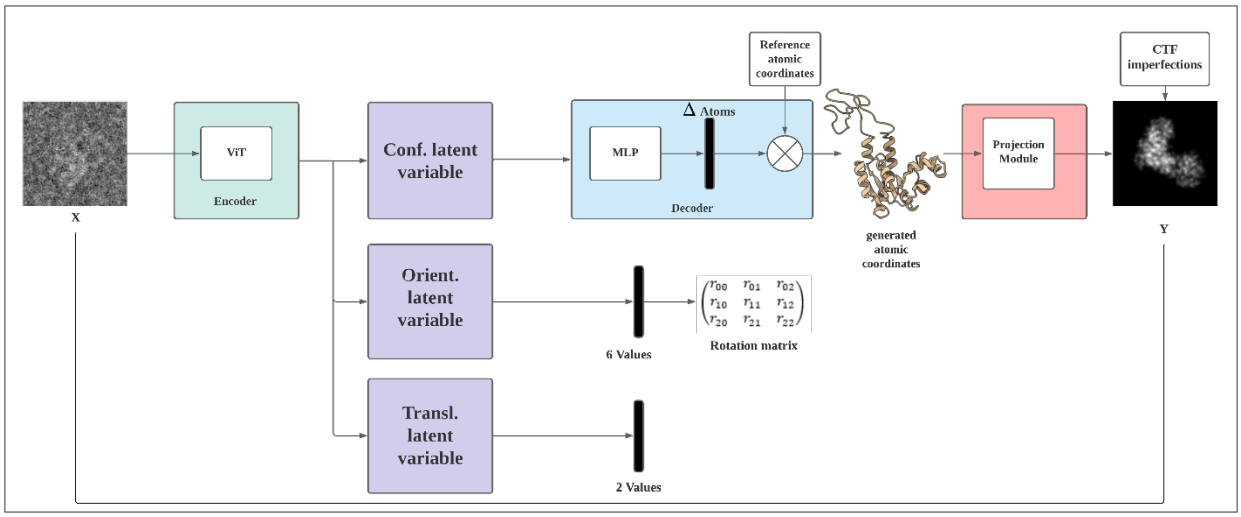


Figure 40 Overview of the proposed unsupervised Cryo-ViT network, a VAE composed of a ViT encoder, three latent variables (conformational, orientational and translational latent variables), a 6-layered MLP decoder, and a non-learnable projection module that takes the estimated atomic coordinates and generates a cryo-EM image affected by the CTF. The conformational latent variable is used to predict the displacement of the atoms while the orientational latent space generates a 3x3 rotation matrix and the translational latent variable predicts the two shifts (x,y) to refine the initial input rigid-body alignment parameters.

VAE for unsupervised learning of continuous conformational variability

Let $X = \{x_1, x_2, x_3, \dots, x_n\}, X \in \mathbb{R}^2$, be a set of cryo-EM single-particle projection images of a specific biomolecular complex, where $\boldsymbol{X}$ follows a mixture of unknown continuous distributions (orientational and conformational heterogeneity), and $S = \{s_1, s_2, s_3, \dots, s_n\}, S \in \mathbb{R}^3$, a set of the corresponding atomic coordinates. Our objective is to determine by projecting back the atomic coordinates s_i given a single particle image x_i:

$$P(s_i) = \widehat{x_i}, \qquad \widehat{x_i} \approx x_i$$

where $P(s_i)$ is the projection function of the i^{th} atomic structure s, and $\hat{\mathrm{x}}_i$ is a 2D projection similar to x_i.

We train the neural network to learn a function $f(x)$: $\mathbb{R}^2 \rightarrow \mathbb{R}^3$, that projects a single particle image back to a displacement (delta) of the reference atomic coordinates, $s \in \mathbb{R}^3$, given the initial non-refined orientations and translation of the image. Cryo-ViT, based on a VAE [144], learns to approximate a posterior distribution $q(z|x)$ to the true posterior distribution $p(z|x)$ of conformations present in input images.

In the first step, the encoder, in this case a ViT, extracts relevant features and approximates the input image's distribution in a lower dimensional latent representation. This latent space embeds the representation of each image in terms of the atomic displacement (delta) with respect to the reference atomic structure. The decoder generates an atomic structure for each single particle image given the latent representation and the reference atomic structure. In the decoder step, a 6-layer MLP takes the latent space as input and generates the delta for each single particle image, and this delta (conformational variation) is then added to the reference atomic coordinates to get a structure that will be projected on an image plane. Before projecting, the obtained structure is rotated and shifted using the input rigid-body alignment parameters. Then the generated atomic structures are projected, and the CTF is applied to the projections. The obtained projection image is compared with the input image.

Encoder: The encoder consists of a vanilla vision transformer (vanilla-ViT) [162]; it encodes an input image represented as a set of patches, typically 16 × 16 pixels, into a feature vector space. It is worth mentioning that the patches are generated internally. In this work, we use a ViT with 16 heads; here heads refer to the number of heads in the multihead attention; the original implementation inspires the choice of 16 heads, and the depth of each one is 6 (the depth here represents the number of Transformer blocks used), and an MLP of size 1024. From the vector space, we sample three latent variables: the conformational latent variable (z_c), the orientational latent variable (z_A), and the translational latent variable z_T where z_c is considered as a lower dimensionality representation of the conformational variability presented in the images and $z_A$and z_T is considered latent variables that refines the error of the initial rigid-body alignment parameters.

Latent variable: The idea of having spereate latent variables z_c, z_o and z_T, as mentioned in the previous section, is to disentangle conformational heterogeneity from orientational and transitional heterogeneity, where the angles are sampled from a continuous gaussian distribution for rigid-body rigid-body alignment parameters refinement. The latent space for conformational variability (z_c) and translations (z_T) is a classical latent space [144] of dimension 32 where z_c is sampled from a standard normal distribution shown in top panel of Figure 41. In contrast to z_c, z_A should be sampled from a symmetrical Gaussian distribution that is invariant to the rotation transformation.

In this case, we use the SO(3)-valued latent space [14], where the classical reparameterization trick is extended to compact the SO(3) Lie group, and therefore we can extract a rotation matrix. Moreover, z_A is sampled from a reparameterizable distribution and mapping it to the Lie group SO(3) using an exponential map.

The top panel of Figure 41 Illustration of the continuous reparameterization trick on SO(3) and the extension of the classical reparameterization trick. Top panel, the steps required to perform the classical reparameterization trick from (a) to (c). Bottom panel, the steps required to extend the

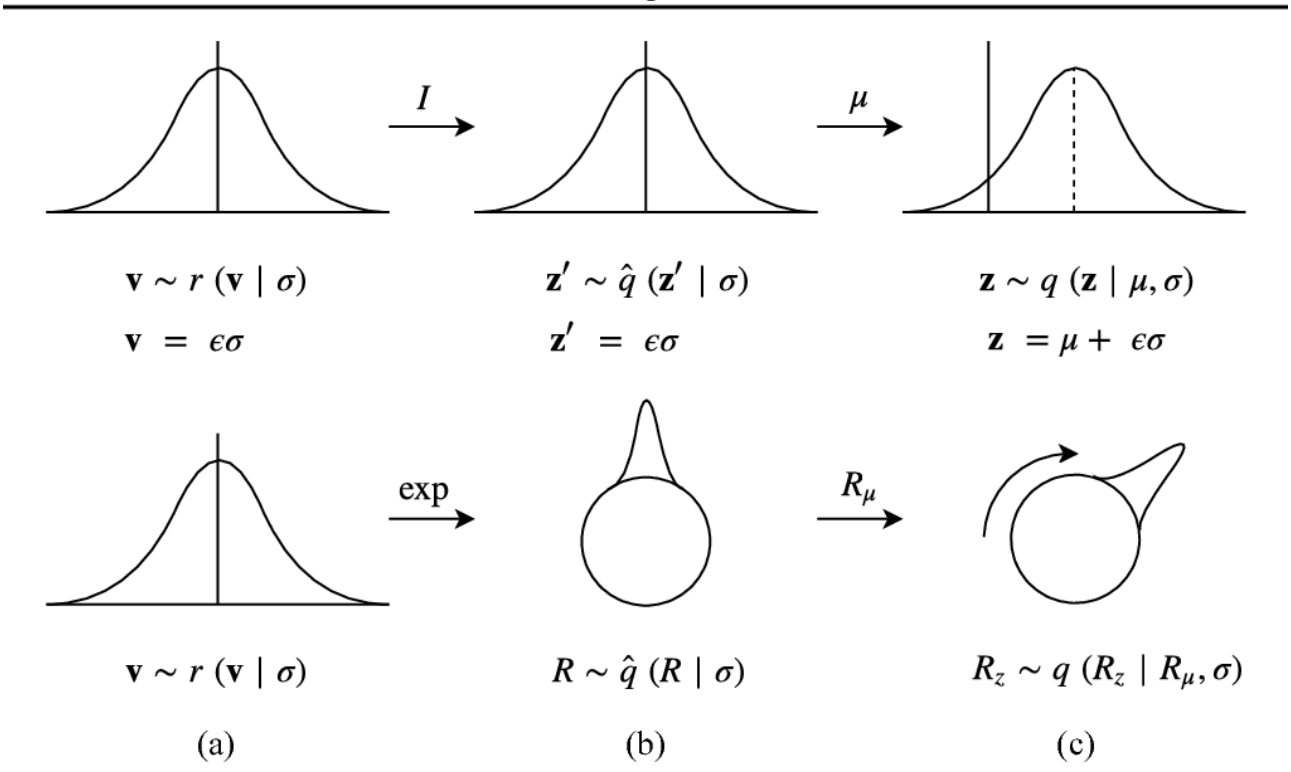


Figure 41 Illustration of the continuous reparameterization trick on SO(3) and the extension of the classical reparameterization trick. Top panel, the steps required to perform the classical reparameterization trick from (a) to (c). Bottom panel, the steps required to extend the classical reparameterization trick from (a) to (c). Adapted from [14].

classical reparameterization trick from (a) to (c). Adapted from [14]. ((a) to (b)) illustrates the building blocks of the classical reparameterization trick used in the conformational latent space z_c, where first the standard deviation σ is multiplied by a noise parameter ϵ, ($v = \epsilon\sigma$), to approximate the intractable true posterior distribution, ($\boldsymbol{v}{\sim}r(\boldsymbol{v}|\sigma)$, and to allow the gradient flow during backpropagation, as explained in Chapter 3, then the mean μ is introduced ((b) to (c)) to the distribution to change its location and fit the input data. The bottom panel of Figure 41 Illustration of the continuous reparameterization trick on SO(3) and the extension of the classical reparameterization trick. Top panel, the steps required to perform the classical reparameterization trick from (a) to (c). Bottom panel, the steps required to extend the classical reparameterization trick from (a) to (c). Adapted from [14]. demonstrates the expansion of the reparameterization trick for the orientational latent variable, where the sampled distribution is first mapped using an exponential map ((a) to (b)); note that in the classical reparameterization trick, the mapping is done using an identity matrix I, then, the pushforward of the mean R_μ, obtained by applying the exponential to the mean, is left multiplied ((b) to (c)) by the mapped distribution, $R{\sim}\hat{q}(R|\sigma)$, to change the location of the distribution. The reader is referred to [14] for further details on the extension of the reparameterization trick.

Decoder: The decoder is a 6-layered non-linear Multi-Layer Perceptron (MLP), in this case, LeakyReLU [163] activation function, where each layer has 256 hidden units followed by a final linear layer. This final layer has M hidden units ($M = number\ of\ atoms \times 3$). For each image, the decoder generates from the approximated posterior distribution $q(z|x)$ of the latent space z_c the delta (conformational variations) and applies it to the initial reference atomic coordinates. z_A serves us here to correct the angular and shift errors by generating eight parameters, six values will be used to generate the rotation matrix [164] (6 parameters are enough to determine the other three parameters of the rotation matrix) and correct the angular errors, and the two values will be used to correct shifts in-plane. The predicted atomic structures must be projected to 2D images to train the neural network in an end-to-end unsupervised fashion.

Projection module: The projection module is a non-trainable differential block. The atomic coordinates are first rotated and translated using the refined angles, represented as a rotation matrix, and in-plane (x and y axes) shifts. The atomic coordinates are then converted into pseudo-atoms where a gaussian function with a standard deviation, typically 1, replaces each atom. The pseudo-atoms are then projected into 2D images along the z-axis, and a CTF model is applied to add imperfections to

the projections to be compared to the input images. It is worth mentioning that additive Gaussian noise was tested; however, it did not affect the results.

The projection module consists of the following:

$$P_{sim} = P_{ideal} * h(r); \tag{26}$$

$$P_{ideal}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2\pi\sigma^2} \sum_{\{i=1\}}^{N} e^{-\frac{1}{2\pi\sigma^2}\left(-(\boldsymbol{x}-\mathrm{a}_{\mathrm{x}}^{\mathrm{n}})^2+\left(\boldsymbol{y}+a_y^n\right)^2\right)}$$

$$\begin{pmatrix} a_x^n \\ a_y^n \\ a_z^n \end{pmatrix} = R^{-1} \begin{pmatrix} x^n \\ \mathrm{y}^{\mathrm{n}} \\ \mathrm{z}^{\mathrm{n}} \end{pmatrix} - T$$

Where σ is the standard deviation of the Gaussian functions, N is the number of atoms in the model with the coordinates $X_n = (x^n, y^n, z^n), n = 1, N$, and a_x^n and a_y^n are the coordinates of the n_{th} rotated and translated atoms, respectively. To add CTF effect $(h(r))$, we use the the same CTF model described in Eq. (3).

Exploring the conformational space: Two ways can be used to visualize the conformational variability. The first one is to explore the conformational landscape at the atomic level; this can be achieved either by (1) using PCA on the predicted atomic structures and creating trajectories of the atomic structure from this PCA low-dimensional space to visualize the motion of the protein complex; or (2) by using the conformational latent variable z_c either in the original dimensions (where each dimension is considered as a single continuous motion) or in a low-dimensional space of z_c (in the next section, we explain what dimensionality reduction methods we tested) to obtain possible trajectories of the protein complex. A second way to explore possible conformations is to perform 3D reconstructions. To do this, one can either group points (1) from the PCA low-dimensional space of the predicted atomic coordinates, or (2) from the low-dimensional space of the latent variable z_c. Each point in the low-dimensional space is considered an image, and close points correspond to similar conformations.

Training: During training, the image log-likelihood and the KL divergence terms must be optimized. In addition to the loss computed from reconstructing images (mean squared error between the 2D projection of the predicted atomic coordinates and the given input image) and the KL divergence term, we add a loss term based on the difference between the center of mass of the reference atomic

coordinates and the center of mass of the predicted atomic coordinates. The latter loss helps significantly in keeping the predictions physically plausible (without this center of mass term, the model does not train successfully). As explained in [145], we weigh the contribution of the KL term with a parameter (β) in order to control the tradeoffs between the quality of predicted atomic coordinates and the posterior approximation. Additionally, we weigh the contribution of the center of the mass term with a parameter $\gamma = 10^{-3}$. Hence, the objective function per image is:

$$-log\left[\frac{1}{N}\sum_{n=1}^{N} P_\theta(x_i|z_n)\right] + \beta KL\left[\, Q_\phi(z|x_i)||P_\theta(z)\right] + \gamma\frac{1}{N}\sum_{n=1}^{N}|MX - MX'| \tag{27}$$

Where the first term, the image log-likelihood, is optimized with the mean squared error between the input image and the image reconstructed by the decoder given z_i ... z_n sampled from the posterior $P_\theta(z|x_i)$, and the second term is the weighted KL divergence between the posterior and the prior, and the third term is the mean absolute difference between the center of mass of the reference atomic structure X $(\mathrm{x},\mathrm{y},\mathrm{z})$ and the aligned atomic structure $X'(\mathrm{x}',\mathrm{y}',\mathrm{z}')$ decoded from the latent variable z, M is the mass of atoms, and it is assumed to be 1. We find that the initial rigid-body alignment parameters must be provided to train the model successfully.

Noteworthy, the results of the experiments with synthetic datasets were obtained using a potential energy term instead of the center of mass, which restricted the movement in the tests with experimental datasets.

Inference: Once the network is trained, we first predict the 32-dimensional latent representation for each new image, then the predicted atomic coordinates (the values of delta added to the given reference atomic coordinates). Afterward, we perform 3D reconstructions or produce molecular movies using either a lower dimensional latent representation or a lower dimensional representation of the predicted atomic coordinates to visualize the conformational changes. However, in our experiments, we use the lower dimensional latent representation because the β parameter selected penalizes the reconstructions and affects the quality of the predicted atomic coordinates.

2. Results

This section presents the experiments and discusses the results of several experiments we conducted on simulated and experimental data.

Performance of the method on simulated data

This section describes the experiments and results obtained on synthetic datasets.

Dataset simulation model

We use an NMA-based image generation approach to generate a realistic synthetic cryo-EM image set. To synthesize the images, we (1) displace the atomic model using a combination of a set of lower-frequency non-rigid-body normal modes (different amplitudes of normal modes were used depending on the complex), (3) rotate and shift each displaced atomic model with random angles and shifts and (4) project to a 2D single particle image, (5) apply a CTF model and add noise as explained in chapter 2.

The rotations are represented by Euler angles that follow the ZYZ angular convention, with the first and third rotation angles (around the z-axis) between 0° and 180° and the second rotation angle (around the y-axis) between 0° and 360° (random uniform angular distribution). The shifts were [-5, +5] pixels in the x and y directions (random normal shift distribution). In the synthetic experiments of this method, we used an SNR of 0.1, and the CTF was simulated for a 200-kV microscope with a spherical aberration of 2 mm and a defocus of -0.5 μm. The size of the synthesized images and the pixel size was 128 × 128 pixels.

Adenylate kinase chain A (4AKE)

This experiment aims to retrieve the ground truth of continuous conformational variability from a synthesized dataset by training the proposed approach (Cryo-VIT) in an unsupervised fashion. Also, we would like to reproduce the results obtained with the network trained in a supervised manner, described in Chapter 4. We first conducted an experiment to test the ability of the method to retrieve the conformational heterogeneity when the orientation and shifts are precise. Therefore, we used the synthesized 10,000 images set of the supervised experiment, illustrated in Chapter 4, generated using amplitudes of normal modes 7-9. However, instead of using the original size of images, 128 × 128 pixels, the images are down-sampled to 64 × 64 pixels to reduce the processing time of the projection module. Then we trained our neural network on 8,000 images for 60 epochs using the ground truth rigid-body parameters and the reference atomic coordinates 4AKE used to synthesize the images and trained to determine only the conformational heterogeneity without performing a refinement of the rigid-body parameters. After training, we inferred on 2,000 images that the network did not use in training to predict the corresponding atomic coordinates.

Then, we applied a PCA on the predicted 32-dimensional latent representation to visualize the 2D conformational space. Figure 42 The predicted conformational variability for Chain A of AK synthetic dataset with unsupervised cryo-ViT. ***a*** *2D scatter plot of amplitudes of normal mode 7-9 used to synthesize the images (a) 2D conformational space obtained from PCA on the predicted 32-dimensional latent space with cryo-ViT from synthetic AK data, colored by the clusters obtained from K-means. b) Atomic coordinates displayed in side, forward and top view obtained from the centroid of each class of K-means.*

.a illustrates the ground-truth amplitudes of normal modes 7-8 used to add heterogeneity to synthesize the images. Figure 42 The predicted conformational variability for Chain A of AK synthetic dataset with unsupervised cryo-ViT. ***a*** *2D scatter plot of amplitudes of normal mode 7-9 used to synthesize the images (a) 2D conformational space obtained from PCA on the predicted 32-dimensional latent space with cryo-ViT from synthetic AK data, colored by the clusters obtained from K-means. b) Atomic coordinates displayed in side, forward and top view obtained from the centroid of each class of K-means.*

.b illustrates the k-means clustering in 10 clusters on the predicted 2D conformational space alongside each cluster's centroid atomic structures. Indeed, the predicted conformational heterogeneity follows the ground-truth continuous conformational transition introduced by a relationship between the normal mode amplitudes when synthesizing the images, as shown in Figure 42 The predicted conformational variability for Chain A of AK synthetic dataset with unsupervised cryo-ViT. ***a*** *2D scatter plot of amplitudes of normal mode 7-9 used to synthesize the images (a) 2D conformational space obtained from PCA on the predicted 32-dimensional latent space with cryo-ViT from synthetic AK data, colored by the clusters obtained from K-means. b) Atomic coordinates displayed in side, forward and top view obtained from the centroid of each class of K-means.*

.b. Although it is possible to produce 3D reconstructions for each cluster, here, we draw ten centroids of each group to visualize the conformational variability at the atomic level for straightforward qualitative assessment (compare with Figure 39).

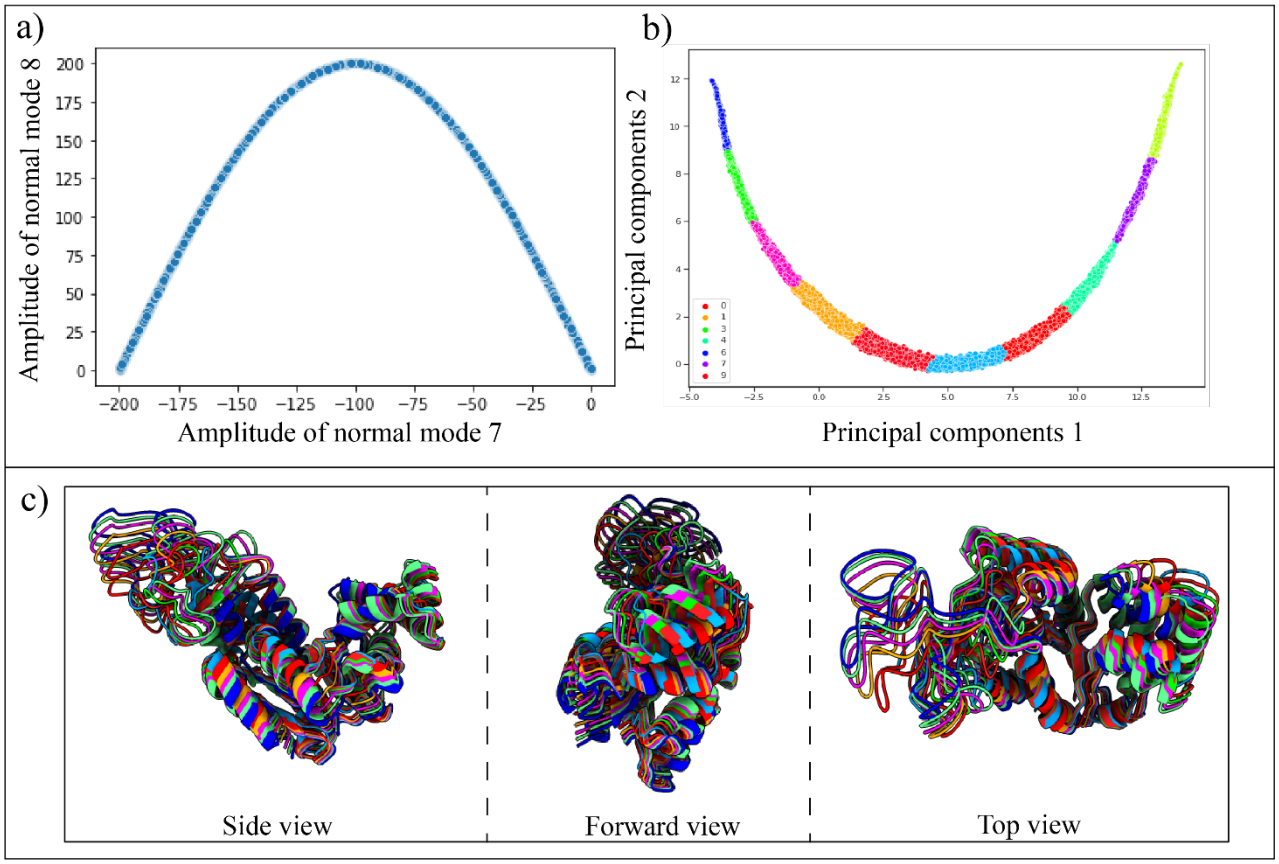


Figure 42 The predicted conformational variability for Chain A of AK synthetic dataset with unsupervised cryo-ViT. ***a*** *2D scatter plot of amplitudes of normal mode 7-9 used to synthesize the images (a) 2D conformational space obtained from PCA on the predicted 32-dimensional latent space with cryo-ViT from synthetic AK data, colored by the clusters obtained from K-means. b) Atomic coordinates displayed in side, forward and top view obtained from the centroid of each class of K-means.*

RMSD	Aligned atomic coordinates [Å]	
	Mean	Std
Inferred *vs.* Ground-truth	0.8	0.6

Table 11 Mean and standard deviation (Std) of the distance between inferred and ground-truth atomic coordinates (predicted on 2000 images), expressed in terms of RMSD.

The RMSDs between the inferred and ground-truth aligned atomic coordinates are shown in Table 11 . Comparing Table 11 and Table 10, it is evident that unsupervised cryo-ViT improved the amplitude of the extracted conformational variability.

Tmr ABC exporter (6RAH)

This experiment aims to retrieve the ground-truth continuous conformational variability from a synthesized dataset by training the proposed approach (Cryo-VIT) in an unsupervised fashion. For this experiment, we synthesized 20,000 images of size of 128 × 128 pixels of Tmr ABC exporter (PDB:6RAH) where the pixel size was 2.154 × 2.154 Å. To synthesize images with conformational heterogeneity, we use the same data synthesis protocol used in DeepHEMNMA using only the two first lowest normal modes (modes 7-8) following:

$$q_7(r) = -150 \cdot r, \;\; q_8(r) = 150 \cdot sin\ (\pi \cdot r) \qquad 28$$

where r is a random variable uniformly distributed between 0 and 1. We conducted two experiments. In the first experiment, the goal was to test the ability of our approach to retrieve the conformational heterogeneity when using the ground truth rigid-body parameters; therefore, we used the ground truth rigid-body parameters and trained our neural network on 10,000 down-sampled images of size 64 x 64 pixels for 60 epochs to determine only the conformational heterogeneity without performing any rigid-body alignment parameters refinement. In the second experiment, the goal was to test the ability of cryo-VIT to determine the conformational variability in the presence of rigid-body parameters estimations errors, and thus, we trained our neural network to retrieve the conformational heterogeneity and refine the rigid-body parameters simultaneously. Hence, we used a random Gaussian variable to introduce an angular error with the following parameters $\mu = 0, \sigma = 3$ to the ground truth angles, and a random Gaussian error with $\mu = 0, \sigma = 2$ to the ground truth shifts.

In both experiments, a continuous distribution of conformational variability is estimated, globally similar to the ground truth (Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent

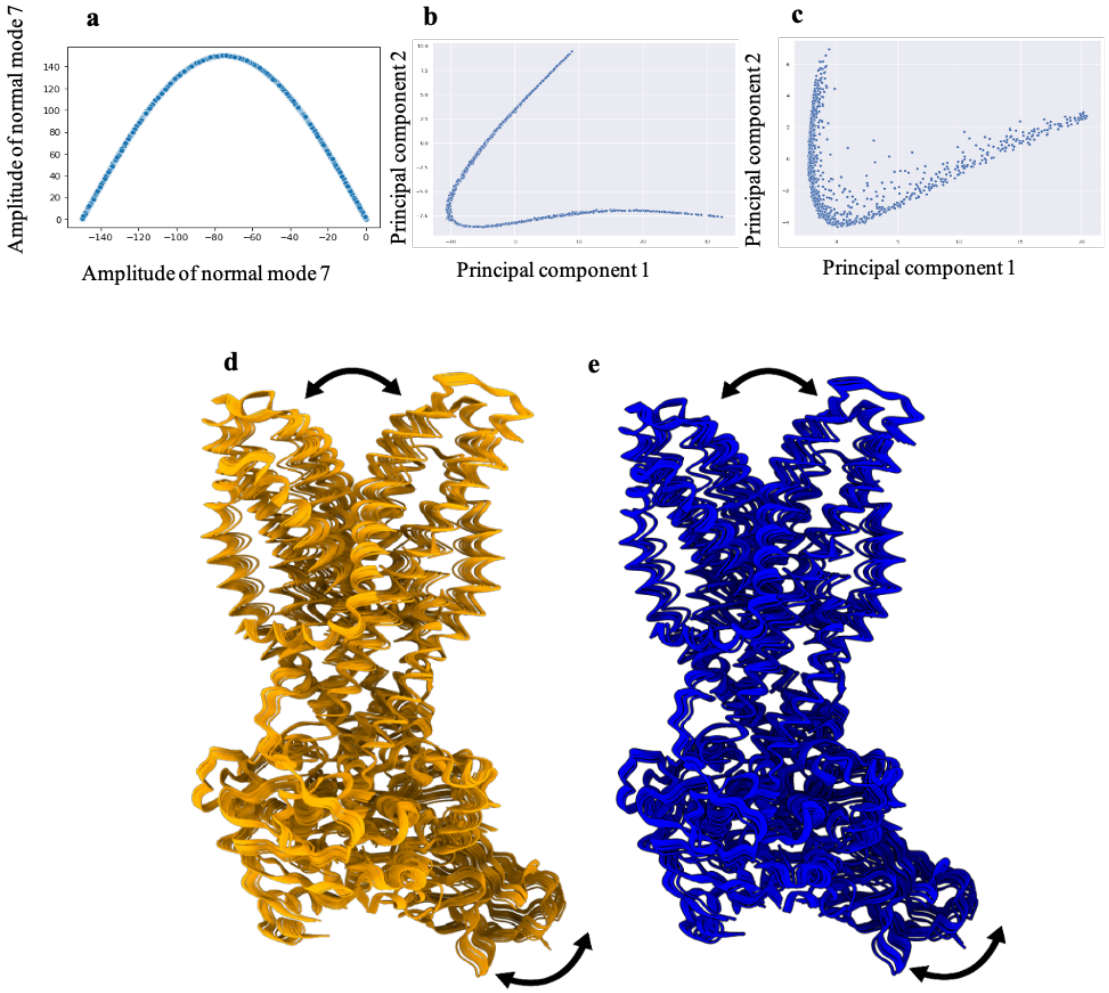


*Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**.*

representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..a) conformational variability introduced by a relationship between the normal mode

amplitudes when synthesizing the images. The results show a slightly less compact distribution with some outliers in the second case, as shown in Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**c**. This was expected due to the errors added to the rigid body parameters.

Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**b** shows the 2D conformational space obtained by applying PCA on the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**b** shows the 2D conformational space obtained by applying PCA on the predicted 32-dimensional latent representation from the network trained to refine rigid-body alignment parameters. Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space

obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**d** shows 50 atomic models sampled from a trajectory of the 2D conformational space where the network was trained using ground-truth rigid-body alignment parameters (Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**b**). Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**e** shows the same 50 atomic models but sampled from the 2D conformational space of the experiment where the network was trained to refine rigid-body alignment while learning the conformational heterogeneity (Figure 43 The predicted conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**c**). comparing the atomic models in Figure 43 The predicted

conformational variability for Tmr ABC exporter synthetic dataset with cryo-ViT. **a** the amplitudes of normal mode 7-8 used to synthesize the images. **b** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained with ground-truth rigid-body alignment parameters. **c** The 2D conformational space obtained by applying PCA to the predicted 32-dimensional latent representation for the network trained to refine the rigid-body alignment parameters.**d** Atomic model sampled from the 2D conformational space **b**. **e** Atomic models sampled from the 2D conformational space **c**..**d-e,** one can notice that the network can predict the conformational variability in the presence of 3 degrees of orientations errors and 2 pixels of shifts.

Performance of the method on Experimental data

To test our method, we consider two publicly available EMPIAR datasets, each exhibiting different types of heterogeneity. The observed variability is well-known in each case, validating the method.

Yeast 80S Ribosome

In this experiment, we use the same Yeast 80s ribosome used in DeepHEMNMA experiment. The goal is to check if we obtain coherent results with what was already published [158] and the finding of DeepHEMNMA, allowing quantitative comparative analysis. The most significant motion in this data is the rotation of the two sub-units, the 40S, and 60S, which is related to the presence and absence of an additional tRNA in one of the conformations. As highlighted in DeepHEMNMA experiments, the orientations and shifts were provided with the images and estimated using FREALIGN [161]. Two atomic models, 3j77 and 3j78, were fitted into the provided density maps, EMD-5977 nonrotated conformation with two tRNA and EMD-5976 rotated conformation with one tRNA, respectively. Since unsupervised cryo-ViT needs the initial rigid-body alignment parameters beforehand, we trained unsupervised cryo-ViT for 60 epochs on 10,000 down-sampled images (size 128 × 128 pixels) using the rigid-body alignment parameters estimated by FREALIGN, together with the provided atomic model 3j78 (as the reference model). After training, we performed the inference on 10,000 new images that were not used during training. As mentioned before, the conformational space can be extracted either from the latent representation or a low-dimensional representation of the predicted atomic coordinates.

Hence, we applied PCA on the 32-dimensional latent representation of the predicted 10,000 images and explored the first two components, as shown in Figure 44 a) 2D conformational space obtained from PCA of the low dimensional latent representation. b) superposition of two 3D reconstructions

obtained by grouping images of two clusters c) 3D reconstruction obtained by grouping images from cluster 1, d) 3D reconstruction obtained by grouping images from cluster 2.
. Figure 44 a) 2D conformational space obtained from PCA of the low dimensional latent representation. b) superposition of two 3D reconstructions obtained by grouping images of two clusters c) 3D reconstruction obtained by grouping images from cluster 1, d) 3D reconstruction obtained by grouping images from cluster 2.

.a illustrates the first dimension of the latent space alongside the obtained 3D reconstructions. The 2D PCA space obtained for the set of 10,000 test images was split along the first principal axis into two groups of images, and two reconstructed maps were produced using 1,000 particles per map following the first axis of the latent space. The resulting structures show a slight rotation of the two sub-units, the 60S and 40S.

The amplitude of the motion extracted by the neural network is relatively small compared to the motion learned by DeepHEMNMA. We suspect that the size of the data set used to train the network is too small to accurately learn the rotation conformation amplitude. Additionally, the image size of 64 x 64 pixels and the size of the atomic model may limit the quality of the features learned by the network. More experiments will be performed in the future to improve the results for this specific dataset.

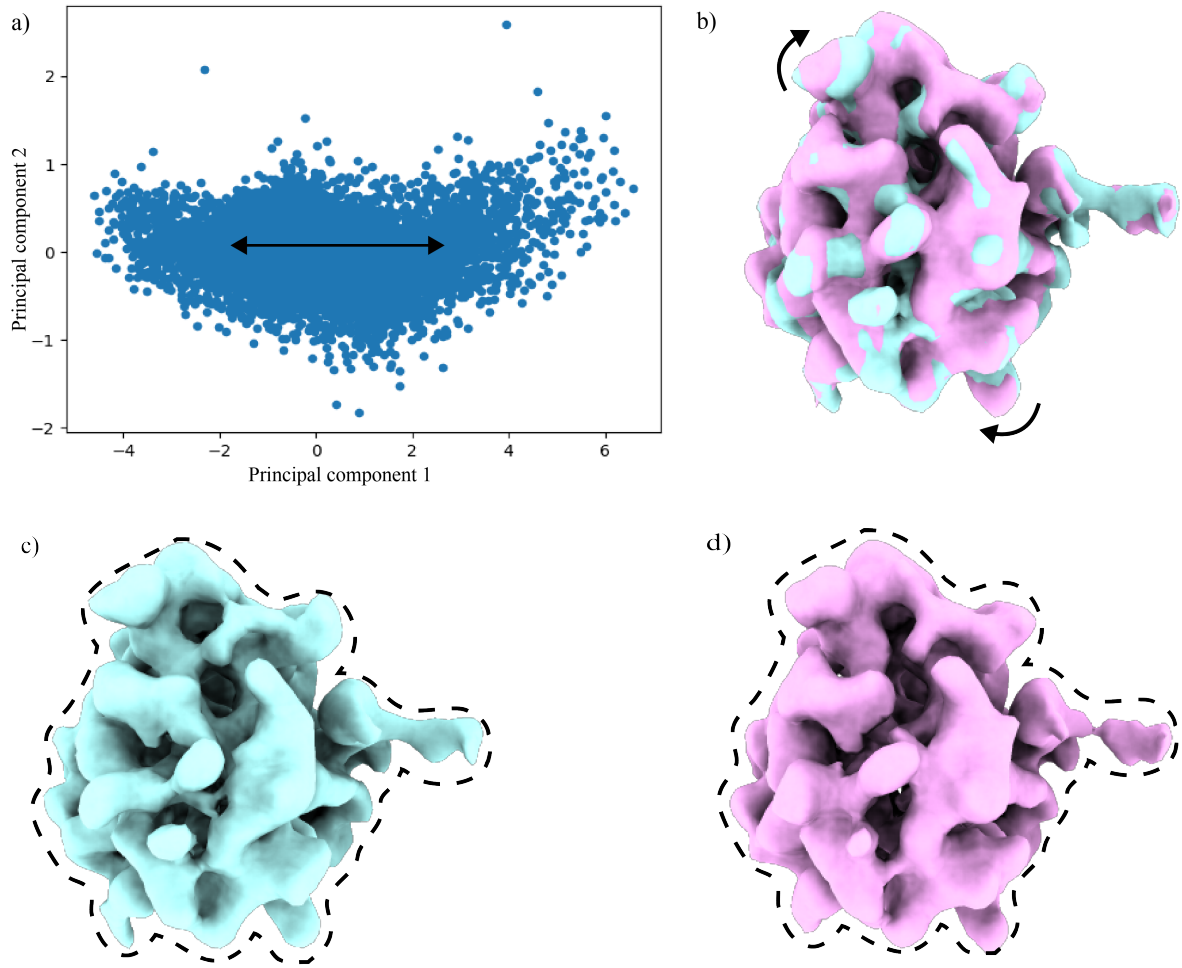


Figure 44 a) 2D conformational space obtained from PCA of the low dimensional latent representation. b) superposition of two 3D reconstructions obtained by grouping images of two clusters c) 3D reconstruction obtained by grouping images from cluster 1, d) 3D reconstruction obtained by grouping images from cluster 2.

Transmembrane TMEM16F

Dataset: In this experiment, we used the public dataset consisting of transmembrane ion channels, namely TMEM16F, in which the channel is activated by calcium binding. To collect such dataset, Feng et al.[165] prepared cryo-EM experiments using different Ca^{+2} conditions to resolve TMEM16F's Ca^{+2} bound and unbound conformations and their effect on lipid scrambling. The dataset consists of a stack of images of 256 × 256 pixels with a pixel size of 1.059 Å. Moreover, RELION [73] was used to process 1.2 million images, and only 13% of the particles (170,827) were kept and used to generate the final reconstruction of the Ca^{+2}-bound conformation provided with the images at a resolution of 3.5 Å. Figure 45 shows a representation of the TMEM16F conformational changes from the closed to open states in the absence and presence of Ca^{+2}.

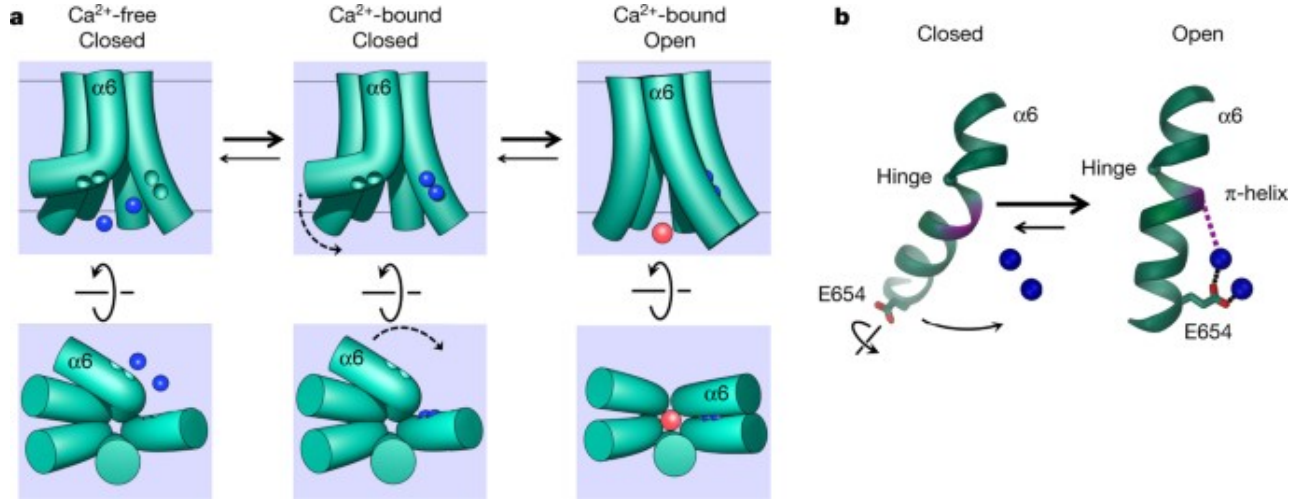


Figure 45 conformations of transmembrane TMEM16F. ***a*** *Schematic representation of the TMEM16 closed and open states in the absence and presence of Ca^{+2}.* ***b*** *Closed and open states of the helix TM6. Adapted from [166]*

In this work, we repeat the same experiment established in Cryo-BIFE [99]. The goal was to extract the Ca^{+2}-unbound state from the 13% particles used to generate the averaged reconstruction of the Ca^{+2} bound conformation, as demonstrated [99]. We used the same setup as the previous systems in this experiment to train the neural network.

Here, we used the rigid-body alignment parameters provided with the data, estimated using RELION. First, we down-sampled the images to 64 × 64 pixels and trained the network using 15,000 images for 60 epochs. As the reference model, we used the atomic model (PDB:6p46) fitted to the provided averaged reconstruction. Furthermore, we use only the Cα atomic model to simplify

and speed up the image projection step. After training, we performed the inference on 20,000 new images (not used during training). Then, we applied PCA to reduce the latent representation from 32 to 2 dimensions to explore the obtained conformational variability. We then grouped the images into two clusters, each reconstructed in 3D. The observed conformational differences are coherent with the known states, as demonstrated in [99] and shown in Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions. **C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B. **F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively. **I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H. **L** the superposed reconstructions from J and K.*
***M** The superposition of reconstruction A with reconstruction G. **N** The superposition of reconstruction B with reconstruction H. **P** The superposition of reconstruction D with reconstruction J. **Q** The superposition of reconstruction E and K.*
O-R superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.

The 2D PCA space obtained for the set of 20,000 test images was split along the first principal axis into two groups of images, one with 4,488 images and the other with 4,394 images. The two 3D reconstructions obtained from these two groups (Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.

***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions. **C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B. **F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively. **I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H. **L** the superposed reconstructions from J and K.*
***M** The superposition of reconstruction A with reconstruction G. **N** The superposition of reconstruction B with reconstruction H. **P** The superposition of reconstruction D with reconstruction J. **Q** The superposition of reconstruction E and K.*
***O-R** superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.**A-F**) indicate two different average conformations, with a closed hydrophilic groove and lipid-conductive cavity (Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from*

EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions.* ***C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B.* ***F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively.* ***I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H.* ***L** the superposed reconstructions from J and K.*
***M** The superposition of reconstruction A with reconstruction G.* ***N** The superposition of reconstruction B with reconstruction H.* ***P** The superposition of reconstruction D with reconstruction J.* ***Q** The superposition of reconstruction E and K.*
***O-R** superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.**A,D**) and an open hydrophilic groove and lipid-conductive cavity (Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.*
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions.* ***C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B.* ***F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively.* ***I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H.* ***L** the superposed reconstructions from J and K.*
***M** The superposition of reconstruction A with reconstruction G.* ***N** The superposition of reconstruction B with reconstruction H.* ***P** The superposition of reconstruction D with reconstruction J.* ***Q** The superposition of reconstruction E and K.*
***O-R** superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.**B,E**). The reconstructions were obtained using RELION metadata files from EMPIAR-10278 (Ca^{+2} bound conformation) and EMPIAR-10279 (Ca^{+2} unbound conformation)[165] (Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.*
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions.* ***C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B.* ***F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively.* ***I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H.* ***L** the superposed reconstructions from J and K.*

M *The superposition of reconstruction A with reconstruction G.* ***N*** *The superposition of reconstruction B with reconstruction H.* ***P*** *The superposition of reconstruction D with reconstruction J.* ***Q*** *The superposition of reconstruction E and K.*
O-R *superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.****G,J*** *and Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.*
A-B *Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions.* ***C*** *the superposed reconstructions from A and B.*
D-E *Top view of the same two reconstruction from A and B.* ***F*** *the superposed reconstructions from D and E.*
G-H *Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively.* ***I*** *the superposed reconstructions from G and H.*
J-K *Top view of the same two reconstruction from G and H.* ***L*** *the superposed reconstructions from J and K.*
M *The superposition of reconstruction A with reconstruction G.* ***N*** *The superposition of reconstruction B with reconstruction H.* ***P*** *The superposition of reconstruction D with reconstruction J.* ***Q*** *The superposition of reconstruction E and K.*

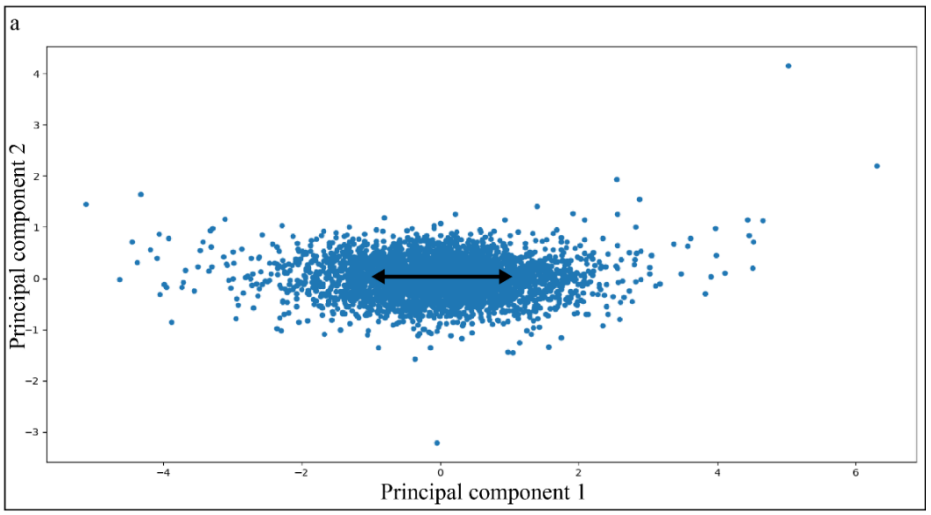
O-R superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.**H,K**) show similarity with those obtained with our neural network. It is worth mentioning that the two reconstructions from RELION classification were obtained using 170,827 and 324,627 images (related to EMD-20244 and EMD-20245 maps), respectively, whereas the 3D reconstructions were obtained using a much smaller number of images with our method.

Moreover, we validated the findings of [99], where the network successfully extracted the Ca^{+2}-unbound state from the dataset that mainly was used to produce the Ca^{+2}-bound state. However, the bending of α-helices representing the channel bending inside the membrane as demonstrated in [165] and presented in EMD-20244 and EMD-20245 maps, were not resolved due to the low-resolution maps reconstructed from the predicted conformational space (Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.
A-B *Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions.* ***C*** *the superposed reconstructions from A and B.*
D-E *Top view of the same two reconstruction from A and B.* ***F*** *the superposed reconstructions from D and E.*
G-H *Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively.* ***I*** *the superposed reconstructions from G and H.*
J-K *Top view of the same two reconstruction from G and H.* ***L*** *the superposed reconstructions from J and K.*
M *The superposition of reconstruction A with reconstruction G.* ***N*** *The superposition of reconstruction B with reconstruction H.* ***P*** *The superposition of reconstruction D with reconstruction J.* ***Q*** *The superposition of reconstruction E and K.*

O-R superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.**A-B**). As the channel is covered by the nanodisk, more images are needed to reconstruct high-resolution maps to visualize the α helices bending, more specifically, the TM6 α helix.

This experiment demonstrates that the neural network can be useful for extensive analyses of conformational variability of biomolecular complexes but that larger sets of images than those used

here may be needed to train the neural network for better accuracy of learning and higher resolution 3D reconstructions of the different conformational states.



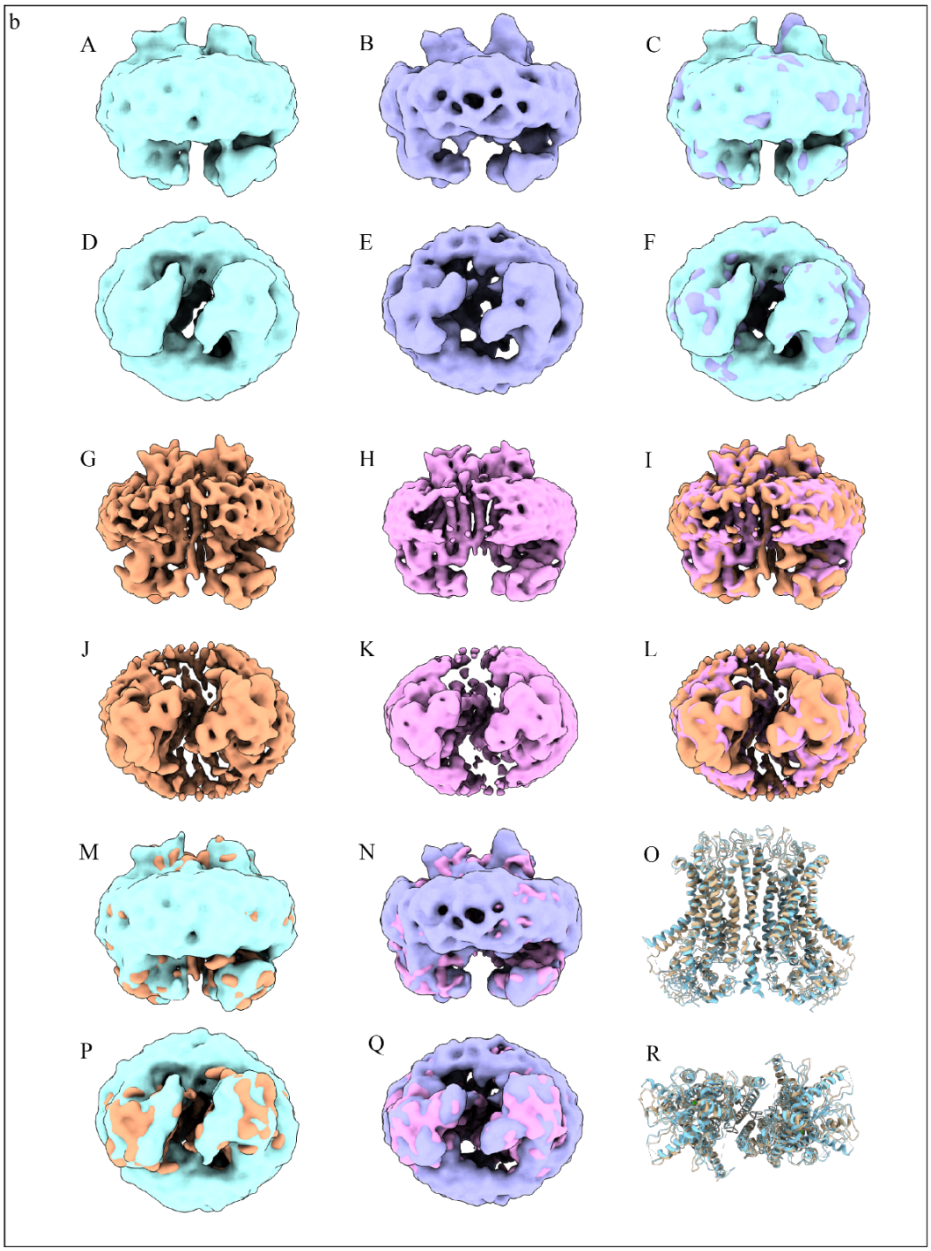


Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions. **C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B. **F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively. **I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H. **L** the superposed reconstructions from J and K.*
***M** The superposition of reconstruction A with reconstruction G. **N** The superposition of reconstruction B with reconstruction H. **P** The superposition of reconstruction D with reconstruction J. **Q** The superposition of reconstruction E and K.*
***O-R** superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.*

3. Discussion

This chapter introduced cryo-ViT, an unsupervised end-to-end neural network approach to studying the continuous conformational variability of biomolecular complexes from single-particle cryo-EM images. The neural network aims to learn a continuous conformational landscape from cryo-EM images. Unsupervised cryo-ViT determines the corresponding atomic coordinates variation of the biomolecular complex for each single-particle image. Initial rigid-body alignment parameters can be obtained with a standard discrete classification method. Then, these rigid-body alignment parameters are refined while training the neural network to learn the atomic coordinates variation from images. Unsupervised cryo-ViT is a VAE, where the encoder is a ViT feature extractor, and the decoder is an MLP. The trained network is then used to predict the parameters from the remaining images (unseen during the training). Finally, the conformational landscape is obtained by mapping either the inferred atomic coordinates onto a lower-dimensional space or by mapping the latent representation onto a lower-dimensional space, allowing 3D reconstructions using the refined angles and shifts. Also, this space allows animations of a model displacement and identification of possible conformations.

We detailed this new approach and highlighted its potential with synthetic and experimental data. The validation of unsupervised cryo-ViT is still a work in progress and requires more investigation (e.g., the use of larger datasets for training and larger images). Unsupervised cryo-ViT software will be publicly available as part of ContinuousFlex[26] plugin for Scipion V3[167].

Chapter 6: Software contributions

DeepHEMNMA software is publicly available as part of the cryo-EM/ET data processing of the open-source software package ContinuousFlex[26,168], which is also available as a plugin of Scipion software[167] whose backend is based on Xmipp software[169,170]. On the other hand, Cryo-ViT is under in-depth tests and will be integrated into ContinuousFlex after finalizing all the experiments (validation).

This chapter presents the software development contributed to ContinuousFlex software package during the work on this thesis.

1. ContinuousFlex

ContinuousFlex is an open-source software package designed for conformational heterogeneity analysis of macromolecules by conducting an exhaustive analysis of their continuous conformational variability in cryo-EM/ET data. It also provides methods for MD-simulation-based flexible fitting of cryo-EM maps and cryo-EM images with atomic models. ContinuousFlex is available as a plugin of Scipion. This pluginization allows for better maintenance, faster development, and more frequent releases of bug fixes and developed methods. As a plugin of Scipion, ContinuousFlex enables reproducible research, as all the data processing steps used in experiments are automatically stored on the disk (with their parameters) and can be reproduced at any moment using the same or modified parameters. Additionally, the project containing all the data processing steps can be directly uploaded to EMPIAR, as allowed by Scipion[168].

2. HEMNMA and DeepHMEMNA in ContinuousFlex

ContinuousFlex allows performing all the steps of HEMNMA, as shown in Figure 46 2D conformational space obtained from PCA (first two component shown) of 32-dimensioanl latent variable predicted for 20,000 cryo-EM images of TMEM16F complexes from EMPIAR-10278, and two average conformations of TMEM16F obtained by 3D reconstruction by Cryo-ViT and original RELION classification results.
***A-B** Forward view of two reconstructions obtained from the conformational space from which groups of 4,488 and 4,394 images were used for the reconstructions. **C** the superposed reconstructions from A and B.*
***D-E** Top view of the same two reconstruction from A and B. **F** the superposed reconstructions from D and E.*
***G-H** Forward view of two reconstructions obtained using RELION parameters for 170,827 and 324,627 images resulting in EMD-20244 and EMD-20245 maps, respectively. **I** the superposed reconstructions from G and H.*
***J-K** Top view of the same two reconstruction from G and H. **L** the superposed reconstructions from J and K.*

M *The superposition of reconstruction A with reconstruction G.* ***N*** *The superposition of reconstruction B with reconstruction H.* ***P*** *The superposition of reconstruction D with reconstruction J.* ***Q*** *The superposition of reconstruction E and K.*

O-R superposition of atomic structures (PDB: 6p46 and PDB:6p47) of TMEM16F.. The steps performed through the graphical interface are as follows:

1- Importing an atomic structure or an EM map. If an EM map is imported, it is converted to a pseudoatomic structure.

2- Performing NMA.

3- Reminding the user that HEMNMA software may also be used for NMA and visualization only.

4- Importing images, synthesizing test images, and resizing the images. We note here that the protocol "Synthesize particles" provides test data for the method and is not a part of data processing. Whereas the protocol "Resize images" is an optional data preprocessing that can be useful if the size of images is too large.

5- Performing rigid-body and elastic alignment of images based on NMA.

6- Obtaining and analyzing the conformational space based on the alignment result.

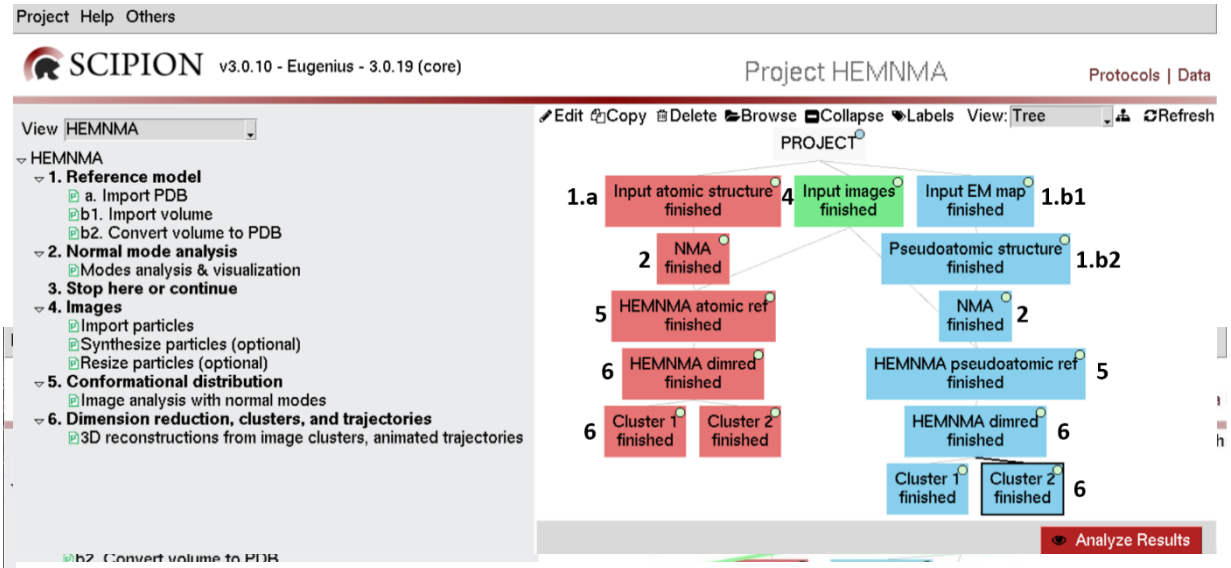


Figure 47 The graphical interface of HEMNMA in ContinuousFlex. The green box represents the input single-particle images. The red branch is for an input reference choice as an atomic structure, whereas the blue branch is an input reference choice for an EM map.

Figure 48 Graphical interface of HEMNMA in ContinuousFlex showing the use of deep learning in combination with HEMNMA for faster data processing (DeepHEMNMA method). Green box: input dataset of single particle images. Red boxes: HEMNMA steps. Blue boxes: Additional steps required to accelerate HEMNNA with deep learning.

DeepHEMNMA is part of the Continuousflex plugin, incorporated via optional steps of HEMNMA data processing pipeline, as shown in Figure 2. These optional steps allow:

• Splitting an input single particle dataset into training/inference sets

• Performing HEMNMA (up to step 5/6 in Figure 1) on the training set.

• Training DeepHEMNMA on the results of HEMNMA.

• Using DeepHEMNNA to produce results on the inference set.

• Obtaining and analyzing the conformational space based on HEMNMA and DeepHEMNNA results for the complete input set.

Chapter 7. Discussion and perspectives

Among the various approaches proposed in the literature, deep learning has emerged as a technique that could revolutionize cryo-EM data analysis. Generative deep learning was recently extensively used in the context of continuous conformational variability analysis, in combination with standard cryo-EM methods or as a standalone approach. Although generative deep learning helped overcome several challenges of standard cryo-EM methods and offered flexibility in dealing with discrete and continuous conformational variability, it still suffers from a few challenges that limit its potential. One of the main drawbacks of current deep learning approaches is a manual selection of hyperparameters, such as the learning rate, the activation functions, and when to stop the training, which is prone to subjectivity. Hyperparameter optimization through automated machine learning (AutoML) approaches[171,172], such as grid search[173] or Bayesian optimization[174] can be helpful in to search for the best hyperparameter values to achieve the highest accuracy.

With the growing interest in the on-the-fly [175] approaches, several single-particle cryo-EM steps required before conformational variability analysis have been automated to minimize human interference. However, most conformational variability analysis approaches, including generative deep learning, are time-consuming iterative processes, and in the majority of cases, human interference is needed to validate the findings. Therefore, fully automating the conformational variability analysis step is an important direction for future work in the cryo-EM field. This thesis used deep learning to develop new methods for continuous conformational variability of single-particle cryo-EM data. Two supervised learning methods (DeepHEMNMA and supervised cryo-ViT), and one unsupervised learning method (unsupervised cryo-ViT) have been developed. DeepHEMNMA uses a convolutional neural network to accelerate HEMNMA by training the network in a supervised manner to learn the values of a set of continuous variables (specifying the rigid-body and conformational parameters) that were previously determined by HEMNMA, for each single particle image. The parameter estimation by HEMNMA is based on a prior simulation of the motion direction by NMA. NMA provides information on the simulated motion direction by a set of vectors (called normal modes) along which atoms move, but it does not provide information on the motion amplitude. The normal-mode amplitudes are estimated for each image by HEMNMA through image alignment with a given 3D model of the molecular complex under study (the given 3D model

is either an atomic model or an EM map). The image normal-mode amplitudes determine the conformation of the complex in the image, and the conformations obtained from a given set of images determine the conformational space. The combination of HEMNMA with a CNN makes the conformational space determination faster.

Although DeepHEMNMA can simultaneously determine the normal-mode amplitudes and the rigid-body parameters, selecting the representative normal modes for an unknown motion is challenging, and this can lead to misrepresentation of the actual movement of the bimolecular complex in the data. Recently, our team developed another approach that, similarly to HEMNMA, combines simulation with image analysis. This new hybrid approach is named MDSPACE and uses a combination of NMA and classical MD simulation to obtain the continuous conformational landscape from single-particle images. An advantage of MDSPACE with respect to HEMNMA is its flexibility in exploring motion directions around those fixed by the given set of normal-mode vectors. A disadvantage of MDSPACE is a slow image processing based on MD simulation. Therefore, in this thesis, we have worked on developing a supervised deep learning approach (currently named supervised cryo-ViT) that could imitate image analysis by MDSPACE to speed it up.

This supervised cryo-ViT learns the relationships between a set of images and the corresponding atomic coordinates determined by MDSPACE from these images prior to learning. The number of the parameters to learn in the case of supervised cryo-ViT is much larger (three times the number of atoms) than in the DeepHEMNMA case (2 shifts, 3 angles, and M normal-mode amplitudes, where M is much smaller than the number of atoms). The supervised cryo-ViT method has shown encouraging results in the experiments performed during this thesis. However, additional experiments are needed to show its efficiency with experimental data. As the development of the MDSPACE method (Ph.D. thesis of Rémi Vuillemot) was done in parallel with this thesis work, and the MDSPACE validations were finalized only very recently (the article published in January 2023), the time allocated for this thesis did not permit to go further with the validations of the supervised cryo-ViT method. Instead, considering recent advances in unsupervised deep learning for cryo-EM, we considered the possibility of altering the supervised cryo-ViT method to make it learn in an unsupervised fashion. In this context, a new approach was developed recently, currently named unsupervised cryo-ViT.

The unsupervised CryoViT is an unsupervised VAE that estimates the corresponding displaced 3D atomic coordinates for every single-particle image. The conformational heterogeneity analysis with this method is data-driven; it is not helped by molecular mechanics simulation, but it requires an initial atomic 3D model (the atomic displacements with respect to this initial model are estimated for each single particle image). The results of the heterogeneity analysis can be further analyzed either from the estimated latent representation of the data or from the lower-dimensional representation of the estimated (predicted) atomic coordinates. The current version of the method learns to refine the given initial values of the rigid-body alignment parameters while learning to map the images to 3D atomic coordinates.

DeepHEMNMA was tested using synthetic and experimental yeast ribosome 80S data and produced results coherent with previous findings. Unsupervised Cryo-ViT showed encouraging results on synthetic and experimental datasets, but more experiments, and potentially improvements of the method, are required before the method can be used in practice. With the experimental dataset of yeast ribosome 80S, unsupervised cryo-ViT showed that it detects the rotation between the two ribosome subunits, but the obtained rotation amplitude is smaller than the expected rotation amplitude (from the literature and obtained by DeepHEMNMA). With the experimental dataset of transmembrane TMEM16F channel, unsupervised cryo-ViT detected conformational changes coherent with previous findings on this complex, although at a lower resolution. The lower resolutions of the reconstructed conformational states, in both cases TMEM16F and 80S ribosome, is likely due to the use of much smaller subsets of images for 3D reconstructions in this work. The experiments with unsupervised cryo-ViT presented in this thesis show the potential of the unsupervised cryo-ViT method. Once the method is fully tested and possibly improved, benchmarking should also be done on other complexes used in the literature, such as Tri-snRNP spliceosome and Spliceosome on precursor mRNA, on which other cryo-EM methods have been tested [6,8,114].

To obtain a low-dimensional conformational landscape, DeepHEMNMA and Cryo-ViT use a dimensionality reduction method, typically PCA, on their output. For instance, one can explore the conformational landscape in terms of animations of molecular flexibility by displacing the atomic coordinates of the reference 3D model in this conformational landscape (interpolation). Another way to explore the conformational landscape is to group points that correspond to similar conformations in this conformational landscape to produce 3D reconstructions. DeepHEMNMA and Cryo-ViT allow

the discovery of both more dominant and less dominant conformations as the conformational space describes the full conformational variability and allows an interactive selection of denser and less dense regions, respectively. This conformational space allows for deriving models of continuous and discrete conformational changes in the case of Cryo-ViT (e.g., yeast ribosome 80S binding with different tRNAs, transmembrane TMEM16F channel binding with Ca^{2+}), and only those that can be described by normal modes in the case of DeepHEMNMA (e.g., ribosome binding with different tRNAs).

DeepHEMNMA and Cryo-ViT are deep learning-based approaches that suffer from the manual selection of hyperparameters, which is a limitation of all deep learning approaches proposed for cryo-EM so far.

The dependency of DeepHEMNMA on HEMNMA can be crucial as it relies on a partly subjective selection of a small subset of normal modes (usually, up to 6 modes are selected to work with HEMNMA). Different selections of normal modes may yield different results of the conformational variability analysis in the images. The selection of the set of normal modes is usually done based on their collectivity and frequency (the lowest-frequency and highest-collectivity normal modes are usually selected). However, some preliminary knowledge on the conformational changes of the studied system can also be used to facilitate the selection, which introduces subjectivity. Cryo-ViT overcomes this problem as it uses a standalone VAE without any strong prior about the conformational variability. However, Cryo-ViT suffers from the rigid-body pre-alignment bias, as it requires the initial orientation and shift parameter values to be determined beforehand. Using synthetic data, we have shown that Cryo-ViT was able to extract the conformational variability in presence of errors in the initial rigid-body alignment that was up to 3 degrees off the ground-truth alignment. However, in case of larger distances between the initial and ground-truth rigid-body alignments, Cryo-VIT may not be efficient. On the contrary, DeepHEMNMA does not require any initial rigid-body prealignment (an exhaustive search at each iteration of the search for the conformational parameters in HEMNMA determines the rigid-body parameter values).

A possible future work regarding DeepHEMNMA is to extend it to analyze cryo-electron tomography (cryo-ET) 3D subtomogram data. An advantage of cryo-ET over single-particle cryo-EM is that it allows analyzing the conformational variability of the complexes in their cellular environment. Our team has recently proposed an extension of HEMNMA to analyze continuous

conformational variability of molecular complexes in cryo-ET subtomograms (the method named HEMNMA-3D [132]). In this context, a version of DeepHEMNMA to process 3D volumes instead of 2D images could be developed, combining HEMNMA-3D with a CNN. However, cryo-ET still produces small data sets compared to single-particle cryo-EM, but this may change with instrumental advances. At present, trials to use DeepHEMNMA on 2D per-particle-tilt images (larger data sets) instead of 3D subtomograms could be attempted, although the outcome is uncertain because of extremely high noise in the tilt images.

Regarding Cryo-ViT, the work with supervised Cryo-ViT could be continued in order to allow the acceleration of MDSPACE [28]. Regarding unsupervised Cryo-ViT, a near-future work direction is to improve its robustness and generalization and improve the network's ability to refine the rigid-body alignment parameters. A mid-future work is to prevent the rigid-body alignment bias by allowing a simultaneous estimation of conformations and rigid-body alignment parameters (overcome the need for an initial rigid-body alignment). Another future direction would be to explore the potential of Cryo-ViT to analyze continuous conformational variability either from cryo-ET subtomograms (an extension of Cryo-ViT to work with input 3D volumes) or from cryo-ET per-particle-tilt images (an adaptation of Cryo-ViT to work with higher levels of noise or to include input data preprocessing to reduce noise).

Bibliography

1 Carlson, D. B. & Evans, J. E.

2 Contributors. *Variational autoencoder. (2023, January 29). In Wikipedia.*, < https://en.wikipedia.org/wiki/Variational_autoencoder> (

3 Williams, D. B. & Carter, C. B.

4 Chung, S.-C., Niu, P.-Y., Huang, S. Y., Chang, W.-H. & Tu, I.-P. 2SDR: Applying Kronecker Envelope PCA to denoise Cryo-EM Images. *ArXiv* **abs/1911.09816** (2019).

5 Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

6 Punjani, A. & Fleet, D. J. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *bioRxiv*, 2021.2004.2022.440893 (2021). https://doi.org/10.1101/2021.04.22.440893

7 Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *ArXiv* **abs/1906.02691** (2019).

8 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods* **18**, 176-185 (2021).

9 Jin, Q. *et al.* Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure* **22**, 496-506 (2014). https://doi.org/https://doi.org/10.1016/j.str.2014.01.004

10 Celler, K., Koning, R., Koster, A. & Wezel, G. Multidimensional View of the Bacterial Cytoskeleton. *Journal of bacteriology* **195** (2013). https://doi.org/10.1128/JB.02194-12

11 Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging* **9**, 611-629 (2018). https://doi.org/10.1007/s13244-018-0639-9

12 Tama, F. & Sanejouand, y.-h. Conformational change of proteins arising from normal mode analysis. *Protein engineering* **14**, 1-6 (2001). https://doi.org/10.1093/protein/14.1.1

13 Heimowitz, A., Andén, J. & Singer, A. APPLE Picker: Automatic Particle Picking, a Low-Effort Cryo-EM Framework. *Journal of structural biology* **204 2**, 215-227 (2018).

14 Falorsi, L. *et al.* Explorations in homeomorphic variational auto-encoding. *arXiv preprint arXiv:1807.04689* (2018).

15 Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).

16 Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. arXiv 2020. *arXiv preprint arXiv:2010.11929* (2010).

17 Behrmann, E. *et al.* Structural Snapshots of Actively Translating Human Ribosomes. *Cell* **161**, 845-857 (2015). https://doi.org/https://doi.org/10.1016/j.cell.2015.03.052

18 Elad, N., Clare, D. K., Saibil, H. R. & Orlova, E. V. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. *Journal of Structural Biology* **162**, 108-120 (2008). https://doi.org/https://doi.org/10.1016/j.jsb.2007.11.007

19 Frank, J. & Ourmazd, A. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61-67 (2016). https://doi.org/https://doi.org/10.1016/j.ymeth.2016.02.007

20 Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **16**, 673-683 (2008). https://doi.org/https://doi.org/10.1016/j.str.2008.03.005

21 Noble, A. J. *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *eLife* **7** (2017).

22 Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* **177**, 630-637 (2012). https://doi.org/https://doi.org/10.1016/j.jsb.2012.02.003

23 Victor Ikechukwu, A., Murali, S., Deepu, R. & Shivamurthy, R. C. ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. *Global Transitions Proceedings* **2**, 375-381 (2021). https://doi.org/https://doi.org/10.1016/j.gltp.2021.08.027

24 Hamitouche, I. & Jonic, S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Frontiers in Molecular Biosciences* **9** (2022). https://doi.org/10.3389/fmolb.2022.965645

25 Hamitouche, I. & Jonić, S. in *2021 29th European Signal Processing Conference (EUSIPCO)*. 1251-1255 (IEEE).

26 Harastani, M., Sorzano, C. O. S. & Jonić, S. Hybrid electron microscopy normal mode analysis with scipion. *Protein Science* **29**, 223-236 (2020).

27 Zhong, E. D., Bepler, T., Davis, J. H. & Berger, B. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. *arXiv preprint arXiv:1909.05215* (2019).

28 Vuillemot, R. *et al.* MDSPACE: Extracting Continuous Conformational Landscapes from Cryo-EM Single Particle Datasets Using 3D-to-2D Flexible Fitting based on Molecular Dynamics Simulation. *Journal of Molecular Biology*, 167951 (2023). https://doi.org/https://doi.org/10.1016/j.jmb.2023.167951

29 Wollman, A. J. M., Nudd, R., Hedlund, E. G. & Leake, M. C. From Animaculum to single molecules: 300 years of the light microscope. *Open Biology* **5** (2015).

30 Knoll, M. & Ruska, E. Das Elektronenmikroskop. *Zeitschrift für Physik* **78**, 318-339 (1932). https://doi.org/10.1007/BF01342199

31 Porter , K. R., Claude , A. & Fullam , E. F. A STUDY OF TISSUE CULTURE CELLS BY ELECTRON MICROSCOPY : METHODS AND PRELIMINARY OBSERVATIONS. *Journal of Experimental Medicine* **81**, 233-246 (1945). https://doi.org/10.1084/jem.81.3.233

32 Dubochet, J. & McDowall, A. W. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *Journal of Microscopy* **124** (1981).

33 Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* **21**, 129 - 228 (1988).

34 Danev, R., Yanagisawa, H. & Kikkawa, M. Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends in Biochemical Sciences* **44**, 837-848 (2019). https://doi.org/https://doi.org/10.1016/j.tibs.2019.04.008

35 Krivanek, O. L. & Mooney, P. E. Applications of slow-scan CCD cameras in transmission electron microscopy. *Ultramicroscopy* **49**, 95-108 (1993). https://doi.org/https://doi.org/10.1016/0304-3991(93)90216-K

36 Bai, X.-c., McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences* **40**, 49-57 (2015). https://doi.org/https://doi.org/10.1016/j.tibs.2014.10.005

37 Erickson, H. P. & Klug, A.

38 Cong, Y. & Ludtke, S. J. in *Methods in Enzymology* Vol. 482 (ed Grant J. Jensen) 211-235 (Academic Press, 2010).

39 Sorzano, C. O. S., Jonic, S., Núñez-Ramírez, R., Boisset, N. & Carazo, J. M. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *Journal of Structural Biology* **160**, 249-262 (2007). https://doi.org/https://doi.org/10.1016/j.jsb.2007.08.013

40 Jonić, S., Sorzano, C. O. S., Cottevieille, M., Larquet, E. & Boisset, N. A novel method for improvement of visualization of power spectra for sorting cryo-electron micrographs and their local areas. *Journal of Structural Biology* **157**, 156-167 (2007). https://doi.org/https://doi.org/10.1016/j.jsb.2006.06.014

41 Velázquez-Muriel, J. A., Sorzano, C. O. S., Fernández, J. J. & Carazo, J. M. A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment. *Ultramicroscopy* **96**, 17-35 (2003). https://doi.org/https://doi.org/10.1016/S0304-3991(02)00377-7
42 van Heel, M. *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* **33**, 307-369 (2000). https://doi.org/10.1017/S0033583500003644
43 Frank, J. & Penczek, P. A. On the correction of the contrast transfer function in biological electron microscopy. *Optik* **98**, 125-129 (1995).
44 Grigorieff, N. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 å in ice11Edited by R. Huber. *Journal of Molecular Biology* **277**, 1033-1046 (1998). https://doi.org/https://doi.org/10.1006/jmbi.1998.1668
45 (eds Robert M. Glaeser, Eva Nogales, & Wah Chiu) (IOP Publishing, 2021).
46 Sindelar, C. V. & Grigorieff, N. Optimal noise reduction in 3D reconstructions of single particles using a volume-normalized filter. *Journal of Structural Biology* **180**, 26-38 (2012). https://doi.org/https://doi.org/10.1016/j.jsb.2012.05.005
47 Sorzano, C. O. S., Marabini, R., Herman, G. T., Censor, Y. & Carazo, J. M. Transfer function restoration in 3D electron microscopy via iterative data refinement. *Physics in Medicine & Biology* **49**, 509 - 522 (2004).
48 Skoglund, U., Öfverstedt, L.-G., Burnett, R. M. & Bricogne, G. Maximum-Entropy Three-Dimensional Reconstruction with Deconvolution of the Contrast Transfer Function: A Test Application with Adenovirus. *Journal of Structural Biology* **117**, 173-188 (1996). https://doi.org/https://doi.org/10.1006/jsbi.1996.0081
49 Baker, L. A. & Rubinstein, J. L. in *Methods in Enzymology* Vol. 481 (ed Grant J. Jensen) 371-388 (Academic Press, 2010).
50 Kelly, J. J. in *Advances in Nuclear Physics* (eds J. W. Negele & Erich Vogt) 75-294 (Springer US, 2002).
51 Egerton, R. F. Mechanisms of radiation damage in beam-sensitive specimens, for TEM accelerating voltages between 10 and 300 kV. *Microscopy Research and Technique* **75**, 1550-1556 (2012). https://doi.org/https://doi.org/10.1002/jemt.22099
52 Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3-15 (2016). https://doi.org/https://doi.org/10.1016/j.ymeth.2016.02.017
53 Frechard, A., Sharov, G., Werderer, M. & Schultz, P. Optimization of Sample Preparation for the Observation of Macromolecular Complexes by Electron (cryo-)Microscopy. *Methods in molecular biology* **2247**, 243-256 (2021).
54 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290-296 (2017). https://doi.org/10.1038/nmeth.4169
55 Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nature Communications* **11**, 5208 (2020). https://doi.org/10.1038/s41467-020-18952-1
56 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216-221 (2015). https://doi.org/https://doi.org/10.1016/j.jsb.2015.08.008
57 Nicholson, W. V. & Glaeser, R. M. Review: Automatic Particle Detection in Electron Microscopy. *Journal of Structural Biology* **133**, 90-101 (2001). https://doi.org/https://doi.org/10.1006/jsbi.2001.4348
58 Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014). https://doi.org/10.7554/eLife.03665

59 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331-332 (2017). https://doi.org/10.1038/nmeth.4193

60 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584-590 (2013). https://doi.org/10.1038/nmeth.2472

61 Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). https://doi.org/10.7554/eLife.06980

62 Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17 (2019). https://doi.org/https://doi.org/10.1107/S205225251801463X

63 Frank, J. & Wagenknecht, T. Automatic selection of molecular images from electron micrographs. *Ultramicroscopy* **12**, 169-175 (1983). https://doi.org/https://doi.org/10.1016/0304-3991(83)90256-5

64 Harauz, G. & Fong-Lochovsky, A. Automatic selection of macromolecules from electron micrographs by component labelling and symbolic processing. *Ultramicroscopy* **31**, 333-344 (1989). https://doi.org/https://doi.org/10.1016/0304-3991(89)90331-8

65 Ogura, T. & Sato, C. Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages: a new reference free method for single-particle analysis. *Journal of structural biology* **145 1-2**, 63-75 (2004).

66 Wang, F. *et al.* DeepPicker: a Deep Learning Approach for Fully Automated Particle Picking in Cryo-EM. *Journal of structural biology* **195 3**, 325-336 (2016).

67 Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics* **18**, 348 (2017). https://doi.org/10.1186/s12859-017-1757-y

68 Al-Azzawi, A. *et al.* DeepCryoPicker: fully automated deep neural network for single protein particle picking in cryo-EM. *BMC Bioinformatics* **21**, 509 (2020). https://doi.org/10.1186/s12859-020-03809-7

69 Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology* **2**, 218 (2019). https://doi.org/10.1038/s42003-019-0437-z

70 Zhao, Z. & Singer, A. Rotationally invariant image representation for viewing direction classification in cryo-EM. *Journal of Structural Biology* **186**, 153-166 (2014). https://doi.org/https://doi.org/10.1016/j.jsb.2014.03.003

71 Sigworth, F. J. Principles of cryo-EM single-particle image processing. *Microscopy* **65**, 57-67 (2016). https://doi.org/10.1093/jmicro/dfv370

72 van Heel, M. & Frank, J. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy* **6**, 187-194 (1981). https://doi.org/https://doi.org/10.1016/S0304-3991(81)80197-0

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

74 Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2**, 37-52 (1987). https://doi.org/https://doi.org/10.1016/0169-7439(87)80084-9

75 Jonić, S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Current Opinion in Structural Biology* **43**, 114-121 (2017). https://doi.org/https://doi.org/10.1016/j.sbi.2016.12.011

76 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal* **478**, 4169-4185 (2021). https://doi.org/10.1042/bcj20210708

77 Gao, H., Valle, M., Ehrenberg, M. & Frank, J. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *Journal of Structural Biology* **147**, 283-290 (2004). https://doi.org/https://doi.org/10.1016/j.jsb.2004.02.008
78 Jonić, S. Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes. *Computational and Structural Biotechnology Journal* **14**, 385-390 (2016). https://doi.org/https://doi.org/10.1016/j.csbj.2016.10.002
79 Valle, M. *et al.* Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *The EMBO Journal* **21**, 3557-3567 (2002). https://doi.org/https://doi.org/10.1093/emboj/cdf326
80 Heymann, J. B. *et al.* Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nature Structural & Molecular Biology* **10**, 334-341 (2003). https://doi.org/10.1038/nsb922
81 Galkin, V. E., Orlova, A., Vos, M. R., Schröder, G. F. & Egelman, E. H. Near-atomic resolution for one state of F-actin. *Structure* **23**, 173-182 (2015).
82 Yang, S. *et al.* Flexibility of the Rings: Structural Asymmetry in the DnaB Hexameric Helicase. *Journal of Molecular Biology* **321**, 839-849 (2002). https://doi.org/https://doi.org/10.1016/S0022-2836(02)00711-8
83 Scheres, S. H. W. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods* **4**, 27-29 (2007). https://doi.org/10.1038/nmeth992
84 Sigworth, F. J., Doerschuk, P. C., Carazo, J.-M. & Scheres, S. H. W. in *Methods in Enzymology* Vol. 482 (ed Grant J. Jensen) 263-294 (Academic Press, 2010).
85 Moon, T. K. The expectation-maximization algorithm. *IEEE Signal Processing Magazine* **13**, 47-60 (1996). https://doi.org/10.1109/79.543975
86 Sigworth, F. J. A Maximum-Likelihood Approach to Single-Particle Image Refinement. *Journal of Structural Biology* **122**, 328-339 (1998). https://doi.org/https://doi.org/10.1006/jsbi.1998.4014
87 Pascual-Montano, A. *et al.* A Novel Neural Network Technique for Analysis and Classification of EM Single-Particle Images. *Journal of Structural Biology* **133**, 233-245 (2001). https://doi.org/https://doi.org/10.1006/jsbi.2001.4369
88 Scheres, S. H. W., Valle, M. & Carazo, J.-M. Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics* **21**, ii243-ii244 (2005). https://doi.org/10.1093/bioinformatics/bti1140
89 Marabini, R., Herman, G. T. & Carazo, J. M. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy* **72 1-2**, 53-65 (1998).
90 Dashti, A. *et al.* Trajectories of the ribosome as a Brownian nanomachine. *Proceedings of the National Academy of Sciences* **111**, 17492-17497 (2014). https://doi.org/doi:10.1073/pnas.1419276111
91 Schwander, P., Fung, R. & Ourmazd, A. Conformations of macromolecules and their complexes from heterogeneous datasets. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369** (2014).
92 Maji, S. *et al.* Propagation of Conformational Coordinates Across Angular Space in Mapping the Continuum of States from Cryo-EM Data by Manifold Embedding. *Journal of Chemical Information and Modeling* **60**, 2484-2491 (2020). https://doi.org/10.1021/acs.jcim.9b01115
93 Dashti, A. *et al.* Retrieving functional pathways of biomolecules from single-particle snapshots. *Nature Communications* **11**, 4734 (2020). https://doi.org/10.1038/s41467-020-18403-x
94 Seitz, E., Acosta-Reyes, F., Maji, S., Schwander, P. & Frank, J. Recovery of Conformational Continuum From Single-Particle Cryo-EM Images: Optimization of ManifoldEM Informed

by Ground Truth. *IEEE Transactions on Computational Imaging* **8**, 462-478 (2022). https://doi.org/10.1109/TCI.2022.3174801

95 Calero, D. H. *et al.* Continuous heterogeneity analysis of CryoEM images through Zernike polynomials and spherical harmonics. *Microscopy and Microanalysis* **27**, 1680-1682 (2021). https://doi.org/10.1017/S1431927621006176

96 Herreros, D. *et al.* Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials. *IUCrJ* **8**, 992-1005 (2021). https://doi.org/doi:10.1107/S2052252521008903

97 Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1-27 (1964).

98 Tama, F., Gadéa, F. X., Marques, O. & Sanejouand, Y. H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure* **41** (2000).

99 Giraldo-Barreto, J. *et al.* A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments. *Scientific Reports* **11**, 13657 (2021). https://doi.org/10.1038/s41598-021-92621-1

100 Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-420 (2008). https://doi.org/10.1038/nature07192

101 Hill, M. O. Correspondence analysis: a neglected multivariate method. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **23**, 340-354 (1974).

102 Lee, D. & Seung, H. S. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems* **13** (2000).

103 Frank, J. & van Heel, M. Correspondence analysis of aligned images of biological particles. *Journal of Molecular Biology* **161**, 134-137 (1982). https://doi.org/https://doi.org/10.1016/0022-2836(82)90282-0

104 White, H. E., Saibil, H. R., Ignatiou, A. & Orlova, E. V. Recognition and Separation of Single Particles with Size Variation by Statistical Analysis of their Images. *Journal of Molecular Biology* **336**, 453-460 (2004). https://doi.org/https://doi.org/10.1016/j.jmb.2003.12.015

105 van Heel, M. & Stöffler-Meilicke, M. Characteristic views of E. coli and B. stearothermophilus 30S ribosomal subunits in the electron microscope. *The EMBO journal* **4**, 2389-2395 (1985).

106 Falke, S., Tama, F., Brooks, C. L., Gogol, E. P. & Fisher, M. T. The 13Å Structure of a Chaperonin GroEL–Protein Substrate Complex by Cryo-electron Microscopy. *Journal of Molecular Biology* **348**, 219-230 (2005). https://doi.org/https://doi.org/10.1016/j.jmb.2005.02.027

107 Jin, X. & Han, J. in *Encyclopedia of Machine Learning* (eds Claude Sammut & Geoffrey I. Webb) 563-564 (Springer US, 2010).

108 Liu, W. & Frank, J. Estimation of variance distribution in three-dimensional reconstruction. I. Theory. *J. Opt. Soc. Am. A* **12**, 2615-2627 (1995). https://doi.org/10.1364/JOSAA.12.002615

109 Penczek, Pawel A., Kimmel, M. & Spahn, Christian M. T. Identifying Conformational States of Macromolecules by Eigen-Analysis of Resampled Cryo-EM Images. *Structure* **19**, 1582-1590 (2011). https://doi.org/https://doi.org/10.1016/j.str.2011.10.003

110 Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H. & Rao, M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *Journal of Structural Biology* **191**, 245-262 (2015). https://doi.org/https://doi.org/10.1016/j.jsb.2015.05.007

111 Katsevich, G., Katsevich, A. & Singer, A. Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. *SIAM journal on imaging sciences* **8 1**, 126-185 (2013).

112 Liao, Hstau Y., Hashem, Y. & Frank, J. Efficient Estimation of Three-Dimensional Covariance and its Application in the Analysis of Heterogeneous Samples in Cryo-Electron Microscopy. *Structure* **23**, 1129-1137 (2015). https://doi.org/https://doi.org/10.1016/j.str.2015.04.004

113 Melero, R. *et al.* Continuous flexibility analysis of SARS-CoV-2 spike prefusion structures. *IUCrJ* **7**, 1059-1069 (2020). https://doi.org/doi:10.1107/S2052252520012725

114 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213**, 107702 (2021). https://doi.org/https://doi.org/10.1016/j.jsb.2021.107702
115 McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
116 Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 4066-4075.
117 Levy, A., Wetzstein, G., Martel, J. N. P., Poitevin, F. & Zhong, E. D. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *ArXiv* **abs/2210.07387** (2022).
118 Gupta, H., Phan, T. H., Yoo, J. & Unser, M. in *Computer Vision – ECCV 2020 Workshops.* (eds Adrien Bartoli & Andrea Fusiello) 429-444 (Springer International Publishing).
119 Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods* **18**, 930-936 (2021). https://doi.org/10.1038/s41592-021-01220-5
120 Donnat, C., Levy, A., Poitevin, F. & Miolane, N. Deep Generative Modeling for Volume Reconstruction in Cryo-Electron Microscopy. *Journal of structural biology*, 107920 (2022).
121 Collier, T. A., Piggot, T. J. & Allison, J. R. in *Protein Nanotechnology: Protocols, Instrumentation, and Applications* (eds Juliet A. Gerrard & Laura J. Domigan) 311-327 (Springer US, 2020).
122 ben-Avraham, D. & Tirion, M. M. Normal modes analyses of macromolecules. *Physica A: Statistical Mechanics and its Applications* **249**, 415-423 (1998). https://doi.org/https://doi.org/10.1016/S0378-4371(97)00501-3
123 Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters* **77**, 1905 (1996).
124 Sanejouand, Y.-H. in *Biomolecular Simulations: Methods and Protocols* (eds Luca Monticelli & Emppu Salonen) 601-616 (Humana Press, 2013).
125 Bahar, I., Lezon, T. R., Bakan, A. & Shrivastava, I. H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chemical Reviews* **110**, 1463-1497 (2010). https://doi.org/10.1021/cr900095e
126 Koehl, P., Orland, H. & Delarue, M. Parameterizing elastic network models to capture the dynamics of proteins. *Journal of Computational Chemistry* **42**, 1643-1661 (2021). https://doi.org/https://doi.org/10.1002/jcc.26701
127 Tama, F., Miyashita, O. & Brooks III, C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *Journal of molecular biology* **337**, 985-999 (2004).
128 Kulik, M., Mori, T. & Sugita, Y. Multi-Scale Flexible Fitting of Proteins to Cryo-EM Density Maps at Medium Resolution. *Frontiers in Molecular Biosciences* **8** (2021). https://doi.org/10.3389/fmolb.2021.631854
129 Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174-180 (2009). https://doi.org/https://doi.org/10.1016/j.ymeth.2009.04.005
130 Kim, D. N. *et al.* Cryo_fit: Democratization of flexible fitting for cryo-EM. *Journal of Structural Biology* **208**, 1-6 (2019). https://doi.org/https://doi.org/10.1016/j.jsb.2019.05.012
131 Vuillemot, R., Miyashita, O., Tama, F., Rouiller, I. & Jonic, S. NMMD: Efficient Cryo-EM Flexible Fitting Based on Simultaneous Normal Mode and Molecular Dynamics atomic displacements. *Journal of Molecular Biology* **434**, 167483 (2022). https://doi.org/https://doi.org/10.1016/j.jmb.2022.167483
132 Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes. *Front Mol Biosci* **8**, 663121 (2021). https://doi.org/10.3389/fmolb.2021.663121

133 Canziani, A., Paszke, A. & Culurciello, E. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678* (2016).
134 Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* **36**, 193-202 (1980). https://doi.org/10.1007/BF00344251
135 Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278-2324 (1998). https://doi.org/10.1109/5.726791
136 Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **60**, 84-90 (2017).
137 Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* **115**, 211-252 (2014).
138 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770-778.
139 Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv* **abs/1810.04805** (2019).
140 Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
141 Trockman, A. & Kolter, J. Z. Patches Are All You Need? *ArXiv* **abs/2201.09792** (2022).
142 Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. in *North American Chapter of the Association for Computational Linguistics.*
143 Choi, Y. *et al.* in *Proceedings of the seventh ACM conference on learning@ scale.* 341-344.
144 Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *CoRR* **abs/1312.6114** (2013).
145 Higgins, I. *et al.* in *International Conference on Learning Representations.*
146 Jimenez Rezende, D., Mohamed, S. & Wierstra, D. in *International Conference on Machine Learning.*
147 Hu, M., Zhang, Q., Yang, J. & Li, X. Unit quaternion description of spatial rotations in 3D electron cryo-microscopy. *Journal of Structural Biology* **212**, 107601 (2020). https://doi.org/https://doi.org/10.1016/j.jsb.2020.107601
148 Jonić, S. & Sorzano, C. Ó. S. Coarse-graining of volumes for modeling of structure and dynamics in electron microscopy: Algorithm to automatically control accuracy of approximation. *IEEE Journal of Selected Topics in Signal Processing* **10**, 161-173 (2015).
149 Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **13**, 373-380 (2005). https://doi.org/https://doi.org/10.1016/j.str.2005.02.002
150 Tama, F. & Charles L. Brooks, I. SYMMETRY, FORM, AND SHAPE: Guiding Principles for Robustness in Macromolecular Machines. *Annual Review of Biophysics and Biomolecular Structure* **35**, 115-133 (2006). https://doi.org/10.1146/annurev.biophys.35.040405.102010
151 Skjaerven, L., Hollup, S. M. & Reuter, N. Normal mode analysis for proteins. *Journal of Molecular Structure: THEOCHEM* **898**, 42-48 (2009).
152 Bahar, I., Lezon, T. R., Yang, L.-W. & Eyal, E. Global Dynamics of Proteins: Bridging Between Structure and Function. *Annual Review of Biophysics* **39**, 23-42 (2010). https://doi.org/10.1146/annurev.biophys.093008.131258
153 López-Blanco, J. R. & Chacón, P. New generation of elastic network models. *Current Opinion in Structural Biology* **37**, 46-53 (2016). https://doi.org/https://doi.org/10.1016/j.sbi.2015.11.013
154 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019). https://doi.org/10.1038/s41592-019-0580-y
155 Rappez, L., Rakhlin, A., Rigopoulos, A., Nikolenko, S. & Alexandrov, T. DeepCycle reconstructs a cyclic cell cycle trajectory from unsegmented cell images using convolutional neural networks. *Mol Syst Biol* **16**, e9474 (2020). https://doi.org/10.15252/msb.20209474
156 Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 (2014). <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K>.

157 Müller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147-156 (1996). https://doi.org/https://doi.org/10.1016/S0969-2126(96)00018-4
158 Svidritskiy, E., Brilot, A. F., Koh, C. S., Grigorieff, N. & Korostelev, A. A. Structures of yeast 80S ribosome-tRNA complexes in the rotated and nonrotated conformations. *Structure* **22**, 1210-1218 (2014). https://doi.org/10.1016/j.str.2014.06.003
159 Peng, L.-M., Ren, G., Dudarev, S. & Whelan, M. Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallographica Section A: Foundations of Crystallography* **52**, 257-276 (1996).
160 Mahalanobis, P. C. (National Institute of Science of India).
161 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* **183**, 377-388 (2013).
162 Dosovitskiy, A. *et al.* An Image is Worth 16× 16 Words: Transformers for Image Recognition at Scale. (2020).
163 Maas, A. L., Hannun, A. Y. & Ng, A. Y. in *Proc. icml.* 3 (Atlanta, Georgia, USA).
164 Zhou, Y., Barnes, C., Lu, J., Yang, J. & Li, H. On the Continuity of Rotation Representations in Neural Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5738-5746 (2018).
165 Feng, S. *et al.* Cryo-EM Studies of TMEM16F Calcium-Activated Ion Channel Suggest Features Important for Lipid Scrambling. *Cell Reports* **28**, 567-579.e564 (2019). https://doi.org/https://doi.org/10.1016/j.celrep.2019.06.023
166 Paulino, C., Kalienkova, V., Lam, A. K. M., Neldner, Y. & Dutzler, R. Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM. *Nature* **552**, 421-425 (2017). https://doi.org/10.1038/nature24652
167 de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology* **195**, 93-99 (2016). https://doi.org/https://doi.org/10.1016/j.jsb.2016.04.010
168 Harastani, M., Vuillemot, R., Hamitouche, I., Moghadam, N. B. & Jonic, S. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *Journal of Structural Biology* **214**, 107906 (2022). https://doi.org/https://doi.org/10.1016/j.jsb.2022.107906
169 Sorzano, C. O. S. *et al.* XMIPP: a new generation of an open-source image processing package for electron microscopy. *Journal of Structural Biology* **148**, 194-204 (2004). https://doi.org/https://doi.org/10.1016/j.jsb.2004.06.006
170 Strelak, D. *et al.* Advances in Xmipp for Cryo–Electron Microscopy: From Xmipp to Scipion. *Molecules* **26**, 6224 (2021).
171 Yu, T. & Zhu, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *ArXiv* **abs/2003.05689** (2020).
172 Feurer, M. & Hutter, F. in *Automated Machine Learning: Methods, Systems, Challenges* (eds Frank Hutter, Lars Kotthoff, & Joaquin Vanschoren) 3-33 (Springer International Publishing, 2019).
173 Bergstra, J. & Bengio, Y. Random search for hyper-parameter optimization. *Journal of machine learning research* **13** (2012).
174 Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE* **104**, 148-175 (2015).
175 Pichkur, E. *et al.* Towards on-the-fly Cryo-Electron Microscopy Data Processing by High Performance Data Analysis. *Journal of Physics: Conference Series* **955**, 012005 (2018). https://doi.org/10.1088/1742-6596/955/1/012005

Table of Figures

Table of Tables